Integrative Analysis of shared Genetic Pathogenesis by Autism Spectrum Disorder and Obsessive-Compulsive Disorder

CURRENT STATUS: POSTED

Dongbai Liu
The first people's hospital affiliated to soochow university

Hongbao Cao
Elsevier Inc

Kamil Can Kural
George Mason University

Qi Fang
The First People's Hospital Affiliated to Soochow University

f.qi@gousinfo.com Corresponding Author

Fuquan Zhang
Nanjing Medical University

DOI:
10.21203/rs.2.9334/v1

SUBJECT AREAS
Internal Medicine Specialties

KEYWORDS
Obsessive-compulsive disorder, autism spectrum disorder, Pathway enrichment analysis, Network connectivity analysis
Abstract

Background

Many common pathological features have been observed for both autism spectrum disorders (ASD) and obsessive-compulsive disorder (OCD). However, no systematic analysis of the common gene markers associated both ASD and OCD has been conducted so far.

Results

Here, two batches of large-scale literature based disease-gene relation data (updated in 2017 and 2019, respectively) and gene expression data were integrated to study the possible association between OCD and ASD at the genetic level. Genes linked to OCD and ASD present significant overlap (p-value<2.64e-39). A genetic network of over 20 genes was constructed, through which OCD and ASD may exert influence on each other. The 2017-based analysis suggested six potential common risk genes for OCD and ASD (CDH2, ADCY8, APOE, TSPO, TOR1A, and OLIG2), and the 2019-based study identified two more genes (DISP1 and SETD1A). Notably, the gene APOE identified by the 2017-based analysis has been implicated to have an association with ASD in a recently study (2018) with DNA methylation analysis.

Conclusions

Our results support the possible complex genetic associations between OCD and ASD. Genes linked to one disease is worthy of further investigation as potential risk factors for the other.

Background

Autism spectrum disorders (ASD) are common, highly heritable neuro-developmental conditions characterized by language impairments, social deficits, and repetitive behaviors. So far, many articles have reported that a number of core pathological features of ASD are also commonly observed in obsessive-compulsive disorder (OCD) [1,2], and similar brain abnormalities have also been suggested among ASD and OCD patients [3]. Additionally, considerable amounts of evidence demonstrate that patients with ASD are at an increased risk of comorbid anxiety disorders [4-7]. For instance, van Steensel et al. reported that about 40% of patients with ASD are assigned at least one comorbid diagnosis of anxiety and approximately 17% of children with ASD meet criteria for OCD [8].
Diagnostic characteristics of these patients include persistent and distressing thoughts and behaviors used to “cope with” those thoughts.

In recent years, genetic studies using both genome-wide association study (GWAS) and gene expression data have revealed hundreds of genes associated with both ASD and OCD [9-12]. However, as far as we know, there has been no systematical study performed to investigate the common genes between both diseases.

In this study, we integrated gene expression data and large-scale literature knowledge database to study the association between OCD and ASD at the genetic level, with the purpose to gain a better understanding of the possible common genetic basis and to identify novel common genes associated with both diseases. The disease-related genes were identified using Pathway Studio (http://www.pathwaystudio.com/), which has been widely used to study modeled relationships between proteins, genes, complexes, cells, tissues, and diseases [13] (http://pathwaystudio.gousinfo.com/Mendeley.html). Updated weekly, the Pathway Studio possesses the largest database among known competitors in the field [14].

Results

Common genes for OCD and ASD

Within the curated ASD_OCD database (2017-based-analysis), there were 81 genes associated with OCD, supported by 450 scientific references from 1992 to 2016 (see ASD_OCD: OCD Related Genes and Ref for OCD Related Genes). For ASD, there were 529 related genes supported by 2,098 references from 2000 to 2017 (ASD_OCD: ASD Related Genes and Ref for ASD Related Genes). A significant overlap of 47 genes was identified for both diseases (Right tail Fisher’s Exact test, p-value= 1.66e-48), as shown in Fig. 1. The p-value here means that, for two random gene sets with the size of 81 and 529, respectively, the probability that they present an overlap of 47 or more is less than 1.66e-48. More information on these 47 genes is presented in ASD_OCD→47 common genes.

Within the ASD_OCD_2019 database (2019-based-analysis), there were 86 OCD genes and 624 ASD genes, with an overlap of 43 genes. The decrease of the common genes was due to the removal of the references of low confidence level in the new analysis. In this case, the overlap present a
significance p-value = 2.64e-39. Please refer to ASD_OCD_2019 for the relevant information.

