

Genomics4RD: An integrated platform to share Canadian deep-phenotype and multiomic data for international rare disease gene discovery

Hannah G. Driver¹  | Taila Hartley¹  | E. Magda Price¹  | Andrei L. Turinsky²  | Orion J. Buske³  | Matthew Osmond¹  | Arun K. Ramani²  | Emily Kirby⁴  | Kristin D. Kernohan^{1,5,6} | Madeline Couse²  | Hillary Elrick²  | Kevin Lu² | Pouria Mashouri²  | Aarthi Mohan²  | Delvin So² | Conor Klamann² | Hannah G. B. H. Le² | Andrea Herscovich⁶ | Christian R. Marshall^{6,7}  | Andrew Statia⁶ | Care4Rare Canada Consortium¹ | Bartha M. Knoppers⁴  | Michael Brudno^{3,6,8,9}  | Kym M. Boycott^{1,6} 

¹Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Canada

²Centre for Computational Medicine, The Hospital for Sick Children, Toronto, Canada

³PhenoTips, The Hospital for Sick Children, Toronto, Canada

⁴Centre of Genomics and Policy, McGill University, Montreal, Canada

⁵Newborn Screening Ontario, Children's Hospital of Eastern Ontario, Ottawa, Canada

⁶Genomics4RD Steering Committee, Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada

⁷Genome Diagnostics, The Hospital for Sick Children, Toronto, Canada

⁸Techna Institute, University Health Network, Toronto, Canada

⁹Department of Computer Science, University of Toronto, Toronto, Canada

Correspondence

Kym M. Boycott, Department of Genetics, Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, 401 Smyth Rd, Ottawa, ON K1H 8L1, Canada.
Email: kboycott@cheo.on.ca

Funding information

Genome Canada, Canadian Institutes of Health Research, Ontario Genomics, Ontario Research Fund, Genome Alberta, Genome British Columbia, Genome Quebec, and the Children's Hospital of Eastern Ontario Foundation, Grant/Award Number: OGI-147; Frederick Banting and Charles Best Canada Graduate Scholarship Doctoral Award from CIHR (T.H.); CIHR Foundation Grant and a Tier 1 Canada Research Chair in Rare Disease Precision

Abstract

Despite recent progress in the understanding of the genetic etiologies of rare diseases (RDs), a significant number remain intractable to diagnostic and discovery efforts. Broad data collection and sharing of information among RD researchers is therefore critical. In 2018, the Care4Rare Canada Consortium launched the project C4R-SOLVE, a subaim of which was to collect, harmonize, and share both retrospective and prospective Canadian clinical and multiomic data. Here, we introduce Genomics4RD, an integrated web-accessible platform to share Canadian phenotypic and multiomic data between researchers, both within Canada and internationally, for the purpose of discovering the mechanisms that cause RDs. Genomics4RD has been designed to standardize data collection and processing, and to help users systematically collect, prioritize, and visualize participant information. Data storage,

Hannah G. Driver, Taila Hartley, Michael Brudno and Kym M. Boycott contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Human Mutation* published by Wiley Periodicals LLC.

Health (K.M.B.), Grant/Award Number: FDN-154279

authorization, and access procedures have been developed in collaboration with policy experts and stakeholders to ensure the trusted and secure access of data by external researchers. The breadth and standardization of data offered by Genomics4RD allows researchers to compare candidate disease genes and variants between participants (i.e., matchmaking) for discovery purposes, while facilitating the development of computational approaches for multiomic data analyses and enabling clinical translation efforts for new genetic technologies in the future.

KEYWORDS

data sharing, deep phenotyping, genomic data, Genomics4RD, matchmaking, multiomic data, rare disease database

1 | INTRODUCTION

Although each rare disease (RD) affects only a small fraction of the population, RDs are estimated to collectively affect more than 450 million people globally (Ferreira, 2019). Approximately, 72% of all RDs are considered to be genetic in origin (i.e., resulting from disease-causing variation in genes; Nguengang Wakap et al., 2020). Over the past decade, advances in genomic technologies have contributed greatly to our understanding of the etiologies of RDs. Genome-wide sequencing (GWS) strategies, like exome and whole genome sequencing (ES and WGS, respectively), are widely credited with ushering in a new era of both disease-gene discovery (Bamshad et al., 2019; Boycott et al., 2013) and diagnostic test development, through which approximately 40% of RD families receive a diagnostic result (Clark et al., 2018). Along with the genomic sequencing technologies that are actively transitioning to the health-care system, other omic technologies, like long-read genome, transcriptome, and deep sequencing, as well as lipidomic, metabolomic, and epigenomic profiling, have begun to contribute to the discovery of the molecular mechanisms of genetic RDs (Boycott et al., 2019). Despite this, it is believed that approximately half of all monogenic diseases remain unsolved (Bamshad et al., 2019).

Over the past decade, a number of international organizations have been orchestrating a global push for data sharing to ensure the continued progress of our understanding of the molecular etiologies of RDs, and more generally, to enable precision medicine. Frameworks and standards to enable the responsible, voluntary, and secure sharing of genomic and health-related data have been created by the Global Alliance for Genomics in Health (GA4GH; www.ga4gh.org; Knoppers, 2014). Furthermore, one of the goals of the International Rare Diseases Research Consortium (IRDiRC) is that all undiagnosed RD patients have the option to be entered into a “globally coordinated diagnostic and research pipeline” (Austin et al., 2018). This push for a concerted effort to share RD data at a global level will help to enable the diagnosis of all RD patients.

Currently, one of the central approaches to gene discovery through data sharing is matchmaking (Figure 1). Matchmaking is the process by which two or more parties are connected who have

contributed data from RD patients with overlapping phenotypes and deleterious-appearing variant(s) in the same gene. Thus far, the primary approach to matchmaking has been “two-sided,” where both parties have flagged the same compelling candidate gene and have input the gene, with or without the phenotype and inheritance, into an existing matchmaking database, such as those connected through the Matchmaker Exchange (MME; www.matchmakerexchange.org; Philippakis et al., 2015). However, the efficacy of two-sided matchmaking is limited by the need for both parties to flag the same gene of interest before comparing RD patients. Thus, other forms of matchmaking, including one-sided (i.e., only the querier must flag a candidate gene and examine the variants in the same gene held in a database to identify connections) and eventually zero-sided (i.e., computational software identifies connections between phenotypically and/or genotypically similar individuals) matchmaking are imperative for RD gene discovery (Hartley et al., 2020).

While many databases geared toward RD gene discovery currently exist, they are typically limited in the types and specificity of data that they collect or display. For example, many of the federated databases connected by the MME only allow for the deposition of candidate genes or variants and do not require the input of standardized phenotypic information, such as specific human phenotype ontology (HPO) terms (Köhler et al., 2014). In contrast, other databases may contain extensive participant data, but restrict data sharing externally for a number of reasons including social, ethical, financial, and legal hurdles. Some of these hurdles may be overcome by implementing a centralized database with clearly defined rules and regulations, careful curation of data, and dedicated personnel for database management and bioinformatic processing. However, the centralized approach, with its concomitant oversight, needs to be carefully structured based on data sharing use cases to avoid making the process of accessing data a barrier in and of itself.

