

Diagnostic testing laboratories are valuable partners for disease gene discovery: 5-year experience with GeneMatcher

Meghan Towne¹, Mari Rossi¹, Bess Wayburn¹, Jennifer Huang¹, Kelly Radtke¹, Wendy Alcaraz¹, Kelly Farwell¹, and Deepali Shinde¹

¹Ambry Genetics Corp

October 26, 2021

Abstract

Clinical and research laboratories extensively use exome sequencing due to its high diagnostic rates, cost savings, impact on clinical management, and efficacy for disease gene discovery. While the rates of disease gene discovery have steadily increased, only ~16% of genes in the genome have confirmed disease associations. Here we describe our diagnostic laboratory's disease gene discovery and ongoing data-sharing efforts with GeneMatcher. In total, we submitted 246 candidates from 243 unique genes to GeneMatcher, of which 45.93% are now clinically characterized. Submissions with at least one case meeting our candidate genes reporting criteria were significantly more likely to be characterized as of October 2021 compared to genes with no candidates meeting our reporting criteria ($p=0.025$). We reported relevant findings related to these gene-disease associations for 480 probands. In 219 (45.63%) instances, these results were reclassifications after an initial candidate gene (uncertain) or negative report. Since 2013, we have co-authored 105 publications focused on delineating gene-disease associations. Diagnostic laboratories are pivotal for disease gene discovery efforts and can screen phenotypes based on genotype matches, contact clinicians of relevant cases, and issue proactive reclassification reports. GeneMatcher is a critical resource in these efforts.

Diagnostic testing laboratories are valuable partners for disease gene discovery: 5-year experience with GeneMatcher

Meghan C. Towne, Mari Rossi, Bess Wayburn, Jennifer Huang, Kelly Radtke, Wendy Alcaraz, Kelly D. Farwell Hagman, Deepali N. Shinde Ambry Genetics 1 Enterprise Aliso Viejo CA, USA Corresponding authors: Meghan C. Towne - mtowne@ambrygen.com Deepali N. Shinde - dshinde@ambrygen.com No grant funding was used for this work

Conflict of interestAll authors are salaried employees of Ambry Genetics.

Abstract

Clinical and research laboratories extensively use exome sequencing due to its high diagnostic rates, cost savings, impact on clinical management, and efficacy for disease gene discovery. While the rates of disease gene discovery have steadily increased, only ~16% of genes in the genome have confirmed disease associations. Here we describe our diagnostic laboratory's disease gene discovery and ongoing data-sharing efforts with GeneMatcher. In total, we submitted 246 candidates from 243 unique genes to GeneMatcher, of which 45.93% are now clinically characterized. Submissions with at least one case meeting our candidate genes reporting criteria were significantly more likely to be characterized as of October 2021 compared to genes with no candidates meeting our reporting criteria ($p=0.025$). We reported relevant findings related to these gene-disease associations for 480 probands. In 219 (45.63%) instances, these results were reclassifications after an initial candidate gene (uncertain) or negative report. Since 2013, we have co-authored 105 publications focused on delineating gene-disease associations. Diagnostic laboratories are pivotal for disease gene discovery

efforts and can screen phenotypes based on genotype matches, contact clinicians of relevant cases, and issue proactive reclassification reports. GeneMatcher is a critical resource in these efforts.

Keywords (5-6): Disease gene discovery; gene-disease validity; clinical validity; data-sharing; GeneMatcher; Matchmaker Exchange; exome sequencing

1 | INTRODUCTION

The past decade has been a time of rapid disease gene discovery, driven by the rise in popularity of next-generation sequencing (NGS) technologies and the increasing use of web-based collaborative data-sharing initiatives, such as the Matchmaker Exchange (Chong, 2015; Sobreira, 2015; Sobreira, 2017; Boycott, 2019; Bamshad, 2019). Matchmaker Exchange enhances data-sharing and characterization of novel gene-disease associations by connecting multiple genomic and phenotypic databases through a common application programming interface (API) (Sobreira, 2017). One of the components of Matchmaker Exchange is GeneMatcher (<http://www.genematcher.org>) which launched in 2013 to connect scientists and clinicians to share standardized data on candidate genes of interest and the associated phenotypes of individuals with presumed but unidentified Mendelian disorders. By sharing candidate gene information through GeneMatcher, researchers can assemble a critical mass of probands to support the characterization of new gene-disease associations (Sobreira, 2015). As more genes are associated with Mendelian disorders, the overall diagnostic rate of genomic technologies and the potential to identify new therapeutic targets inherently increase (Myers, 2018; Bamshad, 2019). More directly, disease gene discovery impacts patients by ending the notorious ‘diagnostic odyssey,’ providing more tailored clinical care, and informing reproductive risks.

