ONLINE UPDATES

Emily Vardell, Column Editor

MedGen: NCBI's Portal to Information on Medical Conditions with a Genetic Component

Diana Nelson Louden

ABSTRACT. MedGen serves as a portal to information on genetic aspects of human health and disease. Created and maintained by the National Center for Biotechnology Information (NCBI), it aggregates clinically-relevant content from both NCBI and non-NCBI databases. MedGen summaries and curated links are designed to be particularly useful to health care professionals considering genetic aspects of patient care.

KEYWORDS. Genetics, health care, online database, genomic medicine, MedGen, NCBI

AUTHORS.

Diana Nelson Louden, MLib (dknl@uw.edu) is Biomedical and Translational Sciences Librarian, University of Washington, Seattle, WA 98195.

Comments and suggestions should be sent to the Column Editor: Emily Vardell (evardell@emporia.edu).

INTRODUCTION

Since its founding in 1988, the National Center for Biotechnology Information (NCBI), a division of the National Library of Medicine, has created and maintained numerous databases of genetic and molecular biology information. Initially NCBI databases were used primarily by scientists. As knowledge increased about genetic factors in human diseases, new capabilities arose in genomic medicine, and pharmacogenomics information began to be incorporated in prescribing practices, the need grew for improved access to genetic information for clinicians. In response, in 2012 NCBI began working on MedGen (<https://www.ncbi.nlm.nih.gov/medgen/>) to aggregate genetic information relating to clinical attributes, disease states, and drug metabolism – or what the NCBI Handbook refers to as "human disorders and other phenotypes having a genetic component."[1]

A person's phenotype – physical attributes and functions – is determined by a combination of a person's genes and environmental factors affecting the expression of those genes. By saying that MedGen includes information on phenotypes having a genetic component, they are referring to medical conditions that always have a genetic cause, such as sickle cell disease or Marfan syndrome, as well as conditions that may arise due to a combination of genetic and environmental factors, such as colon cancer or postmenopausal osteoporosis.

As suits a clinically-oriented resource, MedGen includes practice guidelines and resources to assist with diagnosis, risk assessment, and treatment. It also includes links to clinical trials in ClinicalTrials.gov, the medical literature in PubMed, and consumer health information from a variety of sources. All of the information in MedGen is freely available as part of the federally-funded suite of resources provided on the NCBI platform.

METHODOLOGY

Developers created MedGen in order to organize information from many disparate sources relating to human medical genetics. As a result, MedGen records contain content from and links to key NCBI and non-NCBI resources using a combination of computational and human curation methods. Some of the key sources of content connected to an information-rich MedGen record are *GeneReviews*®, **Online Mendelian Inheritance in Man (OMIM)**®, ClinVar (genetic variants from patient samples), Genetic Testing Registry® (GTR), practice guidelines, **Pharmacogenomics Knowledge Base (PharmGKB)**, and Genetics Home Reference (GHR).

In order to integrate information from different sources that use different terminology, developers have mapped related terms to single concepts and assigned MedGen concept unique identifiers (CUI). Preferred concept terms and definitions are taken from (in order of preference) Genetic Testing Registry, ClinVar, the Human Phenotype Ontology, the National Cancer Institute thesaurus, SNOMED CT, Orphanet, and the Unified Medical Language System (UMLS).[2]

Through these linkages, information is extracted or pointed to. For example, definitions, clinical features, and associated genes are extracted from resources such as Human Phenotype Ontology and NCBI's Gene database to populate a MedGen record. In addition, links within a MedGen record may point to ongoing trials in ClinicalTrials.gov or available clinical tests in Genetic Testing Registry. Curated professional guidelines are embedded in individual records, and the entire collection of guidelines is also linked to directly from the home page (see Figure 1).

[PLACE FIGURE 1 HERE]

Legend: FIGURE 1. MedGen home page.

