

SCIENTIFIC REPORTS



OPEN

Knowledge-Guided Bioinformatics Model for Identifying Autism Spectrum Disorder Diagnostic MicroRNA Biomarkers

Li Shen^{1,2,*}, Yuxin Lin^{1,*}, Zhandong Sun¹, Xuye Yuan¹, Luonan Chen³ & Bairong Shen¹

Received: 12 October 2016
Accepted: 24 November 2016
Published: 21 December 2016

Autism spectrum disorder (ASD) is a severe neurodevelopmental disease with a high incidence and effective biomarkers are urgently needed for its diagnosis. A few previous studies have reported the detection of miRNA biomarkers for autism diagnosis, especially those based on bioinformatics approaches. In this study, we developed a knowledge-guided bioinformatics model for identifying autism miRNA biomarkers. We downloaded gene expression microarray data from the GEO Database and extracted genes with expression levels that differed in ASD and the controls. We then constructed an autism-specific miRNA–mRNA network and inferred candidate autism biomarker miRNAs based on their regulatory modes and functions. We defined a novel parameter called the autism gene percentage as autism-specific knowledge to further facilitate the identification of autism-specific biomarker miRNAs. Finally, 11 miRNAs were screened as putative autism biomarkers, where eight miRNAs (72.7%) were significantly dysregulated in ASD samples according to previous reports. Functional enrichment results indicated that the targets of the identified miRNAs were enriched in autism-associated pathways, such as Wnt signaling (in KEGG and IPA), cell cycle (in KEGG), and glioblastoma multiforme signaling (in IPA), thereby supporting the predictive power of our model.

Autism spectrum disorder (ASD) is a severe neurodevelopmental disorder, which has attracted increasing public attention in recent years. ASD has been defined as a common developmental disorder that occurs in children according to the *Diagnostic and Statistical Manual IV (DSM-IV)*. In general, it is characterized by defective communication and social interaction, abnormal narrow interests, and repetitive behavior¹. A recent study showed that the incidence rate of ASD was 1.47% in the USA during 2010². Due to the heterogeneous nature of this disorder, a large number of studies have aimed to understand its mechanisms at different levels. In recent years, an increasing number of studies have found hundreds of genes that may contribute to autism and relevant phenotypes³. However, only a few of these studies considered autism-associated microRNAs (miRNAs), especially miRNA biomarkers.

miRNAs are small non-coding regulatory RNAs that measure 20–25 nucleotides in length. Abnormal changes in miRNA expression are known to be associated with the occurrence and development of human diseases⁴. Mundalil *et al.* used serum samples without drug treatment to conduct miRNA expression profiling and found several significant differentially expressed miRNAs that could be potential biomarkers⁵. Marrale *et al.* and Vaishnavi *et al.* reported that copy number variants or single nucleotide polymorphisms in miRNAs were correlated with ASD^{6,7}. Nguyen *et al.* reported four miRNAs that are deregulated in ASD⁸. Hicks *et al.* screened 14 salivary miRNAs as accurate indicators of ASD⁹. Wu *et al.* investigated the miRNA expression profiles in post-mortem brains from ASD patients and confirmed the function of miRNAs in ASD pathophysiology¹⁰.

We investigated ASD-associated miRNAs in our previous study and found that almost no consensus miRNA sets could be extracted. Thus, similar to prostate cancer¹¹, ASD seems to be a complex disease where the dysregulation of miRNAs is generally heterogeneous due to different genetic and non-genetic risk factors. Many

¹Center for Systems Biology, Soochow University, Suzhou, 215006, China. ²Institute of Biological Sciences and Biotechnology, Donghua University, Shanghai, 201620, China. ³Key laboratory of Systems Biology, Shanghai Institute of Biological Sciences, Chinese Academy of Sciences, Shanghai, 200031, China. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to B.S. (email: bairong.shen@suda.edu.cn)

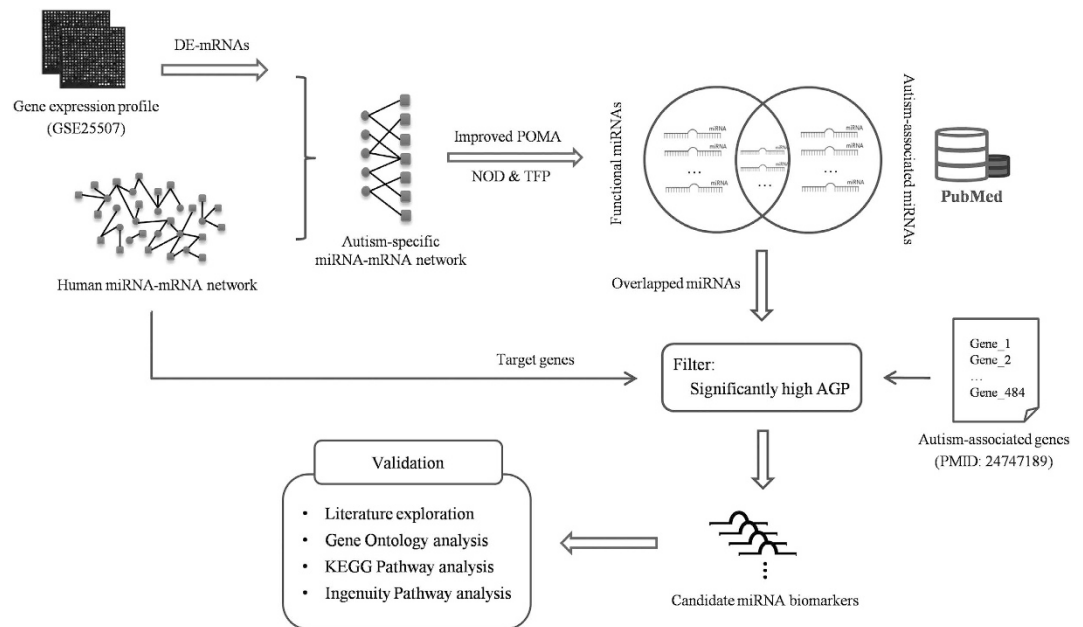


Figure 1. Schematic of the pipeline employed for identifying miRNAs as candidate autism biomarkers.

mechanisms can lead to ASD and diverse miRNA expression profiles are used as factors to identify associations with ASD. In order to discover diagnostic biomarkers, an association with the disease is the only necessary but not sufficient condition for a biomarker.

