Databases and ontologies

SNP2NMD: A database of human single nucleotide polymorphisms causing nonsense-mediated mRNA decay

Areum Han1,†, Woo-Yeon Kim1,† and Seong-Min Park1,2,*

1Korean Bioinformation Center, KIBB, Daejeon 305-806, Korea and 2Functional Genomics Research Center, KIBB, Daejeon 305-806, Korea

Received on September 21, 2006; revised and accepted on November 18, 2006
Advance Access publication November 22, 2006
Associate Editor: Joaquin Dopazo

SUMMARY

Elucidating the effects of genetic polymorphisms on genes and gene networks is an important step in disease association studies. We developed the SNP2NMD database for human SNPs (single nucleotide polymorphisms) that result in PTCs (premature termination codons) and trigger nonsense-mediated mRNA decay (NMD). The SNP2NMD Web interfaces provide extensive genetic information on and graphical views of the queried SNP, gene, and disease terms.

Availability: SNP2NMD is available from http://biportal.kobic.re.kr/SNP2NMD

Contact: kimplove@kribb.re.kr, lastmhc@kribb.re.kr

Supplementary information: http://biportal.kobic.re.kr/SNP2NMDWiki.jsp?page=Statistics

1 INTRODUCTION

Annotating SNPs (single nucleotide polymorphisms) is becoming important in biology since information on their effects on genes and gene networks enables biologists to select and interpret disease-associated SNPs. An expert-curated SNP catalog (SNP@Web; http://biportal.kobic.re.kr/SNPatWEB/Wiki.jsp?page=Annotation) has reported over 20 SNP-annotation services. However, no application is available for nonsense-mediated mRNA decay (NMD) triggered by SNPs so far. Among SNPs, nonsense SNPs resulting in premature termination codons (PTCs) should be identified systematically since they go beyond altering protein sequences, by being able to trigger the NMD pathway and eliminate the production of proteins (Hentze and Kulozik, 1999). Recent research suggests that NMD controls cellular function as well as corrects biosynthetic errors, since one-third of human transcripts contain PTCs and are potential targets of NMD (Nagy and Maquat, 1998). We therefore denote the ‘NMD distance’ as the distance between an SNP and the 3′-most exon–exon junction, and is set to a default NMD rule as an SNP with an NMD distance >50 nt downstream in a spliced mRNA generally elicit NMD. The chromosomal positions and alleles of SNPs were parsed from a public SNP database (dbSNP ver. 125; Sherry et al., 2001) and mapped to exon and intron structures downloaded from the gene annotation files of UCSC (refGene, hg17; Karolchik et al., 2003) whilst considering the version of the human genome build (ver. 34) and the chromosomal strand. As a result, we identified 1301 (1%) SNPs producing stop codons from 123 908 coding SNPs.

2 METHODS AND RESULTS

The SNP2NMD database was developed in the three steps described below (Fig. 1a).

2.1 Identification of nonsense SNPs

We identified human SNPs resulting in PTCs by integrating transcript structure annotation and positional information of the SNPs of human genes. The chromosomal positions and alleles of SNPs were parsed from a public SNP database (dbSNP ver. 125; Sherry et al., 2001) and mapped to exon and intron structures downloaded from the gene annotation files of UCSC (refGene, hg17; Karolchik et al., 2003) whilst considering the version of the human genome build (ver. 34) and the chromosomal strand. As a result, we identified 1301 (1%) SNPs producing stop codons from 123 908 coding SNPs.

2.2 Application of an NMD rule

According to mammalian NMD research, PTCs followed by an exon–exon junction that is located more than ~50–55 nt downstream in a spliced mRNA generally elicit NMD (Nagy and Maquat, 1998). We therefore denote the ‘NMD distance’ as the distance between an SNP and the 3′-most exon–exon junction, and is set to a default NMD rule as an SNP with an NMD distance >50 nt (Notice that search interfaces of SNP2NMD are also able to accept a user-defined NMD distance). With the default NMD rule, we detected 765 SNP-mRNA pairs with 635 mRNAs as NMD candidates. The mean, median, and standard deviation of the NMD distance were ~1999.4, 775 and 5079.7 nt, respectively.

2.3 Integration with functional annotation

The SNP2NMD database adopted various gene annotations including pathways (KEGG; Kanehisa and Goto, 2000), gene ontology (GOA; Camon et al., 2004), and disease information (GAD, Becker et al., 2004; HGMD, David et al., 2005; and OMIM, McKusick, 1998). The raw data files were integrated into the SNP2NMD database based on a gene synonym table from HGNC (HUGO Gene Nomenclature Committee; Eyre et al., 2006). These annotations provide insight into the effects of SNP2 resulting in NMD and help to characterize target genes of NMD caused from SNPs. To find
significant associations of Gene Ontology terms with target genes, we assigned genes into gene ontology categories and selected the categories, which were significantly enriched \( (P < 0.01) \) with minimum number of observed gene number \( > 5 \) (GOTM; Zhang et al., 2004). The analysis showed that NMD-candidate genes were associated with gene ontology categories including ‘cell adhesion’, ‘physical interaction between organisms’, ‘reproductive physiological process’, ‘nucleotide binding’, ‘protein binding’, and ‘extra cellular matrix’ (More information is available on the website supplement page).

3 WEB INTERFACES

As shown in Figure 1b, users can search the SNP2NMD database using three entries: (1) SNP identifier (rs number from dbSNP), (2) gene ID (refSeq ID starting with ‘NM’) or name/symbol and (3) a disease term. In the case of a user submitting a gene or disease term, SNP2NMD returns a gene list related to the query that provides summary and gene details. The summary shows the number of NMD candidates within the resultant gene set. When users access the detailed view of NMD candidates by searching an SNP or selecting a gene, NMD information including the NMD distance and 2D views as well as SNP and gene details are shown. The 2D view was developed using Gbrowse (Stein et al., 2002), and provides a graphical view of the gene structure and SNP position simultaneously (Fig. 1c). The summary and external links to SNP, gene and disease information are designed to assist in further research.

ACKNOWLEDGEMENTS

The authors thank Jong Bhak for editing the manuscript and the SNP Pipeline Team at KOBIC (Korean Bioinformation Center). Also, we thank Young Il Yeom for discussion about the validation of SNP2NMD data. This project was supported by the Korean Ministry of Science and Technology (MOST) under grant number M10407010001-06N0701-00110, which also provided funding to pay for Open Access publication charges.

Conflict of Interest: none declared.

REFERENCES


