GENETIC AND EPIGENETIC MECHANISMS UNDERPINNING VULNERABILITY TO DEVELOPING PSYCHIATRIC DISORDERS

EDITED BY: Elena Martín-García, Patrick E. Williams and Noèlia Fernàndez-Castillo

PUBLISHED IN: Frontiers in Psychiatry, Frontiers in Genetics and Frontiers in Neuroscience

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88976-515-7
DOI 10.3389/978-2-88976-515-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

GENETIC AND EPIGENETIC MECHANISMS UNDERPINNING VULNERABILITY TO DEVELOPING PSYCHIATRIC DISORDERS

Topic Editors:
Elena Martín-García, Pompeu Fabra University, Spain
Patrick E. Williams, University of Kansas, United States
Noèlia Fernàndez-Castillo, Centre for Biomedical Network Research (CIBER), Spain

Citation: Martín-García, E., Williams, P. E., Fernàndez-Castillo, N., eds. (2022). Genetic and Epigenetic Mechanisms Underpinning Vulnerability to Developing Psychiatric Disorders. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-515-7

Table of Contents

frontiers | Frontiers in Psychiatry

EDITORIAL
published: 13 June 2022
doi: 10.3389/fpsyt.2022.917198

Editorial: Genetic and Epigenetic Mechanisms Underpinning Vulnerability to Developing Psychiatric Disorders

Noèlia Fernández-Castillo [1,2,3,4] *and Elena Martín-García* [5,6*]

[1] *Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Catalonia, Spain,* [2] *Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Catalonia, Spain,* [3] *Institut de Biomedicina de la Universitat de Barcelona (IBUB), Catalonia, Spain,* [4] *Institut de Recerca Sant Joan de Déu (IR-SJD), Esplugues de Llobregat, Barcelona, Spain,* [5] *Laboratory of Neuropharmacology-Neurophar, Department of Medicine and Life Sciences, Universitat Pompeu Fabra (UPF), Barcelona, Spain,* [6] *Hospital del Mar Medical Research Institute (IMIM), Barcelona, Spain*

Keywords: behavioral genetics, epigenetics, vulnerability, psychiatric disorders, multifactorial, animal models

Editorial on the Research Topics

Genetic and Epigenetic Mechanisms Underpinning Vulnerability to Developing Psychiatric Disorders

The study of genetic and epigenetic mechanisms underlying vulnerability to developing psychiatric disorders has obtained significant outcomes from the behavioral genetics area, providing increased knowledge of individual differences in behavior. Classical research in behavioral genetics with humans was applied in studies of twins, family investigations, and adoption. These studies produced a significant advancement in behavioral genetics, and studies of twins and adoption pointed out that the environment and the interaction between genes and the environment are essential. In addition, molecular genetics has confirmed that the susceptibility to psychiatric disorders is heritable, polygenic, and complex (1). Thus, psychiatric disorders have a complex multifactorial pattern of inheritance (**Figure 1**). This multifactorial genetic model explains how the interaction between multiple gene networks and environmental factors strongly impacts brain function early during development and later throughout adulthood, influencing behavior (2). Hence, genes are not unique direct triggers of psychiatric disorders but contribute to confer risk for pathological behavior development, accounting for a fraction of total variation (2). An interesting open question is why some individuals are vulnerable to developing a psychiatric disease while others are resilient to certain environmental risk factors, even if their access is not prevented (3).

To answer this question, research in psychiatric disorders has dramatically improved thanks to the use of new omic technologies, which allowed hypothesis-free approaches and have impacted the understanding of these disorders and their underlying mechanisms. Articles published in this Research Topics focus on the relevance of using genomic and epigenomic approaches to disentangle the genetic and epigenetic mechanisms underpinning vulnerability to developing psychiatric disorders. In a review article, Singh et al. highlight the contribution of postzygotic somatic mutations to neurodevelopmental and mental disorders, particularly schizophrenia. Postzygotic mutations, which can be detected nowadays with novel genomic and epigenomic technologies, are relatively common in the brain and have been reported in patients with neurodevelopmental disorders. Spencer and Kisby (4) made a commentary article to this review agreeing with their analysis and arguing that these somatic mutations or epimutations may be sufficient to develop a neurodegenerative disorder. Pol-Fuster et al. applied a system genomic approach to identify rare genetic variants involved in psychosis in a family with several members affected. They

OPEN ACCESS

Edited and reviewed by:
Jacob Peedicayil, Christian Medical College & Hospital, India

***Correspondence:**
Elena Martín-García
elena.martin@upf.edu

Specialty section:
This article was submitted to Behavioral and Psychiatric Genetics, a section of the journal Frontiers in Psychiatry

Received: *10 April 2022*
Accepted: *17 May 2022*
Published: *13 June 2022*

Citation:
Fernández-Castillo N and Martín-García E (2022) Editorial: Genetic and Epigenetic Mechanisms Underpinning Vulnerability to Developing Psychiatric Disorders. Front. Psychiatry 13:917198. doi: 10.3389/fpsyt.2022.917198

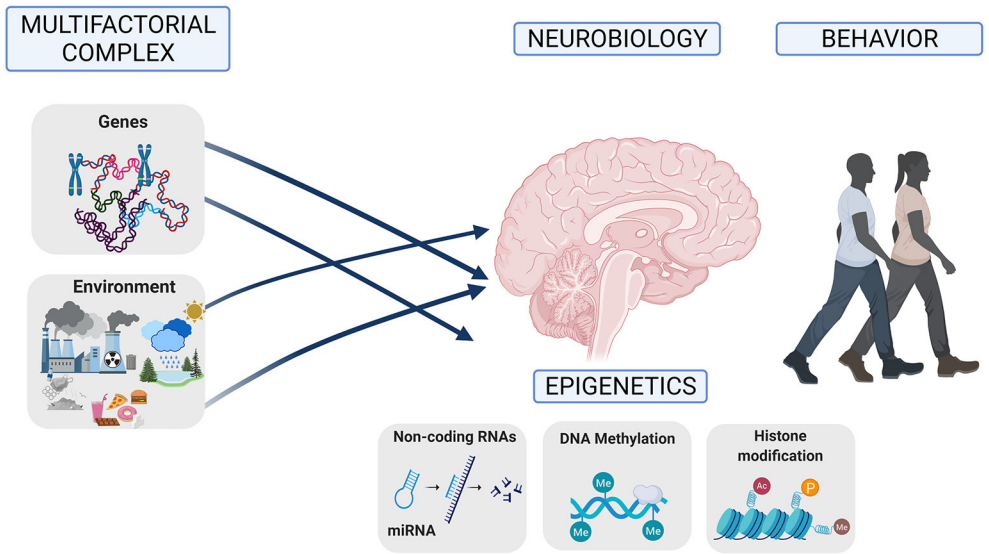


FIGURE 1 | Psychiatric disorders have a complex multifactorial origin with an interaction between multiple genes and multiple environments that influence neurobiology and modify epigenetics during development and adulthood, influencing behavior.

identified a rare missense variant in the *MACF1* gene and two rare CNVs affecting the genes *CNTN6* and *CDH13* that segregated with psychosis in this family. Liu et al. also performed whole-exome sequencing identifying novel variants in *PGM3* in patients with idiopathic focal epilepsy. Odintsova et al. show the value of combining different omics information for predicting models. For a multi-omics prediction model, the authors combined polygenic scores with methylation scores for smoking status, prenatal maternal exposure, birth weight, and BMI. They found that methylation scores could predict current and former smoking, prenatal maternal smoking, and BMI.

In addition, transcriptomic studies have contributed to understanding the mechanisms involved in psychiatric disorders by directly assessing gene expression alterations and pathways. Using brain samples would be ideal, but unfortunately, this tissue cannot be accessed in patients, and only *post-mortem* samples can be used, which are very difficult to obtain. On the other hand, investigating gene expression alterations in more accessible tissues, such as blood, has become a helpful approach. Bountress et al. investigated differences in gene expression associated with post-stress syndrome disorder (PTSD) and trauma exposure (TE). Peripheral blood mononuclear cells identified 283 genes differentially expressed between PTSD and TE and 7 gene modules enriched in pathways such as focal adhesion or neuroactive ligand-receptor interaction. Zhang et al. used data from *post-mortem* brain samples of patients to identify relevant genes and pathways for bipolar disorders. Three modules of differentially expressed genes were associated with bipolar disorder, with 4 genes identified as hub genes: *NOTCH1, POMC, NGF*, and *DRD2*, being *NOTCH1* replicated in an independent sample. Bipolar disorder has also been associated with an increased prevalence of minor physical anomalies. The brain and the skin share a common origin since they are derived from the ectoderm. Thus, minor physical anomalies may be physical markers of aberrant neurodevelopment. In the work of Csulak and coauthors, it has been demonstrated for the first time that there is an increased prevalence of minor physical anomalies among the first-degree unaffected relatives of bipolar I patients. The study supports the concept that minor physical anomalies can be endophenotypic markers of bipolar I affective disorder (Csulak et al.). The endophenotypes reported were minor physical anomalies in the eye, head, mouth, and trunk regions that were found to be prevalent among the relatives of bipolar I patients compared to normal controls. The individual analyses showed that one minor malformation (sole crease) and one phenogenetic variant (high arched palate) were also more prevalent in the relative group compared to the normal control group (Csulak et al.). It is important to remark that the endophenotype conception of bipolar disorder signifies an essential method in exploring the pathogenesis of the disease. Gottesman (5) described this concept of endophenotype as an intermediate phenotype that fills the gap between genes and diseases.

Candidate gene studies were more frequently used in the past, before the advent of omic technologies, which showed that larger clinical samples were required for obtaining reliable findings in association studies. In this Research Topics, some studies have evaluated the contribution of genetic variants in particular genes. For instance, Shang et al. meta-analyzed previous studies testing the association with an obsessive-compulsive disorder of a single nucleotide polymorphism (SNP) in *BDNF*, an important neurotrophic factor. They found negative results, but gender-specific analyses showed a nominal association in females. Likewise, Komatsu et al. investigated an SNP in the *OLIG2* gene, involved in oligodendrocyte functions, associated with self-schema, a susceptibility trait factor for major depressive disorder. In addition, Tao et al. found a nominal association

between temporal lobe epilepsy and an SNP in *DTNBP2*, a gene related to schizophrenia. Finally, Chen et al. found that patients with the first psychosis episode had higher complement components C3 and C4 levels in serum and found that a SNP in the C4 gene was associated.

On the other hand, animal models have helped us understand the role of critical genes in behavioral genetics. For example, Rivero et al. article demonstrates that the haploinsufficiency of the attention-deficit/hyperactivity disorder risk gene *St3gal3* in mice causes alterations in cognition and expression of genes involved in myelination and sialylation. Indeep, the authors found gender differences because male *St3gal3* heterozygous mice showed cognitive deficits, while female heterozygous animals displayed increased activity and cognitive control compared to their wild-type littermates. In addition, expression of several markers implicated in oligodendrogenesis, myelin formation, protein sialylation, and cell adhesion/synaptic target glycoproteins of *ST3GAL3* in a brain region- and sex-specific manner were found. In another article on this issue, the gender differences were highlighted by the work of Reever and colleagues. Their work showed that maternal immune activation during gestation has a prolonged and sex-dependent impact on the transcriptome of the amygdala in pigs (Keever et al.). Specifically, they observed differential gene expression associated with neurodevelopmental illnesses, including schizophrenia and autism spectrum disorders. In schizophrenia, other authors using animal models to study the genetic basis of psychiatric disorders revealed that *IL15Rα* KO mice exhibit altered expression of multiple pathways, including lipid biosynthesis and metabolism in the central nervous system (He et al.).

Finally, an interesting view of genomic variation and the paradox of mental illness can be found in this Research Topics (Gualtieri). In this hypothesis and theory article, Gualtieri raises the question of how disorders such as autism and schizophrenia persist and are high heritable, although they confer low reproductive success. The article integrates different perspectives such as heritability and reproductive success and a throughout the view of the contribution of genomic variability at different levels, bringing a novel view to the genetics of psychiatric disorders.

This Research Topic has contributed to a better understanding that psychiatric disorders are complex multifactorial disorders in which nature and nurture interact. Nurture is the personal experience of the interactions with the environment as a raising background, whereas genetic factors define nature. Furthermore, in animal models, nurture has been translationally modeled using procedures that include childhood separation after birth or stress induction (6). Also, the study of the influence of genetic factors in animals elucidated critical landscapes, and it is an evolving field with a promising future to untangle how genetic susceptibility affects the neurobiological substrates of psychiatric disorders. In summary, the new advances in understanding the complex mechanisms involved in the epigenetic control of gene expression and the novel technological tools to manipulate these mechanisms will allow us to understand the ability of environmental influences to alter gene expression and their impact on the progress of the psychiatric disorders. Furthermore, identifying these epigenetic factors opens new opportunities to define accurate biomarkers of vulnerability and resilience to psychiatric disorders and may open pioneering beneficial perspectives that are still unemployed.

AUTHOR CONTRIBUTIONS

Both authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported by Plan Nacional Sobre Drogas of the Spanish Ministry of Health (#PNSD-2019I006) and by European Union NextGenerationEU/PRTR, the Spanish Ministry of Science and Innovation – MICIN, and the Research State Agency – AEI, through Plan de Recuperación, Transformación y Resiliencia funding program and the ERA-NET neuron translational research projects on neurodevelopmental disorders (MCIN/AEI/UE - PCI2021-122073-2A/10.13039/501100011033) to EM-G. This work was also supported by Plan Nacional Sobre Drogas of the Spanish Ministry of Health (#PNSD-2020I042) to NF-C. Figures are created with BioRender.com.

REFERENCES

1. Plomin R, Daniels D. Why are children in the same family so different from one another? *Int J Epidemiol.* (2011) 40:563–82. doi: 10.1093/ije/dyq148
2. Hamer D. Genetics. Rethinking behavior genetics. *Science.* (2002) 298:71–2. doi: 10.1126/science.1077582
3. Maldonado R, Calve P, García-Blanco A, Domingo-Rodríguez L, Senabre E, Martín-García E. Genomics and epigenomics of addiction. *Am J Med Genet.* (2021) 186:128–39. doi: 10.1002/ajmg.b.32843
4. Spencer PS, Kisby GE. Commentary on Singh et al. (2020) Postzygotic somatic mutations in the human brain expand the threshold-liability model of schizophrenia. *Front Psychiatry.* (2021) 12:2020–2. doi: 10.3389/fpsyt.2021.653624
5. Gottesman I. The endophenotype concept in psychiatry: etymology and strategic intentions. *Int J Dermatol.* (2005) 44:1031–4. doi: 10.1111/j.1365-4632.2005.02252.x
6. Meaney MJ, Szyf M. Environmental programming of stress responses through DNA methylation: life at the interface between a dynamic environment and a fixed genome. *Dialogues Clin Neurosci.* (2005) 7:103–23. doi: 10.31887/DCNS.2005.7.2/mmeaney

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 11 August 2020
doi: 10.3389/fnins.2020.00774

Lasting and Sex-Dependent Impact of Maternal Immune Activation on Molecular Pathways of the Amygdala

*Marissa R. Keever[1], Pan Zhang[2], Courtni R. Bolt[1], Adrienne M. Antonson[1], Haley E. Rymut[1], Megan P. Caputo[1], Alexandra K. Houser[1], Alvaro G. Hernandez[3], Bruce R. Southey[1], Laurie A. Rund[1], Rodney W. Johnson[1,4] and Sandra L. Rodriguez-Zas[1,2,4,5,6]**

[1] Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, United States, [2] Illinois Informatics Institute, University of Illinois at Urbana-Champaign, Urbana, IL, United States, [3] High-throughput Sequencing and Genotyping Unit, Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign, Urbana, IL, United States, [4] Neuroscience Program, University of Illinois at Urbana-Champaign, Urbana, IL, United States, [5] Department of Statistics, University of Illinois at Urbana-Champaign, Urbana, IL, United States, [6] Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, United States

The prolonged and sex-dependent impact of maternal immune activation (MIA) during gestation on the molecular pathways of the amygdala, a brain region that influences social, emotional, and other behaviors, is only partially understood. To address this gap, we investigated the effects of viral-elicited MIA during gestation on the amygdala transcriptome of pigs, a species of high molecular and developmental homology to humans. Gene expression levels were measured using RNA-Seq on the amygdala for 3-week-old female and male offspring from MIA and control groups. Among the 403 genes that exhibited significant MIA effect, a prevalence of differentially expressed genes annotated to the neuroactive ligand–receptor pathway, glutamatergic functions, neuropeptide systems, and cilium morphogenesis were uncovered. Genes in these categories included corticotropin-releasing hormone receptor 2, glutamate metabotropic receptor 4, glycoprotein hormones, alpha polypeptide, parathyroid hormone 1 receptor, vasointestinal peptide receptor 2, neurotensin, proenkephalin, and gastrin-releasing peptide. These categories and genes have been associated with the MIA-related human neurodevelopmental disorders, including schizophrenia and autism spectrum disorders. Gene network reconstruction highlighted differential vulnerability to MIA effects between sexes. Our results advance the understanding necessary for the development of multifactorial therapies targeting immune modulation and neurochemical dysfunction that can ameliorate the effects of MIA on offspring behavior later in life.

Keywords: immune activation, pigs, RNA-seq, neuropeptides, glutamatergic pathway, GABAergic pathway

OPEN ACCESS

Edited by:
Noèlia Fernàndez-Castillo, Centre for Biomedical Network Research (CIBER), Spain

Reviewed by:
Silvia Pellegrini, University of Pisa, Italy
Tewarit Sarachana, Chulalongkorn University, Thailand

****Correspondence:***
Sandra L. Rodriguez-Zas
rodrgzzs@illinois.edu

Specialty section:
This article was submitted to Neurogenomics, a section of the journal Frontiers in Neuroscience

Received: *01 May 2020*
Accepted: *01 July 2020*
Published: *11 August 2020*

Citation:
Keever MR, Zhang P, Bolt CR, Antonson AM, Rymut HE, Caputo MP, Houser AK, Hernandez AG, Southey BR, Rund LA, Johnson RW and Rodriguez-Zas SL (2020) Lasting and Sex-Dependent Impact of Maternal Immune Activation on Molecular Pathways of the Amygdala. Front. Neurosci. 14:774. doi: 10.3389/fnins.2020.00774

INTRODUCTION

The maternal immune response triggered by pathogens and other environmental stressors during gestation can also elicit an indirect response by the fetal immune cells (Kroismayr et al., 2004; Odorizzi and Feeney, 2016; Prins et al., 2018). Viral infection during gestation, for example, activates a cytokine-related signaling cascade, and molecules from this process can cross the

placenta and reach the fetal brain. The resulting maternal immune activation (MIA) can impact fetal developmental processes and exert long-term postnatal effects in the offspring (Rutherford et al., 2014). The relationship between MIA and neurodevelopmental disorders, including schizophrenia spectrum disorders (SSD) and autism spectrum disorders (ASD), and neurodegenerative disorders, such as Alzheimer's disease (AD), in offspring has been established (Knuesel et al., 2014; Canetta et al., 2016; Mattei et al., 2017). These diseases share some behavior symptoms, comorbidities such as eating disorders, and genetic and environmental (i.e., MIA) agents (Canitano and Pallagrosi, 2017). The previous neurological disorders have been associated with abnormal structure and dysregulation of the amygdala (Schumann et al., 2011; Fernandez-Irigoyen et al., 2014) and share genes and molecular mechanisms including histocompatibility complex (MHC) genes (Anders and Kinney, 2015), glutamatergic and GABAergic-associated genes (Bourgeron, 2009; Marin, 2012; Li et al., 2016), and mitochondrial activity processes (Pieczenik and Neustadt, 2007; Sragovich et al., 2017).

The fetal amygdala is susceptible to inflammatory signals, and the plasticity of this brain structure to MIA can lead to alterations of the developmental trajectory. These disruptions may have long-lasting and maladaptive consequences for the offspring, due to the significant role that the amygdala plays in many neurological pathways. Located in the forebrain, the amygdala influences social interaction, cognition, neuroendocrine, behavior, learning, memory, emotion, and autonomic systems. The amygdala also modulates the response of these processes to stressors, including pathogenic infections and those resulting from management practices, such as weaning (Tian et al., 2015). The amygdala experiences high uptake of gonadal hormones and is anatomically connected to other sexually dimorphic nuclei. Therefore, this brain region is involved in regulation of several dimorphic functions such as aggression, sexual behavior, gonadotropin secretion, and integration of olfactory information (Hines et al., 1992). Evidence supports the differential activation of the amygdala to stimuli between males and females (Killgore and Yurgelun-Todd, 2001), including differences in the sexual responses and emotional memory (Hamann, 2005), and differential vulnerability to insult (Baird et al., 2007). Due to the interconnected and multi-regulatory nature of this brain structure, insults to the amygdala can impact the individual's social, locomotor, and feeding behavior (Petrovich and Gallagher, 2003); growth and reproductive physiology; health status; and immunological response to secondary stressors.

Recent studies lend support to the link between MIA and altered amygdala function (Carlezon et al., 2019). In mice, MIA elicited by polyinosinic:polycytidylic acid [Poly(I:C)] increased the synaptic strength of glutamatergic projections from the prefrontal cortex to the amygdala (Li et al., 2018). In open-field tests, mice exposed to MIA spent less time in the center and traveled a higher distance, indicative of a higher anxiety behavior incidence than the control counterparts. These findings suggest that the change in the balance between excitation (glutamatergic) and inhibition (feedforward GABAergic) modified the spike output of amygdala neurons, therefore affecting brain circuits that could regulate behavior in SSD and ASD. A candidate gene study of the effects of social stress during gestation reported that the expression of a corticotropin-releasing hormone receptor in the amygdala of 10-week-old pigs was higher in females than in males (Rutherford et al., 2014). This study concluded that prenatal stress substantially increased anxiety-related behaviors in female pigs. Studies of the impact of maternal stressors during gestation on specific amygdala molecular profiles and associated neurological or behavioral disorders in the offspring later in life highlight the complexity of the molecular mechanisms underlying the pathophysiology of MIA.

Research on the lasting effects of MIA in pigs complements the insights offered by rodent models (Antonson et al., 2019). The advantages of studying a pig model stem from the greater homology of humans to pigs, rather than to rodents, when considering organ physiology, size, development and, in particular, brain growth and development processes (Murphy et al., 2014). A pig model that has offered insights into MIA employs porcine reproductive and respiratory syndrome virus (PRRSV) to elicit MIA. This immune challenge activates the microglia (i.e., macrophage-like cells in the brain) and is associated with behavioral changes in neonatal pigs (Antonson et al., 2017, 2018).

The study of MIA elicited by PRRSV allows for the characterization of the impact of a live viral pathogen that self-replicates in the host, evoking extended activation of immune pathways. PRRSV challenge during gestation is a well-characterized, replicable, and effective method for inducing MIA in pigs (Antonson et al., 2017, 2018). In addition, PRRSV outbreaks impose a major economic burden to the livestock industry. PRRSV is an enveloped single-stranded RNA virus that infects alveolar macrophages, causing interstitial pneumonia and increased serum levels of the cytokines interleukin 1 beta, interleukin 6, and tumor necrosis factor alpha (Antonson et al., 2017). The persistent repercussions of MIA on the molecular pathways of the pig amygdala are yet to be investigated. Moreover, the potentially distinct vulnerability to the prolonged effects of MIA between sexes remains unknown.

The overarching goal of the present study is to advance the understanding of the impact of MIA on the molecular mechanisms of the amygdala. Three supporting objectives are explored: (a) characterization of prolonged transcriptome changes elicited by viral MIA in pigs, a species that has high neurodevelopmental homology with humans, and food production value; (b) identification of molecular pathways that present differential vulnerability to MIA between sexes; and (c) understanding the effect of MIA on molecular interactions assisted by gene network inference. The findings from these complementary analyses support the use of multiple therapeutic targets to ameliorate the potential detrimental effect of MIA on the offspring physiology and behavior.

MATERIALS AND METHODS

Animal Experiments

All experimental procedures used published protocols (Antonson et al., 2017, 2018). The animal studies were approved by the Illinois Institutional Animal Care and Use Committee (IACUC) at the University of Illinois and are in compliance with the USDA Animal Welfare Act and the NIH Public Health Service Policy on the Humane Care and Use of Animals.

Camborough gilts born and raised at the University of Illinois at Urbana-Champaign herd were inseminated at 205 days of age using PIC 359 boar sperm (Antonson et al., 2017, 2018). All gilts were PRRSV negative and were moved at gestation day (GD) 69 into disease-containment chambers maintained at 22°C and a 12 h light/dark cycle with lights on at 7:00 AM. The gilts were fed daily 2.3 kg of a gestational diet and had *ad libitum* water access. One week after acclimation, four gilts were intranasally inoculated with live PRRSV strain P129-BV (School of Veterinary Medicine at Purdue University, West Lafayette, IN, United States) using 5 mL of 1×10^5 median tissue culture infectious dose ($TCID_{50}$) diluted in sterile Dulbecco's modified Eagle medium (DMEM; 5 mL total volume). The four gilts in the Control group were intranasally inoculated with an equal volume of sterile DMEM. PRRSV inoculation corresponded to the last third of gestation in pigs and humans, during initiation of rapid fetal brain growth (Antonson et al., 2017, 2018). PRRSV and Control groups were housed in separate containment chambers.

The rectal temperatures and diet consumption of the gilts were recorded daily until farrowing (Antonson et al., 2017, 2018). The PRRSV-inoculated gilts were offered the maximum fed daily, and feed refusal was measured. The Control gilts were fed the same amount consumed by the PRRSV-inoculated gilts on the previous day. The daily body temperature and feed intake levels were compared using a mixed-effects model analyzed with PROC MIXED (SAS Institute Inc., Cary, NC, United States). The model included the effects of gilt treatment and replicate while accommodating for heterogeneity of variance between MIA groups.

Farrowing was induced with an intramuscular injection of 10 mg of Lutalyse (dinoprost tromethamine, Pfizer, New York, NY, United States) on GD 113 in consideration that the average gestation length is approximately 114 days (Antonson et al., 2017, 2018). Gilts farrowed in individual farrowing crates of standard dimensions (1.83 × 1.83 m). After farrowing, the gilts were fed twice a day up to 5 kg of a nutritionally complete diet for the lactating period and water remained available *ad libitum*. Pigs received intramuscular injections of iron dextran (100 mg/pig, Butler Schein Animal Health, Dublin, OH, United States) and Excede for Swine (25 mg/pig; Zoetis, Parsippany, NJ, United States) to control for respiratory diseases. The pigs remained with their mothers until PD 22. The body weight of pigs was measured daily and analyzed using the mixed-effects model in SAS, PROC MIXED (SAS Institute Inc., Cary, NC, United States). The model included the effect of MIA and the random effect of gilt, accommodating for heteroscedasticity between pig treatment and sex groups. The impact of MIA was studied at PD 22 because this is a common age to wean pigs. The study of transcriptome profiles from older pigs could be confounded with changes in diet and environment associated with weaning, while profiles from younger pigs would hinder the assessment of the prolonged effects of MIA.

RNA Extraction and Sequencing

A balanced experimental design was studied, including 24 pigs evenly distributed between maternal PPRSV activated (MPA group of pigs) and Control gilts (CON group of pigs), each group encompassing males and females (denoted Ma and Fe, respectively). At PD 22, pigs were removed from the farrowing crate and anesthetized intramuscularly using a telazol:ketamine:xylazine drug cocktail (50 mg of tiletamine; 50 mg of zolazepam) reconstituted with 2.5 mL ketamine (100 g/L) and 2.5 mL xylazine (100 g/L; Fort Dodge Animal Health, Fort Dodge, IA, United States) at a dose of 0.03 mL/kg body weight, following protocols (Antonson et al., 2017).

Following anesthetization, pigs were euthanized using an intracardiac injection of sodium pentobarbital (86 mg/kg body weight, Fata Plus, Vortech Pharmaceuticals, Dearborn, MI, United States). Pig brains were extracted, the amygdalae were recognized using the stereotaxic atlas of the pig brain (Felix et al., 1999), dissected out, flash frozen on dry ice, and stored at −80°C following published protocols (Antonson et al., 2019). RNA was isolated using EZNA isolation kit following the manufacturer's instructions (Omega Biotek, Norcross, GA, United States). The RNA integrity numbers of the samples were above 7.5, indicating low RNA degradation. The RNA-Seq libraries were prepared with TruSeq Stranded mRNAseq Sample Prep kit (Illumina Inc., San Diego, CA, United States). The libraries were quantitated by qPCR and sequenced on one lane on a NovaSeq 6000 for 151 cycles from each end of the fragments using NovaSeq S4 reagent kit. FASTQ files were generated and demultiplexed with the bcl2fastq v2.20 conversion software. Paired-end reads (150 nt long) were obtained, and the FASTQ files are available in the National Center for Biotechnology Information Gene Expression Omnibus (GEO) database (experiment accession number GSE149695).

RNA Sequence Mapping and Differential Expression Analysis

The average Phred quality score of the reads assessed using FastQC (Andrews, 2010) was > 35 across all read positions, and therefore, no reads were trimmed. The paired-end reads from the individual samples were aligned to the *Sus scrofa* genome (version Sscrofa 11.1; Pruitt et al., 2007) using kallisto v0.43.0 (Bray et al., 2016) with default settings. The normalized (trimmed mean of *M*-values) gene expression values were described using a generalized linear model encompassing the effects of the MIA group (MPA or CON levels), sex (Fe or Ma levels), and MIA-by-sex interaction and analyzed using edgeR (version 3.14.0) in the R v. 3.3.1 environment (Robinson et al., 2010). Genes supported by > 5 transcripts per million (TPM) by each MIA–sex combination were analyzed to ensure adequate representation across comparisons.

Orthogonal pairwise contrasts between MIA and sex groups were evaluated in addition to testing for the effects of MIA-by-sex interaction and main effects of MIA and sex. The four groups compared in the contrasts, identified by treatment followed by the sex levels, are: MPA_Fe, MPA_Ma, CON_Fe, and CON_Ma. The *P*-values were adjusted for multiple testing using the Benjamini–Hochberg false discovery rate (FDR) approach (Benjamini and Hochberg, 1995).

Functional Enrichment and Network Inference

Two complementary approaches were used to identify over-represented functional categories among the genes exhibiting differential expression across MIA and sex groups (Caetano-Anollés et al., 2015, 2016; Gonzalez-Pena et al., 2016a,b). Functional categories investigated included Gene Ontology (GO) biological processes (BPs), GO molecular functions (MF), and KEGG pathways. The Gene Set Enrichment Analysis (GSEA) approach implemented in the software package GSEA-P 2.0 (Subramanian et al., 2007) was used to identify category over-representation with gene over- and under-expressed while considering all genes analyzed. The normalized enrichment score (NES) of the categories in the Molecular Signature Database (MSigDB) was calculated using the maximum deviation of the cumulative sum based on the signed and standardized fold change. The statistical significance of the enrichment was assessed using the FDR-adjusted *P*-value computed from 1000 permutations.

The over-representation of functional categories was also evaluated among genes that exhibited a significant MIA-by-sex interaction or main effect using the Database for Annotation, Visualization and Integrated Discovery (DAVID 6.8) (Huang et al., 2009). The enrichment of Direct GO categories in the DAVID database was assessed. The *Sus scrofa* genome was used as the background for enrichment testing, and enrichment is reported using the Expression Analysis Systematic Explorer (EASE) score that was computed using a one-tailed jackknifed Fisher hypergeometric exact test. Functional categories were clustered based on gene annotation, and the statistical significance of clusters is summarized as the geometric mean of the -$\log_{10}$ EASE scores of the categories (Delfino et al., 2011; Serao et al., 2011; Delfino and Rodriguez-Zas, 2013).

Weighted Gene Co-expression Network Analysis and Gene Network Visualization

An approach complementary to the identification of differentially expressed genes was used to uncover co-expression networks using Weighted Gene Co-expression Network Analysis (WGCNA) version 1.68 (Langfelder and Horvath, 2008). The input data were voom-transformed read count values generated using the limma package (version 3.40.2) (Ritchie et al., 2015) in R (version 3.6.1). Genes were filtered to remove those with low expression levels or no variation across samples per developer recommendations. The number of genes used for network analysis was 16,175 genes. Considering potential for interaction patterns, a sex-dependent soft-thresholding power was used to call for network topology analysis. The lowest power values that support a scale-free topology power used were 15 for the CON_Ma-MPA_Ma contrast and 27 for the MPA_Fe-MPA_Ma contrast. The Pearson correlation coefficient of the normalized expression values was used to identify modules of connected genes. The minimum module size was set to 30, with the deepSplit set to 2, and the mergeCutHeight set to 0.15. Module profiles were identified using the correlation between the eigengene of each module and pig group. Enrichment of functional categories among the genes in each module profile was explored with DAVID using the *Sus scrofa* genome as background, and testing included an FDR multiple test adjustment.

Further understanding of the impact of the MIA-by-sex interaction was gained through the reconstruction of gene networks using the BisoGenet package (Martin et al., 2010) in the Cytoscape platform (Shannon et al., 2003). Information from gene and protein interactions annotated in databases including BIOGRID, HPRD, DIP, BIND, INTACT, and MINT was used to visualize relationships between genes (Salwinski et al., 2004; Alfarano et al., 2005; Mishra et al., 2006; Stark et al., 2006; Kerrien et al., 2007; Licata et al., 2012). Networks highlighting differences in gene levels associated with MIA within sex (i.e., the contrasts MPA_Ma-CON_Ma and MPA_Fe-CON_Fe) were compared. The network framework includes genes that exhibited a significant MIA-by-sex interaction effect (FDR-adjusted $P < 0.1$) and are annotated to enriched functional categories. The framework genes were identified by full nodes with size reflecting the differential expression level between the MPA and CON groups. The network edges depict known molecular relationships curated in the BisoGenet databases. The framework genes were connected through correlated genes listed in the BisoGenet database of molecular interactions that did not reach significant MIA-by-sex interaction effect. The comparison of these networks offered insights into the simultaneous effect of MIA across interacting genes and enabled the detection of shared and distinct co-regulation patterns between MPA and CON pigs across sexes.

RESULTS

Maternal Immune Activation and Sequencing Metrics

The differences between MPA and CON gilts in rectal temperatures and daily diet consumption indicated the activation of the maternal immune system in response to PRRSV. The difference in body temperature between CON and MPA gilts on GD 87 was -1.00°C (standard error 0.35°C; $P < 0.005$). The difference in feed refusal between CON and MPA gilts on GD 88 was -927.6 g (standard error 201.2 g; $P < 0.0001$). A significant increase in rectal temperatures and decrease in feed intake ($P < 0.001$) was observed within 48 h of inoculation and returned to baseline levels within 10 days for body temperature and within 14 days for feed intake. At 21 days of age, CON pigs were 1.20 kg

heavier than MPA pigs (standard error = 0.5673; $P < 0.089$) while no significant sex or interaction effects were detected.

The sequencing of the 24 RNA samples produced 6.6 billion sequenced reads, and 69 million paired-end reads per sample. The number of reads was consistent across MIA and sex groups (coefficient of variation < 0.1), and the effects of MIA, sex, and MIA-by-sex interaction were tested on 16,175 genes that surpassed the minimum number of reads per MIA–sex combination.

Transcriptome Changes Associated With Maternal Immune Activation That Are Sex-Dependent

Overall, 328 genes exhibited a significant (FDR-adjusted $P < 0.1$) MIA-by-sex interaction effect, and among these, 273 genes had a significant effect at FDR-adjusted $P < 0.05$. The profile of these genes indicated that the effect of MIA differed between females and males. Forty-six genes that presented a MIA-by-sex interaction effect are listed in **Table 1** together with their expression pattern and *P*-value. The majority of the genes in **Table 1**, including neurotensin (NTS), displayed a reversal in the expression level between CON and MPA groups across sexes (i.e., opposite Log_2[fold change] sign across sexes). An extended list including 328 genes that exhibited a MIA-by-sex interaction effect at FDR-adjusted $P < 0.1$ is provided in **Supplementary File S1 Table A**.

Another frequent pattern among the genes that displayed a MIA-by-sex interaction effect was characterized by a consistent expression profile between CON and MPA across sexes, albeit the magnitude differed between sexes (**Table 1**). For example, glycoprotein hormones, alpha polypeptide (CGA) was over-expressed in CON relative to MPA, but the differential was higher in males than in females. Other genes presenting this pattern included guanylate-binding protein 1 (GBP1), transthyretin (TTR), aldehyde dehydrogenase 1 family member A2 (ALDH1A2), hemoglobin subunit beta (HBB), and basic helix-loop-helix family member e22 (BHLHE22).

Notable is the significant MIA-by-sex interaction effect on genes associated with neuropeptides and hormones, and genes that participate in glutamatergic processes. Genes under-expressed in MPA relative to CON males while presenting the opposite pattern in females (**Table 1**) included NTS, the neuropeptide gene proenkephalin (PENK), the neuropeptide gene gastrin-releasing peptide (GRP), the neuropeptide-related gene vasoactive intestinal peptide receptor 2 (VIPR2), corticotropin releasing hormone receptor 2 (CRHR2), neuron-derived neurotrophic factor (NDNF), reelin (RELN), glutamate metabotropic receptor 4 (GRM4), solute carrier family 17 member 6 (SLC17A6), calcium voltage-gated channel auxiliary subunit alpha 2 delta 3 (CACNA2D3), EF-hand domain family member D1 (EFHD1), glutathione peroxidase 3 (GPX3), parathyroid hormone 1 receptor (PTH1R), thyroid hormone responsive (THRSP), and CGA. The CGA gene codes for the alpha subunit protein of the hormones chorionic gonadotropin (CG), luteinizing hormone (LH), follicle-stimulating hormone (FSH), and thyroid-stimulating hormone (TSH).

Functional and Network Analysis of Genes That Exhibit Sex-Dependent Associations With Maternal Immune Activation

The genes expressing significant MIA-by-sex interaction effects were analyzed for functional enrichment. **Table 2** presents the clusters of most enriched and informative categories from the DAVID analysis, and the complete list of categories is in **Supplementary File S1 Table B**. The categories in **Table 2** encompass genes presenting the most frequent interaction profile characterized by under-expression in CON females relative to males but over-expression in MPA females relative to males. These genes include KEGG Autoimmune thyroid disease (Cluster 1) and BP brain development (GO:0007420) (Cluster 4).

Enrichment results from GSEA complemented the findings from DAVID. Highly enriched informative categories among genes that have a MIA-by-sex interaction effect are presented in **Table 3**, and the extended list of categories is presented in **Supplementary File S1 Table C**. The categories in **Table 3** support pathways in **Table 2** including ion homeostasis (**Table 2**) and regulation of voltage-gated calcium channel activity processes (**Table 3**). Notably, the enrichment of the neuroactive ligand receptor interaction pathway and the hormone and neuropeptide activity processes include genes such as CGA and VIPR2 that were identified in **Table 1**.

Network visualization furthered the understanding of the impact of MIA on the relationships among genes that exhibited a significant MIA-by-sex interaction effect. The networks in **Figures 1**, **2** depict the relationships between genes in the enriched neuroactive ligand receptor pathway that highlight the differential expression between CON and MPA in males and females (i.e., CON_Ma-MPA_Ma and CON_Fe-MPA_Fe contrasts), respectively. Red and blue rectangular nodes represent framework genes, and edges represent the known associations between genes based on curated databases of molecular interactions. Red and blue nodes denote over- or under-expression of the gene in CON relative to MPA, and the size is an inverse logarithmic function of the differential expression *P*-value. The simultaneous study of the differential expression pattern and connectivity among genes highlights the discrepancy in network modules elicited by MIA between the sexes.

Transcriptome Changes Associated With Maternal Immune Activation

Overall, genes exhibited differential (FDR-adjusted $P < 0.1$) expression between MPA and CON pigs, irrespective of sex. **Table 4** lists notable highly differentially expressed genes, and the complete list is in **Supplementary File S1 Table D**. The majority of these genes were over-expressed in MPA relative to CON pigs. Among the genes over-expressed in MPA compared to CON pigs were islet amyloid polypeptide (IAPP),

TABLE 1 | Genes exhibiting significant (FDR-adjusted P-value < 0.1) maternal immune activation-by-sex interaction effect.

Gene symbol	*P*-value	[a]CON Fe-CON Ma	MPA Fe-MPA Ma	CON Fe-MPA Fe	CON Ma-MPA Ma	CON Fe-MPA Ma	CON Ma-MPA Fe
RGS16	<5E-11	−3.25	2.24	−2.44	3.06	−0.19	0.81
CGA	<5E-11	−5.86	0.36	0.27	6.49	0.63	6.13
POMC	<5E-11	−2.79	0.33	0.06	3.19	0.39	2.86
GPX3	<5E-11	−1.19	0.99	−0.37	1.81	0.63	0.82
RELN	<5E-11	−0.17	0.66	−0.74	0.09	−0.08	−0.57
VIPR2	<5E-11	−1.17	1.04	−1.03	1.18	0.00	0.14
ANKRD34C	<5E-11	−0.77	0.87	−0.71	0.93	0.16	0.06
GBP1	<5E-11	0.92	−0.32	−0.75	−1.99	−1.07	−1.66
GRM4	<5E-11	−0.97	0.91	−0.70	1.17	0.20	0.27
CCDC136	5.3E-09	−0.39	0.43	−0.15	0.67	0.29	0.24
SLC17A6	1.1E-08	−0.40	0.54	−0.16	0.77	0.37	0.23
BTBD11	4.4E-08	−0.62	0.45	−0.40	0.67	0.05	0.22
TTR	5.0E-08	−0.47	0.24	0.29	1.00	0.53	0.76
CACNA2D3	4.8E-07	−0.50	0.44	−0.26	0.68	0.18	0.23
CRHR2	2.8E-06	−0.24	1.09	−0.36	0.97	0.74	−0.12
NDNF	2.8E-06	−0.09	0.88	−0.70	0.27	0.18	−0.61
CXCL12	6.2E-06	−0.41	−0.01	−0.10	0.29	−0.11	0.30
USP43	6.4E-06	−0.64	0.43	−0.34	0.72	0.09	0.29
CCDC17	7.1E-06	0.74	−0.30	0.26	−0.40	0.03	0.12
KCNIP4	7.2E-06	−0.01	0.35	0.11	0.47	0.46	0.11
CAMK2N2	9.6E-06	−0.24	0.16	−0.23	0.18	−0.07	0.01
ALDH1A2	1.4E-05	−1.35	0.50	0.48	2.32	0.98	1.83
GRP	1.5E-05	−0.89	0.73	−0.61	1.02	0.12	0.28
PENK	1.6E-05	−0.10	0.46	−0.04	0.51	0.42	0.06
SYT12	2.9E-05	−0.16	0.32	−0.11	0.37	0.21	0.05
PTH1R	3.7E-05	−0.33	0.52	−0.51	0.34	0.01	−0.18
HBB	6.7E-05	−0.48	0.10	0.12	0.70	0.22	0.60
ESYT1	8.5E-05	−0.47	0.29	−0.27	0.49	0.02	0.20
EFHD1	9.6E-05	−0.31	0.42	−0.37	0.36	0.05	−0.06
BHLHE22	1.0E-04	0.06	−0.59	0.05	−0.61	−0.55	−0.01
ZFP37	1.2E-04	0.60	0.60	−0.19	−0.18	0.41	−0.78
SLC2A2	1.5E-04	0.08	0.50	−0.04	0.38	0.46	−0.12
THRSP	3.1E-04	−0.72	0.51	−1.24	−0.01	−0.73	−0.52
NR4A3	3.3E-04	−0.12	−0.01	−0.08	0.04	−0.08	0.04
LOC396781	4.4E-04	−0.75	1.36	−5.38	−3.27	−4.01	−4.63
C1QTNF1	4.5E-04	−0.37	0.32	−0.23	0.46	0.09	0.14
RAB27A	5.6E-04	−0.97	0.05	−0.25	0.77	−0.20	0.72
NTS	7.9E-04	−1.43	2.81	−1.78	2.47	1.03	−0.35
GVIN1	7.9E-04	0.80	−0.41	−1.01	−2.22	−1.42	−1.81
SSTR1	8.6E-04	0.11	−0.38	0.16	−0.33	−0.22	0.05
CCDC9B	8.8E-04	−0.24	0.35	−0.34	0.25	0.01	−0.10
CCDC33	1.4E-03	0.48	−0.39	0.49	−0.38	0.10	0.01
CCDC162P	1.4E-03	0.11	−0.48	0.36	−0.23	−0.12	0.25
PTH	1.4E-03	−0.03	0.29	−0.68	−0.36	−0.39	−0.65
SYNPO2L	1.5E-03	−0.28	0.59	−0.44	0.43	0.15	−0.16
CHGB	1.8E-03	−0.58	−0.16	0.37	0.79	0.21	0.96

[a] *Log_2[fold change] between two maternal immune activation-sex groups: MPA, PRRSV-induced maternal immune activation; CON, control; Fe, females; Ma, males.*

ankyrin repeat domain 24 (ANKRD24), interferon-induced transmembrane protein 1 (IFITM1) and 3 (IFITM3), cathepsin C (CTSC), mitogen-activated protein kinase kinase 7 (MAP2K7), heparan sulfate-glucosamine 3-sulfotransferase 5 (HS3ST5), secreted phosphoprotein 1 (SPP1), immunoglobulin heavy chain (IGHG), and transforming acidic coiled-coil-containing protein 1 (TACC1). Among the genes under-expressed in MPA relative to CON pigs are insulin-like growth factor 2 (IGF2), cellular retinoic acid-binding protein 2 (CRABP2), and aldehyde dehydrogenase 1 family member A1 (ALDH1A1).

Functional Analysis of Genes Associated With Maternal Immune Activation

Table 5 presents the top significant clusters of informative enriched categories from the DAVID analysis of genes differentially expressed between MPA and CON groups across sexes (the extended list of categories is presented in **Supplementary File S1 Table E**). Some categories identified by the DAVID analysis are consistent with the categories detected at more significant levels among the genes presenting an MIA-by-sex interaction effect (**Table 2**) and include the BP angiogenesis (GO:0001525) and KEGG autoimmune thyroid disease and Epstein–Barr virus infection pathways (**Table 5**). Also enriched (**Supplementary File S1 Table E**) were the BP homeostatic (GO:0042592), MF ion binding (GO:0043167), and BP anatomical structure formation in morphogenesis (GO:0048646).

The GSEA enrichment results within the gene expression patterns of CON relative to MPA groups complemented the findings from DAVID. The most informative enriched categories are presented in **Table 6**, and the extended list of categories is presented in **Supplementary File S1 Table F**. Enriched clusters of genes over-expressed in CON relative to MPA detected by GSEA were the BP enrichment of microtubule bundle formation (GO:0001578) and cilium morphogenesis (GO:0060271).

TABLE 2 | Most enriched DAVID clusters and supporting functional categories (enrichment score ES > 1.3) among the genes presenting significant maternal immune activation-by-sex interaction effect.

[a]Category	Category identifier and name	*P*-value	[b]FDR *P*-value
Cluster 1	**ES = 2.74**		
KEGG	ssc05320:Autoimmune thyroid disease	2.90E-06	4.50E-04
BP	GO:000250~Antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	1.90E-03	3.40E-01
KEGG	ssc04514:Cell-adhesion molecules (CAMs)	2.30E-03	3.20E-02
KEGG	ssc05323:Rheumatoid arthritis	4.80E-03	5.60E-02
KEGG	ssc05164:Influenza A	2.00E-02	1.70E-01
Cluster 2	**ES = 1.9**		
BP	GO:0051050~Positive regulation of transport	4.40E-03	3.50E-01
BP	GO:0051049~Regulation of transport	4.50E-03	3.20E-01
BP	GO:0050801~Ion homeostasis	1.30E-02	3.60E-01
BP	GO:0048878~Chemical homeostasis	2.80E-02	4.80E-01
BP	GO:0030001~Metal ion transport	4.50E-02	5.70E-01
Cluster 3	**ES = 1.75**		
BP	GO:0048871~Multicellular organismal homeostasis	2.00E-04	3.70E-01
Cluster 4	**ES = 1.69**		
BP	GO:0061564~Axon development	6.40E-05	1.70E-01
BP	GO:0007420~Brain development	1.30E-03	3.10E-01

[a]BP, biological process; KEGG, KEGG pathway. [b]False discovery rate adjusted P-value.

TABLE 3 | Enriched informative categories (NES > |1.3|) using GSEA among the genes based on the overall maternal immune activation-by-sex interaction.

[a]Category	Category identifier and name	[b]NES	*P*-value	[c]FDR *P*-value
KEGG	ssc04080:Neuroactive ligand receptor interaction	−1.84	<1.0E-10	8.3E-02
KEGG	ssc04912:GnRH signaling pathway	−1.83	<1.0E-10	9.9E-02
MF	GO:0005179~Hormone activity	−1.80	<1.0E-10	2.2E-01
BP	GO:0006970~Response to osmotic stress	−1.79	<1.0E-10	3.5E-01
BP	GO:0019221~Cytokine mediated signaling pathway	−1.33	9.1E-02	5.4E-01
BP	GO:1901385~Regulation of voltage gated calcium channel activity	−1.37	1.2E-01	5.4E-01
KEGG	ssc04020:Calcium signaling pathway	−1.34	1.2E-01	5.4E-01
BP	GO:0085029~Extracellular matrix assembly	−1.31	1.8E-01	5.5E-01

[a]MF, molecular function; KEGG, KEGG pathway; BP, biological process. [b]Normalized enrichment score; negative values indicate genes under-expression in CON females relative to males but over-expression in MPA females relative to males. [c]False discovery rate adjusted P-value.

Transcriptome Differences Between Sexes Independent of Maternal Immune Activation

Overall, 150 genes were differentially expressed between males and females (FDR-adjusted $P < 0.05$). These genes exhibited a consistent differential expression between sexes, irrespective of the MIA group. The complete list of genes differentially expressed between sexes at FDR-adjusted $P < 0.1$ is available in **Supplementary File S1 Table G**, and the majority were over-expressed in males relative to females. Among the previous genes, excluding those that presented MIA-by-sex interaction effect, a selection of informative genes is listed in **Table 7**. Genes over-expressed in males relative to females included eukaryotic translation initiation factor 1A, Y-linked (EIF1AY), leptin receptor (LEPR), luteinizing hormone beta polypeptide (LHB), LIM homeobox 9 (LHX9), luteinizing hormone beta polypeptide (LHB), and immunoglobulin family member 1 (IGSF1).

Informative categories among the DAVID clusters of enriched categories for the genes differentially expressed between sexes are listed in **Table 8** (a complete list is available in **Supplementary File S1 Table H**). The previous categories include BP gland development (GO:0048732), response to hormone (GO:0009725), and brain development (GO:0007420).

Weighted Gene Co-expression Network Analysis

The WGCNA study of the correlation between expression and experimental factors was based on the log-transformed TPM expression level of 16,175 genes. This study identified 62 modules of expression correlated with MIA groups among males. The number of genes in each module ranges from 32 to 1381; **Supplementary File S1 Figure A** depicts the relationship between gene modules using a dendrogram. The correlation (corr) between the eigengene expression profile and maternal treatment in males is depicted in **Supplementary File S1 Figure B**. A fairly

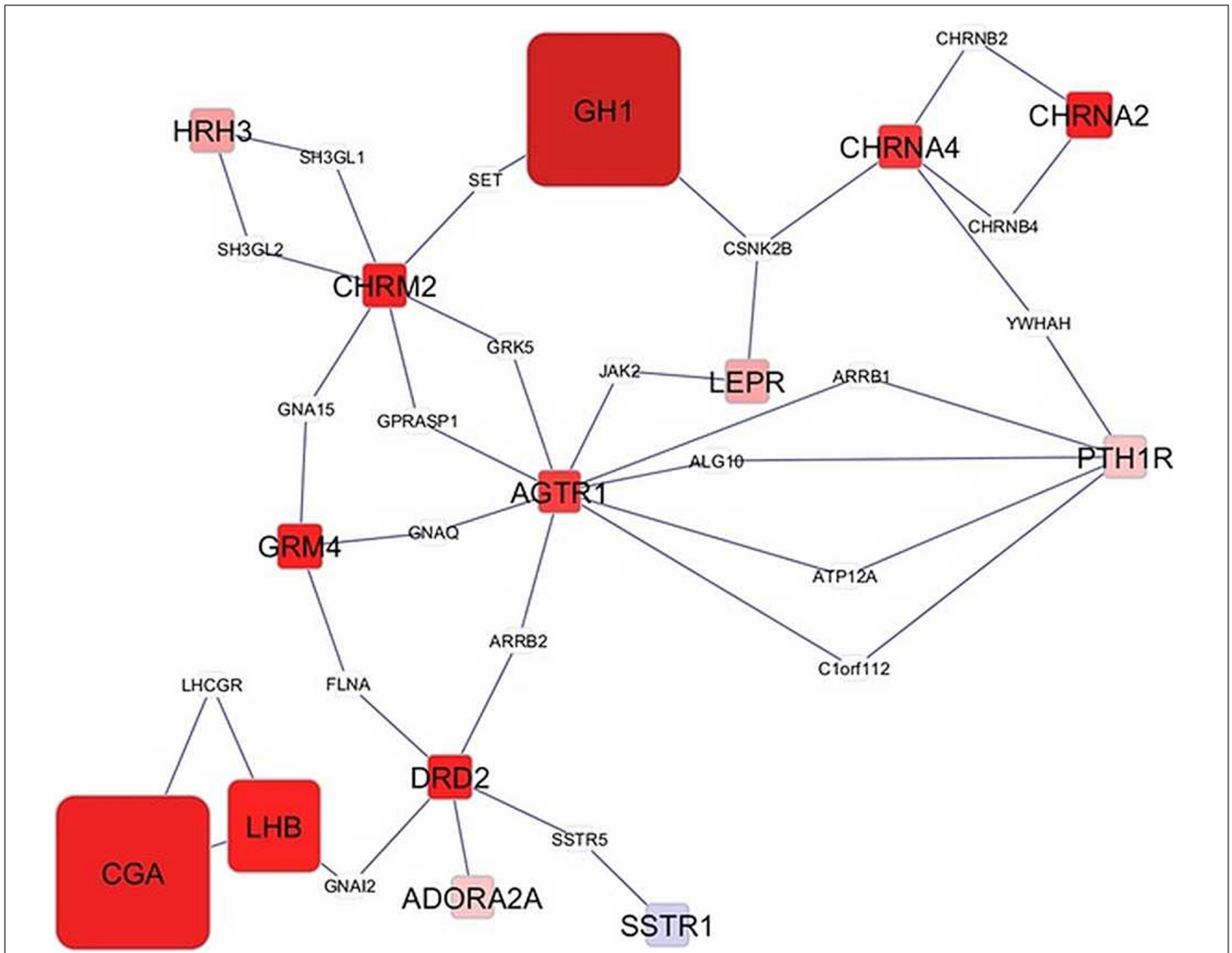


FIGURE 1 | Network of genes differentially expressed in the amygdala of males (Ma) from control (CON) relative to maternal immune activation (MPA) (contrast between CON_Ma and MPA_Ma). Framework square node color: red and blue denote framework genes over- and under-expressed in CON relative to MPA pigs, respectively; framework node size: -Log_{10}[*P*-value]; other genes connecting framework nodes were not differentially expressed in this study.

even distribution of positive and negative correlation estimates was observed. The genes in the modules pale violet red (corr = –0.58, $P < 0.03$) and gray60 (corr = –0.59, $P < 0.06$) presented a strong negative correlation, indicating that low expression levels were associated with MPA.

The enrichment analysis of the genes in the WGCNA modules provided additional insights into the processes impacted by MIA. The extended list of enriched categories across modules is available in **Supplementary File S1 Table I**. **Table 9** lists representative clusters of enriched categories in the gene modules that were highly correlated with MIA differences in males. Complementary categories that were enriched in modules encompassing gene patterns negatively correlated with MPA include the KEGG AD and oxidative phosphorylation pathways in the pale violet red3 and the gray60 modules, MF NADH dehydrogenase activity (GO:0003954) in the gray60 module, and the KEGG ribosomal pathway and associated GO processes in the light yellow module. Confirming the enrichment results from the previous differential expression analysis (**Table 2** and **Supplementary File S1 Table B**), categories enriched in modules encompassing gene patterns negatively correlated with MPA include ion transport (also in **Table 2** and **Supplementary File S1 Table H**) in the pale violet red3 and gray60 modules, and oxidoreductase activity in the gray60 module (also in **Supplementary File S1 Table H**).

The WGCNA study identified 61 modules of expression correlated with sex among MPA pigs. The number of genes in each module range from 34 to 2082 and **Supplementary File S1 Figure C** depicts the dendrogram of modules. Positive correlation estimates that indicate higher expression in male than female were identified in the ivory (corr = 0.65, $P < 0.03$) and antique maroon (corr = 0.62, $P < 0.04$) modules, whereas negative correlations denoting lower expression in male were identified in sienna3 (corr = –0.65, $P < 0.03$)

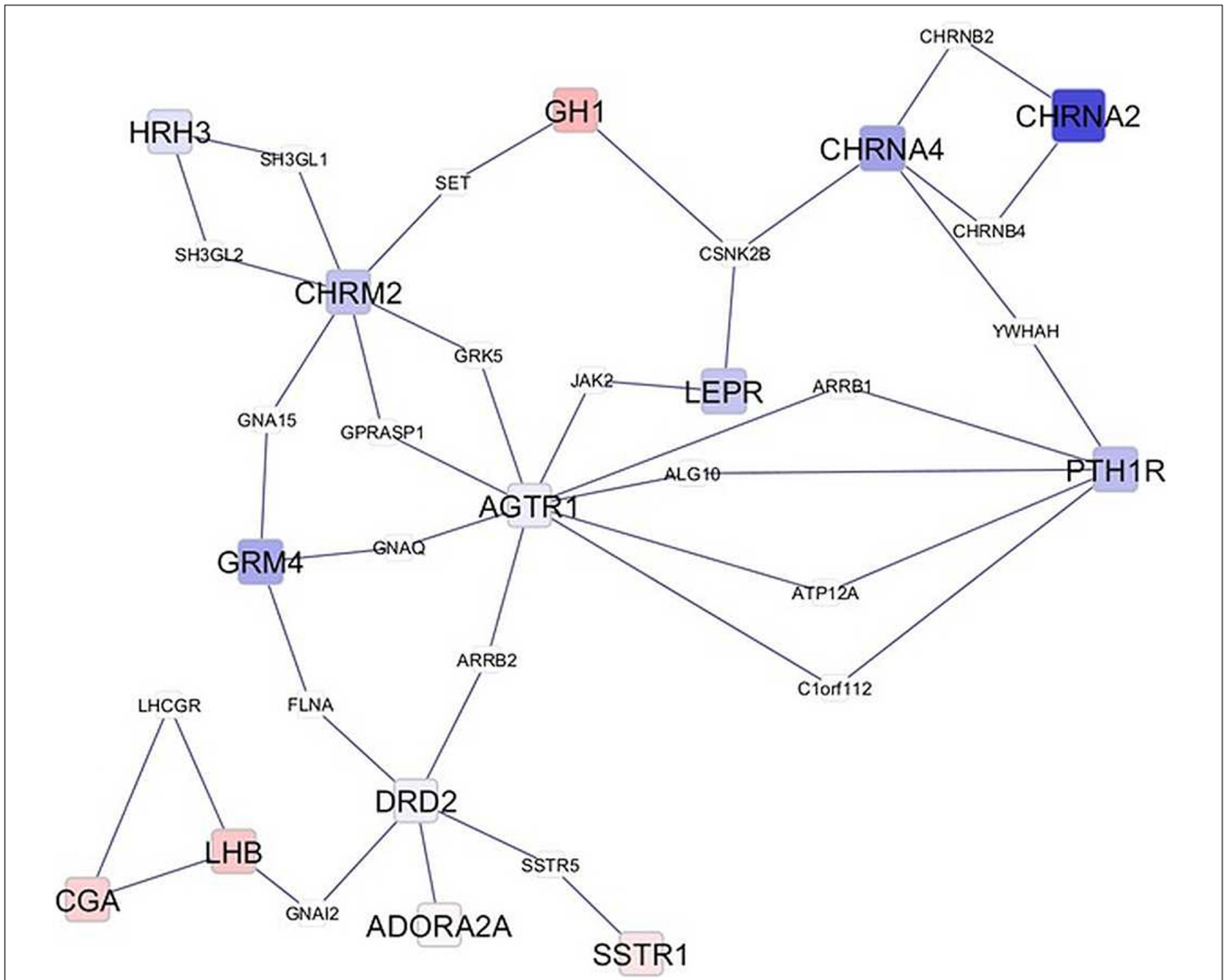


FIGURE 2 | Network of genes differentially expressed in the amygdala of females (Fe) from control (CON) relative to maternal immune activation (MPA) (contrast between CON_Fe and MPA_Fe). Framework square node color: red and blue denote framework genes over- and under-expressed in CON relative to MPA pigs, respectively; framework node size: -Log_{10}[*P*-value]; other genes connecting framework nodes were not differentially expressed in this study.

(**Supplementary File S1 Figure D**). **Table 10** lists representative clusters of enriched functional categories in these modules. The extended list of categories and modules is in **Supplementary File S1 Table J**. Within the ivory module of positively correlated gene patterns, the BP synapse (GO:0007416) and neural development (GO:0007399) categories were enriched.

DISCUSSION

The present study uses a validated model of MIA triggered by a live viral (i.e., PRRSV) infection during a critical neurodevelopmental stage (Antonson et al., 2017) to gain innovative insights into sex-dependent molecular changes in the amygdala. The changes in gilt body weight and temperature within 2 weeks postinfection were consistent with the mode of action of the PRRSV and previous reports (Antonson et al., 2017). These results indicate that the PRRSV-infected gilts experienced extended activation of inflammatory pathways during gestation. The present study characterized the prolonged effect of the MIA, 60 days after exposure, on 3-week-old offspring.

Our study identified patterns of differential gene expression and prevalently dysregulated gene networks and processes, some of which have been reported in clinical and preclinical studies of AD, ASD, and SSD (Malkova et al., 2012; Xuan and Hampson, 2014; Aavani et al., 2015). These disorders have been associated with MIA and amygdala functions, yet the corresponding neurological and molecular changes have been studied mostly using pathogen mimetic challenges on rodents (Xuan and Hampson, 2014; Aavani et al., 2015). Moreover, our study identified sex-dependent molecular patterns that are consistent with the differential prevalence and symptoms of

TABLE 4 | Representative genes differentially expressed (FDR-adjusted *P*-value < 0.1) between pigs from control (CON) relative to PRRSV-treated (MPA) gilts.

Gene symbol	Gene name	[a]CON-MPA	*P*-value	[b]FDR *P*-value
IGHG	IgG heavy chain	−4.74	1.1E-31	<1.0E-10
IFITM3	Interferon induced transmembrane prot 3	−1.27	7.6E-11	6.2E-08
IGF2	Insulin-like growth factor 2	1.28	8.1E-07	2.7E-04
PAQR6	Progestin and adipoQ receptor family member 6	−1.02	1.9E-06	5.6E-04
RGS8	Regulator of G protein signaling 8	−0.88	8.6E-06	2.0E-03
NDNF	Neuron-derived neurotrophic factor	−1.27	1.5E-05	3.3E-03
HS3ST5	Heparan sulfate-glucosamine 3-sulfotransferase 5	−1.24	3.5E-05	6.8E-03
CTSC	Cathepsin C	−1.03	4.1E-05	7.7E-03
SPP1	Secreted phosphoprotein 1	−0.89	4.1E-05	7.7E-03
TACC1	Transforming acidic coiled-coil-containing protein 1	−1.55	5.5E-05	9.7E-03
IFITM1	Interferon induced transmembrane prot 1	−1.15	7.4E-05	1.2E-02
ALDH1A1	Aldehyde dehydrogenase 1 fam mem A1	0.79	9.7E-05	1.5E-02
PLEKHD1	Pleckstrin homology coiled-coil domain-containing D1	−0.99	5.6E-04	6.5E-02
HERC5	HECT and RLD domain-containing E3 ubiquitin ligase 5	−1.05	5.7E-04	6.6E-02
IAPP	Islet amyloid polypeptide	−1.36	7.0E-04	7.7E-02
ISLR	Immunoglobulin superfamily contain leucine rich repeat	1.13	8.3E-04	8.5E-02
ANKRD24	Ankyrin repeat domain 24	−0.73	8.7E-04	8.9E-02
MAP2K7	Mitogen-activated protein kinase kinase 7	−0.64	9.9E-04	1.0E-01

[a]Log_2[fold change] between CON and MPA pigs. [b]False discovery rate adjusted P-value.

TABLE 5 | Clusters of enriched functional categories (enrichment score ES > 1.3) among the genes presenting significant maternal immune activation effect, and representative categories identified using DAVID.

[a]Category	Category identifier and name	*P*-value	[b]FDR *P*-value
Cluster 1	**ES = 1.51**		
BP	GO:0048646∼anatomic structure formation in morphogenesis	3.7E-03	9.7E-01
BP	GO:0001525∼angiogenesis	8.4E-02	9.8E-01
Cluster 2	**ES = 1.40**		
KEGG	ssc05330:Allograft rejection	1.7E-02	8.0E-01
KEGG	ssc05169:Epstein–Barr virus infection	2.3E-02	6.7E-01
KEGG	ssc05320:Autoimmune thyroid disease	3.0E-02	6.1E-01

[a]BP, biological process; KEGG, KEGG pathway. [b]False discovery rate adjusted P-value.

ASD, SSD, and MIA-related disorders between sexes reported in rodent and human studies (Wischhof et al., 2015). For example ASD tends to be more prevalent in young males (Kirsten et al., 2010; Haida et al., 2019), while SSD tends to be more prevalent in females (Bale et al., 2010; Bale, 2011). Similarly, lower sociability and preference for social novelty were observed in 2 weeks old pigs from gilts inoculated with PRRSV at GD 76 than from control gilts (Antonson et al., 2017). A discussion of the molecular mechanisms impacted by MIA can offer insights into therapies to ameliorate the lasting effects on physiology and behavior.

TABLE 6 | Enriched informative categories (NES > |1.3|) among the genes differentially expressed between pigs from the control relative to the maternal immune activation group using GSEA.

[a]Category	Category identifier and name	[b]NES	*P*-value	[c]FDR *P*-value
BP	GO:0001578∼Microtubule bundle formation	2.03	<1.0E-10	<1.0E-08
BP	GO:0060271∼Cilium morphogenesis	2.01	<1.0E-10	<1.0E-08
BP	GO:0044782∼Cilium organization	1.97	<1.0E-10	<1.0E-08
BP	GO:0035082∼Axoneme assembly	1.94	<1.0E-10	1.8E-04
BP	GO:0003341∼Cilium movement	1.94	<1.0E-10	1.4E-04

[a]BP, biological process. [b]Normalized enrichment score, positive values refer to genes under-expressed in maternal immune activated relative to control pigs. [c]False discovery rate adjusted P-value.

TABLE 7 | Informative genes presenting significant differential expression between males (Ma) and females (Fe).

Gene symbol	Gene name	[a]Ma-Fe	*P*-value	[b]FDR *P*-value
EIF1AY	Eukaryotic translation initiation factor 1A, Y-linked	13.68	<1.0E-08	<1.0E-08
LHX9	LIM homeobox 9	3.38	<1.0E-08	<1.0E-08
LHB	Luteinizing hormone beta polypeptide	3.92	<1.0E-08	<1.0E-08
TRPC3	Transient receptor potential cation channel C3	1.83	1.7E-07	4.4E-05
WNT3	Wnt family member 3	1.72	9.0E-07	2.0E-04
IGSF1	Immunoglobulin superfamily member 1	0.85	6.4E-06	1.3E-03
NPM2	Nucleophosmin/nucleoplasmin 2	1.79	1.8E-05	2.8E-03
RORA	RAR-related orphan receptor A	0.74	2.0E-04	2.4E-02
LEPR	Leptin receptor	0.62	8.8E-04	8.4E-02

[a]Log_2[fold change] between male and female offspring. [b]False discovery rate adjusted P-value.

Sex-Dependent Transcriptome Changes Associated With Maternal Immune Activation

The evaluation of the 328 genes presenting a significant MIA-by-sex interaction effect augmented the understanding of the differential response of transcripts to MIA between sexes (**Table 1** and **Supplementary File S1 Table A**). The majority of the previous genes were under-expressed in MPA relative

TABLE 8 | Clusters of informative enriched functional categories (enrichment score ES > 1.3) among the genes differentially expressed between sexes identified using DAVID.

[a]Category	Category identifier and name	*P*-value	[b]FDR *P*-value
Cluster 1	**ES = 1.82**		
MF	GO:0005201~extracellular matrix structural constituent	3.0E-02	7.9E-01
Cluster 2	**ES = 1.74**		
KEGG	ssc04080:Neuroactive ligand-receptor interaction	2.7E-06	2.4E-04
BP	GO:0009725~response to hormone	7.3E-04	9.4E-01
Cluster 3	**ES = 1.74**		
BP	GO:0048732~gland development	1.7E-03	5.7E-01
Cluster 4	**ES = 1.45**		
BP	GO:0007420~brain development	1.1E-01	7.9E-01
BP	GO:0030900~forebrain development	1.9E-01	8.7E-01
Cluster 5	**ES = 1.41**		
MF	GO:0020037~heme binding	6.2E-02	9.1E-01
MF	GO:0005506~iron ion binding	9.3E-02	9.3E-01
Cluster 6	**ES = 1.35**		
BP	GO:0009790~embryo development	5.8E-03	5.4E-01
BP	GO:0090596~sensory organ morphogenesis	8.9E-03	5.1E-01

[a]BP, biological process; MF, molecular function; KEGG, KEGG pathway. [b]False discovery rate adjusted P-value.

TABLE 9 | Clusters of enriched functional categories (enrichment score ES > 1.3) among the genes in modules presenting a significant correlation with maternal immune activation (MPA) relative to control within males using DAVID.

[a]Category	Category identifier and name	*P*-value	[b]FDR *P*-value
MODULE	**Gray60 (low expression in MPA)**		
Cluster 1	**ES = 14.5**		
KEGG	Ssc05010:Alzheimer's disease	1.3E-18	9.1E-17
KEGG	Ssc05012:Parkinson's disease	9.8E-18	4.7E-16
Cluster 2	**ES = 7.32**		
MF	GO:0016651~oxidoreductase activity, acting on NAD(P)H	3.6E-10	1.5E-07
MF	GO:0003954~NADH dehydrogenase activity	3.4E-09	6.8E-07
Cluster 3	**ES = 6.41**		
BP	GO:0042775~mitochondrial ATP synthesis electron	4.0E-10	8.6E-07
BP	GO:0046034~ATP metabolic process	8.0E-10	8.7E-07
BP	GO:0009116~nucleoside metabolic process	7.8E-08	9.4E-06
MODULE	**Light yellow (low expression in MPA)**		
Cluster 1	**ES = 4.18**		
KEGG	ssc03010:Ribosome	3.7E-06	5.9E-05
BP	GO:0006412~translation	9.5E-06	1.7E-02
Cluster 1	**ES = 2.49**		
BP	GO:0007006~mitochondrial membrane organization	5.1E-04	7.8E-02
BP	GO:0007007~inner mitochondrial organization	1.1E-02	3.5E-01

[a]BP, biological process; MF, molecular function; KEGG, KEGG pathway. [b]False discovery rate adjusted P-value.

TABLE 10 | Clusters of enriched functional categories (enrichment score ES > 1.3) among the genes in modules presenting a significant correlation with sex within the maternal immune activation treatment using DAVID.

[a]Category	Category identifier and name	*P*-value	[b]FDR *P*-value
MODULE	**Sienna3 (low expression in males)**		
Cluster 1	**ES = 1.58**		
BP	GO:0044255~cellular lipid metabolic process	1.1E-02	1.0E+00
BP	GO:0044283~small molecule biosynthetic process	3.3E-02	9.8E-01
MODULE	**Ivory (high expression in males)**		
Cluster 1	**ES = 1.47**		
BP	GO:0007416~synapse assembly	8.1E-03	1.0E+00
BP	GO:0007399~nervous system development	2.7E-02	9.1E-01

[a]BP, biological process; MF, molecular function; KEGG, KEGG pathway. [b]False discovery rate adjusted P-value.

to CON males, and the profile in females was opposite or less extreme. Many genes presenting a significant MIA-by-sex effect code for neuropeptides and hormones, or participate in glutamatergic processes.

The lower NTS levels in the amygdala of male rats associated with lower conditioned place preference (Laszlo et al., 2010) is consistent with the lower level of NTS transcripts in MPA males observed in the present study (**Table 1**). The under-expression of neuropeptide gene POMC in MPA relative to CON males (**Table 1**) may be associated with the changes in POMC-related peptide transmission that has been reported in the brains of patients diagnosed with SSD (Jamali and Tramu, 1997). The under-expression of the neuropeptide gene PENK in MPA relative to CON males (**Table 1**) is in agreement with lower levels of PENK expression in the brains of mice models of SSD (Gottschalk et al., 2013). PENK was also differentially expressed in the amygdala of an ASD rat model using prenatal valproic acid exposure (Oguchi-Katayama et al., 2013). Valproic acid treatment during pregnancy was associated with a sevenfold increase in ASD incidence, social difficulties, and reduced attention (Roullet et al., 2013). Also, CACNA2D3 was under-expressed in the amygdala of rats exposed prenatally to valproic acid and coincided with social behavior abnormalities including heightened anxiety (Barrett et al., 2017). In the present study, the interaction pattern of CACNA2D3 was characterized by under-expression in MPA relative to CON males (**Table 1**).

The over-expression of the neuropeptide receptor VIPR2 in the amygdala of MPA relative to CON females (**Table 1**) is consistent with reports that duplications in this gene confer a significant risk for SSD (Morris and Pratt, 2014). The differential expression of the VIP receptor is particularly important because GABAergic interneurons in the amygdala express the neuropeptide VIP that facilitates cell firing (Rhomberg et al., 2018) and maintains the balance of pro- and anti-inflammatory

cytokines (Martinez et al., 2020). The under-expression of GPX3 in MPA relative to CON males (**Table 1**) is consistent with the association between GPX3 gene expression and SSD (Zhao et al., 2018).

The sex-dependent response to MIA of hormone receptor CRHR genes detected in our study (**Table 1**) has been also reported by others. The expression of CRHR1 in the amygdala of 10 weeks old female pigs from sows exposed to a social stressor during mid-gestation was higher than in pigs from control gilts, whereas no stressor effects were observed in males nor in the expression of CRHR2 (Rutherford et al., 2014). Also, the expression of CRHR was associated with SSD (Mistry et al., 2013). The alignment between the profiles of the CRHR2 and neuropeptide GRP genes observed in our study (**Table 1**) is in agreement with reports of simultaneous release of CRH and GRP in the amygdala of rats elicited by the stress hormone corticosterone (Merali et al., 2008). The over-expression of the parathyroid hormone receptor PTH1R gene in response to MIA (**Table 1**) is consistent with the over-expression of this gene in SSD and AD (Ibanez et al., 2014). The over-expression of the functionally related thyroid hormone responsive gene THRSP in MPA relative to CON females further supports the PTH1R pattern (Munshi and Rosenkranz, 2018).

Several genes in the glutamatergic and GABAergic pathways displayed sex-dependent MIA effects (**Table 1**). This shared pattern may stem from pro-inflammatory cytokines intensifying glutamatergic release by the amygdala in a sex-dependent manner. The glutamate receptor GRM4 was under-expressed in MPA relative to CON males, while a less extreme and opposite trend was detected in females (**Supplementary File S1 Table A**). Supporting our finding, the expression of glutamate receptor genes was lower in SSD brains (Meador-Woodruff and Healy, 2000). The pattern differences between sexes could be connected with differences in gene expression across sexes in multiple GRM genes including GRM4 in association with behavior disorders (Gray et al., 2015). The expression pattern of SLC17A6, a gene in the glutamatergic pathway, was consistent with that of GRM4 (**Table 1**). Moreover, NDNF and RELN (two genes in the GABAergic pathway) displayed similar expression patterns in our study in agreement with previous reports (Hou and Capogna, 2018). NDNF interneurons evoke inhibitory postsynaptic potentials mediated by GABA receptors. Also, the GABAergic gene RELN has been associated with ASD and SSD (Canitano and Pallagrosi, 2017) and is under-expressed in the prefrontal cortex of subjects diagnosed with SSD (Fatemi et al., 2005). Consistent with our results, genes that regulate neural migration of GABAergic interneurons were under-expressed in the brain of offspring from rats exposed to lipopolysaccharide (LPS)-induced MIA (Oskvig et al., 2012). The activity of GABA and glutamate on serotonergic neurons is modulated by chemokine ligands such as CXCL12 (Stuart and Baune, 2014), and consistent with this interaction, CXCL12 and CXCL13 were under-expressed in MPA relative to CON males (**Table 1** and **Supplementary File S1 Table A**). Our results are also consistent with findings that LPS injection of adult mice dysregulated CXCL12, which in turn increased glutamatergic release in the amygdala, and anxiety-like behavior occurrence (Yang et al., 2016). EFHD1, a calcium binding protein associated with synaptic transmission and levels of gamma glutamyltransferase was over-expressed in MPA relative to CON females (**Supplementary File S1 Table A**). This gene was also over-expressed in the amygdala of patients diagnosed with SSD (Chang et al., 2017).

Among the 16,175 genes tested for differential expression in the present study, the profile of several proinflammatory and neuroinflammatory genes were consistent with those from a candidate gene study of the amygdala from 9 days old mice exposed to MIA (Carlezon et al., 2019). Consistent with the patterns observed in mice exposed to viral mimetic Poly(I:C) MIA (Carlezon et al., 2019), an interaction between PRRSV-elicited MIA and sex was detected in pigs for the genes glial fibrillary acidic protein (GFAP), nitric oxide synthases 1 and 2 (NOS1 and NOS2, respectively), and translocator protein (TSPO)-associated protein ($0.006 < P < 0.02$). These patterns, albeit consistent, failed to surpass the FDR-adjusted $P < 0.05$ threshold. Also consistent with the MIA study of mice amygdala (Carlezon et al., 2019), the differences in expression between MPA and CON males for tumor necrosis factor alpha (TNF-α), and interleukin 1 beta (IL-1β), were not statistically significant. The expression levels for interleukin 6 (IL-6) and interleukin 1 beta (IL-1β) among MPA males were below the minimum threshold for testing, and therefore, the interaction effects for these genes are not reported. IL-6 and IL-1β were differentially expressed between MIA groups in female mice (Carlezon et al., 2019); however, the reported relative abundances for these genes suggests that MIA male mice, like MPA male pigs, presented the lowest levels of IL-6 and IL-1β abundance of all groups studied.

The pattern of several genes presenting a MIA-by-sex interaction effect was characterized by the same relative abundance between MIA groups, albeit sexes differed in magnitude (**Table 1**). Males presented a more extreme under-expression of CGA in MPA relative to CON than females (**Table 1**). This difference could be associated with the participation of CGA in multiple hormone processes that regulate female reproductive performance. The lower impact of MIA on the CGA levels in females may prevent the dysregulation of multiple downstream processes associated with reproductive function. Likewise, the under-expression of TTR in MPA relative to CON was more acute in males than in females (**Table 1**). Over-expression of TTR was noted in the amygdala of rats treated with MK-801, a N-methyl-D-aspartate antagonist that elicits SSD-like behavior (Matsuoka et al., 2008). The differential expressions of HBB and GBP1 among MIA groups detected in our study are also observed in the amygdala of SSD cases (Chang et al., 2017).

Functional Analysis of Sex-Dependent Maternal Immune Activation Transcriptome

The study of over-represented functional categories among the genes presenting sex-dependent profiles between MPA and CON pigs (**Tables 2**, **3** and **Supplementary File S1 Tables B**, **C**) identified categories consistent with previous studies of MIA and amygdala inflammation. The enrichment of the KEGG

pathways related to autoimmune disease and antigen processing and presentation via histocompatibility complex (MHC) (**Table 2** and **Supplementary File S1 Table B**) are in agreement with the reported association between autoimmune diseases, SSD, and variants in the MHC gene family (Anders and Kinney, 2015). Autoimmune reactions are capable of inducing psychiatric symptoms that are mediated by the amygdala such as those associated with SSD (Lennox et al., 2012). Genes annotated to MHC receptor activity are over-expressed in the amygdala of individuals that have SSD (Chang et al., 2017). The cytokine-mediated signaling pathway was also over-represented among the genes under-expressed in MPA relative to CON pigs (**Table 3**). Consistent with our results, the expression of 28 genes annotated to immune stimulus had MIA-by-sex interaction effects in the microglia of GD 97 fetuses after GD 76 PRRSV injection (Antonson et al., 2019).

Multiple BPs associated with homeostasis and extracellular matrix assembly were enriched among the genes presenting a significant MIA-by-sex effect (**Tables 2**, **3** and **Supplementary File S1 Tables B**, **C**). This finding is in agreement with the deficit in perineuronal nets in the amygdala of patients diagnosed with SSD (Paylor et al., 2016). These nets are extracellular matrix structures that support the high metabolic demand of the interneurons, and contribute to ion homeostasis around them.

The enrichment of BP axon and brain development among the genes presenting an MIA-by-sex effect (**Table 2** and **Supplementary File S1 Table B**) is consistent with a report that LPS-elicited MIA is associated with under-expression of neurodevelopmental genes in the rat fetal brain, including genes linked to ASD (Oskvig et al., 2012). Furthermore, the amygdala of rats prenatally exposed to valproic acid, a stressor that leads to ASD, presented activation of neuron development pathways (Barrett et al., 2017). Similarly, the KEGG pathway of cell-adhesion molecules (CAMs), molecules that are fundamental for nervous system development and maintenance, was enriched among the genes presenting a sex-dependent MIA effect. This pathway was also enriched among genes under-expressed in the brain of rats exposed to LPS-triggered MIA and among genes under-expressed in the cortex of patients diagnosed with ASD (Lombardo et al., 2018).

The KEGG pathway neuroactive ligand receptor interaction was enriched among the genes presenting a significant interaction effect characterized by under-expression in the amygdala of CON relative to MPA males (**Table 3**). This pattern is aligned with findings that Poly(I:C)-elicited MIA augmented the synaptic strength of glutamatergic projections from the frontal cortex into the amygdala of mice (Li et al., 2018). The neuroactive ligand receptor pathway encompasses neuroreceptor genes such as dopamine, serotonin, GABA, and glutamate receptors.

The impact of sex-dependent MIA effects on neuropeptide and hormone genes (e.g., CGA, POMC, and SSTR1) expressed in the amygdala is evidenced by the enrichment of GnRH signaling pathway and hormone activity among the genes under-expressed in MPA relative to CON pigs (**Table 3** and **Supplementary File S1 Table C**). Our results suggest that the disruption of glucocorticoid hormone balance on the HPA axis initiated by MIA can have long-lasting effects because amygdala processes are regulated by glucocorticoid receptors and glucocorticoids repress GnRH secretion.

Impact of Maternal Immune Activation on Gene Networks Within Sex

Further understanding of sex-dependent effects of MIA on the co-expression of gene sets was gained from the identification of WGCNA modules of genes that share relative expression profiles between the CON and MPA groups in males (**Table 9**) or that share relative expression profiles between sexes in MPA pigs (**Table 10**). The WGCNA gene modules profiling changes between MIA groups in males uncovered enrichment of genes annotated to AD (**Table 9**). This result is consistent with findings of common molecular mechanisms shared between SSD and AD (Prestia, 2011; Sumitomo et al., 2018). Likewise, the enrichment of ATP metabolic processes among the module of genes associated with MIA effects in males is in agreement with evidence that mitochondrial dysfunction associated with SSD (Prabakaran et al., 2004).

Insights into the distinct vulnerability to MIA between sexes on the interplay between critical genes was gained from the study of the network of genes that had a significant MIA-by-sex effect in the enriched neuroactive ligand receptor pathway. The comparison of **Figures 1**, **2** highlights the distinct interaction between genes in response to PRRSV in males and females, respectively. Notably, males present a strong and consistent over-expression (i.e., red color) of genes in CON relative to MPA, with the exception of SSTR1. On the other hand, females present a weaker expression differential with a slight majority of genes under-expressed (i.e., blue color) in CON relative to MPA pigs. These results suggest lower vulnerability to MIA effects on gene expression in females than in males at 3 weeks of age. An example of this pattern is the module of the neuropeptide receptors, angiotensin II receptor type 1 (AGTR1), and PTH1R. The co-expression of these two G-coupled receptors is consistent with the shared metabolic function (Regard et al., 2008). Distinct to the opposite patterns between sexes observed in the previous network cluster, the highly connected CGA and GH1 are over-expressed in CON relative to MPA in both sexes, albeit the differences are more extreme in males than in females. Correlated under-expression of GH1 and CGA was observed in the cerebellum and prefrontal cortex of rats that presented altered depressive-like behavior (Yamamoto et al., 2015).

Sex-Independent Associations Between Maternal Immune Activation and Transcriptome Changes

The 161 differentially expressed genes between MPA and CON pigs (**Table 4** and **Supplementary File S1 Table D**) included genes supported by previous studies of MIA. Additionally, the differential expression of several genes in **Table 4** has been linked

to neurological disorders such as SSD, ASD, and AD. The over-expression of ANKRD24 in MPA relative to CON pigs (**Table 4**) is supported by the over-expression of an ankyrin repeat domain family member (ANKRD32) in rat fetal brains exposed to MIA elicited by LPS (Oskvig et al., 2012). IFITM3, IFITM1, and CTSC were over-expressed in the amygdala of MPA relative to CON pigs (**Table 4**). Similarly, these genes were over-expressed in the amygdala of individuals diagnosed with SSD (Takao et al., 2012; Chang et al., 2017). Consistent with our results, IFITM3 was over-expressed in the hippocampi of neonatal mice treated with Poly(I:C) that resulted in developmental impairment of the central nervous system and lasting brain dysfunction. Conversely, *Ifitm3*$^{-/-}$ mice treated with Poly(I:C) exhibited normal neural development and did not present neural deficiencies (Ibi et al., 2013). CTSC was also over-expressed in the hippocampus of Shn-2 KO mice that exhibited SSD-like behaviors (Takao et al., 2012). MAP2K7 was over-expressed in MPA relative to CON pigs (**Table 4**), and this gene has been implicated in SSD incidence. MAP2K7 exclusively activates c-Jun N-terminal kinases (JNK) (Lisnock et al., 2000; Zeke et al., 2016), a mediator of the MIA response in the developing fetus that likely contributes to the neurological abnormalities in SSD (Openshaw et al., 2019).

HS3ST5 and SPP1 were over-expressed in the amygdala of MPA compared to CON pigs (**Table 4**), and both genes have been linked to ASD. Consistent with the patterns in our study, the expression of SPP1 in the temporal cortex of humans was higher in individuals diagnosed with ASD compared to controls (Garbett et al., 2008). Indeed, SPP1 participates in multiple immuno-related pathways in neural tissues (Carecchio and Comi, 2011; Brown, 2012). Genome-wide association studies identified a genetic variation near HS3ST5 that was significantly associated with ASD while a single-nucleotide polymorphism within this gene has been associated with SSD (Wang et al., 2009, 2015).

TACC1, CRABP2, and ALDH1A1 participate in the retinoid signaling and metabolic pathways and were differentially expressed in MPA compared to CON pigs (**Table 4** and **Supplementary File S1 Table D**). Dysregulation of retinoid pathways may disrupt neural development leading to SSD (Goodman, 1996), and retinoid toxicity and deficiency are associated with central nervous system abnormalities (Maden, 1994). The differential expression of genes in the retinoid pathways is consistent with the previously described changes in the expression of genes in the thyroid hormone cascades. Defective cross talks between the retinoid and/or thyroid hormone processes have been associated with the development of SSD (Palha and Goodman, 2006). CRABP2 was under-expressed in MPA compared to CON pigs (**Supplementary File S1 Table D**), and mutations in this gene have been implicated in SSD (Goodman, 1995). Aldehyde dehydrogenase 1 family member A1 (ALDH1A1) was under-expressed in MPA compared to CON pigs (**Table 4**). ALDH1A1 was under-expressed in the amygdala of neonatal rats exposed to odor-shock conditioning, mimicking the effects of unpredictable early life trauma and resulting in amygdala dysfunction (Sarro et al., 2014).

IGF2 was under-expressed in MPA compared to CON pigs (**Table 4**), and consistent with our results, the systemic administration of IGF2 reduces ASD phenotypes; promotes normal social interaction, cognition, and executive function; and reduces repetitive behavior (Steinmetz et al., 2018). TTR was over-expressed in the amygdala of MPA compared to CON pigs. Over-expression of TTR was noted in the amygdala of rats treated with MK-801, a N-methyl-D-aspartate antagonist that elicits SSD-like behavior (Matsuoka et al., 2008).

Functional Analysis of Sex-Independent Maternal Immune Activation Transcriptome

The functional categories over-represented among the genes differentially expressed between MPA and CON pigs (**Table 5** and **Supplementary File S1 Tables E**) were supported by previous studies of MIA and associated neurodevelopmental disorders. Among these categories was the BP anatomical structure formation involved in morphogenesis and angiogenesis (**Table 5**). This result is consistent with the under-expression of early growth response 1 (EGR1, a regulator of angiogenic factors) in the GD 97 fetal microglia of MPA relative to CON pigs of both sexes after the GD 76 PRRSV challenge (Antonson et al., 2019). Angiogenesis was also enriched among differentially expressed genes in the frontal cortex of rats that are exposed to valproic acid *in utero* (Hill et al., 2015) to model ASD (Favre et al., 2013). The observed enrichment of the BP angiogenesis was also reported among genes dysregulated in the hippocampal microglia of mice exposed to Poly(I:C) relative to control mice (Mattei et al., 2017).

