OVERVIEW

Human Mutation

OFFICIAL JOURNAL HGVS HUMAN GENOME VARIATION SOCIETY www.hgvs.org

The Matchmaker Exchange: A Platform for Rare Disease Gene Discovery

Anthony A. Philippakis,[1,2,3] Danielle R. Azzariti,[4] Sergi Beltran,[5] Anthony J. Brookes,[6] Catherine A. Brownstein,[3,7] Michael Brudno,[8,9,10] Han G. Brunner,[11,12] Orion J. Buske,[8,9,10] Knox Carey,[13] Cassie Doll,[14] Sergiu Dumitriu,[10] Stephanie O.M. Dyke,[15] Johan T. den Dunnen,[16] Helen V. Firth,[17] Richard A. Gibbs,[18] Marta Girdea,[8,10] Michael Gonzalez,[19] Melissa A. Haendel,[20] Ada Hamosh,[21] Ingrid A. Holm,[3,7] Lijia Huang,[22] Matthew E. Hurles,[23] Ben Hutton,[23] Joel B. Krier,[3,24] Andriy Misyura,[10] Christopher J. Mungall,[25] Justin Paschall,[26] Benedict Paten,[27] Peter N. Robinson,[28,29,30,31] François Schiettecatte,[32] Nara L. Sobreira,[21] Ganesh J. Swaminathan,[23] Peter E. Taschner,[16,33] Sharon F. Terry,[34] Nicole L. Washington,[2] Stephan Züchner,[35] Kym M. Boycott,[36] and Heidi L. Rehm[1,3,4,37]*

[1]The Broad Institute of Harvard and MIT, Cambridge, Massachusetts; [2]Department of Cardiology, Brigham & Women's Hospital, Boston, Massachusetts; [3]Harvard Medical School, Boston, Massachusetts; [4]Laboratory for Molecular Medicine, Partners Personalized Medicine, Boston, Massachusetts; [5]Centro Nacional de Análisis Genómico, Barcelona, Spain; [6]Department of Genetics, University of Leicester, Leicester, UK; [7]Division of Genetics and Genomics and the Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, Massachusetts; [8]Department of Computer Science, University of Toronto, Toronto, Canada; [9]Genetics and Genome Biology Program, The Hospital for Sick Children, Toronto, Canada; [10]Centre for Computational Medicine, The Hospital for Sick Children, Toronto, Canada; [11]Radboud University Medical Center, Department of Human Genetics, Nijmegen 6500 HB, The Netherlands; [12]Maastricht University Medical Center, Department of Clinical Genetics, Maastricht 6202AZ, The Netherlands; [13]Gene Cloud, California; [14]Google Inc., Mountain View, California; [15]Centre of Genomics and Policy, Faculty of Medicine, McGill University, Quebec, Canada; [16]Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands; [17]East Anglian Medical Genetics Service, Box 134, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK; [18]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030; [19]The Genesis Project Inc., Miami, Florida; [20]Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, Oregon; [21]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland; [22]The Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada; [23]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK; [24]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02115; [25]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California; [26]European Molecular Biology Laboratory European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK; [27]UC Santa Cruz Genomics Institute, Santa Cruz, California; [28]Institute for Medical Genetics and Human Genetics, Charité-Universitätsmedizin Berlin, Berlin 13353, Germany; [29]Max Planck Institute for Molecular Genetics, Berlin 14195, Germany; [30]Institute for Bioinformatics, Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin 14195, Germany; [31]Berlin Brandenburg Center for Regenerative Therapies, Berlin 13353, Germany; [32]FS Consulting LLC, Salem, Massachusetts 01970; [33]Generade Center of Expertise Genomics, University of Applied Sciences Leiden, Leiden, The Netherlands; [34]Genetic Alliance, Washington, District of Columbia; [35]Dr. John T. Macdonald Foundation Department of Human Genetics and John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, Florida; [36]Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; [37]Department of Pathology, Brigham & Women's Hospital, Boston, Massachusetts

For the Matchmaker Exchange Special Issue

Received 25 May 2015; accepted revised manuscript 21 July 2015.
Published online in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22858

ABSTRACT: There are few better examples of the need for data sharing than in the rare disease community, where patients, physicians, and researchers must search for "the needle in a haystack" to uncover rare, novel causes of disease within the genome. Impeding the pace of discovery has been the existence of many small siloed datasets within individual research or clinical laboratory databases and/or disease-specific organizations, hoping for serendipitous occasions when two distant investigators happen to learn they have a rare phenotype in common and can "match" these cases to build evidence for causality.

