



HHS Public Access

Author manuscript

Hum Mutat. Author manuscript; available in PMC 2023 June 01.

Published in final edited form as:

Hum Mutat. 2022 June ; 43(6): 743–759. doi:10.1002/humu.24364.

ModelMatcher: A scientist-centric online platform to facilitate collaborations between stakeholders of rare and undiagnosed disease research

J. Michael Harnish^{1,2}, Lucian Li^{2,3}, Sanja Rogic^{4,5}, Guillaume Poirier-Morency^{4,5}, Seon-Young Kim^{2,3}, Undiagnosed Diseases Network, Kym M. Boycott⁶, Michael F. Wangler^{1,2,7}, Hugo J. Bellen^{1,2,7,8}, Philip Hieter⁴, Paul Pavlidis^{4,5}, Zhandong Liu^{2,3,9}, Shinya Yamamoto^{1,2,7,8,#}

¹Department of Molecular and Human Genetics, Baylor College of Medicine (BCM), Houston, TX, 77030, USA

²Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital (TCH), Houston, TX, 77030, USA

³Department of Pediatrics, Division of Neurology and Developmental Neuroscience, BCM, Houston, TX, 77030, USA

⁴Michael Smith Laboratories, University of British Columbia, Vancouver, BC V6T1Z4, Canada

⁵Department of Psychiatry, University of British Columbia, Vancouver, BC V6T1Z4, Canada

⁶Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, ON K1H8L1, Canada

⁷Development, Disease Models & Therapeutics Graduate Program, BCM, Houston, TX, 77030, USA

⁸Department of Neuroscience, BCM, Houston, TX, 77030, USA

⁹Quantitative and Computational Biosciences Graduate Program, BCM, Houston, TX, 77030, USA

Abstract

Next-generation sequencing is a prevalent diagnostic tool for undiagnosed diseases and has played a significant role in rare disease gene discovery. While this technology resolves some cases, others are given a list of possibly damaging genetic variants necessitating functional studies. Productive collaborations between scientists, clinicians, and patients (affected individuals) can help resolve such medical mysteries, and provide insights into *in vivo* function of human genes. Furthermore, facilitating interactions between scientists and research funders, including non-profit organizations or commercial entities, can dramatically reduce the time to translate discoveries from bench to bedside. Several systems designed to connect clinicians and researchers with a shared gene of interest have been successful. However, these platforms exclude some stakeholders based on their role or geography. Here we describe ModelMatcher, a global online matchmaking

#Corresponding Author: Shinya Yamamoto (yamamoto@bcm.edu, phone: +1-832-824-8119).

tool designed to facilitate cross-disciplinary collaborations, especially between scientists and other stakeholders of rare and undiagnosed disease research. ModelMatcher is integrated into the Rare Diseases Models and Mechanisms Network and Matchmaker Exchange, allowing users to identify potential collaborators in other registries. This living database decreases the time from when a scientist or clinician is making discoveries regarding their genes of interest, to when they identify collaborators and sponsors to facilitate translational and therapeutic research.

Keywords

Collaboration; Functional Studies; Matchmaker Exchange (MME); Matchmaking; Model Organisms; Rare Diseases; Rare Diseases Models and Mechanisms (RDMM) Network; Undiagnosed Diseases; Undiagnosed Diseases Network (UDN); Variants of Unknown Significance (VUS)

Introduction:

Rare and undiagnosed diseases are prevalent and costly issues facing the modern healthcare landscape. In the United States alone, 30 million individuals are estimated to be suffering from a rare disease, over 80% of which are estimated to have strong genetic components (Nguengang Wakap et al. 2020). The use of whole-exome sequencing (WES) or whole-genome sequencing (WGS) technologies as diagnostic tools has been increasing steadily. This increase is in part due to the decreasing cost of high-throughput sequencing. What remains a challenge is that many affected individuals who undergo sequencing are still not given a specific diagnosis (Coventry et al., 2010; Lupski et al., 2011). Affected individuals with an inconclusive WES or WGS result typically receive a list of several identified rare variants that are potentially responsible for the individual's disease. While some variants can be ruled out by comparing the disease-gene association to the affected individual's clinical presentation, this method still often leaves many candidates. A number of these genetic changes are variants of unknown significance (VUS) in known disease-causing genes whose associate phenotypes have some overlap with the affected individuals' clinical presentation. Some are rare variants in genes of uncertain significance (GUS), which have yet to be associated with a genetic disorder in human. Missense variants in particular are difficult to interpret as modern *in silico* tools are still not able to accurately predict their functional consequences (Ghosh et al. 2017). To resolve these cases, experiments must be performed in a laboratory setting to better understand the functional consequence of the variant allele (Harnish et al. 2019). However, most clinicians and affected individuals do not have contact with scientists who can carry out this task.

Collaborations between scientists and clinicians are increasingly commonplace, especially between model organism biologists who work on fundamental biological questions and physicians who see undiagnosed patients (Wangler et al., 2017). Several countries and regions have created research programs designed to facilitate these endeavors. These programs include the Undiagnosed Diseases Network (UDN, <https://undiagnosed.hms.harvard.edu/>) in the United States (Gahl et al., 2016) and the Rare Diseases Models and Mechanisms (RDMM, <http://www.rare-diseases-catalyst-network.ca/>) Network in Canada (Boycott et al., 2020). Over the last five years, these platforms have

been demonstrated to be very effective in facilitating diagnosis as well as identifying potential therapeutic avenues for many rare diseases. Functional studies in the UDN are carried out through the centralized Model Organisms Screening Centers (MOSC) by a relatively small number of model organism experts with broad biological knowledge. The MOSC sites utilize sophisticated genetic techniques in fruit flies, nematode worms, and zebrafish to characterize the effects of variant alleles found in rare disease patients to ultimately aid in the diagnosis and obtain mechanistic insights to further understand disease pathogenesis (Baldrige et al., 2021). While this UDN-MOSC system has many advantages including economy of scale, pipeline standardization, and establishment of strong trust relationships between scientists and clinicians over time, it doesn't take full advantage of the entirety of scientific expertise that various scientists in the field possess. In contrast, RDMM adopts a decentralized model in which a Coordination Center that is supported by Clinical and Scientific Advisory Committees performs matchmaking between model organism scientists and clinicians with shared genetic interests within Canada (Boycott et al., 2020). Once a match is made, scientists are given seed grants (Can\$25,000/case) to carry out experiments. Over the past three years, similar platforms were established in Europe [Solve-RD/RDMM-Europe (<https://solve-rd.eu/>)], Japan [J-RDMM (<https://j-rdmm.org/indexEn.html>)] and Australia [Australian Functional Genomics Network (AFGN, <https://www.functionalgenomics.org.au/>)] to facilitate collaborative research within each country or region. The RDMM model catalyzes and supports collaborations within each network, and there has been some effort to facilitate collaborations across different RDMM-like networks (Boycott et al., 2020). However, there are many scientists and clinicians who cannot participate (e.g., in countries like the United States), primarily because there are large administrative barriers in transferring grant funds across borders. Since clinicians with undiagnosed patients and scientists with expertise in specific genes may be found in different parts of the world, there is an unmet need to establish an open platform designed to facilitate collaborations on a global scale. In addition, while these existing platforms have been successful in facilitating collaborations between model organism biologists and clinical groups, scientists with other types of expertise such as cell biologists, biochemists, and structural biologists who primarily carry out *in vitro* experiments are under-represented.

In addition to clinicians, other stakeholders of rare and undiagnosed diseases such as affected individuals, families, non-profit entities, public funding agencies, and biotech or pharmaceutical companies can benefit from having direct connections with scientists. In contrast to clinicians who often interact with these entities, only a handful of scientists, most of whom are affiliated with medical schools, have experience in interacting with these types of stakeholders. Considering that there are many scientists in basic biology departments who have the capacity to conduct sophisticated gene and variant functional studies in a timely manner, it will be valuable to establish a network in which both clinical and non-clinical stakeholders can match with scientists who possess relevant expertise. Such a network will enable all stakeholders to find scientists with expertise in specific genes and disciplines to collaborate with them to diagnose, understand, and ultimately manage the rare disease of interest. Scientists also stand to benefit from such a system because collaborations demonstrating the relevance of their research to human disease can often increase the significance and impact of their ongoing research projects, which can in turn

help secure additional sources of funding for their laboratories. Furthermore, data obtained from studying rare variants identified in affected individuals can also provide scientists with new information about their genes of interest, expanding fundamental knowledge that has a broad impact beyond the specific affected individual and disease.