Several KEGG pathways associated with inflammation and infection, including allograft rejection, Epstein–Barr virus infection, and autoimmune thyroid disease, were enriched, among the genes differentially expressed between MPA and CON pigs (**Table 5** and **Supplementary File S1 Table E**). This finding is consistent with the significant effect of MIA from GD 76 PRRSV injection on expression of 12 genes in the amygdala of GD 97 fetuses from both sexes (Antonson et al., 2019). At this fetal stage, most of the genes differentially expressed between MPA and CON gilts were annotated to the BP immune response, including genes annotated to the BP cytokine-mediated signaling pathway, and to the toll-like receptor pathway (Antonson et al., 2019). This enrichment is in agreement with reports of amygdala inflammation and transcriptome changes in glial cells in response to MIA elicited by LPS administration in mice that persist into adulthood (O'Loughlin et al., 2017).

Many BPs that modulate neurodevelopment were enriched among genes under-expressed in MPA compared to CON pigs, including microtubule bundle formation, cilium morphogenesis, cilium organization, cilium movement, and axoneme assembly (**Table 6** and **Supplementary File S1 Table F**). Reduced neuronal primary cilia can reduce cellular communication during development, the number of dendrites (Goetz and Anderson, 2010; Guadiana et al., 2013), and can hinder neurogenesis in the adult brain (Amador-Arjona et al., 2011). Furthermore, primary cilia participate in the development of the circuitry of GABAergic interneurons, and disrupted cilia formation leads to dysregulated

excitatory/inhibitory signaling between neurons (Guo et al., 2017). This finding is consistent with the glutamatergic and GABAergic-associated genes that presented a significant MIA-by-sex interaction effect previously discussed (**Table 1**). Disruption of this circuitry underlies the neurological disorders associated with ASD and SSD (Bourgeron, 2009; Marin, 2012; Li et al., 2016), and olfactory neuronal precursors collected from SSD patients were found to have less primary cilia growth *in vitro* compared to controls (Munoz-Estrada et al., 2018).

Main Effect of Sex on Gene Expression and Functional Enrichment

The amygdala is a sexually dimorphic area of the brain that is highly responsive to signaling from gonadal steroid hormones (Hines et al., 1992; Cooke and Woolley, 2005). Supporting this, 150 genes were differentially expressed (**Table 7** and **Supplementary File S1 Table G**), and most were over-expressed in males compared to females. The activity of the amygdala has been related to cortisol response, immune responses, and hormonal physiology that differs between sexes (Goldstein et al., 2019). The amygdala is a brain region dense in sex steroid, glucocorticoid, and cytokine receptors that are key co-activators of the HPA axis. The amygdala of male mice has more excitatory synapses per neuron compared to female mice (Cooke and Woolley, 2005; Guneykaya et al., 2018).

The enriched KEGG pathway neuroactive ligand–receptor interaction and biological process response to hormone (**Table 8** and **Supplementary File S1 Table H**) are supported by the evidence indicating that the amygdala development is greatly influenced by sex hormones (Cooke and Woolley, 2005). LHX9 and LHB were strongly over-expressed, while IGSF1 was slightly over-expressed in males compared to females (**Table 7**). The importance of LHX9 in the sexual dimorphism of males and females is highlighted by evidence that *Lhx9*$^{-/-}$ mice fail to produce gonads, and male *Lhx9*$^{-/-}$ resemble females phenotypically (Birk et al., 2000). IGSF1-deficient males have later testosterone secretion than individuals with normal IGSF1 expression (Joustra et al., 2013).

Several developmental BP including gland, embryo, sensory organ, and brain development, along with the MF extracellular matrix structural constituent, were enriched among the genes differentially expressed between sexes (**Table 8**). Supporting these categories, EIF1AY was over-expressed in the amygdala of males compared to females (**Table 7**). Consistent with our results, EIF1AY was over-expressed in the male brain at early childhood, puberty, and adulthood and contributes to the structural sexual dimorphism of the amygdala (Shi et al., 2016).

Additional Considerations

The results of our study of long-lasting changes in the amygdala transcriptome profiles at approximately 60 days after exposure to PRRSV advanced the understanding of the impact of MIA on molecular pathways associated with neurodevelopmental processes, neurodegenerative diseases, and behavior disorders. Biological processes impacted at in the amygdala of PD 22 pigs were also impacted at other developmental stages such as fetal GD 83, 96, or 111 (Antonson et al., 2018, 2019) and included antigen processing and presentation of peptide or polysaccharide antigen via MHC class II and immune response. Likewise, neurodevelopmental processes impacted in the present study were also reported in the amygdala of rats prenatally exposed to MIA by means of valproic acid (Barrett et al., 2017).

Notably, we identified sex-dependent vulnerability to the effects of MIA on gene expression in the amygdala of pigs at PD 22. Sex-independent effects dominated in GD 96 fetuses (Antonson et al., 2019). This comparison supports the hypothesis that sex-dependent effects of MIA in the amygdala become stronger as the animal develops. Additional understanding of MIA effects on the amygdala requires the evaluation of female and male pigs at older ages, in consideration that female pigs reach puberty at 5 months of age. Also, evaluation of the impact of MIA on other brain structures that interact with the amygdala will enable the determination of broader molecular profiles.

CONCLUSION

The present study advances the understanding of the prolonged effects of MIA in the molecular pathways of the amygdala, a brain structure key to social, feeding, and other behaviors. The RNA-Seq profiling of 3-week-old female and male pigs, 2 months after viral infection during gestation, offered insights into MIA-associated neurodevelopmental diseases in humans such as ASD and SSD and potential effects in livestock health. The prevalent and sex-dependent dysregulation of genes in immune pathways was detected, supporting established immunotherapies to alleviate the pathophysiology of SSD, ASD, and AD (Busche et al., 2015; Miller and Buckley, 2016).

Our study detected lesser explored molecular processes affected by MIA including the neuroactive ligand–receptor, glutamatergic, amyloid peptides, neuropeptide, retinoid, and ciliogenesis systems. The detection of the previous processes backs the integration of therapies based on immune modulation together with therapies that target neurochemical dysfunction in MIA-associated disorders. The effectiveness of these therapies may be further advanced by our innovative identification of frequently disrupted neuropeptide systems. Our functional and network analyses solidify the promise of multifactorial therapeutic strategies combining immune and neurochemical targets to ameliorate MIA-associated neurodevelopmental and neurodegenerative disorders.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, GSE149695.

ETHICS STATEMENT

The animal studies were approved by the Illinois Institutional Animal Care and Use Committee (IACUC) at the University of Illinois, and are in compliance with the USDA Animal Welfare Act and the NIH Public Health Service Policy on the Humane Care and Use of Animals.

AUTHOR CONTRIBUTIONS

RJ and SR-Z contributed to the conception and design of the study. CB, AA, LR, HR, MC, AKH, and AGH organized the animal experiments and collected the data. MK, PZ, and BS performed the bioinformatics analyses. MK and SR-Z interpreted the results and wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This study was supported by the USDA NIFA AFRI (grant no. 2018-67015-27413).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2020.00774/full#supplementary-material

REFERENCES

Aavani, T., Rana, S. A., Hawkes, R., and Pittman, Q. J. (2015). Maternal Immune Activation Produces Cerebellar Hyperplasia and Alterations in Motor and Social Behaviors in Male and Female Mice. *Cerebellum* 14, 491–505. doi: 10.1007/s12311-015-0669-5

Alfarano, C., Andrade, C. E., Anthony, K., Bahroos, N., Bajec, M., Bantoft, K., et al. (2005). The Biomolecular Interaction Network Database and related tools 2005 update. *Nucleic Acids Res.* 33, D418–D424. doi: 10.1093/nar/gki051

Amador-Arjona, A., Elliott, J., Miller, A., Ginbey, A., Pazour, G. J., Enikolopov, G., et al. (2011). Primary cilia regulate proliferation of amplifying progenitors in adult hippocampus: implications for learning and memory. *J. Neurosci.* 31, 9933–9944. doi: 10.1523/Jneurosci.1062-11.2011

Anders, S., and Kinney, D. K. (2015). Abnormal immune system development and function in schizophrenia helps reconcile diverse findings and suggests new treatment and prevention strategies. *Brain Res.* 1617, 93–112. doi: 10.1016/j.brainres.2015.02.043

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: http://www.bioinformatics.babraham.ac.uk?/projects/fastqc/ (accessed October 6, 2011).

Antonson, A. M., Balakrishnan, B., Radlowski, E. C., Petr, G., and Johnson, R. W. (2018). Altered Hippocampal Gene Expression and Morphology in Fetal Piglets following Maternal Respiratory Viral Infection. *Dev. Neurosci.* 40, 104–119. doi: 10.1159/000486850

Antonson, A. M., Lawson, M. A., Caputo, M. P., Matt, S. M., Leyshon, B. J., and Johnson, R. W. (2019). Maternal viral infection causes global alterations in porcine fetal microglia. *Proc. Natl. Acad. Sci. U.S.A.* 116, 20190–20200. doi: 10.1073/pnas.1817014116

Antonson, A. M., Radlowski, E. C., Lawson, M. A., Rytych, J. L., and Johnson, R. W. (2017). Maternal viral infection during pregnancy elicits anti-social behavior in neonatal piglet offspring independent of postnatal microglial cell activation. *Brain Behav. Immun.* 59, 300–312. doi: 10.1016/j.bbi.2016.09.019

Baird, A. D., Wilson, S. J., Bladin, P. F., Saling, M. M., and Reutens, D. C. (2007). Neurological control of human sexual behaviour: insights from lesion studies. *J. Neurol. Neurosurg. Psychiatry* 78, 1042–1049. doi: 10.1136/jnnp.2006.107193

Bale, T. L. (2011). Sex differences in prenatal epigenetic programing of stress pathways. *Stress Int. J. Biol. Stress* 14, 348–356. doi: 10.3109/10253890.2011.586447

Bale, T. L., Baram, T. Z., Brown, A. S., Goldstein, J. M., Insel, T. R., McCarthy, M. M., et al. (2010). Early Life Programming and Neurodevelopmental Disorders. *Biol. Psychiatry* 68, 314–319. doi: 10.1016/j.biopsych.2010.05.028

Barrett, C. E., Hennessey, T. M., Gordon, K. M., Ryan, S. J., McNair, M. L., Ressler, K. J., et al. (2017). Developmental disruption of amygdala transcriptome and socioemotional behavior in rats exposed to valproic acid prenatally. *Mol. Autism.* 8:42. doi: 10.1186/s13229-017-0160-x

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Birk, O. S., Casiano, D. E., Wassif, C. A., Cogliati, T., Zhao, L. P., Zhao, Y. G., et al. (2000). The LIM homeobox gene Lhx9 is essential for mouse gonad formation. *Nature* 403, 909–913. doi: 10.1038/35002622

Bourgeron, T. (2009). A synaptic trek to autism. *Curr. Opin. Neurobiol.* 19, 231–234. doi: 10.1016/j.conb.2009.06.003

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification (vol 34, pg 525, 2016). *Nat. Biotechnol.* 34:888. doi: 10.1038/nbt0816-888d

Brown, A. (2012). Osteopontin: a key link between immunity, inflammation and the central nervous system. *Transl. Neurosci.* 3, 288–293. doi: 10.2478/s13380-012-0028-7

Busche, M. A., Grienberger, C., Keskin, A. D., Song, B., Neumann, U., Staufenbiel, M., et al. (2015). Decreased amyloid-β and increased neuronal hyperactivity by immunotherapy in Alzheimer's models. *Nat. Neurosci.* 18, 1725–1727. doi: 10.1038/nn.4163

Caetano-Anollés, K., Mishra, S., and Rodriguez-Zas, S. L. (2015). Synergistic and antagonistic interplay between myostatin gene expression and physical activity levels on gene expression patterns in triceps brachii muscles of C57/BL6 Mice. *PLoS One* 10:e0116828. doi: 10.1371/journal.pone.0116828

Caetano-Anollés, K., Rhodes, J. S., Garland, T. Jr., Perez, S. D., Hernandez, A. G., Southey, B. R., et al. (2016). Cerebellum transcriptome of mice bred for high voluntary activity offers insights into locomotor control and reward-dependent behaviors. *PLoS One* 11:e0167095. doi: 10.1371/journal.pone.0167095

Canetta, S., Bolkan, S., Padilla-Coreano, N., Song, L. J., Sahn, R., Harrison, N. L., et al. (2016). Maternal immune activation leads to selective functional deficits in offspring parvalbumin interneurons. *Mol. Psychiatry* 21, 956–968. doi: 10.1038/mp.2015.222

Canitano, R., and Pallagrosi, M. (2017). Autism spectrum disorders and schizophrenia spectrum disorders: excitation/inhibition imbalance and developmental trajectories. *Front. Psychiatry* 8:69. doi: 10.3389/fpsyt.2017.00069

Carecchio, M., and Comi, C. (2011). The role of osteopontin in neurodegenerative diseases. *J. Alzheimers Dis.* 25, 179–185. doi: 10.3233/Jad-2011-102151

Carlezon, W. A., Kim, W., Missig, G., Finger, B. C., Landino, S. M., Alexander, A. J., et al. (2019). Maternal and early postnatal immune activation produce sex-specific effects on autism-like behaviors and neuroimmune function in mice. *Sci. Rep.* 9:16928. doi: 10.1038/s41598-019-53294-z

Chang, X., Liu, Y., Hahn, C. G., Gur, R. E., and Sleiman, P. M. A. (2017). Hakonarson H. RNA-seq analysis of amygdala tissue reveals characteristic expression profiles in schizophrenia. *Transl. Psychiatr.* 7:e1203. doi: 10.1038/tp.2017.154

Cooke, B. M., and Woolley, C. S. (2005). Sexually dimorphic synaptic organization of the medial amygdala. *J. Neurosci.* 25, 10759–10767. doi: 10.1523/Jneurosci.2919-05.2005

Delfino, K. R., and Rodriguez-Zas, S. L. (2013). Transcription factor-microRNA-target gene networks associated with ovarian cancer survival and recurrence. *PLoS One* 8:e58608. doi: 10.1371/journal.pone.0058608

Delfino, K. R., Serao, N. V., Southey, B. R., and Rodriguez-Zas, S. L. (2011). Therapy-, gender- and race-specific microRNA markers, target genes and networks related to glioblastoma recurrence and survival. *Cancer Genom. Proteom.* 8, 173–183.

Fatemi, S. H., Stary, J. M., Earle, J. A., Araghi-Niknam, A., and Egan, E. (2005). GABAergic dysfunction in schizophrenia and mood disorders as reflected by decreased levels of glutamic acid decarboxylase 65 and 67 kDa and reelin proteins in cerebellum (vol 72, pg 109, 2004). *Schizophr. Res.* 74:287. doi: 10.1016/j.schres.2004.12.001

Favre, M. R., Barkat, T. R., LaMendola, D., Khazen, G., Markram, H., and Markram, K. (2013). General developmental health in the VPA-rat model of autism. *Front. Behav. Neurosci.* 7:88. doi: 10.3389/fnbeh.2013.00088

Felix, B., Leger, M. E., and Albe-Fessard, D. (1999). Stereotaxic atlas of the pig brain. *Brain Res Bull* 49, 1–137.

Fernandez-Irigoyen, J., Zelaya, M. V., and Santamaria, E. (2014). Applying mass spectrometry-based qualitative proteomics to human amygdaloid complex. *Front. Cell Neurosci.* 8:80. doi: 10.3389/fncel.2014.00080

Garbett, K., Ebert, P. J., Mitchell, A., Lintas, C., Manzi, B., Mirnics, K., et al. (2008). Immune transcriptome alterations in the temporal cortex of subjects with autism. *Neurobiol. Dis.* 30, 303–311. doi: 10.1016/j.nbd.2008.01.012

Goetz, S. C., and Anderson, K. V. (2010). The primary cilium: a signalling centre during vertebrate development. *Nat. Rev. Genet.* 11, 331–344. doi: 10.1038/nrg2774

Goldstein, J. M., Hale, T., Foster, S. L., Tobet, S. A., and Handa, R. J. (2019). Sex differences in major depression and comorbidity of cardiometabolic disorders: impact of prenatal stress and immune exposures. *Neuropsychopharmacology* 44, 59–70. doi: 10.1038/s41386-018-0146-1

Gonzalez-Pena, D., Nixon, S. E., O'Connor, J. C., Southey, B. R., Lawson, M. A., McCusker, R. H., et al. (2016a). Microglia transcriptome changes in a model of depressive behavior after immune challenge. *PLoS One* 11:e0150858. doi: 10.1371/journal.pone.0150858

Gonzalez-Pena, D., Nixon, S. E., Southey, B. R., Lawson, M. A., McCusker, R. H., Hernandez, A. G., et al. (2016b). Differential Transcriptome Networks between IDO1-knockout and wild-type mice in brain microglia and macrophages. *PLoS One* 11:e0157727. doi: 10.1371/journal.pone.0157727

Goodman, A. B. (1995). Chromosomal Locations and Modes of Action of Genes of the Retinoid (Vitamin-a) system support their involvement in the etiology of schizophrenia. *Am. J. Med. Genet.* 60, 335–348. doi: 10.1002/ajmg.1320600415

Goodman, A. B. (1996). Congenital anomalies in relatives of schizophrenic probands may indicate a retinoid pathology. *Schizophr. Res.* 19, 163–170. doi: 10.1016/0920-9964(96)88523-9

Gottschalk, M. G., Sarnyai, Z., Guest, P. C., Harris, L. W., and Bahn, S. (2013). Estudos traducionais de neuropsiquiatria e esquizofrenia: modelos animais genéticos e de neurodesenvolvimento. *Arch. Clin. Psychiatr.* 40, 41–50. doi: 10.1590/s0101-60832012005000007

Gray, A. L., Hyde, T. M., Deep-Soboslay, A., Kleinman, J. E., and Sodhi, M. S. (2015). Sex differences in glutamate receptor gene expression in major depression and suicide (vol 20, pg 1057, 2015). *Mol. Psychiatry* 20:1139. doi: 10.1038/mp.2015.114

Guadiana, S. M., Semple-Rowland, S., Daroszewski, D., Madorsky, I., Breunig, J. J., Mykytyn, K., et al. (2013). Arborization of Dendrites by Developing Neocortical Neurons Is Dependent on Primary Cilia and Type 3 Adenylyl Cyclase. *J. Neurosci.* 33, 2626–2638. doi: 10.1523/Jneurosci.2906-12.2013

Guneykaya, D., Ivanov, A., Hernandez, D. P., Haage, V., Wojtas, B., Meyer, N., et al. (2018). Transcriptional and translational differences of microglia from male and female brains. *Cell Rep.* 24, 2773–+. doi: 10.1016/j.celrep.2018.08.001

Guo, J., Otis, J. M., Higginbotham, H., Monckton, C., Cheng, J., Asokan, A., et al. (2017). Primary cilia signaling shapes the development of interneuronal connectivity. *Dev. Cell* 42:286. doi: 10.1016/j.devcel.2017.07.010

Haida, O., Al Sagheer, T., Balbous, A., Francheteau, M., Matas, E., Soria, F., et al. (2019). Sex-dependent behavioral deficits and neuropathology in a maternal immune activation model of autism. *Transl. Psychiatry* 9, 124. doi: 10.1038/s41398-019-0457-y

Hamann, S. (2005). Sex differences in the responses of the human amygdala. *Neuroscientist* 11, 288–293. doi: 10.1177/1073858404271981

Hill, D. S., Cabrera, R., Schultz, D. W., Zhu, H. P., Lu, W., Finnell, R. H., et al. (2015). Autism-like behavior and epigenetic changes associated with autism as consequences of in utero exposure to environmental pollutants in a mouse model. *Behav. Neurol.* 2015:426263. doi: 10.1155/2015/426263

Hines, M., Allen, L. S., and Gorski, R. A. (1992). Sex differences in subregions of the medial nucleus of the amygdala and the bed nucleus of the stria terminalis of the rat. *Brain Res.* 579, 321–326. doi: 10.1016/0006-8993(92)90068-k

Hou, W. H., and Capogna, M. (2018). Dendritic inhibition in layer 1 cortex gates associative memory. *Neuron* 100, 516–519. doi: 10.1016/j.neuron.2018.10.029

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4:44. doi: 10.1038/nprot.2008.211

Ibanez, K., Boullosa, C., Tabares-Seisdedos, R., Baudot, A., and Valencia, A. (2014). Molecular evidence for the inverse comorbidity between central nervous system disorders and cancers detected by transcriptomic meta-analyses. *PLoS Genet.* 10:e4173. doi: 10.1371/journal.pgen.1004173

Ibi, D., Nagai, T., Nakajima, A., Mizoguchi, H., Kawase, T., Tsuboi, D., et al. (2013). Astroglial IFITM3 mediates neuronal impairments following neonatal immune challenge in mice. *Glia* 61, 679–693. doi: 10.1002/glia.22461

Jamali, A. K., and Tramu, G. (1997). Daily cycle of fos expression within hypothalamic POMC neurons of the male rat. *Brain Res.* 771, 45–54. doi: 10.1016/S0006-8993(97)00767-1

Joustra, S. D., Schoenmakers, N., Persani, L., Campi, I., Bonomi, M., Radetti, G., et al. (2013). The IGSF1 deficiency syndrome: characteristics of male and female patients. *J. Clin. Endocrinol. Metab.* 98, 4942–4952. doi: 10.1210/jc.2013-2743

Kerrien, S., Alam-Faruque, Y., Aranda, B., Bancarz, I., Bridge, A., Derow, C., et al. (2007). IntAct–open source resource for molecular interaction data. *Nucleic Acids Res.* 35, D561–D565. doi: 10.1093/nar/gkl958

Killgore, W. D., and Yurgelun-Todd, D. A. (2001). Sex differences in amygdala activation during the perception of facial affect. *Neuroreport* 12, 2543–2547. doi: 10.1097/00001756-200108080-00050

Kirsten, T. B., Taricano, M., Florio, J. C., Palermo-Neto, J., and Bernardi, M. M. (2010). Prenatal lipopolysaccharide reduces motor activity after an immune challenge in adult male offspring. *Behav. Brain Res.* 211, 77–82. doi: 10.1016/j.bbr.2010.03.009

Knuesel, I., Chicha, L., Britschgi, M., Schobel, S. A., Bodmer, M., Hellings, J. A., et al. (2014). Maternal immune activation and abnormal brain development across CNS disorders. *Nat. Rev. Neurol.* 10, 643–660. doi: 10.1038/nrneurol.2014.187

Kroismayr, R., Baranyi, U., Stehlik, C., Dorfleutner, A., Binder, B. R., and Lipp, J. (2004). HERC5, a HECT E3 ubiquitin ligase tightly regulated in LPS activated endothelial cells. *J. Cell Sci.* 117, 4749–4756. doi: 10.1242/jcs.01338

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformat.* 9:559. doi: 10.1186/1471-2105-9-559

Laszlo, K., Toth, K., Kertes, E., Peczely, L., and Lenard, L. (2010). The role of neurotensin in positive reinforcement in the rat central nucleus of amygdala. *Behav. Brain Res.* 208, 430–435. doi: 10.1016/j.bbr.2009.12.022

Lennox, B. R., Coles, A. J., and Vincent, A. (2012). Antibody-mediated encephalitis: a treatable cause of schizophrenia. *Br. J. Psychiatry* 200, 92–94. doi: 10.1192/bjp.bp.111.095042

Li, Y., Missig, G., Finger, B. C., Landino, S. M., Alexander, A. J., Mokler, E. L., et al. (2018). Maternal and Early Postnatal Immune Activation Produce Dissociable Effects on Neurotransmission in mPFC-Amygdala Circuits. *J. Neurosci.* 38, 3358–3372. doi: 10.1523/JNEUROSCI.3642-17.2018

Li, Y. F., Sun, H., Chen, Z. C., Xu, H. X., Bu, G. J., and Zheng, H. (2016). Implications of GABAergic Neurotransmission in Alzheimer's disease. *Front. Aging Neurosci.* 8:31. doi: 10.3389/fnagi.2016.00031

Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E., et al. (2012). MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.* 40, D857–D861. doi: 10.1093/nar/gkr930

Lisnock, J., Griffin, P., Calaycay, J., Frantz, B., Parsons, J., O'Keefe, S. J., et al. (2000). Activation of JNK3 alpha 1 requires both MKK4 and MKK7: kinetic characterization of in vitro phosphorylated JNK3 alpha 1. *Biochemistry* 39, 3141–3148. doi: 10.1021/bi992410

Lombardo, M. V., Moon, H. M., Su, J., Palmer, T. D., Courchesne, E., and Pramparo, T. (2018). Maternal immune activation dysregulation of the fetal brain transcriptome and relevance to the pathophysiology of

autism spectrum disorder. *Mol. Psychiatry* 23, 1001–1013. doi: 10.1038/mp.2017.15

Maden, M. (1994). Vitamin-a in embryonic-development. *Nutr. Rev.* 52, 3–12.

Malkova, N. V., Yu, C. Z., Hsiao, E. Y., Moore, M. J., and Patterson, P. H. (2012). Maternal immune activation yields offspring displaying mouse versions of the three core symptoms of autism. *Brain Behav. Immun.* 26, 607–616. doi: 10.1016/j.bbi.2012.01.011

Marin, O. (2012). Interneuron dysfunction in psychiatric disorders. *Nat. Rev. Neurosci.* 13, 107–120. doi: 10.1038/nrn3155

Martin, A., Ochagavia, M., Rabasa, L., Miranda, J., Fernandez-de-Cossio, J., and Bringas, R. (2010). BisoGenet: a new tool for gene network building, visualization and analysis. *BMC Bioinform.* 11:91. doi: 10.1186/1471-2105-11-91

Martinez, C., Juarranz, Y., Gutierrez-Canas, I., Carrion, M., Perez-Garcia, S., Villanueva-Romero, R., et al. (2020). A clinical approach for the use of VIP axis in inflammatory and autoimmune diseases. *Int. J. Mol. Sci.* 21:65. doi: 10.3390/ijms21010065

Matsuoka, T., Tsunoda, M., Sumiyoshi, T., Takasaki, I., Tabuchi, Y., Tanaka, K., et al. (2008). Effect of MK-801 on gene expressions in the amygdala of rats. *Int. J. Neuropsychopharmacol.* 11:265.

Mattei, D., Ivanov, A., Ferrai, C., Jordan, P., Guneykaya, D., Buonfiglioli, A., et al. (2017). Maternal immune activation results in complex microglial transcriptome signature in the adult offspring that is reversed by minocycline treatment. *Transl. Psychiatry* 7:e1120. doi: 10.1038/tp.2017.80

Meador-Woodruff, J. H., and Healy, D. J. (2000). Glutamate receptor expression in schizophrenic brain. *Brain Res. Rev.* 31, 288–294. doi: 10.1016/S0165-0173(99)00044-2

Merali, Z., Anisman, H., James, J. S., Kent, P., and Schulkin, J. (2008). Effects of corticosterone on corticotrophin-releasing hormone and gastrin-releasing peptide release in response to an aversive stimulus in two regions of the forebrain (central nucleus of the amygdala and prefrontal cortex). *Eur. J. Neurosci.* 28, 165–172. doi: 10.1111/j.1460-9568.2008.06281.x

Miller, B. J., and Buckley, P. F. (2016). The case for adjunctive monoclonal antibody immunotherapy in schizophrenia. *Psychiatr. Clin.* 39, 187–198. doi: 10.1016/j.psc.2016.01.003

Mishra, G. R., Suresh, M., Kumaran, K., Kannabiran, N., Suresh, S., Bala, P., et al. (2006). Human protein reference database–2006 update. *Nucleic Acids Res.* 34, D411–D414. doi: 10.1093/nar/gkj141

Mistry, M., Gillis, J., and Pavlidis, P. (2013). Genome-wide expression profiling of schizophrenia using a large combined cohort. *Mol. Psychiatry* 18, 215–225. doi: 10.1038/mp.2011.172

Morris, B. J., and Pratt, J. A. (2014). Novel treatment strategies for schizophrenia from improved understanding of genetic risk. *Clin. Genet.* 86, 401–411. doi: 10.1111/cge.12485

Munoz-Estrada, J., Lora-Castellanos, A., Meza, I., Elizalde, S. A., and Benitez-King, G. (2018). Primary cilia formation is diminished in schizophrenia and bipolar disorder: a possible marker for these psychiatric diseases. *Schizophr. Res.* 195, 412–420. doi: 10.1016/j.schres.2017.08.055

Munshi, S., and Rosenkranz, J. A. (2018). Effects of peripheral immune challenge on in vivo firing of basolateral amygdala neurons in adult male rats. *Neuroscience* 390, 174–186. doi: 10.1016/j.neuroscience.2018.08.017

Murphy, E., Nordquist, R. E., and van der Staay, F. J. (2014). A review of0020behavioural methods to study emotion and mood in pigs. *Sus scrofa*. *Appl. Anim. Behav. Sci.* 159, 9–28. doi: 10.1016/j.applanim.2014.08.002

Odorizzi, P. M., and Feeney, M. E. (2016). Impact of In Utero Exposure to Malaria on Fetal T Cell Immunity. *Trends Mol. Med.* 22, 877–888. doi: 10.1016/j.molmed.2016.08.005

Oguchi-Katayama, A., Monma, A., Sekino, Y., Moriguchi, T., and Sato, K. (2013). Comparative gene expression analysis of the amygdala in autistic rat models produced by pre- and post-natal exposures to valproic acid. *J. Toxicol. Sci.* 38, 391–402. doi: 10.2131/jts.38.391

O'Loughlin, E., Pakan, J. M. P., Yilmazer-Hanke, D., and McDermott, K. W. (2017). Acute in utero exposure to lipopolysaccharide induces inflammation in the pre- and postnatal brain and alters the glial cytoarchitecture in the developing amygdala. *J. Neuroinflam.* 14:212. doi: 10.1186/s12974-017-0981-8

Openshaw, R. L., Kwon, J., McColl, A., Penninger, J. M., Cavanagh, J., Pratt, J. A., et al. (2019). JNK signalling mediates aspects of maternal immune activation: importance of maternal genotype in relation to schizophrenia risk. *J. Neuroinflam.* 16:18. doi: 10.1186/s12974-019-1408-5

Oskvig, D. B., Elkahloun, A. G., Johnson, K. R., Phillips, T. M., and Herkenham, M. (2012). Maternal immune activation by LPS selectively alters specific gene expression profiles of interneuron migration and oxidative stress in the fetus without triggering a fetal immune response. *Brain Behav. Immun.* 26, 623–634. doi: 10.1016/j.bbi.2012.01.015

Palha, J. A., and Goodman, A. B. (2006). Thyroid hormones and retinoids: a possible link between genes and environment in schizophrenia. *Brain Res. Rev.* 51, 61–71. doi: 10.1016/j.brainresrev.2005.10.001

Paylor, J. W., Lins, B. R., Greba, Q., Moen, N., De Moraes, R. S., Howland, J. G., et al. (2016). Developmental disruption of perineuronal nets in the medial prefrontal cortex after maternal immune activation. *Sci. Rep.* 6:37580. doi: 10.1038/srep37580

Petrovich, G. D., and Gallagher, M. (2003). Amygdala subsystems and control of feeding behavior by learned cues. *Amygdala Brain Funct.* 985, 251–262. doi: 10.1111/j.1749-6632.2003.tb07086.x

Pieczenik, S. R., and Neustadt, J. (2007). Mitochondrial dysfunction and molecular pathways of disease. *Exp. Mol. Pathol.* 83, 84–92. doi: 10.1016/j.yexmp.2006.09.008

Prabakaran, S., Swatton, J. E., Ryan, M. M., Huffaker, S. J., Huang, J. T. J., Griffin, J. L., et al. (2004). Mitochondrial dysfunction in schizophrenia: evidence for compromised brain metabolism and oxidative stress. *Mol. Psychiatry* 9, 684–697. doi: 10.1038/sj.mp.4001511

Prestia, A. (2011). Alzheimer's disease and schizophrenia: evidence of a specific, shared molecular background. *Future Neurol.* 6, 17–21. doi: 10.2217/fnl.10.61

Prins, J. R., Eskandar, S., Eggen, B. J. L., and Scherjon, S. A. (2018). Microglia, the missing link in maternal immune activation and fetal neurodevelopment; and a possible link in preeclampsia and disturbed neurodevelopment? *J. Reprod. Immunol.* 126, 18–22. doi: 10.1016/j.jri.2018.01.004

Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *NAR* 35, D61–D65. doi: 10.1093/nar/gkl842

Regard, J. B., Sato, I. T., and Coughlin, S. R. (2008). Anatomical profiling of G protein-coupled receptor expression. *Cell* 135, 561–571. doi: 10.1016/j.cell.2008.08.040

Rhomberg, T., Rovira-Esteban, L., Vikor, A., Paradiso, E., Kremser, C., Nagy-Pal, P., et al. (2018). Vasoactive intestinal polypeptide-immunoreactive interneurons within circuits of the mouse basolateral amygdala. *J. Neurosci.* 38, 6983–7003. doi: 10.1523/Jneurosci.2063-17.2018

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y. F., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *NAR* 43:e47. doi: 10.1093/nar/gkv007

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Roullet, F. I., Lai, J. K. Y., and Foster, J. A. (2013). In utero exposure to valproic acid and autism - A current review of clinical and animal studies. *Neurotoxicol. Teratol.* 36, 47–56. doi: 10.1016/j.ntt.2013.01.004

Rutherford, K. M., Piastowska-Ciesielska, A., Donald, R. D., Robson, S. K., Ison, S. H., Jarvis, S., et al. (2014). Prenatal stress produces anxiety prone female offspring and impaired maternal behaviour in the domestic pig. *Physiol. Behav.* 129, 255–264. doi: 10.1016/j.physbeh.2014.02.052

Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U., and Eisenberg, D. (2004). The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* 32, D449–D451. doi: 10.1093/nar/gkh086

Sarro, E. C., Sullivan, R. M., and Barr, G. (2014). Unpredictable neonatal stress enhances adult anxiety and alters Amygdala gene expression related to serotonin and Gaba. *Neuroscience* 258, 147–161. doi: 10.1016/j.neuroscience.2013.10.064

Schumann, C. M., Bauman, M. D., and Amaral, D. G. (2011). Abnormal structure or function of the amygdala is a common component of neurodevelopmental disorders. *Neuropsychologia* 49, 745–759. doi: 10.1016/j.neuropsychologia.2010.09.028

Serao, N. V., Delfino, K. R., Southey, B. R., Beever, J. E., and Rodriguez-Zas, S. L. (2011). Cell cycle and aging, morphogenesis, and response to stimuli genes are individualized biomarkers of glioblastoma progression and survival. *BMC Med. Genom.* 4:49. doi: 10.1186/1755-8794-4-49

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Shi, L., Zhang, Z., and Su, B. (2016). Sex biased gene expression profiling of human brains at major developmental stages. *Sci. Rep.* 6:21181. doi: 10.1038/srep21181

Sragovich, S., Merenlender-Wagner, A., and Gozes, I. (2017). ADNP plays a key role in autophagy: from autism to schizophrenia and Alzheimer's disease. *Bioessays* 39:1700054. doi: 10.1002/bies.201700054

Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006). BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 34, D535–D539. doi: 10.1093/nar/gkj109

Steinmetz, A. B., Stern, S. A., Kohtz, A. S., Descalzi, G., and Alberini, C. M. (2018). Insulin-Like Growth Factor II Targets the mTOR Pathway to Reverse Autism-Like Phenotypes in Mice. *J. Neurosci.* 38, 1015–1029. doi: 10.1523/Jneurosci.2010-17.2017

Stuart, M. J., and Baune, B. T. (2014). Chemokines and chemokine receptors in mood disorders, schizophrenia, and cognitive impairment: a systematic review of biomarker studies. *Neurosci. Biobehav. Rev.* 42, 93–115. doi: 10.1016/j.neubiorev.2014.02.001

Subramanian, A., Kuehn, H., Gould, J., Tamayo, P., and Mesirov, J. P. (2007). GSEA-P: a desktop application for gene set enrichment analysis. *Bioinformatics* 23, 3251–3253. doi: 10.1093/bioinformatics/btm369

Sumitomo, A., Horike, K., Hirai, K., Butcher, N., Boot, E., Sakurai, T., et al. (2018). A mouse model of 22q11.2 deletions: molecular and behavioral signatures of Parkinson's disease and schizophrenia. *Sci. Adv.* 4:eaar6637. doi: 10.1126/sciadv.aar6637

Takao, K., Kobayashi, K., Esaki, K., Furuya, S., Takagi, T., Walton, N., et al. (2012). Deficiency of schnurri-2, an MHC enhancer binding protein, induces mild chronic inflammation in the brain and confers molecular, neuronal, and behavioral phenotypes related to schizophrenia. *Int. J. Neuropsychopharmacol.* 15:136.

Tian, J., Dai, H. M., Deng, Y. Y., Zhang, J., Li, Y., Zhou, J., et al. (2015). The effect of HMGB1 on sub-toxic chlorpyrifos exposure-induced neuroinflammation in amygdala of neonatal rats. *Toxicology* 338, 95–103. doi: 10.1016/j.tox.2015.10.010

Wang, K., Zhang, H. T., Ma, D. Q., Bucan, M., Glessner, J. T., Abrahams, B. S., et al. (2009). Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 459, 528–533. doi: 10.1038/nature07999

Wang, Q., Xiang, B., Deng, W., Wu, J., Li, M., and Ma, X. (2015). Genome-wide association analysis with gray matter volume as a quantitative phenotype in first-episode treatment-naive patients with schizophrenia (vol 8, e75083, 2013). *PLoS One* 10:e0122945. doi: 10.1371/journal.pone.0122945

Wischhof, L., Irrsack, E., Osorio, C., and Koch, M. (2015). Prenatal LPS-exposure - a neurodevelopmental rat model of schizophrenia - differentially affects cognitive functions, myelination and parvalbumin expression in male and female offspring. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* 57, 17–30. doi: 10.1016/j.pnpbp.2014.10.004

Xuan, I. C. Y., and Hampson, D. R. (2014). Gender-dependent effects of maternal immune activation on the behavior of mouse offspring. *PLoS One* 9:e104433. doi: 10.1371/journal.pone.0104433

Yamamoto, Y., Ueyama, T., Ito, T., and Tsuruo, Y. (2015). Downregulation of growth hormone 1 gene in the cerebellum and prefrontal cortex of rats with depressive-like behavior. *Physiol. Genomics* 47, 170–176. doi: 10.1152/physiolgenomics.00119.2014

Yang, L., Wang, M., Guo, Y. Y., Sun, T., Li, Y. J., Yang, Q., et al. (2016). Systemic inflammation induces anxiety disorder through CXCL12/CXCR4 pathway. *Brain Behav. Immun.* 56, 352–362. doi: 10.1016/j.bbi.2016.03.001

Zeke, A., Misheva, M., Remenyi, A., and Bogoyevitch, M. A. J. N. K. (2016). Signaling: regulation and functions based on complex protein-protein partnerships. *Microbiol. Mol. Biol. Rev.* 80, 793–835. doi: 10.1128/Mmbr.00043-14

Zhao, Y., He, A. W., Zhu, F., Ding, M., Hao, J. C., Fan, Q. R., et al. (2018). Integrating genome-wide association study and expression quantitative trait locus study identifies multiple genes and gene sets associated with schizophrenia. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* 81, 50–54. doi: 10.1016/j.pnpbp.2017.10.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 15 October 2020
doi: 10.3389/fgene.2020.559080

Heterozygous *PGM3* Variants Are Associated With Idiopathic Focal Epilepsy With Incomplete Penetrance

Xiao-Rong Liu, Wen-Jun Bian, Jie Wang, Ting-Ting Ye, Bing-Mei Li, De-Tian Liu, Bin Tang, Wei-Wen Deng, Yi-Wu Shi, Tao Su, Yong-Hong Yi and Wei-Ping Liao on behalf of the China Epilepsy Gene 1.0 Project*

Key Laboratory of Neurogenetics and Channelopathies of Guangdong Institute, Department of Neurology of the Second Affiliated Hospital of Guangzhou Medical University, of Neuroscience, Province and the Ministry of Education of China, Guangzhou, China

Introduction: Idiopathic focal epilepsy (IFE) is a group of self-limited epilepsies. The etiology for the majority of the patients with IFE remains elusive. We thus screened disease-causing variants in the patients with IFE.

Methods: Whole-exome sequencing was performed in a cohort of 323 patients with IFE. Protein modeling was performed to predict the effects of missense variants. The genotype–phenotype correlation of the newly defined causative gene was analyzed.

Results: Four novel heterozygous variants in *PGM3*, including two *de novo* variants, were identified in four unrelated individuals with IFE. The variants included one truncating variant (c.1432C > T/p.Q478X) and three missense variants (c.478C > T/p.P160S, c.1239C > G/p.N413K, and c.1659T > A/p.N553K), which had no allele frequency in the gnomAD database. The missense variants were predicted to be damaging and affect hydrogen bonds with surrounding amino acids. Mutations Q478X, P160S, and N413K were associated with benign childhood epilepsy with centrotemporal electroencephalograph (EEG) spikes. P160S and N413K were located in the inner side of the enzyme active center. Mutation N553K was associated with benign occipital epilepsy with incomplete penetrance, located in the C-terminal of Domain 4. Further analysis demonstrated that previously reported biallelic *PGM3* mutations were associated with severe immunodeficiency and/or congenital disorder of glycosylation, commonly accompanied by neurodevelopmental abnormalities, while monoallelic mutations were associated with milder symptoms like IFE.

Conclusion: The genetic and molecular evidence from the present study implies that the *PGM3* variants identified in IFE patients lead to defects of the *PGM3* gene, suggesting that the *PGM3* gene is potentially associated with epilepsy. The genotype–phenotype relationship of *PGM3* mutations suggested a quantitative correlation between genetic impairment and phenotypic severity, which helps explain the mild symptoms and incomplete penetrance in individuals with IFE.

Keywords: whole-exome sequencing, congenital disorder of glycosylation, immunodeficiency, *PGM3* gene, idiopathic focal epilepsy

OPEN ACCESS

Edited by:
Patrick E. Williams, University of Kansas, United States

Reviewed by:
Daniele Ghezzi, Carlo Besta Neurological Institute (IRCCS), Italy
Alysia Vrailas-Mortimer, Illinois State University, United States

****Correspondence:***
Wei-Ping Liao
wpliao@163.net

Specialty section:
This article was submitted to Neurogenomics, a section of the journal Frontiers in Genetics

Received: *05 May 2020*
Accepted: *17 September 2020*
Published: *15 October 2020*

Citation:
Liu X-R, Bian W-J, Wang J, Ye T-T, Li B-M, Liu D-T, Tang B, Deng W-W, Shi Y-W, Su T, Yi Y-H and Liao W-P (2020) Heterozygous PGM3 Variants Are Associated With Idiopathic Focal Epilepsy With Incomplete Penetrance. Front. Genet. 11:559080. doi: 10.3389/fgene.2020.559080

INTRODUCTION

Idiopathic focal epilepsy (IFE), also named as localization-related idiopathic epilepsy (G40.0 in ICD-10 2016, WHO), is a common group of self-limited childhood-onset epilepsies. It contains two major subtypes, i.e., benign childhood epilepsy with centrotemporal electroencephalograph (EEG) spikes (BECTS, OMIM# 117100) and benign occipital epilepsy (BOE, OMIM# 132090). BECTS is affecting 0.2% of the population (Strug et al., 2009), while the prevalence of BOE is unknown but is considered to affect around 6% of children aged 1–15 years (Weir et al., 2018). Although both BECTS and BOE are generally regarded as autosomal dominant inheritance with age-dependent penetrance, rare causative genes are identified in the patients. Recent studies have shown that several genes, including *GRIN2A*, *ELP4*, *SRPX2*, and *DEPDC5*, are associated with BECTS (Strug et al., 2009; Lemke et al., 2013; Lal et al., 2014; Reinthaler et al., 2014) *GRIN2A* mutations were regarded as the most relevant causative gene for BECTS but accounted for only 4.9% of the patients with BECTS (Lemke et al., 2013). So far, no causative gene has been reported to be associated with BOE. Thus, the etiology for majority of patients with IFE remains unknown.

In this study, we performed whole-exome sequencing (WES) in a cohort of patients with IFE. Four novel heterozygous variants in *PGM3* were identified. Previously, *PGM3* mutations were associated with immunodeficiency-23 (IMD23, OMIM# 615816) and/or congenital disorder of glycosylation (CDG) that is occasionally accompanied by seizures. Our further analysis showed that previously reported homozygous and compound heterozygous mutations in *PGM3* were associated with severe immunodeficiency and glycosylation disturbance, while monoallelic mutations were potentially associated with mild phenotypes like IFE, suggesting that *PGM3* is potentially a candidate causative gene of epilepsy.

MATERIALS AND METHODS

Subjects

A total of 323 subjects with IFE were recruited in the Epilepsy Center of the Second Affiliated Hospital of Guangzhou Medical University in China from January 2014 to 2019. These subjects were epilepsy patients with partial seizures and/or focal discharges but without acquired causes. Their EEGs showed focal abnormalities with features of idiopathic epilepsies, including shifting, bilateral, or multiple focal discharges with normal backgrounds. Eligible subjects had a clinical diagnosis of IFE according to the International League Against Epilepsy (ILAE-1989, 2006, and 2017) (Commission on Classification and Terminology of the International League Against Epilepsy, 1989; Engel, 2006; Scheffer et al., 2017) after appropriate investigations, including long-term video electroencephalography (VEEG), brain MRI, cognitive and behavioral evaluation, and neurometabolic testing. All of the subjects were followed up for at least 1 year.

The studies adhered to the guidelines of the International Committee of Medical Journal Editors with regard to patient consent for research or participation and received approval from local ethics committees of the participating hospitals. The Ethics Committee of the Second Affiliated Hospital of Guangzhou Medical University provided ethics approval.

Whole-Exome Sequencing

Genomic DNA was extracted from peripheral blood using a QuickGene DNA whole blood kit (Fujifilm, Tokyo, Japan). Blood samples of parents and siblings were used for linkage and segregation analysis. Trio-based whole-exome sequencing was conducted on the Illumina HiSeq 2500/4000 platform by BGI-Shenzhen (Shenzhen, China). Paired-end reads with a length of 90bp were generated by massive parallel sequencing with more than 125 times average depth and more than 98% coverage of the target region. The raw data were aligned to the reference human genome (GRCh37) using the Burrows–Wheeler Alignment (BWA). Potential pathogenic variants were derived by stepwise filtering. Population-based filtration removed common variants presenting a minor allele frequency $\geq$0.005 in the gnomAD database, except for those previously reported to be associated with disease in the Human Gene Mutation Database (HGMD) and/or the Online Mendelian Inheritance in Man database. Potential pathogenic variants were flagged if predicted as damaging by multiple *in silico* programs (**Supplementary Table 1**). Inheritance-based filtration filtered variants based on family history and possible inheritance models. Assessment of pathogenicity of the variants was performed following the ACMG guidelines (Richards et al., 2015). Conservation of mutated positions was evaluated using sequence alignment of different species. All *PGM3* variants were annotated based on the transcript ENST00000512866.1. Sanger sequencing was used to validate the positive findings and the variant origination.

Mutation Analysis

To evaluate the damaging effect of candidate variants, protein modeling was performed using the Iterative Threading ASSEmbly Refinement (I-TASSER) software (Zhang, 2009; Yang and Zhang, 2015). The confidence of each modeling was quantitatively measured by a C-score of 0.14. The three-dimensional structures were shown using PyMOL 1.7.

In an attempt to evaluate the genotype–phenotype correlation, *PGM3* mutations and their related phenotypes were systematically retrieved on the PubMed database till November 2019.

RESULTS

Identification of *PGM3* Variants

Four variants in the *PGM3* gene (OMIM* 172100) were identified, including one nonsense variant (c.1432C > T/p.Q478X) and three missense variants (c.478C > T/p.P160S, c.1239C > G/p.N413K, and c.1659T > A/p.N553K). Three variants (P160S, N413K, and Q478X) were identified in the cases with BECTS, and one variant (N553K) was in the case with BOE. Variants N413K and Q478X were *de novo*, while variants P160S and N553K originated

from their asymptomatic parents, respectively (**Figures 1A,B**). These variants in *PGM3* had no allele frequency in the gnomAD database. Variant P160S was predicted to be damaging or probably damaging by all used *in silico* programs, variant N413K was predicted to be damaging or probably damaging by 11 out of 16 prediction tools, and variant N553K was predicted to be damaging by SIFT and fitCons (**Supplementary Table 1**). The amino acid sequence alignment showed that residues P160 and N413 were highly conserved across vertebrates. Residue N553 was less conserved in vertebrates and did not appear in the transcript of several species (**Figure 1C**). Variant Q478X was estimated as a pathogenic variant, and variant N413K was estimated as a likely pathogenic variant by ACMG assessment, while variants P160S and N553K were uncertain significant variants. All cases had no other pathogenic or likely pathogenic variants in genes known to be associated with seizure disorders (Wang et al., 2017).

Structural Alteration of *PGM3* Protein

PGM3 contains four domains (Domains 1–4) (**Figure 2A**). The catalytic cleft is formed by the active serine loop in Domain 1, the metal-binding loop in Domain 2, the sugar-binding loop in Domain 3, and the phosphate-binding loop in Domain 4 (Nishitani, 2006). The truncating variant Q478X is located in Domain 4 and would lead to haploinsufficiency. Variant P160S is located in Domain 1, and variant N413K is located in Domain 3. The three-dimensional structural model of *PGM3* indicated that residues P160 and N413 lay in the inner side of the enzyme active center and were expected to be more "deleterious." Variant N553K, located in Domain 4, which lay in the C-terminal of Domain 4 and a distance away from the enzyme active center (**Figure 2B**), potentially led to mild damage effect. The molecular effects of the missense variants analyzed by protein modeling using I-TASSER showed that all the three missense variants led to the alterations of hydrogen bonds and may affect the protein steric configuration. Originally, residue P160 formed a hydrogen bond with Y164. When proline was replaced by serine at residue 160, additional two hydrogen bonds were formed with Y183 and T158 (**Figure 2C**). Residual N413 formed five hydrogen bonds with T376, A415, A416, D418, and S421, respectively. When asparagine was replaced by lysine, the hydrogen bonds with A415, A416, and D418 were destroyed, and only the hydrogen bonds with T376 and S421 were kept (**Figure 2D**). Residue N553 originally formed hydrogen bonds with I551, N552, A554, G558, and T559, respectively. When the asparagine was replaced by lysine at 553, the hydrogen bonds with N552, A554, and G558 were destroyed, and a new hydrogen bond with F555 was formed instead (**Figure 2E**).

Clinical Features of the Cases With *PGM3* Variants

All affected cases showed childhood-onset focal epilepsy. **Table 1** summarized the main clinical features of the four cases with *PGM3* variants. The cases with variants P160S, N413K, and Q478X were diagnosed as BECTS. The three cases started seizures at preschool or school age with generalized tonic–clonic seizure (GTCS) during sleep. The GTCSs occurred with approximately one to two attacks per year. No developmental delay, intellectual or speech disability, or visual abnormalities were observed. Their electroencephalograph (EEG) showed unilateral or bilateral independent centrotemporal discharges, dominantly during sleep (**Figures 3A–C**). The brain MRIs were normal. They all got seizure-free by low-dose monotherapy of valproic acid or lamotrigine. The case with variant N553K was diagnosed as BOE. He was a 3-year-old boy with normal development and had two GTCSs at the age of 2 years. The EEG obtained at the age of 3 years showed right occipital discharges (**Figure 3D**). His brain MRI was normal. He was seizure-free for 9 months without any treatment of antiepileptic drugs. Except seizures, no manifestation of immunodeficiency and CDG was observed in the four cases.

Genotype–Phenotype Correlation of *PGM3* Mutations

Previously, 18 mutations in *PGM3* were identified (**Table 2**). Sixteen out of 18 mutations, including 10 homozygous mutations and 3 compound heterozygous mutations, were identified in 16 families, involved 28 individuals (Sassi et al., 2014; Stray-Pedersen et al., 2014; Zhang et al., 2014; Lundin et al., 2015; Bernth-Jensen et al., 2016; Pacheco-Cuellar et al., 2017; Lundin et al., 2018; Abolhassani et al., 2019). All the individuals carrying biallelic mutations had the symptoms of immunodeficiency and/or congenital disorder of glycosylation (CDG). The neurological features included cognitive disability (12/28), developmental delay (12/28), seizures (10/28), characteristic facies (8/28), and hypotonia (4/28). The mutations included 11 missense, two small deletions, one small insertion, one splicing mutation, and one gross deletion including the whole gene of *PGM3*.

In addition, one heterozygous mutation c.626A > G/p.K209R was identified in a patient with fetal alcohol syndrome (de la Morena-Barrio et al., 2018); the other heterozygous mutation c.1369G > A/p.A457T was identified in a patient from a big cohort of developmental disorders by WES (Deciphering Developmental Disorders Study, 2017). No manifestation about immunodeficiency and glycosylation disturbance was mentioned.

DISCUSSION

PGM3 gene is located on chromosome 6q14.1–q15 (chr6:83164873-83193936) and encodes phosphoglucomutase 3 (*PGM3*). Previous studies have demonstrated that *PGM3* mutations were associated with severe immunodeficiency and glycosylation disorder, commonly accompanied by neurodevelopmental disorders and occasionally by seizures. In this study, four novel heterozygous *PGM3* variants, including two *de novo* variants, were identified in the cases with IFE and without the manifestation of immunodeficiency or glycosylation disorders. Further analysis indicated that biallelic mutations were associated with severer immunodeficiency or glycosylation disorders, whereas heterozygous mutations were potentially associated with mild phenotypes like IFE. The evidence

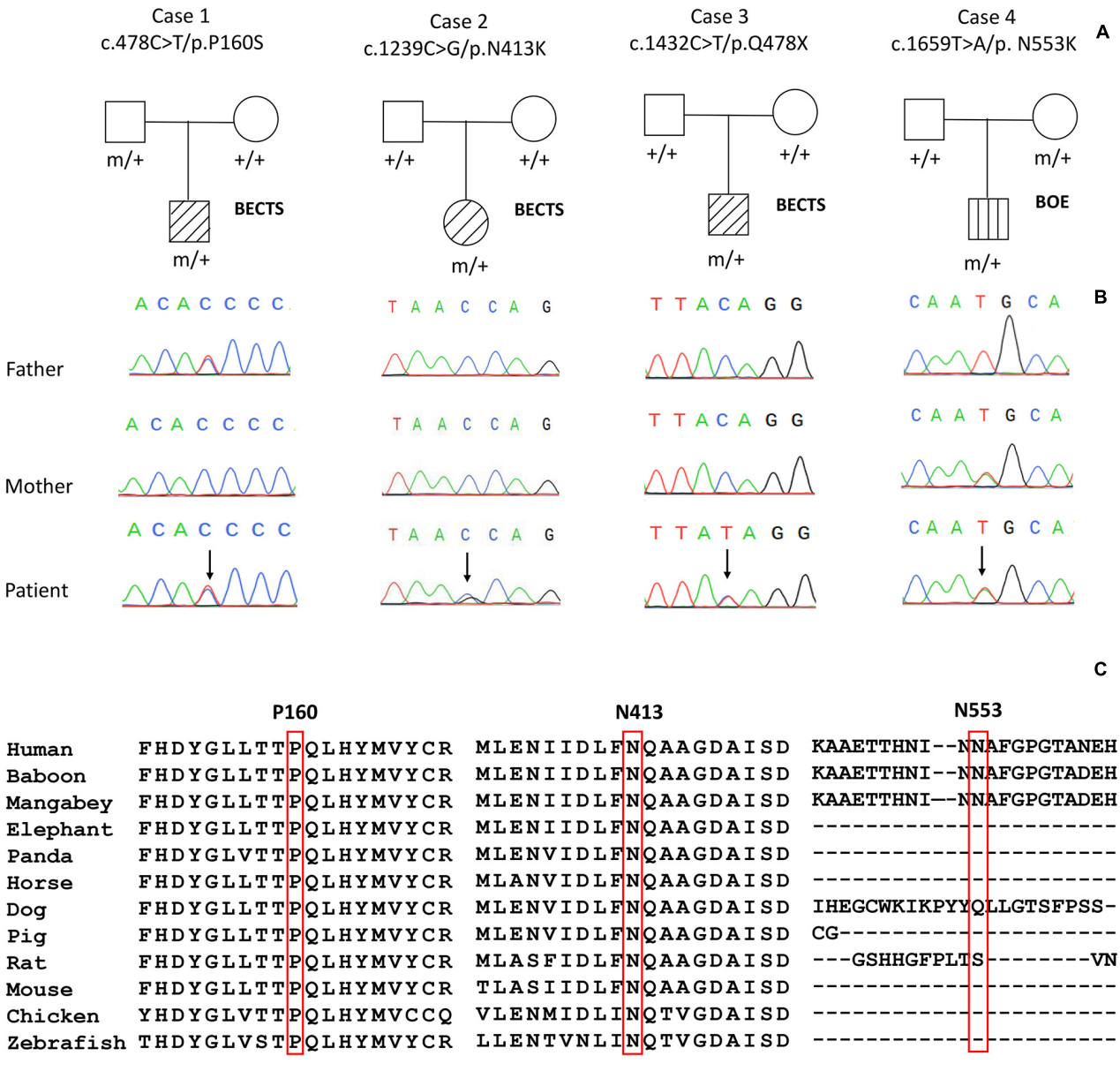


FIGURE 1 | Genetic data on the patients with idiopathic focal epilepsy with *PGM3* mutations. **(A)** Pedigrees of the four cases with *PGM3* mutations and their corresponding phenotypes. **(B)** DNA sequence chromatogram of the *PGM3* mutations. Arrows indicate the positions of the mutations. **(C)** The amino acid sequence alignment of the three missense mutations shows that residues P160 and N413 are highly conserved across vertebrates. Residue N553 is less conserved and does not appear in the transcript of several species.

suggests that *PGM3* gene is potentially a candidate pathogenic gene of epilepsy.

PGM3, which catalyzes the conversion of N-acetyl-glucosamine (GlcNAc)-6-phosphate into GlcNAc-1-phosphate, is universally required for the synthesis of UDP-GlcNAc, a sugar nucleotide critical to multiple glycosylation pathways (Stray-Pedersen et al., 2014). The *PGM3* gene is ubiquitously expressed across the whole lifespan (including embryonic period)[1]. In *PGM3*-deficiency mice, homozygous or compound heterozygous hypomorphic alleles caused trilineage cytopenias, even embryonic lethality (Greig et al., 2007), suggesting that *PGM3* is a vital gene for fetal development. In contrast, the heterozygous $PGM3^{+/gt}$ and $PGM3^{+/mld1}$ mice presented subclinical alterations like reduction of *PGM3* activity in multiple tissues (Greig et al., 2007), indicating a genetic dose effect (quantitative correlation). In humans, *PGM3* mutations were initially associated with IMD-23 characterized by recurrent respiratory, skin infections beginning in early childhood, and notably increased serum IgE in laboratory studies. Recently, *PGM3* mutations have also been identified as a cause of

[1]http://www.gdap.org.cn/genedetail.php?symbol=PGM3

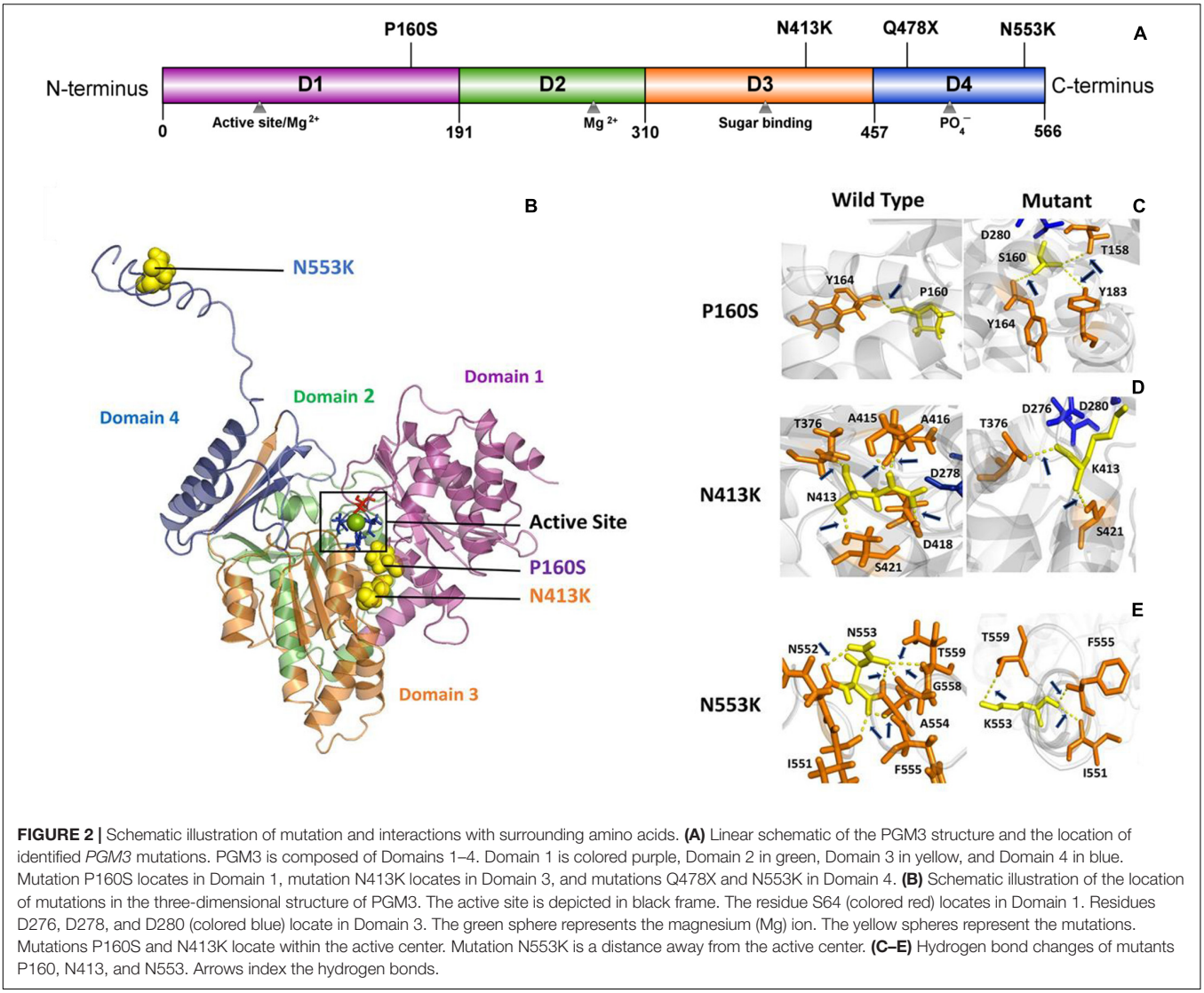


FIGURE 2 | Schematic illustration of mutation and interactions with surrounding amino acids. **(A)** Linear schematic of the PGM3 structure and the location of identified *PGM3* mutations. PGM3 is composed of Domains 1–4. Domain 1 is colored purple, Domain 2 in green, Domain 3 in yellow, and Domain 4 in blue. Mutation P160S locates in Domain 1, mutation N413K locates in Domain 3, and mutations Q478X and N553K in Domain 4. **(B)** Schematic illustration of the location of mutations in the three-dimensional structure of PGM3. The active site is depicted in black frame. The residue S64 (colored red) locates in Domain 1. Residues D276, D278, and D280 (colored blue) locate in Domain 3. The green sphere represents the magnesium (Mg) ion. The yellow spheres represent the mutations. Mutations P160S and N413K locate within the active center. Mutation N553K is a distance away from the active center. **(C–E)** Hydrogen bond changes of mutants P160, N413, and N553. Arrows index the hydrogen bonds.

CDG, usually exhibiting severe skeletal dysplasia, congenital malformations, and developmental delay. Among the patients with IMD-23 and CDG, neurodevelopmental abnormalities, including developmental delay, cognitive disability, seizures, and hypotonia, were common. Many patients with *PGM3* gene deficiency need hematopoietic stem cell therapy to save their life. Our analysis demonstrated that the cases with severe immunodeficiency and/or glycosylation disturbance all carried homozygous or compound heterozygous mutations, consistent with the results from biallelic deficiency animal models. Previously, two monoallelic mutations (K209R and A457T) were reported in the patients with fetal alcohol syndrome and developmental disorder, respectively (Deciphering Developmental Disorders Study, 2017; de la Morena-Barrio et al., 2018), and no immunodeficiency was mentioned. In the present study, four monoallelic mutations, including two *de novo* mutations, were identified in patients with pure IFE. One truncating mutation led to haploinsufficiency of *PGM3*. The other three missense mutations were highly conserved across the vertebrates and were predicted to alter hydrogen bond formation with surrounding amino acids. All mutations had no allele frequency in the gnomAD database and were predicted to be damaging by multiple software. Taking together the evidence that *PGM3* gene is expressed in brain and associated with neurodevelopmental disturbance, the *PGM3* gene is suggested to be potentially a candidate pathogenic gene of epilepsy. The genotype–phenotype correlations indicate that biallelic mutations in the *PGM3* gene are associated with severe immunodeficiency and glycosylation disturbance, while heterozygous mutations are potentially associated with mild phenotypes like IFE, in accordance with the findings in the *PGM3* deficiency mice.

PGM3 protein has four domains arranged in an overall heart shape similar to phosphoglucomutase 1 (PGM1) (Stiers et al., 2017). The catalytic cleft is formed by four loops from each domain. The active site of *PGM3* is located in

TABLE 1 | Clinical and genetic features of the cases with *PGM3* mutations.

	Case 1	Case 2	Case 3	Case 4
PGM3 mutation	c.478C > T/p.P160S	c.1239C > G/p.N413K	c.1432C > T/p.Q478X	c.1659T > A/p.N553K
Origin	Paternal	*De novo*	*De novo*	Maternal
Epilepsy syndrome	BECTS	BECTS	BECTS	BOE
Gender	Male	Female	Male	Male
Present age	12 yr	4 yr	17 yr	3 yr 9 mo
Age of seizure onset	7 yr 10 mo	3 yr	12 yr 4mo	1 yr 6 mo
Seizure types	sGTCS	GTCS	GTCS	GTCS
Frequency of seizure	1/yr	1/yr	2/mo	2/yr
Family history of seizure	None	None	None	None
Intelligence	Normal	Normal	Normal	Normal
Developmental delay	No	No	No	No
Speech	Normal	Stutter	Normal	Normal
Behavior	Normal	Salivation	Normal	Normal
Vision	Normal	Normal	Normal	Normal
Movement disorder	None	None	None	None
EEG discharges	Rt centrotemporal	Rt centroparietal	Bilateral temporal	Rt occipital
Brain MRI	Normal	Normal	Normal	Normal
Treatment	VPA 10 mg/kg/day	LTG 5 mg/kg/day	VPA 250 mg/bid	NA
Seizure outcome	Free for 3 yr	Free for 0.5 yr	Free for 4 yr	Free for 1.5 yr
MAF in gnomAD	None	None	None	None
ACMG assessment	Uncertain significant (PM2 + PP3)	Likely pathogenic (PS2 + PM2 + PP3)	Pathogenic (PVS1 + PS2 + PM2)	Uncertain significant (PM2)

GTCS, generalized tonic–clonic seizures; sGTCS, secondary GTCS; BECTS, benign epilepsy of childhood with centro-temporal spikes; BOE, benign occipital epilepsy; NA, not available; yr, year; mo, month; VPA, valproic acid; LTG, lamotrigine; MAF, minor allelic frequency; Rt, right.

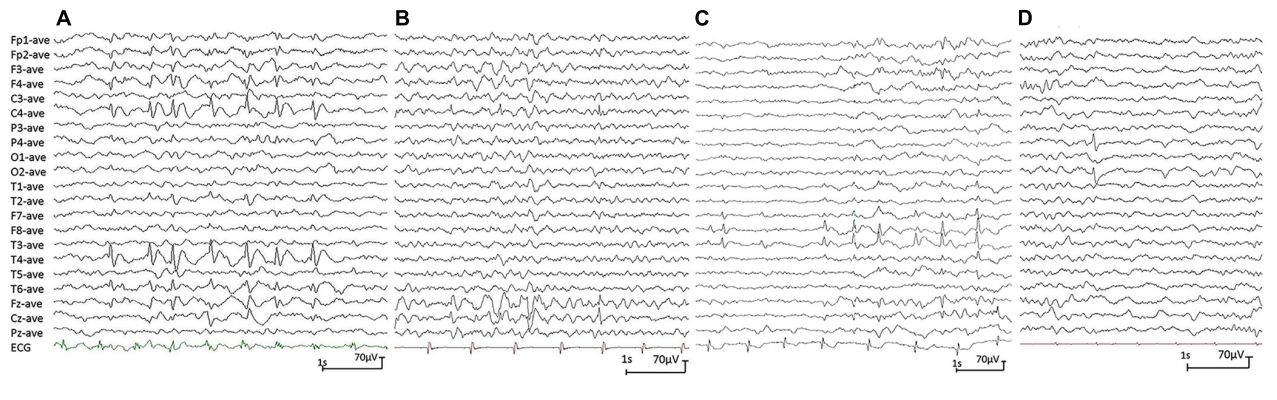


FIGURE 3 | Electroencephalograph (EEG) changes during non-rapid-eye-movement sleep in the cases with idiopathic focal epilepsy with *PGM3* mutations (referential montage with average reference was used in all EEGs). **(A)** Interictal EEG of case 1 showed right centrotemporal spike and slow waves (obtained at the age of 8 years). **(B)** Interictal EEG of case 2 showed right central spike and slow waves (at the age of 3 years). **(C)** Interictal EEG of case 3 showed bilateral temporal spike and slow waves (at the age of 12 years). **(D)** Interictal EEG of case 4 showed right occipital spike waves (at the age of 3 years).

a large central cleft, formed at the confluence of the four structural domains. The cleft contains (i) phosphoserine 64 that participates in phosphoryl transfer, (ii) the metal-binding loop including the three coordinating aspartates (residues 276, 278, and 280), (iii) a sugar-binding loop, and (iv) the phosphate-binding site (residue 505) (Uniprot database) (Stiers et al., 2016). In the present study, besides the truncating mutation Q478X, all missense mutations were predicted to change the hydrogen bonds with their surrounding amino acids and affect the protein structural stability, potentially impairing enzyme catalytic activity. Mutations P160S and N413K were located within the active center and were associated with typical BECTS. Instead, mutation N553K was identified in a family with BOE with incomplete penetrance. N553K was a distance away from the active center and was less conserved, suggesting less damage effect. Our recent study showed that molecular subregional location was a critical factor to determine the damaging effect that presented a continuous distribution and correlated with the severity of phenotypes (Tang et al., 2019), providing one of the explanations for the less damaging effect and atypical symptoms/incomplete penetrance associated with

TABLE 2 | Summary of neurological features in the cases/families with *PGM3* mutations.

Allele 1	Allele 2	Phenotype	Age of onset	Cognitive disability	Developmental delay	Seizure	Characteristic facies	Hypotonia	References
Homozygous mutations									
c.248T > C/p.L83S	c.248T > C/p.L83S	IMD	3–4 mo	1/4	3/4	–	4/4	1/4	Sassi et al., 2014
c.631G > C/p.D211H	c.631G > C/p.D211H	IMD	1–8 yr	NA	NA	NA	NA	NA	Abolhassani et al., 2019
c.737A > G/p.N246S	c.737A > G/p.N246S	CDG, IMD	Birth	2/2	2/2	–	2/2	–	Stray-Pedersen et al., 2014; Bernth-Jensen et al., 2016
c.877 + 3A > G	c.877 + 3A > G	IMD	3 w	–	–	–	–	–	Lundin et al., 2018
c.891T > G/p.D297E	c.891T > G/p.D297E	IMD, NI	10–12 yr	2/3	3/3	2/3	–	1/3	Zhang et al., 2014
c.965T > C/p.I322T	c.965T > C/p.I322T	IMD	13 mo	–	–	–	–	–	Lundin et al., 2015
c.1019_1021delAAG/p.E340del	c.1019_1021delAAG/p.E340del	IMD	1–7 mo	2/4	2/4	1/4 (3 yr)	–	2/4	Sassi et al., 2014
c.1135T > C/p.F379L	c.1135T > C/p.F379L	IMD, CM	Birth	NA	NA	–	+	–	Pacheco-Cuellar et al., 2017
c.1504G > T/p.D502Y	c.1504G > T/p.D502Y	IMD	6.5 yr	–	1/1	1/1	–	–	Sassi et al., 2014
c.1558G > A/p.A520T	c.1558G > A/p.A520T	IMD	1 yr	NA	NA	NA	NA	NA	Abolhassani et al., 2019
Compound heterozygous mutations									
c.715G > C/p.D239H	1.2 Mb incl. entire gene	CDG, IMD, SD	Birth	–	–	–	–	–	Stray-Pedersen et al., 2014
c.1352A > G/p.Q451R	c.737dupA/p.N246K fsX7	CDG, IMD, SD	Birth	1/1	1/1	1/1	1/1	–	Stray-Pedersen et al., 2014
c.1501G > C/p.E501Q	c.1354_1358delCTTAA/p.L452SfsX10	IMD, NI	32 yr	4/5	–	1/5	–	–	Zhang et al., 2014
Monoallelic mutations									
c.478C > T/p.P160S	–	BECTS	7 yr	–	–	1/2	–	–	The present study
c.626A > G/p.K209R	–	FAS	1 yr	–	–	–	1/1	–	de la Morena-Barrio et al., 2018
c.1239C > G/p.N413K	–	BECTS	3 yr	–	–	1/1	–	–	The present study
c.1369G > A/p.A457T	–	DD	NA	NA	NA	NA	NA	NA	The Deciphering Developmental Disorders Study, 2017
c.1432C > T/p.Q478X	–	BECTS	12 yr	–	–	1/1	–	–	The present study
c.1659T > A/p.N553K	–	BOE	1 yr	–	–	1/2	–	–	The present study

The fractions in the table represent the ratio of the number of cases with corresponding symptoms to the total number of cases. IMD, immunodeficiency; CDG, congenital disorder of glycosylation; NI, neurocognitive impairment; CM, congenital malformations; SD, skeletal dysplasia; FAS, fetal alcohol syndrome; DD, developmental disorder; BECTS, benign epilepsy of childhood with centrotemporal spikes; BOE, benign occipital epilepsy; NA, not available; yr, year; mo, month; w, week.

N553K. However, further studies are required to determine its pathogenicity and association with BOE.

IFE is a group of self-limited epilepsies with infrequent partial seizures and a normal background of EEG. Although both BECTS and BOE are regarded as genetically determined epileptic syndromes, big pedigrees are rare. Majority of patients with BECTS and BOE are sporadic. Incomplete penetrance is potentially one of the explanations. The present study demonstrated that less impaired functions of the *PGM3* gene, i.e., heterozygous mutations, were associated with IFE. Additionally, heterozygous mutations with a less damaging effect were potentially associated with the incomplete penetrance. These findings were consistent with the evidence of quantitative correlation from *PGM3* deficiency models, in which the heterozygous damaging resulted in subclinical abnormalities, such as decrease in enzyme activity and hematological alterations (Greig et al., 2007). In the literature, clinical information about parents in the biallelic cases was not available. Generally, in the patients with BECTS and BOE, epileptiform discharges in EEG will disappear, and epileptic seizures will stop in adulthood. Even in childhood, the epileptic seizures in BECTS and BOE are infrequent and usually occur during sleep, which are usually overlooked. These could be the reasons of incomplete penetrance in the parents with PGM3 heterozygous mutations. It was also found that most early-onset patients with immunodeficiency had no seizure, while the late-onset patients had seizure manifestations, suggesting an age-dependent penetrance. The early-onset patients usually had more severe symptoms and rarely survived to the age of epilepsy occurrence. This may be one of the explanations for a portion of the patients with biallelic mutations having no seizure.

Clinically, the features of BECTS and BOE are often overlapped. Up to 20% patients with BOE have rolandic spikes as those with BECTS (Lada et al., 2003). BECTS and BOE may appear in the same patient at different life periods. In the present study, *PGM3* mutations were identified in both BECTS and BOE, suggesting that BECTS and BOE potentially have similar etiologies. To date, multiple genes have been implicated in IFE. The role of multiple rare heterozygous variants in IFE warrants further study.

In conclusion, we identified four novel heterozygous *PGM3* mutations in patients with IFE without apparent manifestation of immunodeficiency and glycosylation disorder. The genetic and molecular evidence implies that the mutations lead to defect of the *PGM3* gene. Taking together the data from the gene expression profile and the *PGM3* deficiency animal model, it is suggested that *PGM3* gene is potentially a candidate pathogenic gene of epilepsy. Further analysis revealed that previously reported biallelic mutations in the *PGM3* gene were associated with severe immunodeficiency and glycosylation disturbance, while monoallelic mutations were associated with mild phenotypes like IFE, suggesting a quantitative correlation between genetic impairment and phenotypic severity, which help explain the mild symptoms and incomplete penetrance in individuals with IFE.

DATA AVAILABILITY STATEMENT

The datasets for this article are not publicly available due to concerns regarding participant/patient anonymity. Requests to access the datasets should be directed to the corresponding author.

ETHICS STATEMENT

The Ethics Committee of the Second Affiliated Hospital of Guangzhou Medical University provided ethics approval. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individuals and minors' legal guardian for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

X-RL and W-PL contributed to the conception of the study. X-RL contributed to the interpretation of clinical data and drafting of the figures and the manuscript. W-JB, JW, and T-TY examined the patient and participated in drafting of the manuscript. B-ML, D-TL, BT, and W-WD contributed to the collection and analysis of clinical data. Y-WS and TS helped perform the analysis with constructive discussions. W-PL provided critical review and substantially revised the manuscript. All the authors read and approved the manuscript before sending the manuscript to the journal for publication.

FUNDING

This work was supported by the National Natural Science Foundation of China (81871015), China Association Against Epilepsy-UCB Pharma Ltd., Joint Science Research Foundation (2018024), Science and Technology Project of Guangdong Province (2017B09090436 and 2017B030314159), Natural Science Foundation of Guangdong Province (2020A1515010108), and Science and Technology Project of Guangzhou (201904020028).

ACKNOWLEDGMENTS

We are deeply grateful to the patients and clinicians who participated in this work and appreciate the supporting of the China Epilepsy Gene 1.0 Project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.559080/full#supplementary-material

REFERENCES

Abolhassani, H., Aghamohammadi, A., Fang, M., Rezaei, N., Jiang, C., Liu, X., et al. (2019). Clinical implications of systematic phenotyping and exome sequencing in patients with primary antibody deficiency. *Genet. Med.* 21, 243–251. doi: 10.1038/s41436-018-0012-x

Bernth-Jensen, J. M., Holm, M., and Christiansen, M. (2016). Neonatal-onset T(-)B(-)NK(+) severe combined immunodeficiency and neutropenia caused by mutated phosphoglucomutase 3. *J. Allergy Clin. Immunol.* 137, 321–324. doi: 10.1016/j.jaci.2015.07.047

Commission on Classification and Terminology of the International League Against Epilepsy. (1989). Proposal for revised classification of epilepsies and epileptic syndromes. Commission on Classification and Terminology of the International League Against Epilepsy. *Epilepsia* 30, 389–399. doi: 10.1111/j.1528-1157.1989.tb05316.x

de la Morena-Barrio, M. E., Ballesta-Martinez, M. J., Lopez-Galvez, R., Anton, A. I., Lopez-Gonzalez, V., Martinez-Ribot, L., et al. (2018). Genetic predisposition to fetal alcohol syndrome: association with congenital disorders of N-glycosylation. *Pediatr. Res.* 83, 119–127. doi: 10.1038/pr.2017.201

Deciphering Developmental Disorders Study, (2017). Prevalence and architecture of de novo mutations in developmental disorders. *Nature* 542, 433–438. doi: 10.1038/nature21062

Engel, J. (2006). ILAE classification of epilepsy syndromes. *Epilepsy Res.* 70, 5–10. doi: 10.1016/j.eplepsyres.2005.11.014

Greig, K. T., Antonchuk, J., Metcalf, D., Morgan, P. O., Krebs, D. L., Zhang, J. G., et al. (2007). Agm1/Pgm3-mediated sugar nucleotide synthesis is essential for hematopoiesis and development. *Mol. Cell. Biol.* 27, 5849–5859. doi: 10.1128/mcb.00802-07

Lada, C., Skiadas, K., Theodorou, V., Loli, N., and Covanis, A. (2003). A Study of 43 patients with panayiotopoulos syndrome, a common and benign childhood seizure susceptibility. *Epilepsia* 44, 81–88. doi: 10.1046/j.1528-1157.2003.32602.x

Lal, D., Reinthaler, E., Schubert, J., Muhle, H., Riesch, E. G. K., Kluger, G., et al. (2014). DEPDC5 mutations in genetic focal epilepsies of childhood. *Ann. Neurol.* 75, 788–792.

Lemke, J. R., Lal, D., Reinthaler, E. M., Steiner, I., Nothnagel, M., Alber, M., et al. (2013). Mutations in GRIN2A cause idiopathic focal epilepsy with rolandic spikes. *Nat. Genet.* 45, 1067–1072. doi: 10.1038/ng.2728

Lundin, K. E., Hamasy, A., Backe, P. H., Moens, L. N., Falk-Sorqvist, E., Elgstoen, K. B., et al. (2015). Susceptibility to infections, without concomitant hyper-IgE, reported in 1976, is caused by hypomorphic mutation in the phosphoglucomutase 3 (PGM3) gene. *Clin. Immunol.* 161, 366–372. doi: 10.1016/j.clim.2015.10.002

Lundin, K. E., Wang, Q., Hamasy, A., Marits, P., Uzunel, M., Wirta, V., et al. (2018). Eleven percent intact PGM3 in a severely immunodeficient patient with a novel splice-site mutation, a case report. *BMC Pediatr.* 18:285. doi: 10.1186/s12887-018-1258-9

Pacheco-Cuellar, G., Gauthier, J., Desilets, V., Lachance, C., Lemire-Girard, M., Rypens, F., et al. (2017). A novel PGM3 mutation is associated with a severe phenotype of bone marrow failure, severe combined immunodeficiency, skeletal dysplasia, and congenital malformations. *J. Bone Miner. Res.* 32, 1853–1859. doi: 10.1002/jbmr.3173

Reinthaler, E. M., Lal, D., Jurkowski, W., Feucht, M., Steinbock, H., Gruber-Sedlmayr, U., et al. (2014). Analysis of ELP4, SRPX2, and interacting genes in typical and atypical rolandic epilepsy. *Epilepsia* 55, e89–e93. doi: 10.1111/epi.12712

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. *Genet. Med.* 17, 405–424. doi: 10.1038/gim.2015.30

Sassi, A., Lazaroski, S., Wu, G., Haslam, S. M., Fliegauf, M., Mellouli, F., et al. (2014). Hypomorphic homozygous mutations in phosphoglucomutase 3 (PGM3) impair immunity and increase serum IgE levels. *J. Allergy Clin. Immunol.* 133, 1410-9, 1419.e1-e13. doi: 10.1016/j.jaci.2014.02.025

Scheffer, I. E., Berkovic, S., Capovilla, G., Connolly, M. B., French, J., Guilhoto, L., et al. (2017). ILAE classification of the epilepsies: position paper of the ILAE Commission for Classification and Terminology. *Epilepsia* 58, 512–521. doi: 10.1111/epi.13709

Stiers, K. M., Graham, A. C., Kain, B. N., and Beamer, L. J. (2017). Asp263 missense variants perturb the active site of human phosphoglucomutase 1. *FEBS J.* 284, 937–947. doi: 10.1111/febs.14025

Stiers, K. M., Kain, B. N., Graham, A. C., and Beamer, L. J. (2016). Induced structural disorder as a molecular mechanism for enzyme dysfunction in phosphoglucomutase 1 deficiency. *J. Mol. Biol.* 428, 1493–1505. doi: 10.1016/j.jmb.2016.02.032

Stray-Pedersen, A., Backe, P. H., Sorte, H. S., Mørkrid, L., Chokshi, N. Y., Erichsen, H. C., et al. (2014). PGM3 mutations cause a congenital disorder of glycosylation with severe immunodeficiency and skeletal dysplasia. *Am. J. Hum. Genet.* 95, 96–107. doi: 10.1016/j.ajhg.2014.05.007

Strug, L. J., Clarke, T., Chiang, T., Chien, M., Baskurt, Z., Li, W., et al. (2009). Centrotemporal sharp wave EEG trait in rolandic epilepsy maps to Elongator Protein Complex 4 (ELP4). *Eur. J. Hum. Genet.* 17, 1171–1181. doi: 10.1038/ejhg.2008.267

Tang, B., Li, B., Gao, L. D., He, N., Liu, X. R., Long, Y. S., et al. (2019). Optimization of in silico tools for predicting genetic variants: individualizing for genes with molecular sub-regional stratification. *Brief. Bioinform.* 2019, bbz115. doi: 10.1093/bib/bbz115

Wang, J., Lin, Z. J., Liu, L., Xu, H. Q., Shi, Y. W., Yi, Y. H., et al. (2017). Epilepsy-associated genes. *Seizure* 44, 11–20. doi: 10.1016/j.seizure.2016.11.030

Weir, E., Gibbs, J., and Appleton, R. (2018). Panayiotopoulos syndrome and benign partial epilepsy with centro-temporal spikes: a comparative incidence study. *Seizure* 57, 66–69. doi: 10.1016/j.seizure.2018.03.002

Yang, J., and Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* 43, W174–W181. doi: 10.1093/nar/gkv342

Zhang, Y. (2009). I-TASSER: fully automated protein structure prediction in CASP8. *Proteins* 77, 100–113. doi: 10.1002/prot.22588

Zhang, Y., Yu, X., Ichikawa, M., Lyons, J. J., Datta, S., Lamborn, I. T., et al. (2014). Autosomal recessive phosphoglucomutase 3 (PGM3) mutations link glycosylation defects to atopy, immune deficiency, autoimmunity, and neurocognitive impairment. *J. Allergy Clin. Immunol.* 133, 1400-9, 1409.e1-e5. doi: 10.1016/j.jaci.2014.02.013

Conflict of Interest: This study received funding from UCB Pharma Ltd. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychiatry

REVIEW
published: 22 October 2020
doi: 10.3389/fpsyt.2020.587162

Postzygotic Somatic Mutations in the Human Brain Expand the Threshold-Liability Model of Schizophrenia

Shiva M. Singh, Christina A. Castellani*† and Kathleen A. Hill**

Molecular Genetics Unit, Department of Biology, The University of Western Ontario, London, ON, Canada

The search for what causes schizophrenia has been onerous. This research has included extensive assessment of a variety of genetic and environmental factors using ever emerging high-resolution technologies and traditional understanding of the biology of the brain. These efforts have identified a large number of schizophrenia-associated genes, some of which are altered by mutational and epi-mutational mechanisms in a threshold liability model of schizophrenia development. The results, however, have limited predictability and the actual cause of the disease remains unknown. This current state asks for conceptualizing the problem differently in light of novel insights into the nature of mutations, the biology of the brain and the fine precision and resolution of emerging technologies. There is mounting evidence that mutations acquired during postzygotic development are more common than germline mutations. Also, the postzygotic somatic mutations including epimutations (PZMs), which often lead to somatic mosaicism, are relatively common in the mammalian brain in comparison to most other tissues and PZMs are more common in patients with neurodevelopmental mental disorders, including schizophrenia. Further, previously inaccessible, detection of PZMs is becoming feasible with the advent of novel technologies that include single-cell genomics and epigenomics and the use of exquisite experimental designs including use of monozygotic twins discordant for the disease. These developments allow us to propose a working hypothesis and expand the threshold liability model of schizophrenia that already encompasses familial genetic, epigenetic and environmental factors to include somatic *de novo* PZMs. Further, we offer a test for this expanded model using currently available genome sequences and methylome data on monozygotic twins discordant for schizophrenia (MZD) and their parents. The results of this analysis argue that PZMs play a significant role in the development of schizophrenia and explain extensive heterogeneity seen across patients. It also offers the potential to convincingly link PZMs to both nervous system health and disease, an area that has remained challenging to study and relatively under explored.

Keywords: neurodevelopment, *de novo* mutations, postzygotic somatic mutations, epimutations, neuronal diversity, mosaicism, neurological disorders, threshold liability model

OPEN ACCESS

Edited by:
Elena Martín-García, Pompeu Fabra University, Spain

Reviewed by:
Mark Samuels, Université de Montréal, Canada
Pippa Ann Thomson, Medical Research Council Institute of Genetics and Molecular Medicine (MRC), United Kingdom

****Correspondence:***
Shiva M. Singh
ssingh@uwo.ca
Christina A. Castellani
ccastel3@uwo.ca
Kathleen A. Hill
khill22@uwo.ca

†Present address:
Christina A. Castellani, Department of Genetic Medicine, McKusick-Nathans Institute, Johns Hopkins University School of Medicine, Baltimore, MD, United States

Specialty section:
This article was submitted to Behavioral and Psychiatric Genetics, a section of the journal Frontiers in Psychiatry

Received: *25 July 2020*
Accepted: *22 September 2020*
Published: *22 October 2020*

Citation:
Singh SM, Castellani CA and Hill KA (2020) Postzygotic Somatic Mutations in the Human Brain Expand the Threshold-Liability Model of Schizophrenia. Front. Psychiatry 11:587162. doi: 10.3389/fpsyt.2020.587162

INTRODUCTION

Mutations provide the foundation for individual genetic differences. They also play a critical role in health and disease as their effect on an individual may range from being lethal to beneficial. Although most genetic variations in populations and families are passed on from generation to generation, some may be acquired *de novo*. Traditionally, *de novo* mutations of interest to diseases have been identified by the genomic difference between parents and the progeny with the disease. The assumption is that the mutation causing the disease in the progeny must have arisen in one of the parental gametes used to generate the zygote and the resulting progeny. Such *de novo* germ line mutations may become part of the familial gene pool. The occurrence of *de novo* mutations however is not restricted to the germ line, only. Most represent errors in DNA replication that may follow meiosis to generate gametes or mitosis to increase the number of somatic cells during development and differentiation of the zygote. Theoretically, postzygotic *de novo* mutations may originate at any time during development, from the early two-cell dividing embryo to any stage of an individual's prenatal and postnatal development. Also, cells carrying *de novo* postzygotic mutations may become part of an individual's heterogeneous genomic composition. As such, they have been difficult, even daunting to identify and appraise. Improved technologies and innovative experimental designs, such as single-cell genome sequencing, have revolutionized this research. For the first time it is possible to identify postzygotic mutations, with increasing sensitivity using ongoing revolutions in genomic technologies (1–7).

