Canadian Open Genetics Repository (COGR): a unified clinical genomics database as a community resource for standardising and sharing genetic interpretations

ORIGINAL ARTICLE

Canadian Open Genetics Repository (COGR): a unified clinical genomics database as a community resource for standardising and sharing genetic interpretations

Jordan Lerner-Ellis,[1,2] Marina Wang,[3] Shana White,[4,5] Matthew S Lebo,[4,6] and the Canadian Open Genetics Repository Group

▶ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jmedgenet-2014-102933).

[1]Laboratory Medicine and Pathobiology, University of Toronto & Mount Sinai Hospital, Toronto, Ontario, Canada
[2]Ontario Institute for Cancer Research, Toronto, Ontario, Canada
[3]Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada
[4]Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine, Cambridge, Massachusetts, USA
[5]Massachusetts General Hospital, Boston, Massachusetts, USA
[6]Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA

Correspondence to
Jordan Lerner-Ellis, Laboratory Medicine and Pathobiology, University of Toronto & Mount Sinai Hospital, 600 University Ave, Toronto, Ontario, Canada M5G 1X5; jlerner-ellis@mtsinai.on.ca

Received 4 December 2014
Revised 9 March 2015
Accepted 15 March 2015
Published Online First 22 April 2015

To cite: Lerner-Ellis J, Wang M, White S, *et al*. *J Med Genet* 2015;**52**: 438–445.

ABSTRACT

Background The Canadian Open Genetics Repository is a collaborative effort for the collection, storage, sharing and robust analysis of variants reported by medical diagnostics laboratories across Canada. As clinical laboratories adopt modern genomics technologies, the need for this type of collaborative framework is increasingly important.

Methods A survey to assess existing protocols for variant classification and reporting was delivered to clinical genetics laboratories across Canada. Based on feedback from this survey, a variant assessment tool was made available to all laboratories. Each participating laboratory was provided with an instance of GeneInsight, a software featuring versioning and approval processes for variant assessments and interpretations and allowing for variant data to be shared between instances. Guidelines were established for sharing data among clinical laboratories and in the final outreach phase, data will be made readily available to patient advocacy groups for general use.

Results The survey demonstrated the need for improved standardisation and data sharing across the country. A variant assessment template was made available to the community to aid with standardisation. Instances of the GeneInsight tool were provided to clinical diagnostic laboratories across Canada for the purpose of uploading, transferring, accessing and sharing variant data.

Conclusions As an ongoing endeavour and a permanent resource, the Canadian Open Genetics Repository aims to serve as a focal point for the collaboration of Canadian laboratories with other countries in the development of tools that take full advantage of laboratory data in diagnosing, managing and treating genetic diseases.

INTRODUCTION

Many individual Canadian and international databases offer a rich store of data on DNA variants. However, most of these resources have been of limited utility for clinical laboratories due to a number of serious shortcomings, including a lack of clinically related information on phenotypical consequences. Moreover, many public-access genetic databases (eg, the Human Gene Mutation Database (HGMD) and dbSNP) are limited in their scope (eg, limited to locus-specific data), lack clinically approved interpretations, and/or are hampered by clinical and technical false positives and negatives.

In recent years, enormous efforts have been devoted to understanding the various components of the human genome and how it relates to biology and physiology.[1–3] Implicit in these projects is the need to translate research discoveries into guidelines for clinical practice and related areas like population health impact, areas that have traditionally been lacking in funding.[4] Clinical and academic laboratories today face the onset of next-generation sequencing with the attendant need to sort and interpret large numbers of variants.

The work carried out by molecular geneticists must be done with extreme care and thoroughness, as their findings and advice to attending physicians will often be determinant in major surgical and chemotherapeutic decisions. As they are identified, variants undergo clinical assessment—a process used to classify variants based on established criteria. Information about the variant, including type of change, location, frequency, previous reports in the literature, segregation in families, conservation, biochemical properties and computational predictions are compiled and used in assigning the appropriate classification. This lab-based work is very time-consuming: the clinical interpretation of a single variant can take an average of 40 min, and in some complex cases far longer.[5]

The variant assessment process typically involves hundreds of data points drawn from laboratory work, genomic resources, the scientific literature, and the expertise and experience of the geneticist undertaking the particular assessment in question. The typical clinical lab-based structure and decision trees for classifying variants are extremely complex. Furthermore, some of the rules involved don't always apply; for example, silent variants can be pathogenic and loss-of-function variants can be benign. The understanding of variants in the context of phenotypical consequences adds further complexity and rules may also change to reflect varying degrees of heterogeneity or penetrance. Variants identified during testing for dominant hereditary cancers may be treated or classified differently than those discovered during tumour testing or testing of a recessive condition. Until recently, much of this work was carried out using data and expertise housed within individual laboratories.

However, as laboratories expand their testing menus to include more extensive panels and exome and genome sequencing, variants associated with phenotypes that lie outside areas of disease expertise will be detected with regularity.

In July 2013, Genome Canada funded a 3-year bioinformatics project whose overall goal was to implement a unified, open, Canadian database of clinical genetic variants. The Canadian Open Genetics Repository (COGR) marks a first in Canada, with the potential to unlock the benefits of Canada's database resources on a more systematic, robust, and community-wide basis. In this paper we present how the COGR project was explicitly designed to assist with challenges currently facing the clinical genetics community.

