Bioinformatics, 32(12), 2016, 1914–1916
doi: 10.1093/bioinformatics/btw086
Advance Access Publication Date: 15 February 2016
Applications Note

OXFORD

Database and ontologies

dbDSM: a manually curated database for deleterious synonymous mutations

Pengbo Wen[1], Peng Xiao[1] and Junfeng Xia[1,2,*]

[1]Institute of Health Sciences, School of Life Sciences, Anhui University, Hefei, Anhui 230601, China and [2]Co-Innovation Center for Information Supply and Assurance Technology, School of Computer Science and Technology, Anhui University, Hefei, Anhui 230601, China

*To whom correspondence should be addressed.
Associate Editor: Jonathan Wren

Received on November 18, 2015; revised on January 23, 2016; accepted on February 6, 2016

Abstract

Motivation: Synonymous mutations (SMs), which changed the sequence of a gene without directly altering the amino acid sequence of the encoded protein, were thought to have no functional consequences for a long time. They are often assumed to be neutral in models of mutation and selection and were completely ignored in many studies. However, accumulating experimental evidence has demonstrated that these mutations exert their impact on gene functions via splicing accuracy, mRNA stability, translation fidelity, protein folding and expression, and some of these mutations are implicated in human diseases. To the best of our knowledge, there is still no database specially focusing on disease-related SMs.

Results: We have developed a new database called dbDSM (database of Deleterious Synonymous Mutation), a continually updated database that collects, curates and manages available human disease-related SM data obtained from published literature. In the current release, dbDSM collects 1936 SM-disease association entries, including 1289 SMs and 443 human diseases from ClinVar, GRASP, GWAS Catalog, GWASdb, PolymiRTS database, PubMed database and Web of Knowledge. Additionally, we provided users a link to download all the data in the dbDSM and a link to submit novel data into the database. We hope dbDSM will be a useful resource for investigating the roles of SMs in human disease.

Availability and implementation: dbDSM is freely available online at http://bioinfo.ahu.edu.cn:8080/dbDSM/index.jsp with all major browser supported.

Contact: jfxia@ahu.edu.cn

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

Point mutations can be classified as either non-synonymous or synonymous based on whether they change the amino acid sequence of a protein (Gotea *et al.*, 2015). Historically, synonymous mutations (SMs) have been consistently overlooked because they do not alter the protein sequence and are thus considered functionally irrelevant. Many studies completely ignore SMs, whereas others use them to build background models for the estimation of deleterious mutations (Cartegni *et al.*, 2002; Kimura, 1980). But 'synonymous' is different from 'same', and accumulating evidence suggested that SMs could have functional consequences (Bali and Bebok, 2015; Buske *et al.*, 2013; Chaney and Clark, 2015; Hunt *et al.*, 2014; Sauna and Kimchi-Sarfaty, 2011; Shabalina *et al.*, 2013; Supek *et al.*, 2014). Over 50 diseases have been associated with SMs in human genes so far (Chamary and Hurst, 2009). Studies analyzing the consequences of SMs have revealed that they play an important role in multiple biological processes, such as mRNA splicing (Supek *et al.*, 2014), mRNA folding (Cartegni *et al.*, 2002), protein translation and folding (Spencer *et al.*, 2012) and microRNA–mRNA interaction (Brest *et al.*, 2011). In summary,

ignoring SMs in human genes would cause incorrect estimation of mutation rate in genome analysis (background mutation would be overestimated and functional mutation would be underestimated; Supek *et al.*, 2014).

Currently, there are many databases that focus on point mutation, such as dbNSFP (Liu *et al.*, 2013) and miRNASNP (Gong *et al.*, 2015). However, to the best of our knowledge, there is no database that can provide comprehensive resources for the disease-related SMs. To fill this gap, we developed a literature-based disease-related SMs database called dbDSM (database of Deleterious Synonymous Mutation), which is proposed to assist with investigating the mechanism of SM-disease association and facilitate translational research on disease genetics. We hope it will be a useful resource for the SM research community.

2 Methods

The flowchart of the construction of dbDSM is illustrated in Fig. 1. We firstly performed an extensive literature query of PubMed database and Web of Knowledge using a list of key words, such as 'synonymous', 'neutral', 'mutation', 'disease' and 'deleterious' (for details, please see Supplementary Material) on December 31, 2015. More than 500 articles were obtained from PubMed and Web of Knowledge in this way. After we obtained the literature, we read through and interpreted each paper by collecting the important information, including the diseases examined, mutation position, amino acid residue and publication information. In our extraction process, if some information was not available, it was labeled as missing.

In addition, we collected known SM-disease association data from five public databases, including Clinvar (Landrum *et al.*, 2014), GRASP 2.0 (Eicher *et al.*, 2015), PolymiRTS Database 3.0 (Bhattacharya *et al.*, 2014), NHGRI GWAS Catalog (Welter *et al.*, 2014) and GWASdb database (Li *et al.*, 2011). We retrieved 1573 SM-disease associations from these databases on June 15, 2015.

