miRNASNP-v3: a comprehensive database for SNPs and disease-related variations in miRNAs and miRNA targets

Chun-Jie Liu, Xin Fu, Mengxuan Xia, Qiong Zhang, Zhifeng Gu, and An-Yuan Guo

1Research Center of Clinical Medicine, Affiliated Hospital of Nantong University, Nantong 226001, China; 2Center for Artificial Intelligence Biology, Hubei Bioinformatics and Molecular Imaging Key Laboratory, Key Laboratory of Molecular Biophysics of the Ministry of Education, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China and 3Department of Rheumatology, Affiliated Hospital of Nantong University, Nantong 226001, China

Received August 08, 2020; Revised September 02, 2020; Editorial Decision September 06, 2020; Accepted September 09, 2020

ABSTRACT

MicroRNAs (miRNAs) related single-nucleotide variations (SNVs), including single-nucleotide polymorphisms (SNPs) and disease-related variations (DRVs) in miRNAs and miRNA-target binding sites, can affect miRNA functions and/or biogenesis, thus to impact on phenotypes. miRNASNP is a widely used database for miRNA-related SNPs and their effects. Here, we updated it to miRNASNP-v3 (http://bioinfo.life.hust.edu.cn/miRNASNP) with tremendous number of SNVs and new features, especially the DRVs data. We analyzed the effects of 7,161,741 SNPs and 505,417 DRVs on 1897 pre-miRNAs (2630 mature miR-UTRs of 18,152 genes). miRNASNP-v3 provides a one-stop resource for miRNA-related SNVs research with the following functions: (i) explore associations between miRNA-related SNPs/DRVs and diseases; (ii) browse the effects of SNPs/DRVs on mRNA-target binding; (iii) functional enrichment analysis of miRNA target gain/loss caused by SNPs/DRVs; (iv) investigate correlations between drug sensitivity and miRNA expression; (v) inquire expression profiles of miRNAs and their targets in cancers; (vi) browse the effects of SNPs/DRVs on pre-miRNA secondary structure changes; and (vii) predict the effects of user-defined variations on miRNA-target binding or pre-miRNA secondary structure. miRNASNP-v3 is a valuable and long-term supported resource in functional variation screening and miRNA function studies.

INTRODUCTION

MicroRNAs (miRNAs) are key post-transcriptional regulators of gene expression in various biological and disease processes (1,2). Genetic variations in miRNA genes or miRNA-target binding sites could influence the biogenesis of miRNAs and miRNA-mediated regulatory functions, thus further contributing to the alteration of regulatory networks and abnormal physiological processes (3,4). For example, the single-nucleotide polymorphism (SNP) rs61764370 in the three prime untranslated region (3′UTR) of kirsten rat sarcoma viral oncogene homolog (KRAS) gene could disrupt the let-7a:KRAS interaction and promote metastasis of osteosarcoma (5). SNP rs7911488 contributes to the initiation of colorectal cancer through decreasing the expression of miR-1307 and increasing the expression of its target B-cell lymphoma 2 (Bcl2) via blocking the Dicer processing (6). Thus, a comprehensive survey and annotation of these miRNA-related variations will be a key step for studying miRNA regulation and functional variation selection.

In recent years, several resources have been developed to assess the effects of variations on the alterations of miRNA secondary structure and targeting, such as miRdSNP (7), dbMTS (8), PolymiRTS (9), SomamiR (10), MSDD (11) and miRNASNP (12,13). miRdSNP mainly focused on SNPs on 3′UTRs of coding genes. PolymiRTS analyzed the effects of SNPs on miRNAs and their targets with the latest data at 2014 (14). While SomamiR focused on the effects of cancer somatic mutations on miRNAs and their potential interactions with ceRNAs, dbMTS is a database for variations on miRNA target sites and their functional annotations. MSDD is a manually curated database containing experimentally supported associations among miRNA-
related SNVs and human diseases. Our miRNASNP is a database for miRNA-related SNP and SNP effects, which was built at 2011 and updated to v2 at 2015. Since the variation number identified in human genome was tremendously increased in last 5 years, a state-of-the-art database with comprehensive investigation of potential effects of SNPs and mutations on miRNAs and their binding sites is very necessary.

