





Variant-level matching for diagnosis and discovery: Challenges and opportunities

Eliete da S. Rodrigues¹ | Sean Griffith¹ | Renan Martin¹ | Corina Antonescu¹ | Jennifer E. Posey²  | Zeynep Coban-Akdemir² | Shalini N. Jhangiani^{2,3} | Kimberly F. Doheny¹ | James R. Lupski^{2,3,4,5}  | David Valle¹ | Michael J. Bamshad^{6,7} | Ada Hamosh¹  | Assaf Sheffer⁸ | Jessica X. Chong^{6,7}  | Yaron Einhorn⁸ | Miro Cupak⁹ | Nara Sobreira¹ 

¹McKusick-Nathans Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

²Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA

³Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA

⁴Department of Pediatrics, Baylor College of Medicine, Houston, Texas, USA

⁵Texas Children's Hospital, Houston, Texas, USA

⁶Division of Genetic Medicine, Department of Pediatrics, University of Washington, Seattle, Washington, USA

⁷Brotman Baty Institute for Precision Medicine, Seattle, Washington, USA

⁸Franklin by Genoox, Palo Alto, California, USA

⁹DNASTack, Toronto, Ontario, Canada

Correspondence

Nara Sobreira, McKusick-Nathans Department of Genetic Medicine, Johns Hopkins University School of Medicine, 733 North Broadway St., Suite 569, Baltimore, MD 21205, USA.
Email: nsobrei2@jhmi.edu

Funding information

NHGRI/NHLBI, Grant/Award Number: UM1 HG006542; National Human Genome Research Institute/National Heart, Lung, and Blood Institute, Grant/Award Number: UM1 HG006542

Abstract

Here we describe MyGene2, Geno2MP, VariantMatcher, and Franklin; databases that provide variant-level information and phenotypic features to researchers, clinicians, healthcare providers and patients. Following the footsteps of the Matchmaker Exchange project that connects exome, genome, and phenotype databases at the gene level, these databases have as one goal to facilitate connection to one another using Data Connect, a standard for discovery and search of biomedical data from the Global Alliance for Genomics and Health (GA4GH).

KEYWORDS

Data Connect, Franklin, Geno2MP, Matching Tools, MyGene2, variant-level, VariantMatcher

1 | INTRODUCTION

Since 2009, whole-exome sequencing (WES) and whole-genome sequencing (WGS) have become the most prominent genomic tools for the discovery of novel disease genes and variants related to rare Mendelian phenotypes (Chong et al., 2015). Using this approach,

progress has accelerated so that the number of genes with known phenotype-causing variants has expanded from 2346 in 2009 to 4588 currently or ~22% of the total protein-encoding genes in the genome (OMIM, 2021). That leaves nearly 80% of the predicted ~20,000 protein-encoding genes yet to be connected to a disease phenotype. Similarly, 50%–75% of clinical and research WES do not

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

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

identify a responsible variant(s) even in families that present Mendelian segregation of human disease traits (Baxter et al., 2022; Chong et al., 2015; Posey et al., 2019; Retterer et al., 2016; Yang et al., 2014).

Possible explanations for the modest molecular diagnostic rate include unappreciated phenotypic and genetic heterogeneity; causative variants is not yet recognized disease genes (Liu et al., 2019); high locus heterogeneity; complex molecular mechanisms underlying incomplete penetrance; technical limitations in the applied sequencing approach; and limitations in the variant analyses and classification. One particular limitation, that we focus on here, is the lack of accurate analytical tools to interpret and classify variants in known or novel disease genes.

Variant classification in the research or clinical setting is a complex process that takes into consideration many different features related to the individual, the phenotype, the variant, the population, the gene and the environment. In 2015, Richards et al. (2015), published a guideline for variant interpretation and classification based on criteria using typical types of variant evidence (e.g., population data, computational data, functional data, segregation data, etc.). To apply these criteria, research and clinical laboratories use many different databases with different types of information and evidence, but very few them allow the laboratories to have access to detailed phenotypic information related to the specific variants being investigated. Knowing the phenotypic features of other individuals that carry the variant of interest can be a critical step in variant classification, but detailed phenotypic information linked to putatively-causal variants is rarely available in public or even controlled-access databases because of the difficulty in obtaining detailed phenotype data, rarity of the candidate variants, and challenges and uncertainty due to potential regulatory requirements to maintain the confidentiality and privacy of individuals who carry these rare variants.

Here we describe several databases that have made variant-level information together with phenotype or phenotypic features available to researchers, clinicians, healthcare providers and patients; and, their goal to facilitate patient care by connecting to each other following the footsteps of the Matchmaker Exchange project that connects gene-level databases (Azzariti & Hamosh, 2020; Sobreira et al., 2017).

