# 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

**Nara Sobreira**[1,*], **François Schiettecatte**[2], **Corinne Boehm**[1], **David Valle**[1,3], and **Ada Hamosh**[1,3]

[1]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

[2]FS Consulting LLC, Salem, Massachusetts 01970

[3]Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

## Abstract

Identifying the causative variant from among the thousands identified by whole-exome sequencing or whole-genome sequencing is a formidable challenge. To make this process as efficient and flexible as possible, we have developed a Variant Analysis Module coupled to our previously described Web-based phenotype intake tool, PhenoDB (http://researchphenodb.net and http://phenodb.org). When a small number of candidate-causative variants have been identified in a study of a particular patient or family, a second, more difficult challenge becomes proof of causality for any given variant. One approach to this problem is to find other cases with a similar phenotype and mutations in the same candidate gene. Alternatively, it may be possible to develop biological evidence for causality, an approach that is assisted by making connections to basic scientists studying the gene of interest, often in the setting of a model organism. Both of these strategies benefit from an open access, online site where individual clinicians and investigators could post genes of interest. To this end, we developed GeneMatcher (http://genematcher.org), a freely accessible Website that enables connections between clinicians and researchers across the world who share an interest in the same gene(s).

### Keywords

whole-exome sequencing; whole-genome sequencing; next-generation sequencing; Mendelian disease; mutation

## Introduction

The number of Mendelian disorders whose molecular basis is known has increased steadily over the last 5 years with the advent of whole-exome and whole-genome sequencing (WES

---

and WGS, respectively) and with the development of appropriate analytic strategies [Ng et al., 2010; Sobreira et al., 2010; Boycott et al., 2013; Beaulieu et al., 2014] (Fig. 1). Initially, these approaches were utilized in research laboratories to identify the genes responsible for Mendelian disorders and, more recently, in clinical laboratories as WES became a cost-effective diagnostic tool to solve puzzling clinical cases [Jacob et al., 2013; Yang et al., 2013; Lee et al., 2014; Yang et al., 2014].

Despite this progress, recognition of a causative variant from among thousands of sequence variants identified either by WES and WGS remains a challenge. For example, ~75% of cases evaluated by clinical WES are not solved [Yang et al., 2013; Lee et al., 2014; Yang et al., 2014]. Previously, as part of the Baylor–Hopkins Center for Mendelian Genomics (BHCMG), we developed PhenoDB, a Web-based system for managing and analyzing phenotypic/clinical and sequencing information [Hamosh et al., 2013]. Here, we describe the WES/WGS Sample Tracking and Variant Analysis Modules that have been added to PhenoDB to assist in the process of variant filtering and prioritization strategies. PhenoDB is now available in two versions: http://researchphenodb.net and http://phenodb.org. At http://researchphenodb.net, the entirety of PhenoDB, including the Sample Module and ELSI module for consent deliberation [Hamosh et al., 2013] is available for download at no cost. This version of PhenoDB is suggested for large projects with large numbers of samples to be sequenced. A simpler tool, http://phenodb.org, is also freely available for download or use as an online tool. Phenodb.org includes identifiers, associated clinicians, and different outcomes: in progress, solved, or unsolved, but does not include the Sample or ELSI Modules. We created phenodb.org for centers wishing to store and reanalyze VCFs from clinical WES sent on their patients. We suggest that to try the Variant Analysis Module, the user should create an account in phenodb.org.

Ultimately, to define a variant as being causative, we often require multiple unrelated individuals with a similar phenotype who have mutations in the same gene. This goal can be difficult to achieve because many Mendelian disorders are quite rare. Thus, sharing phenotypic and genotypic information about specific candidate genes can facilitate rapid and unambiguous identification of the causative variant and disease gene from a set of candidates. Such data sharing also connects basic scientists working on a particular gene, gene networks, and/or classes of phenotypes in model organisms with clinical investigators interested in the orthologous human phenotypes and genes. With this in mind, we developed GeneMatcher (http://genematcher.org), a freely accessible Web-based resource designed to enable connections between clinicians and basic scientists around the world who share an interest in the same or orthologous gene(s). GeneMatcher allows investigators to post genes of interest and connect with others posting the same genes. When a match occurs, each submitter receives an automatic email notification. Further communication is at the discretion of the submitters. GeneMatcher also has an option to match on OMIM phenotype numbers and genomic position. In the near future, we expect to add tools enabling matching based on phenotypic features. As part of the Matchmaker Exchange project (http://matchmakerexchange.org/), we have also developed an application programing interface (API; available upon request) that is being implemented and allows submitters to query other databases of genetic variants and phenotype information (e.g., PhenomeCentral [https://phenomecentral.org/], DECIPHER [https://decipher.sanger.ac.uk/], etc.).