Most current methods for miRNA biomarker discovery are based on experimental screening. High-throughput expression profiling is often used to identify differentially expressed genes in disease samples and controls, before low-throughput experimental validation is employed to further refine the candidate sets. However, discovering miRNA biomarkers in the laboratory is expensive and time-consuming due to the complexity and diversity of disease progression. Given the accumulated data and knowledge regarding ASD, it is now possible as well as necessary to develop a computational and bioinformatics model to identify biomarkers that may support the precise diagnosis of ASD.

Computational or bioinformatics models for biomarker discovery are often based on machine learning and statistical inference. For example, Cogill and Wang trained a support vector machine model using developmental brain gene expression data to prioritize candidate autism genes¹², but the models built using a training-and-testing procedure were often overfitted and their generalizability was poor because the data used for training and testing were not always representative of the heterogeneity of this complex disease. However, models built according to knowledge-guided and mechanism-based analyses are more useful for biomarker discovery in complex diseases. To discover miRNA biomarkers, we previously proposed a model called Pipeline of Outlier MicroRNA Analysis (POMA) based on an integrative analysis of the substructure of a miRNA–mRNA regulatory network and disease-specific expression profiles, and this model was applied successfully to miRNA biomarker discovery in different cancers^{13–16}. In the present study, we extended and improved the POMA model to discover miRNA biomarkers for autism.

Materials and Methods

Knowledge-guided model (improved POMA model). The POMA model enriches for weak sites in a miRNA–mRNA regulatory network and considers the biological functions of miRNA targets. Two feature parameters, i.e., the novel out degree (NOD) and transcription factor gene percentage (TFP), were defined to quantify the regulatory power of miRNAs in an autism-specific miRNA–mRNA network, where the former represents the numbers of genes that are independently regulated by an individual miRNA and the latter is the percentage of transcription factor (TF) genes targeted by given miRNAs. Statistical evidence indicates that miRNAs with high NOD and TFP values are likely to be biomarkers, and applications to miRNA biomarker discovery in prostate cancer^{15,16}, clear cell renal cell carcinoma¹³, sepsis¹⁷, and pediatric acute myeloid leukemia¹⁴ support the predictive power.

To employ autism-specific knowledge to improve our ability to discover autism biomarkers, we defined a novel parameter called the autism gene percentage (AGP), which is numerically equivalent to the percentage of autism-associated genes in the miRNA targets. We assumed that miRNAs with high AGP values are more closely related to the initiation and progression of autism, and thus they are more suitable for use as biomarkers. The novel AGP index made the model more specific to autism miRNA biomarker discovery, thereby improving the precision of the predictions, and we refer to it as the improved POMA model. A schematic illustrating the pipeline of our method is shown in Fig. 1.

Dataset collection. Gene expression profile GSE25507¹⁸ was generated by an Affymetrix Human Genome U133 Plus 2.0 Array was selected from the Gene Expression Omnibus (GEO)¹⁹. The dataset contained 82 autism samples and 64 controls, where the gene samples were extracted from peripheral blood lymphocytes, and the normalized data were downloaded directly from the GEO database.

Differentially expressed gene extraction. The Limma R package was utilized to identify differentially expressed genes (DE-mRNAs) based on the selected dataset²⁰. The expression levels of genes with multiple probes equaled their average probe intensity. The Student's *t*-test was used to calculate the significant differences (*p*-value) and fold changes were determined according to the average expression level of each gene in the autism and control groups, respectively. A *p*-value <0.05 and |fold-change|>2 were used as the cut-off criteria.

Autism-specific miRNA–mRNA network construction. A human miRNA–mRNA regulatory network was built based on miRNA–mRNA pairs mined from experimentally validated (i.e., miRTarBase²¹, miRecords²², miR2Disease²³, and TarBase²⁴) and computationally predicted databases (i.e., starBase²⁵, HOCTAR²⁶, and ExprTargetDB²⁷). This network was used as a reference for autism-specific miRNA–mRNA network construction. In the next step, DE-mRNAs were mapped onto the network and the unaltered mRNAs were trimmed, where the remainder comprised the autism-specific miRNA–mRNA network.

Functional miRNA detection. We employed our in-house POMA model to detect key miRNAs with functions in autism progression (see the Introduction). First, we extracted all of the miRNAs with significantly high NOD values (*p*-value <0.05, Wilcoxon signed-rank test) in the autism-specific miRNA–mRNA network. These miRNAs had higher independent regulatory abilities so they were assumed to be sufficiently powerful to affect the system's stability. Next, we selected the miRNAs with significantly high TFP values (*p*-value <0.05, Wilcoxon signed-rank test) from those screened in the first step. These miRNAs regulated more TF genes in addition to their independent regulatory power, and thus they were considered as autism functional outliers.

Autism-associated miRNA mining. We manually searched for citations in PubMed in order to determine whether the identified functional miRNAs had reported associations with autism. In this case, “association” meant: 1) miRNAs with differential expression levels in autism samples compared with the normal controls^{28,29}; OR 2) putative autism functional miRNAs inferred by computational methods³⁰.

Candidate miRNA biomarker identification. The targets of the miRNAs were extracted from the human miRNA–mRNA regulatory network. To further investigate the specificity of these genes, we defined a novel parameter called AGP, which represents the percentage of autism-associated genes targeted by a single miRNA. Finally, miRNAs with significantly high AGP values (*p*-value <0.05, Wilcoxon signed-rank test) were identified as candidate biomarkers.

Functional enrichment analyses. Gene Ontology (GO) and pathway analysis were performed to validate the relationships between the targets of identified miRNA biomarkers and autism. The online Database for Annotation, Visualization and Integrated Discovery (DAVID, version 6.7)³¹ was used for GO annotation and the Kyoto Encyclopedia of Genes and Genomes (KEGG)³² pathway analysis. In addition, the commercial Ingenuity Pathway Analysis (IPA) program was used for pathway enrichment analyses. The Bonferroni method was used to adjust raw *p*-values and terms with adjusted *p*-values (adj.*p*-value) <0.05 were considered statistically significant and subjected to further analysis. Pathways highly connected with cancers were removed and the top 10 remaining significantly enriched pathways were selected and their correlations with autism were further confirmed by PubMed literature searches.

