BMC
Genomics

miRdSNP: a database of disease-associated SNPs and microRNA target sites on 3'UTRs of human genes

Andrew E Bruno[1], Li Li[2], James L Kalabus[3], Yuzhuo Pan[3], Aiming Yu[3] and Zihua Hu[1,4,5*]

Abstract

Background: Single nucleotide polymorphisms (SNPs) can lead to the susceptibility and onset of diseases through their effects on gene expression at the posttranscriptional level. Recent findings indicate that SNPs could create, destroy, or modify the efficiency of miRNA binding to the 3'UTR of a gene, resulting in gene dysregulation. With the rapidly growing number of published disease-associated SNPs (dSNPs), there is a strong need for resources specifically recording dSNPs on the 3'UTRs and their nucleotide distance from miRNA target sites. We present here miRdSNP, a database incorporating three important areas of dSNPs, miRNA target sites, and diseases.

Description: miRdSNP provides a unique database of dSNPs on the 3'UTRs of human genes manually curated from PubMed. The current release includes 786 dSNP-disease associations for 630 unique dSNPs and 204 disease types. miRdSNP annotates genes with experimentally confirmed targeting by miRNAs and indexes miRNA target sites predicted by TargetScan and PicTar as well as potential miRNA target sites newly generated by dSNPs. A robust web interface and search tools are provided for studying the proximity of miRNA binding sites to dSNPs in relation to human diseases. Searches can be dynamically filtered by gene name, miRBase ID, target prediction algorithm, disease, and any nucleotide distance between dSNPs and miRNA target sites. Results can be viewed at the sequence level showing the annotated locations for miRNA target sites and dSNPs on the entire 3'UTR sequences. The integration of dSNPs with the UCSC Genome browser is also supported.

Conclusion: miRdSNP provides a comprehensive data source of dSNPs and robust tools for exploring their distance from miRNA target sites on the 3'UTRs of human genes. miRdSNP enables researchers to further explore the molecular mechanism of gene dysregulation for dSNPs at posttranscriptional level. miRdSNP is freely available on the web at http://mirdsnp.ccr.buffalo.edu.

Background

Single nucleotide polymorphisms (SNPs) underlie disease susceptibility through their effects on protein function and gene expression. Most identified mutations are non-synonymous SNPs that result in amino acid changes in proteins. It is well known that non-coding disease-associated SNPs (dSNPs) within regulatory regions of the genome can result in gene dysregulation at either transcriptional or posttranscriptional level. One potential source for the latter is SNPs which create,

destroy, or modify the efficiency of miRNA binding to the 3'UTR of a gene. Supporting this idea, SNPs within the miRNA target sites of genes have been implicated in hippocampal sclerosis [1], parkinson disease [2], tourette's syndrome [3], asthma [4], cardiovascular disease [5], neurodegenerative disease [6], periodontal diseases [7], tumor susceptibility [8], and various types of cancers [9-12]. Other than SNPs within miRNA target sites, SNPs outside miRNA binding site can affect miRNA function. One recent finding [13] demonstrates that a polymorphism outside the miR-24 binding site in the 3'UTR of human dihydrofolate reductase (DHFR) affects DHFR expression by interfering with miR-24 function, resulting in DHFR over-expression and methotrexate resistance. There is also a report suggesting that SNPs

* Correspondence: zihuahu@ccr.buffalo.edu
[1]Center for Computational Research, New York State Center of Excellence in Bioinformatics & Life Sciences, State University of New York at Buffalo, Buffalo NY 14260, USA
Full list of author information is available at the end of the article

within a certain region on both sides of miRNA target sites have the highest influence on miRNA binding to the target sites and that SNPs on the rest of 3'UTR sequences have impact on miRNA function as well [14].

A few databases have been built to aid researchers in exploring the impact of SNPs on the binding of miRNA and targets. While polymiRTS [15] represents the polymorphism in putative miRNA target sites and their involvement in quantitative trait locus effects, Patrocles database compiles DNA sequence polymorphisms in the 3'UTR of genes in seven vertebrate species that perturb miRNA-mediated gene regulation [16]. The findings of dSNPs on the 3'UTRs have been growing rapidly during the past few years. Furthermore, a few dSNPs [1-5,7-12] have been proven to alter gene expression through modifying specific miRNA target sites, however the molecular mechanism causing diseases for majority of dSNPs on the 3'UTRs is largely not known. There is a strong need to have a database specifically recording dSNPs and tools for capturing their proximity to miRNA target sites on the 3'UTRs so that researchers can explore further the molecular mechanism of gene dysregulation for dSNPs at posttranscriptional level.

