Published in final edited form as:

Hum Mutat. 2015 October ; 36(10): 928–930. doi:10.1002/humu.22844.

GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene

Nara Sobreira^{1,*}, François Schiettecatte², David Valle¹, and Ada Hamosh¹

¹Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland

²FS Consulting, Salem, Massachusetts

Abstract

Here, we describe an overview and update on GeneMatcher (<http://www.genematcher.org>), a freely accessible Web-based tool developed as part of the Baylor-Hopkins Center for Mendelian Genomics. We created GeneMatcher with the goal of identifying additional individuals with rare phenotypes who had variants in the same candidate disease gene. We also wanted to facilitate connections to basic scientists working on orthologous genes in model systems with the goal of connecting their work to human Mendelian phenotypes. Meeting these goals will enhance the identification of novel Mendelian genes. Launched in September, 2013, Gene-Matcher now has 2,178 candidate genes from 486 submitters spread across 38 countries entered in the database (June 1, 2015). GeneMatcher is also part of the Match-maker Exchange (<http://matchmakerexchange.org/>) with an Application Programming Interface enabling submitters to query other databases of genetic variants and phenotypes without having to create accounts and data entries in multiple systems.

Keywords

whole exome sequencing; whole genome sequencing; next-generation sequencing; Mendelian disease; variant analysis; Matchmaker Exchange

Introduction

In the last few years, whole-exome sequencing (WES) has been the main method used to search for Mendelian disease genes. Identifying the pathogenic mutation from among thousands of genomic variants in typical WES is a challenge. In about 75% of the individuals who have clinical WES, the responsible gene and variant(s) cannot be determined (Yang et al., 2014). Reasons for this relatively low yield include phenotypic variation, the small number of currently known disease genes (~3,400 or only 15% of the total protein coding genes), uncertainty regarding functional consequences of identified variants, and limited connections between clinicians with clinical WES data and basic scientists. Structured, comprehensive phenotypic data in a widely shared format as well as

*Correspondence to: Nara Sobreira, 733 North Broadway Suite 512, Baltimore, Maryland 21205. nsobrei2@jhmi.edu.
For the Matchmaker Exchange Special Issue

improvements in searching for patients or model organisms with variants in specific candidate genes are important for the success of clinical and research WES analysis. Here, we describe an overview and update on GeneMatcher (<http://www.genematcher.org>; Sobreira et al., 2015), a freely accessible Web-based tool designed to enable connections between clinicians and researchers from around the world who share an interest in the same gene(s) with the goal of connecting genes to Mendelian phenotypes and increasing our understanding of these rare disorders.

GeneMatcher Overview

GeneMatcher was developed almost entirely in Python with a few Perl support scripts. The Website was developed using the Django Web application framework and uses MySQL as the underlying database. The entire stack runs on a virtual machine running Red-Hat Enterprise Linux. Python was selected because it is a popular, robust language with a wide variety of external libraries, and Django was a natural choice as a Web application framework since it is itself developed in Python. GeneMatcher is hosted in a Johns Hopkins University data center specifically designed to host sensitive systems and data. External access to GeneMatcher is only possible through secure HTTP, and access to the machine itself is only possible from within the Johns Hopkins University network. The machine operating system is kept up to date with patches as needed, and backups are encrypted.

GeneMatcher does not collect identifiable data. After creating an account, the site allows investigators to post a gene(s) (by gene symbol, Entrez- or Ensembl-Gene ID) of interest (Fig. 1A) and will connect investigators who post the same gene. The match is done automatically upon submission and both submitters receive an email notification with each other's contact information, including name, affiliation, email address, and matched gene. Follow-up is at the discretion of the submitters. In order to generate a match, each submitter is required to enter at least the name of the gene(s) of interest (Fig. 1A). The database is not searchable by other mechanisms. Submitters have access to their own data and may edit it or delete it at will (Fig. 1A); they can also search among their own entries based on MIM number, features (Fig. 1B), gene name, or genomic location (Fig. 1C). There is also an option to provide diagnosis based upon OMIM® number (Fig. 1D) and variant based on base pair position (Fig. 1A) and match based on this information, but this is not required. If a match is not identified at the time of submission, the entered genes will continue to be queried as new entries are added to the database. Basic scientists working on orthologous genes in model systems are also encouraged to enter their gene(s) of interest and indicate what organism they are working with (Fig. 1E). GeneMatcher can also accept phenotypic features (Fig. 1F). Phenotypic features, genes, and variants can be submitted directly from any implementation of PhenoDB [Hamosh et al., 2013; Sobreira et al., 2015], allowing easier and faster submissions by clinicians and researchers using it.

