

OPEN

An integrated transcriptomic analysis of autism spectrum disorder

Yi He^{1,2}, Yuan Zhou¹, Wei Ma³ & Juan Wang^{1,2}

Autism spectrum disorder (ASD) is not a single disease but a set of disorders. To find clues of ASD pathogenesis in transcriptomic data, we performed an integrated transcriptomic analysis of ASD. After screening based on several standards in Gene Expression Omnibus (GEO) database, we obtained 11 series of transcriptomic data of different human tissues of ASD patients and healthy controls. Multidimensional scaling analysis revealed that datasets from the same tissue had bigger similarity than from different tissues. Functional enrichment analysis demonstrated that differential expressed genes were significantly enriched in inflammation/immune response, mitochondrion-related function and oxidative phosphorylation. Interestingly, genes enriched in inflammation/immune response were up-regulated in the brain tissues and down-regulated in the blood. In addition, drug prediction provided several compounds which might reverse gene expression profiles of ASD patients. And we also replicated the methods and criteria of transcriptomic analysis with datasets of ASD animal models and healthy controls, the results from animal models consolidated the results of transcriptomic analysis of ASD human tissues. In general, the results of our study may provide researchers a new sight of understanding the etiology of ASD and clinicians the possibilities of developing medical therapies.

Autism spectrum disorder (ASD) represents a set of neurodevelopmental disorders characterized by two core symptoms, impaired social interaction and restrictive and repetitive behaviors. It is sex-biased that ASD affects boys four to five times more than girls¹. The pathogenesis of ASD still remains perplexing. Biological researches supported that it was a set of disorders with multiple non-genetic and genetic factors as well as their interactions rather than a single disease^{2–6}. About 10–20% of ASD patients have a definite genetic risk⁷. However, the genetic etiology of ASD is heterogeneous. Up till now, hundreds of genes have been found associated with ASD⁶. For example, synaptic genes such as neuroligin 3 (*NLGN3*), neuroligin 4X-linked (*NLGN4X*)⁸, SH3 and multiple ankyrin repeat domains 3 (*SHANK3*)^{9,10} were certified as the ASD genes. Monogenetic diseases like Fragile-X syndrome and Rett Syndrome associated with the mutations of fragile X mental retardation 1 (*FMR1*)/methyl-CpG binding protein 2 (*MECP2*) were also found with autistic symptoms¹¹. Ubiquitin protein ligase E3A (*UBE3A*) gene coding for E3 ubiquitin-protein ligase was linked to both ASD and Angelman Syndrome². The results above came from the researches of DNA sequence, but they cannot identify the cause of ASD in a large number of cases¹¹.

Transcriptomic analysis plays an important role in exploring genome structure and function and identifying genetic networks in aspect of gene transcription. Gene expression patterns in autistic population have been demonstrated in different tissues including lymphoblastoid cell lines, peripheral blood, brain cells and pluripotent stem cell-derived neurons^{6,12–14}. Several researches pointed out that aberrant gene expression in the blood of children with ASD involved in transcriptional regulation, biosynthesis of protein, processing of ribosomal RNA and neural-related pathways^{13,15,16}. Transcriptome analysis between autistic brain and normal brain identified discrete modules by gene co-expression network analysis: a neuronal module and a module enriched for immune genes and glial markers¹². Another finding of brain transcriptional association with ASD revealed two abnormal areas of metabolism: mitochondrial oxidative phosphorylation and protein translation¹⁷.

Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) has collected the raw and processed data for studies of high-throughput gene expression and genomics¹⁸. To date, GEO has gene expression datasets coming from hundreds of studies related to ASD. Although widely assumed, it remains unknown whether the

¹Department of Biomedical Informatics, School of Basic Medical Sciences, Peking University, Beijing, 100191, China.

²Autism Research Center of Peking University Health Science Center, Beijing, 100191, China. ³The Sixth Medical Center, Chinese PLA General Hospital, Beijing, 100048, China. Correspondence and requests for materials should be addressed to J.W. (email: wjuan@hsc.pku.edu.cn)

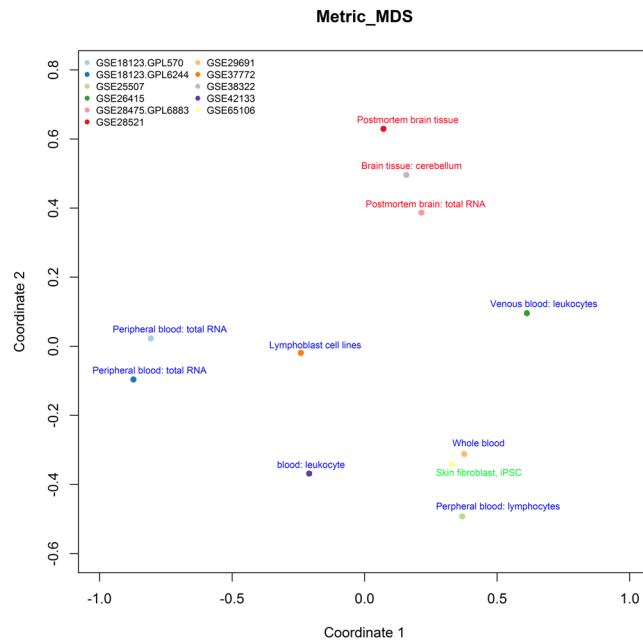


Figure 1. Multidimensional scaling map of 11 series of datasets. Multidimensional scaling analysis was performed among these datasets based on the distance matrix. Each dot represented a series of dataset. For each dataset, the tissue source was noted on the dot, brain tissues were presented by red, blood was presented by blue. Spatial distance among datasets derived from same tissue was smaller than that derived from different tissues ($P = 0.0013$). That meant datasets coming from same tissue had bigger similarity.

transcriptomic signatures were consistent among ASD population. Evidence certainly denoted that differential expressed genes in the blood could be an indicator of ASD, however, whether the transcriptomic signatures in different tissues were consistent with each other hasn't been explored. Whether we could find some potential drugs based on transcriptomic signatures of ASD. The purpose of our study is to systematically explore the status of gene expression between ASD and healthy control by integrating several human datasets. In this work, we chose the transcriptomic datasets from GEO based on several criteria and screened the differential expressed genes by a computational method and carried on a series of bioinformatic analysis. And we attempted to predict potential drugs based on these differential expressed genes. In addition, we also replicated the strategy and criteria of the analysis in transcriptomic datasets of animal model to validate the results of human tissues.

