MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs

Chenxing Liu1,2, Fuquan Zhang1,2, Tingting Li4,5, Ming Lu4, Lifang Wang1,2, Weihua Yue1,2* and Dai Zhang1,2,3*

Abstract

Background: Numerous single nucleotide polymorphisms (SNPs) associated with complex diseases have been identified by genome-wide association studies (GWAS) and expression quantitative trait loci (eQTLs) studies. However, few of these SNPs have explicit biological functions. Recent studies indicated that the SNPs within the 3’UTR regions of susceptibility genes could affect complex traits/diseases by affecting the function of miRNAs. These 3’UTR SNPs are functional candidates and therefore of interest to GWAS and eQTL researchers.

Description: We developed a publicly available online database, MirSNP (http://cmbi.bjmu.edu.cn/mirsnp), which is a collection of human SNPs in predicted miRNA-mRNA binding sites. We identified 414,510 SNPs that might affect miRNA-mRNA binding. Annotations were added to these SNPs to predict whether a SNP within the target site would decrease/break or enhance/create an miRNA-mRNA binding site. By applying MirSNP database to three brain eQTL data sets, we identified four unreported SNPs (rs3087822, rs13042, rs1058381, and rs1058398), which might affect miRNA binding and thus affect the expression of their host genes in the brain. We also applied the MirSNP database to our GWAS for schizophrenia: seven predicted miRNA-related SNPs \( (p < 0.0001) \) were found in the schizophrenia GWAS. Our findings identified the possible functions of these SNP loci, and provide the basis for subsequent functional research.

Conclusion: MirSNP could identify the putative miRNA-related SNPs from GWAS and eQTLs researches and provide the direction for subsequent functional researches.

Keywords: microRNA, Single nucleotide polymorphism (SNP), Genome-wide association study (GWAS), Expression quantitative trait loci (eQTLs), MirSNP

Background

MicroRNAs (miRNAs) are small non-coding RNA molecules of ~22 nucleotides that primarily mediate post-transcriptional gene silencing processes in animals [1,2]. MiRNAs inactivate specific mRNAs and interfere with the translation of target proteins [3]. In mammals, miRNAs are predicted to control the activities of ~50% of all protein-coding genes [4]. As key post-transcriptional regulators, miRNAs have an important role in a wide range of biological processes, including cell proliferation, differentiation, apoptosis and metabolism [2,3]. Evidence indicates that miRNAs are also involved in the pathogenesis of complex diseases, such as cancer and mental disorders [4,5].

Complementarity to bases 2-8 of the miRNA (the seed site) is important in miRNA-mRNA binding [6,7]. MiRNAs are key regulators of gene expression; therefore, SNPs in the seed sites of miRNAs targets may create, as well as destroy, miRNA binding sites, and further affect phenotypes and disease susceptibility [8]. Identifying these seed-site SNPs could help in the further exploration of the molecular mechanism of gene dysregulation. In addition, genetic variants in miRNA genes may also have important roles by affecting miRNA maturation, which may affect disease susceptibility [8]. Certain polymorphisms in miRNA genes have been found to be associated with...
various complex diseases, including cancers, mental diseases, cardiomyopathy, and asthma (Additional file 1: Table S1).

GWAS and eQTLs are powerful methods for identifying genetic variants contributing to disease risk and gene expression. In a GWAS of schizophrenia (SCZ), Ripke et al. found the most significant SNP ($p < 1.6 \times 10^{-11}$) within an intron of a putative primary transcript for hsa-mir-137 and found four other SNPs achieving genome-wide significance that were located in predicted target sites of hsa-mir-137 [9]. It was estimated that more than 50% of the protein-coding genes are regulated by miRNAs, and each miRNA may regulate hundreds of potential targets [10,11]. Taking into account large scale of biological significance shown by miRNAs, miRNA-related SNPs could be important variations leading to traits and diseases. Identifying allele-specific miRNA-mRNA interactions may indicate the functional roles of the SNPs from GWAS and eQTLs that presently lack obvious known function.

To help identifying putative miRNA-related SNPs from researchers’ own GWAS and cis-acting eQTLs data set, we have developed a freely available database, named “MirSNP”, which provides SNPs located in predicted miRNA target sites. This database has minor allele frequency (MAF) and linkage disequilibrium (LD) information for the SNPs and has annotations concerning their creative or disruptive effects on putative target sites. The MirSNP database comprises over 414,510 predicted miRNA-related SNPs, enabling users to identify potential miRNA-related SNPs from their own GWAS or eQTLs data. In this work, we applied MirSNP to our schizophrenia GWAS data and several brain eQTLs data as examples.