Due to the high volume of testing, diagnostic laboratories that offer diagnostic exome sequencing (DES) are valuable partners for disease gene discovery (Bamshad, 2019). However, most of the data generated by DES are not adequately available for data-sharing and matchmaking (Boycott, 2019). Some diagnostic laboratories evaluate and report rare variants in uncharacterized genes as part of their DES protocol (Retterer, 2016; Farwell Hagman, 2017). At our laboratory, we have a standardized and validated scoring metric for evaluating gene-disease validity (GDV) (Smith, 2017). Genes with no clinical evidence or limited evidence are considered uncharacterized and those with a GDV score of moderate or higher are considered characterized. Both characterized and uncharacterized genes may be reported if meeting our DES reporting criteria and have strong evidence for their association with a proband’s phenotype (Farwell Hagman, 2017). Reporting criteria for uncharacterized candidate genes can vary widely between diagnostic laboratories with published reports of 5.8-24.2% of DES cases having a reported candidate gene (Farwell Hagman, 2017; Retterer, 2016).

Because GDV scores are based on a gene-disease relationship, having access to comprehensive phenotypic data ideally in the form of clinical notes that summarize the salient points of the medical history are crucial for accurately assessing what genes may be relevant for a proband (Seaby, 2020). Genes that meet reporting criteria for our uncharacterized genes are entered into GeneMatcher on a rolling basis. This process allows us to enter high-confidence, potentially disease-causing variants representing the strongest candidates and is consistent with the “gene-to-patient” model proposed by Seaby et al. (2021) to reduce the burden of sifting through large volumes of unvetted variants (“analytical noise”). A thoughtful approach to identifying what variants are the most likely to be disease-causing in a proband is needed before submitting to GeneMatcher to ensure the highest positive outcomes to matches. This in turn leads to newly published data which ultimately can lead to gene characterization (Figure1).

Rates of disease gene discovery have steadily increased over time with a spike occurring as the adoption of NGS technologies became more prominent. However, the rates of publications reporting these discoveries are not keeping up (Bamshad, 2019). The elusive gene-disease relationships that remain to be described may be due to several factors, including complex inheritance or the difficulty in ascertaining probands with extremely rare disorders. Publications may be delayed until the collection of a large enough cohort with robust clinical data curation and paired functional studies. This may be hindering the characterization of gene-disease associations in extremely rare cases with highly specific clinical findings that are less conducive to cohort studies. Moving forward, participation by commercial laboratories in these data-sharing initiatives

is even more imperative to help identify the elusive, ultrarare diagnoses.

Here, we report our laboratory’s experience with GeneMatcher, how it has impacted characterization of gene-disease associations, and collaborations for additional research investigations into the clinical validity of the reported gene-disease associations.

2 | MATERIALS AND METHODS

2.1 | Gene disease validity assessment and gene characterization

Our diagnostic laboratory has established standardized guidelines for curating and scoring GDV which have been independently validated (Smith, 2017). Briefly, the tiered system is similar to the ClinGen model and scores evidence of gene-disease relationships as definitive, strong, moderate, limited, no reported evidence, and conflicting evidence reported (Strande, 2017). A team of scientists reviews published peer-reviewed literature on an ongoing basis and determines the GDV score for each gene-disease relationship. Scores of “moderate” or higher are considered characterized, and those with a score of “limited” or lower are uncharacterized. Some genes have evidence to support associations with more than one phenotype, inheritance pattern or disease mechanism. This could include genes with multiple characterized associations or more a combination of characterized and uncharacterized associations. Gene-disease relationships classified as “limited” by GDV assessment may be included as an uncharacterized candidate for clinical reporting purposes. Genes with supporting evidence, but that do not meet reporting criteria are not included on reports but may be reported as ‘notable’ findings and are available to clinicians by way of filtered variant lists (Farwell Hagman, 2017).

2.2 | Exome Sequencing

DES was performed as previously described (Farwell, 2015; Farwell Hagman, 2017). In brief, samples were prepared using either the SureSelect Target Enrichment System (Agilent Technologies, Santa Clara, CA), SeqCap EZ VCRome 2.0 (Roche NimbleGen, Madison, WI), or the IDT xGen Exome Research Panel V1.0 (Integrated DNA Technologies, Coralville, IA). Sequencing was performed using paired-end, 100 or 150-cycle chemistry on the Illumina HiSeq, NovaSeq, or NextSeq (Illumina, San Diego, CA). Bioinformatics filtering removed common benign variants, intergenic and 3’/5’ UTR variants, intronic variants outside ± 6 , and non-splice-related synonymous variants. Family history-based filtering and inheritance models were applied to the data.

2.3 | Candidate gene analysis for diagnostic exome cases

DES cases submitted as informative trios, defined as having samples from family members representative of both the maternal and paternal lineages, and that were negative after review of characterized gene-disease relationships are eligible for uncharacterized candidate gene analysis as previously described (Farwell-Hagman, 2017). If an identified characterized gene finding only accounts for part of a patient’s phenotype, uncharacterized genes can be analyzed for causes of the unaccounted-for clinical features. Uncharacterized genes that meet reporting criteria are considered candidate genes with uncertain clinical significance (Richards, 2015). Alterations in uncharacterized genes identified during analysis as having potential relevance undergo dual scientist and director review to determine if there is adequate evidence to report as a candidate finding (Smith, 2017).