Results

Datasets derived from same tissue had bigger similarity. To obtain the spatial or geometric representations of the datasets, multidimensional scaling (MDS) was utilized in our study. As shown in Fig. 1, three series of datasets (GSE28475, GSE28521, GSE38322) were derived from brain tissues gathered together, and they were apart from other datasets spatially. In other words, the dimensional distances between any two of three datasets derived from brain tissue were smaller than that between any of them and dataset which came from blood. Similarly, datasets from blood (GSE18123, GSE25507, GSE29691, GSE37772, GSE42133) gathered together. Given the above, datasets derived from the same tissue had the bigger similarity than from different tissues. (The average distance of datasets derived from brain tissues = 1.18, the average distance of datasets derived from blood = 1.40, $P = 0.0013$).

Gene expression level between brain tissues and blood. Based on the fold change of expression level, we found that differential expressed genes of ASD patients/controls in brain tissues and blood, had different gene expression patterns. The number of up-regulated genes in brain tissues were apparently more than that in blood, meanwhile the number of down-regulated genes in brain tissues was less than that in blood (Supplementary Fig. S1). Especially, several genes up-regulated in brain tissue were down-regulated in blood, like TIMP metalloproteinase inhibitor 1 (*TIMP1*), phospholipase A and acyltransferase 4 (*RARRES3*), DNA damage inducible transcript 4 (*DDIT4*), cytochrome b-245 alpha chain (*CYBA*), bone marrow stromal cell antigen 2 (*BST2*) (Fig. 2). Namely, there existed a transcriptomic difference in ASD patients' tissues between brain tissues and blood.

Functional enrichment of the differential expressed genes. In terms of fold change and P -value of Wilcoxon rank-sum test, each dataset had two gene lists, one for up-regulated genes and the other for down-regulated genes. We input them into the DAVID respectively, and obtained two sets of functional annotations in each dataset. Pathway analysis indicated that multiple pathways associated with inflammation/immune response, mitochondrion-related function and other significantly meaningful pathways (Figs 3 and 4, Supplementary Fig. S2), which were consistent with previous studies^{12,17,19,20}. In this work, genes enriched in the KEGG pathway of inflammation/immune related disease, e.g. systemic lupus erythematosus (KEGG

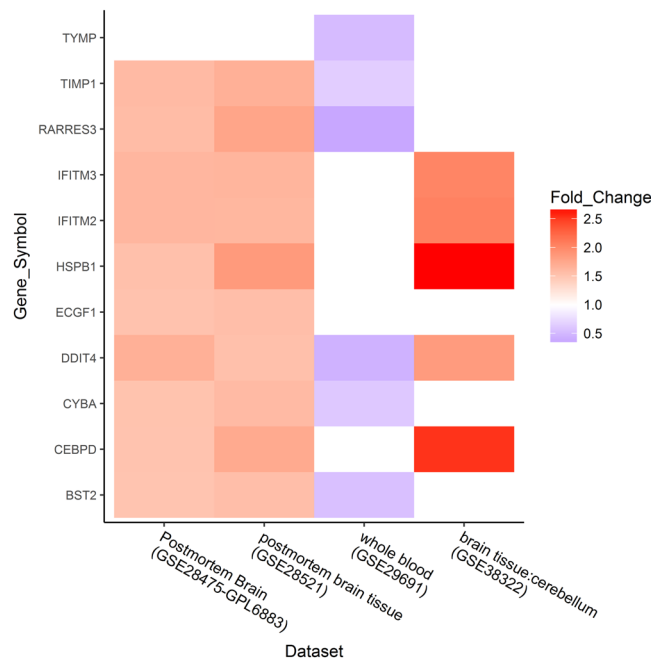


Figure 2. Gene expression levels in different tissues. Each dataset had a differential expressed gene list, gene symbols of y-coordinate were present in at least three datasets. Up-regulated genes were represented by red, and down-regulated genes were blue. The shade of the color reflected the value of fold change of each gene.

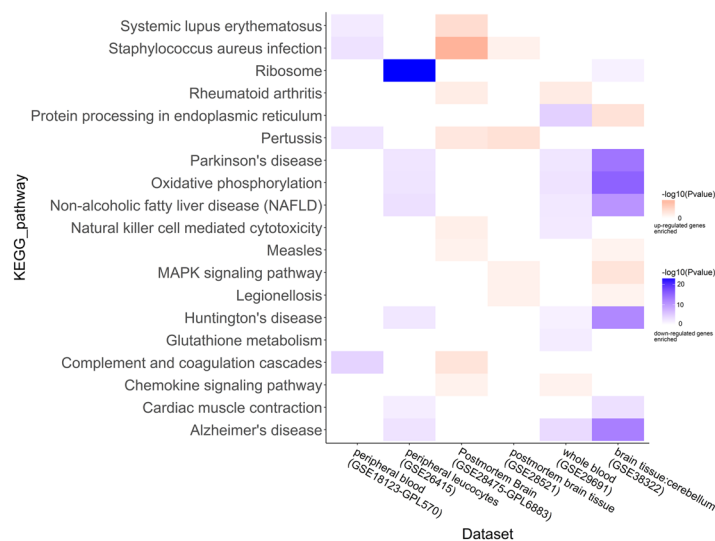


Figure 3. Heat-map for the enriched KEGG pathways. Functional enrichment analysis were performed by DAVID. P value of each KEGG pathways was <0.05 . KEGG pathways of y-coordinate were the pathways which overlapped in two datasets or more. Up-regulated genes enriched KEGG pathways were represented by red, and down-regulated genes enriched KEGG pathways were blue. The shades of the colors reflected the $-\log(P$ value) of the enrichment analysis.