Construction and content
Data sources
To store the mRNA sequences, miRNA data and SNPs, a local Structured Query Language (SQL) database was built using MySQL software. Human miRNAs were downloaded from miRBase18.0 [12]. For SNP catalogs, four tables from dbSNP135 [13] were used (see URLs). To maximize the consistency between different databases, mRNA sequence files were obtained from NCBI (see URLs) rather than from the UCSC genome browser. In total, 42,733 mRNA sequences and 513,249 SNPs located in the 3′UTRs of mRNA were eligible for subsequent analysis.

Identifying polymorphisms in pre-miRNA genes
We identified SNPs either located in human pre-miRNA genes or in their adjacent upstream/downstream 200 bp regions by comparing the coordinates of both SNPs and related miRNAs (Figure 1). There were 12,846 polymorphisms, including 1,940 SNPs in pre-miRNA sequences. An SQL script was written to calculate the SNP density (Figure 2). The SNP density in pre-miRNA regions declined rapidly to 0.43 SNPs/kb if we considered only SNPs having a MAF above 0.01 in four populations (CEU, HCB, JPT, YRI).

Identifying polymorphisms in miRNA target sites
The method of identifying polymorphisms in mRNAs affecting miRNA-mediated processes is shown in Figure 3. The information as to whether or not a SNP is located in the 3′UTR of a mirRNA came from dbSNP135. Only SNPs located in mRNA 3′UTR areas were recorded in the local SQL database. Preforming sequence alignment between 20-bp DNA sequences surrounding 3′UTR SNPs and the corresponding mRNA sequences, variants were mapped onto their mRNA sequences. Subsequently, each SNP in our database had two to four mRNA sequence records corresponding to different alleles. Using the mRNA sequences of SNPs and miRNA mature sequences, we obtained the predicted target results using an miRNA target prediction algorithm, miRanda, which has highest sensitivity among eight tested algorithms [14].

Although there are examples that imperfect 7-nt seed site pairing can be functional, there is overwhelming evidence to support the hypothesis that Watson-Crick pairing to the miRNA seed site is the most important feature for miRNA prediction and function [6,7]. Therefore, we adopted the 7-nt seed region in the miRNA as the major criterion in the miRanda algorithm [15]. In detail, to predict miRNA binding sites, we applied miRanda v3.3a with the default pairing score cutoff of “+140” and “–strick” command, which only considered stringent 7-nt seed pairing (requiring an uninterrupted match of at least seven nucleotides from the 5′-end of the miRNA).

Additional information about MirSNP
A requirement for perfect 7-nt seed site pairing improves the reliability of miRNA target prediction, particularly when seed motif is conserved in the UTR regions of whole-genome alignments [16,17]. We downloaded the conservative information of whole-genome alignments (phastCons 46way vertebrates from UCSC ftp site, see URLs) [18] and then added the average value of conservative scores of 7-nt seed motif to our database. The score of mirSVR methodology, a machine learning method for ranking miRNA target sites [19], were also added to the MirSNP database as annotation. However, the data of MirSNP are not identical to that of the mirSVR. Imperfect overlap may traced back to SNPs (the mirSVR methodology didn’t consider the impact of SNPs on miRNA-mRNA bindings), the use of different UTR
not all results in MirSNP should have the score of mirSVR. SNPs become more important if they have a high frequency or are undergoing positive selection. Therefore, we have added MAF information from dbSNP135 (see URLs) for the SNPs into MirSNP. Based on this information, the results of MirSNP filtered by MAF could be displayed. Analysis of the MAF data revealed that there were 32,822 SNPs located in miRNA-mRNA binding sites with MAFs greater than 0.01 in the four populations. In addition, 122 SNPs in pre-miRNA genes had MAF data and the remaining 1,818 SNPs lacked MAF data.

LD information was obtained for each SNP from the HapMap project [20]. The Phase2 HapMap fileset (see URLs) was downloaded for the four populations and the linkage of the SNPs was calculated using a threshold of $r^2$ greater than 0.8 using the PLINK software [21].
and MAF information of SNPs were stored into separate tables in download page (see URLs).

