

Variant-level Matching Tools

Eliete Rodrigues¹, Sean Griffith¹, Renan Martin¹, Corina Antonescu¹, Jennifer Posey², Zeynep Coban-Akdemir², Shalini Jhangiani², Kim Doheny³, James R. Lupski², David Valle⁴, Michael Bamshad⁵, Ada Hamosh⁶, Assaf Sheffer⁷, Jessica Chong⁵, Yaron Einhorn⁷, Miro Cupak⁸, and Nara Sobreira¹

¹Johns Hopkins University School of Medicine

²Baylor College of Medicine

³Johns Hopkins School of Medicine

⁴Johns Hopkins

⁵University of Washington

⁶Johns Hopkins University

⁷Franklin by Genoox

⁸DNASTack

November 9, 2021

Abstract

Here we describe MyGene2, Geno2MP, VariantMatcher, and Franklin; databases that have made variant-level information together with phenotype or phenotypic features available to researchers, clinicians, health care providers and patients. Following in the footsteps of the Matchmaker Exchange project that connects exome, genome, and phenotype databases at the gene level, these databases plan to connect to each other using Data Connect, a standard for discovery and search of biomedical data from the Global Alliance for Genomics and Health (GA4GH).

Introduction

Since 2009, whole exome sequencing (WES) and whole genome sequencing (WGS) have become the main 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 2,346 in 2009 to 4,532 currently or ~22% of the total protein-encoding genes in the genome (OMIM). 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 identify a responsible variant(s) even in families that present Mendelian segregation of human disease traits (Chong et al. 2015; Posey et al. 2019; Retterer et al. 2016; Yang et al. 2014).

Possible explanations for the modest diagnostic rate include: unappreciated phenotypic and genetic heterogeneity; causative variants in 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 analysis 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 gene and the environment.

In 2015, Richards and colleagues (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 evidence, but very few of 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 is 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, health care providers and patients; and, their plan to connect to each other following in the footsteps of the Matchmaker Exchange project that connects gene-level databases (Sobreira et al. 2017).

MyGene2 and Geno2MP

MyGene2 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 and genotype/phenotype relationship studies. Families can share their own identifiable health and genetic information and clinicians and researchers can use MyGene2 to share de-identified 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, a search engine across the global network of Beacons enabling discovery of genetic variants of interest around the world (Fiume et al. 2019). Beacon Network users can visit the MyGene2 profiles for any variant identified through a search of the Network.

Geno2MP 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 de-identified 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 dataset will eventually be integrated into MyGene2 in order 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.

VariantMatcher

VariantMatcher (variantmatcher.org) was developed to connect individuals (researchers and health care 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 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 6,235 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 health care providers. 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, a standard for discovery of genetic variants, developed by the Global Alliance for Genomics and Health (GA4GH) (Global Alliance for 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. Through the Beacon protocol, VariantMatcher was connected to the Beacon Network (<https://beacon-network.org/>) (Fiume et al. 2019). If a user finds out 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 1 October 2021, VariantMatcher had 695 submitters from 44 countries. 4,406 variants had been queried and 153 variants matched to 1,248 individuals in VariantMatcher.

To further develop VariantMatcher’s capabilities to support variant classification and facilitate discovery of disease-causing variants, we plan to implement additional query capabilities, 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).

Franklin

Franklin by Genoox (<https://franklin.genoox.com>) (Genoox) 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 short list 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 ACMG-based classification for SNPs and CNVs, 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 1,700 health organizations and is available at no charge. The platform is widely used in hospitals, laboratories, and medical facilities throughout 44 markets across the globe.

Data Connect

Data Connect is a standard for discovery and search of biomedical data from the GA4GH (<https://github.com/ga4gh-discovery/data-connect/blob/master/SPEC.md>). 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.

Through Data Connect, databases harboring variant-level data with or without phenotypic feature data will be able to connect in a federated network, answer more complex questions, and communicate while preserving their respective data models.

Databases can connect in the network by implementing the Data Connect application programming interface (API). The API consists of three parts:

- 1) Table API, through which each database describes its data models to enable their discovery as well as fetching of associated data;
- 2) Service Info API for discovery of metadata about the database, and;
- 3) Search API allowing 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, where each database connects to one or more databases within the network. Because of the sensitivity of the information being shared, most databases will require requests from other databases to be authenticated with a pre-shared key (PSK). These keys are usually shared via encrypted email messages. This process of connecting to other databases can be time-consuming, but it assures each database full control over who it shares data with.

