RBP-Var: a database of functional variants involved in regulation mediated by RNA-binding proteins

Fengbiao Mao¹,², Luoyuan Xiao³, Xianfeng Li⁴, Jialong Liang¹,², Huajing Teng¹,⁵, Wanshi Cai¹,⁻ and Zhong Sheng Sun¹,⁶,⁺

¹Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China, ²University of Chinese Academy of Sciences, Beijing 100049, China, ³Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China, ⁴State Key Laboratory of Medical Genetics, Central South University, Changsha, Hunan, 410078, China, ⁵Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China and ⁶Institute of Genomic Medicine, Wenzhou Medical University, Wenzhou 325035, China

Received August 13, 2015; Revised November 01, 2015; Accepted November 09, 2015

ABSTRACT

Transcription factors bind to the genome by forming specific contacts with the primary DNA sequence; however, RNA-binding proteins (RBPs) have greater scope to achieve binding specificity through the RNA secondary structure. It has been revealed that single nucleotide variants (SNVs) that alter RNA structure, also known as RiboSNitches, exhibit 3-fold greater local structure changes than replicates of the same DNA sequence, demonstrated by the fact that depletion of RiboSNitches could result in the alteration of specific RNA shapes at thousands of sites, including 3’ UTRs, binding sites of microRNAs and RBPs. However, the network between SNVs and post-transcriptional regulation remains unclear. Here, we developed RBP-Var, a database freely available at http://www.rbp-var.biols.ac.cn/, which provides annotation of functional variants involved in post-transcriptional interaction and regulation. RBP-Var provides an easy-to-use web interface that allows users to rapidly find whether SNVs of interest can transform the secondary structure of RNA and identify RBPs whose binding may be subsequently disrupted. RBP-Var integrates DNA and RNA biology to understand how various genetic variants and post-transcriptional mechanisms cooperate to orchestrate gene expression. In summary, RBP-Var is useful in selecting candidate SNVs for further functional studies and exploring causal SNVs underlying human diseases.

INTRODUCTION

A single nucleotide variant (SNV) can result in complex diseases through various molecular mechanisms, including amino-acid changes at the protein level, alteration of regulatory elements at the transcription level and disruption of RNA–protein interaction at the post-transcription level (1–4). Previous studies mainly concentrated on the effect of SNVs on protein structure (5,6) and the misregulation of transcriptional regulation (7,8), known as cSNVs and rSNVs, respectively. However some SNVs, which are located in non-coding regions and have no relationship with transcriptional regulation (8), can have an observable phenotype or result in a disease. Therefore, some SNVs may play crucial roles in post-transcriptional gene regulation by altering micro (mi)RNA targets, RNA secondary structure and RNA–protein interactions (1,4,9–12).

Intensive attempts have been undertaken to better understand RNA-binding proteins (RBPs) and RNA–protein interactions (13). The advent of next-generation sequencing and relevant molecular techniques have allowed the development of systematic experimental protocols for the identification of protein-binding sites on RNA using the immunoprecipitation of RBPs, with or without in vivo RNA–protein crosslinking, known as crosslinking immunoprecipitation sequencing (CLIP-seq) and RNA immunoprecipitation sequencing (RIP-seq), respectively (14,15). CLIP-seq that combines high-throughput sequencing (HITS)-CLIP (16), photoactivatable ribonucleoside-enhanced (PAR)-CLIP (17) and individual nucleotide resolution crosslinking immunoprecipitation (iCLIP) (18), is a powerful tool for investigating the in vivo binding sites of RBPs on a transcriptome-wide scale, providing higher resolution than previous technologies, such as RIP-seq, for identifying protein-binding sites on RNAs. Identifying changes in RBP binding induced by SNVs for a messenger RNA (mRNA) in vivo remains chal...
lenging because RBPs have much more scope to achieve specificity through the secondary structure of RNA and SNVs could change intramolecular bonds which may twist RNA molecules into hairpins, stem-loops and various other bumps and bulges (13). The impact of SNVs on RNA structure and gene regulation has been widely studied (1–4). Wan et al. (19) studied the initial landscape and variation of RNA secondary structures in a human family trio (mother, father, offspring) and revealed that \( \approx 15\% \) of all transcribed SNVs are RiboSNitches that alter the local RNA structure. These RiboSNitches were also significantly depleted around predicted miRNA target sites and protein-binding sites on RNA. Moreover, Maticzka et al. (20) demonstrated the importance of RNA secondary structure for conferring binding specificity in a subset of RBPs using both RNA structure and sequence features to develop a machine learning-based approach for predicting protein-binding sites on RNA from CLIP data. Fukunaga et al. (21) developed the CapR software that determines the probability of secondary structure throughout an RNA molecule for RBP binding and found that the human RBP Pumilio-2 has a preference for hairpin loop structures. Overall, RNA secondary structures exert a significant effect on RBP binding, and the RNA interactome network can be disrupted once the RNA structures are transformed.