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Heidi L. Rehm, Partners Laboratory for Molecular Medicine, 65 Landsdowne St., Cambridge, MA 02139. E-mail: hrehm@partners.org

Contract grant sponsors: NIH (grants U41HG006834, T32GM007748, R01NS075764, 5R01NS072248, U54NS065712, U54HG006542, N01CO42400-80, HG007530, HG007690, U54HG007990, HD077671, 5R24OD011883, U54HG003273); BioSHaRE-EU project (EC FP7, #261433); PCORI (contract PPRN-1306-04899); Robert Wood Johnson Foundation (grant 71636); PXE International; Genome Canada; Canadian Institutes of Health Research (grants EP1-120608, EP2-120609); the Ontario Genomics Institute; NSERC/CIHR Collaborative Health Research Project (CHRP); the Garron Family Cancer Centre and Hospital for Sick Children Foundation Student Scholarship Program; Broad Ignite Award; NCI Cloud Pilot (grant N01CO42400-80); Wellcome Trust (grant number WT098051); Director, Office of Science, Office of Basic Energy Sciences of the US Department of Energy (contract no. DE-AC02-05CH11231); European Union Seventh Framework Programme (FP7/2007-2013) (grant agreement no. 305444); U54 HG003273; Canada Research Chair in Law and Medicine; Public Population Project in Genomics and Society (P3G).

However, serendipity has never proven to be a reliable or scalable approach in science. As such, the Matchmaker Exchange (MME) was launched to provide a robust and systematic approach to rare disease gene discovery through the creation of a federated network connecting databases of genotypes and rare phenotypes using a common application programming interface (API). The core building blocks of the MME have been defined and assembled. Three MME services have now been connected through the API and are available for community use. Additional databases that support internal matching are anticipated to join the MME network as it continues to grow.

Hum Mutat 36:915–921, 2015.

KEY WORDS: matchmaking; rare disease; genomic API; gene discovery; Matchmaker Exchange; GA4GH, IRDiRC

Introduction

The content of genetic tests has gradually expanded over the years, with major leaps happening recently with the introduction of exome and genome sequencing. Although the rate of solving monogenic "Mendelian" disorders has increased with the ability to query all genes, a large fraction of patients still remain without a diagnosis. A portion of these unsolved cases harbor suspicious variants in candidate disease genes. For such cases, finding just a single additional unrelated case with a deleterious variant in the same gene and overlapping phenotype may provide sufficient evidence to causally implicate the gene, enabling a diagnosis for the patient. Methods for identifying these additional cases have evolved over time. From word of mouth between colleagues to sharing published case reports, laboratory diagnosticians and clinicians have worked to uncover connections between patients [Loucks et al., 2015]. In a world of rapidly evolving information technologies, however, a more efficient solution is needed that can scale with the exploding growth in genomic sequencing.

Multiple projects have addressed this need by developing platforms that use genotype and phenotype-driven matching algorithms to identify cases with common phenotypes and disrupted genes [Washington et al., 2009; Gonzalez et al., 2012, Swaminathan et al., 2012, Gonzalez et al., 2013, Robinson et al., 2014; Zemojtel et al., 2014; Buske et al., 2015a; Lancaster et al., 2015; Sobreira et al., 2015a]. However, no organized system existed to facilitate the interaction between these multiple disconnected projects (Fig. 1) before the Matchmaker Exchange (MME). To unify these efforts and harness the collective data across all of the databases, groups representing rare disease repositories held a meeting in October 2013 to launch an open collaboration later named the MME (http://www.matchmakerexchange.org). This collaborative effort has launched a federated platform (exchange) to facilitate the identification of cases with similar phenotypic and genotypic profiles (matchmaking) through a standardized application programming interface (API) and procedural conventions. The MME enables searches of multiple databases (matchmaker services) from another, connected matchmaker service, without having to separately query all services, or deposit data in each one. The queries are designed to allow a gene or genotype, combined with a condition or phenotypic features, to be sent as a query in order to get a returned response containing any similar or "matched" cases. Matching algorithms are defined by the matchmaker services and will evolve over time as described below.