## PhenoDB Variant Analysis Tool

### Overview

The PhenoDB variant analysis tool enables efficient and adjustable sequence analysis coupled to the clinical information housed in PhenoDB [Hamosh et al., 2013]. To enter PhenoDB in any capacity, it is necessary to register as a user. User authorizations are granted by a system administrator and are required for access to any of the different modules of the database (Fig. 2). The Variant Analysis Module, accessible from any other Module or Function in the submission, stores the clinical summary, variant call file (VCF) and ANNOVAR files [Wang and Hakonarson, 2010] from each member of a family under investigation, any information derived from a SNP array analysis, including any of the following file types: PLINK, CNV report, LOH report, B_Allele_Freq and LogR ratio chromosome plots, PCA plot, relatedness check, .ped, QC report. The analysis module also stores the deliberations and final conclusions regarding the analyses; the final results file including genes and variants that are likely causative for the disorder under consideration; and, the variant genotyping file which includes variants that were validated by an orthologous sequencing method and their segregation among other family members genotyped but not submitted for WES and/or WGS. We structured the data into fields that can be used to generate a report to the submitter and be displayed in summary tables.

### Analyze

The Analyze function, also accessible from any other Module or Function in the submission, allows the design of the data filtering strategy and variant prioritization and a PDF showing detailed steps of how to access and use it is found in the Supporting Information. The Analyze function page will display a table with the family members and their IDs, their affection status (adjustable if desired), and the ANNOVAR files for each individual sequenced. ANNOVAR is a software tool that functionally annotates genetic variants detected from genome sequence. Given a list of variants from WES or WGS, the TABLE ANNOVAR function will generate an Excel-compatible file with gene annotation, amino acid change annotation, SIFT scores, PolyPhen scores, LRT scores, MutationTaster scores, PhyloP conservation scores, GERP++ conservation scores, dbSNP identifiers, 1000 Genomes Project allele frequencies, NHLBI-ESP 6500 exome project allele frequencies, and other information [Wang et al., 2010]. The submitter selects the ANNOVAR files to be used in an analysis of a particular family and is also able to change the affected status of the chosen individuals. This is useful in cases where incomplete penetrance is suspected. Next, the submitter selects the inheritance pattern for each analysis trial (autosomal-recessive compound heterozygous; autosomal-recessive homozygous; X-linked recessive; autosomal-dominant new mutation; autosomal-dominant inherited mutation; autosomal-dominant variants). Each inheritance pattern follows a different set of rules:

- Autosomal-recessive compound heterozygous includes only the heterozygous variants identified in the proband (assumed to be affected); if there is more than one affected family members, the analysis include only the variants that are identified in all affected members; next, the analysis includes only the genes that have more than one variant in the proband but if the same set of variants in a gene is found in one

of the parents or in other unaffected family member then this gene (and its variants) is excluded of the analysis (Fig. 3A) [Hoover-Fong et al., 2014; Sobreira et al., 2015];

- Autosomal-recessive homozygous identifies homozygous variants that are shared by all affected individuals and excludes variants that are homozygous in an unaffected individual [Hoover-Fong et al., 2014;Migliavacca et al., 2014;Moldenhauer Minillo et al., 2014];

- X-linked recessive excludes X-linked variants found in a related unaffected male but retains X-linked heterozygous variants present in unaffected females;

- Autosomal-dominant new mutation excludes heterozygous variants that are also identified in a parent [Gripp et al., 2015];

- Autosomal–dominant inherited mutation retains heterozygous variants that are shared by affected individuals and excludes those found in unaffected individuals. Here, the submitter also has the option of looking for variants shared not only by all the affected individuals but by a subset of them, accounting for errors including misclassification as affected or failure to identify a variant by the WES/WGS in one of the affected individuals;

- Autosomal-dominant variants retains heterozygous variants with a minor allele frequency (MAF) less than the threshold selected for the 1000 Genome and Exome Variant Server and excludes variants found in a particular version(s) of dbSNP database, if any is selected [Gripp et al., 2015]. It should be used when only one individual is being analyzed.

The submitter then selects: (1) the types of variants to include in the analysis (protein coding exonic [missense, nonsense, indels, synonymous], splice site [20 bases into the introns], nonprotein coding exonic [3′UTR, 5′UTR]), with the option to exclude variants found in dbSNP 126, 129, and/or 131; (2) the MAF cutoff value for variant exclusion in the 1000 Genomes Project [The 1000 Genomes Project Consortium, 2012] and Exome Variant Server (release ESP6500SI-V2) databases; and (3) exclude/include X chromosome variants (Fig. 4). The analysis runs in less than a minute and generates tab-delimited and Excel files that can be stored indefinitely in the Analyze function page along with the date of creation and a log detailing each step and the options selected in the analysis process (Fig. 5). The analysis log succinctly and unambiguously describes the analysis as well as serving to control for and identify analytic errors (Fig. 3A). Analysis result lists can be deleted or retained and selected to populate the final results table provided in the submission view (Fig. 5). In each analysis result list, biological information from external databases such as OMIM (http://www.omim.org/); Mouse Genome Informatics (http://www.informatics.jax.org/); Gepis Tissue (http://research-public.gene.com/Research/genentech/genehubgepis/index.html); Gene Cards (http://www.genecards.org/) information including interaction network data (http://string905.embl.de/newstring_cgi/show_input_page.pl?UserId=MpGrSTzkKX50&sessionId=yfk6AXmEWLK1, http://ophid.utoronto.ca/ophidv2.204/index.jsp, http://mint.bio.uniroma2.it/mint/Welcome.do, http://www.uniprot.org/) and Gene Ontology data (http://amigo.geneontology.org/amigo); Pubmed