Finally, we consolidated 1289 deleterious SM from ClinVar, GRASP, GWAS Catalog, GWASdb, PolymiRTS database, PubMed database and Web of Knowledge. For extracted SM-disease associations, we collected related information regarding the SM from NCBI dbSNP/Ensembl database. In addition, we provided a consensus classification of molecular mechanisms by which SMs contribute to disease phenotype (Supplementary Material) for each deleterious SM in our database. We stored and managed all the data in MySQL (version 5.1.73), which is a popular and open-source relationship database management system that has been widely used in biomedical research.

3 Results

In the current version, dbDSM collected 1936 SM-disease association entries, including 1289 SMs and 443 human diseases from ClinVar, GRASP, GWAS Catalog, GWASdb, PolymiRTS database, PubMed database and Web of Knowledge.

The data in dbDSM can be easily accessed in a variety of ways. Users can query the database in 'Search' page by gene name, dbSNP ID, disease name or chromosome number, which results in display of SM-centered information. Users can also explore all entries of the database through the 'Browse' page. All the results are available to download for further analysis.

Aside from data retrieval from dbDSM, users are encouraged to submit novel data into the database. Users may first search dbDSM to check whether their data have already been deposited into the database. If not, they may upload the related data information, which will be stored in dbDSM. The novel data will be forwarded to the dbDSM developer via email and will become available after a manual check and confirmation.

All the descriptions above are available in the 'Document' page.

In conclusion, dbDSM is the first attempt to establish a literature-based resource of deleterious SMs. It is a resource for better understanding SM-disease associations and developing useful information for investigating the roles of SMs in human disease. This database could also be used for the annotation of RNA-seq data (Supek *et al.*, 2014). The collection and curation of deleterious SMs will keep on going and more valuable resources will be integrated into the dbDSM in the future. We believe that dbDSM will be a useful tool for the biomedical research community.

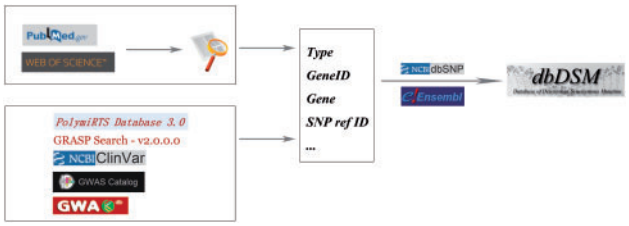

Fig. 1. The flowchart of dbDSM construction. The flowchart showing data processing and information integration

Acknowledgement

The authors thank the members of our laboratory for their valuable contributions to dbDSM.

Funding

This work was supported by the grants from the National Natural Science Foundation of China (31301101, 61272339 and 61272333), the Anhui Provincial Natural Science Foundation (1408085QF106 and 1508085QF135), the Specialized Research Fund for the Doctoral Program of Higher Education (20133401120011) and the Co-Innovation Center for Information Supply and Assurance Technology of Anhui University (Y01002337).

Conflict of Interest: none declared.

References

Bali,V. and Bebok,Z. (2015) Decoding mechanisms by which silent codon changes influence protein biogenesis and function. *Int. J. Biochem. Cell Biol.*, **64**, 58–74.

Bhattacharya,A. *et al.* (2014) PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res.*, **42**, D86–D91.

Brest,P. *et al.* (2011) A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nat. Genet.*, **43**, 242–245.

Buske,O.J. *et al.* (2013) Identification of deleterious synonymous variants in human genomes. *Bioinformatics*, **29**, 1843–1850.

Cartegni,L. *et al.* (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.*, **3**, 285–298.

Chamary,J.V. and Hurst,L.D. (2009) The price of silent mutations. *Sci. Am.*, **300**, 46–53.

Chaney,J.L. and Clark,P.L. (2015) Roles for synonymous codon usage in protein biogenesis. *Annu. Rev. Biophys.*, **44**, 143–166.

Eicher,J.D. *et al.* (2015) GRASP v2.0: an update on the Genome-Wide Repository of Associations between SNPs and phenotypes. *Nucleic Acids Res.*, **43**, D799–D804.

Gong,J. *et al*. (2015) An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools. *Database*, **2015**, bav029.

Gotea,V. *et al*. (2015) The functional relevance of somatic synonymous mutations in melanoma and other cancers. *Pigment Cell Melanoma Res*., **28**, 673–684.

Hunt,R.C. *et al*. (2014) Exposing synonymous mutations. *Trends Genet*., **30**, 308–321.

Kimura,M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol*., **16**, 111–120.

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

Li,M.J. *et al*. (2011) GWASdb: a database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res*., **40**, D1047–D1054.

Liu,X. *et al*. (2013) dbNSFP v2.0: A Database of Human Non-synonymous SNVs and Their Functional Predictions and Annotations. *Hum. Mutat*., **34**, E2393–E2402.

Sauna,Z.E. and Kimchi-Sarfaty,C. (2011) Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet*., **12**, 683–691.

Shabalina,S.A. *et al*. (2013) Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity. *Nucleic Acids Res*., **41**, 2073–2094.

Spencer,P.S. *et al*. (2012) Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *J. Mol. Biol*., **422**, 328–335.

Supek,F. *et al*. (2014) Synonymous mutations frequently act as driver mutations in human cancers. *Cell*, **156**, 1324–1335.

Welter,D. *et al*. (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*., **42**, D1001–D1006.