Results

Biomarker miRNAs for autism diagnosis. We selected 2,768 genes with different expression levels in ASD and the controls. The autism-specific miRNA–mRNA network contained 6,719 miRNA–mRNA pairs among 472 miRNAs and 1,138 genes. We identified 66 autism functional miRNAs based on our POMA model, 30 of which had reported associations with autism according to citations in PubMed (see Table S1).

We employed the specific AGP parameter to measure the power of miRNAs that regulate autism-associated genes. In total, 484 autism-associated genes were extracted from the study by Vaishnavi *et al.*⁷. We also screened candidate miRNA biomarkers from the 30 miRNAs based on our improved POMA model, i.e., considering the AGP distribution in the miRNA targets, we finally obtained 11 miRNAs (i.e., miR-193b-3p, miR-186-5p, miR-486-5p, miR-129-5p, miR-106b-5p, miR-181b-5p, miR-34a-5p, miR-96-5p, miR-211-5p, miR-205-5p, and miR-195-5p, as listed in Table 1) with significantly high AGP values as biomarkers. As shown in Fig. 2, key genes (e.g., PTEN, MeCP2, and SHANK3^{33–35}) with important roles during autism development were regulated directly by these miRNAs.

Literature validation of identified miRNA biomarkers. All 11 miRNAs were experimentally supported or they had inferred connections with autism according to our research pipeline. In particular, miR-181b-5p, miR-106b-5p, and miR-195-5p were validated by Mundalil Vasu *et al.* as candidate miRNA biomarkers for autism⁵. In their experiment, total RNA was extracted from the serum samples of autism patients and control groups, where they examined the expression levels of miRNAs and finally selected 14 differentially expressed miRNAs as autism candidates. The experimental result indicated that miR-181b-5p was downregulated in autism serum samples whereas the other two miRNAs were significantly upregulated.

Wu *et al.* performed genome-wide miRNA expression profiling and screened key miRNAs with differential expression in ASD brain samples using a linear mixed-effects model¹⁰. We found that seven of the miRNAs in

No.	miRNA ID	NOD (p-value)	TFP (p-value)	AGP (p-value)
1	miR-193b-3p	1 (6.01E-30)	0.308 (8.30E-34)	0.0714 (1.86E-09)
2	miR-186-5p	1 (6.01E-30)	0.3 (4.85E-33)	0.0566 (6.15E-08)
3	miR-486-5p	1 (6.01E-30)	0.2 (2.13E-12)	0.056 (1.66E-07)
4	miR-129-5p	1 (6.01E-30)	0.286 (2.63E-31)	0.0556 (3.15E-07)
5	miR-106b-5p	2 (8.82E-79)	0.157 (3.63E-03)	0.0508 (3.79E-06)
6	miR-181b-5p	1 (6.01E-30)	0.167 (5.07E-05)	0.05 (9.19E-06)
7	miR-34a-5p	3 (1.97E-108)	0.28 (7.02E-31)	0.0444 (2.78E-03)
8	miR-96-5p	1 (6.01E-30)	0.267 (4.88E-29)	0.0442 (3.47E-03)
9	miR-211-5p	1 (6.01E-30)	0.2 (2.13E-12)	0.0436 (6.99E-03)
10	miR-205-5p	3 (1.97E-108)	0.25 (6.55E-26)	0.0429 (1.22E-02)
11	miR-195-5p	1 (6.01E-30)	0.226 (5.59E-20)	0.0417 (4.42E-02)

Table 1. Candidate miRNA biomarkers identified by our proposed model. Note: The miRNAs are shown in descending order according to their AGP values.

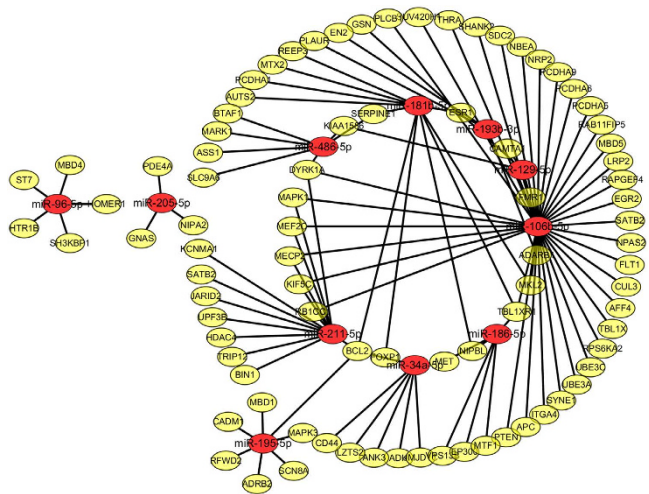


Figure 2. Autism-associated genes targeted by the identified miRNA biomarkers. Ellipses in red and yellow represent miRNAs and genes, respectively. In the network, the genes in the inner circle are regulated by at least two of the miRNAs.

our results, i.e., miR-193b-3p, miR-486-5p, miR-129-5p, miR-181b-5p, miR-34a-5p, miR-96-5p, and miR-195-5p, were also dysregulated in the study by Wu *et al.* (p -value ≤ 0.05 ; see Table S2 and Figure S1), which demonstrated the diagnostic value of our identified miRNA biomarkers.

In summary, eight of our miRNAs (72.7%, 8/11), i.e., miR-106b-5p, miR-193b-3p, miR-486-5p, miR-129-5p, miR-181b-5p, miR-34a-5p, miR-96-5p, and miR-195-5p, have been reported as significantly dysregulated in ASD samples according to previous studies.

GO analysis. The online tool DAVID was used for GO analysis at three levels: biological process (BP), cellular component, and molecular function (MF). The results are summarized in Table S3. The top 10 significantly enriched items were selected at each level, as shown in Fig. 3. Most of the genes regulated by the identified biomarker miRNAs mapped well onto autism-associated processes or biological factors. For example, at the BP level, recent studies have demonstrated that the regulation of retinoic acid-related orphan receptor γ (ROR γ) is connected with the cholesterol biosynthetic pathway³⁶. RORs have also been demonstrated to play key roles in several neural disorders, including autism³⁷. At the MF level, if we consider TF activity as an example, evidence indicates that several TFs, such as th1 and th2, are dysregulated in autism groups^{38–39}. All the results supported the functional importance of the identified autism miRNA biomarkers.