Finally, interdisciplinary collaborations between scientists are becoming more and more critical to make significant advances in the field of biomedical sciences. In contrast to several decades ago when single laboratory papers were commonly published in high-impact journals, collaborative manuscripts with contributions from multiple laboratories with a wide range of expertise now dominate these journals (Sekara et al., 2018). Collaborations between scientists are often established locally, or through personal networks that are built based on face-to-face meeting opportunities including scientific conferences. While social media and virtual scientific networks such as Google Scholar (<https://scholar.google.com/>) and ResearchGate (<https://www.researchgate.net/>) can also be utilized to identify potential collaborators, the vast amount of information on these sites makes identifying the right collaborator a daunting task. Facilitating collaborations between different types of scientists with interests in the same or orthologous genes in multiple species has been referred to as ‘vertical integration’ (https://orip.nih.gov/sites/default/files/Validation_Session_IX_Meeting_Report_Final_508.pdf). As an example of vertical integration, studies of orthologous genes in fruit flies, zebrafish, and human have provided complementary information to provide a diagnosis for rare disease patients (Barish et al., 2020; Ravenscroft et al., 2021). Furthermore, multidisciplinary studies by model organism biologists, cell biologists and biochemists have successfully classified the pathogenicity of variants in evolutionarily conserved genes in the context of both rare and common diseases (Post et al., 2020). While open services that include scientist-clinician matchmaking as part of their larger scope already exist such as GeneMatcher (<https://genematcher.org/>) (Sobreira et al. 2015), scientist-scientist matchmaking is rarely performed on such platforms.

After considering the above points carefully and discussing with many stakeholders, we concluded that a new platform was needed to address multiple gaps in the current biomedical collaboration ecosystem. To this end, we developed ModelMatcher (<https://www.modelmatcher.net/>) to address the unmet need of an open, international, decentralized network devoted to pursuing functional studies relevant to rare and undiagnosed diseases which is catered to all stakeholders (Table 1). The core of ModelMatcher is an open registry of scientists who input their genes of interest as well as their expertise in experimental and organ systems. Most information on this database is publicly searchable, but we also offer a matchmaking system that allows registered users to reach out to scientists with unpublished or confidential information. ModelMatcher is further connected to external scientist and clinical registries through the RDMM International and Matchmaker Exchange (MME) networks, respectively. These developments will expand the network paradigm for biomedical research by extending the collaboration ecosystem worldwide and including rare disease stakeholders, such as affected individuals, their family members and support groups, who are currently un- or under-represented in the current collaborative ecosystem.

Methods:

ModelMatcher is hosted on an AWS EC2 virtual machine running Ubuntu 18.04 and consists of four components: the landing/information page, the Scientist Registry, the Match Dashboard Registry, and MME API endpoints. The Scientist Registry operates a modified version of the RDMM server application (Boycott et al., 2020). The landing/information page, Match Dashboard Registry, and API endpoints operate as a single Node.js/Express application. Domain hosting and DNS services are provided by domain.com. A list of URLs referred to in this manuscript can be found in Supp. Table S1.

Landing/Information Page

The landing page of ModelMatcher is an HTML-based web page that is opened when a user visits our website using internet browsers (<https://www.modelmatcher.net/>). The page contains general information about the project (“About us” tab), a description of how the platform works (“Instructions” tab), answers to frequently asked questions (“FAQ” tab), contact form to reach the administrative team (“Contact” tab), a page with hyperlinks to our partner registries and related websites (“Links” tab), a link to page to register or log in as a Scientist or a Match Dashboard user (“Register/Login” tab), and a “Search” button to search information on the Scientist Register. The “Links” tab also contains a hyperlink to a webpage that allows users to download icons representing model organisms, organ systems and scientific categories that were developed for this project (<https://www.modelmatcher.net/downloads.html>). These icons can be freely used under Creative Commons license BY-NC 4.0. Hyperlinks to social media accounts are also provided, which provides tutorial videos (YouTube: https://www.youtube.com/channel/UC2o1FuMMZoa6oe_ekphd0sw) and updated news (Twitter: <https://twitter.com/modelmatchernet>; Facebook: <https://www.facebook.com/modelmatcher>).

Scientist Registry and the Search feature

The Scientist Registry of ModelMatcher is based on version 1.4.7 of the RDMM registry application hosted on GitHub (<https://github.com/pavlidisLab/rgr>). The program is primarily written in Java, with Spring and Thymeleaf providing the server functionality. The backend database service utilizes MySQL 8. Gene lists and metadata for each model organism were downloaded from the NCBI gene information database (<https://www.ncbi.nlm.nih.gov/gene>) and entrez gene IDs are used for internal identification purposes. Uberon is used for organ system metadata information (Mungall et al., 2012). Additional information can be found on the linked GitHub page shown above, or in the RDMM publication (Boycott et al., 2020).

We implemented the RDMM registry application for ModelMatcher with very limited changes. The only major modification was the implementation of a stricter ortholog candidate filtering criteria, necessitated by the lack of manual review in ModelMatcher compared to RDMM. We implemented a filter based on “DIOPT scores greater than 5” and “either best forward or reverse match” based on DIOPT version 8.0 (Hu et al., 2011). Some ortholog pairs with low or no DIOPT scores were manually entered. The latest ortholog candidate conversion table can be downloaded from our website (<https://>

www.modelmatcher.net/assets/homologs.csv). All other modifications were minor, involving either text or visual styling changes.

The Search function of ModelMatcher is based on the RDMM's initial implementation. Through the RDMM API, ModelMatcher users can discover scientists in partner registries that are part of the RDMM International Network (currently RDMM, Solve-RD and AFGN. See), and ModelMatcher scientists can also be discovered by users of our partner registries.

Match Dashboard Registry and the Match Request Form

The backend of the Match Dashboard Registry uses MongoDB for user registration and authentication and accesses the Scientist Registry MySQL database for matching and API queries. We use the Node.js Passport package to handle user authentication and login and the Nodemailer package to provide email server capabilities.

Match Dashboard is a user-based system that requires login to ensure security and allow users to maintain a long-term profile complete with genes of interest. Unlike the Scientist Registry, genes of interest stored in the Match Dashboard Registry are not shared with other users, unless the user actively submits a match request to MME. To identify a potential scientist collaborator, users enter their human genes of interest in the -Match with Scientists- tab. ModelMatcher converts human genes of interest entered by the Match Dashboard users to appropriate ortholog candidate using the conversion table as described above and access the Scientist Registry SQL database to query for matches. These matches are presented to the user in a table format. The user can either manually send an email based on the contact information displayed, or use the provided Match Request Form to send an email through the backend of ModelMatcher. User profile state is maintained so that match requests sent through the backend have the most vital information (e.g., name, contact email) automatically populated. The account-based structure also allows future development of additional features such as sending notification emails to users when a new scientist has registered a gene of interest.

Matchmaker Exchange API endpoints

The MME API endpoints use version 1.1 of the MME standard specification (<https://github.com/ga4gh/mme-apis>). For incoming queries, we use only the human gene symbol for matching purposes with our match strength score determined by DIOPT scores. We query the Scientist Registry for matches based on the previously described ortholog candidate conversion table. We return the corresponding information of our Scientist Registry users, and also provide a link to their profile information. If the gene of interest is set to the Restricted privacy category by the user, we provide more limited profile information (PI status, model system, model organism, organ system of interest) and a link to a contact form (similar to the Match Request Form) to email the Anonymous Scientist through the backend without needing to create a ModelMatcher account. Whenever we provide data to an external partner, we also notify our own users and provide them with the matching information we receive through the query. Match Dashboard users can also request matches to participants of partner MME registries (currently GeneMatcher and MyGene2) through the -Match with Clinicians and Patients- tab. When a match is initiated, their

profiles and case information are shared through the API. We notify any matches based on a standardized email template. The match history is also stored in their user profile and displayed on the -Match with Clinicians and Patients- tab.

Results:

The scientist registry serves as the starting point for all collaborations through ModelMatcher

The primary function of ModelMatcher is to provide users with a list of collaborative scientists who have expertise in specific human genes or their orthologs in genetic model organisms (Figure 1). ModelMatcher accomplishes this by maintaining an active Scientist Registry based on a software that was originally developed for the RDMM (Boycott et al., 2020). The information stored in this registry includes the scientist's name and contact information, affiliation, principal investigator (PI) status, research interests, relevant publications, types of experiments that are performed in their laboratories (e.g., *in vitro*, *in vivo*, *in silico*), organisms they use to conduct gene function studies, organ systems they have expertise in, and their genes of interest (Figure 2A). While the original RDMM software was primarily designed to recruit model organism scientists who conduct *in vivo* experiments, other types of scientists including cell biologists, molecular biologists, biochemists, structural biologists, and computational biologists are also encouraged to participate in ModelMatcher (Figure 2B, B'). If a scientist works with one of the ten supported model organisms [human (*Homo sapiens*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), frog (*Xenopus tropicalis* or *X. laevis*), zebrafish (*Danio rerio*), fruit fly (*Drosophila melanogaster*), nematode worm (*Caenorhabditis elegans*), budding yeast (*Saccharomyces cerevisiae*), fission yeast (*Schizosaccharomyces pombe*), or enterobacteria (*Escherichia coli*)], they can enter their genes of interest based on official gene symbols for each species (Figure 2C-D'). If they work with other species, scientists are encouraged to provide the gene symbol of the orthologous human gene, which could be identified based on information from databases such as Ensembl (<https://useast.ensembl.org/>) or homology alignment programs such as BLAST (<https://blast.ncbi.nlm.nih.gov/>). If a scientist works on multiple model organisms, this information is also captured in the Scientist Registry.