pathway identifier: hsa05322), staphylococcus aureus infection (KEGG pathway identifier: hsa05150), etc., were up-regulated in the brain tissues and down-regulated in the blood (Fig. 3). It was similar with the Gene Ontology (GO) terms. Genes enriched in inflammation/immune response were up-regulated in the brain tissues and down-regulated in blood (Fig. 4). Increasing evidence indicated that neuropsychiatric diseases were associated with brain inflammation, such as ASD, schizophrenia²¹. And the immune dysfunction might be the key point for the genetic and environmental factors to develop ASD²². However, we were the first to point out that genes enriched in inflammation/immune response were differential expressed in brain and blood. Moreover, we also found down-regulated genes both in the brain tissues and blood were enriched in some significant pathways, such as oxidative phosphorylation and mitochondrion-related function (Fig. 4, Supplementary Fig. S2). This suggested that there might exist energy metabolism disorders in ASD population. Previous studies have shown

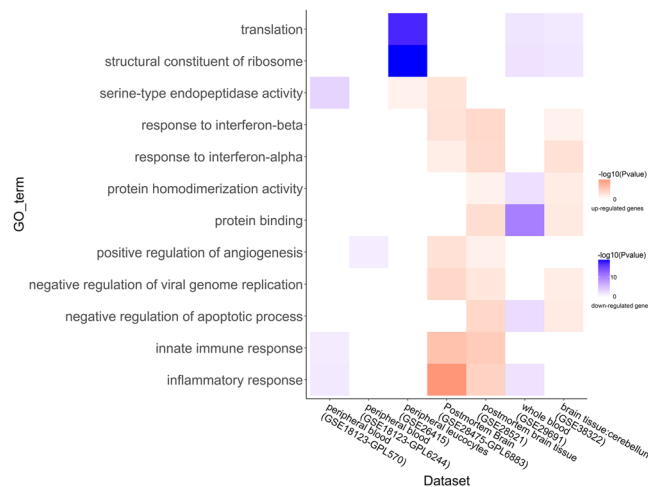


Figure 4. Heat-map for the enriched GO terms. Functional enrichment analysis were performed by DAVID. *P* value of each GO term was <0.05 . GO terms of *y*-coordinate were the pathways which overlapped in three datasets or more. Up-regulated genes enriched GO terms were represented by red, and down-regulated genes enriched GO terms were blue. The shades of the colors reflected the $-\log(P \text{ value})$ of the enrichment analysis.

Compounds	Description
efavirenz	HIV protease inhibitor
SB-525334	TGF beta receptor inhibitor
RHO-kinase-inhibitor-III[rockout]	Rho associated kinase inhibitor
rilmenidine	Adrenergic receptor agonist
PKCbeta-inhibitor	PKC inhibitor
deferiprone	Chelating agent
nifedipine	Calcium channel blocker
cholic-acid	Bile acid

Table 1. Compounds overlapping in four datasets.

that ASD might be associated with mitochondrial disorders. About twenty years ago, researchers reported a boy whose origin of the ASD might be the mitochondrial DNA mutation²³. A systematic review has reported that the prevalence of the mitochondrial disease in general population of ASD was 5.0%, which was much higher than the prevalence of general population²⁴. Functional analysis also demonstrated that differential expressed genes enriched in protein translation and ribosome related functions were down-regulated both in brain tissues and blood. Previous study has confirmed that changes in the efficiency of protein translation were related to ASD by using a published set of 1,800 autism quartets and genome-wide variants²⁵.

Drug prediction based on the transcriptomic signatures. The connectivity map (CMap) provides a way to find inverse relationship between disease signature and compound-based signature. In this context, the compounds could potentially present an opportunity to reverse the status of differential expressed genes in ASD population. On account of technical limitation of the online tool, we conducted CMap queries in 5 datasets. And we chose the compounds whose connectivity score ranged from -80 to -100 . One dataset (GSE18123-GPL570) was excluded to conduct subsequent analysis because it only had one compound meeting the condition. The compounds overlapping in four datasets were shown in Table 1. Whether these compounds could improve the symptoms of ASD needs to be further explored.

To assess the targets and pathways of the compounds obtained from CMap, we chose the compounds overlapping in more than three datasets as the list of interested drugs derived from human tissues, then used the DrugPattern to conduct enrichment analysis of these drugs. DrugPattern is an online tool for drug set enrichment analysis and contains 7019 drug sets, including indications, adverse reactions, targets, pathways, etc.²⁶. The results of DrugPattern were detailed in Supplementary Table S1. The significant drug targets that our interested drugs enriched were 30S ribosome protein, gamma-aminobutyric acid receptor, sodium channel protein, voltage-dependent calcium channel and etc.

