Weak Association between Autism Spectrum Disorder and Two Genes YBX3 and HSPA1A – A Meta-Analysis

Cheng Xu[1], Fuquan Zhang[2], Rachel Amey[3*], Yin Yao[4*]

[1] Department of Magnetic Resonance Imaging, Shanxi Province People's Hospital, Taiyuan, 030001, People's Republic of China;

[2] Wuxi Mental Health Center, Nanjing Medical University, Wuxi, 214151, China;

[3] Department of Psychological and Brain Sciences, University of Delaware, Newark, De;

[4] Unit on Statistical Genomics, Intramural Research Programs, National Institute of Mental Health, National Institutes of Health, NIH, Bethesda, MD.

***Corresponding to:** Yin Yao. Email: yinyao21043@gmail.com. Rachel Amey, Email: ramey@psych.udel.edu.

ABSTRACT

Background: Recently, HSPA1A and YBX3 have been suggested as influential genes for Autism Spectrum Disorder (ASD). However, the results have shown a lack of replication.

Method: Here we performed a systematic review and a meta-analysis to explore the gene expression activity of HSPA1A and YBX3 in ASD. Fifteen RNA expression datasets of ASD case vs. control studies were acquired from Gene Expression Omnibus, including 1,534 samples (830 ASD cases, and 704 healthy controls). Both fixed-effect and random-effects models were employed for meta-analysis. Multiple linear regression (MLR) were employed to study the possible influence of the following factors: tissue, sample size, sex and racial region.

Results: Utilizing all 15 datasets, both genes presented no significant log fold change (LFC) in ASD: YBX3 (LFC = 0.001; p-value = 0.49); HSPA1A (LFC = 0.115; p-value = 0.16), but weak effect sizes for the 3 brain tissue sampled datasets: YBX3 (LOR = 0.89; p-value = 0.11); HSPA1A (LOR = 0.91; p-value = 0.086). Sample tissue is a major factor for both genes (p-value < 1.5e-04) and sex is for HSPA1A only (p-value < 0.01). Literature based pathway analysis show that YBX3 and HSPA1A may play inhibitory roles again ASD.

Conclusion: Our results suggest that YBX3 and HSPA1A may present weak protective regulation within the brain of ASD patients.

Keywords: Autism Spectrum Disorder; RNA expression data; Random model; Multiple linear regression model

1 INTRODUCTION

Autism Spectrum Disorder (ASD) describes a range of conditions classified as neurodevelopmental disorders with high heritability [1-3]. Both family and twin studies suggest that ASDs are highly heritable neuropsychiatric disorders [4-5]. Although the specific causes of ASD have yet to be found, it is believed that the genetic or environmental risk factors perturb common underlying molecular pathways in the brain that may contribute to the ASD pathogenesis [6].

Data from brain regions and blood at the gene expression levels were well-employed in efforts to identify ASD genetic determinants [7-9]. Recently, two novel genes, YBX3 and HSPA1A, were proposed to play significant roles in the pathogenesis of ASD [10]. The study was conducted using gene expression and large scale literature data, with pathway enrichment analysis (PEA) and network connectivity analysis (NCA) conducted to analyze the potential functional association between these two genes and ASD. However, the results were lack of replication, raising the necessity of further validation using larger datasets.

Due to limited sample sizes and specifications in terms of phenotype characteristics, results from independent studies may engender bias in their conclusions. Because meta-analyses are often used to explore the validity of findings, we propose a meta-analysis of these genes to provide insight into their effects on ASD with higher statistical reliability. Moreover, we utilize multiple linear regression (MLR) to test potential influential factors of ASD.

2 METHODS AND MATERIALS

2.1 Data screening

A systematic search was conducted by using Illumine BaseSpace Correlation Engine (http://www.illumina.com) and public functional genomics database--Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/). The initial search with keywords set as 'Autism Spectrum Disorders'. Further filter criteria includes: 1) The data organism is Homo sapien; 2) The data type is RNA expression by array or by sequencing; and 3) The study design is limited to ASD vs. healthy control cases. To note, there is no date limitation. In total, 15 studies satisfied the selection criteria and were included in this meta-analysis [11-22]. The total samples included in this study was 1,534, with 830 ASD cases and 704 controls, as shown in Table 1. The data age (current year - collection date) was from 0 to 10 years (2007 to 2017).