The COGR draws from existing data holdings at clinical laboratories across the country and upload these data holdings on a common software platform to enhance and promote data-sharing, collaboration, and constant improvement in the quality and clinical applications of these data. There are three aims of the COGR project: (1) the design of standardised variant assessment procedures; (2) data extraction and transfer from Canadian clinical laboratories into a central repository; (3) data access and dissemination of consensus agreement variant interpretations through a publicly available database.

METHODS

A survey of 34 questions was designed and fielded to 76 potential project participants from 41 institutions across Canada. The purpose of the survey was to understand the current landscape of genetic testing within laboratories, the state of data holdings, and specifically to identify areas of strength or weakness where additional support and resources could be provided by the COGR. The three aims, supported by the survey results, were designed to address the needs of Canadian laboratories.

Aim 1. Design of variant assessment procedures

To allow clinical and research laboratories alike to classify human genetic variants of all kinds and from all sources in a scalable, robust and automated manner, one key short-term scientific objective was to design and build a variant assessment tool (VAT). Towards that aim, a VAT, developed and updated at Partners HealthCare[5] within the framework of published guidelines,[6] was made freely available to all working group members (http://opengenetics.ca/resources) with the intent of having all participants carry out the variant assessment process in a similar fashion. Having multiple stakeholders assess variant significance in a systematic, comprehensive and consistent manner will foster knowledge aggregation. The overall effort is to facilitate the process of transforming data-variant holdings into a unified format, while eliminating discrepancies, omissions and duplication of effort.

Aim 2. Data extraction and transfer

The project team supports the extraction of the variant data currently held within participating laboratories in Canada. Our bioinformatics team was made available to work with each laboratory to ensure that their data are transmitted safely and efficiently to a central repository.

Aim 3. Data access and dissemination

Future methods will be developed to make the data holdings extremely accurate and readily accessible by all interested parties, including participating labs, clinicians, geneticists and scientists engaged in basic research. The tools necessary to carry out this phase will be developed in close collaboration with the US-based National Center for Biotechnology Information clinical genetics repository (ClinVar). To maximise the value of this resource to the community at large, our project team will put plans in place to encourage adoption of a unified platform, as well as to train and educate stakeholders.

The COGR workgroups were created for project participants based on their area of expertise. The Executive Committee was charged with overall supervision of the project and is responsible for budget, scheduling and workgroup coordination. The Bioinformatics and IT workgroup was created for sourcing and developing software and other tools to interpret genetic variation; extracting data from participating labs; and making all project information available to the community at large. The Data Collection and Standards workgroup was set up to define and monitor the operational goals of the project and is made up of representatives from each participating laboratory. Members of this workgroup decide how much information their laboratory is willing to share or is comfortable with sharing and in what capacity. The Outreach and Patient Advocacy workgroup was created for the purpose of using data flowing from the project for patient care as well as to create awareness of and interest in the long-term benefits of the project. This workgroup will become more prominent as the number of COGR participants uploading variant interpretations and sharing data increases. A full list of workgroups and institutions involved is provided in online supplementary appendix 1.

Data sharing between multiple institutions will be facilitated using GeneInsight, a web-based tool for the collection, storage, tracking, and sharing of human DNA variant information.[7–9] GeneInsight is a proprietary software, for which the COGR holds a 3-year licensing agreement to provide each institutional member of the COGR with a web-based instance of the software platform at no additional cost. Each GeneInsight instance is provided as a site-license which can be accessed by any number of users at the institution. DNA variant information, including variant classification and disease ontology, from the laboratory's database can be imported to GeneInsight via the user-interface. The bioinformatics team within the COGR will support laboratories with complex data holdings or limited resources. Sharing variant information with other COGR members is optional, and a decision that is made independently by each laboratory. Once a laboratory chooses to share variant information, they will be able to see variant information from other laboratories that have opted to share information. Consensus agreements on shared data from the project will be made publicly available via web (http://opengenetics.ca/database/) and through ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/) (figures 1–3).

Sharing policies and agreements

Currently, data for the COGR is temporarily stored in a secure data centre at Partners HealthCare based in Boston with the intention of moving it to a centralised Canadian site in the near future. See online supplementary appendix 2 for a description of our data sharing policies. In compliance with the Personal Health Information Protection Act, patient identifying information is not included as part of the COGR project.