The Care4Rare Canada Consortium, formerly known as FORGE Canada, is a pan-Canadian multisite research consortium that has been using next-generation sequencing (NGS) technologies for gene discovery and clinical translation purposes for the past decade (Beaulieu et al., 2014; www.care4rare.ca). Over the past 5 years, the majority of Canada's 13 provincial and territorial health-care systems

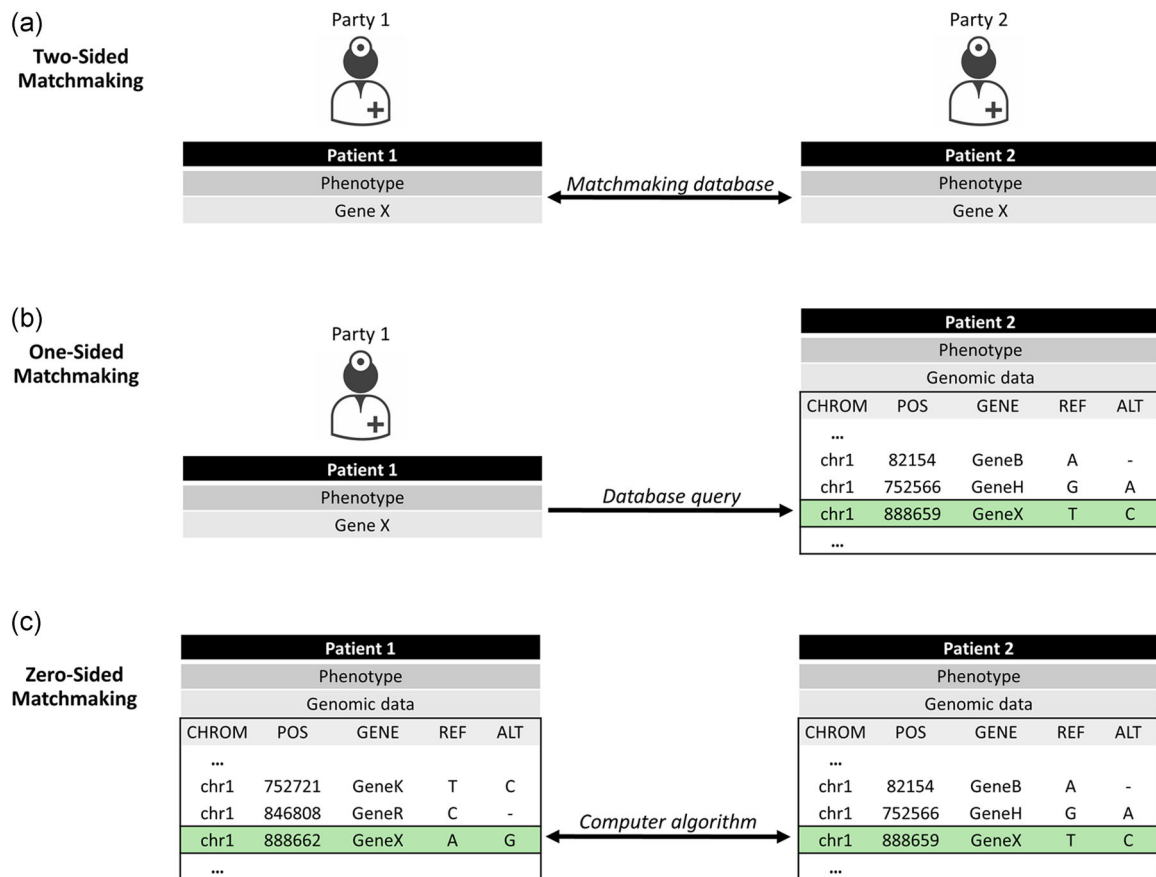


FIGURE 1 Approaches to patient matchmaking. (a) Two-sided matchmaking: parties flag the same candidate gene and subsequently exchange more detailed information related to patient phenotypes and DNA variation to resolve possible matches. (b) One-sided matchmaking: one party flags a candidate gene and queries a database of genomic variation to examine DNA variants present in the data along with phenotypic and inheritance information to identify matches. (c) Zero-sided matchmaking: automated computational approaches compare patient phenotypes and genotypes within or between databases to identify matches

have implemented access to publicly funded GWS for RD diagnosis. The challenge now becomes how to leverage these data for secondary use, particularly for undiagnosed patients with a suspected RD, in the context of multiple institutional, jurisdictional, and provincial clinical data custodians in Canada. To address these rapidly emerging data silos, advance RD gene discovery, and facilitate precision medicine for patients with RDs, Care4Rare Canada launched Genomics4RD, a national integrated web-accessible research platform to share Canadian phenotypic and multiomic data, from consented patients, with researchers both within Canada and internationally.

2 | METHODS AND RESULTS

2.1 | Genomics4RD overview

We set out to build Genomics4RD as a centralized repository to collocate, harmonize, and share clinical information and multiomic data both nationally and internationally for the purpose of RD gene discovery, diagnosis, and other related research (Figure 2). Genomics4RD is designed so that data entry is compatible with both prospective and

retrospective (i.e., previously collected) research-consented Canadian RD data. Using a cloud-based infrastructure, Genomics4RD includes a web portal where users can access resources, as well as collect, store, visualize, and share data. We aimed to maximize standardized data sharing via ethically and legally harmonized data collection, use, and governance. Genomics4RD first launched for prospective data collection in February 2019, and retrospective cohorts began to be transferred into the database in January 2021. As of September 2021, there were more than 3300 participant records in Genomics4RD, representing approximately 1000 participants with largely undiagnosed RDs and their family members.

2.2 | Governance and storage

At the core of the Genomics4RD database is its unambiguously structured governance framework, prepared by the Centre of Genomics and Policy at McGill University. The framework aims to provide oversight mechanisms regarding the administration, stewardship of, and access to data. The framework describes criteria for the identification of eligible retrospective and prospective RD

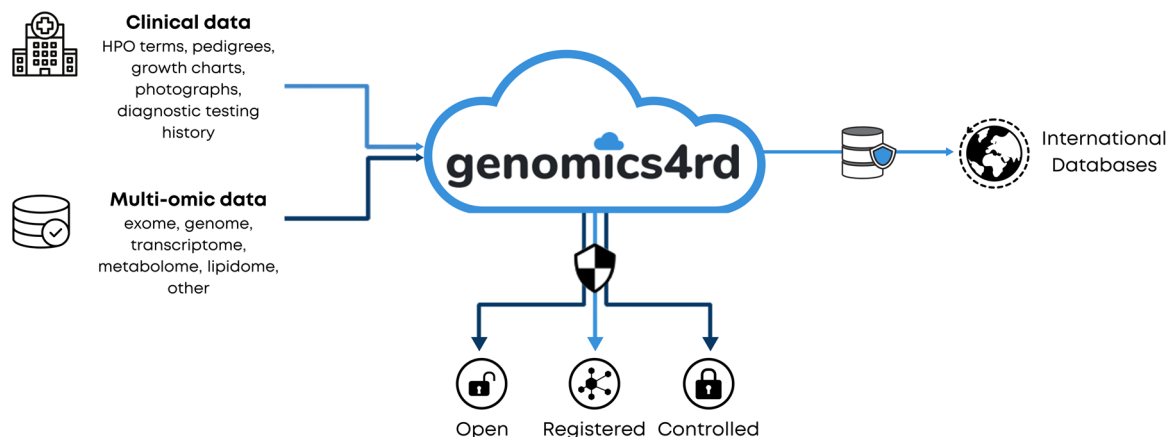


FIGURE 2 Genomics4RD vision. Data depositors from across Canada contribute both clinical and multiomic data to Genomics4RD, which harbors a variety of tools within its web-accessible portal and makes data available to external researchers via various data sharing processes

cohorts and outlines the process by which data use restrictions or conditions are attributed to different cohorts using the GA4GH's Automatable Discovery and Access Matrix (ADA-M) codes (Woolley et al., 2018). Cohorts are defined as sets of data that have been collected using a given research ethics board (REB) protocol and therefore have unique data sharing permissions. An ADA-M profile is created for all data related to a given cohort that is deposited into Genomics4RD and acts as a form of structured metadata (i.e., data about the cohort's data sets), which allows for standardized and effective communication of access and use conditions.