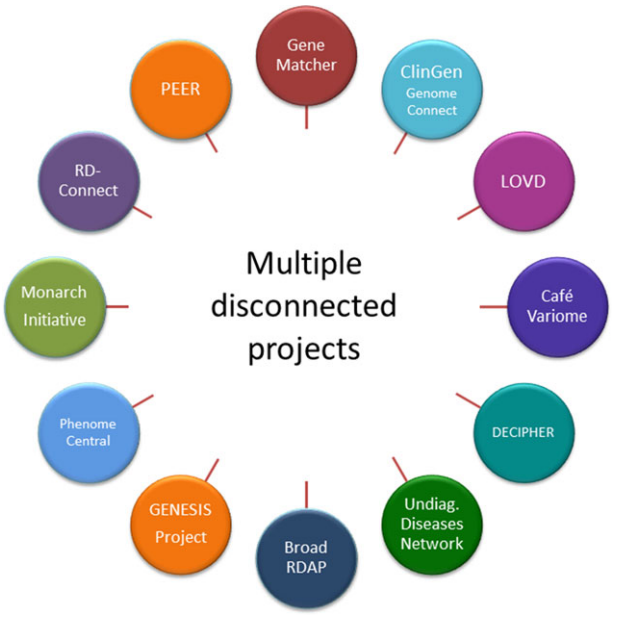

Figure 1. Databases and programs that gathered to form the basis for the MME. The MME includes representatives from the founding organizations and databases supporting or intending to support matchmaking services (Tables 1 and 2). Collaborative work has focused on both the technical aspects of data sharing, as well as policy considerations. This work has resulted in version 1.0 of a MME API [Buske et al., 2015b], a set of requirements for qualifying as a MME service, and a user agreement for querying the MME. The MME has been identified as a demonstration project for the Global Alliance for Genomics and Health (GA4GH) and the MME has been leveraging the expertise of the GA4GH working groups for guidance on pertinent aspects of the project.

Federated Versus Centralized Approaches to Data Sharing

Historically, most genetic and genomic data sharing has been accomplished through the aggregation of data in a single "centralized" site, such as the National Center for Biotechnology Information's (NCBI) Database of Genotypes and Phenotypes (dbGaP) [Tryka et al., 2013] or other large data centers such as those employed for the International Cancer Genome Consortium (ICGC) [Zhang et al., 2011] and the Cancer Genome Atlas (TCGA) [Weinstein et al., 2013]. This approach allows for easy data analysis given that a data holder is in complete control of the entire dataset; however, a higher regulatory burden must be overcome to allow data to be shared with another entity, putting its security and privacy management entirely in the hands of the database owner. In addition, users may only wish to share certain datasets with others and only under certain circumstances that can be better controlled by the use of an API to enable data access. Finally, data annotations such as phenotype are dynamic within a patient, but static within a disconnected database, where they can be difficult to capture longitudinally. A federated system makes it easier to support longitudinal connections to patient phenotype and updated genomic interpretations.

An alternative approach is the use of a federated network in which multiple distributed databases are connected through APIs, whereby each database supports queries of other databases in the network. This allows each database to be autonomous with respect

to its own data schema, maintain ongoing control of its own data, and continuously innovate at its own pace. In this model, no single database acts as the "central" database, nor does a single database take on the privacy and security requirements of the whole network.

It is this latter federated model that was chosen to support the MME, though some data contributors may prefer to deposit data into an existing matchmaker service for participation in the MME instead of setting up their own matchmaker. This initial approach allows each participating matchmaker service to maintain their autonomy and primary purpose, while contributing valuable data to the MME and the genomics community. Data contributors no longer need to deposit the same datasets into multiple databases in order to find matches, and they will have more options for databases in which to deposit data, including databases in their own jurisdiction if certain regulations prohibit data from leaving a region. Also, data contributors may decide to put some cases into one database and other cases into another database depending on the focus of each database. The decision of where to start may be based upon a variety of factors as described below, including the database's supported content and algorithms for matching. However, in the MME, data contributors are discouraged from depositing the same dataset into multiple databases in order to minimize data duplication.

Building Blocks to Support the MME

To promote responsible data sharing, the founding members of the MME have established a set of requirements for participating matchmaking services, a user agreement for those wishing to use the MME, and a steering committee (SC) to govern the program. The SC is composed of a representative from each approved MME service, as well as program organizers and representation from Global Alliance for Genomics and Health (GA4GH) and the International Rare Diseases Research Consortium (IRDiRC). The SC is charged with maintaining the service requirements, user agreement, and oversight of the API to ensure the MME meets the needs of the rare disease community and reflects consensus standards and best practices as set forth by the GA4GH and IRDiRC. The MME also supports a monthly conference call and periodic in-person meetings, most of which are open to the community to encourage active participation by all stakeholders.

