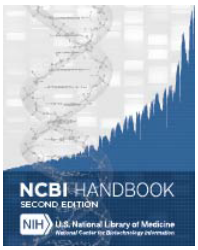
U.S. National Library of Medicine
National Center for Biotechnology Information

NLM Citation: Halavi M, Maglott D, Gorelenkov V, et al. MedGen. 2013 May 28 [Updated 2018 Dec 11]. In: The NCBI Handbook [Internet]. 2nd edition. Bethesda (MD): National Center for Biotechnology Information (US); 2013-.
Bookshelf URL: https://www.ncbi.nlm.nih.gov/books/



MedGen

Maryam Halavi, MD, PhD,[1] Donna Maglott, PhD,[1] Viatcheslav Gorelenkov, MS,[1] and Wendy Rubinstein, MD, PhD[1]

Created: May 28, 2013; Updated: December 11, 2018.

Scope

MedGen is NCBI's portal to information about human disorders and other phenotypes having a genetic component. MedGen is structured to serve health care professionals, the medical genetics community, and other interested parties by providing centralized access to diverse types of content. For example, because MedGen aggregates the plethora of terms used for particular disorders into a specific concept, it provides a Rosetta stone for stakeholders who may use different names for the same disorder. Maintaining a clearly defined set of concepts and terms for phenotypes is essential to support efforts to characterize genetic variation by its effects on specific phenotypes. The assignment of identifiers for those concepts allows computational access to phenotypic information, an essential requirement for the large-scale analysis of genomic data.

Once a concept is defined, MedGen offers a growing collection of attributes about that concept including a definition or description, clinical findings, causative genetic variants and the genes in which they occur, available clinical and research tests, molecular resources, professional guidelines, original and review literature, consumer resources, clinical trials, and Web links to other related NCBI and non-NCBI resources. Convenient access to such a range of supporting data allows MedGen's users to synthesize and apply the latest knowledge to important clinical and biological questions.

History

There are multiple public databases, ontologies, or tools that provide terms, definitions, and other information about human diseases and phenotypes. None, however, is focused on maintaining an up-to-date information resource that both harmonizes terminology data from others and also provides an interface to use those harmonized terms to identify related information in the current public arena.

Recognizing this broad challenge and the collective interest to address it, MedGen was initiated in 2012 as a public resource in the National Center for Biotechnology Information (NCBI) at the National Institutes of Health (NIH), National Library of Medicine (NLM). Seeded from terms established from NLM's Unified Medical Language system (UMLS®), the NIH Genetic Testing Registry (GTR®), and ClinVar; MedGen continues to add terms, types of content, and services.

As of August 2013, more than 170,000 concepts had been integrated in MedGen. Many aspects of MedGen and its data gathering procedures are evolving, so users' suggestions and comments are welcomed.

Author Affiliation: 1 NCBI; Email: maglott@ncbi.nlm.nih.gov; Email: viatcheslav.gorelenkov@nih.gov.

Data Model

MedGen has a very simple data model. Once a concept is identified, categories of information that relate to that concept are identified and reported. These categories may be descriptors, links to other databases, or concept-concept relationships.

A MedGen record has several important components including:

Concept-related

***Concept*:** An aggregation of terms from multiple vocabulary sources, which have been determined to be comparable (i.e., represent the same concept).

***ConceptID (CUI—Concept Unique Identifier)*:** A unique stable identifier assigned to each concept.

***Semantic type*:** A defined set of categories of concepts, so that concepts that share terms can be differentiated by scope. An example is the distinction between 'autism' as a synonym for the diagnosis Autism spectrum disorders (C1510586) and 'autism' as a clinical feature or finding (CN000674) identified in multiple disorders.

Descriptors

MedGen maintains multiple types of descriptors, including names, acronyms, abbreviations, sources of descriptors, attribution for and identifiers used by those sources, cytogenetic locations, mode of inheritance, textual definitions or descriptions, types of genetic testing registered in the NIH Genetic Testing Registry (GTR®), genes for which aberrant function may be related to the disorder, related variations, professional guidelines, consumer resources, and more as detailed below (see Table 1). Sources of terms are called 'vocabularies', consistent with the usage of UMLS® (1). The descriptors are often presented in the query interface as distinct fields or as concepts' properties.

Table 1. A list of data elements aggregated in MedGen and their sources

#	Data Type	Source
1	Clinical and research tests	GTR®
2	Clinical features	HPO
3	Clinical trials®	ClinicalTrials.gov
4	Concept ID	UMLS® or GTR®
5	Consumer resources	Genetics Home Reference, Genetic Alliance, Genetic and Rare Diseases Information Center, MedlinePlus
6	Cytogenetic location	NCBI annotation
7	Gene	NCBI's Gene
8	Links to other NCBI's resources	NCBI's resources such as Gene, MeSH®, ClinVar, Bookshelf, BioSystems, etc.
9	MedGen Identifier	MedGen
10	Mode of inheritance	OMIM® /ClinVar/GTR®
11	Molecular resources	Coriell Institute for Medical Research
12	Professional Guidelines	NCBI curation
13	RefSeqGene	RefSeqGene
14	Reviews	*GeneReviews*™, PubMed Clinical Queries
15	Semantic type	UMLS®

Table 1. continued from previous page.