RESULTS

We received a total of 31 responses to the survey, representing 23 laboratories from around Canada. Out of the 31 responses, 19 were from laboratory directors; the rest were clinicians, genetic counsellors or those involved in genetic research. The survey was designed to provide feedback on several areas of relevance to our project planning. In particular, we wanted to develop a detailed profile of the current working methods in

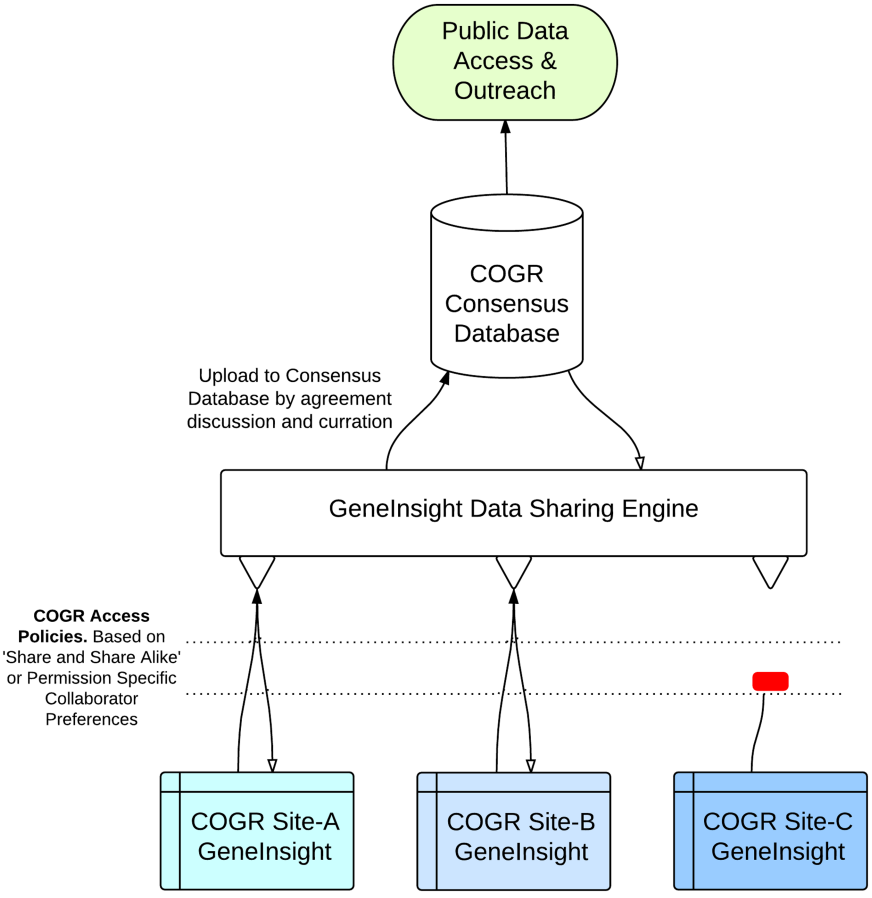

Figure 1 Canadian Open Genetics Repository (COGR) schematic. Black arrowheads show data being uploaded from individual GeneInsight instances into the Data Sharing Engine and then into the Consensus Database for upload to a public database. White arrowheads illustrate sharing of data from the consensus database or between labs via the sharing engine. In this example Site-C has chosen not to upload data to the sharing engine.

Canadian clinical genetics, a crucial starting point for the development and implementation of new tools and procedures. Selected responses to the survey are shown in table 1 and complete survey questions and results are in online supplementary appendix 3.

Survey respondents cited a wide range of over 50 genetic variant data fields that they collect in their labs on a routine basis. Of these, the most widely collected field, cited by 19 respondents, was cDNA nomenclature, followed by protein nomenclature (18); PubMed search (17); variant coding effect, for example, missense, nonsense, frameshift (17); variant type, for example, substitution, insertion (16); and variant location, for example, exon 1 (16).

Responses to questions C2, C5, C6 and C11 (table 1) show that there is a considerable lack of consistency in the handling of variant classification across laboratories. While the majority of laboratories are using consistent and standardised classification terms, a large number of laboratories (31%) do not. In addition, there is a lack of formalised written rules for variant classification in the Canadian molecular genetics community.

Questions C19–21 asked about connections between classification results and patient files. The responses indicated that not all laboratories connected their variant information to patient or disease information. Further, the responses to question C23 and C24, suggest that clinical reports and individuals having a particular variant are tracked in a database, but overall families associated with a variant are not. Whether this reflects a failure to track this information or is related to privacy issues cannot be determined from this survey.

To date, we have received support from over 40 laboratories and interested parties, including the majority of clinical diagnostic laboratories across Canada. The COGR has the support of the Canadian College of Medical Geneticists and has recently become the interim country node for the Human Variome Project (HVP).[10] Many clinical diagnostic laboratories across Canada have started using their GeneInsight instance to store and share variant information. Currently, variant data is only accessed by participating laboratories. Validated consensus variant data will be uploaded to a public access database, for wider use, by year 3 of the project. More sensitive discrepant data will only be accessible by laboratory directors until a consensus clinical assertion has been reached.

As of this writing, there are 19 sites across Canada that are actively involved with the project and have been provided an instance of GeneInsight. Six sites are currently sharing variant data in real time. A total of 3877 intragenic variants (3242 unique) from 25 genes and 3339 copy number variants are being shared. Up-to-date information can be obtained from: http://opengenetics.ca/current-stats/.

The creation of laboratory-specific uploading scripts is a continuing effort that will streamline the process of updating database instances as new variants are found, and as reported variants are reclassified. Moving forward, project participants will need to make decisions on how disagreements between variant classifications will be resolved and how consensus-level variant information will be determined (eg, by group or by individual laboratories with disagreements) before data uploaded into the COGR will be made available to the wider community.