2.4 | Data-sharing to GeneMatcher

Our laboratory began submitting to GeneMatcher in March 2016. Alterations reported as a candidate gene finding or classified as “insufficient evidence” but with substantiation for disease relevance are entered in GeneMatcher on a rolling basis. A new entry is submitted for each case with a novel mechanism or inheritance model associated with a gene.

2.5 | Follow-up to Matches

Matches are sent to a designated inbox (genematcher@ambrygen.com) which is monitored by research staff

who oversee both our rare disease research efforts and recruitment for external research studies. Upon receiving a match notification, all cases with rare variants and consistent genotype in a matched gene of interest are reviewed for clinical overlap. High-level phenotypic data and variant information are shared in a HIPAA-compliant manner with GeneMatcher collaborators to assess the integrity of the match. Ordering providers for cases with consistent matches are contacted by email with a general description of the collaboration, investigator’s contact information, and gene of interest. A second follow-up message to the ordering provider is sent by encrypted email and contains the proband’s identifiable information, variant of interest, and other relevant information, for example, why the alteration did not meet our reporting criteria at the time of the initial report, and therefore there is no documentation of the alteration or if it was not Sanger confirmed. We do not share the clinician or patient’s contact information with the collaborating investigator without consent. Once the connection is established, the clinician works with the family to obtain appropriate consent for inclusion in any resulting work. We continue as a partner to share additional data as requested (Sanger or IGV imaging, testing methodology details), DNA (if available), and additional family testing for variant co-segregation. In certain cases, our laboratory issues research-grade single site analysis reports to document alterations that were not previously reported to aid in the clinician’s discussion with the family.

2.6 | Data Analysis

To assess the overall impact of our laboratory’s involvement with GeneMatcher, we analyzed the outcomes of our submissions to date. For purposes of this analysis, duplicate entries for a candidate gene-disease relationship with the same inheritance, zygosity, and alteration type were consolidated and counted as a single entry. We retrospectively reviewed all our DES cases from 2011 through January 2021 and tabulated the number of probands with reported alterations in genes with a GeneMatcher entry. For genes with more than one disease association, we only counted cases associated with the phenotype related to the GeneMatcher entry. We assessed the classification of the gene at the time of the initial report (uncharacterized, characterized, or notable), the current GDV score (uncharacterized, moderate, strong, or definitive) and the number of probands who received a reclassification report due to gene characterization. Fisher exact test was used for statistical analysis.

To quantify our laboratory’s contribution to disease gene discovery efforts, we queried PubMed (<https://pubmed.ncbi.nlm.nih.gov/>) for peer-reviewed publications focused on gene characterization efforts which listed authors affiliated with Ambry Genetics. These articles were reviewed for content and categorized as (1) description of new gene-disease association, (2) follow-up case report or supportive evidence (i.e., functional or animal model evidence), (3) follow-up descriptive cohort, and (4) follow-up expansion of phenotype. Publications which included a broadening and further defining of the phenotypic spectrum but not a major expansion in the phenotype were counted as a follow-up descriptive cohort. We cross referenced this published gene list with our GeneMatcher submissions.

3 | RESULTS

In total, 246 candidate entries representing 243 unique genes were submitted to GeneMatcher from our laboratory between March 2016 and September 2021 [Supplement Table 1]. Four gene entries (*EIF2C*, *ITGA11*, *LRFN2*, and *UNC45B*) were subsequently marked as ‘suspended’ because they no longer met candidate gene criteria per our laboratory’s GDV scoring metric. Of note, one of these genes, *UNC45B* was subsequently characterized for an autosomal dominant congenital myopathy in December 2020; however, we had originally submitted as an autosomal recessive cause of Joubert syndrome with skeletal and ocular anomalies. There were three genes (*ANK3*, *RYR3*, and *SPTBN4*) that each had two entries representing distinct phenotypes of interest including differences in inheritance models or suspected mechanism of disease.

As of October 2021, 45.93% of entries have been characterized [Table 1]. Most candidate gene entries were for autosomal dominant *de novo* occurrences; however, there are no significant differences for characterization based on zygosity, inheritance pattern or alterations type. Of the 246 entries, 153 (62.20%) had at least one case with a reported uncharacterized gene finding, meaning the gene, variant, proband combination met candidate gene reporting criteria described in Farwell Hagman, et al. (2017). Submissions that had at least

one case meeting our candidate criteria were significantly more likely to be characterized (79/153; 51.63%) as of October 2021 compared to genes with no candidates meeting reporting criteria (34/93; 36.56%; $p=0.025$).

In total, 196 probands received a candidate gene report involving one of the genes entered in GeneMatcher. Of these, 32 genes had more than one uncharacterized report [range 2-4]. Currently, 480 probands received a reportable finding after the gene was characterized, which accounts for 11.96% (480/4012) of all characterized gene reports from our laboratory (data not shown). Of these 480 cases, 219 (45.63%) were reclassifications of an original result [Figure 2]. Reclassifications include upgrades from candidate gene results (uncertain) or the addition of the finding to a previous report. 75 genes had more than one reported characterized case [range of 2-31].