(http://www.ncbi.nlm.nih.gov/); ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/); UniProt (http://www.uniprot.org/); Intolerance Score and Percentile [Petrovski et al., 2013]; and CADD Score [Kircher et al., 2014] are added and can be used to guide the interpretation and prioritization of the candidate variants together with the information provided by ANNOVAR.

In the Family, Sample & VCF/ANNOVAR Files section of the Submission View, the submitter can upload the VCF and the program automatically converts it to an ANNOVAR file in the format utilized by the analytic program in both versions of PhenoDB (http:// researchphenodb.net and http://phenodb.org). If the submitter is using the online tool of phenodb.org, the VCF file is erased as soon as it is converted to the ANNOVAR file and the ANNOVAR file is available for analysis for 24 hr after the last analysis, after that it will also be erased but all the analysis result lists are stored in the Analyze function page. If the submitter uploads a combined VCF, the program offers a tool (tools section) that generates an individual VCF file for each individual in the combined file.

The analysis log also includes a list of up to 20 possible diagnoses generated by a search in OMIM based on the phenotypic features of the proband added by the submitter (Fig. 3B). PhenoDB terms have been mapped to human phenotype ontology (HPO) (http:// www.human-phenotype-ontology.org) terms to provide feature definition and synonyms. HPO terms where originally extracted from OMIM Clinical Synopses and match them closely. This allows PhenoDB to perform an OMIM search by taking the PhenoDB features selected for an individual of interest, constructing a search using the matching HPO terms, running that search against OMIM and returning the top 20 diagnoses for that search. The OMIM search is restricted to the OMIM phenotype entries and/or a specific inheritance pattern if it is specified in the submission and if the HPO terms are searched for as phrases. The matching OMIM documents are ranked using standard TF.IDF (term frequency/inverse document frequency); a document will be more relevant if a term occurs more frequently [Amberger et al., 2011; Köhler et al., 2014]. If a gene in the analysis result list is known to cause any of the top 20 suggested diagnoses, the gene is flagged for the submitter, thereby facilitating the recognition of a known disease gene(s) possibly responsible for the phenotype being investigated [Meloni et al., 2014; Migliavacca et al., 2014].

An alternative search tool based on phenotypic features is found in the Tools section where the submitter can also find a Diagnosis Search function created to allow for a diagnoses search independent of the entry being saved in PhenoDB. This search will not be saved.

**Filter**

From any other Module or Function in the submission (Fig. 2), the submitter can invoke the Filter function designed to facilitate the search for new and known disease genes and incidental findings. For example, the ANNOVAR file or analysis result list can be filtered to retain only the variants in the 56 genes in the ACMG incidental findings list [Green et al., 2013] or the variants in a selected gene list defined by the submitter. Any ANNOVAR file or analysis result list can also be filtered to select variants in an existing list of genes associated with a phenotypic series from OMIM. The submitter can also select for variants in genes that interact (first-, second-, or third-order interactions) with other genes of interest. For example,

if the submitter is interested on the RAS-MAP kinase pathway, they can filter any file to retain only variants in gene(s) that have a known second-order interaction with one or more selected genes in this pathway.

### Analyses

From the homepage, the submitter can access the Analyses section where he/she can create an analysis sandbox to upload and analyze ANNOVAR files from individuals not previously submitted as part of the original project. The sandbox is only available at http://researchphenodb.net. The submitter can also display all the analysis result lists available in the database and select a number of different lists from unrelated families and compare them asking which genes are mutated in any number of lists out of the total number of lists selected; this tool is very useful in the analysis of cohorts with or without locus heterogeneity, and it has proved to be very efficient [Hoover-Fong et al., 2014; Gripp et al., 2015].

A polygenic analysis is also available in the Analyses section. For this, the submitter can create multiple gene lists derived from unrelated probands/families and compare them asking for what set of probands/families have mutations in the same set of genes (two or more). For example, comparing the autosomal-dominant file of 10 unrelated probands, a possible result would be that of these 10 probands, four have mutations (different or not) in the same two genes.