#	Data Type	Source
16	SNOMED CT® terms	SNOMED CT®
17	Source Identifiers	Various sources, such as OMIM®, HPO, etc.
18	Terms definitions	*GeneReviews*™, Medical Genetics Summaries, etc.
19	Terms, acronyms, synonyms	Defined vocabularies
20	Terms hierarchies	GTR®, MedGen
21	Variations	ClinVar

Definitions/Descriptions

Terms imported to MedGen may have been associated with a definition by their source or in other sources. MedGen includes a single definition for each concept to be displayed in the search summaries. In case of multiple definitions, a simple prioritization is used: the first priority is assigned to the definitions from GTR®, followed by SNOMED CT®. Subsequent to that order, MedGen adheres to the UMLS® prioritization of source vocabularies and term types. In the full report for a concept MedGen may report multiple definitions. Any definition displayed in MedGen includes attribution and a link to the source.

Identifiers from source databases

MedGen maintains and displays identifiers from source databases not only to provide attribution, but also to support interactive and programmatic searching and links to sources' websites. MedGen also maintains alternative IDs from the Human Phenotype Ontology (HPO) in the current record of that HPO concept.

Gene-phenotype relationships

A disease concept in MedGen reports the symbol(s) of genes that are reported to be causative. Links are also provided to NCBI's Gene database and OMIM®.

Cytogenetic locations

Cytogenetic locations for each concept are reported from ClinVar based on location of the genes that contribute to that disorder.

Inter-concept relationships

MedGen maintains two major types of inter-concept relationships. The first is the diagnosis-clinical feature relationship, so that the user can see all features reported for a disorder and find all disorders that share a clinical feature. The second type is hierarchies. MedGen currently provides three types:

- Clinical features, from top-level to each child, no matter what the final level. This is used to group clinical features by type
- Disorders: computed from parent-child relationships
- Disorders: curated by GTR®/ClinVar staff.

Published literature

MedGen's Web interface provides information about related publications in multiple ways. One is curation, by NCBI staff, of professional guidelines related to a disorder. The second is aggregation of publications from contributors ClinVar and GTR®, which are reflected in the links to PubMed. The Recent clinical studies section

uses PubMed's Clinical Queries logic. Finally, the preferred name of the MedGen record is used to query PubMed and identify highly related literature.

Dataflow

Data in MedGen are acquired both programmatically and manually via curation, depending in part on the type of information and the sources for that information. This section summarizes those flows organized by type of information.

The first step in organizing the information is to establish a concept that defines a disorder or phenotype, classify that concept by type, and then assign that concept a stable unique identifier. With that framework established, data are then aggregated around that concept. The extent of metadata attributed to each concept may vary based on the availability. However, to ensure maximized benefit of having current, correct, and complete metadata, relevant metadata are actively managed in MedGen.

Concept UID

Concepts' unique identifiers (Concept UID or CUI) are assigned to each concept to facilitate connecting different terms from various vocabularies to that concept. The Concept UIDs in MedGen are either derived from UMLS® or assigned by MedGen (starting with CN) if a match based on term and semantic type in UMLS® cannot be identified. UMLS® is maintained by NLM and provided to researchers on the terms of license agreement without any charge. Terms in UMLS® are classified based on broad categories of semantic types and term relations (2). Unique identifiers assigned in UMLS® are permanent Concept UIDs, however, in each semi-annual UMLS® release, some of the Concept UIDs can be merged or deleted. If so, concept UIDs calculated by MedGen are deprecated in favor of those from UMLS® for that concept. Concept UIDs also may be merged or deleted either because of vocabulary changes or because of NCBI internal curation. If a concept semantic type is in the scope for MedGen, Concept UID is imported to MedGen programmatically. MedGen maintains the history of the UMLS® Concept UID merges and deletions to ensure stable and permanent identifiers.

Names, acronyms, abbreviations (terms)

MedGen integrates large sets of terms, their relationships, and their definitions (if available), as well as additional supplementary information from a variety of sources (termed vocabularies).

- ClinVar, daily
- Human Phenotype Ontology (HPO), weekly
- Genetic Testing Registry (GTR®), daily
- Medical Subject Headings Thesaurus (MeSH®), semi-annually via UMLS®
- National Cancer Institute Thesaurus (NCIt), semi-annually via UMLS®
- Online Mendelian Inheritance in Man (OMIM®), daily
- Systemized Nomenclature of Medicine—Clinical Terms (SNOMED-CT®), semi-annually via UMLS®

Names and acronyms for each concept are aggregated from these vocabulary sources. The preferred name and preferred acronym are either internally curated or selected based on the UMLS® standards. Alternate terms derived from other vocabularies are reported as synonyms for each concept.

MedGen restricts the processing of concepts from UMLS® to a subset of semantic types (disease or syndrome, abnormality and dysfunction, sign and symptom, finding, molecular and pathological function, pharmacologic substance, neoplastic process, etc.). However, one concept can be associated with more than one semantic type and reported more than once with different semantic types. A semantics-aware mapping approach is used to maintain useful associations between concepts and support their bi-directional multi-level relationships.

Names from OMIM® (3) are processed from both UMLS® and from daily updates directly from OMIM®. Terms from HPO, as a primary source for clinical features of Mendelian disorders, are updated weekly. Terms from GTR® are mainly based on what was provided by the submitters during a test registration, but curators will review the evidence for each submission. Because SNOMED-CT® has many concepts that are not in scope for MedGen, MedGen does not represent all of SNOMED CT®, but only SNOMED CT's terms for concepts in MedGen.