Postzygotic somatic mutations may contribute to mosaicism and add another layer of genetic variation across individuals, potentially affecting physiology, function and phenotype. Although it was once assumed that such mutations are rare and inconsequential they may be almost two orders of magnitude more frequent than germline mutations (8). Yet, the degree and consequences of mosaicism resulting from somatic mutations have not been adequately assessed, primarily due to inaccessibility of needed samples and lack of necessary technologies. As such, somatic mutations represent the latest addition to measures of biological diversity, an area of research that until recently relied almost exclusively on germline mutations. Generally, postzygotic development has been characterized by an increase in cell number via mitosis and strictly regulated dynamics in gene expression. The cells produced by mitosis during ontogeny have long-since been considered primarily genomic clones of the zygote. Although mutations representing DNA sequence changes may theoretically arise with every cycle of mitosis (9), the fate of cells carrying such mutations is not assured. If a mutation occurs very early in development, the new mutation may be incorporated in most tissues of the progeny. However, if the mutation occurs later in development, fewer cells will carry the mutation, and the mutation-carrying cells may be confined to a given tissue and/or cell type. Additionally, cells carrying new mutation(s) may be selected against during development, while others may have mutations that result in little to no phenotypic effect (low-level or micro-mosaicism). Alternatively, some such mutations can have a positive effect on cell proliferation and subsequently drive the accumulation of mutant cells. These driver mutations may result in significant mosaicism with potential to affect the phenotype. Additionally, somatic mutations can arise as random events, programmed events, or as the product of an inherited genotype as in the case of a mutator phenotype (10). A long-term consequence of ongoing somatic mutations may be that some/most non-germ cells (soma) will differ in their genomes and thus in theory no two somatic cells may harbor 100% identical genomes (11–13). The consequence is that every individual will embody a composite mosaic of genetically distinct cells. The contributions of somatic mutations and somatic mosaicism were previously not considered biologically significant, and the phenomenon of postzygotic mutation has remained largely underexplored due to the biological complexity of mutagenesis, mosaicism, access of required cells and tissues for evaluation and the technical challenges of single-cell "omics." Despite these hurdles, the potential of postzygotic mutations has been implicated in a number of diseases [see review, (14)]. Specifically and most relevant to this discussion they have been implicated in a number of mental disorders [see recent review, (15)]. In this overview, rather than offering yet another extensive review on the subject, we offer a measured perspective on the phenomenon of developmental postzygotic mutations relevant to mental disorders and specifically expand upon the threshold model of schizophrenia, a devastating life-changing neurodevelopmental disorder of poorly understood etiology despite the increasing appreciation of the contribution of postzygotic events in complex diseases (reviewed elsewhere).

SOMATIC MUTATIONS ARE HETEROGENEOUS AND ACCUMULATE OVER TIME WITH EFFECTS ON NEUROLOGICAL PHENOTYPES

Empirically, somatic mutations leading to mosaics with different genotypes within an individual have been recognized over the decades (16), but the nature and extent of the mosaicism is almost never known (2). Rare studies that have begun to characterize the origin, nature, and consequence of somatic mutations and mosaicism show that the rate of *de novo* somatic mutations leading to mosaicism differs with age and across tissues, individuals, and families (2, 4, 17–20). Also, somatic mutations may arise as single-nucleotide substitutions, transpositions, insertions, deletions (including copy number change), and aneuploidies. These events appear at random and are likely caused by a variety of mutational mechanisms. Some of the mutational mechanisms may be development-specific while others may constantly add new mutations throughout life in a clock-like manner (9). The rate and propensity of somatic mutations may also differ across genes and gene sequences (20). Interestingly, despite the assumption that such mutations are stochastic, there is emerging evidence that some somatic mutations may be programmed, as in the case of somatic hypermutation involving immunoglobulin

genes (21–25). Developmental windows for sensitivity to mutational events have also been demonstrated in mouse neuronal cells (26). Such results allow us to hypothesize that postzygotic somatic mutations have the potential to contribute to a broad spectrum of neurological phenotypes from health to disease but remain unexplored. To this end, the better understood dominant role of somatic mutations in cancer could be used as a working model that can be applied to the involvement of PZMs in neurological phenotypes and disorders.

POSTZYGOTIC SOMATIC MUTATIONS CONTRIBUTE TO DISEASE: THE CANCER MODEL

Initiation and progression of different forms of cancer represent two of the most explored areas of *de novo* genetic changes in somatic cells over the last few decades (27, 28). This extensive research has identified a long list of critical genes that are reported to undergo somatic changes in variety of cancers (http://cancer.sanger.ac.uk). The results offer a comprehensive resource for exploring the impact of gene-specific somatic mutations, both individually and in combination, across different forms of cancers. Follow-up comprehensive studies on such genes have provided the precise role of these mutations in oncogenic transformations. Some of these insights have been possible via increased resolution involving studies on single cells that include single-cell sequencing (29) and single-cell multiomics (30). The results of these studies have assisted in uncovering molecular insights relevant to the initiation and progression of different types of cancers (31–33). Of special interest to this discussion is the role of *de novo* mutations acting as the "second hit" that may initiate and subsequently serve as the "driving force" behind unchecked proliferation and progression of different cancer phenotypes (34), including metastasis (35). Also, *de novo* mutations have been observed to increase at a constant rate with each genome replication in mutator phenotypes (36) or increase sporadically with transient hypermutability (31, 32) in different forms and stages of cancer (37–39). These results have allowed characterization of subtypes of cancers based on signatures of *de novo* mutations and identification of critical players at different stages in carcinogenesis (40). Translationally, such insights have helped to redefine cancer from an untreatable, poorly understood disease to one that can be classified, diagnosed, treated, and in some cases, cured (40), thus adding new hope to what would otherwise be a bleak diagnosis just a few years ago. We note that neurological aberrations are currently poorly understood but may involve postzygotic somatic mutations that can be detected and classified. As in the case of cancer, they may provide a valuable starting point toward gaining insight into the etiology of specific neurological disorders. It may lead to the characterization of gene specific changes including somatic mosaicism that may translate to treatment successes for neurological disorders as exemplified by the cancer model.

POSTZYGOTIC SOMATIC MUTATIONS OFTEN ARISE DURING NEURODEVELOPMENT

Often it has been assumed that all cells in the brain have identical genomes. However, postzygotic somatic mutations arising during the development of the brain have the potential to affect large and small clonal lineages depending upon the developmental timing of the mutation and the resultant cell populations carrying the *de novo* mutation (41). They may generate differences between the neuronal structures of monozygotic twins that started life as a single zygote (**Figure 1**). Also, they could contribute to the complex spectrum of brain phenotypes across individuals in the population including some disease phenotypes (**Figure 2**). Heterogeneous PZMs are often reported in the brain (41–43), but it is not known whether they arise through a random process, if they are inherently directed, or if they occur in response to environmental cues. What is understood, however, is that new mutations are somatic, present in the brain and are not directly transmitted to the next generation. PZMs however, may ensure that neuronal genomes in an individual are not singular, homogeneous, or static, but instead establish mosaic, heterogeneous and dynamic populations of neural genomes, with new neurons arising throughout life (42). Among other processes, neurons are known to undergo *de novo* long interspersed nuclear element (LINE-1/L1) retrotransposition (44), as has been reported in adult neurons (45, 46) and during embryonic development (43, 47, 48). Interestingly, the number of retrotranspositions has been reported as significantly higher in brain than non-brain tissue samples (49). Researchers have argued that every cell in the human brain may contain a number of somatic insertions and that retrotransposition may play an important role in reshaping the genetic circuitry (50). This phenomenon can lead to neuron-to-neuron variation, a neuron-specific transcriptome, and a neurobiological phenotype (51–53). Besides retrotranspositions, neuronal variations may also arise from additional genetic alterations in whole chromosome numbers (54), rearrangement of mobile elements (50), insertions/deletions (55), and single nucleotide variants (SNVs) (18). Theoretically, there is no limit to the kind and amount of genomic variation within an individual, and therefore every gene in the human neural genome may be mutated in some neurons (56). The potential exists to uniquely define nearly 100 billion neurons and over 100 trillion neuronal connections. In this context, PZMs will add an extra level of variation for plasticity, adaptability, and resilience to the dynamics of environmental change and insults, which is specifically operational and sensitive in the brain (53). Such results argue that PZMs may be key contributors to intra-individual variability that may affect neuronal phenotypes. In doing so, they also provide a novel and unique insight in the biology of the brain, each representing a unique composite of mosaics that has remained unexplored. What is being realized is that postzygotic transposition events are higher in some mental disorders (49). These events preferentially affect genes associated with neuronal functions, and an uncontrolled

retrotransposition may increase the risk of mutations leading to disorders. Indeed, pre-existing retrotransposons may act as "lightning rods" for novel insertions, which may modulate gene expression (49). Specifically, *de novo* PZMs have been reported in a number of neurodegenerative (57, 58) as well as neurodevelopmental diseases (59–61). Of special interest to this discussion is the high frequency of *de novo* mutations reported in patients for neurodevelopmental disorders such as autism (62) and schizophrenia (63–66). Here, we will specifically focus on schizophrenia as a model of neurodevelopmental disorders in assessing the involvement of PZMs in brain disorders.

POSTZYGOTIC MUTATIONS ARE CANDIDATE CONTRIBUTORS TO SCHIZOPHRENIA

Schizophrenia is a complex neurodevelopmental disorder that is present worldwide at a relatively stable frequency (∼1%). The disease is heterogeneous, often expresses in early adulthood and difficult to early diagnose in absence of any biological test. It has high heritability (80%), as well as high discordance in monozygotic twins (50%) (67). The search for genes and inherited factors causing schizophrenia has been long and exhaustive but the identification of causal gene(s) has been elusive. However, this meticulous and long research has identified a large number of schizophrenia-associated single nucleotide polymorphisms (SNPs) and copy number variations (CNVs) in hundreds of genes and non-coding loci (http://www.szdb.org). As it stands, most of the findings reported in this database have been difficult to replicate and are not unique to this disease (68). Some of these genes and mutations could provide predisposition diagnostic information including clinical spectrum for schizophrenia and other related disorders. Further, an assessment of results by the Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) has identified 108 of the most common gene variants that have met a high degree of significance but explain a small fraction of the disease risk. Yet additional analysis has led to the prioritization of 145 most common schizophrenia variants that may serve as a foundation for patient specific genetic liability for this disease (69). These variants may directly or indirectly contribute to a disease threshold that could be achieved by inherited variants in concert with still-uncharacterized random events, including the effect of the environment. Indeed, most research on the development of schizophrenia is compatible with this long-standing threshold model (70, 71).

Of the random events reported in schizophrenia, two biological features appear noteworthy. The first involves *de novo* DNA sequence changes including CNVs reported in the brain (72) and blood (73–75). These often affect synaptic genes, and associated CNVs tend to occur in genome regions that are prone to recurrent mutations (76) and are implicated in schizophrenia (77, 78). The second random feature implicated in schizophrenia is epimutations, particularly changes in DNA methylation (79). They may undergo such changes randomly or in response to external exposures and be passed on through succeeding mitotic cycles (9, 16). Epimutations, that often affect gene expression may complicate familial tendencies with random and environmentally-responsive events. Additionally, there may be a role for epigenetically-regulated human endogenous retroviral HERV and related sequences (80). HERVs are retroviruses that have the potential to transpose and facilitate copy number changes. They may also affect epigenetic features (80). Such sequences have been uniquely isolated from the genomes of the affected members of the monozygotic twin pairs discordant for schizophrenia (MZD) (81). Additionally, HERV-related DNA/RNA sequences that were detected in the genomes of the affected members of MZD for schizophrenia have been reported to be elevated in schizophrenia patients (82). These sequences may represent copy number changes and/or increased expression with potential to increase transposition. Additionally, the HERV-associated C4 locus within the major histocompatibility complex on chromosome 6 may affect C4 expression, which is involved in the pruning of the dendritic spine in schizophrenia (83). These results support the hypothesis that schizophrenia may involve mutations representing DNA sequence changes as well as changes representing any epigenetic modifications that may affect gene expression. Together, they may function as the sum total of postzygotic somatic modifications (PZMs) that may contribute to the development of schizophrenia. Some of these modifications may be hereditary and transmitted through the parent(s), while others may represent postzygotic changes that are now technologically identifiable (using exceptional methods) and allow for proposition of an expansion of the threshold-liability model for schizophrenia.

AN EXPANDED-THRESHOLD-LIABILITY MODEL FOR SCHIZOPHRENIA MAY INCLUDE POSTZYGOTIC EVENTS

There is mounting evidence that every brain is a unique mosaic representing a composite of genetically distinct cells [see review (83)]. It may apply to a "normal" brain as well as a brain from an individual with a neurological disorder, including schizophrenia (56). Here, the brain from a patient may carry many more mutations/modifications affecting pathway(s) defective in schizophrenia. Also, given extensive heterogeneity, the number and type of genes affected may be unique across all/most patients with schizophrenia. Brains from different patients will be expected to carry heterogeneous sets of mutations/epimutations and these differences may account for the highly variable disease manifestation across patients. In some patients, the set of causal mutations/modifications needed for disease manifestation may all be inherited from parents, while in others they may represent inherited plus somatic *de novo* events that may include DNA sequence changes and/or epimutations (**Figure 3**). Here, the *de novo* events may or may not be needed to raise the liability and eventually to cross the threshold of liability for the manifestation of the disease (84, 85). Naturally, the acquired *de novo* neuronal mutations and epimutations will not be transmitted to the next generation even though any predisposition for such mutation(s), if present, may follow a

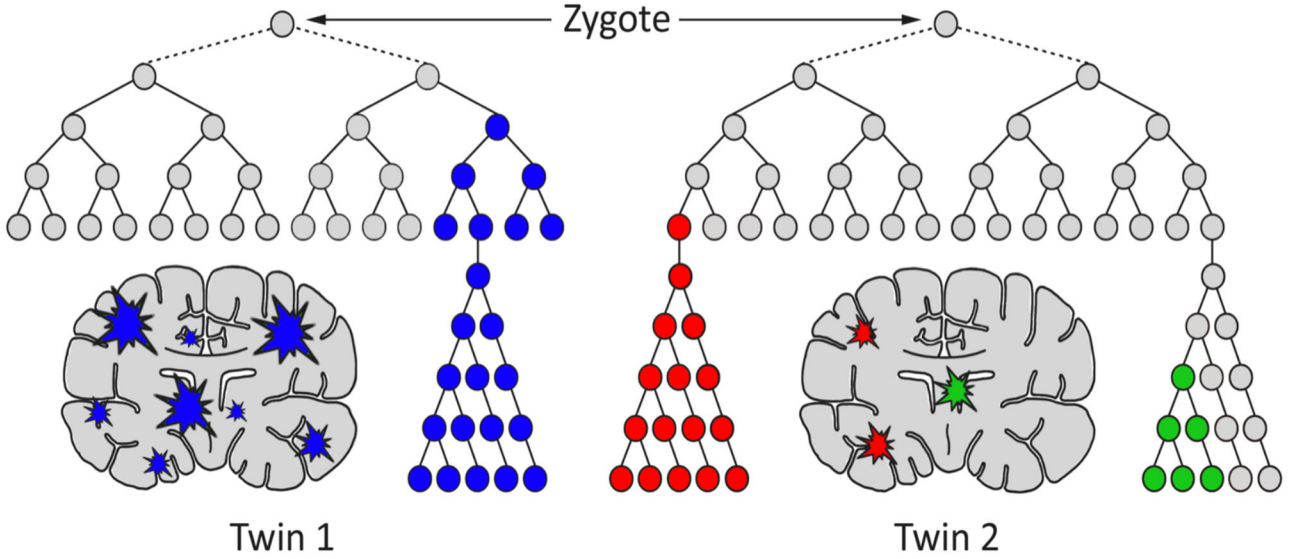


FIGURE 1 | A hypothetical model for the generation of genetic differences between brains of monozygotic twins by postzygotic mutations (PZM). The timing, genes affected, clonality, and brain region(s) affected may cause the twins to develop discordance for neurodevelopmental diseases, including schizophrenia.

familial transmission. The inclusion of such postzygotic events adds a novel and acquired somatic genetic/epigenetic change during the life of the individual to the causation of the disease.

The expanded-threshold-liability model for schizophrenia that includes postzygotic mutations and epimutations (PZMs) is compatible with a number of unusual features of this disease. These include high discordance (∼50%) of the disease among monozygotic twins, adult onset of the disease, high heterogeneity, and a spectrum of manifestations including overlapping endophenotypes involving different diagnostic entities. For the first time, this model also provides the most logical biological explanation for a comparable risk of transmission of schizophrenia by the members of a monozygotic twin pair discordant for the disease (MZD) (86, 87). The model argues that the affected and unaffected members of the MZD pair would inherit a comparable level of genetic predisposition and pass it on to their respective offspring. This level of inherited liability is expected to be below the disease threshold. However, the addition of *de novo* somatic PZMs (by chance alone) in the disease twin will lead to the development of disease (**Figure 4**) while maintaining a comparable risk of transmission to the next generation by both members of the MZD twin pair. Here, the risk of transmission of the disease by the ill twin will not be any higher than the risk of transmission by the well (unaffected) twin. Also, this expanded-threshold-liability model being presented for schizophrenia will be applicable to most neurological disorders with high discordance in monozygotic twins. It will be particularly relevant in disorders with neurodegeneration and neurodevelopmental manifestation. Although logical, the expanded model remains a theoretical concept and needs to be tested and established experimentally.

TESTING THE EXPANDED-THRESHOLD-LIABILITY MODEL FOR SCHIZOPHRENIA

Testing the proposed expanded-threshold-liability model in schizophrenia that incorporates *de novo* neuronal postzygotic mutations and epimutations, acquired during the life of an individual will be challenging for two reasons. First, the somatic *de novo* events that add to the threshold for the disease will not be transmitted to the next generation; they are acquired by an individual during his/her lifetime and are eliminated from the population with their death. Second, the somatic nature of *de novo* mutations will generate complex mosaicism, making every brain genetically heterogeneous, different and unique (41). This cellular heterogeneity will present a special challenge in any attempt of characterization as it will require assessment of cell-specific, rather than brain-specific, features. To this end, a variety of evolving methods and experimental designs including single cell genomic and multiomics (61, 90) may offer novel strategies and resolution to this complexity. Indeed, it will be challenging to identify and obtain desirable cells for such studies. Taking just any cell from any individual or brain will not be satisfactory. There will be a need to compare the neuronal zygotic and postzygotic genome that is not always practical or ethically acceptable. Although the use of multiple tissues is a practical way to identify postzygotic DNA sequence changes, others including epigenetic changes and transpositions may present problems as they are expected to be tissue specific. The challenges in ascertaining the extent and nature of acquired liability will also relate to the timing of mutation, the size of mutant clones, the accessibility of the mutant cells, the cell types affected, and the positional localization of the mutant cells. Additionally, it will be challenging to

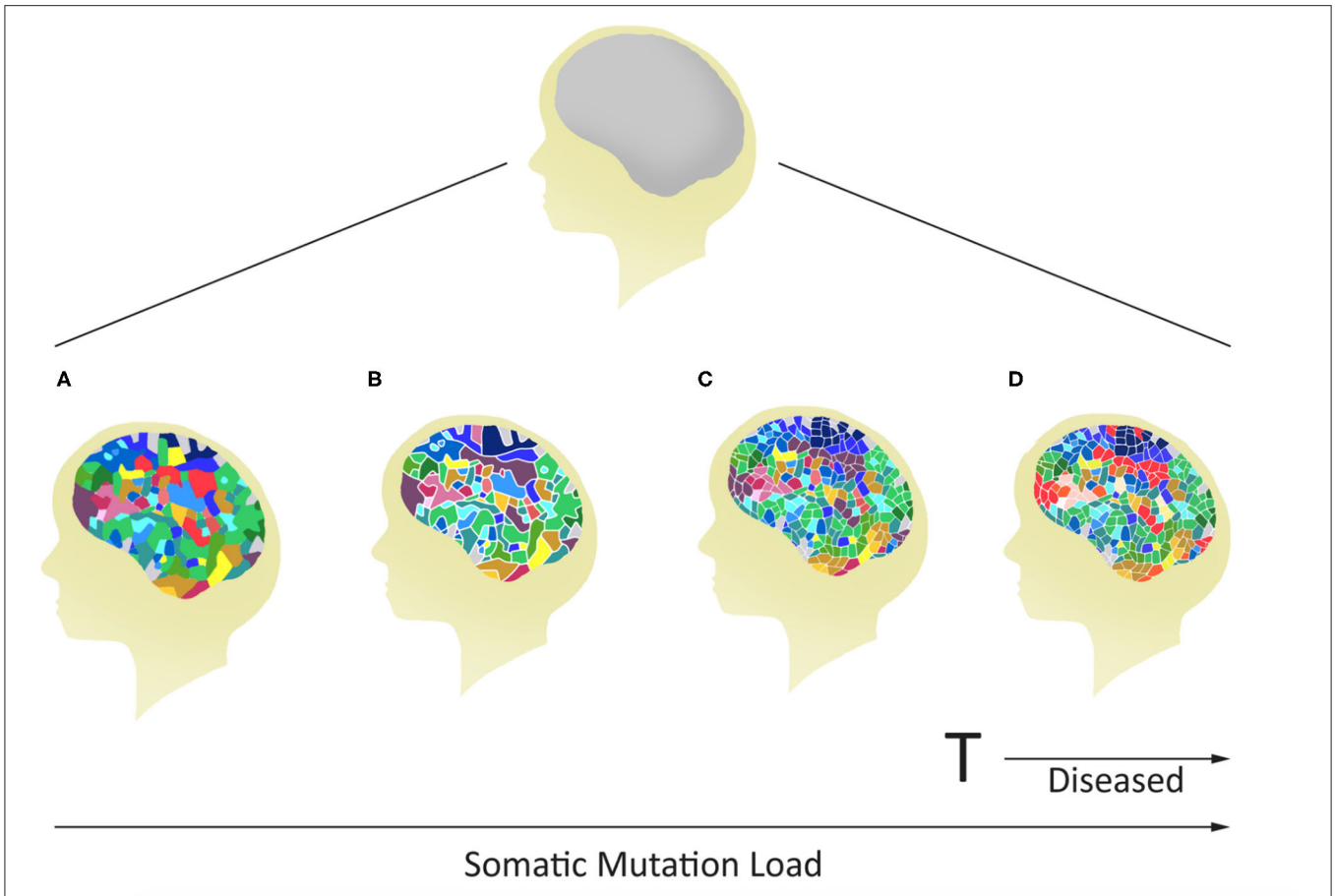


FIGURE 2 | Potential for variable mosaicism in an individual brain based on differences in postzygotic somatic *de novo* mutations during development. The outcome will make every brain unique, some below **(A–C)**, and others above **(D)** a genomic threshold for the development of a mental disease.

distinguish disease-relevant mutations from background *de novo* variants necessitating interpretation using a tested statistical framework (91). Under these circumstances, there is an imperfect alternative that may help assess the feasibility of the expanded-threshold-liability model in schizophrenia. Here, monozygotic twin pairs discordant for schizophrenia (MZD) may be used to represent an appropriate resource for testing the expanded-threshold-liability model for schizophrenia (88, 89). Recall that members of the MZD pair begin life as a single zygote but develop independently as two different individuals (somas). They share common familial predispositions and acquire independent PZMs, making the two twins distinct (**Figure 1**). Here, the most logical experiment involving a direct assessment of brain regions potentially carrying different somatic mutations/epimutations in the two brains will be difficult for a variety of reasons including accessibility. However, any PZM that occurs very "early" during embryonic development may be expected to be maintained in neuronal as well as (some) non-neuronal cells during ontogeny. As such, some of the easily accessible non-neuronal cells (blood, cheek swabs, etc.) may serve as a proxy for occurrence of "early" somatic mutations that may also be present in neuronal cells. Indeed, DNA from the blood of MZD twins has been shown to be not 100% identical (74, 89, 92). They carry a variety of sequence differences, epigenetic and genetic, that must result from postzygotic events. Here, any use of blood DNA from MZD twins could overcome challenges faced by the inaccessibility of live neuronal cells for evaluation, even though such results will be limited and may miss out on "late" mutations that may have occurred after the differentiation of a neuronal lineage. Also, the genome of the unaffected twin may be used to represent the zygotic genome that started the two members of the MZD twin toward identification of any/all schizophrenia specific postzygotic changes in the disease twin. This is by no means a perfect experiment. However, the results will have the potential to implicate involvement of somatic mutations in the genes that are known to be involved in the disorder (69), particularly those that occurred "early" during the independent development of the two twins.

Interestingly, the complete DNA sequence and genome-wide DNA methylation (88, 89) data on the blood DNA from two MZ twin pairs discordant for schizophrenia and their parents available in the literature allows us to assess any involvement of PMZs in this disease. The analysis of these data has allowed an assessment of a large number of disease associated mutations

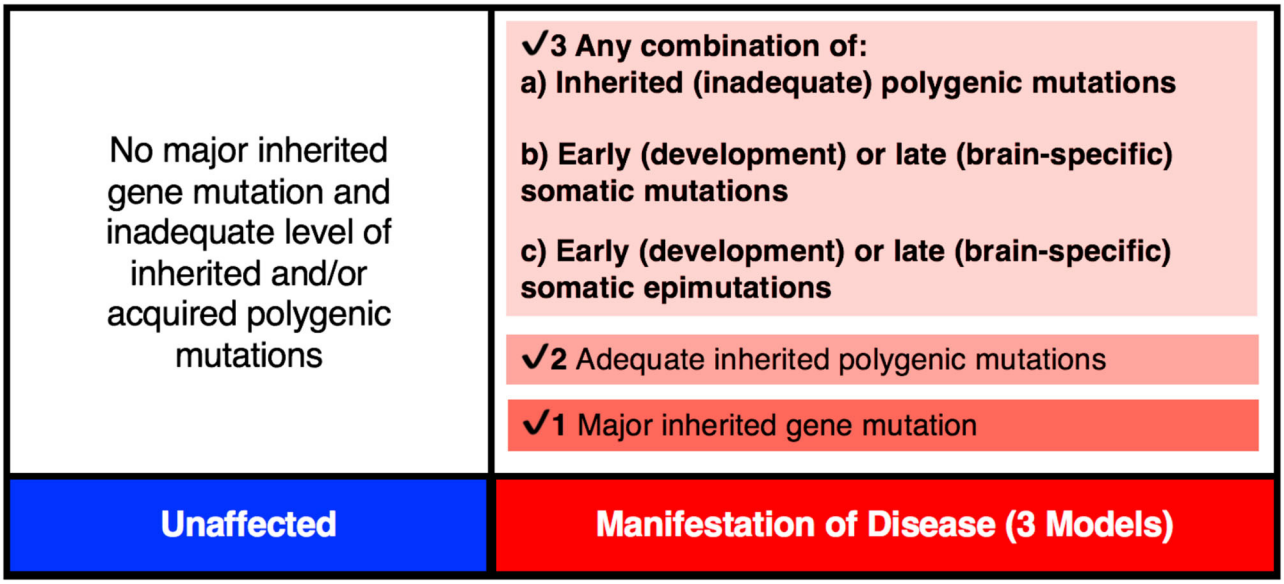


FIGURE 3 | The genetic threshold required for the development of a neurodevelopmental disease may be met by (1) A major gene insult; (2) An adequate level of polygenic mutations; or (3) A combination of inherited (a) plus acquired somatic mutations (b) and/or somatic epimutations (c).

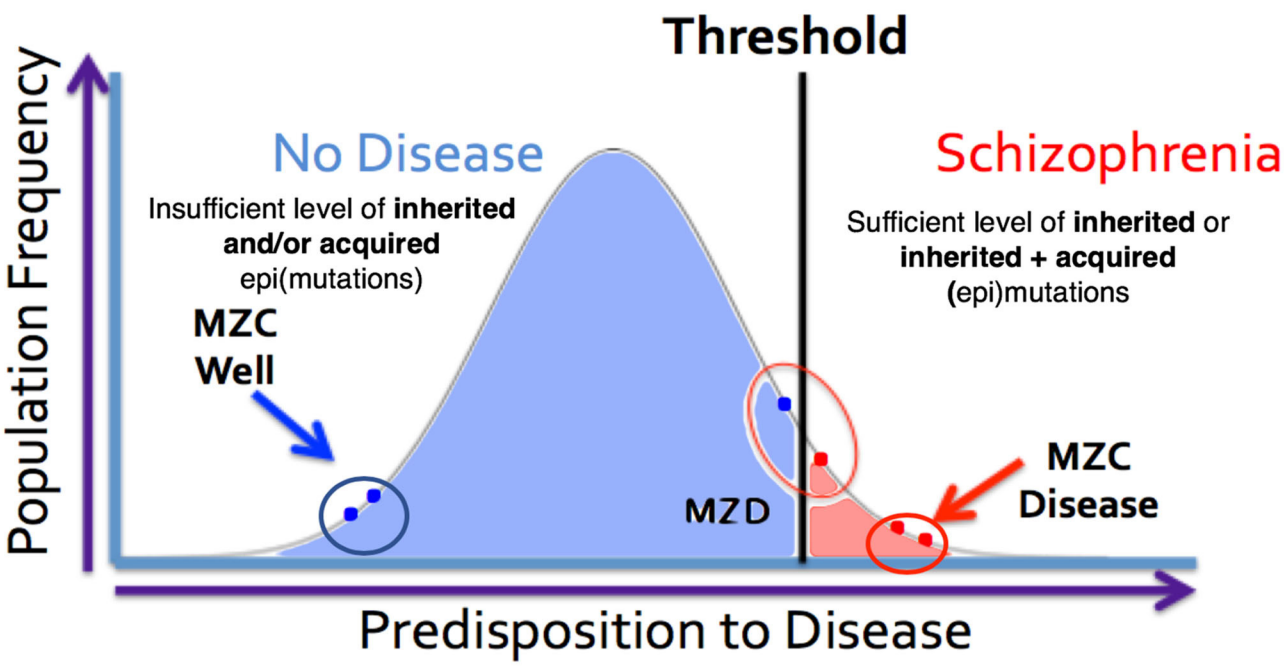


FIGURE 4 | MZ twins as a model for assessing the genetic threshold liability hypothesis. Representation of three sets of twins [MZC well (monozygotic twins both unaffected), MZC disease (monozygotic twins both diseased) and MZD (monozygotic twins discordant)] within the threshold liability hypothesis (unaffected = blue, disease affected = red). In this model, the affected twin in the MZD pair must acquire somatic mutation/epimutation to cross the threshold and develop the disease. Genomic and epigenomic assessment of these exceptional twin pairs will allow for identification of pair-specific postzygotic somatic event(s) [Adapted from Castellani et al. (88, 89)].

(http://www.szdb.org) reported in the blood DNA. Specifically, it has allowed assessment of postzygotic changes in the 145 most common schizophrenia associated gene variants (69) in the disease twin that is not present in his/her unaffected counterpart. The family specific genome sequence results (89) show that every member of the two families carried a subset of

common schizophrenia-associated gene variants (69). Although the majority of the schizophrenia-associated gene mutations present in the twins are shared and present in one or the other parent (inherited), the remainder are not seen in either parent and are unique to only one member of the MZD twin pair (89). The latter are most compatible with their postzygotic origin. More important, they could have not originated in any parental gamete that produced the zygote. The results also show that although both members of the MDZ twin pairs carry some disease-related mutations, the affected members have acquired additional disease-associated mutations in glutamate and dopamine pathway genes in both patients [see details in (89)]. It is argued that such mutations will have the potential to help cross the disease liability threshold and develop schizophrenia in the diseased twin only (**Figures 3**, **4**). It is important to suggest that any dataset that is based on blood DNA is limited to known schizophrenia-associated *de novo* events that may have occurred "early" in development. As such, they may not include *de novo* mutations that may have occurred "late" in development, not apparent in the blood and would be/are confined to the brain.

It is important to note that not all mutations that may contribute to schizophrenia involve sequence changes. Some other forms may involve epigenetic changes including DNA methylation. The details of genome-wide methylation results on the same two MZD pairs that were studied for genome sequence difference reported in Castellani et al. also show methylation aberrations in the blood DNA of the patient as compared to the well member of the twin pair (88). The results show that *de novo* events involving genome sequence and epigenetic DNA methylation changes may be independent and add to the disease liability in an expanded-liability-threshold model of schizophrenia. To the best of our understanding such results provide among the most comprehensive account of all/most gene mutations and DNA methylation changes that may have led to the development of schizophrenia in the two affected members of the two unrelated MZD twins. They may represent a near complete list of disease-causing mutation in individual patients for the first time. It has been made possible by assessment of the monozygotic twins discordant for schizophrenia, where the genotype and epigenotype (methylation specificity) of the well-twin is used to serve as the perfect-matched control. It allows a reliable assessment of inherited as well as postzygotic somatic changes during the independent development of the members of the MZD twin pair that started life as a single zygote. Finally, it is critical to note that these results are based on only two patients. There is a need to replicate such results on a larger sample size. We note that it will be challenging to find and study such patients belonging to well-characterized MZD pairs. Indeed access and availability of brain DNA from such MZD twins for such studies will be most valuable. To this end, we recognize the challenge and our inability to acquire a more comprehensive result that includes all genetic and epigenetic variants, inherited, acquired, and present in different regions of the disease brain. Such results however, will be needed to fully define the nature of genetic predisposition in individuals, including the unaffected member of the MZD pair. Despite such concerns, the limited results discussed offer potential involvement of PZMs (89) that includes epimutations (88) in the development of schizophrenia.

It is important to point out that the PZMs implicated in schizophrenia affect a relatively large number of genes, and that not all patients will require postzygotic changes during development in order to reach the disease threshold. In some cases, all of the changes necessary for the manifestation of the disorder will be acquired via familial transmission. This may be the case in monozygotic twin pairs that are concordant (MZC-disease) for the disease (**Figure 4**). The proposed expanded model is testable using MZD twin pairs that have inherited some, but not sufficient, genetic liability to reach the disease threshold. Unlike the unaffected member of this MZD pair, the affected MZD patient is shown to have acquired additional PZMs leading to a threshold necessary for disease manifestation (**Figures 3**, **4**). The addition of *de novo* somatic mutations and epimutations during early or late development (and present in the brain) is a timely addition to the revised threshold-liability model for schizophrenia. This expansion recognizes that almost all neurodevelopmental disorders are multifactorial and have a heterogeneous causation. These may include any of the three options: major familial mutation(s), adequate inherited polygenic mutations/epimutations, or a combination of inherited and acquired *de novo* (early and/or late) mutations and epimutations that help cross the liability threshold for the development of schizophrenia. This expanded model should be applicable to most disorders with complex genetic and epigenetic etiology and involvement of postzygotic changes.

One of the major challenges in neurologic disease research is the accessibility of appropriate target biological sample and use of perfectly matched control. Such samples are particularly critical for characterizing neurological disorders, including schizophrenia. These disorders involve a large number of genes, where each gene itself contributes relatively small effects. Most of the genes involved are expected to be polymorphic in the population, and many of the genes may undergo postzygotic genetic and/or epigenetic changes, particularly in the brain. This extensive polygenicity and inherited and non-inherited heterogeneity makes the investigation of the etiology of neurological disorders one of the next great challenges in biomedical science. Of special concern is the complex genetic and epigenetic somatic mosaicism being increasingly reported in the brain. Although logical and timely, research on *de novo* mutations and epimutations will demand novel approaches and high-resolution technologies, clever experiments, exceptional patients, precious and ethically sensitive biological samples, ample time and resources. Additionally, many of the specific PZMs reported to date have not yet been replicated, cataloged and curated. Here, the proposed Brain Somatic Mosaicism Network (93) has the potential to open novel avenues that have remained unexplored. This initiative will include refinement of technologies (94) that will permit characterization of every neuronal type (95) and its connectivity with other neurons in the mammalian brain (96). The anticipated results will be revolutionary and provide a long-awaited breakthrough for the field of precision medicine in neurodevelopmental disorders, and in particular, schizophrenia. Among the

most immediate application of precision corrections may include *in vivo* neuroepigenome editing for treating brain pathology (97).

CONCLUDING REMARKS

The mammalian brain is a dynamic organ with a high degree of mosaicism likely caused by postzygotic genetic and epigenetic alterations that may contribute to most multifactorial and complex neurological disorders. Traditionally, researchers have used an established threshold-liability model that incorporated the sum total of all inherited mutations in combination with environmental factors to define a liability scale sufficient for the development of disease (**Figure 4**). As it stands, this model was postulated before there was any realization of any role for postzygotic genomic/epigenomic changes and mosaicism. The timely revision of this model included in this discussion incorporates two observations. First, postzygotic changes are rather widespread and make every brain a unique genetic mosaic (**Figures 1**, **2**). Second, depending on the number and type of genes affected, the postzygotic changes may contribute to the liability scale toward the threshold of the development of a disorder like schizophrenia (**Figure 3**). The occurrence of differential postzygotic changes (both DNA methylation and DNA sequence) has been demonstrated in the MZD twin blood DNA (88, 89). It may serve as a proxy for their differential presence in the brain. The observed PZMs seen in the blood of MZD pairs are expected to have occurred "early" during development and differentiation but this must be shown empirically.

The results discussed here are compatible with the involvement of postzygotic somatic changes in schizophrenia. A comprehensive assessment of this phenomenon will only be possible if all early, late, and ongoing postzygotic alterations present in the brain are identified. Further studies will require the assessment of postmortem or surgery-derived samples of brain regions relevant to the disorder. Some of these may become feasible given the need for relatively small number of cells needed for this assessment as future experimental work must use more refined methods including single-cell multiomics at the neuronal level. An informative experiment may involve single-cell neurons from monozygotic twins discordant for the disease. Also, there may be a need to add the genomic location, cells representing relevant brain region(s), and the genetic background of the individual to this the "expanded-threshold-liability model." Finally, the revised model explains the extensive variability in brain phenotypes in the general population and the high discordance of monozygotic twins (MZD) including the fact that both members of the MZD pair have comparable risk of passing on the predisposition to their progeny. Experimental validation, although logical, timely and promising, continues to be challenging. The results, however, will bring about a long-awaited revolution in the understanding of neurological disorders, and their management, prevention and treatment that may include *in vivo* neuroepigenome editing (97). There is every reason to envision that the outcomes of the expanded research direction on schizophrenia will be comparable and even surpass the revolution of precision medicine being realized in the treatment of different forms of cancers. There are major challenges with neurological disorders, but breakthroughs are possible with the right ideas.

AUTHOR CONTRIBUTIONS

SS and KH wrote the first draft of the manuscript. CC, SS, and KH contributed equally to the formatting, editing, revision, figures, and intellectual content of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by grants from the Canadian Institutes of Health Research (CIHR), Ontario Mental Health Foundation, Schizophrenia Society of Ontario, Natural Sciences and Engineering Research Council (NSERC) of Canada to SS, and Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery (Grant No. R3511A12 to KH). CC was supported by a CIHR Doctoral Research Award and CIHR Postdoctoral Fellowship.

ACKNOWLEDGMENTS

The authors acknowledge the assistance of Xiao Min Pearl Hung in the construction of **Figures 1**, **2**.

REFERENCES

1. Lupski JR. Genome mosaicism–one human, multiple genomes. *Science.* (2013) 341:358–9. doi: 10.1126/science.1239503
2. Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, et al. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science.* (2015) 348:880–6. doi: 10.1126/science.aaa6806
3. Jung S-H, Yim S-H, Oh HJ, Park JE, Kim MJ, Kim GA, et al. *De novo* copy number variations in cloned dogs from the same nuclear donor. *BMC Genomics.* (2013) 14:863. doi: 10.1186/1471-2164-14-863
4. Sugiura M, Kasama Y, Araki R, Hoki Y, Sunayama M, Uda M, et al. Induced pluripotent stem cell generation-associated point mutations arise during the initial stages of the conversion of these cells. *Stem Cell Rep.* (2014) 2:52–63. doi: 10.1016/j.stemcr.2013.11.006
5. Bae T, Tomasini L, Mariani J, Zhou B, Roychowdhury T, Franjic D, et al. Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science.* (2017) 555:1–10. doi: 10.1126/science.aan8690
6. Leija-Salazar M, Piette CL, Proukakis C. Somatic mutations in neurodegeneration. *Neuropathol Appl Neurobiol.* (2018) 44:267–85. doi: 10.1111/nan.12465
7. Dou Y, Gold HD, Luquette LJ, Park PJ. Detecting somatic mutations in normal cells. *Trends Genet.* (2018) 34:545–57. doi: 10.1016/j.tig.2018.04.003
8. Milholland B, Dong X, Zhang L, Hao X, Suh Y, Vijg J. Differences between germline and somatic mutation rates in humans and mice. *Nat Commun.* (2017) 8:15183. doi: 10.1038/ncomms15183
9. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Peter J. Clock-like mutational processes in human somatic cells. *Nat Publ Gr.* (2015) 47:1402–7. doi: 10.1038/ng.3441
10. Liu P, Yuan B, Carvalho CMB, Wuster A, Walter K, Zhang L, et al. An organismal CNV mutator phenotype restricted to early human development. *Cell.* (2017) 168:830–42.e7. doi: 10.1016/j.cell.2017.01.037
11. Forsberg LA, Gisselsson D, Dumanski JP. Mosaicism in health and disease — clones picking up speed. *Nat Rev Genet.* (2016) 18:128–42. doi: 10.1038/nrg.2016.145

12. De S. Somatic mosaicism in healthy human tissues. *Trends Genet.* (2011) 27:217–23. doi: 10.1016/j.tig.2011.03.002
13. Cotterman CW. Somatic mosaicism for antigen A2. *Acta Genet Stat Med.* (1956) 6:520–21. doi: 10.1159/000150891
14. Biesecker LG, Spinner NB. A genomic view of mosaicism and human disease. *Nat Rev Genet.* (2013) 14:307–20. doi: 10.1038/nrg3424
15. Nishioka M, Bundo M, Iwamoto K, Kato T. Somatic mutations in the human brain: implications for psychiatric research. *Mol Psychiatry.* (2018) 24:839–56. doi: 10.1038/s41380-018-0129-y
16. Hall JG. Review and hypotheses: somatic mosaicism: observations related to clinical genetics. *Am J Hum Genet.* (1988) 43:355–63.
17. Lodato MA, Rodin RE, Bohrson CL, Coulter ME, Barton AR, Kwon M, et al. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science.* (2017) 4426:1–8. doi: 10.1101/221960
18. Lodato MA, Woodworth MB, Lee S, Evrony GD, Mehta BK, Karger A, et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science.* (2015) 350:94–97. doi: 10.1126/science.aab1785
19. Verheijen BM, Vermulst M, van Leeuwen FW. Somatic mutations in neurons during aging and neurodegeneration. *Acta Neuropathol.* (2018) 135:811–26. doi: 10.1007/s00401-018-1850-y
20. Hodgkinson A, Eyre-Walker A. Variation in the mutation rate across mammalian genomes. *Nat Rev Genet.* (2011) 12:756–66. doi: 10.1038/nrg3098
21. Livnat A. Interaction-based evolution: how natural selection and nonrandom mutation work together. *Biol Direct.* (2013) 8:24. doi: 10.1186/1745-6150-8-24
22. Papavasiliou FN, Schatz DG. Somatic hypermutation of immunoglobulin and genes: merging mechanisms for genetic diversity. *Cell.* (2002) 109:S35–44. doi: 10.1016/S0092-8674(02)00706-7
23. Tsunoyama K, Bellgard MI, Gojobori T. Intragenic variation of synonymous substitution rates is caused by nonrandom mutations at methylated CpG. *J Mol Evol.* (2001) 53:456–64. doi: 10.1007/s002390010235
24. Francioli LC, Polak PP, Koren A, Menelaou A, Chun S, Renkens I, et al. Genome-wide patterns and properties of *de novo* mutations in humans. *Nat Genet.* (2015) 47:822–6. doi: 10.1038/ng.3292
25. Briney BS, Crowe JE. Secondary mechanisms of diversification in the human antibody repertoire. *Front Immunol.* (2013) 4:42. doi: 10.3389/fimmu.2013.00042
26. Konstantinidou C, Taraviras S, Pachnis V. Geminin prevents DNA damage in vagal neural crest cells to ensure normal enteric neurogenesis. *BMC Biol.* (2016) 14:94. doi: 10.1186/s12915-016-0314-x
27. Sabel M, Osborn M, Aronson M, Kratz CP, Dunn GP, de Borja R, et al. Comprehensive analysis of hypermutation in human cancer. *Cell.* (2017) 171:1042–56.e10. doi: 10.1016/j.cell.2017.09.048
28. Acuna-Hidalgo R, Veltman JA, Hoischen A. New insights into the generation and role of *de novo* mutations in health and disease. *Genome Biol.* (2016) 17:241. doi: 10.1186/s13059-016-1110-1
29. Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. *Nat Rev Genet.* (2016) 17:175–88. doi: 10.1038/nrg.2015.16
30. Macaulay IC, Ponting CP, Voet T. Single-cell multiomics: multiple measurements from single cells. *Trends Genet.* (2017) 33:155–68. doi: 10.1016/j.tig.2016.12.003
31. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature.* (2013) 500:415–21. doi: 10.1038/nature12477
32. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell.* (2012) 149:979–93. doi: 10.1016/j.cell.2012.04.024
33. Chin L, Hahn WC, Getz G, Meyerson M. Making sense of cancer genomic data. *Genes Dev.* (2011) 25:534–55. doi: 10.1101/gad.2017311
34. Bolli N, Avet-Loiseau H, Wedge DC, Van Loo P, Alexandrov LB, Martincorena I, et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat Commun.* (2014) 5:2997. doi: 10.1038/ncomms3997
35. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature.* (2010) 464:999–1005. doi: 10.1038/nature08989
36. Loeb LA, Bielas JH, Beckman RA, Bodmer IW. Cancers exhibit a mutator phenotype: clinical implications. *Cancer Res.* (2008) 68:3551–7. doi: 10.1158/0008-5472.CAN-07-5835
37. Drake JW, Bebenek A, Kissling GE, Peddada S. Clusters of mutations from transient hypermutability. *Proc Natl Acad Sci USA.* (2005) 102:12849–54. doi: 10.1073/pnas.0503009102
38. Chen JM, Férec C, Cooper DN. Transient hypermutability, chromothripsis and replication-based mechanisms in the generation of concurrent clustered mutations. *Mutat Res Rev Mutat Res.* (2012) 750:52–59. doi: 10.1016/j.mrrev.2011.10.002
39. Wang J, Gonzalez KD, Scaringe WA, Tsai K, Liu N, Gu D, et al. Evidence for mutation showers. *Proc Natl Acad Sci USA.* (2007) 104:8403–8. doi: 10.1073/pnas.0610902104
40. Bailey P, Chang DK, Nones K, Johns AL, Patch A-M, Gingras M-C, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature.* (2016) 531:47–52. doi: 10.1038/nature16965
41. Rohrback S, Siddoway B, Liu CS, Chun J. Genomic mosaicism in the developing and adult brain. *Dev Neurobiol.* (2018) 78:1026–48. doi: 10.1002/dneu.22626
42. Eriksson PS, Perfilieva E, Björk-Eriksson T, Alborn AM, Nordborg C, Peterson DA, et al. Neurogenesis in the adult human hippocampus. *Nat Med.* (1998) 4:1313–17. doi: 10.1038/3305
43. Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, et al. L1 retrotransposition in human neural progenitor cells. *Nature.* (2009) 460:1127–31. doi: 10.1038/nature08248
44. Muotri AR, Chu VT, Marchetto MCN, Deng W, Moran JV, Gage FH. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature.* (2005) 435:903–10. doi: 10.1038/nature03663
45. Doyle GA, Crist RC, Karatas ET, Hammond MJ, Ewing AD, Ferraro TN, et al. Analysis of LINE-1 elements in DNA from postmortem brains of individuals with schizophrenia. *Neuropsychopharmacology.* (2017) 42:2602–11. doi: 10.1038/npp.2017.115
46. MacIa A, Widmann TJ, Heras SR, Ayllon V, Sanchez L, Benkaddour-Boumzaouad M, et al. Engineered LINE-1 retrotransposition in nondividing human neurons. *Genome Res.* (2017) 27:335–48. doi: 10.1101/gr.206805.116
47. van den Hurk JAJM, Meij IC, del Carmen Seleme M, Kano H, Nikopoulos K, Hoefsloot LH, et al. L1 retrotransposition can occur early in human embryonic development. *Hum Mol Genet.* (2007) 16:1587–92. doi: 10.1093/hmg/ddm108
48. Richardson SR, Gerdes P, Gerhardt DJ, Sanchez-Luque FJ, Bodea GO, Muñoz-Lopez M, et al. Heritable L1 retrotransposition in the mouse primordial germline and early embryo. *Genome Res.* (2017) 27:1395–405. doi: 10.1101/gr.219022.116
49. Jacob-Hirsch J, Eyal E, Knisbacher BA, Roth J, Cesarkas K, Dor C, et al. Whole-genome sequencing reveals principles of brain retrotransposition in neurodevelopmental disorders. *Cell Res.* (2018) 28:187–203. doi: 10.1038/cr.2018.8
50. Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, et al. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature.* (2011) 479:534–7. doi: 10.1038/nature10531
51. Singer T, Mcconnell MJ, Marchetto MCN, Coufal NG, Gage FH. LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? *Trends Neurosci.* (2011) 33:345–54. doi: 10.1016/j.tins.2010.04.001
52. Harbom LJ, Chronister WD, McConnell MJ. Single neuron transcriptome analysis can reveal more than cell type classification. *BioEssays.* (2016) 38:157–61. doi: 10.1002/bies.201500097
53. Hiesinger PR, Hassan BA. The evolution of variability and robustness in neural development. *Trends Neurosci.* (2018) 41:577–86. doi: 10.1016/j.tins.2018.05.007
54. Rehen SK. Constitutional aneuploidy in the normal human brain. *J Neurosci.* (2005) 25:2176–80. doi: 10.1523/JNEUROSCI.4560-04.2005
55. Evrony GD, Cai X, Lee E, Hills LB, Elhosary PC, Lehmann HS, et al. Single-neuron sequencing analysis of l1 retrotransposition and somatic mutation in the human brain. *Cell.* (2012) 151:483–96. doi: 10.1016/j.cell.2012.09.035
56. Curran T. Deconstructing the mosaic brain. *The Scientist.* (2011) Available online at: https://www.the-scientist.com/thought-experiment/deconstructing-the-mosaic-brain-42134
57. Frank SA. Somatic mosaicism and disease. *Curr Biol.* (2014) 24:R577–81. doi: 10.1016/j.cub.2014.05.021
58. Nicolas G, Veltman JA. The role of *de novo* mutations in adult-onset neurodegenerative disorders. *Acta Neuropathol.* (2019) 137:183–207. doi: 10.1007/s00401-018-1939-3

59. Veltman JA, Brunner HG. De novo mutations in human genetic disease. *Nat Rev Genet.* (2012) 13:565–75. doi: 10.1038/nrg3241
60. Pujana MA, Han J-DJ, Starita LM, Stevens KN, Tewari M, Ahn JS, et al. Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nat Genet.* (2007) 39:1338–49. doi: 10.1038/ng.2007.2
61. Wang W, Corominas R, Lin GN. *De novo* mutations from whole exome sequencing in neurodevelopmental and psychiatric disorders: from discovery to application. *Front Genet.* (2019) 10:258. doi: 10.3389/fgene.2019.00258
62. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of *de novo* coding mutations to autism spectrum disorder. *Nature.* (2014) 515:216–21. doi: 10.1038/nature13908
63. Singh SM, Castellani CA, O'Reilly RL. Copy number variation showers in schizophrenia: an emerging hypothesis. *Mol Psychiatry.* (2009) 14:356–8. doi: 10.1038/mp.2008.149
64. Rees E, Kirov G, O'Donovan MC, Owen MJ. *De novo* mutation in schizophrenia. *Schizophr Bull.* (2012) 38:377–81. doi: 10.1093/schbul/sbs047
65. McCarthy SE, Gillis J, Kramer M, Lihm J, Yoon S, Berstein Y, et al. *De novo* mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Mol Psychiatry.* (2014) 19:652–8. doi: 10.1038/mp.2014.29
66. Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of *de novo* copy number mutations with sporadic schizophrenia. *Nat Genet.* (2008) 40:880–5. doi: 10.1038/ng.162
67. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait. *Arch Gen Psychiatry.* (2003) 60:1187–92. doi: 10.1001/archpsyc.60.12.1187
68. Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnusdottir B, Morgen K, Arnarsdottir S, et al. CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature.* (2014) 505:361–6. doi: 10.1038/nature12818
69. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet.* (2018) 50:381–9. doi: 10.1038/s41588-018-0059-2
70. Kelly J, Murray RM. What risk factors tell us about the causes of schizophrenia and related psychoses. *Curr Psychiatry Rep.* (2000) 2:378–85. doi: 10.1007/s11920-000-0019-1
71. McGue M, Gottesman II, Rao DC. The transmission of schizophrenia under a multifactorial threshold model. *Am J Hum Genet.* (1983) 35:1161–78.
72. Fullard JF, Charney AW, Voloudakis G, Uzilov AV, Haroutunian V, Roussos P. Assessment of somatic single-nucleotide variation in brain tissue of cases with schizophrenia. *Transl Psychiatry.* (2019) 9:21. doi: 10.1038/s41398-018-0342-0
73. Sakai M, Watanabe Y, Someya T, Araki K, Shibuya M, Niizato K, et al. Assessment of copy number variations in the brain genome of schizophrenia patients. *Mol Cytogenet.* (2015) 8:46. doi: 10.1186/s13039-015-0144-5
74. Maiti S, Kumar KHBG, Castellani CA, O'Reilly R, Singh SM. Ontogenetic *de novo* copy number variations (CNVs) as a source of genetic individuality: studies on two families with MZD twins for schizophrenia. *PLoS ONE.* (2011) 6:e17125. doi: 10.1371/journal.pone.0017125
75. Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet.* (2016) 49:27–35. doi: 10.1038/ng1017-1558d
76. Glessner JT, Reilly MP, Kim CE, Takahashi N, Albano A, Hou C, et al. Strong synaptic transmission impact by copy number variations in schizophrenia. *Proc Natl Acad Sci USA.* (2010) 107:10584–9. doi: 10.1073/pnas.1000274107
77. Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kahler AK, Akterin S, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet.* (2013) 45:1150–59. doi: 10.1038/ng.2742
78. Yurov YB, Iourov IY, Vorsanova SG, Demidova IA, Kravetz VS, Beresheva AK, et al. The schizophrenia brain exhibits low-level aneuploidy involving chromosome 1. *Schizophr Res.* (2008) 98:139–47. doi: 10.1016/j.schres.2007.07.035
79. Jiang YH, Bressler J, Beaudet AL. Epigenetics and human disease. *Annu Rev Genomics Hum Genet.* (2004) 5:479–510. doi: 10.1146/annurev.genom.5.061903.180014
80. Hurst TP, Magiorkinis G. Epigenetic control of human endogenous retrovirus expression: focus on regulation of long-terminal repeats (LTRs). *Viruses.* (2017) 9:1–13. doi: 10.3390/v9060130
81. Deb-Rinker P, Klempan T, O'Reilly RL, Torrey EF, Singh SM. Molecular characterization of a MSRV-like sequence identified by RDA from monozygotic twin pairs discordant for schizophrenia. *Genomics.* (1999) 61:133–44. doi: 10.1006/geno.1999.5946
82. Bundo M, Toyoshima M, Okada Y, Akamatsu W, Ueda J, Nemoto-Miyauchi T, et al. Increased L1 retrotransposition in the neuronal genome in schizophrenia. *Neuron.* (2014) 81:306–13. doi: 10.1016/j.neuron.2013.10.053
83. Sekar A, Bialas AR, de Rivera H, Davis A, Hammond TR, Kamitaki N, et al. Schizophrenia risk from complex variation of complement component 4. *Nature.* (2016) 530:177–83. doi: 10.1038/nature16549
84. Haegert DG. Analysis of the threshold liability model provides new understanding of causation in autoimmune diseases. *Med Hypotheses.* (2004) 63:257–61. doi: 10.1016/j.mehy.2004.02.015
85. Wandstrat A, Wakeland E. The genetics of complex autoimmune diseases: non-MHC susceptibility genes. *Nat Immunol.* (2001) 2:802–9. doi: 10.1038/ni0901-802
86. Fischer M. Psychoses in the offspring of schizophrenic monozygotic twins and their normal co-twins. *Br J Psychiatry.* (1971) 118:43–52. doi: 10.1192/bjp.118.542.43
87. Gottesman II, Bertelsen A. Confirming unexpressed genotypes for schizophrenia. *Arch Gen Psychiatry.* (1989) 46:867–72. doi: 10.1001/archpsyc.1989.01810100009002
88. Castellani CAA, Melka MGG, Gui JLL, O'Reilly RLL, Singh SMM. Integration of DNA sequence and DNA methylation changes in monozygotic twin pairs discordant for schizophrenia. *Schizophr Res.* (2015) 169:433–40. doi: 10.1016/j.schres.2015.09.021
89. Castellani CA, Melka MG, Gui JL, Gallo AJ, O'Reilly RL, Singh SM. Post-zygotic genomic changes in glutamate and dopamine pathway genes may explain discordance of monozygotic twins for schizophrenia. *Clin Transl Med.* (2017) 6:43. doi: 10.1186/s40169-017-0174-1
90. D'Gama AM, Walsh CA. Somatic mosaicism and neurodevelopmental disease. *Nat Neurosci.* (2018) 21:1504–14. doi: 10.1038/s41593-018-0257-3
91. Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, et al. A framework for the interpretation of *de novo* mutation in human disease. *Nat Genet.* (2014) 46:944–50. doi: 10.1038/ng.3050
92. Tang J, Fan Y, Li H, Xiang Q, Zhang DF, Li Z, et al. Whole-genome sequencing of monozygotic twins discordant for schizophrenia indicates multiple genetic risk factors for schizophrenia. *J Genet Genomics.* (2017) 44:295–306. doi: 10.1016/j.jgg.2017.05.005
93. McConnell MJ, Moran JV, Abyzov A, Akbarian S, Bae T, Cortes-Ciriano I, et al. Intersection of diverse neuronal genomes and neuropsychiatric disease: the brain somatic mosaicism network. *Science.* (2017) 356:eaal1641. doi: 10.1126/science.aal1641
94. Macosko EZ, Basu A, Satija R, Nemesh J, Goldman M, Tirosh I, et al. HHS public access. *Cell.* (2016) 161:1202–14. doi: 10.1016/j.cell.2015.05.002
95. Shekhar K, Lapan SW, Whitney IE, Tran NM, Evan Z, Kowalczyk M, et al. Neurons by single-cell transcriptomics. *Cell.* (2017) 166:1308–23. doi: 10.1016/j.cell.2016.07.054
96. Fletcher RB, Das D, Gadye L, Street KN, Baudhuin A, Wagner A, et al. Deconstructing olfactory stem cell trajectories at single-cell resolution. *Cell Stem Cell.* (2017) 20:817–30.e8. doi: 10.1016/j.stem.2017.04.003
97. Yim YY, Teague CD, Nestler EJ. *In vivo* locus-specific editing of the neuroepigenome. *Nat Rev Neurosci.* (2020) 21:471–84. doi: 10.1038/s41583-020-0334-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychiatry

GENERAL COMMENTARY
published: 06 August 2021
doi: 10.3389/fpsyt.2021.653624

Commentary on Singh et al. (2020) Postzygotic Somatic Mutations in the Human Brain Expand the Threshold-Liability Model of Schizophrenia

Peter S. Spencer[1]* *and Glen E. Kisby*[2]

[1] *Department of Neurology, School of Medicine, and Oregon Institute of Occupational Health Sciences, Oregon Health & Science University, Portland, OR, United States,* [2] *Department of Basic Medical Sciences, College of Osteopathic Medicine of the Pacific Northwest, Western University of Health Sciences, Lebanon, OR, United States*

Keywords: schizophrenia, amyotrophic lateral sclerosis, environmental somatic mutagen, brain DNA damage, methylazoxymethanol

A Commentary on

Postzygotic Somatic Mutations in the Human Brain Expand the Threshold-Liability Model of Schizophrenia
by Singh, S. M., Castellani, C. A., and Hill, K. A. (2020). Front. Psychiatry 11:587162. doi: 10.3389/fpsyt.2020.587162

OPEN ACCESS

Edited by:
Patrick E. Williams, University of Kansas, United States

Reviewed by:
Jurjen Luykx, University Medical Center Utrecht, Netherlands

***Correspondence:**
Peter S. Spencer
spencer@ohsu.edu

Specialty section:
This article was submitted to Behavioral and Psychiatric Genetics, a section of the journal Frontiers in Psychiatry

Received: *14 January 2021*
Accepted: *14 July 2021*
Published: *06 August 2021*

Citation:
Spencer PS and Kisby GE (2021) Commentary on Singh et al. (2020) Postzygotic Somatic Mutations in the Human Brain Expand the Threshold-Liability Model of Schizophrenia. Front. Psychiatry 12:653624. doi: 10.3389/fpsyt.2021.653624

INTRODUCTION

Singh and colleagues note that the mammalian brain has a high degree of mosaicism likely caused by postzygotic genetic and epigenetic alterations that may contribute to most multifactorial and complex neurological disorders, for which the authors use schizophrenia to exemplify. They suggest that schizophrenia arises from a sufficient level of inherited, or inherited plus acquired, brain somatic mutations and/or epimutations, a model with particular relevance they suggest to disorders with neurodegeneration and neurodevelopmental manifestation. We wish to extend this model by noting that acquired somatic mutations/epimutations *in the absence of any inherited predisposition* may be sufficient to trigger a neurodegenerative disorder with neurodevelopmental manifestations and links to schizophrenia. Additionally, in accord with the authors' analysis, we note the importance of environmental mutagens as potential triggers of somatic mutation/epimutations that result not only in cancer but also neurodegenerative disease.

MOTOR SYSTEM DEGENERATION

The neurodegenerative disease of interest here is the Western Pacific Amyotrophic Lateral Sclerosis and Parkinsonism-Dementia Complex (ALS/PDC), which is known principally in residents of and migrants to and from the former disease hot spots of Guam Island (among Chamorros) and Kii Peninsula, Honshu Island (among Japanese) (1). Once thought to have a genetic origin, later a changing environmental component that explained its decline, and subsequently, with the absence of a characteristic genotype and disappearance of the disease on Guam, considered to have a prominent if not exclusive exogenous etiology. Evidence points to the traditional use of cycad seed for food and/or medicine, practices that disappeared in concert with post-WW II modernization.

On Guam, cycad seed, and the food derived therefrom, contain methylazoxymethanol (MAM) β-D-glucoside (cycasin) and β-*N*-methylamino-L-alanine (L-BMAA), an uncommon amino acid (2). These substances are both metabolized to the mutagen formaldehyde (1), which from endogenous sources and/or from exogenous exposure is linked to both brain cancer and sporadic ALS (3). Accordingly, MAM, the active form of the major cycad toxin, is a genotoxic agent with both carcinogenic and neurotoxic properties (1). There is a strongly significant correlation between the concentration of cycasin (but not L-BMAA) in cycad flour prepared Chamorro-style and the incidence of ALS and P-D among males and females on Guam (4). Additionally, MAM disrupts retinal and cerebellar development, which respectively anticipate and attend the adult onset of ALS/PDC (5). Most significantly, for present purposes, MAM is widely used experimentally to produce a rodent model of schizophrenia. In female rats, the administration of MAM on gestational day 17 disrupts brain development leading to histological, neurophysiological and behavioral deficits analogous to those of schizophrenia (6, 7). How closely the rodent MAM model reproduces the neuropathological and behavioral features of schizophrenia is unknown.

SCHIZOPHRENIA AND ALS

While rare cases have linked schizophrenia and Western Pacific ALS/PDC (8, 9), there is growing clinical, epidemiological and biological evidence of an association between ALS and psychotic illness (10), particularly schizophrenia (11). Westphal (12) observed that the paranoid and manic-depressive states of schizophrenia were associated with ALS but considered the neuropsychological and motor system disorders to be unrelated. Wechsler and Davison (13) reported that mental symptoms were due to cortical degenerative changes associated with ALS. Turner and colleagues (14) found that schizophrenia may represent a risk factor for ALS (OR 5.0). Howland (15) noted several cases in which schizophrenia occurred in ALS patients. Register-based nationwide studies show a higher occurrence of schizophrenia up to 1–5 years before and 2–5 years after ALS diagnosis (16). The coexistence of ALS and schizophrenia has been interpreted as indicative of a shared polygenic basis (11), and GWAS studies support a genetic correlation between the two conditions (11, 17). Neuropsychiatric symptoms other than schizophrenia, including obsessive-compulsive disorder, autism, and alcoholism, occur more frequently in first- or second-degree relatives of ALS patients with and without *C9orf72* expanded repeats (18, 19). Disturbances in motor neuron function have been demonstrated in schizophrenia (20–22).

MOLECULAR MECHANISMS

We have discussed elsewhere evidence that MAM experimentally induces early epigenetic changes that coincide with DNA damage and cell-cycle reactivation, evidence for which is seen in the ALS/PDC brain (23–25). MAM disrupts the cell cycle presumably by inducing DNA damage *via* methylation of guanine (i.e., *N*7 methyl and/or O^6-methyl adducts) that inhibits DNA replication during S phase (26) and disrupts neuroepithelial cells undergoing their final mitosis (27). Some of the early changes induced by MAM in somatic cells include nucleoprotein structural alterations, mitotic abnormalities, and induction of polyploidy (28) as well as retinoblastoma (Rb) gene mutations, which lead to the development of intraocular neoplasms (29, 30). Expression of the retinoblastoma gene is also altered in the prefrontal cortex of rats treated developmentally with MAM (31) and in human neuroprogenitor cells (hNPCs) 24 h after acute treatment with the genotoxin (32). L-BMAA also induces cell-cycle dysregulation in embryonic rat striatal neurons (1, 33).

DISCUSSION

Whether MAM-induced DNA damage and/or epigenetic changes are the initial event(s) that trigger the cell-cycle changes is presently unknown, but it is clear that this genotoxin induces somatic cell changes that are linked with both experimental schizophrenia and neurodegeneration in the form of Western Pacific ALS/PDC. Given the absence of any known genetic susceptibility factor for this prototypical neurodegenerative disease—one that often has subclinical evidence of developmental cerebellar and retinal dysplasia (5)—it is reasonable to propose that exposure to an environmental mutagen/epimutagen alone (notably MAM) is sufficient to trigger the disorders. Given this conclusion, we extend the model proposed by Singh and colleagues to include *environmentally acquired* somatic mutations/epimutations as sufficient to trigger a neurodegenerative disorder with neurodevelopmental manifestations and links to schizophrenia. The corollary emphasizes the need to search for early-life exposure to environmental mutagens/epimutagens in related spontaneous neurodegenerative disorders, including ALS, atypical Parkinson syndromes such as Progressive Supranuclear Palsy, and Alzheimer disease (34).

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

Research was supported in part by NIH-NIEHS grant U19ES011384 (PS and GK) and NIH 1R01ES026225 (GK). Stem cell research was supported by Illumina and Western University of Health Sciences (GK).

REFERENCES

1. Spencer PS, Palmer VS, Kisby GE. Western Pacific ALS-PDC: evidence implicating cycad genotoxins. *J Neurol Sci.* (2020) 419:117185. doi: 10.1016/j.jns.2020.117185
2. Kisby GE, Ellison M, Spencer PS. Content of the neurotoxins cycasin (methylazoxymethanol β-D-glucoside) and BMAA (β-*N*-methylamino-L-alanine) in cycad flour prepared by Guam Chamorros. *Neurology*. (1992) 42:1336–40. doi: 10.1212/WNL.42.7.1336
3. Rana I, Rieswijk L, Steinmaus C, Zhang L. Formaldehyde and brain disorders: a meta-analysis and bioinformatics approach. *Neurotox Res.* (2021) 20:320. doi: 10.1007/s12640-020-00320-y
4. Román GC. Neuroepidemiology of amyotrophic lateral sclerosis: clues to aetiology and pathogenesis. *J Neurol Neurosurg Psychiatry.* (1996) 61:131–7. doi: 10.1136/jnnp.61.2.131
5. Spencer PS. Etiology of retinal and cerebellar pathology in Western Pacific Amyotrophic Lateral Sclerosis and Parkinsonism-Dementia Complex. *Eye Brain.* (2020) 12:97–104. doi: 10.2147/EB.S260823
6. Lodge DJ, Grace AA. Gestational methylazoxymethanol acetate administration: a developmental disruption model of schizophrenia. *Behav Brain Res.* (2009) 204:306–12. doi: 10.1016/j.bbr.2009.01.031
7. Jones CA, Watson DJ, Fone KC. Animal models of schizophrenia. *Br J Pharmacol.* (2011) 164:1162–94. doi: 10.1111/j.1476-5381.2011.01386.x
8. Yase Y, Matsumoto M, Azuma K, Nakai Y, Shiraki H. Amyotrophic lateral sclerosis: association with schizophrenic symptoms and showing Alzheimer's tangles. *Arch Neurol.* (1972) 27:118–28. doi: 10.1001/archneur.1972.00490140022005
9. Shindo A, Ueda Y, Kuzuhara S, Kokubo Y. Neuropsychological study of amyotrophic lateral sclerosis and parkinsonism-dementia complex in Kii peninsula, Japan. *BMC Neurol.* (2014) 14:151. doi: 10.1186/1471-2377-14-151
10. Zucchi E, Ticozzi N, Mandrioli J. Psychiatric symptoms in Amyotrophic Lateral Sclerosis: beyond a motor neuron disorder. *Front Neurosci.* (2019) 13:175. doi: 10.3389/fnins.2019.00175
11. McLaughlin R, Schijven D, van Rheenen W, van Eijk KR, O'Brien M, Kahn RS, et al. Project MiniE GWAS Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium Genetic correlation between amyotrophic lateral sclerosis and schizophrenia. *Nat Commun.* (2017) 8:14774. doi: 10.1038/ncomms14774
12. Westphal A. Schizophrene Krankheitsprozesse nnd amyotrophische Lateralsklerose. *Arch fiir Psychlat.* (1925) 74:810. doi: 10.1007/BF01814189
13. Wechsler IS, Davison C. Amyotrophic lateral sclerosis with mental symptoms. A clinicpathologic study. *Arch Neurol Psychiatr.* (1932) 27:859–80. doi: 10.1001/archneurpsyc.1932.02230160100010
14. Turner MR, Goldacre R, Talbot K, Goldacre MJ. Psychiatric disorders prior to amyotrophic lateral sclerosis. *Ann Neurol.* (2016) 80:935–8. doi: 10.1002/ana.24801
15. Howland RH. Schizophrenia and amyotrophic lateral sclerosis. *Comp Psychiatry.* (1990) 31:327–36. doi: 10.1016/0010-440X(90)90039-U
16. Longinetti E, Mariosa D, Larsson H, Ye W, Ingre C, Almqvist C, et al. Neurodegenerative and psychiatric diseases among families with amyotrophic lateral sclerosis. *Neurology.* (2017) 89:578–85. doi: 10.1212/WNL.0000000000004179
17. Restuadi R, Garton FC, Benyamin B, Lin T, Williams KL, Vinkhuyzen A, et al. Polygenic risk score analysis for amyotrophic lateral sclerosis leveraging cognitive performance, educational attainment and schizophrenia. *Eur J Hum Genet.* (2021) 21:885. doi: 10.1038/s41431-021-00885-y
18. Byrne S, Heverin M, Elamin M, Bede P, Lynch C, Kenna K, et al. Aggregation of neurologic and neuropsychiatric disease in amyotrophic lateral sclerosis kindreds: a population-based case-control cohort study of familial and sporadic amyotrophic lateral sclerosis. *Ann Neurol.* (2013) 74:699–708. doi: 10.1002/ana.23969
19. O'Brien M, Burke T, Heverin M, Vajda A, McLaughlin R, Gibbons J, et al. Clustering of neuropsychiatric disease in first-degree and second-degree relatives of patients with amyotrophic lateral sclerosis. *JAMA Neurol.* (2017) 74:1425–30. doi: 10.1001/jamaneurol.2017.2699
20. Crayton JW, Meltzer HY, Goode DJ. Motoneuron excitability in psychiatric patients. *Biol Psychiatry.* (1977) 12:545–61.
21. Crayton JW, Meltzer HY. Degeneration and regeneration of motor neurons in psychotic patients. *Biol Psychiatry.* (1979) 14:803–19.
22. Goode DJ, Manning AA. Specific imbalance of right and left sided motor neuron excitability in schizophrenia. *J Neuro Neurosurg Psychiatry.* (1988) 51:626–9. doi: 10.1136/jnnp.51.5.626
23. Wang W, Bu B, Xie M, Zhang M, Yu Z, Tao D, Neural cell cycle dysregulation and central nervous system diseases. *Prog Neurobiol.* (2009) 89:1–17. doi: 10.1016/j.pneurobio.2009.01.007
24. Stone JG, Siedlak SL, Tabaton M, Hirano A, Castellani RJ, Santocanale C, et al. The cell cycle regulator phosphorylated retinoblastoma protein is associated with tau pathology in several tauopathies. *J Neuropathol Exp Neurol.* (2011) 70:578–87. doi: 10.1097/NEN.0b013e3182204414
25. Husseman JW, Nochlin D, Vincent I. Mitotic activation: a convergent mechanism for a cohort of neurodegenerative diseases. *Neurobiol Aging.* (2000) 21:815–28. doi: 10.1016/S0197-4580(00)00221-9
26. Matsumoto H, Spatz M, Laqueur GL. Quantitative changes with age in the DNA content of methylazoxymethanol-induced microencephalic rat brain. *J Neurochem.* (1972) 19:297–306. doi: 10.1111/j.1471-4159.1972.tb01339.x
27. Cattaneo E, Reinach B, Caputi A, Cattabeni F, Di Luca M. Selective *in vitro* blockade of neuroepithelial cells proliferation by methylazoxymethanol, a molecule capable of inducing long lasting functional impairments. *J Neurosci Res.* (1995) 41:640–7. doi: 10.1002/jnr.490410510
28. Zedeck MS, Sternberg SS, Yataganas X, McGowan J. Early changes induced in rat liver by methylazoxymethanol acetate: mitotic abnormalities and polyploidy. *J Natl Cancer Inst.* (1974) 53:719–24. doi: 10.1093/jnci/53.3.719
29. Hawkins WE, Fournie JW, Overstreet RM, Walker WW. Intraocular neoplasms induced by methylazoxymethanol acetate in Japanese medaka (*Oryzias latipes*). *Natl Cancer Inst.* (1986) 76:453–65.
30. Rotchell JM, Blair JB, Shim JK, Hawkins WE, Ostrander GK. Cloning of the Retinoblastoma cDNA from the Japanese medaka (*Oryzias latipes*) and preliminary evidence of mutational alterations in chemically-induced retinoblastomas. *Gene.* (2001) 263:231–7. doi: 10.1016/S0378-1119(00)00566-7
31. Merker R, Kisby GE, Moore H. Abnormal neonatal and adult gene expression patterns in the hippocampus and prefrontal cortex of offspring of rat dams exposed to methylazoxymethanol acetate at embryonic day 17. *Soc. Neurosci*. (2009).
32. Kisby GE, Chlebowski AC, Grygoryev D, Carbone L, Davis B, Nevonen KA, et al. The neurodevelopmental toxin methylazoxymethanol (MAM) induces DNA methylation changes in differentiated human iPSC-derived neuroprogenitor cells (hNPCs). *Soc Neurosci*. (2018) 196:03/A3.
33. Pierozan P, Karlsson O. Mitotically heritable effects of BMAA on striatal neural stem cell proliferation and differentiation. *Cell Death Dis.* (2019) 10:478. doi: 10.1038/s41419-019-1710-2
34. Spencer PS. Hypothesis: etiologic and molecular mechanistic leads for sporadic neurodegenerative diseases based on experience with Western Pacific ALS/PDC. *Front Neurol.* (2019) 10:754. doi: 10.3389/fneur.2019.00754

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychiatry

HYPOTHESIS AND THEORY
published: 21 January 2021
doi: 10.3389/fpsyt.2020.593233

Genomic Variation, Evolvability, and the Paradox of Mental Illness

*Camillo Thomas Gualtieri**

North Carolina Neuropsychiatry, PA, Chapel Hill, NC, United States

Twentieth-century genetics was hard put to explain the irregular behavior of neuropsychiatric disorders. Autism and schizophrenia defy a principle of natural selection; they are highly heritable but associated with low reproductive success. Nevertheless, they persist. The genetic origins of such conditions are confounded by the problem of variable expression, that is, when a given genetic aberration can lead to any one of several distinct disorders. Also, autism and schizophrenia occur on a spectrum of severity, from mild and subclinical cases to the overt and disabling. Such irregularities reflect the problem of missing heritability; although hundreds of genes may be associated with autism or schizophrenia, together they account for only a small proportion of cases. Techniques for higher resolution, genomewide analysis have begun to illuminate the irregular and unpredictable behavior of the human genome. Thus, the origins of neuropsychiatric disorders in particular and complex disease in general have been illuminated. The human genome is characterized by a high degree of structural and behavioral variability: DNA content variation, epistasis, stochasticity in gene expression, and epigenetic changes. These elements have grown more complex as evolution scaled the phylogenetic tree. They are especially pertinent to brain development and function. Genomic variability is a window on the origins of complex disease, neuropsychiatric disorders, and neurodevelopmental disorders in particular. Genomic variability, as it happens, is also the fuel of evolvability. The genomic events that presided over the evolution of the primate and hominid lineages are over-represented in patients with autism and schizophrenia, as well as intellectual disability and epilepsy. That the special qualities of the human genome that drove evolution might, in some way, contribute to neuropsychiatric disorders is a matter of no little interest.

Keywords: autism, schizophrenia, genomic variability, evolvability, missing heritability, copy number variation, neural Darwinism

OPEN ACCESS

Edited by:
Noèlia Fernàndez-Castillo, Centre for Biomedical Network Research (CIBER), Spain

Reviewed by:
Owen Murray Rennert, Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), United States
Richard S. Lee, Johns Hopkins University, United States

***Correspondence:**
Camillo Thomas Gualtieri
tg@ncneuropsych.com

Specialty section:
This article was submitted to Behavioral and Psychiatric Genetics, a section of the journal Frontiers in Psychiatry

Received: *10 August 2020*
Accepted: *27 November 2020*
Published: *21 January 2021*

Citation:
Gualtieri CT (2021) Genomic Variation, Evolvability, and the Paradox of Mental Illness. Front. Psychiatry 11:593233. doi: 10.3389/fpsyt.2020.593233

EVOLVABILITY AND THE PARADOX OF MENTAL ILLNESS

> There remains a gaping hole in Darwinian psychiatry's account of mental disorders: there are no good explanations of why human brains seem to malfunction so often, and why these malfunctions are both heritable and disastrous to survival and reproduction. That is, there is still no good answer for why such susceptibility alleles have persisted despite thousands of generations of natural selection for adaptive human behavior (1).

Nothing in biology makes sense except in the light of evolution (2) but, in light of evolution, mental illness does not make sense. Consider autism and schizophrenia. They begin early in

life and are disabling during the reproductive years. In the language of natural selection, they compromise reproductive fitness. They are associated with decreased fertility, yet their prevalence is undiminished. The heritability of autism and schizophrenia is high (3–6), yet both are characterized by severe social impairment and are associated with low reproductive success.

The persistence of such disorders from one generation to the next is only one of several instances where mental illness is a challenge to Darwinian principles. For example, the problem of missing heritability (7). Autism and schizophrenia are thought to be the consequence of multiple genes of small effect; yet genome-wide association studies association studies have identified hundreds of such genes, and together they account for only a small proportion of cases (6, 8). The contribution of individual genetic variants and their cumulative action to mental disorders is disconcertingly small, usually <10% (9).

Another problem is variable expression, that is, the same genetic variant with various phenotypic expressions. The same genetic aberration may be associated with autism or schizophrenia or intellectual disability or epilepsy, or other neuropsychiatric disorders or combinations thereof (10). Variable expression is illustrated by the fact that mental disorders do not "breed true" but occur in different forms in family members (11–13).

Even within defined diagnostic boundaries, patients differ in virtually every salient characteristic, including symptoms, intellectual and functional abilities, neurocognitive strengths and weaknesses, neuropathological correlates, prognosis, and response to treatment. This observation has led to the idea of taxonomic "spectra," e.g., the autism spectrum, the schizophrenia spectrum, the spectrum of mood disorders, etc. Implicit to the idea is that within every diagnostic category, mental disorders show a continuum of severity, from mild and subclinical cases to the overt and disabling. Most mental disorders occur in pure form in small numbers of individuals and in partial or subclinical forms in a great many more. One presumes a dosage effect, but a dose of what? (14–17).

New developments in genomic medicine have begun to illuminate the irregular behavior of complex diseases in general and mental disorders in particular. They are particularly salient to the neurodevelopmental disorders (NDD) and the focus here is on autism, a prototypical NDD, and schizophrenia which is increasingly recognized as such (18–21). NDD are conditions that originate during gametogenesis/embryogenesis and affect neural development (22). As a family, NDD are diverse in their clinical characteristics and prone to high rates of co-occurrence with other NDD. They are highly heritable but, in most instances, studies have only discovered multiple genes that are probabilistic in their association. Many if not most NDDs arise from structural changes to DNA; e.g., aneuploidy (Down syndrome), simple sequence repeats (Fragile X syndrome), and copy number variants (intellectual disability, epilepsy, autism and schizophrenia, and other mental disorders).

Structural variation, however, is only one chapter in an evolving story. It is just one of many irregularities that characterize the human genome, which we are learning to be uniquely dynamic and mutable. Techniques for higher resolution genomewide analysis highlight the irregular and unpredictable behavior of the genome, endowed as it is with a high degree of variability. It has served the hominid lineage for better and worse. Genomic variability accounts for no small proportion of the missing heritability of complex diseases (7, 23). It has also presided over the runaway evolution of our lineage over the past two million years, and especially the past hundred-thousand. The complex and adaptable human brain reflects a genome that is uniquely mutable and responsive to challenging environments.

The relevant principle is *evolvability*, a species trait that describes the capacity to generate heritable variations (24). The essence of evolvability is inter-generational and inter-individual variability. Phenotypic variation drives natural selection, but variation ultimately derives from the variability of individual genotypes; evolvability describes a genome that can generate a spectrum of phenotypes ranging from major evolutionary innovations to small changes between generations (25, 26). A dynamic, mutable genome is also unstable; it is responsive to life events, especially early ones, and vulnerable to insults of various kind throughout life (24, 27).

Perhaps mental disorders do not make sense in light of traditional genetics, but they make good sense in terms of evolvability. Variability is a highly evolved characteristic of the human genome. It rendered the hominid lineage especially *evolvable*, and humans uniquely adaptable. A dynamic genome is evolvable because it can generate a wide range of phenotypic variations in a comparatively short period of time. It is able to change not only at random but also in response to local exigencies. It has more tools in its kit than random point mutations; if mutational events that directly affected protein-coding sequences were the only available molecular mechanism to generate new variants, adaptive evolution would be ponderous and slow (25, 28). Primates would still be waiting in the trees for successive mutations to occur. Point mutations are insufficient for explaining the runaway evolution of the hominid brain. They can't explain why humans are so different from chimpanzees, why selection might favor genes for post-reproductive longevity, how signal human traits evolved as quickly as they have and how the extraordinary diversity of the human condition can arise from fewer than 25,000 protein-coding genes.

In the special case of hominid evolution, evolvability has been both an independent and a dependent variable. It is a general characteristic that promotes variation, but in our lineage it has been trait under positive selection (29). To support the trait of adaptability, hominids' special traits were agents of persistent and ongoing selective pressure. Social cooperation, abstract intelligence, language, speech and tool-making, and post-reproductive longevity rendered hominids uniquely adaptable. At the same time, they also generated selection pressures in favor of those very traits. Brain evolution, therefore, made hominids adaptable and also positively selected the trait of evolvability. The virtuous circle has been described as runaway evolution (30).

That the special qualities of the human genome that drove our evolution might, in some way, contribute to complex diseases, NDDs and mental disorders is a matter of no little interest.

Origin Stories

The problem of High Heritability and Low Reproductive Success (HHLRS) is paradigmatic. It has been addressed many times and from different perspectives, especially with respect to schizophrenia, a condition that occurs with the same frequency in all the nations and, as far as we know, always has. The quality of our understanding has steadily progressed, beginning with explanations that posited adaptation, thence to those that emphasize adaptability; beginning with evolution as an origin story and later concerned with evolvability as the consequence of a mutable and dynamic genome.

"Adaptationist" theories suggest that a trait embedded in the schizophrenic genome generates phenotypic variations that are advantageous to individuals who do not express the full phenotype. Perhaps the unique characteristics of psychotic individuals were more valued in ancestral times. Psychotic individuals may have been charismatic leaders or shamans (31). Their lower threshold for threat perception may have been useful in times when life was nasty and brutish (32)—*Just because you're paranoid doesn't mean they aren't after you*[1]. The relatives of schizophrenic individuals may be more resistant to disease (33, 34). An idea still current is that mental disorders are associated with creativity (35)—*The romantic view is that illness exacerbates consciousness. Once that illness was TB; now it is insanity*[2]. It was supported by a recent Swedish study that reported individuals with bipolar disorder and the healthy siblings of people with schizophrenia were overrepresented in the creative professions (36) and an Icelandic study that found the same association based on polygenic risk scores (37).

In a similar vein, the occurrence of autistic styles of thinking in the first degree relatives of autistic individuals suggests traits like hyper-systematizing, preference for visuo-spatial relationships and detail-focused processing are adaptive, not only for individuals, but as we learned from the accomplishments of scientists and mathematicians with autistic traits, for the distributed intelligence of society. If autism were a single-gene disorder, it would suggest heterozygote advantage. If autism were a polygenic disorder, a Gaussian model might be relevant, with traits for autistic thinking normally distributed in the population, at the tail end of the curve "hyper-systematizers" would cluster, and beyond a certain threshold reside autism (38, 39).

Adaptationist theories propose that susceptibility alleles are maintained by antagonistic pleiotropy or balancing selection. In that case, probands and/or non-affected siblings would have higher fitness than the general population. In fact, individuals with autism and schizophrenia have lower fertility rates, and their siblings' are the same or lower than the general population (1, 40–44).

Faustian Bargains

Many theories share the metaphor of a Faustian bargain; the human brain is complex, and complex systems are "intrinsically and irreducibly hazardous" (45). For example:

- As specialized functions evolved, biological "trial and error" produced individuals with advanced abilities and others with abnormalities, including schizophrenia (46).
- In the course of evolution, positive selection for cerebral flexibility allowed language to emerge, but the "by-product" was variation in psychological functioning, personality disorders and schizophrenia (47).
- The human brain, with its complex and recently evolved circuitry for social cognition, matured over a long span of time, a span that rendered it susceptible to genetic insults. "This susceptibility was the trade-off for the advantages gained in social cognition" (48).
- The same key genes that have been major contributors to the rapid evolutionary expansion of the human brain and its exceptional cognitive capacity also, in different combinations, are significant contributors to autism and schizophrenia (49, 50).

"By-products" and "trade-offs" refer to antagonistic pleiotropy, when alleles increase the evolutionary fitness payoff of one trait while simultaneously reducing it for another. Such theories posit that a species' predisposition to mental disorder has a selection advantage, albeit indirect. However, interesting, they lack empirical support.

More recently, theories of the HHLRS problem have implicated genomic variability. A high mutation rate characterizes primates in general and humans in particular; multiple, independent *de novo* mutations in many different vulnerable genes and genomic regions (25, 51–60). A theory based on polygenetic mutation-selection balance is that complex brain functions are the work of multiple genes; the genome already harbors a large number of mutations, and new mutations are occurring all the time. Every generation, therefore, carries the burden of old and new mutations, some favorable and some not. Because the distribution of mutations is assumed to be continuous in the population, the expression of positive and negative phenotypes is also continuous. The distribution of negative traits reflects the continuous nature of most (if not all) mental disorders. The "cliff-edge" model captures the non-linear amplification of negative traits that leads to severe and disabling mental illness (61).

Theories derived from a high mutation rate, unlike the others mentioned above, do not propose evolutionary trade-offs; i.e., the cost of complexity is met by the occurrence of neurodevelopmental disorders. The theory of polygenetic mutation-selection balance explains how fitness-reducing genetic variation is maintained in the population; harmful mutations are removed from the gene pool at a rate proportional to their effect on fitness, but "novel mutations occur all the time" (1, 53, 62–65). Thus, the occurrence of mental disorders is inevitable, fueled by deleterious mutations, rare at individual loci but ubiquitous in genomes (1, 66–68).

Addressing the HHLRS problem from the perspective of evolvability is not quite so pessimistic. Novel mutations do, in fact, occur all the time, including rare alleles with large effect, genomic transformations like copy number variants, mutations in non-coding regions of the genome and epigenetic changes in

[1] Joseph Heller, Catch-22.