### Analyses Search

From the homepage, the submitter can also access the Analyses Search section where he/she can search for a gene or variant (using bp position number) among all the analysis result lists generated for all the entries or narrow the search specifying the inheritance pattern of the analysis result lists or searching only among the final result lists.

### Samples Module

From the Submission or Analysis Modules, the submitter can also go to the PhenoDB Samples Module (Fig. 2). This provides a site where sample ID numbers are assigned and samples are tracked from the point of their receipt through their transfer to the sequencing laboratory. A sample_ID number is created for any subject who has been defined in the "Family and Samples" section of the PhenoDB entry. The system suggests a sample ID based on the individual's PhenoDB member_ID plus a sequentially numbered suffix. The user indicates the type of sample (e.g., blood, saliva, and tumor) associated with the sample_ID. Multiple sample_IDs can be assigned for each individual. The user can also enter his/her own independent sample ID. The system provides for the creation and storage of three types of manifests: intake, repository, and sequencing. It also provides for recording date of sample receipt, date sample sent to repository, date sent to sequencing laboratory, type of sequencing requested, laboratory doing the sequencing, and data returned from the sequencing laboratory. This information can be entered manually or by parsing data in a manifest file. Bulk data upload is available using an Excel spreadsheet. This can be used both to create new samples for the same or additional subjects and to add data to previously entered samples. All sample data can be downloaded as a tab-delimited file or Excel

spreadsheet. Standard queries allow sample data to be accessed based on several attributes, including submitter name, sample type, sample_ID, family or cohort_ID, and manifest.

## GeneMatcher

GeneMatcher (http://genematcher.org) is a freely accessible Website designed to enable connections between clinicians and researchers from around the world who share an interest in the same gene(s) or orthologous genes. The principle goal for making GeneMatcher available is to help solve "unsolved" exomes. This may be done with cases from research or clinical sources. No identifiable data are collected. GeneMatcher is also useful for basic scientists who have an interest in a gene or set of genes characterized in model organisms and who now wish to connect with clinical geneticists with human patients with mutations in the orthologous gene or genes. GeneMatcher was developed with support from the BHCMG as part of the Centers for Mendelian Genomics network. The site allows investigators to post a gene (or genes) of interest and connects investigators who post the same gene. The match is done automatically and, to protect privacy, the database is not searchable. When a match occurs, the submitters will automatically receive email notification. Follow-up is at the discretion of the submitters. Aside from the site administrator, no one has access to all the information in the database. Submitters have access to their own data and may edit it or delete it at will. Users create an account and submit gene(s) of interest (by gene symbol or base pair position). Alternatively, if the submitter has an account in any instance of PhenoDB, the genes of interest can be submitted directly from PhenoDB to GeneMatcher. There is also an option to provide and match diagnosis based upon OMIM® number, but this is not required. If a match is not identified at the time of submission, the genes of interest will continue to be queried by new entries. Genes or gene lists may also be left on the site even after a match has been identified. We have also developed an API (available upon request) that allows submitters to query other databases of genetic variants and phenotype information (e.g., PhenomeCentral [https://phenomecentral.org/]). In the future, we expect to enable matching based upon phenotypic features, to enable search for individuals with very rare or new Mendelian conditions with or without candidate genes.

## Discussion

Currently, there are several other WES analysis tools freely available that perform variant filtering and prioritization including eXtasy (http://homes.esat.kuleuven.be/_bioiuser/eXtasy/); Exomiser (http://www.sanger.ac.uk/resources/databases/exomiser/); PhenGen (http://phen-gen.org/index.html); PHEVOR (http://weatherby.genetics.utah.edu/cgi-bin/Phevor/PhevorWeb.html); VAAST (http://www.yandell-lab.org/software/vaast.html); KGGSEQ (http://statgenpro.psychiatry.hku.hk/limx/kggseq/); GeneTalk (https://www.gene-talk.de/dashboard); wANNOVAR (http://wannovar.usc.edu/); and VariantMaster (http://sourceforge.net/projects/variantmaster/).

Some of these tools use the phenotypic information as one of the ways to prioritize variants. For most, phenotypic features are only accepted if provided as an HPO term. PhenoDB uses its own PhenoDB terms that are mapped to HPO, Elements of Morphology (http://elementsofmorphology.nih.gov/), and ICHPT (data not published); the mapping of the terms

is available at https://www.phenodb.org/help/features. If the submitter wants to add features not found in the database, they can be written in free text boxes what allows for a more complete description of the patient. In PhenoDB, the features are stored and can be accessed or downloaded anytime during the variant analysis. The program uses the terms entered to suggest the 20 most likely diagnoses and the terms entered in the free text boxes are also used in the search. The genes associated with these suggested diagnoses are then flagged if present in an analysis result list. The diagnosis search and flagging of a final candidate gene associated with a suggested diagnosis allows for identification of known genes associated with known phenotypes. In the research setting, the causative gene may not be associated with any phenotype or features because not enough is known about the gene.