MME service requirements

To become a MME service, each new site must achieve the following:

1. Require users to deposit case data to undertake a federated query across the MME service providers
2. Establish a minimum of two point-to-point API connections to other MME services
3. Contain content that is considered by the MME SC to be useful for matching, including the flagging of, or ability to prioritize, candidate genes
4. Successfully implement matching algorithms using test data
5. During user queries, enable dual notification of data requester (i.e., the querier) and prior data depositor (i.e., the queried) including sharing the identities and contact information for each
6. For each database to which a MME service is connected by an API, the connected database's disclaimers should be posted on the MME service's website and displayed with query results. Disclaimers can be found on GitHub (https://github.com/ga4gh/mme-apis)
7. Store queries sent and received between MME sites only for the purpose of auditing, defining query statistics, and following up queries to understand rates of validated gene discovery
8. Attest to database security requirements as defined by the GA4GH Security WG (forthcoming)
9. Advance the goals of the MME project through active participation in meetings and conference calls including defining a representative for the MME SC

MME end user agreement

To use the MME, each data querier agrees to the following:

1. To make no attempt to identify individual patients in any MME database
2. To enable all cases submitted for querying to be stored in the query-initiating database for future matching
3. To obtain permission from the source of the matching data before publishing or presenting the results of queries
4. To acknowledge the MME, and the specific MME service that supported any discoveries in publications, as appropriate

MME API for genotypes and phenotypes

APIs define protocols for how components of computer systems communicate, and are a crucial part of the modern information technology landscape. In particular, web APIs have enabled the creation of our modern ecosystem of automatic communication between computer programs or services. APIs represent a defined protocol between technology services, such that a given input results in an expected output in a standardized format.

Participating matchmaker services are required to implement a standardized API, consistent with standards developed by the GA4GH Data Working Group, for exchanging genotypic and phenotypic information. The API supports queries, where a query is a patient record, and where the receiving system decides how best to process a specific query. Thus, the system does not support queries such as "Do you have any patients with a deleterious variant in *CASQ2*?" or "Do you have any patients with hypertelorism and arachnodactyly?," but instead supports a query of "Do you have any patients *similar* to one who has hypertelorism and arachnodactyly with a deleterious variant in *CASQ2*," where the definition of similarity is at the discretion of the receiving system. This API is described in greater detail in a companion article of this journal issue [Buske et al., 2015b]. In brief, the core elements of each query that are transferred through the API include several mandatory elements: case ID, submitter information, and candidate gene(s) and/or phenotype terms. The API also accommodates additional fields to increase the specificity of queries including gender, age of onset, mode of inheritance, condition name (e.g., OMIM or Orphanet ID), chromosome, chromosome region, zygosity, and variant type (e.g., frameshift, missense, etc.).

Federated authentication and authorization

The MME recognizes the importance of authentication (validation of a user) and authorization (approval of a user to initiate a query) and has begun working closely with the GA4GH Security Working Group to define minimum standards to which each MME service must adhere in order to participate. Currently, these

practices are defined by systems developed by the initial set of linked matchmaker services but is expected to develop more formally over time and in collaboration with the GA4GH Security WG.

Informed consent policy

The MME worked closely with the GA4GH's Regulatory and Ethics Working Group and Consent Task Team on developing a proposal for informed consent for data sharing in the context of genomic matchmaking within the MME. We have distinguished two levels of matchmaking and different consent requirements based on the data shared and the probability of reidentifying the patient:

Level 1: No additional consent required. This level of matchmaking involves a data requester querying on a broad phenotype description or disease name using standardized terms or codes (Human Phenotype Ontology [HPO], OMIM, Orphanet) and/or candidate gene names ± variant type. This level of sharing is consistent with current clinical practice with low risk of possible reidentification and therefore specific patient consent for this activity is not required.

Level 2: Consent required. This level of matchmaking involves a data requester querying on a unique or sensitive phenotype description and/or sequence level and related information, such as defined variants and/or genomic datasets. This level of sharing requires consent from the patient. If the patient had previously consented to data being shared in an open or registered access database whose declared purpose involves data sharing for purposes consistent with those of this matchmaking, no additional consent is required.

The MME service in which data are deposited is responsible for ensuring patient data used in matchmaking is consented appropriately.

Matching Algorithms: Optimizing for Success

A key component of the success of the MME is implementing matching algorithms that balance sensitivity with specificity when executing matching algorithms. For example, if a case is annotated with a single candidate gene (gene X) and a defined condition (disease Y), a highly specific matching algorithm would require the gene and condition to be an exact match to return the result. However, matching algorithms could increase their sensitivity by allowing a case with any phenotype term that is a component of disease Y to also be returned. At the start of this program, when the number of MME services is few and the number of cases in each database is still limited, data contributors who are querying the MME may prefer matching algorithms that are less specific in hopes of having the highest sensitivity. However, as the MME scales and the number of cases deposited into each participating MME database grows, increased specificity and sophistication of matching algorithms will become critical.

It is also likely that data contributors will have different tolerances for being notified of matches on their data, with some only wishing to be notified of high-probability matches and others more tolerant of a range of results. To achieve this balance, some MME services have developed algorithms that have associated scores that can quantitate the specificity of a match. This allows contributors to specify their own threshold for notification of matches. It also allows the query results to be provided in a rank order.