In miRNASNP-v3, we updated it with comprehensive variation data, new features and analysis tools. miRNASNP-v3 provides an almost one-stop solution for miRNA-related variation research. Users can browse and inquire the relationship between miRNA-related SNVs and diseases, explore their effects on miRNA-target bindings and energy changes in pre-miRNA secondary structures. Four new features, such as expression profiles and correlations of miRNA–target pairs, correlations between drug sensitivity and miRNA expression, online analysis tools, as well as functional changes of biological processes by enrichment analysis are integrated.

DATA SUMMARY AND OVERVIEW OF MIRNASNP-V3

In miRNASNP-v3, we obtained SNVs from dbSNP v151 (15), GWAS Catalog (16), ClinVar (17) and COSMIC v88 (18). The miRNAs’ annotations and 3'UTRs of RefSeq genes were downloaded from miRBase v22 (2) and UCSC genome browser (19), respectively (Figure 1). To better understand associations of functional miRNA-related SNVs with phenotypes, we categorized disease-related variations (DRVs) with diseases/trait annotations. After removal of redundancies, we identified 7 161 741 SNPs and 505 417 DRVs on 1897 pre-miRNAs (2630 mature miRNAs) and 3'UTRs of 18 152 genes. Specially, 5891 DRVs characterized on pre-miRNAs (including 738 DRVs on miRNA seed regions) and 499 526 DRVs on 3'UTRs are associated with 35 tissues and over 4000 diseases/trait. Besides, 46 826 SNPs on the pre-miRNAs (including 6229 SNPs on miRNA seed regions) may affect the miRNA secondary structure, biogenesis and target binding. Additionally, 7 115 596 SNPs on 3'UTRs may affect miRNA target binding (Table 1). The experimentally verified miRNA–target pairs were integrated from TarBase (422 390), miRecords (240), miRTarBase (37490), miR2Disease (641) and starBase (370 058) (20–24).

miRNASNP provided functional modules involved in human miRNAs biogenesis, predicting the effects of SNVs on pre-miRNA secondary structure changes and miRNA-target binding alterations. In this updated miRNASNP-v3, besides the expanding number of identified miRNA-related SNVs, we integrated several useful functional modules including miRNA-related DRVs and diseases/trait information, online prediction tools and functional enrichment analysis of gained/lost targets by variations in miRNA seed regions (Figure 1). We also offered expression profiles of miRNAs and correlations between miRNAs and
Table 1. Data summary and comparisons with other databases

