Bioinformatics, 34(13), 2018, 2325–2326
doi: 10.1093/bioinformatics/bty054
Advance Access Publication Date: 1 February 2018
Applications Note

Databases and ontologies

MutHTP: mutations in human transmembrane proteins

A. Kulandaisamy[1], S. Binny Priya[1], R. Sakthivel[1], Svetlana Tarnovskaya[2], Ilya Bizin[2], Peter Hönigschmid[3], Dmitrij Frishman[2,3] and M. Michael Gromiha[1,4,*]

[1]Department of Biotechnology, Bhupat and Jyoti Mehta School of BioSciences, Indian Institute of Technology Madras, Chennai 600 036, Tamilnadu, India, [2]Department of Bioinformatics, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation, [3]Department of Bioinformatics, Technische Universität München, WissenschaftszentrumWeihenstephan, Freising, Germany and [4]Advanced Computational Drug Discovery Unit (CADD), Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, Kanagawa 226-8501, Japan

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on November 18, 2017; revised on January 8, 2018; editorial decision on January 28, 2018; accepted on January 31, 2018

Abstract

Motivation: Existing sources of experimental mutation data do not consider the structural environment of amino acid substitutions and distinguish between soluble and membrane proteins. They also suffer from a number of further limitations, including data redundancy, lack of disease classification, incompatible information content, and ambiguous annotations (e.g. the same mutation being annotated as disease and benign).

Results: We have developed a novel database, MutHTP, which contains information on 183 395 disease-associated and 17 827 neutral mutations in human transmembrane proteins. For each mutation site MutHTP provides a description of its location with respect to the membrane protein topology, structural environment (if available) and functional features. Comprehensive visualization, search, display and download options are available.

Availability and implementation: The database is publicly available at http://www.iitm.ac.in/bioinfo/MutHTP/. The website is implemented using HTML, PHP and javascript and supports recent versions of all major browsers, such as Firefox, Chrome and Opera.

Contact: gromiha@iitm.ac.in

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Transmembrane proteins have diverse functional roles, featuring important functions such as cell signalling, transport of molecules and ions across membranes and many others. Approximately, 20–30% of human genes encode membrane proteins and ~60% of them act as drug targets (Almeida *et al.*, 2017; Gromiha and Ou, 2014). Mutations or aberrant activity in membrane proteins cause many different developmental disorders and diseases, including several types of cancer, neurodegenerative diseases, etc. The experimental mutation data are available in various databases such as Humsavar (http://www.uniprot.org/docs/Humsavar), SwissVar (Mottaz *et al.*, 2010), 1000 Genomes (Auton *et al.*, 2015), COSMIC (Forbes *et al.*, 2015) and ClinVar (Landrum *et al.*, 2014). The existing databases of experimental mutation data have a number of limitations: (i) information is available only for mutations and diseases without any sequence/structure based features (ii) data redundancy (iii) no disease classification (iv) information content in different databases is not compatible and (v) annotation is often ambiguous (eg. same mutation is annotated as disease and benign).

Two databases—Mpstruc (http://blanco.biomol.uci.edu/mpstruc/) and PDBTM (Kozma *et al.*, 2013) collect information on experimentally determined 3D structures of membrane proteins. TMFunction (Gromiha *et al.*, 2009) provides functionally important residues of membrane proteins and MeMotif (Marsico *et al.*, 2009)

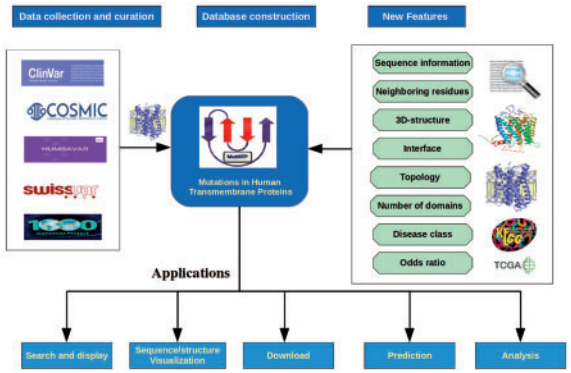

Fig. 1. Schematic diagram describing data collection, features and applications of MutHTP

gives structurally and functionally important motifs in α-helical membrane proteins. The experimentally determined topology information of membrane proteins is available in TOPDB resource (Dobson *et al.*, 2015). However, a database of disease causing and neutral mutations for membrane proteins is currently not available. In order to overcome these limitations, we present MutHTP (Mutations in Human Transmembrane Proteins), a mutation database of human transmembrane proteins with various sequence and structural information, membrane protein features and cross-links to other related sequence and structure databases.