Table 1. Fifteen studies satisfied the selection criteria and were included in meta-analysis

Study Name	GEO ID	Sample Organism	Data Type	Tissue Name	N Cases	N Controls	Country
Hu *et al* 2009B [11]	GSE15402	Homo sapiens	RNA expression	lymphoblastoid	87	29	USA
Hu *et al* 2009 [12]	GSE15451	Homo sapiens	RNA expression	lymphoblastoid	21	17	USA
Alter *et al* 2011 [13]	GSE25507	Homo sapiens	RNA expression	blood	82	64	USA
Kuwano *et al* 2011 [14]	GSE26415	Homo sapiens	RNA expression	blood	21	63	Japan
Chow *et al* 2012 [9]	GSE28475	Homo sapiens	RNA expression	brain tissue	52	61	USA
Luo *et al* 2012 [16]	GSE37772	Homo sapiens	RNA expression	blood	233	206	USA

Ginsberg *et al* 2012 [7]	GSE38322	Homo sapiens	RNA expression	cerebellum	18	18	USA
Pramparo *et al* 2015 [17]	GSE42133	Homo sapiens	RNA expression	leukocyte	91	56	USA
Voineagu *et al* 2011 [8]	GSE28521	Homo sapiens	RNA expression	postmortem brain tissue representing Cerebellum (C), Frontal cortex (F), and Temporal cortex (T)	39	40	USA
Liu *et al* 2017 [18]	GSE65106	Homo sapiens	RNA expression	induced pluripotent stem cells (iPSCs)	21	38	USA
Gregg *et al* 2008 [19]	GSE6575	Homo sapiens	RNA expression	blood	35	12	USA
Nishimura *et al* 2007 [20]	GSE7329	Homo sapiens	RNA expression	blood	15	15	USA
Inoue *et al* 2017	GSE77103	Homo sapiens	RNA expression	blood	4	4	Japan
Walker *et al* 2016 [21]	GSE87847	Homo sapiens	RNA expression	blood	45	48	USA
Kong *et al* 2012 [22]	GSE18123	Homo sapiens	RNA expression	blood	66	33	USA

2.2 Meta-analysis models

Both fixed-effect model and random-effects model were employed to study the effect size of the two genes, YBX3 and HSPA1A, in a case vs. control expression comparison. The expression log fold change (LFC) were used as an index of effect size. Results from both models were reported and compared. The heterogeneity of the meta-analysis was analyzed to study the variance within and between different studies.

2.3 Multiple linear regression analysis

A MLR model was employed to study the possible influence of four factors on the gene expression change in ASD: tissue, sample size, sex and population. P-values and 95 % confidence interval (CI) were reported for each of the four factors.

3 RESULTS

3.1 Meta-analysis results

Table 2 and Table 3 present the effect sizes and related statistics from the fifteen studies and the meta-analysis results for gene HSPA1A and YBX3, respectively.

Table 2. The effects of two models for HSPA1A

| Study Name	Effect Size	Lower Limit of 95% CI	Upper Limit of 95% CI	Z_Value	pValue (one tailed)	pValue (two tailed)	Weight_ Fixed	Weight_ Random
Inoue *et al* 2017	-0.20	-2.72	2.39	-0.10	0.55	0.9	0.59	0.59
Liu *et al* 2017[18]	0.15	-1.37	1.68	0.20	0.42	0.84	1.65	1.65
Walker *et al* 2016[21]	0.30	-2.02	2.62	0.25	0.40	0.80	0.72	0.72
Pramparo *et al* 2015[17]	-0.02	-0.61	0.56	-0.10	0.53	0.94	11.2	11.2
Chow *et al* 2012[9]	2.00	-4.19	8.19	0.63	0.26	0.53	0.10	0.10
Ginsberg *et al* 2012[7]	1.28	-0.73	3.29	1.25	0.11	0.21	0.95	0.95
Kong *et al* 2012[22]	0.12	-0.55	0.78	0.34	0.37	0.73	8.59	8.59
Luo *et al* 2012[16]	0.11	-2.29	2.51	0.09	0.46	0.93	0.67	0.67
Alter *et al* 2011[13]	0.05	-0.82	0.92	0.11	0.45	0.91	5.03	5.03
Kuwano *et al* 2011[14]	0.09	-1.24	1.42	0.13	0.45	0.90	2.17	2.17
Voineagu *et al* 2011[8]	0.53	-1.26	2.32	0.58	0.28	0.56	1.20	1.20
Gregg *et al* 2008[19]	0.22	-0.62	1.06	0.51	0.30	0.61	5.43	5.43
Nishimura *et al* 2007[20]	0.10	-0.21	0.42	0.64	0.26	0.52	39.3	39.3
Fixed Model	0.11	-0.11	0.34	1.01	0.16	0.31		
Random Model	0.11	-0.11	0.34	1.01	0.16	0.31		