While most scientists are open to publicly declaring their genes of interest, some scientists with unpublished data may only wish to share such information with potential collaborators to avoid being scooped by their competitors. Therefore, we integrated a gene level privacy feature in which a scientist can classify their genes of interest into Public and Restricted categories on a case-by-case basis. Public genes will be searchable by anyone that has access to the open ModelMatcher website, whereas scientists themselves can control access to information regarding their Restricted genes. How these two types of information are used within ModelMatcher and shared with partner registries will be discussed in more detail in the following section. In addition, scientists are also asked to classify their genes of interest into Primary and Secondary priority tiers. Primary tier genes are genes that a scientist's lab actively works on, and has protocols and reagents ready to perform functional assays. In contrast, Secondary tier genes are genes that a scientist's lab has worked on in the past or has a strong interest in working on in the future, but would need additional time to establish

protocols or acquire/develop reagents to conduct experiments. Such information can be used to estimate how quickly a scientist may be able to generate data, which could become important when a quick turnaround of functional data is required.

The Search function allows anyone to query public data from ModelMatcher and partner scientist registries

The Scientist Registry of ModelMatcher can be queried by anyone using the Search feature that does not require user registration (<https://www.modelmatcher.net:8443/search>, Figure 3). Users of this feature can input a gene of interest and optionally narrow their results via placing filters on model organism, organ system, or gene tier if desired (Figure 3A). The default setting prompts the users to input a human gene of interest based on HGNC (HUGO Gene Nomenclature Committee, <https://www.genenames.org/>) nomenclature (Tweedie et al., 2021) and returns a list of scientists with expertise in this human gene or their ortholog candidates across all supported model organisms. When a search is performed using a human gene symbol, we provides hyperlinks to corresponding pages on NCBI Gene (<https://www.ncbi.nlm.nih.gov/gene>), MARRVEL (<http://marrvel.org/>), Alliance of Genome Resources (AGR, <https://www.alliancegenome.org/>), Gene2Function (G2F, <https://www.gene2function.org/>) and Monarch Initiative (<https://monarchinitiative.org/>) so that the users can gather additional information about their genes of interest and related genes. Ortholog candidates of human genes are primarily identified based on the DIOPT (DRSC Integrative Ortholog Prediction Tool, https://www.flyrnai.org/cgi-bin/DRSC_orthologs.pl), an integrative ortholog prediction tool that currently uses 18 different algorithms and provides a score that reflects the likelihood of certain gene pairs to be orthologous to one another (Hu et al., 2011) (Figure 3B). Some gene pairs that are true orthologs with low DIOPT scores have been manually included, in addition to a list of orthologs for model organisms that are not currently supported by DIOPT [e.g., *E coli*]. As improved lists of true orthologs become available through community efforts, such as the Alliance of Genome Resources (Agapite et al., 2020) and Quest for Orthologs (Altenhoff et al., 2021), we plan to adopt such datasets to improve the accuracy of ortholog identification.

Basic information for each scientist identified as having expertise in the specific human gene or their ortholog candidates will be displayed in a simple search output table (Figure 3C). In addition to performing searches based on gene symbols, searches can also be performed based on specific terms (Figure 3A'). When user searches ModelMatcher using specific medical (e.g. cancer, diabetes, Alzheimer disease) or scientific terms (e.g. transcription, ion channel, glia), a list of scientists who included these terms in their profile pages will be generated and displayed. Users can further navigate to the scientist's profile page to find additional information about their research in order to identify the ideal collaborator candidate. Contact information is listed under each user's profile page, which can be used to initiate a conversation via email. Some scientists may be listed more than once if they work on orthologous genes in multiple species. When there is a scientist who has classified the gene of interest in their Restricted category, such information will be displayed as information from an Anonymous Scientist. In order to reach out to an Anonymous Scientist, users will need to create a Match Dashboard account, which is discussed below.

In addition to querying the Scientist Registry of ModelMatcher, our Search feature allows users to query the list of scientists within the RDMM International Network (Figure 1, <https://rdmmglobal.org/>). This network consists of multiple scientist registries which have been built using the same RDMM software. These registries communicate with each other via APIs (application programming interfaces) in real time when a query is made. Currently, information that is classified as being public from RDMM, Solve-RD, and AFGN is displayed below the table of ModelMatcher scientists in a similar manner (Figure 3D). By clicking the “View User Profile” button for each entry, users will be able to access information about scientists in partner registries and reach out to them through the contact information listed there. Therefore, in addition to identifying scientists who registered their information in ModelMatcher, users can identify potential collaborators in Canada, Europe, and Australia who are already part of a regional rare disease research network. Currently (as of 1/31/2022), 205 scientists have registered a total of 971 unique model organism genes that correspond to 1,053 human genes (5.5% of protein coding genes in the human genome) in ModelMatcher. When we combine this with searchable information available through our partner registries of the RDMM International, ModelMatcher users have immediate access to information from 495 scientists who have registered a total of 5,383 unique model organism genes that correspond to 6,018 unique human genes (31.3% of all protein coding genes in the human genome) (Supp. Table S2).

Match Dashboard provides additional features to query the Scientist Registry and initiate communication

In addition to the Search feature that allows anyone to query the ModelMatcher and partner registries as discussed above (Figure 3), we built the Match Dashboard feature that provides additional benefits to registered users (Figure 4–5). A Match Dashboard account, which is independent of a Scientist Registry account, can be created by anyone who wishes to collaborate with scientists through a separate registration webpage (<https://www.modelmatcher.net/login.html>). Upon registration, users gain access to their Match Dashboard, which is a personalized webpage that facilitates communication with potential collaborators. In addition, the Match Dashboard can be used to identify potential clinical collaborators through MME, which is discussed in the following section.

In their profile page, Match Dashboard users input their information including their user category (e.g., patient/family member, clinician, scientist, industry, etc.), affiliation (if any), contact information, and a general statement regarding their interest in rare and undiagnosed diseases (Figure 4A-B). Once this is complete, the user will then navigate either to the -Match with Scientists- (Figure 4C) or -Match with Clinicians and Patients- (Figures 4D, 6) tabs. When the user clicks the -Match with Scientists- tab and enters the human genes they are interested in based on HGNC nomenclature, ModelMatcher generates a table of scientists who registered these human genes or their ortholog candidates from our Scientist Registry (Figures 4C', 5A). This display in the Match Dashboard is similar to what is returned via the public Search function, although it provides additional features. First, the genes of interest list for each user is stored in our Match Dashboard registry, which is kept separately from the Scientist Registry, and queried against the latest Scientist Registry in real time. Therefore, whenever a user logs into their Match Dashboard account, they will be able

to identify all of the matching scientists for their genes of interest without having to perform individual searches. This is especially useful if a user has multiple genes of interest that they would like to keep an eye on longitudinally, considering that additional scientists with expertise in the gene of interest may appear in the future. Second, the user will have access to the Match Request Form, which is a messaging system that facilitates the communication process. Upon identification of a scientist whom the user wishes to contact based on the scientist's profile (Figure 5B), the user can open the Match Request Form by clicking the "Send Match Request" button that is shown for each scientist (Figure 5C, arrow). A new tab will open that contains a contact form addressed towards the specific scientist regarding the specific gene (Figure 5C'). The user may add a specific reason for which they want to contact this scientist if this is not documented in the general user profile, and click the "Send" button to send a message. ModelMatcher will then generate a standardized email that will be sent to the scientist. The user's profile from Match Dashboard is automatically included in this email, which saves the user from entering the same information every time they contact a new scientist. More importantly, Match Request Forms can also be used to contact Anonymous Scientists who classified specific genes as Restricted and whose contact information is not publicly displayed (Figure 5D, 5D'). When a message is sent to the scientist, the system logs the last contact date and displays this information on the Match Dashboard (Figures 5A), providing the user with an easy way to track which scientists they contacted, and how long ago that attempt to contact took place. In summary, the -Match with Scientists- feature of Match Dashboard allows users to reach out to scientists regardless of the gene privacy level using a standardized email template, and provides a tracking system to log all of the communication attempts.