Additionally, for most of these other tools, the submitter can analyze the ANNOVAR file of only one individual per family. For the few that allow the analysis and comparison of multiple individuals per family, unrelated families cannot be analyzed and compared in a single analysis, for example, as a cohort. In a few of the other analysis tools, programing knowledge is necessary for the tool installation.

The PhenoDB variant analysis tool has several features that enable flexibility and connectivity to existing resources in variant analysis. Variant filtering is achieved, by analysis of one or more (as many as desired) VCF files in the same family or from different families according to the chosen Mendelian model. Moreover, PhenoDB, as in other software, allows flexible filtering according to any model of inheritance, type of variants (protein coding exonic [missense, nonsense, indels, and synonymous], splice site [20 bases into the introns], nonprotein coding exonic [3′UTR, 5′UTR]), with the option to exclude variants found in dbSNP 126, 129, and/or 131 and flexible MAF cutoff values for variant exclusion in the 1000 Genomes Project [The 1000 Genomes Project Consortium, 2012] and Exome Variant Server (release ESP6500SI-V2) databases. PhenoDB provides links to many publically accessible databases (described above) that provide known biological information about the genes and variants in the final variant list and can be used by the submitter in variant prioritization. Functions like filter, cohort analysis, and polygenic analysis (described above) are not found in the other variant analysis tools.

The PhenoDB variant analysis tool has proved to be easy, efficient, flexible, and fast. We developed it for the Centers for Mendelian Genomics project, an NHGRI/NHLBI funded initiative to ascertain the causal gene for unsolved Mendelian disorders using WES and WGS and as of November 2014, it has been used by BHCMG to analyze 462 sequenced samples (332 families). Using this tool, we have identified 32 novel disease genes and 41 known disease genes.

The utility of the PhenoDB variant analysis tool extends beyond this initial intent and is likely to benefit any laboratory performing or clinic analyzing WES/WGS to identify novel and known disease-causing genes and variants. This tool is freely available for institutions (toggle with or without personal health information and/or research centers [http://phenodb.org and http://researchphenodb.net], respectively). As of November 2014, 199 researchers and clinicians have downloaded PhenoDB (185 downloads

—

fromresearchphenodb.net and 14 downloads from phenodb.org) and 100 accounts have been created in the online version (http://phenodb.org).

In GeneMatcher as of November 2014, there are 713 genes from 200 submitters from 28 countries and 25 matches have been made enabling collaboration between clinicians and researchers from different countries and different backgrounds but with interest in the same genes. We performed a follow up of these 25 matches by personal communication and while most are still in progress, at least one successful match is described in a paper under review and refers to a match connecting a human phenotype to a mouse phenotype (Laura Reinholdt, personal communication). In our laboratory, we are also performing functional studies to confirm the pathogenicity of two de novo variants identified in the same gene in unrelated individuals with overlapping phenotype matched by GeneMatcher.

## Summary

In summary, we have developed two important tools (http://phenodb.org, http://researchphenodb.net and http://genematcher.org) to facilitate analysis of WES/WGS data and identification of the variants and genes responsible for rare Mendelian disease. These tools are freely available and will continue to be upgraded in the future.

## Acknowledgments

## References

Amberger J, Bocchini C, Hamosh A. A new face and new challenges for Online Mendelian Inheritance in Man (OMIM®). Hum Mutat. 2011; 32:564–567. [PubMed: 21472891]

Beaulieu CL, Majewski J, Schwartzentruber J, Samuels ME, Fernandez BA, Bernier FP, Brudno M, Knoppers B, Marcadier J, Dyment D, Adam S, Bulman DE, et al. FORGE Canada Consortium: outcomes of a 2-year national rare-disease gene-discovery project. Am J Hum Genet. 2014; 94:809–817. [PubMed: 24906018]

Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet. 2013; 14:681–691. [PubMed: 23999272]

Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O'Daniel JM, Ormond KE, Rehm HL, Watson MS, Williams MS, Biesecker LG. American College of Medical Genetics and Genomics. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013; 15:565–574. [PubMed: 23788249]

Gripp KW, Robbins KM, Sobreira NL, Witmer PD, Bird LM, Avela K, Makitie O, Alves D, Hogue JS, Zackai EH, Doheny KF, Stabley DL, Sol-Church K. Truncating mutations in the last exon of NOTCH3 cause lateral meningocele syndrome. Am J Med Genet A. 2015; 167:271–281. [PubMed: 25394726]

Hamosh A, Sobreira N, Hoover-Fong J, Sutton VR, Boehm C, Schiettecatte F, Valle D. PhenoDB: a new web-based tool for the collection, storage, and analysis of phenotypic features. Hum Mutat. 2013; 34:566–571. [PubMed: 23378291]

Hoover-Fong J, Sobreira N, Jurgens J, Modaff P, Blout C, Moser A, Kim OH, Cho TJ, Cho SY, Kim SJ, Jin DK, Kitoh H, et al. Mutations in PCYT1A, encoding a key regulator of phosphatidylcholine metabolism, cause spondylometaphyseal dysplasia with cone-rod dystrophy. Am J Hum Genet. 2014; 94:105–112. [PubMed: 24387990]

Jacob HJ, Abrams K, Bick DP, Brodie K, Dimmock DP, Farrell M, Geurts J, Harris J, Helbling D, Joers BJ, Kliegman R, Kowalski G, et al. Genomics in clinical practice: lessons from the front lines. Sci Transl Med. 2013; 5:194cm5.