DISCUSSION

The responses to the initial COGR survey bore out the assumption that many laboratories have yet to implement sophisticated

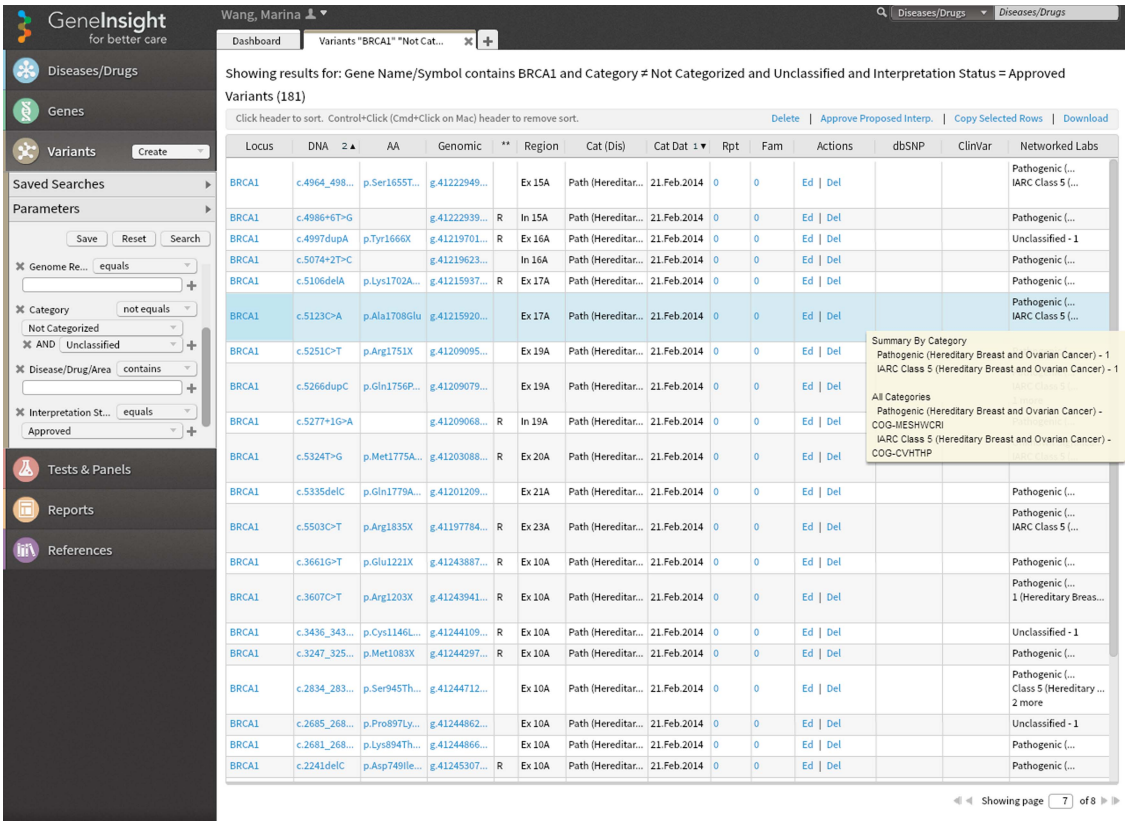

Figure 2 Example of how laboratory directors can view variant data in their instance of the GeneInsight platform. This example shows variants in the *BRCA1* gene as well as accompanying variant information such as variant location and classification. Shown in the right hand column are variant classifications from other labs.

formal systems of the kind we intend to introduce as part of the COGR project. Smaller laboratories may have protocols and systems for variant classification in place that are simply not well documented, making standardisation and collaborations between laboratories more difficult; this is a key point that the COGR project aims to address. The COGR was created on the basis of three aims: (1) Design of variant assessment procedures; (2) Data extraction and transfer; (3) Data access and dissemination. Upon completion of all three aims of the project, laboratories across Canada will be able to store variant information in a standardised and formal fashion. Currently, variant information is available for sharing between participating COGR laboratories, allowing these clinical laboratories to compare variant classifications and interpretations with one another directly. Those variants that have consensus group agreements across multiple laboratories will be added to the central COGR database, which will then be made publicly available on our website and available to other public databases in the international community. The COGR team is currently actively considering several long-term funding options to ensure ongoing access and participation.

For the purposes of this project, a structured, standardised process for individual groups to derive their variant interpretations is critical. The ultimate goal for creation and usage of the VAT is to enable laboratories to provide an interpretation of the clinical significance of a variant with clear, accurate and consistent evidence, including language for reporting to attending physicians. The data classes to be captured in the VAT include: the nature of genetic change(s) and consequence (eg, premature stop codon and other loss of function); the number of times a variant is observed in probands with a particular disease or phenotype; co-occurrence with other pathogenic variants; ethnicity and absence from race-matched control populations; whether a variant is informative or uninformative for segregation (number of meioses); additional published information on probands such as age of onset; zygosity; functional data; nucleotide and amino acid conservation; location in functional protein domain; in silico predictions on the effect of the variant; and variations in nomenclature used in literature. Much of this information will come from primary literature and/or from variant databases, for example, ClinVar.[11] Ethnic frequencies and frequencies in control proband populations will be obtained from publicly accessible projects like 1000 Genomes project,[12] HapMap,[13] and the Exome Aggregation Consortium (URL: http://exac.broadinstitute.org).