Transcriptomic analysis of ASD animal models validated the results of human tissues. After screening in GEO database, we obtained 8 datasets of ASD animal models and healthy controls (Supplementary Table S2). One dataset was from *rattus norvegicus*, and other seven datasets were from *mus musculus*. All these samples in the datasets were derived from brain tissues. We defined differential expressed genes and performed functional enrichment analysis and made drug prediction by using the similar strategy and criteria which were

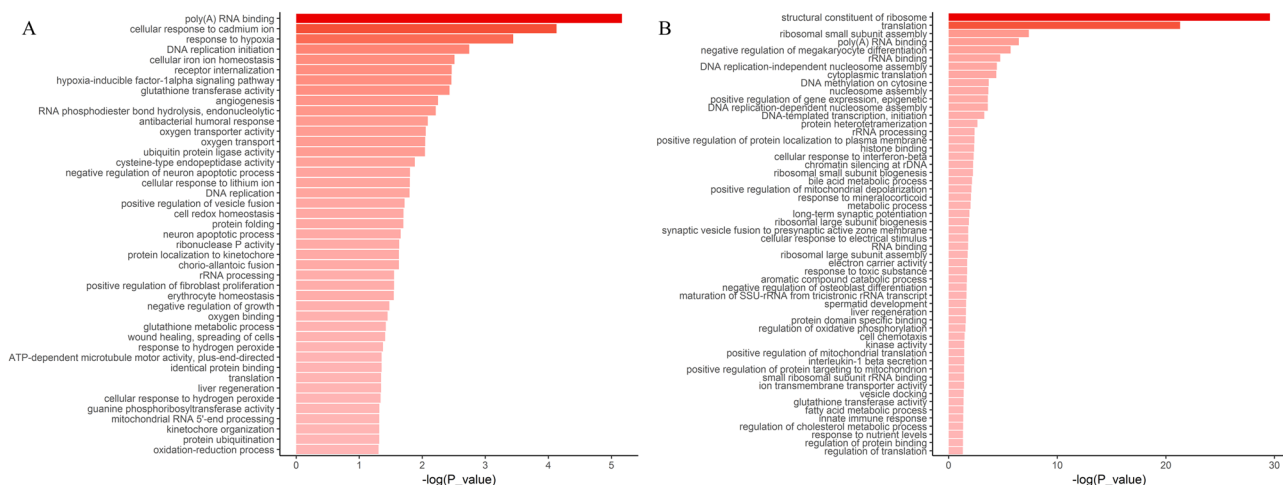


Figure 5. Functional enrichment analysis of differential expressed genes derived from animal models. Functional enrichment analysis were separately performed in differential expressed genes derived from rats and mice. P value of each GO term was < 0.05 . (A) *Rattus norvegicus*, (B) *Mus musculus*.

used in transcriptomic analysis of human tissues. Our results demonstrated all differential expressed genes derived from animal models were up-regulated (fold change > 1.5 , P value < 0.05 , Supplementary Table S2), and none of differential expressed genes was down-regulated. And we separately performed enrichment analysis for differential expressed genes of *mus musculus* and *rattus norvegicus*. One dataset (GSE77972) had only one differential expressed gene, so we performed functional enrichment analysis in other 7 datasets and chose statistically significant GO terms. Functional enrichment analysis demonstrated that genes were enriched in oxidative stress related pathway, inflammation/immune response, ribosome and protein translation (Fig. 5), which were consistent with the results of human tissues.

On account of technical limitation of the online tool, we conducted CMap queries in 6 datasets derived from animal models. After CMap query, the compounds whose connectivity score ranged from -80 to -100 overlapping in more than four datasets were defined as the list of interested drugs derived from animal models. There were 18 drugs overlapping in interested drug lists of human tissues and animal models (Supplementary Table S3). Finally, we performed enrichment analysis of interested drugs of animal models by DrugPattern. The results also suggested that there were several significant drug targets overlapping with the results of human tissues, namely epidermal growth factor receptor, sodium channel protein type5 and type9 subunit alpha, tubulin alpha-4A chain (Fig. 6, Supplementary Table S3).

To validate whether the similar results between human tissues and animal models were generated by random or on account of autism spectrum disorder, we separately performed ten thousand random experiments for enrichment analysis and drug prediction. Random experiment of enrichment analysis and drug prediction based on genes were similar to that based on functional annotations and drugs. We calculated the number of same GO terms derived from randomly enrichment analysis for human tissues and animal models based on the same criteria in our study in each random experiment. Random experiments suggested that the number of same GO terms was mostly zero, which was statistically different from our result ($P < 0.0001$). According to the total number of drugs in CMap database, each random experiment accounted the number of same drugs derived from randomly drug prediction for human tissues and animal models based on the same criteria in our study. Finally, our result that there were 18 drugs overlapping in interested drug lists of human tissues and animal models was statistically different from ten thousand random experiments ($P < 0.0001$) (Supplementary Fig. S3). In general, the results from animal models statistically validated the results of transcriptomic analysis of human tissues.

Discussion

As a complex disorder which has significant individual differences, a majority of researchers pay more attention to ASD. Mutations in a limited number of genes could not account for a large number of cases in ASD population. Changes in the shape of the entire gene expression distribution suggest alterations at global levels of gene expression regulation. Here we obtained 11 series of datasets from GEO database to explore the gene expression distribution derived from different human tissues. In the previous studies, differential expressed genes were defined according to different fold change. We screened the differential expressed genes based on the same criterion among datasets included in our study. We performed multidimensional scaling analysis among them and concluded that datasets of the same tissue had the bigger similarity compared with the datasets of different tissues. Subsequently, we reinforced the different role of transcriptional regulation between brain tissues and blood by demonstrating with enrichment analysis that, genes enriched in some significant pathways, like inflammation/immune response, were up-regulated in the brain tissues and down-regulated in blood. And we exploited two online tools, CMap and DrugPattern, to find several compounds which may reverse the expression pattern of ASD and explore the possible therapy of ASD. And we replicated the methods and criteria of transcriptomic analysis with animal models and healthy controls, the results were similar to that of human tissues.