Aiming to provide a comprehensive data source of dSNPs affecting posttranscriptional regulation of disease-related genes and tools for exploring the nucleotide distance between miRNA target sites and dSNPs, we present here miRdSNP, a database of manually curated dSNPs on the 3'UTRs of human genes from available publications in PubMed. A robust web interface and advanced search tools are provided showing the nucleotide distance between dSNPs and predicted miRNA target sites from the most popular algorithms, namely TargetScan [17] and PicTar [18]. We also incorporated all SNPs on 3'UTRs of individual genes into the database so the relationship of SNPs with both dSNPs and miRNA binding sites can be analyzed using the web interface. In addition, we also include predicted miRNA target sites generated by dSNPs based on our analysis and annotate genes with experimentally confirmed targeting by miRNAs from four separate curated databases.

Construction and content
Data Sources
miRdSNP provides a manually curated dataset of dSNPs on the 3'UTRs of human genes from all PubMed articles linked in Entrez. These include 786 dSNP-disease associations from 630 unique dSNPs and 204 types of diseases. Out of these diseases 97 (47.5%) of them are associated with only 1 dSNP and 153 (75%) are associated with no more than 3 dSNPs (Figure 1). Breast cancer has the highest number with 52 associated dSNPs, followed by type 2 diabetes, schizophrenia, rheumatoid arthritis, obesity, and colorectal cancer with 42,

38, 24, 21 and 20 dSNPs, respectively. The database also incorporates reference sequence (RefSeq) genes, predicted miRNA target sites, and SNP sequence data into a single consolidated resource.

Database Construction
We obtained 3'UTRs of human RefSeq genes (hg18 March 2006 assembly) from the UCSC Genome browser [19]. A total of 19,834 genes (including introns) were parsed and loaded into miRdSNP and the chromosomal coordinates for each gene were indexed along with the exon lengths. If a gene had multiple transcripts we selected the one with the longest sequence length. We obtained the SNP dataset from UCSC Genome browser (NCBI dbSNP [20] Build 130). Of 18,833,531 SNPs we indexed the genomic coordinates for a subset of 175,351 located on 3'UTRs of 16,810 genes. SNPs aligning to more than 1 locus or mapped to intron regions were excluded. We then annotated dSNPs using an in-house developed data pipeline which searches for PubMed articles linked to SNPs. The data pipeline queries ELink from the Entrez Programming Utilities to find all PubMed IDs linked to SNPs via the "snp pubmed" link. We queried all 3'UTR SNPs and found 2,785 PubMed publications linked to 16,447 unique SNPs. We then manually reviewed these literatures and identified 630 dSNPs for 204 human diseases from 754 publications. The data pipeline for harvesting PubMed-SNP associations from Entrez is automated and the results are displayed in a web interface, allowing multiple users to manually review articles in parallel. This enables us to update the curated dSNP dataset frequently as the literature evolves.

We captured linkage disequilibrium (LD) information for each dSNP using the latest data provided by the HapMap project [21] (version 2009-04_rel27). We downloaded the raw LD files for each population and searched for pairs of genetic variants that included dSNPs. We then indexed all variants in strong LD of each dSNP using an $R^2 \geq 0.80$ threshold. Regional LD plots were also generated for each dSNP using a modified version of the R script provided by SNAP [22] and data from the CEU (CEPH Utah Residents with Northern and Western European Ancestry) population.