Table 3. The effects of two models for YBX3

Study Name	Effect Size	Lower Limit of 95% CI	Upper Limit of 95% CI	Z_Value	pValue (one tailed)	pValue (two tailed)	Weight_ Fixed	Weight_ Random
Liu *et al* 2017 [18]	-0.10	-0.62	0.52	-0.20	0.57	0.85	11.80	11.80
Pramparo *et al* 2015 [17]	0.05	-0.56	0.65	0.16	0.44	0.88	10.50	10.52
Chow *et al* 2012 [9]	1.55	-3.86	6.95	0.56	0.29	0.58	0.13	0.13
Ginsberg *et al* 2012 [7]	0.94	-1.06	2.94	0.92	0.18	0.36	0.96	0.96
Kong *et al* 2012 [22]	-0.1	-0.87	0.66	-0.3	0.61	0.78	6.61	6.61
Luo *et al* 2012 [16]	0.01	-0.37	0.39	0.04	0.49	0.97	26.5	26.51
Alter *et al* 2011 [13]	-0.02	-0.73	0.69	-0.1	0.52	0.95	7.68	7.68
Kuwano *et al* 2011 [14]	0.31	-0.54	1.16	0.72	0.23	0.47	5.31	5.31
Voineagu *et al* 2011 [8]	0.71	-1.52	2.94	0.63	0.27	0.53	0.77	0.77
Gregg *et al* 2008 [19]	-0.10	-0.8	0.65	-0.2	0.58	0.84	7.39	7.39
Nishimura *et al* 2007 [20]	-0.02	-0.23	0.19	-0.2	0.58	0.85	86.2	86.18
Fixed Model	0.001	-0.15	0.15	0.01	0.49	0.99		
Random Model	0.001	-0.15	0.15	0.01	0.49	0.99		

For HSPA1A, 13 out of 15 studies included the gene and were employed for the meta-analysis. Only 11 studies included gene YBX3 and were used for its meta-analysis. We noted that the weights from the random-effects model and the fix-effect model are the same for both genes (See Table 2 and Table 3, columns 'Weight_Fixed' and 'Weight_Random'). This was due to the fact that the between study variance Tau-squared, , was calculated as 0 for both genes, indicating no significant between-study variance. The total variance, Q, for the two genes were 2.22 and 2.30, with the expected variance df (under the assumption that all studies have the same effect size) of 12 and 10, respectively. This leads to p-values of 1.00 and 0.99 for the hypothesis that Q was from within-study variances only. Therefore, the following analyses will focus on results from the fix-effect model.