Integration of ModelMatcher into the Matchmaker Exchange network allows our users to discover clinical collaborators and vice versa

While the Search and Match Dashboard features allow ModelMatcher users to discover scientists with expertise in certain genes and experimental/organ systems, they do not allow our scientists to effectively discover clinical collaborators since they need to passively wait to be contacted by a clinician. In order to allow scientists to actively match with clinicians who have access to affected individuals with rare, potentially pathogenic variants in their genes of interest, ModelMatcher became part of MME (<https://www.matchmakerexchange.org/>), a large network of federated databases for rare disease gene discovery (Philippakis et al., 2015). So far, ModelMatcher has developed partnerships with three nodes of MME: GeneMatcher, MyGene2 and PhenomeCentral. GeneMatcher (<https://genematcher.org/>) is a matchmaking platform provided by the Baylor-Hopkins Center of Mendelian Genomics (CMG) in the United States (Sobreira et al. 2015). As of September 2021, more than 10,000 users, most of whom are clinicians or clinical researchers, from >90 countries have submitted >50,000 match requests corresponding to >13,000 human genes. MyGene2 (<https://mygene2.org/>) is a patient-centric matchmaking platform hosted by the University of Washington CMG in the United States (Chong et al., 2016). As of September 2021, more than 4,000 families with affected individuals with rare diseases have joined the network, many of whom are sharing their profiles with users from other MME nodes. PhenomeCentral (<https://www.phenomecentral.org/>) is a service developed in Canada that integrates both genotype and phenotype information to match

clinicians with rare disease patients (Buske et al., 2015). As of September 2021, over 12,000 cases have been entered into this registry from research initiatives such as FORGE (Finding of Rare Disease Genes in Canada) (Beaulieu et al., 2014), Care4Rare (Care for Rare Canada) (<http://care4rare.ca/>), UDN (Brownstein et al., 2015) and UDN International (Taruscio et al., 2020). Using standardized MME APIs that allow the connection of ModelMatcher to these and other collaboration focused network nodes of MME, users of ModelMatcher can identify clinicians and affected individuals who share interest in the same or orthologous genes and vice versa.

ModelMatcher users can identify clinicians and affected individuals from other MME nodes using their Match Dashboard account. Currently, this feature is supported to actively identify potential collaborators in GeneMatcher and MyGene2 but not for PhenomeCentral based on the different conditions each nodes has regarding data sharing. By creating a Match Dashboard account as described above (Figure 4A), both scientists and non-scientists can take advantage of this feature. Once the user logs in to their Match Dashboard page, they can navigate to the -Match with Clinicians and Patients- tab to access this feature (Figure 6). From here, the user enters their human gene of interest and send Match Requests to other parties. The genes entered in the -Match with Scientists- tab will be automatically loaded into the -Match with Clinicians and Patients- so that the list of genes of interest will be synchronized within the Match Dashboard. When the user clicks the “Exchange Info” button shown for each gene, they will be taken to a Match Request Form that allows them to match with clinicians or other stakeholders interested in this gene from other MME nodes (Figure 6B). Users will be given the option to opt-out of matching with a specific registry if they feel that it is not applicable to their use case. When an MME match request is executed by clicking the “Send” button (Figure 6C), the user’s profile and additional information provided for the specific gene will be sent to the MME partner registries, which will then return information regarding their matching cases to ModelMatcher. We provide this match data to our users in two ways. First, users will receive an email detailing the case information of any external matches they have received. Even if the match results are negative, they will receive an email documenting this information, together with a recommendation to repeat the MME Match Request at a later time point (e.g., after ~6 months). Second, a history of Match Requests for each gene is stored and displayed on the Match Dashboard (Figure 6C’), allowing users to access the information of previous matches without having to sort through their emails. This prevents the users from performing multiple redundant Match Requests in a short time frame, and allows them to keep track of previous matches for their genes of interest.

Reciprocally, MME nodes can send Match Requests to ModelMatcher if their users wish to identify scientists with expertise in certain human genes or their ortholog candidates. Currently, such requests can be initiated by users of GeneMatcher, MyGene2 and PhenomeCentral using the specific features provided on each website. ModelMatcher uses the information stored within the Scientist Registry to identify matches and provides information to both the requester and the matching scientist. When users from other registries send Match Requests through the MME, they are typically given an option to opt-in to match with a ModelMatcher scientist to discuss collaboration opportunities. Using the aforementioned ortholog conversion standard, ModelMatcher converts human gene queries

from MME into the corresponding model organism genes to identify matching scientists. Depending on the privacy level set for the specific gene, we send back the full user profile or an anonymized contact form (similar to a Match Request Form to an Anonymous Scientist) to the MME partner registries. Users in our Scientist Registry will also immediately receive an email if they appear in an external Match Request, providing them with the information on the originating registry, relevant case information, and contact information of the other party to help facilitate a decision on potential collaborations.

Discussion:

Implementation of next generation sequencing technologies into clinical care over the past decade has revolutionized rare disease diagnosis. However, as clinicians and researchers sequence more affected individuals, they find more and more VUS in known disease-causing genes, or rare potentially damaging variants in GUS. These cases need functional assessment through experimental strategies to evaluate the pathogenicity of the disease-associated genetic variants. This necessitates the creation of collaborative research platforms that can connect affected individuals, clinicians, and scientists. While multiple efforts to facilitate experimental studies of genetic variants identified in rare disease patients have been launched in different parts of the world with great success, many scientists with expertise in specific genes and biological processes who can make significant contributions do not have the opportunity to contribute. This often happens because the scientists are not part of the collaborative consortia or networks due to several barriers to entry. ModelMatcher is a network designed to place individual scientists at the center of our registry. We specifically designed this platform so clinicians, affected individuals, and other stakeholders of rare and undiagnosed disease research can easily access scientists, who may even reside in other countries, and initiate the collaborations needed to solve medical mysteries. Reciprocally, by connecting ModelMatcher to MME, we provide a way for scientists to reach out to clinicians and affected individuals, which can help increase the clinical significance of their research projects. Through these efforts, we hope that more researchers working in pure-scientific laboratory settings will gain better access and stronger connections to the medical community. The integration of ModelMatcher to the larger RDMM International Network further increases the number of scientists who can participate and contribute to this ecosystem. This will have long-term beneficial effects by both bolstering rare disease research and advancing the fields of basic or fundamental biology.

There is well established data that shows how beneficial an accurate diagnosis is to the mental health of affected individuals and family members, regardless of how it might change patient management (McConkie-Rosell et al., 2018). Additionally, there are cases in which identifying the affected gene has led to new treatment strategies being implemented. Over the years, many patient and patient support groups have also formed as people with a diversity of undiagnosed diseases receive genetic diagnoses and reach out to other families who received a similar diagnosis. Family conferences hosted by these groups allow for a more holistic view of symptoms across affected individuals, which can be variably expressed and incompletely penetrant, to deepen our understanding of the newly documented condition. In addition, diagnosis opens up the opportunity for coalition building around a certain genetic condition, forming a community that can support one another

and grow as a group as more affected individuals become recognized with the same or related disorders. This, combined with group lobbying and fund-raising efforts, can facilitate resource allocation to support biomedical research on the genes and diseases of interest, increasing biological knowledge that can be used to design effective treatments and therapy. ModelMatcher can facilitate this effort by matching patients', clinicians', and scientists' interests in the same or orthologous genes to make appropriate connections in order to boost the speed of translational and preclinical research.

Clinicians, patients, and support organizations are not the only groups to benefit from connecting with scientists. There are several stakeholders that are not often considered in the current collaborative matchmaking paradigm. Government agencies, non-profit organizations, biotech and pharmaceutical companies, and venture capital funds all desire to make informed decisions about who to contract for specific projects, and where to allocate their funds. A common strategy for these groups is to mine publications to find scientists and clinicians working on specific genes or pathways related to their project using online search tools such as PubMed (<https://pubmed.ncbi.nlm.nih.gov/>). This is suboptimal for two reasons: scientists who have worked on a gene in the past do not necessarily continue to work on it, and there could be scientists currently studying a gene who have not yet published their findings (Table 1). Having a central, up-to-date database like ModelMatcher will streamline collaborator identification and fund allocation to the appropriate scientists to accelerate specific projects.

While collaborations can bring tremendous benefits to biomedical research and clinical care, there are pitfalls that must be carefully avoided when working with others. Those unfamiliar with collaboration, especially across disciplines, should be mindful of expectations; both the expectations they have of others and the expectations others place on them. This is especially true for any users coming into ModelMatcher, as this service is designed solely to facilitate collaboration initiation, but does not provide guidance to maintain and expand collaborations once they are established. Collaborations between scientists and clinicians will not universally lead to diagnosis or treatment due to several factors. Even when they do, the timeline for these developments can be years long. All parties entering a collaboration should discuss realistic timelines, expectations, and what information will need to be shared prior to embarking on a project. All parties should be responsible for analyzing and vetting clinical and scientific information provided by other groups to ensure the project stands on solid foundations. Communication amongst members of the collaboration should be honest, open, and frequent. Considering that many ModelMatcher users are likely to be collaborating with people with different backgrounds and interests for the first time, it is crucial that everyone puts a significant amount of effort and energy into the project so the collaboration goes smoothly. Each participant must remember that a bad collaboration experience may not only ruin the current project but could also poison the collaborative atmosphere around them and make others hesitant to take part in future cross-disciplinary collaborations. Finally, it is important to note that, unlike other RDMM-like networks that can directly provide funding to scientists, ModelMatcher does not provide financial support to kickstart a collaborative research project. Therefore, each research group must identify financial support to conduct preliminary experiments. Once some positive data are obtained from pilot experiments, scientists can leverage this to apply for additional funding

from public and private funding organizations such as the NIH and individual patient organizations.