We obtained miRNA target site datasets from two miRNA target prediction algorithms, namely TargetScan 5.1 and PicTar which were found to have the highest precision and sensitivity out of eight commonly used algorithms [23]. The data for each prediction algorithm was downloaded from the respective source, and the genomic coordinates for each target site were indexed and mapped to RefSeq genes. In addition, each miRNA target site was cross referenced with miRBase [24] and targets referencing dead or non-existent miRNAs were
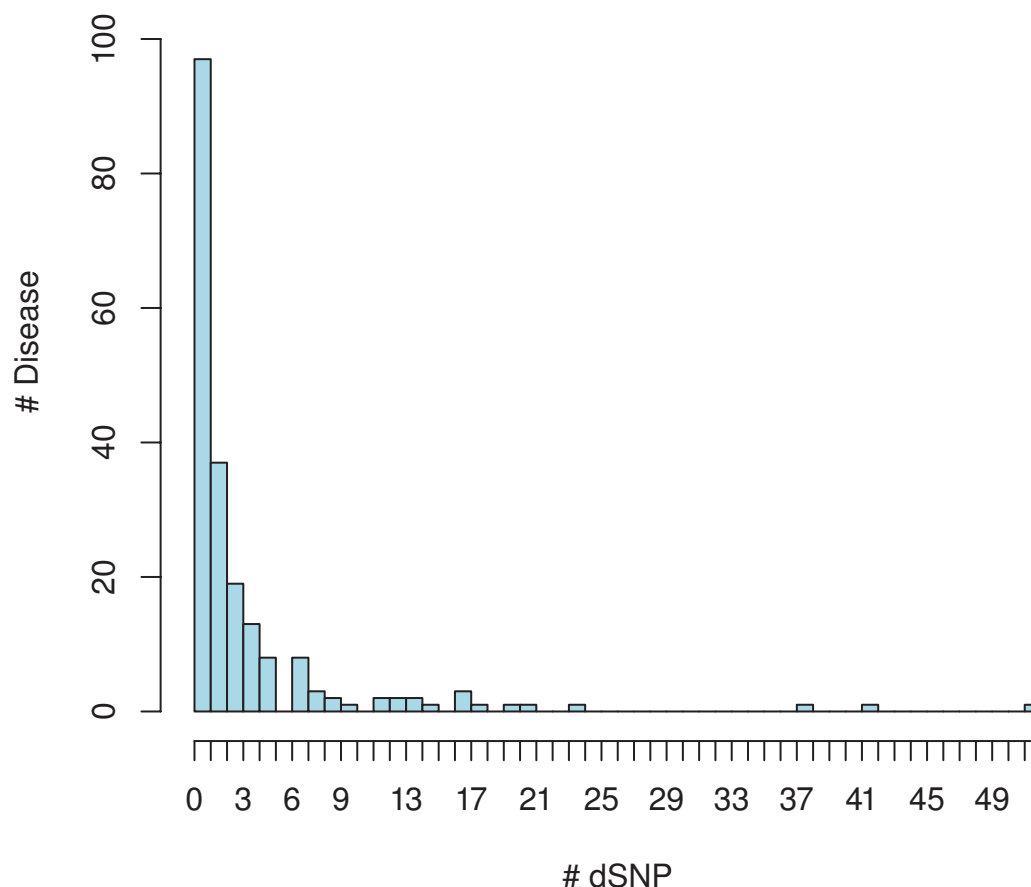
Figure 1 Histogram of the number of dSNPs associated with individual diseases.