Possible co-regulations between OCD and ASD

2017-based analysis using PS showed that 25 out of the 47 common genes present both downstream and upstream regulation relationship with both ASD and OCD (influenced by and influencing both OCD and ASD), as shown in Fig. 2 (a). The detailed information of the network presented in Fig. 2 (a) can be found in ASD_OCD→ Co-Regulation Network, including the type of the relationship, supporting references and related sentences from the references where the relationship has been identified.

For 2019-batch related analysis, 23 out of 43 common genes present both downstream and upstream regulation relationship with both ASD and OCD, as shown in Fig. 2 (b). More information is presented in ASD_OCD_2019→ Co-Regulation Network. Fig. 2 showed that OCD and ASD may influence the pathogenic development of each other through these genetic networks.

Gene expression analysis

Although there was a significant overlap between ASD-genes and OCD-genes, some genes were linked to one disease only. Specifically, From ASD_OCD, there are 34 genes linked to OCD but not to ASD, this number changed to 43 for ASD_OCD_2019. Here we tested the correlation between these OCD-specific genes and ASD, using expression data (GSE28521 and GSE38322). Fig. 3 elucidates the ‘-log10’ transferred p-values (q=0.05 for FDR) of each gene tested. The detailed results are presented in ASD_OCD and ASD_OCD_2019 (GSE28521 and GSE38322, respectively), including the p-values and FDR correction status.

Fig. 3 showed that two genes (TSPO and APOE) from GSE28521 passed the p=0.05 FDR (Fig. 3a), and seven genes (CDH2, ADCY8, APOE, TOR1A, OLIG2, DISP1, and SETD1A) from GSE38322 passed the FDR. To note, two genes (DISP1 and SETD1A) were newly identified in the 2019-based analysis, and the gene APOE were replicated by a recently published study [15], which showed that APOE presented a significant association with ASD in a DNA methylation analysis. The replication
demonstrated the effectiveness of the workflow proposed in this study.

PPI and Shorted-path analysis

By using a shorted-path approach (conducted by using Pathway Studio) we explored possible pathways between the 8 identified genes and ASD, as shown in Fig. 4. The shorted path analysis was conducted to identify entities (e.g., drugs and proteins) that were linked to both a gene and ASD in a directed path (e.g., ADCY8→BDNF→ASD). The detailed information of the relationships in Fig. 4 is presented in ASD_OCD_2019(Shortest_Path).

Put Figure 4 above here.

Discussion

Previous studies showed that OCD is closely related to ASD [1-3]. In this study, we integrated large-scale literature based relation data and gene expression data to test the hypothesis that ASD and OCD display significant shared genetic basis in terms of common related genes. Gene expression data analysis suggested novel potential common genes for both diseases, supported by the functional network analysis. To note, we used literature data from both 2017 and 2019 to generate solid results using the proposed workflow.

Genes linked to OCD and ASD reveals significant overlaps (see Fig. 1, p-value=6.76e-34 and p-value=2.64e-39 for the 2017-based and 2019-based analysis, respectively). Moreover, we observed an co-regulation network between OCD and ASD, composed of more than 20 genes (Fig. 2). The genes within the network are downstream targets of ASD/OCD, while they are also the upstream regulators of OCD/ASD. Our findings support the genetic association between OCD and ASD.

To explore the possible linkage between the genes that have only been implicated with OCD but not ASD, we used two ASD gene expression dataset (GSE28521 and GSE38322) to explore the OCD-specific genes in case of ASD. For the 2017-based analysis, results revealed 6 OCD genes also present significant differences (FDR corrected p-value<0.05) between ASD cases and healthy controls, including 2 genes (TSPO and APOE) from GSE28521 and 5 genes (CDH2, ADCY8, APOE, TOR1A, and OLIG2) from GSE38322. Notably, the gene APOE that has been identified in both datasets in the 2017-based analysis were reported to have a potential association with ASD in a recent study [15], with
supported the effectiveness of the proposed workflow. It has been shown that APOE methylation in pediatric patients with ASD was significantly higher than that in the healthy controls (median PMR, 33 vs. 11%; \( P=2.36\times10^{-10} \)). Thus, APOE hypermethylation in peripheral blood DNA may be used as a diagnostic biomarker for ASD. In addition, the 2019-based analysis suggested two more common genes for both OCD and ASD (DISP1 and SETD1A), as highlighted by the green circle in Fig. 3 b. Functional network analysis showed that these 8 OCD genes also presented functional correlation with ASD, forming a genetic network supported by over 1,600 scientific reports (Fig. 4; see ASD_OCD_2019: Shortest_Path). These results suggested multiple genetic paths through which these genes play roles for the pathological development of ASD.