SEARCHING THE DATABASE

Like other NCBI databases, MedGen provides both a general search box on the home page and an advanced search page. There are several points of entry to MedGen records. Common search terms are:

- Medical conditions, diseases, or syndromes (e.g., cystic fibrosis or Alzheimer disease);
- Symptoms or clinical features (e.g., failure to thrive or dementia);
- Drugs (to find pharmacogenomics information, e.g., warfarin or clopidogrel); and
- Gene symbols, e.g., APOE or BRCA1.

Also like other NCBI databases, terms can be searched with or without field tags. Commonly-used field tags include [Clinical Features] and [Gene]. When searching for a gene identified as being associated with a clinical condition, it is most effective to search for the official HUGO symbol (<https://www.genenames.org/>) of the gene and to limit the search to the gene field (e.g. APOE[gene]).

When reviewing search results, it is helpful to know that MedGen contains several types of records, including: disease or syndrome, sign or symptom, finding, laboratory result, and pharmacologic substance. A search for a gene, for example, will retrieve a list of diseases or syndromes associated with mutations or variations in that gene. A search for a clinical feature will retrieve a list of clinical findings or diseases where that feature can be a symptom.

SEARCHING THE DATABASE

To find a summary of clinically-oriented information on genetic aspects of breast cancer, it is more effective to search for the widely-used term "breast cancer" rather than the MeSH term "breast neoplasms." The search results will be a relevance-ranked list of clinical conditions or phenotypes relating to breast cancer that have a genetic component. In MyNCBI settings, consider activating the MedGen filter described as a "subset of diseases and clinical features with rich support." That filter will then appear on the right side of the search screen with the short description "Recommended for clinicians" and can be used to limit search results. It can be difficult to decide which records in the search results to view, but if a patient care summary is wanted, it can be helpful to start with a record that includes content from *GeneReviews* – as indicated by whether *GeneReviews* is greyed out or linked to (see Figure 2).

[PLACE FIGURE 2 HERE]

Legend: FIGURE 2. Results from search for breast cancer. Record 6 "Familial cancer of breast" contains content from GTR, ClinVar, Gene, OMIM, and *GeneReviews* .

Choosing the sixth record in the search results "Familial cancer of breast" yields the record seen in Figure 3. The table of contents in the upper right hand corner is useful for directing the reader to the most essential parts of the record. Each MedGen record includes the concept ID from the UMLS and links to computationally-derived related references in PubMed. Content-rich records also contain synonyms for the concept, the gene or genes associated with that phenotype,

excerpts from other resources such as *GeneReviews* or OMIM, a section on Clinical Features, and a Term Hierarchy. There may also be a curated section of Suggested Reading or links to highly relevant non-NCBI websites. Links to available tests in Genetic Testing Registry and clinically-significant gene variants in ClinVar are provided on the right side of the record for people with those specialized information needs.

[PLACE FIGURE 3 HERE]

Legend: FIGURE 3. MedGen record for Familial cancer of breast.

The manually-curated Term Hierarchy (see Figure 4) can be particularly helpful for additional searching because it outlines how concepts are related within MedGen. Click on "Breast-ovarian cancer, familial 1," for example, to get to the MedGen record on BRCA1-associated breast cancer. Each term in the hierarchy also contains links to relevant genetic tests in Genetic Testing Registry (GTR) and to genetic variants identified in patient samples from ClinVar. GTR and ClinVar will be described in more detail below.

A final important way to identify relevant information in MedGen is to identify PubMed references describing medical conditions with a genetic component. When relevant, the Related Information abstract portlet will link to MedGen summaries for the selected citation.

[PLACE FIGURE 4 HERE]

Legend: FIGURE 4. Term hierarchy.

ADDITIONAL RESOURCES IN THIS DATABASE

MedGen's value lies in the fact that it provides a single point of entry to integrated content from multiple clinically-oriented resources. Because of the way the database is organized, a person does not need a lot of specialized expertise to identify useful information on conditions with a genetic component. Some of the key resources accessible through MedGen follow.

[PLACE FIGURE 5 HERE]

Legend: FIGURE 5. Alzheimer disease chapter from *GeneReviews*.