Recently, great progress has been made in identifying mutations in RBPs associated with disease risk (22) such as mutations in the RBPs FUS and TDP-43 that can cause amyotrophic lateral sclerosis (23). However, it remains unclear what changes occur to RNA substrates in diseases and their effect on RNA structure and RNA–protein interaction induced by SNVs or RNA editing sites (24). Therefore, a comprehensive database is required for the annotation of functional variants involved in post-transcriptional interactions. Here, we developed RBP-Var (Figure 1), an integrated database for the annotation of functional variants involved in RBP-mediated post-transcriptional regulation and interaction. We also determined trait/disease-associations and deciphered gene regulation for SNVs located in potential protein-binding sites on RNA. Furthermore, we developed a heuristic scoring system based on the functional confidence of a variant, including SNV or RNA edit, to assist comparison among annotations.

**DATA COLLECTION AND PROCESSING**

**Data sources for RBP-Var**

RBP-Var provides extensive annotation for the functional variants involved in post-transcriptional interactions. We collected all the available processed data on RBPs from starBase (25) and CLIPdb (26) and extracted sequencing data from Gene Expression Omnibus (GEO) (27). Genomic coordinates of RBPs derived from GEO were converted from human genome assembly hg18 into hg19 using the University of California, Santa Cruz (UCSC) LiftOver tool (http://genome.ucsc.edu). We also retrieved a collection of expression quantitative trait loci (eQTL) from previous studies (28–37), human trait-associated SNVs from the genome-wide association study (GWAS) catalog (38), somatic mutations in cancer from COSMIC (39) and disease-associated mutations from ClinVar (40). In addition, RBP-Var includes collection of RBP binding SNVs (rbSNVs) which may disrupt the binding of RBPs, derived from 112 CLIP-seq data sets (60 RBPs) and 319 motif matching sites (153 RBPs). These functional RBP binding SNVs include deletions, insertions, single nucleotide polymorphism (SNPs), multiple nucleotide polymorphisms (MNs), microsatellites, insertions/deletions from dbSNP database (version 142) as well as RNA editing (A-to-I editing sites) from RADAR and DARNED (Table 1).

**Position weight matrix (PWM) motif matching**

Many RBPs interact with mRNAs via a limited set of modular RNA-binding domains, including the RNA recognition motif, heterogeneous nuclear ribonucleoprotein K-homology domain and zinc fingers (11,41). The RNA-binding domains of RBPs initially determine the specificity and preferences of RNA binding with specific sequence motifs (42). Therefore, all PWMs from the catalog of inferred sequence binding preferences of RNA binding proteins (CIS-BP-RNA) (42) and the RNA-binding protein database (RBPD) (43) deposited in the AURA database (44) were used to call motif matches in the transcriptome. The possible k-mers were aligned to the human transcriptome using MAST in the MEME suite (45) to give a final motif mapping with a match score >0 and \( P \)-values < 0.0001.