PPI showed that TSPO regulates both APOE and CDH2, and these three genes presented multiple common pathways regulating ASD (Fig 4a). On the other hand, we see no connection between the rests of five genes (Fig 4b). Based on the hypothesis that, if both pair of genes play roles within OCD and ASD, they were more likely functionally linked to each other than not. Thus, our results suggested that more attention should be paid to the three genes (TSPO, APOE, and CDH2).

Specifically, the suggested potential APOE-ASD association has also been proposed by a recent study [15]. APOE is a widely-studied and well-known gene, primarily produced by the liver and macrophages, and mediates cholesterol metabolism in an isoform-dependent manner. APOE is the principal cholesterol carrier in the brain [16], with three major alleles: APOE2 (cys112, cys158), APOE3 (cys112, arg158), and APOE4 (arg112, arg158) [17]. The APOE4 variant was frequently reported to be the largest known genetic risk factor for late-onset sporadic Alzheimer's disease [18].

The genetic paths found in the functional network connecting APOE with ASD (Fig. 4a) may provide new insights for a possible linkage between APOE and ASD. For instance, APOE is associated with the production of NO in macrophages [19], while increased NO synthase plays roles in the pathologic development of ASD [20]. APOE can also significantly inhibit RELN binding [21], while RELN has been reported to play an important pathophysiological mechanism in ASD [22]. These findings suggested a clue for the possible association between APOE and ASD. More of these genetic paths can be identified from ASD_OCD_2019_Shortest_Path.
To note, the ASD and OCD related genes employed in this study were identified from a literature review. However, a reported relationship in the publication does not necessarily guarantee a true biological gene-disease linkage. Therefore, findings from this study are more suggestive than confirmative. Further study including biological experiments is preferred to confirm the results identified in this study.

Conclusion

To sum up, results from this study support the hypothesis that OCD and ASD present significant association at the genetic level, which may explain their common pathological features in the clinic. Additionally, 8 genes were suggested as common genes for both OCD and ASD, and one of them has been recently confirmed by another study. To our knowledge, this is the first study to integrate large-scale literature relation data and gene expression data for a systematical evaluation of the associations between OCD and ASD at the genetic level. Findings here may add new insights into the current field of OCD-ASD correlation study, and guarantee further studies using more data sets to test novel potential risk genes for both ASD and OCD.

Methods

The large-scale literature data based ASD-gene and OCD-gene relations were studied targeting identification of common genes associated with both diseases. These disease-implicated genes were then tested using two ASD expression datasets to discover possible novel common genes. After that, functional network analysis was conducted to study the pathogenic significance of the identified genes in ASD.

To validate the stability of the proposed workflow, two batches of literature data (updated in August 2017 and March 2019, respectively) were analyzed and results were compared. For the 2017-batch-based analysis, all results were organized in a databased ASD_OCD. For the 2019-batch-based analysis, it’s in ASD_OCD_2019. The downloadable format of these two databases is available at gousinfo.com/database/Data_Genetic/ASD_OCD.xlsx and http://gousinfo.com/database/Data_Genetic/ASD_OCD_2019.xlsx, respectively. The two files are also available as Supplementary files: ASD_OCD.xlsx and ASD_OCD_2019.xlsx.
Disease-Gene relation data

Disease-gene relation data for both ASD and OCD were acquired through large-scale literature data analysis assisted by Pathway Studio (www.pathwaystudio.com) and presented in ASD_OCD and ASD_OCD_2019. Besides the full lists of genes linked to both diseases, we also presented the information of supporting references for each disease-gene relation (ASD_OCD and ASD_OCD_2019: Ref for OCD/ASD Related Genes), including titles of the references and the related sentences where the disease-gene relationship was identified. The information could be used to locate a detailed description of how a candidate gene is associated with OCD and/or ASD.