**Database construction**

All the useful data were stored in a local MySQL database. We used Django, a web application framework written in Python, to build a user-friendly online website (http://cmbi.bjmu.edu.cn/mirsnp).

**Utility**

The website

We obtained over one million records for 414,510 miRNA-related SNPs. These SNPs were classified into four groups, labeled as create, enhance, decrease or break. To display the records in the MirSNP database, we designed a user-friendly MirSNP web site (see URLs) to allowing searching for SNPs in putative miRNA target sites. On the frame “single search” (Figure 4A), the system allows users to search records by entering a RefGene name, mRNA id, SNP RSid or mature miRNA name. All identifiable RefGene names and mRNA ids for this search can be found by clicking the hyperlinks “Refgene name” and “mRNA id” on the page. After pressing the “Search” button, the results are presented on a new page (Figure 4B). The record in Figure 4B shows that the A-allele of rs56352346 may promote the binding of CYP4B1 gene and hsa-let-7a-2-3p and when rs56352346 is the C-allele, the mature miRNA and gene cannot combine (Figure 4B). The specific explanation of each column is provided in the help page. Additionally, on the “single search” frame, users can choose whether to display the binding alignment as well as whether to filter the output by MAF.

We would like to recommend the frame named “Query disease & trait associated SNPs” (Figure 5). This frame permits the search of SNPs from GWAS and eQTL studies. The acceptable input is a list of SNP RSid in a text file. Here, users can query not only the submitted SNPs, but also their linked SNPs (r^2 > 0.8). For example, searching for associated SNPs in our schizophrenia GWAS data (p < 0.0001) in this frame, four additional SNPs were identified compared to using the “batch search” frame that does not consider linked SNPs.

**The use of MirSNP for brain eQTL data sets**

Many studies have implicated the association between SNPs and the expression of their host genes. We speculate that it will be of great significance to combine our miRNA-related SNPs data and eQTL data. We used the MirSNP database for three human brain eQTL data sets, including human cortical samples from 193 individuals [22], cortex samples from 269 individuals [23] and samples obtained from 150 subjects [24]. SNPs located in miRNA-mRNA binding sites could affect the expression of their host genes; therefore, we only considered eQTLs that had a cis-effect on the host genes. Four putative miRNA-related SNPs (rs3087822, rs13042, rs1058381, rs1058398) were selected and they were statistically significant, genome-wide, in the three brain eQTL data (Table 1). Based on our in silico analysis, we hypothesize that these SNPs may affect the miRNA mechanism and thus affect the mRNA expression of their host genes in...
Further experiments are necessary to confirm our speculation concerning these SNPs.

The use of MirSNP for a schizophrenia GWAS data set

Our previous GWAS data [25], involving 746 SCZ cases and 1,599 healthy controls, identified a set of 7,705 SNPs having a statistical significance of \( p < 0.01 \). Here, we conducted a genome-wide analysis for these GWAS SNPs falling within computationally predicted miRNA targets. We combined putative miRNA-related SNPs and GWAS SNPs with SNP id as the key. To increase the range of combination, we used HapMap data and the software Plink to calculate \( r^2 \) between pairwise SNPs. The GWAS data are from Chinese Han population; therefore, we chose 90 Asian individuals from the HapMap project for LD analyses. A subset of 4,997 SNPs in predicted miRNA target sites were in our GWAS analyses. Hence, we set \( 1.0 \times 10^{-5} \) (0.05/4997) as the threshold of statistical significance. Three polymorphisms were identified (Table 2). The SNP that showed the strongest association with schizophrenia was found in the TBC1D15 gene (\( p = 4.0 \times 10^{-6} \) in the Chinese Han population). The in silico analysis implied that three SNPs (rs17110432, rs11178988, and rs11178989) in 3’UTR area of TBC1D15 may affect the miRNA-mRNA binding process. However, further experiments are necessary.

We also overlapped the SNPs around microRNA genes in our GWAS data of schizophrenia. A subset of 108 SNPs around miRNA genes was identified. We found one SNP (rs10173558), which is only 5 bp away from the start site of "hsa-mir-1302-4", having a statistical significance lower than 0.001. Yuan et al. reported that the...
highest expression level of miR-1302’ target genes was in the nervous system and the genes were enriched in both synapses and intracellular membrane-bounded organelles [26]. This finding implied a potential relevance of miR-1304 to psychiatric disorders.