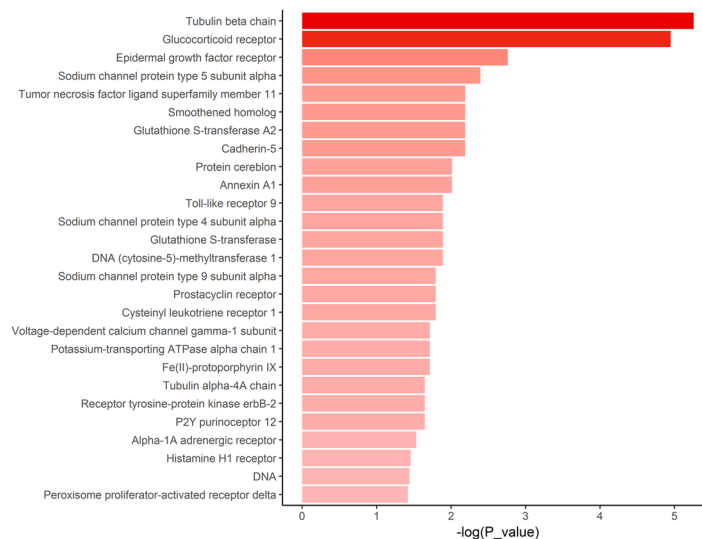


Figure 6. Same drug targets of interested drugs derived from animal models. The compounds overlapping in more than four datasets were defined as the list of interested drugs derived from animal models. Then they were inputted into DrugPattern to perform drug set enrichment. The shade of the color reflected the value of $-\log(P_value)$ of each term.

Alterations of the transcriptomic regulation could have tissue specificity. For instance, in our study, gene expression patterns in the brain and blood were apparently different. The number of up-regulated genes in the brain tissues was significantly higher than that in the blood. Multidimensional scaling analysis demonstrated datasets derived from brain tissues gathered together spatially. However, the postmortem brain tissues of ASD are scarce, the brain samples of three datasets were procured from the Harvard Brain Tissue Resource Center (www.brainbank.mclean.org). If there were any sample shared by different studies, it would contribute to the dataset similarity. In addition, animal experiment demonstrated that gene expression level might alter along with the fixation times and storage conditions²⁷. Blood samples were usually processed fresh, however, human brain tissues were obtained post-mortem. Since lack of the information of sample fixation in GEO database, whether the differences between brain and blood could be at least partially explained by tissue processing need to be further researched.

Immune dysregulation or inflammation, oxidative stress, mitochondrial dysfunction and environmental toxicant exposures are the four major areas implicated in ASD and other psychiatric disorders²⁸. A large percentage of publications implicated an association between ASD and these four major areas²⁸. Transcriptomic profiling may be particularly important in understanding the pathogenesis of diseases such as ASD where multiple systems are involved. Thus, it is possible that changes of transcriptomic regulation could offer a unifying understanding of multi-systemic effects of ASD. Our results were consistent with previous studies. Differentially expressed genes were enriched mainly in inflammation/immune response, mitochondrion-related functions and oxidative phosphorylation. Interestingly, genes enriched in inflammation/immune response were up-regulated in the brain tissues and down-regulated in the blood. ASD children are vulnerable to environmental factors, such as infection, stress or other toxicants exposure⁶. Corticotropin-releasing hormone (CRH) is secreted under stress, and it can stimulate mast cells and microglia along with neurotensin, which results in brain inflammation and neurotoxicity²⁹. Mast cells and microglia were found to be activated in brains of children with ASD^{30,31}. Increasing evidence indicated that mast cells activation was related to the disruption of blood-brain barrier^{29,32,33}. So, inflammation related cytokines may enter into blood across the blood-brain barrier. In fact, alterations in immune function have been reported in ASD patients, such as increased pro-inflammatory cytokine profiles in the cerebrospinal fluid and blood, elevated brain-specific auto-antibodies³⁴. Thus, we speculated that when inflammatory cytokines entered the blood, it would inhibit the related-genes expression, which may provide an explanation why genes enriched in inflammation/immune response were up-regulated in brain tissues but down-regulated in the blood.

Risperidone and aripiprazole are the only two drugs which have been approved by the US Food and Drug Administration for the treatment of ASD, however, these two drugs both target irritability rather than social deficits and repetitive behavior³⁵. In this work, we described eight overlapping compounds which have the reverse expression profiles of ASD patients, such as efavirenz, rilmenidine, PKCbeta-inhibitor, deferiprone, nifedipine and etc. Extended-release guanfacine, alpha2 agonists, has an effect on hyperactivity, impulsiveness and distractibility in children with ASD³⁶. Rilmenidine, a alpha2A-adrenoceptors agonist, has an effect on cardiovascular system in D79N alpha2A-adrenoceptor transgenic mice³⁷. Whether rilmenidine could improve some symptoms in children with ASD, like guanfacine, needs to be further explored. Moreover, other compounds also need to conduct some experiments to testify their clinical effects on ASD. In addition, DrugPattern provides us some important information based on the compounds overlapping in more than three datasets. These compounds targeted 30S ribosome protein, gamma-aminobutyric acid (GABA) receptor, sodium channel protein, voltage-dependent calcium channel and etc., which might be potential therapies of ASD. In fact, previous study has verified that

GEO accession number	Platform	Tissue	ASD:healthy control
GSE18123 ²⁰	GPL570	peripheral blood	104:82
GSE18123 ²⁰	GPL6244	peripheral blood	66:33
GSE25507 ¹³	GPL570	peripheral lymphocytes	82:64
GSE26415 ¹⁵	GPL6480	peripheral leucocytes	21:42
GSE28475 ⁴⁶	GPL6883	postmortem brain tissues	52:71
GSE28521 ¹²	GPL6883	postmortem brain tissues	39:40
GSE29691	GPL570	lymphoblastoid cell lines (LCLs)	2:13
GSE37772 ¹⁶	GPL6883	lymphoblast cell lines	233:206
GSE38322 ¹⁷	GPL10558	postmortem brain tissues	18:18
GSE42133 ^{19,47}	GPL10558	peripheral leucocytes	91:56
GSE65106 ¹⁴	GPL6244	skin fibroblast, iPSC, iPSC-derived neural progenitors, and iPSC-derived neurons	21:38

Table 2. Datasets from GEO database.

mutations in GABA_A receptor subunit was associated with epilepsy, autism and other neuropsychiatric disorders³⁸. Taken together, our results provided some information to explore the etiology and therapy of ASD.