excluded. The genomic coordinates from PicTar were converted from hg17 assembly to hg18 assembly using the LiftOver utility from UCSC. The exon index for each miRNA target site was computed and used to calculate the nucleotide distance between SNPs and predicted miRNA target sites. This distance is calculated from the start or end location of the miRNA target "seed" region (~7nt), depending on whether the SNP is upstream or downstream of the miRNA target site. SNPs which fall inside the miRNA target "seed" region have a distance of 0. To address the low prediction specificity of the miRNA target prediction algorithms we incorporate data from four curation databases (TarBase [25], miRTarBase [26], miRecords [27], and miR2disease [28]) which collect experimentally confirmed miRNA target interactions. Genes with experimentally confirmed targeting by miRNAs were annotated in miRdSNP and displayed with a green check icon within the user interface.
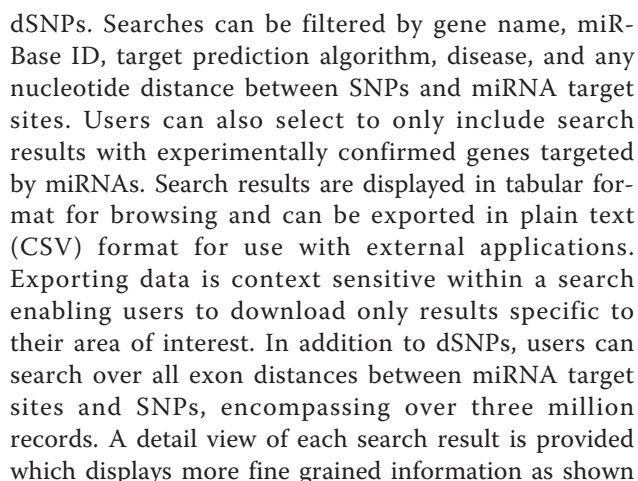
To predict new miRNA target sites created by dSNPs, we first generated candidate sequences using a 6-nt flank up/down stream from each dSNP location replacing the dSNP with the observed allele. Using these new candidate sequences we searched for perfect match 7-mer seed regions from miRBase mature sequence data. We then extracted 25-nt flank upstream of the matching seed region and used the miRNA target prediction program miRanda [29] with default cutoff to further eliminate false positives. We were able to identify 180 newly created miRNA target sites from 138 dSNPs.

Loading and indexing new data in miRdSNP is an automated process allowing for streamlined updates to the database as new RefSeq, miRNA target prediction, and SNP data become available. All data in miRdSNP is available for download in raw text format (CSV, BED) with access to previous versions.

Utility and discussion

miRdSNP provides a publicly accessible web interface for interrogating the database. The advanced search tool as shown in Figure 2A allows the user to perform proximity searches between miRNA targets and

Figure 2 Overview of miRdSNP web interface. (A) The search results in tabular format, (B) detail view for miRNA target and dSNP, and (C) sequence view for the selected gene.

dSNPs. Searches can be filtered by gene name, miRBase ID, target prediction algorithm, disease, and any nucleotide distance between SNPs and miRNA target sites. Users can also select to only include search results with experimentally confirmed genes targeted by miRNAs. Search results are displayed in tabular format for browsing and can be exported in plain text (CSV) format for use with external applications. Exporting data is context sensitive within a search enabling users to download only results specific to their area of interest. In addition to dSNPs, users can search over all exon distances between miRNA target sites and SNPs, encompassing over three million records. A detail view of each search result is provided which displays more fine grained information as shown

in Figure 2B. Here one can view the SNPs in strong LD and the Regional LD plot for the dSNP of interest. Individual search results can also be viewed at the sequence level showing the annotated locations for miRNA target sites, dSNPs, and SNPs (Figure 2C). Each annotated location on the sequence is click-able, showing detailed information such as mature miRNA sequence, UTR index, and links to miRBase, dbSNP, and PubMed. The sequence views display miRNA target site annotations from multiple prediction algorithms allowing the user to dynamically toggle between them for easy comparison. Along with searching, the miRdSNP web interface provides the ability to browse by gene, displaying all miRNA targets, SNPs, and diseases for every RefSeq gene in the database.
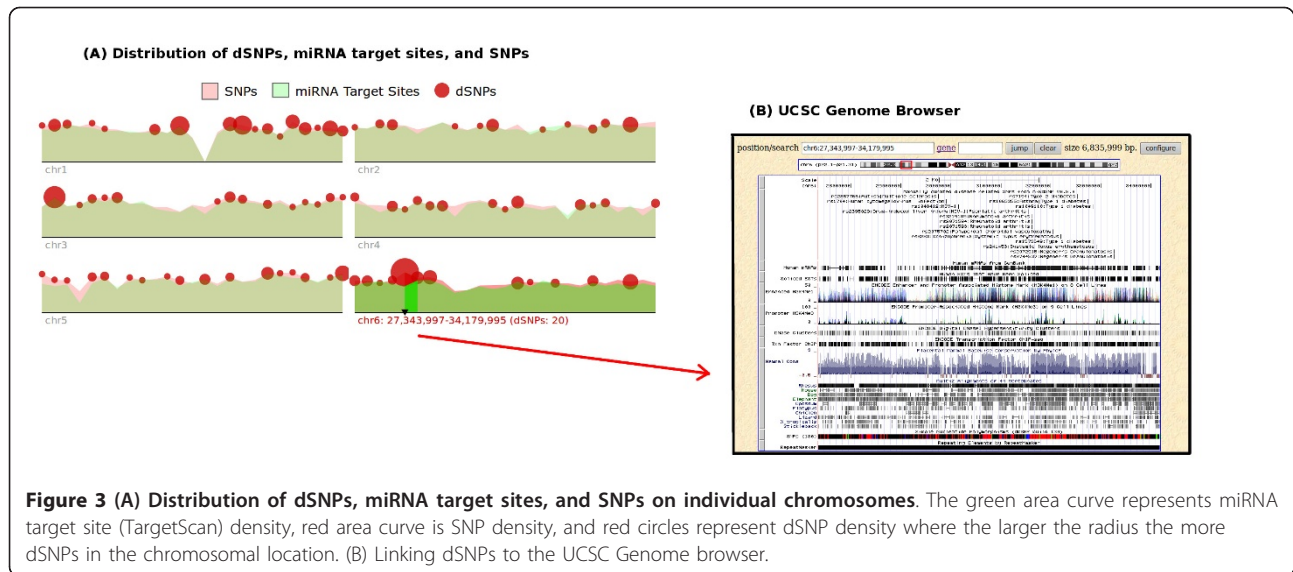
Figure 3 (A) Distribution of dSNPs, miRNA target sites, and SNPs on individual chromosomes. The green area curve represents miRNA target site (TargetScan) density, red area curve is SNP density, and red circles represent dSNP density where the larger the radius the more dSNPs in the chromosomal location. (B) Linking dSNPs to the UCSC Genome browser.