2 | METHODOLOGY

2.1 | MyGene2 and Geno2MP

MyGene2 (<https://mygene2.org/>) (Figure 1) is a web-based platform that enables families, clinicians, and researchers to share genetic information such as candidate genes/variants and phenotypic data publicly with the goal of facilitating rare disease research such as novel disease gene discovery, allelic series, and genotype/phenotype relationship studies. Families can share their own identifiable health and genetic information and clinicians and researchers can use

MyGene2 to share deidentified health and candidate gene/variant information on behalf of families with Mendelian conditions. Information about each family is organized into a "Profile." All profiles are publicly searchable, and all stakeholders have the same level of access to the data shared through the site. To help families share their data and make matches with others interested in the same candidate genes or Mendelian conditions, MyGene2 uses concept recognition software to automatically extract structured phenotype terms from narratives submitted by families and variant validation software to assist families in submitting their candidate variants to the site. MyGene2 currently hosts ~5000 family profiles that share candidate gene/variant and reported clinical findings and shares all variants through Beacon Network (<https://beacon-network.org/>), a search engine across the global network of Beacons enabling discovery of genetic variants of interest around the world (Fiume et al., 2019). The connection to the Beacon Network is enabled by the Beacon protocol, a standard for discovery of genetic variants, developed by the Global Alliance for Genomics and Health (GA4GH) (The Global Alliance for Genomics and Health, 2016). Beacon allows users, including anonymous users, to query a database for the presence of a variant of interest by asking a question of the form "Do you have information about the following variant?," to which a Beacon responds either "yes" or "no," optionally with additional aggregate information. Beacon Network users can visit the MyGene2 profiles for any variant identified through a search of the Network.

Geno2MP (<https://geno2mp.gs.washington.edu/>) (Figure 2) is a web-accessible, searchable database containing rare variant genotypes linked to phenotypic information developed by the University of Washington Center for Mendelian Genomics to publicly share all rare (<2.5% alternate allele frequency in gnomAD v2.1 or v3.0) sequence variants identified in individuals affected by suspected Mendelian conditions and/or in their putatively unaffected relatives alongside deidentified phenotype data describing the affected status and original phenotype of interest for which each family was ascertained. As of October 1, 2021, Geno2MP shares variant zygosity and phenotype data from 19,344 individuals sequenced by one of the University of Washington, Broad, or Yale Centers for Mendelian Genomics. Geno2MP can be searched by gene, chromosomal coordinates, or HPO term, and a bulk sites-only VCF is available for download. The Geno2MP data set will eventually be integrated into MyGene2 to make its data available for matchmaking/querying via the future variant-level Matchmaker Exchange and the Beacon Network. This will also enable users to sign up to be notified about future variant matches.

2.2 | VariantMatcher

VariantMatcher (<https://variantmatcher.org/>) (Figure 3) was developed to connect individuals (researchers and healthcare providers) around the globe with interest in a specific variant. It enables sharing of variant-level and phenotypic data from participants in research projects for discovery of novel disease genes including the

784 | WILEY-Human Mutation | RODRIGUES ET AL.

Home | My Dashboard | Manage Account | Logout | Contact | FAQ | Donate

Genes | Families | Care Providers | Researchers | News & Events | About

MyGene²

Search by gene (e.g., KDM1A), Family ID, or clinical finding

Family Profiles

4483

Empower discovery by connecting and sharing data

Share, Unite, and Discover

Watch video

Haven't yet created a family profile?

Get started

Spotlight Family Profile

FIGURE 1 MyGene2 homepage

Baylor-Hopkins Center for Mendelian Genomics project. VariantMatcher (Wohler et al., 2021) contains the rare (MAF < 1% in gnomAD), coding (including missense, nonsense, stop-loss, and splice site variants; synonymous are excluded), single nucleotide variants identified in 6235 VCF files (896,847 unique variants) of affected and unaffected individuals sequenced as part of multiple projects and their detailed phenotypic information.

To comply with patient privacy and security regulations, users of the site must register and be approved by site administrators after verification that the user is a clinician, researcher or healthcare provider. Users may upload up to 10 genomic coordinates per day to the site and are notified of any match. The query format is "chr:coordinate refAllele > altAllele" (e.g., chr2:1234567G>T) and is available for genomic builds hg18, hg19, or hGRCh38. The current VariantMatcher data is on the hg19 build of the human reference assembly. Queries that use hg18 or hg38 are lifted over to hg19 before the match occurs. Phenotypic features can also be added, but the match is based only on the genomic location. If the user adds phenotype information (minimum of three features and maximum of six), however, and there is a match based on genomic location, the

phenotype information from the matched entries is shared in the email notifying the users of the match. When there is a match, both parties are notified by simultaneous emails so that they can choose to exchange additional information about their cases. If a match is not made, the queried coordinates can be stored for future matching.