Our laboratory has co-authored 105 peer-reviewed papers describing gene-disease associations for 104 unique genes between 2013 and 2021 [Supplement Table 2]. 39.42% (41/104) of these genes were candidates we had submitted to GeneMatcher. Majority of publications were the initial description of a gene-disease association, but also follow-up publications that further defined rare Mendelian disorders [Figure 3].

4 | DISCUSSION

Identifying patients is a critical but limiting step in the disease gene discovery process for rare disorders. Diagnostic laboratories provide meaningful contributions to these efforts by screening large testing cohorts for relevant phenotypes based on genotype matches, contacting clinicians of appropriate cases, and issuing proactive reclassification reports. Over the last 5 years, our laboratory has entered 246 candidate entries into GeneMatcher of which 45.93% are now characterized. There was no significant difference for the likelihood to be characterized based on inheritance type, zygosity, or alteration type. However, entries which met our candidate gene reporting criteria were significantly more likely to be characterized, suggesting that while *de novo* and biallelic alterations are easier to identify and represent a large percentage of our GeneMatcher entries (56.91% and 33.74%, respectively), our process of evaluating existing evidence for a gene-disease association is a strong predictor of disease association regardless of mechanism.

By integrating standardized methods for evaluating the evidence available for gene-disease associations into the clinical DES analysis framework, diagnostic laboratories can be valuable partners in screening genes for potential disease associations. All probands with informative trios and negative results after characterized gene review undergo additional candidate gene analysis at our laboratory. Analysts consider the published literature and in-house DES data from previous cases. In some instances, we amass sufficient evidence to characterize a gene-disease association based on internal data. For example, we identified six internal probands with *de novo* rare *ZNF292* alterations and similar neurodevelopmental phenotypes more than one year before the manuscript describing the association was published. Our laboratory submits data on genes with substantial evidence for a disease association into GeneMatcher. This creates a more meaningful contribution to data-sharing compared to automated ‘data dumping’ practices where every rare variant in an uncharacterized gene is entered. Broad data deposits without evaluating the potential clinical relevance can create analytical noise and superfluous communications (Seaby, 2021). Follow-up communications to screen matches are laborious, so contributors to GeneMatcher should have standardized methods to appropriately vet candidate genes before submitting.

One advantage of DES is the ability to analyze the expanding catalog of characterized genes over time. Gene characterization has a cascading effect with the potential for multiple patients to receive a reclassified diagnostic finding once a gene-disease relationship is established. Here, we found GeneMatcher entries ultimately resulted in 480 probands receiving a reportable finding. This accounts for a substantial portion (11.96%) of our total DES cohort with reported characterized findings. Importantly, nearly half of the probands with reported variants in these genes received a reclassification report. Laboratories should have processes in place to curate and score new evidence GDV in real-time, and retrospectively evaluate cases for proactive reclassification. This is an ongoing effort and several cases reported here are under review for clinical overlap and reporting [Figure 2]. Without a process to re-evaluate previous cases, nearly half of this cohort would not have received an updated report. Establishing an accurate molecular diagnosis has pro-

found impacts on patients, especially for individuals with rare disorders. A diagnosis can precipitate tailored clinical management, clearer recurrence risk counseling, and in some cases, eligibility for clinical trials.

Further, because of the ability to reclassify previously reported DES cases, diagnostic laboratories can quickly identify additional probands and further power studies to clarify genotype-phenotype correlations and phenotypic spectrum expansions. Often after an initial cohort describes a new gene-disease relationship there is a follow-up publication of a larger cohort to further delineate the newly characterized disorder. For example, we co-authored Peng et al. (2017) which described the association between biallelic *FDXR* alterations and auditory neuropathy and optic atrophy. The following year, we collaborated with Slone et al. (2018) on a follow-up study to include additional patient reports and functional evidence. We participated in 24 follow-up cohort studies as well as 14 publications describing a single case report or other data in support of the association. Lastly, there were 5 publications we participated in that described a phenotypic expansion for a gene to include major features not previously reported [citations listed in supplement Table 2]. Diagnostic laboratories can serve as a recruitment hub for researchers where more than one proband of interest can be identified through a single source. We can screen the phenotype of individuals with rare variants in a candidate gene of interest, identify the probands with phenotypic overlap, and reach out to multiple clinicians. In some cases, a candidate gene may not have met reporting criteria. In these instances, we can provide reports documenting the presence of an alteration, but not commenting on the clinical relevance until sufficient information is amassed and reported that meets our characterization status. Clinicians find these reports useful to facilitate discussion with families about an alteration which was not originally included on a DES report and the option to enroll in a study further investigating a disease association. Laboratories can further assist the research process by sharing remaining banked DNA (if available and with the appropriate permissions) with researchers and providing cascade testing to other family members for purposes of risk prediction, recurrence risk estimates, and variant segregation within a family.