Oversight of the database is provided by a Steering Committee which consists of a minimum of two Genomics4RD project leaders, two cohort project leaders, and one RD community stakeholder. The Steering Committee is responsible for overseeing Genomics4RD progress, operations, administration, budget, knowledge translation, and interaction with end-users. The Steering Committee receives guidance from an independent Advisory Board, composed of internationally recognized experts in RD, molecular genetics, systems biology, bioinformatics, ethics, commercialization, health technology, and health policy. Furthermore, the Steering Committee provides oversight for the activities of the Data Access Committee (DAC), which consists of five members including a patient advocate, an RD researcher, a technical expert, and individuals with expertise in legal/ethical issues in RD research, privacy, and pediatrics. The DAC adjudicates requests for registered and controlled access to data in accordance with the ADA-M profiles of the cohorts within Genomics4RD.

Finally, the framework sets out ethical considerations pertaining to the storage of data in Genomics4RD. For example, privacy and security protections are detailed such that the infrastructure must be hosted on a cloud-based server located in Canada and meet the relevant security and privacy standards for storage of health-related data. As such, the Genomics4RD platform is hosted in a dedicated cloud tenancy provided by HPC4Health; a large-scale cloud infrastructure for health informatics built and operated jointly by the

Hospital for Sick Children and the University Health Network in Toronto since 2015 (www.hpc4health.ca). Hosting the Genomics4RD platform on HPC4Health provides a number of important benefits, such as enterprise-level governance, management and technical support, and the capability to handle both research and clinical data processing with full compliance to all internal and external data handling policies, including multiple third-party security audits.

Additional portions of the framework describe the closure/transfer of the database, management of participant withdrawal, and potential risks and benefits of participating in the database. In doing so, it assists potentially eligible cohorts and their respective REBs in assessing the oversight mechanisms and data sharing opportunities presented by Genomics4RD. We have linked the framework to a library of consent templates and REB authorizations (e.g., approvals or consent waivers, where appropriate) to aid in sharing data and depositing it into Genomics4RD, while ensuring that participant data are collected, stored, tagged with sharing permissions, accessed, and, where applicable, withdrawn appropriately. A copy of the governance framework is available upon request.

2.3 | Data collection, visualization, and access for data depositors

Genomics4RD data depositors can collect, store, visualize, and share their data using the platform. Genomics4RD users who deposit data into the database receive full access and edit permissions for their respective participant records. Thus, data depositors may enter and subsequently edit their participants' data and metadata, as well as download their raw and processed multiomic data. Data depositors can also examine and search across other participant profiles within Genomics4RD, but they may not edit or download these data. For privacy and security reasons, all participant data entered into Genomics4RD are coded without standard personal health information.

2.3.1 | Multiomic data collection

The Genomics4RD platform is backed by a fully integrated bioinformatics system for data management, analysis, and tracking. The platform supports a growing range of multiomic data, such as exome, genome, transcriptome, and eventually metabolome and lipidome, that originate from both research studies and clinical testing with research consent for secondary use. To submit multiomic data into Genomics4RD, a user must supply the following minimal metadata: research IDs, tissue type from which the data were derived, data type (e.g., ES and WGS), and sequencing date. Data sets can be further annotated with sequencing-specific attributes such as the library preparation kit, capture kit, and sequencing center that were used. Data depositors securely transfer their data files into the Genomics4RD data repository using MinIO, an open-source Amazon S3-compatible object storage server (www.min.io). Data transfer is enabled both through a web browser interface and through a command-line client. The latter is especially well-suited for large file sharing as it verifies checksums upon transfer completion, so that interrupted or failed transfer operations can be resumed from the point of failure.

Concurrently, the related sample information, family relationships, analysis requests, and other metadata are tracked using a specialized laboratory information management system called Sample Tracking and Genomics Resource (STAGeR), developed by our team. STAGeR allows users to input genomic metadata and edit annotations, link them to the relevant data sets, and request specific analyses and track their statuses. The Genomics4RD backend can thus support both structured and unstructured data types through the combined use of MinIO and STAGeR. MinIO offers a number of security features for user identity and access management, also supported via STAGeR. In particular, the data from separate sources or subprojects are deposited into their specialized data folders, called buckets, where secure access to each bucket is provided only to the relevant project participants.

Once the data are transferred to Genomics4RD, analyses are enabled using a suite of bioinformatics pipelines for the corresponding data types. The pipelines are built using Snakemake, a modern workflow management system that supports reproducible and scalable data analysis (Köster & Rahmann, 2012; links to the available pipelines can be found under Web Resources). Data are processed in a standardized manner according to the NGS type. For example, all exomes are processed using the exome subworkflow of the *crg2* pipeline (see Web Resources), while all genomes are processed using the genome subworkflow of the *crg2* pipeline. All pipelines use the GRCh37 reference with decoy sequence. Variants are annotated with predicted gene impacts, population frequencies, and disease-phenotype association using Variant Effect Predictor (McLaren et al., 2016), *vcfanno* (Pedersen et al., 2016), and custom *crg2* scripts. Rare variants (gnomAD maximum population allele frequency < 1%) impacting protein-coding sequences and splice regions are uploaded to PhenoTips and are made available to the users of Genomics4RD through the variant viewer.

2.3.2 | Clinical data collection

All clinical data are collected via the Genomics4RD web portal using a PhenoTips platform (<https://phenotips.com>) that has been customized to include additional variables to support gene discovery and clinical translation efforts (Girdea et al., 2013). The PhenoTips platform enables users to record demographics, medical history, family history (in the form of a pedigree that links to relevant family members within the database), prenatal history, physical and laboratory measurements, and physical findings represented as HPO terms, along with additional metadata (onset, laterality, and pace of progression), and supporting materials (images, documents; Figure S1). The platform also records specific genetic information for a given participant and allows users to annotate the status of particular genes, whether they have been ruled out, identified as a candidate, or confirmed causal. In addition, Genomics4RD supports recording clinical and molecular diagnoses using terms from the Online Mendelian Inheritance in Man (OMIM) and Orphanet Rare Disease ontology (ORDO) technologies. At a minimum, cohorts that are deposited in Genomics4RD are required to include phenotypes in the form of HPO terms and genetic information in the form of either a candidate gene or genomic data set. Additional benefits from using the PhenoTips platform include being able to upload participant data (including phenotypic information) in bulk into Genomics4RD using the PhenoTips REST application programming interfaces (APIs) in either the PhenoTips JSON or GA4GH Phenopacket formats. Bulk data can be exported or transferred to other systems through several interoperable formats: (1) GA4GH Phenopackets, (2) JSON export, (3) Excel export, (4) User-triggered or automated synchronization of participant records to another instance that supports the PhenoTips Patient Push API.