GeneReviews (Figure 5) provides point-of-care information for clinicians on inherited conditions. Created and edited by the University of Washington, each expert-authored, peer-reviewed, extensively referenced chapter covers diagnosis, management, and genetic counseling information for a gene, phenotype, or common condition. Each chapter is regularly updated, incorporates relevant practice guidelines, and lists support organizations and patient registries. Recently *GeneReviews* added educational information for clinicians, such as Detailed Information for Clinicians Ordering Genetic Tests.

Online Mendelian Inheritance in Man (OMIM) is a long-standing compendium of all known Mendelian disorders. OMIM is authored and edited by Johns Hopkins University and focuses on gene-phenotype relationships. Overviews contain a clinical synopsis and extracted findings from the literature on aspects of the disorders such as gene mapping, pathogenesis, and population genetics.

Professional Guidelines are identified by NCBI staff that relate to disorders, genes, or genetic variants. These professional practice guidelines, position statements, and recommendations are linked to the appropriate MedGen summaries and are also listed in a collection available at: https://www.ncbi.nlm.nih.gov/medgen/docs/guideline/.

ClinVar compiles reports on the relationships between specific human genetic variants and associated phenotypes. Submitters to NCBI's ClinVar database report on variants found in patient samples, assertions regarding their clinical significance (e.g., benign, pathogenic, uncertain signifcance), and evidence supporting those assertions. For each genetic variant, ClinVar uses a star system to indicate the level of evidence supporting a particular interpretation and whether there are conflicting interpretations.

Genetic Testing Registry (GTR) provides information about specific genetic tests offered by clinical and commercial labs that have chosen to register with GTR. Both clinical and research tests are included in NCBI's GTR, and tests may be designed for disease diagnosis, prenatal testing, risk prediction, drug response, and precision medicine (e.g., targeted cancer therapy).

Pharmacogenomics Knowledge Base (PharmGKB) is the premier resource for pharmacogenomics, the effect of human genetic variation on drug responses. PharmGKB editors at Stanford University curate clinical practice guidelines and FDA drug labels containing pharmacogenomic information. They also publish pharmacogenomic-based drug dosing guidelines.

Genetics Home Reference (GHR), written by National Library of Medicine staff, provides consumer-oriented information on genetic concepts and the effects of genetics on human health

conditions. An excellent overview of Genetics Home Reference was published by Koos and Bassett in 2018.[3]

PubMed references listed in the "recent clinical studies" and "recent systematic reviews" sections of MedGen records are computationally extracted from PubMed using the Clinical Queries tool. PubMed references listed in "professional guidelines" and "suggested reading" sections of MedGen records are curated by NCBI staff.

CONCLUSION

Because MedGen aggregates information from many important resources relating to human genetics in health and disease, it is an excellent place to start when it is not clear what single resource might be most valuable or what terminology to use. It is also a highly-recommended starting point for clinicians looking for guidance on incorporating genetic information in patient care. The MedGen portal provides easy access to freely-available information on genetic aspects of medical conditions, genetic effects on drug metabolism, and practice guidelines on the use of genetic information for diagnosis, risk assessment, and treatment.

FOR MORE INFORMATION

For additional information on MedGen, please see the help documentation at https://www.ncbi.nlm.nih.gov/medgen/docs/help/, or contact:

National Center for Biotechnology Information

U.S. National Library of Medicine

8600 Rockville Pike

Bethesda MD, 20894 USA

E-mail: info@ncbi.nlm.nih.gov

REFERENCES

1. Halavi M Maglott D, Gorelenkov V, Rubinstein W. MedGen Bethesda (MD): National Center for Biotechnology Information (US); [updated 2018 Dec 11; cited 2019]. 2nd:[Available from: https://www.ncbi.nlm.nih.gov/books/NBK159970/.

2. MedGen Help Documentation Bethesda (MD): National Center for Biotechnology Information (US); [updated 2017 October 30; cited 2019]. Available from: https://www.ncbi.nlm.nih.gov/medgen/docs/help/.

3. Koos JA, Bassett A. Genetics Home Reference: A Review. Medical reference services quarterly. 2018;37(3):292-9. doi: 10.1080/02763869.2018.1477716.