The ultimate goal of variant assessment is to provide an interpretation of the clinical significance of a variant that results in clear and accurate reporting to the requesting physician. The general outline of a variant interpretation takes the following form on a patient report: a literature search to determine if the variant has been previously published with associated phenotypical information; database(s) where the variant is identified; if previously detected by the lab; description of relevant data; number of probands (out of how many tested); presence or absence in healthy control data; population frequencies; segregation and/or co-occurrence with pathogenic variants; nature of change and consequences; conservation and in silico analyses; functional data if available; conflicting information and reconciliation if possible, and resulting classification. The molecular geneticist carrying out an analysis may add summary sentences

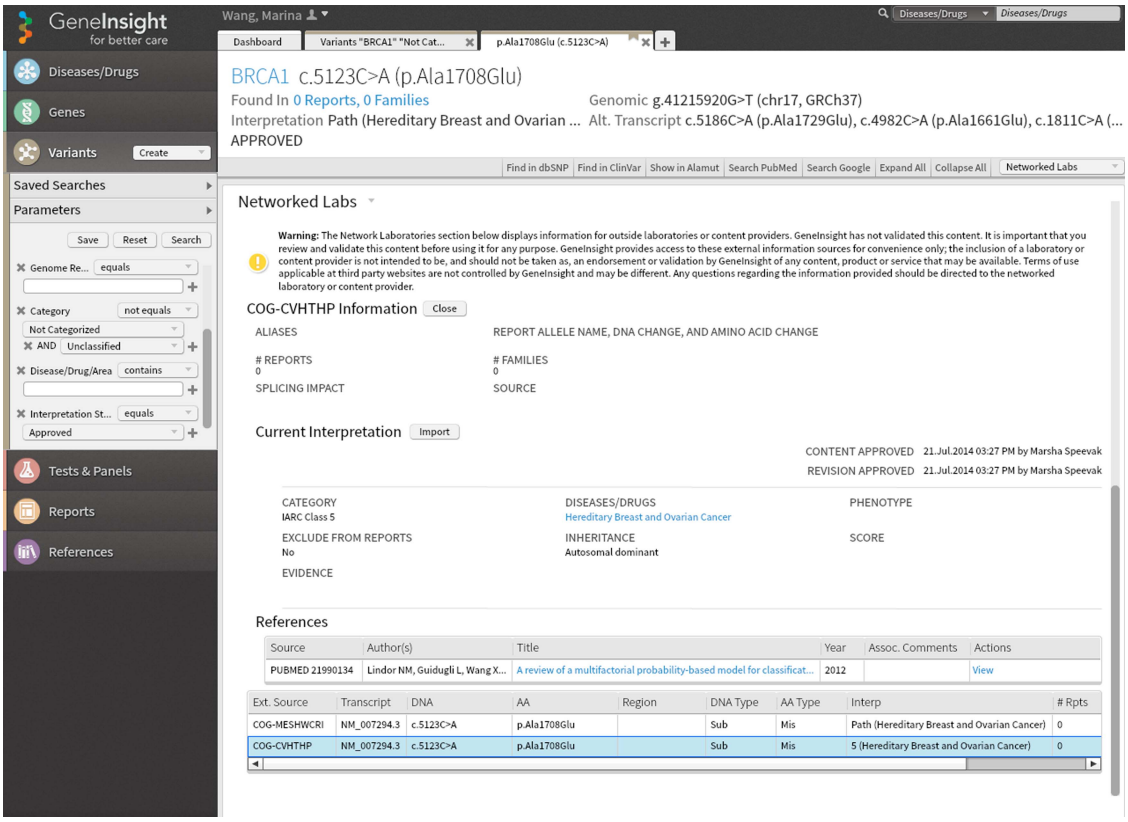

Figure 3 This example shows how variants may be viewed individually. This view shows the complete variant information including revision history and variant interpretation and evidence from other networked (sharing) laboratories. Not shown here, you may also zoom in to additional variant information including gene details such as genomic alignments, transcripts and gene regions, assessments, annotations, laboratory interpretation information and references.

Table 1 Sample of survey results on working methods for variant collection and classification from laboratory directors

Question	Laboratory directors only	Total
(C2) Use formal tracking system for variants in the literature	**Yes: 9 (56%)** No: 7 (44%) n=16	Yes: 11 (46%) **No: 13 (54%)** n=24
(C5) Use the the American College of Medical Genetics (ACMG) system for classifying sequence variants	**Yes: 10 (67%)** No: 5 (33%) n=15	**Yes: 12 (63%)** No: 7 (37%) n=19
(C6) Laboratory uses consistent set of terms for classification	**Yes: 11 (69%)** No: 5 (31%) n=16	**Yes: 13 (65%)** No: 7(35%) n=20
(C11) Laboratory has written rules for evidence-based classification of variants	Yes: 6 (40%) **No: 9 (60%)** n=15	Yes: 8 (40%) **No: 12 (60%)** n=20
(C19) Variant data are linked to all patients with particular variant	**Yes: 9 (69%)** No: 4 (31%) n=13	**Yes: 12 (67%)** No: 6 (33%) n=18
(C21) Variant data are linked to disease type	**Yes: 8 (62%)** No: 5 (38%) n=13	**Yes: 11 (61%)** No: 7 (39%) n=18
(C23) Maintains database tracking individuals with particular variant	**Yes: 11 (73%)** No: 3 (20%) DK: 1 (7%) n=15	**Yes: 15 (68%)** No: 6 (27%) DK: 1 (5%) n=22
(C24) Maintains database tracking families associated with particular variant	Yes: 7 (47%) **No: 8 (53%)** n=15	**Yes: 12 (55%)** No: 10 (45%) n=22