[2] Sontag S. Illness as metaphor and AIDS and its metaphors. Toronto: Doubleday; 1990, page 36.

gene expression. They may persist for only a few generations but are continually replenished by virtue of high mutation rates (69). Mutations, however, are not necessarily random, but appear to occur in the human genome in accord with principles that are only now coming into focus. The complex, evolvable genome is prone to devastating errors but they occur in particular, knowable ways that open possibilities for anticipation and prevention.

Genomic Variability and the Problem of Missing Heritability

Although single-nucleotide polymorphisms (SNPs) are the most abundant form of DNA variation in the human genome (67), new technologies have shown that individual variation is also the consequence of structural variants involving larger segments of DNA (8, 70–72). Two randomly selected human genomes differ by 0.1% when only SNPs are measured, but when structural variants are also measured, they differ by at least 1% (73, 74).

Structural variants may be rare, compared to 37 million or so SNPs but their large size increases the potential to affect gene expression (4, 75). Genes containing regulatory regions, exons and introns occupy about 5% of the genome and protein-coding exons only about 1% (76), while structural variants comprise no <55% of DNA (77) and perhaps as much as two-thirds (78). Thirty per cent of the human genome are microsatellites (79) and about a third are copy number variants (80–82). Structural variation is more common in humans than other mammals and occurs several times more frequently in neurons than other cells (54–58). Evidence from a wide range of common diseases indicates that genetic heterogeneity is a key characteristic of the human genome and that "most genetic control is due to rare variants" (4). The challenges that genetic analysis poses to mental disorders are well in accord with our new and developing appreciation of genomic behavior. Genomic variability is not only the fuel of evolvability but a window to understanding the origins of complex disease, mental disorders and NDDs.

THE NATURE OF GENOMIC VARIABILITY

1. Structural Variation

The hominid genome is not a static blueprint but a source of "gene nurseries" that play a role in gene innovation and adaptability (83). It is something that happened during the later course of evolution. Pre-primate evolution was probably driven by point mutations or whole-genome duplications (59, 84, 85). Ascending the phylogenetic ladder, mutations by genetic rearrangement have been progressively more important. Almost all of the genetic differences between humans and other primates are a result of duplications, deletions, inversions, insertions, and transformations (86, 87). For example, about 35 million nucleotide substitutions distinguish humans from chimpanzees. About half are transposable element insertions (88).

The high-resolution molecular scanning tools developed during this century have revealed a genome that is no less than restless. For example, aneuploidy traditionally referred to supernumerary copies of whole chromosomes (e.g., trisomy 21, Down syndrome). In recent years, novel aneuploidy syndromes have been identified, and the definition has been extended to include deletions, insertions and duplications of subchromosomal regions (89). A new term that includes classical aneuploidy and other structural variants is DNA content variation (DCV). Such variants are dynamic, fluid and unstable, both through germlines and in somatic events and they are ubiquitous in the human genome (74).

Structural variation is most likely to occur in non-coding regions of the DNA molecule. Non-coding regions are comparatively unstable and generate mutations and rearrangements at a high rate (90–92). They are especially prone to genomic rearrangements. As it happens, the human genome contains more non-coding DNA than any other animal or plant. Among microorganisms, <25% of DNA is non-coding. In plants and lower animals, about 60% is non-coding; in primates, the proportion is higher. In humans, it is 98.5% (93).

The human genome is particularly enriched in both number and length of retrotransposons (94). The propagation of *Alu* elements is coincidental with the fast evolution of segmental duplications in the primate genome, which grew as a result of a major burst in *Alu* activity 25–55 MYA; ours have continued to expand (56, 59). Compared to chimpanzees, humans have nearly three times as many *Alu* elements (88).

Copy number variants (CNVs) are duplications or deletions that are >1,000 base pairs in length. At least half of the CNVs thus far detected include protein-coding regions and affect the behavior of the relevant genes. The functional impact of CNVs extends across the full range of biology, from gene expression to cellular phenotypes (70, 95) and to all classes of human disease with an underlying genetic basis, whether inherited or sporadic (74, 96). They are especially relevant to autism and schizophrenia.

Aneuploidy Is Common in Brain

The human brain is a genetic mosaic. In the adult cerebral cortex aneuploidy, broadly defined, is estimated to occur in no fewer than 30–50% of neurons, many times more frequently than in somatic cells (97). Fetal neurons develop over a longer span of time than those of most other animals, and undergo more cell-divisions along the way. A prolonged span of mitotic activity, one presumes, makes such rearrangements more likely (98).

How the brain accommodates events that are potentially deleterious is a mystery. In spite of cellular variability and diversity and the consequent differences in gene expression, the functionality of the CNS is usually not compromised (99, 100). Aneuploid cells are capable of surviving the massive cell death that accompanies neurogenesis (101). They can differentiate into neuronal lineages and are integrated into active neural circuitry with the potential to influence normal brain functions (102, 103). DCV shows regional variation within the human brain, more in the frontal cortex and less in the cerebellum. Neuronal DCV concentration also differs among individuals (104). Genetically mosaic neural circuitries are part of the normal brain organization (105, 106) and reflect the structural and functional mosaicism of brain itself, its neuronal diversity and expansive range of behavior (97).

The prevalence of neuronal aneuploidy captures two principles that we shall rely on. One is that *brain is an evolutionary system*. From this perspective, maintenance of neutral or beneficial aneuploidies is the end-result of selective pressures (107). Extreme forms of aneuploidy are eliminated during fetal development by programmed cell death, while other neurons survive and contribute to neural organization. One assumes that the loss or gain of genetic material may render some cells more "fit" than others, perhaps by increasing stress resistance or enhancing functional capacity (97).

The second principle is that *high mutation rates confer adaptability*. Genomic diversity prepares developing brain for the multitude of tasks before it. A more diverse neuronal population is better-equipped to adapt to challenging environments; a lineage thus endowed is under positive selection for genomic variability and evolvability.

Although mature neurons are terminally differentiated cells, they remain capable of generating structural variants. When they are stressed, neurons seem to be able to reactivate elements of the replication mechanism (103). Such "cell-cycle events" may be pathological (108); they have been observed in several human neurodegenerative diseases, including Alzheimer's disease and ataxia telangiectasia. Alternatively, they may be an attempt at neuronal self-protection. The ectopic expression of cell cycle markers may be "a desperate attempt of a neuron under stress to protect itself" (109). It is not a paradox but another principle of evolvability: *mechanisms to increase robustness are also, in some circumstances, fragile, and vulnerable to pathology*.

That genetic plasticity of post-mitotic neurons is mostly adaptive is captured by the frequency with which mobile elements can change their position within the genome, either by a DNA-based (transposition) or an RNA-based (retrotransposition) mechanism. The latter is of particular interest because it is a form of plasticity that responds to early life experiences (110, 111).

Structural Variation in Autism and Schizophrenia

DCV contributes to phenotypic diversity, adaptability and individual differences in brain organization. It also plays a role in disease, albeit one that is quite variable (89, 97). A number of recent studies have provided compelling evidence that autistic, schizophrenic, and bipolar patients are more likely to possess CNVs in their genome, especially deletions of genomic regions (112, 113); no one CNV in particular, however, but rather any of a number of rare CNVs (11, 72, 114–120).

Compared to healthy controls, individuals with schizophrenia are three times more likely to harbor rare structural mutations. The risk is even higher in subjects with early onset schizophrenia. Each rare mutation disrupts a different gene or genes and the disrupted genes are disproportionally involved with signaling and neurodevelopment (4, 113). The burden of CNVs in schizophrenic patients ranges is low, from 2.4 to 10% (121, 122). However, CNV analyses with better genome coverage will probably discover many more relevant associations (123). When microdeletions and microduplications >100 kb were identified by microarray comparative genomic hybridization and validated by high-resolution platforms, novel deletions and duplications were found in 15% of individuals with schizophrenia and 20% of young-onset cases (vs. 5% in controls) (124).

The prevalence of structural variants in autistic patients, on the basis of less sensitive genomic scans, is usually given as 5–10% but may be as high as 28% (112, 114, 120, 122, 125). Although the vast majority of CNVs are inherited, CNVs that occur as *de novo* mutations are more commonly associated with autism (and schizophrenia). *De novo* CNVs are found with higher frequency among sporadic cases, whereas inherited CNVs are more common in familial cases (121, 126–129). With better technology, the prevalence of structural variants will likely be higher (95, 128, 130, 131).

It probably will be but what will it mean? Structural variants, interesting as they are, may simply join the long list of "multiple genes of small effect" and the even longer list of epigenetic and non-genetic factors associated with autism and schizophrenia, with no one making more than a small contribution (132). Many CNVs are functionally neutral; healthy individuals carry, on average, about 11 CNVs (57). Even the CNVs that are known to be associated with a neuropsychiatric disorder are present in low concentrations in healthy controls (49, 128, 133–138). CNVs are not, as a rule, highly penetrant (139–142). Further, when CNVs are expressed, the phenotypes are variable: schizophrenia or autism, but more often developmental disabilities, congenital malformations, or other mental disorders (11, 125, 140).

Genomic Variability: 2. Variable Expression

Reduced penetrance refers to individuals with a specific genotype, but the clinical phenotype is not expressed or is expressed in a lesser form. *Variable expression* is the degree of variation in a clinical phenotype in individuals who carry a specific genotype (143). CNVs manifest both.

CNVs known to be associated with autism or schizophrenia are also found to occur in association with other psychiatric or neurodevelopmental conditions. A good example is the 22q11.2 deletion. The deletion occurs in only about 1% of schizophrenic patients, yet it is the strongest DNA-based risk factor for schizophrenia identified so far. Individuals with the 22q11.2 deletion have a 20-fold increase in risk for schizophrenia. Nevertheless, individuals with the deletion may be perfectly normal; or they may have developmental delay, congenital malformations (144, 145), generalized epilepsy (146, 147), intellectual disability (148), learning disability, autism, ADHD, anxiety, depression, OCD, or bipolar disorder (11, 97, 113, 118, 149–155). Variable expression is certainly the case for many other deletions and duplications [e.g., 1q21, 15q13.3, and 16p11.2; (113)]. Even when a structural variant and the alleles therein are well-defined, or when a repeat number is correctly counted, the clinical consequences are unpredictable. In a kindred harboring a translocation disrupting *DISC1*, carriers had schizophrenia, bipolar disorder, major depressive disorder, cognitive impairment or no mental disorders at all (116, 125). It is perplexing but typical not only of complex diseases but complex traits in general (115).

Genomic Variability: 3. Epistasis

Most phenotypes result from intricate gene interactions. These interactions, recognized as deviations from additive genetic effects on the phenotype, and collectively called epistasis (156). It is a universal characteristic of complex genetic traits. It is one more aspect of genomic behavior that accounts for the phenotypic variations that occur with mutations in the same gene (157). A mutation may be benign or beneficial to one individual but deleterious to another individual; it is contingent on an individual's "genetic background" (158). "Sign epistasis" means that a mutation is beneficial on some genetic backgrounds but deleterious on others (159).

"Epistasis" was coined to describe the suppression of an allelic phenotype by an allele at another locus (160). It is more complicated than gene B influencing the expression of gene A; "higher order" epistasis involves interactions among multiple mutations (161, 162). Antagonistic epistasis among deleterious mutations and synergistic epistasis among beneficial mutations represent positive epistasis, whereas the opposite situations represent negative epistasis. Intra-gene epistasis results from effects of mutations on RNA stability and enzyme activity and inter-gene epistasis may result from protein interactions and the structure of metabolic networks (156, 163).

To say that a disorder is the consequence of "multiple genes of small effect" understates the magnitude of the problem. The results of multiple genes acting together may be additive or multiplicative, linear or non-linear. Epistasis is also affected by events in the environment of the cell (164, 165) and the developmental stage at which they occur (166). Within a gene, different alleles can interact epistatically with different gene sets (167).

Divergent phenotypes emanating from identical genetic variation(s) are consequences not only of epistasis but also (in varying degrees and combinations) pleiotropy and locus heterogeneity, environmental factors, epigenetic mechanisms, stochastic events, dose and timing of gene expression, and RNA regulatory elements (4, 118, 138, 168–172). Some variants may only affect risk if they co-occur with other genetic or environmental risk factors (115).

The Darwinian "problems" we cited earlier were the heterogeneity of disorders even within well-defined categories, their occurrence in pure form in small numbers of individuals and in partial or subclinical forms in a great many more, the fact that most such conditions do not "breed true" but are present in different forms in family members, comorbidity, and the HHLRS problem (11–17). The question is not why they happen but, knowing what we now know, how could they happen in any other way.

Genomic Variability: 4. Gene Regulation

The main effects of DNA structural variants are commonly attributed to changes in gene expression and its regulation. The latter, however, represent an entirely different dimension of genomic variability. Gene expression is the work of multiple genes and other DNA segments, proteins, and RNAs of varied stripe, as well as epigenetic changes to DNA and chromatin. Together, they comprise the genome's control architecture and participate in countless gene regulatory networks (GRNs). Networks of regulatory genes occupy more DNA than protein-coding genes do (173, 174).

Variation in gene expression levels is abundant within and among populations; quantitative differences in gene expression are responsible for a significant amount of the variation represented by individual differences (89, 175). Gene regulation is itself a heritable trait. When variation in gene expression phenotypes is compared among unrelated individuals, among siblings within families and between monozygotic twins, there is a strong genetic contribution to variation in the level of gene expression (176, 177).

Gene expression microarrays and transcriptome sequencing have revealed remarkable natural variation in gene expression levels within populations as well as between species (178). Regulatory variation within and between species is thought to explain a large proportion of phenotypic diversity of life. It is believed that most complex traits originate in non-coding regions that affect gene regulation (179).

- *GRNs are the basis of organization and stability*. They are models of dynamic complexity with modular structures that were present at the base of the metazoan tree (93). Their stability is a function of modularity, feedback loops and the redundancy of the genetic material upon which they operate. They are able to withstand gene disruptions due to mutation or environmental stress (93, 180).
- *GRNs are organized hierarchically*. GRNs are organized hierarchically and control the nature of available variation by "packaging" genetic modules for selection (181). Thus, they are agents not only of stability but also evolvability.
- *GRNs also reflect the fact that the organism is a complex adaptive system*. In GRNs, information from the cell state and the outside environment is translated into the correctly timed expression of the appropriate gene products (182). GRNs have been called the "nexus of physiological adaptation" (183).
- *GRNs are a source of endless variation*. For most genes, transcript structure and expression level are not only highly variable but functionally independent (175).

Genomic Variability: 5. Epigenetics

Epigenetic changes participate in GRNs by influencing whether a gene is expressed, when it is and to what degree. Histone (or chromatin) modification changes the structure of a core octamer that contains two copies of the histones H2A, H2B, H3, and H4 (humans have them and chimpanzees don't). Histone remodeling ensures that DNA remains accessible to the transcriptional machinery (184–187).

DNA methylation is another epigenetic mechanism that can affect gene expression by making cytosine binding sites more or less accessible to transcription factors; or it may attract proteins that are themselves gene repressors. Variable DNA methylation is the loss or gain of methylation at CpG dinucleotides. Variable DNA methylation is the loss or gain of

CpG dinucleotides; levels of methylation vary among individuals and change at different stages of life. Patterns of methylation or hypomethylation at different sites appear to be associated with longevity or, alternatively, aging-related diseases like cancer and Alzheimer's (188–191).

Non-coding RNAs (ncRNA) are a third epigenetic mechanism. ncRNAs are generated by intergenic or antisense transcription, usually at introns, and comprise more than 90% of transcriptional output from the genome. ncRNAs remain tethered to their transcription site, serving as allelic markers and lending spatial and temporal specificity to gene expression (173, 192). An example of temporal specificity are the 35 ncRNAs that govern the development of dopaminergic neurons from neural stem cells. They are differentially expressed between progenitor and mature states and, in all probability at different stages of the life cycle (193).

Although epigenetic mechanisms are most active during embryogenesis and early life, they remain active throughout life, meaning that life experiences can make enduring changes in the genome. Further, epigenetic changes can be passed on. Phenotypic variations unrelated to variations in DNA base sequences may be transmitted to subsequent generations of cells or organisms; an epigenetic trait is "a stably heritable phenotype" (194–196). Epigenetic inheritance is not quite the same as the Lamarckian theory of inheritance of acquired characteristics, but it's not all that different. It allows genetic variants that do not change the *mean* phenotype change the *variability* of phenotype. However, epigenetic modifications can also affect the probability that a region of the genome will mutate [e.g., single-base and transposon-mediated mutations and translocation; (197)]. Therefore, not only may epigenetic modification promote heritable phenotypic variation, but can also facilitate genetic evolution by modulating mutation rates across the genome (198, 199). Epigenetic variation, therefore, is a powerful mechanism for evolutionary adaptation to changing environments.

Epigenetic participation in GRNs is illustrated by the remarkable diversity of genetically identical cells and organisms even when they have identical environmental exposures. Cloned animals, for example, may have different phenotypes at birth. In monozygotic human twins, gene expression is four times more dissimilar in older subjects (50-year-old monozygotic twins) in comparison to younger subjects (3-year-old monozygotic twins). Autism and schizophrenia both show surprisingly high frequencies of phenotypic discordance in monozygotic twins. In one study, genomic DNA extracted from leukocytes of male MZ twins discordant for schizophrenia was found to have significant differences between the twins at sites that are closely associated with CpG islands and gene expression (200, 201). Similar differences in methylation patterns have been noted in MZ twins discordant for autism (202).

Epigenetics reflects three principles long familiar to psychiatrists:

1. Experience influences development and the effects may be long-lasting.
2. The maternal contribution to development is important on many levels, including the transmission of gene expression patterns by epigenetic programing *in utero* and early life.
3. Epigenetic programming is particularly robust in early life but epigenetic re-programming, modified by experience, occurs throughout the lifespan.

Genomic Variability: 6. Stochasticity, Noise

Another principle that psychiatrists encounter, perhaps more often than they like, is captured by the stochastic behavior of gene regulatory networks. The generation of gene products is necessarily sensitive to unpredictable fluctuations. Gene regulation is intrinsically "noisy" (203).

Transcription factors are proteins expressed by genes which, in turn, control the expression of genes. Their dynamics are constrained by a highly structured, densely tangled intracellular environment where DNA, RNA, and proteins may be present and active with only a few copies per cell (181). Since typically 30–100 regulatory proteins per gene are used as transcription factors (204), a corresponding number of genes must go through their individual cycles of expression in a perfectly synchronized manner; otherwise, a shortage of a few transcription factors may lead to drop-out from the regulatory process and a halting of big sections of transcription machinery (205). When a transcription factor initiates gene expression, its effects are amplified by cycles of epigenetic reprogramming; normal cell activity involves thousands of transcription factors operating in parallel with epigenetic mechanisms. Thus, there are ample opportunities for "noise" or stochasticity—to arise. A small number of epigenetic changes, or a single one, may have a net effect on multiple downstream targets. Environmental signals can affect the activity of transcription factors and epigenetic complexes to regulate gene expression (93, 192).

The modular structure of GRNs is a stabilizing factor but there is no overlying regulatory system, in the sense that "regulation" is used in systems control theory (181). If anything, GRNs are self-organized, responding to differences in the internal states of cells; to the effects of subtle environmental differences, such as morphogen gradients during development; to predictable processes such as cell cycle progression; to random processes such as partitioning of mitochondria during cell division; and to ongoing genetic mutations (206). They also manifest the inherent stochasticity of biochemical processes that are dependent on infrequent molecular events involving small numbers of molecules. Like people, they encounter innumerable opportunities to do something unpredictable.

Intellectual disability, autism and schizophrenia have been related to GRN's that regulate neurogenesis, neuronal connectivity, cell signaling, axon guidance, presynaptic pathways, post-synaptic protein complexes, cytoskeleton dynamics, intracellular signal transduction pathways, transcription regulation, and epigenetic modulation of the chromatin structure (207–211). Clinical studies of gene networks, however, are theoretical, relying on statistical associations among genes that are expressed in particular areas or that have known functions.

Mapping the "interactome" directly has only been done in model organisms such as yeast, fruit flies, and roundworms.

The Origins of Genomic Variability

It may be reassuring to clinicians to learn that the genetics of complex disease are no less complex than the phenotypes they generate but may wonder how it came to be that way. Genomic variability, in all its dimensions, has increased as evolution scaled the phylogenetic tree; a high mutation rate characterizes primates in general and humans in particular. It is telling that, of all the great apes, humans have less genetic diversity defined by the number of SNPs, but much more in terms of structural variants (212).

DNA content variation is more common in humans than other mammals; because it occurs several times more frequently in neurons than other cells we may assume it favors the evolution of a complex and adaptable brain (25, 54–60). Most of the structural rearrangements in the mammalian lineage are believed to have occurred in brain-specific genes (55). The duplication rate accelerated at the time of the common hominoid ancestor (57) and our own species is remarkable for numerous large segmental duplications (25, 60).

Structural rearrangements allow multiple forms of a gene to co-evolve and to rapidly reorganize the genome. Variation in the amounts and types of repetitive DNA varies between organisms and reflects how rapidly a species is capable of evolving in response to changes in its environment (213). It is believed to foster inter-individual genetic variability and variation from one generation to the next (168, 214, 215). The creation of novel genes by genomic transformation is said to have been "the major driving force in hominid evolution" (91) and "the *sine qua non* for evolvability" (24).

Changes in the gene regulatory machinery are another creative force in morphological evolution (55, 216). The proliferation of new regulatory genes coincided with the emergence of increasing organismal complexity, and they enabled organisms to develop new functionalities. Differences in gene expression are probably the real divisor between humans and chimpanzees; our genes are largely the same but the difference is accelerated gene expression changes in the human brain (58). During hominid evolution, mechanisms of gene expression have been elaborated to an extraordinary degree and this is especially true of genes that govern brain development (28, 58, 217, 218). The adaptability of the genome is further enhanced by epigenetic mechanisms and, as it happens, humans are the most epigenetically complex species (219, 220).

The convergence of genomic variability, morphologic evolution and the evolution of a complex, highly adaptable brain is contained within the concept of evolvability. Animals possess two systems with which to address the challenges of their environment: a neural system, that directs behavior and a genomic system that provides the wherewithal for behavior to occur. Both systems are composed of networks that are at once inordinately complex and also flexible. Their complexity and flexibility increase as one ascends the phylogenetic ladder. We humans are the carriers of genomic and neural systems that make us uniquely adaptable and, one may say, evolvable.

We also know that genomic events that conferred evolvability to the primate/hominid lineage are over-represented in patients with autism and schizophrenia, as well as intellectual disability and epilepsy. The irregular behavior of the genome is probably reflected in the evolution and organization of neural connectivity; it is certainly reflected in the irregular behavior of complex diseases in general, autism and schizophrenia in particular. The pictures are gradually coming into focus: the mechanisms of genomic variation, the vicissitudes of gene expression and their variable phenotypic consequences. As the particulars accrue, so do underlying principles. Genomic variability is clearly related to evolvability and to NDDs, but are the latter two connected? The question has been answered: "The same genes that were responsible for the evolution of the human brain are also a significant cause of autism and schizophrenia" (50). There are hazards, however, to facile answers. The issue isn't "genes" at all but the principles that govern their behavior. Three principles relevant to evolvability are reiterated in a clinical arena where the idea of evolvability is current, in studies of cancer.

Evolvability in Real Life

Cancer as an evolutionary system is neither metaphor nor theory but an observation that has implications for diagnosis and treatment (221). Cancers are dynamic entities that never cease to evolve. They don't only grow; they "evolve according to well-understood principles of somatic selection, along trajectories that can be described by established methods for tracing phylogenies" (222). Individual cancer cells are reproductive units within a population and compete not only with non-neoplastic cells but also with other cancer cells that possess different genotypes. Neoplastic cells, especially solid cancers, generate additional mutations with new phenotypes, such as the ability to invade adjacent tissues, recruit blood supply, overcome nutritional deficiencies and resist immune attack (223). Their mutational patterns evolve stochastically and are highly diverse. Cancers contain, on average, 50 non-silent mutations in the coding regions of different genes; only a small fraction of the mutations are common to all tumors in a given class (224–226).

In cancers, genetic heterogeneity is "the fuel that drives selection" (223, 227). Cancer cells are notorious for their genomic instability. They show genomic rearrangements at the microscopic and submicroscopic level; mutations in coding regions, genomic loss or amplification, transpositions and transformations, aberrant methylation and expression profiles. Such mutations render cancer cells evolvable and, all too often, difficult to treat[3] (228, 229).

The evolvability of cancer cells captures three relevant principles. The first is that *high mutation rates confer adaptability*. Cancer cells are subject to stresses—hypoxia, nutrient depletion, immune surveillance, chemotherapy, —and they show that high mutation rates occur when cells are stressed. Mutations accelerate adaptation and cells with the requisite phenotypes are selected (230–232). This observation has wider import; it belies the

[3]Mutant clones are often the source of treatment resistance. Mutant clones that are resistant to chemotherapy may persist in a dormant state, only to erupt at a later time.

assumption that mutagenesis is random, constant, and gradual. Mutations occur more frequently when cells are maladapted to their environments, and the mechanisms that drive mutation tend to target specific genomic structures (233).

The second principle is that the genome is balanced between two *evolutionary traits, stability, and variability*. Stability, or robustness, is the ability to remain adapted to existing conditions in spite of perturbations. Cells that are genetically unstable—for example, those with certain CNVs—are predisposed to neoplastic transformation (229, 234). Cancer cells are unusually robust because high mutation rates render them more variable, more likely to advance in the face of hostile environments (235).

The third is that *the human genome has achieved a good balance.* The hominid lineage has evolved successfully while maintaining an array of DNA monitoring and repair enzymes that render most mutations an evolutionary dead-end. That is why cancers, like severe, disabling mental disorders, are comparatively rare events. Our genome derives disproportionally from individuals who had effective mechanisms for suppressing the conditions (223, 236). Nevertheless, mutations occur and continue to exercise effects, not always good ones. The balance is not perfect. Cancers, like mental disorders, continue to occur. Mechanisms for mutation prevention and suppression are imperfect. The accumulation of new genetic variants may fuel evolution of the species but not all variants are beneficial to individuals.

Neural Darwinism

Neoplastic cells compete not only with their host but with each other. The evolutionary expansion of neoplastic cells is a far cry from what happens among post-mitotic neurons in autism and schizophrenia, but the point is that evolution is more than an historical event; it is the product of biological processes and principles that remain active. So, if brain were regarded as an evolutionary system, the evolvable units would not be neurons but the connections that form among them—at a fundamental level, synapses and at the next, neural networks. If autism and schizophrenia are, in fact, disruptions of neural circuitry (237, 238), they may be conceptualized as aberrations in one or more evolutionary processes.

The development and maturation of brain is characterized by, among other things, the evolution of discrete interconnected groups of active neurons––"cell assemblies" or neural networks (239). "Synaptic Darwinism" refers to synapses "replicating" by strengthening their connections and "mutating" by connecting new cells. "Neural Darwinism" is more than a metaphor. During the course of development and throughout life, there is always a vast array of potential connections to be made, but only the fittest survive (240). The axonal arbors of neurons tend to be in close proximity; competition results in strengthening the connection with one neuron and withdrawal of the rest. Onle certain connection patterns are *selected* in order to form optimal configurations (241); alternatively, one may say that neurons *compete* to form effective connections (242) "What could be analogous in the case of synapses to genetic mutations? We believe the obvious analog to genetic mutation is structural synaptic change" (243).

"Competition" among neurons is reflected at the level of molecules; in living cells different molecular species compete for binding to the same molecular target. A relevant example is the competition of genes for the transcription machinery or the competition of mRNAs for the ribosome. In transcription and translation networks, competition takes place within large-scale networks that typically have hundreds to thousands of competitors, all at a relatively low concentration. Many different genes compete for RNA polymerase and transcription factors, microRNAs compete for mRNA and thousands of different transcripts compete for a common set of ribosomes and translation factors. Translated proteins compete for a common folding and transport machinery and when they are to be removed they compete for a common degradation machinery (244–246).

Competition among alternative pathways occurs in the behavior of neural networks and the mechanisms of cognitive control (247, 248). In the lateral amygdala, neurons with increased excitability during training outcompete their neighbors for allocation to an engram (249). Representation in the visual system is competitive; both top-down and bottom-up bias influences the ongoing competition (250). Awareness itself is a bottleneck through which only a small amount of neural activity is successful in coming to one's attention (251, 252).

The evolvable human genome gives rise to neural systems that are themselves evolvable. The principle that mutations confer adaptability is equally pertinent to the genome and the neural connectome; variability in the former is a homolog of plasticity in the latter. So also does the principle of balance. The neural connectome depends on the expression of many genes involved in neural activity, including the behavior of receptors, membrane transporters, enzymes for neurotransmitter synthesis or degradation, cytoskeletal and vesicular proteins, signaling and effector proteins, and regulators of transcription and translation (253). The variable expression of so many genes lends bias to the development of neural connectivity. Autism and schizophrenia originate in networks of DNA, nucleotides and proteins that preside over neurogenesis and neural migration, synaptogenesis, and arborization. One can say that autism and schizophrenia are "disruptions of neural circuitry," or one might better say that certain genes lend bias to the evolution neural circuity. "Disruption" conveys either/or; "bias" is consistent with the clinical heterogeneity of those disorders. The competition at every level to form optimal connectivity is biased against a perfect balance.

Robustness Is the Balance Between Stability and Variability

Complex adaptive systems, such as the genome and the neural connectome, are networks comprised of a hierarchy of networks. Coherent behavior in such systems arises from the interplay among many individual agents (254, 255). Networks and the agents within them are subject to events in the environment and to inevitable stochastic variations, and to all are available alternative, competitive pathways. A system is robust if it optimizes the balance between cooperation and competition,

stability and variability, randomness, and regularity; thus, a species is robust if it can maintain essential functions yet maintains the potential to evolve (256, 257). The optimal balance is a robust system that is stable but also flexible and adaptable. Thus, the genome responds to stresses with adaptive mutations and the neural connectome responds with adaptive plasticity. How well cells, organisms, and species achieve that balance in the face of innumerable unforeseen events is the essence of evolvability.

The capacity of systems to maintain essential functions when exposed to challenges is called phenotypic robustness. Robustness is central to evolvability, because it allows an evolving population to explore new genotypes without detrimentally affecting essential phenotypes. Related terms are developmental stability, the ability to produce a robust phenotype when faced with challenges during development; and canalization, when genetic systems under long-term stabilizing selection evolve to a state of increased stability (157, 258–260). To that end, organisms have evolved systems to buffer the impact of mutations, noise and untoward environmental events.

The trait of robustness is pervasive in biology at every organizational level including protein folding, gene expression, and the neural connectome, as well as in physiological homeostasis, development, organism survival, species persistence, and ecological resilience (256, 261–265). In successful organisms, however, the trait of robustness exists in balance with trait evolvability. It seems a paradox; a robust system is resistant to generating new phenotypes while an evolvable system takes advantage of mutations to generate phenotypic variations (266). However, a robust system is only static with respect to essential morphological characteristics, that are said to be "deeply canalized." If perturbations always led back to the native state, organisms would find it difficult to contend to unfamiliar challenges (262). A better evolutionary strategy is to protect essential native functions in the face of unexpected tumult while evolving new ones to adjust to challenges. This kind of robustness is not a barrier to evolution, but enhances it. It enables genotypic variability and new phenotypes without untoward functional consequences (24, 256, 267).

Robust but Fragile

The mechanisms that successful organisms employ to ensure robustness are sufficiently flexible to support change when circumstances demand. Yet the balance between robustness and evolvability is tenuous and contains the germs of fragility. Robustness supports evolvability but the genome and the neural connectome are vulnerable to devastating disturbances. It is the balance between robustness and evolvability that may be relevant to complex diseases and mental disorders. The balance is maintained by two kinds of buffering systems, one structural, the other dynamic.

Modularity, for example, is a way most complex systems are structured. Modularity promotes stability by containing perturbations, reducing the interdependence of events and minimizing system-wide impact. Modules optimize network function because they are energy-efficient (268). They also confer evolvability by reducing constraints on change (269). Modular systems, however, are vulnerable to unexpected perturbations. This "robust yet fragile" trade-off is fundamental to complex dynamic systems (262, 270). Other mechanisms that are at once stabilizing and supportive of evolvability, such as redundancy, gene duplication, feedback control, and bow-tie architecture, are also vulnerable to catastrophic failure in the face of unusual stressors (267, 269). Gene duplication, for example, is a way to lower the intrinsic noise in gene expression. Increased copy number provides stability by preserving the native function of the gene when copies happen to mutate (206). We have already met some of the catastrophic failures that cometimes accompany copy number variation.

When new genes arise, systems exist to suppress their expression. Novel genes are not necessarily eliminated but survive from one generation to the next. The result is a vast pool for potential change, known as cryptic genetic variation (CGV). It is invisible under normal conditions but when circumstances change it is fuel for evolution (271). The hidden genes are a massive cache of adaptive potential. On the other hand, they are also a reservoir of potentially deleterious alleles (272).

When a robust system is stressed beyond its level of tolerance, phenotypic expression may be decanalized, which leads to increased phenotypic variability (273). A population moving beyond its adaptive niche challenges its genome to respond by increasing trait variability; decanalization is thus an agent of evolvability. At an individual level, crossing the threshold of stability opens the opportunity for cryptic alleles to express themselves, sometimes in untoward ways. Decanalization has been proposed to explain missing heritability in complex disease (273–275).

In diverse model organisms, the threshold of robustness differs among individuals; those with decreased robustness show increased penetrance of mutations and express previously cryptic genetic variation. It is not unlikely that phenotypic robustness also differs among humans. Individuals with lower robustness are thus more responsive to genetic and environmental perturbations and more susceptible to disease (157, 275). This is particularly relevant to disorders like autism and schizophrenia, where environmental events such as intrauterine exposure or obstetrical suboptimality may evoke a latent genetic proclivity.

Brain may be particularly vulnerable to decanalization because its development and activity are instructed by more than half of the genome, each allele with a different level of robustness. The hominid neocortex has expanded considerably, compared with closely related species; perhaps there hasn't been sufficient time to evolve robust cortical developmental trajectories. Brain has tightly-regulated critical windows of development; thus, there are few opportunities to compensate for perturbations (274).

Versatile Proteins

Heat-shock proteins (HSP) are examples of a second type of buffering mechanism, dynamic and dependent on the expression of versatile proteins. HSP are protective against a wide range of environmental stresses, notably those known to be prenatal risk factors for neurological and psychiatric disorders, such as viral infection, hypoxia, inflammation, irradiation, alcohol, maternal seizure, and methylmercury (276–279). They are also genetic buffers, residing at the boundary between evolutionary stasis and change (264, 280).

HSP are molecular chaperones that play an essential role by protecting their "client" proteins from misfolding, for example, in the face of heat stress (281, 282). Like many genes and proteins, they are so-named after their earliest reported role; HSP are induced by high temperatures. They are an essential element of the biological stress response (281, 283). All species, from prokaryotes on up, have HSP genes; HSP expression is correlated with resistance to stress; and species' thresholds for HSP expression are correlated with the levels of stress that they naturally undergo (284). One HSP, hsp90, is necessary for the glucocorticoid receptor to develop (285) and regulates receptor activity (286). It participates in hormone signaling (280, 287) and restores hypothalamic–pituitary–adrenal homeostasis after stressful events (288). The aging-related decline in stress tolerance is associated with a lower capacity to generate stress proteins in general and HSP in particular (289, 290).

Genetic mutations are an important source of abnormal and misfolded proteins, and HSP buffer their effects as well (291). When an abnormal protein is expressed by a mutant gene, chaperones such as hsp90 participate in its degradation. Thus, HSP can buffer (i.e., suppress) phenotypic variation (283, 292–295). When genetic variations are "decoupled" from phenotypic expression, however, cryptic mutations accumulate (262, 293, 294, 296, 297). In light of this action, demonstrated in studies of fruit flies, zebrafish, bacteria, yeast, fungi and plants, hsp90 is called a "genetic capacitor" (282, 292, 294, 298, 299). It buffers against mutations and thus contribute to the robustness of the phenotype. When buffering operates normally, it prevents the development of abnormal phenotypes; if the potential phenotype were autism or schizophrenia, an effective buffering system might prevent the condition even in individuals with genetic or environmental risk factors (299, 300). Buffering capacity is finite, however, in some individuals more so than others. When an individual's buffering capacity is exceeded by excessive or unusual perturbations the result is increased phenotypic variation (292, 301, 302). In populations, this stress-sensitive storage and release of suppressed alleles may favor adaptive evolution (299, 303). In individuals, the consequences may not be quite so sanguine.

Fetal development has to be insulated from the damaging impacts of environmental and genetic perturbations to produce highly predictable phenotypes (300). The relevance of HSP to NDD is highlighted by their known effects on development; hsp90, for example, occupies a critical position in development because most of its client proteins are signal transducers (282, 292, 304–306). Because cell division is such an active process during fetal life, the opportunities for genetic mistakes to occur—especially in developing neurons—are legion. From studies in various organisms we have learned that many HSP buffer developmental perturbations on morphological traits (300, 305). In embryonic mice, exposure to subthreshold levels of environmental toxins induces HSP activity; inhibition of HSP leads to structural brain abnormalities and epileptogenesis (307).

In experimental animals, HSP activity can be inhibited in various ways. In real life, HSP is naturally variable and individuals differ in their capacity to generate buffering activity. There is substantial inter-individual variation in HSP induction during embryonic development of the central nervous system, where chaperones are essential to neuronal differentiation and survival (308). Embryos with stronger induction of HSP are less likely to be affected by inherited mutations; their development is more robust because mutations are less likely to be expressed (309). Clinically, Individual differences in HSP buffering are also associated with vulnerability to heart disease (310, 311) and the likelihood of extreme longevity (312, 313).

Throughout life, HSP concentrations are sensitive to tissue damage or destruction; concentrations are higher in patients with cardiovascular and autoimmune disease (314). HSP are induced in response to brain pathology; e.g., stroke, neurodegenerative disease, epilepsy, and trauma. One in particular, hsp90, is constitutively expressed in brain throughout life (315) and is especially abundant in limbic system-related structures such as the hippocampus (316). It is necessary for efficient neurotransmitter release at the presynaptic terminal and the development of receptors in the post-synaptic membrane (317).

The decline of HSP response with aging may be a cause of neurodegenerative disease (318). If unfolded or misfolded proteins are not recognized by HSP, they are capable of forming aggregates (319). Conversely, a vigorous HSP response reduces amyloid production (320, 321) and inhibits the aggregation of tau protein (322), alpha-synuclein (323), and huntingtin (324, 325). It induces the clearance of aggregates by autophagy (323, 326, 327).

Studies of HSP in autism and schizophrenia show higher levels of hsp70 (279, 328, 329) and increased auto-antibodies to hsp60, 70, and 90 and CRP40, a catecholamine-regulated heat-shock-like protein (33–278, 278–328, 328–334). In neural stem cells from schizophrenic patients, there is higher variability in the levels of HSF1, an HSP transcription factor (307). Interestingly, HSP levels are elevated in patients with temporal lobe epilepsy and lupus who are also psychotic, but not in those who are not (287, 307, 335).

The cited studies are by no means definitive. They are compromised by inconsistent findings and results that may be affected by the medications patients are taking. Nevertheless, it seems to be a promising area to pursue. Perhaps there is an intrinsic or acquired weakness in normal neuroprotective mechanisms in autism and schizophrenia (307, 328). Perhaps, too, the proper approach to disorders arising from the interactions of so many genes and genomic variants might dwell in the "hubs" (291, 295) and "bow-ties" (336, 337) that reside along the trajectory from genotype to phenotype, and represented by the heat-shock proteins (327).

Variable Buffering

The approach may be promising but it won't be easy. HSP are an extended family of more than 100 proteins, traditionally identified by their molecular weight (from 8 to 110 kDa). Each category includes multiple proteins, many of which have multiple isoforms, not all of which are easily measurable (338–340). Individual HSP vary considerably in their expression, protein structure, localization, and ability to be induced. Complicating matters, the buffering capacity of an individual HSP is guided by multiple "co-chaperones" (291).

The behavior of HSP is not consistent. Hsp90, for example, protects the cell from genetic variation; or it may have the opposite effect, rescuing proteins that arise from mutations with folding or stability defects (341). Thus, they may maintain mutated proteins in a partially active state, permitting them to persist within the cell, to aggregate or cause some other mischief (327).

Buffering mechanisms also operate at different levels among the multiple components that contribute to every polygenic trait (342). The subunits that participate in a complex trait—single genes, or sets of well-integrated genes—may be robust or not. Subunits most likely to be robust are ones with high mutation rates, often at the expense of reduced robustness of genes or subunits with lower mutation rates. This may pose constraints or lead to conflicts that influence the buffering of the unit as a whole (256).

Nor do HSP operate in a vacuum. Whether a chaperone acts as a buffer, lessening mutational effects, or as a potentiator, increasing mutational effects, is not a fixed property of the protein itself but is influenced by the different mutations with which it interacts (343). The actions of HSP are also influenced by their genetic background (157, 344, 345). We don't know much about the genes that participate in the heat-shock response in humans, but 59 genes (7 positive activators and 52 negative regulators) participate in the heat-shock response in *Caenorhabditis elegans* (346). In *Saccharomyces cerevisiae*, no fewer than a thousand genes have been identified that alter sensitivity to heat shock (347).

The thresholds of robustness mechanisms are not only *a priori* variable, but are influenced by epigenetic changes that occur throughout one's life and that occurred in generations past (172, 283, 293, 301, 309, 348–350). Genetic variation present in one generation can influence phenotypic traits in the next, even if individuals do not inherit the variation. The environment experienced by one generation can influence phenotypic variation in the next several (351).

The critical balance between stability and variability, therefore, hovers on the fine edge of criticality. The cell has developed an exquisite system to prevent or mitigate destabilizing events and thus preserve its integrity and that of the organism. Understanding and possibly manipulating the agents that buffer cells from destabilizing agents holds at least some promise as a therapeutic approach to complex phenotypes like cancer and autoimmune disease (352–355). It may be a frail reed in the face of the complexity of genomic behavior but is one that is close to hand.

If Things Were Simple, Word Would Have Gotten Around

Individual differences in buffering capacity and genetic background are two of many elements that contribute to the variable and unpredictable expression of clinical phenotypes. Is it possible, therefore, to make useful predictions about the phenotypes of individuals from their complete genome sequences? The "typical" phenotypic outcome of an individual's genome may well be predictable, but it is much more difficult to predict the actual outcome for a particular individual. Individual outcomes are no more than probabilistic. The genome, like the neural connectome, exists in a metastable state, a critical, narrow edge between order and chaos. In such states, irreducible uncertainties and unexpected hazards are always present. The premises of "personalized medicine" look more like Laplace's demon every day.

The problem, for a personalized medicine demon, is that the relation of genotype and the phenotype of complex traits is decidedly non-linear. Non-linearity is characterized by sudden changes in phenotype with small changes in genotype; thus, not all genes are equally correlated with the trait whose ontogeny they control (356). Non-linearities are a ubiquitous feature of development and gene expression networks (357–364). Non-linearities are given to sudden discontinuities and can lapse, unexpectedly, into catastrophe (365–367).

Non-linearity is a bother to demons but is necessary for robustness to co-exist with evolvability. In model genotypes with low levels of non-linearity, stability is the rule; those with non-linear dynamics allow expression levels to be robust to small perturbations, while generating high diversity under larger perturbations; that is, evolvability (368, 369).

Genetic variation influences the phenotype by processes that act at different scales, times, and locations within the organism (182, 364, 370). The trajectory from genotype to phenotype, therefore, is not only complicated, it is complex. At every step, one is confronted with complex systems composed of parts that are complex systems in their own right and complex systems tend to show surprising and unexpected behavior. The behavior of every system is affected by interactions, direct and indirect, with all the others. The individual and his destiny can't be understood in terms of one or two such sub-systems, or even all of them together.

The individual components of complex systems interact in manifold ways, including highly dynamic regulatory and feedback mechanisms (371). Within this framework, a single cause can produce multiple and unpredictable effects and even small fluctuations can have unexpected consequences. Linear casual explanations—that conceive reality as a linear succession of elementary events from cause to effect—are usually unable to describe how complex systems behave (372, 373).

One struggles, therefore, to generate clinical insights from the information we have gleaned from years of study of our two most important adaptive systems:

- Complex systems may never be given to complete descriptions, or unchanging and non-provisional rules to control their behavior (374).
- We still know frustratingly little about how changes in genotype determine the changes in phenotype (341).
- The "sequence space" of the genome is so vast that an exhaustive functional mapping and characterization of epistasis for any protein or gene is nigh-on impossible (375).
- The genetic information we are lacking about traits and diseases is potentially immense (376).
- The phenotype of each individual is usually considered as an interaction between two variables: the genes each individual

carries and the environment that they experience… (but) the evidence suggests this is not always the case (351).
- It is futile to seek the basis of autism and schizophrenia amidst the profound complexity and variability of genetic and neural networks. Our present nosological constructs are imperfect, only "umbrella terms" that comprise a heterogenous mix of "real" conditions (377, 378).
- The complexity of brain function and structure is not reflected in current psychiatric disease nosology (238).

If such were really the case, we would be left with nothing but truisms: complex systems like the genome and neural connectome are "intrinsically and irreducibly hazardous" (45). Or rhetorical flourishes: "The same "genes" that drive us mad have made us human" (379). I prefer to think that our "umbrella terms," fuzzy sets as they are, may be the best way to translate the complexities of genomics and connectomics in a meaningful way. Mental disorders, I think, are as variable and mutable as the neural networks and the genome whence they arise. They are as variable and unpredictable as people are.

I prefer to leave the reader with more than an intellectual dead-end. Appreciating the complexity of the genome is an opportunity to revise our expectations of how it generates complex disease and how we may address the crucial issue of prevention.

About genes:

- The genome is not a static blueprint but a dynamic participant in the affairs of the cell.
- There is more to genomic exploration than SNPs and QTLs.
- What matters about the genome is not only its base-pair sequence but its behavior, especially its interactions with the environment, which is mediated by proteins and RNAs of various stripe.
- Genomic expression is intrinsically noisy, stochastic, and unpredictable.
- The dynamic and mutable nature of gene expression may be a source of species adaptability but it is also a potential source of individual vulnerability.

About mutations:

- They occur frequently, especially in the human genome.
- Mutations occur more often in specific regions of the genome and are more likely to occur in some individuals.
- Mutations that are only slightly deleterious (or beneficial) are subject to weak selection (380).
- Therefore, mutations accumulate, a reservoir of potential adaptions for the species but, on occasion, of disastrous events in the lives of individuals (381).

About phenotypes:

- From transcription to RNA processing, translation, and protein folding and all the way up to protein activity and cellular fitness, there are many layers of biological organization where the effects of mutation can be transformed (382).
- To make predictions about the phenotypes of individuals, it is clear that knowledge of genome sequencing is usually insufficient. Rather, we need to consider how genetic, environmental and stochastic variation, together with transgenerational effects, combine to determine the phenotypes of individuals (351).
- The processes that govern the trajectory from phenotype to genotype hover on a critical edge between stability and variability.
- The spectrum of outcomes can be related to continuous functions in variable elements.
- Catastrophic outcomes may be related to non-linear functions.

Complex traits, including complex diseases, are attributable to "multiple genes of small effect"-–among which are included all the sources of genomic, proteomic and neuronal variability. Nevertheless, complex traits exist and although quite variable, they occur with sufficient regularity to allow reliable descriptions and exhibit behavior that is more-or-less consistent. It is this, rather than their variability, that should entertain our interest. A current explanation is that mutations occurring in many different forms and at hundreds of targets converge on a much smaller number of molecular, cellular and anatomical pathways critical to the development and functioning of the CNS (383–385). The assumption is that aberrations in the expression of a great many genes affect a much smaller number of developmental pathways. It reflects a characteristic of complex systems: hierarchical networks are more concentrated higher up the scale. Although we suspect that developmental pathways are fewer in number than the genes involved, this is merely an assumption (360). However, we know that gene expression is processed through a much smaller number of "bow tie" processes that govern the generation of phenotypes and that maintain robustness without compromising evolvability. They deserve critical analysis.

At the level of pure theory, we may entertain a speculation about how primate evolution and then evolution of the hominids was so successful in the face of such a high degree of variability—to a point where even mechanisms designed to buffer it from perturbations are as variable and unpredictable. Ironically, an answer may be gleaned from computer simulations and studies of evolution among cancer cells and bacteria. From such studies, we learn that reproductive fitness is not the only goal of natural selection, nor is adaptation. Populations with lower initial fitness systematically adapt more rapidly than populations with higher initial fitness. Genotypes with lower fitness are more adaptable, over the long run, than those with higher fitness (381, 386–388).

"It is tempting to suggest that the promiscuity inherent in biology is tolerated with minimal detriment rather than corrected at high cost" (389). The promiscuity inherent in biology is not only tolerated, it is put to good use. If anything, it prevents a species from achieving a "genetic optimum." The hydra have achieved a genetic optimum, I suppose, and so have crocodilians and turtles. But their optimum is really only a genetic ceiling. Hominids have achieved adaptability, not a genetic optimum. As a species, we are fit but not perfect. Adaptability is not a perfect fit to one's niche but the capacity to accommodate a wide range of potential environments. It is the genomic flexibility that allows small and large changes to be made in response to change. To that end we have evolved the trait of evolvability, which is manifest

in variability at every point on the genotype-to-phenotype map. Trait evolvability, however, exists in balance with trait robustness. The balance is not perfectly balanced, however. It is distributed unevenly. It confers stability to most of us, albeit in graded fashion; less to many and very little to an unfortunate few.

AUTHOR'S NOTE

My hypothesis is that evolvability is relevant to the problems of psychiatric genetics: the problem of missing heritability, for example; of the variable expression of neurodevelopmental disorders arising from a particular genetic aberration; and the intriguing problem of autism and schizophrenia, disorders that are highly heritable but persist in spite of conferring low reproductive success. Genomic variability drove the evolution of neural complexity; it was also fuel for the evolvability of the hominid lineage. The paper is a review of mechanisms of genomic variability that are well-developed, scientifically, and also clinically relevant. The genetic elements that conferred variability and evolvability happen to be over-represented in autism and schizophrenia. The relevance of evolvability to neuropsychiatric disorders is illustrated by the robustness of the human genome, its ability to maintain stability in the face of genomic variability; yet it is also uniquely evolvable. The balance between robustness and evolvability is illustrated by buffering mechanisms that are structural or dynamic. Both, however, are fragile and subject to a high degree of inter-individual variation. The paper introduces a novel way to think about the genetics of neuropsychiatric disorders in particular, complex traits and complex disease in general.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This work was funded by North Carolina Neuropsychiatry, PA.

REFERENCES

1. Keller MC, Miller G. Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best? *Behav Brain Sci.* (2006) 29:385–404. doi: 10.1017/S0140525X06009095
2. Dobzhansky T. Nothing in biology makes sense except in the light of evolution. *Am Biol Teach.* (1975) 35:125–9.
3. Rosenberg RE, Law JK, Yenokyan G, McGready J, Kaufmann WE, Law PA. Characteristics and concordance of autism spectrum disorders among 277 Twin Pairs. *Arch Pediatr Adolesc Med.* (2009) 163:907–14. doi: 10.1001/archpediatrics.2009.98
4. McClellan J, King MC. Genetic heterogeneity in human disease. *Cell.* (2010) 141:210–17. doi: 10.1016/j.cell.2010.03.032
5. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature.* (2014) 511:421–7. doi: 10.1038/nature13595
6. Bray NJ, O'Donovan MC. The genetics of neuropsychiatric disorders. *Brain Neurosci Adv.* (2019) 2:2398212818799271. doi: 10.1177/2398212818799271
7. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature.* (2009) 461:747–53. doi: 10.1038/nature08494
8. Goldstein DB. Common genetic variation and human traits. *N Engl J Med.* (2009) 360:1696–8. doi: 10.1056/NEJMp0806284
9. Nadeau JH. Transgenerational genetic effects on phenotypic variation and disease risk. *Hum Mol Genet.* (2009) 18:R202–10. doi: 10.1093/hmg/ddp366
10. Doornbos M, Sikkema-Raddatz B, Ruijvenkamp CAL, Dijkhuizen T, Bijlsma EK, Gijsbers ACJ, et al. Nine patients with a microdeletion 15q11.2 between breakpoints 1 and 2 of the prader-willi critical region, possibly associated with behavioural disturbances. *Eur J Med Genet.* (2009) 52:108–15. doi: 10.1016/j.ejmg.2009.03.010
11. Cook EH Jr, Scherer SW. Copy-number variations associated with neuropsychiatric conditions. *Nature.* (2008) 455:919–23. doi: 10.1038/nature07458
12. Sebat J, Levy DL, McCarthy SE. Rare structural variants in schizophrenia: one disorder, multiple mutations; one mutation, multiple disorders. *Trends Genet.* (2009) 25:528–35. doi: 10.1016/j.tig.2009.10.004
13. Lichtenstein P, Yip BH, Björk C, Pawitan Y, Cannon TD, Sullivan PF, et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet.* (2009) 373:234–9. doi: 10.1016/S0140-6736(09)60072-6
14. Maier W, Lichtermann D, Minges J, Hallmayer J, Heun R, Benkert O, et al. Continuity and discontinuity of affective disorders and schizophrenia: results of a controlled family study. *Arch Gen Psychiatry.* (1993) 50:871–83. doi: 10.1001/archpsyc.1993.01820230041004
15. Cardno AG, Rijsdijk FV, Sham PC, Murray RM, McGuffin P. A twin study of genetic relationships between psychotic symptoms. *AJP.* (2002) 159:539–45. doi: 10.1176/appi.ajp.159.4.539
16. Steinhausen HC. The heterogeneity of causes and courses of attention-deficit/hyperactivity disorder. *Acta Psychiatr Scand.* (2009) 120:392–9. doi: 10.1111/j.1600-0447.2009.01446.x
17. Mortensen PB, Pedersen MG, Pedersen CB. Psychiatric family history and schizophrenia risk in Denmark: which mental disorders are relevant? *Psychol Med.* (2010) 40:201–10. doi: 10.1017/S0033291709990419
18. Lieberman JA. Is schizophrenia a neurodegenerative disorder? A clinical and neurobiological perspective. *Biol Psychiatry.* (1999) 46:729–39.
19. Rapoport JL, Addington AM, Frangou S, Psych MRC. The neurodevelopmental model of schizophrenia: update 2005. *Mol Psychiatry.* (2005) 10:434–49. doi: 10.1038/sj.mp.4001642
20. Fatemi SH, Folsom TD. The neurodevelopmental hypothesis of schizophrenia, revisited. *Schizophr Bull.* (2009) 35:528–48. doi: 10.1093/schbul/sbn187
21. Powell SB. Models of neurodevelopmental abnormalities in schizophrenia. *Curr Top Behav Neurosci.* (2010) 4:435–81. doi: 10.1007/7854_2010_57
22. Ross CA, Margolis RL, Reading SAJ, Pletnikov M, Coyle JT. Neurobiology of schizophrenia. *Neuron.* (2006) 52:139–53. doi: 10.1016/j.neuron.2006.09.015
23. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA.* (2009) 106:9362–7. doi: 10.1073/pnas.0903103106
24. Kirschner M, Gerhart J. Evolvability. *Proc Natl Acad Sci USA.* (1998) 95:8420–7.
25. Feder ME. Evolvability of physiological and biochemical traits: evolutionary mechanisms including and beyond single-nucleotide mutation. *J Exp Biol.* (2007) 210:1653–60. doi: 10.1242/jeb.02725
26. Draghi J, Wagner GP. The evolutionary dynamics of evolvability in a gene network model. *J Evol Biol.* (2009) 22:599–611. doi: 10.1111/j.1420-9101.2008.01663.x
27. Jones AG, Arnold SJ, Bürger R. The mutation matrix and the evolution of evolvability. *Evolution.* (2007) 61:727–45. doi: 10.1111/j.1558-5646.2007.00071.x

28. Gommans WM, Mullen SP, Maas S. RNA editing: a driving force for adaptive evolution? *Bioessays.* (2009) 31:1137–45. doi: 10.1002/bies.200900045
29. Pigliucci M. Is evolvability evolvable? *Nat Rev Genet.* (2008) 9:75–82. doi: 10.1038/nrg2278
30. Wills C. *The Runaway Brain: The Evolution of Human Uniqueness*. New York, NY: Basic Books (1993).
31. Polimeni J, Reiss JP. Evolutionary perspectives on schizophrenia. *Can J Psychiatry.* (2003) 48:34–9. doi: 10.1177/070674370304800107
32. Dodgson G, Gordon S. Avoiding false negatives: are some auditory hallucinations an evolved design flaw? *Behav Cogn Psychother.* (2009) 37:325–34. doi: 10.1017/S1352465809005244
33. Carter M, Watts CA. Possible biological advantages among schizophrenics' relatives. *BJP.* (1971) 118:453–60. doi: 10.1192/bjp.118.545.453
34. Spyropoulos B. Tay-sachs carriers and tuberculosis resistance. *Nature.* (1988) 331:666. doi: 10.1038/331666a0
35. Nettle D, Clegg H. Schizotypy, creativity and mating success in humans. *Proc R Soc B.* (2006) 273:611–15. doi: 10.1098/rspb.2005.3349
36. Kyaga S, Lichtenstein P, Boman M, Hultman C, Långström N, Landén M. Creativity and mental disorder: family study of 300 000 people with severe mental disorder. *Br J Psychiatry.* (2011) 199:373–9. doi: 10.1192/bjp.bp.110.085316
37. Power RA, Steinberg S, Bjornsdottir G, Rietveld CA, Abdellaoui A, Nivard MM, et al. Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nat Neurosci.* (2015) 18:953–5. doi: 10.1038/nn.4040
38. Baron-Cohen S. The hyper-systemizing, assortative mating theory of autism. *Prog Neuropsychopharmacol Biol Psychiatry.* (2006) 30:865–72. doi: 10.1016/j.pnpbp.2006.01.010
39. Ploeger A, Galis F. Evolutionary approaches to autism- an overview and integration. *Mcgill J Med.* (2011) 13:28.
40. MacCabe JH, Koupil I, Leon DA. Lifetime reproductive output over two generations in patients with psychosis and their unaffected siblings: the Uppsala 1915-1929 Birth cohort multigenerational study. *Psychol Med.* (2009) 39:1667–76. doi: 10.1017/S0033291709005431
41. Laursen TM, Munk-Olsen T. Reproductive patterns in psychotic patients. *Schizophr Res.* (2010) 121:234–40. doi: 10.1016/j.schres.2010.05.018
42. Bundy H, Stahl D, MacCabe JH. A systematic review and meta-analysis of the fertility of patients with schizophrenia and their unaffected relatives. *Acta Psychiatr Scand.* (2011) 123:98–106. doi: 10.1111/j.1600-0447.2010.01623.x
43. Oksenberg N, Stevison L, Wall JD, Ahituv N. Function and regulation of AUTS2, a gene implicated in autism and human evolution. *PLoS Genet.* (2013) 9:e1003221. doi: 10.1371/journal.pgen.1003221
44. Power RA, Kyaga S, Uher R, MacCabe JH, Långström N, Landen M, et al. Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA Psychiatry.* (2013) 70:22–30. doi: 10.1001/jamapsychiatry.2013.268
45. Cook RI. How complex systems fail. *Cognitive Technologies Laboratory.* University of Chicago: Chicago IL. (1998). Available online at: https://www.researchgate.net/profile/Richard_Cook3/publication/228797158_How_complex_systems_fail/links/0c96053410db96a89c000000.pdf (accessed April 6, 2017).
46. Randall PL. Schizophrenia, abnormal connection, and brain evolution. *Med Hypoth.* (1983) 10:247–80. doi: 10.1016/0306-9877(83)90114-7
47. Crow TJ. Is schizophrenia the price that Homo sapiens pays for language? *Schizophr Res.* (1997) 28:127–41.
48. Burns JK. An evolutionary theory of schizophrenia: cortical connectivity, metarepresentation, and the social brain. *Behav Brain Sci.* (2004) 27:831–55. doi: 10.1017/s0140525x04000196
49. Dumas L, Sikela JM. DUF1220 domains, cognitive disease, and human brain evolution. *Cold Spring Harb Symp Quant Biol.* (2009) 74:375–82. doi: 10.1101/sqb.2009.74.025
50. Sikela JM, Searles Quick VB. Genomic trade-offs: are autism and schizophrenia the steep price of the human brain? *Hum Genet.* (2018) 137:1–13. doi: 10.1007/s00439-017-1865-9
51. Book JA. Schizophrenia as a gene mutation. *Acta Genet Stat Med.* (1953) 4:133–9.
52. Penrose LS. Mutation in man. *Acta Genet Stat Med.* (1956) 6:169–82.
53. McClellan JM, Susser E, King MC. Schizophrenia: a common disease caused by multiple rare alleles. *Br J Psychiatry.* (2007) 190:194–9. doi: 10.1192/bjp.bp.106.025585
54. Rubinsztein DC, Amos B, Cooper G. Microsatellite and trinucleotide-repeat evolution: evidence for mutational bias and different rates of evolution in different lineages. *Philos Trans R Soc Lond B Biol Sci.* (1999) 354:1095–9. doi: 10.1098/rstb.1999.0465
55. Burnashev N, Rozov A. Genomic control of receptor function. *Cell Mol Life Sci.* (2000) 57:1499–507. doi: 10.1007/pl00000634
56. Liu G, Program NCS, Zhao S, Bailey JA, Sahinalp SC, Alkan C, et al. Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. *Genome Res.* (2003) 13:358–68. doi: 10.1101/gr.923303
57. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-scale copy number polymorphism in the human genome. *Science.* (2004) 305:525–8. doi: 10.1126/science.1098918
58. Enard W, Khaitovich P, Klose J, Zöllner S, Heissig F, Giavalisco P, et al. Intra- and interspecific variation in primate gene expression patterns. *Science.* (2002) 296:340–3. doi: 10.1126/science.1068996
59. Antonell A, de Luis O, Domingo-Roura X, Pérez-Jurado LA. Evolutionary mechanisms shaping the genomic structure of the williams-beuren syndrome chromosomal region at human 7q11.23. *Genome Res.* (2005) 15:1179–88. doi: 10.1101/gr.3944605
60. Bailey JA, Eichler EE. Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat Rev Genet.* (2006) 7:552–64. doi: 10.1038/nrg1895
61. Nesse RM. Cliff-edged fitness functions and the persistence of schizophrenia. *Behav Brain Sci.* (2004) 27:6.
62. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet.* (2001) 69:124–37. doi: 10.1086/321272
63. Ng MYM, Levinson DF, Faraone SV, Suarez BK, DeLisi LE, Arinami T, et al. Meta-analysis of 32 genome-wide linkage studies of schizophrenia. *Mol Psychiatry.* (2009) 14:774–85. doi: 10.1038/mp.2008.135
64. van Dongen J, Boomsma DI. The evolutionary paradox and the missing heritability of schizophrenia. *Am J Med Genet B Neuropsychiatr Genet.* (2013) 162B:122–36. doi: 10.1002/ajmg.b.32135
65. Keller MC. Evolutionary perspectives on genetic and environmental risk factors for psychiatric disorders. *Annu Rev Clin Psychol.* (2018) 14:471–93. doi: 10.1146/annurev-clinpsy-050817-084854
66. Risch NJ. Searching for genetic determinants in the new millennium. *Nature.* (2000) 405:847–56. doi: 10.1038/35015718
67. Hinds DA. Whole-genome patterns of common DNA variation in three human populations. *Science.* (2005) 307:1072–9. doi: 10.1126/science.1105436
68. Uher R, Zwicker A. Etiology in psychiatry: embracing the reality of poly-gene-environmental causation of mental illness. *World Psychiatry.* (2017) 16:121–9. doi: 10.1002/wps.20436
69. Rees E, Moskvina V, Owen MJ, O'Donovan MC, Kirov G. *De novo* rates and selection of schizophrenia-associated copy number variants. *Biol Psychiatry.* (2011) 70:1109–14. doi: 10.1016/j.biopsych.2011.07.011
70. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science.* (2007) 315:848–53. doi: 10.1126/science.1136678
71. Beckmann JS, Estivill X, Antonarakis SE. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat Rev Genet.* (2007) 8:639–46. doi: 10.1038/nrg2149
72. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev.* (2009) 19:212–19. doi: 10.1016/j.gde.2009.04.010
73. The International HapMap Consortium. The international hapmap project. *Nature.* (2003) 426:789–96. doi: 10.1038/nature02168
74. Buchanan JA, Scherer SW. Contemplating effects of genomic structural variation. *Genet Med.* (2008) 10:639–47. doi: 10.1097GIM.0b013e318183f848
75. Marian AJ. molecular genetic studies of complex phenotypes. *Transl Res.* (2012) 159:64–79. doi: 10.1016/j.trsl.2011.08.001
76. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science.* (2001) 291:1304–51. doi: 10.1126/science.1058040

77. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature.* (2001) 409:860–921. doi: 10.1038/35057062
78. de Koning APJ, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* (2011) 7:e1002384. doi: 10.1371/journal.pgen.1002384
79. Budworth H, McMurray CT. A brief history of triplet repeat diseases. *Methods Mol Biol.* (2013) 1010:3–17. doi: 10.1007/978-1-62703-411-1_1
80. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. *Nat Genet.* (2004) 36:949–51. doi: 10.1038/ng1416
81. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, et al. The diploid genome sequence of an individual human. *PLoS Biol.* (2007) 5:e254. doi: 10.1371/journal.pbio.0050254
82. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature.* (2008) 452:872–6. doi: 10.1038/nature06884
83. Nahon JL. Birth of "human-specific" genes during primate evolution. *Genetica.* (2003) 118:193–208. doi: 10.1023/a:1024157714736
84. Ohno S, Wolf U, Atkin NB. Evolution from fish to mammals by gene duplication. *Hereditas.* (2009) 59:169–87. doi: 10.1111/j.1601-5223.1968.tb02169.x
85. Morrow EM. genomic copy number variation in disorders of cognitive development. *J Am Acad Child Adolesc Psychiatry.* (2010) 49:1091–104. doi: 10.1016/j.jaac.2010.08.009
86. Yunis JJ, Prakash O. The origin of man: a chromosomal pictorial legacy. *Science.* (1982) 215:1525–30.
87. Jauch A, Wienberg J, Stanyon R, Arnold N, Tofanelli S, Ishida T, et al. Reconstruction of genomic rearrangements in great apes and gibbons by chromosome painting. *Proc Natl Acad Sci USA.* (1992) 89:8611–15.
88. Häsler J, Strub K. Alu elements as regulators of gene expression. *Nucleic Acids Res.* (2006) 34:5491–7. doi: 10.1093/nar/gkl706
89. Dierssen M, Herault Y, Estivill X. Aneuploidy: from a physiological mechanism of variance to down syndrome. *Physiol Rev.* (2009) 89:887–920. doi: 10.1152/physrev.00032.2007
90. Lupski JR. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.* (1998) 14:417–22.
91. Stankiewicz P, Lupski JR. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* (2002) 18:74–82. doi: 10.1016/s0168-9525(02)02592-1
92. Kondrashov AS. Direct estimates of human per nucleotide mutation rates at 20 loci causing mendelian diseases. *Hum Mutat.* (2003) 21:12–27. doi: 10.1002/humu.10147
93. Aravind L, Anantharaman V, Venancio TM. Apprehending multicellularity: regulatory networks, genomics, and evolution. *Birth Defects Res C Embryo Today.* (2009) 87:143–64. doi: 10.1002/bdrc.20153
94. Biémont C. A brief history of the status of transposable elements: from junk DNA to major players in evolution: figure 1.—. *Genetics.* (2010) 186:1085–93. doi: 10.1534/genetics.110.124180
95. Hurles ME, Dermitzakis ET, Tyler-Smith C. The functional impact of structural variation in humans. *Trends Genet.* (2008) 24:238–45. doi: 10.1016/j.tig.2008.03.001
96. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. *Nature.* (2010) 464:704–12. doi: 10.1038/nature08516
97. Bushman DM, Chun J. The genomically mosaic brain: aneuploidy and more in neural diversity and disease. *Semin Cell Dev Biol.* (2013) 24:357–69. doi: 10.1016/j.semcdb.2013.02.003
98. Arendt T, Mosch B, Morawski M. Neuronal aneuploidy in health and disease: a cytomic approach to understand the molecular individuality of neurons. *Int J Mol Sci.* (2009) 10:1609–27. doi: 10.3390/ijms10041609
99. Torres EM, Williams BR, Amon A. Aneuploidy: cells losing their balance. *Genetics.* (2008) 179:737–46. doi: 10.1534/genetics.108.090878
100. Sheltzer JM, Torres EM, Dunham MJ, Amon A. Transcriptional consequences of aneuploidy. *Proc Natl Acad Sci USA.* (2012) 109:12644–9. doi: 10.1073/pnas.1209227109
101. Blaschke AJ, Staley K, Chun J. Widespread programmed cell death in proliferative and postmitotic regions of the fetal cerebral cortex. *Development.* (1996) 122:1165–74.
102. Kingsbury MA, Friedman B, McConnell MJ, Rehen SK, Yang AH, Kaushal D, et al. Aneuploid neurons are functionally active and integrated into brain circuitry. *Proc Natl Acad Sci USA.* (2005) 102:6143–7. doi: 10.1073/pnas.0408171102
103. Faggioli F, Vijg J, Montagna C. Chromosomal aneuploidy in the aging brain. *Mech Ageing Dev.* (2011) 132:429–36. doi: 10.1016/j.mad.2011.04.008
104. Westra JW, Rivera RR, Bushman DM, Yung YC, Peterson SE, Barral S, et al. Neuronal DNA content variation (DCV) with regional and individual differences in the human brain. *J Comp Neurol.* (2010) 518:3981–4000. doi: 10.1002/cne.22436
105. Kaushal D, Contos JJA, Treuner K, Yang AH, Kingsbury MA, Rehen SK, et al. Alteration of gene expression by chromosome loss in the postnatal mouse brain. *J Neurosci.* (2003) 23:5599–606. doi: 10.1523/JNEUROSCI.23-13-05599.2003
106. Rehen SK, Yung YC, McCreight MP, Kaushal D, Yang AH, Almeida BSV, et al. Constitutional aneuploidy in the normal human brain. *J Neurosci.* (2005) 25:2176–80. doi: 10.1523/JNEUROSCI.4560-04.2005
107. Peterson SE, Yang AH, Bushman DM, Westra JW, Yung YC, Barral S, et al. Aneuploid cells are differentially susceptible to caspase-mediated death during embryonic cerebral cortical development. *J Neurosci.* (2012) 32:16213–22. doi: 10.1523/JNEUROSCI.3706-12.2012
108. Verheijen BM, Vermulst M, van Leeuwen FW. Somatic mutations in neurons during aging and neurodegeneration. *Acta Neuropathol.* (2018) 135:811–26. doi: 10.1007/s00401-018-1850-y
109. Yang Y, Herrup K. Cell division in the CNS: protective response or lethal event in post-mitotic neurons? *Biochim Biophys Acta.* (2007) 1772:457–66. doi: 10.1016/j.bbadis.2006.10.002
110. Bedrosian TA, Quayle C, Novaresi N, Gage FH. Early life experience drives structural variation of neural genomes in mice. *Science.* (2018) 359:1395–9. doi: 10.1126/science.aah3378
111. Song S, Gleeson JG. Early life experience shapes neural genome. *Science.* (2018) 359:1330–1. doi: 10.1126/science.aat3977
112. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong association of *de novo* copy number mutations with autism. *Science.* (2007) 316:445–9. doi: 10.1126/science.1138659
113. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science.* (2008) 320:539–43. doi: 10.1126/science.1155174
114. Jacquemont M-L, Sanlaville D, Redon R, Raoul O, Cormier-Daire V, Lyonnet S, et al. Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. *J Med Genet.* (2006) 43:843–9. doi: 10.1136/jmg.2006.043166
115. Burmeister M, McInnis MG, Zöllner S. Psychiatric genetics: progress amid controversy. *Nat Rev Genet.* (2008) 9:527–40. doi: 10.1038/nrg2381
116. Chubb JE, Bradshaw NJ, Soares DC, Porteous DJ, Millar JK. The DISC locus in psychiatric illness. *Mol Psychiatry.* (2008) 13:36–64. doi: 10.1038/sj.mp.4002106
117. International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature.* (2008) 455:237–41. doi: 10.1038/nature07239
118. Stefansson H, Rujescu D, Cichon S, Pietiläinen OPH, Ingason A, Steinberg S, et al. Large recurrent microdeletions associated with schizophrenia. *Nature.* (2008) 455:232–6. doi: 10.1038/nature07229
119. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, et al. Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet.* (2009) 41:1223–7. doi: 10.1038/ng.474
120. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, et al. Functional impact of global rare copy number variation in autism spectrum disorder. *Nature.* (2010) 466:368–72. doi: 10.1038/nature09146
121. Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of *de novo* copy number mutations with sporadic schizophrenia. *Nat. Genet.* (2008) 40:880–5. doi: 10.1038/ng.162

122. Kirov G. CNVs in neuropsychiatric disorders. *Hum Mol Genet.* (2015) 24:R45–49. doi: 10.1093/hmg/ddv253
123. Mitchell KJ, Porteous DJ. Rethinking the genetic architecture of schizophrenia. *Psychol Med.* (2011) 41:19–32. doi: 10.1017/S003329171000070X
124. Szatkiewicz JP, O'Dushlaine C, Chen G, Chambert K, Moran JL, Neale BM, et al. Copy number variation in schizophrenia in Sweden. *Mol Psychiatry.* (2014) 19:762–73. doi: 10.1038/mp.2014.40
125. Guilmatre A, Dubourg C, Mosca AL, Legallic S, Goldenberg A, Drouin-Garraud V, et al. Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. *Arch Gen Psychiatry.* (2009) 66:947–56. doi: 10.1001/archgenpsychiatry.2009.80
126. Chow ML, Pramparo T, Winn ME, Barnes CC, Li HR, Weiss L, et al. Age-dependent brain gene expression and copy number anomalies in autism suggest distinct pathological processes at young versus mature ages. *PLoS Genet.* (2012) 8:e1002592. doi: 10.1371/journal.pgen.1002592
127. Hosak L, Silhan P, Hosakova J. Genomic copy number variations: a breakthrough in our knowledge on schizophrenia etiology? *Neuro Endocrinol Lett.* (2012) 33:183–90.
128. Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell.* (2012) 148:1223–41. doi: 10.1016/j.cell.2012.02.039
129. Rees E, Kirov G, O'Donovan MC, Owen MJ. *De novo* mutation in schizophrenia. *Schizophr Bull.* (2012) 38:377–81. doi: 10.1093/schbul/sbs047
130. Ionita-Laza I, Rogers AJ, Lange C, Raby BA, Lee C. Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. *Genomics.* (2009) 93:22–6. doi: 10.1016/j.ygeno.2008.08.012
131. Loeb LA. Human cancers express mutator phenotypes: origin, consequences and targeting. *Nat Rev Cancer.* (2011) 11:450–7. doi: 10.1038/nrc3063
132. Chanock S, Manolio T, Boehnke M, Hunter D, Thomas G. Replicating genotype–phenotype associations. *Nature.* (2007) 447:655–60. doi: 10.1038/447655a
133. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet.* (2006) 7:85–97. doi: 10.1038/nrg1767
134. Zhao X, Leotta A, Kustanovich V, Lajonchere C, Geschwind DH, Law K, et al. A unified genetic theory for sporadic and inherited autism. *Proc Natl Acad Sci USA.* (2007) 104:12831–6. doi: 10.1073/pnas.0705803104
135. Ben-Shachar S, Lanpher B, German JR, Qasaymeh M, Potocki L, Nagamani SCS, et al. Microdeletion 15q13.3: a locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders. *J Med Genet.* (2009) 46:382–8. doi: 10.1136/jmg.2008.064378
136. Beunders G, Kamp JM van de, Veenhoven RH, Hagen JM van, Nieuwint AWM, Sistermans EA. A triplication of the williams–beuren syndrome region in a patient with mental retardation, a severe expressive language delay, behavioural problems and dysmorphisms. *J Med Genet.* (2010) 47:271–5. doi: 10.1136/jmg.2009.070490
137. Levitt P, Campbell DB. The genetic and neurobiologic compass points toward common signaling dysfunctions in autism spectrum disorders. *J Clin Invest.* (2009) 119:747–54. doi: 10.1172/JCI37934
138. Schaaf CP, Sabo A, Sakai Y, Crosby J, Muzny D, Hawes A, et al. Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders. *Hum Mol Genet.* (2011) 20:3366–75. doi: 10.1093/hmg/ddr243
139. Bassett AS, Chow EWC, Husted J, Weksberg R, Caluseriu O, Webb GD, et al. Clinical features of 78 adults with 22q11 deletion syndrome. *Am J Med Genet A.* (2005) 138:307–13. doi: 10.1002/ajmg.a.30984
140. Kirov G, Rees E, Walters JTR, Escott-Price V, Georgieva L, Richards AL, et al. The penetrance of copy number variations for schizophrenia and developmental delay. *Biol Psychiatry.* (2014) 75:378–85. doi: 10.1016/j.biopsych.2013.07.022
141. Vassos E, Collier DA, Holden S, Patch C, Rujescu D, St Clair D, et al. Penetrance for copy number variants associated with schizophrenia. *Hum Mol Genet.* (2010) 19:3477–81. doi: 10.1093/hmg/ddq259
142. Rosenfeld JA, Coe BP, Eichler EE, Cuckle H, Shaffer LG. Estimates of penetrance for recurrent pathogenic copy-number variations. *Genet Med.* (2013) 15:478–81. doi: 10.1038/gim.2012.164
143. Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet.* (2013) 132:1077–130. doi: 10.1007/s00439-013-1331-2
144. Brunetti-Pierri N, Berg JS, Scaglia F, Belmont J, Bacino CA, Sahoo T, et al. Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat Genet.* (2008) 40:1466–71. doi: 10.1038/ng.279
145. Mefford HC, Sharp AJ, Baker C, Itsara A, Jiang Z, Buysse K, et al. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med.* (2008) 359:1685–99. doi: 10.1056/NEJMoa0805384
146. Helbig I, Mefford HC, Sharp AJ, Guipponi M, Fichera M, Franke A, et al. 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. *Nat Genet.* (2009) 41:160–162. doi: 10.1038/ng.292
147. Dibbens LM, Mullen S, Helbig I, Mefford HC, Bayly MA, Bellows S, et al. Familial and sporadic 15q13.3 microdeletions in idiopathic generalized epilepsy: precedent for disorders with complex inheritance. *Hum Mol Genet.* (2009) 18:3626–31. doi: 10.1093/hmg/ddp311
148. Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA, et al. Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet.* (2006) 38:1038–42. doi: 10.1038/ng1862
149. Gothelf D, Presburger G, Levy D, Nahmani A, Burg M, Berant M, et al. Genetic, developmental, and physical factors associated with attention deficit hyperactivity disorder in patients with velocardiofacial syndrome. *Am J Med Genet B Neuropsychiatr Genet.* (2004) 126B:116–21. doi: 10.1002/ajmg.b.20144
150. Lee JA, Lupski JR. Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron.* (2006) 52:103–21. doi: 10.1016/j.neuron.2006.09.027
151. Gothelf D, Feinstein C, Thompson T, Gu E, Penniman L, Van Stone E, et al. Risk factors for the emergence of psychotic disorders in adolescents with 22q11.2 deletion syndrome. *Am J Psychiatry.* (2007) 164:663–9. doi: 10.1176/appi.ajp.164.4.663
152. Kobrynski LJ, Sullivan KE. Velocardiofacial syndrome, DiGeorge syndrome: the chromosome 22q11.2 deletion syndromes. *Lancet.* (2007) 370:1443–52. doi: 10.1016/S0140-6736(07)61601-8
153. Ramocki MB, Zoghbi HY. Failure of neuronal homeostasis results in common neuropsychiatric phenotypes. *Nature.* (2008) 455:912–18. doi: 10.1038/nature07457
154. Ousley O, Rockers K, Dell ML, Coleman K, Cubells JF. A review of neurocognitive and behavioral profiles associated with 22q11 deletion syndrome: implications for clinical evaluation and treatment. *Curr Psychiatry Rep.* (2007) 9:148–58. doi: 10.1007/s11920-007-0085-8
155. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, et al. Autism Consortium. Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med.* (2008) 358:667–75. doi: 10.1056/NEJMoa075974
156. de Visser JAGM, Cooper TF, Elena SF. The causes of epistasis. *Proc R Soc B Biol Sci.* (2011) 278:3617–24. doi: 10.1098/rspb.2011.1537
157. Suliman R, Ben-David E, Shifman S. Chromatin regulators, phenotypic robustness, and autism risk. *Front Genet.* (2014) 5:81. doi: 10.3389/fgene.2014.00081
158. Breen MS, Kemena C, Vlasov PK, Notredame C, Kondrashov FA. Epistasis as the primary factor in molecular evolution. *Nature.* (2012) 490:535–8. doi: 10.1038/nature11510
159. Weinreich DM, Watson RA, Chao L. Perspective: sign epistasis and genetic costraint on evolutionary trajectories. *Evolution.* (2005) 59:1165–74. doi: 10.1111/j.0014-3820.2005.tb01768.x
160. Bateson W, Saunders E, Punnett R, Hurst C. *Reports to the Evolution Committee of the Royal Society, Report II*. London: Harrison & Sons (1905)
161. Weinreich DM, Lan Y, Wylie CS, Heckendorn RB. Should evolutionary geneticists worry about higher-order epistasis? *Curr Opin Genet Dev.* (2013) 23:700–7. doi: 10.1016/j.gde.2013.10.007
162. Sailer ZR, Harms MJ. Detecting high-order epistasis in nonlinear genotype-phenotype maps. *Genetics.* (2017) 205:1079–88. doi: 10.1534/genetics.116.195214
163. Sackman AM, Rokyta DR. Additive phenotypes underlie epistasis of fitness effects. *Genetics.* (2018) 208:339–48. doi: 10.1534/genetics.117.300451