The most common data model for representing a participant's clinical phenotype for research and GWS-based analysis is an unordered collection of HPO terms. However, when a participant presents with many clinical features, it can be difficult for both RD researchers and computational algorithms to determine which are the most relevant. Our research suggests that just five distinctive features may be sufficient for prioritizing GWS-based sequencing results (Kernohan et al., 2018). To support the curation of these distinctive features, the Genomics4RD user interface allows users to "star" the curated HPO terms, which are then displayed more prominently.

While clinical GWS tests have demonstrated utility in providing diagnoses for unsolved RD patients, details such as the ideal time within the diagnostic odyssey to initiate such testing, the cost effectiveness of GWS versus targeted diagnostic tests, and impact on clinical management have not been well characterized. In addition, new and emerging technologies will require similar evidence to support translation and reimbursement in the future. Therefore, in addition to the collection of standardized phenotypic terms, Genomics4RD also supports collection of data related to a participant's diagnostic and management care pathway. The care pathway data collection form consists of two modules, one that is to be filled in before testing and a second that is to be collected after testing. The pretesting module includes: (a) patient characteristics, (b) DNA-based diagnostic tests performed to date, (c)

non-DNA-based diagnostic tests performed to date, (d) specialists involved in ongoing care, and (e) a record of the diagnostic tests that the responsible clinician would have ordered if the test was not available.

The post testing module includes: (a) laboratory details, testing details, and laboratory interpretation of testing results; (b) clinical interpretation of testing results; (c) diagnostic testing ordered if the test was nondiagnostic; (d) diagnostic testing averted if the testing was diagnostic; and (e) medical management activities triggered by the test results. All of these data are captured using structured questions that have suggested, standardized responses, including a data dictionary that contains over 250 medical tests grouped into 6

categories (biochemistry and metabolic testing, cytogenetic testing, electrical activity, imaging, molecular testing, and pathology). Collecting data related to participant care pathways and outcomes will be instrumental in translating new diagnostic testing to the clinic.

2.3.3 | Data visualization

Genomics4RD data depositors can visualize and search clinical and genomic data, either at an individual level or across the database (Figure 3). The participant variant viewer is available within individual

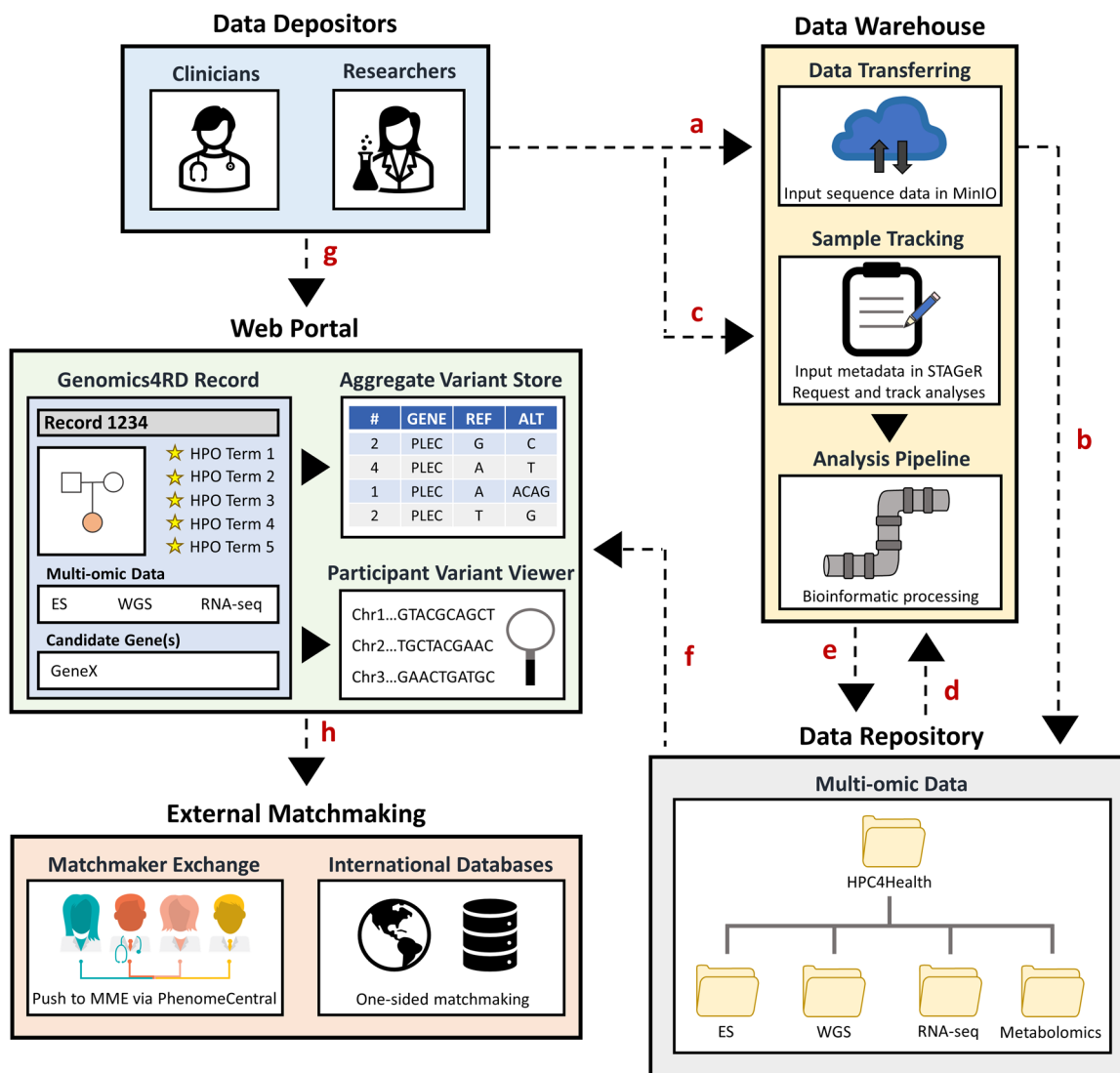


FIGURE 3 Genomics4RD data input and visualization. (a) Raw multiomic sequencing files are uploaded into the secure MinIO server for (b) deposition into the data repository. (c) Metadata about participants are entered into a laboratory management system called STAGeR. (d) Raw multiomic data files are run through their respective bioinformatic pipelines for processing and (e) stored in the Genomics4RD data repository. (f) The catalog of available multiomic data sets will be displayed for each participant record, and the filtered genomic results are directly browsable through the variant viewer interfaces. (g) Data depositors may access the Genomics4RD user interface via the web portal, where they complete coded records for their participants (including inputting clinical data, pedigrees, prioritized HPO terms, and candidate genes). Within the web portal, users can visualize their participants' processed files via the participant variant viewer and query phenotypic and variant-level data across Genomics4RD. (h) Further matchmaking opportunities can be explored by pushing Genomics4RD records to the PhenomeCentral node of the Matchmaker Exchange (for two-sided matchmaking), and through connections with other international databases (for one-sided matchmaking). HPO, human phenotype ontology