ASD expression data

Two ASD expression datasets were acquired from Illumine BaseSpace Correlation Engine (http://www.illumina.com). After the initial search with target set as ‘Autism Spectrum Disorders’, the expression datasets were screened by the following criteria, including: 1) The data organism is Homo sapiens; 2) The data type is RNA expression; 3) The samples of studies come from brain tissues; and 4) The studies are limited to ASD case vs. healthy control study (almost equal number of case and control). The top two gene expression datasets (GSE28521: 39 ASD vs. 40 healthy controls; GSE38322: 18 ASD vs. 18 healthy controls) were selected to test the genes linked to OCD but not with ASD. For a gene to be tested, one-way ANOVA was performed to compare the expression of this gene between ASD controls and cases. The genes that passed an FDR corrected (q=0.05) will be identified as significant potential ASD target genes for further analysis.

Shorted-path analysis of the target risk genes

For the significant genes identified through expression analysis described above, shorted-path-based network analysis was conducted between the target genes and the disease (ASD/OCD) to identify potential biological connections. The analysis was performed using the ‘Shortest Path’ module of Pathway Studio (www.pathwaystudio.com).

Protein-protein interaction analysis

To explore the relationships between any identified potential common genes for ASD and OCD, we conducted a literature-based protein-protein interaction analysis (PPI). Two genes were identified to
have an association if they have been reported as such in one or more scientific reports. The underlying hypothesis is that, if these genes were associated with both diseases, they may present functional associations between each other.

List Of Abbreviations
ASD: Autism spectrum disorder
OCD: obsessive-compulsive disorder
GWAS: genome-wide association study
GSEA: Gene Set Enrichment Analysis

Declarations
Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Availability of data and materials
All the data used in this study are provided in two online cross-disease genetic databases: **ASD OCD 2019** and **ASD OCD**.

Competing interests
There is no conflict of interest.

Funding
This work was supported by the National Natural Science Foundation of China (81471364, 81571319) and National Key Research and Development Program of China (2016YFC1307004), including design of the study, collection, analysis, and interpretation of data, and writing of the manuscript. Publication charges for this article have also been covered by the first grant.

Authors’ contributions
DL, FQ and FZ designed the study and collected the data. DL, FZ, HC and KK performed the data analysis and contributed to the writing of the manuscript. DL and FQ contributed to the acquisition of funding that supported this study. All authors read and approved the final manuscript.
Acknowledgments

Not applicable.

References

1. Fineberg NA, Brown A, Reghunandanan S, Pampaloni I. Evidence-based pharmacotherapy of obsessive-compulsive disorder. Int J Neuropsychopharmacol. 2012; 15(8), 1173-1191.

2. Hurwitz R, Blackmore R, Hazell P, Williams K, Woolfenden S. Tricyclic antidepressants for autism spectrum disorders (ASD) in children and adolescents. Cochrane Database Syst Rev. 2012; (3): CD008372.

3. Gross-Isseroff R, Hermesh H, Weizman A. Obsessive compulsive behaviour in autism—towards an autistic-obssesive compulsive syndrome? World J Biol Psychiatry. 2001; 2(4): 193-197.

4. de Bruin EI, Ferdinand RF, Meester S, de Nijs PF, Verheij F. High rates of psychiatric co-morbidity in PDD-NOS. J Autism Dev Disord. 2007; 37(5): 877-886.

5. MacNeil BM, Lopes VA, Minnes PM. Anxiety in children and adolescents with autism spectrum disorders. Research in Autism Spectrum Disorders. 2009; 3(2009): 1-21.

6. Simonoff E, Pickles A, Charman T, Chandler S, Loucas T, Baird G. Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. J Am Acad Child Adolesc Psychiatry. 2008; 47(8): 921-9.

7. White SW, Oswald D, Ollendick T, Scahill L. Anxiety in children and adolescents with autism spectrum disorders. Clin Psychol Rev. 2009; 29 (3):216-29.

8. van Steensel FJ, Bogels SM, Perrin S. Anxiety disorders in children and adolescents with autistic spectrum disorders: a meta-analysis. Clin Child Fam Psychol Rev. 2011; 14: 302-17.

9. Liu X, Shimada T, Otowa T, Wu YY, Kawamura Y, Tochigi M, et al.. Genome-wide Association Study of Autism Spectrum Disorder in the East Asian Populations. Autism Res. 2016; 9(3): 340-49.

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

11. Mathews CA, Badner JA, Andresen JM, Sheppard B, Himle JA, Grant JE, et al. Genome-wide linkage analysis of obsessive-compulsive disorder implicates chromosome 1p36. Biol Psychiatry. 2012; 72:
12. Jaffe AE, Deep-Soboslay A, Tao R, Hauptman DT, Kaye WH, Arango V, et al.. Genetic neuropathology of obsessive psychiatric syndromes. Transl Psychiatry. 2014; 4: e432.