GeneMatcher is best used for candidate genes that are not yet known to have an associated human phenotype. Optimally, submitters are encouraged to enter only strong candidate genes with validated candidate-causative variants, for example, by an orthogonal sequencing method. Following these two rules will limit matches to those that are more likely to identify new Mendelian disease genes.

GeneMatcher Update

GeneMatcher was launched on September, 2013, and since then, 2,433 individual genes from 539 submitters spread across 49 countries have been submitted to the database and 450 matches have been generated enabling collaboration between clinicians and researchers from different countries and distinct backgrounds but with interest in the same genes (July 1, 2015). Of the 2,178 genes, 307 are related to animal models. The growth in the number of submitted genes and matches is shown in Fig. 2. Because only the submitters have access to the details of the matches, we can only follow up on our own matches or by personal communication. Baylor-Hopkins Center for Mendelian Genomics (BHCMG) has submitted >180 candidate genes from 62 unrelated families. We have had at least 69 matches, corresponding to 43 genes and 30 unrelated families. At least 16 matches (corresponding to nine novel Mendelian genes) are also a phenotype match and three of them are being submitted for publication including *SPATA5*, *HNRNPK*, and *TELO2*. A match on *AP3B2* connected a human phenotype to a mouse model with similar phenotypic features [Fairfield et al., 2015]. At least 26 of 69 BHCMG matches (18 genes) were in instances in which the phenotypes of the affected individuals were not completely overlapping. The remaining matches are still under evaluation. GeneDx, a clinical diagnostic laboratory, has also started using GeneMatcher by uploading their class 3 genes (genes with likely pathogenic variants and not yet associated with disease), generating dozens of matches between clinicians and researchers.

To enhance interpretation of clinical and research exome sequence data, we are now testing different algorithms to match entries using phenotypic features, with or without candidate genes. This will enable identification of individuals with rare phenotypes from around the world that have not been previously described.

As part of the Matchmaker Exchange (<http://matchmakerexchange.org/>), we have also participated in the development of an Application Programming Interface now being implemented by GeneMatcher, PhenomeCentral, and DECIPHER, and open to others, allowing submitters to query other datasets of genetic variants and phenotypes without having to create different accounts and entering data in distinct systems. GeneMatcher can send out queries to these databases based on gene, genomic location, OMIM number, and soon phenotypic features using PhenoDB terms that match to ICHPT (International Consortium for Human Phenotype Terminologies) terms and/or HPO terms [Köhler et al., 2014].

Discussion

GeneMatcher was developed as part of the BHCMG because of our need to identify additional individuals with rare phenotypes sharing variants in the same gene and to make connections to basic scientists working on orthologous genes in model systems. For many of the families investigated in the project by WES, we have a small list of candidate variants and choosing only one candidate gene can be difficult. For many of the candidate genes identified, little is known about their biology and animal models are often not available, especially for the missense variants. Because of the rarity of most of the phenotypes being

investigated, proving the causality of a gene by identifying multiple unrelated probands with pathogenic variants in the same gene can be a challenge. Identifying similar individuals by phenotypic description is not a simple task, many times the individuals are not completely investigated or described and the lack of description of an anomaly does not mean that the individual does not have it. Some features can be described with different terms and many dysmorphic features are subjective and the precise description of these features depends on the experience of the clinician performing the evaluation. Defining what degree of phenotypic overlap is required to identify similar individuals is challenging and may vary from one phenotype to the other. Moreover, most Mendelian phenotypes are characterized by variable expressivity and in many instances the “same” phenotype can be caused by multiple genes. For all these reasons, we chose to search for matches based on the genes rather than phenotypes. Searching by gene also eliminates the need for consenting since no individual identifiers are required, making the search process simpler and faster.

The follow-up of the BHCMG data has revealed gene matches from individuals with discordant phenotypes; at least 26 of 69 BHCMG matches (corresponding to 18 genes) did not have matching phenotypic features. There are at least three explanations for this result. The first is that the gene is one in which different variants produce distinct phenotypes, for example, as occurs for *FGFR3*, *FBN1*, and many others. Second, the gene was entered prior to Sanger validation of the variants and/or segregation testing and represents a false positive. Third, the entered gene is one that harbors a large number of rare benign variants, that is, a valid gene match but the variants are not responsible for the phenotype.

The GeneMatcher approach has resulted in connection of novel Mendelian genes (e.g., *SPATA5*, *HNRNPK*, and *TELO2*) to novel phenotypes. We suggest that informed use of GeneMatcher will enable many new gene/phenotype connections. We anticipate the full impact of GeneMatcher will be revealed in the published literature over the next years.