It should be noted that the more detailed the query sent by the requester, the more information the recipient services will have at their disposal to sort cases in their database by relevance to the patient under query. With this additional detail, the query is more likely to result in successful and accurate matches, leading to a virtuous cycle that incentivizes data requestors to provide the greatest level of detail on their samples.

Table 1. MME Services

DECIPHER	https://decipher.sanger.ac.uk/
GeneMatcher	https://genematcher.org/
PhenomeCentral	https://phenomecentral.org/

Table 2. Databases Intending to Launch the MME API

Cafe Variome-based networks	http://www.cafevariome.org/
Broad Institute Rare Disease Analysis Portal	https://atgu.mgh.harvard.edu/xbrowse
ClinGen's GenomeConnect	http://genomeconnect.org/
GENESIS (GEM.app)	https://genomics.med.miami.edu/
Leiden Open Variation Database (LOVD)	http://www.lovd.nl/3.0
Monarch Initiative	http://monarchinitiative.org/
Platform for Engaging Everyone Responsibly (PEER)	http://www.geneticalliance.org/peer
RD-Connect	http://rd-connect.eu

At the start of the program, MME services have defined their own algorithms for matching. This allows groups to constantly innovate on approaches to matching, yet MME services will be able to provide their algorithms on GitHub for other sites to adopt. In addition, allowing each site to control their own algorithms is necessary given the unique data schemas that support each MME database. For example, some MME databases have not yet implemented the flagging of candidate genes and instead simply store variant call format (vcf) files containing all variation on each case. In this scenario, most cases would result in a match with any executed query given the presence of variation in most genes in the genome. As such, matching algorithms can be further specified, for example, to require the optional field of variant type that would only return matches if a gene contains a predicted truncating or *de novo* variant.

Launching the MME

Defining the key approaches and requirements for supporting the initial intended purpose of the MME has been a critical step in launching this program. However, equally important is the execution of the project to launch a functionally connected federated network of matchmaker services that can demonstrate the identification and return of useful and successful matches in response to user-initiated queries. Such success enables the ongoing discovery of novel genetic causes of disease. Listed here, and detailed in the Supporting Information, are the steps that have been achieved in launching the MME: (1) goals of the MME defined, (2) MME API developed, (2) MME core policies developed, (3) MME website launched, (4) matching algorithm principles defined, (5) API test phase, (6) MME test dataset developed, and (7) user interfaces developed to support queries.

These steps have resulted in the current status of the MME in which three of the participating databases, PhenomeCentral, GeneMatcher, and DECIPHER, are now capable of returning the results of queries from API-supported connections to other MME services (Table 1; Fig. 2). The next areas of focus for the MME are to aid in bringing new MME services onto the network (Table 2) and promoting use of the MME by the broader community. In addition, MME services will continue refining the matching algorithms and integrate additional supporting evidence for why a candidate gene has been flagged in a given case.

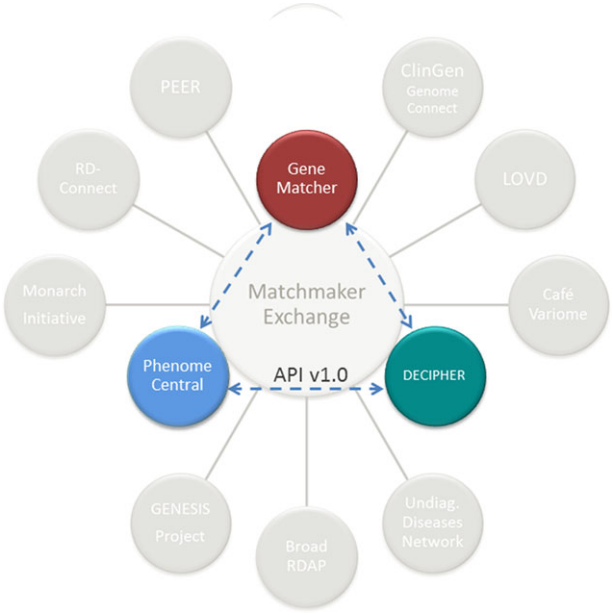

Figure 2. The current state of API-connected MME services. The figure depicts those databases that have implemented the MME API and satisfied the MME service requirements as described above. Additional databases are in the process of implementing the API and other MME service requirements. Progress can be monitored via the MME website (http://www.matchmakerexchange.org).