Discussion

In the last 10 years, 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 (<https://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 de-identified 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, and/or late onset.

To overcome this limitation, databases such DECIPHER (Firth et al. 2009), MyGene2/Geno2MP (Chong et al. 2016), VariantMatcher (Wohler et al. 2021) and Franklin (Genoox) have created a public way to share genomic and phenotypic data from individuals with rare phenotypes that is easily accessible to researchers, clinicians, health care providers, and patients. While they are each queried in a slightly different way, they all harbor accessible genomic and phenotypic information from patients 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. To date, most of the VUSs investigated in databases such 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. This will facilitate data sharing, the identification of individuals harboring the same variant, and the exchange of phenotypic information making variant classification more specific. 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 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 only 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 start 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 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. To enhance the likelihood of a match on pathogenic variant databases such VariantMatcher and Geno2MP only harbor rare coding variants.

We will follow the recommendations of the Consent Task Team from the GA4GH Regulatory and Ethics Working Group Since and individual written informed consent will typically be required since variant-level data and/or phenotypic data will be provided. Each database will manage the security and privacy of the data they harbor.

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

Conflict of Interest

Michael J. Bamshad is the Chair of the Scientific Advisory Committee for GeneDx.

The other authors have no conflict of interest to declare.

Data Availability

Data sharing not applicable – no new data generated.

References

Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, Harrell TM, McMillin MJ, Wiszniewski W, Gambin T, Coban Akdemir ZH, Doheny K, Scott AF, Avramopoulos D, Chakravarti A, Hoover-Fong J, Mathews D, Witmer PD, Ling H, Hetrick K, Watkins L, Patterson KE, Reinier F, Blue E, Muzny D, Kircher M, Bilguvar K, Lopez-Giraldez F, Sutton VR, Tabor HK, Leal SM, Gunel M, Mane S, Gibbs RA, Boerwinkle E, Hamosh A, Shendure J, Lupski JR, Lifton RP, Valle D, Nickerson DA, Centers

for Mendelian G, Bamshad MJ (2015) The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *Am J Hum Genet* 97: 199-215. doi: 10.1016/j.ajhg.2015.06.009

Chong JX, Yu JH, Lorentzen P, Park KM, Jamal SM, Tabor HK, Rauch A, Saenz MS, Boltshauser E, Patterson KE, Nickerson DA, Bamshad MJ (2016) Gene discovery for Mendelian conditions via social networking: de novo variants in KDM1A cause developmental delay and distinctive facial features. *Genet Med* 18: 788-95. doi: 10.1038/gim.2015.161

ExomeVariantServer NHLBI GO Exome Sequencing Project (ESP). <http://evs.gs.washington.edu/EVS/>. Accessed 4 Nov 2021

Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, Van Vooren S, Moreau Y, Pettett RM, Carter NP (2009) DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet* 84: 524-33. doi: 10.1016/j.ajhg.2009.03.010

Fiume M, Cupak M, Keenan S, Rambla J, de la Torre S, Dyke SOM, Brookes AJ, Carey K, Lloyd D, Goodhand P, Haeussler M, Baudis M, Stockinger H, Dolman L, Lappalainen I, Tornroos J, Linden M, Spalding JD, Ur-Rehman S, Page A, Flicek P, Sherry S, Haussler D, Varma S, Saunders G, Scollen S (2019) Federated discovery and sharing of genomic data using Beacons. *Nat Biotechnol* 37: 220-224. doi: 10.1038/s41587-019-0046-x

Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR (2015) A global reference for human genetic variation. *Nature* 526: 68-74. doi: 10.1038/nature15393

Genoox Franklin by Genoox. <https://franklin.genoox.com>. Accessed 4 Nov 2021

Global Alliance for G, Health (2016) GENOMICS. A federated ecosystem for sharing genomic, clinical data. *Science* 352: 1278-80. doi: 10.1126/science.aaf6162

Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Genome Aggregation Database C, Neale BM, Daly MJ, MacArthur DG (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581: 434-443. doi: 10.1038/s41586-020-2308-7

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation C (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536: 285-91. doi: 10.1038/nature19057