An interactive visualization tool is provided for viewing the chromosomal distribution of dSNPs, miRNA target sites (from TargetScan), and SNPs (Figure 3A). Using this tool a user can view the density relationship between miRNA target sites, dSNPs, and SNPs across chromosomes. dSNPs are shown as red circles, the larger the radius the more dSNPs found at that chromosomal location. miRNA target site and SNP densities are displayed as log-normalized area curves. Hovering over a particular region of the chromosome shows the coordinates along with the number of dSNPs in that region. Integration with the UCSC Genome browser is provided using a custom track BED file. Any chromosomal region containing dSNPs is linked directly to the UCSC Genome browser (Figure 3B), allowing the user to view more detailed information for genes around the dSNP location. The interactive tool requires a SVG (Scalable Vector Graphics) compliant browser but we also provide a non-interactive version for users without SVG support.

Conclusion

miRdSNP is an ongoing effort to create a comprehensive data source for exploring the effect of SNPs on miRNA binding in relation to human diseases. We are working on importing data from other miRNA target prediction algorithms such as DIANA-microT v3.0 [30] and ElMMo [31]. Since the accuracy of the manually curated database of dSNPs is an integral part of miRdSNP, we aim to further broaden the amount of data captured from the manual process. Data such as study design, sample size, and p-values would further enhance the ability to determine the disease-SNP association. We aim to update the dSNP curation database yearly and as

new versions of the miRNA target prediction algorithms become available.

Availability and requirements

miRdSNP is freely available on the web at http://mirdsnp.ccr.buffalo.edu.

Author details
[1]Center for Computational Research, New York State Center of Excellence in Bioinformatics & Life Sciences, State University of New York at Buffalo, Buffalo NY 14260, USA. [2]Department of Pharmacy, Beijing Jishuitan Hospital, China. [3]Department of Pharmaceutical Sciences, State University of New York at Buffalo, Buffalo NY 14260 USA. [4]Department of Ophthalmology, Department of Biostatistics, Department of Medicine, State University of New York at Buffalo, Buffalo NY 14260 USA. [5]SUNY Eye Institute, Buffalo NY 14260 USA.

Authors' contributions
ZH conceived of the project. LL, JLK, YP, AY, and ZH participated in the manual curation of disease-associated SNPs. AEB designed and implemented the database, front end web interface, and implemented all data processing. ZH and AEB drafted the manuscript. All authors read and approved the final manuscript.