Access tier	Features
Data depositor	<ul style="list-style-type: none"> • View all participant records across the database • Access to aggregate variant store • Edit metadata for cohorts entered by the data depositor • Download raw and processed multiomic data for cohorts entered by the data depositor • Push functionality to PhenomeCentral for participants entered by the data depositor • Access to external one-sided matchmaking platform
Open access user	<ul style="list-style-type: none"> • View aggregate database information <ul style="list-style-type: none"> ◦ number of records ◦ number of diagnosed participants ◦ distribution of HPO terms ◦ list of flagged candidate genes ◦ list of unsolved recognizable disorders ◦ number of multiomic data sets
Registered access user	<ul style="list-style-type: none"> • Access to external one-sided matchmaking platform
Controlled access user	<ul style="list-style-type: none"> • Download appropriate clinical information from participant records • Download raw and processed multiomic data for approved participants

Abbreviation: HPO, human phenotype ontology.

participant profiles and provides a user-friendly interface that allows data depositors to view and filter variants across more than 50 DNA variant annotations. The variant viewer integrates the preselected HPO, and HPO to gene annotations (<https://hpo.jax.org/app/download/annotation>) to enable users to easily filter for variants in genes associated with the participant's phenotype. Genomics4RD also features an aggregate variant store that allows users to search and visualize an aggregation of variant-level information across the database (Figure S2). This aggregate-level view of Genomics4RD variant data can be multifiltered for a variety of genotypic factors, including specific genes of interest, allele frequency, and variant zygosity. Data depositors can therefore use the aggregate variant store to search for variants of interest in their queried candidate gene and then directly view the full profiles of participants with the identified variants to answer their specific research question.

2.3.4 | Technology stack

Genomics4RD comprises two separate components that represent the main user-accessible portal based on PhenoTips, and the backend data analysis system, respectively, with an API-based communication established between the two.

The main user-accessible web portal is implemented as a dedicated instance of the PhenoTips platform which has been customized to support additional data entry and presentation. The PhenoTips frontend is written in a combination of React, Javascript, and server-rendered HTML. Authentication uses the OpenID Connect (OIDC) protocol with user identities managed by Auth0 (<https://auth0.com/>). The backend server is implemented as a Java web application and is

TABLE 1 Genomics4RD features available to users in accordance with their level of access

containerized with Docker. Data persistence uses Java Hibernate and a local MariaDB server, with Apache Solr indices for search capability.

The internal data analysis ecosystem is accessed through STAGeR, which operates independently from PhenoTips. The frontend is written in TypeScript and React, with Material-UI for theming. The backend is written in Python 3 and Flask, which uses SQLAlchemy to communicate with a MySQL database for storing metadata and analysis statuses, and is containerized with Docker. MinIO is used to store the uploaded multiomic data and the results of their analyses. GitHub Actions continuous integration manages software deployment. The bioinformatics analysis pipelines are written in Python using the Snakemake workflow management system. They can be run on high-performance clusters with the Slurm scheduler as is done in Genomics4RD, or with the TORQUE-Mohab scheduler. Source code for both is hosted on GitHub.

2.4 | Genomics4RD data sharing

In accordance with the governance framework, and the identified data use restrictions or conditions of the different cohorts, data within Genomics4RD are made available to the international RD research community through three tiers of data access: open, registered, and controlled (Table 1 and Figure 2).

2.4.1 | Open access

Genomics4RD displays several types of open data on our website (<https://www.genomics4rd.ca/>) including the number of participant

records, the number of diagnosed participants, the distribution of HPO terms, a list of candidate genes flagged in all participants, a list of unsolved recognizable disorders, and the number of participant files related to each multiomic approach stored in Genomics4RD. These publicly available, aggregate data are updated quarterly to reflect the addition of new participants and changes in the database. Given that these data are displayed on our website, open access is available globally, and only requires an internet connection.

2.4.2 | Registered access

RD researchers who are not data depositors but would like to query the data within Genomics4RD will be able to do so using registered access (Dyke et al., 2016, 2018) for specific use cases. With this type of access, researchers will sign a registered-access agreement, which mandates appropriate use of the data including confidentiality, security, respect of consent withdrawal, and the return of information to Genomics4RD. The main use case for registered access at this time is one-sided matchmaking, which we describe further below. Registered access can be granted to researchers who apply to, and are approved by, the Genomics4RD DAC and have study approval at their own institution. Once this type of access is available, registered access requests can be emailed to Genomics4RD@cheo.on.ca.

2.4.3 | Controlled access

The final level of data access available in Genomics4RD is the controlled access tier. Controlled access may be granted to external researchers who wish to perform research related to RDs using appropriately consented data hosted in Genomics4RD, based on ADA-M access permissions. Users granted controlled access are able to download participant-level clinical data, structured multiomic data files (e.g., FASTQ, BAM, and VCF sequencing files; and processed VCF files), as well as other unstructured multiomic data (e.g., transcriptomes and metabolomes) from the Genomics4RD data repository through a separate sample management interface. Controlled access can be granted to bonafide third-party researchers who apply to, and are approved by, the Genomics4RD DAC as well as their respective institutional REBs. Expressions of interest for controlled access to Genomics4RD, including questions related to the types of data available, or further clarification as to the application process, may be sent to Genomics4RD@cheo.on.ca.

2.5 | External matchmaking approaches enabled by Genomics4RD

We have designed Genomics4RD to facilitate matchmaking with international RD researchers for the discovery of mechanisms causing RD.

2.5.1 | Integration with phenomecentral for two-sided matchmaking via the MME

For RD participants with a manually identified candidate variant(s) in a novel disease-gene, the simplest and most effective form of data sharing for the purposes of gene discovery is through two-sided matchmaking using databases affiliated with the MME (Philippakis et al., 2015). The databases are configured such that connections are made between any two parties that have flagged the same gene of interest. Information on phenotype and mode of inheritance shared by the database, or more often by subsequent communication between the two parties, is used to reject the candidate gene or suggest that more discussion and collaboration is the next step. There are now over 150,000 entries and 13,000 unique genes from 11,000 international contributors connected via the eight RD databases connected by the MME, making it an incredible source of potential information for discovery (Azzariti & Hamosh, 2020). In our experience, 15% of candidate genes entered into the MME lead to collaboration and subsequent publication of a novel discovery (Osmond et al., 2022). As such, we wanted to encourage the submission of candidate genes to the MME by integrating Genomics4RD with PhenomeCentral (Buske et al., 2015). It was important that we reduce the burden of double data entry and ensure consistency in the submissions between the two sites (Genomics4RD and PhenomeCentral). To do so, a “push” functionality was added to Genomics4RD to allow data depositors to migrate their records into PhenomeCentral. Once records are pushed to PhenomeCentral, the phenotype (as a list of HPO terms) and candidate genes (including auxiliary details about the variants and their inheritance pattern) are available for matching with other nodes of the MME. The Genomics4RD record maintains a link to the corresponding record in PhenomeCentral so that any updates to the phenotypes and candidate genes in Genomics4RD can be easily synchronized with PhenomeCentral. As of September 2021, over 200 candidate genes within Genomics4RD have been shared with the MME.