Acknowledgments

Contract grant sponsor: NHGRI (1U54HG006542).

References

- Fairfield H, Srivastava A, Ananda G, Liu R, Kircher M, Lakshminarayana A, Harris BS, Karst SY, Dionne LA, Kane CC, Curtain M, Berry ML, et al. Exome sequencing reveals pathogenic mutations in 91 strains of mice with Mendelian disorders. 2015. *Genome Res.* 2015; 25:948–957. [PubMed: 25917818]
- Hamosh A, Sobreira N, Hoover-Fong J, Sutton VR, Boehm C, Schiettecatte F, Valle D. PhenoDB: a new web-based tool for the collection, storage, and analysis of phenotypic features. 2013. *Hum Mutat.* 2013; 34:566–571. [PubMed: 23378291]
- Köhler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Bailleul-Forestier I, Black GCM, Brown DL, Brudno M, Campbell J, FitzPatrick DR, Eppig JT, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res.* 2014; 42(Database issue):D966–D974. [PubMed: 24217912]
- Sobreira N, Schiettecatte F, Boehm C, Valle D, Hamosh A. 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. *Hum Mutat.* 2015; 36:425–431. [PubMed: 25684268]

Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, Ward P, Braxton A, Wang M, Buhay C, Veeraraghavan N, Hawes A, et al. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA. 2014; 312:1870–1879. [PubMed: 25326635]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

A

Results (human genes only) :

Gene Name	Ensembl Gene ID	Entrez Gene ID	Chromosome	Start Position	End Position	Assembly
The Gene name or the Ensembl Gene ID or the Entrez Gene ID is required. The Chromosome is required if the Start/End Positions are set. Both Start/End Positions are required if either are set. The Assembly defaults to hg19 if the Start/End Positions are set, valid values are hg18, hg19, and hg20.						

Upload results file: no file selected. You can [download a sample results spreadsheet](#).

You can add results to the Results section either by entering them in the table above or by uploading a results spreadsheet file. Results uploaded with a results spreadsheet file will be appended to the existing results. You can [download a sample results spreadsheet](#) in CSV. You can clear existing results by selecting the Delete results table checkbox above.

☒ Select this checkbox if you want to run a match against other submissions when saving.

B

Submissions Identifier Search Diagnosis Search Gene Search

MIM number:

Disorder type: Select optional disorder here...

Presumed inheritance pattern: Select optional inheritance pattern...

Features:

☐ Submissions must match all features

C

Submissions Identifier Search Diagnosis Search Gene Search

Gene names: ☐ Submissions must match all gene names

Genomic locations: ☐ Submissions must match all genomic locations

Assembly: (defaults to hg19)

Gene name:

Genomic location:

Assembly:

☐ Optional match on genomic location

D

Genetic Disorder (optional) :

Diagnosis Category:

☒ A Mendelian disorder described in OMIM for which the responsible gene has not been identified (example: 223170, Dubowitz syndrome)

MIM number:

☐ A Mendelian disorder with locus heterogeneity (LO) described in OMIM for which the known responsible gene(s) explain only a fraction of the cases and those accounting for more than 25% have been ruled out in your case, (example: 192650, cardiomyopathy, familial hypertrophic)

☐ An unknown disorder (not described in OMIM) but with segregation in your family consistent with Mendelian inheritance

Inheritance pattern: Pick best fit...

The Genetic Disorder section is optional, and all the fields in this section are optional.

E

Organism :

Organism:

Select the organism.
You should only enter Human genes in the Results section.

F

Family Members (optional) :

Family Member	Affected	Deceased Phenotypes	Family Member ID
Female Patient	☒ Yes ☐ No ☐ Unk.	☐	Add features (optional)
Mother	☐ Yes ☒ No ☐ Unk.	☐	Add features (optional)
Father	☐ Yes ☒ No ☐ Unk.	☐	Add features (optional)
Select relation	☐ Yes ☐ No ☐ Unk.	☐	Add features (optional)
Unk. = Unknown			

The Family Members section is optional, and all the fields in this section are optional.

Figure 1.

A: A user can post a gene(s) by gene symbol, Entrez- or Ensembl-Gene ID or a variant based on base pair position. Entries can be deleted at will. **B** and **C:** A user can search among their own entries based on MIM number, clinical features (**B**), gene name or genomic location (**C**). **D:** A user also has the option to provide a diagnosis based upon OMIM® number and match on that. **E:** Basic scientists working on orthologous genes in model systems can indicate the organism they are studying. **F:** GeneMatcher can also accept phenotypic features and, in the near future, users will also be able to match on features.

Figure 2.
The growth in the number of genes submitted to GeneMatcher and the number of matches.