If a new vocabulary source is identified, MedGen will integrate the data based on evaluation of terms and semantic types, maintaining the source of the data (vocabulary) and the identifier used by the source. If a term matched an existing concept, the term and source will be added; otherwise a new CUI will be established.

Genetic/genomic characteristics (MOI, cytogenetic location)

All modes of inheritance for a disorder are extracted from the resources reporting the term. Gene symbols and cytogenetic locations associated with a disorder concept are derived from NCBI's Gene database (which uses the HUGO Gene Nomenclature Committee (HGNC) standard) based on the gene-disorder relationships. Figure 1 shows an example of how these data are displayed on the website.

Clinical feature-disorder relationships

Clinical features describing the sign and symptoms characteristic of a disorder are provided from HPO (4) and updated weekly. The phenotype abnormalities in HPO are categorized in 20 organ abnormality groups and MedGen uses the same categorization for its reporting of clinical features in a hierarchical format. MedGen enhances access to all conditions with this clinical feature by providing an option to search on a specific clinical feature. Figure 2 illustrates the Clinical features section in MedGen.

Term hierarchies

Term hierarchies are constructed based on relationships reported for each concept as direct or indirect links between terms from the vocabulary sources. This enables users to expand their search queries and browse terms relationships. MedGen represents hierarchies as trees in which the terms are arranged having a root (top level node) and many branches (children), which can be in the same level or below their parent. The MedGen hierarchy is constructed based on either direct links for each concept or extending links vertically on the hierarchy tree toward the parents and the children (traveling 3 levels upward to find a common parent and then downward to find related children). Alternative hierarchies such as GTR® are also provided by a tabbed navigation, as it is shown in Figure 3. The concepts in GTR® hierarchy are displayed alongside any available links to Clinical tests, Research tests, OMIM®, or *GeneReviews*™. Figure 3 illustrates the display of both GTR® hierarchy (A) and MedGen hierarchy (B) in the Term Hierarchy section.

Available testing

The clinical tests and research tests registered in the Genetic Testing Registry (GTR®) are mapped to each relevant disorder concept and hyperlinked to full test records in GTR®. This convenient access, as displayed in Figure 4, improves users' ability to view the test's purpose, methodology, validity, evidence of a test's usefulness, and laboratory contacts and credentials. The tests are grouped according to the primary method used, as reported to GTR® (5).

Professional guidelines

The relevant clinical practice guidelines, position statements, and recommendations from various sources, such as American College of Medical Genetics and Genomics (ACMG), Evaluation of Genomic Applications in Practice and Prevention (EGAPP), American Congress of Obstetricians and Gynecologists (ACOG), The

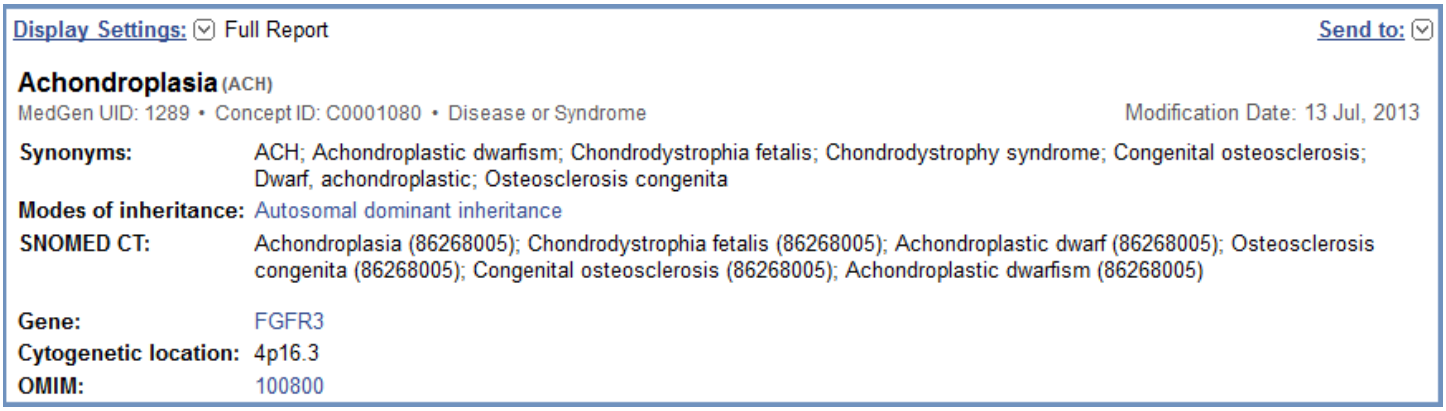


Figure 1. Names, acronyms, identifiers, MOI, and Cytogenetic location displayed on MedGen website. Top section of the full report in MedGen includes the title for the concept, IDs associated with the concept, semantic type, synonyms, mode of inheritance, related terms from SNOMED CT®, gene, cytogenetic locations, and if available MIM identifier from OMIM®.