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014; 46:310–315. [PubMed: 24487276]

Köhler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Bailleul-Forestier I, Black GC, Brown DL, Brudno M, Campbell J, FitzPatrick DR, Eppig JT, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res. 2014; 42:D966–D974. [PubMed: 24217912]

Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, Das K, Toy T, Harry B, Yourshaw M, Fox M, Fogel BL, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. JAMA. 2014; 312:1880–1887. [PubMed: 25326637]

Meloni VA, Guilherme RS, Oliveira MM, Migliavacca M, Takeno SS, Sobreira NL, deFatima Faria Soares M, deMello CB, Melaragno MI. Cytogenomic delineation and clinical follow-up of two siblings with an 8.5 Mb 6q24.2-q25.2 deletion inherited from a paternal insertion. 2014. Am J Med Genet A. 2014; 164A:2378–2384. [PubMed: 24898331]

Migliavacca MP, Sobreira NL, Antonialli GP, Oliveira MM, Melaragno MI, Casteels I, deRavel T, Brunoni D, Valle D, Perez AB. Sclerocornea in a patient with van den Ende-Gupta syndrome homozygous for a SCARF2 microdeletion. Am J Med Genet A. 2014; 164A:1170–1174. [PubMed: 24478002]

Moldenhauer, Minillo R.; Sobreira, N.; deFatima, de Faria Soares M.; Jurgens, J.; Ling, H.; Hetrick, KN.; Doheny, KF.; Valle, D.; Brunoni, D.; Alvarez, Perez AB. Novel deletion of SERPINF1 causes autosomal recessive osteogenesis imperfecta type VI in two Brazilian families. Mol Syndromol. 2014; 5:268–275. [PubMed: 25565926]

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet. 2010; 42:30–35. [PubMed: 19915526]

Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet. 2013; 9:e1003709. [PubMed: 23990802]

Sobreira NL, Cirulli ET, Avramopoulos D, Wohler E, Oswald GL, Stevens EL, Ge D, Shianna KV, Smith JP, Maia JM, Gumbs CE, Pevsner J, et al. Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene. PLoS Genet. 2010; 6:e1000991. [PubMed: 20577567]

Sobreira N, Modaff P, Steel G, You J, Nanda S, Hoover-Fong J, Valle D, Pauli RM. An anadysplasia-like, spontaneously remitting spondylometaphyseal dysplasia secondary to lamin B receptor (LBR) gene mutations: further definition of the phenotypic heterogeneity of LBR-bone dysplasias. Am J Med Genet A. 2015; 167A:159–163. [PubMed: 25348816]

The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1092 human genomes. Nature. 2012; 491:56–65. [PubMed: 23128226]

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from next-generation sequencing data. Nucleic Acids Res. 2010; 38:e164. [PubMed: 20601685]

Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z, Hardison M, Person R, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med. 2013; 369:1502–1511. [PubMed: 24088041]