VariantMatcher recently added support for Beacon and through the Beacon protocol, VariantMatcher was connected to the Beacon Network (Fiume et al., 2019). If a user is informed that the variant queried through the Beacon Network is present in VariantMatcher, they have the option to create an account in VariantMatcher to obtain further information about the variant of interest. As of October 1, 2021, VariantMatcher had 695 submitters from 44 countries. 4406 variants had been queried and 153 variants matched to 1248 individuals in VariantMatcher.

To further develop VariantMatcher's capabilities to support variant classification and facilitate discovery of disease-causing variants, additional query capabilities are planned, including: (1) indels; (2) variants by zygosity state; (3) specific variant with feature (s); and (4) specific group of variants/gene (e.g., individuals with nonsense variants in gene X).

Click here to contribute data to Geno₂MP

HPO Browser
Need to find the HPO term for a clinical finding?

Search for a gene, chromosomal position, region or HPO term

Search

Gene (MYH3); chrom. position of variant (17:10534960); chrom. region of variants (17:10534960-10535460); dbSNP rsID (rs34393601)

HPO term (oral cleft); HPO number (0000202)

About the Genotype to Mendelian Phenotype (Geno₂MP v2.6) Browser

Geno₂MP is a web-based query tool that searches a database of rare variants from exome sequencing data linked to phenotypic information from a wide variety of Mendelian gene discovery projects. Specifically, each rare genotype is linked to individual-level phenotypic profiles defined by human phenotype ontology (HPO) terms. Thus, it enables users to link “Genotypes to Mendelian Phenotypes” to facilitate new gene discovery efforts.

Currently, Geno₂MP contains data from ~19,344 individuals (updated September 2021), including both persons affected with a Mendelian condition and unaffected individuals who are relatives of persons with a Mendelian condition. Geno₂MP shows phenotypic profiles for affected individuals and, for unaffected individuals, the phenotypic profile of their affected relative(s). As with most genetic variation in humans, most of the variants in Geno₂MP are not causal for a Mendelian condition.

This database is a collaborative, shared resource for the human genetics community. In fact, the power of Geno₂MP will only grow as thousands more samples and phenotypes are added, and we invite the entire Mendelian genetics community to join us in contributing to this resource.

A sites-only vcf is available [for download!](#)

FIGURE 2 Geno2MP homepage

2.3 | Franklin

Franklin by Genoox (<https://franklin.genoox.com>) (Genoox, 2021) (Figure 4) connects clinicians, genetic counselors, and healthcare organizations, enabling its users to make impactful discoveries using the most advanced genomic tools and applications. The power of the Franklin community provides actionable insights from the largest real-time real-life genomic database serving professionals at the point of care.

Franklin is built on top of an advanced AI-based interpretation engine and provides an automated workflow from raw sequencing data (FASTQ/VCF) to a shortlist of candidate variants for a final clinical report. The interpretation engine supports multiple genetic applications including rare diseases, oncology, hereditary cancer, and carrier screening. The engine provides in-depth variant evidence, literature and text-mining evidence, automated American College of Medical Genetics and Genomics -based classification for single nucleotide variants and copy number variations, and a wealth of annotations and assessment tools.

Franklin community members can share their own evidence and insights regarding a variant or a gene with the rest of the community, as well as start a discussion on a case/variant level, to reach a consensus. In addition, members are able to contact other members in an anonymous way to inquire about a specific variant of interest that surfaced during their analysis. This feature assists in resolving uncertainties and providing more accurate classification.

Franklin is used today by over 1700 healthcare organizations and is available at no charge. The platform is widely used in hospitals, laboratories, and medical facilities throughout 44 countries across the globe.

3 | DISCUSSION

During the last decade, multiple public databases such as Exome Variant Server (<https://evs.gs.washington.edu/EVS/>) (ExomeVariantServer), 1000 Genome Project (<https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/>) (Genomes Project et al., 2015), ExAC (gs://gnomad-public/legacy) (Lek et al., 2016), gnomAD (<https://gnomad.broadinstitute.org/>) (Karczewski et al., 2020), and ABraOM (<https://abraom.ib.usp.br/search.php>) (Naslavsky et al., 2017) have facilitated the identification of rare pathogenic variants in the represented populations by publicly sharing the genomic data of mostly healthy individuals in an accessible way. However, these databases are deidentified and do not harbor phenotypic information of included individuals. Therefore, these databases are not optimized for the interpretation of rare variants possibly associated with rare phenotypes, particularly those characterized by mild presentation, incomplete penetrance, mosaicism, and/or late-onset.