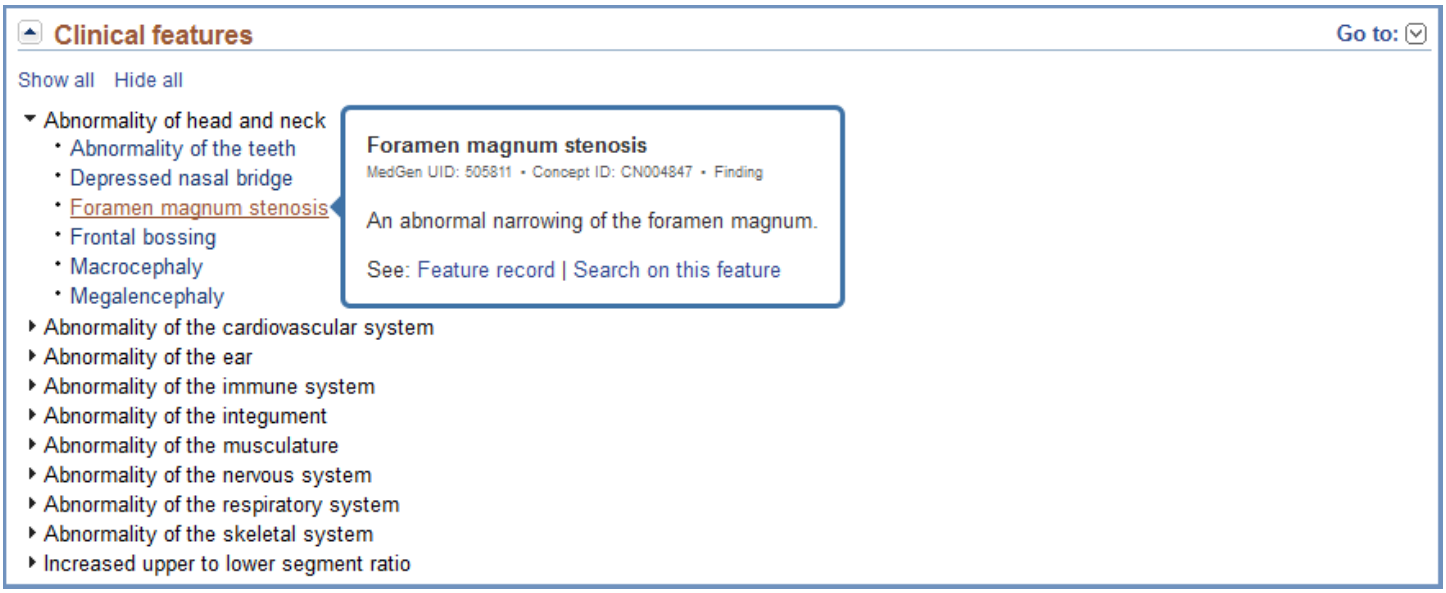


Figure 2. Clinical feature section in MedGen. Clinical features are reported under top nodes established by HPO (top nodes are immediate children of the HPO term Phenotypic abnormality - HP:0003812). Each node can be expanded or collapsed to improve viewing of all items. A click on any item in the list displays a pop-up window with the definition of that feature and links to either the full report in MedGen for that feature (Feature record) or other conditions reported to have that feature (Search on this feature).

Clinical Pharmacogenetics Implementation Consortium (CPIC), The National Society of Genetic Counselors (NSGC), etc., are curated and associated with related concepts (currently, 258 guidelines have been curated for 462 conditions). To facilitate access to these guidelines, MedGen has a dedicated section in its full report (Professional guidelines section), which provides hyperlinks to either PubMed or PMC (if available) or to other online source for each guideline.

Publications

Highly relevant literature and publications are useful in assessing the nature and importance of a concept as well as expanding a user's perspective about a concept. In MedGen, these highly relevant literatures are aggregated through use of comprehensive and specialized queries on PubMed and PubMed Central (PMC). The citations are not directly stored in MedGen, but are retrieved dynamically to keep the content and links up-to-date.