Discussion

In recent years, increasing numbers of databases have been published to aid researchers to explore the impact of SNPs on miRNA targets [27-33]. Some researchers, such as Richardson [31] and Ziebarth [32], have provided links between SNPs in miRNA target sites, cis-acting eQTLs and the results of GWAS. Previous works summarized the characteristics of miRNA-related SNPs and showed the potential of applying such databases in GWAS and eQTL researches. However, these databases may not be suitable for a single GWAS or eQTL data set. Some databases cannot perform batch search for numbers of SNPs and some cannot provide effective miRNA-related information for strongly-associated loci. MirSNP was developed to identify putative miRNA-related SNPs from single data sets of GWAS or eQTL, especially from newly published data sets. First, our analysis covered 513,249 known 3’UTR SNPs based on dbSNP135 and we used highly consistent data sources to avoid data loss.

Table 1 Brain eQTLs found in predicted miRNA-mRNA binding sites

| mirSNP | Gene | Involved miRNAs | Allele | eQTL | Linkagea (Myers, A.J. et al)b | p-value | p-value (Colantuoni, C. et al)c | p-value (Gibbs, J.R. et al)d |
|--------|------|-----------------|--------|------|------------------------------|---------|-------------------------------|-------------------------------|
| rs3087822 | CRIP1 | hsa-mir-200a-3p, hsa-mir-20b-3p, hsa-mir-29b-1-5p, hsa-mir-335-3p | A/G | rs3087822 | 1 | 8.45E-05 | 9.54E-18 | 3.20E-13 |
| rs13042 | FAM82B | hsa-mir-202-5p, hsa-mir-362-5p, hsa-mir-500b, hsa-mir-1914-5p | A/Gc | rs13042/ rs4961193 | 1/0.864 | 5.03E-06 | 8.52E-16 | 1.30E-08 |
| rs1058381 | RABEP1 | hsa-mir-4760-3p | A/G | rs1058381/ rs1065483 | 1/0.954 | 0.00254 | 7.81E-19 | 2.89E-08 |
| rs1058398 | RABEP1 | hsa-mir-134, hsa-mir-3118, hsa-mir-943, hsa-mir-192-3p, hsa-mir-5002-3p | A/G | rs1065483 | 0.955 | 0.00254 | 7.81E-19 | 3.92E-10 |

a|Linkage analysis of predicted miRNA-related SNPs and eQTLs in 90 CEU individuals from HapMap project.
b|A survey of gene expression of 193 normal human brain tissues (cortex) [22]. The p-values have been adjusted by Bonferroni correction.
c|A survey of gene expression of 269 normal human brain tissues (prefrontal cortex) [23]. Genome-wide Bonferroni corrected p = 0.05 (2.6e-12).
d|A survey of gene expression of 150 normal human brain tissues (cerebellum, frontal cortex, pons, and temporal cortex) [24].
e|Although rs13042 has four allelotypes in MirSNP, only A and G alleles have frequencies in populations.

Eleven putative miRNA-related SNPs link to SCZ-GWAS SNPs with a significance of p < 0.0001.