For HSPA1A, the effect size (LFC) from the meta-analysis was 0.11 (95 % CI: [-0.11, 0.34], *p* = 0.16). For YBX3, the effect size (LFC) from the meta-analysis was 0.001 (95 % CI: [-0.15, 0.15], *p* = 0.49). These results suggest that both genes present no significant gene expression fold change in ASD.
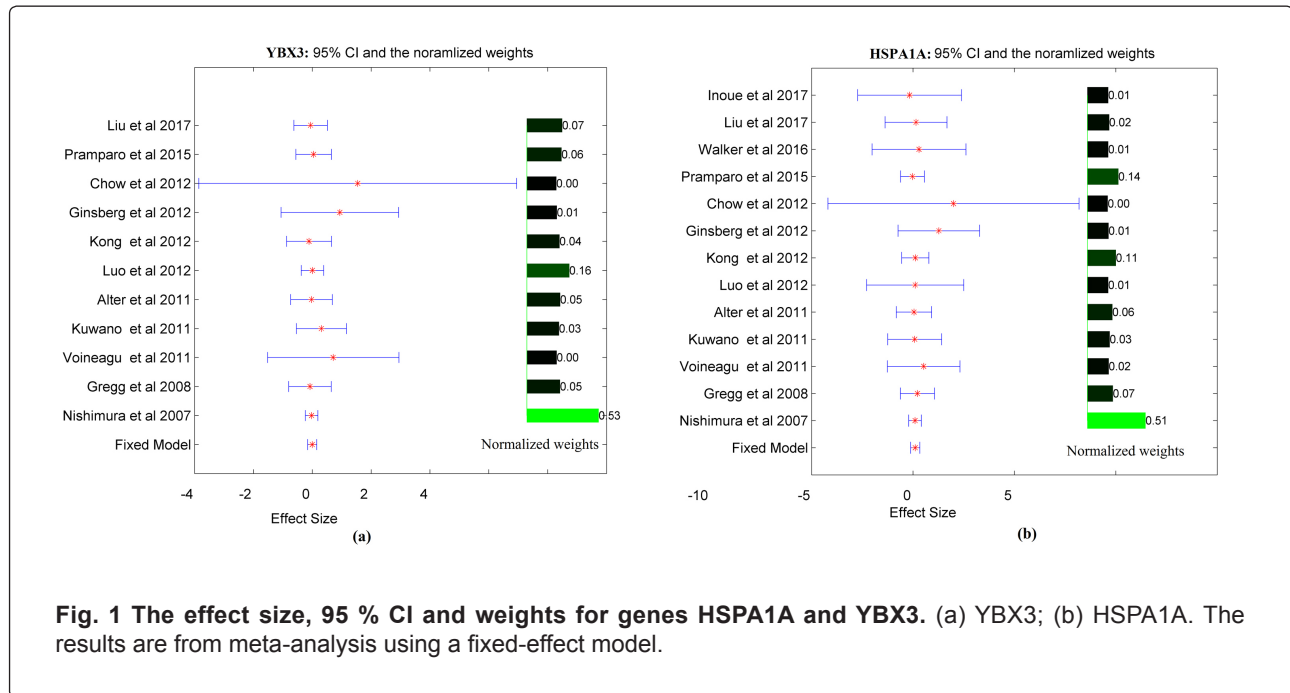
Fig. 1 The effect size, 95 % CI and weights for genes HSPA1A and YBX3. (a) YBX3; (b) HSPA1A. The results are from meta-analysis using a fixed-effect model.

3.2 MLR analysis results

As shown in Table 4, results from the MLR models showed that tissue was a significant influential factor for the expression fold change of both HSPA1A and YBX3 (p = 1.5E-4 and 1.0E-5, respectively) in regards to ASD. Sex was a significant factor for HSPA1A (p < 0.01) but not for YBX3 (p = 0.33). The case / control ratio and Country were not significant influential factors.

Table 4. Multiple linear regression analysis results

Gene Name	MLR parameter	Sex	Brain Tissue	Case/Control Ratio	Country
HSPA1A	Beta	1.043	1.12	-0.04	0.20
	Lower Limit	-0.91	0.53	-0.44	-0.52
	Upper Limit	3.00	1.71	0.37	0.91
	p-value	0.01	1.5 E-4	0.60	0.23
YBX3	Beta	0.17	1.074	-0.04	0.30
	Lower Limit	-8.08	0.66	-0.33	-0.38
	Upper Limit	8.42	1.49	0.25	0.98
	p-value	0.33	1.0E-5	0.676	0.102

Fig. 2 presents the linear regression relation (LRR) between the parameter 'Brain Tissue' and the expression log fold change for the two genes: HSPA1A and YBX3. From Fig. 2, it can be seen that the expression LFC of both genes were much higher in brain tissue than non-brain tissue.