While diagnostic laboratories can be an efficient tool for finding matches given there is a warehouse of clinical data to screen, the level of clinical data available is dependent on the information submitted at the time of testing. Comprehensive clinic notes that summarize the salient points of a proband’s phenotype should be submitted with all DES testing (Seaby, 2020). A complete clinical description of a proband’s presentation is not only important for variant classification, but also for characterizing new disease genes. In reviewing previously reported cases for collaborators, those with more specific and complete clinical information are more evident as positive matches compared to vague descriptions. Further, robust clinical data can help identify links between a proband’s features and a gene with limited evidence-based on animal models and other functional data.

There is still much work to be done, and many suspected gene-disease associations with insufficient evidence for characterization exist. In total, we have 70 genes that have been reported as a candidate, but remained uncharacterized for more than 2 years, indicating that there was sufficient experimental evidence at the time to associate the gene with disease; however, robust patient data has not yet become available despite the use of GeneMatcher. More than one novel candidate report was issued in only nine of these genes, making these true “n-of-one” cases in need of additional patient reports to corroborate the finding. For example, in 2015, compound heterozygous predicted loss-of-function alterations in the *NMNAT2* gene were identified in a fetus with multiple congenital brain anomalies, contractures, and absence of muscle, which matched perfectly with a mouse model published years earlier (Hicks, 2012). Follow-up in-house structural studies and functional studies through a collaborator support the pathogenicity of these alterations (Lukacs, 2019); however, this gene remains uncharacterized as no additional patient reports have been published. Another example of the importance of robust phenotypic data is the *CHD3* gene, which was reported as a gene of interest in patients with neurodevelopmental disorders in several large cohort studies (O’Roark, 2012; Iossifov, 2014; DDD, 2016), but did not meet our criteria for characterization until a case series with detailed phenotypic data five years after the first published report (Snijders Blok, 2017). Robust case series like this are needed to garnish enough evidence to meet the strict criteria for gene-disease characterization. However, these manuscripts are more likely to take longer to publish, delaying the necessary evidence for characterization to be available. This delay is most profound for gene-disease associations in extremely rare cases that are less conducive to

cohort studies and disorders with clinical heterogeneity which require large cohorts to identify the phenotypic commonalities. Collaborative efforts such as GeneMatcher are needed to power these studies.

By involving diagnostic laboratories in research collaborations there is also a direct clinical impact that results in diagnostic findings going to patients faster. We previously found that clinically relevant reclassification results were issued faster for publications on which we were collaborators compared to genes for which we were not included on the publication (internal data). Involvement of diagnostic laboratories in data-sharing platforms like the Matchmaker Exchange may shorten the time from publication to reclassification report issuance.

Next steps

Trio testing is powerful for the identification of *de novo* and biallelic alterations confirmed *in trans*. In fact, 90.53% of our GeneMatcher entries were for autosomal dominant *de novo* occurrences or autosomal recessive reports. Future identification of elusive disease genes, such as those with lower penetrance or complex inheritance patterns, will require extensive involvement in data-sharing initiatives to further power disease gene discovery efforts. This will be even more important as clinical DES use extends into varied clinical settings. For example, expanding the clinical uses into the prenatal clinic may ascertain more severe phenotypes not compatible with life and therefore yet to be identified. As more commercially available panels are run on an exome backbone, the ability to tailor gene content based on the latest literature and proactively provide reclassification results can extend to other testing platforms. Ongoing participation in GeneMatcher and other Matchmaker Exchange platforms will heighten the diagnostic and disease gene discovery power of DES.

Conclusion

Due to the high volumes of DES data generated and the clinical data curation efforts needed for accurate variant classifications, diagnostic laboratories are a powerful partner for clinical research initiatives, especially those involving disease gene discovery efforts. GeneMatcher is a data-sharing initiative for identifying and collecting genotype-phenotype paired data for gene-disease characterization and has been crucial in our laboratory's gene characterization efforts. In turn, this has a direct and meaningful impact on the diagnostic rates for clinical genetic testing both for prospective tests and retrospectively providing reclassification of variants to previously tested individuals.

Acknowledgments

We are thankful for all the families for their participation, and for the clinicians for sharing the clinical histories and playing a critical role in recruitment, and researchers for their ongoing collaboration. Thank you to Sha Tang, Zoe Powis, and Taylor Cain for their work with GeneMatcher entry evaluations and submissions.

Data Availability

Ambry Genetic supports open data sharing. All data supporting the work presented here can be found at www.genematcher.org or in the supplemental files attached. Further, classified alterations identified during diagnostic testing are deposited into ClinVar.

Ethics approval

All HIPAA-covered patient identifiers were removed. Western Institutional Review Board determined the study to be exempt from the Office for Human Research Protections Regulations for the Protection of Human Subjects (45 CFR 46) under category 4. Retrospective data analysis of anonymized data exempted the study from the requirement to receive consent from patients.

