

# PhenomeCentral: 7 years of rare disease matchmaking

Matthew Osmond<sup>1</sup>, Taila Hartley<sup>1</sup>, Brittney Johnstone<sup>2</sup>, Sasha Andjic<sup>3</sup>, Marta Girdea<sup>3</sup>, Meredith Gillespie<sup>1</sup>, Orion Buske<sup>4</sup>, Sergiu Dumitriu<sup>3</sup>, Veronika Koltunova<sup>3</sup>, Arun Ramani<sup>5</sup>, Kym Boycott<sup>6</sup>, and Michael Brudno<sup>7</sup>

<sup>1</sup>Children's Hospital of Eastern Ontario Research Institute

<sup>2</sup>Sunnybrook Health Sciences Centre

<sup>3</sup>University Health Network

<sup>4</sup>Phenotips

<sup>5</sup>University of Toronto Department of Computer Science

<sup>6</sup>Children's Hospital of Eastern Ontario Research Institute

<sup>7</sup>University of Toronto

February 22, 2024

## Abstract

A major challenge in validating genetic causes for patients with rare diseases (RDs) is the difficulty in identifying other RD patients with overlapping phenotypes and variants in the same candidate gene. This process, known as matchmaking, requires robust data sharing solutions in order to be effective. In 2014 we launched PhenomeCentral, a RD data repository capable of collecting computer-readable genotypic and phenotypic data for the purposes of RD matchmaking. Over the past 7 years PhenomeCentral's features have been expanded and its dataset has consistently grown. There are currently 1,615 users registered on PhenomeCentral, which have contributed over 12,000 patient cases. Most of these cases contain detailed phenotypic terms, with a significant portion also providing genomic sequence data or other forms of clinical information. Matchmaking within PhenomeCentral, and with connections to other data repositories in the Matchmaker Exchange, have collectively resulted in over 60,000 matches, which have facilitated multiple gene discoveries. The collection of deep phenotypic and genotypic data has also positioned PhenomeCentral well to support next generation of matchmaking initiatives that utilize genome sequencing data, ensuring that PhenomeCentral will remain a useful tool in solving undiagnosed RD cases in the years to come.

# PhenomeCentral: 7 years of rare disease matchmaking

Matthew Osmond<sup>1</sup>, Taila Hartley<sup>1</sup>, Brittney Johnstone<sup>2</sup>, Sasha Andjic<sup>3</sup>, Marta Girdea<sup>3</sup>, Meredith Gillespie<sup>1</sup>, Orion Buske<sup>5</sup>, Sergiu Dumitriu<sup>3</sup>, Veronika Koltunova<sup>3</sup>, Arun Ramani<sup>6</sup>, Kym M. Boycott<sup>1, 7</sup>, Michael Brudno<sup>3, 4, 6\*</sup>

<sup>1</sup>Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, ON, Canada

<sup>2</sup>Cancer Genetics and High Risk Program, Sunnybrook Health Sciences Centre and Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

<sup>3</sup>DATA Team and Techna Institute, University Health Network, Toronto, ON, Canada

<sup>4</sup>Department of Computer Science, University of Toronto, ON, Canada

<sup>5</sup>PhenoTips, Toronto, ON, Canada

<sup>6</sup>Hospital for Sick Children, Toronto, ON, Canada

<sup>7</sup>Department of Genetics, Children’s Hospital of Eastern Ontario, ON, Canada

\* To whom communication should be addressed: [brudno@cs.toronto.edu](mailto:brudno@cs.toronto.edu)

**Grant Numbers:** Ontario Genomics Institute (OGI-147), CIHR Foundation Grant (FDN-154279)

## Abstract

A major challenge in validating genetic causes for patients with rare diseases (RDs) is the difficulty in identifying other RD patients with overlapping phenotypes and variants in the same candidate gene. This process, known as matchmaking, requires robust data sharing solutions in order to be effective. In 2014 we launched PhenomeCentral, a RD data repository capable of collecting computer-readable genotypic and phenotypic data for the purposes of RD matchmaking. Over the past 7 years PhenomeCentral’s features have been expanded and its dataset has consistently grown. There are currently 1,615 users registered on PhenomeCentral, which have contributed over 12,000 patient cases. Most of these cases contain detailed phenotypic terms, with a significant portion also providing genomic sequence data or other forms of clinical information. Matchmaking within PhenomeCentral, and with connections to other data repositories in the Matchmaker Exchange, have collectively resulted in over 60,000 matches, which have facilitated multiple gene discoveries. The collection of deep phenotypic and genotypic data has also positioned PhenomeCentral well to support next generation of matchmaking initiatives that utilize genome sequencing data, ensuring that PhenomeCentral will remain a useful tool in solving undiagnosed RD cases in the years to come.