To overcome this potential limitation, databases such as DECIPHER (Firth et al., 2009), MyGene2/Geno2MP (Chong et al., 2016), Variant-Matcher (Wohler et al., 2021), and Franklin (Genoox) have created a public way to share genomic and phenotypic data from individuals with

VariantMatcher Home About Contact Us Statistics Editor Help My Account Log Out

Search Saved Searches

Search

Variants : 1:100100 A>T ... NCBI36 GRCh37 GRCh38

The variant search format is as follows:
chr:coord refAllele>altAllele

For example :
1:100100 A>T

Allele can be one of ATCG, both reference and alternate alleles are required.

You may use ? as a wildcard to represent any alleles, for example.
1:100100 A>?

You can search up to 10 variants per day.

Features :

Features entered here will be shared with submitters if there is a match, and submission features will be shared with you. This is optional, at least three need to be entered to be shared, up to six can be specified.

Save Search : Save this search to be automatically re-run monthly

Search

FIGURE 3 VariantMatcher search page

rare phenotypes that is easily accessible to researchers, clinicians, healthcare providers, and patients. While they are each queried in a slightly different way, they all harbor accessible genomic and phenotypic information from individuals with rare phenotypes. The use of these databases has supported the identification of novel disease-causing variants and the more precise classification of many variants of uncertain significance (VUS). To date, most of the VUSs investigated in databases such as MyGene2/Geno2MP, VariantMatcher, and Franklin could be classified as benign after close comparison of the phenotypes facilitating identification of stronger candidate causative variants for the phenotypes being investigated (Wohler et al., 2021).

We plan to follow the successful Matchmaker Exchange (MME) model to connect these databases and others in a federated network using the GA4GH Data Connect standard (<https://github.com/ga4gh-discovery/data-connect/>).

4 | DATA CONNECT

GA4GH provides a suite of technologies for supporting federated networks, with two complementary standards for discovery or search of networked data—Data Connect and Beacon. While Beacon lacks the flexibility for searching beyond position-based lookups of

variants, its extended version 2 is under development. The current version 2 draft proposes a data model for eight commonly represented concepts, such as variants and sequencing analyses, and introduces a new query language allowing for simple filtering of instances of these concepts shared through a Beacon.

Data Connect is a GA4GH approved standard for discovery and search of biomedical data. It provides data custodians with a mechanism to organize and semantically describe their data and its data model, and data consumers with a mechanism to construct flexible queries and search the described data. Unlike other data-sharing technologies, Data Connect does not prescribe a data model, thus allowing arbitrary data to be discovered and searched “as is,” without potentially expensive transformations. It relies on the JSON Schema standard (<https://json-schema.org/>) for describing data models, and the SQL standard for querying.

Contrasted with a specialized application programming interface (API) like Beacon, Data Connect is a more general framework, providing a suitable foundation for building Beacons, Matchmakers, and other networked applications. The flexibility of its query language allows the construction of queries combining different types of data within and across databases, as well as queries leveraging custom functions for matching. The model-agnostic nature of the standard allows one to connect the databases without having