2 Contents

Membrane proteins were retrieved from the UniProt (Boutet *et al.*, 2016) database using keyword search with 'Transmembrane' in subcellular localization. We have collected mutation data from Humsavar, SwissVar, 1000 Genomes, COSMIC and ClinVar databases and extracted only insertions, deletions and missense mutations occurring in membrane proteins. The detailed, workflow and features are shown in Figure 1. MutHTP contains 173 757 missense mutations, 2455 insertions and 7183 deletions associated with 2695 diseases, as well as 17 516 missense mutations, 39 insertions and 272 deletions which are neutral. It provides various sequence, structure and specific membrane protein features of the mutation sites. The features include gene name, chromosome number and genome position, type of mutation, protein and nucleotide level mutation, origin, Uniprot ID, PDB ID, effect of the mutation, interface information of mutation site in protein-protein complexes, neighboring residues, conservation score, number of domains, topology, protein spanning the membrane, disease name, disease class, odds ratio of the mutation site and source database. An example for the records of A124G mutation in IL1R1 gene is shown in Supplementary Table S1.

The experimentally determined membrane protein features such as the number of domains and topology are obtained from the TOPDB and UniProt annotations. All the 2587 unique diseases are classified into 14 classes based on KEGG human diseases. Odds ratio is a cohort from The Cancer Genome Atlas (TCGA) and from non-cancer population. Location of mutation sites in protein structure is visualized using the JSmol applet (Hanson *et al.*, 2013) and the mutation site is highlighted in the protein sequence.

3 Unique features

The existing mutation databases provide mutation data for both globular and membrane proteins with limited search and display options. MutHTP provides comprehensive characteristics of mutations, such as conservation score, neighboring residues, 3D structure, mutation site in the protein-protein interface, disease class, minor allele frequency as well as a number of membrane protein related features, including topology and the number of domains. Users can also visualize the location of mutation site at the sequence and structure level. In addition, entries are cross-linked to Gene Cards, UniProt, PDB and all the source databases mentioned above. The feature comparison of MutHTP and existing databases are tabulated in Supplementary Table S2.

4 Applications

MutHTP database will be useful for performing a wide-range of analyses related to mutations in transmembrane proteins, such as developing algorithms for the discrimination of disease-causing and neutral mutations and exploring the preferred mutations in various disease classes. This integrated database will help researchers to obtain insights into the role of mutations in disease and to design and develop personalized treatment strategies.

Acknowledgements

We thank the Bioinformatics facility, Department of Biotechnology and Indian Institute of Technology Madras for computational facilities.

Funding

We acknowledge the support of the Department of Science and Technology, Government of India (INT/RUS/RSF/P-09) and of the Russian Science Foundation for a research grant (16-44-02002). A.K. was supported by the Ministry of Human Resource and Development, India. S.B.P. thanks SERB for NPDF.

Conflict of Interest: none declared.

References

Almeida,J.G. *et al.* (2017) Membrane proteins structures: a review on computational modeling tools. *Biochim. Biophys. Acta*, **1859**, 2021–2039.

Auton,A. *et al.* (2015) 1000 Genomes Project Consortium: a global reference for human genetic variation. *Nature*, **526**, 68–74.

Boutet,E. *et al.* (2016) UniProtKB/Swiss-Prot, the manually annotated section of the UniProtKnowledgeBase: how to use the entry view. *Methods Mol. Biol.*, **1374**, 23–54.

Dobson,L. *et al.* (2015) Expediting topology data gathering for the TOPDB database. *Nucleic Acids Res.*, **43**, D283–D289.

Forbes,S.A. *et al.* (2015) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.*, **43**, D805–D811.

Gromiha,M.M. *et al.* (2009) TMFunction: database for functional residues in membrane proteins. *Nucleic Acids Res.*, **37**, D201–D204.

Gromiha,M.M. and Ou,Y.Y. (2014) Bioinformatics approaches for functional annotation of membrane proteins. *Brief Bioinform.*, **15**, 155–168.

Hanson,R.M. *et al.* (2013) JSmol and the next-generation web-based representation of 3D molecular structure as applied to proteopedia. *Isr. J. Chem.*, **53**, 207–216.

Kozma,D. *et al.* (2013) PDBTM: protein Data Bank of transmembrane proteins after 8 years. *Nucleic Acids Res.*, **41**, D524–D529.

Landrum,M.J. *et al.* (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.*, **42**, D980–D985.

Marsico,A. *et al.* (2009) MeMotif: a database of linear motifs in α-helical transmembrane proteins. *Nucleic Acids Res.*, **38**, D181–D189.

Mottaz,A. *et al.* (2010) Easy retrieval of single amino-acid polymorphisms and phenotype information using SwissVar. *Bioinformatics*, **26**, 851–852.