164. Remold SK, Lenski RE. Pervasive joint influence of epistasis and plasticity on mutational effects in *Escherichia coli*. *Nat Genet.* (2004) 36:423–6. doi: 10.1038/ng1324
165. Hoekstra LA, Julick CR, Mika KM, Montooth KL. Energy demand and the context-dependent effects of genetic interactions underlying metabolism. *Evol Lett.* (2018) 2:102–13. doi: 10.1002/evl3.47
166. Senner NR, Conklin JR, Piersma T. An ontogenetic perspective on individual differences. *Proc Biol Sci.* (2015) 282:1050. doi: 10.1098/rspb.2015.1050
167. Xu L, Barker B, Gu Z. Dynamic epistasis for different alleles of the same gene. *Proc Natl Acad Sci USA.* (2012) 109:10420–5. doi: 10.1073/pnas.1121507109
168. Inoue K, Dewar K, Katsanis N, Reiter LT, Lander ES, Devon KL, et al. The 1.4-Mb CMT1A duplication/HNPP deletion genomic region reveals unique genome architectural features and provides insights into the recent evolution of new genes. *Genome Res.* (2001) 11:1018–33. doi: 10.1101/gr.180401
169. Eichler EE, Nickerson DA, Altshuler D, Bowcock AM, Brooks LD, Carter NP, et al. Completing the map of human genetic variation. *Nature.* (2007) 447:161–5. doi: 10.1038/447161a
170. Geschwind DH. Autism: many genes, common pathways? *Cell.* (2008) 135:391–5. doi: 10.1016/j.cell.2008.10.016
171. Carvalho CMB, Zhang F, Lupski JR. Genomic disorders: a window into human gene and genome evolution. *Proc Natl Acad Sci USA.* (2010) 107(Suppl. 1):1765–71. doi: 10.1073/pnas.0906222107
172. Raj A, Rifkin SA, Andersen E, van Oudenaarden A. Variability in gene expression underlies incomplete penetrance. *Nature.* (2010) 463:913–18. doi: 10.1038/nature08781
173. Mattick JS, Gagen MJ. The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol Biol Evol.* (2001) 18:1611–30. doi: 10.1093/oxfordjournals.molbev.a003951
174. Potapova TA, Zhu J, Li R. Aneuploidy and chromosomal instability: a vicious cycle driving cellular evolution and cancer genome chaos. *Cancer Metastasis Rev.* (2013) 32:377–89. doi: 10.1007/s10555-013-9436-6
175. Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PA, Monlong J, Rivas MA, Gonzàlez-Porta M, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature.* (2013) 501:506–11. doi: 10.1038/nature12531
176. Cheung VG, Conlin LK, Weber TM, Arcaro M, Jen K-Y, Morley M, et al. Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat Genet.* (2003) 33:422–5. doi: 10.1038/ng1094
177. Deutsch S, Lyle R, Dermitzakis ET, Attar H, Subrahmanyan L, Gehrig C, et al. Gene expression variation and expression quantitative trait mapping of human chromosome 21 genes. *Hum Mol Genet.* (2005) 14:3741–9. doi: 10.1093/hmg/ddi404
178. Hodgins-Davis A, Rice DP, Townsend JP. Gene expression evolves under a house-of-cards model of stabilizing selection. *Mol Biol Evol.* (2015) 32:2130–40. doi: 10.1093/molbev/msv094
179. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell.* (2017) 169:1177–86. doi: 10.1016/j.cell.2017.05.038
180. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc Natl Acad Sci USA.* (2006) 103:17973–8. doi: 10.1073/pnas.0605938103
181. Erwin DH, Davidson EH. The evolution of hierarchical gene regulatory networks. *Nat Rev Genet.* (2009) 10:141–8. doi: 10.1038/nrg2499
182. Crombach A, Hogeweg P. Evolution of evolvability in gene regulatory networks. *PLoS Computat Biol.* (2008) 4:e1000112. doi: 10.1371/journal.pcbi.1000112
183. Papo D, Zanin M, Pineda-Pardo JA, Boccaletti S, Buldú JM. Functional brain networks: great expectations, hard times and the big leap forward. *Philos Trans R Soc B Biol Sci.* (2014) 369:20130525. doi: 10.1098/rstb.2013.0525
184. Jones PA, Takai D. The role of DNA methylation in mammalian epigenetics. *Science.* (2001) 293:1068–70. doi: 10.1126/science.1063852
185. Jablonka E, Lamb MJ. The changing concept of epigenetics. *Ann N Y Acad Sci.* (2006) 981:82–96. doi: 10.1111/j.1749-6632.2002.tb04913.x
186. Tsankova N, Renthal W, Kumar A, Nestler EJ. Epigenetic regulation in psychiatric disorders. *Nat Rev Neurosci.* (2007) 8:355–67. doi: 10.1038/nrn2132
187. Feinberg AP, Irizarry RA. Evolution in health and medicine Sackler colloquium: stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci USA.* (2010) 107(Suppl. 1):1757–64. doi: 10.1073/pnas.0906183107
188. Bollati V, Galimberti D, Pergoli L, Dalla Valle E, Barretta F, Cortini F, et al. DNA methylation in repetitive elements and alzheimer disease. *Brain Behav Immun.* (2011) 25:1078–83. doi: 10.1016/j.bbi.2011.01.017
189. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* (2013) 14:R115. doi: 10.1186/gb-2013-14-10-r115
190. Gentilini D, Mari D, Castaldi D, Remondini D, Ogliari G, Ostan R, et al. Role of epigenetics in human aging and longevity: genome-wide DNA methylation profile in centenarians and centenarians' offspring. *Age.* (2013) 35:1961–73. doi: 10.1007/s11357-012-9463-1
191. Durso DF, Bacalini MG, Sala C, Pirazzini C, Marasco E, Bonafé M, et al. Acceleration of leukocytes' epigenetic age as an early tumor and sex-specific marker of breast and colorectal cancer. *Oncotarget.* (2017) 8:23237–45. doi: 10.18632/oncotarget.15573
192. Lee JT. Epigenetic regulation by long noncoding RNAs. *Science.* (2012) 338:1435–9. doi: 10.1126/science.1231776
193. Ng S-Y, Johnson R, Stanton LW. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J.* (2012) 31:522–33. doi: 10.1038/emboj.2011.459
194. Alcazar RM, Lin R, Fire AZ. Transmission dynamics of heritable silencing induced by double-stranded RNA in caenorhabditis elegans. *Genetics.* (2008) 180:1275–88. doi: 10.1534/genetics.108.089433
195. Jablonka E, Raz G. Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Q Rev Biol.* (2009) 84:131–76. doi: 10.1086/598822
196. Nelson VR, Nadeau JH. Transgenerational genetic effects. *Epigenomics.* (2010) 2:797–806. doi: 10.2217/epi.10.57
197. Jablonka E, Lamb M. *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. Cambridge: The MIT Press (2005).
198. Skinner MK. Environmental epigenetics and a unified theory of the molecular aspects of evolution: a neo-lamarckian concept that facilitates neo-darwinian evolution. *Genome Biol Evol.* (2015) 7:1296–302. doi: 10.1093/gbe/evv073
199. O'Dea RE, Noble DWA, Johnson SL, Hesselson D, Nakagawa S. The role of non-genetic inheritance in evolutionary rescue: epigenetic buffering, heritable bet hedging and epigenetic traps. *Environ Epigenet.* (2016) 2:dvv014. doi: 10.1093/eep/dvv014
200. Petronis A. The origin of schizophrenia: genetic thesis, epigenetic antithesis, and resolving synthesis. *Biol Psychiatry.* (2004) 55:965–70. doi: 10.1016/j.biopsych.2004.02.005
201. Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature.* (2007) 447:433–440. doi: 10.1038/nature05919
202. Wong CCY, Meaburn EL, Ronald A, Price TS, Jeffries AR, Schalkwyk LC, et al. Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Mol Psychiatry.* (2014) 19:495–503. doi: 10.1038/mp.2013.41
203. Li P, Lam J, Shu Z. On the transient and steady-state estimates of interval genetic regulatory networks. *IEEE Trans Syst Man Cybern B Cybern.* (2010) 40:336–49. doi: 10.1109/TSMCB.2009.2022402
204. Kadonaga JT. Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell.* (2004) 116:247–57. doi: 10.1016/S0092-8674(03)01078-X
205. Rosenfeld S. Mathematical descriptions of biochemical networks: stability, stochasticity, evolution. *Prog Biophys Mol Biol.* (2011) 106:400–9. doi: 10.1016/j.pbiomolbio.2011.03.003
206. Raser JM, O'Shea EK. Noise in gene expression: origins, consequences, and control. *Science.* (2005) 309:2010–13. doi: 10.1126/science.1105891
207. van Bokhoven H. Genetic and epigenetic networks in intellectual disabilities. *Annu Rev Genet.* (2011) 45:81–104. doi: 10.1146/annurev-genet-110410-132512
208. Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D. Rare *de novo* variants associated with autism implicate a large functional network

of genes involved in formation and function of synapses. *Neuron.* (2011) 70:898–907. doi: 10.1016/j.neuron.2011.05.021
209. Kou Y, Betancur C, Xu H, Buxbaum JD, Ma'ayan A. Network- and attribute-based classifiers can prioritize genes and pathways for autism spectrum disorders and intellectual disability. *Am J Med Genet C Semin Med Genet.* (2012) 160C:130–42. doi: 10.1002/ajmg.c.31330
210. Gilman SR, Chang J, Xu B, Bawa TS, Gogos JA, Karayiorgou M, et al. Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. *Nat Neurosci.* (2012) 15:1723–8. doi: 10.1038/nn.3261
211. Zoghbi HY, Bear MF. Synaptic dysfunction in neurodevelopmental disorders associated with autism and intellectual disabilities. *Cold Spring Harb Perspect Biol.* (2012) 4:a009886. doi: 10.1101/cshperspect.a009886
212. Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, et al. Great ape genetic diversity and population history. *Nature.* (2013) 499:471–5. doi: 10.1038/nature12228
213. Marques-Bonet T, Eichler EE. The evolution of human segmental duplications and the core duplicon hypothesis. *Cold Spring Harb Symp Quant Biol.* (2009) 74:355–62. doi: 10.1101/sqb.2009.74.011
214. Inoue K, Lupski JR. Molecular mechanisms for genomic disorders. *Annu Rev Genomics Hum Genet.* (2002) 3:199–242. doi: 10.1146/annurev.genom.3.032802.120023
215. Samonte RV, Eichler EE. Segmental duplications and the evolution of the primate genome. *Nat Rev Genet.* (2002) 3:65–72. doi: 10.1038/nrg705
216. Muller GB. Evo–devo: extending the evolutionary synthesis. *Nat Rev Genet.* (2007) 8:943–9. doi: 10.1038/nrg2219
217. Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature.* (2005) 437:69–87. doi: 10.1038/nature04072
218. Dorus S, Vallender EJ, Evans PD, Anderson JR, Gilbert SL, Mahowald M, et al. Accelerated evolution of nervous system genes in the origin of homo sapiens. *Cell.* (2004) 119:1027–40. doi: 10.1016/j.cell.2004.11.040
219. Ho MW. How development directs evolution. In: *Developmental Systems Theory & Methodology*. New York, NY: Guildford Press (2013). p. 131–53.
220. Hernando-Herraez I, Prado-Martinez J, Garg P, Fernandez-Callejo M, Heyn H, Hvilsom C, et al. Dynamics of DNA methylation in recent human and great ape evolution. *PLoS Genet.* (2013) 9:e1003763. doi: 10.1371/journal.pgen.1003763
221. Nowell PC. The clonal evolution of tumor cell populations. *Science.* (1976) 194:23–8. doi: 10.1126/science.959840
222. Aktipis CA, Nesse RM. Evolutionary foundations for cancer biology. *Evol Appl.* (2013) 6:144–59. doi: 10.1111/eva.12034
223. Salk JJ, Fox EJ, Loeb LA. Mutational heterogeneity in human cancers: origin and consequences. *Annu Rev Pathol.* (2010) 5:51–75. doi: 10.1146/annurev-pathol-121808-102113
224. Sjöblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science.* (2006) 314:268–74. doi: 10.1126/science.1133427
225. Wood LD, Parsons DW, Jones S, Lin J, Sjöblom T, Leary RJ, et al. The genomic landscapes of human breast and colorectal cancers. *Science.* (2007) 318:1108–13. doi: 10.1126/science.1145720
226. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, et al. Patterns of somatic mutation in human cancer genomes. *Nature.* (2007) 446:153–8. doi: 10.1038/nature05610
227. Konkel MK, Batzer MA. A mobile threat to genome stability: the impact of non-LTR retrotransposons upon the human genome. *Semin Cancer Biol.* (2010) 20:211–21. doi: 10.1016/j.semcancer.2010.03.001
228. Tian T, Olson S, Whitacre JM, Harding A. The origins of cancer robustness and evolvability. *Integr Biol.* (2011) 3:17–30. doi: 10.1039/C0IB00046A
229. Lee J-K, Choi Y-L, Kwon M, Park PJ. Mechanisms and consequences of cancer genome instability: lessons from genome sequencing studies. *Annu Rev Pathol.* (2016) 11:283–312. doi: 10.1146/annurev-pathol-012615-044446
230. Sniegowski PD, Gerrish PJ, Johnson T, Shaver A. The evolution of mutation rates: separating causes from consequences. *BioEssays.* (2000) 22:1057–66. doi: 10.1002/1521-1878(200012)22:12<1057::AID-BIES3>3.0.CO;2-W
231. Fitzgerald DM, Hastings PJ, Rosenberg SM. Stress-induced mutagenesis: implications in cancer and drug resistance. *Annu Rev Cancer Biol.* (2017) 1:119–40. doi: 10.1146/annurev-cancerbio-050216-121919
232. Ram Y, Hadany L. Stress-induced mutagenesis and complex adaptation. *Proc R Soc B Biol Sci.* (2014) 281 :20141025. doi: 10.1098/rspb.2014.1025
233. Rosenberg SM, Queitsch C. Combating evolution to fight disease. *Science.* (2014) 343:1088–9. doi: 10.1126/science.1247472
234. Colnaghi R, Carpenter G, Volker M, O'Driscoll M. The consequences of structural genomic alterations in humans: genomic disorders, genomic instability and cancer. *Semin Cell Dev Biol.* (2011) 22:875–85. doi: 10.1016/j.semcdb.2011.07.010
235. Loeb LA. Mutator phenotype may be required for multistage carcinogenesis. *Cancer Reasearcg.* (1991) 51:6.
236. Leigh Egbert Giles. Natural selection and mutability. *Am Nat.* (1970) 104:301–5. doi: 10.1086/282663
237. Insel TR. Rethinking schizophrenia. *Nature.* (2010) 468:187–93. doi: 10.1038/nature09552
238. Geschwind DH, Flint J. Genetics and genomics of psychiatric disease. *Science.* (2015) 349:1489–94. doi: 10.1126/science.aaa8954
239. Hebb DO. *The Organization of Behaviour: A Neuropsychological Theory*. New York, NY: John Wiley & Sons (1949).
240. Gómez-Robles A, Hopkins WD, Schapiro SJ, Sherwood CC. Relaxed genetic control of cortical organization in human brains compared with chimpanzees. *Proc Natl Acad Sci USA.* (2015) 112:14799–804. doi: 10.1073/pnas.1512646112
241. Adams P. Hebb and darwin. *J Theoretical Biol.* (1998) 195:419–38. doi: 10.1006/jtbi.1997.0620
242. Edelman GM. Neural Darwinism: Selection and reentrant signaling in higher brain function. *Neuron.* (1993) 10:115–25. doi: 10.1016/0896-6273(93)90304-A
243. Fernando CT, Szathmary E, Husbands P. Selectionist and evolutionary approaches to brain function: a critical appraisal. *Front Comput Neurosci.* (2012) 6:24. doi: 10.3389/fncom.2012.00024
244. Vos DD, Bruggeman FJ, Westerhoff HV, Bakker BM. How molecular competition influences fluxes in gene expression networks. *PLoS ONE.* (2011) 6:e28494. doi: 10.1371/journal.pone.0028494
245. Karreth FA, Pandolfi PP. ceRNA crosstalk in cancer: when ce-bling rivalries go awry. *Cancer Discov.* (2013) 3:1113–21. doi: 10.1158/2159-8290.CD-13-0202
246. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. *Nature.* (2014) 505:344–352. doi: 10.1038/nature12986
247. Tiesinga PHE, Sejnowski TJ. Rapid temporal modulation of synchrony by competition in cortical interneuron networks. *Neural Comput.* (2004) 16:251–75. doi: 10.1162/089976604322742029
248. Cho RY, Konecky RO, Carter CS. Impairments in frontal cortical gamma synchrony and cognitive control in schizophrenia. *Proc Natl Acad Sci USA.* (2006) 103:19878–83. doi: 10.1073/pnas.0609440103
249. Rashid AJ, Yan C, Mercaldo V, Hsiang HL, Park S, Cole CJ, et al. Competition between engrams influences fear memory formation and recall. *Science.* (2016) 353:383–7. doi: 10.1126/science.aaf0594
250. Beck DM, Kastner S. Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Res.* (2009) 49:1154–65. doi: 10.1016/j.visres.2008.07.012
251. Buffelli M, Burgess RW, Feng G, Lobe CG, Lichtman JW, Sanes JR. Genetic evidence that relative synaptic efficacy biases the outcome of synaptic competition. *Nature.* (2003) 424:430–4. doi: 10.1038/nature01844
252. Kasthuri N, Lichtman JW. The role of neuronal identity in synaptic competition. *Nature.* (2003) 424:426–30. doi: 10.1038/nature01836
253. Belmonte MK, Bourgeron T. Fragile X syndrome and autism at the intersection of genetic and neural networks. *Nat Neurosci.* (2006) 9:1221–5. doi: 10.1038/nn1765
254. Holland J. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Application to Biology*. Cambridge: MIT Press (1992).
255. Whitacre JM, Bender A. Networked buffering: a basic mechanism for distributed robustness in complex adaptive systems. *Theoretical Biol Med Modell.* (2010) 7:20. doi: 10.1186/1742-4682-7-20
256. Visser JAGM de, Hermisson J, Wagner GP, Meyers LA, Bagheri-Chaichian H, Blanchard JL, et al. Perspective: evolution and detection of genetic robustness. *Evolution.* (2003) 57:1959–72. doi: 10.1111/j.0014-3820.2003.tb00377.x

257. Lenski RE, Barrick JE, Ofria C. Balancing robustness and evolvability. *PLoS Biol.* (2006) 4:e428. doi: 10.1371/journal.pbio.0040428
258. Waddington CH. Canalization of development and the inheritance of acquired characters. *Nature.* (1942) 150:563–5. doi: 10.1038/150563a0
259. Gibson G, Wagner G. Canalization in evolutionary genetics: a stabilizing theory? *BioEssays.* (2000) 22:372–80. doi: 10.1002/(SICI)1521-1878(200004)22:4<372::AID-BIES7>3.0.CO;2-J
260. Payne JL, Wagner A. The causes of evolvability and their evolution. *Nat Rev Genet.* (2019) 20:24–38. doi: 10.1038/s41576-018-0069-z
261. Holling CS. Understanding the complexity of economic, ecological, and social systems. *Ecosystems.* (2001) 4:390–405. doi: 10.1007/s10021-001-0101-5
262. Kitano H. Biological robustness. *Nat Rev Genet.* (2004) 5:826–37. doi: 10.1038/nrg1471
263. Kitano H. Towards a theory of biological robustness. *Mol Syst Biol.* (2007) 3:137. doi: 10.1038/msb4100179
264. Whitacre JM. Biological robustness: paradigms, mechanisms, and systems principles. *Front Genet.* (2012) 3:67. doi: 10.3389/fgene.2012.00067
265. Aerts H, Fias W, Caeyenberghs K, Marinazzo D. Brain networks under attack: robustness properties and the impact of lesions. *Brain.* (2016) 139:3063–83. doi: 10.1093/brain/aww194
266. Wagner A. Robustness and evolvability: a paradox resolved. *Proc R Soc B.* (2008) 275:91–100. doi: 10.1098/rspb.2007.1137
267. Stelling J, Sauer U, Szallasi Z, Doyle FJ, Doyle J. Robustness of cellular functions. *Cell.* (2004) 118:675–85. doi: 10.1016/j.cell.2004.09.008
268. Welch JJ, Waxman D. Modularity and the cost of complexity. *Evolution.* (2003) 57:1723–34. doi: 10.1111/j.0014-3820.2003.tb00581.x
269. Friedlander T, Mayo AE, Tlusty T, Alon U. Evolution of bow-tie architectures in biology. *PLoS Computat Biol.* (2015) 11:e1004055. doi: 10.1371/journal.pcbi.1004055
270. Sporns O. The non-random brain: efficiency, economy, and complex dynamics. *Front Comput Neurosci.* (2011) 5:5. doi: 10.3389/fncom.2011.00005
271. Paaby AB, Rockman MV. Cryptic genetic variation, evolution's hidden substrate. *Nat Rev Genet.* (2014) 15:247–58. doi: 10.1038/nrg3688
272. Ibáñez-Marcelo E, Alarcón T. The topology of robustness and evolvability in evolutionary systems with genotype–phenotype map. *J Theoretical Biol.* (2014) 356:144–62. doi: 10.1016/j.jtbi.2014.04.014
273. Gibson G. Decanalization and the origin of complex disease. *Nat Rev Genet.* (2009) 10:134–40. doi: 10.1038/nrg2502
274. McGrath JJ, Hannan AJ, Gibson G. Decanalization, brain development and risk of schizophrenia. *Transl Psychiatry.* (2011) 1:e14. doi: 10.1038/tp.2011.16
275. Queitsch C, Carlson KD, Girirajan S. Lessons from model organisms: phenotypic robustness and missing heritability in complex disease. *PLoS Genet.* (2012) 8:e1003041. doi: 10.1371/journal.pgen.1003041
276. Melkonyan HS, Ushakova TE, Umansky SR. Hsp70 gene expression in mouse lung cells upon chronic γ-irradiation. *Int J Radiat Biol.* (1995) 68:277–80. doi: 10.1080/09553009514551201
277. Bates PR, Hawkins A, Mahadik SP, McGrath JJ. Heat stress lipids and schizophrenia. *Prostaglandins Leukot Essent Fatty Acids.* (1996) 55:101–7. doi: 10.1016/s0952-3278(96)90153-2
278. Kim JJ, Lee SJ, Toh KY, Lee CU, Lee C, Paik IH. Identification of antibodies to heat shock proteins 90 kDa and 70 kDa in patients with schizophrenia. *Schizophrenia Research.* (2001) 52:127–35. doi: 10.1016/S0920-9964(00)00091-8
279. Kowalczyk M, Owczarek A, Suchanek R, Paul-Samojedny M, Fila-Danilow A, Borkowska P, et al. Heat shock protein 70 gene polymorphisms are associated with paranoid schizophrenia in the Polish population. *Cell Stress Chaperones.* (2014) 19:205–15. doi: 10.1007/s12192-013-0446-7
280. Rutherford SL. From genotype to phenotype: buffering mechanisms and the storage of genetic information. *BioEssays.* (2000) 22:1095–105. doi: 10.1002/1521-1878(200012)22:12<1095::AID-BIES7>3.0.CO;2-A
281. Lindquist S, Craig EA. The heat-shock proteins. *Annu Rev Genet.* (1988) 22:631–77. doi: 10.1146/annurev.ge.22.120188.003215
282. Jarosz DF, Lindquist S. Hsp90 and environmental stress transform the adaptive value of natural genetic variation. *Science.* (2010) 330:1820–4. doi: 10.1126/science.1195487
283. Queitsch C, Sangster TA, Lindquist S. Hsp90 as a capacitor of phenotypic variation. *Nature.* (2002) 417:618–24. doi: 10.1038/nature749
284. Feder ME, Hofmann GE. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annu Rev Physiol.* (1999) 61:243–82. doi: 10.1146/annurev.physiol.61.1.243
285. Baker JD, Ozsan I, Rodriguez Ospina S, Gulick D, Blair LJ. Hsp90 heterocomplexes regulate steroid hormone receptors: from stress response to psychiatric disease. *Int J Mol Sci.* (2019) 20:79. doi: 10.3390/ijms20010079
286. Riggs DL, Roberts PJ, Chirillo SC, Cheung-Flynn J, Prapapanich V, Ratajczak T, et al. The Hsp90-binding peptidylprolyl isomerase FKBP52 potentiates glucocorticoid signaling *in vivo*. *EMBO J.* (2003) 22:1158–67. doi: 10.1093/emboj/cdg108
287. Kandratavicius L, Hallak JE, Carlotti CG, Assirati JA, Leite JP. Hippocampal expression of heat shock proteins in mesial temporal lobe epilepsy with psychiatric comorbidities and their relation to seizure outcome. *Epilepsia.* (2014) 55:1834–43. doi: 10.1111/epi.12787
288. Criado-Marrero M, Rein T, Binder EB, Porter JT, Koren J, Blair LJ. Hsp90 and FKBP51: complex regulators of psychiatric diseases. *Philos Trans R Soc B Biol Sci.* (2018) 373:20160532. doi: 10.1098/rstb.2016.0532
289. Rea IM, McNerlan S, Pockley AG. Serum heat shock protein and anti-heat shock protein antibody levels in aging. *Exp Gerontol.* (2001) 36:341–52. doi: 10.1016/s0531-5565(00)00215-1
290. Proctor CJ, Soti C, Boys RJ, Gillespie CS, Shanley DP, Wilkinson DJ, et al. Modelling the actions of chaperones and their role in ageing. *Mech Ageing Dev.* (2005) 126:119–31. doi: 10.1016/j.mad.2004.09.031
291. Taipale M, Jarosz DF, Lindquist S. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat Rev Mol Cell Biol.* (2010) 11:515–28. doi: 10.1038/nrm2918
292. Rutherford SL, Lindquist S. Hsp90 as a capacitor for morphological evolution. *Nature.* (1998) 396:336–42. doi: 10.1038/24550
293. Sangster TA, Lindquist S, Queitsch C. Under cover: causes, effects and implications of Hsp90-mediated genetic capacitance. *BioEssays.* (2004) 26:348–62. doi: 10.1002/bies.20020
294. Tokuriki N, Tawfik DS. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature.* (2009) 459:668–73. doi: 10.1038/nature08009
295. Rutherford S, Knapp JR, Csermely P. Hsp90 and developmental networks. In: *Landes Bioscience.* Austin, TX (2013).
296. Kitami T, Nadeau JH. Biochemical networking contributes more to genetic buffering in human and mouse metabolic pathways than does gene duplication. *Nat Genet.* (2002) 32:191–4. doi: 10.1038/ng945
297. Hartman IV JL. Principles for the buffering of genetic variation. *Science.* (2001) 291:1001–4. doi: 10.1126/science.291.5506.1001
298. Yeyati PL, Bancewicz RM, Maule J, van Heyningen V. Hsp90 selectively modulates phenotype in vertebrate development. *PLoS Genet.* (2007) 3:e43. doi: 10.1371/journal.pgen.0030043
299. Specchia V, Piacentini L, Tritto P, Fanti L, D'Alessandro R, Palumbo G, et al. Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons. *Nature.* (2010) 463:662–5. doi: 10.1038/nature08739
300. Takahashi KH, Rako L, Takano-Shimizu T, Hoffmann AA, Lee SF. Effects of small Hspgenes on developmental stability and microenvironmental canalization. *BMC Evol Biol.* (2010) 10:284. doi: 10.1186/1471-2148-10-284
301. Sollars V, Lu X, Xiao L, Wang X, Garfinkel MD, Ruden DM. Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. *Nat Genet.* (2003) 33:70–4. doi: 10.1038/ng1067
302. Levy SF, Siegal ML. Network hubs buffer environmental variation in saccharomyces cerevisiae. *PLoS Biol.* (2008) 6:e264. doi: 10.1371/journal.pbio.0060264
303. Cowen LE, Lindquist S. Hsp90 potentiates the rapid evolution of new traits: drug resistance in diverse fungi. *Science.* (2005) 309:2185–9. doi: 10.1126/science.1118370
304. Davidson SM, Loones M-T, Duverger O, Morange M. The developmental expression of small HSP. In: Arrigo AP, Müller WEG, editors. *Small Stress Proteins. Progress in Molecular and Subcellular Biology*. Berlin: Springer (2002). p. 103–128. doi: 10.1007/978-3-642-56348-5_6

305. Christians ES, Zhou Q, Renard J, Benjamin IJ. Heat shock proteins in mammalian development. *Semin Cell Dev Biol.* (2003) 14:283–90. doi: 10.1016/j.semcdb.2003.09.021
306. Zabinsky RA, Mason GA, Queitsch C, Jarosz DF. It's not magic - Hsp90 and its effects on genetic and epigenetic variation. *Semin Cell Dev Biol.* (2019) 88:21–35. doi: 10.1016/j.semcdb.2018.05.015
307. Hashimoto-Torii K, Torii M, Fujimoto M, Nakai A, El Fatimy R, Mezger V, et al. Roles of heat shock factor 1 in neuronal response to fetal environmental risks and its relevance to brain disorders. *Neuron.* (2014) 82:560–72. doi: 10.1016/j.neuron.2014.03.002
308. Reed-Herbert K, Samali A, Gorman A. The role of heat shock proteins in neuronal differentiation and development. In: *Heat Shock Proteins in Neural Cells*. Berlin: Landes Bioscience and Springer Science + Business Media (2006). p. 25–37.
309. Burga A, Casanueva MO, Lehner B. Predicting mutation outcome from early stochastic variation in genetic interaction partners. *Nature.* (2011) 480:250–3. doi: 10.1038/nature10665
310. Pockley AG, Georgiades A, Thulin T, de Faire U, Frostegård J. Serum heat shock protein 70 levels predict the development of atherosclerosis in subjects with established hypertension. *Hypertension*. (2003) 42:235–38. doi: 10.1161/01.HYP.0000086522.13672.23
311. Li RH, Churchill GA. Epistasis contributes to the genetic buffering of plasma HDL cholesterol in mice. *Physiol Genomics.* (2010) 42A:228–34. doi: 10.1152/physiolgenomics.00044.2010
312. Clarke GM. The genetic basis of developmental stability. V. Inter- and intra-individual character variation. *Heredity.* (1998) 80:562–7. doi: 10.1046/j.1365-2540.1998.00294.x
313. Bergman A, Atzmon G, Ye K, MacCarthy T, Barzilai N. Buffering mechanisms in aging: a systems approach toward uncovering the genetic component of aging. *PLoS Computat Biol.* (2007) 3:e170. doi: 10.1371/journal.pcbi.0030170
314. Terry DF, McCormick M, Andersen S, Pennington J, Schoenhofen E, Palaima E, et al. Cardiovascular disease delay in centenarian offspring: role of heat shock proteins. *Ann N Y Acad Sci.* (2004) 1019:502–5. doi: 10.1196/annals.1297.092
315. D'Souza SM, Brown IR. Constitutive expression of heat shock proteins Hsp90, Hsc70, Hsp70 and Hsp60 in neural and non-neural tissues of the rat during postnatal development. *Cell Stress Chaperones.* (1998) 3:188–99. doi: 10.1379/1466-1268(1998)003<0188:ceohsp>2.3.co;2
316. Izumoto S, Herbert J. Widespread constitutive expression of HSP90 messenger RNA in rat brain. *J Neurosci Res.* (1993) 35:20–28. doi: 10.1002/jnr.490350104
317. Gerges NZ, Tran IC, Backos DS, Harrell JM, Chinkers M, Pratt WB, et al. Independent functions of hsp90 in neurotransmitter release and in the continuous synaptic cycling of AMPA receptors. *J Neurosci.* (2004) 24:4758–66. doi: 10.1523/JNEUROSCI.0594-04.2004
318. Calderwood SK, Murshid A, Prince T. The shock of aging: molecular chaperones and the heat shock response in longevity and aging – a mini-review. *GER.* (2009) 55:550–8. doi: 10.1159/000225957
319. Wilhelmus MMM, de Waal RMW, Verbeek MM. Heat shock proteins and amateur chaperones in amyloid-beta accumulation and clearance in alzheimer's disease. *Mol Neurobiol.* (2007) 35:203–16. doi: 10.1007/s12035-007-0029-7
320. Hoshino T, Nakaya T, Araki W, Suzuki K, Suzuki T, Mizushima T. Endoplasmic reticulum chaperones inhibit the production of amyloid-beta peptides. *Biochem J.* (2007) 402:581–9. doi: 10.1042/BJ20061318
321. Kumar P, Ambasta RK, Veereshwarayya V, Rosen KM, Kosik KS, Band H, et al. CHIP and HSPs interact with beta-APP in a proteasome-dependent manner and influence Abeta metabolism. *Hum Mol Genet.* (2007) 16:848–64. doi: 10.1093/hmg/ddm030
322. Salminen A, Ojala J, Kaarniranta K, Hiltunen M, Soininen H. Hsp90 regulates tau pathology through co-chaperone complexes in alzheimer's disease. *Prog Neurobiol.* (2011) 93:99–110. doi: 10.1016/j.pneurobio.2010.10.006
323. Riedel M, Goldbaum O, Schwarz L, Schmitt S, Richter-Landsberg C. 17-AAG induces cytoplasmic α-synuclein aggregate clearance by induction of autophagy. *PLoS ONE.* (2010) 5:e8753. doi: 10.1371/journal.pone.0008753
324. Sittler A, Lurz R, Lueder G, Priller J, Hayer-Hartl MK, Hartl FU, et al. Geldanamycin activates a heat shock response and inhibits huntingtin aggregation in a cell culture model of Huntington's disease. *Hum Mol Genet.* (2001) 10:1307–15. doi: 10.1093/hmg/10.12.1307
325. Waza M, Adachi H, Katsuno M, Minamiyama M, Tanaka F, Doyu M, et al. Modulation of Hsp90 function in neurodegenerative disorders: a molecular-targeted therapy against disease-causing protein. *J Mol Med.* (2006) 84:635–46. doi: 10.1007/s00109-006-0066-0
326. Peterson LB, Blagg BS. To fold or not to fold: modulation and consequences of Hsp90 inhibition. *Future Med Chem.* (2009) 1:267–83. doi: 10.4155/fmc.09.17
327. Koren J, Jinwal UK, Lee DC, Jones JR, Shults CL, Johnson AG, et al. Chaperone signalling complexes in alzheimer's disease. *J Cell Mol Med.* (2009) 13:619–30. doi: 10.1111/j.1582-4934.2008.00557.x
328. Schwarz MJ, Riedel M, Gruber R, Ackenheil M, Müller N. Antibodies to heat shock proteins in schizophrenic patients: implications for the mechanism of the disease. *Am J Psychiatry.* (1999) 156:1103–4. doi: 10.1176/ajp.156.7.1103
329. El-Ansary A, Al-Ayadhi L. Neuroinflammation in autism spectrum disorders. *J Neuroinflammation.* (2012) 9:265. doi: 10.1186/1742-2094-9-265
330. Evers M, Cunningham-Rundles C, Hollander E. Heat shock protein 90 antibodies in autism. *Mol Psychiatry.* (2002) 7:S26–8. doi: 10.1038/sj.mp.4001171
331. Schwarz MJ, Riedel M, Gruber R, Müller N, Ackenheil M. Autoantibodies against 60-kDa heat shock protein in schizophrenia. *Eur Arch Psychiatry Clin Neurosci.* (1998) 248:282–8.
332. Leykin I, Spivak B, Weizman A, Cohen IR, Shinitzky M. Elevated cellular immune response to human heat-shock protein-60 in schizophrenic patients. *Eur Arch Psychiatry Clin Neurosci.* (1999) 249:238–46.
333. Kilidireas K, Latov N, Strauss DH, Gorig AD, Hashim GA, Gorman JM, et al. Antibodies to the human 60 kDa heat-shock protein in patients with schizophrenia. *Lancet.* (1992) 340:569–72. doi: 10.1016/0140-6736(92)92105-o
334. Wang X-F, Wang D, Zhu W, Delrahim KK, Dolnak D, Rapaport MH. Studies characterizing 60 kda autoantibodies in subjects with schizophrenia. *Biol Psychiatry.* (2003) 53:361–35. doi: 10.1016/s0006-3223(02)01972-8
335. Latchman DS, Isenberg DA. The role of hsp90 in SLE. *Autoimmunity.* (1994) 19:211–18. doi: 10.3109/08916939408995696
336. Csete M, Doyle J. Bow ties, metabolism and disease. *Trends Biotechnol.* (2004) 22:446–450. doi: 10.1016/j.tibtech.2004.07.007
337. Tieri P, Grignolio A, Zaikin A, Mishto M, Remondini D, Castellani GC, et al. Network, degeneracy and bow tie. Integrating paradigms and architectures to grasp the complexity of the immune system. *Theor Biol Med Model.* (2010) 7:32. doi: 10.1186/1742-4682-7-32
338. Kregel KC. Invited review: heat shock proteins: modifying factors in physiological stress responses and acquired thermotolerance. *J Appl Physiol.* (2002) 92:2177–86. doi: 10.1152/japplphysiol.01267.2001
339. Hoter A, El-Sabban ME, Naim HY. The HSP90 family: structure, regulation, function, and implications in health and disease. *Int J Mol Sci.* (2018) 19:2560. doi: 10.3390/ijms19092560
340. Miller DJ, Fort PE. Heat shock proteins regulatory role in neurodevelopment. *Front Neurosci.* (2018) 12:821. doi: 10.3389/fnins.2018.00821
341. Jarosz DF, Taipale M, Lindquist S. Protein homeostasis and the phenotypic manifestation of genetic diversity: principles and mechanisms. *Annu Rev Genet.* (2010) 44:189–216. doi: 10.1146/annurev.genet.40.110405.090412
342. Hermisson J, Hansen TF, Wagner GP. Epistasis in polygenic traits and the evolution of genetic architecture under stabilizing selection. *Am Nat.* (2003) 161:708–34. doi: 10.1086/374204
343. Geiler-Samerotte K, Sartori FMO, Siegal ML. Decanalizing thinking on genetic canalization. *Semin Cell Dev Biol.* (2019) 88:54–66. doi: 10.1016/j.semcdb.2018.05.008
344. Takahashi KH. Multiple capacitors for natural genetic variation in Drosophila melanogaster. *Mol Ecol.* (2013) 22:1356–65. doi: 10.1111/mec.12091
345. Takahashi KH. Novel genetic capacitors and potentiators for the natural genetic variation of sensory bristles and their trait specificity in Drosophila melanogaster. *Mol Ecol.* (2015) 24:5561–72. doi: 10.1111/mec.13407

346. Guisbert E, Czyz DM, Richter K, McMullen PD, Morimoto RI. Identification of a tissue-selective heat shock response regulatory network. *PLoS Genet.* (2013) 9:e1003466. doi: 10.1371/journal.pgen.1003466
347. Jarolim S, Ayer A, Pillay B, Gee AC, Phrakaysone A, Perrone GG, et al. Saccharomyces cerevisiae genes involved in survival of heat shock. *G3.* (2013) 3:2321–33. doi: 10.1534/g3.113.007971
348. Peng JC, Karpen GH. Epigenetic regulation of heterochromatic DNA stability. *Curr Opin Genet Dev.* (2008) 18:204–11. doi: 10.1016/j.gde.2008.01.021
349. Peng JC, Karpen GH. Heterochromatic genome stability requires regulators of histone H3 K9 methylation. *PLoS Genet.* (2009) 5:e1000435. doi: 10.1371/journal.pgen.1000435
350. Grummt I. The nucleolus—guardian of cellular homeostasis and genome integrity. *Chromosoma.* (2013) 122:487–97. doi: 10.1007/s00412-013-0430-0
351. Burga A, Lehner B. Beyond genotype to phenotype: why the phenotype of an individual cannot always be predicted from their genome sequence and the environment that they experience. *FEBS J.* (2012) 279:3765–75. doi: 10.1111/j.1742-4658.2012.08810.x
352. Sharp S, Workman P. Inhibitors of the HSP90 molecular chaperone: current status. *Adv Cancer Res.* (2006) 95:323–48. doi: 10.1016/S0065-230X(06)95009-X
353. Rappa F, Farina F, Zummo G, David S, Campanella C, Carini F, et al. HSP-molecular chaperones in cancer biogenesis and tumor therapy: an overview. *Anticancer Res.* (2012) 32:5139–50.
354. Shrestha L, Bolaender A, Patel HJ, Taldone T. Heat shock protein (HSP) drug discovery and development: targeting heat shock proteins in disease. *Curr Top Med Chem.* (2016) 16:2753–64. doi: 10.2174/1568026616666160413141911
355. Hendriks LEL, Dingemans A-MC. Heat shock protein antagonists in early stage clinical trials for NSCLC. *Expert Opin Investig Drugs.* (2017) 26:541–50. doi: 10.1080/13543784.2017.1302428
356. Nijhout HF. The nature of robustness in development. *Bioessays.* (2002) 24:553–63. doi: 10.1002/bies.10093
357. Lander AD, Nie Q, Wan FYM. Do morphogen gradients arise by diffusion? *Dev Cell.* (2002) 2:785–96. doi: 10.1016/S1534-5807(02)00179-X
358. Polev D. Transcriptional noise as a driver of gene evolution. *J Theor Biol.* (2012) 293:27–33. doi: 10.1016/j.jtbi.2011.10.001
359. Torres-Sosa C, Huang S, Aldana M. Criticality is an emergent property of genetic networks that exhibit evolvability. *PLoS Comput Biol.* (2012) 8:e1002669. doi: 10.1371/journal.pcbi.1002669
360. Chakravarti A, Clark AG, Mootha VK. Distilling pathophysiology from complex disease genetics. *Cell.* (2013) 155:21–26. doi: 10.1016/j.cell.2013.09.001
361. Pechenick DA, Moore JH, Payne JL. The influence of assortativity on the robustness and evolvability of gene regulatory networks upon gene birth. *J Theor Biol.* (2013) 330:26–36. doi: 10.1016/j.jtbi.2013.03.019
362. Gonze D, Abou-Jaoudé W. The goodwin model: behind the Hill function. *PLoS ONE.* (2013) 8:e69573. doi: 10.1371/journal.pone.0069573
363. Frank TD, Cavadas MAS, Nguyen LK, Cheong A. Non-linear dynamics in transcriptional regulation: biological logic gates. In: Carballido-Landeira J, Escribano B, editors. *Nonlinear Dynamics in Biological Systems. SEMA SIMAI Springer Series*. Berlin: Springer International Publishing. (2016). p. 43–62. doi: 10.1007/978-3-319-33054-9_3
364. Green RM, Fish JL, Young NM, Smith FJ, Roberts B, Dolan K, et al. Developmental nonlinearity drives phenotypic robustness. *Nat Commun.* (2017) 8:1970. doi: 10.1038/s41467-017-02037-7
365. Barton S. Chaos, self-organization, and psychology. *Am Psychol.* (1994) 49:5–14.
366. Yates FE. Order and complexity in dynamical systems: Homeodynamics as a generalized mechanics for biology. *Math Comput Model.* (1994) 19:49–74.
367. Dekker S, Cilliers P, Hofmeyr JH. The complexity of failure: implications of complexity theory for safety investigations. *Saf Sci.* (2011) 49:939–45. doi: 10.1016/j.ssci.2011.01.008
368. Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL. Genetic properties influencing the evolvability of gene expression. *Science.* (2007) 317:118–21. doi: 10.1126/science.1140247
369. Steinacher A, Bates DG, Akman OE, Soyer OS. Nonlinear dynamics in gene regulation promote robustness and evolvability of gene expression levels. *PLoS ONE.* (2016) 11:e0153295. doi: 10.1371/journal.pone.0153295
370. Garfield DA, Runcie DE, Babbitt CC, Haygood R, Nielsen WJ, Wray GA. The impact of gene expression variation on the robustness and evolvability of a developmental gene regulatory network. *PLoS Biol.* (2013) 11:e1001696. doi: 10.1371/journal.pbio.1001696
371. Regenmortel MHVV. Reductionism and complexity in molecular biology. *EMBO Rep.* (2004) 5:1016–20. doi: 10.1038/sj.embor.7400284
372. Kellenberger E. The evolution of molecular biology. *EMBO Rep.* (2004) 5:546–9. doi: 10.1038/sj.embor.7400180
373. Mazzocchi F. Complexity in biology. *EMBO Rep.* (2008) 9:10–14. doi: 10.1038/sj.embor.7401147
374. Heylighen F, Cilliers P, Gershenson C. Complexity and philosophy. *arXiv[Preprint].arXiv:cs/0604072.* (2006). Available online at: https://arxiv.org/abs/cs/0604072 (accessed April 21, 2017).
375. Anderson DW, McKeown AN, Thornton JW. Intermolecular epistasis shaped the function and evolution of an ancient transcription factor and its DNA binding sites. *Elife.* (2015) 4:e7864. doi: 10.7554/eLife.07864
376. de Magalhaes JP, Wang J. The fog of genetics: Known unknowns and unknown unknowns in the genetics of complex traits and diseases. *bioRxiv [Preprint].* (2019) doi: 10.1101/553685
377. Zachar P. Psychiatric disorders are not natural kinds. *Philos Psychiatry Psychol.* (2000) 7:167–82.
378. Adriaens PR. Debunking evolutionary psychiatry's schizophrenia paradox. *Med Hypotheses.* (2008) 70:1215–22. doi: 10.1016/j.mehy.2007.10.014
379. Faulks S. *Human Traces: A Novel*. New York, NY: Random House (2005).
380. Bataillon T, Bailey SF. The year in evolutionary biology. *Ann N Y Acad Sci.* (2014) 1320:76–92. doi: 10.1111/nyas.12460
381. Kryazhimskiy S, Rice DP, Jerison ER, Desai MM. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science.* (2014) 344:1519–22. doi: 10.1126/science.1250939
382. Domingo J, Baeza-Centurion P, Lehner B. The causes and consequences of genetic interactions (Epistasis). *Annu Rev Genom Hum Genet.* (2019) 20:433–60. doi: 10.1146/annurev-genom-083118-014857
383. State MW, Levitt P. The conundrums of understanding genetic risks for autism spectrum disorders. *Nat Neurosci.* (2011) 14:1499–506. doi: 10.1038/nn.2924
384. Fuller T, Reus V. Shared genetics of psychiatric disorders. *F1000Res.* (2019) 8:F1000 Faculty Rev-1626. doi: 10.12688/f1000research.18130.1
385. Guan J, Cai JJ, Ji G, Sham PC. Commonality in dysregulated expression of gene sets in cortical brains of individuals with autism, schizophrenia, and bipolar disorder. *Transl Psychiatry.* (2019) 9:152. doi: 10.1038/s41398-019-0488-4
386. Sniegowski PD, Gerrish PJ, Lenski RE. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature.* (1997) 387:703–5. doi: 10.1038/42701
387. Taddei F, Radman M, Maynard-Smith J, Toupance B, Gouyon PH, Godelle B. Role of mutator alleles in adaptive evolution. *Nature.* (1997) 387:700–2. doi: 10.1038/42696
388. Woods RJ, Barrick JE, Cooper TF, Shrestha U, Kauth MR, Lenski RE. Second-order selection for evolvability in a large *Escherichia coli* population. *Science.* (2011) 331:1433–6. doi: 10.1126/science.1198914
389. Atkins WM. Biological messiness vs. biological genius: mechanistic aspects and roles of protein promiscuity. *J Steroid Biochem Mol Biol.* (2015) 151:3–11. doi: 10.1016/j.jsbmb.2014.09.010

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 03 February 2021
doi: 10.3389/fnins.2020.582279

Phenotypic Resemblance to Neuropsychiatric Disorder and Altered mRNA Profiles in Cortex and Hippocampus Underlying *IL15Rα* Knockout

*Yi He[1,2], Yuxin Yu[3], Yanan Li[4], Weicheng Duan[5], Zuoli Sun[1,2], Jian Yang[1,2], Abba J. Kastin[6], Weihong Pan[7], Yan Zhang[8] and Kang Wang[5]**

[1] The National Clinical Research Center for Mental Disorders, Beijing Key Laboratory of Mental Disorders, Beijing Anding Hospital, Capital Medical University, Beijing, China, [2] Advanced Innovation Center for Human Brain Protection, Capital Medical University, Beijing, China, [3] Department of Endocrinology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [4] Department of Gastrointestinal Surgery, Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China, [5] Department of Forensic Medicine, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [6] Pennington Biomedical Research Center, Baton Rouge, LA, United States, [7] BioPotentials Consult, Sedona, AZ, United States, [8] Department of Clinical Laboratory, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

OPEN ACCESS

Edited by:
Patrick E. Williams,
University of Kansas, United States

Reviewed by:
Xiaojun Wu,
Shanghai University of Traditional Chinese Medicine, China
Rupert W. Overall,
Helmholtz Association of German Research Centers (HZ), Germany

***Correspondence:**
Kang Wang
wangkjustice@hust.edu.cn

Specialty section:
This article was submitted to Neurogenomics, a section of the journal Frontiers in Neuroscience

Received: 11 July 2020
Accepted: 23 December 2020
Published: 03 February 2021

Citation:
He Y, Yu Y, Li Y, Duan W, Sun Z, Yang J, Kastin AJ, Pan W, Zhang Y and Wang K (2021) Phenotypic Resemblance to Neuropsychiatric Disorder and Altered mRNA Profiles in Cortex and Hippocampus Underlying IL15Rα Knockout. Front. Neurosci. 14:582279. doi: 10.3389/fnins.2020.582279

Background: Previous studies of the functions of *IL15Rα* have been limited to immune activities and skeletal muscle development. Immunological factors have been identified as one of the multiple causes of psychosis, and neurological symptoms have been described in *IL15Rα* knockout (KO) mice. Seeking to explore possible mechanisms for this in the *IL15Rα*$^{-/-}$ mouse brain, we analyzed gene expression patterns in the cortex and hippocampus using the RNA-seq technique.

Methods: *IL15Rα* KO mice were generated and littermate wildtype (WT) mice were used as a control group. A Y-maze was used to assess behavior differences between the two groups. The cortex and hippocampus of 3-month-old male mice were prepared and RNA-seq and transcriptome analysis were performed by gene set enrichment analysis (GSEA).

Results: Compared with the WT group, *IL15Rα* KO animals showed higher speed in the novel arm and more entrance frequency in the old arm in the Y-maze experiment. GSEA indicated that 18 pathways were downregulated and 13 pathways upregulated in both cortex and hippocampus from the GO, KEGG, and Hallmark gene sets. The downregulated pathways formed three clusters: respiratory chain and electron transport, regulation of steroid process, and skeletal muscle development.

Conclusion: *IL15Rα* KO mice exhibit altered expression of multiple pathways, which could affect many functions of the brain. Lipid biosynthesis and metabolism in the central nervous system (CNS) should be investigated to provide insights into the effect of *IL15Rα* on psychosis in this murine model.

Keywords: *IL15Rα*, RNA-seq, gene set enrichment analysis, neuroscience, overlap analysis, lipid metabolism

INTRODUCTION

Psychosis is a group of complex diseases with various unknown etiologies. There is a large economic burden to patients with disease such as schizophrenia and depression, associated with increased direct healthcare costs, lost productivity, changes in employment status, and suicide (Jin and Mosweu, 2017; Amos et al., 2018). Resulting sociological issues include violence, crime, and other problems. A focus of psychiatric research has been the relationships between the central nervous, immune, and endocrine systems (Steinberg et al., 2015). Depression and schizophrenia patients exhibit dysregulation in their immune function (Arolt et al., 2002), and IL-1β, IL-2r, IL-6, and TNF-α have been identified as biomarkers of schizophrenia (Upthegrove et al., 2014).

The immunomodulatory cytokine interleukin-15 (IL-15) exhibits functional similarities with IL-2 and is ubiquitously expressed in the normal CNS as well as other organs (He et al., 2010a). Compared with IL-2/IL-2Rα, IL-15/IL-15Rα is a more broadly expressed bio-regulator that affects a wider range of target cell populations, giving it the ability to regulate the homeostasis and growth of various non-immune cell types and tissues (Budagian et al., 2006). *IL-2Rα* KO mice have manifestations of autoimmunity and 25% die of severe anemia (Willerford et al., 1995). By contrast, *IL-15Rα* KO mice exhibit several unique alterations in immune cell development and function without increased mortality (Lodolce et al., 1998).

mRNA for *IL-15Rα* isoforms is widely expressed in microglia, astrocytes, and neurons, and it has a recognized major signaling role in cerebral function (Lee et al., 1996; Hanisch et al., 1997; Kurowska et al., 2002). *IL-15Rα* KO mice show hyperactivity, depressive-like behavior, reduced anxiety, and impaired memory (He et al., 2010a; Pan et al., 2013). Recent studies reported a rare loss-of-function *IL-15Rα* mutation in a schizophrenic patient and his siblings by sequencing—the clinical manifestations were similar to *IL-15Rα* KO mice (Pan et al., 2019). KO mice have further abnormalities such as a leaner body composition with lower fat (He et al., 2010b), altered skeletal motor activation, elevation of body temperature, and increased energy expenditure (He et al., 2010b).

How lipid metabolism can influence CNSs is attracting a rising interest. Altered lipid metabolism was linked to many neurological disorders, such as autism spectrum disorder and schizophrenia (Khan et al., 2002; Bell et al., 2004; Chauhan et al., 2004). The *IL-15* was proved to inhibit lipid deposition in murine 3 T3-L1 preadipocytes, and global *IL-15Rα* KO mice were observed with lower body fat accompanied by a series of neurological disorders (Budagian et al., 2006; He et al., 2010a; Loro et al., 2015). In this study, we sought a holistic view of the potential molecular mechanisms or pathways within the brain affected by the deletion of *IL-15Rα*. With the help of NGS technology, we studied the transcriptome in the cortex and hippocampus of *IL15Rα*$^{-/-}$ mice. GSEA was used to explore common mechanisms of the two brain regions, with validation by qRT-PCR. In addition, we describe connections among enriched pathways and cluster them into three themes.

Abbreviations: CNS, central nervous system; DEGs, differentially expressed genes; FDR, false discovery rate; FPKM, fragments per kilobase of exon model per million reads mapped; GO, gene ontology; GO-BP, gene ontology biological process; GO-CC, gene ontology cellular component; GO-MF, gene ontology molecular function; GSEA, gene set enrichment analysis; HISAT, hierarchical indexing for spliced alignment of transcripts; KEGG, Kyoto Encyclopedia of Genes and Genomes; KO, knockout; MSigDB, Molecular Signatures Database; NES, normalized enrichment score; NGS, next-generation sequencing; qRT-PCR, quantitative real-time polymerase chain reaction; RNA-Seq, whole transcriptome sequencing; RSEM, RNA-Seq by expectation-maximization; SZDB, schizophrenia database; WT, wildtype.

MATERIALS AND METHODS

Animal Maintenance

Male *IL15Rα*$^{-/-}$ mice were purchased from the Jackson Laboratory (003723) and were crossed to C57BL/6 females from Beijing Charles River Laboratories. After six generations of reproduction, heterozygous mice were identified and maintained for the next experiments. All mice were housed under 12/12 h light/dark conditions with free access to food and water in the animal facility of Capital Medical University. Littermate homozygous mice were obtained by pairing heterozygous mice. Genotyping was performed to differentiate *IL15Rα*$^{-/-}$ mice from WT animals. All experiments were conducted under 3R principles (reducing, reusing, and recycling) according to local ethical and safety rules. Animal work was approved by the Animal Use and Care Committee at Capital Medical University.

Animal Behavior Observations in a Y-Maze

The Y-maze (470 × 160 × 460 mm) consisted of three arms (start arm, novel arm, and the other arm) at an angle of 120° to each other, interconnected by a central zone. Mice underwent two trials: (1) with the novel arm blocked, mice were introduced to the start arm and allowed to freely explore the start arm and other arm for 10 min and (2) after a 1-h interval, mice were placed into the start arm again and had access to the start arm, novel arm, and other arm for 5 min. All trials were conducted in the same environment without human interference. The Y-maze was cleaned with a 5% alcohol solution between each trial to eliminate possible biasing effects from odor cues left behind. All trials were recorded by the camera above the maze, and subsequent video recordings were analyzed by the Supermaze animal behavior analysis system. Two-tailed unpaired Student's *t*-tests were applied using IBM SPSS Statistics for statistical data analysis, with p-value < 0.05 considered significant.

Total RNA Extraction and Library Construction

The cortex and hippocampus were microdissected from 3-month-old *IL15Rα*$^{-/-}$ and WT brains ($n = 3$) and then soaked and stored in TRIzol Reagent (Invitrogen, Catalog 15596026). RNA extraction was conducted according to the manufacturer's instructions and quantification was performed using Qubit (Invitrogen). Library construction and sequencing

were conducted by Igenecode Technology, Beijing, China. Following the manufacturer's instructions, mRNA was captured and isolated by magnetic beads coupled with Oligo (dT). The fragmented mRNA was then converted into double-stranded (ds) cDNA, ligated with a sequencing adapter, and amplified by PCR.

RNA Sequencing

After RNA quality control, RNA-seq was carried out with the HiSeq sequencing strategy using the Illumina second-generation high-throughput sequencing platform. Low-quality data were removed, and subsequently the filtered reads were aligned to the *Mus musculus* genome (GRCm38) using HISAT (Kim et al., 2015). The mapped reads were counted and converted to FPKM (Mortazavi et al., 2008) by RSEM (Li and Dewey, 2011).

Correlation Analysis

Biological replicate samples were grouped into one module in which FPKMs of each gene were averaged. The processed data were imported into GraphPad Prism and subjected to linear regression and Pearson correlation analysis, comparing cortex and hippocampus in both mutant and control groups.

Differential Gene Expression Analysis and qRT-PCR Validation

The matrix including raw counts from mutant and control groups was imported into R (v.3.6.3). The DESeq2 (Love et al., 2014, 2) (v.1.26.0) package was applied to analyze whether the expression pattern had changed due to *IL15Rα* KO and to identify significantly up-/downregulated genes in cortex or hippocampus. Only genes with an absolute value of log2 (fold change) ≥ 1 and p-value < 0.05 were regarded as differentially expressed between mutant and control groups.

Quantitative real-time PCR (qRT-PCR) was performed to evaluate and validate the expression pattern of DEGs in the cortex and hippocampus (**Supplementary Table S1**). cDNA was synthesized using the cDNA Reverse Transcription Kit (TIANGEN). All reactions were run in duplicate PCR amplification cycles in a real-time PCR instrument ABI7500 (ABI). Target mRNA was quantified relative to the β-*actin* gene as a reference molecule and analyzed with two-tailed unpaired Student's t-tests to assess statistical significance of gene expression levels between the *IL15Rα*$^{-/-}$ and control groups in cortex or hippocampus.

Gene Set Enrichment Analysis

Pathway enrichment analysis was implemented in the GSEA software (v.4.0.3) by comparing expression patterns between the *IL15Rα*$^{-/-}$ and control groups. Annotated gene sets were downloaded from the MSigDB (Liberzon et al., 2011) and included H (Hallmark gene sets) (Liberzon et al., 2015), KEGG (KEGG subset of CP), BP (GO biological process), CC (GO cellular component), and MF (GO molecular function). Only pathways with nominal p-values < 0.05 in both cortex and hippocampus were taken into consideration. The subsets were visualized by bubble diagrams created in R (v.3.6.3). Overlapped matrix of subsets was uploaded into Circos (Krzywinski et al., 2009) (v.0.63-9[1]) to perform online visualization.

Network Visualization and Analysis

To construct a union enrichment map, both cortex and hippocampus data were imported into Cytoscape (Doncheva et al., 2019) (v3.8.0) software and clusters were obtained after pruning the merged networks. ClusterMaker2 (Morris et al., 2011) (v.1.3.1) and AutoAnnotate (Kucera et al., 2016) (v.1.3.3) plugins helped to build clusters based on the topological structure of visualized networks and a summary of these annotations was generated by the WordCloud (Oesper et al., 2011) (v.3.1.3) plugin as a tag cloud over the clusters.

RESULTS

IL15Rα expression data from brains of both schizophrenia patients and healthy controls were extracted from the SZDB database (**Supplementary Table S2**) (Wu et al., 2017). There is a significant association of hippocampal *IL15Rα* expression with schizophrenia ($p = 8.32 \times 10^{-5}$), with a significant FDR ($p = 0.01046$). Schizophrenia patients seem to have higher *IL15Rα* expression levels in the hippocampus than controls (**Supplementary Figure S1**).

The Y-maze test was chosen for behavioral testing to assess cognitive functions of mice after *IL15Rα* KO. Utilizing the innate tendency to explore a novel environment (Deacon and Rawlins, 2006), we recorded the distance, time, speed, and entrance frequency in each arm, respectively. *IL15Rα*$^{-/-}$ mice showed a higher speed in the novel arm (**Figure 1A**) and more entrance frequency in the old arm (**Figure 1E**). With regard to the total arm (novel arm, start arm, and other arm), the two groups showed no significant difference in speed or entrance frequency (**Figures 1C,F**).

Next, we extracted total RNA samples from cortex and hippocampus in adult WT control and *IL15Rα* KO mice. Transcriptional profiles were assessed by NGS RNA-seq. After data quality control, we compared expression patterns between cortex and hippocampus in both the *IL15Rα*$^{-/-}$ group (**Figure 2A**) and control group (**Figure 2B**), which showed Pearson's correlations of 0.9648 and 0.9691, respectively. This finding reveals similar expression patterns between cortex and hippocampus. Therefore, we compared enrichment outcomes of cortex and hippocampus and selected overlapping features for further analysis in the following steps, to provide explanations of transformation in these two brain regions.

DEG Identification and qPCR Validation

Compared with control mice, 28 genes were expressed differentially in the *IL15Rα*$^{-/-}$ group in the cortex and 22 in the hippocampus, with FDR < 0.05 for both—these included nine upregulated and 19 downregulated genes or nine upregulated and 13 downregulated genes, respectively (**Figure 3A**). DEGs were listed (**Supplementary Table S3**),

[1]http://circos.ca/

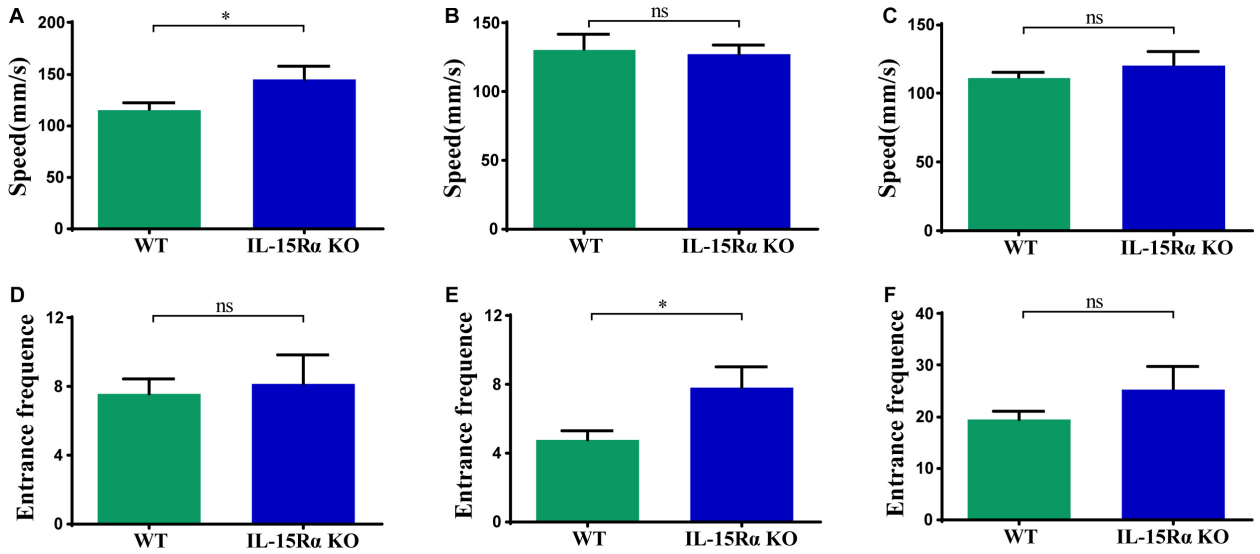


FIGURE 1 | Speed and entrance frequency behavior abnormalities after *IL15Rα* knockout. **(A)** Histogram presents the average speed of IL15Rα mice and wildtype mice in novel arm. **(B)** Histogram summarizes the average speed of IL15Rα mice and wildtype mice in old arm (start arm and the other arm). **(C)** Histogram reveals the average speed of IL15Rα mice and wildtype mice in total arm (novel arm, start arm, and the other arm). **(D)** Histogram presents the entrance frequency of IL15Rα mice and wildtype mice in novel arm. **(E)** Histogram summarizes the entrance frequency of IL15Rα mice and wildtype mice in old arm (start arm and the other arm). **(F)** Histogram reveals the entrance frequency of IL15Rα mice and wildtype mice in total arm (novel arm, start arm, and the other arm). Average speed is computed from total distance traveled in certain region divided by total time in certain region. The results are depicted as the mean ± SD from eight biological replicates in IL15Rα group and 12 biological replicates in control group. The level of significance is calculated by two-tailed unpaired Student's *t*-test. $^*p < 0.05$.

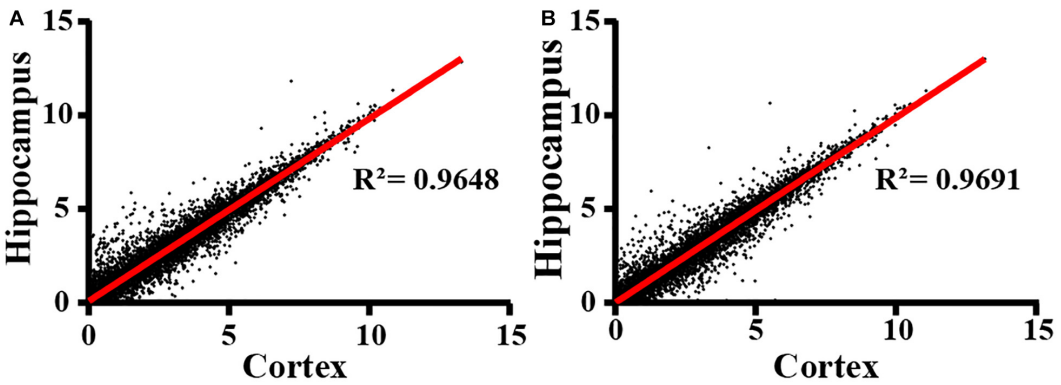


FIGURE 2 | Correlation of RNA expression profiles between cortex and hippocampus. Cortex and hippocampus are computed for sample correlation by Pearson's correlation analysis in *IL15Rα*$^{-/-}$ group **(A)** and control group **(B)**, respectively. Pearson's correlation coefficient is represented by R^2 values.

and 11 interesting listed genes were chosen to validate the transcript sequencing results, including three genes in the cortex and eight in the hippocampus. Expression changes were detected by qPCR using reverse transcript cDNA from total RNA samples (**Figures 3B–L**). These 11 genes are involved in lipid metabolism, oligodendrocytic function, GTP binding, and cytoskeletal function. The results demonstrate that transcription levels of *Gem*, *Opalin*, and *Pllp* are downregulated in the cortex and *Arc*, *Egr4*, *Fos*, *Fosb*, *Dsp*, *Npas4*, and *Junb* are downregulated and *Plin4* upregulated in the hippocampus. Quantitative PCR analysis was in accordance with the NGS results (**Supplementary Tables S3, S4**).

GSEA Generated Gene Set Enrichment and Functional Annotation

As a computational method that determines whether a defined set of genes shows statistically significant concordant differences between two biological states, GSEA focuses on groups of genes that share common biological function or regulation (Subramanian et al., 2005). We chose three widespread and frequently used gene sets from MSigDB, namely, Hallmark, GO, and KEGG gene sets for further enrichment analyses by GSEA. The procedure was performed in cortex and hippocampus separately and the overlap region identified with the criteria p-value < 0.05. In total, 625 pathways were identified as

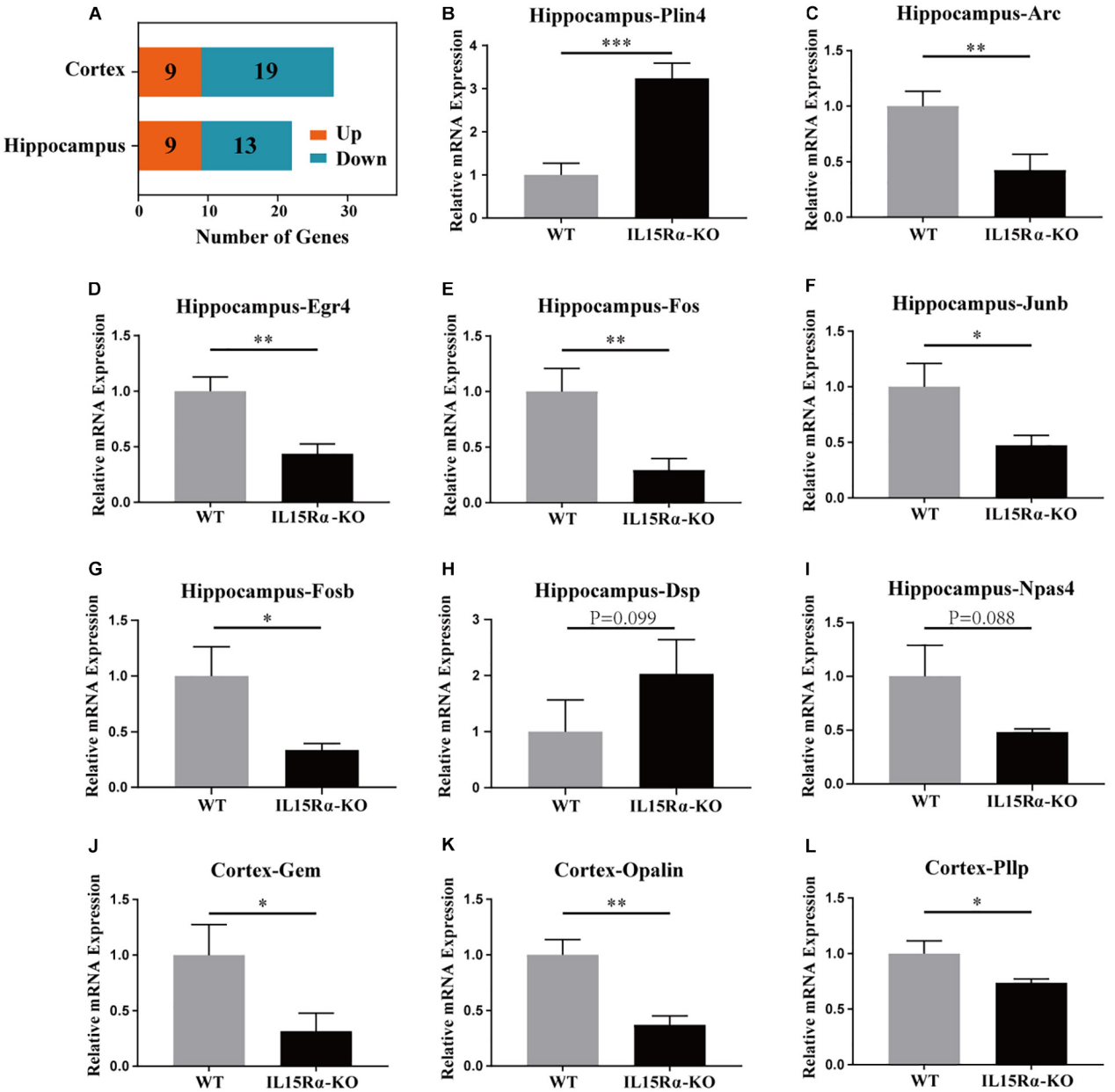


FIGURE 3 | DEG analysis in cortex and hippocampus and qRT-PCR validation of certain DEGs. **(A)** Stacked diagram reveals the amount of up-/downregulated DEGs in cortex and hippocampus. The genes with absolute value of log2 (fold change) ≥ 1 and *p*-value < 0.05 between *IL15Rα*$^{-/-}$ group and control group are defined as DEGs. DEGs possessing log2(fold change) > 0 are considered to be significantly upregulated, whereas the other are significantly downregulated. **(B–I)** Six DEGs (*Plin4*, *Arc*, *Egr4*, *Fos*, *Fosb*, and *Junb*) in hippocampus are evaluated to be differently expressed between *IL15Rα*$^{-/-}$ group and control group by qRT-PCR. **(J–L)** Three DEGs (*Gem*, *Opalin*, and *Pllp*) in cortex are validated to be differently expressed between *IL15Rα*$^{-/-}$ group and control group by qRT-PCR. Expression is determined by the relative quantification method with β-*actin* gene as internal reference. Relative mRNA expression is calculated from three biological replicates and depicted as the mean ± SD after normalizing to average expression in control group. The level of significance is calculated by two-tailed unpaired Student's *t*-test. **p* < 0.05; ***p* < 0.01; ****p* < 0.001.

downregulated and 565 as upregulated in the three pre-defined gene sets (**Figures 4A,B**). Based on previous studies, IL-15 shares many biological properties with interleukin-2 (IL-2) (Grabstein et al., 1994). Furthermore, IL15Rα and IL2Rα are tightly linked in both mouse and humans with respect to gene size and organization (Anderson et al., 1995; Tagaya et al., 1996). In accordance with an expectation of a compensation mechanism, significantly overlapping upregulated enriched pathways were mapped to the regulation, biosynthesis, or production of *IL-2* and inflammatory response or negative regulation of immune system processes (**Supplementary Table S5**). Next, we confirmed four existing representative enriched pathways

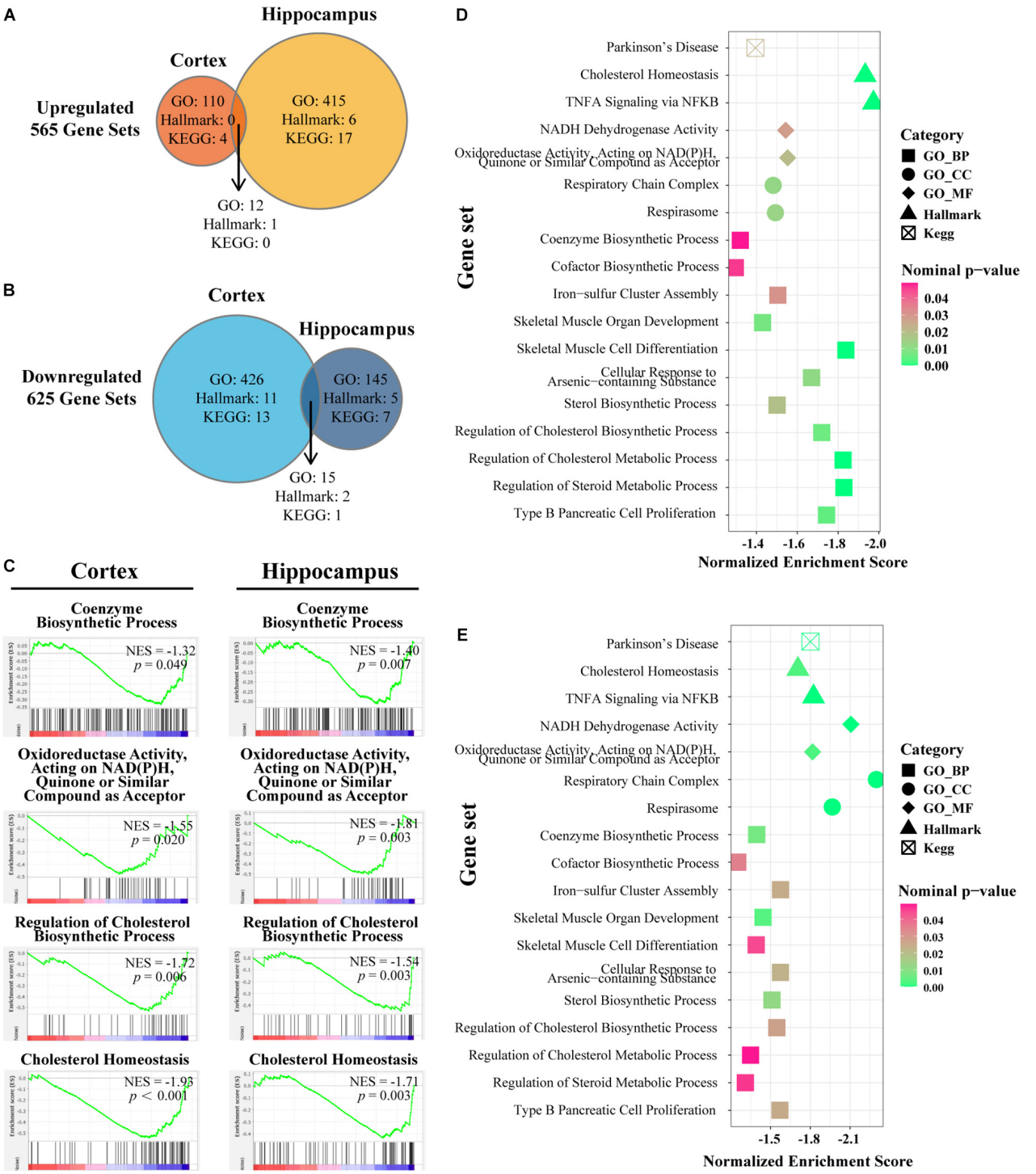


FIGURE 4 | Gene set enrichment analyses in cortex and hippocampus and the intersection of the enriched pathways. **(A,B)** Venn diagram depicts the amount of significantly up-/downregulated pathways with nominal *p*-value < 0.05 in the cortex or hippocampus of *IL15Rα*$^{-/-}$ samples. Enriched pathways are retrieved from following sets of genes: H (Hallmark gene sets), KEGG (KEGG subset of CP), BP (GO biological process), CC (GO cellular component), and MF (GO molecular function). Overlapped areas in the diagram count the commonly contributing pathways in both cortex and hippocampus. **(C)** Enrichment plot provides a graphical view of the enrichment score for pathways chosen from downregulated overlapped pathways. **(D)** Bubble diagram conveys the core characters of certain pathways in cortex. **(E)** Bubble diagram presents the state of certain pathways in hippocampus. The shape, color, and abscissa value of points indicate the category of gene sets, magnitude of nominal *p*-value, and normalized enrichment score, respectively. Pathways are selected from overlapped downregulation.

based on the aforementioned results: coenzyme biosynthetic process, oxidoreductase activity acting on NADPH quinone or similar compound as acceptor, regulation of cholesterol biosynthetic process, and cholesterol homeostasis. Both cortex and hippocampus showed significant enrichment at the head of the targeted gene sets with NES < -1 (**Figure 4C**). In the next section, we concentrated our attention on overlapping downregulated enriched pathways and cluster terms of both cortex and hippocampus were identified (**Figures 4D,E**). A total of 18 downregulated pathways were composed of 11 GO-BP ontology, two GO-CC ontology, two GO-MF ontology pathways, two Hallmark pathways, and one KEGG pathway (**Supplementary Table S5**). For the *IL15Rα*$^{-/-}$ group, cortex showed highly NES terms on cholesterol homeostasis (Hallmarks), skeletal muscle cell differentiation (GO-BP), and regulation of steroid or cholesterol biosynthetic or metabolic process (GO-BP) (**Figure 4D**). At the same time, pathways related to catalysis of an oxidation–reduction reaction in NADH or NADPH (GO-MF), respiratory chain activity (GO-CC), and cholesterol homeostasis (Hallmarks) were markedly enriched in the hippocampus (**Figure 4E**).

Overlapped Pathways Connection Analysis and Unified Construction of Enrichment Map

To make a specific cross-examination between the 18 overlapped pathways in cortex and hippocampus, we defined a new gene set composed of these gene set names and specific continents from MSigDB of the BROAD Institute (Liberzon et al., 2011, 0). To extract the core members of the chosen pathways and explore connections between them, leading-edge analysis was used to grasp leading-edge subsets. The link intensity among the 18 pathways was drawn based on the similarity of contributors. Cortex (**Figure 5A**) and hippocampus (**Figure 5B**) showed analogous connection patterns within the 18 pathways but differed when analyzed in detail. There were complicated connections of different strengths (represented by width of the lines) in the circle, showing close or alienated relationships among pathways. Both cortex and hippocampus samples revealed dense connections from TNF-α signaling via NFκB pathways and many others. For the respiration-related and electron transport pathways, such as respirasome, respiratory chain complex, or NADPH dehydrogenase activity, hippocampus samples had more connective links with cofactor biosynthetic process and coenzyme biosynthetic process. Another difference is a larger connection weight between sterol biosynthetic with regulation of cholesterol biosynthetic and metabolic process, in cortex samples.

Designed for visualization networks among enriched pathways, "enrichment map" is a cytoscape plugin that has erected a bridge between GSEA software and cytoscape. An enrichment map helps to identify interesting pathways and themes (Reimand et al., 2019). Here, we detected clusters of enriched pathways based on the visualized enrichment map of the 18 pathways (**Figure 5C**). The largest cluster, labeled "respiratory chain and electron transport," was composed of five pathways, followed by the second biggest cluster labeled "regulation steroid process" and the last cluster labeled "skeletal muscle development."

DISCUSSION

With an extensive number of genes reported to be associated with various diseases, the genetic mechanisms of neurological abnormalities are regarded as highly heterogeneous. Accumulated evidence confirms that *IL15Rα* insufficiency results in several disorders, including but not limited to immunodeficiency, skeletal muscle variation, and a series of neurological symptoms (Lodolce et al., 1998; He et al., 2010a,b; Pan et al., 2013; O'Connell et al., 2015). In this study, we first performed Y-maze tests to examine mouse behavioral properties after *IL15Rα* KO. Y-maze is a behavioral test based on the willingness of rodents to explore new environments. The task-involved brain regions include the hippocampus, septum, basal forebrain, and prefrontal cortex. Overall, *IL15Rα*$^{-/-}$ mice showed a little higher speed and a little more entrance frequency (without significance) throughout the testing process (**Figures 1C,F**). However, *IL15Rα*$^{-/-}$ mice moved faster in the novel arm (**Figure 1A**) and moved equally in the old arm (**Figure 1B**). Although previous studies have demonstrated that lack of *IL15Rα* influences muscle mitochondrial structure and function and results in improvement of athletic performance (Loro et al., 2018), *IL15Rα*$^{-/-}$ mice only exhibited higher speed in novel environments without the old arm. We suppose this was due to the high neuroexcitability that brings greater alertness in the new environment, rather than changes in skeletal muscle. In addition, compared with the similar entrance frequency of the novel arm (**Figure 1D**), *IL15Rα*$^{-/-}$ mice return to the old arm more frequently (**Figure 1E**). Mice are born with curiosity to explore new environments and the Y-maze test is designed to detect it. In a former anxiety evaluation trial, *IL15Rα* KO mice showed reduced anxiety both in the open field test and elevated plus maze test (Wu et al., 2010). Among the identified DEGs, the immediate–early gene *Arc* was reported as dynamically regulated in the hippocampus (Guzowski et al., 2006). An increase of *Arc* has been observed in the fear memory (Clifton et al., 2017). In this study, the expression of Arc was found to decrease in *IL15Rα* KO mice, supporting the behavior with less anxiety. The increasing curiosity may result in a higher speed in the novel arm, and the lowered anxiety may bring mice more frequently returning to the old arm, similar to the increase in the inner grids crossed in the open field test. Our previous study discovered that the *IL15Rα* KO mice did worse in memory retention than the controls 1 week after the initial trial by stone 14-unit T-maze (He et al., 2010a). However, mouse performance in a Y-maze indicated no deficit in short-term memory after 1 h of the initial trial. The difference between the two experiments may be due to the time interval after training, the different rationale of the two experiments, or to genetic differences. Next, we aimed to explore mechanisms of neurological abnormality in *IL15Rα* insufficiency and identify new gene sets and pathways in *IL15Rα*$^{-/-}$ mouse brain. Three-month-old male littermate mice housed in a barrier environment were selected to exclude interference from genetic

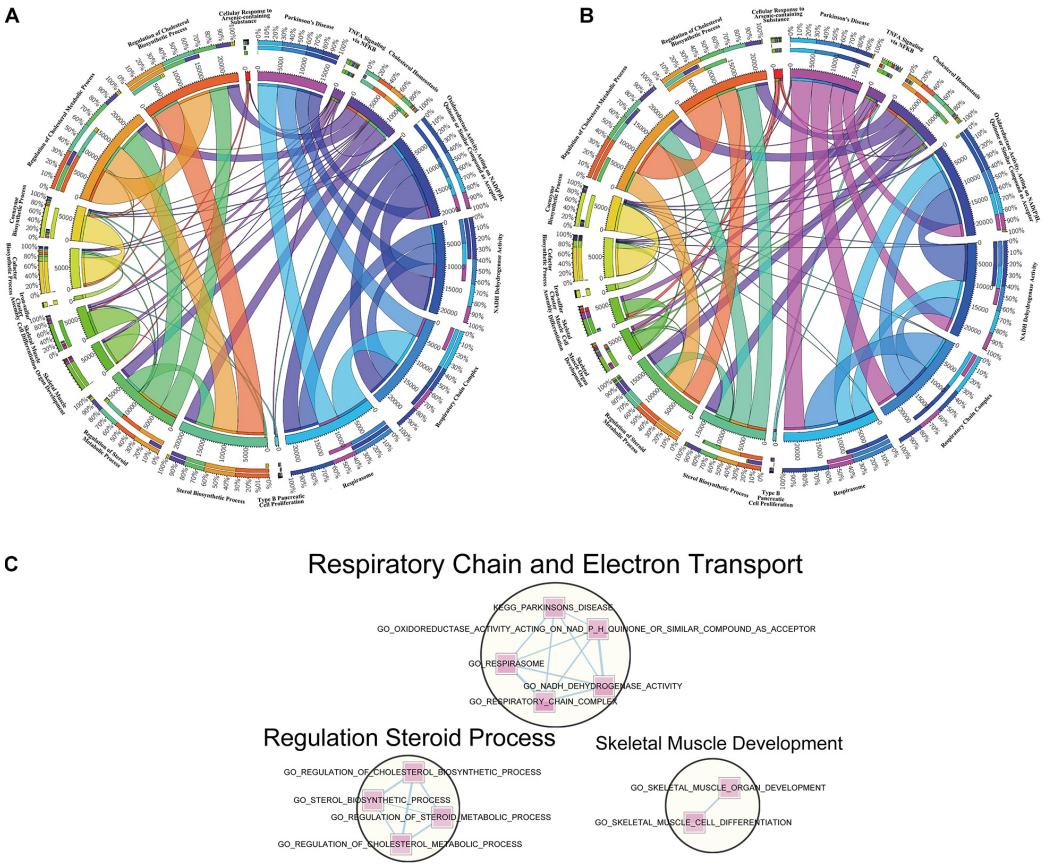


FIGURE 5 | Leading-edge analysis of downregulated overlapped pathways. **(A)** Circos diagram presents a circular visualization of connection patterns in cortex within the 18 pathways chosen from overlapped downregulation. **(B)** Circos diagram utilizes a circular layout to facilitate the display of interaction patterns in hippocampus within the 18 pathways chosen from overlapped downregulation. The width of line represents various degrees to the relevance among pathways. **(C)** Enrichment networks reveal the important clusters achieved from the 18 overlapped downregulated gene sets. The pathways in circles represent the functionally associated sub-modules in the overall networks based on the topological structures, with the functional annotation displayed above.

background, age, and feeding conditions. This can ensure a more reliable analysis of differences caused by *IL15Rα* deletion compared with WT gene transcription. Our analysis clearly demonstrates that the cortex and hippocampus response to *IL15Rα* deletion involves alterations in multiple pathways related to the respiratory chain, along with electron transport, regulation of steroid processes, skeletal muscle development, and regulation and biosynthesis or production of *IL-2*.

High-throughput sequencing technology can produce big data but has the disadvantage of inaccuracy. To verify the transcriptome results, qRT-PCR was used to quantitate the expression levels of 11 DEGs (**Supplementary Tables S3, S4**). Among our validated DEGs, *Fos* is a family of transcription factors consisting of four members: *FOS*, *FOSB*, *FOSL1*, and *FOSL2*. The *Fos* family is involved in biological processes including neurogenesis, neuron differentiation, and neuron development. *Npas4* is also a kind of transcriptional regulator, which is involved in a wide range of physiologic and developmental events (Ooe et al., 2004). *Egr1*, *Egr2*, and *Egr4* belong to the Egr family of transcriptional regulatory factors, which is implicated in neuronal plasticity. These identified abnormalities in transcription factors indicated an association with numerous mental health disorders in *IL15Rα* KO mice, including less anxiety or the memory disorders. Subsequently, GSEA was utilized to assist analysis and interpretation of the long lists of gene expression data into more easily interpretable biological pathways (Powers et al., 2018). Both cortex and hippocampus have been widely reported to be involved in psychosis, including autism and schizophrenia (Schultz and Andreasen, 1999; Carper and Courchesne, 2005; Duan et al., 2020). Exploring the transcriptomic changes in these organs in the *IL15Rα* insufficiency mouse model will help to achieve a better understanding of the molecular pathways involved in this disease and benefit future novel diagnosis and treatment methods.

Our GSEA screening characterized the alterations of expression profiles and identified overlapping pathways enriched in both cortex and hippocampus (**Figure 4**). We found that skeletal muscle organ development and cell differentiation in the two brain regions were enriched, consistent with previous studies (Loro et al., 2018). Another enrichment module was related to steroid processes, including regulation of steroid

biosynthesis and metabolism, and cholesterol biosynthetic and metabolic processes. Sterol lipids are an essential class of lipids in all mammals, synthesized in many forms with the major form being cholesterol, and are of particular importance to the brain (Tracey et al., 2018). Abnormal lipid metabolism has been identified as hallmarks of neurodevelopmental and neurodegenerative disorders (Hebbar et al., 2017), for example, dysfunction in lipid metabolism and clearance were observed to be implicated in Alzheimer's disease pathogenesis (Di Paolo and Kim, 2011; Visan, 2017). Considering the fact that neurons have a lower capacity for cholesterol synthesis (Zhang and Liu, 2015), sterol or cholesterol disorders may induce abnormalities in the brain. Cholesterol synthesis or metabolism requires acetyl CoA and electron input at multiple steps, utilizing both NADH and NADPH as the electron source (Porter, 2015). Indeed, this was exactly the pathway we enriched (**Figures 4C–E**) accompanied by steroid processes pathways. Furthermore, the pathway cluster of chain and electron transport involved in respiratory was also enriched (**Figure 5C**).

There were also some findings to be revisited. Steroid and cholesterol biosynthetic and metabolic pathway was more enriched in cortex than hippocampus in *IL15Rα*$^{-/-}$ mice (**Figure 5A**), whereas the pathway related to respiratory chain or electron transfer showed more prominence in the hippocampus (**Figure 5B**). More downregulated pathways were enriched in the cortex; by contrast, more upregulated pathways were enriched in the hippocampus (**Figures 4A,B**). These subtle differences may indicate divergence of mechanisms in the two brain regions, which is worth further investigation in future research.

To summarize, we analyzed gene expression patterns in *IL15Rα* KO mice by RNA-seq and validated some of the DEGs by qPCR. Transcriptional changes in overlapping pathways emerged and clustered into three groups. The data from *IL15Rα*$^{-/-}$ mice led us to propose that sterol biosynthetic and metabolic in the brain may be a key factor in schizophrenia-like etiology.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://db.cngb.org/cnsa/, CNP0001191.

ETHICS STATEMENT

The animal study was reviewed and approved by the Ethics Committee of Animal Use and Care Committee at Capital Medical University.

AUTHOR CONTRIBUTIONS

YH, ZS, and JY designed and conducted the experiments. YH provided biological samples. YL and YZ supported animal care. YH and KW conducted the statistical analysis. YY and YL analyzed and visualized high-throughput sequencing data. YY, WD, and KW cooperated in writing the original draft. AK and WP helped to revise the article and proposed constructive opinions. All authors approved the original draft of the article.

FUNDING

This study was supported by Beijing Municipal Administration of Hospitals Incubating Program (PX2018065) and National Natural Science Foundation of China (31970918, 81801339, and 81701317).

ACKNOWLEDGMENTS

We are particularly grateful to Jingming Liu for providing language help.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2020.582279/full#supplementary-material

REFERENCES

Amos, T. B., Tandon, N., Lefebvre, P., Pilon, D., Kamstra, R. L., Pivneva, I., et al. (2018). Direct and indirect cost burden and change of employment status in treatment-resistant depression: a matched-cohort study using a us commercial claims database. *J. Clin. Psychiatry* 79, 17m11725. doi: 10.4088/JCP.17m11725

Anderson, D. M., Kumaki, S., Ahdieh, M., Bertles, J., Tometsko, M., Loomis, A., et al. (1995). Functional characterization of the human interleukin-15 receptor alpha chain and close linkage of IL15RA and IL2RA genes. *J. Biol. Chem.* 270, 29862–29869. doi: 10.1074/jbc.270.50.29862

Arolt, V., Rothermundt, M., Peters, M., and Leonard, B. (2002). immunological research in clinical psychiatry: report on the consensus debate during the 7th expert meeting on psychiatry and immunology. *Mol. Psychiatr.* 7, 822–826. doi: 10.1038/sj.mp.4001115

Bell, J. G., MacKinlay, E. E., Dick, J. R., MacDonald, D. J., Boyle, R. M., and Glen, A. C. A. (2004). Essential fatty acids and phospholipase A2 in autistic spectrum disorders. *Prostaglandins Leukot Essent Fatty Acids* 71, 201–204. doi: 10.1016/j.plefa.2004.03.008

Budagian, V., Bulanova, E., Paus, R., and Bulfone-Paus, S. (2006). IL-15/IL-15 receptor biology: a guided tour through an expanding universe. *Cytokine Growth Factor Rev.* 17, 259–280. doi: 10.1016/j.cytogfr.2006.05.001

Carper, R. A., and Courchesne, E. (2005). Localized enlargement of the frontal cortex in early autism. *Biol. Psychiatry* 57, 126–133. doi: 10.1016/j.biopsych.2004.11.005

Chauhan, A., Chauhan, V., Brown, W. T., and Cohen, I. (2004). Oxidative stress in autism: increased lipid peroxidation and reduced serum levels of ceruloplasmin and transferrin–the antioxidant proteins. *Life Sci.* 75, 2539–2549. doi: 10.1016/j.lfs.2004.04.038

Clifton, N. E., Cameron, D., Trent, S., Sykes, L. H., Thomas, K. L., and Hall, J. (2017). Hippocampal regulation of postsynaptic density homer1 by associative learning. *Neural Plast.* 2017, 5959182. doi: 10.1155/2017/5959182

Deacon, R. M. J., and Rawlins, J. N. P. (2006). T-maze alternation in the rodent. *Nat. Protoc.* 1, 7–12. doi: 10.1038/nprot.2006.2

Di Paolo, G., and Kim, T.-W. (2011). Linking lipids to Alzheimer's disease: cholesterol and beyond. *Nat. Rev. Neurosci.* 12, 284–296. doi: 10.1038/nrn3012

Doncheva, N. T., Morris, J. H., Gorodkin, J., and Jensen, L. J. (2019). Cytoscape StringApp: network analysis and visualization of proteomics data. *J. Proteome Res.* 18, 623–632. doi: 10.1021/acs.jproteome.8b00702

Duan, W., Wang, K., Duan, Y., Chu, X., Ma, R., Hu, P., et al. (2020). Integrated transcriptome analyses revealed key target genes in mouse models of autism. *Autism Res.* 13, 352–368. doi: 10.1002/aur.2240

Grabstein, K. H., Eisenman, J., Shanebeck, K., Rauch, C., Srinivasan, S., Fung, V., et al. (1994). Cloning of a T cell growth factor that interacts with the beta chain of the interleukin-2 receptor. *Science* 264, 965–968. doi: 10.1126/science.8178155

Guzowski, J. F., Miyashita, T., Chawla, M. K., Sanderson, J., Maes, L. I., Houston, F. P., et al. (2006). Recent behavioral history modifies coupling between cell activity and Arc gene transcription in hippocampal CA1 neurons. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1077–1082. doi: 10.1073/pnas.0505519103

Hanisch, U. K., Lyons, S. A., Prinz, M., Nolte, C., Weber, J. R., Kettenmann, H., et al. (1997). Mouse brain microglia express interleukin-15 and its multimeric receptor complex functionally coupled to Janus kinase activity. *J. Biol. Chem.* 272, 28853–28860. doi: 10.1074/jbc.272.46.28853

He, Y., Hsuchou, H., Wu, X., Kastin, A. J., Khan, R. S., Pistell, P. J., et al. (2010a). Interleukin-15 receptor is essential to facilitate GABA transmission and hippocampal-dependent memory. *J. Neurosci.* 30, 4725–4734. doi: 10.1523/JNEUROSCI.6160-09.2010

He, Y., Wu, X., Khan, R. S., Kastin, A. J., Cornelissen-Guillaume, G. G., Hsuchou, H., et al. (2010b). IL-15 receptor deletion results in circadian changes of locomotor and metabolic activity. *J. Mol. Neurosci.* 41, 315–321. doi: 10.1007/s12031-009-9319-z

Hebbar, S., Khandelwal, A., Jayashree, R., Hindle, S. J., Chiang, Y. N., Yew, J. Y., et al. (2017). Lipid metabolic perturbation is an early-onset phenotype in adult spinster mutants: a *Drosophila* model for lysosomal storage disorders. *Mol. Biol. Cell* 28, 3728–3740. doi: 10.1091/mbc.E16-09-0674

Jin, H., and Mosweu, I. (2017). The societal cost of schizophrenia: a systematic review. *Pharmacoeconomics* 35, 25–42. doi: 10.1007/s40273-016-0444-6

Khan, M. M., Evans, D. R., Gunna, V., Scheffer, R. E., Parikh, V. V., and Mahadik, S. P. (2002). Reduced erythrocyte membrane essential fatty acids and increased lipid peroxides in schizophrenia at the never-medicated first-episode of psychosis and after years of treatment with antipsychotics. *Schizophr. Res.* 58, 1–10. doi: 10.1016/s0920-9964(01)00334-6

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Kucera, M., Isserlin, R., Arkhangorodsky, A., and Bader, G. D. (2016). AutoAnnotate: a Cytoscape app for summarizing networks with semantic annotations. *F1000Res.* 5, 1717. doi: 10.12688/f1000research.9090.1

Kurowska, M., Rudnicka, W., Maślińska, D., and Maśliński, W. (2002). Expression of IL-15 and IL-15 receptor isoforms in select structures of human fetal brain. *Ann. N. Y. Acad. Sci.* 966, 441–445. doi: 10.1111/j.1749-6632.2002.tb04245.x

Lee, Y. B., Satoh, J., Walker, D. G., and Kim, S. U. (1996). Interleukin-15 gene expression in human astrocytes and microglia in culture. *Neuroreport* 7, 1062–1066. doi: 10.1097/00001756-199604100-00022

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* 12:323. doi: 10.1186/1471-2105-12-323

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* 1, 417–425. doi: 10.1016/j.cels.2015.12.004

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27, 1739–1740. doi: 10.1093/bioinformatics/btr260

Lodolce, J. P., Boone, D. L., Chai, S., Swain, R. E., Dassopoulos, T., Trettin, S., et al. (1998). IL-15 receptor maintains lymphoid homeostasis by supporting lymphocyte homing and proliferation. *Immunity* 9, 669–676. doi: 10.1016/s1074-7613(00)80664-0

Loro, E., Bisetto, S., and Khurana, T. S. (2018). Mitochondrial ultrastructural adaptations in fast muscles of mice lacking IL15RA. *J. Cell. Sci.* 131, jcs218313. doi: 10.1242/jcs.218313

Loro, E., Seifert, E. L., Moffat, C., Romero, F., Mishra, M. K., Sun, Z., et al. (2015). IL-15Rα is a determinant of muscle fuel utilization, and its loss protects against obesity. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 309, R835–R844. doi: 10.1152/ajpregu.00505.2014

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8

Morris, J. H., Apeltsin, L., Newman, A. M., Baumbach, J., Wittkop, T., Su, G., et al. (2011). clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinform.* 12:436. doi: 10.1186/1471-2105-12-436

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226

O'Connell, G., Guo, G., Stricker, J., Quinn, L. S., Ma, A., and Pistilli, E. E. (2015). Muscle-specific deletion of exons 2 and 3 of the IL15RA gene in mice: effects on contractile properties of fast and slow muscles. *J. Appl. Physiol.* 118, 437–448. doi: 10.1152/japplphysiol.00704.2014

Oesper, L., Merico, D., Isserlin, R., and Bader, G. D. (2011). WordCloud: a Cytoscape plugin to create a visual semantic summary of networks. *Source Code Biol. Med.* 6, 7. doi: 10.1186/1751-0473-6-7

Ooe, N., Saito, K., Mikami, N., Nakatuka, I., and Kaneko, H. (2004). Identification of a novel basic helix-loop-helix-PAS factor, NXF, reveals a Sim2 competitive, positive regulatory role in dendritic-cytoskeleton modulator drebrin gene expression. *Mol. Cell Biol.* 24, 608–616. doi: 10.1128/mcb.24.2.608-616.2004

Pan, W., Wu, X., He, Y., Hsuchou, H., Huang, E. Y.-K., Mishra, P. K., et al. (2013). Brain interleukin-15 in neuroinflammation and behavior. *Neurosci. Biobehav. Rev.* 37, 184–192. doi: 10.1016/j.neubiorev.2012.11.009

Pan, Y., Wang, Z., Zhang, G., Guo, J., Zhu, X., Zhou, J., et al. (2019). Schizophrenia patient shows a rare interleukin 15 receptor alpha variant disrupting signal transduction. *Curr. Mol. Med.* 19, 560–569. doi: 10.2174/1566524019666190617172054

Porter, T. D. (2015). Electron transfer pathways in cholesterol synthesis. *Lipids* 50, 927–936. doi: 10.1007/s11745-015-4065-1

Powers, R. K., Goodspeed, A., Pielke-Lombardo, H., Tan, A.-C., and Costello, J. C. (2018). GSEA-InContext: identifying novel and common patterns in expression experiments. *Bioinformatics* 34, i555–i564. doi: 10.1093/bioinformatics/bty271

Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., et al. (2019). Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.* 14, 482–517. doi: 10.1038/s41596-018-0103-9

Schultz, S. K., and Andreasen, N. C. (1999). Schizophrenia. *Lancet* 353, 1425–1430. doi: 10.1016/s0140-6736(98)07549-7

Steinberg, H., Kirkby, K. C., and Himmerich, H. (2015). The historical development of immunoendocrine concepts of psychiatric disorders and their therapy. *Int. J. Mol. Sci.* 16, 28841–28869. doi: 10.3390/ijms161226136

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102

Tagaya, Y., Burton, J. D., Miyamoto, Y., and Waldmann, T. A. (1996). Identification of a novel receptor/signal transduction pathway for IL-15/T in mast cells. *EMBO J.* 15, 4928–4939.