References

Bamshad, M. J., Nickerson, D. A., & Chong, J. X. (2019). Mendelian Gene Discovery: Fast and

- Furious with No End in Sight. *American journal of human genetics* , 105 (3), 448–455. <https://doi.org/10.1016/j.ajhg.2019.07.011>
- 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, I. K., Visel, A., & Baynam, G. (2019). A Diagnosis for All Rare Genetic Diseases: The Horizon and the Next Frontiers. *Cell* , 177 (1), 32–37. <https://doi.org/10.1016/j.cell.2019.02.040>
- 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., Hetrick, K., ... 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>
- Deciphering Developmental Disorders Study (2017). Prevalence and architecture of de novo mutations in developmental disorders. *Nature* , 542 (7642), 433–438. <https://doi.org/10.1038/nature21062>
- Farwell, K. D., Shahmirzadi, L., El-Khechen, D., Powis, Z., Chao, E. C., Tippin Davis, B., Baxter, R. M., Zeng, W., Mroske, C., Parra, M. C., Gandomi, S. K., Lu, I., Li, X., Lu, H., Lu, H. M., Salvador, D., Ruble, D., Lao, M., Fischbach, S., Wen, J., ... Tang, S. (2015). Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genetics in medicine : official journal of the American College of Medical Genetics* , 17 (7), 578–586. <https://doi.org/10.1038/gim.2014.154>
- Farwell Hagman, K. D., Shinde, D. N., Mroske, C., Smith, E., Radtke, K., Shahmirzadi, L., El-Khechen, D., Powis, Z., Chao, E. C., Alcaraz, W. A., Helbig, K. L., Sajan, S. A., Rossi, M., Lu, H. M., Huether, R., Li, S., Wu, S., Nuñez, M. E., & Tang, S. (2017). Candidate-gene criteria for clinical reporting: diagnostic exome sequencing identifies altered candidate genes among 8% of patients with undiagnosed diseases. *Genetics in medicine : official journal of the American College of Medical Genetics* , 19 (2), 224–235. <https://doi.org/10.1038/gim.2016.95>
- Hicks, A. N., Lorenzetti, D., Gilley, J., Lu, B., Andersson, K. E., Miligan, C., Overbeek, P. A., Oppenheim, R., & Bishop, C. E. (2012). Nicotinamide mononucleotide adenylyltransferase 2 (Nmnat2) regulates axon integrity in the mouse embryo. *PLoS one* , 7 (10), e47869. <https://doi.org/10.1371/journal.pone.0047869>
- Iossifov, I., O’Roak, B. J., Sanders, S. J., Ronemus, M., Krumm, N., Levy, D., Stessman, H. A., Witherspoon, K. T., Vives, L., Patterson, K. E., Smith, J. D., Paepers, B., Nickerson, D. A., Dea, J., Dong, S., Gonzalez, L. E., Mandell, J. D., Mane, S. M., Murtha, M. T., Sullivan, C. A., ... Wigler, M. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature* , 515 (7526), 216–221. <https://doi.org/10.1038/nature13908>
- Myers, K. A., Johnstone, D. L., & Dymont, D. A. (2019). Epilepsy genetics: Current knowledge, applications, and future directions. *Clinical genetics* , 95 (1), 95–111. <https://doi.org/10.1111/cge.13414>
- Lukacs, M., Gilley, J., Zhu, Y., Orsomando, G., Angeletti, C., Liu, J., Yang, X., Park, J., Hopkin, R. J., Coleman, M. P., Zhai, R. G., & Stottmann, R. W. (2019). Severe biallelic loss-of-function mutations in nicotinamide mononucleotide adenylyltransferase 2 (NMNAT2) in two fetuses with fetal akinesia deformation sequence. *Experimental neurology* , 320 , 112961. <https://doi.org/10.1016/j.expneurol.2019.112961>
- O’Roak, B. J., Stessman, H. A., Boyle, E. A., Witherspoon, K. T., Martin, B., Lee, C., Vives, L., Baker, C., Hiatt, J. B., Nickerson, D. A., Bernier, R., Shendure, J., & Eichler, E. E. (2014). Recurrent de novo mutations implicate novel genes underlying simplex autism risk. *Nature communications* , 5 , 5595. <https://doi.org/10.1038/ncomms6595>
- Peng, Y., Shinde, D. N., Valencia, C. A., Mo, J. S., Rosenfeld, J., Truitt Cho, M., Chamberlin, A., Li, Z., Liu, J., Gui, B., Brockhage, R., Basinger, A., Alvarez-Leon, B., Heydemann, P., Magoulas, P. L., Lewis, A. M., Scaglia, F., Gril, S., Chong, S. C., Bower, M., ... Huang, T. (2017). Biallelic mutations in the ferredoxin

reductase gene cause novel mitochondriopathy with optic atrophy. *Human molecular genetics* , 26 (24), 4937–4950. <https://doi.org/10.1093/hmg/ddx377>

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., Chung, W. K., ... Bale, S. (2016). Clinical application of whole-exome sequencing across clinical indications. *Genetics in medicine : official journal of the American College of Medical Genetics* , 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., & ACMG 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 : official journal of the American College of Medical Genetics* , 17 (5), 405–424. <https://doi-org/10.1038/gim.2015.30>