Table 2 Putative miRNA-related SNPs associated with schizophrenia

| mirSNP | Involved gene | MAFa | Involved miRNAs | Alleleb | SCZ-associated SNP | Linkagec p value OR |
|--------|---------------|------|-----------------|---------|---------------------|-------------------|
| rs11178988 | TBC1D15 | 0.156 | hsa-mir-145-3p, hsa-mir-3680-3p, hsa-mir-5689, hsa-mir-1294 | C/T | rs17110426 | 0.945521 | 4.06E-06 | 0.6449 |
| rs11178899 | TBC1D15 | 0.156 | hsa-mir-4501 | A/C | rs17110426 | 0.945521 | 4.06E-06 | 0.6449 |
| rs17110432 | TBC1D15 | 0.167 | hsa-mir-1193, hsa-mir-335-3p, hsa-mir-29b-2-5p | A/G | rs17110426 | 1 | 4.06E-06 | 0.6449 |
| rs11544338 | FAM117B | 0.305 | hsa-mir-409-3p, hsa-mir-653 | A/G/ T | rs11544338 | 1 | 5.48E-06 | 0.6611 |
| rs11680951 | FAM117B | 0.291 | hsa-mir-516a-3p, hsa-mir-376a-5p, hsa-mir-1287, hsa-mir-145-3p, hsa-mir-5191, hsa-mir-516b-3p | A/G/ T | rs11544338 | 1 | 5.48E-06 | 0.6611 |
| rs6058896 | DNM3T3B | 0.078 | hsa-mir-30b-5p, hsa-mir-3686, hsa-mir-30c-5p, hsa-mir-4773, hsa-mir-2278, hsa-mir-589-3p | C/T | rs6058894 | 0.946356 | 2.31E-05 | 1.584 |
| rs10484565 | TAP2 | 0.078 | hsa-mir-3689b-5p, hsa-mir-3177-5p, hsa-mir-3689a-5p, hsa-mir-3689f, hsa-mir-190b, hsa-mir-3689e, hsa-mir-190a | A/G | rs10484565 | 1 | 3.58E-05 | 1.448 |
| rs670358 | CDC42BPG | 0.395 | hsa-mir-4305, hsa-mir-331-3p, hsa-mir-4638-3p, hsa-mir-4705, hsa-mir-3690, hsa-mir-5195-5p, hsa-mir-4308 | A/G | rs670358 | 1 | 4.68E-05 | 0.7677 |
| rs11563929 | STEAP2 | 0.1 | hsa-mir-16-2-3p, hsa-mir-195-3p, hsa-mir-3942-5p, hsa-mir-4703-5p, hsa-mir-4766-3p | G/T | rs11563929 | 1 | 6.81E-05 | 0.7281 |
| rs16870907 | TAP2 | 0.047 | hsa-mir-4760-3p, hsa-mir-330-3p | C/T | rs16870907 | 1 | 8.48E-05 | 1.589 |
| rs12178 | ZBTB34 | 0.456 | hsa-mir-1267, hsa-mir-501-5p, hsa-mir-3613-3p, hsa-mir-653 | A/G/ T | rs12178 | 1 | 9.66E-05 | 1.277 |

aMinor allele frequencies (MAF) are from an HCB population.
bAllele reports are from NCBI.
cLinkage analysis of putative miRNA-related SNPs and SCZ-associated SNPs in 90 Asian individuals from the HapMap project.
while integrating different data. Furthermore, sequence alignments between surrounding DNA sequences of SNPs and the corresponding mRNA sequences were used to map variants into their mRNA sequences. Finally, our web site was designed to directly use GWAS or eQTL data sets in a batch query, particularly considering the linked SNPs in different populations.

The MirSNP database stores a large number of records of SNPs in predicted miRNA targets sites and we concentrated on providing a convenient search platform so that recent GWAS or eQTL results can be placed on this platform for batch retrieval. In MirSNP, all 3’UTR SNPs stored in dbSNP135 database (513,249) were analyzed. Compared with other databases, MirSNP obtained more results for both common and rare SNPs that might influence miRNA processes. In Table 3 and Figure 6, a comparison between MirSNP and three existing databases (PolymiRTS, Mirsnpscore and Patrocles) are presented. To compare the sensitivity of MirSNP prediction, we identified 13 validated miRNA-related SNPs by literature review: most of the SNPs coming from the table 2 in Sethupathy and Collins [34]. Of the 13 cases (Table 4), nine were identified using MirSNP. The SNPs that not found were either not located in the 3’UTR based on our records (rs34764978, rs13212041) or do not have a perfect 7-nt binding in the seed site (rs2735383, rs9341070). For the other three databases, from thirteen cases, no more than five validated SNPs were identified. Therefore, the MirSNP database could cover majority of validated miRNA-related SNPs.

MiRNAs can downregulate gene expression by two posttranscriptional mechanisms: mRNA cleavage or translational repression. In animals, miRNA is thought to have a repressive effect that influences protein expression, not the mRNA levels. However, it has been estimated that over 80 percent of miRNAs acted to lower mRNA levels, which shows that mRNA destabilization is the primary action mode of miRNAs on target mRNAs [35].

Table 3 Comparison among four databases

| Data source | MirSNP | PolymiRTS 2.0 | Mirsnpscore* | Patrocles |
|-------------|--------|---------------|--------------|-----------|
| Predicting strategy | NCBI | UCSC | UCSC | UCSC |
| miRNA-related SNPs | 414510 | 117037 | 19513 | 26578 |
| miRNA-related SNPs with MAF > 0.01 | 32822 | 21938 | 19208 | 7334 |
| Involved Genes | 17569 | 13129 | 8495 | 11314 |
| Involved mature miRNAs | 1921 | 1738 | 1222 | 846 |
| Records (SNPs with MAF > 0.01) | 121796 | 78351 | 69026 | 10908 |
| Overlap with MirSNP | | | | |
| Number of SNPs (MAF > 0.01) coinciding with MirSNP | 15913 | 17481 | 5759 |