Fig. 2 Scatter plot of the linear regression relationship between 'Brain Tissue' samples and the expression log fold change for the two genes. (a) HSPA1A; (b) YBX3.
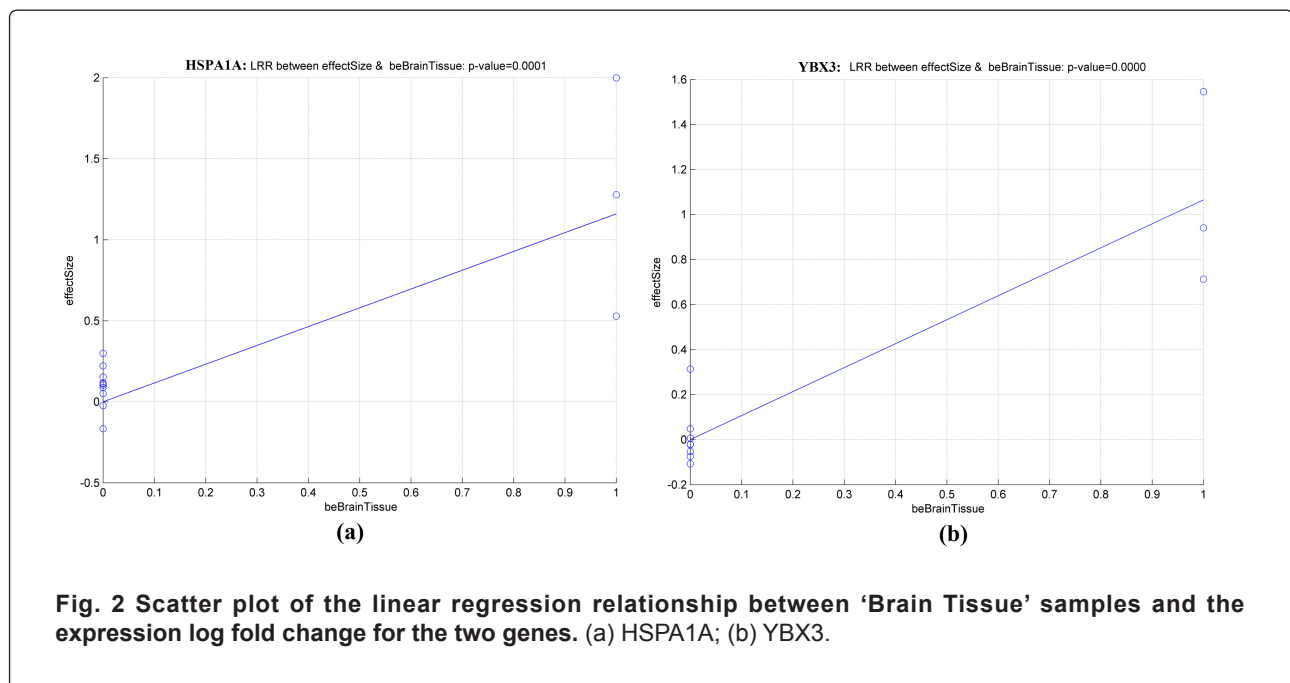
3.3 Meta-analysis results using brain-tissue samples

Considering the significant difference between expression data acquired using brain-tissue samples and non-brain tissue samples, we conducted a meta-analysis using only brain-tissue samples. Results showed that, for HSPA1A, an effect size of 0.91 (95 % CI: [-0.40, 2.22], *p* = 0.09) was acquired. While for gene YBX3, the effect size was 0.89 (95 % CI: [-0.55, 2.32], *p* = 0.11). These results indicated an existing but weak expression variation of these two genes in the brain regions of ASD patients.

4 DISCUSSION

This study conducted an RNA expression data based meta-analysis to explore the gene expression activity of two genes (HSPA1A and YBX3) in case of ASD. Both fixed-effect model and random-effects model were employed with heterogeneity analysis conducted. Furthermore, MLR analysis was performed to study the influence of four potential factors, tissue, sample size, sex and population, towards the effect size of these two genes in ASD.

Fifteen studies were acquired for this study,

including 1,534 from human (830 ASD cases, and 704 controls), as shown in Table 1. Among these 15 expression datasets, 3 were acquired from brain tissue, and the other 12 were acquired from non-brain tissue. Using all available data for a meta-analysis, no significant fold change was observed for neither HSPA1A nor YBX3 (p-value = 0.49 and 0.16, respectively). Results were presented from fixed-effect model, as no significant between-study variance was observed in the heterogeneity analysis (HSPA1A: Q = 2.22; p-value = 1.00; YBX3: Q = 2.30; p-value = 0.99).

MLR results revealed that gene expressions sampled from brain-tissue and non-brain tissue were significantly different, as shown in Fig. 2. For both genes, the p-values were less than 0.0001 for the 'Brain Tissue' parameter. Considering ASD is a brain-dysfunction disease, we conducted a further meta-analysis using brain-tissue sampled data only. Results showed weak effective size for the two genes in case of ASD: (p-value = 0.086 and 0.11 for HSPA1A and YBX3, respectively). These results suggested that, in about 91.4 % of the total ASD cases, the expression of HSPA1A in the brain will present increased fold change averaged as 2^0.91 = 1.88. While for YBX3, about 89 % ASD cases will present increased expression levels averaged around 2^0.89 = 1.85 fold. These results suggested

weak but possible association between ASD and HSPA1A and YBX3.