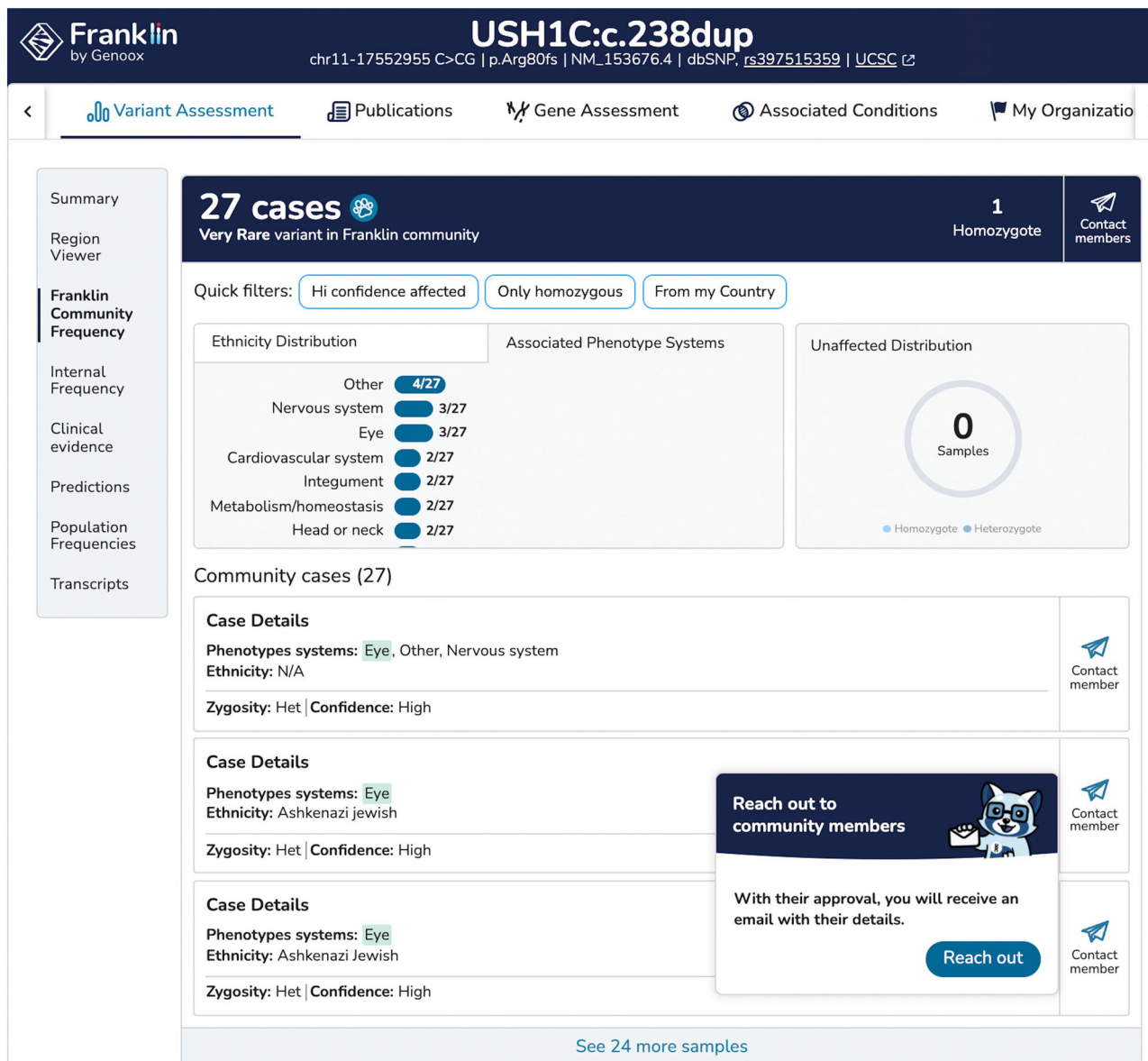


FIGURE 4 Franklin by Genoox result page

to transform their data into a shared, agreed-upon data model, with the ability to describe the data semantically and harmonize gradually over time. Data Connect's choice of SQL as the query language enables one to rely on a number of existing tools, and integrates well with these databases. Combined with a growing ecosystem of tools compatible with Data Connect, this network can be implemented with minimal effort.

Databases are connected via the Data Connect API that consists of three parts:

- 1) Table API, through which each database describes its data models to enable data discovery and retrieval;
- 2) Service Info API for discovery of metadata about the database, and;
- 3) Search API which allows other databases to search the database for similar variants using rich and flexible queries.

The algorithm that decides similarity is defined by the database being queried. The database evaluates the query, applies the matching function, and replies with a list of other similar cases it hosts.

We plan to establish a peer-to-peer federated network based on Data Connect to facilitate data sharing, identification of individuals harboring the same variant, and exchange of phenotypic information, making variant classification more specific. We will rely on GA4GH Passport standard (Voisin et al., 2021) to facilitate access in the network in a way that allows each database to control access to its data. To enable this, each participating database will be set up to issue Passport visas for the users it has approved according to its policies, and make decisions on access to its data based on Passport visas issued by other connected databases.

TABLE 1 Databases' metrics as of October 1st 2021

	Number of variants	Number of Genes Corresponding to the Variants	Number of Individuals Corresponding to the Variants	Number of Submitters	Number of Countries Corresponding to the Submitters
MyGene2	2973	622	4041	1130	40
Geno2MP	38,026,951	>20,000	19,344	>300	>55
VariantMatcher	896,847	20,484	6235	695	44
Franklin	297,278,579	NA	> 150,000	>1700 organizations	44

Users will be able to choose the most appropriate database to share their data and easily query other connected databases for similar cases. When a match occurs among connected databases, the users will automatically and simultaneously receive an email notification informing them of the presence or absence of a match in the queried database(s). A matching email contains the matching data (genomic ± phenotypic features), contact information of the users to whom they matched, and additional metadata that will only be shared at the discretion of the databases that harbor the matching cases. Subsequently, the matched users can choose to contact each other to exchange further information about their cases including detailed phenotypic information.