**Impacts of SNVs on RNA structures**

Based on human reference transcripts retrieved from UCSC (hg19), we changed the corresponding allele in a given sequence of the reference transcript to an alternative allele as an altered transcript for each SNV. Then, the RNAfold program was employed (46) with default parameters to illustrate the secondary structure and calculate the minimal free energy (MFE, \( \Delta G \)) for both transcript sequence of wild-type and mutant. Energy changes in RNA structures (\( \Delta \Delta G = \))
Impact of SNPs on miRNA–RNA interactions

Single nucleotide polymorphisms (SNPs) in the target sites of miRNAs may destroy or create miRNA-binding sites on RNAs, which lead to loss-of-function and/or gain-of-function miRNA–RNA interactions. To assess the impact of each SNP on miRNA–RNA interaction, TargetScan (47) and miRanda (48) were used to predict target sites around each SNP with ±25 bp sequences for Ref-transcripts and Alt-transcripts, separately. In addition, we compared miRNA target score versus target score induced by SNVs by 201 SNPs in miRNA seed regions in total 556 008 pairs miRNA–RNA interactions. We analyzed target alteration of miRNA by SNPs in 3′-UTR of mRNA and miRNA target alteration by SNPs in miRNA seed regions. We found that 888 674 pairs miRNA–RNA interactions was altered significantly (P-value < 0.01) by 253 428 SNPs in 3′-UTR of mRNA in a total of 1 654 644 pairs miRNA–RNA interactions. Meanwhile, we found that 302 723 pairs miRNA–RNA interactions was altered significantly (P-value < 0.01) by 201 SNPs in miRNA seed regions in total 556 008 pairs miRNA–RNA interactions.

Classifying functional variants

To better interpret and comprise the catalog of rbSNV information, we developed a heuristic scoring system based on the functional confidence of variants. The scoring system represents with increasing confidence that a variant lies in a functional location and probably results in a functional consequence (i.e. alteration of RBP binding and a gene regulatory effect) (Table 2). For example, we consider variants that are known eQTLs as significant and label them as Category

Table 1. Database content of RBP-Var as of 10 October, 2015

| Data type                  | Sources                                                                 | Description                                                                 | Records                      | Genomic coverage (bp) |
|----------------------------|-------------------------------------------------------------------------|----------------------------------------------------------------------------|------------------------------|------------------------|
| CLIP-seq                   | starBase, CLIPdb, GEO                                                   | 112 data sets for 60 RBPs                                                 | 52,528,288                   | 847,717,483            |
| Predicted binding (PWMs)   | CISBP-RNA, RBPDB                                                       | 319 PWMs for 153 RBPs                                                     | 113,777,778                 | 822,156,225            |
| dbSNP                      | dbSNP v142                                                             | including SNP, InDel and MNP                                              | 113,128,211                 | 113,128,211            |
| RNA editing                | Radar, DARNED                                                          | A->G editing                                                               | 3,361,915                    | 26,958,420             |
| miRNA targets              | TargetScan, miRanda, miRNASENP                                         | miRNA-SNP pairs                                                           | 2,576,460                    | 2,576,460              |
| eQTLs                      | MuTher, SCAN, seeQTL, GTEx, Harvard                                     | SNP-gene pairs                                                            | 4,546,890                    | 4,546,890              |
| dsQTLs                     | dsQTL Browser                                                          | SNP-gene pairs                                                            | 214,522                      | 214,522                |
| LD proxies                 | HapMap, 1000Genome                                                     | r² > 0.8 by Plink                                                          | 4,600,670                    | 4,600,670              |
| 6mA                        | Mety-DB                                                                 | 6mA peaks called by MACS including 7 splicing patterns                    | 46,037                       | 127,588,574           |
| Splicing sites             | H-DBAS                                                                  | SNP-trait or gene-trait pairs                                             | 1,669,641                    | 158,324,909           |
| trait/disease associations | GWAS, COSMIC, ClinVar, TCGA, DISEASES                                  | SNP-trait or gene-trait pairs                                             | 8,955,125                    | 8,955,125             |
Moreover, we applied the JBrowse Genome Browser (http://jbrowse.org) to establish the ‘Browser’ page to visualize genome-wide signals of RBP binding for each CLIP-seq dataset. Users could browse all binding signals across specific genomic regions of their interests. The searching result notations ranging from the most informational variants (1a, variant may change the motif for RBP binding) to the least informational variants (1e, variant only has a motif for RBP binding).