| Data item (human)               | miRNASNP-v3 | miRNASNP-v2 | PolymRTS 3.0 | miRdSNP | SomamiR 2.0 | MSDD   | dbMTS   |
|---------------------------------|-------------|-------------|--------------|---------|-------------|--------|---------|
| Number of SNPs on miRNA seed regions | 6229        | 227         | 294          | N.A.    | N.A.        | N.A.   | N.A.    |
| Number of SNPs on pre-miRNAs    | 46 826      | 2257        | N.A.         | N.A.    | N.A.        | N.A.   | N.A.    |
| Number of SNPs on 3’UTRs         | 7 115 796   | 566 176     | 427 096      | 175 351 | N.A.        | 5 318 919 | N.A. |
| Number of DRVs on miRNA seed regions | 809        | N.A.        | N.A.         | 181     | 5           | N.A.   |         |
| Number of DRVs on pre-miRNAs    | 4 712       | N.A.        | N.A.         | N.A.    | 1779        | 72     | N.A.    |
| Number of DRVs on 3’UTRs         | 265 697     | N.A.        | N.A.         | 4 844   | 54          | 16 212 |         |
| Associated with GWAS SNPs        | 235 085     | 131 686     | 4830         | 630     | N.A.        | 1571   |         |
| Experimental validated miRNA targets | 1 027      | 393 396     | 2070         | N.A.    | 356 384     | N.A.   |         |
| Number of SNPs on 3’UTRs         | 14 491 855  | 162 441     | 1579         | N.A.    | N.A.        | N.A.   | N.A.    |
| Number of SNPs on pre-miRNAs    | 13 384 176  | 153 290     | 181          | N.A.    | N.A.        | N.A.   | N.A.    |
| Number of SNPs on miRNA seed regions | 35 816 938 | 509 791    | 906 703      | 174     | N.A.        | N.A.   | N.A.    |
| Loss of targets by SNVs on 3’UTRs | 37 427 973 | 445 737     | 923 082      | N.A.    | 182 258     | N.A.   |         |
| Number of mature miRNAs          | 2632        | 2042        | 2578         | N.A.    | 987         | 182    | 2588    |
| Secondary structure              | Yes         | Yes         | No           | No      | No          | No     | No      |
| Drug sensitivity                 | Yes         | Yes         | No           | No      | No          | No     | No      |
| Diseases/Drugs information       | Yes         | Yes         | Yes          | Yes     | Yes         | Yes    |         |
| INDELS                          | Yes         | Yes         | Yes          | Yes     | No          | No     | No      |
| miRNA/miRNA expression           | Yes         | Yes         | No           | No      | No          | No     | No      |
| Year of update                   | 2020        | 2015        | 2014         | 2012    | 2016        | 2017   | 2020    |

N.A. Not available.

The number of SNP in the table refers to SNPs and INDELs.

their targets in cancers. Moreover, miRNASNP-v3 allows users to explore the correlations between drug sensitivity and miRNA expression. Compared with the miRNASNP-v2 and other resources (Table 1), miRNASNP-v3 provides comprehensive resources and useful modules for users to investigate the roles of SNVs on miRNAs and their target sites underlying biological processes and diseases.

**IMPROVED CONTENT AND NEW FEATURES**

To reduce false positives, miRNASNP-v3 adopted the combination of two popular tools, TargetScan v7.2 and miRmap v1.1 (25–27), to determine the miRNA target binding sites. A miRNA target binding site predicted by both TargetScan and miRmap was considered as a putative target binding site. The effects of variations on the alteration of miRNA targets (gain or loss) were determined by our previous strategy (12,13). In brief, the seed region (2–8 nt from 5’ end) of a mature miRNA and 3’UTR of a gene were employed to predict the gain/loss effects of a variation on miRNA-target binding. We predicted miRNA targets by both of TargetScan and miRmap on both wild-type sequence and sequence with the variation allele. If a putative target of a mutated miRNA was not predicted by any tools on the wild-type miRNA, the miRNA-target relation was a gain one, vice versa. Similarly, the effects of variations on 3’UTRs in miRNA-target binding were predicted by the same procedure.

miRNASNP-v3 built useful modules for users to browse and explore the effects of SNPs/DRVs on the alterations of miRNA-target binding through multiple approaches (Figure 2A–C). For a mature miRNA, in Figure 2A, miRNASNP-v3 displays the number of SNPs/DRVs identified in the seed region of hsa-let-7a-2-3p (top of Figure 2A), annotations of hsa-let-7a-2-3p (middle of Figure 2A) and the number of gained/lost target genes (bottom of Figure 2A). To explore the effects of variations on gene 3’UTRs, searching an interested gene will be redirected to the gene searching result page as Figure 2B. miRNASNP-v3 presents the number of SNPs/DRVs on the gene 3’UTR and their effects on the alteration of miRNA-target binding sites (Figure 2B). In addition, users could investigate the effects of SNPs/DRVs on the changes of energy and secondary structure for a pre-miRNA (Figure 2C).