Liu P, Meng L, Normand EA, Xia F, Song X, Ghazi A, Rosenfeld J, Magoulas PL, Braxton A, Ward P, Dai H, Yuan B, Bi W, Xiao R, Wang X, Chiang T, Vetrini F, He W, Cheng H, Dong J, Gijavanekar C, Benke PJ, Bernstein JA, Eble T, Eroglu Y, Erwin D, Escobar L, Gibson JB, Gripp K, Kleppe S, Koenig

MK, Lewis AM, Natowicz M, Mancias P, Minor L, Scaglia F, Schaaf CP, Streff H, Vernon H, Uhles CL, Zackai EH, Wu N, Sutton VR, Beaudet AL, Muzny D, Gibbs RA, Posey JE, Lalani S, Shaw C, Eng CM, Lupski JR, Yang Y (2019) Reanalysis of Clinical Exome Sequencing Data. *N Engl J Med* 380: 2478-2480. doi: 10.1056/NEJMc1812033

Naslavsky MS, Yamamoto GL, de Almeida TF, Ezquina SAM, Sunaga DY, Pho N, Bozoklian D, Sandberg TOM, Brito LA, Lazar M, Bernardo DV, Amaro E, Jr., Duarte YAO, Lebrao ML, Passos-Bueno MR, Zatz M (2017) Exomic variants of an elderly cohort of Brazilians in the ABraOM database. *Hum Mutat* 38: 751-763. doi: 10.1002/humu.23220

OMIM Online Mendelian Inheritance in Man. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). <https://www.omim.org/>. Accessed 4 Nov 2021

Posey JE, O'Donnell-Luria AH, Chong JX, Harel T, Jhangiani SN, Coban Akdemir ZH, Buyske S, Pehlivan D, Carvalho CMB, Baxter S, Sobreira N, Liu P, Wu N, Rosenfeld JA, Kumar S, Avramopoulos D, White JJ, Doheny KF, Witmer PD, Boehm C, Sutton VR, Muzny DM, Boerwinkle E, Gunel M, Nickerson DA, Mane S, MacArthur DG, Gibbs RA, Hamosh A, Lifton RP, Matisse TC, Rehm HL, Gerstein M, Bamshad MJ, Valle D, Lupski JR, Centers for Mendelian G (2019) Insights into genetics, human biology and disease gleaned from family based genomic studies. *Genet Med* 21: 798-812. doi: 10.1038/s41436-018-0408-7

Retterer K, Juusola J, Cho MT, Vitazka P, Millan F, Gibellini F, Vertino-Bell A, Smaoui N, Neidich J, Monaghan KG, McKnight D, Bai R, Suchy S, Friedman B, Tahiliani J, Pineda-Alvarez D, Richard G, Brandt T, Haverfield E, Chung WK, Bale S (2016) Clinical application of whole-exome sequencing across clinical indications. *Genet Med* 18: 696-704. doi: 10.1038/gim.2015.148

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, Committee ALQA (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. *Genet Med* 17: 405-24. doi: 10.1038/gim.2015.30

Sobreira NLM, Arachchi H, Buske OJ, Chong JX, Hutton B, Foreman J, Schiettecatte F, Groza T, Jacobsen JOB, Haendel MA, Boycott KM, Hamosh A, Rehm HL, Matchmaker Exchange C (2017) Matchmaker Exchange. *Curr Protoc Hum Genet* 95: 9 31 1-9 31 15. doi: 10.1002/cphg.50

Wohler E, Martin R, Griffith S, Rodrigues EDS, Antonescu C, Posey JE, Coban-Akdemir Z, Jhangiani SN, Doheny KF, Lupski JR, Valle D, Hamosh A, Sobreira N (2021) PhenoDB, GeneMatcher and VariantMatcher, tools for analysis and sharing of sequence data. *Orphanet J Rare Dis* 16: 365. doi: 10.1186/s13023-021-01916-z

Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, Ward P, Braxton A, Wang M, Buhay C, Veer-araghavan N, Hawes A, Chiang T, Leduc M, Beuten J, Zhang J, He W, Scull J, Willis A, Landsverk M, Craigen WJ, Bekheirnia MR, Stray-Pedersen A, Liu P, Wen S, Alcaraz W, Cui H, Walkiewicz M, Reid J, Bainbridge M, Patel A, Boerwinkle E, Beaudet AL, Lupski JR, Plon SE, Gibbs RA, Eng CM (2014) Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA* 312: 1870-9. doi: 10.1001/jama.2014.14601