Simultaneously, our study had some limitations. First, we didn't conduct false discovery rate (FDR) correction for multiple testing, because the number of differential expressed genes we obtained after FDR correction was too small to perform analysis. Thus we screened the differential expressed genes based on fold change and *P* value of Wilcoxon rank-sum test. Second, we can hardly perform further experimental study to explore the compounds found in our study and ASD on account of limited resources. Because of the complexity of ASD, it is important to take advantage of the bioinformatics methods to explore the pathogenesis and therapy of this disorder.

Methods and Materials

Transcriptomic datasets of ASD. We searched the transcriptomic datasets from GEO database based on the key words "autism", "autism spectrum disorder" and "ASD", then picked up the datasets by using the following criteria: (1) samples derived from human tissues, (2) samples in each dataset must include ASD and healthy controls, (3) gene expression datasets which are tested by DNA microarray must be on one channel. Finally, we acquired 11 series of transcriptomic datasets (Table 2).

The concept of differential expressed genes is the genes showing different expression level in human tissues between ASD patients and healthy controls. We used R software to perform our analysis. Specially, each gene's differential expression level can be evaluated by fold change and Wilcoxon rank-sum test. If we defined differential expressed genes according to FDR, the number of differential expressed genes was too small to perform the sequent analysis (Supplementary Table S4), so we chose the genes meeting the conditions that fold change >1.5 or <1/1.5 with *P* value < 0.05 as the differential expression genes. Calculated differential expressed genes in different datasets have been summarized in Supplementary Table S5.

Multidimensional scaling analysis of datasets. MDS is a method of quantitatively estimating the similarity among groups of items, and it refers to a set of statistical techniques that are used to reduce the dimensions of the data, so as to find out the visual appreciation of the underlying relational structures³⁹. The result of MDS is a "map" which spatially denotes the relationships among items, wherein similar items are located proximal to one another, and dissimilar items are located proportionately further apart³⁹. Given that, we performed MDS to quantify similarity judgements among datasets. First, we separately calculated fold change of all genes for each dataset. Then we carried out the spearman rank correlation analysis between each two datasets based on fold change. And we obtained the Euclidean distance between each two datasets according to correlation coefficients. Finally we chose 2 as the value of dimension and used distance matrix of these datasets to perform MDS by R3.4.4 software stats package.

Enrichment analysis of differential expressed genes. The enrichment analysis of interesting gene lists which are derived from microarray or next-generation RNA sequencing (RNA-seq) is a basic method to find significant information about the biological pathways⁴⁰. Compared with the analysis of all differential expressed genes together, enrichment analysis of up- and down-regulated genes separately was proven to be more powerful⁴¹. For each dataset in our study, we conducted enrichment analysis of up- and down-regulated gene lists respectively by the DAVID v6.8 (<https://david.ncicfcrf.gov/>)^{40,42}.

Drug prediction. The CMap database contains over 7000 expression profiles tested in five human cell lines, and utilizes gene expression profiling to reveal the associations among genes, chemicals and biological conditions such as disease^{43,44}. The next generation connectivity map termed as L1000 is more than a 1000-fold scale-up of the CMap (<https://clue.io/L1000>)⁴⁵. The set of differential expressed genes in our study can be used to query and compare against a large reference catalogue of gene expression profiles derived from drug or other perturbation treatment cell lines in the connectivity map. The CMap provides an online tool to perform CMap queries against the chemical reference catalogues. For the query CMap, up-regulated genes are essential, however, down-regulated genes are optional. The number of valid genes inputted into the query CMap should be limited to

10–150. So we chose the datasets whose number of valid up-regulated genes exceeded 10. If the valid up-regulated genes exceeded 150, we chose the top 150 valid genes according to fold change to conduct CMap query. Finally, we acquired 5 datasets meeting the conditions to perform CMap query (GSE18123-GPL570, GSE28475-GPL6883, GSE28521, GSE29691, GSE38322). The connectivity score ranges from +100 to –100. A positive connectivity score represents a positive correlation and a negative connectivity score denotes a negative correlation between our differential expressed gene list and a reference profile derived from an individual chemical perturbation. Thus a negative score may imply that exposure to a specific chemical may reverse the expression pattern of ASD. We chose the compounds whose connectivity score were between –80 to –100 in each dataset. The compounds overlapping in more than three datasets were inputted into DrugPattern (<http://www.cuilab.cn/drugpattern>) in order to mine regular rules and patterns behind a list of drugs²⁶.

Transcriptomic analysis of animal model. Increasing number of datasets will help to better understand the gene expression profiles of ASD, so we also analyzed transcriptomic data derived from ASD animal models and healthy controls. Similarly, we screened gene expression datasets from the GEO database according to the following criteria: (1) the samples are from animals, (2) there are ASD animal models and healthy controls in each dataset, (3) the number of ASD animal models and healthy controls both should be greater than four in each dataset. Finally, we obtained 8 series of datasets (Supplementary Table S2). Then, we identified differential expressed genes between ASD models and healthy controls according to fold change >1.5 or fold change <1/1.5 and *P* value < 0.05, and performed enrichment analysis of these differential expressed genes by using DAVID, and made drug prediction by using CMap and DrugPattern. The number of valid genes inputted into the query CMap should be also limited to 10–150. So we performed CMap query in 6 datasets (GSE34058, GSE62594, GSE63303, GSE77971, GSE99277, GSE117327). Similarly, we chose the compounds whose connectivity score were between –80 to –100 in each dataset. The compounds overlapping in more than four datasets were defined as the list of interested drugs derived from animal models. Then they were inputted into DrugPattern to perform drug set enrichment.