**The miRNA expression and correlations with target genes in cancers**

The expression level of a miRNA will greatly influence its functions, and the expression correlation of a miRNA and its target is an important factor for miRNA target prediction. In miRNASNP-v3, we integrated the expression data of miRNAs and their target genes in 33 cancer types from The Cancer Genome Atlas (TCGA). We used our previous tool GSCA lite (28) to extract and correlate the miRNA and gene expression in each cancer type. In brief, in each cancer type, we performed Pearson correlation analysis between miRNA and its’ predicted targets. Those correlations with a P-value <0.05 will be displayed in miRNASNP-v3. In the ‘miRNA detail’ target loss panels on website, as Figure 2D illustrated, users could click ‘Gene exp.’ to show the miRNA expression level across TCGA cancer types, and click ‘Exp. cor.’ detail button to present the correlations of miRNA and this target across cancer types. The ‘Predict’ detail button on the right of each row will show the detailed information of the miRNA binding site.

**Enrichment analysis for gained/lost genes by SNVs on miRNA**

One SNV occurs on the miRNA seed region could lead to gain and/or lose a set of target genes. This gained/lost set of target genes could be enriched in special biological pathways so that the SNV could affect the miRNA function. To further benefit users investigating biological functions of the altered miRNA targets caused by SNVs, miRNASNP-v3 provides functional modules to present Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis for gained/lost targets. After predicting gained/lost target genes of a miRNA with a SNP in seed region, the functional enrichment analysis for this set of gained/lost targets was performed via an
Figure 2. The utilities and new modules of miRNASNP-v3. (A) The SNVs in the seed regions and their effects on the target gains and losses. (B) The information of SNVs on gene 3′UTRs and their effects on the alteration of miRNA target binding. (C) The effects of SNVs on the changes of minimum free energy and secondary structure for pre-miRNAs. (D) The detail information of miRNA-target gain and loss, expression profiles, expression correlations and binding site details. (E) GO and KEGG pathway enrichment analysis for the gained or lost targets. (F) Correlations between drug sensitivity and the expression level of miRNAs. (G) Online prediction tools for potential alterations of miRNA target binding and secondary structure of a pre-miRNA for user customized sequences.
R-package clusterProfiler (29) with default parameters, and $P$-value was corrected by false discovery rate (FDR). For instance, SNP rs550418881 (G>C) in hsa-miR-1181 seed region could introduce a gain of 662 target genes. These gained target genes are mainly enriched in glutamatergic synapse, Hedgehog signaling and Wnt signaling pathways that have effect on cell fate, proliferation and differentiation (Figure 2E). Meanwhile, this SNP rs550418881 also locates in the linkage disequilibrium (LD) region of a tagSNP rs35251378 that is associated with the number of lymphocytes (30). Thus, results of this module moderately associate with a trait identified by GWAS. This could help users to link miRNA variation with diseases, further for better experimental design.

Correlations between drug sensitivity and miRNA expression

miRNASNP-v3 integrated the expression profiles of miRNAs with the drug sensitivity data. The correlations between drug sensitivity and miRNA expression were estimated by NCI-60 dataset downloaded from CellMiner Cross Database (31). The NCI-60 dataset contains the data of half cell growth inhibition concentration (GI50) of 18 724 drug/compounds and corresponding expression profiles of 335 miRNAs. The correlations between the miRNAs and drug sensitivity were evaluated by Pearson correlation coefficient between GI50 value and miRNA expression level. The $P$-value of Pearson correlation coefficient were corrected with FDR, and miRNA-drug with a FDR <0.05 was kept in the ‘miRNA detail’ panel (Figure 2F) of miRNASNP-v3.

Online prediction tools

To satisfy the demands of customized miRNA-related SNVs prediction, we integrated the miRNA target prediction methods we used in miRNASNP-v3 and built an online webserver, which is the ‘Predict’ page. On the ‘Predict’ page, there are three prediction panels. The first two panels can predict the effect of a SNV on miRNA and 3’UTR, respectively, by inputting both wild and mutated sequences. The backend methods and strategies are the same as we used in the data analysis of miRNASNP-v3. The third panel (Figure 2G) is to predict the energy change and secondary structure of a mutation on a pre-miRNA by the software ViennaRNA RNAfold v2 (32) with default parameters.