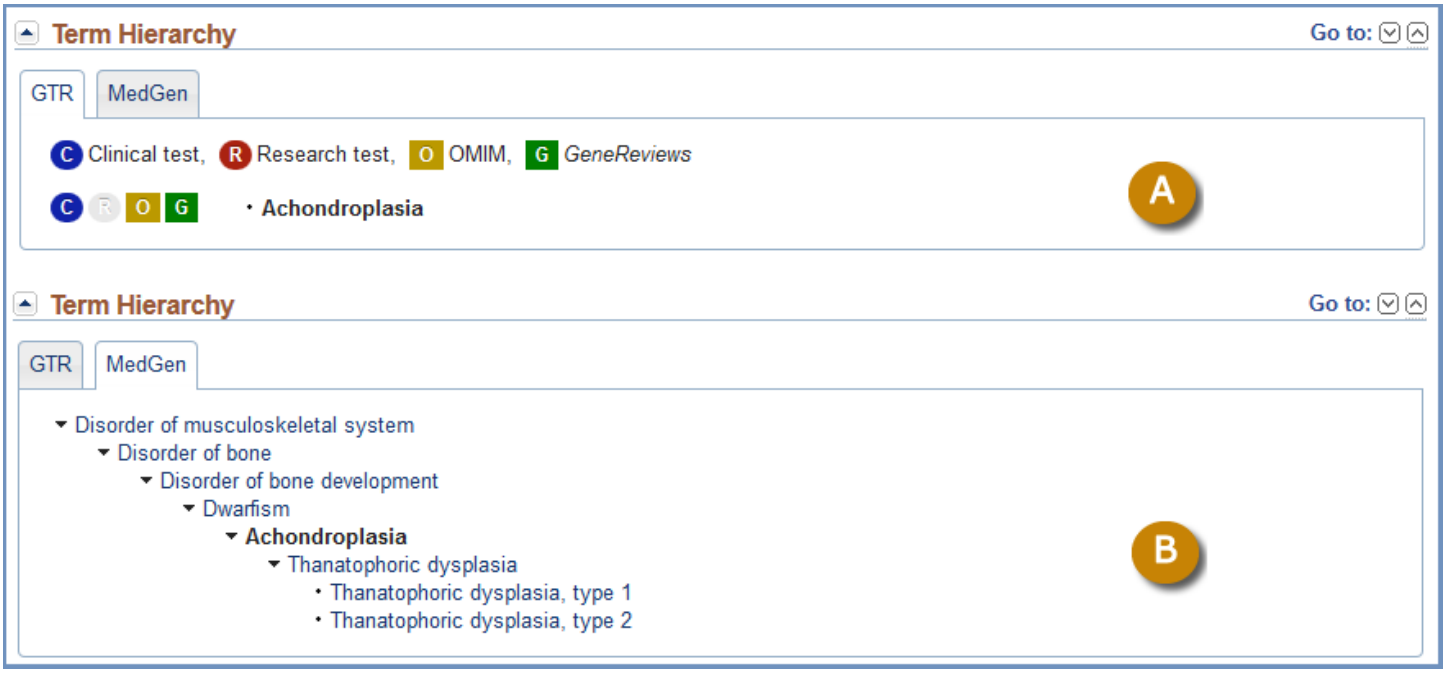


Figure 3. Term hierarchies in MedGen. The hierarchies are illustrated in tabbed navigation format. (A) Illustrate the GTR® hierarchy alongside with icons, which provide links to corresponding Clinical tests, Research tests, OMIM®, or *GeneReviews*™ Records. (B) Illustrate the hierarchy reported in MedGen. Small arrows on the left allow for expansion and contraction of the branches of a large hierarchical tree to ease navigation.

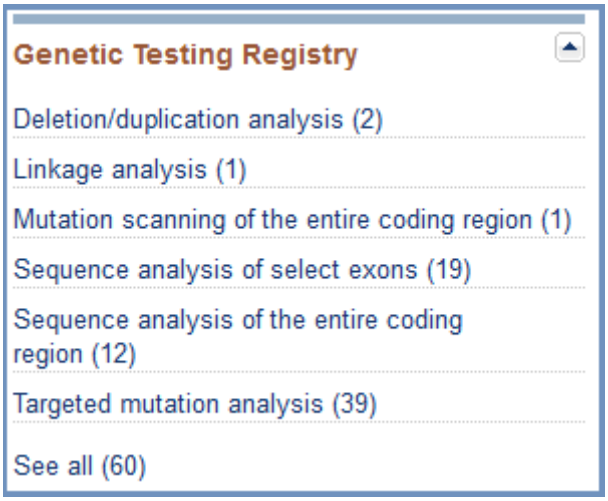


Figure 4. The clinical and research tests associated with each concept. The tests are listed according to the primary method used. The total number of available tests is shown and the option "See all" enables access to all test records registered for each concept.

Direct queries to PubMed and PMC are executed either based on the relevant terms from Medical Subject Headings (MeSH®) as query term; or if there is no MeSH® term connection, by using the preferred name of the concept, its synonyms, or its acronyms as query terms. Both PubMed and PMC use a ranking system for pushing more relevant results to the top of the results set. However, to increase the specificity of the results MedGen uses filters such as English language, human species, having abstract, genetics subheading, etc. If the number of the articles returned by these direct queries is very large, an arbitrary number may be used as a cutoff point to keep the results manageable. If no result is generated by using these queries, then a graphical model is used. In this graphical model MedGen combines logical structures and probabilities to create a flexible framework for finding

related articles based on relationship to associated disorders, genetic mechanism of the disorders, body sites, or unique ingredients for drugs.

Although not strictly part of its dataflow, MedGen also provides access to publications at the time of the display of a full record. In other words, the name of the record is submitted to PubMed via PubMed Clinical Queries. These clinical queries are filtered by five different Clinical Study Categories (Etiology, Diagnosis, Prognosis, Therapy, and Clinical prediction guides) and Systematic Reviews. Use of the broad scope option provides maximum coverage of the relevant literature and additional filters (namely English language and human, and not comment publication types or letter publication types) increase specificity. The results are displayed in "Recent clinical studies" and "Recent systematic reviews" sections, respectively. In each section, the title, list of authors, and journal information is displayed. Figure 5 illustrates the display of the literature in various sections.

Gene-disorder relationships

Gene symbols in MedGen are limited to current or previous official symbols from the HGNC (Data sources for Gene). If there is no official symbol from the HGNC, then NCBI Gene's preferred symbol is used. The data supporting the gene-phenotype relationship is built primarily from OMIM® with review from NCBI staff.

Molecular resources

Concepts in MedGen are mapped to several resources in the area of molecular medicine based on data in ClinVar and Gene. This allows users to explore additional molecular information such as related sequence, its location, variations, etc. For example, MedGen provides access to relevant genomic sequences, which are reference standards for well-defined genes reported in RefSeqGene; link to relevant records in Coriell Institute for Medical Research, which provides essential research reagents to the scientific community and links to ClinVar and Gene.

Consumer resources

MedGen also actively seeks and provides direct access to available consumer-friendly information related to each concept. Submitters provide URLs connected either directly or indirectly to Concept IDs. Figure 6 illustrates how these sections are displayed in MedGen.