Percentages are based on the total number of responses received per question. Question numbers are stated in brackets. For complete survey questions and results see online supplementary appendix 3. For each question, the majority response has been bolded.
DK: don't know.

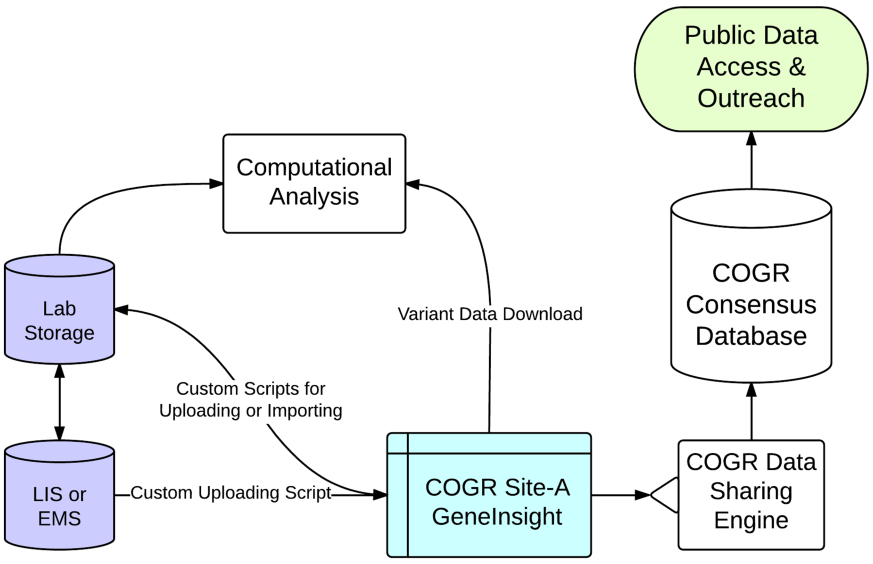

Figure 4 Schematic of data uploading from laboratory databases and/or laboratory information system (LIS)/electronic medical system (EMS) databases into GeneInsight. Custom scripts may be used to facilitate uploading. Variant data, specific to each laboratory, can also be downloaded from GeneInsight directly for analysis purposes.

stating major reasons for the variant classification. Such assertions must be reconciled with any applicable patient phenotypes, and additional supporting evidence may be added if variants in a gene have not been previously reported.

Currently, the VAT works as an excel spreadsheet. Our bioinformatic and other technical staff are in the process of automating the VAT and implementing it to be coupled with GeneInsight in subsequent versioned releases such that additional variant information will be available directly through GeneInsight.

Aim 2 of the project seeks to capture variant data from Canadian laboratories, including structural data and sequence variants, and phenotypical and clinical information of all kinds, including from patient files, research studies and family histories. Rather than creating new software, the GeneInsight application was chosen for its ability to capture and share variant data so that resources could be directed towards addressing the primary needs of the Canadian genetics community. A full list of variant data fields for capture from participating laboratory can be found on the web (http://opengenetics.ca/communities/policiesguidelines; see online supplementary appendix 2).

A key consideration for the project work plan was ensuring that during data submission, a common, standardised set of procedures, nomenclatures and annotations was used. This requirement reflects the long-term objective of the project to eliminate the many structural discrepancies between existing databases. Each laboratory database is unique in its use of database software and field names for variant information. In addition, laboratories have different systems in place for the integration of their database with their specific Laboratory Information System or patient Electronic Medical Record(s). For example, variant classification, interpretation, classification date, patients and families tested, may be stored in one laboratory database system while other details remain strictly within the patient report within the Laboratory Information System.

To further the standardisation of collecting variant information, the Disease Ontology[14] was chosen to be the standard disease collection system in the COGR database. All clinically significant variants are linked to a disease listed in Disease Ontology as to avoid redundant disease naming, for example, 'glycogen storage disease type II' versus 'Pompe disease'.

The COGR will automate the process of data reformatting and uploading to ensure that each institution's instance of the COGR database is up-to-date, especially when it is not used as the laboratory's system of record for variant information. Custom scripts are being developed so that variant data can be routinely and automatically transferred from individual laboratories' systems to the COGR with minimal manual effort while validating the information and ensuring the proper disease ontology data is included (Schematic: figure 4).