Table 3. Parameters Used for Matching and Score Output

MME service	Gene	Diagnosis	Phenotypic features	Provides match score output
PhenomeCentral	√	√	√	√
GeneMatcher	√	√		
DECIPHER	√		√	√

Guiding Community Use of the MME

The MME is a true federated system and as such, there is no single centralized entry point. Instead, users must choose one of the existing MME services as a starting point. In addition, in order to build the content of the MME over time, users must deposit their data in the point of entry into the MME. To guide users in where to deposit their data, Tables 3 and 4 provide a summary of the data fields that are maintained for each of the participating MME services and the parameters used for matching. Users may wish to deposit data in one system or another depending on the type of genotype and phenotype data associated with cases and how queries are supported.

Current State of the MME

The success of matching is directly related to the volume of cases that are deposited into the MME services and therefore, to identify all causes of rare disease, we will need to engage the community broadly in encouraging deposition of cases into the system. Building off the birthday paradox, the probability of a match increases with number of patient records that are matchable [Krawitz et al., 2015]. As such, even a small number of cases will begin yielding matches as has been demonstrated in the accompanying papers in this issue. After connecting these databases through the MME API, several additional matches have already been made between the Phenome Central and Gene Matcher Systems, including two promising hits undergoing further evaluation [Buske et al., 2015b]. Furthermore, implementation of the API is underway in other systems that will collectively bring on thousands of additional cases and model organism data from databases that have already been serving as matchmakers inside their own systems [Lancaster et al., 2015; Gonzalez et al., 2015; Mungall et al., 2015].

Evolving the MME

As outlined above, the initial launch of the MME is focusing on the simple matching of unsolved rare disease cases that share a common phenotype and candidate gene. However, additional uses of a federated case-level database containing genotypic and phenotypic data have not escaped the view of the MME. Large, shared datasets have been leveraged throughout the genomics era to identify the genetic basis of common and rare diseases. This has been through both hypothesis-free approaches such as GWASs [Altshuler et al., 2008] or PheWASs [Denny et al., 2010], as well as targeted approaches in Mendelian diseases.

As such, one goal of the MME is to expand the scope of discovery to allow matching in the absence of an identified candidate gene within the genomic dataset. Enabling broader, hypothesis-free approaches to discovery requires MME services to support deeper queries that can return data from entire genomic datasets as opposed to a small number of genes or variants flagged as potentially causal.

A second future goal of the MME is to expand the scope of analysis to genes and genomic variation already implicated in genetic disorders. In this scenario, the goal is to better define the phenotypic spectrum associated with individual genes as well as facilitate the understanding of specific variants identified in known disease genes. Use of sophisticated deep phenotyping approaches, combined with databases like the MME, can better objectively define the phenotypic spectrum of diseases. To support this, solved cases of Mendelian disease must be added and remain in the databases to gradually build larger datasets.

A third goal is to more effectively support the role of patient-initiating matchmaking in the MME. There are already examples of patients who have played such roles in identifying causes of rare disease [Lambertson et al., 2015] and the MME intends to better support their efforts. Two manuscripts in this special issue describe

Table 4. Types of Data Maintained by each MME Service

	Phenotype				Genotype				Candidates	
MME service	Name of condition	Diagnosis code	Phenotypic terms	Nonhuman models	Gene name	Chromosomal coordinates	Variants	VCF files	Flagged gene candidates	Evidence for gene candidates
PhenomeCentral (Canada)	√	√	√		√	√	√	√	√	
GeneMatcher (USA)	√	√	√	√	√	√			√	
DECIPHER (UK)	√		√		√	√	√		√	

how patients themselves have taken an interest in matchmaking and are creating their own systems both within and apart from the MME [Kirkpatrick et al., 2015, Lambertson et al., 2015].

A fourth goal of the MME effort is to contribute to the growing array of tools and strategies for broader data sharing and use. The first iteration of the MME enables investigators with unsolved rare disease cases to submit their patient data and thereby find each other and undertake selective data sharing. This balances support for gene discovery with a researcher's desire to protect resource investment in identifying candidate genes. Alternative methods could be used for matchmaking within controlled access and open access environments, some of which would allow researchers to query databases even without patient data in hand (or in situations where submission of patient data is not permitted). Many argue for a far more open environment for data sharing, which would drive scientific discovery in many more ways. For example, a researcher studying a biological pathway may hypothesize that genes in that pathway, when mutated, could cause disorders affecting a certain organ system and wish to validate that hypothesis in the absence of having access to real cases. If that researcher could query MME services for cases with relevant phenotypes and deleterious variants in pathway genes, such a hypothesis could be more quickly validated and form the basis for future studies. Similarly, researchers may wish to perform meta-analyses of large datasets to arrive at generalized conclusions as well as have access to large datasets to train algorithms for pathogenicity detection. To enable these types of investigations, MME systems will need to designate datasets and provide services that allow searching without requiring data deposition of a patient case. Some MME services already have apportioned some or all of their data for open interrogation such as DECIPHER [Chatzimichali et al., 2015] and the Monarch Initiative [Mungall et al., 2015], or enable direct searches within private networks as in the case of Cafe Variome [Lancaster et al., 2015]. Others services are committed to supporting such activities in the future.