### WEB INTERFACE

Data retrieving in RBP-Var could be implemented based on SNP-centric or gene-centric. SNP-centric data retrieving was appropriate to analyze results from genetic studies, especially the results of high-throughput studies, and then provided evidence for further functional studies to identify causal SNPs and shed light on underlying molecular mechanisms. Gene-centric searches were useful in candidate SNP selections that are based on genes of interest. In addition, RBP-Var provided RNA-editing retrieving, which could decipher the potential function of RNA editing in involved in post-regulation and gene expression (24). Furthermore, we applied the JBrowse Genome Browser (http://jbrowse.org) to establish the ‘Browser’ page to visualize genome-wide signals of RBP binding for each CLIP-seq data set. Users could browse all binding signals across specific genomic region of their interests. The searching result for SNP rs1802295 in RBP-Var database was used as an example of web-interface (Figure 2).

We found that 874 160 (0.77%) SNVs in the SNP database were overlapped with RNA binding sites as identified by CLIP-seq. These SNVs included 835 130 SNPs, 26 854 deletions, 11 323 insertions, 166 insertions/deletions, 679 multiple nucleotide polymorphisms (SNPs) and 8 microsatellites. Subsequently, we employed our heuristic scoring system to annotate these SNVs, and we found 828 210 SNVs classified into Category 1/2 (Table 2), which were composed of 802 765 SNPs, 23 570 deletions, 1052 insertions, 152 insertions/deletions, 663 MNPs and 8 microsatellites (Figure 3A).

### APPLICATIONS

To verify the heuristic system, we compared the functional rbSNVs of RBP-Var with regulatory SNVs from RegulomeDB (8) and deleterious SNVs predicted by SIFT (5) and PolyPhen2 (6). A total of 1 712 307 SNVs were identified as deleterious or functional SNVs by at least one of the four classification methods. Among 828 210 functional rbSNPs identified by RBP-Var, 40 904 were classified into functional SNVs by RegulomeDB, 166 647 were presented and predicted as deleterious by PolyPhen2 and 109 364 by SIFT (Figure 3B), whereas, the rest 651 784 rbSNVs were identified uniquely in RBP-Var. Then, we compared the rbSNVs uniquely identified in RBP-Var with GWAS-lead SNVs to determine causal disease variants. As a result, we identified 93 unique rbSNVs that were associated with human diseases (Supplemental Table S1). Similarly, we identified 7108 and 80 928 unique rbSNVs with RBP-Var score 1/2 in ClinVar and COSMIC databases, respectively (Supplemental Tables S2 and S3). These observations indicated that RBP-Var could identify more rbSNVs which could be implicated in human disease and biological functions even though they did not affect protein coding.

As one of the non-coding regulatory regions, untranslated regions (UTRs) are known to have many sequence and structural motifs that can regulate translational and transcriptional efficiency and the stability of mRNAs through interactions with RBPs and other non-coding RNAs like microRNAs (miRNAs) (4). Here, we found that 38 GWAS-lead SNPs located in the UTRs were classified with RBP-Var score 1/2 (Supplemental Table S1). For example, SNP rs1802295 (C/T) which is associated with type 2 diabetes in individuals of South Asian ancestry (53), but its underlying mechanism participated in type 2 diabetes is still unknown (54). Our result revealed SNP rs1802295 is located in the 3’-UTR of VPS26A, a region containing motif of binding by protein PABPC1 and PABPC3 based on PWM matching. This motif is likely disrupted by this SNP as the core component C in the motif would be changed to U (Figure 3A). Meanwhile, the experimental binding intensity around this SNP on mRNA of VPS26A includes binding by protein AGO2, PTBP1 from HeLa cell line, and IGF2BP3, WDR33, ELAVL1 from HEK293 cell line (Figure 4A). All