2.5.2 | External one-sided matchmaking solutions for Genomics4RD data

Given the potential for discovery that remains in genomic data that has not had a candidate gene flagged, we considered it very important to facilitate one-sided matchmaking on data housed in Genomics4RD. We have thus developed a specific platform to facilitate this use case by enabling multilevel filtering for both genetic variation and phenotypic information within a single user interface and ensuring that compound heterozygous variants in a single participant are identifiable. Queriers will specify their gene of interest, and then they will be given a list of variants that occur within that gene across all participants in Genomics4RD. These query results can be expanded to view additional details on the variant (such as allele frequency and in silico pathogenicity scores) as well as case-specific details (i.e., variant zygosity, read depth, and HPO phenotypes). The

platform also provides queriers with the ability to contact the data depositor of any match of interest for follow-up and potential collaboration. External users coming to Genomics4RD for one-sided matchmaking will access this platform using the registered access model described above, and a pilot is currently ongoing to evaluate the utility of this approach. We also developed this platform to support linking Genomics4RD with other similar databases, with an aim to onboard at least one other RD database within the next year. We anticipate that this platform will enable one-sided matchmaking queries across large numbers of unsolved cases, maximizing the chances of gene discovery.

2.5.3 | Future directions to enable zero-sided matchmaking in Genomics4RD

In addition to one- and two-sided matchmaking, we aim to develop tools in Genomics4RD that will facilitate zero-sided matchmaking. Using computational approaches, phenotypic and multiomic data will be compared between participants with the goal of identifying similar mechanisms of disease. For example, phenotypes will be inspected to identify similarly affected participants, and their multiomic data will be subsequently compared in an effort to pinpoint a shared disease gene or variant. Alternatively, multiomic data may first be scanned to identify shared, putative pathogenic variants, with matches then being narrowed down based on phenotypic comparisons. Without the need for human input, the efficiency of zero-sided matchmaking is expected to have huge implications for disease-gene discovery and diagnostic yields in RD patients in the future.

3 | DISCUSSION

Genomics4RD has been designed to address current and future data sharing needs to enable diagnoses for Canadian families with RDs. It operates as a versatile database that is able to collect a broad range of data types and share portions of its data in accordance with the data sharing permissions or restrictions of the individual cohorts, effectively allowing data to be shared to its maximum potential. The immediate utility of the data in Genomics4RD is in addressing the remaining discovery potential of the coding portion of the genome. For participants with candidate genes identified, Genomics4RD provides access to the MME, which links to tens of thousands of patients across the world (Azzariti & Hamosh, 2020).

For participants without candidate genes identified, we have developed a means of querying our database for rare, high-quality variants with the aim of identifying new candidate genes not yet highlighted in a participant's data (one-sided matchmaking), ensuring the maximum discovery potential of existing data. We have piloted one-sided matchmaking within the Care4Rare Canada cohort, and since October 2019, we have queried our cohort for 361 candidate genes (including new disease-gene candidates identified through Care4Rare analyses, collaborator requests, and new OMIM disease-

gene associations), of which 12 have resulted in new diagnoses. This pilot demonstrated that 3% of one-sided matchmaking queries resulted in a diagnosis in our cohort of approximately 1000 undiagnosed participants. In one instance we were able to diagnose multiple additional families with a single query. During a reanalysis of ES data, we identified a homozygous variant in the gene *ABHD16A* in two French Canadian families with complicated hereditary spastic paraplegia, which had been missed by our previous bioinformatics pipeline. Using one-sided matchmaking in Genomics4RD, we identified biallelic variants in two additional families (Lemire et al., 2021). Overall, reviewing variants and corresponding participant phenotypes for one-sided matches does take time and yields fewer discoveries so far than two-sided matchmaking, for which half of the genes in the human genome have been flagged and input into matchmaking databases (Azzariti & Hamosh, 2020). However, in our experience, one-sided matchmaking can lead to diagnoses in participants for whom a diagnosis would be otherwise unlikely – particularly for those with singleton exome data where the potential for novel discovery is lower. While the current means of data sharing will continue to yield discoveries, they are by no means sufficient. Broader data sharing internationally is critical to driving mechanistic discovery for RDs. To that end, we contributed as the Canadian Anchor Project for the World Economic Forum's Breaking Barriers to Health Data Project, with partners such as Genomics England, the Australian Genomics Health Alliance, and Intermountain Health (Chediak, 2020). Building on the valuable work done by this group, it is our intention to expand one-sided matchmaking in Genomics4RD such that users can link out to similar RD databases for the purposes of one-sided matchmaking.

Despite the continued high research value of the coding portion of the genome, the genetic mechanism of disease for some patients will be intractable to the methods described above. A number of new technologies have shown potential utility in identifying a small portion of additional diagnoses. Consequently, it was important that broad data types be captured by Genomics4RD. By harmonizing and storing these data, and making them available to researchers, we aim to enable new computational approaches to identify novel genetic mechanisms. Different data types may be examined individually or in combination, either manually or eventually using automated computation. These analyses can provide additional functional evidence for known genetic mechanisms (e.g., examining splicing effects for an intronic variant), or, in contrast, can be used to explore new genetic mechanisms (e.g., transcriptomic data can be looked at in combination with genomic data to identify noncoding RNA or regulatory DNA disease-causing variation).

Ultimately, the goal for any research focused on delineating the diagnostic potential of a new technology is to inform the appropriate translation of such technology into the health-care system. To that end, we felt it was important to ensure that Genomics4RD could facilitate the data collection necessary to support the clinical translation of the next wave of genetic technologies. As such, Genomics4RD has been set up to collect health services data that can be used to build evidence for reimbursement by funders of healthcare. These customizations are actively being used to collect clinical utility

and cost-effectiveness data related to exome and genome sequencing in collaboration with two health-care funders in Canada. Similarly, Genomics4RD is set up to collect the necessary health services data for the evaluation of transcriptome sequencing when the time comes to capture the cost-effectiveness, value-for-money, and impact on patient care in a real-world setting. Such data will prove useful in supporting health technology assessments and the development of best-practice guidelines.

Finally, we have positioned Genomics4RD to capture the data from clinical GWS in Canada for secondary use. The pace of GWS for RD diagnosis within the health-care system is now outpacing research. It is our intention that every patient that receives GWS in Canada will have the opportunity to contribute their data to this platform. Recently, Canadian clinical laboratories have begun offering provincially performed GWS and we are working nationally to build a recontact registry to facilitate secondary use of this clinically generated data within Genomics4RD.

4 | CONCLUSIONS

Despite recent advances in multiomic sequencing technologies, there are still a considerable number of RD patients who remain without a known molecular etiology for their disease. The future of RD gene discovery and diagnosis will thus rely on broad-scale sharing of both clinical and research generated data of diverse types. With this in mind, we designed Genomics4RD to harmonize and colocate multiomic and phenotypic data from all consented Canadian RD patients for the purpose of disease-gene discovery. Genomics4RD users can capitalize on various matchmaking approaches available through the database, to gain additional evidence of molecular RD mechanisms and increase confidence in RD diagnoses. With its internal tools and external connections to international databases, Genomics4RD will help to facilitate often long-awaited diagnoses for RD patients around the world.