Finally in addition to its impact on the rare and undiagnosed research community, ModelMatcher can be used to facilitate research on common diseases. WES and WGS are being applied to identify genetic causes or risk factors of a variety of common diseases such as type 2 diabetes (Xue et al., 2018), autism spectrum disorders (Grove et al., 2019) and Alzheimer disease (Jansen et al., 2019). While these disorders tend to be more complicated than rare diseases because multiple genetic and environmental factors often contribute to disease risk and expressivity, functional studies of disease associated variants are often desired or required. Moreover, interpretation of somatic mutations identified in cancer patients can also be facilitated by functional studies performed by scientists, including model organism biologists who can perform gene and protein specific assays *in vivo* (Ablain et al., 2018; Bangi et al., 2019). Importantly, studies of rare diseases can provide important biological insights into related common diseases since some rare and common diseases converge on the same or similar pathogenic mechanisms. Finally, individuals with rare genetic diseases often provide valuable knowledge about *in vivo* gene function in human, which cannot be obtained through model organism research or through *in vitro* or *ex vivo* studies performed in test tubes. The informatic resource available through ModelMatcher has numerous applications, which we hope will benefit the larger biomedical research community to improve healthcare and advance science for mankind.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

The ModelMatcher project is supported by the Jan and Dan Duncan Neurological Institute at Texas Children's Hospital, a grant from the National Institutes of Health (NIH, U54NS093793) to M.F.W., H.J.B, Z.L. and S.Y., and support from the Hamill Foundation to L.L.. The Undiagnosed Diseases Network (UDN) is supported by the NIH Common Fund. H.J.B received support from the Howard Hughes Medical Institute. The Rare Diseases Models and Mechanisms (RDMM) Network is funded by the Canadian Institutes of Health Research (RCN-137793 and RCN-160422) with additional support from Genome Canada and Genome British Columbia to K.M.B. and P.H.. We thank the leadership and technology teams of Solve-RD, Australian Functional Genomics Network, GeneMatcher, MyGene2 and PhenomeCentral for data sharing and technical advice. We thank Dr. Huda Y. Zoghbi and the leadership teams of Alliance of Genome Resources (AGR), UDN Participant Engagement and Empowerment Resource (UDN PEER), and Matchmaker Exchange for valuable advice and useful discussions. We also acknowledge individual scientists for their active participation in ModelMatcher and the RDMM International Network, and clinicians, affected individuals, and family members for their valuable contributions to research initiatives that are part of the Matchmaker Exchange. A list of members of the Undiagnosed Diseases Network can be found in Supp. Table S3. The authors have no conflict of interest to declare.

References Cited:

- Ablain J, Xu M, Rothschild H, Jordan RC, Mito JK, Daniels BH, Bell CF, Joseph NM, Wu H, Bastian BC, Zon LI, & Yeh I (2018). Human tumor genomics and zebrafish modeling identify SPRED1 loss as a driver of mucosal melanoma. *Science*, 362(6418), 1055–1060. 10.1126/science.aau6509 [PubMed: 30385465]
- Agapite J, Albou LP, Aleksander S, Argasinska J, Arnaboldi V, Attrill H, Bello SM, Blake JA, Blodgett O, Bradford YM, Bult CJ, Cain S, Calvi BR, Carbon S, Chan J, Chen WJ, Cherry JM, Cho J, Christie KR, ... Yook K (2020). Alliance of Genome Resources Portal: Unified model organism

research platform. *Nucleic Acids Research*, 48(D1), D650–D658. 10.1093/nar/gkz813 [PubMed: 31552413]

- Altenhoff AM, Garrayo-Ventas J, Cosentino S, Emms D, Glover NM, Hernández-Plaza A, Nevers Y, Sundesha V, Szklarczyk D, Fernández JM, Codó L, Li Gelpi J, Huerta-Cepas J, Iwasaki W, Kelly S, Lecompte O, Muffato M, Martin MJ, Capella-Gutierrez S, ... Dessimoz C (2021). The Quest for Orthologs benchmark service and consensus calls in 2020. *Nucleic Acids Research*, 48(W1), W538–W545. 10.1093/NAR/GKAA308
- Baldrige D, Wangler MF, Bowman AN, Yamamoto S, Schedl T, Pak SC, Postlethwait JH, Shin J, Solnica-Krezel L, Bellen HJ, & Westerfield M (2021). Model organisms contribute to diagnosis and discovery in the undiagnosed diseases network: Current state and a future vision. *Orphanet Journal of Rare Diseases*, 16(1), 206. 10.1186/s13023-021-01839-9 [PubMed: 33962631]
- Bangi E, Ang C, Smibert P, Uzilov AV, Teague AG, Antipin Y, Chen R, Hecht C, Gruszczynski N, Yon WJ, Malyshev D, Laspina D, Selkridge I, Rainey H, Moe AS, Lau CY, Taik P, Wilck E, Bhardwaj A, ... Cagan RL (2019). A personalized platform identifies trametinib plus zoledronate for a patient with KRAS-mutant metastatic colorectal cancer. *Science Advances*, 5(5), eaav6528. 10.1126/sciadv.aav6528 [PubMed: 31131321]
- Barish S, Barakat TS, Michel BC, Mashtalir N, Phillips JB, Valencia AM, Ugur B, Wegner J, Scott TM, Bostwick B, Murdock DR, Dai H, Perenthaler E, Nikoncuk A, van Slegtenhorst M, Brooks AS, Keren B, Nava C, Mignot C, ... Bellen HJ (2020). BICRA, a SWI/SNF Complex Member, Is Associated with BAF-Disorder Related Phenotypes in Humans and Model Organisms. *American Journal of Human Genetics*, 107(6), 1096–1112. 10.1016/j.ajhg.2020.11.003 [PubMed: 33232675]
- Beaulieu CL, Majewski J, Schwartzentruber J, Samuels ME, Fernandez BA, Bernier FP, Brudno M, Knoppers B, Marcadier J, Dymont D, Adam S, Bulman DE, Jones SJM, Avard D, Nguyen MT, Rousseau F, Marshall C, Wintle RF, Shen Y, ... Boycott KM (2014). FORGE Canada consortium: Outcomes of a 2-year national rare-disease gene-discovery project. *American Journal of Human Genetics*, 94(6), 809–817. 10.1016/j.ajhg.2014.05.003 [PubMed: 24906018]
- Boycott KM, Campeau PM, Howley HE, Pavlidis P, Rogic S, Oriel C, Berman JN, Hamilton RM, Hicks GG, Lipshitz HD, Masson JY, Shoubridge EA, Junker A, Leroux MR, McMaster CR, Michaud JL, Turvey SE, Dymont D, Innes AM, ... Hieter P (2020). The Canadian Rare Diseases Models and Mechanisms (RDMM) Network: Connecting Understudied Genes to Model Organisms. *American Journal of Human Genetics*, 106(2), 143–152. 10.1016/j.ajhg.2020.01.009 [PubMed: 32032513]
- Brownstein CA, Holm IA, Ramoni R, Goldstein DB, Bacino C, Lee B, Eng C, Veeraghavan N, Bernick D, Brownstein C, Holm I, Kohane I, McCray A, Splinter K, Shashi V, Goldstein D, Beggs A, Loscalzo J, MacRae C, ... Phillips J (2015). Data Sharing in the Undiagnosed Diseases Network. *Human Mutation*, 36(10), 985–988. 10.1002/humu.22840 [PubMed: 26220576]
- Buske OJ, Girdea M, Dumitriu S, Gallinger B, Hartley T, Trang H, Misyura A, Friedman T, Beaulieu C, Bone WP, Links AE, Washington NL, Haendel MA, Robinson PN, Boerkoel CF, Adams D, Gahl WA, Boycott KM, & Brudno M (2015). PhenomeCentral: A Portal for Phenotypic and Genotypic Matchmaking of Patients with Rare Genetic Diseases. *Human Mutation*, 36(10), 931–940. 10.1002/humu.22851 [PubMed: 26251998]
- 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. *Genetics in Medicine*, 18(8), 788–795. 10.1038/gim.2015.161 [PubMed: 26656649]
- Coventry A, Bull-Otterson LM, Liu X, Clark AG, Maxwell TJ, Crosby J, Hixson JE, Rea TJ, Muzny DM, Lewis LR, Wheeler DA, Sabo A, Lusk C, Weiss KG, Akbar H, Cree A, Hawes AC, Newsham I, Varghese RT, ... Sing CF (2010). Deep resequencing reveals excess rare recent variants consistent with explosive population growth. *Nature Communications*, 1, 131. 10.1038/ncomms1130
- Gahl WA, Mulvihill JJ, Toro C, Markello TC, Wise AL, Ramoni RB, Adams DR, & Tiftt CJ (2016). The NIH Undiagnosed Diseases Program and Network: Applications to modern medicine. *Molecular Genetics and Metabolism*, 117(4), 393–400. 10.1016/j.ymgme.2016.01.007 [PubMed: 26846157]