Figure 6 Comparison between MirSNP and three similar databases.
studies have shown that SNPs in seed-sites region can significantly change the expression of the target mRNA and protein [36-40]. eQTL analysis is an experimental method for exploring the relationship between SNPs and mRNA expression at a high throughput. Genetic variants that create or destroy an miRNA binding site may be the causal cis-acting eQTLs. The combination of our MirSNP database and eQTL data will provide possible explanations for the eQTLs. We have already identified four SNPs in predicted miRNA target sites that were proven to be brain eQTLs in three independent studies. Further experiments are needed to prove that these eQTLs affect gene expression through an miRNA mechanism. On the other hand, using GWAS, enormous numbers of associations between SNPs and diseases have been reported. There are many disease-associated SNPs that function by miRNA regulation. Overlapping miRNA-related SNPs and the existing GWAS data could identify possible biological mechanisms for these disease-associated variations and provide the in silico basis for further studies.

Our aim was to merge the MirSNP database with high throughput SNP experimental data. We identified many SNPs in predicted miRNA targets and indicated their potential functions based on sequence algorithms.

| MiRNA Target gene | Functional SNP | Reported associated disease | MirSNP | PolymiRTS2 | Mirsnpscore | Patrocles | Pubmedid | Journal | Author |
|-------------------|----------------|----------------------------|--------|------------|-------------|-----------|---------|---------|--------|
| hsa-mir-629       | NBS1 rs2735383 | Lung cancer                | ✓      | ✓          |             |           | 22114707 Carcinogenesis Yang L |
| hsa-mir-184       | TNFAIP2 rs8126 | Squamous cell carcinoma of the head and neck | ✓      | ✓          |             |           | 21934093 Carcinogenesis Liu Z |
| hsa-mir-1827      | MYCL1 rs3134615 | Small-cell lung cancer     | ✓      | ✓          | ✓           |           | 21676885 Cancer Res Xiong F |
| hsa-mir-148a      | HLA-C rs67384697 | HIV                       | ✓      |            |             |           | 21499264 Nature Kulkarni S |
| hsa-mir-191       | MDM4 rs4245739 | Ovarian carcinomas         | ✓      | ✓          |             |           | 21084273 Cancer Res Wynendaele J |
| hsa-mir-125b      | BMPR1B rs1434536 | Breast cancer             | ✓      | ✓          | ✓           | ✓         | 19738052 Cancer Res Saetrom P |
| hsa-mir-510       | HTR3E rs56109847 | Diarrhea predominant irritable bowel syndrome | ✓      |            |             |           | 18614545a Hum. Mol. Genet Kapeller J |
| hsa-mir-96        | HTR1B rs13212041 | Arson or property damage  | ✓      |            |             |           | 18283276a Mol. Psychiatry Jensen KP |
| hsa-mir-433       | FGF20 rs12720028 | Parkinson’s disease        | ✓      |            |             |           | 18252210a Am. J. Hum. Genet Wang G |
| hsa-mir-148/152   | HLA-G rs10632320 | Childhood asthma           | ✓      | ✓          | ✓           | ✓         | 17847008a Am. J. Hum. Genet Tan Z |
| hsa-mir-24        | DHFR rs34764978 | Methotrexate resistance    | ✓      |            |             |           | 17686970a PNAS Mishra PJ |
| hsa-mir-155       | AGTR1 rs5186     | Hypertension               | ✓      |            |             |           | 17668390a Am. J. Hum. Genet Sethupathy P |
| hsa-mir-206       | ESR1 rs9341070   | Breast cancer              | ✓      |            |             |           | 17312270a Mol. Endocrinol Adams BD |
| hsa-mir-24        | SUTR1 rs193302862 | Tourette’s syndrome        | ✓      |            |             |           | 16224024a Science Abelson JF |

Some cases are not in MirSNP because the SNPs were not located in the 3’UTR based on our records (rs34764978, rs13212041) or do not have a perfect 7-nt binding in the seed site (rs2735383, rs9341070).

*These cases are from Table 2 in Sethupathy and Collins [34].
Unfortunately, the expression information of miRNAs and mRNAs are not supplied in our database. We considered the possibility of an miRNA and mRNA combining, but under the complex mechanism of spatial and temporal expression, we had no idea if the two molecules would encounter each other in vivo. Combining MirSNP with additional databases containing expression information, such as miRGator [41], would improve the functionality of the database.