WEB RESOURCES

Auth0: <https://auth0.com/>

Care4Rare: <http://www.care4rare.ca>

Centre of Genomics and Policy: <https://www.genomicsandpolicy.org/>

Global Alliance for Genomics and Health: <https://www.ga4gh.org/>

Genomics4RD: <https://www.genomics4rd.ca>

HPC4Health: <http://www.hpc4health.ca>

HPO annotations: <https://hpo.jax.org/app/download/annotation>

International Rare Diseases Research Consortium: <https://irdirc.org/>

Matchmaker Exchange: <https://www.matchmakerexchange.org/>

MinIO documentation: <https://docs.min.io/docs/minio-client-complete-guide.html>

PhenomeCentral: <https://phenomecentral.org/>

PhenoTips: <https://phenotips.com/>

STAGeR documentation: <https://github.com/ccmbioinfo/stager>

WGS and WES Snakemake pipeline documentation: <https://github.com/ccmbioinfo/crg2>

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Care4Rare Canada Consortium (funded by Genome Canada, the Canadian Institutes of Health Research [CIHR], Ontario Research Fund, Genome Alberta, Genome British Columbia, Genome Quebec, and Children's Hospital of Eastern Ontario Foundation under grant OGI-0147) for providing the opportunity to collectively study undiagnosed RDs. The authors would like to thank Élisabeth Soubry for her help in producing Figure 2. T.H. was supported by a Frederick Banting and Charles Best Canada Graduate Scholarship Doctoral Award from CIHR. K.M.B. was supported by a CIHR Foundation Grant (FDN-154279) and a Tier 1 Canada Research Chair in Rare Disease Precision Health. The study was supported by Grant Numbers Genome Canada-OGI-147.

CONFLICTS OF INTEREST

Orion J. Buske and Michael Brudno have an equity interest in, and Orion J. Buske is an employee of PhenoTips, which licenses software used in the Genomics4RD database.

AUTHOR CONTRIBUTIONS

Hannah G. Driver, Taila Hartley, Andrei L. Turinsky, Orion J. Buske, Matthew Osmond, and Kym M. Boycott drafted the manuscript. Emily Kirby, Bartha M. Knoppers, Taila Hartley, Michael Brudno, and Kym M. Boycott developed the Genomics4RD Governance Framework and related documents. E. Magda Price, Orion J. Buske, Taila Hartley, Matthew Osmond, Arun K. Ramani, Kristin D. Kernohan, Michael Brudno, and Kym M. Boycott have been responsible for the overall vision and project management (both overall and site-specific) of Genomics4RD. Orion J. Buske customized the PhenoTips software. Andrei L. Turinsky, Arun K. Ramani, Madeline Couse, Hillary Elrick, Kevin Lu, Pouria Mashouri, Aarthi Mohan, and Delvin So have been integral to the development of components of the Genomics4RD backend. Conor Klamann, Hannah G. B. H. Le, Matthew Osmond, Orion J. Buske, E. Magda Price, Taila Hartley, Arun K. Ramani, Andrei L. Turinsky, Hannah G. Driver, Kym M. Boycott, and Michael Brudno worked on development of the external one-sided matchmaking platform. E. Magda Price and Emily Kirby have overseen retrospective data deposition and ADA-M coding. Andrea Herscovich, Christian R. Marshall, Andrew Statia, Kristin D. Kernohan, Michael Brudno, and Kym M. Boycott provided valuable guidance as part of the Genomics4RD Steering Committee. All authors critically revised the manuscript.

DATA AVAILABILITY STATEMENT

Genomics4RD open access data are available on our website, at <https://www.genomics4rd.ca/>. Inquiries about obtaining registered or controlled access to our database can be emailed to Genomics4RD@cheo.on.ca.

ORCID

Hannah G. Driver  <http://orcid.org/0000-0002-6780-6899>

Taila Hartley  <http://orcid.org/0000-0002-1957-9820>

E. Magda Price  <http://orcid.org/0000-0002-9447-2310>
 Andrei L. Turinsky  <http://orcid.org/0000-0002-5555-8461>
 Orion J. Buske  <http://orcid.org/0000-0002-9064-092X>
 Matthew Osmond  <http://orcid.org/0000-0002-5886-873X>
 Arun K. Ramani  <http://orcid.org/0000-0002-8162-8523>
 Emily Kirby  <http://orcid.org/0000-0002-9419-6600>
 Madeline Couse  <http://orcid.org/0000-0002-3373-3214>
 Hillary Elrick  <http://orcid.org/0000-0001-6178-1670>
 Pouria Mashouri  <http://orcid.org/0000-0003-4332-530X>
 Aarthi Mohan  <http://orcid.org/0000-0003-2556-9730>
 Christian R. Marshall  <http://orcid.org/0000-0002-4003-7671>
 Bartha M. Knoppers  <http://orcid.org/0000-0001-7004-2722>
 Michael Brudno  <http://orcid.org/0000-0001-7947-2243>
 Kym M. Boycott  <http://orcid.org/0000-0003-4186-8052>