Finally, now that a core federated network has been formed with successful implementation of the MME API v1.0, efforts will turn toward encouraging use of the MME and bringing new MME services into the network. We hope that the MME will grow into a large and vibrant community of commercial, clinical, and academic users who are committed to a federated model of data sharing for the advancement of science and medicine.

Conclusions

In summary, this paper provides an overview of the MME, from its founding principles and goals to the steps required to launch it as a robust platform for rare disease discovery. The ensuing papers in this special issue of Human Mutation define many of the individual matchmaker services already connected [Buske, et al., 2015a; Chatzimichali et al., 2015; Sobreira et al., 2015b), or intending to connect to the federated network [Lancaster et al., 2015; Kirkpatrick et al., 2015; Lambertson et al., 2015; Mungall et al., 2015], as well as other core components [Buske et al., 2015b] and concepts [Akle et al., 2015; Krawitz et al., 2015] that support genomic matchmaking. A few case examples of discoveries already made through use of matchmaking approaches are highlighted to add further support for this robust approach to rare disease gene discovery [Au et al., 2015; Jurgens et al., 2015; Loucks et al., 2015]. It is our hope that the success of the MME will serve as a model and foundation for innovative data sharing that leverages the increasing role of computational infrastructure to support the scaling of genomics as we collectively advance medicine and improve human health.

Acknowledgments

Members of the MME acknowledge the contributions of GA4GH and IRDiRC in advancing this collaborative initiative.

Disclosure statement: The following authors have a commercial conflict of interest: S. Zuchner is Chair of the Scientific Advisory Board of the not-for-profit charity The Genesis Foundation (501(c)(3); R. Gibbs is the acting C.S.O. of Baylor-Miraca Genetics Laboratories; A. Philippakis is a Venture Partner at Google Ventures.

References

Akle S, Chun S, Jordan DM, Cassa CA. 2015. Mitigating false-positive associations in rare disease gene discovery. Hum Mutat 36:998–1003.

Altshuler DM, Daly MJ, Lander ES. 2008. Genetic mapping in human disease. Science 322:881–888.

Au PYB, You J, Caluseriu O, Schwartzentruber J, Majewski J, Bernier FP, Ferguson M, Care for Rare Canada Consortium, Valle D, Parboosingh JS, Sobreira S, Innes AM, Kline AD. 2015. GeneMatcher aids in the identification of a new malformation syndrome with intellectual disability, unique facial dysmorphisms, and skeletal and connective tissue caused by de novo variants in *HNRNPK*. Hum Mutat 36:1009–1014.

Brownstein CA, Holm I, Ramoni R, Goldstein DB. 2015. Data sharing in the undiagnosed disease network. Hum Mutat 36:985–988.

Buske OJ, Girdea M, Dumitriu S, Gallinger B, Hartley T, Trang H, Misyura A, Friedman T, Beaulieu C, Bone WP, Links AE, Washington NL, et al. 2015a. PhenomeCentral: a portal for phenotypic and genotypic matchmaking of patients with rare genetic diseases. Hum Mutat 36:931–940.

Buske OJ, Schiettecatte F, Hutton B, Dumitriu S, Misyura A, Huang L, Hartley T, Girdea M, Sobreira N, Mungall C, Brudno M. 2015b. The matchmaker exchange API: automating patient matching through the exchange of structured phenotypic and genotypic profiles. Hum Mutat 36:922–927.

Chatzimichali E, Brent S, Hutton B, Perrett D, Wright CF, Bevan AP, Hurles ME, Firth HV, Swaminathan GJ. 2015. Facilitating collaboration in rare genetic disorders through effective matchmaking in DECIPHER. Hum Mutat 36:941–949.

Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, Wang D, Masys DR, Roden DM, Crawford DC. 2010. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics 26:1205–1210.

Gonzalez MA, VanBooven D, Hulme W, Ulloa RH, Lebrigio RF, Osterloh J, Logan M, Freeman M, Zuchner S. 2012. Whole genome sequencing and a new bioinformatics platform allow for rapid gene identification in D. melanogaster EMS screens. Biology 1:766–777.

Gonzalez MA, Lebrigio RFA, VanBooven D, Ulloa RH, Powell E, Speziani F, Tekin M, Schule R, Zuchner S. 2013. GEnomes Management Application (GEM.app): a new software tool for large-scale collaborative genome analysis. Hum Mutat 34:842–846.