- Ghosh R, Oak N, & Plon SE (2017). Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biology*, 18(1), 225. 10.1186/s13059-017-1353-5 [PubMed: 29179779]
- Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, Pallesen J, Agerbo E, Andreassen OA, Anney R, Awasthi S, Belliveau R, Bettella F, Buxbaum JD, Bybjerg-Grauholm J, Bækvad-Hansen M, Cerrato F, Chambert K, Christensen JH, ... Børglum AD (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics*, 51(3), 431–444. 10.1038/s41588-019-0344-8 [PubMed: 30804558]
- Hu Y, Flockhart I, Vinayagam A, Bergwitz C, Berger B, Perrimon N, & Mohr SE (2011). An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics*, 12, 357. 10.1186/1471-2105-12-357 [PubMed: 21880147]
- Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, Sealock J, Karlsson IK, Hägg S, Athanasiu L, Voyle N, Proitsi P, Witoelar A, Stringer S, Aarsland D, Almdahl IS, Andersen F, Bergh S, Bettella F, ... Posthuma D (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature Genetics*, 51(3), 404–413. 10.1038/s41588-018-0311-9 [PubMed: 30617256]
- Lupski JR, Belmont JW, Boerwinkle E, & Gibbs RA (2011). Clan genomics and the complex architecture of human disease. *Cell*, 147(1), 32–43. 10.1016/j.cell.2011.09.008 [PubMed: 21962505]
- McConkie-Rosell A, Hooper SR, Pena LDM, Schoch K, Spillmann RC, Jiang YH, Cope H, Palmer C, & Shashi V (2018). Psychosocial Profiles of Parents of Children with Undiagnosed Diseases: Managing Well or Just Managing? *Journal of Genetic Counseling*, 27(4), 935–946. 10.1007/s10897-017-0193-5 [PubMed: 29297108]
- Michael Harnish J, Deal SL, Chao HT, Wangler MF, & Yamamoto S (2019). In vivo functional study of disease-associated rare human variants using drosophila. *Journal of Visualized Experiments*, (150), e59658. 10.3791/59658
- Mungall CJ, Torniai C, Gkoutos GV, Lewis SE, & Haendel MA (2012). Uberon, an integrative multi-species anatomy ontology. *Genome Biology*, 13(1), R5. 10.1186/gb-2012-13-1-r5 [PubMed: 22293552]
- Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, Murphy D, Le Cam Y, & Rath A (2020). Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *European Journal of Human Genetics*, 28(2), 165–173. 10.1038/s41431-019-0508-0 [PubMed: 31527858]
- Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, Brudno M, Brunner HG, Buske OJ, Carey K, Doll C, Dumitriu S, Dyke SOM, den Dunnen JT, Firth HV, Gibbs RA, Girdea M, Gonzalez M, Haendel MA, Hamosh A, ... Rehm HL (2015). The Matchmaker Exchange: A Platform for Rare Disease Gene Discovery. *Human Mutation*, 36(10), 915–921. 10.1002/humu.22858 [PubMed: 26295439]
- Post KL, Belmadani M, Ganguly P, Meili F, Dingwall R, McDiarmid TA, Meyers WM, Herrington C, Young BP, Callaghan DB, Rogic S, Edwards M, Niciforovic A, Cau A, Rankin CH, O'Connor TP, Bamji SX, Loewen CJR, Allan DW, ... Haas K (2020). Multi-model functionalization of disease-associated PTEN missense mutations identifies multiple molecular mechanisms underlying protein dysfunction. *Nature Communications*, 11(1), 2073. 10.1038/s41467-020-15943-0
- Ravenscroft TA, Phillips JB, Fieg E, Bajikar SS, Peirce J, Wegner J, Luna AA, Fox EJ, Yan YL, Rosenfeld JA, Zirin J, Kanca O, Acosta MT, Adam M, Adams DR, Agrawal PB, Alejandro ME, Alvey J, Amendola L, ... Bellen HJ (2021). Heterozygous loss-of-function variants significantly expand the phenotypes associated with loss of GDF11. *Genetics in Medicine Online ahead of print*. 10.1038/s41436-021-01216-8
- Sekara V, Deville P, Ahnert SE, Barabási AL, Sinatra R, & Lehmann S (2018). The chaperone effect in scientific publishing. *Proceedings of the National Academy of Sciences of the United States of America*, 115(50), 12603–12607. 10.1073/pnas.1800471115 [PubMed: 30530676]
- 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. 10.1002/humu.22769 [PubMed: 25684268]

- Sobreira N, Schiettecatte F, Valle D, & Hamosh A (2015). GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene. *Human Mutation*, 36(10), 928–930. 10.1002/humu.22844 [PubMed: 26220891]
- Taruscio D, Baynam G, Cederroth H, Groft SC, Klee EW, Kosaki K, Lasko P, Melegh B, Riess O, Salvatore M, & Gahl WA (2020). The Undiagnosed Diseases Network International: Five years and more! *Molecular Genetics and Metabolism*, 129(4), 243–254. 10.1016/j.ymgme.2020.01.004 [PubMed: 32033911]
- Tweedie S, Braschi B, Gray K, Jones TEM, Seal RL, Yates B, & Bruford EA (2021). [Genenames.org](https://www.genenames.org/): The HGNC and VGNC resources in 2021. *Nucleic Acids Research*, 49(D1), D939–D946. 10.1093/nar/gkaa980 [PubMed: 33152070]
- Wangler MF, Yamamoto S, Chao HT, Posey JE, Westerfield M, Postlethwait J, Hieter P, Boycott KM, Campeau PM, Bellen HJ, Adams DR, Hanchard NA, Papp JC, Alejandro ME, Handley LH, Parker NH, Allard P, Herzog MR, Pena LDM, ... Palmer CG (2017). Model organisms facilitate rare disease diagnosis and therapeutic research. *Genetics*, 207(1), 9–27. 10.1534/genetics.117.203067 [PubMed: 28874452]
- Xue A, Wu Y, Zhu Z, Zhang F, Kemper KE, Zheng Z, Yengo L, Lloyd-Jones LR, Sidorenko J, Wu Y, Agbessi M, Ahsan H, Alves I, Andiappan A, Awadalla P, Battle A, Beutner F, Bonder MJJ, Boomsma D, ... Yang J (2018). Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nature Communications*, 9(1), 2941. 10.1038/s41467-018-04951-w

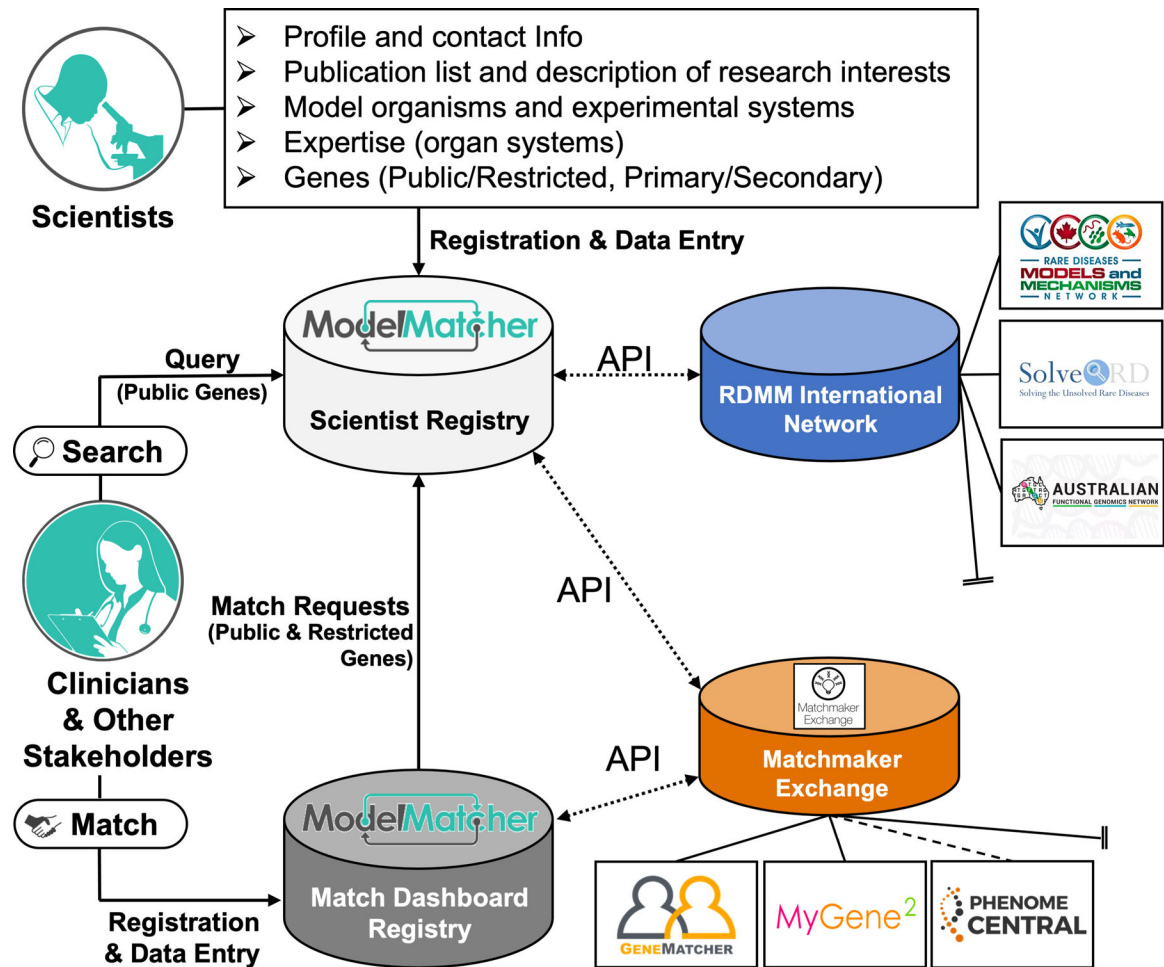


Figure 1. Basic structure of ModelMatcher and its relationship to partner registries.