**Keywords:** PhenomeCentral, Matchmaker Exchange, rare disease, patient matchmaking, HPO, data-sharing

## Introduction

After genome-wide sequencing, the majority of patients with rare disease (RD) remain without an identified genetic cause for their condition (Lee et al., 2014; Wright et al., 2018; Yang et al., 2013). Oftentimes, finding even one other unrelated patient with a similar phenotype and available genomic data can lead to the identification of a shared genetic cause (Bamshad et al., 2011). Given that RDs are just that, rare, for many disorders no one clinician will see two patients with the same unsolved RD. Instead, the “matching” patients (having similar manifestations with the same genetic cause) may be located across the globe.

A significant barrier to identifying matching patients has been the storage of clinical and genomic data in isolated silos. Furthermore, much of the clinical data is collected in unstructured, non-standardized formats, which impedes computation and sharing across groups. Data is shared by clinicians in numerous ways including case discussion at conferences and publication of case reports. However, given the rarity of the conditions being discussed, these means of sharing are ineffective and inefficient, which translates to delays in gene discovery and answers for patients.

In the field of genetics, data generation is occurring at unprecedented rates in both the clinical and research settings. The subsequent abundance of data in different siloes has led to an interpretation bottleneck and data sharing plays a crucial role in addressing this issue. Broad data sharing will lead to better data quality and interpretation, faster answers for patients, and rapid advancements in the field of genetics. There is consensus among professional societies, patients, and experts that responsible data sharing is imperative (ACMG Board of Directors, 2017; Bush et al., 2018; Darquy et al., 2016; Rehm, 2017). Consideration must

be given to the social, societal, privacy, and policy challenges to implementing international data sharing responsibly, and these are being tackled by groups such as the Global Alliance for Genomics and Health (Rahimzadeh, Dyke, & Knoppers, 2016). There is no question that data sharing will be essential to solving the currently unsolved patients with rare diseases. Furthermore, it will also support better understanding of complex, non-Mendelian conditions and facilitate the shift toward personalized medicine.

To being to address the problem of responsible data sharing, in 2014 we launched PhenomeCentral, a web portal designed for the matchmaking of cases entered by clinicians and researchers working with rare diseases (Buske et al., 2015a). Since its entrance into the rare disease space, PhenomeCentral has been facilitating matches and the identification of second cases or case series leading to gene discovery and answers for families often experiencing a diagnostic odyssey. PhenomeCentral is a founding member of the MME (Philippakis et al., 2015), a collaborative effort to solve genetic disorders by building an international network of rare disease databases connected by a common application interface (API; Buske et al., 2015b). In this manuscript we describe the growth of PhenomeCentral over the past 7 years, improvements we have made for the portal, and goals of our future work.

## Materials and Methods

### Original functionality

The PhenomeCentral web portal is based on the PhenoTips software (Girdea et al., 2013), which is used in both clinical and genomics research applications across the world. Clinicians or researchers (users) can create detailed patient records (cases) which enables the sharing of both genotypic and phenotypic information with other PhenomeCentral cases and across the MME. When a case is first created, users are prompted to select what level of data sharing a patient has consented for, which can include medical and family history, genome-wide sequencing data, or photographs. Genotypic information can be entered as manually curated gene entries and/or uploaded variant call format (VCF) files. Any VCF data is processed by the Exomiser software to produce a computationally prioritized list of candidate genes (Smedley et al., 2015). Case phenotypes are entered as lists of phenotypic terms, which are then mapped to a standardized phenotype vocabulary known as the Human Phenotype Ontology (HPO; Köhler et al., 2021). These HPO terms can be used to generate a list of “Suggested Genes”, which lists all known gene associations for each of a case’s terms. Finally, users have the ability to provide additional clinical information including a pedigree, growth charts and measurements, and other medical history details using free text boxes.

Once a case has been added to PhenomeCentral, internal matching is automatically performed and a list of similar PhenomeCentral cases is displayed. Users also have the option to indicate that their case has consented to share across the MME, which adds similar matches from other rare disease repositories in the MME (called nodes) to the results.