| Category Description | # of SNV | # of RNA editing |
|----------------------|----------|-----------------|
| Likely to affect RBP binding, RNA secondary structure and expression |          |                |
| 1a eQTL+RBP binding+matched RBP motif+riboSNitch+miRNA targets | 1        | 0               |
| 1b eQTL+RBP binding+any RBP motif +riboSNitch+miRNA targets | 1,059    | 0               |
| 1c eQTL+RBP binding+any RBP motif +riboSNitch | 2,417    | 1               |
| 1d eQTL+RBP binding+any RBP motif | 334      | 0               |
| 1e eQTL+RBP binding | 916      | 0               |
| Likely to affect RBP binding, RNA secondary structure |          |                |
| 2a RBP binding+matched RBP motif +riboSNitch+miRNA targets | 23       | 1               |
| 2b RBP binding+any RBP motif +riboSNitch+miRNA targets | 29,415   | 483             |
| 2c RBP binding+any RBP motif +riboSNitch | 177,806  | 5,758           |
| 2d RBP binding+riboSNitch | 616,239  | 46,963          |
| Less likely to affect RBP binding |        |                |
| 3a RBP binding+any RBP motif+miRNA targets | 15       | 49              |
| 3b RBP binding+matched RBP motif | 28,599   | 0               |
| Minimal possibility to affect RBP binding |          |                |
| 4 RBP binding+any RBP motif | 844      | 534             |
| 5 RBP binding+miRNA targets | 85       | 209             |
| 6 RBP binding or RBP motif hit | 29,415   | 14              |
Figure 2. Web-interface of RBP-Var. (A–H) The snapshot of searching result for SNP rs1802295 in RBP-Var database.
Figure 3. The features of RBP-Var. (A) The component of rbSNVs derived from dbSNP. (B) Comparison of SNVs with category score 1/2 deposited in RBP-Var and RegulomeDB with damaging SNVs predicted by SIFT and PolyPhen2. (C) Comparison of RNA editing events with category score 1/2 deposited in RBP-Var with damaging RNA editing events predicted by SIFT and PolyPhen2.
Figure 4. Application of RBP-Var for an SNV related to type-2 diabetes. (A) The RBP binding signal for SNP rs1802295, which is a functional rbSNV associated with type 2 diabetes. The motif of binding by protein PABPC1L and PABPC3 are showed on the top of rs1802295. And the binding intensity around this SNP on mRNA of VPS26A by protein AGO2, PTBP1 from HeLa cell line, and IGF2BP3, WDR33, ELAVL1 from HEK293 cell line is illustrated in different tracks. The targeting site of miRNA hsa-miR-510 is marked by yellow vertical line. (B) The optimal secondary structure of RNA sequence from 200 bp of flanking regions on either side of SNP rs1802295 for wild-type of VPS26A and mutant of VPS26A induced by rs1802295. The green and red nucleotides represent the local sequence in interval of 50 bp for wild-type and mutant of VPS26A, respectively.
DISCUSSION

In this study, we included extensive information on annotated and computed post-regulatory elements in the human genome, and we combined a large array of data sources into a single and integrated database to quickly generate prioritized hypotheses for the function of variants affecting both coding and non-coding regions in a genome. This novel database has a convenient interface that allows the easy query submission and provides an instant classification of significant variants. The SNV summary page allows users to quickly form a hypothesis regarding the true RBP-mediated post-transcriptional consequence of a variant.

Guo et al. (61) constructed the rSNPbase database, which provides a similar annotation of SNVs integrated with RBP-mediated post-transcriptional regulation, using data sets of four RBPs derived from RIP-chip or RIP-seq analysis. Our database provides additional information by prioritizing SNVs within general regulatory regions based on specific RBPs, RNA secondary structure and PWM information.

Our simple heuristic scoring system may be improved over time, as more functional rbSNVs will be validated. Our analysis method and database are centered on a likely disruption of a protein-RNA interaction via alteration of RNA secondary structure and regulation of gene expression by rbSNVs if they are eQTLs. It is likely that there are additional sources of data that reinforce each other in a different manner that need to be explored; however, the scoring system provides significant enrichment along with better category scores, as shown when compared to GWAS-lead, ClinVar and COSMIC rbSNVs. Hence, our database and scoring system provide the best current method for annotating and prioritizing variants involved in RBP-mediated post-transcriptional regulation. Overall, we comprehensively integrated eQTLs, protein-binding sites on RNA, RNA secondary structure, miRNA, and SNV information along with GWAS, COSMIC, ClinVar, 6mA, alternative splicing, and mRNA expression profiles to seek potential post-functional SNVs.