The project is being developed based on the understanding that access will eventually be granted to several different categories of end users, among them research scientists, attending physicians, clinical geneticists, patient advocacy groups and patients, with differing levels of detail and complexity provided to laboratory personnel than to patients. Aim 3 will be completed in due course as more laboratory data has been assembled.

Analyses of variant data being shared by participating laboratories are done on a monthly basis. Variants identified by more than one laboratory are compared to determine concordance versus discordance among the various categories. Under discussion is how the COGR will handle discordant interpretations between sites. Considerations include having a working group or committee look over discordances or a notification based system that facilitates site-specific based review of the variant to resolve discrepancies.

The COGR project is not the first initiative that has attempted to create database resources for clinical molecular genetics with greater consistency and potential for collaboration. However, COGR differs from other efforts like the HGMD and locus specific mutation databases (LSDBs) in several ways. Unlike other databases, it will allow sharing between genetic specialists who will have the ability to monitor and update new information as it becomes available. Further, the consensus variant-level information will be made available to many different groups and individuals with an interest in clinical genetics without any associated cost. The database was also designed to accept information from many different sources such as from new research findings and has advanced capabilities including the ability to link genes to disease and DNA variants to drug response. As described above, the COGR features a VAT designed to help scientists interpret variants, a task becoming increasingly time-consuming as the use of new sequencing technologies becomes standard in clinical laboratories. The COGR database will not be merely a static record but a highly functional working

resource featuring attribution of variant data, tracking and versioning of information, an advanced hierarchical approval system, and consensus agreement system for variant interpretations, all while being able to maintain individual lab interpretations.

The COGR is collaborating with similar international efforts. The Clinical Genome Resource (ClinGen), is one such effort and is aimed at sharing and evaluating genomic variants and disease associations (http://clinicalgenome.org; http://www.nih.gov/news/health/sep2013/nhgri-25.htm). The ClinGen project includes the ClinVar database (http://www.ncbi.nlm.nih.gov/clinvar), operated by the National Center for Biotechnology Information as their depository of record. To date, over 139 000 variants have been deposited into ClinVar by over 288 laboratories and consortia (numbers as of 06 Mar 2015, http://www.ncbi.nlm.nih.gov/clinvar/submitters). As with the COGR project, the overall goal is to amass existing information spread across multiple sources and combine them in a single common resource so that all the scientific and clinical information contained therein can be shared with all potential users. Collaborations with the ClinGen project are inherently linked with the GeneInsight platform, which allows for direct access to the ClinVar database. At this point in time, the COGR is not yet uploading consensus variant data to ClinVar but can facilitate the upload process for individual laboratories. Once the COGR consensus database is firmly established, discussion will take place to decide what variant information will be pushed to other sources.

Like ClinGen, the HVP is another collaboration within the clinical genetics community aimed to improve international collaboration.[10] The organisation's focus is bringing together 'local' variant databases by creating standards and guidelines for genetic interpretations globally. Much like the COGR, the HVP puts emphasis on the standardisation of genetic variant interpretation across different laboratories, but at the international level. COGR has taken a grass-roots approach and will first standardise variant information at the national level, in order to better facilitate the entry of Canadian genetic information into the international community. As the first Canadian initiative for the sharing and standardisation of genetic variant information and the interim Canadian node for the HVP, COGR will play an active role in how data generated in Canada is shared with the international community.

As of April 2014, the COGR became a founding member of VariantWire, a clinical consortium of clinical laboratories across the USA and Canada that are sharing human genetic variant data in real time via the GeneInsight platform. Members of the COGR have the option to apply to and join the VariantWire network, and indeed multiple laboratories have already done so.

COGR will continue to collaborate with other established international projects and make these resources available to others in the clinical genetics field. In this way, the COGR initiative will contribute to the understanding of clinical genetic information and help facilitate integration of genomics into healthcare.

Summary

The COGR project aims to amalgamate existing variant data from individual clinical genetic laboratories across Canada into a centralised repository for the purposes of sharing and collaboration. By pooling variant information currently stored in individual clinical laboratories, the interpretation of human genetic variants can be made more clinically useful. There are many obstacles when sharing genetic variant information between different laboratories, including lack of a standardised variant classification system and differences in clinical reporting protocols. The COGR seeks to resolve these issues in three steps. First, to resolve differences in variant classification, a standardised VAT was developed and made freely available to all participating laboratories. Second, to bring genetic data from different laboratories together and facilitate the sharing process, a web-based instance of the GeneInsight platform was provided to all participating laboratories, making inherent use of its structure and sharing capabilities. Finally, using the shared data from participating laboratories, the project will create a publicly available repository of consensus interpretations for variants, including their classifications and implications for disease. In this way, consensus variant data that has been approved by different institutions in Canada will be presented to stakeholders at an appropriately detailed level. The COGR endeavours to serve as a focal point for the collaboration of Canadian laboratories with themselves and other countries in the development of tools and methods that leverage laboratory data in diagnosing, managing and treating genetic diseases. As more laboratories worldwide share data, knowledge will improve and ultimately lead to better patient care.