KEGG pathway analysis. To further investigate the functional mechanisms of these 11 miRNAs, we performed pathway enrichment analysis based on their targets. The enriched pathways are listed in Table S4. We removed the pathways with strong connections to cancer. Next, we selected the top 10 pathways with highly significant levels, as shown in Fig. 4. Seven pathways had experimentally validated relationships with autism, including the Wnt signaling pathway, cell cycle, p53 signaling pathway, MAPK signaling pathway, glioma signaling pathway, endocytosis, and neurotrophin signaling pathway^{40–43}. In addition, two pathways (i.e., adherens junction and TGF-beta signaling pathway) were confirmed by computational methods⁴⁴.

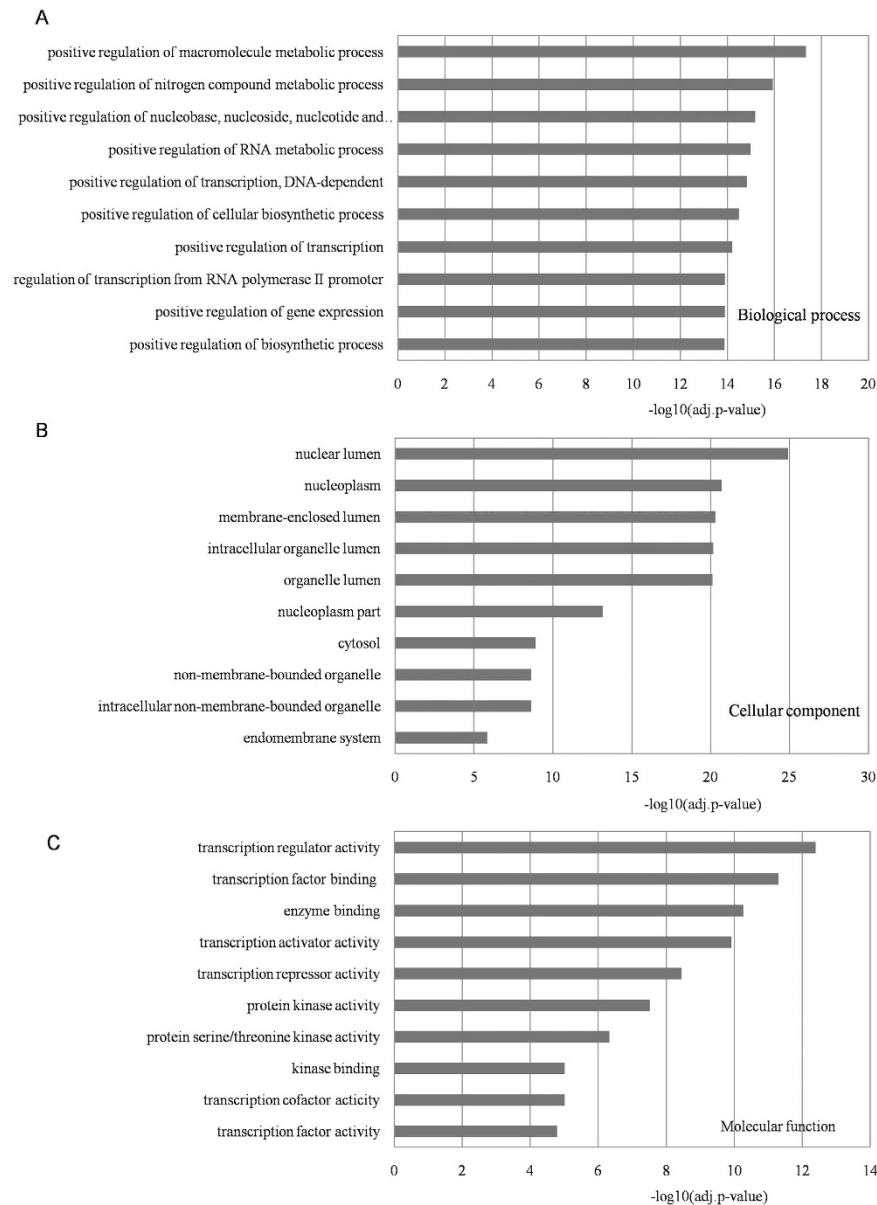


Figure 3. Gene ontology (GO) annotations for the targets of identified miRNA biomarkers. The targets of the identified miRNA biomarkers were extracted from the human miRNA–mRNA network and the online tool DAVID was used to perform GO annotation at three levels: biological process, cellular component, and molecular function, as shown in subfigures (A–C), respectively. Each statistical significance value (adj.p-value) was negative log-10 base transformed and the top 10 significantly enriched items at each level are listed.

For example, the En2 gene, which is one of the targets of miR-193b-3p, was enriched in the Wnt signaling pathway⁴⁵. Evidence indicates that its location, 7q36, is a specific chromosomal region that is strongly associated with autism^{46,47}. There is also a pvuII polymorphism in En2, which differs clearly between autistic and non-autistic children⁴⁸. An investigation of En2 null mutant mice detected autism-like symptoms, such as repetitive behavior and a communication barrier⁴⁹. The results all support a potential relationship between the Wnt signaling pathway and autism pathogenesis.

The MAPK signaling pathway transmits signals from cell surface receptors to DNA. Using KEGG, it was demonstrated that the dysregulated network built by Ghahramani Seno *et al.* was associated with nervous system development. It is interesting that the network was nucleated by several kinases from the MAPK signaling pathway. These kinases, such as MAPK and AKT, are systematically dysregulated in the development of autism. Thus, their roles in cell signal transmission may contribute to the pathogenesis of autism⁴⁰.

In the p53 signaling pathway, evidence indicates that in the brains of autistic subjects, the Bcl2 gene expression level is significantly decreased whereas the expression of p53 is increased⁵⁰. The p53 signaling pathway is also involved with the cell cycle. In addition, a recent study found that dysregulated gene networks in children affected early brain development, thereby influencing the development of autism⁵¹.