Seaby, E. G., & Ennis, S. (2020). Challenges in the diagnosis and discovery of rare genetic disorders using contemporary sequencing technologies. *Briefings in functional genomics* , 19 (4), 243–258. <https://doi-org/10.1093/bfpg/elaa009>

Seaby, E. G., Rehm, H. L., & O'Donnell-Luria, A. (2021). Strategies to Uplift Novel Mendelian Gene Discovery for Improved Clinical Outcomes. *Frontiers in genetics* , 12 , 674295. <https://doi-org/10.3389/fgene.2021.674295>

Slone, J., Peng, Y., Chamberlin, A., Harris, B., Kaylor, J., McDonald, M. T., Lemmon, M., El-Dairi, M. A., Tchapyjnikov, D., Gonzalez-Krellwitz, L. A., Sellars, E. A., McConkie-Rosell, A., Reinholdt, L. G., & Huang, T. (2018). Biallelic mutations in FDXR cause neurodegeneration associated with inflammation. *Journal of human genetics* , 63 (12), 1211–1222. <https://doi.org/10.1038/s10038-018-0515-y>

Smith, E. D., Radtke, K., Rossi, M., Shinde, D. N., Darabi, S., El-Khechen, D., Powis, Z., Helbig, K., Weller, K., Grange, D. K., Tang, S., & Farwell Hagman, K. D. (2017). Classification of Genes: Standardized Clinical Validity Assessment of Gene-Disease Associations Aids Diagnostic Exome Analysis and Reclassifications. *Human mutation* , 38 (5), 600–608. <https://doi-org/10.1002/humu.23183>

Snijders Blok, L., Rousseau, J., Twist, J., Ehresmann, S., Takaku, M., Venselaar, H., Rodan, L. H., Nowak, C. B., Douglas, J., Swoboda, K. J., Steeves, M. A., Sahai, I., Stumpel, C., Stegmann, A., Wheeler, P., Willing, M., Fiala, E., Kochhar, A., Gibson, W. T., Cohen, A., ... Campeau, P. M. (2018). CHD3 helix-se domain mutations cause a neurodevelopmental syndrome with macrocephaly and impaired speech and language. *Nature communications* , 9 (1), 4619. <https://doi.org/10.1038/s41467-018-06014-6>

Sobreira, N., Arachchi, H., Buske, O. J., Chong, J. X., Hutton, B., Foreman, J., Schiettecatte, F., Groza, T., Jacobsen, J., Haendel, M. A., Boycott, K. M., Hamosh, A., Rehm, H. L., & Matchmaker Exchange Consortium (2017). Matchmaker Exchange. *Current protocols in human genetics* , 95 , 9.31.1–9.31.15. <https://doi-org/10.1002/cphg.50>

Sobreira, N., Schiettecatte, F., Boehm, C., Valle, D., & Hamosh, A. (2015). New tools for Mendelian disease gene identification: PhenoDB variant analysis module; and GeneMatcher, a web-based tool for linking investigators with an interest in the same gene. *Human mutation* , 36 (4), 425–431. <https://doi-org/10.1002/humu.22769>

Strande, N. T., Riggs, E. R., Buchanan, A. H., Ceyhan-Birsoy, O., DiStefano, M., Dwight, S. S., Goldstein, J., Ghosh, R., Seifert, B. A., Sneddon, T. P., Wright, M. W., Milko, L. V., Cherry, J. M., Giovanni, M. A., Murray, M. F., O'Daniel, J. M., Ramos, E. M., Santani, A. B., Scott, A. F., Plon, S. E., ... Berg, J. S. (2017). Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. *American journal of human genetics* , 100 (6), 895–906. <https://doi-org./10.1016/j.ajhg.2017.04.015>

Figure Titles and Captions

Figure 1: Uncharacterized gene reporting and reclassification workflow

Figure 1: Up until 2020, uncharacterized gene analysis was performed for informative trios with an uncertain or negative result following characterized gene analysis. Currently, uncharacterized genes are analyzed for negative cases or for probands where the identified characterized gene finding only accounts for part their phenotype ('partial overlap'). Genes are continually evaluated based on the latest literature to determine if the published evidence is sufficient to characterize the gene-disease association. Once characterized, all previously reported DES cases with rare variants in a gene are assessed for clinical overlap and proactive reclassification reports are issued to appropriate cases. Our laboratory has several workflows are addressing uncertain results for reclassification. New gene characterizations result in our highest rate of reclassifications. Participation in data-sharing initiatives include GeneMatcher, ClinVar, the Undiagnosed Disease Network, and collaborations established directly with investigators. Only gene and variant information are shared on these platforms with follow-up clinical details provided on a case-by-case basis. Ordering clinicians are contacted for a promising lead and connected to collaborators for additional data sharing and patient consent to participate in publications or additional studies. Other efforts for reclassification include performing reanalysis by clinician request and implementing bioinformatic pipeline upgrades and other technology upgrades, both of which are largely time-dependent. Family studies and additional lab studies (i.e. RNA analysis, structural modeling) are variant-dependent and to date not widely available for most alterations.