Additional information

MedGen provides links to clinical resources such as ClinicalTrials.gov (see Figure 6). This web-based resource is maintained by the National Library of Medicine (NLM) at the National Institutes of Health (NIH) and provides patients, their family members, health care professionals, researchers, and the public with easy access to information on publicly and privately supported clinical studies of human participants conducted around the world. Figure 7 illustrates how these sections are displayed in MedGen.

MedGen data is further supplemented by providing links to curated relevant reviews for each concept. The reviews can be selected from *GeneReviews*™ or reviews in PubMed. The *GeneReviews*™ are exclusively published online on NCBI's Bookshelf and are expert-authored, peer-reviewed disease descriptions presented in a standardized format and focused on clinically relevant and medically actionable information on the diagnosis, management, and genetic counseling of patients and families with specific inherited conditions. PubMed Clinical Queries provides an interactive interface to discover citations for medical genetics content such as systematic reviews, meta-analyses, reviews of clinical trials, evidence-based medicine, consensus development conferences, and guidelines. Other reviews in PubMed are retrieved by querying the concept term in PubMed and limiting the results to human species and setting review as publication type.

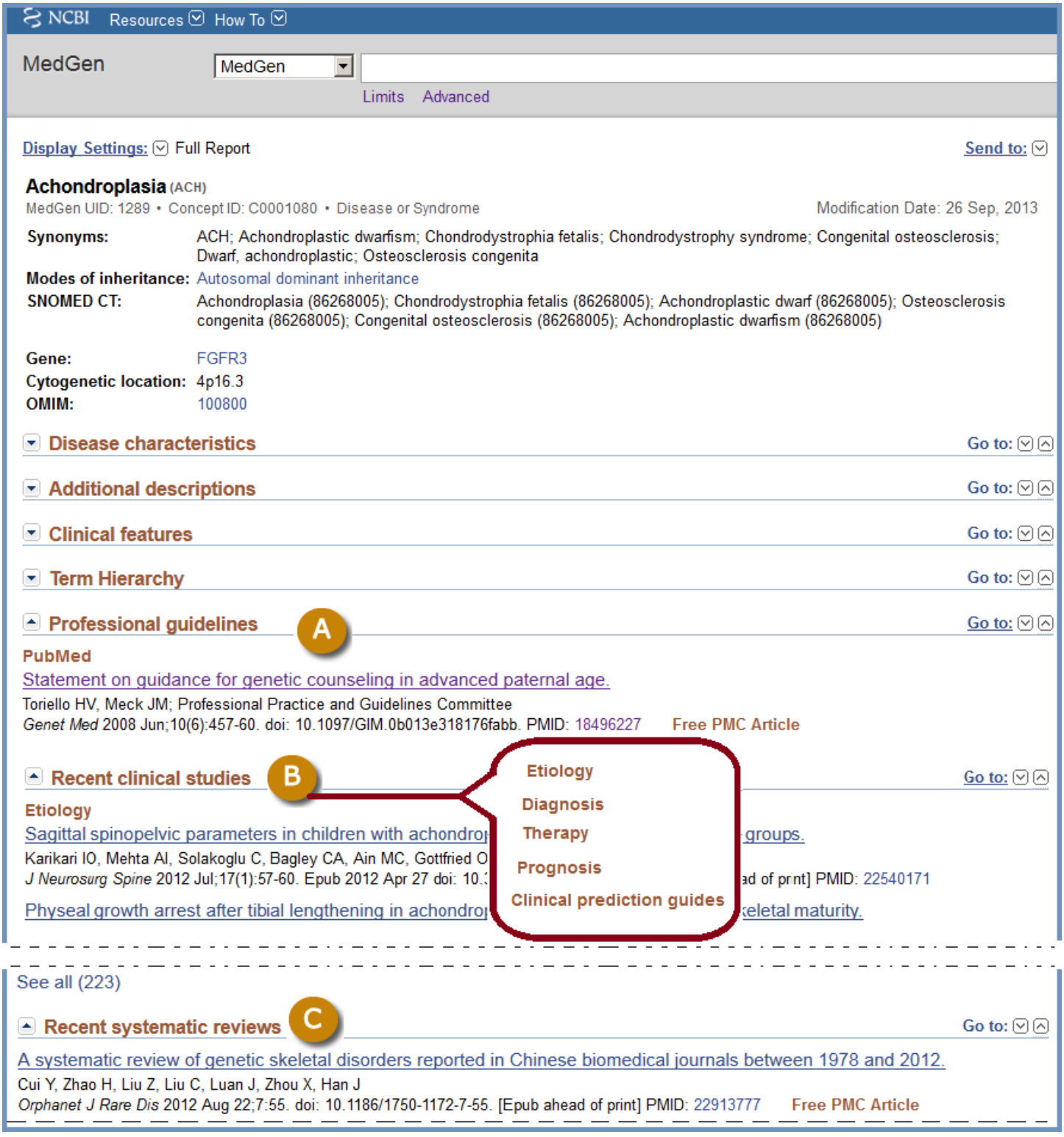


Figure 5. The professional guidelines and clinical literature sections in MedGen. (A) For each professional guideline the title, list of authors, and the journal information are presented in a "Professional guidelines" section, which include PMID and links to PMC, if available. The results from clinical queries in PubMed are presented in a (B) "Recent clinical studies" section under five Clinical Study Categories: Etiology, Diagnosis, Prognosis, Therapy, and Clinical prediction guides. The results of using the systematic reviews filter in the PubMed Clinical Queries are displayed in (C) Recent Systematic Reviews section with similar format. If available the link to the free article in PMC is included.