Conclusions
MirSNP is a database of SNPs in predicted miRNA target sites, based on information from dbSNP135 and mirBASE18. MirSNP is highly sensitive and covers most experiments confirmed SNPs that affect miRNA function. The results suggest that our prediction and data processing are full-scale. MirSNP may be combined with researchers’ own GWAS or eQTL positive data sets to identify the putative miRNA-related SNPs from traits/diseases associated variants. We aim to update the MirSNP database as new versions of mirBASE and dbSNP database become available.

Availability and requirements
MirSNP is publicly available on the internet at http://cmbi.bjmu.edu.cn/mirsnp.

References
1. He L, Hannon GJ: MicroRNAs: small RNAs with a big role in gene regulation. Nat Rev Genet 2004, 5(7):522–531.
2. Mendell JT: MicroRNAs: critical regulators of developmental, cellular physiology and malignancy. Cell Cycle 2005, 4(9):1179–1184.
3. Bartel DP: MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 2004, 116(2):281–297.
4. Krol J, Loedige I, Filipowicz W: The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet 2010, 11(9):597–610.
5. Gao F: Posttranscriptional control of neuronal development by microRNA networks. Trends Neurosci 2008, 31(1):20–26.
6. Brennecke J, Stark A, Russell RB, Cohen SM: Principles of microRNA-target recognition. PLoS Biol 2005, 3(3):e85.
7. Lai EC: Micro RNAs are complementary to 3’ UTR sequence motifs that mediate negative post-transcriptional regulation. Nat Genet 2002, 30(4):363–364.
8. Ryan BM, Robles AI, Harris CC: Genetic variation in microRNA networks: the implications for cancer research. Nat Rev Cancer 2010, 10(6):389–402.
9. Ripke S, Sanders AR, Levinson DF, Sklar P, Holmes PA, Lin DY, Duan J, Ophoff RA, Andreasen OA, et al: Genome-wide association study identifies five new schizophrenia loci. Nat Genet 2011, 43(10):969–976.
10. Bartel DP: MicroRNAs: target recognition and regulatory functions. Cell 2009, 136(2):215–223.
11. Vinnnet O: Origin, biogenesis, and activity of plant microRNAs. Cell 2009, 136(4):669–687.
12. Kozomara A, Griffiths-Jones S: miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 2011, 39(Database issue): D152–D157.
13. Sherry ST, Ward MH, Kholodov M, Sklar P, Lee L, Shulman S, Janet U, Grant S, Garver A, Baker M, Phan L, Smigielski EM, Sirotkin K: dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001, 29(1):308–311.
14. Alexiou P, Maragakis M, Papadopoulos GL, Rezcko M, Hatzigeorgiou AG: Lost in translation: an assessment and perspective for computational microRNA target identification. Bioinformatics 2009, 25(23):3049–3055.
15. Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. Nucleic Acids Res 2008, 36(Database issue):D149–D153.

16. Berezhkov E. Evolution of microRNA diversity and regulation in animals. Nat Rev Genet 2011, 12(1):846–860.

17. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 2005, 120(1):15–20.

18. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenblum K, Clawson H, Spieht J, Hillier LW, Richards S, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 2005, 15(8):1034–1050.

19. Betel D, Koppal A, Agius P, Sander C, Leslie C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol 2010, 11(8):R90.

20. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007, 449(7164):851–861.

21. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association analysis. Am J Hum Genet 2007, 81(3):559–575.

22. Myers AJ, Gibbs JR, Webster JA, Rohrer K, Zhao A, Marlowe L, Kaleem M, Yue WH, Wang HF, Sun LD, Tang FL, Liu ZH, Zhang HX, Li WQ, Zhang YL, Arepalli S, Dillman A, Rafferty IP, Troncoso J, et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genet 2010, 6(3):e1000952.

23. Yue WH, Wang HF, Sun LD, Tang FL, Liu ZH, Zhang HX, Li WQ, Zhang YL, Zhang Y, Ma CC, et al. Genome-wide association study identifies functional polymorphisms in the 3′ untranslated region of MDM4 affects ovarian cancer progression. Genome Biol 2007, 860:R221.