As a matter of fact, previous studies suggested that HSPA1A can activate nitric oxide (NO) syntheses [23]. Increases in NO production increases the risks of developing ASD [24]. This indicates that HSPA1A could play an inhibitive role for the development of ASD. On the other hand, it has been suggested that increased cytokine can impact neural activity and mediate the behavioral symptoms of ASD [25], while YBX3 has been shown to directly bind the nuclear factor, κB-binding element, to repress the activation of several cytokines [26]. This suggest that increased YBX3 may work as a protector against the development of ASD through a cytokines regulation channel.

5 CONCLUSION

The results of this study demonstrate that genes HSPA1A and YBX3 may act as protecting factors against the development of ASD, although the influential effect may result from a indirect weak regulation. The altered gene expression changes of HSPA1A and YBX3 is more likes happen in the brain tissues than non-brain tissues in case of ASD.

CONFLICT OF INTERESTS

Authors claim no conflict of interests.

REFERENCES

1. Betancur C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. Brain Res. 2011; 1380: 42-77.

2. Szatmari P, Jones MB, Zwaigenbaum L,MacLean JE. Genetics of autism: Overview and new directions. J Autism Dev Disord. 1998; 28: 51-68.

3. Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, Fauchereau F, Nygren G, Rastam M, Gillberg IC, Anckarsäter H, Sponheim E, Goubran-Botros H, Delorme R, Chabane N, Mouren-Simeoni MC, de Mas P, Bieth E, Rogé B, Héron D, Burglen L, Gillberg C, Leboyer M, Bourgeron T. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. Nat Genet. 2007; 39: 25-27.

4. Bolton PF, Pickles A, Murphy M, Rutter M. Autism, affective and other psychiatric disorders: Patterns of familial aggregation. Psychol Med. 1998; 28: 385-395.

5. Ronald A, Hoekstra RA. Autism spectrum disorders and autistic traits: A decade of new twin studies. Am J Med Genet B Neuropsychiatr Genet. 2011; 156B: 255-274.

6. Geschwind DH. Autism: many genes, common pathways? Cell. 2008; 135: 391-395.

7. Ginsberg MR, Rubin RA, Falcone T, Ting AH, Natowicz MR. Brain transcriptional and epigenetic associations with autism. PLoS One. 2012; 7(9): e44736.

8. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, Mill J, Cantor RM, Blencowe BJ, Geschwind DH. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature. 2011; 474(7351): 380-384.

9. Chow ML, Winn ME, Li HR, April C, Wynshaw-Boris A, Fan JB, Fu XD, Courchesne E, Schork NJ. Preprocessing and Quality Control Strategies for Illumina DASL Assay-Based Brain Gene Expression Studies with Semi-Degraded Samples. Front Genet. 2012; 3: 11.

10. Dongbai Liu, Zeya Wang. Identification and Validation Novel Risk Genes for Autism Spectrum Disorder – A Meta-Analysis. JPBS. 2017; 2(1): 2.

11. Hu VW, Sarachana T, Kim KS, Nguyen A, Kulkarni S, Steinberg ME, Luu T, Lai Y, Lee NH. Gene expression profiling differentiates autism case-controls and phenotypic variants of autism spectrum disorders: evidence for circadian rhythm dysfunction in severe autism. Autism Res. 2009; 2(2): 78-97.

12. Hu VW, Nguyen A, Kim KS, Steinberg ME, Sarachana T, Scully MA, Soldin SJ, Luu T, Lee NH. Gene expression profiling of lymphoblasts from autistic and nonaffected sib pairs: altered pathways in neuronal development and steroid biosynthesis. PLoS One. 2009; 4(6): e5775.

13. Alter MD, Kharkar R, Ramsey KE, Craig DW, Melmed RD, Grebe TA, Bay RC, Ober-Reynolds S, Kirwan J, Jones JJ, Turner JB, Hen R, Stephan DA. Autism and increased paternal age related changes in global levels of gene expression regulation. PLoS One. 2011; 6(2): e16715.