Tracey, T. J., Steyn, F. J., Wolvetang, E. J., and Ngo, S. T. (2018). Neuronal lipid metabolism: multiple pathways driving functional outcomes in health and disease. *Front. Mol. Neurosci.* 11:10. doi: 10.3389/fnmol.2018.00010

Upthegrove, R., Manzanares-Teson, N., and Barnes, N. M. (2014). Cytokine function in medication-naive first episode psychosis: a systematic review

and meta-analysis. *Schizophr. Res.* 155, 101–108. doi: 10.1016/j.schres.2014.03.005

Visan, I. (2017). Alzheimer's disease microglia. *Nat. Immunol.* 18, 876. doi: 10.1038/ni.3802

Willerford, D. M., Chen, J., Ferry, J. A., Davidson, L., Ma, A., and Alt, F. W. (1995). Interleukin-2 receptor alpha chain regulates the size and content of the peripheral lymphoid compartment. *Immunity* 3, 521–530. doi: 10.1016/1074-7613(95)90180-9

Wu, X., He, Y., Hsuchou, H., Kastin, A. J., Rood, J. C., and Pan, W. (2010). Essential role of interleukin-15 receptor in normal anxiety behavior. *Brain Behav. Immun.* 24, 1340–1346. doi: 10.1016/j.bbi.2010.06.012

Wu, Y., Yao, Y.-G., and Luo, X.-J. (2017). SZDB: a database for schizophrenia genetic research. *Schizophr. Bull.* 43, 459–471. doi: 10.1093/schbul/sbw102

Zhang, J., and Liu, Q. (2015). Cholesterol metabolism and homeostasis in the brain. *Protein Cell* 6, 254–264. doi: 10.1007/s13238-014-0131-3

Conflict of Interest: WP was employed by the company BioPotentials Consult.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer XW declared a past co-authorship with several of the authors YH, ZS, JY, AK, and WP to the handling editor.

frontiers in Genetics

ORIGINAL RESEARCH
published: 19 February 2021
doi: 10.3389/fgene.2021.553974

Genetic Effects of the Schizophrenia-Related Gene *DTNBP1* in Temporal Lobe Epilepsy

Hua Tao[1,2†], Xu Zhou[2†], Jun Chen[1], Haihong Zhou[1], Lidan Huang[2], Yujie Cai[2], Jiawu Fu[1], Zhou Liu[2], Yanyan Chen[2], Chaowen Sun[1], Bin Zhao[2], Wangtao Zhong[1*] and Keshen Li[3,4*]*

[1] Department of Neurology, Affiliated Hospital of Guangdong Medical University, Zhanjiang, China, [2] Guangdong Key Laboratory of Age-related Cardiac and Cerebral Diseases, Guangdong Medical University, Zhanjiang, China, [3] Institute of Neurology, Guangdong Medical University, Zhanjiang, China, [4] Neurology and Neurosurgery Division, Stroke Center, The First Affiliated Hospital, Clinical Medicine Research Institute, Jinan University, Guangzhou, China

OPEN ACCESS

Edited by:
Elena Martín-García,
Pompeu Fabra University, Spain

Reviewed by:
Gilberto Vargas Alarcón,
Instituto Nacional de Cardiologia Ignacio Chavez, Mexico
Aurelijus Burokas,
Vilnius University, Lithuania

***Correspondence:**
Bin Zhao
binzhaoe@sohu.com
Wangtao Zhong
zhongwangtao512@aliyun.com
Keshen Li
likeshen1971@126.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Behavioral and Psychiatric Genetics, a section of the journal Frontiers in Genetics

Received: 20 April 2020
Accepted: 15 January 2021
Published: 19 February 2021

Citation:
Tao H, Zhou X, Chen J, Zhou H, Huang L, Cai Y, Fu J, Liu Z, Chen Y, Sun C, Zhao B, Zhong W and Li K (2021) Genetic Effects of the Schizophrenia-Related Gene DTNBP1 in Temporal Lobe Epilepsy. Front. Genet. 12:553974. doi: 10.3389/fgene.2021.553974

Recent studies have reported patients who concurrently exhibit conditions of epilepsy and schizophrenia, indicating certain shared pathologies between them. This study aimed to investigate the genetic effects of the schizophrenia-related gene *DTNBP1* in temporal lobe epilepsy (TLE). A total of 496 TLE patients and 528 healthy individuals were successfully genotyped for six *DTNBP1* polymorphisms (rs760665, rs1011313, rs2619528, rs2619522, rs909706, and rs2619538), including 335 TLE patients and 325 healthy controls in cohort 1, and 161 TLE patients and 203 healthy controls in cohort 2. The frequency of the TT genotype at rs909706 T > C was lower in TLE patients than in normal controls in the initial cohort (cohort 1), which was confirmed in an independent cohort (cohort 2). However, the intronic T allele failed to be in linkage disequilibrium (LD) with any functional variations nearby; thus, together with the CCAC and TCAT haplotypes (rs1011313-rs2619528-rs2619522-rs909706) observed in the study, this allele acts only as a protective factor against susceptibility to TLE. Meanwhile, a *novo* mutant allele rs2619538 T > A was exclusively observed in TLE patients, and a dual-luciferase assay revealed that the mutant allele was increased by approximately 22% in the *DTNBP2* promoter compared with the wild-type allele. Together with the trend of increasing *DTNBP1* expression in epilepsy patients and animal models in this study, these are the first findings to demonstrate the genetic association of *DTNBP1* with TLE. Homozygous mutation of rs2619538 T > A likely promotes *DTNBP1* expression and facilitates subsequent processes in epilepsy pathologies. Thus, the role of *DTNBP1* in TLE deserves further exploration in the future.

Keywords: *DTNBP1*, schizophrenia, genetic mutation, temporal lobe epilepsy, polymorphism

INTRODUCTION

Epilepsy is a brain disorder characterized by recurrent seizures. Many gene mutations have been identified via exon sequencing as being responsible for epileptic seizures; these genes involve many ion and non-ion channels, such as *KCNA2, KCNT1, GRIN2A,* and *CHD2* (He et al., 2019). Nevertheless, epilepsy pathologies are difficult to fully identify, making the development of

efficient therapies challenging. Thus, almost one-third of patients, especially those with temporal lobe epilepsy (TLE), a main drug-resistant epilepsy, still suffer from recurrent seizures because of resistance to available treatments. Hence, more efforts should be made to explore epilepsy pathologies, and disclosing epileptic seizures in atypical conditions might be a novel strategy.

In clinical practice, typical symptoms of a disease are often seen as complications due to other diseases, underscoring certain shared mechanisms between them. Considering this phenomenon, some previous studies have explored the manifestation of epileptic seizures in Alzheimer's disease and found that insufficient expression of *ADAM10* is a driving force in epilepsy pathologies (Clement et al., 2008; Prox et al., 2013). The prevalence of epileptic seizures in psychosis has also attracted increased attention in recent years. According to a retrospective cohort study based on the Oxford Record Linkage Study and English national linked Hospital Episode Statistics, patients admitted to the hospital with schizophrenia have an elevated risk of epileptic seizures, and likewise, patients with epilepsy have an elevated risk of schizophrenia (Wotton and Goldacre, 2012). In addition, male adolescents with severe, refractory epilepsy have a high risk of schizophrenia (Andersen et al., 2019). Obviously, this evidence supports a bidirectional relationship and some shared mechanisms between epilepsy and schizophrenia.

Expanding experimental evidence supports the notion of shared mechanisms between epilepsy and schizophrenia. For example, Michelson et al. reported that familial partial trisomy 15q11-13 presented either as intractable epilepsy or as psychiatric illness (Michelson et al., 2011), and Stewart et al. observed that the frequency of copy number abnormalities and maternal duplication 15q11-q13 in patients with combined schizophrenia and epilepsy is significantly higher than in patients with schizophrenia (Stewart et al., 2011). Furthermore, schizophrenia- and epilepsy-related alterations were simultaneously observed in a mouse model with a microdeletion within 15q13.3 (Fejgin et al., 2014). Other studies revealed that abnormal expression of CYFIP1, VRK2, and metabotropic glutamate 2/3 (mGlu2/3) receptor is involved in both schizophrenia and epilepsy (Nebel et al., 2016; Azimi et al., 2018).

The *DTNBP1* gene encodes dystrobrevin binding protein 1, which participates in organelle biogenesis, and recent evidence indicates that it plays key roles in brain development, neuronal excitability, and schizophrenia-related pathologies (Chen et al., 2008; Saggu et al., 2013; Wang et al., 2017; Konopaske et al., 2018; Mohamed et al., 2018), i.e., decreased expression of *DTNBP1* was observed to reduce exocytosis of brain-derived neurotrophic factor (BDNF) from cortical excitatory neurons. The reduction in BDNF exocytosis resulted in a decreasing number of inhibitory synapses located on excitatory neurons, while the application of exogenous BDNF rescued these inhibitory synaptic deficits (Yuan et al., 2016). Interestingly, the hippocampus is an enriched region for *DTNBP1* under normal conditions, and postmortem studies have shown a decreased level of *DTNBP1* expression in the hippocampus of patients with schizophrenia (Wang et al., 2017), indicating that decreased *DTNBP1* is a pathological condition in schizophrenia. Since the hippocampus is also crucial in TLE pathologies characterized by deregulation of neuronal excitability, it is likely that *DTNBP1* plays a key role in both TLE and schizophrenia.

Considering that single-nucleotide polymorphisms are among the most important genetic variations that regulate gene expression and subsequent biological activity, this study selected candidate variations within the *DTNBP1* gene region on the basis of biological plausibility in previous research, including rs2619522 involved in hippocampal gray matter volume (Trost et al., 2013) and rs2619538 in volumes of both gray and white matter (Tognin et al., 2011) in healthy subjects, rs760665 and rs909706 involved in the hippocampal glutamate concentration of healthy individuals (Wirth et al., 2012), and rs1011313 involved in volume reduction of multiple brain regions (Thirunavukkarasu et al., 2014) and rs2619528 in prefrontal function (Fallgatter et al., 2010) of schizophrenia patients. We aimed to clarify the genetic effects of *DTNBP1* in TLE based on these six SNPs.

MATERIALS AND METHODS

Human Subject Enrollment

All experimental protocols involving human subjects were approved by the Ethics Committees of the Affiliated Hospital of Guangdong Medical University, The First Affiliated Hospital of Harbin Medical University, the Central People's Hospital of Zhanjiang, the First People's Hospital of Pingdingshan, and Beijing Tongren Hospital affiliated with Capital Medical University and were executed in accordance with the Declaration of Helsinki. Informed consent documents were signed by participants at the time of human subject enrollment.

A total of 335 TLE patients and 325 healthy controls from the First Affiliated Hospital of Harbin Medical University, the Affiliated Hospital of Guangdong Medical University and Beijing Tongren Hospital affiliated with Capital Medical University were consecutively enrolled in cohort 1. To confirm the initial statistics observed in cohort 1, a validating cohort (cohort 2), including 161 TLE patients and 203 healthy controls from the First People's Hospital of Pingdingshan and the Central People's Hospital of Zhanjiang, was consecutively established. The human subjects in both cohorts 1 and 2 were Han Chinese, and the combined cohort (cohorts 1 + 2) was composed of 496 TLE patients and 528 healthy controls.

Gender, age, age at onset, and severity of disease (seizure frequency and drug response) of all participants were recorded via field investigation or telephone interviews. The inclusion criteria for TLE were based on typical temporal auras and temporal discharges at seizure onset observed by video-electroencephalograph, and subjects who failed to be genotyped were excluded from the study. According to the consensus about drug-resistant epilepsy proposed in 2010 by the Commission of the International League Against Epilepsy (Kwan et al., 2010), the responses to drug treatments were grouped as follows: drug-resistant patients are subjects with the absence of a significant change, an insufficient reduction in seizure frequency (<60%), or even an augmentation after 1 year treatment with a

schedule of not less than two tolerated and properly collected anti-epileptic medicines; the remaining were considered drug-sensitive patients.

DNA Preparation and Genotyping

First, 2 ml peripheral blood samples were collected from each human participant. DNA samples were extracted from blood samples using a Genomic DNA Extraction Kit (Tiangen Biotech, Beijing, China) and then used for genotyping the six *DTNBP1* SNPs (rs760665, rs1011313, rs2619528, rs2619522, rs909706, and rs2619538) via an ABI PRISM SNaPshot system (Applied Biosystems, Carlsbad, CA, United States). The primers used in a multiplex polymerase chain reaction (PCR) for the amplification of the six target fragments (338, 354, 274, 164, 278, and 183 nucleotides, respectively) were as follows: rs760665, 5′- cagccttggcctccaaacat-3′ (forward primer) and 5′-ttagccccagaaatgttattgatttga-3′ (reverse primer); rs1011313, 5′-ccagcatggagacgccaagt-3′ (forward primer) and 5′-tctgaagccttgaacccctcaa-3′ (reverse primer); rs2619528, 5′-tggttgcctgggatggagag-3′ (forward primer) and 5′-cgcactgaaacag tcttttcccatt-3′ (reverse primer); rs2619522, 5′-tgtttgggctctta tgtctacctttcc-3′ (forward primer) and 5′-gcccatttatctgttttccagtgc-3′ (reverse primer); rs909706, 5′-ggtgacaggctaatggcacaca-3′ (forward primer) and 5′-ttctggagtgggctctggactg-3′ (reverse primer); and rs2619538, 5′-ttggatgaggccagtgaggtaa-3′ (forward primer) and 5′-gaggtggtcccccagctacaag-3′ (reverse primer). The primers used for one-nucleotide extension in SNaPshot PCR were 5′-ttttttttttttttttttttttttttttttacaatagccttgctactatttaggggataaaca-3′ (rs760665), 5′-ttttttttttttttcacaggctacagaatggatgttgc-3′ (rs10113-13), 5′-tttggattggatgcacaaccatgtgaa-3′ (rs2619528), 5′-ttttttttttttttttttttttttgcagaagcagtgagtgagagctgaca-3′ (rs2619522), 5′-ttttttttttttttttgtcaagtcagtttccaaggggttctaac-3′ (rs909706), and 5′-ttttttttttttttttttttttttgcaaagcttaataagacagagcagtttacatc-3′ (rs2619538).

The multiplex PCR mix was composed of 1x HotStarTaq buffer, 0.3 mM dNTP, 3.0 mM Mg^{2+}, 1 μl of primer mix, 1 U of HotStarTaq polymerase, and 1 μl of the DNA template. The reaction mix was then run as follows: 1 cycle for 95°C/120 s; 11 cycles for 94°C/20 s, 65°C/40 s, and 72°C/90 s; 24 cycles for 94°C/20 s, 59°C/30 s, and 72°C/90 s; and 1 cycle for 72°C/120 s. Subsequently, the reaction products were purified using shrimp alkaline phosphatase and exonuclease I. The SNaPshot PCR mix contained 5 μl of SNaPshot Multiplex Kit, 1 μl of primer mix, 2 μl of ultrapure H_2O, and 2 μl of the purified reaction products. The experimental program for a one-nucleotide extension was as follows: 1 cycle of 96°C/60 s; 28 cycles of 96°C/10 s, 55°C/5 s, and 60°C/30 s; and 1 cycle of 4°C/120 s. After further purification by shrimp alkaline phosphatase, the final products were analyzed using an ABI 3730xl DNA Analyzer and GeneMapper 4.1 (Applied Biosystems, Carlsbad, CA, United States).

Dual-Luciferase Reporter Assay

In light of the *Homo sapiens* chromosome 6, GRCh37.p13 primary assembly, a 2 kb sequence upstream of the *DTNBP1* transcription start site harboring the T or A alleles at rs2619538 was cloned and individually ligated into pGL3 basic to create p*DTNBP1*-Promoter-Wildtype and p*DTNBP1*-Promoter-Mutant plasmids. In addition to the plasmids of the negative control (NC) and the positive control (PC), both of these plasmids were amplified in DH5α cells. Their positive clones were confirmed by further sequencing. Subsequently, HEK-293T cells were plated at 2×10^4 cells per well in 24-well dishes, and 24 h later, the cells were individually cotransfected with four types of plasmids with the assistance of X-tremeGENE HP reagent (Roche, Basel, Switzerland). These four plasmids included 1 μg of pGL3 basic as the NC, 1 μg of pGL3 promoter as the PC, 1 μg of pSMG6-Promoter-Wildtype (Wild-type), and 1 μg of pSMG6-Promoter-Mutant (Mutant). Each cotransfection reaction was replicated four times, and the pRL plasmid served as an internal reference. After 24 h, the cotransfected cells were identified using a Dual-Luciferase® Reporter Assay System (Promega, Madison, WI, United States), and both the firefly and Renilla luciferase activities were measured using a microplate luminometer (BioTek, VT, United States). The mutant impact of rs2619538 T > A was evaluated by dividing the averaged firefly/Renilla ratio of the mutant construct by the averaged firefly/Renilla ratio of the wild-type construct.

Epilepsy Rat Model

Referring to our previous research (Tao et al., 2018), a total of 15 male Sprague Dawley (SD) rats from the Animal Center of Guangdong Medical University, Zhanjiang, China, were bred for adapting to the experimental environment at a temperature between 22 and 26°C and a humidity between 55 and 65%. The light-dark cycle was synchronous with a natural day-and-night cycle. After a 1 week adaptation period with free access to food and water, the SD rats (310 ± 32 g) were used in the following experiments: 10 rats were randomly selected to be administered pentetrazol (PTZ, 60 mg/kg body weight, i.p.; Sigma-Aldrich, St. Louis, MO, United States). The severity of epileptic seizures was classified in five levels according to the Racine scale: (1) twitching of facial muscle; (2) nodding of the head; (3) unilateral forelimb with lifting or clonus; (4) bilateral forelimb with clonus when standing; (5) falling when standing or twisting (Glen et al., 2014; Orozco et al., 2014). The administration of PTZ was repeated every 20 min during a 2 h observation period (10 mg/kg body weight, i.p.) until seizures increased to level 4 or the total dose of PTZ reached 90 mg/kg body weight. Finally, with the exception of one rat reaching level 4 for a duration of only 3 min, nine rats exhibited seizures that reached level 4 with a duration of 60 min and were immediately enrolled in the experimental group, followed by the administration of diazepam (10 mg/kg body weight, i.p.; Sigma-Aldrich, St. Louis, MO, United States) every 5 min until seizure cessation to reduce unexpected deaths before being sacrificed. The remaining five rats were classified as the control group. All animal experiments in this study were performed in accordance with the "Guide for the Care and Use of Laboratory Animals" (He et al., 2006), which was approved by the Animal Ethics Committee of Guangdong Medical University, Zhanjiang, China.

Molecular Experiments

After a 2 h observation period, all rats in the experimental and control groups were sacrificed through decapitation

under deep anesthesia (3% chloral hydrate, 10 ml/kg body weight, i.p.; Sigma-Aldrich, St. Louis, MO, United States). Then, the left and right hippocampi of nine experimental rats and four control rats were rapidly collected for use in real-time quantitative PCR (qPCR) and enzyme-linked immunosorbent assay (ELISA), respectively: (a) real-time qPCR: total RNA was extracted via an RNA extraction kit (QIAGEN Sciences, Germantown, United States), and reverse transcription was performed by using a First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, Waltham, United States) according to the manufacturer's instructions. Subsequently, the cDNA products were amplified using a Light-Cycler 480 sequence detector system (Roche Applied Science, Penzberg, Germany), and the specific primers used for real-time qPCR were as follows: *DTNBP1* forward primer, 5′-ttagcaggtatgaggatgcgt-3′, and reverse primer, 5′-ggtgcagcaaatggttctctac-3′. Relative expression levels were calculated by the $2^{-\Delta\Delta CT}$ method. And (b) ELISA: *DTNBP1* concentrations in the hippocampi were measured using an ELISA kit (R&D Systems, Minneapolis, MN, United States) according to the manufacturer's instructions. Absorbance was detected using an ELISA reader (Bio-Rad Laboratories, Hercules, CA, United States).

In addition to qPCR and ELISA, immunohistochemistry (IHC) was performed as follows: one experimental rat and one control rat under deep anesthesia were cardially perfused with physiological saline, followed by 4% paraformaldehyde, and their brains were then isolated and fixed in 4% paraformaldehyde at 4°C for 20 h. After dehydration and paraffin embedding, the specimens were cut into 4-μm-thick slices for subsequent histological staining. Paraffin-embedded slices were dewaxed with xylene and ethanol. Deparaffinized slices were rinsed using water, placed in a tissue slide, and heated in a microwave oven for antigen retrieval. Endogenous peroxidase was then blocked using 3% hydrogen peroxide. After three 5 min washes with phosphate buffer saline (PBS), the slices were blocked using 3% bovine serum albumin. Subsequently, the primary antibody (*DTNBP1* polyclonal antibody, ABclonal, Boston, MA, United States) was diluted in blocking solution and incubated overnight at 4°C. The next day, the slices were washed three times for 5 min each in PBS and then incubated with the secondary antibody, washed, developed with 3-3′ diaminobenzidine tetrachloride, and washed to stop the reaction until brown color appeared. The slices were counterstained with hematoxylin, dehydrated and mounted, and finally photographed for biological analysis.

DTNBP1 Expression in TLE Patients

To identify whether *DTNBP1* is abnormally expressed in human epilepsy patients, we used the online GSE datasets because of the lack of pathological brain samples of epilepsy patients in our laboratory. Unfortunately, after a series of searches in the online GSE datasets, we could not identify a single study that could be directly used. However, we found that GSE63808 was based on surgically acquired hippocampi from 129 TLE patients, and GSE29378 contained 32 autopsy samples of normal hippocampi. Moreover, both of these GSE datasets were acquired via the same technique, Illumina HumanHT-12 V3.0 expression BeadChip arrays (Johnson et al., 2015), and therefore, combining these two human transcriptome arrays makes it possible to study the expression of *DTNBP1* between epilepsy patients and normal controls. Hence, GSE63808 and GSE29378 were downloaded from the online GSE datasets. In total, 129 TLE patients from the GSE63808 dataset and 32 normal controls form the GSE29378 dataset were individually enrolled in the TLE and control groups, respectively, in the present study. After normalization and testing of multiple hypotheses to reduce the false-positive rate, the average signals of *DTNBP1* mRNA expression in the hippocampi of the TLE and control groups were compared for further analysis.

Statistical Analyses

Variable data, reported as the mean ± SD, and attribute data were analyzed with a Student's *t*-test and a chi-square test, respectively. Logistic regression was employed to correct for bias related to confounding factors, including age and gender, and q-values were used to adjust the false-positive results from multiple statistics by the Bonferroni correction. Statistical tests were performed using SPSS 19.0 (IBM, New York, NY, United States), and $p \leq 0.05$ was considered statistically significant. In addition, haplotype construction and power analyses were carried out via Haploview 4.2 (Daly Lab, Cambridge, MA, United States) and Quanto 1.2 (University of Southern California, Los Angeles, CA, United States), respectively.

RESULTS

General Characteristics of the Enrolled Cohorts

This study enrolled a total of 496 TLE patients and 528 healthy individuals. No significant differences in gender or age were observed between the TLE patients and the healthy controls in cohorts 1, 2, and 1 + 2 (all $p > 0.05$). Gender, age, duration of disease and severity of disease in cohorts 1 + 2 are listed in **Table 1**.

In the present study, all enrolled patients and controls were successfully genotyped for the six *DTNBP1* SNPs (rs760665, rs1011313, rs2619528, rs2619522, rs909706, and rs2619538). The frequency distributions of these SNPs complied with the Hardy-Weinberg equilibrium. Power analyses indicated that this study had 12.9% power for rs760665, 99.7% power for rs1011313, 97.5% power for rs2619528, 97.5% power for rs2619522, 99.6% power for rs909706 and 87.2% power for rs2619538 for detecting a recessive inheritance model with a genetic effect of 1.8 at a significance level of 0.05 for two-sided type 1 error.

Case-Control Analyses of the Six *DTNBP1* SNPs

In cohorts 1 and 2, the frequencies of the *TT* genotype at rs909706 T > C were consistently lower in the patients than in the controls (**Table 2**: $p = 0.000$ and 0.018, respectively). The same trend was observed after the Bonferroni correction for multiple comparisons in cohorts 1 + 2 (**Table 3**; OR = 1.770, $p = 0.000$ and

TABLE 1 | General characteristics of enrolled cohorts.

	Cases	Controls	*p*-values
Sex (male/female, n)			
Cohort 1	170/165	185/140	0.112
Cohort 2	79/82	104/99	0.682
Cohorts 1 + 2	249/247	289/239	0.147
Age (mean ± SD, years)			
Cohort 1	31.2 ± 14.7	31.0 ± 9.7	0.862
Cohort 2	33.5 ± 14.9	32.7 ± 15.6	0.617
Cohorts 1 + 2	31.5 ± 14.8	31.7 ± 12.3	0.743
Other characteristics in cohorts 1 + 2			
Age at onset (mean ± SD, years)	20.8 ± 13.8	–	–
Severity of disease			
Seizure frequencies (mean ± SD, times/month)	6.7 ± 2.7	–	–
Drug response (sensitive/resistant patients, n)	250/246	–	–

$q = 0.050$), which confirmed the findings in cohorts 1 and 2. These results indicate that the *TT* genotype is associated with protection against TLE. In addition, no consistent differences were found for the remaining five SNPs (rs760665, rs1011313, rs2619528, rs2619522, and rs2619538) in cohorts 1, 2, and 1 + 2 (**Tables 2**, **3**).

Through the use of Haploview 4.2, a haplotype block (rs1011313-rs2619528-rs2619522-rs909706) approximately 27 kb in length was constructed based on all six SNPs including four haplotypes (*CCAT, CCAC, TCAT*, *and CTCC*). Their frequencies were then compared between the patients and the controls in cohorts 1 + 2. As displayed in **Table 4**, the patient ratios of the CCAC and TCAT haplotypes were lower than the control ratio (16.3 vs. 19.2%, $p = 0.000$, and 13.7 vs. 23.8%, $p = 0.006$), which was further confirmed by the Bonferroni correction for multiple comparisons (q = 0.000 and 0.025). Thus, the CCAC and TCAT haplotypes could be genetically protective markers against susceptibility to TLE.

Genetic Significance of the TT Genotype at rs909706

Given the findings of case-control analysis, we further investigated the genetic impact of the wild-type homozygote at rs909706 T > C on age at onset, seizure frequency, and drug response in cohorts 1 + 2. As shown in **Figure 1**, no significant differences in age at onset, seizure frequency or drug response in the TLE patients were found between carriers with the *TT* genotype and those with the *TT* + *TG* genotypes ($p = 0.221$, 0.349, and 0.132), and similar results occurred after adjusting for gender and age ($p = 0.202$, 0.254, and 0.130).

According to the assembly of the human genome GRCh38.p7, rs909706 is located in the first intron of the DTNBP1 gene. However, introns are non-coding sequences, and variations within them rarely modulate gene function. Thus, we decided to investigate whether other functional SNPs are in linkage disequilibrium (LD) with variations at rs909706. As shown in **Figure 1**, rs13195978 was found to be in LD with rs909706, but rs13195978 is still non-functional due to its locus in the second intron of the *DTNBP1* gene. Therefore, there are no clues to explain the protective role of the *TT* genotype at rs909706 against susceptibility to TLE.

Impact of Homozygous Mutation of rs2619538 T > A on *DTNBP1* Transcription

As shown in **Figure 2**, among all the SNPs genotyped in the study, rs2619538 T > A is the only SNP located in the predicted promoter region of the *DTNBP1* gene; thus, its variation might affect *DTNBP1* transcription. Notably, homozygous mutation of rs2619538 T > A was rarely observed in the patient group (1/325) but not in the control group (0/335), suggesting that rs2619538 T > A might be a functional site involved in modulation of *DTNBP1* transcription. Given these clues, we further constructed a genealogical chart of the mutant patient and genotyped the same site of his parents and brothers/sisters, and we identified a *novo* mutant allele in his father. Hence, we precluded its potential impact on the activity of the *DTNBP1* gene promoter.

In support of the mutant impact of rs2619538 T > A on the *DTNBP1* gene promoter, we synthesized reporter gene constructs containing *T* (wild-type) or *A* (mutant) alleles in the context of the full-length promoter of *DTNBP1* (2 kb sequence upstream of the transcription start site). The dual-luciferase assay revealed that both wild-type and mutant constructs have biological activities in the full-length promoter of *DTNBP1*, and the latter increased by approximately 22% in comparison with the former (**Figure 2**). This finding indicates that the homozygous mutation at rs2619538 T > A could upregulate *DTNBP1* transcription.

Expression Patterns of *DTNBP1* in an Epilepsy Rat Model and in TLE Patients

To determine whether *DTNBP1* expression is abnormal in epileptic activity, we established a rat epilepsy model according to a PTZ-induced protocol. *DTNBP1* mRNA and protein expression was detected in the hippocampi by qPCR and ELISA. Compared with control rats, the epilepsy model showed an approximately 87% increase ($p = 0.000$) in the expression of *DTNBP1* mRNA, similar to the increased rate of *DTNBP1* protein (approximately 35%, $p = 0.044$). In addition, IHC confirmed the abnormal expression of *DTNBP1* in the epilepsy model, indicating that hyperfunction of *DTNBP1* is involved in epileptic seizures (**Figure 3**).

To further identify whether *DTNBP1* is abnormally expressed in human epilepsy patients, we used online GSE datasets of pathological brain samples from TLE patients and normal controls that were acquired by the same technique, Illumina HumanHT-12 V3.0 expression BeadChip arrays. Despite the small amplitude increase of 1.74%, the level of *DTNBP1* expression in the hippocampi of the TLE patients was significantly higher than that of the normal controls ($p = 9.316 \times 10^{-9}$), suggesting that homozygous mutation at rs2619538 T > A might be involved in epileptic activity by upregulating *DTNBP1* expression.

TABLE 2 | Inheritance models of the *DTNBP1* SNPs in cohorts 1 and 2.

	Cohort 1				Cohort 2			
	Cases n (%)	**Controls n (%)**	**ORs (95% CI)**	***p*-values**	**Cases n (%)**	**Controls n (%)**	**ORs (95% CI)**	***p*-values**
rs760665 T > A								
TT/TA/AA	335 (100.0)/0.0)/0 (0.0)	323 (99.4)/2 (0.6)/0 (0.0)	–	0.113	161 (100.0)/0 (0.0)/0 (0.0)	202 (99.5)/1 (0.5)/0 (0.0)	–	1.000
TT/TA + AA	335 (100.0)/0 (0.0)	323 (99.4)/2 (0.6)	0.78 (0.57–1.06)	0.999	161 (100.0)/0 (0.0)	202 (99.5)/1 (0.5)	–	1.000
TT + TA/AA	335 (100.0)/0 (0.0)	325 (100.0)/0 (0.0)	–	0.999	161 (100.0)/0 (0.0)	203 (99.5)/0 (0.5)	–	–
rs1011313 C > T								
CC/CT/*TT*	221 (66.0)/98 (29.3)/16 (4.8)	182 (56.0)/131 (40.3)/12 (3.7)	0.75 (0.57–0.98)	0.034	105 (65.2)/48 (29.8)/8 (5.0)	114 (56.2)/81 (39.9)/8 (3.9)	0.78 (0.54–1.12)	0.172
CC/CT + *TT*	221 (66.0)/114 (34.0)	182 (56.0)/143 (4.0)	0.64 (0.47–0.88)	0.006	105 (65.2)/56 (34.8)	114 (56.2)/89 (43.8)	0.68 (0.44–1.04)	0.073
CC + CT/*TT*	319 (95.2)/16 (4.8)	313 (96.3)/12 (3.7)	0.81 (0.37–1.74)	0.585	153 (95.0)/8 (5.0)	195 (96.1)/8 (3.9)	0.79 (0.29–2.19)	0.656
rs2619528 C > T								
CC/CT/*TT*	275 (82.1)/58 (17.3)/2 (0.6)	266 (81.8)/55 (16.9)/4 (1.2)	0.95 (0.66–1.37)	0.772	131 (81.4)/29 (18.0)/1 (0.6)	169 (83.3)/32 (15.8)/2 (1.0)	1.11 (0.67–1.84)	0.697
CC/CT + *TT*	275 (82.1)/60 (17.9)	266 (81.8)/59 (18.2)	0.98 (0.66–1.46)	0.923	131 (81.4)/30 (18.6)	169 (83.3)/34 (16.7)	1.15 (0.67–1.98)	0.617
CC + CT/*TT*	333 (99.4)/2 (0.6)	321 (98.8)/4 (1.2)	2.10 (0.38–11.57)	0.395	160 (99.4)/1 (0.6)	201 (99.0)/2 (1.0)	1.54 (0.14–17.16)	0.727
rs2619522 A > C								
AA/AC/CC	277 (82.7)/56 (16.7)/2 (0.6)	266 (81.8)/55 (16.9)/4 (1.2)	1.09 (0.75–1.58)	0.641	132 (82.0)/28 (17.4)/1 (0.6)	169 (83.3)/32 (15.8)/2 (1.0)	0.94 (0.56–1.56)	0.807
AA/AC + CC	277 (82.7)/58 (17.3)	266 (81.8)/59 (18.2)	0.94 (0.63–1.41)	0.774	132 (82.0)/29 (18.0)	169 (83.3)/34 (16.7)	1.10 (0.64–1.90)	0.729
AA + AC/CC	333 (99.4)/2 (0.6)	321 (98.8)/4 (1.2)	2.10 (0.38–11.57)	0.395	160 (99.4)/1 (0.6)	201 (99.0)/2 (1.0)	1.54 (0.14–17.16)	0.727
rs909706 T > C								
TT/TC/CC	128 (38.2)/169 (50.4)/38 (11.3)	172 (52.9)/121 (37.2)/32 (9.8)	0.69 (0.55–0.87)	0.002	64 (39.8)/78 (48.4)/19 (11.8)	106 (52.2)/76 (37.4)/21 (10.3)	0.74 (0.54–1.00)	0.052
TT/TC + CC	128 (38.2)/207 (61.8)	172 (52.9)/153 (47.1)	1.83 (1.34–2.49)	0.000	64 (39.8)/97 (60.2)	106 (52.2)/97 (47.8)	1.66 (1.09–2.53)	0.018
TT + TC/CC	297 (88.7)/38 (11.3)	293 (90.2)/32 (9.8)	0.88 (0.53–1.45)	0.612	142 (88.2)/19 (11.8)	182 (89.7)/21 (10.3)	0.88 (0.45–1.69)	0.691
rs2619538 T > A								
TT/TA/AA	307 (91.6)/27 (8.1)/1 (0.3)	291 (89.5)/34 (10.5)/0 (0.0)	1.28 (0.75–21.17)	0.363	147 (91.3)/14 (8.7)/0 (0.0)	185 (91.1)/18 (8.9)/0 (0.0)	1.00 (0.48–2.08)	0.993
TT/TA + AA	307 (91.6)/28 (8.4)	291 (89.5)/34 (10.5)	–	–	147 (91.3)/14 (8.7)	185 (91.1)/18 (8.9)	–	–
TT + TA/AA	334 (99.7)/1 (0.3)	323 (100.0)/0 (0.0)	0.78 (0.46-1.33)	0.363	146 (100.0)/0 (0.0)	203 (100.0)/0 (0.0)	1.00 (0.48–2.10)	0.993

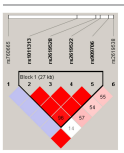
p-values have been adjusted for sex and age.

TABLE 3 | Inheritance models of the SRR SNPs in cohorts 1 + 2.

	Cases n (%)	Controls n (%)	ORs (95% CI)	*p*-values	*q*-values
rs760665 A > T					
AA/AT/*TT*	496 (100.0)/0.0)/0 (0.0)	525 (99.4)/3 (0.6)/0 (0.0)	–	0.999	17.982
AA/AT + *TT*	496 (100.0)/0 (0.0)	525 (99.4)/3 (0.6)	–	0.999	17.982
AA + AT/*TT*	496 (100.0)/0 (0.0)	528 (100.0)/0 (0.0)	–	–	–
rs1011313 C > T					
CC/CT/*TT*	326 (65.7)/146 (29.4)/24 (4.8)	296 (56.1)/212 (40.2)/20 (3.8)	0.76 (0.61–0.94)	0.011	0.198
CC/CT + *TT*	326 (65.7)/170 (34.3)	296 (56.1)/232 (43.9)	0.66 (0.51–0.84)	0.001	0.018
CC + CT/*TT*	472 (95.2)/24 (4.8)	508 (96.2)/20 (3.8)	0.81 (0.44–1.49)	0.497	8.946
rs2619528 C > T					
CC/CT/*TT*	406 (82.1)/87 (17.3)/3 (0.6)	435 (82.4)/87 (16.5)/6 (1.1)	1.00 (0.74–1.35)	0.994	17.892
CC/CT + *TT*	406 (82.1)/90 (17.9)	435 (82.4)/93 (17.6)	1.04 (0.75–1.43)	0.833	14.994
CC + CT/*TT*	493 (99.4)/3 (0.6)	522 (98.9)/6 (1.1)	1.87 (0.47–7.54)	0.377	6.786
rs2619522 A > C					
AA/AC/CC	409 (82.5)/84 (16.9)/3 (0.6)	435 (82.4)/87 (16.5)/6 (1.1)	1.04 (0.77–1.40)	0.814	14.652
AA/AC + CC	409 (82.5)/87 (17.5)	435 (82.4)/93 (17.6)	0.99 (0.72–1.37)	0.973	17.514
AA + AC/CC	493 (99.4)/3 (0.6)	522 (98.9)/6 (1.1)	1.87 (0.47–7.54)	0.377	6.786
rs909706 T > C					
TT/TC/CC	192 (38.7)/247 (49.8)/57 (11.5)	278 (52.7)/197 (37.3)/53 (10.0)	0.71 (0.59–0.85)	0.000	0.005
TT/TC + CC	192 (38.7)/304 (61.3)	278 (52.7)/250 (47.3)	1.77 (1.38–2.27)	0.000	0.000
TT + TC/CC	439 (88.5)/57 (11.5)	475 (90.0)/53 (10.0)	0.88 (0.59–1.30)	0.517	9.306
rs2619538 T > A					
TT/TA/AA	454 (91.5)/41 (8.3)/1 (0.2)	476 (90.2)/52 (9.8)/0 (0.0)	1.14 (0.75–1.74)	0.539	9.702
TT/TA + AA	454 (91.5)/42 (8.5)	476 (90.2)/52 (9.8)	0.85 (0.56–1.31)	0.464	8.352
TT + TA/AA	495 (99.8)/1 (0.2)	528 (100.0)/0 (0.0)	–	1.000	18.000

ORs and p-values have been adjusted for sex and age; considering the number of statistical comparisons reached 18 times, Bonferroni correction was employed to adjust the false positive rate, and the q-value of each comparison was equal to 18 times its p-value.

TABLE 4 | Haplotypes of the *DTNBP1* SNPs in cohorts 1 + 2.

rs1011313-rs2619528-rs2619522-rs909706	Haplotypes	Frequency ratios (%)	Case ratios (%)	Control ratios (%)	*p*-values	*q*-values
	CCAT	44.8	47.5	44.3	0.213	0.852
	CCAC	26.6	19.2	16.3	0.000	0.000
	TCAT	18.8	23.8	13.7	0.006	0.025
	CTCC	9.1	9.9	9.4	0.813	3.253

q-values were calculated with Bonferroni correction; considering the number of statistical comparisons reached 4 times, Bonferroni correction was employed to adjust the false positive rate, and the q-value of each comparison was equal to 4 times its p-value.

DISCUSSION

This study first observed that the frequencies of the *TT* genotype at rs909706 T > C and the *CCAC* and *TCAT* haplotypes (rs1011313-rs2619528-rs2619522-rs909706) were significantly lower in TLE patients than in controls. A previous study reported that the concentration of hippocampal glutamate was significantly affected by variations in rs909706 T > C (Wirth et al., 2012), whereas in our analyses, the intronic *TT* genotype was not related to age at onset or severity of disease and was not in LD with any functional variations nearby. Hence, these findings are considered to indicate protective markers against genetic susceptibility to TLE.

Meanwhile, the homozygous mutation of rs2619538 T > A was rarely observed in TLE patients, and the *novo* mutant allele came from the father. The dual-luciferase assay revealed that *DTNBP2* promoter activity in patients with the mutant allele increased by approximately 22% compared with patients with the wild-type allele. Together with the trend of increased *DTNBP1* expression in epilepsy patients and animal models in the present study, we hypothesized that the homozygous mutation of rs2619538 T > A functions by promoting DTNBP1 expression and subsequent processes in epilepsy pathologies.

According to previous research, the homozygous mutation at rs2619538 T > A is associated with reduced volumes of both gray and white matter in healthy children as young as 10–12 years old (Tognin et al., 2011). Structural volume deficits were observed in cortical regions, the subiculum and dentate gyrus, and the striatum of *DTNBP1* mutant mice (Wirth et al., 2012). These findings indicate a key role of *DTNBP1* in brain development, and abnormalities of brain development often lead to cortical dysplasia and subsequent early conset

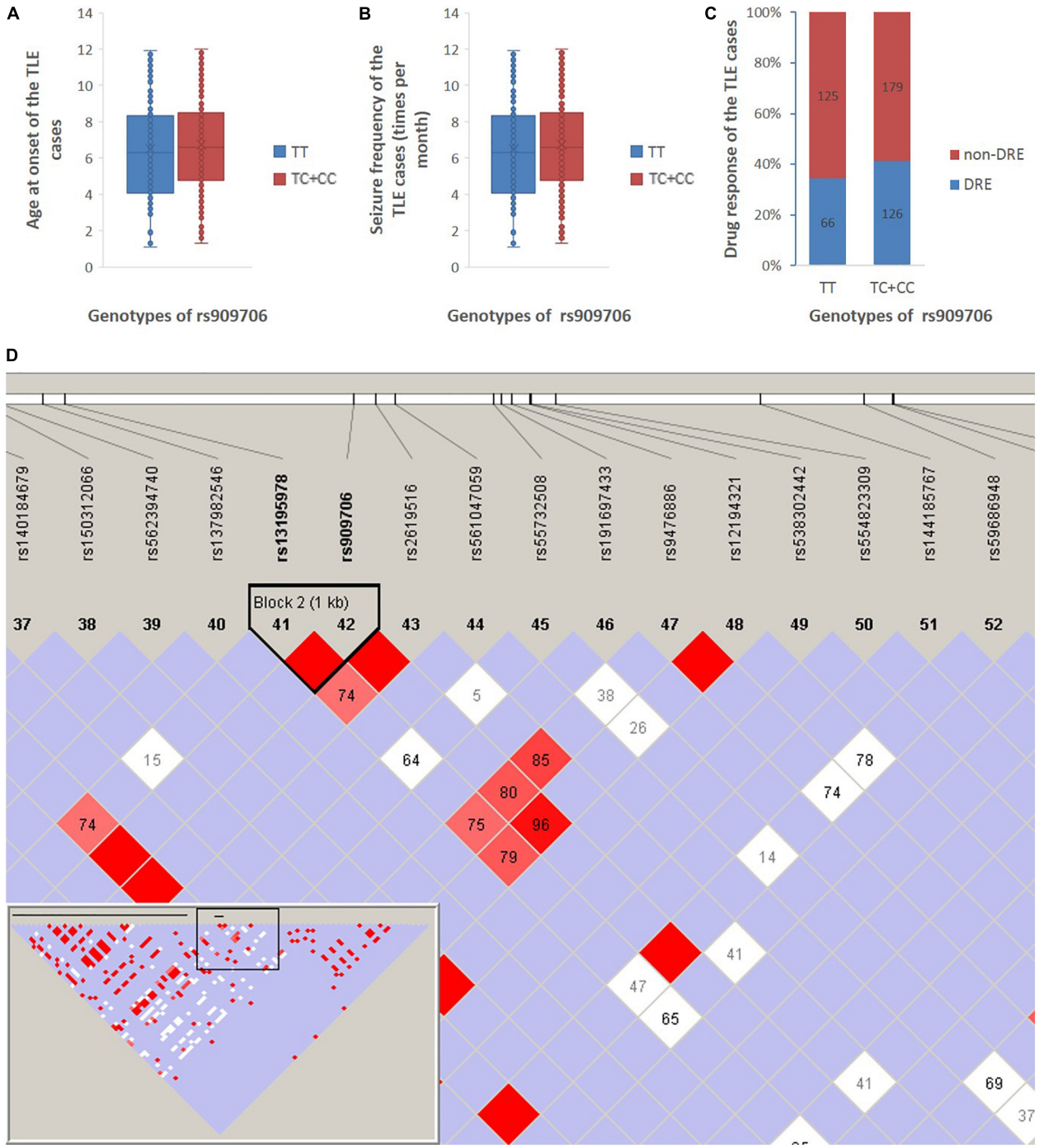


FIGURE 1 | Age at onset, seizure frequency, and drug response between the *TT* genotype and the *TC* + *CC* genotypes in all TLE patients in cohort 1 + 2. **(A)** The age at onset of the TLE patients was 22 ± 16 and 20 ± 13 years among those with the *TT* genotype and the *TC* + *CC* genotypes, respectively. **(B)** Seizure frequency of patients with the *TT* genotype and the *TC* + *CC* genotypes was 6.5 ± 2.7 and 6.7 ± 2.6 per month, respectively. **(C)** Drug-resistant incidence of patients with the *TT* genotype and the *TC* + *CC* genotypes was 34.4 and 41.1%, respectively. **(D)** According to the 1,000 Genomes Project, SNPs are downloaded between 10,000 nucleotides downstream and 10,000 nucleotides upstream of the rs909706 locus (Chromosome 6, NC_000006.11 (15650871..15670871), and according to the analysis using Haploview 4.2, only rs13195978 was observed to be in LD with rs909706 (*D'* = 1.0).

seizures (Palmini and Holthausen, 2013; Liu et al., 2015; Shaker et al., 2016; Maynard et al., 2017). On the other hand, *DTNBP1* inhibits the release of glutamate (Chen et al., 2008; Saggu et al., 2013), which is crucial for AMPAR-mediated synaptic transmission and plasticity and NMDAR-dependent synaptic potentiation in the hippocampus (Glen et al., 2014;

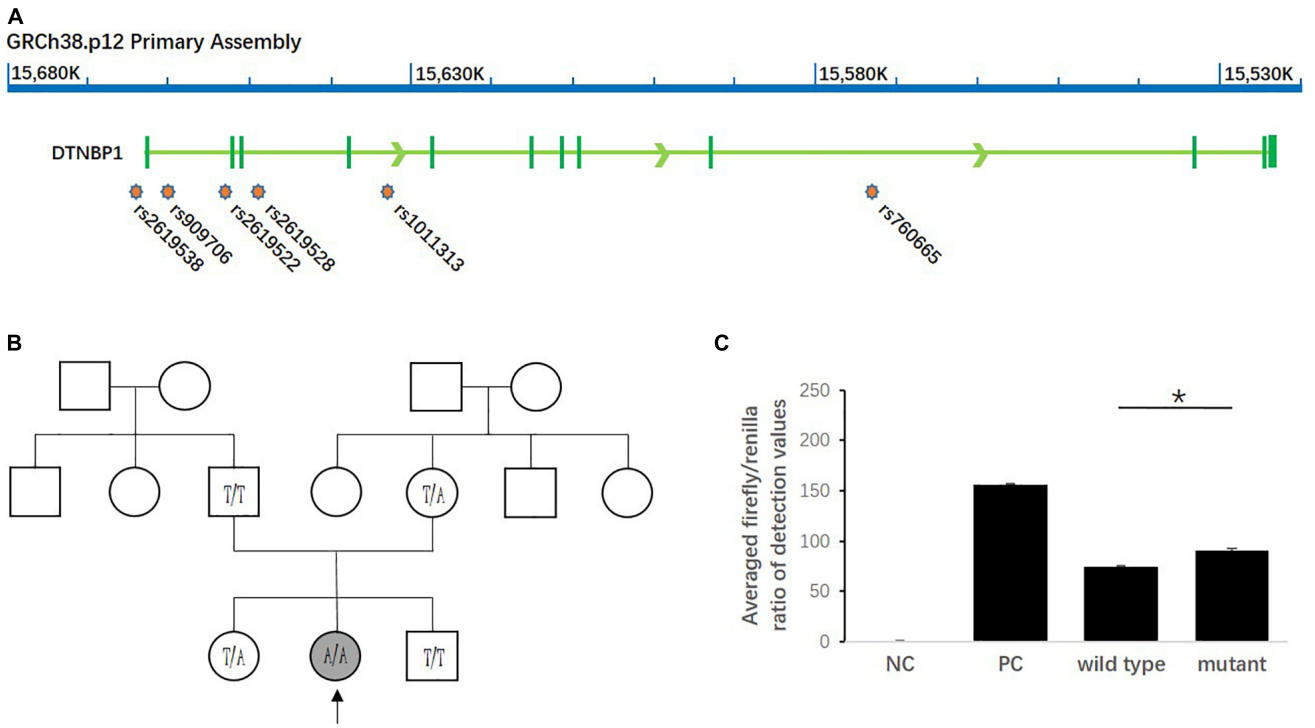


FIGURE 2 | Impact of the homozygous mutation of rs2619538 T > A on DTNBP1 transcription. **(A)** Relative loci of rs2619538 T > A and other SNPs genotyped in this study are shown in the DNTBP1 gene region. Only rs2619538 T > A is potentially functional because of its location in the predicted promoter of the DNTBP1 gene, and yet the remaining SNPs are located within several introns of the DTNBP1 gene. **(B)** The TLE patient carrying the homozygous mutation of rs2619538 T > A was a 33-years-old male (age at onset, 32 years). After observation of the treatment for 1 year, seizure frequencies decreased from three times per month to one time per half year. A genealogical chart of his family was also constructed, and his parents consented to genotyping of the same site at rs2619538 T > A; a *novo* mutant allele probably coming from his father was observed. **(C)** The effects of rs2619538 T > A mutant on the activity of the DTNBP1 gene promoter. Compared with the averaged firefly/renilla ratios of the pGL3 basic negative control (NC), the ratios of the pGL3 promoter positive control (PC), pDTNBP1-Promoter-Wildtype (Wildtype) and pDTNBP1-Promoter-Mutant (Mutant) for rs2619538 T > A were 155.74 ± 1.43, 74.07 ± 0.99, and 90.48 ± 2.09, respectively. *$p < 0.05$.

Orozco et al., 2014). In addition, *DTNBP1* null gene mutation was observed to influence the developmental switch between GluN2B and GluN2A in the mouse cortex and hippocampus (Sinclair et al., 2016). These findings further indicate that *DTNBP1* should be a critical modulator in excitatory signal transduction of glutamate, as well as brain development. Considering that an epilepsy model induced by PTZ is usually used to observe changes of excitatory/inhibitory regulatory molecules during the period of acute convulsions, this study successfully used the model to confirm increased DTNBP1 expression in the PTZ model, which indicates that the model can be used for disclosing potential mechanisms of DTNBP1 in future experiments.

Notably, *DTNBP1* facilitates neurite outgrowth by promoting the transcriptional activity of p53 (Ma et al., 2011), and nucleocytoplasmic shuttling of *DTNBP1* regulates synapsin I expression (Fei et al., 2010). However, neither overexpression (Dys1A-Tg) nor underexpression (Sandy) of *DTNBP1* cause epileptic seizures in mice (Shintani et al., 2014). Because these prior studies focused on schizophrenia, more attention was paid to the effects of low-level *DTNBP1* expression. In contrast, the expression of *DTNBP1* was demonstrated to be increased in TLE patients and an animal model in this study; thus, behavioral observations and electrophysiological results of epileptic seizures in Dys1A-Tg mice and the related mechanisms should be further evaluated in the future.

Several limitations of the present study should be noted. First, *DTNBP1* encodes dysbindin-1, which is composed of three spliceosomes, but only dysbindin-1B, not dysbindin-1A or dysbindin-1C, displays a tendency for toxic aggregation. In postmortem brains, dysbindin-1B not only aggregates with itself but also aggregates with proteins that interact with it (Xu et al., 2015; Zhu et al., 2015; Yang et al., 2016). Another study observed that increased expression of dysbindin-1A resulted in a selective deficit in NMDA receptor signaling in the hippocampus (Jeans et al., 2011). Additionally, dysbindin-1C is required for the survival of hilar mossy cells and the maturation of adult newborn neurons in the dentate gyrus (Wang et al., 2014), and dysbindin-1C deficiency could result in impaired autophagy (Yuan et al., 2015). Therefore, the expression patterns of the three spliceosomes most likely interfere with the function of *DTNBP1* and influence genetic susceptibility to TLE. Second, when we consider 1,000 genomes, the minor allele frequencies of rs2619538 T > A are 0.027 among those of east Asian ancestry, 0.645 among those of African ancestry, 0.549 among those of European ancestry and 0.42 among those of American ancestry, which indicates that the mutant effects

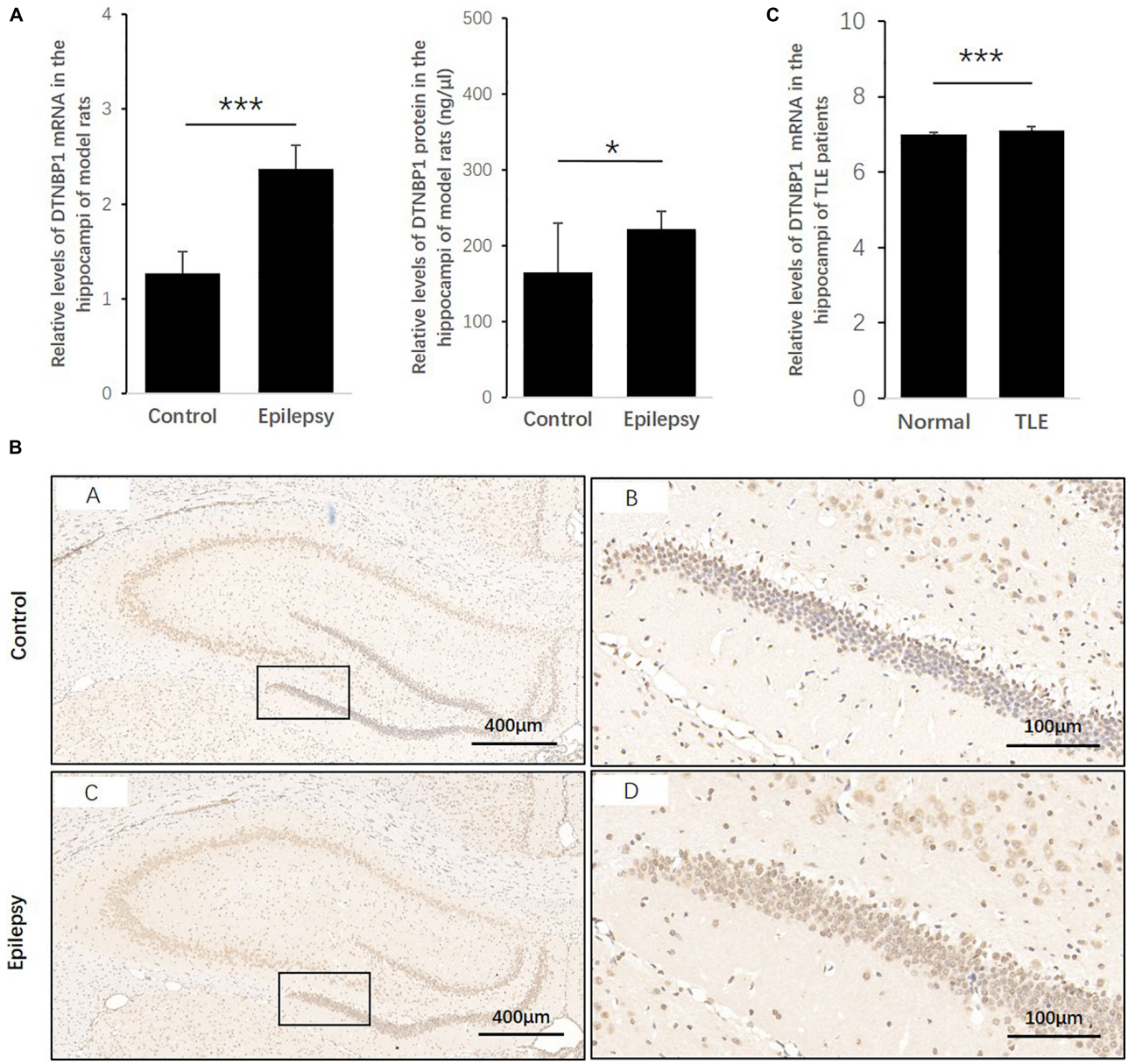


FIGURE 3 | Expression patterns of *DTNBP1* in an epilepsy rat model and TLE patients. **(A)** According to qPCR, the relative levels of *DTNBP1* mRNA in the hippocampi of normal rats and epilepsy model rats were 1.27 ± 0.23 and 2.37 ± 0.25, respectively. **(B)** According to ELISA, the concentrations of *DTNBP1* protein in the hippocampi of normal rats and epilepsy model rats were 164.60 ± 64.93 and 222.42 ± 23.20 ng/μl, respectively. **(C)** A total of 129 TLE patients from the GSE63808 dataset and 32 normal controls from the GSE29378 dataset were enrolled in the TLE and control groups. After normalization and multiple hypothesis testing to reduce the false-positive rate, the average *DTNBP1* mRNA expression in the hippocampi of the TLE and control groups was 7.11 ± 0.09 and 6.99 ± 0.07, respectively. **(D)** IHC staining is shown individually for global and local regions of the hippocampi. Hematoxylin staining was used to mark the cell nuclei (blue), and immunostaining was used to show *DTNBP1* expression (brown). $^{*}p < 0.05$, $^{***}p < 0.001$.

of rs2619538 T > A in the Han Chinese population included in the present study probably represent a mechanism that is distinct from those of other ancestries. Thus, one should be cautious in generalizing our findings to other races. Third, in addition to SNPs at rs2619538 T > A, DNA methylation of the *DTNBP1* promoter also plays a key role in the expression of *DTNBP1* and related pathological activities (Wockner et al., 2014; Abdolmaleky et al., 2015); thus, DNA methylation is another potential factor that could interfere with the genetic effects of rs2619538 T > A.

CONCLUSION

This study first demonstrated the association of *DTNBP1* with TLE from a genetic perspective. In particular, the homozygous

mutation rs2619538 T > A was observed in TLE patients but not in healthy controls. The increased activities of the *DTNBP1* promoter with the A allele in dual-luciferase assays and increased *DTNBP1* expression in epilepsy patients and animal models suggest that the mutation likely functions by promoting transcription of the *DTNBP1* gene and facilitating subsequent processes in epilepsy pathologies. Hence, the role of *DTNBP1* in TLE deserves further exploration in the future.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

All experimental protocols involving human subjects were approved by the Ethics Committees of the Affiliated Hospital of Guangdong Medical University, The First Affiliated Hospital of Harbin Medical University, the Central People's Hospital of Zhanjiang, the First People's Hospital of Pingdingshan, and Beijing Tongren Hospital affiliated with Capital Medical University. The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by the Ethics Committees of Guangdong Medical University. Written informed consent was obtained from the owners for the participation of their animals in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

HT and XZ undertook data analyses and wrote the manuscript. JC, HZ, LH, and YCa carried out biological experiments. JF, ZL, and YCh carried out specimen collection. CS carried out epilepsy model. BZ, WZ, and KL conceptualized the hypothesis and designed the study. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the Postdoctoral Science Foundation of Guangdong Medical University (Grant No. 701z20150002) and the Natural Science Foundation of Guangdong Province (Grant No. 2017A030310658).

REFERENCES

Abdolmaleky, H., Pajouhanfar, S., Faghankhani, M., Joghataei, M., Mostafavi, A., and Thiagalingam, S. (2015). Antipsychotic drugs attenuate aberrant DNA methylation of DTNBP1 (dysbindin) promoter in saliva and post-mortem brain of patients with schizophrenia and Psychotic bipolar disorder. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* 168, 687–696. doi: 10.1002/ajmg.b.32361

Andersen, K., Petersen, L., Vestergaard, M., Pedersen, C., and Christensen, J. (2019). Premature mortality in persons with epilepsy and schizophrenia: a population-based nationwide cohort study. *Epilepsia* 60, 1200–1208. doi: 10.1111/epi.15158

Azimi, T., Ghafouri-Fard, S., Davood Omrani, M., Mazdeh, M., Arsang-Jang, S., Sayad, A., et al. (2018). Vaccinia related Kinase 2 (VRK2) expression in neurological disorders: schizophrenia, epilepsy and multiple sclerosis. *Multi. Sclerosis Relat. Disord.* 19, 15–19. doi: 10.1016/j.msard.2017.10.017

Chen, X., Feng, Y., Hao, C., Guo, X., He, X., Zhou, Z., et al. (2008). DTNBP1, a schizophrenia susceptibility gene, affects kinetics of transmitter release. *J. Cell Biol.* 181, 791–801. doi: 10.1083/jcb.200711021

Clement, A. B., Hanstein, R., Schröder, A., Nagel, H., Endres, K., Fahrenholz, F., et al. (2008). Effects of neuron-specific ADAM10 modulation in an in vivo model of acute excitotoxic stress. *Neuroscience* 152, 459–468.

Fallgatter, A. J., Ehlis, A., Herrmann, M., Hohoff, C., Reif, A., Freitag, C., et al. (2010). DTNBP1 (dysbindin) gene variants modulate prefrontal brain function in schizophrenic patients–support for the glutamate hypothesis of schizophrenias. *Genes Brain Behav.* 9, 489–497. doi: 10.1111/j.1601-183X.2010.00574.x

Fei, E., Ma, X., Zhu, C., Xue, T., Yan, J., Xu, Y., et al. (2010). Nucleocytoplasmic shuttling of dysbindin-1, a schizophrenia-related protein, regulates synapsin I expression. *J. Biol. Chem.* 285, 38630–38640. doi: 10.1074/jbc.M110.107912

Fejgin, K., Nielsen, J., Birknow, M., Bastlund, J., Nielsen, V., Lauridsen, J., et al. (2014). A mouse model that recapitulates cardinal features of the 15q13.3 microdeletion syndrome including schizophrenia- and epilepsy-related alterations. *Biol. Psychiatry* 76, 128–137. doi: 10.1016/j.biopsych.2013.08.014

Glen, W., Horowitz, B., Carlson, G., Cannon, T., Talbot, K., Jentsch, J., et al. (2014). Dysbindin-1 loss compromises NMDAR-dependent synaptic plasticity and contextual fear conditioning. *Hippocampus* 24, 204–213. doi: 10.1002/hipo.22215

He, N., Lin, Z., Wang, J., Wei, F., Meng, H., Liu, X., et al. (2019). Evaluating the pathogenic potential of genes with de novo variants in epileptic encephalopathies. *Genet. Med.* 21, 17–27. doi: 10.1038/s41436-018-0011-y

He, Z., Li, G., Zhu, D., and Lu, S. (2006). *Guide for the Care and Use of Laboratory Animals, Science Press*, 1–469.

Jeans, A., Malins, R., Padamsey, Z., Reinhart, M., and Emptage, N. (2011). Increased expression of dysbindin-1A leads to a selective deficit in NMDA receptor signaling in the hippocampus. *Neuropharmacology* 61, 1345–1353. doi: 10.1016/j.neuropharm.2011.08.007

Johnson, M. R., Behmoaras, J., Bottolo, L., Krishnan, M. L., Pernhorst, K., Santoscoy, P. L. M., et al. (2015). Systems genetics identifies Sestrin 3 as a regulator of a proconvulsant gene network in human epileptic hippocampus. *Nat. Commun.* 6:6031. doi: 10.1038/ncomms7031

Konopaske, G., Balu, D., Presti, K., Chan, G., Benes, F., and Coyle, J. (2018). Dysbindin-1 contributes to prefrontal cortical dendritic arbor pathology in schizophrenia. *Schizophren. Res.* 201, 270–277. doi: 10.1016/j.schres.2018.04.042

Kwan, P., Arzimanoglou, A., Berg, A. T., Brodie, M. J., Allen Hauser, W., Mathern, G., et al. (2010). Definition of drug resistant epilepsy: consensus proposal by the ad hoc Task Force of the ILAE commission on therapeutic strategies. *Epilepsia* 51, 1069–1077. doi: 10.1111/j.1528-1167.2009.02397.x

Liu, W., An, D., Xiao, J., Li, J., Hao, N., and Zhou, D. (2015). Malformations of cortical development and epilepsy: a cohort of 150 patients in western China. *Seizure* 32, 92–99.

Ma, X., Fei, E., Fu, C., Ren, H., and Wang, G. (2011). Dysbindin-1, a schizophrenia-related protein, facilitates neurite outgrowth by promoting the transcriptional activity of p53. *Mol. Psychiatry* 16, 1105–1116. doi: 10.1038/mp.2011.43

Maynard, L., Leach, J., Horn, P., Spaeth, C., Mangano, F., Holland, K., et al. (2017). Epilepsy prevalence and severity predictors in MRI-identified focal cortical dysplasia. *Epilepsy Res.* 132, 41–49.

Michelson, M., Eden, A., Vinkler, C., Leshinsky-Silver, E., Kremer, U., Lerman-Sagie, T., et al. (2011). Familial partial trisomy 15q11-13 presenting as intractable epilepsy in the child and schizophrenia in the mother. *Eur. J. Paediatr. Neurol.* 15, 230–233. doi: 10.1016/j.ejpn.2010.11.001

Mohamed, Z., Tee, S., and Tang, P. (2018). Association of functional polymorphisms in 3′-untranslated regions of COMT, DISC1, and DTNBP1 with schizophrenia: a meta-analysis. *Psychiatr. Genet.* 28, 110–119. doi: 10.1097/ypg.0000000000000210

Nebel, R., Zhao, D., Pedrosa, E., Kirschen, J., Lachman, H., Zheng, D., et al. (2016). Reduced CYFIP1 in human neural progenitors results in dysregulation of schizophrenia and epilepsy gene networks. *PLoS One* 11:e0148039. doi: 10.1371/journal.pone.0148039

Orozco, I. J., Koppensteiner, P., Ninan, I., and Arancio, O. (2014). The schizophrenia susceptibility gene DTNBP1 modulates AMPAR synaptic transmission and plasticity in the hippocampus of juvenile DBA/2J mice. *Mol. Cell. Neurosci.* 58, 76–84. doi: 10.1016/j.mcn.2013.12.003

Palmini, A., and Holthausen, H. (2013). Focal malformations of cortical development: a most relevant etiology of epilepsy in children. *Handb. Clin. Neurol.* 111, 549–565.

Prox, J., Bernreuther, C., Altmeppen, H., Grendel, J., Glatzel, M., D'Hooge, R., et al. (2013). Postnatal disruption of the disintegrin/metalloproteinase ADAM10 in brain causes epileptic seizures, learning deficits, altered spine morphology, and defective synaptic functions. *J. Neurosci.* 33, 12915–12928.

Saggu, S., Cannon, T., Jentsch, J., and Lavin, A. (2013). Potential molecular mechanisms for decreased synaptic glutamate release in dysbindin-1 mutant mice. *Schizophren. Res.* 146, 254–263. doi: 10.1016/j.schres.2013.01.037

Shaker, T., Bernier, A., and Carmant, L. (2016). Focal cortical dysplasia in childhood epilepsy. *Semin. Pediatr. Neurol.* 23, 108–119.

Shintani, N., Onaka, Y., Hashimoto, R., Takamura, H., Nagata, T., Umeda-Yano, S., et al. (2014). Behavioral characterization of mice overexpressing human dysbindin-1. *Mol. Brain* 7:74. doi: 10.1186/s13041-014-0074-x

Sinclair, D., Cesare, J., McMullen, M., Carlson, G., Hahn, C., and Borgmann-Winter, K. (2016). Effects of sex and DTNBP1 (dysbindin) null gene mutation on the developmental GluN2B-GluN2A switch in the mouse cortex and hippocampus. *J. Neurodev. Disord.* 8:14. doi: 10.1186/s11689-016-9148-7

Stewart, L., Hall, A., Kang, S., Shaw, C., and Beaudet, A. (2011). High frequency of known copy number abnormalities and maternal duplication 15q11-q13 in patients with combined schizophrenia and epilepsy. *BMC Med. Genet.* 12:154. doi: 10.1186/1471-2350-12-154

Tao, H., Zhou, X., Xie, Q., Ma, Z., Sun, F., Cui, L., et al. (2018). SRR intronic variation inhibits expression of its neighbouring SMG6 gene and protects against temporal lobe epilepsy. *J. Cell. Mol. Med.* 22, 1883–1893. doi: 10.1111/jcmm.13473

Thirunavukkarasu, P., Vijayakumari, A., John, J., Halahalli, H., Paul, P., Sen, S., et al. (2014). An exploratory association study of the influence of dysbindin and neuregulin polymorphisms on brain morphometry in patients with schizophrenia and healthy subjects from South India. *Asian J. Psychiatry* 10, 62–68. doi: 10.1016/j.ajp.2014.04.002

Tognin, S., Viding, E., McCrory, E., Taylor, L., O'Donovan, M., McGuire, P., et al. (2011). Effects of DTNBP1 genotype on brain development in children. *J. Child Psychol. Psychiatry All. Discipl.* 52, 1287–1294. doi: 10.1111/j.1469-7610.2011.02427.x

Trost, S., Platz, B., Usher, J., Scherk, H., Wobrock, T., Ekawardhani, S., et al. (2013). The DTNBP1 (dysbindin-1) gene variant rs2619522 is associated with variation of hippocampal and prefrontal grey matter volumes in humans. *Eur. Archiv. Psychiatry Clin. Neurosci.* 263, 53–63. doi: 10.1007/s00406-012-0320-0

Wang, H., Xu, J., Lazarovici, P., and Zheng, W. (2017). Dysbindin-1 involvement in the etiology of schizophrenia. *Intern. J. Mol. Sci.* 18:2044. doi: 10.3390/ijms18102044

Wang, H., Yuan, Y., Zhang, Z., Yan, H., Feng, Y., and Li, W. (2014). Dysbindin-1C is required for the survival of hilar mossy cells and the maturation of adult newborn neurons in dentate gyrus. *J. Biol. Chem.* 289, 29060–29072. doi: 10.1074/jbc.M114.590927

Wirth, C., Schubert, F., Lautenschlager, M., Brühl, R., Klär, A., Majic, T., et al. (2012). DTNBP1 (dysbindin) gene variants: in vivo evidence for effects on hippocampal glutamate status. *Curr. Pharm. Biotechnol.* 13, 1513–1521. doi: 10.2174/138920112800784952

Wockner, L., Noble, E., Lawford, B., Young, R., Morris, C., Whitehall, V., et al. (2014). Genome-wide DNA methylation analysis of human brain tissue from schizophrenia patients. *Transl. Psychiatry* 4:e339. doi: 10.1038/tp.2013.111

Wotton, C. J., and Goldacre, M. (2012). Coexistence of schizophrenia and epilepsy: record-linkage studies. *Epilepsia* 53, e71–e74. doi: 10.1111/j.1528-1167.2011.03390.x

Xu, Y., Sun, Y., Ye, H., Zhu, L., Liu, J., Wu, X., et al. (2015). Increased dysbindin-1B isoform expression in schizophrenia and its propensity in aggresome formation. *Cell Discov.* 1:15032. doi: 10.1038/celldisc.2015.32

Yang, W., Zhu, C., Shen, Y., and Xu, Q. (2016). The pathogenic mechanism of dysbindin-1B toxic aggregation: BLOC-1 and intercellular vesicle trafficking. *Neuroscience* 333, 78–91. doi: 10.1016/j.neuroscience.2016.07.008

Yuan, Q., Yang, F., Xiao, Y., Tan, S., Husain, N., Ren, M., et al. (2016). Regulation of brain-derived neurotrophic factor exocytosis and gamma-aminobutyric acidergic interneuron synapse by the schizophrenia susceptibility gene Dysbindin-1. *Biol. Psychiatry* 80, 312–322. doi: 10.1016/j.biopsych.2015.08.019

Yuan, Y., Wang, H., Wei, Z., and Li, W. (2015). Impaired autophagy in hilar mossy cells of the dentate gyrus and its implication in schizophrenia. *J. Genet. Genom.* 42, 1–8. doi: 10.1016/j.jgg.2014.12.001

Zhu, C., Shen, Y., and Xu, Q. (2015). Propagation of dysbindin-1B aggregates: exosome-mediated transmission of neurotoxic deposits. *Neuroscience* 291, 301–316. doi: 10.1016/j.neuroscience.2015.02.016

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychiatry

ORIGINAL RESEARCH
published: 08 March 2021
doi: 10.3389/fpsyt.2021.631475

Association Between *OLIG2* Gene SNP rs1059004 and Negative Self-Schema Constructing Trait Factors Underlying Susceptibility to Depression

Hiroshi Komatsu[1,2*], *Hikaru Takeuchi*[3], *Chiaki Ono*[4], *Zhiqian Yu*[4], *Yoshie Kikuchi*[4], *Yoshihisa Kakuto*[2], *Shunichi Funakoshi*[2,5], *Takashi Ono*[2], *Ryuta Kawashima*[3,6], *Yasuyuki Taki*[3,7,8] and *Hiroaki Tomita*[1,4,9,10,11]

[1] Department of Psychiatry, Tohoku University Hospital, Sendai, Japan, [2] Miyagi Psychiatric Center, Natori, Japan, [3] Division of Developmental Cognitive Neuroscience, Institute of Development, Aging and Cancer, Tohoku University, Sendai, Japan, [4] Department of Disaster Psychiatry, International Research Institute of Disaster Science, Tohoku University, Sendai, Japan, [5] Department of Community Psychiatry, Tohoku University, Sendai, Japan, [6] Smart Ageing International Research Center, Institute of Development, Aging and Cancer, Tohoku University, Sendai, Japan, [7] Department of Nuclear Medicine and Radiology, Institute of Development, Aging and Cancer, Tohoku University, Sendai, Japan, [8] Division of Medical Neuroimaging Analysis, Department of Community Medical Supports, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan, [9] Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan, [10] Department of Disaster Psychiatry, Graduate School of Medicine, Tohoku University, Sendai, Japan, [11] Department of Psychiatry, Graduate School of Medicine, Tohoku University, Sendai, Japan

Recent evidence has indicated that the disruption of oligodendrocytes may be involved in the pathogenesis of depression. Genetic factors are likely to affect trait factors, such as characteristics, rather than state factors, such as depressive symptoms. Previously, a negative self-schema had been proposed as the major characteristic of constructing trait factors underlying susceptibility to depression. Thus, the association between a negative self-schema and the functional single nucleotide polymorphism (SNP) rs1059004 in the *OLIG2* gene, which influences *OLIG2* gene expression, white matter integrity, and cerebral blood flow, was evaluated. A total of 546 healthy subjects were subjected to genotype and psychological evaluation using the Beck Depression Inventory-II (BDI-II) and the Brief Core Schema Scale (BCSS). The rs1059004 SNP was found to be associated with the self-schema subscales of the BCSS and scores on the BDI-II in an allele dose-dependent manner, and to have a predictive impact on depressive symptoms via a negative-self schema. The results suggest the involvement of a genetic factor regulating oligodendrocyte function in generating a negative-self schema as a trait factor underlying susceptibility to depression.

Keywords: depressive symptoms, negative self-schema, single nucleotide polymorphism, *OLIG2*, oligodendrocyte

OPEN ACCESS

Edited by:
Elena Martín-García, Pompeu Fabra University, Spain

Reviewed by:
Shwetal Mehta, Barrow Neurological Institute (BNI), United States
Chuntao Zhao, Cincinnati Children's Hospital Medical Center, United States

***Correspondence:**
Hiroshi Komatsu
hkomatsu1019@gmail.com

Specialty section:
This article was submitted to Behavioral and Psychiatric Genetics, a section of the journal Frontiers in Psychiatry

Received: *20 November 2020*
Accepted: *05 February 2021*
Published: *08 March 2021*

Citation:
Komatsu H, Takeuchi H, Ono C, Yu Z, Kikuchi Y, Kakuto Y, Funakoshi S, Ono T, Kawashima R, Taki Y and Tomita H (2021) Association Between OLIG2 Gene SNP rs1059004 and Negative Self-Schema Constructing Trait Factors Underlying Susceptibility to Depression. Front. Psychiatry 12:631475. doi: 10.3389/fpsyt.2021.631475

INTRODUCTION

Major depressive disorder (MDD) is a common disease with a 12-month and lifetime prevalence of 6.6 and 16.2%, respectively (1). Depressive disorders, including MDD and dysthymia, are also a leading cause of the global disease burden (2). Although effective treatments (e.g., antidepressants and cognitive behavioral therapy) are available for MDD, remission rates for antidepressants are not currently optimally achieved (3). Therefore, the development of new and more effective drugs is required in the future and elucidating the pathology of MDD could help us develop these novel medicines. Depressive disorders are multifactorial diseases that are thought to develop through interactions between genetic and psychosocial factors. Although the neurobiological mechanism underlying MDD has been extensively investigated, the definitive biological mechanism remains unclear.

Prior studies based on post-mortem brain, neuroimaging, and animal studies have indicated the involvement of disrupted oligodendrocytes in the pathology of MDD. Histological examination of the post-mortem brain revealed that patients with MDD have a substantially reduced density and number of oligodendrocytes in their prefrontal cortex (4, 5). Oligodendrocyte-related genes were also found to be reduced in the temporal frontal cortex and in the white matter of the ventral prefrontal cortex in patients with MDD (6, 7). More recent studies have reported that psychological stress results in the downregulation of oligodendrocyte-related genes and reduced myelination fiber length and density in the brains of mice (8, 9). Cathomas et al. (8) revealed that after psychological stress, emotion, and microglial activity were altered in mice that were heterozygous for the cyclic nucleotide phosphodiesterase (Cnp1) oligodendrocyte gene compared to wild-type mice; this suggests that oligodendrocyte gene expression affects behavioral changes after psychosocial stress. Several diffusion tensor imaging studies have documented decreased white matter integrity in the corpus callosum and several frontal, temporal, and parietal regions in patients with MDD (10–13). In addition, other studies have found microstructural abnormalities in some fiber tracts, such as anterior callosal fibers, in patients with MDD (14, 15). Patients with MDD also show a reduced magnetic transfer ratio, which reflects demyelination in brain regions, such as the frontal and striatal regions, limbic areas, occipital white matter, and the genu and splenium of the corpus callosum (16–18). Evidence found in post-mortem brains, neuroimaging, and animal studies suggests that disrupted oligodendrocyte function may be involved in impaired mood regulation in MDD.

The *OLIG2* gene is a basic helix-loop-helix transcription factor that is expressed exclusively in oligodendrocytes and oligodendrocyte precursors and is involved in oligodendrocyte differentiation (19, 20). One post-mortem brain study found that *OLIG2* gene expression was downregulated in the temporal cortex of patients with MDD (6). The *OLIG2* gene single nucleotide polymorphism (SNP) rs1059004 is a functional SNP that influences *OLIG2* gene expression and influences white matter integrity in Caucasian populations (21–23). A more recent study revealed that SNP rs1059004 affects resting-state cerebral blood flow in a broad region of the brain as well as white matter integrity in the Japanese population (23). Genetic association studies have shown that *the OLIG2* SNP rs1059004 is associated with several psychiatric disorders such as schizophrenia and obsessive-compulsive disorder in a certain population (24–28). However, no previous studies have investigated the genetic association between this SNP and MDD.

Beck proposed the cognitive model of depression, in which negative self-schema individuals are vulnerable to developing depression in the future (29). Previous studies have shown a marked association between the schemata concerning the self and others, and the severity of depressive symptoms (30, 31). Evans et al. revealed that a negative self-schema was a risk factor for the development of depression in women (32). The results of these studies support the cognitive model of depression hypothesized by Beck (29).

Based on prior evidence indicating the involvement of oligodendrocyte abnormalities in the pathology of MDD, the present study hypothesized the association between *OLIG2* SNP rs1059007 and the negative-self core schema, as it plays a major role in constructing trait factors underlying susceptibility to MDD by influencing the function of oligodendrocytes. To verify this hypothesis, we investigated whether *the OLIG2* SNP rs1059007 was associated with self-negative core schema in healthy subjects.

MATERIALS AND METHODS

Subjects

A total of 777 healthy, right-handed individuals were recruited, and their genotyping data, mood states, core beliefs, brain imaging data, cognitive function, aging, genetics, and daily habits were examined as detailed below and elsewhere (33–36). Of the 777 participants, all gene polymorphism data, scores on the Beck Depression Inventory-II (BDI-II), and scores on the Brief Core Schema Scale (BCSS) were successfully obtained from 546 subjects (313 men and 233 women; 20.5 $\pm$ 1.8 years of age). All subjects had normal vision and were university, college or postgraduate students, or subjects who had graduated from these institutions within 1 year prior to the experiment. None of the participants had a history of neurological or psychiatric illness. Handedness was evaluated using the Edinburgh Handedness Inventory (37). High-molecular-weight DNA was isolated from saliva specimens using Oragene containers (DNA Genotek Inc., Ottawa, Canada), according to the manufacturer's instructions. After the study procedures were fully explained, written informed consent was obtained from all participants in accordance with the Declaration of Helsinki (1991). This study was approved by the Ethics Committee of Tohoku University.

Psychological Assessments

Participants were administered the Japanese version of the Brief Core Schema Scale (BCSS) to assess core schemas about the self and others. The BCSS is a 24-item self-report scale developed by Fowler et al. to measure core schemata with regard to the self and others (30). The BCSS includes four dimensions of self- and other-assessment: negative self, positive self, negative others, and

positive others. Each dimension comprised six items that were evaluated on a five-point rating scale (0–4), and respondents were asked to indicate "yes" or "no" for whether they held each belief. If they held the belief, they were then asked to indicate the degree of their conviction on a 4-point scale ranging from 1 to 4 (1: believe slightly, 2: believe it moderately, 3: believe it very much, 4: believe it totally).

Participants were also tested using the Japanese version of the BDI-II to evaluate the degree of depressive symptoms. BDI-II, developed by Beck et al. (38), is a 21-item self-report questionnaire used to evaluate the severity of depression in normal and psychiatric subjects (39). Each item consisted of a 4-point scale from 0 (symptoms absent) to 3 (severe symptoms). The minimum and maximum BDI-II scores were 0 and 63, respectively with higher scores indicating greater severity of depressive symptoms. In normal subjects, scores above 20 have been reported to indicate depression (40).

SNP Genotyping

Genotyping of *OLIG2* SNP rs1059004 was carried out using TaqMan assays (Applied Biosystems, Foster City, CA, USA). Polymerase chain reactions (PCRs) were performed using 20 ng of genomic DNA, 40 × TaqMan Probe Assay Mix (Probe ID: C_2442961_10) (Applied Biosystems, Waltham, MA, USA), 2 × Universal PCR Master Mix (Applied Biosystems), and nuclease-free water in a 10 μL total reaction volume; allele-specific fluorescence was measured using the CFX96 Real-Time System (Bio-Rad, Hercules, CA, USA). Information about the TaqMan Probe sequence can be obtained from https://www.thermofisher.com/order/genome-database/details/genotyping/C___2442961_20?CID=&ICID=&subtype=. The PCR cycle conditions consisted of an initial denaturation at 95°C for 10 min, followed by 50 cycles at 92°C for 15 s and at 57°C for 1 min. Then, PCR products spanning the SNP were amplified with primers (forward: gagcgctgtctggctttaac, reverse: gaggaacggccacagttcta) from two representative subjects for each of the three genotypes of this SNP and were subjected to direct sequencing to validate the TaqMan assay-based genotyping.

Statistical Analysis

Statistical evaluations were performed using SPSS statistics 24 and Amos 23 (Japan IBM, Tokyo, Japan) software packages. Demographic variables among groups were compared using χ^2-tests, analysis of variance, *t*-tests, or Kruskal-Wallis test, where appropriate. Deviations in genotype distribution from the Hardy-Weinberg equilibrium (HWE) were assessed using the χ^2-test for goodness of fit. Spearman's bivariate correlation analysis was performed to examine the correlations between BDI-II scores and the four BCSS subscales. The Jonckheere-Terpstra test was performed to investigate whether *the OLIG2* SNP rs1059004 was associated with the scores on the four BCSS subscales or those of BDI-II in an allele dose-dependent manner. After the Jonckheere-Terpstra test, the *post-hoc* test was used to compare the differences between each genotype using the Bonferroni corrections.