Figure 2: Overall impact of gene characterization following GeneMatcher entry

Figure 2: 480 characterized gene reports were issued as of October 2021 for gene-disease associations entered in GeneMatcher. In total, 219 (45.63%) of these were reclassifications proactively issued by the laboratory. 109 of these reports were upgraded reclassifications from candidate gene reports, and 110 were reclassifications issued to individuals who did not initially meet our candidate gene reporting criteria. After establishing a gene-disease association, 261 additional probands received a relevant finding in one of these genes, illustrating the impact disease gene discoveries have on the overall DES diagnostic rate. For 79 cases for which we issued a candidate gene report, the gene-disease association remains uncharacterized. In 4 cases, further review of a reported candidate gene no longer met our criteria. These findings were removed from the proband's report, and reclassifications were issued. Additionally, these matches were marked as 'suspended' in GeneMatcher. Currently, we have 4 cases pending review for reclassification based on literature published this year.

Figure 3: Co-authored peer-reviewed publications

Figure 3: Since 2013, we have co-authored 105 peer-reviewed papers describing gene-disease associations. These included descriptions of 104 unique genes. 4 genes (*BPTF*, *FDXR*, *IQSEC2*, and *SON*) were the topic of more than one paper. A comprehensive list of these publications can be found in supplement Table 2.

Table Title and Captions

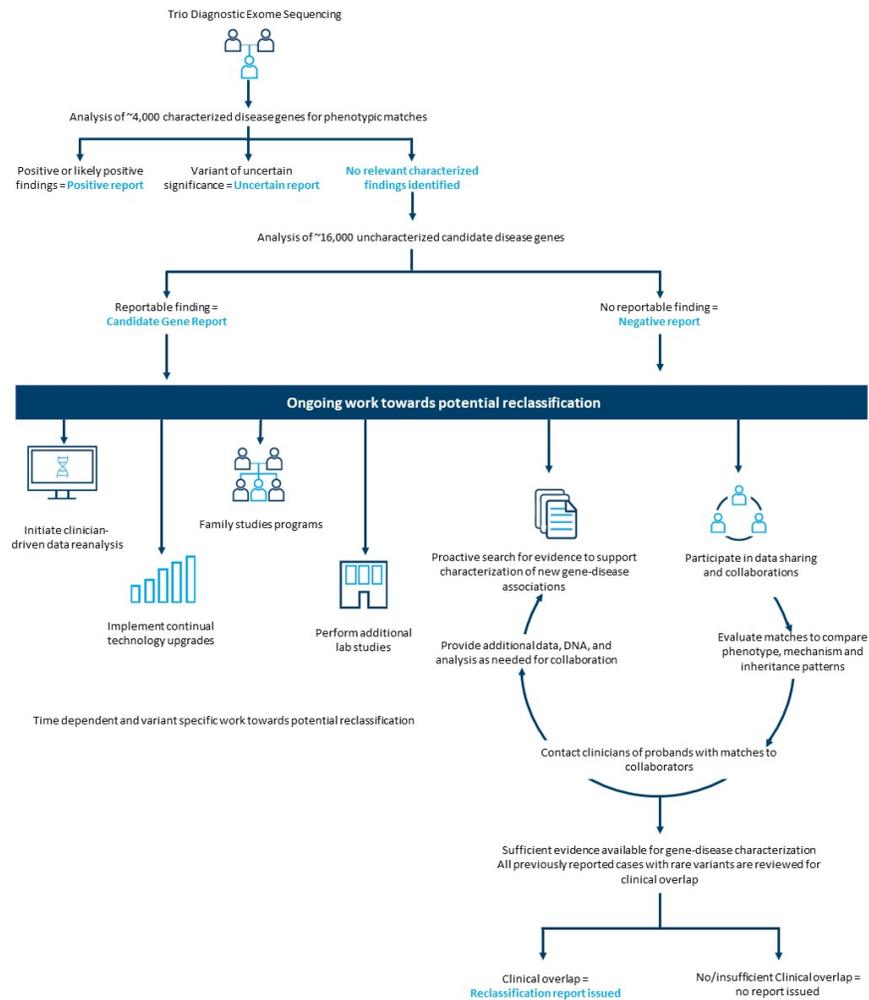
Table 1: Characteristics of GeneMatcher Entries

Table 1: Inheritance, zygosity and alteration types for GeneMatcher entries are listed. Entries with homozygous or compound heterozygous alterations reported were counted for each alteration type to accurately record the percentage of each alteration type. There were no statistical differences within each category.

Supplements

Table 1: All GeneMatcher entries with current gene-disease validity scores and associated case count

Table 2: List of co-authored peer-reviewed publications involving gene-disease descriptions



Hosted file

Table 1.docx available at <https://authorea.com/users/443073/articles/543141-diagnostic-testing-laboratories-are-valuable-partners-for-disease-gene-discovery-5-year-experience-with-genematcher>