References
1. Dickson DW, Baker M, Rademakers R: **Common variant in GRN is a genetic risk factor for hippocampal sclerosis in the elderly.** *Neurodegener Dis* 2010, **7(1-3)**:170-174.
2. Wang G, van der Walt JM, Mayhew G, Li YJ, Züchner S, Scott WK, Martin ER, Vance JM: **Variation in the miRNA-433 binding site of FGF20 confers risk for Parkinson disease by overexpression of alpha-synuclein.** *Am J Hum Genet* 2008, **82(2)**:283-289.

3. Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, Morgan TM, Mathews CA, Pauls DL, Rasin MR, Gunel M, Davis NR, Ercan-Sencicek AG, Guez DH, Spertus JA, Leckman JF, Dure LS, Kurlan R, Singer HS, Gilbert DL, Farhi A, Louvi A, Lifton RP, Sestan N, State MW: **Sequence variants in SLITRK1 are associated with Tourette's syndrome.** *Science* 2005, **310(5746)**:317-320.

4. Tan Z, Randall G, Fan J, Camoretti-Mercado B, Brockman-Schneider R, Pan L, Solway J, Gern JE, Lemanske RF, Nicolae D, Ober C: **Allele-specific targeting of microRNAs to HLA-G and risk of asthma.** *Am J Hum Genet* 2007, **81(4)**:829-834.

5. Martin MM, Buckenberger JA, Jiang J, Malana GE, Nuovo GJ, Chotani M, Feldman DS, Schmittgen TD, Elton TS: **The human angiotensin II type 1 receptor +1166 A/C polymorphism attenuates microrna-155 binding.** *J Biol Chem* 2007, **282(33)**:24262-24269.

6. Rademakers R, Eriksen JL, Baker M, Robinson T, Ahmed Z, Lincoln SJ, Finch N, Rutherford NJ, Crook RJ, Josephs KA, Boeve BF, Knopman DS, Petersen RC, Parisi JE, Caselli RJ, Wszolek ZK, Uitti RJ, Feldman H, Hutton ML, Mackenzie IR, Graff-Radford NR, Dickson DW: **Common variation in the miR-659 binding-site of GRN is a major risk factor for TDP43-positive frontotemporal dementia.** *Hum Mol Genet* 2008, **17(23)**:3631-3642.

7. Schaefer AS, Richter GM, Nothnagel M, Laine ML, Rühling A, Schäfer C, Cordes N, Noack B, Folwaczny M, Glas J, Dörfer C, Dommisch H, Groessner-Schreiber B, Jepsen S, Loos BG, Schreiber S: **A 3' UTR transition within DEFB1 is associated with chronic and aggressive periodontitis.** *Genes Immun* 2010, **11**:45-54.

8. Nicoloso MS, Sun H, Spizzo R, Kim H, Wickramasinghe P, Shimizu M, Wojcik SE, Ferdin J, Kunej T, Xiao L, Manoukian S, Secreto G, Ravagnani F, Wang X, Radice P, Croce CM, Davuluri RV, Calin GA: **Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility.** *Cancer Res* 2010, **70(7)**:2789-2798.

9. Brendle A, Lei H, Brandt A, Johansson R, Enquist K, Henriksson R, Hemminki K, Lenner P, Försti A: **Polymorphisms in predicted microRNA-binding sites in integrin genes and breast cancer: ITGB4 as prognostic marker.** *Carcinogenesis* 2008, **29(7)**:1394-1399.

10. Chin LJ, Ratner E, Leng S, Zhai R, Nallur S, Babar I, Muller RU, Straka E, Su L, Burki EA, Crowell RE, Patel R, Kulkarni T, Homer R, Zelterman D, Kidd KK, Zhu Y, Christiani DC, Belinsky SA, Slack FJ, Weidhaas JB: **A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk.** *Cancer Res* 2008, **68(20)**:8535-8540.

11. He H, Jazdzewski K, Li W, Liyanarachchi S, Nagy R, Volinia S, Calin GA, Liu CG, Franssila K, Suster S, Kloos RT, Croce CM, de la Chapelle A: **The role of microRNA genes in papillary thyroid carcinoma.** *Proc Natl Acad Sci USA* 2005, **102(52)**:19075-19080.