Scientists who wish to identify potential collaborators and sponsors register their information into the Scientist Registry of ModelMatcher. In addition to basic information about their interests and expertise, scientists enter their genes of interest, which are classified into Public or Restricted privacy categories and Primary or Secondary priority tiers (see text for additional information about gene classification). Clinicians and other stakeholders of rare and undiagnosed disease research can mine this data in two ways. The Search feature allows anyone to query the information associated with genes classified as Public without needing to register in ModelMatcher. In addition to identifying scientists in the ModelMatcher Scientist Registry, Search feature users will also receive information about scientists in partner scientist registries that are part of the RDMM International Network including RDMM, Solve-RD, and AFGN. The Match Request feature, which requires users to register for a Match Dashboard account, provides a unique mechanism for users to contact a scientist who not only classified a gene as Public but also Restricted. The Match Dashboard account also enables users to track their matches, provides a communication platform to send messages to scientists, and further provides an entry point to connect with clinicians and affected individuals who participate in clinical registries that are part of Matchmaker Exchange including GeneMatcher, MyGene2 and PhenomeCentral.

scientists can enter information about their specific research topic using each organism and generate a list of genes they are willing to collaborate on using the official gene symbol that is specific to each species. Scientists can categorize each gene into Primary or Secondary priority tiers based on whether they are actively working on the specific gene or not. Furthermore, scientists can classify genes into Public or Restricted privacy categories based on confidentiality levels. Public genes will be associated with the scientist in both the Search and the Match Dashboard features. Restricted genes will show up as a gene that is entered by an Anonymous Scientists, but will not divulge the information of the individual who registered it. The identity of the Anonymous Scientist will only be revealed if the scientist wishes to disclose the information upon receiving a Match Request email from individual Match Dashboard or Matchmaker Exchange users. This allows scientists to identify potential collaborators on unpublished projects that consider as confidential due to potential competitors in their field.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

ModelMatcher Login & Registration

Search Genes, Models & Organs Systems

Genes **Terms** **A'**

Search for researchers that study the following gene or ortholog candidates: Narrow your search by tier (optional):

In this model organism:

Narrow your search by organ system (optional)

Additional info on human **NOTCH1** and its homologs:

MARRVEL²

Showing the results for the specified human gene and the following ortholog candidates: *Mus musculus* (*Notch1*), *Drosophila melanogaster* (*N*), *Xenopus tropicalis* (*notch1*), *Caenorhabditis elegans* (*lin-12*, *glp-1*), *Danio rerio* (*notch1a*), *Rattus norvegicus* (*Notch1*), from the ModelMatcher registry. Displaying results from our partner registries (RDMM, Solve-RD, AFGN) as well, which may use slightly different ortholog prediction algorithms.

Researchers registered in ModelMatcher

	Full Name	Principal Investigator	Organization	Gene	Tier	Organism
<input type="button" value="View User Profile"/>	Yamamoto, Shinya	Yes	Baylor College of Medicine	NOTCH1	Primary	Human/other
<input type="button" value="View User Profile"/>	Yamamoto, Shinya	Yes	Baylor College of Medicine	N	Secondary	Fruit Fly
<input type="button" value="Restricted Gene"/>	Anonymous Scientist	?		notch1	Primary	Frog

Partner registry results

	Full Name	Organization	Gene	Tier	Organism
<input type="button" value="View User Profile"/>	Viemincx, Kris	Ghent University	NOTCH1	Secondary	Human
<input type="button" value="View User Profile"/>	Lopes, Susana	CEDOC - Centro de Estudos de Doenças Crónicas NOVA Medical School / Faculdade de Ciências Médicas Universidade Nova de Lisboa	notch1a	Secondary	Zebrafish
<input type="button" value="View User Profile"/>	Samuel, Buck	Baylor College of Medicine	glp-1	Secondary	Roundworm
<input type="button" value="View User Profile"/>	Boulin, Thomas	Institut NeuroMyoGène - CNRS UMR5310 - INSERM U1217 - Université de Lyon	lin-12	Primary	Roundworm
<input type="button" value="View User Profile"/>	vaccari, thomas	University of Milan	N	Primary	Fruit Fly
<input type="button" value="View User Profile"/>	Verheyen, Esther	Simon Fraser University	N	Primary	Fruit Fly
<input type="button" value="View User Profile"/>	Boulianne, Gabrielle	The Hospital for Sick Children & University of Toronto	N	Primary	Fruit Fly

Figure 3. Search feature identifies scientists registered in ModelMatcher and partner registries. The Search page outputs information on scientists who work on specific human genes or its predicted orthologs in model organisms. **A)** A ‘Gene’ search query is typically initiated by entering a human gene symbol according to the official nomenclature, although one can initiate a search based on model organism genes if desired. If necessary, searches can be refined by which model organism is being used, whether the gene is classified as Primary or Secondary priority tiers, and in which organ system the scientist specializes. **A’)** By clicking the ‘Terms’ link (blue arrow), one can perform a search based on a keyword to identify a user based on specific terms used in their user profile. **B)** Upon clicking the “Search” button in **A)**, a list of genes considered ortholog candidates of the human gene of interest (human *NOTCH1*, in this case) in the ten supported organisms is displayed. When a user clicks a gene symbol listed here, they will be navigated to the specific NCBI Gene page (<https://www.ncbi.nlm.nih.gov/gene>) corresponding to each human or model organism

gene. Hyperlinks to corresponding pages of MARRVEL (<http://marrvel.org>), Alliance of Genome Resources (AGR, <https://www.alliancegenome.org>), Gene2Function (G2F, <https://www.gene2function.org>) and Monarch Initiative (<https://monarchinitiative.org>) are provided here for the users to obtain additional information related to their gene of interest. **C)** When a Search is performed, a table of ModelMatcher scientists who registered the human gene of interest or its ortholog candidates is generated. If the gene is classified as Public, the name of the scientist, their affiliation, and whether they are PIs or not will be displayed along with the matched gene, gene's priority tier and the corresponding model organism. Users can click the "View User Profile" button to access their full user profile page (see Figure 5B). A scientist who classified the specific gene as Restricted will be shown as an Anonymous Scientist, who can be contacted through the Match Dashboard (see Figures 4–5). **D)** When a Search is performed, a table of scientists registered in partner registries in the RDMM International Network (RDMM, Solve-RD and AFGN) who expressed interest in the human gene of interest or its ortholog candidates is also generated.

A

ModelMatcher -Profile- Match with Scientists- Match with Clinicians and Patients-

Match Dashboard Profile

* denotes required fields

Basic Information

User Title Dr.

Name* Shinya Yamamoto

User Type* **B** Scientist

Affiliation Baylor College of Medicine

Website https://www.bcm.edu/people-search/shinya-yamamoto

Contact Information

Preferred Email (if different)

Phone

About Me

I am a Drosophila biologist who works on rare and undiagnosed diseases. My lab is part of the Undiagnosed Diseases Network. We are also interested in more common neurological, neurodevelopmental and neuropsychiatric disorders such as Alzheimer's disease, autism spectrum disorders and mood disorders.

Save

B'

Scientist Clinical Patient/Family Other

Non-Profit Groups Government Industry

C'

Match Dashboard Match with Scientists

My Genes

Please enter your human genes of interest (NCBI gene symbol):

ex. APP, C9orf72, NOTCH4




Show 10 entries Search my genes:

Symbol	Full Gene Name
x NOTCH1	notch receptor 1

Showing 1 to 1 of 1 entries Previous 1 Next

Scientist Matches

Show 10 entries Search matches:

Your matching gene	Symbol	Organism	Researcher	Institution	Principal Investigator	Contact	Last contacted
NOTCH1	N		Yamamoto, Shinya	Baylor College of Medicine	Yes	Send Match Request	2021-9-1
NOTCH1	notch1		Anonymous Scientist	N/A	Yes	Send Match Request	
NOTCH1	NOTCH1		Yamamoto, Shinya	Baylor College of Medicine	Yes	Send Match Request	

Showing 1 to 3 of 3 entries Previous 1 Next

Figure 4. Information collected from rare and undiagnosed disease research stakeholders to populate a Match Dashboard account.