The matchmaking philosophy of PhenomeCentral is that phenotypic data is a valuable, but often overlooked resource (Baynam et al., 2015), and can lead to higher quality matches and decrease time spent exploring false positives (potential matches that are disproven with further exploration). Without phenotypic data, gene level matching may yield many false positive matches (Osmond et al., in press). Consequently, matches in PhenomeCentral are evaluated based on both phenotypic similarity and genotypic similarity. A score for phenotypic similarity is generated using the simGIC algorithm, a semantic similarity measure capable of comparing multiple lists of HPO terms (Pesquita et al., 2008). Genotypic similarity between cases is denoted either as having the same manually curated candidate gene, or in situations where a PhenomeCentral case had VCF data, a match between a manually entered gene and VCF file data with a variant identified in the matching gene (Buske et al., 2015a). After reviewing the matches for a PhenomeCentral case, users are given the option to send an introductory email through PhenomeCentral to other users who have matches of interest.

Any case deposited in PhenomeCentral remains available for future cases to match against. New matches

with other PhenomeCentral cases are updated automatically on the web portal, and users can manually refresh their case to identify new matches across the MME.

## New and Noteworthy

Since PhenomeCentral's initial release, additional features have been added to facilitate a more efficient matchmaking workflow. Matches for a specific case are now displayed under a new "Matching Patients" tab, which initially displays all matches for every candidate gene listed for that case. These matches can then be prioritized using a number of filters including the MME node, genotypic/phenotypic scores, case identifiers, whether a matching case is listed as diagnosed, and whether a match has already been contacted. The communication history is also listed for each match, allowing users to view when each match was contacted by email or manually flag that communication for a match occurred outside of PhenomeCentral. Each match can be flagged with its status (i.e. "Saved" for matches of interest and "Rejected" for matches that were ruled out), and a notes section allows users to record a summary of any communication that has occurred regarding the match. Finally, the features added to filter and track matches for an individual case have also been applied to track larger cohorts within PhenomeCentral. Each user has access to the "My Matches" table, which displays matches for all cases where a user has edit privileges (**Figure 1**).

To make it easier for users to use the portal there is a video series detailing PhenomeCentral's functions for effective genotype and phenotype-driven matchmaking ([https://www.youtube.com/watch?v=...](#)). The videos provide instructions on how to register for an account on PhenomeCentral, how to add genotypic and phenotypic data to a case, how to modify case access permissions for collaborators, how to review matches for an individual case, and how to utilize the My Matches table for larger cohorts.

Additional support has also been provided for facilitating matchmaking connections. While users can still refresh their cases manually to identify new matches across the MME, cases consented for MME matching are now re-queried across the full MME network on a weekly basis. A genetic counselor manually reviews new matches produced by this refresh each week. Case owners are then sent email notifications for any matches with their cases that have a high degree of phenotypic and genotypic similarity.

Finally, as the MME network has expanded to include more rare disease data repositories, PhenomeCentral has added support to match with these newer databases. In addition to matching with similar cases in GeneMatcher and DECIPHER, PhenomeCentral now allows for matching to occur with cases from *seqr* (Arachchi et al., 2018), MyGene2 (Chong et al., 2016), and RD-Connect (Thompson et al., 2014). With these new connections, PhenomeCentral users are now capable of querying over 103,600 cases when making matching requests through the MME ("Matchmaker Exchange Statistic and Publications," 2021).

## Results

PhenomeCentral presently has 1,615 users that have entered 12,292 cases (**Figure 2**) from all over the world. PhenomeCentral has become a major data repository for several large scale rare disease research programs, which have collectively contributed approximately 69% of the total cases within PhenomeCentral. PhenomeCentral is endorsed by multiple rare disease consortia and initiatives and is a recommended resource by the International Rare Diseases Research Consortium (IRDiRC; [https://www.irdirc.org/](#)). The Care for Rare Canada Consortium (C4R; [https://www.careforrare.ca/](#)) has a user base of 159 accounts which have deposited over 2,900 of their cases into PhenomeCentral. Other groups contributing a significant number of cases and using PhenomeCentral regularly include the Genetics of Development Disorders Team (GAD; [https://www.gadteam.org/](#)), the NIH Undiagnosed Diseases Program and Network (UDN; [https://www.undiscovereddiseases.org/](#)), and the Telethon Institute of Genetics and Medicine (TIGEM; [https://www.tigem.it/](#)). The last two are members of the Undiagnosed Diseases Network International (UDNI; [https://www.udni.org/](#)), which has endorsed PhenomeCentral as a data sharing solution. Collectively, the RD research programs mentioned above bring together 389 users that have contributed over 8,400 cases to date.