REFERENCES

- Austin, C. P., Cutillo, C. M., Lau, L. P., Jonker, A. H., Rath, A., Julkowska, D., Thomson, D., Terry, S. F., de Montleau, B., Ardigò, D., Hivert, V., Boycott, K. M., Baynam, G., Kaufmann, P., Tarusico, D., Lochmüller, H., Suematsu, C. I., Draghia-Akli, R., Norstedt, I., ... H. J. S. (2018). Future of rare diseases research 2017–2027: An IRDiRC perspective. *Clinical and Translational Science*, 11, 21–27. <https://doi.org/10.1111/cts.12500>
- Azzariti, D. R., & Hamosh, A. (2020). Genomic data sharing for novel mendelian disease gene discovery: The matchmaker exchange. *Annual Review of Genomics and Human Genetics*, 21, 305–326. <https://doi.org/10.1146/annurev-genom-083118-014915>
- Bamshad, M. J., Nickerson, D. A., & Chong, J. X. (2019). Mendelian gene discovery: Fast and furious with no end in sight. *The American Journal of Human Genetics*, 105, 448–455. <https://doi.org/10.1016/j.ajhg.2019.07.011>
- Beaulieu, C. L., Majewski, J., Schwartzenruber, J., Samuels, M. E., Fernandez, B. A., Bernier, F. P., Brudno, M., Knoppers, B., Marcadier, J., Dymont, D., Adam, S., Bulman, D. E., Jones, S. J., Avar, D., Nguyen, M. T., Rousseau, F., Marshall, C., Wintle, R. F., Shen, Y., ... Boycott, K. M. (2014). FORGE Canada Consortium: Outcomes of a 2-year national rare-disease gene-discovery project. *The American Journal of Human Genetics*, 94, 809–817. <https://doi.org/10.1016/j.ajhg.2014.05.003>
- Boycott, K. M., Hartley, T., Biesecker, L. G., Gibbs, R. A., Innes, A. M., Riess, O., Belmont, J., Dunwoodie, S. L., Jovic, N., Lassmann, T., Mackay, D., Temple, K., Visel, A., & Baynam, G. (2019). A diagnosis for all rare genetic diseases: The horizon and the next frontiers. *Cell*, 177, 32–37. <https://doi.org/10.1016/j.cell.2019.02.040>
- Boycott, K. M., Vanstone, M. R., Bulman, D. E., & MacKenzie, A. E. (2013). Rare-disease genetics in the era of next-generation sequencing: Discovery to translation. *Nature Reviews Genetics*, 14, 681–691. <https://doi.org/10.1038/nrg3555>
- Buske, O. J., Girdea, M., Dumitriu, S., Gallinger, B., Hartley, T., Trang, H., Misyura, A., Friedman, T., Beaulieu, C., Bone, W. P., Links, A. E., Washington, N. L., Haendel, M. A., Robinson, P. N., Boerkoel, C. F., Adams, D., Gahl, W. A., Boycott, K. M., & Brudno, M. (2015). PhenomeCentral: A portal for phenotypic and genotypic matchmaking of patients with rare genetic diseases. *Human Mutation*, 36, 931–940. <https://doi.org/10.1002/humu.22851>
- Chediak, L. (2020). Sharing sensitive health data in a federated data consortium model: An eight step guide. World Economic Forum. http://www3.weforum.org/docs/WEF_Sharing_Sensitive_Health_Data_2020.pdf
- Clark, M. M., Stark, Z., Farnaes, L., Tan, T. Y., White, S. M., Dimmock, D., & Kingsmore, S. F. (2018). Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genomic Medicine*, 3, 1–10. <https://doi.org/10.1038/s41525-018-0053-8>
- Dyke, S. O. M., Kirby, E., Shabani, M., Thorogood, A., Kato, K., & Knoppers, B. M. (2016). Registered access: A 'Triple-A' approach. *European Journal of Human Genetics*, 24, 1676–1680. <https://doi.org/10.1038/ejhg.2016.115>
- Dyke, S. O. M., Linden, M., Lappalainen, I., De Argila, J. R., Carey, K., Lloyd, D., Spalding, J. D., Cabili, M. N., Kerry, G., Foreman, J., Cutts, T., Shabani, M., Rodriguez, L. L., Haeussler, M., Walsh, B., Jiang, X., Wang, S., Perrett, D., Boughtwood, T., & Flicek, P. (2018). Registered access: Authorizing data access. *European Journal of Human Genetics*, 26, 1721–1731. <https://doi.org/10.1038/s41431-018-0219-y>
- Ferreira, C. R. (2019). The burden of rare diseases. *American Journal of Medical Genetics, Part A*, 179, 885–892. <https://doi.org/10.1002/ajmg.a.61124>
- Girdea, M., Dumitriu, S., Fiume, M., Bowdin, S., Boycott, K. M., Chénier, S., Chitayat, D., Faghfoury, H., Meyn, M. S., Ray, P. N., So, J., Stavropoulos, D. J., & Brudno, M. (2013). PhenoTips: Patient phenotyping software for clinical and research use. *Human Mutation*, 34, 1057–1065. <https://doi.org/10.1002/humu.22347>
- Hartley, T., Lemire, G., Kernohan, K. D., Howley, H. E., Adams, D. R., & Boycott, K. M. (2020). New diagnostic approaches for undiagnosed rare genetic diseases. *Annual Review of Genomics and Human Genetics*, 21, 351–372. <https://doi.org/10.1146/annurev-genom-083118-015345>
- Kernohan, K. D., Hartley, T., Alirezaie, N., Care4Rare Canada Consortium, Robinson, P. N., Dymont, D. A., & Boycott, K. M. (2018). Evaluation of exome filtering techniques for the analysis of clinically relevant genes. *Human Mutation*, 39, 197–201. <https://doi.org/10.1002/humu.23374>
- Knoppers, B. M. (2014). Framework for responsible sharing of genomic and health-related data. *The HUGO Journal*, 8, 1–6. <https://doi.org/10.1186/s11568-014-0003-1>
- Köhler, S., Doelken, S. C., Mungall, C. J., Bauer, S., Firth, H. V., Bailleul-Forestier, I., Black, G. C., Brown, D. L., Brudno, M., Campbell, J., FitzPatrick, D. R., Eppig, J. T., Jackson, A. P., Freson, K., Girdea, M., Helbig, I., Hurst, J. A., Jähn, J., Jackson, L. G., ... Robinson, P. N. (2014). The Human Phenotype Ontology project: Linking molecular biology and disease through phenotype data. *Nucleic Acids Research*, 42, D966–D974. <https://doi.org/10.1093/nar/gkt1026>
- Köster, J., & Rahmann, S. (2012). Snakemake—A scalable bioinformatics workflow engine. *Bioinformatics*, 28, 2520–2522. <https://doi.org/10.1093/bioinformatics/bts480>
- Lemire, G., Ito, Y. A., Marshall, A. E., Chrestian, N., Stanley, V., Brady, L., Tarnopolsky, M., Curry, C. J., Hartley, T., Mears, W., Derksen, A., Rioux, N., Laflamme, N., Hutchison, H. T., Pais, L. S., Zaki, M. S., Sultan, T., Dane, A. D., Gleeson, J. G., ... Boycott, K. M. (2021). ABHD16A deficiency causes a complicated form of hereditary spastic paraplegia associated with intellectual disability and cerebral anomalies. *The American Journal of Human Genetics. Advance online publication*, 108, 2017–2023. <https://doi.org/10.1016/j.ajhg.2021.09.005>
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., Flicek, P., & Cunningham, F. (2016). The ensembl variant effect predictor. *Genome Biology*, 17, 1–14. <https://doi.org/10.1186/s13059-016-0974-4>
- Nguengang Wakap, S., Lambert, D. M., Olry, A., Rodwell, C., Gueydan, C., Lanneau, V., Murphy, D., Le Cam, Y., & Rath, A. (2020). Estimating cumulative point prevalence of rare diseases: Analysis of the orphanet database. *European Journal of Human Genetics*, 28, 165–173. <https://doi.org/10.1038/s41431-019-0508-0>
- Osmond, M., Hartley, T., Dymont, D., Kernohan, K., Brudno, M., Buske, O., Innes, M., & Boycott, K. M. (2022). Outcome of 1500 matches through the Matchmaker Exchange for rare disease gene discovery:

- The two year experience of Care4Rare Canada. *Genetics in Medicine*, 24, 100–108. <https://doi.org/10.1016/j.gim.2021.08.014>
- Pedersen, B. S., Layer, R. M., & Quinlan, A. R. (2016). Vcfanno: Fast, flexible annotation of genetic variants. *Genome Biology*, 17, 1–9. <https://doi.org/10.1186/s13059-016-0973-5>
- Philippakis, A. A., Azzariti, D. R., Beltran, S., Brookes, A. J., Brownstein, C. A., Brudno, M., Brunner, H. G., Buske, O. J., Carey, K., Doll, C., Dumitriu, S., Dyke, S. O., denDunnen, J. T., Firth, H. V., Gibbs, R. A., Girdea, M., Gonzalez, M., Haendel, M. A., Hamosh, A., ... Rehm, H. L. (2015). The Matchmaker Exchange: A platform for rare disease gene discovery. *Human Mutation*, 36, 915–921. <https://doi.org/10.1002/humu.22858>
- Woolley, J. P., Kirby, E., Leslie, J., Jeanson, F., Cabili, M. N., Rushton, G., Hazard, J. G., Ladas, V., Veal, C. D., Gibson, S. J., Tassé, A. M., Dyke, S. O. M., Gaff, C., Thorogood, A., Knoppers, B. M., Wilbanks, J., & Brookes, A. J. (2018). Responsible sharing of biomedical data and biospecimens via the “Automatable Discovery and Access Matrix” (ADA-M). *NPJ Genomic Medicine*, 3, 1–6. <https://doi.org/10.1038/s41525-018-0057-4>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Driver, H. G., Hartley, T., Price, E. M., Turinsky, A. L., Buske, O. J., Osmond, M., Ramani, A. K., Kirby, E., Kernohan, K. D., Couse, M., Elrick, H., Lu, K., Mashouri, P., Mohan, A., So, D., Klamann, C., Le, H. G. B. H., Herscovich, A., Marshall, C. R., ... Boycott, K. M. (2022). Genomics4RD: An integrated platform to share Canadian deep-phenotype and multiomic data for international rare disease gene discovery. *Human Mutation*, 43, 800–811. <https://doi.org/10.1002/humu.24354>