Upon creating a Match Dashboard account, users enter information about themselves and the human genes they are interested in. **A)** The -Profile- tab requires users to provide general information, including who they are, where they work (if applicable), and why they are interest in rare and undiagnosed diseases. **B-B')** Upon clicking the “User Type” pull down menu in **A)** (blue arrow), users can select one or more user categories to which they feel they belong. **C-C')** The -Match with Scientists- button in **A)** (orange arrow) will navigate the users to a tab to enter their human genes of interest to identify potential scientific collaborators. Upon entry, a table will be generated in the “Scientist Matches” box at the bottom of the -Match with Scientist- tab that lists all scientists who expressed interest in the same human gene or its ortholog candidates. Scientists who selected the specific gene as Public are identifiable whereas scientists who selected the gene as Restricted are displayed as an Anonymous Scientist.

A

Scientist Matches

Show 10 entries

Your matching gene	Symbol	Organism	Researcher	Institution	Principal Investigator	Contact	Last contacted
NOTCH1	N		Yamamoto, Shinya B	Baylor College of Medicine	Yes	Send Match Request C	2021-9-1
NOTCH1	notch1		Anonymous Scientist	N/A	Yes	Send Match Request D	
NOTCH1	NOTCH1		Yamamoto, Shinya	Baylor College of Medicine	Yes	Send Match Request	

Showing 1 to 3 of 3 entries

Previous 1 Next

B'

User Profile Yamamoto, Shinya

Basic Information

First Name Shinya
Last Name Yamamoto
Organization Baylor College of Medicine
Department Department of Molecular and Human Genetics
Website <https://www.bcm.edu/people-search/shinya-yamamoto-33456>

Model Systems
In vitro (Cell Culture)
In vivo (Model Organism)

Principal Investigator Yes

Contact Information

Email yamamoto@bcm.edu

About My Research

My lab works on variety of human diseases using the fruit fly, *Drosophila melanogaster*. The diseases that we study primarily related to the nervous system ranging from Autism Spectrum Disorders and Alzheimer's Disease to Zika virus mediated microcephaly. My lab also functions as part of the Model Organisms Screening Center (MOSC) of the Undiagnosed Diseases Network (UDN) to study the function of rare genetic variants found in patients with rare and undiagnosed diseases. I have also

Organ Systems of Study

Publications

PubMed	Title
3431705	TNPO2 variants associate with human developmental delays, neurologic defects, and dysmorphic fe...
3413067	Heterozygous loss-of-function variants significantly expand the phenotypes associated with loss o...
3394331	Model organisms contribute to diagnosis and discovery in the undiagnosed diseases network: curru...
33864376	Rare deleterious de novo missense variants in <i>IRX2/Slings2</i> are associated with a neurodevelopmental...

Human/Other Research Focus

We have established cell culture based assays to study the function of rare variants in human NOTCH1 genes (NOTCH1-4).

Human/other

Genes

Symbol	Name	%	Tier
NOTCH1	notch receptor 1		Primary
NOTCH2	notch receptor 2		Primary
NOTCH3	notch receptor 3		Secondary
NOTCH4	notch receptor 4		Secondary

Fruit Fly Research Focus

I primarily use *Drosophila melanogaster* to study the function of variants identified in rare and undiagnosed diseases. We also have a track record of studying genes involved in Notch signaling using files

Fruit Fly

Genes

Symbol	Name	%	Tier
amx	almondex		Primary
N	Notch		Secondary
Scx	Sex combs extra		Secondary

C'

Contact SHINYA YAMAMOTO

Researcher Model Systems

Organ Systems of Study

Basic Information

Human gene of interest* NOTCH1
Model Organism* Fruit Fly
MO Gene* N

Reason for contact

Hi Shinya, I would like to discuss a potential collaboration with you. We have a patient with a variant of unknown significance in NOTCH1 and would like to obtain functional data. Based on your publications, it seems you perform this using fruit flies. Can we have a ZOOM call if you are interested in collaborating with us?

D'

Contact Anonymous Scientist

Researcher Model Systems

Organ Systems of Study

Basic Information

Human gene of interest* NOTCH1
Model Organism* Frog
MO Gene* notch1

Reason for contact

Hello, I am interested in reaching out to you for potential collaboration. We have a patient with a missense variant in NOTCH1 who shows multiple neurological symptoms. This individual does not have Adams-Oliver syndrome and may be a phenotypic expansion. If you have a functional assay to test variants in this gene, I would love to share more information with you.

Figure 5. Match Dashboard allows users to contact scientists through a standardized Match Request Form.

A) Starting from the table of scientists and their genes of interest shown in Figure 4C', Match Dashboard users can obtain information about each scientist and utilize the Match Request Forms to initiate communication. **B-B')** By clicking a scientist's name shown teal color in **A)** (blue arrow), Match Dashboard users can access their full public profile page. Note that the same information can also be obtained from clicking the "View User Profile" button next to the corresponding scientist's name in the Search result window shown in Figure 3C. **C-D)** By clicking the "Send Match Request" link in **A)** (orange arrows), a Match Request Form will be displayed that allow users to contact the specific scientist. **C')** shows a contact form addressed to a scientist who categorized the gene of interest as Public, while **D')** shows a contact form addressed to a scientist who categorized the gene as Restricted. Note that some non-identifiable information about the anonymous user, including whether they are a PI, their experimental model systems, and their organ systems of interest

are displayed. The identity of the Anonymous Scientist will be revealed once the scientist responds to a specific user's match request, and only to the person who made the match request.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

A

B' Match with Matchmaker Exchange partners

Basic Information

Human gene of interest* MYH3

Reason for contact

Hello, I am a Drosophila (fruit fly) researcher working on Mhc (Myosin heavy chain) in the context of the muscular and cardiac systems. This gene is orthologous to a number of human genes including MYH3. I am very interested in establishing collaborations with clinicians who see patients with pathogenic variant for translational studies. Please let me know if you are interested in hearing more about our work. Thank you very much for your attention.

Select partner registries

GENE MATCHER MyGene²

I acknowledge my information including name, methods of contact, gene of interest, and any additional information provided in the field above will be sent via email to the Match Request recipient.

C'

Matchmaker Exchange matches

Show 10 entries Search matches:

Your gene	GeneMatcher	MyGene2	Last matched
KMT2D	0 Users	1 User	08/05/2021
MYH3	5 Users	3 Users	08/05/2021

Showing 1 to 2 of 2 entries Previous 1 Next

Figure 6. ModelMatcher users can actively identify potential clinical collaborators through Matchmaker Exchange.

A) By logging in to their Match Dashboard account and clicking the -Match with Clinicians and Patients- button (blue arrow in **A'**), users will be taken to a tab to identify clinicians and affected individuals who are registered in our partner registries in Matchmaker Exchange (currently GeneMatcher, MyGene2, and PhenomeCentral) who share interests in the same gene. The list of the user's human genes of interest, which were previously entered into the Match Dashboard through the -Match with Scientists- tab, will be automatically loaded. Additional genes can also be added here. **B-B')** By clicking the "Exchange Info" button next to each gene in **A)** (orange arrow), a Match Request Form to exchange information with clinical users of partner registries will appear. The User will be asked to briefly explain why they wish to connect with clinicians and affected individuals who express interests in the specific gene, select the specific databases that they wish to exchange information with (default setting is to attempt to match with all partners), and consent that the information

provided in this form and their user profile will be sent to the other party. C) By clicking the “Send” button (yellow arrow), the Match Request will be sent to the corresponding partner registries and the user will receive an email regarding their matches. C’) The information regarding each Match Request sent by the user will also be archived in the “Clinician and Patient Matches” box at the bottom of the -Match with Clinicians and Patients- tab. Note that the match results displayed here are synthesized data from test servers and do not reflect the actual matches with clinicians and affected individuals who participate in Matchmaker Exchange.

Table 1.
Features and strengths of networks and tools that facilitates collaboration between rare and undiagnosed disease stakeholders.

Comparison of key features and strengths of the Undiagnosed Disease Network (UDN), Rare Disease Models and Mechanism (RDMM) Network, Matchmaker Exchange (MME), PubMed (literature search), and ModelMatcher are listed here.

Collaboration platform	Decentralized structure	Up to date information	No geographic barrier	Funding provided	Participating rare disease stakeholders			
					Clinicians	Scientists	Patients	Research funders
UDN		X		X	X	X	X	X
RDMM	X	X		X	X	X		X
MME	X	X	X		X	X	X	
PubMed	X		X		X	X		
ModelMatcher	X	X	X		X	X	X	X