13. Nikitin A, Egorov S, Daraselia N, Mazo I. Pathway studio–the analysis and navigation of molecular networks. Bioinformatics. 2003;19(16): 2155-2157.

14. Lorenzi PL, Claerhout S, Mills GB, Weinstein JN. A curated census of autophagy-modulating proteins and small molecules: candidate targets for cancer therapy. Autophagy. 2014; 10 (7): 1316-26.

15. Hu Z, Yang Y, Zhao Y, Yu H, Ying X, Zhou D, et al. APOE hypermethylation is associated with autism spectrum disorder in a Chinese population. Exp Ther Med. 2018; 15(6):4749-4754.

16. Puglielli L, Tanzi RE, Kovacs DM. Alzheimer’s disease: the cholesterol connection. Nat Neurosci. 2003; 6: 345-351.

17. Ghebranious N, Ivacic L, Mallum J, Dokken C. Detection of ApoE E2, E3 and E4 alleles using MALDI-TOF mass spectrometry and the homogeneous mass-extend technology. Nucleic Acids Res. 2005;33 (17): e149.

18. Sadigh-Eteghad S, Talebi M, Farhoudi M. Association of apolipoprotein E epsilon 4 allele with sporadic late onset Alzheimer’s disease. A meta-analysis. Neurosciences (Riyadh). 2012; 17(4): 321-326.

19. Colton CA, Czapiga M, Snell-Callanan J, Chernyshev ON, Vitek MP. 2001. Apolipoprotein E acts to increase nitric oxide production in macrophages by stimulating arginine transport. Biochim Biophys Acta. 2001; 1535(2):134-144.

20. Yui K, Kawasaki Y, Yamada H, Ogawa S. Oxidative Stress and Nitric Oxide in Autism Spectrum Disorder and Other Neuropsychiatric Disorders. CNS Neurol Disord Drug Targets. 2016;15 (5): 587-596.

21. Pappas A, Chaiworapongsa T, Romero R, Korzeniewski SJ, Cortez JC, Bhatti G, et al. Transcriptomics of maternal and fetal membranes can discriminate between gestational-age matched preterm neonates with and without cognitive impairment diagnosed at 18-24 months. PLoS One.
22. Macrì S, Biamonte F, Romano E, Marino R, Keller F, Laviola G. Perseverative responding and neuroanatomical alterations in adult heterozygous reeler mice are mitigated by neonatal estrogen administration. Psychoneuroendocrinology. 2010; 35(9):1374-87.

Supplementary File Legend
Filename: ASD_OCD_2019.xlsx; Title of data: The cross-disease genetic database Description of data: The ASD_OCD_2019 database contains autism spectrum disorders (ASD) and obsessive-compulsive disorder (OCD) related genes, pathways, novel genes, info of related supporting references and the analysis results of two gene expression datasets (GSE28521 and GSE38322). The data and results were updated in 2019.

Filename: ASD_OCD.xlsx; Title of data: The cross-disease genetic database Description of data: The ASD_OCD database contains autism spectrum disorders (ASD) and obsessive-compulsive disorder (OCD) related genes, pathways, novel genes, info of related supporting references and the analysis results of two gene expression datasets (GSE28521 and GSE38322). The data and results were updated in 2017.

Figures
Figure 1

Venn diagram between ASD-genes and OCD-genes. (a) Results based on the 2017-updated database; (b) Results based on the 2019-updated database.
Figure 2

Co-regulation Network between ASD and OCD. (a) Results based on the 2017-updated database; (b) Results based on the 2019-updated database.
The p-values of the OCD individual genes for ASD case/control expression comparison in dataset GSE28521 and GSE38322. (a) The p-values of the OCD-specific genes in dataset GSE38322; (b) The p-values of the OCD-specific genes in dataset GSE38322; The p-values have been through FDR correction with q=0.05 and logic transformation using ‘–log10’. Names and corresponding transferred p-values of selected genes passed the FDR correction (q=0.05) were marked at corresponding positions. The two genes (DISP1 and SETD1A) highlighted by green circle were newly identified by 2019-based analysis, and the red-circle highlighted gene (APOE) was replicated in a 2018-published article.
Figure 4

PPI and shortest-path analysis results. (a) Three connected genes (APOE, TSOP, and CDH2) present association with ASD. (b) Five disconnected genes (DISP1, SETD1A, OLIG2, TOR1A, and ADCY8) present association with ASD.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.
ASD_OCD.xlsx
ASD_OCD_2019.xlsx