The associations between *the OLIG2* SNP rs1059004, self-schema, and depressive symptoms were assessed using path analysis. Structural equation modeling was performed to evaluate the association between the above-mentioned three variables using SPSS and Amos 23, and model fits were estimated using the maximum likelihood method. The two initial hypothetical models shown in **Figure 1** were constructed based on the results of the trend test, bivariate correlation analysis, and a previous study (32). In the two initial models shown in **Figure 1**, the self-schema affects depressive symptoms, and *OLIG2* SNP rs1059004 impacts schema and depressive symptoms in one direction. In contrast, a negative schema influences or is affected by a positive schema about the self. We removed paths scoring $p > 0.05$ from the initial models and examined whether the model fit improved. The best-fitting path model was ultimately adopted in the path analysis. Chi-square statistics were used to test the goodness-of-fit model, and the following fit indices were calculated: goodness-of-fit index (GFI), adjusted GFI (AGFI), comparative fit index (CFI), Akaike information criterion (AIC), and root mean square error of approximation (RMSEA).

To investigate whether *the OLIG2* SNP rs1059004 influenced the severity of depressive symptoms and core schema about the self and others in a gender-dependent manner, two-way analysis of covariance (ANCOVA) controlling for age, with genotype and gender as independent variables, was used to assess differences in scores on the BDI-II and on the four BCSS subscales between A-allele carriers (AA genotype + AC genotype) and non-A-allele carriers (CC genotype). As few individuals carried the AA genotype, subjects carrying the AA genotype were combined with those carrying the AC genotype, and the differences in scores on the BDI-II and BCSS subscales were compared between carriers and non-carriers of the A allele. *Post-hoc* analyses were performed using Bonferroni corrections. Statistical significance was defined as a two-tailed $p < 0.05$.

RESULTS

There were significant differences in age and in positive, negative, and other subscale scores between men and women (**Table 1**).

The genotype distribution of the subjects was as follows: homozygous A allele ($n = 14$, 2.5%), heterozygous A/C ($n = 155$, 28.3%), and homozygous C allele ($n = 377$, 69%), which did not deviate from the HWE ($\chi^2 = 0.33$, $p > 0.05$). There were no significant differences in age or sex among the three *OLIG2* genotypic groups (**Table 2**).

Meanwhile, the Kruskal-Wallis test showed significant differences in BDI-II and the negative and positive self subscales of the BCSS among *OLIG2* genotypes ($p = 0.03$, $p = 0.042$, $p = 0.015$, respectively, **Table 2**). The Cronbach's α values for the BDI-II total score, the negative and positive self subscale scores, and the negative and positive other subscale scores were 0.82, 0.80, 0.82, 0.78, and 0.86, respectively.

Bivariate Associations Between BDI-II and Four BCSS Subscale Scores

Spearman's bivariate correlation analysis revealed a positive association between BDI-II score and the negative self

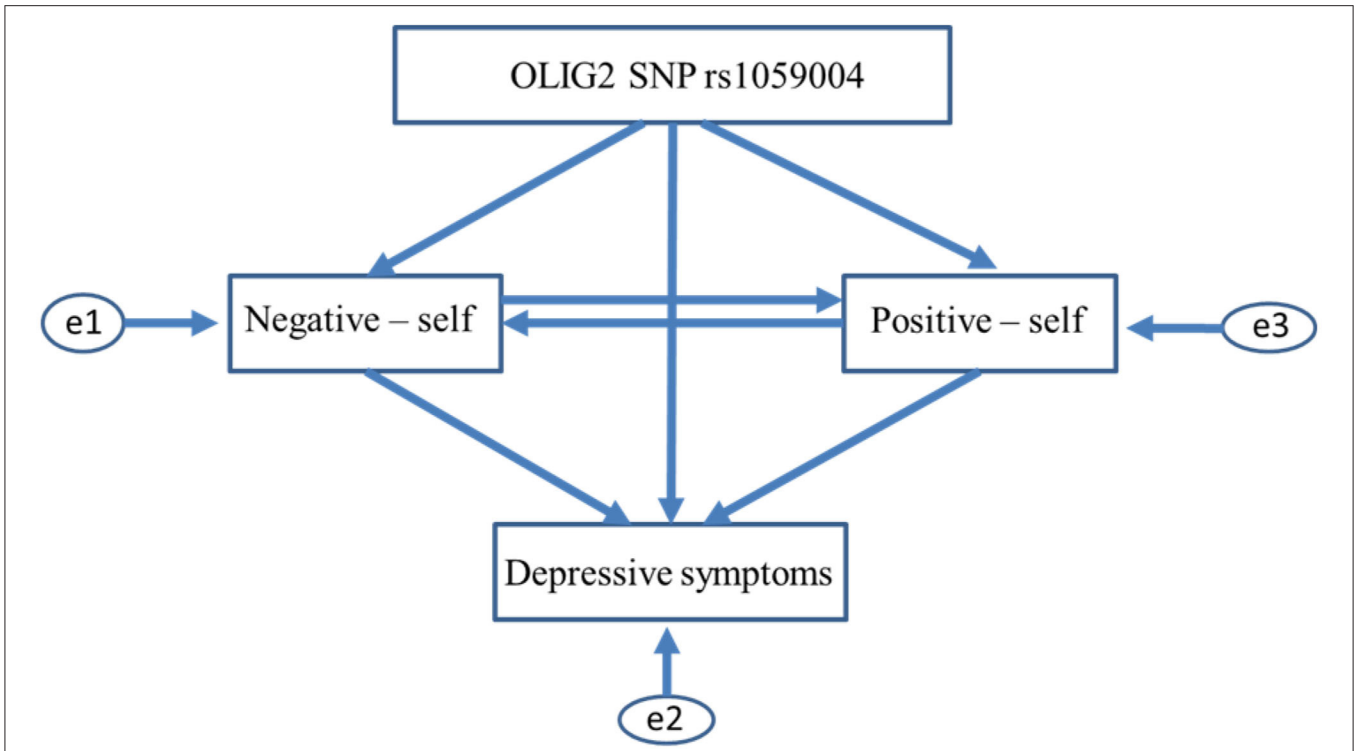


FIGURE 1 | Initial path models between *OLIG2* gene SNP rs1059004, self-core schema, and depressive symptoms. The two initial path models (AIC = 20, CFI = 1, RMSEA = 0.363 in both models) shown in **Figure 1** assume that *OLIG2* SNP rs1059004 affects negative self-schema (negative self), positive self-schema (positive self), and depressive symptoms in one direction. In contrast, hypothetical paths were not created among each self-schema. SNP, single nucleotide polymorphism; AIC, Akaike information criterion; CFI, comparative fit index; RMSEA, root mean square error of approximation.

TABLE 1 | Demographics of all variables in the healthy subjects.

	Men (*N* = 313)	Women (*N* = 233)	*t*-value	*p*-value[a]	All subjects (*N* = 546)
Age (mean ± sd, range)	20.7 ± 1.9, 18–27	20.3 ± 1.5, 18–26	2.622	**0.009**	20.5 ± 1.8, 18–27
BDI-II score (mean ± sd, range)	8.0 ± 6.2, 0–31	8.4 ± 6.3, 0–31	−0.78	0.436	8.2 ± 6.2, 0–31
BCSS subscale scores					
Negative self score (mean ± sd, range)	5.2 ± 4.4, 0–23	4.8 ± 4.3, 0–20	0.887	0.461	5.0 ± 4.3, 0–23
Positive self score (mean ± sd, range)	6.0 ± 4.3, 0–23	5.2 ± 3.8, 0–23	2.219	**0.04**	5.6 ± 4.1, 0–23
Negative other score (mean ± sd, range)	2.3 ± 3.1, 0–19	1.6 ± 2.5, 0–15	2.755	**0.003**	2.0 ± 2.9, 0–19
Positive other score (mean ± sd, range)	8.0 ± 4.4, 0–20	8.9 ± 5.0, 0–24	−2.236	**0.024**	8.4 ± 4.7, 0–24

The table represents the mean ± sd of age and BDI-II and the BCSS scores among males and females, in addition to all other variables for all participants. The p-values and t-values from t-tests are also shown.
BDI-II, Beck Depression Inventory-II; BCSS, Brief Core Schema Scale; sd, standard deviation.
[a]*t-test between men and women.*

and negative other subscales of the BCSS ($r = 0.636$ and $r = 0.296$, respectively, $p < 0.001$, **Table 3**). There was a significant negative association between BDI-II score and the positive self and positive other subscales of the BCSS ($r = -0.459$ and $r = -0.306$, respectively, $p < 0.001$, **Table 3**).

Association Between OLIG2 SNP rs1059004 and the Severity of Depressive Symptoms

The present study found a significant association between *the OLIG2* SNP rs1059004 and BDI-II score in an allele dose-dependent manner (Jonckheere-Terpstra test, $p = 0.022$,

TABLE 2 | The demographics of subjects subdivided according to *OLIG2* genotype.

	Genotype			
	C/C	**C/A**	**A/A**	***p*-value**
N	377 (69.0%)	155 (28.3%)	14 (2.5%)	
Age (mean ± sd, range)	20.5 ± 1.8, 18–27	20.6 ± 1.8, 18–27	20.3 ± 1.6, 18–24	0.864
Gender (male/female)	221/156	87/68	5/9	0.221
BDI-II (mean ± sd, range)	11 ± 5.6, 0–31	8.9 ± 6.6, 0–31	7.9 ± 6.1, 1–20	0.03
Scores on BCSS subscales				
Negative self score (mean ± sd, range)	4.8 ± 4.3, 0–23	5.7 ± 4.5, 0–20	5.8 ± 4.3, 0–13	0.042
Positive self score (mean ± sd, range)	5.9 ± 4.1, 0–23	5.4 ± 4.2, 0–18	3.1 ± 3.6, 0–11	0.015
Negative other score (mean ± sd, range)	2.1 ± 3.0, 0–19	2.0 ± 2.8, 0–15	1.5 ± 2.2, 0–8	0.847
Positive other score (mean ± sd, range)	8.3 ± 4.7, 0–24	10.5 ± 4.9, 0–21	10.4 ± 4.5, 3–17	0.17

The table shows mean ± sd for age and gender for each OLIG2 SNP rs1059004 genotype. P-values for the analysis of variance (ANOVA) are used to compare the differences in age among OLIG2 genotypes, and the chi-square tests between gender and OLIG2 genotype were represented in the table. The table shows p-values for the Kruskal-Wallis test to compare the difference in BDI-II and BCSS subscales among OLIG2 genotypes.
sd, standard deviation.

TABLE 3 | Spearman's bivariate correlation analysis of age, BDI-II score, and BCSS scores.

	Age	**BDI-II**	**Negative self**	**Positive self**	**Negative other**	**Positive other**
Age	–					
BDI-II	−0.057	–				
Negative self	−0.062	0.636**	–			
Positive self	0.096	−0.459**	−0.404**	–		
Negative other	−0.033	0.296**	0.357**	0.009	–	
Positive other	−0.050	−0.306**	−0.202**	0.296**	−0.349**	–

The table exhibits Spearman's correlation coefficients between age and scores on the BDI-II and the four BCSS subscales.
BDI-II, Beck Depression Inventory-II; BCSS, Brief Core Schema Scale.
***Bonferroni-corrected p < 0.001.*

Figure 2A). *Post-hoc* analysis showed that AA genotype carriers had a significantly higher BDI-II score than the CC genotype carriers (Bonferroni-corrected $p = 0.041$, **Figure 2A**). Two-way ANCOVA adjusted for age, with genotype and gender as the fixed factors showed that there was a significant main effect of genotype on BDI-II score (**Table 4**). However, the genotype-gender interaction was not significant for the BDI-II score (**Table 4**).

Association Between OLIG2 SNP rs1059004 and Core Schema About the Self and Others

The Jonckheere-Terpstra test showed that *the OLIG2* SNP rs1059004 was associated with both negative and positive self-schema in an allele dose-dependent manner (Jonckheere-Terpstra test, $p = 0.012$, $p = 0.037$; **Figures 2B,C**). Subjects with the AC genotype had a significantly higher negative-self subscale score than those with the CC genotype (Bonferroni-corrected $p = 0.026$, **Figure 2B**). Individuals carrying the AA genotype had a significantly lower positive-self subscale score than those with the CC genotype and AC genotype, respectively (Bonferroni-corrected $p = 0.009$, $p = 0.049$, **Figure 2C**).

Two-way ANCOVA controlling for age and with genotype and gender as the fixed factors showed that there was a significant main effect of genotype on negative-self subscale score (**Table 4**). However, no significant genotype-gender interactions were observed for scores on the four BCSS subscales (**Table 4**).

Path Analysis of the Relationship Between OLIG2 rs1059004, the Self-Core Schema, and Depressive Symptoms

We removed paths scoring $p > 0.05$ from the initial models (AIC = 20, CFI = 1, RMSEA = 0.363 in both models) and examined whether the model fit improved. Ultimately, the path model shown in **Figure 3** produced the best results in the model fit evaluations (Chi-square = 2.669, df = 2, $P = 0.263$; GFI = 0.998; AGFI = 0.988; CFI = 0.998; AIC = 18.669; RMSEA = 0.025). The number of C alleles of the *OLIG2* gene SNP rs1059004

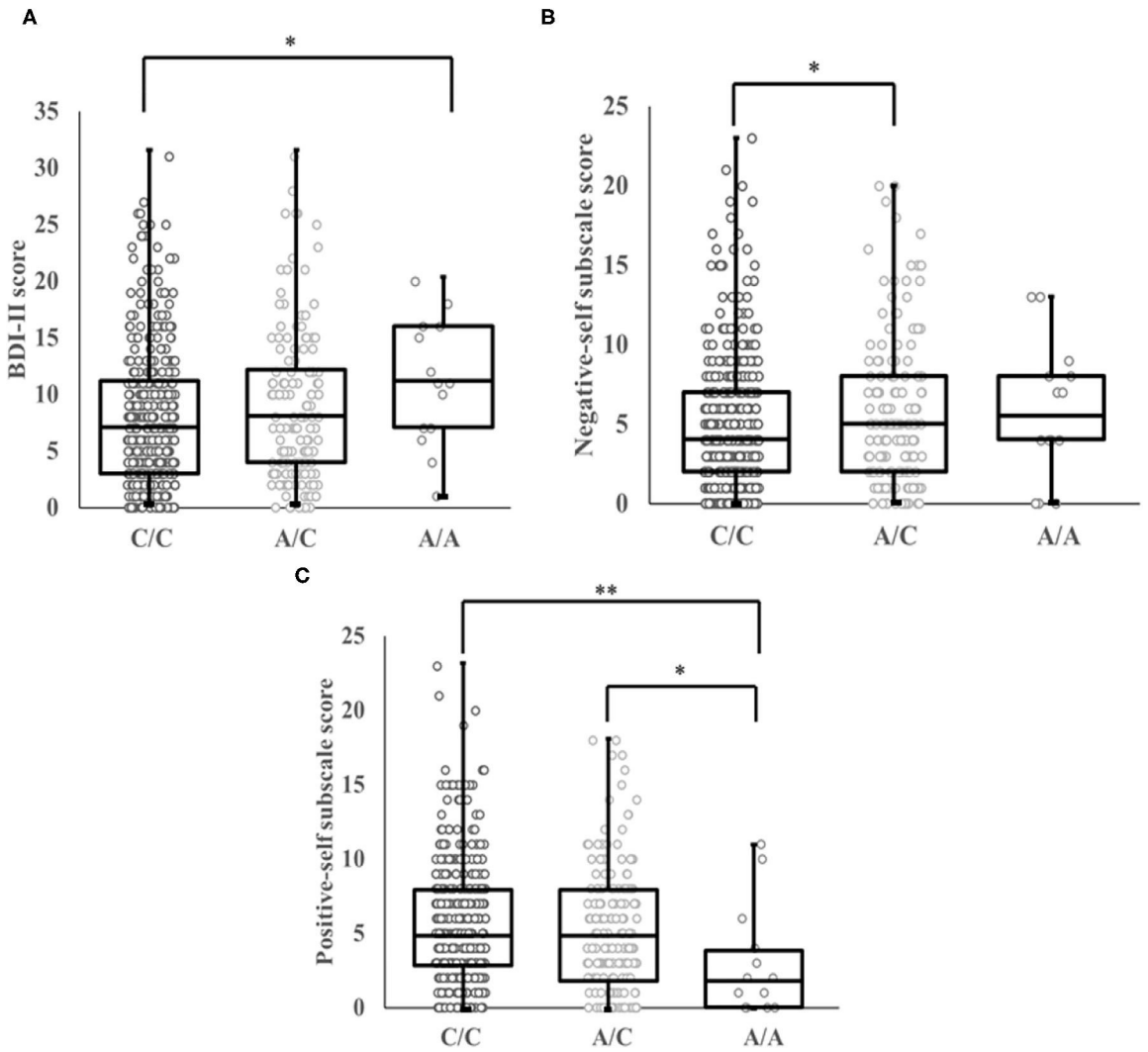


FIGURE 2 | (A) Association between genotype at SNP rs1059004 in the *OLIG2* gene and BDI-II scores. There were significant correlations between A allele dosage and BDI-II scores among the three genotype groups. BDI-II scores decreased in a C-allele dose-dependent manner ($p = 0.022$, Jonckheere-Terpstra test). AA genotype carriers had a significantly higher BDI-II score than the CC genotype carriers (Bonferroni-corrected $p = 0.041$). **(B)** Association between genotype at SNP rs1059004 in the *OLIG2* gene and negative-self subscale scores. Significant gene-dose associations between the SNP rs1089004 and negative-self subscale scores were observed among the three genotype groups. Negative-self subscale scores decreased in a C-allele dose-dependent manner ($p = 0.012$, Jonckheere-Terpstra test). Subjects with AC genotype had a significantly higher negative-self subscale score compared to those with CC genotype ($p = 0.026$). **(C)** Association between genotype at SNP rs1059004 and positive-self subscale scores. The Jonckheere-Terpstra test also showed significant associations between A allele dosage and positive-self subscale scores among the three genotype groups. Positive-self subscale scores increased in a C-allele dose-dependent manner ($p = 0.037$, Jonckheere-Terpstra test). People carrying the AA genotype had a significantly lower positive-self subscale score than those with the CC genotype and AC genotype, respectively (Bonferroni-corrected $p = 0.009$, $p = 0.049$). The bars of the box plot show the range from minimum scores to maximum scores. SNP, single nucleotide polymorphism. $^{*}p < 0.05$, $^{**}p < 0.01$ compared to the CC genotype.

was negatively associated with the negative self-schema, and the schema was assumed to affect the positive self-schema and lead to a higher severity of depressive symptoms (**Figure 3**).

DISCUSSION

Based on prior evidence that the disruption of oligodendrocytes may be implicated in the pathology of MDD, the present study hypothesized an association between the *OLIG2* SNP rs1059004 and the negative self-schema, constructing trait factors underlying susceptibility to MDD. To verify this hypothesis, the associations between *the OLIG2* SNP rs1059004, BCSS and BDI-II scores were examined in 546 healthy subjects. Consistent with the above hypothesis, the number of C alleles in *the OLIG2* gene SNP rs1059004 was associated with decreased BDI-II scores and the negative BCSS self-schema subscale in an allele

TABLE 4 | The results of two-way ANCOVA controlling for age, with genotype and gender as independent variables.

	Men		Women		Genotype		Gender		Genotype × Gender	
	Non-A-allele carriers (CC genotype)	A-allele carriers (AA genotype +AC genotype)	Non-A-allele carriers (CC genotype)	A-allele carriers (AA genotype +AC genotype)	*F*	*p*	*F*	*p*	*F*	*p*
N	221	92	156	77						
BDI-II (mean ± sd)	7.88 ± 6.03	8.48 ± 6.63	7.86 ± 6.30	9.74 ± 6.28	4.569	**0.033**	0.889	0.346	1.219	0.27
Scores on subscales of BCSS										
Negative self schema score (mean ± sd)	4.98 ± 4.33	5.79 ± 4.61	4.51 ± 4.29	5.62 ± 4.41	5.626	**0.018**	0.81	0.368	0.131	0.718
Positive self schema score (mean ± sd)	6.18 ± 4.28	5.59 ± 4.52	5.40 ± 3.78	4.82 ± 3.81	2.449	0.118	2.988	0.084	0	0.998
Negative other schema score (mean ± sd)	2.41 ± 3.20	2.09 ± 2.99	1.56 ± 2.56	1.75 ± 2.42	0.055	0.815	5.444	**0.02**	0.911	0.34
Positive other schema score (mean ± sd)	7.79 ± 4.39	8.73 ± 4.69	8.92 ± 5.05	9.13 ± 5.18	1.722	0.19	2.492	0.115	0.664	0.416

The table displays differences in the scores on the BDI-II and the BCSS by gender between A-allele carriers and non-A allele carriers. F-values and p-values for the main effect of genotype, gender and their interaction effect are also represented in the table.
BDI-II, Beck Depression Inventory-II; BCSS, Brief Core Schema Scale; sd, standard deviation.

dose-dependent manner. Path analysis revealed that a negative self-schema mediated the association between *the OLIG2* SNP rs1059004 and the severity of depressive symptoms. The results of the present study indicate that *the OLIG2* SNP rs1059004 has a predictive impact on depressive symptoms via a negative schema of the self. In addition, the results suggest the involvement of a genetic factor regulating oligodendrocyte function in generating a negative-self schema that plays a major role in constructing factors underlying susceptibility to depression.

Although previous studies have not found significant differences in scores for any BCSS subscale between men and women (30, 31), the current study found that gender created significant differences in positive, negative, and "other" BCSS subscale scores. This discrepancy in the results may be due to differences in the ethnicities, number, and/or ages of the participants. The findings of the present study indicate that the positive self-schema may be affected by a variety of experiences during young adulthood in the Japanese population. In contrast, four of the BCSS subscales were significantly correlated with the severity of depressive symptoms, consistent with previous reports (30, 31).

Previous genetic association studies have exclusively focused on genes related to monoamines, the hypothalamic-pituitary-adrenal axis, and glutamatergic neurotransmitters. For example, polymorphisms of the glucocorticoid receptor gene, monoamine oxidase A gene, and group-2 metabotropic glutamate receptor gene have been previously reported to be associated with MDD (41–44). Although polymorphisms of oligodendrocyte-related genes have been shown to be associated with schizophrenia (24), no previous studies have investigated the genetic association between the polymorphisms of oligodendrocyte-related genes and MDD. This is the first study to reveal that *the OLIG2* SNP rs1059004 was associated with the severity of depressive symptoms and with negative self-schema. Whether oligodendrocyte disruption is a causative factor in mood dysregulation remains unclear, but several studies have suggested that changes in oligodendrocyte function and structure can influence neural circuits that mediate mood regulation in mice (8, 45–47). For example, mice exposed to cuprizone, a mouse model of demyelination, showed deceased anxiety-like behavior. Moreover, mice lacking the Cnp1 oligodendrocyte-related gene were found to show depressive-like behaviors (45). Therefore, *the OLIG2* SNP rs1059004 may be associated with the severity of depressive symptoms and negative self-schema by influencing the function of oligodendrocytes. One functional magnetic resonance imaging-based study indicated that the activities of the anterior cingulate cortex and the inferior frontal cortex are involved in negative self-reflection (48), and prior investigations have found that the *OLIG2* polymorphism rs1059004 affects not only white matter integrity but also resting-state cerebral blood flow in widespread brain regions, including the anterior cingulate and inferior frontal regions (23). Therefore, the effect of SNP rs1059004 on brain perfusion may be a biological mechanism underlying the association between the variant and negative self-schema. Further studies are needed to reveal the biological basis of the association between *the OLIG2* SNP rs1059004 and negative self-schema.

Beck proposed the cognitive model of depression, in which individuals who hold negative self-schemas are vulnerable to developing depression in the future (29). Evans et al. revealed that a negative self-schema was a risk factor for the onset of depression in women in a longitudinal study (32). This finding may support

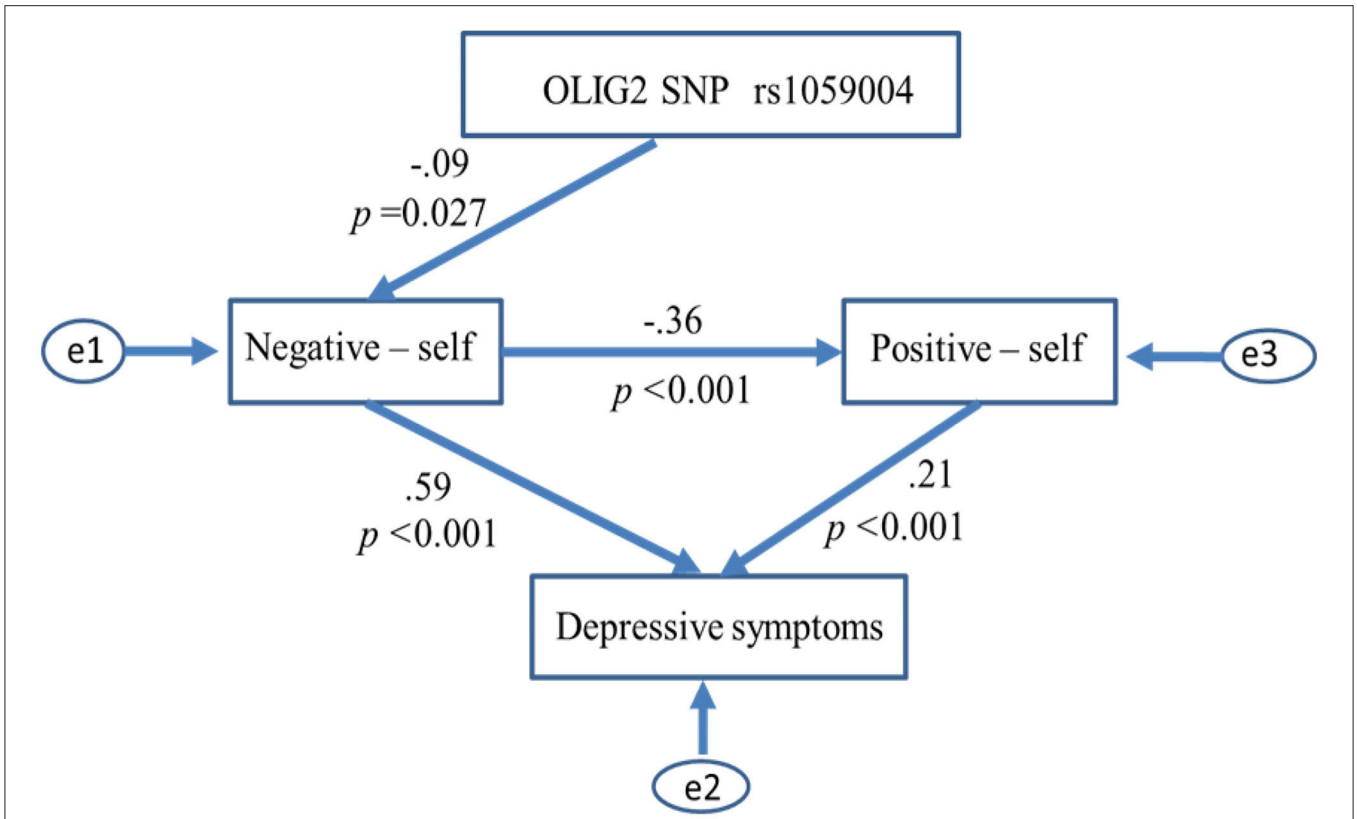


FIGURE 3 | Relationships between *OLIG2* SNP rs1059004, self-core schema and depressive symptoms. Paths scoring $p > 0.05$ were removed from the initial models, and the best-fitting path model was ultimately adopted in the path analysis. The path model produced the best model fit results, as shown below: Chi-square = 2.669, df = 2, P = 0.263; GFI = 0.998; AGFI = 0.988; CFI = 0.998; AIC = 18.669; and RMSEA = 0.025. This path model indicates the predictive effect of *the OLIG2* SNP rs1059004 on depressive symptoms via a negative self-schema (negative self) that influences a positive self-schema (positive self). SNP, single nucleotide polymorphism; GFI, goodness-of-fit index; AGFI, adjusted GFI; CFI, comparative fit index; AIC, Akaike information criterion; RMSEA, root mean square error of approximation.

the cognitive model of depression proposed by Beck. Therefore, negative self-schema could play a major role in constructing trait factors underlying susceptibility to MDD. The results of the present study indicate a significant association between *the OLIG2* SNP rs1059004 and a negative self-schema, constructing trait factors concerning susceptibility to MDD. Considering the significant association between the *OLIG2* SNP rs1059004 and a negative self-schema as the trait factor underlying the development of depression, it would be worthwhile to conduct a longitudinal study investigating whether SNP rs1059004 affects the risk of the onset of MDD by influencing negative self-schema.

The present study has several limitations. First, the participants were limited in terms of age and education history; that is, only young adults and university, college, or postgraduate students were examined; therefore, the results of the present study may not apply to the general population. The second limitation is the cross-sectional nature, which cannot disentangle the direction of effects between variables. The third limitation is that some problems have recently arisen regarding single-gene studies of multifactorial phenotypes, such as depression, and many studies have failed to be replicable. For example, Culverhouse et al. found no proof that the thousands of studies on the relationship between 5-HTTLPR and depression provided evidence of a genuine genetic difference (49). More recently, Border et al. found no evidence for "any candidate gene polymorphism associations with depression phenotypes or any polymorphism-by-environment moderator effects (50)." Although a negative self-schema can relatively reflect trait factors compared with depression, the phenotype is also highly polygenic, and the genetic effect expected from a single gene can be small. However, in the present study, the number of participants was small, the AA genotype was relatively low compared to other genotypes, the power was insufficient, and cross-validation was not performed. Therefore, it will be necessary to replicate these results with a much larger sample size in the future. The fourth limitation is that it is not clear whether the genetic association in the present study is applied to ethnicities other than Japanese. The fifth limitation is that although prior studies indicated that SNP rs1059004 predicted OLIG2 gene expression in Caucasian subjects, it is unclear whether a significant association exists between the SNP rs1059004 and OLIG2 gene expression in Japanese

subjects. This study is the first to indicate a significant genetic association between the *OLIG2* SNP rs1059004 and negative self-schema as a trait factor in susceptibility to MDD in Japanese subjects. The results of the present study suggest an association between *the OLIG2* SNP rs1059004 and negative self-schema by influencing the function of oligodendrocytes. These results may support the involvement of a genetic factor regulating oligodendrocyte function in generating a negative self-schema as a trait factor underlying susceptibility to depression in the pathology of MDD.

DATA AVAILABILITY STATEMENT

The SNP data presented in the study are publicly available. This data can be found in dbSNP (https://www.ncbi.nlm.nih.gov/SNP/snp_viewTable.cgi?handle=TOHOKUPSY).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Ethics Committee of Tohoku University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

HK, HTa, YKi, CO, ZY, and HTo contributed to the acquisition of data or the analysis and interpretation of data. HK and HTa were involved in drafting the manuscript. HTa, YKa, SF, TO, and HTo critically revised the manuscript for important scientific content. HTa, RK, YT, and HTo made substantial contributions to the conception and design of the study. All authors read and approved the final version of the manuscript and agreed on the order in which their names would be listed in the manuscript.

FUNDING

This work was partially supported by a grant-in-aid for scientific research on innovative areas (No. 24116007) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan and the Strategic Research Program for Brain Sciences from the Japan Agency for Medical Research and Development, AMED (JP20dm0107099). This study was also partially supported by JST/RISTEX, JST/CREST, a grant-in-aid for young scientists (B) (KAKENHI 23700306), a grant-in-aid for young scientists (A) (KAKENHI 25700012), and Tohoku University Advanced Research Center for Innovations in Next-Generation Medicine, Japan. None of the funding sources mentioned above were involved in the study design; collection, analysis, or interpretation of data; writing of the report, or decision to submit the article for publication.

ACKNOWLEDGMENTS

We thank Haruka Nouchi for conducting the psychological tests, all other assistants for helping with the experiments in the study, and the study participants and all our other colleagues at Tohoku University for their support.

REFERENCES

1. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA*. (2003) 289:3095–105. doi: 10.1001/jama.289.23.3095
2. Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJ, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLOS Med.* (2013) 10:e1001547. doi: 10.1371/journal.pmed.1001547
3. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry*. (2006) 163:1905–17. doi: 10.1176/ajp.2006.163.11.1905
4. Hayashi Y, Tatebayashi Y. A flow cytometric postmortem brain study for major depressive disorders: implication for oligodendroglial differentiation and functions. *Nihon Shinkei Seishin Yakurigaku Zasshi.* (2012) 32:211–8.
5. Vostrikov VM, Uranova NA, Orlovskaya DD. Deficit of perineuronal oligodendrocytes in the prefrontal cortex in schizophrenia and mood disorders. *Schizophr Res.* (2007) 94:273–80. doi: 10.1016/j.schres.2007.04.014
6. Aston C, Jiang L, Sokolov BP. Transcriptional profiling reveals evidence for signaling and oligodendroglial abnormalities in the temporal cortex from patients with major depressive disorder. *Mol Psychiatry.* (2005) 10:309–22. doi: 10.1038/sj.mp.4001565
7. Rajkowska G, Mahajan G, Maciag D, Sathyanesan M, Iyo AH, Moulana M, et al. Oligodendrocyte morphometry and expression of myelin - related mRNA in ventral prefrontal white matter in major depressive disorder. *J Psychiatr Res.* (2015) 65:53–62. doi: 10.1016/j.jpsychires.2015.04.010
8. Cathomas F, Azzinnari D, Bergamini G, Sigrist H, Buerge M, Hoop V, et al. Oligodendrocyte gene expression is reduced by and influences effects of chronic social stress in mice. *Genes Brain Behav.* (2019) 18:e12475. doi: 10.1111/gbb.12475
9. Lehmann ML, Weigel TK, Elkahloun AG, Herkenham M. Chronic social defeat reduces myelination in the mouse medial prefrontal cortex. *Sci Rep.* (2017) 7:46548. doi: 10.1038/srep46548
10. de Diego-Adeliño J, Pires P, Gómez-Ansón B, Serra-Blasco M, Vives-Gilabert Y, Puigdemont D, et al. Microstructural white-matter abnormalities associated with treatment resistance, severity and duration of illness in major depression. *Psychol Med.* (2014) 44:1171–82. doi: 10.1017/S003329171300158X
11. Jiang J, Zhao YJ, Hu XY, Du MY, Chen ZQ, Wu M, et al. Microstructural brain abnormalities in medication-free patients with major depressive disorder: a systematic review and meta-analysis of diffusion tensor imaging. *J Psychiatry Neurosci.* (2017) 42:150–63. doi: 10.1503/jpn.150341
12. Ma N, Li L, Shu N, Liu J, Gong G, He Z, et al. White matter abnormalities in first-episode, treatment-naive young adults with major depressive disorder. *Am J Psychiatry.* (2007) 164:823–6. doi: 10.1176/ajp.2007.164.5.823
13. Matsuoka K, Yasuno F, Kishimoto T, Yamamoto A, Kiuchi K, Kosaka J, et al. Microstructural differences in the corpus callosum in patients with bipolar disorder and major depressive disorder. *J Clin Psychiatry.* (2017) 78:99–104. doi: 10.4088/JCP.15m09851
14. Osoba A, Hänggi J, Li M, Horn DI, Metzger C, Eckert U, et al. Disease severity is correlated to tract specific changes of fractional anisotropy in MD and CM thalamus–a DTI study in major depressive disorder. *J Affect Disord.* (2013) 149:116–28. doi: 10.1016/j.jad.2012.12.026
15. Yamada S, Takahashi S, Ukai S, Tsuji T, Iwatani J, Tsuda K, et al. Microstructural abnormalities in anterior callosal fibers and their

relationship with cognitive function in major depressive disorder and bipolar disorder: a tract-specific analysis study. *J Affect Disord.* (2015) 174:542–8. doi: 10.1016/j.jad.2014.12.022
16. Chen Z, Zhang H, Jia Z, Zhong J, Huang X, Du M, et al. Magnetization transfer imaging of suicidal patients with major depressive disorder. *Sci Rep.* (2015) 5:9670. doi: 10.1038/srep09670
17. Gunning-Dixon FM, Hoptman MJ, Lim KO, Murphy CF, Klimstra S, Latoussakis V, et al. Macromolecular white matter abnormalities in geriatric depression: a magnetization transfer imaging study. *Am J Geriatr Psychiatry.* (2008) 16:255–62. doi: 10.1097/JGP.0000300628.33669.03
18. Kumar A, Gupta RC, Albert Thomas M, Alger J, Wyckoff N, Hwang S. Biophysical changes in normal-appearing white matter and subcortical nuclei in late-life major depression detected using magnetization transfer. *Psychiatry Res.* (2004) 130:131–40. doi: 10.1016/j.pscychresns.2003.12.002
19. Ligon KL, Fancy SP, Franklin RJ, Rowitch DH. Olig gene function in CNS development and disease. *Glia.* (2006) 54:1–10. doi: 10.1002/glia.20273
20. Meijer DH, Kane MF, Mehta S, Liu H, Harrington E, Taylor CM, et al. Separated at birth? The functional and molecular divergence of OLIG1 and OLIG2. *Nat Rev Neurosci.* (2012) 13:819–31. doi: 10.1038/nrn3386
21. Mitkus SN, Hyde TM, Vakkalanka R, Kolachana B, Weinberger DR, Kleinman JE, et al. Expression of oligodendrocyte-associated genes in dorsolateral prefrontal cortex of patients with schizophrenia. *Schizophr Res.* (2008) 98:129–38. doi: 10.1016/j.schres.2007.09.032
22. Prata DP, Kanaan RA, Barker GJ, Shergill S, Woolley J, Georgieva L, et al. Risk variant of oligodendrocyte lineage transcription factor 2 is associated with reduced white matter integrity. *Hum Brain Mapp.* (2013) 34:2025–31. doi: 10.1002/hbm.22045
23. Komatsu H, Takeuchi H, Kikuchi Y, Ono C, Yu Z, Iizuka K, et al. Ethnicity-dependent effects of schizophrenia risk variants of the OLIG2 gene on OLIG2 transcription and white matter integrity. *Schizophr Bull.* (2020) 46:1619–28. doi: 10.1093/schbul/sbaa049
24. Georgieva L, Moskvina V, Peirce T, Norton N, Bray NJ, Jones L, et al. Convergent evidence that oligodendrocyte lineage transcription factor 2 (OLIG2) and interacting genes influence susceptibility to schizophrenia. *Proc Natl Acad Sci USA.* (2006) 103:12469–74. doi: 10.1073/pnas.0603029103
25. Huang K, Tang W, Tang R, Xu Z, He Z, Li Z, et al. Positive association between OLIG2 and schizophrenia in the Chinese Han population. *Hum Genet.* (2008) 122:659–60. doi: 10.1007/s00439-007-0434-z
26. Sims R, Hollingworth P, Moskvina V, Dowzell K, O'Donovan MC, Powell J, et al. Evidence that variation in the oligodendrocyte lineage transcription factor 2 (OLIG2) gene is associated with psychosis in Alzheimer's disease. *Neurosci Lett.* (2009) 461:54–9. doi: 10.1016/j.neulet.2009.05.051
27. Stewart SE, Platko J, Fagerness J, Birns J, Jenike E, Smoller JW, et al. A genetic family-based association study of OLIG2 in obsessive-compulsive disorder. *Arch Gen Psychiatry.* (2007) 64:209–14. doi: 10.1001/archpsyc.64.2.209
28. Zhang X, Liu J, Guo Y, Jiang W, Yu J. populationAssociation. *Depress Anxiety.* (2015) 32:720–7. doi: 10.1002/da.22394
29. Beck AT. Thinking and depression. ii. theory and therapy. *Arch Gen Psychiatry.* (1964) 10:561–71. doi: 10.1001/archpsyc.1964.01720240015003
30. Fowler D, Freeman D, Smith B, Kuipers E, Bebbington P, Bashforth H, et al. The Brief Core Schema Scales (BCSS): psychometric properties and associations with paranoia and grandiosity in non-clinical and psychosis samples. *Psychol Med.* (2006) 36:749–59. doi: 10.1017/S0033291706007355
31. Uchida T, Kawamura C, Mifune N, Hamaie Y, Matsumoto K, Ambo H, et al. The Japnanese version of the brief core schema scale for schemata concerning the self and others: indentification of Shema patterns and relationschip with depression. *Jpn J Pers.* (2012) 20:143–54. doi: 10.2132/personality.20.143
32. Evans J, Heron J, Lewis G, Araya R, Wolke D, ALSPAC Study Team. Negative self-schemas and the onset of depression in women: longitudinal study. *Br J Psychiatry.* (2005) 186:302–7. doi: 10.1192/bjp.186.4.302
33. Takeuchi H, Taki Y, Nouchi R, Hashizume H, Sekiguchi A, Kotozaki Y, et al. Anatomical correlates of self-handicapping tendency. *Cortex.* (2013) 49:1148–54. doi: 10.1016/j.cortex.2013.01.014
34. Takeuchi H, Taki Y, Nouchi R, Hashizume H, Sekiguchi A, Kotozaki Y, et al. Effects of working memory training on functional connectivity and cerebral blood flow during rest. *Cortex.* (2013) 49:2106–25. doi: 10.1016/j.cortex.2012.09.007
35. Takeuchi H, Taki Y, Nouchi R, Sekiguchi A, Kotozaki Y, Miyauchi CM, et al. Regional gray matter density is associated with achievement motivation: evidence from voxel-based morphometry. *Brain Struct Funct.* (2014) 219:71–83. doi: 10.1007/s00429-012-0485-3
36. Takeuchi H, Taki Y, Sassa Y, Hashizume H, Sekiguchi A, Fukushima A, et al. Brain structures associated with executive functions during everyday events in a non-clinical sample. *Brain Struct Funct.* (2013) 218:1017–32. doi: 10.1007/s00429-012-0444-z
37. Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia.* (1971) 9:97–113. doi: 10.1016/0028-3932(71)90067-4
38. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry.* (1961) 4:561–71. doi: 10.1001/archpsyc.1961.01710120031004
39. Beck AT, Steer RA, Brown GK. *Manual for the Beck Depression Inventory - II.* San Antonio, TX: The Psychological Corporation (1996) doi: 10.1037/t00742-000
40. Kendall PC, Hollon SD, Beck AT, Hammen CL, Ingram RE. Issues and recommendations regarding use of the beck depression inventory. *Cognit Ther Res.* (1987) 11:289–99. doi: 10.1007/BF01186280
41. Fan M, Liu B, Jiang T, Jiang X, Zhao H, Zhang J. Meta-analysis of the association between the monoamine oxidase-A gene and mood disorders. *Psychiatr Genet.* (2010) 20:1–7. doi: 10.1097/YPG.0b013e3283351112
42. Tsunoka T, Kishi T, Ikeda M, Kitajima T, Yamanouchi Y, Kinoshita Y, et al. Association analysis of group II metabotropic glutamate receptor genes (GRM2 and GRM3) with mood disorders and fluvoxamine response in a Japanese population. *Prog Neuropsychopharmacol Biol Psychiatry.* (2009) 33:875–9. doi: 10.1016/j.pnpbp.2009.04.007
43. van Rossum EF, Binder EB, Majer M, Koper JW, Ising M, Modell S, et al. Polymorphisms of the glucocorticoid receptor gene and major depression. *Biol Psychiatry.* (2006) 59:681–8. doi: 10.1016/j.biopsych.2006.02.007
44. Yoon HK, Kim YK. Association between glycogen synthase kinase-3beta gene polymorphisms and major depression and suicidal behavior in a Korean population. *Prog Neuropsychopharmacol Biol Psychiatry.* (2010) 34:331–4. doi: 10.1016/j.pnpbp.2009.12.009
45. Edgar NM, Touma C, Palme R, Sibille E. Resilient emotionality and molecular compensation in mice lacking the oligodendrocyte-specific gene Cnp1. *Transl Psychiatry.* (2011) 1:e42. doi: 10.1038/tp.2011.40
46. Franco-Pons N, Torrente M, Colomina MT, Vilella E. Behavioral deficits in the cuprizone-induced murine model of demyelination/remyelination. *Toxicol Lett.* (2007) 169:205–13. doi: 10.1016/j.toxlet.2007.01.010
47. Xu H, Yang HJ, Zhang Y, Clough R, Browning R, Li XM. Behavioral and neurobiological changes in C57BL/6 mice exposed to cuprizone. *Behav Neurosci.* (2009) 123:418–29. doi: 10.1037/a0014477
48. Ma Y, Li B, Wang C, Shi Z, Sun Y, Sheng F, et al. 5-HTTLPR polymorphism modulates neural mechanisms of negative self-reflection. *Cereb Cortex.* (2014) 24:2421–9. doi: 10.1093/cercor/bht099
49. Culverhouse RC, Saccone NL, Horton AC, Ma Y, Anstey KJ, Banaschewski T, et al. Collaborative meta-analysis finds no evidence of a strong interaction between stress and 5-HTTLPR genotype contributing to the development of depression. *Mol Psychiatry.* (2018) 23:133–42. doi: 10.1038/mp.2017.44
50. Border R, Johnson EC, Evans LM, Smolen A, Berley N, Sullivan PF, et al. No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. *Am J Psychiatry.* (2019) 176:376–87. doi: 10.1176/appi.ajp.2018.18070881

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychiatry

ORIGINAL RESEARCH
published: 17 March 2021
doi: 10.3389/fpsyt.2021.553305

Weighted Gene Coexpression Network Analysis Reveals Essential Genes and Pathways in Bipolar Disorder

Zhen-Qing Zhang[1,2†], Wei-Wei Wu[1†], Jin-Dong Chen[1], Guang-Yin Zhang[3], Jing-Yu Lin[2], Yan-Kun Wu[2], Yu Zhang[4], Yun-Ai Su[2], Ji-Tao Li[2] and Tian-Mei Si[2*]*

[1] Xiamen Xianyue Hospital, Xiamen, China, [2] Peking University Sixth Hospital, Peking University Institute of Mental Health, Peking University, Beijing, China, [3] Department of Psychosomatic Medicine, First Teaching Hospital of Tianjin University of Traditional Chinese Medicine, Tianjin, China, [4] Institute of Mental Health, Hebei North University, Hebei, China

OPEN ACCESS

Edited by:
Elena Martín-García,
Pompeu Fabra University, Spain

Reviewed by:
Gabriel R. Fries,
University of Texas Health Science Center at Houston, United States
Liang Chen,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

***Correspondence:**
Tian-Mei Si
si.tian-mei@163.com
Ji-Tao Li
ljt_102124@163.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Behavioral and Psychiatric Genetics, a section of the journal Frontiers in Psychiatry

Received: 20 April 2020
Accepted: 24 February 2021
Published: 17 March 2021

Citation:
Zhang Z-Q, Wu W-W, Chen J-D, Zhang G-Y, Lin J-Y, Wu Y-K, Zhang Y, Su Y-A, Li J-T and Si T-M (2021) Weighted Gene Coexpression Network Analysis Reveals Essential Genes and Pathways in Bipolar Disorder.
Front. Psychiatry 12:553305.
doi: 10.3389/fpsyt.2021.553305

Bipolar disorder (BD) is a major and highly heritable mental illness with severe psychosocial impairment, but its etiology and pathogenesis remains unclear. This study aimed to identify the essential pathways and genes involved in BD using weighted gene coexpression network analysis (WGCNA), a bioinformatic method studying the relationships between genes and phenotypes. Using two available BD gene expression datasets (GSE5388, GSE5389), we constructed a gene coexpression network and identified modules related to BD. The analyses of Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathways were performed to explore functional enrichment of the candidate modules. A protein-protein interaction (PPI) network was further constructed to identify the potential hub genes. Ten coexpression modules were identified from the top 5,000 genes in 77 samples and three modules were significantly associated with BD, which were involved in several biological processes (e.g., the actin filament-based process) and pathways (e.g., MAPK signaling). Four genes (*NOTCH1*, *POMC*, *NGF*, and *DRD2*) were identified as candidate hub genes by PPI analysis and CytoHubba. Finally, we carried out validation analyses in a separate dataset, GSE12649, and verified *NOTCH1* as a hub gene and the involvement of several biological processes such as actin filament-based process and axon development. Taken together, our findings revealed several candidate pathways and genes (*NOTCH1*) in the pathogenesis of BD and call for further investigation for their potential research values in BD diagnosis and treatment.

Keywords: bipolar disorders, coexpression modules, hub genes, WGCNA, pathway analysis

INTRODUCTION

Bipolar disorder (BD) is a chronic and recurrent severe mental disorder that affect about 1% global population (1). The disease is associated with high heritability, ranging from 70 to 90% (2), but its key genetic and neurobiological mechanisms are still not recognized. With the development of high-throughput sequencing technology, significant progress has been made in the genomics of BD. Recently, some systematic reviews and genome-wide association study (GWAS) findings have revealed more than 40 genes, including *ANK3*, *ERBB2*, *ODZ4*, *CACNA1C*, and *FADS* (3, 4). Pathway analyses of these genes showed that they were involved the regulation of insulin secretion,

apoptosis, immunological response, neuroplasticity, HPA axis dysregulation, and the signal of endocannabinoid, etc., (5, 6). Recent RNA-sequencing studies also highlight dysregulation of neuroplasticity, circadian rhythms, as well as GTPase binding in BD (7). These findings provide clear evidence that BD is associated with an extensive polygenic genetic architecture (5), which calls for the network-level investigation to reveal the correspondence between risk genes and phenotypes. By combining public microarray data with bioinformatic analysis, we can gain an in-depth understanding of the molecular processes and pathogenesis of BD.

Weighted gene coexpression network analysis (WGCNA) is a bioinformatic method to study the relationships between genes and phenotypes. Recently, this technique has been widely applied to neurological and psychiatric disorders, including posttraumatic stress disorder, depression, schizophrenia, Alzheimer's disease, and Huntington's disease (8–13). Different from previous research methods focusing on individual genes, WGCNA transforms gene expression profiles into coexpression networks (modules). By examining modules of highly correlated genes, this method provides insights into the signaling networks that may be responsible for phenotypic traits of interest and the results may help identify the candidate biomarkers or therapeutic targets of many biological processes (14, 15).

The aim of this study is to reveal gene modules related to BD and to identify their functional pathways and potential hub genes. Using two available BD gene expression datasets (GSE5388, GSE5389), we constructed a gene coexpression network using WGCNA. For the modules related to BD, we then performed the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses to reveal the pathways enriched in these modules. In addition, a protein-protein interaction (PPI) network was constructed for the module of interest to identify candidate hub genes. Finally, we carried out validation analyses in a separate dataset GSE12649. These analyses based on gene coexpression profiles would shed light on the polygenic genetic architecture of BD and help to reveal the potential gene markers for BD diagnosis and treatment.

MATERIALS AND METHODS

Microarray Data

Two gene expression datasets of BD patients, GSE5388, and GSE5389, were downloaded from the GEO database in May 2019 (https://www.ncbi.nlm.nih.gov/geo/). The GSE5388 dataset contained 61 samples (30 bipolar and 31 control subjects) from the dorsolateral prefrontal cortex (DLPFC). The GSE5389 dataset contained 21 samples (10 bipolar and 11 control subjects) from the orbitofrontal cortex (OFC). Both datasets were from the study of Ryan et al. (16) and based on the GPL96 platform ([HG-U133A] Affymetrix Human Genome U133A Array). The clinical data of these datasets, such as diagnosis, demographics (e.g., age, sex), and technical information (e.g., post-mortem interval, pH, RNA degradation, Batch), were also obtained. These two datasets were chosen based on prior evidence of involvement of DLPFC and OFC in BD (17). Ryan and colleagues found that gene expression changes in two regions were comparable (16), which suggests that the two datasets could be merged. Below we performed quality control analysis to further make sure that there are no batch effects between the two datasets.

Data Preprocessing and Quality Control

The preprocessing was carried out separately for each dataset, including quality control and normalization, using the *affy* package in R (version 3.6.0; https://www.r-project.org/) (18). Demographic and technical variables were treated as covariates to control for their potential influence on the differences between bipolar and control subjects. The samples with standardized sample network connection Z-scores <-2 were excluded from further analysis. Five samples (GSM123204, GSM123205, GSM123206, GSM123214, and GSM123243) were defined as outliers and removed, resulting in 77 samples for final analysis. The chip scan date extracted from the metadata was used as an experimental batch for each dataset. The ComBat function of *sva* package in R was used to correct for batch effects (19). Annotations to the probes were performed using Ensembl gene IDs (v75; Feb 2014 data freeze) by the *biomaRt* package in R (20). A larger dataset was built by merging the two datasets following previous practics (12). The ComBat function was used to eliminate study batch effects if present (19).

Weighted Gene Coexpression Network Analysis (WGCNA)

The WGCNA was performed on the average expression levels of the top 5,000 genes in 77 samples using the *WGCNA* package in R (14). The sample network connectivity was standardized by function scale according to the distance before WGCNA, which excluded the outlier samples connectivity <-5. The hierarchical clustering of samples was analyzed using the default method (hclust function), and no sample outliers were found. The soft thresholding power was then screened by the pickSoftThreshold function. Candidate powers from 1 to 20 were applied to test the mean connectivity degrees and the independence of modules. The soft thresholding power was selected if the $R^2 \geqq 0.8$. To construct the WGCNA, the blockwiseModules function in R was used, and multiple parameters were defined. Here, the following parameters were used: power = 8, minModuleSize = 50, deepSplit = 1, networkType = "unsigned." The module detection process was performed automatically by BlockwiseModules. Specifically, it can build a correlation network, create a cluster tree, and then merge nearby branches to form modules. A hierarchical clustering tree (dendrogram) was plotted to display hierarchical clustering. DissTOM (1-Topological Overlap Matrix) was calculated, and the relationships among all genes were visualized in R. Finally, the heatmap function in R was used to analyze the correlations between the modules.

Relationship Analysis Between Coexpression Modules and BD

The module eigengene (ME) represents the first principal component in each module and therefore reflects the level of gene expression in the module (14). Pearson's correlation test was applied to assess the correlation between ME and BD and the heatmap package in R was used to visualize the correlations

between modules and BD. Modules with significantly negative or positive correlations between ME and clinical traits were considered as candidate modules.

GO and KEGG Analyses of Coexpression Modules for Bipolar Disorder

The Gene Ontology (GO) analysis, which includes the biological process (BP), cellular component (CC), and molecular function (MF) ontologies, is a standard method for gene functional annotation (21, 22). The Kyoto Encyclopedia of Genes and Genomes (KEGG) database, which stores gene metabolic pathways, is widely used to determine functional enrichment (23). Here, GO and KEGG analyses were performed for the genes in candidate modules in Metascape (http://metascape.org/gp/index.html#/main/step1), which is a reliable and widely used online software for omics-based research (24). The candidate modules were calculated with a *P*-value cutoff of 0.01, a min overlap of 3, and a min enrichment of 1.5.

Protein-Protein Interaction Network Analysis for Selected Modules of Bipolar Disorder

The genes in candidate modules were mapped into the STRING database (Version 11.0, https://string-db.org/) for PPI network analysis. The Cytoscape software (v3.7.1) (25) was used for visualization, and the CytoHubba plug-in (http://hub.iis.sinica.edu.tw/cytohubba/) was applied to find hub genes in each module. To reduce potential errors caused by a complex biological network, it is necessary to use multiple methods to identify essential proteins (26). Hub genes were analyzed by CytoHubba using the following five methods: maximum neighborhood component, node connect degree, closeness, edge percolated component, and radiality (27–29).

Validation Using Another Gene Expression Dataset of BD

The microarray dataset GSE12649, which was deposited by Iwamoto et al. (30), was used to validate our findings. This dataset contained 102 subjects from the prefrontal cortex (BA46), including 33 BD, 35 schizophrenia, and 34 control subjects. The BD and control samples were used for validation analysis. We carried out the same analyses as our main study, including data preprocessing, WGCNA, GO and KEGG analyses, PPI visualization, and hub genes screening. We validated our previous GO enrichment and KEGG pathway results in this dataset, focusing on top five GO enrichment results in each module and the important KEGG pathways. Finally, the hub genes of main and validation analyses were compared with each other to identify shared hub genes across datasets.

RESULTS

Data Preprocessing and Quality Control

As shown in the boxplots and histograms, each microarray dataset indicated valid normalization and quality control for further analyses (**Supplementary Figure 1**). Bipolar and control status were not significantly associated with sex, age, pH, post-mortem interval (PMI), RNA degradation (RNAdeg), and Batch ($P > 0.05$). Five samples (GSM123204, GSM123205, GSM123206, GSM123214, and GSM123243) were defined as outliers and removed, resulting in 77 samples for final analysis. There were a total of 12,300 genes shared between the two datasets and no batch effects were observed ($P > 0.05$; see **Supplementary Figure 2**). The gene coexpression network was constructed by 5,000 genes with the highest average expression values (31).

WGCNA to Identify Modules Critical to BD

The clustering results in **Supplementary Figure 3** showed that all 77 samples were clustered well. With a soft-threshold power of 8 (**Figure 1A**), nine coexpression modules were identified, ranging in size from 192 to 1,419 genes (**Figures 1B,C**). Specifically, there were 1,419 genes in module 1 (MEturquoise), 944 genes in module 2 (MEblue), 749 genes in module 3 (MEbrown), 352 genes in module 4 (MEyellow), 318 genes in module 5 (MEgreen), 272 genes in module 6 (MEred), 211 genes in module 7 (MEblack), 196 genes in module 8 (MEpink), and 192 genes in module 9 (MEmagenta). In addition, 374 genes not assigned to any of the above modules were classified as a gene set Module 0 (MEgrey). The interactions among the ten modules are shown in **Figure 2** and suggest that the modules were relatively independent.

The correlations between the coexpression modules and clinical traits are shown in **Figure 3**. Three modules were significantly associated with BD status (Bonferroni-corrected $P < 0.05$). The MEblue module ($r = 0.39$, $P = 4e\text{-}04$, corrected $P = 4e\text{-}03$) showed a positive correlation, whereas the MEgreen ($r = -0.39$, $P = 5e\text{-}04$, corrected $P = 5e\text{-}03$), and MEturquoise ($r = -0.38$, $P = 6e\text{-}04$, corrected $P = 6e\text{-}03$) modules showed negative correlations. None of the 10 modules was significantly associated with sex or age (*Ps > 0.05*).

Gene Function and Annotation Enrichment (Functional Enrichment) Analysis of Key Modules

As shown in **Figure 4**, the GO and KEGG pathway analyses were performed for the three key modules. The top five GO enrichment results showed that genes in MEblue were mainly enriched in GO:0003012 (muscle system process), GO:1901137 (carbohydrate derivative biosynthetic process), GO:0030029 (actin filament-based process), GO:0061564 (axon development), and GO:0010817 (regulation of hormone levels). Genes in MEgreen were enriched in GO:0071417 (cellular response to organonitrogen compound), GO:0017038 (protein import), GO:0042391 (regulation of membrane potential), GO:0005773 (vacuole), GO:0018105 (peptidyl-serine phosphorylation). Genes in MEturquoise were mainly enriched in GO:0016604 (nuclear body), GO:0006397 (mRNA processing), GO:0006403 (RNA localization), GO:0006753

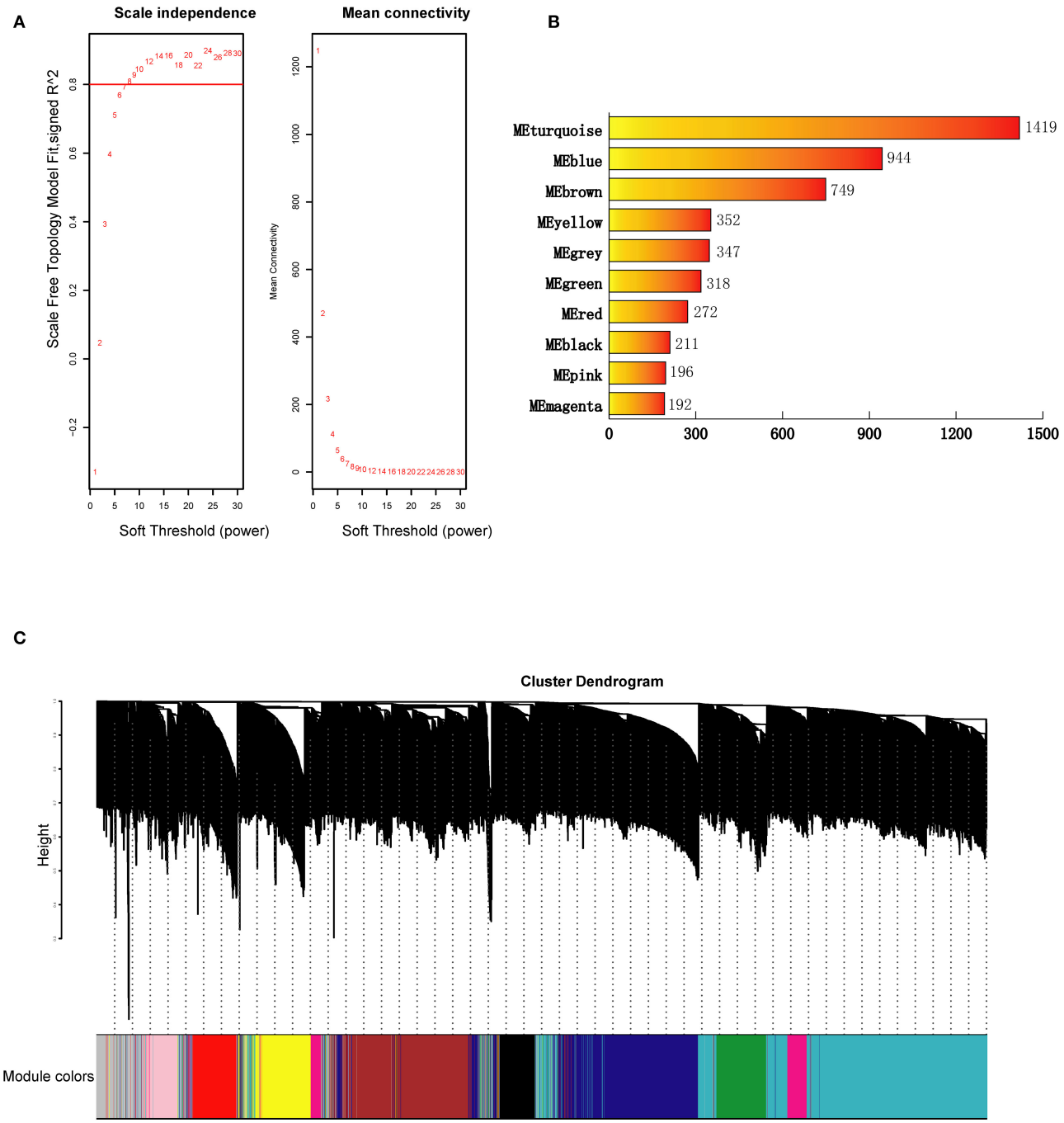


FIGURE 1 | Construction of genes coexpression modules for BD. **(A)** Analysis of network topology for a set of soft-thresholding powers. **(B)** Number of genes in each coexpression module. **(C)** Construction of genes coexpression modules. Each color represents a module and each branch represents a gene. ME, module.

(nucleoside phosphate metabolic process), and GO:0005635 (nuclear envelope). For the top 20 clusters enriched in each module, please refer to **Supplementary Table 1**.

The KEGG pathway analysis showed that genes in MEblue were mainly enriched in hsa04010 (MAPK signaling pathway), hsa01522 (Endocrine resistance), and hsa05414 (Dilated cardiomyopathy), and genes in MEgreen were mainly enriched in hsa03450 (Non-homologous end-joining). No significantly enriched pathways were found for MEturquoise.

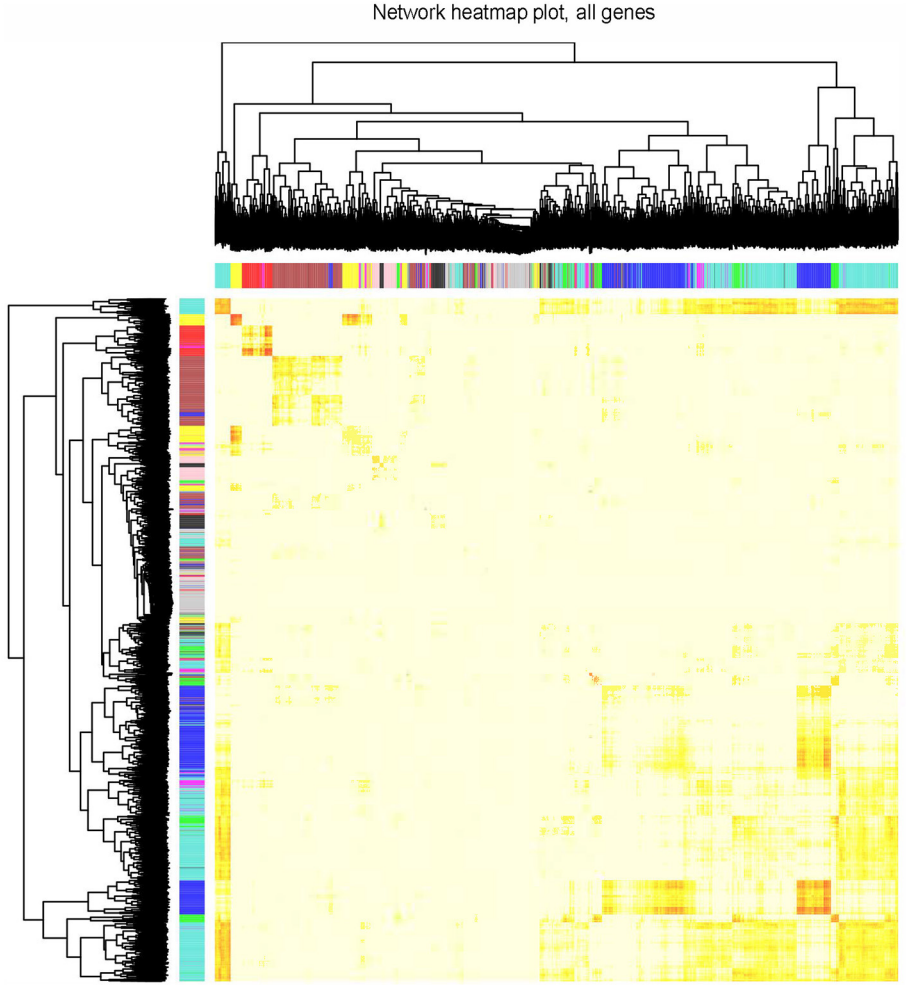


FIGURE 2 | Visualization of TOM of co-expressed genes in different modules by a heat map. Light colors indicate low overlap and dark red indicates high overlap. The darker color blocks along the diagonal are coexpression modules.

Visualization of the PPI Network and Hub Genes

Among the 3 modules, MEblue ($r = 0.39$, $P = 4e\text{-}04$) and MEgreen ($r = -0.39$, $P = 5e\text{-}04$) had the strongest correlations with BD, and the number of genes in MEblue was greater than that in MEgreen. Therefore, we selected the MEblue module for the subsequent analysis (32). The PPI network analysis was performed on the MEblue module using the STRING database. As a result, 853 nodes and 2,687 edges were established in the module and the PPI enrichment *P*-value was 1.0e-12. The interactions of proteins in MEblue were selected and converted into a network, which was visualized with Cytoscape, as shown in **Figure 5**. Each term is represented by a circular node, and its cluster identity is represented by its color. The top 10 hub genes in the PPI network were calculated by CytoHubba (28). As a result, four genes (*NOTCH1*, *POMC*, *NGF*, and *DRD2*), which overlapped among the five methods, were deemed hub genes. Then we analyzed the gene expression levels of these hub genes, and found that, compared with controls, the expression level of *NOTCH1* ($t_{75} = 2.159$, $P = 0.034$), *NGF* ($t_{75} = 3.183$, $P = 0.002$), and *POMC* ($t_{75} = 3.791$, $P = 3e\text{-}04$) presented significant up-regulation in BD (**Supplementary Figure 4**).

Validation of Gene Function, Annotation Enrichment, and Hub Genes

In an independent dataset, GSE12649, we carried out the same analyses as above to validate our results. WGCNA revealed seven coexpression modules in this dataset. Among these modules, MEbrown (including 460 genes) showed a significant positive correlation ($r = 0.27$, $P = 0.04$) with BD (**Supplementary Table 2**). In this module, we then validated the top five GO functional enrichment results of each module and four KEGG pathways we observed previously (**Supplementary Table 3**). The following GO enrichment results appeared in the clusters of MEbrown: GO:0071417 (cellular response to organonitrogen compound), GO:0030029 (actin

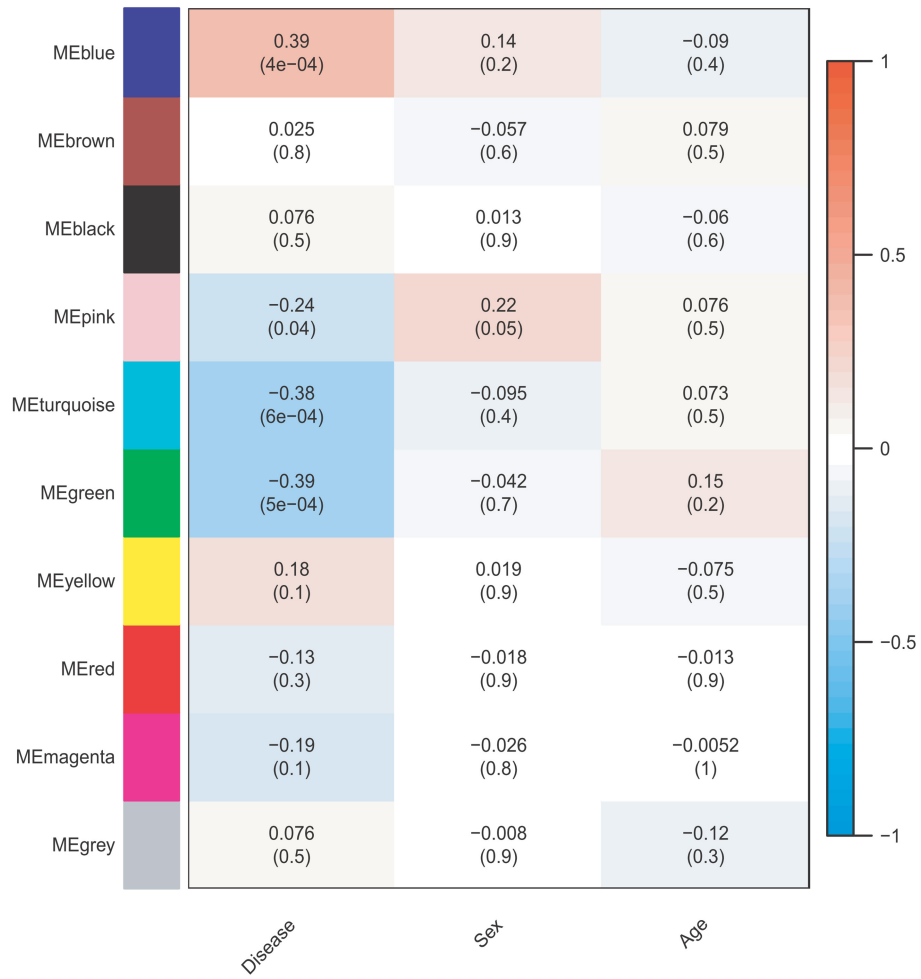


FIGURE 3 | Module-trait relationship. Each row corresponds to a module eigengene, and each column corresponds to one feature. Each cell contains the corresponding correlation and *p*-value.

filament-based process), GO:0003012 (muscle system process), GO:0061564 (axon development), GO:0010817 (regulation of hormone levels), Go:0042391 (regulation of membrane potential). Among the four important KEGG pathways, only the hsa05414 (Dilated cardiomyopathy) was verified in the MEbrown. Finally, among the four hub genes of the main results, *NOTCH1* was verified as a hub gene in this validation dataset (**Supplementary Table 4**).

DISCUSSION

In this study, we applied WGCNA to the BD gene expression profiles GSE5388 and GSE538 and identified 10 modules on the top 5,000 genes from 77 samples. Three out of 10 modules were significantly associated with BD. Functional enrichment and the PPI network of these BD-related modules were explored with Metascape and the STRING database. We then identified four genes, *NOTCH1*, *POMC*, *NGF*, and *DRD2*, as hub genes underlying BD. In a further validation analysis using a separate dataset (GSE12649), we validated several biological processes and pathways (such as actin filament-based process, axon development, hormone level regulation) and *NOTCH1* as a hub gene of BD.

To understand the genetic mechanisms of BD, various studies have been carried out, adopting various genetic techniques such as microarray, single-cell sequencing, RNA-sequencing, and GWAS. The microarray datasets we analyzed were contributed by Ryan and colleagues; their results suggest that BD is associated with the dysregulation of the ubiquitin pathway and synaptic genes in orbitofrontal cortex (16). As we mentioned in the Introduction, given the complex and multifactorial

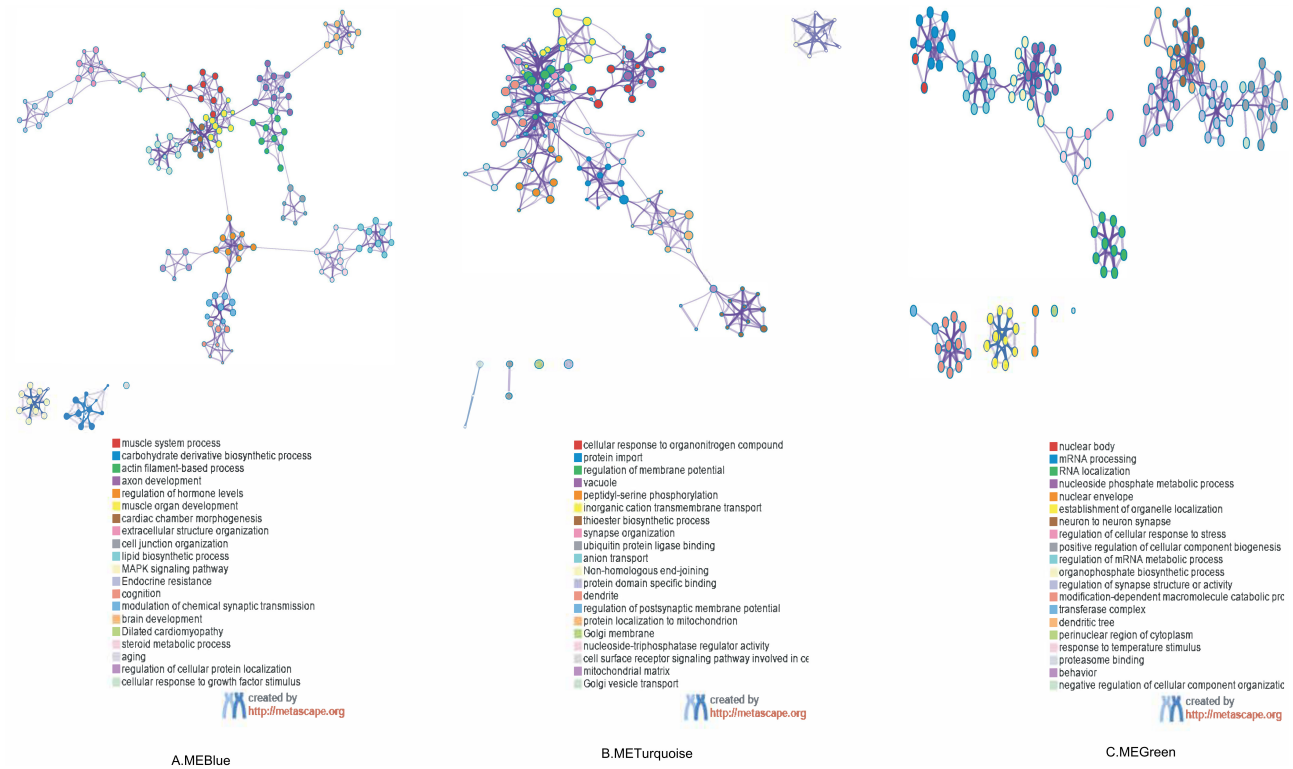


FIGURE 4 | Functional enrichment analysis of three co-expressing network modules: enriched terms are represented by circle nodes, and nodes of the same color belong to the same cluster. **(A)** Colored by cluster ID in MEBlue, where nodes that share the same cluster ID are typically close to each other. **(B)** Colored by cluster ID in METurquoise. **(C)** Colored by cluster ID in MEGreen. ME, module.

nature of BD genetic mechanisms, here we investigated the potential alterations of the interconnection between genes in BD using WGCNA (33). We identified three critical modules in the main analyses, which are associated with various biological processes and pathways, such as muscle system processes, biosynthesis of carbohydrate derivatives, actin filament formation, axon development, hormone level regulation, MAPK signaling pathway, and endocrine resistance. Some of these findings were confirmed in the validation analysis in a separate dataset, such as actin filament-based process, axon development, hormone level regulation. Previous studies have revealed alterations of gene sets mediating synaptic transmission and nervous system development in BD based on microarray technology (16), pathway-based analyses (34) or prioritized gene framework (35) of the genome-wide association datasets. Other studies have highlighted many other pathways, showing convergence on neuroplasticity (5–7). Our findings of actin filament-based process and axon development are in line with consistent observations of neuroplasticity alterations in BD and provide evidence for potential frontal structural plasticity abnormalities in BD.

The most significant KEGG pathways associated with BD were the MAPK signaling pathway and endocrine resistance. A GWAS study has shown that one of the KEGG pathways involved in BD was the MAPK signaling pathway (36), which is involved in mediating entrainment of the circadian system (37). Many circadian genes have been associated with BD (38). The MAPK change in the intracellular signal cascade may be caused by the immune imbalance in BD (39). Endocrine resistance refers to resistance to endocrine therapy agents, such as selective estrogen receptor modulators (e.g., tamoxifen). Some studies have found that tamoxifen was effective against manic episodes (40) and our study lends further support to the potential role of endocrine resistance in BD pathogenesis. Nevertheless, these two pathways were not confirmed in the validation dataset. Instead, both primary and validation datasets revealed that BD-related modules are mainly enriched in has05414 (Dilated cardiomyopathy). Future studies are warranted to examine whether this pathway is involved in the pathophysiology of BD or a by-product of long-term mood stabilizer treatment.

The Notch signaling pathway plays critical roles in neural development and brain homeostasis and is involved in neuronal migration, early differentiation, memory formation, and synaptic plasticity (41). The Notch receptor contains four members (Notch1, Notch2, Notch3, and Notch4), and their expression patterns in the forebrain are as follows: *NOTCH1* in neurons, astrocytes, precursors, ependymal cells, and endothelium; *NOTCH2* in neurons and precursors; *NOTCH3* in precursors; *NOTCH4* in the endothelium (42). The Notch pathway is closely associated with the pathological mechanism of BD (43, 44). For example, the *NOTCH4* gene expression in peripheral blood cells has been found to be upregulated in BD (45). The *NOTCH3*

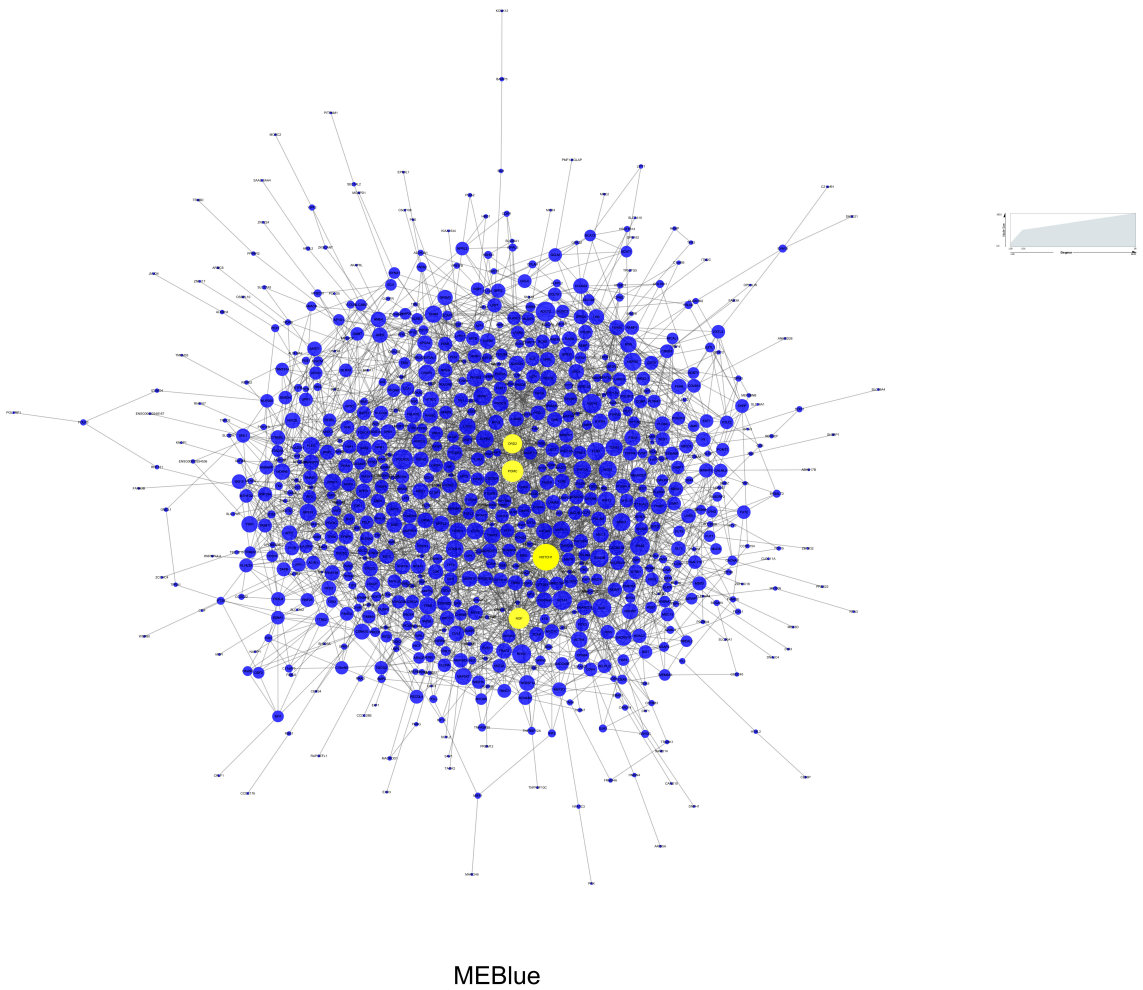


FIGURE 5 | Protein-protein interaction network of MEBlue for BD. Yellow node represents the most important protein of the network and the related gene is defined as hub gene. Node size represents the degree of connectivity.

mutation has also been reported in BD, which however was not consistent across studies (46, 47). A recent WGCNA study with relatively small samples (17 BD vs. 19 controls) from the prefrontal cortex reported *NOTCH2* as *one* of the 30 hub genes in BD (48). Our results highlight *NOTCH1* as a hub gene for BD in both primary and validation datasets. While these results together suggest the involvement of Notch pathway in BD, the inconsistent findings across studies may arise from differences in tissue types, analysis methods, and sample sizes. Intriguingly, the *NOTCH1* signaling could be activated by valproic acid (49), a commonly used mood stabilizer, which suggests that NOTCH1 may serve as a potential treatment target for BD.

Our main results also revealed *POMC*, *NGF*, and *DRD2* as hub genes for BD. Proopiomelanocortin (POMC)-derived peptides are involved in the regulation of energy homeostasis, learning, memory, inflammation, and immune modulation (50, 51). The significantly increased levels of the critical pituitary hormones POMC indicated dysfunction of the HPA axis of BD (52, 53). *POMC* is also one of the shared genes between mood disorder and cardiometabolic diseases (54). In the central nervous system, nerve growth factor (NGF) play key roles in neuroprotection and neural repair (55). Based on published GWASs and candidate gene studies, the *NGF* gene might be a useful biological marker for the manic state and early detection of conversion from significant depression to BD (56–58). Finally, the *DRD2* gene encodes the D2 subtype of the dopamine receptor. Several studies have shown that the *DRD2* gene is associated with BD and that polymorphism in *DRD2* may play a role in BD development (59–61). However, these candidate genes were not observed in the validation dataset and their involvement in BD warrants further scrutiny.

LIMITATIONS

Our study has some limitations. First, the small sample size for WGCNA may affect the robustness of the observed results. Sample sizes in our main and validation datasets met the

minimum requirements for WGCNA (i.e., larger than 15 samples per group) (62), but may be not large enough to detect modules with smaller effect sizes. Future studies with larger sample sizes are needed to validate our findings. Second, compared to RNA-sequencing, the microarray data is limited by the probes pre-defined by the manufacturer and is not sensitive to low-abundance genes. As the RNA-sequencing data storage and analysis become increasingly available to researchers, future studies have the opportunities to reveal other potential BD-related genetic alterations. Third, the brain tissue we analyzed contained multiple cell types, which may have different gene expression profiles. For instance, the *NOTCH1* seems to have higher expression levels in glia cells than in neurons [**Supplementary Figure 5**, based on the single-cell RNA sequence of 466 cells in the healthy human brain in a public scRNA-seq database scRNASeqDB (63, 64)]. Our study may fail to detect BD-related changes in specific cell types; it is also unclear whether our results were driven by one or some cell types. Future investigation is warranted to reveal the cell-type-specific alterations associated with BD. Finally, this was a preliminary study by using public data, and the results need to be further validated with molecular biology experiments. Although some essential genes and pathways were verified in another dataset (GSE12649), the reliability of our study was still insufficient and the findings should be interpreted with caution.

CONCLUSIONS

In summary, we performed WGCNA in independent gene expression datasets and highlighted *NOTCH1* as one candidate gene of BD and the involvement of several biological processes such as actin filament-based process and axon development, which might be targets for BD diagnosis and treatment. These results provide new perspectives for understanding BD pathogenesis and invite further investigations and validations on the candidate genes and biological processes/pathways we observed.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/geo/.