Gonzalez M, Falk M, Gai X, Schüle R, Zuchner S. 2015. Innovative genomic collaboration using the GENESIS (GEM.app) platform. Hum Mutat 36:950–956.

Jurgens J, Sobreira N, Modaff P, Reiser CA, Seo SH, Seong M, Park SS, Kim OH, Cho T, Pauli RM. 2015. Novel COL2A1 variant (c.619G>A, p.Gly207Arg) manifesting as a phenotype similar to progressive pseudorheumatoid dysplasia and spondyloepiphyseal dysplasia, stanescu type. Hum Mutat 36:1004–1008.

Krawitz P, Buske O, Zhu Na, Brudno M, Robinson PN. 2015. The genomic birthday paradox: how much is enough? Hum Mutat 36:989–997.

Kirkpatrick BE, Riggs ER, Azzariti DR, Rangel Miller V, Ledbetter DH, Miller DT, Rehm H, Martin CL, Faucett WA. 2015. GenomeConnect: matchmaking between patients, clinical laboratories and researchers to improve genomic knowledge. Hum Mutat 36:974–978.

Lancaster O, Beck T, Atlan D, Swertz M, Dagleish R, Brookes AJ. 2015. Cafe Variome: general-purpose software for making genotype-phenotype data discoverable in restricted or open access contexts. Hum Mutat 36:957–964.

Lambertson K, Damiani S, Might M, Shelton R, Terry S. 2015. Participant-driven matchmaking in the genomic era. Hum Mutat 36:965–973.

Loucks CM, Parboosingh JS, Shaheen R, Bernier FP, McLeod DR, Seidahmed MZ, Puffenberger EG, Ober C, Hegele RA, Boycott KM, Alkuraya FS, Innes M. 2015. Matching two independent cohorts validates DPH1 as a gene responsible for autosomal recessive intellectual disability with short stature, craniofacial and ectodermal anomalies. Hum Mutat 36:1015–1019.

Mungall C, Washington N, Nguyen Xuan J, Condit C, Smedley D, Köhler S, Groza T, Shefchek K, Hochheiser H, Robinson P, Lewis S, Haendel M. 2015. Use of model organism and disease databases to support matchmaking for human disease gene discovery. Hum Mutat 36:979–984.

Robinson PN, Köhler S, Oellrich A; Sanger Mouse Genetics Project, Wang K, Mungall CJ, Lewis SE, Washington N, Bauer S, Seelow D, Krawitz P, Gilissen C, et al. 2014. Improved exome prioritization of disease genes through cross-species phenotype comparison. Genome Res 24:340–348.

Sobreira N, Schiettecatte F, Boehm C, Valle D, Hamosh A. 2015a. New tools for Mendelian disease gene identification: PhenoDB variant analysis module; and GeneMatcher, a web-based tool for linking investigators with an interest in the same gene. Hum Mutat 36:425–431.

Sobreira N, Schiettecatte F, Valle D, Hamosh A. 2015b. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat 36:928–930.

Swaminathan GJ, Bragin E, Chatzimichali EA, Corpas M, Bevan AP, Wright CF, Carter NP, Hurles ME, Firth HV. 2012. DECIPHER: web-based, community resource for clinical interpretation of rare variants in developmental disorders. Hum Mol Genet 21(R1):R37–R44.

Tryka KA, Hao L, Sturcke A, Jin Y, Kimura M, Wang ZY, Ziyabari L, Lee M, Feolo M. 2013. The NCBI handbook [Internet]. 2nd ed. Bethesda, MD: National Center for Biotechnology Information.

Washington NL, Haendel MA, Mungall CJ, Ashburner M, Westerfield M, Lewis SE. 2009. Linking human diseases to animal models using ontology-based phenotype annotation. PLoS Biol 7:e1000247.

Weinstein JN, Collisson EA, Mills GB, Shaw KM, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM, Cancer Genome Atlas Research Network. 2013. The cancer genome atlas pan-cancer analysis project. Nat Genet 45:1113–1120.

Zemojtel T, Köhler S, Mackenroth L, Jäger M, Hecht J, Krawitz P, Graul-Neumann L, Doelken S, Ehmke N, Spielmann M, Oien NC, Schweiger MR, et al. 2014. Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. Sci Transl Med 6:252ra123.

Zhang J, Baran J, Cros A, Guberman JM, Haider S, Hsu J, Liang Y, Rivkin E, Wang J, Whitty B, Wong-Erasmus M, Yao L, et al. 2011. International Cancer Genome Consortium Data Portal—a one-stop shop for cancer genomics data. Database 2011:bar026.