Data Availability

All data generated or analyzed during this study are included in this published article (and its supplementary information files).

References

- Lai, M.-C., Lombardo, M. V. & Baron-Cohen, S. Autism. *The Lancet* **383**, 896–910, [https://doi.org/10.1016/s0140-6736\(13\)61539-1](https://doi.org/10.1016/s0140-6736(13)61539-1) (2014).
- Chaste, P. & Leboyer, M. Autism risk factors: genes, environment, and gene-environment interactions. *Dialogues in clinical neuroscience* **14**, 281–292 (2012).
- Tordjman, S. *et al.* Gene × Environment Interactions in Autism Spectrum Disorders: Role of Epigenetic Mechanisms. *Frontiers in Psychiatry* **5**, <https://doi.org/10.3389/fpsy.2014.00053> (2014).
- Yuen, R. K. C. *et al.* Whole-genome sequencing of quartet families with autism spectrum disorder. *Nature Medicine* **21**, 185, <https://doi.org/10.1038/nm.3792> (2015).
- Park, H. R. *et al.* A Short Review on the Current Understanding of Autism Spectrum Disorders. *Exp Neurobiol* **25**, 1–13 (2016).
- Vijayakumar, N. T. & Judy, M. V. Autism spectrum disorders: Integration of the genome, transcriptome and the environment. *Journal of the Neurological Sciences* **364**, 167–176, <https://doi.org/10.1016/j.jns.2016.03.026> (2016).
- Betancur, C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res* **1380**, 42–77, <https://doi.org/10.1016/j.brainres.2010.11.078> (2011).
- Jamain, S. *et al.* Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nature Genetics* **34**, 27–29, <https://doi.org/10.1038/ng1136> (2003).
- Moessner, R. *et al.* Contribution of SHANK3 mutations to autism spectrum disorder. *Am J Hum Genet* **81**, 1289–1297, <https://doi.org/10.1086/522590> (2007).
- Gauthier, J. *et al.* Novel de novo SHANK3 mutation in autistic patients. *Am J Med Genet B Neuropsychiatr Genet* **150B**, 421–424, <https://doi.org/10.1002/ajmg.b.30822> (2009).
- Miles, J. H. Autism spectrum disorders—a genetics review. *Genet Med* **13**, 278–294, <https://doi.org/10.1097/GIM.0b013e3181ff67ba> (2011).
- Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384, <https://doi.org/10.1038/nature10110> (2011).
- Alter, M. D. *et al.* Autism and increased paternal age related changes in global levels of gene expression regulation. *PLoS One* **6**, e16715, <https://doi.org/10.1371/journal.pone.0016715> (2011).
- Liu, X. *et al.* Idiopathic Autism: Cellular and Molecular Phenotypes in Pluripotent Stem Cell-Derived Neurons. *Molecular Neurobiology* **54**, 4507–4523, <https://doi.org/10.1007/s12035-016-9961-8> (2017).
- Kuwano, Y. *et al.* Autism-associated gene expression in peripheral leucocytes commonly observed between subjects with autism and healthy women having autistic children. *PLoS One* **6**, e24723, <https://doi.org/10.1371/journal.pone.0024723> (2011).
- Luo, R. *et al.* Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in autism spectrum disorders. *Am J Hum Genet* **91**, 38–55, <https://doi.org/10.1016/j.ajhg.2012.05.011> (2012).
- Ginsberg, M. R., Rubin, R. A., Falcone, T., Ting, A. H. & Natowicz, M. R. Brain transcriptional and epigenetic associations with autism. *PLoS One* **7**, e44736, <https://doi.org/10.1371/journal.pone.0044736> (2012).
- Clough, E. & Barrett, T. The Gene Expression Omnibus Database. *Methods in molecular biology (Clifton, N.J.)* **1418**, 93–110, https://doi.org/10.1007/978-1-4939-3578-9_5 (2016).
- Pramparo, T. *et al.* Prediction of autism by translation and immune/inflammation coexpressed genes in toddlers from pediatric community practices. *JAMA Psychiatry* **72**, 386–394, <https://doi.org/10.1001/jamapsychiatry.2014.3008> (2015).
- Kong, S. W. *et al.* Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. *PLoS One* **7**, e49475, <https://doi.org/10.1371/journal.pone.0049475> (2012).
- Hagberg, H., Gressens, P. & Mallard, C. Inflammation during fetal and neonatal life: implications for neurologic and neuropsychiatric disease in children and adults. *Ann Neurol* **71**, 444–457, <https://doi.org/10.1002/ana.22620> (2012).
- Estes, M. L. & McAllister, A. K. Immune mediators in the brain and peripheral tissues in autism spectrum disorder. *Nat Rev Neurosci* **16**, 469–486, <https://doi.org/10.1038/nrn3978> (2015).
- Graf, W. D. *et al.* Autism Associated With the Mitochondrial DNA G8363A Transfer RNALys Mutation. *Journal of Child Neurology* **15**, 357–361, <https://doi.org/10.1177/088307380001500601> (2000).