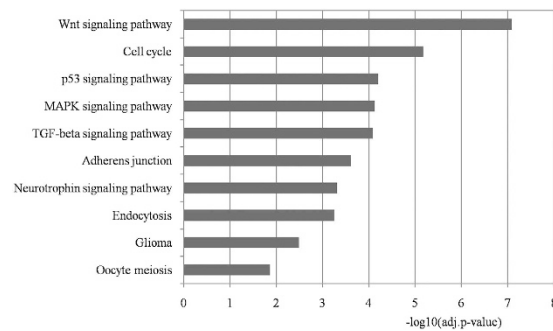


Figure 4. KEGG pathway enrichment analysis for the targets of the identified miRNA biomarkers. Each statistical significance value (adj.p-value) was negative log-10 base transformed. Pathways highly connected with cancers were removed and the top 10 remaining significantly enriched pathways are shown.

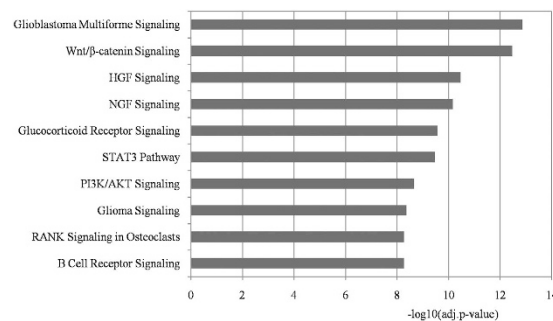


Figure 5. IPA pathway enrichment analysis of the targets of the identified miRNA biomarkers. Each statistical significance value (adj.p-value) was negative log-10 base transformed. Pathways strongly connected with cancers and mouse embryonic stem cell pluripotency were removed. The top 10 remaining significantly enriched pathways are shown.

The other three pathways validated by biological experiments are endocytosis, the neurotrophin signaling pathway, and the glioma signaling pathway. During endocytosis, the SCAMP5 gene is essential when the neuronal activity is high, where it might control the recycling of synaptic vesicles, and thus it could maintain sufficient endocytosis during intense neuronal activity. Thus, the knockdown of SCAMP5 would have negative effects on neuronal functions and lead to synaptic dysfunction⁴³. Moreover, brain-derived neurotrophic factor (BDNF) in the neurotrophin signaling pathway plays vital roles in neurodevelopmental and neurodegenerative diseases⁵². Several recent studies have indicated that reduced cerebrospinal fluid levels of BDNF are associated with Rett syndrome⁵³. It has been demonstrated that the loss of MeCP2 might contribute to decreases in neurotrophic factors, including BDNF, which have strong relationships with autism. A study of C6 glioma cells found that the risperidone targets on glial cells *in vivo* regulate the secretion of S100B from C6 glioma cells, which may be involved in autism⁵⁴.

IPA. IPA is a commercial program for understanding biological systems, including pathway analysis. The significantly enriched IPA pathways are listed in Table S5. Pathways related to cancers and mouse embryonic stem cell pluripotency were removed. The top 10 remaining significantly enriched pathways are summarized in Fig. 5.

We also evaluated the relevance of these pathways to autism by searching for publications in PubMed. In total, 8/10 enriched pathways have been reported to have associations with autism, including the nerve growth factor (NGF) signaling pathway, glioblastoma multiforme signaling pathway, hepatocyte growth factor (HGF) signaling pathway, glucocorticoid receptor signaling pathway, STAT3 pathway, glioma signaling pathway, Wnt/ β -catenin signaling pathway, and PI3K/AKT signaling pathway. Interestingly, the Wnt/ β -catenin and glioma signaling pathway also appeared in the IPA outcome lists, thereby strengthening their relationships with autism. Previous studies have indicated that several differential alternative splicing genes (DAS), such as FXR1, C19orf2, GSN, and SRPK1, are associated with autism. These DAS genes are functional in many signaling pathways, including the NGF signaling pathway⁵⁵.

Omura also identified a relationship between the glioblastoma multiforme signaling pathway and autism⁵⁶. Glioblastoma multiforme is one of the most serious brain tumors and patients with this tumor have often been exposed to a high concentration of asbestos. Asbestos also has relatively strong effects on many other diseases, including autism. This is the only report that has connected autism and glioblastoma multiforme, and thus further experimental validation is needed.

The key gene MET in the HGF signaling pathway, which is regulated by our identified marker miR-186-5p, plays important roles in the pathophysiology of neurodevelopmental disorders, including autism⁵⁷. Campbell *et al.* also reported a strong link between MET in the HGF signaling pathway and autism⁵⁸.

The glucocorticoid receptor signaling pathway is an important pathway *in vivo*. Studies have shown that there is a link between this pathway and relaxin family peptide (RXFP) members⁵⁹. One of the relaxin family peptide members is RXFP-3 and it is a potential drug target for autism treatment⁶⁰.

The remaining three pathways that have been reported as abnormal in autism are the STAT3 pathway, glioma signaling pathway, and PI3K/AKT signaling pathway. Evidence suggests that increased interleukin-6 (IL-6) levels have an important role in influencing brain development. IL-6 induces JAK/STAT3 phosphorylation *in vivo*, which could lead to social and behavior deficits. Inhibiting this process may reduce the risk of autism⁶¹. Bailey *et al.* found that the PI3K/AKT signaling process driven by sAPP- α was aberrant in autistic patients⁶².

Discussion

It has been widely reported that miRNAs can be effective biomarkers for neural disorders, such as Alzheimer's disease and Parkinson's disease^{63,64}. However, few studies have focused on detecting miRNA biomarkers for autism and only a small number of approaches have been proposed based on computational methods. In this study, we improved our miRNA biomarker discovery model by adding a novel autism-specific index called AGP, which we then employed to screen key miRNAs for autism diagnosis. We assumed that miRNAs should have at least two vital characteristics as biomarkers: first, they tend to be dysfunctional in disease samples compared with normal controls, and second, but more importantly, they are sufficiently powerful to indicate the state of the system in health and disease. We inferred 66 functional autism-associated miRNAs using gene expression data based on miRNA-mRNA network analysis and we then selected 30 outliers by manually searching for citations in PubMed. The selected miRNAs satisfied the characteristics of miRNA biomarkers well because they all had reported associations with autism progression, as well as having strong independent regulatory power (high NOD values), and they could regulate more TF genes (high TFP values) according to our predictive model. As mentioned in our previous study¹⁵, disorders in unique regulatory sites are more likely to cause system-level changes, while dysfunctions of TFs may also contribute to disease pathogenesis because they are important regulators that are always located at the center of multiple biological activities¹⁴.