In conclusion, this database provides a user-friendly web interface, allowing users to rapidly find whether SNVs of interest can transform the secondary structure of RNA and identify RBPs whose binding may be subsequently disrupted. Moreover, RBP-Var can assess the impact of each SNV on miRNA–RNA interaction as SNVs may destroy or create miRNA-binding sites, which result in loss-of-function and/or gain-of-function miRNA–RNA interactions. RBP-Var is a useful resource for benchmarking the mutations or RNA-editing events that cause disease by changing post-transcriptional interaction and regulation. In the future, RBP-Var will be updated frequently and extended to other species. We are dedicated to maintaining and improving RBP-Var, since it is a useful resource for the research community.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.
ACKNOWLEDGEMENTS

The authors would like to thank Jinyu Wu and Jinchen Li for helping with website debugging; Baozhen Du and Sen Guo for helping with data collection; Qi Liu and Huigian Chen for all the valuable suggestions on our manuscript.

FUNDING

Funding for open access charge: National High Technology Research and Development Program of China [2012AA02A202]; National Natural Science Foundation of China [31171236/C060503]; Innovation Center of China, AstraZeneca.

Conflict of interest statement. None declared.

REFERENCES

1. Halvorsen,M., Martin,J.S., Broadaway,S. and Laederach,A. (2010) Disease-associated mutations that alter the RNA structural ensemble. PLoS Genet., 6, e1001074.
2. Sabarinathan,R., Tafer,H., Seemann,S.E., Hofacker,I.L., Studler,P.F. and Gorodkin,J. (2013) The RNAasp web server: predicting SNP effects on local RNA secondary structure. Nucleic Acids Res., 41, W475–W479.
3. Corley,M., Solem,A., Qu,K., Chang,H.Y. and Laederach,A. (2015) Detecting riboStitches with RNA folding algorithms: a genome-wide benchmark. Nucleic Acids Res., 43, 1859–1868.
4. Sabarinathan,R., Wenzel,A., Novotny,P., Tang,X.J., Kalari,K.R. and Gorodkin,J. (2014) Transcriptome-wide analysis of UTRs in Non-small cell lung cancer reveals cancer-related genes with SNV-induced changes on RNA secondary structure and miRNA target sites. PLoS One, 9, e82699.
5. Ng,P.C. and Henikoff,S. (2003) SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res., 31, 3812–3814.
6. Adzhubei,I.A., Schmidt,S., Peshkin,L., Ramensky,V.E., Gerasimova,A., Bork,P., Kondrashov,A.S. and Sunyaev,S.R. (2010) A method and server for predicting damaging missense mutations. Nat. Methods, 7, 248–249.
7. Lee,T.I. and Young,R.A. (2013) Transcriptional regulation and its genome-wide benchmark. Nat.Methods, 12, 1518–1522.
8. Li,J.H., Liu,S., Zhou,H., Qu,L.H. and Yang,J.H. (2014) starBase v2.0: decoding miRNA-cRNA, miRNA-nRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res., 42, D92–D97.
9. Yang,Y.C., Di,C., Hu,B., Zhou,M., Liu,Y., Song,N., Li,Y., Umetu,J. and Lu,Z.J. (2015) CLIPdb: a CLIP-seq database for protein-RNA interactions. BMC Genomics, 16, 51.58.
10. Mitchell,S.F. and Parker,R. (2014) Principles and Properties of RNA-binding proteins in Mendelian disease. Trends Genet., 31, 10–16.
11. Lunde,B.M., Moore,C. and Varani,G. (2007) RNA-binding proteins: new genomic technologies and perspectives. Nat.Rev.Genet., 8, 533–543.