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

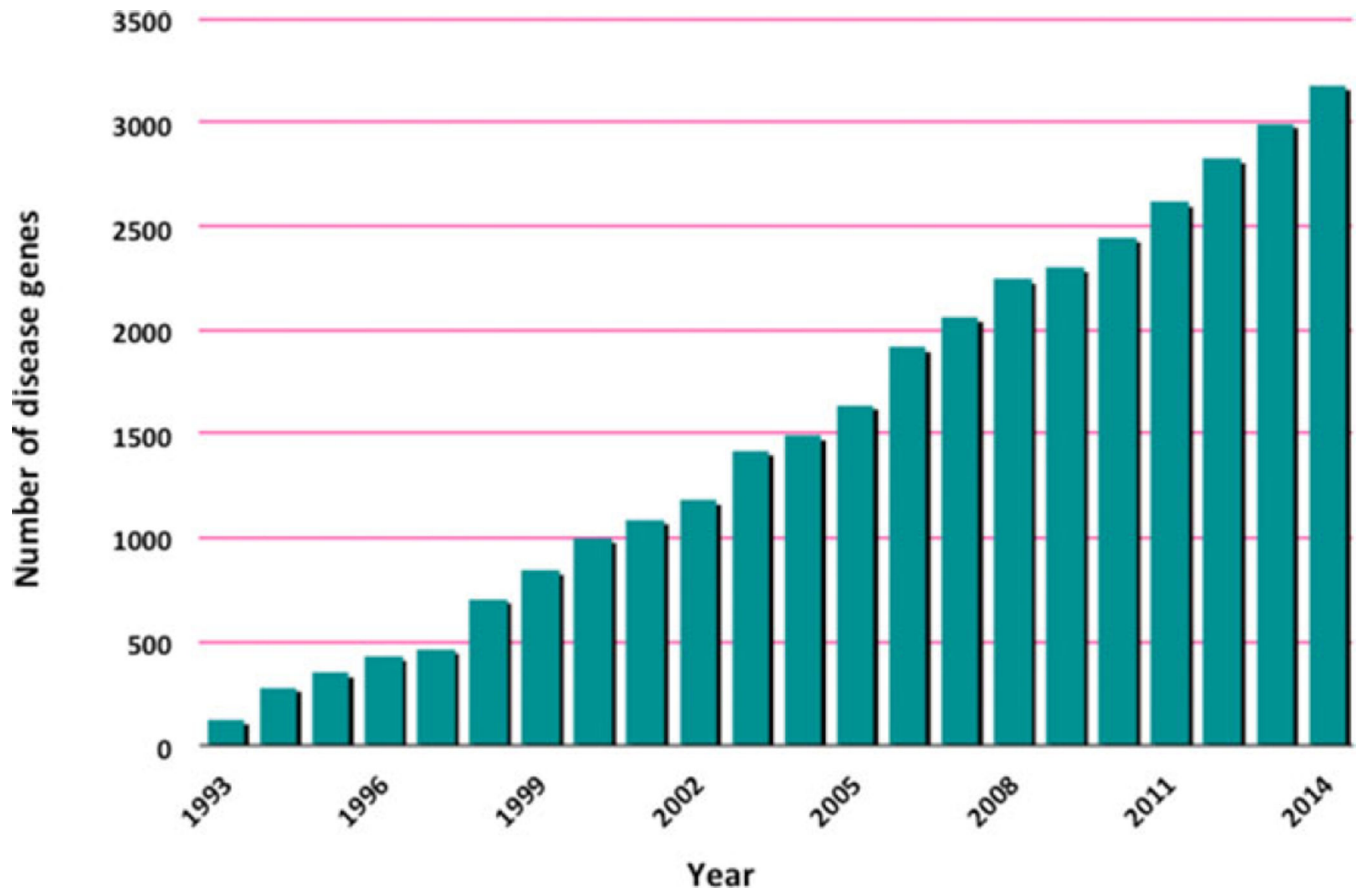**Figure 1.**
Increase in the number of genes identified as responsible for Mendelian phenotypes per year
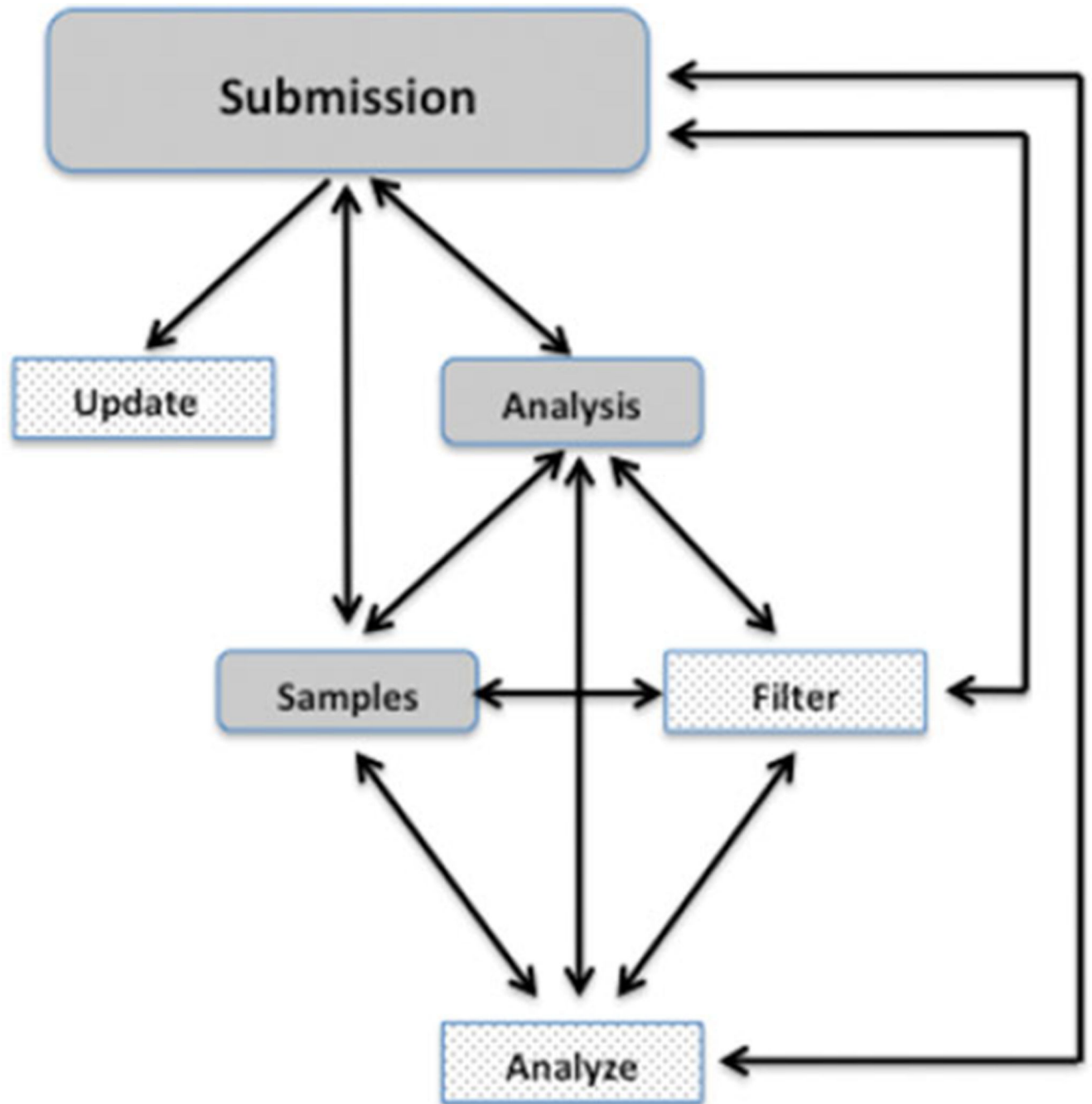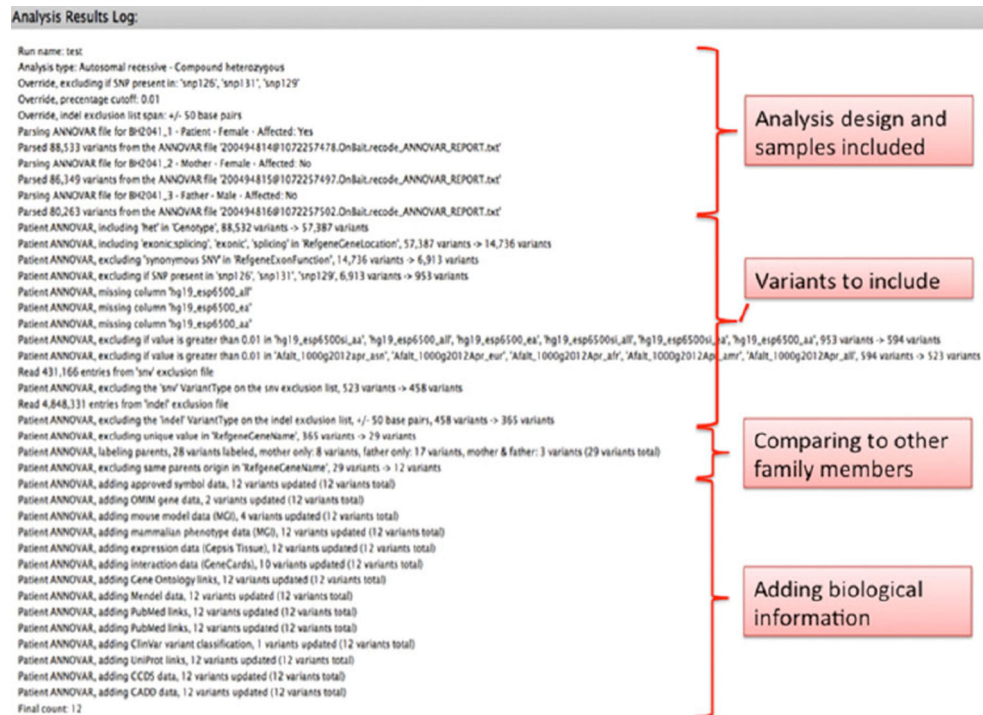(source: Online Mendelian Inheritance in Man, OMIM).

**Figure 2.**
Scheme of PhenoDB Modules (gray-rounded rectangles) and Functions (stippled rectangle).
The arrows show that the submitter can go to any Module or Function from any Module or
Function.

**Figure 3.**
The analysis log of an example autosomal recessive-compound heterozygous analysis (A) and the OMIM diagnosis search results based on the phenotypic features of the proband (B). If any of the candidate genes listed in the "Final count" is responsible for one of the diagnoses suggested by the OMIM search, it will be flagged.

**Figure 4.**
View of the Analyze function page showing the analysis design.

**Figure 5.**
View of the Analyze function page showing the storage of completed analyses.