12. Saetrom P, Biesinger J, Li SM, Smith D, Thomas LF, Majzoub K, Rivas GE, Alluin J, Rossi JJ, Krontiris TG, Weitzel J, Daly MB, Benson AB, Kirkwood JM, O'Dwyer PJ, Sutphen R, Stewart JA, Johnson D, Larson GP: **A risk variant in an miR-125b binding site in BMPR1B is associated with breast cancer pathogenesis.** *Cancer Res* 2009, **69(18)**:7459-7465.

13. Mishra PJ, Humeniuk R, Mishra PJ, Longo-Sorbello GSA, Banerjee D, Bertino JR: **A miR-24 microRNA binding-site polymorphism in dihydrofolate reductase gene leads to methotrexate resistance.** *Proc Natl Acad Sci USA* 2007, **104(33)**:13513-13518.

14. Hu Z, Bruno AE: **The Influence of 3'UTRs on microRNA Function Inferred from Human SNP Data.** *Comparative and Functional Genomics* 2011 [http://dx.doi.org/10.1155/2011/910769], 2011.

15. Bao L, Zhou M, Wu L, Lu L, Goldowitz D, Williams RW, Cui Y: **PolymiRTS Database: linking polymorphisms in microRNA target sites with complex traits.** *Nucleic Acids Res* 2007, **35(Database)**:D51-D54.

16. Hiard S, Charlier C, Coppieters W, Georges M, Baurain D: **Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates.** *Nucleic Acids Res* 2010, **38(Database)**:D640-D651.

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**:15-20.

18. Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N: **Combinatorial microRNA target predictions.** *Nat Genet* 2005, **37(5)**:495-500.

19. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: **The human genome browser at UCSC.** *Genome Res* 2002, **12(6)**:996-1006.

20. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K: **dbSNP: the NCBI database of genetic variation.** *Nucleic Acids Res* 2001, **29**:308-311.

21. Consortium IH, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MMY, Tsui SKW, Xue H, Wong JTF, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallée C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PKH, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PIW, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CDM, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J: **A second generation human haplotype map of over 3.1 million SNPs.** *Nature* 2007, **449(7164)**:851-861.

22. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PIW: **SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap.** *Bioinformatics* 2008, **24(24)**:2938-2939.

23. Alexiou P, Maragkakis M, Papadopoulos GL, Reczko M, Hatzigeorgiou AG: **Lost in translation: an assessment and perspective for computational microRNA target identification.** *Bioinformatics* 2009, **25(23)**:3049-3055.

24. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ: **miRBase: microRNA sequences, targets and gene nomenclature.** *Nucleic Acids Res* 2006, **34(Database)**:D140-D144.

25. Sethupathy P, Corda B, Hatzigeorgiou AG: **TarBase: A comprehensive database of experimentally supported animal microRNA targets.** *RNA* 2006, **12(2)**:192-197.

26. Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, Chan WL, Tsai WT, Chen GZ, Lee CJ, Chiu CM, Chien CH, Wu MC, Huang CY, Tsou AP, Huang HD: **miRTarBase: a database curates experimentally validated microRNA-target interactions.** *Nucleic Acids Res* 2011, **39(Database)**:D163-D169.

27. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T: **miRecords: an integrated resource for microRNA-target interactions.** *Nucleic Acids Res* 2009, **37(Database)**:D105-D110.

28. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, Li M, Wang G, Liu Y: **miR2Disease: a manually curated database for microRNA deregulation in human disease.** *Nucleic Acids Res* 2009, **37(Database)**:D98-104.

29. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS: **MicroRNA targets in Drosophila.** *Genome Biol* 2003, **5**:R1.

30. Maragkakis M, Alexiou P, Papadopoulos GL, Reczko M, Dalamagas T, Giannopoulos G, Goumas G, Koukis E, Kourtis K, Simossis VA, Sethupathy P, Vergoulis T, Koziris N, Sellis T, Tsanakas P, Hatzigeorgiou AG: **Accurate microRNA target prediction correlates with protein repression levels.** *BMC Bioinformatics* 2009, **10**:295.

31. Gaidatzis D, van Nimwegen E, Hausser J, Zavolan M: **Inference of miRNA targets using evolutionary conservation and pathway analysis.** *BMC Bioinformatics* 2007, **8**:69.