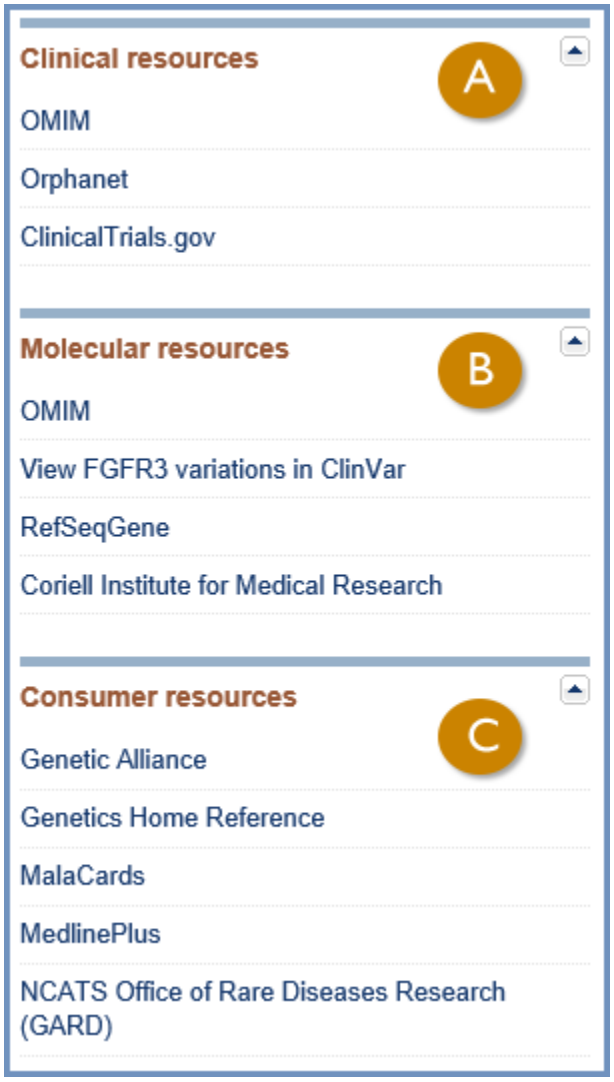


Figure 6. The clinical, molecular, and consumer resources in MedGen. If available, concept specific links to (A) clinical resources, (B) molecular resources and (C) consumer resources are provided in the discovery panel at the right.

Furthermore, MedGen facilitates user's access to related information for a single topic in various NCBI resources by creating reciprocal links between MedGen and other NCBI databases. These links are computed by NCBI's query retrieval system and offer ample opportunities to explore different aspects of a topic based on available related information, such as sequence variations and its relationship to human health (i.e., clinical assertions for a particular disease or particular gene) in ClinVar; gene-specific connections for map, sequence, expression, structure, function, citation, and homology data in Gene; interacting genes, proteins, biomarkers, drugs, and small molecules in Pathways; etc. Some of these links are computed by using a third resource to streamline the connection. For example, the links to PubMed are generated based on general queries to PubMed as well as special queries derived from citations in other sources such as *GeneReviews*™, Medical Genetics Summaries, OMIM®, etc. Figure 7 illustrates how Reviews and Related information sections are displayed in MedGen.

Access

MedGen can be accessed on the web at http://www.ncbi.nlm.nih.gov/medgen/ , which gives users options to find, view information, and learn more about medical genetics by conducting basic and advanced searches. MedGen data can be downloaded based on user specific interest at MedGen FTP site or accessed programmatically via E-utilities.

Figure 7. The Reviews and Related information in MedGen. (A) Concept specific Reviews are presented in three different groups as *GeneReviews*, PubMed Clinical Queries, and Reviews in PubMed. (B) Related information displays a list of available links to various NCBI databases, which varies depending on the concept.

Web interface

MedGen uses a powerful search and retrieval system developed by NCBI (Entrez) to search and retrieve data from MedGen and other integrated databases at NCBI. Basic queries can simply be submitted either by entering free text or entering field values followed by a proper field qualifier such as [gene], [mim], [moddate], etc., in the search bar. MedGen supports constructing complex queries by providing Limits and Advanced options. Users also can take advantage of a spell-check dictionary and Search History function to refine their search terms more precisely. Users can select terms from the suggested list of search terms provided by dictionary as they enter their search keywords. They can use Limits to restrict their query by chromosome, relationships to other NCBI databases, types of content, and/or sources of the terms. Or use the Advanced function to combine concepts and/or different fields in their search terms. Search History allows the users to take advantage of adding more restrictions step by step to a previous query. Detailed instructions on using the search interface are provided in the MedGen Help Documentation.

Search results in MedGen are first presented in a summary format, in order of relevance and with 20 items per page as default. The format is flexible and can be modified from summary to UI List, XML, or text. The number of items per page also can be adjusted based on users' preferences. An easy access to the unique MedGen identifier, Concept ID, and semantic type for each concept is ensured by displaying the identifiers below the name. In a full report users can access all metadata available for a concept and explore all links and

supplementary information aggregated in MedGen. The table of contents in the upper right corner and collapsible sections ease the navigation in accessing desired data (Figure 8).