Currently, 5,554 cases in PhenomeCentral have a manually curated candidate gene listed, spanning a total

of 3,234 unique genes. Most cases have a single candidate gene, with only 12% of cases listing two or more genes as candidates. Additional genotypic data in the form of VCF files are available in 1901 cases, which can be used to generate matches between candidate genes and variants in genomic sequencing data.

The majority of cases in PhenomeCentral also have some degree of computer-readable phenotyping (10250/12292, 83%). Of these cases, 94% have two or more phenotypes documented, with a median of 8 phenotypic terms per case. In total, 118,059 HPO annotations are documented across this dataset, spanning across 7,938 unique HPO terms. When grouped into general phenotypic categories, abnormalities of the nervous system are the most commonly reported type of HPO term, followed by abnormalities of the head or neck and abnormalities of the skeletal system (**Figure 3a**). More specifically, the most prevalent HPO terms within this dataset (**Figure 3b**) include global developmental delay (2867/12292, 23%), seizures (1572/12292, 13%), and short stature (1531/12292, 12%). In addition to HPO terms, many cases have additional information entered, including growth charts (3005/12292, 24%), a pedigree (2555/12292, 21%), and additional medical history information (1972/12292, 16%).

A total of 4,665 matches have been made internally between PhenomeCentral cases. In addition to internal matchmaking with other PhenomeCentral cases, currently 70% of PhenomeCentral cases (8479/12292) have been consented for matching with the other MME nodes. Such cases submitted queries for over 2,900 unique candidate genes, and these 8,479 cases collectively contained over 98,500 HPO annotations. Queries involving PhenomeCentral as a node in the MME have resulted in 61,362 matches with other nodes, with the majority of these originating from cases in GeneMatcher (40701/61362, 66%) and DECIPHER (17097/61362, 28%).

In collaboration with the MME, PhenomeCentral has enjoyed success in facilitating matches leading to gene discovery (e.g., Chelban et al., 2019; den Hoed et al., 2021; Faden et al., 2015; Ito et al., 2018; Johnstone et al., 2017; Kernohan et al., 2017; Y.-R. Lee et al., 2020; Lessel et al., 2020; Martinelli et al., 2018; Oud et al., 2017; Salpietro et al., 2019; Simons et al., 2017; Skraban et al., 2017; Stray-Pedersen et al., 2016; Vavassori et al., 2021), and, ultimately, answers for families.

## Discussion

Assessing the current landscape of the PhenomeCentral dataset shows a steady growth in the deposition of cases for the purposes of matchmaking. Most of these cases are deeply phenotyped, with an average of 11 HPO terms annotated per case, and many cases containing additional medical and family history information. A high amount of genotypic diversity was also observed within the PhenomeCentral dataset, with over 3200 unique candidate genes flagged in total. Finally, all PhenomeCentral cases were subjected to internal matching, and about 70% of cases were also consented to matching with other MME data repositories. Both internal and external matchmaking queries resulted in over 62,000 matches being returned across the entire dataset, ultimately leading to the identification of multiple novel disease-gene associations.

PhenomeCentral is based on the PhenoTips software, which has a number of advantages. Over the past five years, PhenoTips has been actively implemented into hospital systems around the world to enable improved care for RD patients, resulting in a lower barrier to entry as more physicians become familiar with the similar interfaces. Having PhenoTips software at the core of clinical and research databases also expedites the migration of clinical data into research, as all PhenoTips instances support the export and import of the same standardized data files, as well as automated deposition of de-identified cases into PhenomeCentral.

Based on user feedback, we devoted considerable resources to developing the revised matchmaking filters and the My Matches table. As the MME network continues to grow and more data is deposited for matchmaking, we have begun to approach a point where nearly every candidate gene returns matches with other cases (Osmond et al., in press). Combined with the reality that older matchmaking submissions continue to receive new matches years after their initial submission, it is critical that matchmaking nodes provide users with the tools to filter and track up to thousands of matches simultaneously. The new filters and My Matches table

represent initial steps towards providing users with such tools, however additional changes will be required so that matchmaking remains efficient for users.