14. Kuwano Y, Kamio Y, Kawai T, Katsuura S, Inada

N, Takaki A, Rokutan K. Autism-associated gene expression in peripheral leucocytes commonly observed between subjects with autism and healthy women having autistic children. PLoS One. 2011; 6(9): e24723.

15. Chow ML, Winn ME, Li HR, April C, Wynshaw-Boris A, Fan JB, Fu XD, Courchesne E, Schork NJ. Preprocessing and Quality Control Strategies for Illumina DASL Assay-Based Brain Gene Expression Studies with Semi-Degraded Samples. Front Genet. 2012; 3: 11.

16. Luo R, Sanders SJ, Tian Y, Voineagu I, Huang N, Chu SH, Klei L, Cai C, Ou J, Lowe JK, Hurles ME, Devlin B, State MW, Geschwind DH. Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in autism spectrum disorders. Am J Hum Genet. 2012; 91(1): 38-55.

17. Pramparo T, Pierce K, Lombardo MV, Carter Barnes C, Marinero S, Ahrens-Barbeau C, Murray SS3, Lopez L1, Xu R4, Courchesne E. Prediction of autism by translation and immune/inflammation coexpressed genes in toddlers from pediatric community practices. JAMA Psychiatry. 2015; 72(4): 386-394.

18. Liu X, Campanac E, Cheung HH, Ziats MN, Canterel-Thouennon L, Raygada M, Baxendale V, Pang AL, Yang L, Swedo S, Thurm A, Lee TL, Fung KP, Chan WY, Hoffman DA, Rennert OM. Idiopathic Autism: Cellular and Molecular Phenotypes in Pluripotent Stem Cell-Derived Neurons. Mol Neurobiol. 2017; 54(6): 4507-4523.

19. Gregg JP, Lit L, Baron CA, Hertz-Picciotto I, Walker W, Davis RA, Croen LA, Ozonoff S, Hansen R, Pessah IN, Sharp FR. Gene expression changes in children with autism. Genomics. 2008; 91(1): 22-29.

20. Nishimura Y, Martin CL, Vazquez-Lopez A, Spence SJ, Alvarez-Retuerto AI, Sigman M, Steindler C, Pellegrini S, Schanen NC, Warren ST, Geschwind DH. Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways. Hum Mol Genet. 2007; 16(14): 1682-1698.

21. Walker SJ, Beavers DP, Fortunato J, Krigsman A. A Putative Blood-Based Biomarker for Autism Spectrum Disorder-Associated Ileocolitis. Sci Rep. 2016; 6: 35820.

22. Kong SW, Collins CD, Shimizu-Motohashi Y, Holm IA, Campbell MG, Lee IH, Brewster SJ, Hanson E, Harris HK, Lowe KR, Saada A, Mora A, Madison K, Hundley R, Egan J, McCarthy J, Eran A, Galdzicki M, Rappaport L, Kunkel LM, Kohane IS. Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. PLoS One. 2012; 7(12): e49475.

23. Bellmann K, Burkart V, Bruckhoff J, Kolb H, Landry J. p38-dependent enhancement of cytokine-induced nitric-oxide synthase gene expression by heat shock protein 70. J Biol Chem. 2000; 275(24): 18172-18179.

24. Kim HW, Cho SC, Kim JW, Cho IH, Kim SA, Park M, Cho EJ, Yoo HJ. Family-based association study between NOS-I and -IIA polymorphisms and autism spectrum disorders in Korean trios. Am J Med Genet B Neuropsychiatr Genet. 2009; 150B(2): 300-306.

25. Napolioni V, Ober-Reynolds B, Szelinger S, Corneveaux JJ, Pawlowski T, Ober-Reynolds S, Kirwan J, Persico AM, Melmed RD, Craig DW, Smith CJ, Huentelman MJ. Plasma cytokine profiling in sibling pairs discordant for autism spectrum disorder. J Neuroinflammation. 2013; 10: 38.

26. Saito Y,Nakagami H, Azuma N, Hirata S, Sanada F, Taniyama Y, Morishita R, Kaneda Y, Sasajima T. Critical roles of cold shock domain protein A as an endogenous angiogenesis inhibitor in skeletal muscle. Antioxid Redox Signal. 2011; 15(8): 2109-2120.

27. Goines PE, Ashwood P. Cytokine dysregulation in autism spectrum disorders (ASD): possible role of the environment. Neurotoxicol Teratol. 2013; 36: 67-81.