If an old Concept ID or term is used as a search term, information for the new merged record will be retrieved and displayed.

Clicking on the title of a search result provides a full display. For more details of what is accessed from the full display, please see the figures in the Dataflow section.

FTP Download

All Concepts in MedGen can be downloaded via MedGen's FTP site, which is updated weekly every Wednesday. The metadata included with each concept are presented in multiple files. All of these files are in text format and provided as compressed zip files. Vertical bar (|) is used as the delimiter and the column names are declared in the first line of each file.

To assist users in tracking retired or merged Concepts IDs these concepts are reported as paired IDs in MERGED.RRF file. Semantic types for each concept are reported in MGSTY.RRF file and concept definitions alongside of the source of the definition are provided in MGDEF.RRF file. Concept names are stored in NAMES.RRF file. In order to verify if a term is a preferred term from a vocabulary source, the users can look up the ISPREF field (either "Y" or "N") in MGCONSO.RRF file. In this file one can also find any identifier asserted by the source, abbreviation for the source, type of term as defined by the source, etc.

The relationships between concepts are provided in MGREL.RRF. Concepts may have one or multiple relationship labels (i.e. one concept can be a child, a parent, a sibling, etc.). Summary data for each concept identifier is provided in MGCONSO.RRF. Each concept may have many attributes, which all are summarized in MGSAT.RRF file.

By combining data from MGSAT.RRF and MGCONSO.RRF users can work out paths to many of the connections that MedGen has made with its external resources. For example, connections between HPO Primary IDs and MedGen concepts are maintained in MGCONSO.RRF (using the first column (CUI), the 9th column (SAB = HPO), and the 8th column (SDUI) for the ID asserted by HPO). When HPO staff retire or merge a Primary ID in HPO and report it as an Alternative ID, MedGen will report the Alternative ID as an attribute (HPO_ALT_ID) for that concept. Therefore, users can find those Alternative IDs in MGSAT.RRF under ATV in the 8th column. For example, "HP:0003122" is reported as an alternative ID for Glycosuria ("HP:0003076") by HPO. MedGen reports this Alternative ID in MGSAT.RRF as an attribute for Glycosuria ("C0017979"). Thus, in MGSAT.RRF users can find a line for "C0017979", which has HPO_ALT_ID as attribute name (ATN, the 6th column) and "HP:0003122" as attribute value (ATV, the 8th column).

Another example is the connection made between OMIM® IDs and CUIs. OMIM® IDs are reported for concepts having OMIM® as their source vocabulary and either of "term types" (TTY) of "Preferred name" (PT), "synonym" (SYN), or "acronym" (ACR). However, some of the records in OMIM® have a Gene Phenotype Relationships section, which reports MIM numbers for genes associated with that record. Since genes are not in the scope for MedGen, the connections are maintained by gene symbols via the NCBI Gene database.

MedGen assist users in obtaining the mapped data connecting MedGen concepts to HPO and MedGen concepts to HPO and OMIM® by providing two additional data files (MedGen_HPO_Mapping.txt and MedGen_HPO_OMIM_Mapping.txt respectively).

Users also can download a complete list of links between MedGen concepts and literature from PubMed in medgen_pubmed file (i.e., CUI, PMID connection). This allows users to have access to all relevant literature regardless of the rules used to create these connections.

Figure 8. The Table of contents in MedGen. Content of each full report is organized in several sections listed in the table of contents and hyperlinked for easy access.

MedGen's updates and major releases can be followed through MedGen RSS feed announcements.

E-utilities

Entrez provides a series of programming utilities as a stable interface into the Entrez query and database system. These utilities allow using a fixed URL syntax, which translates the input parameters into a query request for search and retrieval. MedGen has enabled use of E-utilities for its database via esearch and summary, but not efetch. For a basic text search users can simply place their query term at the end of the following URL (replacing <query_term>): http://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=medgen&term=<query_term>

References

1. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res. 2004 Jan 1;32(Database issue):D267–70. PubMed PMID: 14681409.
2. Fung KW, Bodenreider O. Utilizing the UMLS for semantic mapping between terminologies. AMIA Annu Symp Proc. 2005.:266–70. PubMed PMID: 16779043.
3. Amberger J, Bocchini CA, Scott AF, Hamosh A. McKusick's Online Mendelian Inheritance in Man (OMIM). Nucleic Acids Res. 2009 Jan;37(Database issue):D793–6. doi: 10.1093/nar/gkn665. Epub 2008 Oct 8. PubMed PMID: 18842627.
4. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. Am J Hum Genet. 2008 Nov;83(5):610–5. doi: 10.1016/j.ajhg.2008.09.017. Epub 2008 Oct 23. PubMed PMID: 18950739.
5. Rubinstein WS, Maglott DR, Lee JM, Kattman BL, Malheiro AJ, Ovetsky M, Hem V, Gorelenkov V, Song G, Wallin C, Husain N, Chitipiralla S, Katz KS, Hoffman D, Jang W, Johnson M, Karmanov F, Ukrainchik A, Denisenko M, Fomous C, Hudson K, Ostell JM. The NIH genetic testing registry: a new, centralized database of genetic tests to enable access to comprehensive information and improve transparency. Nucleic Acids Research. 2013;41(D1):D925–D935. doi: 10.1093/nar/gks1173. PubMed PMID: 23193275.