If there is no match in any of the queried databases, the submission will be stored in the database from which the query originated, not the external databases queried. In the future, if the users would like to repeat the query, they would need to send the submission again. In some databases, such as VariantMatcher, the users have the option to automatically resend the data from their submissions to the other chosen databases on a periodic basis.

Variant information in the format of genomic location is the minimal requirement to initiate a query among the connected nodes. Matching on variant features such as zygosity or phenotypic features in addition to the required variant will also be supported by some of the databases such as VariantMatcher. However, even if some of the databases match only on the variant information, we expect that the users querying the databases through the Data Connect API will also submit zygosity information in addition to detailed phenotypic information so this information can be shared in the email notification which will facilitate further communication among the users who matched. Incomplete penetrance, variable expressivity of the phenotype, age of onset, and zygosity are some of the factors that should be considered when the variants and phenotypes are being compared before the final classification of a candidate variant. To enhance the likelihood of a match on pathogenic variants, databases such as VariantMatcher and Geno2MP harbor only rare coding variants.

We will follow the recommendations of the Consent Task Team from the GA4GH Regulatory and Ethics Working Group and require individual written informed consent, since variant-level data and/or phenotypic data will be provided. If the individual previously consented to data being shared in an open or registered access database whose declared purpose involves data sharing for purposes

consistent with those of this matching network, no additional consent is required. Each database will be responsible for ensuring that the patient is consented appropriately.

Finally, like the MME project, we will follow the federated network model in which multiple distributed databases are connected through APIs and each database will be autonomous with respect to its own data schema, will have ongoing control of its own data, and will be required to manage the security and privacy of the data they harbor and attest to database security requirements as defined by the GA4GH Security Working Group (<https://www.ga4gh.org/how-we-work/2020-2021-roadmap/2020-2021-roadmap-part-ii/data-security-2020-2021-roadmap/>).

By connecting variant-level databases that also facilitate phenotypic data access, we expect to improve the variant classification process in both research and clinical settings and also to increase the discovery rate of novel disease-causing variants by increasing the specificity of matches (Table 1).

ACKNOWLEDGMENT

This article was funded by the National Human Genome Research Institute/National Heart, Lung, and Blood Institute (NHGRI/NHLBI UM1 HG006542).

CONFLICTS OF INTEREST

Michael J. Bamshad is the Chair of the Scientific Advisory Committee for GeneDx. James R. Lupski has stock ownership in 23andMe, is a paid consultant for Regeneron Genetics Center, and is a coinventor on multiple United States and European patents related to molecular diagnostics for inherited neuropathies, eye diseases, genomic disorders and bacterial genomic fingerprinting. The Department of Molecular and Human Genetics at Baylor College of Medicine receives revenue from clinical genetic testing conducted at Baylor Genetics (BG); J.R.L. serves on the Scientific Advisory Board (SAB) of BG. The other authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable—no new data generated.

WEB RESOURCES

<https://evs.gs.washington.edu/EVS/>—Exome Variant Server

<https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/> - 1000 Genome Project
<gs://gnomad-public/legacy> - ExAC
<https://gnomad.broadinstitute.org/> - gnomAD
<https://abramo.ib.usp.br/search.php-ABraOM>
<https://www.ga4gh.org/how-we-work/2020-2021-roadmap/2020-2021-roadmap-part-ii/data-security-2020-2021-roadmap/> - GA4GH Security Working Group
<https://github.com/ga4gh-discovery/data-connect/-GA4GH>
 Data Connect standard
<https://json-schema.org/> - JSON Schema standard
<https://mygene2.org/> - MyGene2
<https://beacon-network.org/> - Beacon Network
<https://geno2mp.gs.washington.edu/> - Geno2MP
<https://variantmatcher.org/> - VariantMatcher
<https://franklin.genoox.com/> - Franklin by Genoox

ORCID

Jennifer E. Posey  <http://orcid.org/0000-0003-4814-6765>
 James R. Lupski  <http://orcid.org/0000-0001-9907-9246>
 Ada Hamosh  <http://orcid.org/0000-0002-1780-5230>
 Jessica X. Chong  <http://orcid.org/0000-0002-1616-2448>
 Nara Sobreira  <http://orcid.org/0000-0002-5228-5613>