To strengthen the bioinformatics model and make it more specific to autism miRNA biomarker discovery, we defined the parameter AGP to guide biomarker discovery. Based on the hypothesis that miRNAs with regulatory effects on autism-associated genes would have stronger relationships with autism, we screened a total of 11 miRNAs with significantly high AGP values from the selected miRNA dataset, which were identified as candidate biomarkers for autism diagnosis. Three of these miRNAs have been reported as potential biomarkers for autism⁵, and the regulated genes, including MeCP2, MET, and EN2, were differentially expressed in samples of autism patients validated by biological experiments^{65–67}. Moreover, seven miRNAs were also shown to be significantly dysregulated in brain samples from autism patients, according to a study by Wu *et al.*¹⁰. We used DAVID and IPA as functional enrichment tools for GO and pathway analysis. The results showed that the genes targeted by the identified miRNA biomarkers were enriched in pathways associated strongly with neural development, such as the Wnt signaling pathway, p53 signaling pathway, and PI3K/AKT signaling pathway^{42,50,62}. For example, previous studies have implicated Wnt signaling in vertebrate neural patterning⁶⁸, and the functional gene EN2, which is located in a chromosomal region that is strongly associated with autism development, is a direct target of this pathway⁴⁵. Several DAS genes, such as FXR1, C19orf2, and GSN, play essential roles in the NGF signaling pathway, and their functional dysregulation may contribute to the initiation of autism in children⁵⁵. In addition to the biological functions of the genes in this pathway, the occurrence of autism is often due to abnormal changes in pathways related to other neural diseases. For example, glioblastoma multiforme signaling is potentially related to autism because exposure to a high concentration of asbestos (which causes glioblastoma multiforme) may promote the initiation of autism⁵⁶. Overall, the pathway enrichment analyses showed that the targets of the identified miRNA biomarkers are strongly related to autism-associated processes, thereby supporting the predictive power of our proposed model. In this study, we proposed a systematic bioinformatics method for identifying autism miRNA biomarkers, which can provide insights into the pathogenesis of autism at the post-transcriptional level. Further clinical validation will be needed in the future in order to translate our findings from theoretical insights into clinical practice.

Conclusion

In this study, we identified 11 miRNAs that could be used as candidate biomarkers for autism diagnosis based on a knowledge-guided bioinformatics model. Literature searches and functional enrichment analyses confirmed our predictions.

References

- Ruhela, R. K., Prakash, A. & Medhi, B. An urgent need for experimental animal model of autism in drug development. *Ann Neurosci* **22**, 44–49, doi: 10.5214/ans.0972.7531.220210 (2015).
- Developmental Disabilities Monitoring Network Surveillance Year Principal, I., Centers for Disease, C. & Prevention. Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *MMWR Surveill Summ* **63**, 1–21 (2014).
- Chang, J., Gilman, S. R., Chiang, A. H., Sanders, S. J. & Vitkup, D. Genotype to phenotype relationships in autism spectrum disorders. *Nature neuroscience* **18**, 191–198, doi: 10.1038/nn.3907 (2015).
- Lossos, I. S. *et al.* Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *N Engl J Med* **350**, 1828–1837, doi: 10.1056/NEJMoa032520 (2004).
- Mundalil Vasu, M. *et al.* Serum microRNA profiles in children with autism. *Molecular autism* **5**, 40, doi: 10.1186/2040-2392-5-40 (2014).