The presence of high quality phenotypic data in cases submitted to matchmaking represents another solution to reducing the time required to resolve an increasing number of matches. The matchmaking experiences of the Care4Rare Canada research team suggest that while most MME nodes support the storage of standardized phenotypic data, more than half of cases in the MME are submitted with little to no information on clinical features (Osmond et al., in press). As a result, most matches are difficult to resolve on an initial review, and require lengthy email exchanges with the matching user to determine whether a given match is of interest. Conversely, matches with cases from nodes where phenotypic data is frequently provided, such as PhenomeCentral, could be ruled out on initial review over 50% of the time, drastically reducing the number of follow-up emails required. As the number of cases and candidate genes submitted to matching continue to grow, it will be critical for nodes to emphasize the importance of submitting phenotypic data to ensure current matchmaking solutions remain practical. The philosophy of PhenomeCentral is that the upfront effort of contributing phenotypic data to cases ultimately saves time in the matchmaking process

We believe that the current design of PhenomeCentral is well positioned for novel approaches to matchmaking, which will utilize genomic sequence data to increase the number of matches made for a given gene. The MME framework is currently based on two-sided matchmaking, an approach where both cases in a match have the same identified candidate gene. An iteration on this approach, called one-sided matchmaking, would instead allow users to directly query the genomic sequence data of patient records in a database for variants in a candidate gene. In the future, zero-sided matchmaking, a process in which algorithms use genomic sequence data and phenotypic information to highlight matches of interest, may also become a reality. One-sided and zero-sided matchmaking, while requiring patient consent to a greater degree of data sharing, both have the potential to increase the number of matches made for a candidate gene. They will also have a greater need for detailed phenotypic data associated with cases to ensure that the larger numbers of matches can be reviewed without resorting to lengthy email exchanges. PhenomeCentral is perfectly positioned for these matchmaking approaches, as it already allows users to upload sequencing data for cases, provides consent checkboxes to indicate which cases may use this data for matchmaking, and can display matches between candidate genes and variants identified in sequencing data.

In summary, PhenomeCentral has continued to grow as a data repository dedicated to gene discovery and finding diagnoses for unsolved rare disease patients through matchmaking. The current dataset consists of cases from both large research groups and individual researchers, and contains a wide variety of candidate genes and computer-readable phenotypes. The development of new features such as robust matching filters and cohort-wide matching tables have helped PhenomeCentral users more efficiently manage an ever-growing number of matches. Finally, an emphasis on contributing high quality genotypic and phenotypic data to matchmaking has both aided MME users in the quick resolution of many matches (Osmond et al., in press), and has positioned PhenomeCentral to contribute to more sophisticated forms of matchmaking in the future.

## Acknowledgements

This work was performed under the Care4Rare Canada Consortium funded by Genome Canada and the Ontario Genomics Institute (OGI-147), the Canadian Institutes of Health Research, Ontario Research Fund, Genome Alberta, Genome British Columbia, Genome Quebec, and Children’s Hospital of Eastern Ontario Foundation. T.H. was supported by a Frederick Banting and Charles Best Canada Graduate Scholarship Doctoral Award from CIHR. K.M.B. was supported by a CIHR Foundation Grant (FDN-154279) and a Tier 1 Canada Research Chair in Rare Disease Precision Health.

## Conflicts of Interest

OJB and MB have an equity interest in, and OJB is an employee of PhenoTips, which provides support for software that underlies the PhenomeCentral database.

## Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

## References

- ACMG Board of Directors. (2017). Laboratory and clinical genomic data sharing is crucial to improving genetic health care: a position statement of the American College of Medical Genetics and Genomics. *Genetics in Medicine* , 19 (7), 721–722. <https://doi.org/10.1038/gim.2016.196>
- Arachchi, H., Wojcik, M. H., Weisburd, B., Jacobsen, J. O. B., Valkanas, E., Baxter, S., ... Rehm, H. L. (2018). matchbox: An open-source tool for patient matching via the Matchmaker Exchange. *Human Mutation* , 39 (12), 1827–1834. <https://doi.org/10.1002/humu.23655>
- Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A., & Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics* , 12 (11), 745–755. <https://doi.org/10.1038/nrg3031>
- Baynam, G., Walters, M., Claes, P., Kung, S., Lesouef, P., Dawkins, H., ... Goldblatt, J. (2015, April 1). Phenotyping: Targeting genotype’s rich cousin for diagnosis. *Journal of Paediatrics and Child Health* , Vol. 51, pp. 381–386. Blackwell Publishing. <https://doi.org/10.1111/jpc.12705>
- Bush, L. W., Beck, A. E., Biesecker, L. G., Evans, J. P., Hamosh, A., Holm, I. A., ... Rehm, H. L. (2018). Professional responsibilities regarding the provision, publication and dissemination of patient phenotypes in the context of clinical genetic and genomic testing: points to consider. A statement of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine* , 20 (2), 169–171. <https://doi.org/10.1038/gim.2017.242>
- Buske, O. J., Girdea, M., Dumitriu, S., Gallinger, B., Hartley, T., Trang, H., ... Brudno, M. (2015a). PhenomeCentral: A portal for phenotypic and genotypic matchmaking of patients with rare genetic diseases. *Human Mutation* , 36 (10), 931–940. <https://doi.org/10.1002/humu.22851>
- Buske, O. J., Schiettecatte, F., Hutton, B., Dumitriu, S., Misyura, A., Huang, L., ... Brudno, M. (2015b). The Matchmaker Exchange API: Automating patient matching Through the exchange of structured phenotypic and genotypic profiles. *Human Mutation* , 36 (10), 922–927. <https://doi.org/10.1002/humu.22850>
- Chelban, V., Wilson, M. P., Warman Chardon, J., Vandrovcova, J., Zanetti, M. N., Zamba-Papanicolaou, E., ... Houlden, H. (2019). PDXK mutations cause polyneuropathy responsive to pyridoxal 5’-phosphate supplementation. *Annals of Neurology* , 86 (2), 225–240. <https://doi.org/10.1002/ana.25524>
- Chong, J. X., Yu, J.-H., Lorentzen, P., Park, K. M., Jamal, S. M., Tabor, H. K., ... Bamshad, M. J. (2016). Gene discovery for Mendelian conditions via social networking: de novo variants in KDM1A cause developmental delay and distinctive facial features. *Genetics in Medicine* , 18 (8), 788–795. <https://doi.org/10.1038/gim.2015.161>
- Darquy, S., Moutel, G., Lapointe, A.-S., D’Auffret, D., Champagnat, J., Guerroui, S., ... Duchange, N. (2016). Patient/family views on data sharing in rare diseases: study in the European LeukoTreat project. *European Journal of Human Genetics* , 24 (3), 338–343. <https://doi.org/10.1038/ejhg.2015.115>