Collaborators Ron Agatep; Peter Ainsworth; Mohammad R Akbari; Melyssa Aronson; Gary D Bader; Raveen Basran; Andre Blavier; Andrea Blumenthal; Kathleen Buckley; Jodi Campbell; Philippe M Campeau; Melanie Care; Nancy Carson; Ronald Carter; George Charames; David Chitayat; George Chong; Edmond Chouinard; Kathy Chun; Kenneth J Craddock; Rod Docking; Andrea Eisen; Hanna Faghfoury; Sandra Farrell; Harriet Feilotter; Bridget Fernandez; Cynthia Forster-Gibson; William Foulkes; Robert Hegele; Spring Holter; Sheri Horsburgh; Lauren Hughes; Stacey Hume; Franny Jewett; Aly Karsan; Sam Khalouei; Joan Knoll; Elena Kolomeitz; Georges Maire; Christian Marshall; Elizabeth McCready; Michael J Moorhouse; Chantal Morel; Tanya Nelson; Brian O'Connor; Francis Ouellette; Jillian Parboosingh; Peter Ray; Heidi Rehm; Christie Riddell; David S Rosenblatt; Andrea Ruchon; Bekim Sadikovic; Kara Semotiuk; Stephen W Scherer; Cheryl Shuman; Josh Silver; Katherine Siminovitch; Lesley Solomon-Izsak; Marsha Speevak; James Stavropoulos; Lincoln Stein; Rhonda Tannenbaum; Deborah Terespolsky; Richard F Wintle; Beatrix Wong; Nora Wong; John S Waye; Michael O Woods; Philip Wyatt; Sean Young.

Contributors JL-E is a co-PI on the project, co-devised the repository, and wrote and revised the majority of the paper. He is also the guarantor. MW organised survey results, developed upload procedures for laboratories, and expanded and revised the paper. SW revised the paper. MSL is a co-PI on the project and revised the paper. For an expanded list of all member involvement, please see the online supplementary appendix 1 or visit the COGR website: http://opengenetics.ca/communities/

Funding This work was funded by the Government of Canada through Genome Canada and the Ontario Genomics Institute (OGI-070).

Competing interests The GeneInsight technology has been licensed to a company named GeneInsight, Inc. the stakeholders of which are Partners HealthCare System and Sunquest. MSL and SW are employees of Partners HealthCare.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Casto AM, Amid C. Beyond the Genome: genomics research ten years after the human genome sequence. *Genome Biol* 2010;11:309.
2 Collins FS, Green ED, Guttmacher AE, Guyer MS; US National Human Genome Research Institute. A vision for the future of genomics research. *Nature* 2003;422:835–47.
3 Green ED, Guyer MS; National Human Genome Research Institute. Charting a course for genomic medicine from base pairs to bedside. *Nature* 2011;470:204–13.
4 Khoury MJ, Bradley LA. Why should genomic medicine become more evidence-based? *Genomic Med* 2007;1:91–3.
5 Duzkale H, Shen J, McLaughlin H, Alfares A, Kelly MA, Pugh TJ, Funke BH, Rehm HL, Lebo MS. A systematic approach to assessing the clinical significance of genetic variants. *Clin Genet* 2013;84:453–63.

6 Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, Lyon E, Ward BE, Molecular Subcommittee of the ALQAC. ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. *Genet Med* 2008;10:294–300.

7 Aronson SJ, Clark EH, Babb LJ, Baxter S, Farwell LM, Funke BH, Hernandez AL, Joshi VA, Lyon E, Parthum AR, Russell FJ, Varugheese M, Venman TC, Rehm HL. The GeneInsight Suite: a platform to support laboratory and provider use of DNA-based genetic testing. *Hum Mutat* 2011;32:532–6.

8 Aronson SJ, Clark EH, Varugheese M, Baxter S, Babb LJ, Rehm HL. Communicating new knowledge on previously reported genetic variants. *Genet Med* 2012.

9 Neri PM, Pollard SE, Volk LA, Newmark LP, Varugheese M, Baxter S, Aronson SJ, Rehm HL, Bates DW. Usability of a novel clinician interface for genetic results. *J Biomed Inform* 2012;45:950–7.

10 Ring HZ, Kwok PY, Cotton RG. Human Variome Project: an international collaboration to catalogue human genetic variation. *Pharmacogenomics* 2006;7:969–72.

11 Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 2014;42(Database issue):D980–5.

12 1000 Genomes Project ConsortiumAbecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061–73.

13 International HapMap CAltshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Peltonen L, Dermitzakis E, Bonnen PE, Altshuler DM, Gibbs RA, de Bakker PI, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Yu F, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Gibbs RA, Muzny DM, Barnes C, Darvishi K, Hurles M, Korn JM, Kristiansson K, Lee C, McCarrol SA, Nemesh J, Dermitzakis E, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Bonnen PE, Gibbs RA, Gonzaga-Jauregui C, Keinan A, Price AL, Yu F, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Schaffner SF, Zhang Q, Ghori MJ, McGinnis R, McLaren W, Pollack S, Price AL, Schaffner SF, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD, McEwen JE. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010;467:52–8.

14 Schriml LM, Arze C, Nadendla S, Chang YW, Mazaitis M, Felix V, Feng G, Kibbe WA. Disease Ontology: a backbone for disease semantic integration. *Nucleic acids Res* 2012;40(Database issue):D940–6.