6. Marrale, M., Albanese, N. N., Cali, F. & Romano, V. Assessing the impact of copy number variants on miRNA genes in autism by Monte Carlo simulation. *PLoS One* **9**, e90947, doi: 10.1371/journal.pone.0090947 (2014).
7. Vaishnavi, V., Manikandan, M. & Munirajan, A. K. Mining the 3'UTR of autism-implicated genes for SNPs perturbing microRNA regulation. *Genomics Proteomics Bioinformatics* **12**, 92–104, doi: 10.1016/j.gpb.2014.01.003 (2014).
8. Nguyen, L. S. *et al.* Profiling olfactory stem cells from living patients identifies miRNAs relevant for autism pathophysiology. *Mol Autism* **7**, 1, doi: 10.1186/s13229-015-0064-6 (2016).
9. Hicks, S. D., Ignacio, C., Gentile, K. & Middleton, F. A. Salivary miRNA profiles identify children with autism spectrum disorder, correlate with adaptive behavior, and implicate ASD candidate genes involved in neurodevelopment. *BMC Pediatr* **16**, 52, doi: 10.1186/s12887-016-0586-x (2016).
10. Wu, Y. E., Parikshak, N. N., Belgard, T. G. & Geschwind, D. H. Genome-wide, integrative analysis implicates microRNA dysregulation in autism spectrum disorder. *Nature neuroscience* **19**, 1463–1476, doi: 10.1038/nn.4373 (2016).
11. Tang, Y. *et al.* Identification of novel microRNA regulatory pathways associated with heterogeneous prostate cancer. *BMC systems biology* **7** Suppl 3, S6, doi: 10.1186/1752-0509-7-S3-S6 (2013).
12. Cogill, S. & Wang, L. Support vector machine model of developmental brain gene expression data for prioritization of Autism risk gene candidates. *Bioinformatics*, doi: 10.1093/bioinformatics/btw498 (2016).
13. Chen, J. *et al.* Clear cell renal cell carcinoma associated microRNA expression signatures identified by an integrated bioinformatics analysis. *J Transl Med* **11**, 169, doi: 10.1186/1479-5876-11-169 (2013).
14. Yan, W. *et al.* MicroRNA biomarker identification for pediatric acute myeloid leukemia based on a novel bioinformatics model. *Oncotarget* **6**, 26424–26436, doi: 10.18632/oncotarget.4459 (2015).
15. Zhang, W. *et al.* Identification of candidate miRNA biomarkers from miRNA regulatory network with application to prostate cancer. *J Transl Med* **12**, 66, doi: 10.1186/1479-5876-12-66 (2014).
16. Zhu, J. *et al.* Screening key microRNAs for castration-resistant prostate cancer based on miRNA/mRNA functional synergistic network. *Oncotarget* **6**, 43819–43830, doi: 10.18632/oncotarget.6102 (2015).
17. Huang, J. *et al.* Identification of microRNA as sepsis biomarker based on miRNAs regulatory network analysis. *Biomed Res Int* **2014**, 594350, doi: 10.1155/2014/594350 (2014).
18. Alter, M. D. *et al.* Autism and increased paternal age related changes in global levels of gene expression regulation. *PLoS One* **6**, e16715, doi: 10.1371/journal.pone.0016715 (2011).
19. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207–210 (2002).
20. Smyth, G. K. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* eds Robert, Gentleman *et al.*) 397–420 (Springer New York, 2005).
21. Hsu, S. D. *et al.* miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res* **42**, D78–85, doi: 10.1093/nar/gkt1266 (2014).
22. Xiao, F. *et al.* miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res* **37**, D105–110, doi: 10.1093/nar/gkn851 (2009).
23. Jiang, Q. *et al.* miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res* **37**, D98–104, doi: 10.1093/nar/gkn714 (2009).
24. Vergoulis, T. *et al.* TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Res* **40**, D222–229, doi: 10.1093/nar/gkr1161 (2012).
25. Li, J. H., Liu, S., Zhou, H., Qu, L. H. & Yang, J. H. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* **42**, D92–97, doi: 10.1093/nar/gkt1248 (2014).
26. Gennarino, V. A. *et al.* HOCTAR database: a unique resource for microRNA target prediction. *Gene* **480**, 51–58, doi: 10.1016/j.gene.2011.03.005 (2011).
27. Gamazon, E. R. *et al.* Exptarget: an integrative approach to predicting human microRNA targets. *PLoS One* **5**, e13534, doi: 10.1371/journal.pone.0013534 (2010).
28. Sarachana, T., Zhou, R., Chen, G., Manji, H. K. & Hu, V. W. Investigation of post-transcriptional gene regulatory networks associated with autism spectrum disorders by microRNA expression profiling of lymphoblastoid cell lines. *Genome Med* **2**, 23, doi: 10.1186/gm144 (2010).
29. Abu-Elneel, K. *et al.* Heterogeneous dysregulation of microRNAs across the autism spectrum. *Neurogenetics* **9**, 153–161, doi: 10.1007/s10048-008-0133-5 (2008).
30. Driessen, T. M. *et al.* Genes showing altered expression in the medial preoptic area in the highly social maternal phenotype are related to autism and other disorders with social deficits. *BMC Neurosci* **15**, 11, doi: 10.1186/1471-2202-15-11 (2014).
31. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57, doi: 10.1038/nprot.2008.211 (2009).
32. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30 (2000).
33. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet* **23**, 185–188, doi: 10.1038/13810 (1999).
34. Berkel, S. *et al.* Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. *Nat Genet* **42**, 489–491, doi: 10.1038/ng.589 (2010).
35. Butler, M. G. *et al.* Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. *J Med Genet* **42**, 318–321, doi: 10.1136/jmg.2004.024646 (2005).
36. Hu, X. *et al.* Sterol metabolism controls T(H)17 differentiation by generating endogenous RORgamma agonists. *Nat Chem Biol* **11**, 141–147, doi: 10.1038/nchembio.1714 (2015).
37. Nguyen, A., Rauch, T. A., Pfeifer, G. P. & Hu, V. W. Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain. *FASEB J* **24**, 3036–3051, doi: 10.1096/fj.10-154484 (2010).
38. Ahmad, S. F. *et al.* Dysregulation of Th1, Th2, Th17, and T regulatory cell-related transcription factor signaling in children with autism. *Mol Neurobiol*. doi: 10.1007/s12035-016-9977-0 (2016).
39. Chen, F. *et al.* The transcription factor NeuroD2 coordinates synaptic innervation and cell intrinsic properties to control excitability of cortical pyramidal neurons. *J Physiol* **594**, 3729–3744, doi: 10.1113/JP271953 (2016).
40. Ghahramani Seno, M. M. *et al.* Gene and miRNA expression profiles in autism spectrum disorders. *Brain Res* **1380**, 85–97, doi: 10.1016/j.brainres.2010.09.046 (2011).
41. Theoharides, T. C., Athanassiou, M., Panagiotidou, S. & Doyle, R. Dysregulated brain immunity and neurotrophin signaling in Rett syndrome and autism spectrum disorders. *J Neuroimmunol* **279**, 33–38, doi: 10.1016/j.jneuroim.2014.12.003 (2015).
42. Zhang, Y., Yuan, X., Wang, Z. & Li, R. The canonical Wnt signaling pathway in autism. *CNS Neurol Disord Drug Targets* **13**, 765–770 (2014).
43. Zhao, H. *et al.* SCAMP5 plays a critical role in synaptic vesicle endocytosis during high neuronal activity. *J Neurosci* **34**, 10085–10095, doi: 10.1523/JNEUROSCI.2156-14.2014 (2014).
44. Huang, F. *et al.* Investigation of Gene Regulatory Networks Associated with Autism Spectrum Disorder Based on MiRNA Expression in China. *PLoS One* **10**, e0129052, doi: 10.1371/journal.pone.0129052 (2015).