- den Hoed, J., de Boer, E., Voisin, N., Dingemans, A. J. M., Guex, N., Wiel, L., ... Vissers, L. E. L. M. (2021). Mutation-specific pathophysiological mechanisms define different neurodevelopmental disorders associated with SATB1 dysfunction. *American Journal of Human Genetics* , 108 (2), 346–356. <https://doi.org/10.1016/j.ajhg.2021.01.007>
- Faden, M., AlZahrani, F., Mendoza-Londono, R., Dupuis, L., Hartley, T., Kannu, P., ... Alkuraya, F. S. (2015). Identification of a recognizable progressive skeletal dysplasia caused by RSPRY1 mutations. *American Journal of Human Genetics* , 97 (4), 608–615. <https://doi.org/10.1016/j.ajhg.2015.08.007>
- Girdea, M., Dumitriu, S., Fiume, M., Bowdin, S., Boycott, K. M., Chenier, S., ... Brudno, M. (2013). PhenoTips: Patient phenotyping software for clinical and research use. *Human Mutation* , 34 (8), 1057–1065. <https://doi.org/10.1002/humu.22347>
- Ito, Y., Carss, K. J., Duarte, S. T., Hartley, T., Keren, B., Kurian, M. A., ... Raymond, F. L. (2018). De novo truncating mutations in WASF1 cause intellectual disability with seizures. *American Journal of Human Genetics* , 103 (1), 144–153. <https://doi.org/10.1016/j.ajhg.2018.06.001>
- Johnstone, D. L., Nguyen, T.-T.-M., Murakami, Y., Kernohan, K. D., Tetreault, M., Goldsmith, C., ... Campeau, P. M. (2017). Compound heterozygous mutations in the gene PIGP are associated with early infantile epileptic encephalopathy. *Human Molecular Genetics* , 26 (9), 1706–1715. <https://doi.org/10.1093/hmg/ddx077>
- Kernohan, K. D., Dymont, D. A., Pupavac, M., Cramer, Z., McBride, A., Bernard, G., ... Thiffault, I. (2017). Matchmaking facilitates the diagnosis of an autosomal-recessive mitochondrial disease caused by biallelic mutation of the tRNA isopentenyltransferase (TRIT1) gene. *Human Mutation* , 38 (5), 511–516. <https://doi.org/10.1002/humu.23196>
- Kohler, S., Gargano, M., Matentzoglou, N., Carmody, L. C., Lewis-Smith, D., Vasilevsky, N. A., ... Robinson, P. N. (2021). The Human Phenotype Ontology in 2021. *Nucleic Acids Research* , 49 (D1), D1207. <https://doi.org/10.1093/nar/gkaa1043>
- Lee, H., Deignan, J. L., Dorrani, N., Strom, S. P., Kantarci, S., Quintero-Rivera, F., ... Nelson, S. F. (2014). Clinical Exome Sequencing for Genetic Identification of Rare Mendelian Disorders. *Journal of American Medical Association* , 312 (18), 1880–1887. <https://doi.org/10.1001/jama.2014.14604>
- Lee, Y.-R., Khan, K., Armfield-Uhas, K., Srikanth, S., Thompson, N. A., Pardo, M., ... Schwartz, C. E. (2020). Mutations in FAM50A suggest that Armfield XLID syndrome is a spliceosomopathy. *Nature Communications* , 11 (1), 3698. <https://doi.org/10.1038/s41467-020-17452-6>
- Lessel, D., Zeitler, D. M., Reijnders, M. R. F., Kazantsev, A., Hassani Nia, F., Bartholomaeus, A., ... Kreienkamp, H.-J. (2020). Germline AGO2 mutations impair RNA interference and human neurological development. *Nature Communications* , 11 (1), 5797. <https://doi.org/10.1038/s41467-020-19572-5>
- Martinelli, S., Krumbach, O. H. F., Pantaleoni, F., Coppola, S., Amin, E., Pannone, L., ... Mirzaa, G. M. (2018). Functional dysregulation of CDC42 causes diverse developmental phenotypes. *American Journal of Human Genetics* , 102 (2), 309–320. <https://doi.org/10.1016/j.ajhg.2017.12.015>
- Matchmaker Exchange Statistic and Publications. (2021). Retrieved September 30, 2021, from <https://www.matchmakerexchange.org/statistics.html>
- Osmond, M., Hartley, T., Dymont, D., Kernohan, K. D., Brudno, M., Buske, O. J., ... Boycott, K. M. (in press). Outcome of over 1500 matches through the Matchmaker Exchange for rare disease gene discovery: the two year experience of Care4Rare Canada. *Genetics in Medicine* .
- Oud, M. M., Tuijnenburg, P., Hempel, M., van Vlies, N., Ren, Z., Ferdinandusse, S., ... Kuijpers, T. W. (2017). Mutations in EXTL3 cause neuro-immuno-skeletal dysplasia syndrome. *American Journal of Human Genetics* , 100 (2), 281–296. <https://doi.org/10.1016/j.ajhg.2017.01.013>