12. Ye,J. and Blelloch,R. (2014) Regulation of Pluripotency by RNA processing.
13. Clark,T.A., Schweitzer,A.C., Blume,J.E., Wang,X.N.
14. Konig,J., Zarnack,K., Rot,G., Turk,T., Kayikci,M., Zupan,B., Turner,D.J., Luscombe,N.M. and Ule,J. (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat. Struct. Mol. Biol., 17, U909–U166.
15. Munschauer,M. and Charisse,J. (2012) lncRNA: Long non-coding RNA in human disease. Bioessays, 259–262.
16. Hasz,R., Walters,G., Garcia,F., Young,N. et al. (2013) The Genotype-Tissue Expression (GTEx) project. Nat. Genet., 45, 580–585.
17. Li,J.B. and Church,G.M. (2013) Deciphering the functions and regulation of brain-enriched a-to-I RNA editing. Nat. Neurosci., 16, 1518–1522.
18. Liu,S., Zhou,H., Qu,L.H. and Yang,J.H. (2014) starBase v2.0: decoding miRNA-cRNA, miRNA-nRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res., 42, D92–D97.
19. Yang,Y.C., Di,C., Hu,B., Zhou,M., Liu,Y., Song,N., Li,Y., Umetu,J. and Lu,Z.J. (2015) CLIPdb: a CLIP-seq database for protein-RNA interactions. BMC Genomics, 16, 51.58.
20. Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashovsky,M., Marshall,K.A., Phillips,K.H., Sherman,P.M., Holko,M. et al. (2013) NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res., 41, D991–D995.
21. Mitchell,S.F. and Parker,R. (2014) Principles and Properties of Eukaryotic miRNAs. Mol. Cell, 54, 547–558.
22. Montgomery,S.B., Sammeth,M., Gutierrez-Arcelus,M., Lach,R.P., Inglé,C., Nisbett,J., Guigo,R. and Dermitzakis,E.T. (2010) A survey of genetic human cortical gene expression. Nat. Genet., 42, 1494–1499.
23. Capriotti,E.T. (2008) Population genomics of human gene expression. FEBS J., 275, 9–9.
24. Veyrieras,J.B., Kasarskis,A., Zhang,B., Wang,S., Suver,C. et al. (2010) Transcriptome-wide identification of target sites. PLoS One, 5, e1001214.
25. Pickrell,J.K., Marioni,J.C., Pai,A.A., Degner,J.F., Engelhardt,B.E., Kistner,E.O., Nicolae,D.L., Dolan,M.E. and Cox,N.J. (2010) SCAN: a searchable database for human eQTLs. Bioinformatics, 26, 451–452.
26. Gaamazon,E.R., Zhang,W., Konkashbaev,A., Duan,S.W., Kistner,E.O., Nicolaie,D.L., Dolan,M.E. and Cox,N.J. (2010) SCAN: SNP and copy number annotation. Bioinformatics, 26, 259–262.
27. Montgomery,S.B., Sammeth,M., Gutierrez-Arcelus,M., Lach,R.P., Ingles,C., Nisbett,J., Guigo,R. and Dermitzakis,E.T. (2010) Transcriptome genomics using second generation sequencing in a Caucasian population. Nature, 464, U773–U151.
28. Zeller,T., Wild,P., Szymczak,S., Retival,M., Schillert,A., Castagne,R., Maouche,S., Germain,M., Lackner,K., Rossmann,H. et al. (2010) Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. PLoS One, 5, e10093.
29. Xia,K., Shabalin,A.A., Huang,S.P., Madar,Y., Zhou,Y.H., Wang,W., Zou,F., Sun,W., Sullivan,F.F. and Wright,F.A. (2012) seeATL: a searchable database for human eQTLs. Bioinformatics, 28, 451–452.
30. Lonsdale,J., Thomas,J., Salvatore,M., Phillips,R., Lo,E., Shad,S., Hasz,R., Walters,G., Garcia,F., Young,N. et al. (2013) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res., 42, D1001–D1006.
39. Forbes.S.A., Beare.D., Gunasekaran.P., Leung.K., Bindal.N., Boutsellakis.H., Ding.M., Bamford.S., Cole.C., Ward.S. et al. (2015) COSMIC: exploring the world’s knowledge of somatic mutations in human cancer. *Nucleic Acids Res.*, 43, D805–D811.