24. Myers AJ, Gibbs JR, Webster JA, Rohrer K, Zhao A, Marlowe L, Kaleem M, Yue WH, Wang HF, Sun LD, Tang FL, Liu ZH, Zhang HX, Li WQ, Zhang YL, Arepalli S, Dillman A, Rafferty IP, Troncoso J, et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genet 2010, 6(3):e1000952.

25. Yue WH, Wang HF, Sun LD, Tang FL, Liu ZH, Zhang HX, Li WQ, Zhang YL, Zhang Y, Ma CC, et al. Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. Nat Genet 2011, 43(12):1228–1231.

26. Yuan Z, Sun X, Jiang D, Ding Y, Lu Z, Gong L, Liu H, Xie J. Origin and evolution of a placental-specific microRNA family in the human genome. BMC Evol Biol 2010, 10:346.

27. Barenboim M, Zoltick BJ, Guo Y, Weinberger DR. MicroSNiPer: a web tool for prediction of SNP effects on putative microRNA targets. Hum Mutat 2010, 31(11):1223–1232.

28. Bruno AE, Li L, Kalabas JL, Pan Y, Yu A, Hu Z. miRdSNP: a database of disease-associated SNPs and microRNA target sites on 3′UTRs of human genes. BMC Genomics 2012, 13:84.

29. Gong J, Tong Y, Zhang HM, Wang K, Hu T, Shan G, Sun J, Guo AY. Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. Hum Mutat 2012, 33(1):254–263.

30. Hard S, Charter C, Coppiters W, Georges M, Baurain D. Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. Nucleic Acids Res 2010, 38(Database issue):D640–D645.

31. Richardson K, Lai CQ, Parnell LD, Lee YC, Ordovas JM. A genome-wide survey for SNPs altering microRNA seed sites identifies functional candidates in GWAS. BMC Genomics 2011, 12:504.

32. Ziebarth JD, Bhattacharya A, Chen A, Cui Y. PolymiRTS Database 2.0: linking polymorphisms in microRNA target sites with human diseases and complex traits. Nucleic Acids Res 2012, 40(Database issue):D216–D221.

33. Thomas LF, Saioto T, Saelens P. Inferring causative variants in microRNA target sites. Nucleic Acids Res 2011, 39(16):e109.

34. Sethupathy P, Collins FS. MicroRNA target site polymorphisms and human disease. Trends Genet 2008, 24(10):489–497.

35. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 2010, 466(7308):835–840.

36. Yang L, Li Y, Cheng M, Huang D, Zheng J, Liu B, Ling X, Li Q, Zhang X, Ji W, et al. A functional polymorphism at microRNA-629-binding site in the 3′-untranslated region of NS1 gene confers an increased risk of lung cancer in Southern and Eastern Chinese population. Carcinogenesis 2012, 33(2):338–347.

37. Liu Z, Wei S, Ma H, Zhao M, Myers JN, Weber RS, Sturgis EM, Wei Q. A functional variant at the mir-184 binding site in TNFAIP2 and risk of squamous cell carcinoma of the head and neck. Carcinogenesis 2011, 32(11):1668–1674.

38. Wynendaele J, Bohnke A, Leucci E, Nielsen SJ, Lambertz I, Hammer S, Srizersny N, Kubitzza D, Wolf A, Gradhand E, et al. An illegitimate microRNA target site within the 3′ UTR of MDMA affects ovarian cancer progression and chemosensitivity. Cancer Res 2010, 70(23):9641–9649.

39. Saelens P, Biesinger J, Li SM, Smith D, Thomas LF, Majzoub K, Rivas GE, Allun J, Rousi J, Kronitsis TC, et al. A risk variant in an mir-125b binding site in BMPR1B is associated with breast cancer pathogenesis. Cancer Res 2009, 69(18):7459–7465.

40. Sun Q, Ku H, Zeng Y, Xia Y, Wang Y, Jing Y, Yang L, Wang B. Hsa-mir-27a genetic variant contributes to gastric cancer susceptibility through affecting mir-27a and target gene expression. Cancer Sci 2010, 101(10):2241–2247.

41. Cho S, Jun Y, Lee S, Choi HS, Jung S, Jang Y, Park C, Kim S, Kim W. miRgator v2.0: an integrated system for functional investigation of microRNAs. Nucleic Acids Res 2011, 39(Database issue):D158–D162.

Cite this article as: Liu et al.: MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. BMC Genomics 2012;13:661.

Cite this article as: Liu et al.: MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. BMC Genomics 2012;13:661.