24. Rossignol, D. A. & Frye, R. E. Mitochondrial dysfunction in autism spectrum disorders: a systematic review and meta-analysis. *Mol Psychiatry* **17**, 290–314, <https://doi.org/10.1038/mp.2010.136> (2012).
25. Rogozin, I. B. *et al.* Genome-Wide Changes in Protein Translation Efficiency Are Associated with Autism. *Genome Biology and Evolution* **10**, 1902–1919, <https://doi.org/10.1093/gbe/evy146> (2018).
26. Huang, C. *et al.* The DrugPattern tool for drug set enrichment analysis and its prediction for beneficial effects of oxLDL on type 2 diabetes. *J Genet Genomics* **45**, 389–397, <https://doi.org/10.1016/j.jgg.2018.07.002> (2018).
27. Sobue, S. *et al.* Characterization of gene expression profiling of mouse tissues obtained during the postmortem interval. *Exp Mol Pathol* **100**, 482–492, <https://doi.org/10.1016/j.yexmp.2016.05.007> (2016).
28. Rossignol, D. A. & Frye, R. E. A review of research trends in physiological abnormalities in autism spectrum disorders: immune dysregulation, inflammation, oxidative stress, mitochondrial dysfunction and environmental toxicant exposures. *Mol Psychiatry* **17**, 389–401, <https://doi.org/10.1038/mp.2011.165> (2012).
29. Theoharides, T. C., Asadi, S. & Patel, A. B. Focal brain inflammation and autism. *J Neuroinflammation* **10**, 46, <https://doi.org/10.1186/1742-2094-10-46> (2013).
30. Theoharides, T. C. Is a subtype of autism an allergy of the brain? *Clin Ther* **35**, 584–591, <https://doi.org/10.1016/j.clinthera.2013.04.009> (2013).
31. Rodriguez, J. I. & Kern, J. K. Evidence of microglial activation in autism and its possible role in brain underconnectivity. *Neuron Glia Biol* **7**, 205–213, <https://doi.org/10.1017/S1740925X12000142> (2011).
32. Ribatti, D. The crucial role of mast cells in blood-brain barrier alterations. *Exp Cell Res* **338**, 119–125, <https://doi.org/10.1016/j.yexcr.2015.05.013> (2015).
33. McKittrick, C. M., Lawrence, C. E. & Carswell, H. V. Mast cells promote blood brain barrier breakdown and neutrophil infiltration in a mouse model of focal cerebral ischemia. *J Cereb Blood Flow Metab* **35**, 638–647, <https://doi.org/10.1038/jcbfm.2014.239> (2015).
34. Onore, C., Careaga, M. & Ashwood, P. The role of immune dysfunction in the pathophysiology of autism. *Brain Behav Immun* **26**, 383–392, <https://doi.org/10.1016/j.bbi.2011.08.007> (2012).
35. Bent, S. & Hendren, R. L. Improving the prediction of response to therapy in autism. *Neurotherapeutics* **7**, 232–240, <https://doi.org/10.1016/j.nurt.2010.05.011> (2010).
36. Scahill, L. *et al.* Extended-Release Guanfacine for Hyperactivity in Children With Autism Spectrum Disorder. *American Journal of Psychiatry* **172**, 1197–1206, <https://doi.org/10.1176/appi.ajp.2015.15010055> (2015).
37. Zhu, Q. M. *et al.* Cardiovascular effects of rilmenidine, moxonidine and clonidine in conscious wild-type and D79N alpha2A-adrenoceptor transgenic mice. *British journal of pharmacology* **126**, 1522–1530, <https://doi.org/10.1038/sj.bjp.0702429> (1999).
38. Kang, J. Q. & Barnes, G. A common susceptibility factor of both autism and epilepsy: functional deficiency of GABA A receptors. *J Autism Dev Disord* **43**, 68–79, <https://doi.org/10.1007/s10803-012-1543-7> (2013).
39. Hout, M. C., Papesch, M. H. & Goldinger, S. D. Multidimensional scaling. *Wiley Interdiscip Rev. Cogn Sci* **4**, 93–103, <https://doi.org/10.1002/wcs.1203> (2013).
40. Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1–13, <https://doi.org/10.1093/nar/gkn923> (2009).
41. Hong, G., Zhang, W., Li, H., Shen, X. & Guo, Z. Separate enrichment analysis of pathways for up- and downregulated genes. *J R Soc Interface* **11**, 20130950, <https://doi.org/10.1098/rsif.2013.0950> (2014).
42. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44, <https://doi.org/10.1038/nprot.2008.211> (2008).
43. Lamb, J. *et al.* The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease. *Science* **313**, 1929, <https://doi.org/10.1126/science.1132939> (2006).
44. Qu, X. A. & Rajpal, D. K. Applications of Connectivity Map in drug discovery and development. *Drug Discovery Today* **17**, 1289–1298, <https://doi.org/10.1016/j.drudis.2012.07.017> (2012).
45. Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437–1452, <https://doi.org/10.1016/j.cell.2017.10.049> (2017).
46. Chow, M. *et al.* Preprocessing and Quality Control Strategies for Illumina DASL Assay-Based Brain Gene Expression Studies with Semi-Degraded Samples. *Frontiers in Genetics* **3**, <https://doi.org/10.3389/fgene.2012.00011> (2012).
47. Pramparo, T. *et al.* Cell cycle networks link gene expression dysregulation, mutation, and brain maldevelopment in autistic toddlers. *Mol Syst Biol* **11**, 841, <https://doi.org/10.15252/msb.20156108> (2015).

Acknowledgements

We thank Prof. Qinghua Cui for valuable comments and suggestions.

Author Contributions

H.Y. performed all data analysis and drafted the manuscript under the supervision of W.J., M.W. and Z.Y. assisted H.Y. in performing drug prediction. W.J. participated in the writing and editing of this manuscript. All the authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-48160-x>.

Competing Interests: The authors declare no competing interests.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019