40. Landrum,M.J., Lee,J.M., Riley,G.R., Jang.W., Rubinstein,W.S., Church,D.M. and Maglott,D.R. (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.*, 42, D980–D985.

41. Castello.A., Fischer.B., Eichelbaum.K., Beckmann.B.M., Strein.C., Davey,N.E., Humphreys,D.T., Preiss.T., Steinmetz,L.M. et al. (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, 149, 1393–1406.

42. Ray.D., Kazan.H., Cook.K.B., Weirauch,M.T., Najafabadi,H.S., Murakawa.Y., Schueler.M., Youngs.N., Penfold-Brown.D., Drew.K., Milek.M. et al. (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell*, 46, 674–690.

43. Cook.K.B., Kazan.H., Zuberi.K., Morris.Q. and Hughes,T.R. (2011) RBPDDB: a database of RNA-binding specificities. *Nucleic Acids Res.*, 39, D301–D308.

44. Dassi.E., Malossini.A., Re.A., Mazza.T., Tebaldi.T., Caputi,L. and Friedman,R.C., Farh,K.K.H., Burge,C.B. and Bartel,D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Nature*, 467, W202–W208.

45. Bailey.T.L., Boden.M., Buske.F.A., Frith.M., Grant.C.E., Clementi,L., Ren.J., Li,W.W. and Noble,W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, 37, W202–W208.

46. Gruber,A.R., Lorenz.R., Bernhart,S.H., Neuboock.R. and Hofacker,I.I. (2008) The Vienna RNA WebSuite. *Nucleic Acids Res.*, 36, W70–W74.

47. Friedman.R.C., Farh.K.K.H., Burge.C.B. and Bartel,D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, 19, 92–105.

48. Betel,D., Wilson.M., Gabow.A., Marks,D.S. and Sander.C. (2008) The microRNA resource: targets and expression. *Nucleic Acids Res.*, 36, D149–D153.

49. Altshuler.D.M., Gibbs.R.A., Peltonen.L., Dermitzakis.E., Schaffner.S.F., Yu.F.L., Bonnen.P.E., de Bakker,P.I.W., Deloukas.P., Gabriel.S.B. et al. (2010) Integrating common and rare genetic variation in diverse human populations. *Nature*, 467, 52–58.

50. Altshuler.D.M., Durbin.R.M., Abecasis.G.R., Bentley.D.R., Chakravarti.A., Clark.A.G., Donnelly.P., Eichler.E.E., Flicek.P., Gabriel.S.B. et al. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491, 56–65.

51. Purcell.S., Neale.B., Todd-Brown,K., Thomas.L., Ferreira,M.A.R., Bender.D., Maller.J., Sklar.P., de Bakker,P.I.W., Daly,M.J. et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, 81, 559–575.

52. Baltz.A.G., Munschauer,M., Schwanhausser.B., Basil,A., Westra,H.J., Peters,M.J., Esko.T., Yaghootkar.H., Schurmann,C., Zuberi,K., Morris.Q. and Hughes,T.R. (2010) Integrating common and rare genetic variation in diverse human populations. *Am. J. Hum. Genet.*, 87, 81–92.

53. Kooper,J.S., Saleheen,D., Sim,X., Sehmi,J., Zhang.W.H., Frossard,P., Been.L.F., Chia,K.S., Dimas.A.S., Hassanali,N. et al. (2011) Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat. Genet.*, 43, U984–U994.

54. Sun.X., Yu.W.H. and Yu.H.C. (2014) Genetics of Type 2 Diabetes: Insights into the Pathogenesis and Its Clinical Application. *Biomed. Res. Int.*, 2014, 926713–926727.

55. Westra,H.J., Peters,M.J., Esko.T., Yaghootkar.H., Schurmann,C., Kettunen,J., Christiansen,M.W., Fairfax,B.P., Schramm,K., Powell,J.E. et al. (2013) Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.*, 45, U1238–U1195.

56. Danecak.P., Nellaker.C., McIntyre,R.E., Buendia-Buendia,J.E., Bumpstead,S., Ponting,C.P., Flint,J., Durbin.R., Keane,T.M. and Adams,D.J. (2012) High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. *Genome Biol.*, 13, 26–37.

57. Garrett.S. and Rosenthal,J.J.C. (2012) RNA editing underlies temperature adaptation in k+ channels from polar octopuses. *Science*, 335, 848–851.

58. Ramaswami.G. and Li.J.B. (2014) RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.*, 42, D109–D113.

59. Kiran.A. and Baranov,P.V. (2010) DARNED: a DAtabase of RNAs EDiting in humans. *Bioinformatics*, 26, 1772–1776.

60. Chen.L.L., Li.Y., Lin.C.H., Chan,T.H.M., Chow,R.K.K., Song,Y.Y., Liu.M., Yuan,Y.F., Fu,L., Kong,K.L. et al. (2013) Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. *Nat. Med.*, 19, 209–216.

61. Guo.L.Y., Du,Y., Chang,S.H., Zhang,K.L. and Wang,J. (2014) rSNPBase: a database for curated regulatory SNPs. *Nucleic Acids Res.*, 42, D1033–D1039.