SUMMARY AND FUTURE PERSPECTIVES

As huge number of human mutation data identified recently, we updated our miRNASNP database to miRNASNP-v3, a state-of-the-art database with comprehensive miRNA and mutation data and new functional modules. In this version, to quickly explore associations between miRNA variations and diseases/traits, we integrated COSMIC, GWAS Catalog and ClinVar variations on miRNA and gene 3’UTR as DRVs. This provides a miRNA-related variation point of view to explain diseases/traits. For example, SNP rs77589182 (C>T) on the recombination signal binding protein for immunoglobulin Kappa J region like (RBPJL) 3’UTR is reported to be associated with lung cancer by GWAS (33). Meanwhile, this C to T mutation leads to a gain of miRNA hsa-miR-10a-3p and loss of hsa-miR-39c-5p target binding that could be potential underlying biological mechanism needing further experimental verification. Associating miRNAs with disease variations could help for better experimental design in disease studies. In miRNASNP-v3, we documented >4000 diseases/trait.

Although several resources dived into miRNA-related SNVs research, the data and functions in our miRNASNP-v3 are remarkable compared with the others (Table 1). miRNASNP-v3 contains the latest and most comprehensive data of SNPs and DRVs that may affect the miRNA related functions. The user-friendly web interface of miRNASNP-v3 offered intuitive views for users to browse the associations between miRNA-related variations and diseases, energy changes in pre-miRNA secondary structure, the alterations of miRNA-target binding and enriched function changes. The new data and functions will make miRNASNP-v3 a useful resource for understanding the influences of SNVs on miRNA biogenesis and miRNA-related regulations and unveiling genetic variants related molecular mechanism of complex diseases and traits.

In summary, our updated miRNASNP-v3 could be an important complementary in silico resource in miRNA-related variation research with updated data, new features and analysis tools. Moreover, our miRNASNP-v3 provided an almost one-stop solution for this area, and we will regularly update the database with new data and functions.

FUNDING

National Natural Science Foundation of China [31822030, 31801113, 31801154, 31771458]; China Postdoctoral Science Foundation [2019M652623, 2018M632830]. Funding for open access charge: National Natural Science Foundation of China 31822030.

Conflict of interest statement. None declared.

REFERENCES

1. Bartel,D.P. (2018) Metazoan MicroRNAs. Cell, 173, 20–51.
2. Kozomara,A., Birgaoanu,M. and Griffiths-Jones,S. (2019) miRBase: from microRNA sequences to function. Nucleic Acids Res., 47, D155–D162.
3. Cammaerts,S., Strazisar,M., De Rijk,P. and Del Favero,J. (2015) Genetic variants in microRNA genes: impact on microRNA expression, function, and disease. Front. Genet., 6, 186.
4. Ghanbari,M., Erkeland,S.J., Xu,L., Coltijn,J.M., Franco,O.H., Dehghan,A., Klaver,C.W. and Meester-Smoor,M.A. (2017) Genetic variants in microRNAs and their binding sites within gene 3’UTRs associate with susceptibility to age-related macular degeneration. Hum. Mutat., 38, 827–838.
5. Zhang,S., Hou,C., Li,G., Zhong,Y., Zhang,J., Guo,X., Li,B., Bi,Z. and Shao,M. (2016) A single nucleotide polymorphism in the 3’-untranslated region of the KRAS gene disrupts the interaction with let-7a and enhances the metastatic potential of osteosarcoma cells. Int. J. Mol. Med., 38, 919–926.
6. Tang,R., Qi,Q., Wu,R., Zhou,X., Wu,D., Zhou,H., Mao,Y., Li,R., Liu,C., Wang,L. et al. (2015) The polymorphic terminal-loop of pre-miR-1307 binding with MBNL1 contributes to colorectal carcinogenesis via interference with Dicer1 recruitment. Carcinogenesis, 36, 867–875.