- Pesquita, C., Faria, D., Bastos, H., Ferreira, A. E., Falcao, A. O., & Couto, F. M. (2008). Metrics for GO based protein semantic similarity: a systematic evaluation. *BMC Bioinformatics* , 9 (S5), S4. <https://doi.org/10.1186/1471-2105-9-S5-S4>
- Philippakis, A. A., Azzariti, D. R., Beltran, S., Brookes, A. J., Brownstein, C. A., Brudno, M., ... Rehm, H. L. (2015). The Matchmaker Exchange: A Platform for Rare Disease Gene Discovery. *Human Mutation* , 36 (10), 915–921. <https://doi.org/10.1002/humu.22858>
- Rahimzadeh, V., Dyke, S. O. M., & Knoppers, B. M. (2016). An international framework for data sharing: Moving forward with the Global Alliance for Genomics and Health. *Biopreservation and Biobanking* ,14 (3), 256–259. <https://doi.org/10.1089/bio.2016.0005>
- Rehm, H. L. (2017). A new era in the interpretation of human genomic variation. *Genetics in Medicine* 2017 19:10 , 19 (10), 1092–1095. <https://doi.org/10.1038/gim.2017.90>
- Salpietro, V., Dixon, C. L., Guo, H., Bello, O. D., Vandrovцова, J., Efthymiou, S., ... Houlden, H. (2019). AMPA receptor GluA2 subunit defects are a cause of neurodevelopmental disorders. *Nature Communications* , 10 (1), 3094. <https://doi.org/10.1038/s41467-019-10910-w>
- Simons, C., Dymont, D., Bent, S. J., Crawford, J., D’Hooghe, M., Kohlschutter, A., ... Wolf, N. I. (2017). A recurrent de novo mutation in TMEM106B causes hypomyelinating leukodystrophy. *Brain* , 140 (12), 3105–3111. <https://doi.org/10.1093/brain/awx314>
- Skraban, C. M., Wells, C. F., Markose, P., Cho, M. T., Nesbitt, A. I., Au, P. Y. B., ... Deardorff, M. A. (2017). WDR26 haploinsufficiency causes a recognizable syndrome of intellectual disability, seizures, abnormal gait, and distinctive facial features. *American Journal of Human Genetics* , 101 (1), 139–148. <https://doi.org/10.1016/j.ajhg.2017.06.002>
- Smedley, D., Jacobsen, J. O. B., Jager, M., Kohler, S., Holtgrewe, M., Schubach, M., ... Robinson, P. N. (2015). Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nature Protocols* , 10 (12), 2004–2015. <https://doi.org/10.1038/nprot.2015.124>
- Stray-Pedersen, A., Cobben, J.-M., Prescott, T. E., Lee, S., Cang, C., Aranda, K., ... Yoon, G. (2016). Biallelic mutations in UNC80 cause persistent hypotonia, encephalopathy, growth retardation, and severe intellectual disability. *American Journal of Human Genetics* ,98 (1), 202–209. <https://doi.org/10.1016/j.ajhg.2015.11.004>
- Thompson, R., Johnston, L., Taruscio, D., Monaco, L., Beroud, C., Gut, I. G., ... Lochmuller, H. (2014). RD-Connect: An integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. *Journal of General Internal Medicine* , 29 (S3), 780–787. <https://doi.org/10.1007/s11606-014-2908-8>
- Vavassori, S., Chou, J., Faletti, L. E., Haunerding, V., Opitz, L., Joset, P., ... Pachlopnik Schmid, J. (2021). Multisystem inflammation and susceptibility to viral infections in human ZNFX1 deficiency. *Journal of Allergy and Clinical Immunology* ,148 (2), 381–393. <https://doi.org/10.1016/j.jaci.2021.03.045>
- Wright, C. F., McRae, J. F., Clayton, S., Gallone, G., Aitken, S., FitzGerald, T. W., ... Firth, H. V. (2018). Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. *Genetics in Medicine* 2018 20:10 , 20 (10), 1216–1223. <https://doi.org/10.1038/gim.2017.246>
- Yang, Y., Muzny, D. M., Reid, J. G., Bainbridge, M. N., Willis, A., Ward, P. A., ... Eng, C. M. (2013). Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *The New England Journal of Medicine* , 369 (16), 1502. <https://doi.org/10.1056/nejmoa1306555>

# Figure Legends

**Figure 1: Screenshot of the My Matches table in PhenomeCentral.**PhenomeCentral users can use this table to keep track of the matches found for all of their cases in one place. Here, they can perform advanced filtering and sorting, compare genotypes and phenotypes between matched cases, write individual notes, initiate contact with other PhenomeCentral or MME users, and access the communication history associated with a specific match.

**Figure 2: Total number of cases deposited in PhenomeCentral.**Since its launch in February 2014, and beginning with only a handful of cases contributed by the Canadian Care for Rare Consortium (<http://care4rare.ca>), PhenomeCentral has seen steady growth and has now surpassed 12,000 patient profiles available for matching.

**Figure 3: Overview of the HPO phenotypes documented in PhenomeCentral cases (N = 12292 cases).** a) Frequency of general categories of HPO terms; abnormalities in the nervous system, abnormalities in the head or neck, and abnormalities in the skeletal system are most prevalent. b) Frequency of specific HPO terms; global developmental delay, seizures, and short stature are most prevalent.