45. McGrew, L. L., Takemaru, K., Bates, R. & Moon, R. T. Direct regulation of the *Xenopus* engrailed-2 promoter by the Wnt signaling pathway, and a molecular screen for Wnt-responsive genes, confirm a role for Wnt signaling during neural patterning in *Xenopus*. *Mechanisms of development* **87**, 21–32 (1999).
46. Gharani, N., Benayed, R., Mancuso, V., Brzustowicz, L. M. & Millonig, J. H. Association of the homeobox transcription factor, ENGRAILED 2, 3, with autism spectrum disorder. *Mol Psychiatry* **9**, 474–484, doi: 10.1038/sj.mp.4001498 (2004).
47. Yonan, A. L. *et al.* A genomewide screen of 345 families for autism-susceptibility loci. *Am J Hum Genet* **73**, 886–897 (2003).
48. Petit, E. *et al.* Association study with two markers of a human homegene in infantile autism. *J Med Genet* **32**, 269–274 (1995).
49. Brielmaier, J. *et al.* Autism-relevant social abnormalities and cognitive deficits in engrailed-2 knockout mice. *PLoS One* **7**, e40914, doi: 10.1371/journal.pone.0040914 (2012).
50. Sheikh, A. M. *et al.* BDNF-Akt-Bcl2 antiapoptotic signaling pathway is compromised in the brain of autistic subjects. *J Neurosci Res* **88**, 2641–2647, doi: 10.1002/jnr.22416 (2010).
51. Pramparo, T. *et al.* Cell cycle networks link gene expression dysregulation, mutation, and brain maldevelopment in autistic toddlers. *Mol Syst Biol* **11**, 841, doi: 10.15252/msb.20156108 (2015).
52. Auty, A. E. & Monteggia, L. M. Brain-derived neurotrophic factor and neuropsychiatric disorders. *Pharmacol Rev* **64**, 238–258, doi: 10.1124/pr.111.005108 (2012).
53. Katz, D. M. Brain-derived neurotrophic factor and Rett syndrome. *Handb Exp Pharmacol* **220**, 481–495, doi: 10.1007/978-3-642-45106-5_18 (2014).
54. Quincozes-Santos, A. *et al.* Effect of the atypical neuroleptic risperidone on morphology and S100B secretion in C6 astroglial lineage cells. *Mol Cell Biochem* **314**, 59–63, doi: 10.1007/s11010-008-9765-x (2008).
55. Stamova, B. S. *et al.* Evidence for differential alternative splicing in blood of young boys with autism spectrum disorders. *Molecular autism* **4**, 30, doi: 10.1186/2040-2392-4-30 (2013).
56. Omura, Y. Asbestos as a possible major cause of malignant lung tumors (including small cell carcinoma, adenocarcinoma & mesothelioma), brain tumors (i.e. astrocytoma & glioblastoma multiforme), many other malignant tumors, intractable pain including fibromyalgia, & some cardio-vascular pathology: Safe & effective methods of reducing asbestos from normal & pathological areas. *Acupuncture & electro-therapeutics research* **31**, 61–125 (2006).
57. Judson, M. C., Bergman, M. Y., Campbell, D. B., Eagleson, K. L. & Levitt, P. Dynamic gene and protein expression patterns of the autism-associated met receptor tyrosine kinase in the developing mouse forebrain. *J Comp Neurol* **513**, 511–531, doi: 10.1002/cne.21969 (2009).
58. Campbell, D. B., Li, C., Sutcliffe, J. S., Persico, A. M. & Levitt, P. Genetic evidence implicating multiple genes in the MET receptor tyrosine kinase pathway in autism spectrum disorder. *Autism Res* **1**, 159–168, doi: 10.1002/aur.27 (2008).
59. Dschietzig, T., Bartsch, C., Baumann, G. & Stangl, K. RXFP1-inactive human relaxin activates human glucocorticoid receptor: further investigations into the relaxin-GR pathway. *Regul Pept* **154**, 77–84, doi: 10.1016/j.regpep.2008.11.010 (2009).
60. Halls, M. L., Bathgate, R. A., Sutton, S. W., Dschietzig, T. B. & Summers, R. J. International Union of Basic and Clinical Pharmacology. XCV. Recent advances in the understanding of the pharmacology and biological roles of relaxin family peptide receptors 1–4, the receptors for relaxin family peptides. *Pharmacol Rev* **67**, 389–440, doi: 10.1124/pr.114.009472 (2015).
61. Parker-Athill, E. *et al.* Flavonoids, a prenatal prophylaxis via targeting JAK2/STAT3 signaling to oppose IL-6/MIA associated autism. *J Neuroimmunol* **217**, 20–27, doi: 10.1016/j.jneuroim.2009.08.012 (2009).
62. Bailey, A. R. *et al.* Aberrant T-lymphocyte development and function in mice overexpressing human soluble amyloid precursor protein-alpha: implications for autism. *FASEB J* **26**, 1040–1051, doi: 10.1096/fj.11-195438 (2012).
63. Muller, M. *et al.* MicroRNA-29a Is a Candidate Biomarker for Alzheimer's Disease in Cell-Free Cerebrospinal Fluid. *Mol Neurobiol* **53**, 2894–2899, doi: 10.1007/s12035-015-9156-8 (2016).
64. Ding, H. *et al.* Identification of a panel of five serum miRNAs as a biomarker for Parkinson's disease. *Parkinsonism Relat Disord* **22**, 68–73, doi: 10.1016/j.parkreldis.2015.11.014 (2016).
65. LaSalle, J. M. & Yasui, D. H. Evolving role of MeCP2 in Rett syndrome and autism. *Epigenomics* **1**, 119–130, doi: 10.2217/epi.09.13 (2009).
66. Peng, Y. *et al.* The autism-associated MET receptor tyrosine kinase engages early neuronal growth mechanism and controls glutamatergic circuits development in the forebrain. *Mol Psychiatry* **21**, 925–935, doi: 10.1038/mp.2015.182 (2016).
67. Genestine, M. *et al.* Engrailed-2 (*En2*) deletion produces multiple neurodevelopmental defects in monoamine systems, forebrain structures and neurogenesis and behavior. *Hum Mol Genet* **24**, 5805–5827, doi: 10.1093/hmg/ddv301 (2015).
68. McGrew, L. L., Hoppler, S. & Moon, R. T. Wnt and FGF pathways cooperatively pattern anteroposterior neural ectoderm in *Xenopus*. *Mechanisms of development* **69**, 105–114 (1997).

Acknowledgements

This study was supported by the National Natural Science Foundation of China (NSFC) (grant nos 31670851, 31470821, and 91530320) and National Key R&D programs of China (2016YFC1306600)

Author Contributions

L.S. and Y.L. contributed equally to this study. L.S., Y.L. and Z.S. performed the computational analyses; L.S., X.Y. and L.C. performed the validation work; Y.L. contributed the analysis tool; L.S., Y.L. and B.S. wrote the manuscript; B.S. conceived and supervised the study.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Shen, L. *et al.* Knowledge-Guided Bioinformatics Model for Identifying Autism Spectrum Disorder Diagnostic MicroRNA Biomarkers. *Sci. Rep.* **6**, 39663; doi: 10.1038/srep39663 (2016).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016