REFERENCES

- 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>
- Baxter, S. M., Posey, J. E., Lake, N. J., Sobreira, N., Chong, J. X., Buyske, S., Blue, E. E., Chadwick, L. H., Coban-Akdemir, Z. H., Doheny, K. F., Davis, C. P., Lek, M., Wellington, C., Jhangiani, S. N., Gerstein, M., Gibbs, R. A., Lifton, R. P., MacArthur, D. G., Matise, T. C., & O'Donnell-Luria, A. (2022). Centers for Mendelian genomics: A decade of facilitating gene discovery. *Genetics in Medicine*. Published online February 8, 2022. <https://doi.org/10.1016/j.gim.2021.12.005>
- Chong, J. X., Buckingham, K. J., Jhangiani, S. N., Boehm, C., Sobreira, N., Smith, J. D., Harrell, T. M., McMillin, M. J., Wiszniewski, W., Gambin, T., Coban Akdemir, Z. H., Doheny, K., Scott, A. F., Avramopoulos, D., Chakravarti, A., Hoover-Fong, J., Mathews, D., Witmer, P. D., Ling, H., ... Bamshad, M. J. (2015). The genetic basis of mendelian phenotypes: Discoveries, challenges, and opportunities. *American Journal of Human Genetics*, 97(2), 199–215. <https://doi.org/10.1016/j.ajhg.2015.06.009>
- Chong, J. X., Yu, J. H., Lorentzen, P., Park, K. M., Jamal, S. M., Tabor, H. K., Rauch, A., Saenz, M. S., Boltshauser, E., Patterson, K. E., Nickerson, D. A., & Bamshad, M. J. (2016). Gene discovery for Mendelian conditions via social networking: De novo variants in KDM1A cause developmental delay and distinctive facial features. *Genetics in Medicine*, 18(8), 788–795. <https://doi.org/10.1038/gim.2015.161>
- ExomeVariantServer. NHLBI GO Exome Sequencing Project (ESP). Retrieved 4 November 2021 from <http://evs.gs.washington.edu/EVS/>
- Firth, H. V., Richards, S. M., Bevan, A. P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R. M., & Carter, N. P. (2009). DECIPHER: Database of chromosomal imbalance and phenotype in humans using ensembl resources. *American Journal of Human Genetics*, 84(4), 524–533. <https://doi.org/10.1016/j.ajhg.2009.03.010>
- Fiume, M., Cupak, M., Keenan, S., Rambla, J., de la Torre, S., Dyke, S. O. M., Brookes, A. J., Carey, K., Lloyd, D., Goodhand, P., Haeussler, M., Baudis, M., Stockinger, H., Dolman, L., Lappalainen, I., Tornroos, J., Linden, M., Spalding, J. D., Ur-Rehman, S., ... Scollen, S. (2019). Federated discovery and sharing of genomic data using Beacons. *Nature Biotechnology*, 37(3), 220–224. <https://doi.org/10.1038/s41587-019-0046-x>
- Genomes Project, C., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., & Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74. <https://doi.org/10.1038/nature15393>
- Genoox. Franklin by Genoox. Retrieved 4 November 2021 from <https://franklin.genoox.com>
- Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alfoldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., ... MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809), 434–443. <https://doi.org/10.1038/s41586-020-2308-7>
- Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., Tukiainen, T., Birnbaum, D. P., Kosmicki, J. A., Duncan, L. E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., ... Exome Aggregation Consortium. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536(7616), 285–291. <https://doi.org/10.1038/nature19057>
- Liu, P., Meng, L., Normand, E. A., Xia, F., Song, X., Ghazi, A., Rosenfeld, J., Magoulas, P. L., Braxton, A., Ward, P., Dai, H., Yuan, B., Bi, W., Xiao, R., Wang, X., Chiang, T., Vetrini, F., He, W., Cheng, H., ... Yang, Y. (2019). Reanalysis of clinical exome sequencing data. *New England Journal of Medicine*, 380(25), 2478–2480. <https://doi.org/10.1056/NEJMc1812033>
- Naslavsky, M. S., Yamamoto, G. L., de Almeida, T. F., Ezquina, S. A. M., Sunaga, D. Y., Pho, N., Bozoklian, D., Sandberg, T. O. M., Brito, L. A., Lazar, M., Bernardo, D. V., Amaro, E., Jr., Duarte, Y. A. O., Lebrao, M. L., Passos-Bueno, M. R., & Zatz, M. (2017). Exomic variants of an elderly cohort of Brazilians in the ABraOM database. *Human Mutation*, 38(7), 751–763. <https://doi.org/10.1002/humu.23220>
- Online Mendelian Inheritance in Man (OMIM). (2021). McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University. Retrieved 4 Nov 2021 from <https://www.omim.org/>
- Posey, J. E., O'Donnell-Luria, A. H., Chong, J. X., Harel, T., Jhangiani, S. N., Coban Akdemir, Z. H., Buyske, S., Pehlivan, D., Carvalho, C. M. B., Baxter, S., Sobreira, N., Liu, P., Wu, N., Rosenfeld, J. A., Kumar, S., Avramopoulos, D., White, J. J., Doheny, K. F., Witmer, P. D., ... Centers for Mendelian Genomics. (2019). Insights into genetics, human biology and disease gleaned from family-based genomic studies. *Genetics in Medicine*, 21(4), 798–812. <https://doi.org/10.1038/s41436-018-0408-7>
- Retterer, K., Juusola, J., Cho, M. T., Vitazka, P., Millan, F., Gibellini, F., Vertino-Bell, A., Smaoui, N., Neidich, J., Monaghan, K. G., McKnight, D., Bai, R., Suchy, S., Friedman, B., Tahiliani, J., Pineda-Alvarez, D., Richard, G., Brandt, T., Haverfield, E., ... Bale, S. (2016). Clinical application of whole-exome sequencing across clinical indications. *Genetics in Medicine*, 18(7), 696–704. <https://doi.org/10.1038/gim.2015.148>
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H. L., & Laboratory Quality Assurance Committee. (2015). Standards and guidelines for the interpretation of sequence variants:

- A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5), 405–424. <https://doi.org/10.1038/gim.2015.30>
- Sobreira, N. L. M., Arachchi, H., Buske, O. J., Chong, J. X., Hutton, B., Foreman, J., Schiettecatte, F., Groza, T., Jacobsen, J. O. B., Haendel, M. A., Boycott, K. M., Hamosh, A., Rehm, H. L., & Matchmaker Exchange Consortium. (2017). Matchmaker exchange. *Current Protocols*, 95, 9.31.1–9.31.15. <https://doi.org/10.1002/cphg.50>
- The Global Alliance for Genomics and Health. (2016). GENOMICS. A federated ecosystem for sharing genomic, clinical data. *Science*, 352(6291), 1278–1280. <https://doi.org/10.1126/science.aaf6162>
- Voisin, C., Linden, M., Dyke, S. O. M., Bowers, S. R., Alper, P., Barkley, M. P., Bernick, D., Chao, J., Courtot, M., Jeanson, F., Konopko, M. A., Kuba, M., Lawson, J., Leinonen, J., Li, S., Ota Wang, V., Philippakis, A. A., Reinold, K., Rushton, G. A., ... Nyronen, T. H. (2021). GA4GH passport standard for digital identity and access permissions, *Cell Genomics*, 1(2) <https://doi.org/10.1016/j.xgen.2021.100030>
- Wohler, E., Martin, R., Griffith, S., Rodrigues, E. D. S., Antonescu, C., Posey, J. E., Coban-Akdemir, Z., Jhangiani, S. N., Doheny, K. F., Lupski, J. R., Valle, D., Hamosh, A., & Sobreira, N. (2021). PhenoDB, GeneMatcher and VariantMatcher, tools for analysis and sharing of sequence data. *Orphanet Journal of Rare Diseases*, 16(1), 365. <https://doi.org/10.1186/s13023-021-01916-z>
- Yang, Y., Muzny, D. M., Xia, F., Niu, Z., Person, R., Ding, Y., Ward, P., Braxton, A., Wang, M., Buhay, C., Veeraraghavan, N., Hawes, A., Chiang, T., Leduc, M., Beuten, J., Zhang, J., He, W., Scull, J., Willis, A., ... Eng, C. M. (2014). Molecular findings among patients referred for clinical whole-exome sequencing. *Journal of the American Medical Association*, 312(18), 1870–1879. <https://doi.org/10.1001/jama.2014.14601>

How to cite this article: Rodrigues, E. d. S., Griffith, S., Martin, R., Antonescu, C., Posey, J. E., Coban-Akdemir, Z., Jhangiani, S. N., Doheny, K. F., Lupski, J. R., Valle, D., Bamshad, M. J., Hamosh, A., Sheffer, A., Chong, J. X., Einhorn, Y., Cupak, M., & Sobreira, N. (2022). Variant-level matching for diagnosis and discovery: Challenges and opportunities. *Human Mutation*, 43, 782–790. <https://doi.org/10.1002/humu.24359>