AUTHOR CONTRIBUTIONS

Z-QZ, W-WW, J-TL, and T-MS designed the study and wrote the manuscript. Z-QZ, J-DC, G-YZ, and J-YL prepared and analyzed the data. Y-KW, YZ, and Y-AS prepared and reviewed the study. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Capital Medical Development Research Fund (2020-2-4113), the National Natural Science Foundation of China (Grant Nos. 81630031, 81771468, 82071528, and 81601184), and the Beijing Brain Project (grant No. Z171100000117016).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2021.553305/full#supplementary-material

REFERENCES

1. Ferrari AJ, Stockings E, Khoo JP, Erskine HE, Degenhardt L, Vos T, et al. The prevalence and burden of bipolar disorder: findings from the Global Burden of Disease Study 2013. *Bipolar Disord*. (2016) 18:440–50. doi: 10.1111/bdi.12423
2. Gordovez FJA, McMahon FJ. The genetics of bipolar disorder. *Mol Psychiatry*. (2020) 25:544–59. doi: 10.1038/s41380-019-0634-7
3. Ikeda M, Saito T, Kondo K, Iwata N. Genome-wide association studies of bipolar disorder: a systematic review of recent findings and their clinical implications. *Psychiatry Clin Neurosci*. (2018) 72:52–63. doi: 10.1111/pcn.12611
4. Hou L, Bergen SE, Akula N, Song J, Hultman CM, Landen M, et al. Genome-wide association study of 40,000 individuals identifies two novel loci associated with bipolar disorder. *Hum Mol Genet*. (2016) 25:3383–94. doi: 10.1093/hmg/ddw181
5. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet*. (2019) 51:793–803. doi: 10.1038/s41588-019-0397-8
6. Sigitova E, Fisar Z, Hroudova J, Cikankova T, Raboch J. Biological hypotheses and biomarkers of bipolar disorder. *Psychiatry Clin Neurosci*. (2017) 71:77–103. doi: 10.1111/pcn.12476
7. Akula N, Barb J, Jiang X, Wendland JR, Choi KH, Sen SK, et al. RNA-sequencing of the brain transcriptome implicates dysregulation of neuroplasticity, circadian rhythms and GTPase binding in bipolar disorder. *Mol Psychiatry*. (2014) 19:1179–85. doi: 10.1038/mp.2013.170
8. Katrinli S, Lori A, Kilaru V, Carter S, Powers A, Gillespie CF, et al. Association of HLA locus alleles with posttraumatic stress disorder. *Brain Behav Immun*. (2019) 81:655–8. doi: 10.1016/j.bbi.2019.07.016
9. Geng R, Li Z, Yu S, Yuan C, Hong W, Wang Z, et al. Weighted gene co-expression network analysis identifies specific modules and hub genes related to subsyndromal symptomatic depression. *World J Biol Psychiatry*. (2018) 21:102–10. doi: 10.1080/15622975.2018.1548782
10. Rangaraju S, Dammer EB, Raza SA, Rathakrishnan P, Xiao H, Gao T, et al. Identification and therapeutic modulation of a pro-inflammatory subset of disease-associated-microglia in Alzheimer's disease. *Mol Neurodegener*. (2018) 13:24. doi: 10.1186/s13024-018-0254-8
11. Mina E, van Roon-Mom W, Hettne K, van Zwet E, Goeman J, Neri C, et al. Common disease signatures from gene expression analysis in Huntington's disease human blood and brain. *Orphanet J Rare Dis*. (2016) 11:97. doi: 10.1186/s13023-016-0475-2
12. Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C, et al. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science*. (2018) 359:693–7. doi: 10.1126/science.aad6469
13. de Jong S, Boks MP, Fuller TF, Strengman E, Janson E, de Kovel CG, et al. A gene co-expression network in whole blood of schizophrenia patients is independent of antipsychotic-use and enriched for brain-expressed genes. *PLoS ONE*. (2012) 7:e39498. doi: 10.1371/journal.pone.0039498
14. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. (2008) 9:559. doi: 10.1186/1471-2105-9-559

15. Wan Q, Tang J, Han Y, Wang D. Co-expression modules construction by WGCNA and identify potential prognostic markers of uveal melanoma. *Exp Eye Res*. (2018) 166:13–20. doi: 10.1016/j.exer.2017.10.007
16. Ryan MM, Lockstone HE, Huffaker SJ, Wayland MT, Webster MJ, Bahn S. Gene expression analysis of bipolar disorder reveals downregulation of the ubiquitin cycle and alterations in synaptic genes. *Mol Psychiatry*. (2006) 11:965–78. doi: 10.1038/sj.mp.4001875
17. Price JL, Drevets WC. Neurocircuitry of mood disorders. *Neuropsychopharmacology*. (2010) 35:192–216. doi: 10.1038/npp.2009.104
18. Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy–analysis of affymetrix genechip data at the probe level. *Bioinformatics*. (2004) 20:307–15. doi: 10.1093/bioinformatics/btg405
19. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. (2012) 28:882–3. doi: 10.1093/bioinformatics/bts034
20. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc*. (2009) 4:1184–91. doi: 10.1038/nprot.2009.97
21. The Gene Ontology C. Expansion of the gene ontology knowledgebase and resources. *Nucleic Acids Res*. (2017) 45:D331–8. doi: 10.1093/nar/gkw1108
22. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. (2000) 25:25–9. doi: 10.1038/75556
23. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. (2016) 44:D457–62. doi: 10.1093/nar/gkv1070
24. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. (2019) 10:1523. doi: 10.1038/s41467-019-09234-6
25. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. (2003) 13:2498–504. doi: 10.1101/gr.1239303
26. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol*. (2014) 8(Suppl. 4):S11. doi: 10.1186/1752-0509-8-S4-S11
27. Sang L, Wang XM, Xu DY, Zhao WJ. Bioinformatics analysis of aberrantly methylated-differentially expressed genes and pathways in hepatocellular carcinoma. *World J Gastroenterol*. (2018) 24:2605–16. doi: 10.3748/wjg.v24.i24.2605
28. Li BQ, Zhang J, Huang T, Zhang L, Cai YD. Identification of retinoblastoma related genes with shortest path in a protein-protein interaction network. *Biochimie*. (2012) 94:1910–7. doi: 10.1016/j.biochi.2012.05.005
29. Zamani-Ahmadmahmudi M, Najafi A, Nassiri SM. Reconstruction of canine diffuse large B-cell lymphoma gene regulatory network: detection of functional modules and hub genes. *J Comp Pathol*. (2015) 152:119–30. doi: 10.1016/j.jcpa.2014.11.008
30. Iwamoto K, Bundo M, Kato T. Altered expression of mitochondria-related genes in postmortem brains of patients with bipolar disorder or schizophrenia, as revealed by large-scale DNA microarray analysis. *Hum Mol Genet*. (2005) 14:241–53. doi: 10.1093/hmg/ddi022
31. Liu FF, Wang J, Hu F, Wei Q, Li K. Gene coexpression networks analysis of sickle stroke risk. *J Cell Biochem*. (2019) 120:15182–9. doi: 10.1002/jcb.28780
32. Li K, Hu F, Xiong W, Wei Q, Liu FF. Network-based transcriptomic analysis reveals novel melatonin-sensitive genes in cardiovascular system. *Endocrine*. (2019) 64:414–9. doi: 10.1007/s12020-019-01925-w
33. Zhao W, Langfelder P, Fuller T, Dong J, Li A, Hovarth S. Weighted gene coexpression network analysis: state of the art. *J Biopharm Stat*. (2010) 20:281–300. doi: 10.1080/10543400903572753
34. Askland K, Read C, Moore J. Pathways-based analyses of whole-genome association study data in bipolar disorder reveal genes mediating ion channel activity and synaptic neurotransmission. *Hum Genet*. (2009) 125:63–79. doi: 10.1007/s00439-008-0600-y
35. Kao C-F, Chuang L-C, Kuo P-H. Risk and information evaluation of prioritized genes for complex traits: application to bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet*. (2014) 165B:596–606. doi: 10.1002/ajmg.b.32263
36. Zhao Y, Liang X, Zhu F, Wen Y, Xu J, Yang J, et al. A large-scale integrative analysis of GWAS and common meQTLs across whole life course identifies genes, pathways and tissue/cell types for three major psychiatric disorders. *Neurosci Biobehav Rev*. (2018) 95:347–52. doi: 10.1016/j.neubiorev.2018.10.005
37. Wang XL, Yuan K, Zhang W, Li SX, Gao GF, Lu L. Regulation of circadian genes by the MAPK pathway: implications for rapid antidepressant action. *Neurosci Bull*. (2019) 36:66–76. doi: 10.1007/s12264-019-00358-9
38. Melo MCA, Abreu RLC, Linhares Neto VB, de Bruin PFC, de Bruin VMS. Chronotype and circadian rhythm in bipolar disorder: a systematic review. *Sleep Med Rev*. (2017) 34:46–58. doi: 10.1016/j.smrv.2016.06.007
39. Wieck A, Grassi-Oliveira R, do Prado CH, Rizzo LB, de Oliveira AS, Kommers-Molina J, et al. Differential neuroendocrine and immune responses to acute psychosocial stress in women with type 1 bipolar disorder. *Brain Behav Immun*. (2013) 34:47–55. doi: 10.1016/j.bbi.2013.07.005
40. Meinhard N, Kessing LV, Vinberg M. The role of estrogen in bipolar disorder, a review. *Nord J Psychiatry*. (2014) 68:81–7. doi: 10.3109/08039488.2013.775341
41. Ho DM, Artavanis-Tsakonas S, Louvi A. The Notch pathway in CNS homeostasis and neurodegeneration. *Wiley Interdiscip Rev Dev Biol*. (2020) 9:e358. doi: 10.1002/wdev.358
42. Ables JL, Breunig JJ, Eisch AJ, Rakic P. Not(ch) just development: Notch signalling in the adult brain. *Nat Rev Neurosci*. (2011) 12:269–83. doi: 10.1038/nrn3024
43. Hoseth EZ, Krull F, Dieset I, Morch RH, Hope S, Gardsjord ES, et al. Attenuated notch signaling in schizophrenia and bipolar disorder. *Sci Rep*. (2018) 8:5349. doi: 10.1038/s41598-018-23703-w
44. Pedroso I, Lourdusamy A, Rietschel M, Nothen MM, Cichon S, McGuffin P, et al. Common genetic variants and gene-expression changes associated with bipolar disorder are over-represented in brain signaling pathway genes. *Biol Psychiatry*. (2012) 72:311–7. doi: 10.1016/j.biopsych.2011.12.031
45. Dieset I, Djurovic S, Tesli M, Hope S, Mattingsdal M, Michelsen A, et al. Up-regulation of NOTCH4 gene expression in bipolar disorder. *Am J Psychiatry*. (2012) 169:1292–300. doi: 10.1176/appi.ajp.2012.11091431
46. Ahearn EP, Speer MC, Chen YT, Steffens DC, Cassidy F, Van Meter S, et al. Investigation of Notch3 as a candidate gene for bipolar disorder using brain hyperintensities as an endophenotype. *Am J Med Genet Suppl*. (2002) 114:652–8. doi: 10.1002/ajmg.10512
47. Wang J, Li J, Kong F, Lv H, Guo Z. Bipolar II disorder as the initial presentation of CADASIL: an underdiagnosed manifestation. *Neuropsychiatr Dis Treat*. (2017) 13:2175–9. doi: 10.2147/NDT.S142321
48. Liu Y, Gu HY, Zhu J, Niu YM, Zhang C, Guo GL. Identification of hub genes and key pathways associated with bipolar disorder based on weighted gene co-expression network analysis. *Front Physiol*. (2019) 10:1081. doi: 10.3389/fphys.2019.01081
49. Greenblatt DY, Cayo MA, Adler JT, Ning L, Haymart MR, Kunnimalaiyaan M, et al. Valproic acid activates Notch1 signaling and induces apoptosis in medullary thyroid cancer cells. *Ann Surg*. (2008) 247:1036–40. doi: 10.1097/SLA.0b013e3181758d0e
50. Carniglia L, Ramirez D, Durand D, Saba J, Turati J, Caruso C, et al. Neuropeptides and microglial activation in inflammation, pain, and neurodegenerative diseases. *Mediators Inflamm*. (2017) 2017:5048616. doi: 10.1155/2017/5048616
51. Segura AG, Mitjans M, Jimenez E, Fatjo-Vilas M, Ruiz V, Saiz PA, et al. Association of childhood trauma and genetic variability of CRH-BP and FKBP5 genes with suicidal behavior in bipolar patients. *J Affect Disord*. (2019) 255:15–22. doi: 10.1016/j.jad.2019.05.014
52. Stelzhammer V, Alsaif M, Chan MK, Rahmoune H, Steeb H, Guest PC, et al. Distinct proteomic profiles in post-mortem pituitary glands from bipolar disorder and major depressive disorder patients. *J Psychiatr Res*. (2015) 60:40–8. doi: 10.1016/j.jpsychires.2014.09.022
53. Belvederi Murri M, Prestia D, Mondelli V, Pariante C, Patti S, Olivieri B, et al. The HPA axis in bipolar disorder: systematic review and meta-analysis. *Psychoneuroendocrinology*. (2016) 63:327–42. doi: 10.1016/j.psyneuen.2015.10.014
54. Amare AT, Schubert KO, Klingler-Hoffmann M, Cohen-Woods S, Baune BT. The genetic overlap between mood disorders and cardiometabolic diseases: a systematic review of genome wide and candidate gene studies. *Transl Psychiatry*. (2017) 7:e1007. doi: 10.1038/tp.2016.261

55. Sofroniew MV, Howe CL, Mobley WC. Nerve growth factor signaling, neuroprotection, and neural repair. *Annu Rev Neurosci*. (2001) 24:1217–81. doi: 10.1146/annurev.neuro.24.1.1217
56. Liu X, Zhang T, He S, Hong B, Chen Z, Peng D, et al. Elevated serum levels of FGF-2, NGF, and IGF-1 in patients with manic episode of bipolar disorder. *Psychiatry Res*. (2014) 218:54–60. doi: 10.1016/j.psychres.2014.03.042
57. Ament SA, Szelinger S, Glusman G, Ashworth J, Hou L, Akula N, et al. Rare variants in neuronal excitability genes influence risk for bipolar disorder. *Proc Natl Acad Sci U S A*. (2015) 112:3576–81. doi: 10.1073/pnas.1424958112
58. Pedrotti Moreira F, Cardoso TC, Mondin TC, Wiener CD, de Mattos Souza LD, Oses JP, et al. Serum level of nerve growth factor is a potential biomarker of conversion to bipolar disorder in women with major depressive disorder. *Psychiatry Clin Neurosci*. (2019) 73:590–3. doi: 10.1111/pcn.12896
59. Zhang L, Hu L, Li X, Zhang J, Chen B. The DRD2 rs1800497 polymorphism increase the risk of mood disorder: evidence from an update meta-analysis. *J Affect Disord*. (2014) 158:71–7. doi: 10.1016/j.jad.2014.01.015
60. Wang Z, Jun C, Gao K, Yang H, Fang Y. Perspective on etiology and treatment of bipolar disorders in china: clinical implications and future directions. *Neurosci Bull*. (2019) 35:608–12. doi: 10.1007/s12264-019-00389-2
61. Xu Y, Wang J, Rao S, Ritter M, Manor LC, Backer R, et al. An integrative computational approach to evaluate genetic markers for bipolar disorder. *Sci Rep*. (2017) 7:6745. doi: 10.1038/s41598-017-05846-4
62. Langfelder P, Horvath S. *WGCNA package FAQ*. (2017). Available online at: https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/faq.html (accessed September 4, 2020).
63. Cao Y, Zhu J, Jia P, Zhao Z. scRNASeqDB: a database for RNA-seq based gene expression profiles in human single cells. *Genes*. (2017) 8:368. doi: 10.3390/genes8120368
64. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, et al. A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci U S A*. (2015) 112:7285–90. doi: 10.1073/pnas.1507125112

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 07 April 2021
doi: 10.3389/fgene.2021.622886

Familial Psychosis Associated With a Missense Mutation at *MACF1* Gene Combined With the Rare Duplications DUP3p26.3 and DUP16q23.3, Affecting the *CNTN6* and *CDH13* Genes

*Josep Pol-Fuster[1,2], Francesca Cañellas[1,3], Laura Ruiz-Guerra[1,4], Aina Medina-Dols[1,4], Bàrbara Bisbal-Carrió[1,2], Víctor Asensio[1,5], Bernat Ortega-Vila[1,5], Diego Marzese[1,4], Carme Vidal[5], Carmen Santos[5], Jerònia Lladó[1,2], Gabriel Olmos[1,2], Damià Heine-Suñer[1,5], Konstantin Strauch[6,7], Antònia Flaquer[6,7] and Cristòfol Vives-Bauzà[1,2,4]**

[1] Health Research Institute of Balearic Islands (IdISBa), Palma, Spain, [2] Department of Biology, University of Balearic Islands (UIB) and Institut Universitari d'Investigacions en Ciències de la Salut, IUNICS, Palma, Spain, [3] Psychiatry Service, University Hospital Son Espases (HUSE), Palma, Spain, [4] Research Unit, HUSE, Palma, Spain, [5] Genomic Service Balearic Islands (GEN-IB), HUSE, Palma, Spain, [6] Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Munich, Germany, [7] Institute of Genetic Epidemiology, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg, Germany

Psychosis is a highly heritable and heterogeneous psychiatric condition. Its genetic architecture is thought to be the result of the joint effect of common and rare variants. Families with high prevalence are an interesting approach to shed light on the rare variant's contribution without the need of collecting large cohorts. To unravel the genomic architecture of a family enriched for psychosis, with four affected individuals, we applied a system genomic approach based on karyotyping, genotyping by whole-exome sequencing to search for rare single nucleotide variants (SNVs) and SNP array to search for copy-number variants (CNVs). We identified a rare non-synonymous variant, g.39914279 C > G, in the *MACF1* gene, segregating with psychosis. Rare variants in the *MACF1* gene have been previously detected in SCZ patients. Besides, two rare CNVs, DUP3p26.3 and DUP16q23.3, were also identified in the family affecting relevant genes (*CNTN6* and *CDH13*, respectively). We hypothesize that the co-segregation of these duplications with the rare variant g.39914279 C > G of *MACF1* gene precipitated with schizophrenia and schizoaffective disorder.

Keywords: copy number variant, whole exome sequencing, schizophrenia, *CDH13*, *CNTN6*, *MACF1*

OPEN ACCESS

Edited by:
Noèlia Fernàndez-Castillo, Centre for Biomedical Network Research (CIBER), Spain

Reviewed by:
Jian-Huan Chen, Jiangnan University, China
Guillaume Huguet, CHU Sainte-Justine, Canada

***Correspondence:**
Cristòfol Vives-Bauzà
cristofol.vives@uib.cat; tofolvb@yahoo.com

Specialty section:
This article was submitted to Behavioral and Psychiatric Genetics, a section of the journal Frontiers in Genetics

Received: *29 October 2020*
Accepted: *15 March 2021*
Published: *07 April 2021*

Citation:
Pol-Fuster J, Cañellas F, Ruiz-Guerra L, Medina-Dols A, Bisbal-Carrió B, Asensio V, Ortega-Vila B, Marzese D, Vidal C, Santos C, Lladó J, Olmos G, Heine-Suñer D, Strauch K, Flaquer A and Vives-Bauzà C (2021) Familial Psychosis Associated With a Missense Mutation at MACF1 Gene Combined With the Rare Duplications DUP3p26.3 and DUP16q23.3, Affecting the CNTN6 and CDH13 Genes. Front. Genet. 12:622886. doi: 10.3389/fgene.2021.622886

INTRODUCTION

Schizophrenia (SCZ), despite a relatively low prevalence, estimated to be around 0.75% (Moreno-Küstner et al., 2018), is ranked among the top 15 leading causes of disability worldwide (GBD 2016 Disease and Injury Incidence and Prevalence Collaborators, 2017). As much as 80% (Lichtenstein et al., 2009; Hilker et al., 2018) of SCZ risk may be explained by genetic factors, including both

common and rare genomic variants. Through genome-wide association studies (GWAS) hundreds of variants have been identified (Manolio et al., 2009; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2018), with an overall contribution of around 25% (Lee et al., 2012). Common SNPs have shown weak individual effects (odds ratios, <1.2); however, copy number variants (CNVs), with a high impact on gene dosage, have been implicated in SCZ with larger effect size compared to SNPs (odds ratios, 2–57) (Rees et al., 2014). Moreover, increased SCZ susceptibility has been associated with increased frequencies of CNVs (The International Schizophrenia Consortium, 2008; Grozeva et al., 2012; Szatkiewicz et al., 2020) and large deletions (Bergen et al., 2012; Szatkiewicz et al., 2020). In addition to CNVs, rare *de novo* or recent single-nucleotide variants (SNVs) with large effects have emerged as highly penetrant contributors (Purcell et al., 2014; Genovese et al., 2016; Singh et al., 2016; Steinberg et al., 2017), thanks to the availability of next-generation whole genome or exome sequencing. Overall, psychotic disorders as SCZ or bipolar disorder (BD) are the consequence of inheriting a genomic architecture where rare variants with high impact precipitate together with common variants with lower impact. Family-based studies are a powerful strategy to identify rare SNVs or CNVs without the need of collecting large cohorts. The purpose of the current study was to conduct a systematic genomic approach to study a Majorcan family enriched for psychosis, having two subjects with SCZ, and two with schizoaffective disorder (SCA).

MATERIALS AND METHODS

Psychiatric and Cognitive Assessments

All family members underwent semi-structured interviews, using the Spanish version of the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) Axis I (SCID-I), the Positive and Negative Syndrome Scale (PANNS), and the Diagnostic Interview for Genetic Studies (DIGS) (**Supplementary Table 1**). The cognitive assessment was performed using the Measurement and Treatment Research to Improve Cognition in Schizophrenia Consensus Cognitive Battery (MATRICS). It includes the assessment of 7 cognitive domains: speed of processing, attention/vigilance, working memory, verbal learning, visual learning, reasoning and problem solving, and social cognition. The same trained psychologist performed all cognitive assessments and most of them were done domiciliary. Medication status was taken into account at the moment of the assessment.

Subject Description

The proband of the Majorcan Family affected by SCA is a mother of 5 children (subject SCA-3, **Figure 1A**). She is characterized by alternating depressive and manic episodes accompanied by delusional ideation of paranoid type. The onset of symptoms was after the birth of her twin daughters. The proband has three sons (**Figure 1A**): SCZ-7 presents SCZ from the age of 19 with predominant negative symptoms. SCZ-8 onset with psychotic symptoms when he was 16 years old; he was diagnosed with paranoid SCZ and died by suicide in 2014. And SCA-9 started to evidence behavioral disorders, anxiety and somatic symptoms at the age of 14, diagnosed later as SCA. The twin daughters do not present psychiatric symptoms (subjects 10 and 11) at the age of 33 years old. The family history of the proband included a brother who died of cirrhosis due to alcoholism (subject 304, **Figure 1A**), an uncle with autism and psychosis (subject 203, **Figure 1A**), another uncle with mental illness, and epilepsy (subject 206, **Figure 1A**), and a cousin with anxiety (subject 6, **Figure 1A**). Subject 6 has a daughter affected by BD I (subject BD-12, **Figure 1A**). Her father (subject 5, **Figure 1A**) is diagnosed with major depressive disorder (MDD). The analysis presented in this article is focused on the nuclear family comprised of SCA-3, her husband, subject 2, and their children, SCZ-7, SCZ-8, SCA-9, and the twins, 10 and 11 (**Figure 1A**). Additionally, some extended family members of this nuclear family could be analyzed and were checked for the variants identified (Subjects 1, 4, 5, 6, 13, 14, BD-12, and 15; **Figure 1**). All studied family members signed the informed consent and the project was approved by the ethics committee of the Balearic Islands (CEI-IB).

Blood Collection and DNA Extraction

Peripheral blood samples were obtained in EDTA tubes for the 7 members of the nuclear family and 8 additional subjects of the extended family. DNA extraction was performed using the DNA Blood Extraction Kit (5Prime®) according to the manufacturer's protocol.

Genotyping by SNP Array

Genotyping was carried in the Genome Analysis Centre (GAC, Helmholtz Zentrum München) using the Infinium Global Screening Array-24 v1.0 (GSA) from Illumina (Illumina Inc., San Diego, CA, United States), which includes 642,824 SNPs. Besides, a pool of 57,254 SNPs previously related to neurological disorders was also genotyped. The genotype calling and CNV analysis were performed using the Genome Studio 2.0 (Illumina Inc. San Diego, CA, United States). Family samples genotyped by SNP array included all the samples of the nuclear family (subjects 2, SCA-3, SCZ-7, SCZ-8, SCA-9, 10, and 11), and extended family members (subjects 1, 4, 5, 6, 13, 14, BD-12 and 15). In total, 15 samples were genotyped by SNP array (**Figure 1**). For the CNV analysis, the *cnv*Partition algorithm, implemented in Genome Studio 2.0, was used (Illumina Inc. San Diego, CA, United States), taking as a reference GRCh37/hg19. This algorithm is based on two parameters: the B allele frequency (BAF) and the LogRRatio (LRR); both can be used to test the genotyping quality of the samples and to check the presence of CNVs across the genome. The BAF is a measure of allelic imbalance. In a normal well-genotyped sample, three genotypes are expected: homozygous AA, heterozygous AB, and homozygous BB. Once referred to the B allele, BAF is expected to have three discrete values: 0, 0.5, and 1 (representing AA, AB, and BB genotypes, respectively). R is defined as the sum of the probe intensities used to genotype the different markers. When R is normalized, it becomes the LRR which is a measure of relative intensity,

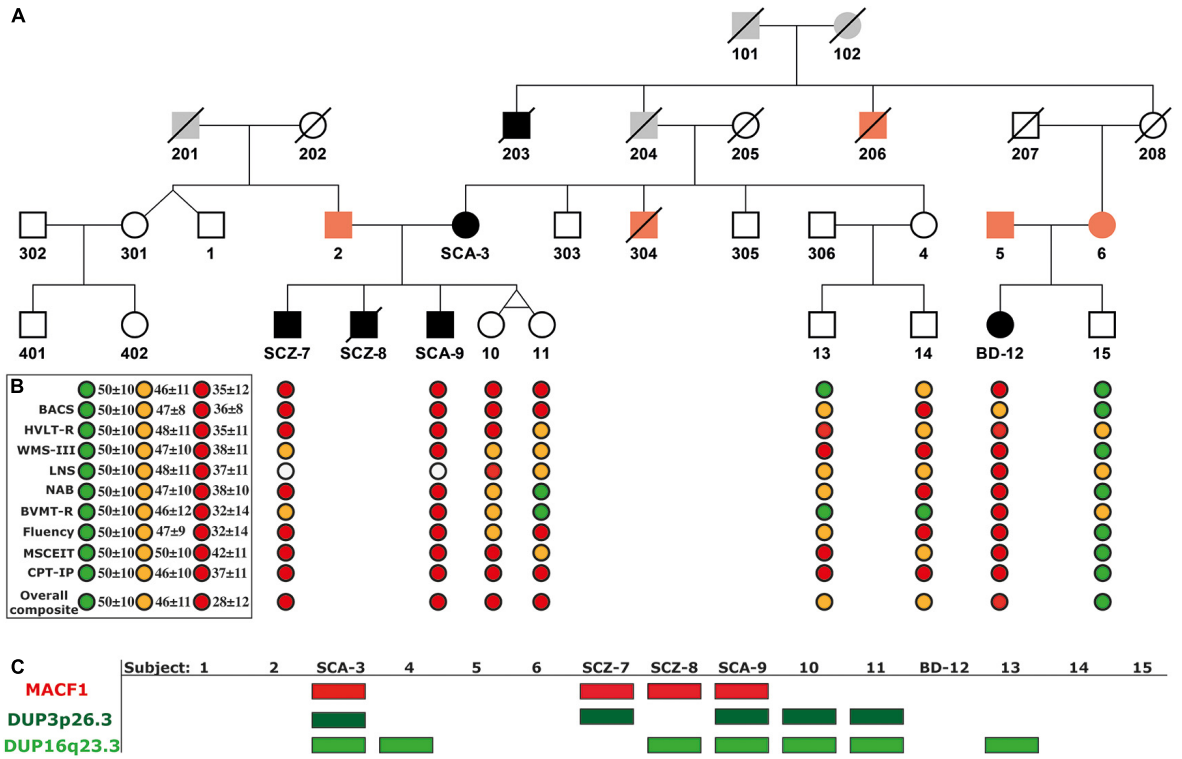


FIGURE 1 | Extended Pedigree of the Majorcan Family with the cognitive evaluation and the segregation of genomic variants associated with psychosis. **(A)** Pedigree. In black, subjects diagnosed with psychosis; in orange, subjects with a history of mental disorders without psychosis; in gray, unknown individuals. **(B)** Cognitive assessment. TMT, Trail Making Test (Part A); BACS, Brief Assessment of Cognition in Schizophrenia (Symbol coding); HVLT-R, Hopkins Verbal Learning Test-Revised; WMS-III, Wechsler Memory Scale-Third Edition (Spatial Span); LNS, Letter-Number Span Test; NAB Mazes, Neuropsychological Assessment Battery Mazes; BVMT-R, Brief Visuospatial Memory Test-Revised; Fluency: Category Fluency Test, Animal naming; MSCEIT, Mayer-Salovey-Caruso Emotional Intelligence Test; CPT-IP, Continuous Performance Test-Identical Pairs; Overall Composite Score, Cognitive index obtained summing all the tests scores and standardizing the result to a T score. In green, values of healthy controls; in orange, values of unaffected relatives with scores below the 25th percentile within their category; in red, values of patients with psychosis. **(C)** Segregation of the c.39,914,279 C > G mutation in the *MACF1* gene and the duplications DUP3p26.3 and DUP16q23.3.

the logarithm (base 2) of the observed value of R (observed probe intensity) divided by the expected value (expected probe intensity) (Peiffer et al., 2006). Before running CNV analysis, array samples were quality controlled considering the missing call rate per subject, the number of null SNPs of a sample over the total number of genotyped SNPs. The missing call rate for all genotyped samples was below 1% (call rate over 99%) except for sample SCZ-9 which had a 10.92% missing call rate (call rate 89%). When the SCZ-9 array sample was visualized in Genome Studio 2.0, it could be observed that both the BAF and the LRR were altered for this sample. These results were consistent with a pattern of contamination, reason why this sample was excluded from the analysis. CNVs < 80,000bp were filtered. CNVs were visually inspected using Genome Studio 2.0, and CNVs carried by more than one subject of the nuclear family were validated by Digital Droplet PCR (ddPCR) (**Supplementary Table 2**).

CNV Analysis by ddPCR

To confirm the CNVs identified by SNP array, Droplet Digital PCR (ddPCR) was performed. Before ddPCR, the genomic DNA of each family member was digested for 1 h at 37°C with the restriction enzyme Hind III. ddPCR was performed using the following set of primers: for DUP16q23.3 5′-TTGGTGTTTGACCCTGTGAA-3′ (Forward) and 5′-TGA GCTAGGGCTCCCACTTA-3′ (Reverse); for DUP3p26.3 5′-TCAGTGAAGTGCCTGGTTTG-3′ (Forward) and 5′-GGCT GTTCCATGAGGAATGT-3′) (Reverse). And the following 5′-FAM probes with a 3′-BHQ1 quencher: for DUP16q23.3 5′-FAM-TTTGGATTGCTTTGCCTACC-BHQ1-3′, and DUP3p 26.3 5′-FAM-CTAGGCTGGGCTCACTTGTC-BHQ1-3′. RPP30 was used as a reference control: 5′GATTTGGACCTG CGAGCG-3′ (Forward), 5′-GCGGCTGTCTCCACAAGT-3′ (Reverse), and the probe VIC-CTGACCTGAAGGCTCT-BHQ1 (**Supplementary Table 3**). The PCR was performed in a C1000 touch thermocycler (Bio-Rad, Hercules, CA, United States) with the following protocol: Initial denaturalization at 95°C for 10 min, followed by 39 cycles at 94°C for 30 s and an extension at 57°C for 1 min, with a final denaturing at 98°C for 10 min. Droplet analysis was performed using the Q200 Droplet Reader (Bio-Rad). CNVs were calculated using the QuantaSoft software (Bio-Rad)[1].

[1]http://www.bio-rad.com/en-us/product/qx200-droplet-digital-pcr-system

Whole Exome Sequencing (WES)

WES included three affected members (SCA-3, SCZ7, and SCZ-8) and one of the unaffected twin sisters (subject 10). Additionally, subject BD-12 from the extended family was also whole-exome sequenced. The SureSelectXT Human All Exon V6 capture library from Agilent Technologies (Santa Clara, CA, United States) was used for target enrichment. Exome sequencing was performed on an Illumina HiSeq4000 sequencer (Illumina Inc. San Diego, CA, United States), using 100 base paired-end sequencing, by BGI Tech Solutions. The workflow to obtain variant call format (VCF) files from raw data (FASTQ) provided by Macrogen was based on GATK Best Practices (Depristo et al., 2011). FASTQ files, containing raw unmapped reads and Phred scores, were quality controlled using the FastQC v0.11.2[2]. Low-quality sequences (Phred score < 20) and adaptors were removed using Cutadapt (v1.4)[3]. QC sequences were aligned against the reference human genome (GRCh37/hg19) using the BWA-MEM algorithm implemented in the Burrows-Wheeler Alignment tool (BWA v0.7.12)[4]. Aligned data in SAM (Sequence Alignment/Map) format were then sorted and converted into BAM files using SAMtools (v1.1)[5]. To generate new BAM files, PCR duplicates were removed using Picard Tools v1.118[6], and realignment around INDELS and base recalibration was performed (BQRS) using Genome Analysis Toolkit (GATK). SNV and INDEL calling was carried from the cleaned BAM files using GATK (v3.3.0)[7] producing unfiltered primary VCF files; which were then hard filtered to generate the definitive VCF files. To verify the exome sequencing results, Sanger sequencing was performed in the Genetic Analyzer AbiPrism 3700 (Applied Biosystems, Foster City, CA, United States) using standard polymerase chain reaction conditions. Results were visualized using Sequencing Analysis 5.1.1. (Applied Biosystems).

Analysis of Rare SNV and CNV Variants

VCF files were annotated using variant annotation and effect prediction tool (SnpEff and SnpSift; Version 4.3)[8] (Cingolani et al., 2012). It allows annotating variants using the dbNSFP database, an integrated database of functional predictions from multiple algorithms. Variants were annotated, based on predictions comprised in the 2.1 version of the dbNSFP database, for protein impact using SIFT (Kumar et al., 2009), PolyPhen2 (Carter et al., 2013), LRT (Chun and Fay, 2009), MutationAssessor (Reva et al., 2007), Meta SVM and LR (Dong et al., 2015), VEST3 (Carter et al., 2013), FATHMM (Shihab et al., 2013), and MutationTaster (Schwarz et al., 2010); and for conservation using SiPhy_29way_logOdds (Garber et al., 2009), PhyloP100way_vertebrate (Pollard et al., 2010), PhastCons100way_vertebrate (Felsenstein and Churchill, 1996) and GERP++ (Davydov et al., 2010). The resulting files were managed with R 3.6.1. Alternatively, VCF files were managed using ENLIS Genome Research version 1.9 (Berkeley, CA, United States). ENLIS uses its annotation pipeline. The RVI Score (Residual Variation Intolerance Score)[9] (Petrovski et al., 2013) was also used to assess the tolerance to variation of the genes affected by a variant in the family. RVIS is a ranked score of genes according to the number of common functional genetic variants they carry compared to the genome-wide expectation. RVIS is based on 6,503 WES samples from the NHLBI Exome Sequencing Project and it is the result of regressing the number of common functional variants on the total number of protein-coding variants. The loss-of-function observed/expected upper bound fraction (LOEUF) (Karczewski et al., 2019) was also considered as an indicator of gene tolerability to predicted loss of function variants. It is based on the continuous metric of the observed/expected (o/e) ratio and its confidence interval[10].

Shared variation among affected individuals was filtered for Read Depth >10 and Minor Allele Frequency (MAF) <0.01. All genomic data for molecular variants in this study were compatible with Genome build GRCh37. Allele frequencies were checked in 1000G[11] (Birney and Soranzo, 2015) and in ExAC and its successor, the Genome Aggregation Database (gnomAD) (see text footnote 10) (Lek et al., 2016). Before validation, SNV variants were checked directly from BAM files using Integrative Genomics Viewer (IGV)[12] (Thorvaldsdottir et al., 2013). The identified SNV variants and the genes affected by them were checked on VarElect[13] (Stelzer et al., 2016), DisGeNET[14] (Piñero et al., 2020), and Schizophrenia Exome Sequencing Genebook (Purcell et al., 2014) to identify potential previous reports of these variants or genes in psychosis. Regarding CNVs, we checked different databases to find previous reports of the CNVs identified in the family and to study their pathogenicity and conservation: DGV (Database of genomic variants)[15] (MacDonald et al., 2014) and DECIPHER (DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources)[16] (Firth et al., 2009). VarSome[17] (Kopanos et al., 2019) and UniProt[18] (Bateman, 2019) were used to predict the levels of variant penetrance.

RESULTS

The Majorcan Family Is Characterized by Poor Cognitive Profiles

Psychotic subjects obtained lower overall composite scores compared to healthy family controls (**Supplementary Table 4**).

[2]https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
[3]https://cutadapt.readthedocs.io/en/stable/
[4]https://github.com/lh3/bwa/releases/tag/0.7.12
[5]https://github.com/samtools/samtools/releases/tag/1.1
[6]https://github.com/broadinstitute/picard/releases/tag/1.118
[7]https://github.com/broadinstitute/gatk/releases
[8]http://snpeff.sourceforge.net/
[9]http://genic-intolerance.org/
[10]https://gnomad.broadinstitute.org/
[11]www.1000genomes.org
[12]http://software.broadinstitute.org/software/igv/
[13]http://varelect.genecards.org/
[14]https://www.disgenet.org
[15]http://dgv.tcag.ca/dgv/app/home
[16]https://decipher.sanger.ac.uk/
[17]http://varsome.com
[18]https://www.uniprot.org

Once scores were compared with normal population values obtained from Mucci et al. (2017) (**Figure 1B**), all family members performed below average. Processing Speed was especially low in subjects SCZ-7 and SCA-9, obtaining both of them the lowest possible scores in TMT and BACS tests (percentiles <0.1 considering their reference group matched for age and sex; **Supplementary Table 4**). Yet, these subjects performed better in fluency evaluation. Their sisters (subjects 10 and 11) and subject BD-12 also had low scores in processing speed (percentiles <5 considering their reference group matched for age and sex; **Supplementary Table 4**). Subjects 13 and 14 had scores in the range of healthy individuals. Working memory was especially low in the affected subject BD-12. LNS was not tested on affected subjects SCZ-7 and SCA-9 because they could not remember the alphabet.

Identification of Two Inherited CNVs Associated With Psychosis

Two rare duplications (DUP3p26.3 and DUP16q23.3) were identified in the nuclear family (**Figures 2A,B** and **Supplementary Table 2**). These CNVs were validated by ddPCR in all studied family members (**Figures 2A,B**, lower panels). DUP3p26.3 is located on chromosome 3 (from 1,159,787 to 1,781,739 bp) and contains the contactin 6 (*CNTN6*) gene. Within these coordinates, 48 CNVs of similar size (under 1 Mb), containing the *CNTN6* gene, have been reported (DECIPHER) (Firth et al., 2009). The gain of dose of this gene has been associated with intellectual disability (ID), delayed speech and language development, cognitive impairment, micro and macrocephaly, and autism (DECIPHER). The mother (SCA-3) carries DUP3p26.3 and transmitted it to her affected sons SCZ-7 and SCA-9 (**Figure 2A**). The healthy twin daughters also carry this duplication (subjects 10 and 11, **Figure 2A**). None of the extended relatives of the family carries it.

DUP16q23.3 is located on chromosome 16 (from 82,180,075 to 83,664,582 bp) and contains the cadherin-13 (*CDH13*) gene. 28 CNVs of similar size (under 2 Mb) have been reported within these coordinates, 18 of which are duplications affecting the *CDH13* gene and are associated with ID, delayed speech, language development, and global developmental delay (DECIPHER). The mother (SCA-3) carries this duplication and she transmitted it to her affected sons SCZ-8 and SCA-9 (**Figure 2B**). Interestingly, the healthy twin daughters also carry this duplication (subjects 10 and 11, **Figure 2B**). Among the extended relatives of the family, DUP16q23.3 is also carried by an unaffected sister of the mother (subject 4), who transmitted it to one of her two unaffected sons, subject 13 (**Figure 2B** and **Supplementary Table 2**).

Contributions of Rare SNVs Identified by WES

To identify rare variants of potentially higher penetrance, WES analysis was restricted to MAF < 1% according to ExAC, 1000G, and gnomAD. The psychotic patients shared 19 rare missense SNVs in conserved regions according to different conservation algorithms that were absent from the healthy twins (**Tables 1**, **2**). Nonsense variants were not identified (**Tables 1**, **2**). Regarding the 19 shared missense variants, all except two (rs146468598 and rs112913396) affect genes that are expressed in the cerebral cortex (GTEx database; **Table 1**). Out of these 17 variants, 9 are predicted "VUS" and 8 "benign" or "likely benign" by VarSome (**Table 1**). Among the 9 variants predicted "VUS," 3 variants stood out because they were predicted "deleterious" by almost all protein prediction algorithms; p. P-56-L in MTMR11, p. P-241-S in TINF2, and p. Y-161-C in ZRANB3 (**Table 3**). Nevertheless, to our knowledge, the genes affected by these variants have no described functions in the central nervous

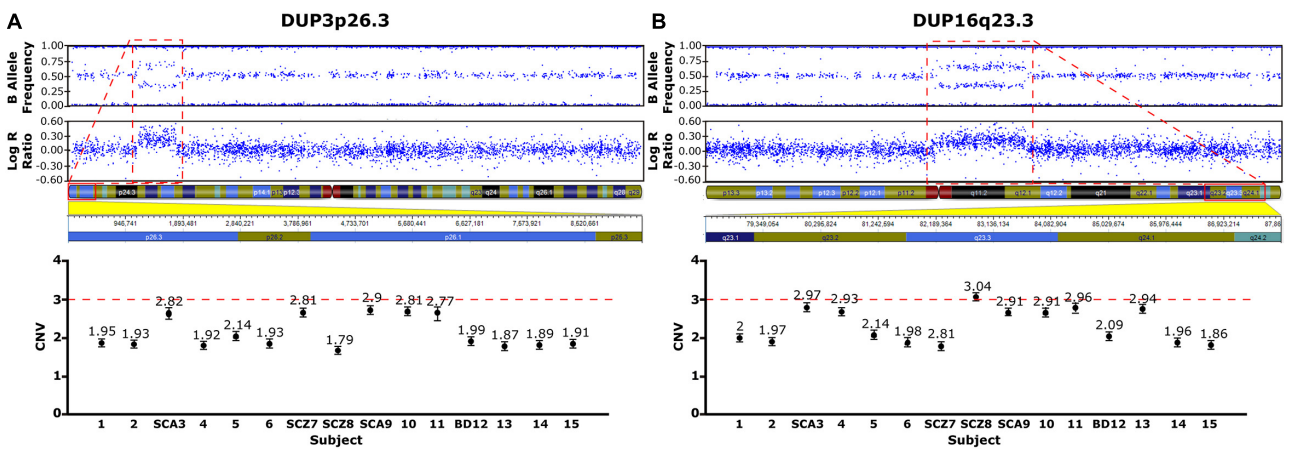


FIGURE 2 | CNVs Identified in the Family Pedigree Associated with Psychosis. Upper panels, CNVs detected using the *cnv*Partition algorithm, based on the B allele frequency and the log R ratio (LRR: logarithm of the observed/expected probe intensity). Duplications identified (red rectangles), 3p26.3 **(A)** and DUP16q23.3 **(B)** are defined by allele frequencies of 0.33 and 0.66. The genetic map of the genomic regions containing the duplications are shown below with the affected genes highlighted in red. Lower panels, CNVs validated by ddPCR. Values close to 3 are indicative of a duplication **(A,B)**.

TABLE 1 | Rare non-synonymous variants identified in conserved regions by WES shared only among affected subjects.

Chr	bp	Ref	Alt	rs	Gene	Protein Impact	aa Change	GTEX	VarSome	DisGeNET	ExAC	gnomAD	1000G
1	39914279	C	G	rs376728302	*MACF1*	MISSENSE	T-4642-S	Yes	VUS	Yes	0.000058	0.000092	NA
1	149907574	G	A	rs202234090	*MTMR11*	MISSENSE	P-56-L	Yes	VUS		0.000115	0.000159	NA
2	136107663	T	C	rs183525970	*ZRANB3*	MISSENSE	Y-161-C	Yes (low)	VUS		0.001391	0.001672	0.001374
2	219528732	C	T	rs138321594	*RNF25*	MISSENSE	R-443-Q	Yes	VUS		0.000069	0.000089	NA
3	102187900	A	G	rs145497598	*ZPLD1*	MISSENSE	K-301-R	Yes (Very low)	VUS		0.000025	0.000021	NA
5	54721089	G	A	rs201703457	*PPAP2A*	MISSENSE	T-268-I	Yes	VUS		0.000016	0.000011	NA
5	175812257	G	A	rs35916027	*NOP16*	MISSENSE	R-120-C	Yes	VUS		0.004062	0.004068	0.005037
14	24709965	G	A	rs17102311	*TINF2*	MISSENSE	P-241-S	Yes	VUS		0.002158	0.002167	0.004121
22	38061637	A	T	rs142028345	*PDXP*	MISSENSE	Y-217-F	Yes	VUS		0.007042	0.000004	0.005952
1	23743859	T	C	rs192746462	*TCEA3*	MISSENSE	E-88-G	Yes (low)	Likely Benign		0.004760	0.004642	0.002289
2	109382936	T	A	rs61748150	*RANBP2*	MISSENSE	S-1981-T	Yes	Benign		0.008895	0.008229	0.006410
2	121748118	G	A	rs138987487	*GLI2*	MISSENSE	R-1543-H	Yes (low)	Benign	Yes	0.000502	0.000534	0.000916
2	230723775	G	A	NA	*TRIP12*	MISSENSE	S-205-L	Yes	Benign		0.000025	0.000028	NA
4	39450295	G	A	rs143809363	*KLB*	MISSENSE	V-1042-I	Yes (low)	Likely Benign		0.003534	0.003275	0.000458
9	79323853	C	T	rs200875180	*PRUNE2*	MISSENSE	D-1113-N	Yes	Likely Benign		0.003781	0.003776	0.000916
10	81925866	T	C	rs72807973	*ANXA11*	MISSENSE	I-278-V	Yes	Benign		0.002990	0.003189	0.003205
15	94841610	G	A	rs149237812	*MCTP2*	MISSENSE	R-39-Q	Yes (low)	Likely Benign	Yes	0.001400	0.001199	0.000916
2	238289767	T	C	rs112913396	*COL6A3*	MISSENSE	D-563-G		VUS		0.002314	0.002079	0.000916
5	56777915	C	T	rs146468598	*ACTBL2*	MISSENSE	R-207-Q		VUS		0.001820	0.002011	0.003205

19 variants with a MAF < 1% were shared by the 3 psychotic patients.
Chr, chromosome; bp, base pair; Ref, reference Allele; Alt, alternate allele; GTEX, GTEX expression; DisGeNET, previous relationship to SCZ; VUS, variant of uncertain significance.
**Low median tpm below 2, **Very low median tpm below 0.05, ***Genetic coordinates are given taking as a reference GRCh37/hg19, ****Variants were annotated based on the 2.1 version of the dbNSFP database, *****Allele frequencies from gnomAD were obtained from the 2.1.1 gnomAD version.*

TABLE 2 | Conservation algorithms.

Chr	bp	Ref	Alt	rs	Gene	SiPhy	GERP_RS	GERP_NR	phyloP100	phastCons100
5	56777915	C	T	rs146468598	*ACTBL2*	15.631	4.91	4.91	7.645	0.989
10	81925866	T	C	rs72807973	*ANXA11*	9.0566	1.76	5.6	3.668	0.879
2	238289767	T	C	rs112913396	*COL6A3*	15.793	5.6	5.6	6.22	1
2	121748118	G	A	rs138987487	*GLI2*	18.456	4.98	4.98	9.657	1
4	39450295	G	A	rs143809363	*KLB*	12.583	4.83	5.68	1.477	0.998
1	39914279	C	G	rs376728302	*MACF1*	20.393	5.95	5.95	7.818	1
15	94841610	G	A	rs149237812	*MCTP2*	9.3437	3.19	5.13	1.658	0.044
1	149907574	G	A	rs202234090	*MTMR11*	12.986	4.68	4.68	3.916	1
5	175812257	G	A	rs35916027	*NOP16*	11.072	3.46	6.17	3.367	0.423
22	38061637	A	T	rs142028345	*PDXP*	10.036	4.54	5.58	4.666	1
5	54721089	G	A	rs201703457	*PPAP2A*	14.027	4.8	5.66	5.819	1
9	79323853	C	T	rs200875180	*PRUNE2*	14.954	5.94	5.94	4.165	0.984
2	109382936	T	A	rs61748150	*RANBP2*	16.063	5.75	5.75	5.214	1
2	219528732	C	T	rs138321594	*RNF25*	8.8576	3.74	5.55	1.989	0.99
1	23743859	T	C	rs192746462	*TCEA3*	11.332	5.02	5.02	1.69	1
14	24709965	G	A	rs17102311	*TINF2*	8.383	3.46	4.36	1.43	0.942
2	230723775	G	A	NA	*TRIP12*	19.489	5.72	5.72	9.434	1
3	102187900	A	G	rs145497598	*ZPLD1*	15.196	5.28	5.28	5.352	1
2	136107663	T	C	rs183525970	*ZRANB3*	15.459	5.42	5.42	6.253	1

Chr, chromosome; bp, base pair; Ref, reference Allele; Alt, alternate allele; SiPhy, SiPhy_29way_logOdds; phyloP100, phyloP100way_vertebrate; phastCons, phastCons100way_vertebrate.
**Genetic coordinates are given taking as a reference GRCh37/hg19, **Conservation scores produced by the different algorithms were obtained from the 2.1 version of the dbNSFP database.*

TABLE 3 | Protein prediction algorithms.

Chr	bp	Ref	Alt	rs	Gene	SIFT	Polyphen2	MutAssessor	SVM	LR	VEST3	FATHMM	LRT	MutTaster
5	56777915	C	T	rs146468598	*ACTBL2*	NA	0.461 (P)	3.37 (M)	0.7735 (D)	0.8058 (D)	0.785	−3.44 (D)	0.00001 (D)	0.999 (D)
10	81925866	T	C	rs72807973	*ANXA11*	0.174 (T)	0.005 (B)	−0.98 (N)	−0.915 (T)	0.0034 (T)	0.171	4.06 (T)	0.011694 (N)	1 (N)
2	238289767	T	C	rs112913396	*COL6A3*	0.005 (D)	1.0 (D)	2.82 (M)	−0.202 (T)	0.3558 (T)	0.932	0.75 (T)	0.000042 (D)	1 (D)
2	121748118	G	A	rs138987487	*GLI2*	0.054 (T)	1.0 (D)	1.89 (L)	−0.8456 (T)	0.1776 (T)	0.415	1.59 (T)	0.000001 (D)	1 (D)
4	39450295	G	A	rs143809363	*KLB*	0.035 (D)	0.18 (B)	0.695 (N)	−1.0599 (T)	0.0642 (T)	0.125	1.67 (T)	0.442178 (N)	0.994 (N)
1	39914279	C	G	rs376728302	*MACF1*	1 (T)	0.589 (P)	0.345 (N)	−1.0558 (T)	0.0883 (T)	0.546	0.78 (T)	0.000008 (D)	1 (D)
15	94841610	G	A	rs149237812	*MCTP2*	0.166 (T)	0.024 (B)	0.46 (N)	−0.8817 (T)	0.1776 (T)	0.161	−0.66 (T)	0.22127 (N)	0.999 (N)
1	149907574	G	A	rs202234090	*MTMR11*	0.02 (D)	0.932 (P)	2.045 (M)	0.2042 (D)	0.6049 (D)	0.568	−1.67 (D)	NA	1 (D)
5	175812257	G	A	rs35916027	*NOP16*	0 (D)	0.297 (B)	1.04 (L)	−1.0548 (T)	0.054 (T)	0.355	NA	0.000101 (N)	0.999 (N)
22	38061637	A	T	rs142028345	*PDXP*	1 (T)	0.087 (B)	−0.715 (N)	−1.0489 (T)	0.0258 (T)	0.25	1.59 (T)	NA	1 (D)
5	54721089	G	A	rs201703457	*PPAP2A*	0.077 (T)	0.263 (B)	1.905 (M)	−1.0907 (T)	0.2078 (T)	0.406	1.5 (T)	0.018731 (N)	1 (D)
9	79323853	C	T	rs200875180	*PRUNE2*	0.0 (D)	1.0 (D)	0.975 (L)	−0.8847 (T)	0.2006 (T)	0.202	0.31 (T)	0.000385 (D)	1 (D)
2	109382936	T	A	rs61748150	*RANBP2*	0.33 (T)	0.39 (B)	2.045 (M)	−1.0432 (T)	0.0603 (T)	0.16	1.56 (T)	NA	0.964 (N)
2	219528732	C	T	rs138321594	*RNF25*	0.005 (D)	0.223 (B)	0.895 (L)	−0.8015 (T)	0.1606 (T)	0.135	−0.21 (T)	0.020699 (N)	0.725 (N)
1	23743859	T	C	rs192746462	*TCEA3*	0.25 (T)	0 (B)	0.695 (N)	−1.0688 (T)	0.0598 (T)	0.379	NA	0.806532 (N)	0.992 (N)
14	24709965	G	A	rs17102311	*TINF2*	0.013 (D)	1.0 (D)	2.045 (M)	0.1966 (D)	0.7489 (D)	0.172	−3.19 (D)	0.000159 (N)	1 (D)
2	230723775	G	A	NA	*TRIP12*	0.017 (D)	0.994 (D)	0.695 (N)	−0.833 (T)	0.1914 (T)	0.684	0.8 (T)	0 (D)	1 (D)
3	102187900	A	G	rs145497598	*ZPLD1*	0.358 (T)	0.004 (B)	0.345 (N)	−0.6888 (T)	0.2761 (T)	0.5	−1.34 (T)	0 (D)	1 (D)
2	136107663	T	C	rs183525970	*ZRANB3*	0.001 (D)	1.0 (D)	3.255 (M)	1.0465 (D)	0.9121 (D)	0.857	−3.21 (D)	0 (D)	1 (D)

system and have not been associated with SCZ or any other mental disorder. Only three variants among the 17 expressed in the cerebral cortex affect genes previously associated with SCZ according to DisGeNET (Piñero et al., 2020): p. R-1543-H in GLI2, p. T-4642-S in MACF1, and p. R-39-Q in MCTP2 (**Table 1**). Out of these three, only the *MACF1* variant is predicted "VUS" (**Table 1**). Moreover, the expression of *MCTP2* [median Transcripts Per Million (TPM) below 1] and *GLI2* (median TPM below 1.5) is low in the cerebral cortex (GTEx database). The variant located in the *MACF1* gene is located in the most conserved region among the 19 variants identified (**Table 2**). The variant in *GLI2* is also located in a very conserved region while the variant located in *MCTP2* is located in a less conserved region (**Table 2**). Moreover, among the 19 genes affected by the variants shared by psychotic individuals, *MACF1* is the most intolerant to variation (RVIS = −3.92 (0.21%); LOEUF = 0.084; **Supplementary Table 5**). Protein prediction algorithms show mixed results for the variants in *MACF1* and *GLI2* while the variant in *MCTP2* is predicted to be benign by all protein prediction algorithms used (**Table 3**). Given these results, we decided to analyze the segregation of the variants in *MACF1* and *GLI2*. The mother (SCA-3) transmitted the g.39914279 C > G (p. T4642S) mutation in *MACF1* to her three affected sons (SCZ-7, SCZ-8, and SCA-9), but not to her healthy twin daughters (subjects 10 and 11) (**Figure 1C**). Besides, this variant was not detected in any of the extended family members that were analyzed. The variant g.121748118 G > A (p. R1543H) in the *GLI2* gene was transmitted by the mother to her affected sons SCZ-7 and SCZ-8. The other affected individual, SCA-9, did not carry this variant which was neither detected in any of the extended family members analyzed. The presence of the duplications DUP3p26.3 and DUP16q23.3 *per se* is not sufficient to precipitate with psychosis, since non-affected subjects carry both duplications (i.e., the twin sisters, subjects 10 and 11). The psychotic phenotype may have precipitated once patients inherited one of these two duplications together with the rare variant c.39,914,279 C > G in the *MACF1* gene.

DISCUSSION

Despite their high heritability, the genetic architecture of psychotic disorders such as SCZ and BD has proven to be complex. Family-based studies, focused mainly on identifying rare inherited or rare *de novo* variants, have allowed identifying rare variants with very high penetrance, including CNVs (Malhotra and Sebat, 2012) and SNVs (Xu et al., 2011). Considering that psychosis is a complex phenotype and that the rare variants contribution is modulated by the genetic background (Nadeau, 2001; Koike et al., 2006; Kearney, 2011; Mitchell, 2011; Kerr et al., 2013) of the individuals carrying them, the more homogeneous the family and the closest the relationships included in the analysis the better.

DUP3p26.3 and DUP16q23.3 Associated With Neurodevelopmental Disorders

The two CNVs identified in this pedigree are highly suggestive to play a significant role in the pathogenesis of psychosis. The duplication DUP3p26.3 has been extensively associated with neuropsychiatric disorders, including SCZ, autism spectrum disorder (ASD), BD, attention-deficit hyperactivity disorder (ADHD), ID, and Tourette syndrome (Kashevarova et al., 2014; Te Weehi et al., 2014; Hu et al., 2015; Huang et al., 2017; Mercati et al., 2017; Juan-Perez et al., 2018). This duplication alters the expression of the *CNTN6* gene, which encodes

the protein contactin-6, also termed NB-3, a member of the contactin family of immunoglobulin domain-containing cell adhesion molecules (IgCAMs). IgCAMs are important signal molecules that mediate cell-cell adhesion and cell-extracellular matrix (ECM) interactions in multiple neural developmental processes, including neuronal migration, neurite outgrowth, axon guidance, and synaptogenesis (Sethi and Zaia, 2017). In mice, *Cntn6* is exclusively expressed in the CNS, where it is needed for proper orientation of dendrite growth in cortical pyramidal neurons (Ye et al., 2008) and synapse formation in the cerebellum (Sakurai et al., 2009). The presence of the duplication in the family may result in a decrease of the CNTN6 expression. Supporting this, a recent study comparing the CNTN6 expression in neurons generated from inducible pluripotent stem cells (iPSCs) derived from fibroblasts of an ID patient carrying a *CNTN6* duplication and two healthy controls found that neurons carrying the duplication had significantly lower expression of CNTN6 compared to wild-type neurons (Gridina et al., 2018). Although the crucial roles of CNTN6 during brain development, the penetrance of this duplication in our pedigree and others previously reported (Kashevarova et al., 2014; Hu et al., 2015) may be low since there are healthy family carriers in different generations.

The second duplication identified in the pedigree is the DUP16q23.3, also very suggestive of playing a major role in psychosis. This duplication alters the expression of *CDH13,* a gene that encodes another adhesion molecule, Cadherin-13, a calcium-dependent class of transmembrane protein that forms adherent junctions and participates in cell-cell recognition and signal transduction, among other functions. Cadherin-13 is highly expressed in the brain, where it may function as a negative regulator of neurite outgrowth and axon guidance (reviewed in Rivero et al., 2013). *CDH13* involvement in neuropsychiatric disorders is well documented, associated with the five major psychiatric disorders: ADHD (Neale et al., 2008), ASD (Sanders et al., 2011), MDD (Terracciano et al., 2010), BD (Xu et al., 2014), and SCZ (Børglum et al., 2014; Otsuka et al., 2015). Despite none of these associations have been significant at a genome-wide level, as pointed by Prata et al. (2019), in their review of GWAS studies from 2012 to 2019, *CDH13* is a highly relevant candidate for psychosis because the GWAS evidence involving this gene with psychosis comes from at least two independent samples. As observed for the DUP3p26.3, the penetrance of DUP16q23.3 in this pedigree also seems low. Both duplications need to co-segregate with other genetic factors to precipitate with psychosis, and the rare variant g.39914279 C > G of *MACF1* is a good candidate. Interestingly, all the carriers of the duplications DUP3p26.3 and DUP16q23.3 evidenced very low cognitive performance; therefore, it seems obvious that both duplications have also an impact not only in psychosis but also in other dimensions related to the disease, like cognition. In fact, other authors found that CNVs conferring risk for SCZ and autism affected cognition not only in patients but also in carrier controls (Stefansson et al., 2014). This supports our hypothesis that the combination of these CNVs with the rare variant in *MACF1* is precipitating with SCZ and SCA in the analyzed pedigree.

Facts to Consider *MACF1* a Genetic Risk Factor for Psychosis

The function of the Microtubule-actin crosslinking factor 1 (MACF1), also known as actin crosslinking factor 7 (ACF7), is to connect the actin and microtubule cytoskeletons with the sites of cell-matrix and cell-cell adhesions (Hu et al., 2016), being an important modulator of cytoskeletal networks (Hu et al., 2016; Moffat et al., 2017). Defects in these cross-talks may have a profound impact on cell proliferation and differentiation, neuronal migration, neurite development, and axon guidance (Sanchez-Soriano et al., 2009; Goryunov et al., 2010; Ka et al., 2014, 2017; Ka and Kim, 2016; Dobyns et al., 2018). Such defects are hypothesized as a pathogenic mechanism for psychosis and several genes involved in them have been associated with SCZ or BD (Lewis et al., 2005; Rossignol, 2011; Muraki and Tanigaki, 2015; Ka et al., 2017). The role of MACF1 in cytoskeletal networks makes it an interesting candidate for psychosis (Ka et al., 2017). Besides, MACF1 also interacts with GSK3β, participating in the Wnt/β-catenin and GSK-3 signaling pathway, during the migration of pyramidal neurons or neurite differentiation (Chen et al., 2006; Ka et al., 2014; Ka and Kim, 2016). Moreover, MACF1 also interacts with DISC1 and DTNBP1 (Camargo et al., 2007; Costas et al., 2013). Interestingly, rare *de novo* variants have been previously described in BD patients (Kataoka et al., 2016; Han et al., 2019) and SCZ patients (Xu et al., 2012; Kenny et al., 2014; Wang et al., 2015) affecting this gene, which is highly intolerant to protein-altering variants. Intriguingly, both the variant identified in the family, g.39914279 C > G (p. T4642S), and other variants previously reported in SCZ patients, g.39827053 C > T (p. R2097W), and g. 39904999 C > T (p. R4033W), are located in the spectrin repeats domain of MACF1 which interacts with GSK3β and other proteins that regulate neuronal migration (Chen et al., 2006).

The rare variant identified in *GLI2* may also contribute to the psychosis phenotype. The transcription factor Gli Family Zinc Finger 2 (GLI2) is a downstream modulator of the Sonic Hedgehog (Shh) signaling pathway. Disruption of the Shh signaling has been associated with several neurodevelopmental disorders (Chen et al., 2018), including SCZ (Betcheva et al., 2013). Using GWAS, the *GLI2* gene has been associated with the manifestation of tardive dyskinesia in SCZ (Greenbaum et al., 2010).

Here we have demonstrated that by combining wide-genomic approaches it is currently feasible to disclose the genomic architecture associated with psychosis in highly penetrant and homogeneous pedigrees. Our results also show that the genomics of psychosis is highly heterogeneous, which makes sense given the involvement of rare variants in its etiology. In this sense, psychosis in the family can be explained by the presence of the variant g.39914279 C > G in *MACF1* in combination with the two rare duplications, but this variant does not explain the psychosis phenotype of their distant relative BD-12. The presence of two highly suggestive duplications, and the cognitive performance

of the individuals carrying them, also suggests that the variant in *MACF1* is modulated by the genetic background of the family. Besides, our results confirm previous results suggesting the involvement of *MACF1, CNTN6,* and *CDH13* in the etiology of psychosis. More research is needed to further elucidate the role of *MACF1* in psychosis as well as its interaction with *CNTN6* and *CDH13*.

DATA AVAILABILITY STATEMENT

The raw data will be available under the accession numbers: EGAS00001004791 and EGAD00010002030 Array_data.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Balearic Islands (CEI-IB). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JP-F performed most of the analyses, including WES and CNV analysis, and wrote the manuscript. FC recruited the patients and performed their clinical characterization. LR-G, AM-D, and BB-C performed INDELS validations and digital droplet PCR. VA and DM performed SNV prioritization analyses. BO-V performed the DNA extraction and the DNA quality control. CV and CS performed Sanger sequencing validations. JL, GO, DH-S, KS, and AF contributed to NGS data analyses and experimental design. CV-B conceived and designed the analysis, supervised the study, contributed to analyze the NGS data and gene ontology, and wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the Carlos III Institute of Health (ISCIII), Ministry of Economy and Competitiveness (Spain) (grants PI15/00809 and PI18/00608), co-funded with European Union ERDF (European Regional Development Fund). JP-F FPU fellowship (FPU/2014-03876) was funded by the FPU fellowship granted by the Ministry of Education, Culture, and Sport (Spain). LR-G, AM-D, and BB-C were funded by the FPI fellowship granted by the General Direction of Innovation and Research, Ministry of Innovation, Research, and Tourism, Balearic Government, co-funded by European Union ERDF. DM was funded by ISCIII (CP17/00188).

ACKNOWLEDGMENTS

The authors thank all the members of the family for their participation. This article is dedicated to subject SCZ-8 who bravely fought the disease and lost his battle.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.622886/full#supplementary-material

Supplementary Table 1 | Clinical characterization. PANSS, Positive and Negative Syndrome Scale.

Supplementary Table 2 | CNVs identified. Chr, Chromosome; bp, base pair; Start (bp), Starting position of the CNV; End (bp), Ending position of the CNV; Size (bp), Size of the CNV in bp; *Genetic coordinates are given taking as a reference GRCh37/hg19.

Supplementary Table 3 | Validation by Digital PCR (ddPCR). CNV name, Copy Number Name; Chr, Chromosome; bp, base pair; *Genetic coordinates are given taking as a reference GRCh37/hg19; **Amplicon, Size of the amplicon in bp and its coordinates.

Supplementary Table 4 | Cognitive assessment. TMT, Trail Making Test (Part A); BACS, Brief Assessment of Cognition in Schizophrenia (Symbol coding); HVLT-R, Hopkins Verbal Learning Test-Revised; WMS-III, Wechsler Memory Scale-Third Edition (Spatial Span); LNS, Letter-Number Span Test; NAB Mazes, Neuropsychological Assessment Battery Mazes; BVMT-R, Brief Visuospatial Memory Test-Revised; Fluency, Category Fluency Test, Animal naming; MSCEIT, Mayer-Salovey-Caruso Emotional Intelligence Test; CPT-IP, Continuous Performance Test-Identical Pairs; Pi, Percentile; Overall Composite Score: It is a cognitive index which is obtained summing all the tests scores and standardizing the result to a T score. *Low scores in comparison to reference group according to Mucci et al. (2017).

Supplementary Table 5 | Residual variation intolerance score (RVIS) and loss-of-function observed/expected upper bound fraction (LOEUF). RVIS, Residual Variation Intolerance Score; LOEUF, Loss-of-function Observed/Expected Upper bound Fraction; *Percentiles are provided in parenthesis.

REFERENCES

Bateman, A. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515.

Bergen, S. E., O'Dushlaine, C. T., Ripke, S., Lee, P. H., Ruderfer, D. M., Akterin, S., et al. (2012). Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. *Mol. Psychiatry* 17, 880–886. doi: 10.1038/mp.2012.73

Betcheva, E. T., Yosifova, A. G., Mushiroda, T., Kubo, M., Takahashi, A., Karachanak, S. K., et al. (2013). Whole-genome-wide association study in the Bulgarian population reveals HHAT as schizophrenia susceptibility gene. *Psychiatr. Genet.* 23, 11–19. doi: 10.1097/ypg.0b013e3283586343

Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium (2018). Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell* 173, 1705–1715.

Birney, E., and Soranzo, N. (2015). The end of the start for population sequencing. *Nature* 526, 52–53. doi: 10.1038/526052a

Børglum, A. D., Demontis, D., Grove, J., Pallesen, J., Hollegaard, M. V., Pedersen, C. B., et al. (2014). Genome-wide study of association and interaction with maternal cytomegalovirus infection suggests new schizophrenia loci. *Mol. Psychiatry* 19, 325–333. doi: 10.1038/mp.2013.2

Camargo, L. M., Collura, V., Rain, J.-C., Mizuguchi, K., Hermjakob, H., Kerrien, S., et al. (2007). Disrupted in Schizophrenia 1 interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia. *Mol. Psychiatry* 12, 74–86. doi: 10.1038/sj.mp.4001880

Carter, H., Douville, C., Stenson, P. D., Cooper, D. N., and Karchin, R. (2013). Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genomics* 14(Suppl. 3):S3. doi: 10.1186/1471-2164-14-S3-S3

Chen, H.-J., Lin, C.-M., Lin, C.-S., Perez-Olle, R., Leung, C. L., and Liem, R. K. H. (2006). The role of microtubule actin in the Wnt (MACF1) signaling pathway. *Genes Dev.* 20, 1933–1945. doi: 10.1101/gad.1411206

Chen, S. D., Yang, J. L., Hwang, W. C., and Yang, D. I. (2018). Emerging roles of sonic hedgehog in adult neurological diseases: neurogenesis and beyond. *Int. J. Mol. Sci.* 19:2423. doi: 10.3390/ijms19082423

Chun, S., and Fay, J. C. (2009). Identification of deleterious mutations within three human genomes. *Genome Res.* 19, 1553–1561. doi: 10.1101/gr.092619.109

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6, 80–92. doi: 10.4161/fly.19695

Costas, J., Suárez-Rama, J. J., Carrera, N., Paz, E., Páramo, M., Agra, S., et al. (2013). Role of DISC1 interacting proteins in schizophrenia risk from genome-wide analysis of missense SNPs. *Ann. Hum. Genet.* 77, 504–512. doi: 10.1111/ahg.12037

Davydov, E. V., Goode, D. L., Sirota, M., Cooper, S. M., Sidow, A., and Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.* 6:e1001025. doi: 10.1371/journal.pcbi.1001025

Depristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–501.

Dobyns, W. B., Aldinger, K. A., Ishak, G. E., Mirzaa, G. M., Timms, A. E., Grout, M. E., et al. (2018). MACF1 mutations encoding highly conserved zinc-binding residues of the GAR domain cause defects in neuronal migration and axon guidance. *Am. J. Hum. Genet.* 103, 1009–1021. doi: 10.1016/j.ajhg.2018.10.019

Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., et al. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* 24, 2125–2137. doi: 10.1093/hmg/ddu733

Felsenstein, J., and Churchill, G. A. (1996). A hidden markov model approach to variation among sites in rate of evolution. *Mol. Biol. Evol.* 13, 93–104. doi: 10.1093/oxfordjournals.molbev.a025575

Firth, H. V., Richards, S. M., Bevan, A. P., Clayton, S., Corpas, M., Rajan, D., et al. (2009). DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. *Am. J. Hum. Genet.* 84, 524–533. doi: 10.1016/j.ajhg.2009.03.010

Garber, M., Guttman, M., Clamp, M., Zody, M. C., Friedman, N., and Xie, X. (2009). Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics* 25, 54–62.

GBD 2016 Disease and Injury Incidence and Prevalence Collaborators (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the global burden of disease study 2016. *Lancet* 390, 1211–1259.

Genovese, G., Fromer, M., Stahl, E. A., Ruderfer, D. M., Chambert, K., Landén, M., et al. (2016). Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* 19, 1433–1441. doi: 10.1038/nn.4402

Goryunov, D., He, C. Z., Lin, C. S., Leung, C. L., and Liem, R. K. H. (2010). Nervous-tissue-specific elimination of microtubule-actin crosslinking factor 1a results in multiple developmental defects in the mouse brain. *Mol. Cell. Neurosci.* 44, 1–14. doi: 10.1016/j.mcn.2010.01.010

Greenbaum, L., Alkelai, A., Rigbi, A., Kohn, Y., and Lerer, B. (2010). Evidence for association of the GLI2 gene with tardive dyskinesia in patients with chronic schizophrenia. *Mov. Disord.* 25, 2809–2817. doi: 10.1002/mds.23377

Gridina, M. M., Matveeva, N. M., Fishman, V. S., Menzorov, A. G., Kizilova, H. A., Beregovoy, N. A., et al. (2018). Allele-specific biased expression of the CNTN6 gene in iPS cell-derived neurons from a patient with intellectual disability and 3p26.3 microduplication involving the CNTN6 gene. *Mol. Neurobiol.* 55, 6533–6546. doi: 10.1007/s12035-017-0851-5

Grozeva, D., Conrad, D. F., Barnes, C. P., Hurles, M., Owen, M. J., O'Donovan, M. C., et al. (2012). Independent estimation of the frequency of rare CNVs in the UK population confirms their role in schizophrenia. *Schizophr. Res.* 135, 1–7. doi: 10.1016/j.schres.2011.11.004

Han, M. R., Han, K. M., Kim, A., Kang, W., Kang, Y., Kang, J., et al. (2019). Whole-exome sequencing identifies variants associated with structural MRI markers in patients with bipolar disorders. *J. Affect. Disord.* 249, 159–168. doi: 10.1016/j.jad.2019.02.028

Hilker, R., Helenius, D., Fagerlund, B., Skytthe, A., Christensen, K., Werge, T. M., et al. (2018). Heritability of schizophrenia and schizophrenia spectrum based on the nationwide Danish twin register. *Biol. Psychiatry* 83, 492–498. doi: 10.1016/j.biopsych.2017.08.017

Hu, J., Liao, J., Sathanoori, M., Kochmar, S., Sebastian, J., Yatsenko, S. A., et al. (2015). CNTN6 copy number variations in 14 patients: a possible candidate gene for neurodevelopmental and neuropsychiatric disorders. *J. Neurodev. Disord.* 7:26.

Hu, L., Su, P., Li, R., Yin, C., Zhang, Y., Shang, P., et al. (2016). Isoforms, structures, and functions of versatile spectraplakin MACF1. *BMB Rep.* 49, 37–44. doi: 10.5483/bmbrep.2016.49.1.185

Huang, A. Y., Yu, D., Davis, L. K., Sul, J. H., Tsetsos, F., Ramensky, V., et al. (2017). Rare copy number variants in NRXN1 and CNTN6 increase risk for Tourette syndrome. *Neuron* 94, 1101–1111.e7.

Juan-Perez, C., Farrand, S., and Velakoulis, D. (2018). Schizophrenia and epilepsy as a result of maternally inherited CNTN6 copy number variant. *Schizophr. Res.* 202, 111–112. doi: 10.1016/j.schres.2018.06.062

Ka, M., and Kim, W.-Y. (2016). Microtubule-actin crosslinking factor 1 is required for dendritic arborization and axon outgrowth in the developing brain. *Mol. Neurobiol.* 53, 6018–6032. doi: 10.1007/s12035-015-9508-4

Ka, M., Jung, E. M., Mueller, U., and Kim, W. Y. (2014). MACF1 regulates the migration of pyramidal neurons via microtubule dynamics and GSK-3 signaling. *Dev. Biol.* 395, 4–18. doi: 10.1016/j.ydbio.2014.09.009

Ka, M., Moffat, J. J., and Kim, W. Y. (2017). MACF1 controls migration and positioning of cortical GABAergic interneurons in mice. *Cereb. Cortex* 27, 5525–5538.

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., et al. (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* [Preprint] doi: 10.1101/531210

Kashevarova, A. A., Nazarenko, L. P., Schultz-Pedersen, S., Skryabin, N. A., Salyukova, O. A., Chechetkina, N. N., et al. (2014). Single gene microdeletions and microduplication of 3p26.3 in three unrelated families: CNTN6 as a new candidate gene for intellectual disability. *Mol. Cytogenet.* 7:97.

Kataoka, M., Matoba, N., Sawada, T., Kazuno, A. A., Ishiwata, M., Fujii, K., et al. (2016). Exome sequencing for bipolar disorder points to roles of de novo loss-of-function and protein-altering mutations. *Mol. Psychiatry* 21, 885–893. doi: 10.1038/mp.2016.69

Kearney, J. A. (2011). Genetic modifiers of neurological disease. *Curr. Opin. Genet. Dev.* 21, 349–353. doi: 10.1016/j.gde.2010.12.007

Kenny, E. M., Cormican, P., Furlong, S., Heron, E., Kenny, G., Fahey, C., et al. (2014). Excess of rare novel loss-of-function variants in synaptic genes in schizophrenia and autism spectrum disorders. *Mol. Psychiatry* 19, 872–879. doi: 10.1038/mp.2013.127

Kerr, T. M., Muller, C. L., Miah, M., Jetter, C. S., Pfeiffer, R., Shah, C., et al. (2013). Genetic background modulates phenotypes of serotonin transporter Ala56 knock-in mice. *Mol. Autism* 4, 1–11.

Koike, H., Arguello, P. A., Kvajo, M., Karayiorgou, M., and Gogos, J. A. (2006). Disc1 is mutated in the 129S6/SvEv strain and modulates working memory in mice. *Proc. Natl. Acad. Sci. U.S.A.* 103, 3693–3697. doi: 10.1073/pnas.0511189103

Kopanos, C., Tsiolkas, V., Kouris, A., Chapple, C. E., Aguilera, M. A., Meyer, R., et al. (2019). VarSome: the human genomic variant search engine. *Bioinformatics* 35, 1978–1980. doi: 10.1093/bioinformatics/bty897

Kumar, P., Henikoff, S., and Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4, 1073–1082. doi: 10.1038/nprot.2009.86

Lee, S. H., DeCandia, T. R., Ripke, S., Yang, J., Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC-SCZ), International Schizophrenia Consortium (ISC), et al. (2012). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat. Genet.* 44, 247–250. doi: 10.1038/ng.1108

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291.

Lewis, D. A., Hashimoto, T., and Volk, D. W. (2005). Cortical inhibitory neurons and schizophrenia. *Nat. Rev. Neurosci.* 6, 312–324. doi: 10.1038/nrn1648

Lichtenstein, P., Yip, B. H., Björk, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F., et al. (2009). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* 373, 234–239. doi: 10.1016/s0140-6736(09)60072-6

MacDonald, J. R., Ziman, R., Yuen, R. K. C., Feuk, L., and Scherer, S. W. (2014). The database of genomic variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.* 42, 986–992.

Malhotra, D., and Sebat, J. (2012). CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* 148, 1223–1241. doi: 10.1016/j.cell.2012.02.039

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., et al. (2009). Finding the missing heritability of complex diseases. *Nature* 461, 747–753.

Mercati, O., Huguet, G., Danckaert, A., André-Leroux, G., Maruani, A., Bellinzoni, M., et al. (2017). CNTN6 mutations are risk factors for abnormal auditory sensory perception in autism spectrum disorders. *Mol. Psychiatry* 22, 625–633. doi: 10.1038/mp.2016.61

Mitchell, K. J. (2011). The genetics of neurodevelopmental disease. *Curr. Opin. Neurobiol.* 21, 197–203.

Moffat, J. J., Ka, M., Jung, E.-M., Smith, A. L., and Kim, W.-Y. (2017). The role of MACF1 in nervous system development and maintenance. *Semin. Cell Dev. Biol.* 69, 9–17. doi: 10.1016/j.semcdb.2017.05.020

Moreno-Küstner, B., Martín, C., and Pastor, L. (2018). Prevalence of psychotic disorders and its association with methodological issues. A systematic review and meta-analyses. *PLoS One* 13:e0195687. doi: 10.1371/journal.pone.0195687

Mucci, A., Galderisi, S., Green, M. F., Nuechterlein, K., Rucci, P., Gibertoni, D., et al. (2017). Familial aggregation of MATRICS consensus cognitive battery scores in a large sample of outpatients with schizophrenia and their unaffected relatives. *Psychol. Med.* 48, 1359–1366. doi: 10.1017/S0033291717002902

Muraki, K., and Tanigaki, K. (2015). Neuronal migration abnormalities and its possible implications for schizophrenia. *Front. Neurosci.* 9:74. doi: 10.3389/fnins.2015.00074

Nadeau, J. H. (2001). Modifier genes in mice and humans. *Nat. Rev. Genet.* 2, 165–174. doi: 10.1038/35056009

Neale, B. M., Lasky-Su, J., Anney, R., Franke, B., Zhou, K., Maller, J. B., et al. (2008). Genome-wide association scan of attention deficit hyperactivity disorder. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* 147, 1337–1344.

Otsuka, I., Watanabe, Y., Hishimoto, A., Boku, S., Mouri, K., Shiroiwa, K., et al. (2015). Association analysis of the Cadherin13 gene with schizophrenia in the Japanese population. *Neuropsychiatr. Dis. Treat.* 11, 1381–1393. doi: 10.2147/ndt.s84736

Peiffer, D. A., Le, J. M., Steemers, F. J., Chang, W., Jenniges, T., Garcia, F., et al. (2006). High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res.* 16, 1136–1148. doi: 10.1101/gr.5402306

Petrovski, S., Wang, Q., Heinzen, E. L., Allen, A. S., and Goldstein, D. B. (2013). Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet.* 9:e1003709. doi: 10.1371/journal.pgen.1003709

Piñero, J., Ramírez-Anguita, J. M., Saüch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., et al. (2020). The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* 48, D845–D855.

Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* 20, 110–121. doi: 10.1101/gr.097857.109

Prata, D. P., Costa-Neves, B., Cosme, G., and Vassos, E. (2019). Unravelling the genetic basis of schizophrenia and bipolar disorder with GWAS: a systematic review. *J. Psychiatr. Res.* 114, 178–207. doi: 10.1016/j.jpsychires.2019.04.007

Purcell, S. M., Moran, J. L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., et al. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 506, 185–190. doi: 10.1038/nature12975

Rees, E., Walters, J. T. R., Georgieva, L., Isles, A. R., Chambert, K. D., Richards, A. L., et al. (2014). Analysis of copy number variations at 15 schizophrenia-associated loci. *Br. J. Psychiatry* 204, 108–114. doi: 10.1192/bjp.bp.113.131052

Reva, B., Antipin, Y., and Sander, C. (2007). Determinants of protein function revealed by combinatorial entropy optimization. *Genome Biol.* 8:R232.

Rivero, O., Sich, S., Popp, S., Schmitt, A., Franke, B., and Lesch, K.-P. (2013). Impact of the ADHD-susceptibility gene CDH13 on development and function of brain networks. *Eur. Neuropsychopharmacol.* 23, 492–507. doi: 10.1016/j.euroneuro.2012.06.009

Rossignol, E. (2011). Genetics and function of neocortical GABAergic interneurons in neurodevelopmental disorders. *Neural Plast.* 2011:649325.

Sakurai, K., Toyoshima, M., Ueda, H., Matsubara, K., Takeda, Y., Karagogeos, D., et al. (2009). Contribution of the neural cell recognition molecule NB-3 to synapse formation between parallel fibers and Purkinje cells in mouse. *Dev. Neurobiol.* 69, 811–824. doi: 10.1002/dneu.20742

Sanchez-Soriano, N., Travis, M., Dajas-Bailador, F., Gonçalves-Pimentel, C., Whitmarsh, A. J., and Prokop, A. (2009). Mouse ACF7 and *Drosophila* short stop modulate filopodia formation and microtubule organisation during neuronal growth. *J. Cell Sci.* 122, 2534–2542. doi: 10.1242/jcs.046268

Sanders, S. J., Ercan-Sencicek, A. G., Hus, V., Luo, R., Murtha, R. T., Moreno-De-Luca, D., et al. (2011). Multiple recurrent De Novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 70, 863–885.

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421–427. doi: 10.1038/nature13595

Schwarz, J. M., Rödelsperger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* 7, 575–576. doi: 10.1038/nmeth0810-575

Sethi, M. K., and Zaia, J. (2017). Extracellular matrix proteomics in schizophrenia and Alzheimer's disease. *Anal. Bioanal. Chem.* 409, 379–394. doi: 10.1007/s00216-016-9900-6

Shihab, H. A., Gough, J., Cooper, D. N., Stenson, P. D., Barker, G. L. A., Edwards, K. J., et al. (2013). Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using Hidden Markov models. *Hum. Mutat.* 34, 57–65. doi: 10.1002/humu.22225

Singh, T., Kurki, M. I., Curtis, D., Purcell, S. M., Crooks, L., McRae, J., et al. (2016). Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat. Neurosci.* 19, 571–577.

Stefansson, H., Meyer-Lindenberg, A., Steinberg, S., Magnusdottir, B., Morgen, K., Arnarsdottir, S., et al. (2014). CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature* 505, 361–366. doi: 10.1038/nature12818

Steinberg, S., Gudmundsdottir, S., Sveinbjornsson, G., Suvisaari, J., Paunio, T., Torniainen-Holm, M., et al. (2017). Truncating mutations in RBM12 are associated with psychosis. *Nat. Genet.* 49, 1251–1254. doi: 10.1038/ng.3894

Stelzer, G., Plaschkes, I., Oz-Levi, D., Alkelai, A., Olender, T., Zimmerman, S., et al. (2016). VarElect: the phenotype-based variation prioritizer of the GeneCards suite. *BMC Genomics* 17:444. doi: 10.1186/s12864-016-2722-2

Szatkiewicz, J. P., Fromer, M., Nonneman, R. J., Ancalade, N., Johnson, J. S., Stahl, E. A., et al. (2020). Characterization of single gene copy number variants in schizophrenia. *Biol. Psychiatry* 87, 736–744.

Te Weehi, L., Maikoo, R., Mc Cormack, A., Mazzaschi, R., Ashton, F., Zhang, L., et al. (2014). Microduplication of 3p26.3 implicated in cognitive development. *Case Rep. Genet.* 2014:295359.

Terracciano, A., Tanaka, T., Sutin, A. R., Sanna, S., Deiana, B., Lai, S., et al. (2010). Genome-wide association scan of trait depression. *Biol. Psychiatry* 68, 811–817. doi: 10.1016/j.biopsych.2010.06.030

The International Schizophrenia Consortium (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455, 237–241. doi: 10.1038/nature07239

Thorvaldsdottir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178–192. doi: 10.1093/bib/bbs017

Wang, Q., Li, M., Yang, Z., Hu, X., Wu, H. M., Ni, P., et al. (2015). Increased co-expression of genes harboring the damaging de novo mutations in Chinese schizophrenic patients during prenatal development. *Sci. Rep.* 5:18209.

Xu, B., Ionita-Laza, I., Roos, J. L., Boone, B., Woodrick, S., Sun, Y., et al. (2012). De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat. Genet.* 44, 1365–1369. doi: 10.1038/ng.2446

Xu, B., Roos, J. L., Dexheimer, P., Boone, B., Plummer, B., Levy, S., et al. (2011). Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat. Genet.* 43, 864–868. doi: 10.1038/ng.902

Xu, W., Cohen-Woods, S., Chen, Q., Noor, A., Knight, J., Hosang, G., et al. (2014). Genome-wide association study of bipolar disorder in Canadian and UK populations corroborates disease loci including SYNE1 and CSMD1. *BMC Med. Genet.* 15:2. doi: 10.1186/1471-2350-15-2

Ye, H., Tan, Y. L. J., Ponniah, S., Takeda, Y., Wang, S.-Q., Schachner, M., et al. (2008). Neural recognition molecules CHL1 and NB-3 regulate apical dendrite orientation in the neocortex via PTPα. *EMBO J.* 27, 188–200. doi: 10.1038/sj.emboj.7601939

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychiatry

ORIGINAL RESEARCH
published: 08 April 2021
doi: 10.3389/fpsyt.2021.581093

Gene Expression Differences Between Young Adults Based on Trauma History and Post-traumatic Stress Disorder

Kaitlin E. Bountress[1†], Vladimir Vladimirov[2,3†], Gowon McMichael[1], Z. Nathan Taylor[1], Gary Hardiman[4], Dongjun Chung[5], Zachary W. Adams[6], Carla Kmett Danielson[7,8‡] and Ananda B. Amstadter[1‡]*

[1] Virginia Institute for Psychiatry and Behavioral Genetics, Virginia Commonwealth University (VCU), Richmond, VA, United States, [2] Department of Psychiatry and Behavioral Sciences, College of Medicine Texas A&M University, Richmond, VA, United States, [3] Lieber Institute for Brain Development, Johns Hopkins University, Baltimore, MD, United States, [4] Institute for Global Food Security, Queens University Belfast, Belfast, United Kingdom, [5] Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, United States, [6] Department of Psychiatry, Indiana University of Medicine, Indianapolis, IN, United States, [7] National Crime Victim Research and Treatment Center, Medical University of South Carolina, Charleston, SC, United States, [8] Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina, Charleston, SC, United States

OPEN ACCESS

Edited by:
Noèlia Fernàndez-Castillo, Centre for Biomedical Network Research (CIBER), Spain

Reviewed by:
Murat Altinay, Cleveland Clinic, United States
Matthew Girgenti, Yale University, United States

***Correspondence:**
Kaitlin E. Bountress
kaitlin.bountress@vcuhealth.org

†These authors share first authorship

‡These authors share last authorship

Specialty section:
This article was submitted to Mood and Anxiety Disorders, a section of the journal Frontiers in Psychiatry

Received: 07 July 2020
Accepted: 15 March 2021
Published: 08 April 2021

Citation:
Bountress KE, Vladimirov V, McMichael G, Taylor ZN, Hardiman G, Chung D, Adams ZW, Danielson CK and Amstadter AB (2021) Gene Expression Differences Between Young Adults Based on Trauma History and Post-traumatic Stress Disorder. Front. Psychiatry 12:581093. doi: 10.3389/fpsyt.2021.581093

Background: The purpose of this study was to identify gene expression differences associated with post-traumatic stress disorder (PTSD) and trauma exposure (TE) in a three-group study design comprised of those with and without trauma exposure and PTSD.

Methods: We conducted gene expression and gene network analyses in a sample ($n = 45$) composed of female subjects of European Ancestry (EA) with PTSD, TE without PTSD, and controls.

Results: We identified 283 genes differentially expressed between PTSD-TE groups. In an independent sample of Veterans ($n = 78$) a small minority of these genes were also differentially expressed. We identified 7 gene network modules significantly associated with PTSD and TE (Bonferroni corrected $p \leq 0.05$), which at a false discovery rate (FDR) of $q \leq 0.2$, were significantly enriched for biological pathways involved in focal adhesion, neuroactive ligand receptor interaction, and immune related processes among others.

Conclusions: This study uses gene network analyses to identify significant gene modules associated with PTSD, TE, and controls. On an individual gene level, we identified a large number of differentially expressed genes between PTSD-TE groups, a minority of which were also differentially expressed in the independent sample. We also demonstrate a lack of network module preservation between PTSD and TE, suggesting that the molecular signature of PTSD and trauma are likely independent of each other. Our results provide a basis for the identification of likely disease pathways and biomarkers involved in the etiology of PTSD.

Keywords: gene expression, gene network analyses, trauma exposure, post-traumatic stress disorder, civilian trauma, veterans

INTRODUCTION

Between 50 and 70% of individuals experience at least one trauma in their lifetimes (1). Of those individuals, 8–32% develop post-traumatic stress disorder [PTSD; (2, 3)]. PTSD is a serious disorder associated with medical ailments, suicide, and early mortality, but major gaps remain in our understanding of its etiology (4–6). PTSD is heritable, with estimates ranging between 24 and 72% (7–9). Large-scale genome wide association studies (GWAS) are revealing key influences on PTSD.

Recent GWAS of PTSD, PTSD total symptoms, and/or its symptom clusters (e.g., re-experiencing symptoms) (10–12) have revealed a few hits that differentiate PTSD cases and controls, and more when examining PTSD symptoms as a quantitative trait and/or when examining continuous symptoms within clusters. One of the most consistent findings is that genes implicated in cell cycle control and other mental health conditions such as schizophrenia and bipolar (*MAL1L1*) are associated with PTSD phenotypes. The effects found in both samples involve genes implicated in intracellular protein transport associated with risk-taking (*TSNARE1*) as well as nucleic acid binding which is important for cognitive abilities (*EXD3*). Additionally, genes important for steroid signaling, hormone metabolism, and stress response (CRHR1, HSD17B11), central nervous system development (TCF4), and transcriptional activity and enhancer functions (i.e., ZDHHC14, PARK2, chr13: 55,652,129–55,759,209; 11) were also found to be important.A review of PTSD GWAS summarizes key findings, with several significant SNPs differentiating cases and controls, coming from genes in systems important for regulating circadian rhythm, synaptic processes, immune function, and neuroplasticity [e.g., LINC01090, BC036345, ZNRD1-AS1, RORA, NLGN1, TLL1; (13)]. Despite these advances, questions remain about how genetic factors relate to PTSD and whether observed patterns are specific to PTSD or are related to trauma exposure broadly. Because PTSD is more prevalent in women than men and genetic influences on PTSD are stronger among women than men (7, 14), studies are needed to characterize the genetic correlates of PTSD and trauma exposure (TE) in women.

Gene Expression Studies

Studying gene expression is essential for our understanding of the PTSD pathophysiology (15, 16). While trauma exposure precedes all cases of PTSD, not all people with TE will develop PTSD. Therefore, expression studies are needed that include participants with or without TE and with or without PTSD to disentangle whether observed genetic differences are influenced by TE or PTSD. Previous post-mortem expression studies have reported expression differences between subjects with PTSD and controls in the dorsolateral pre-frontal cortex, which regulates fear-based responses (17). Research examining the post-mortem brains of people who died of suicide with and without child abuse suggests differences in the expression of genes important for myelination (18). A study of transcriptome-wide analyses of gene expression changes in post-mortem brains of individuals with PTSD identified genes involved in synapse, neuron and axon terms, as well as glia formation, actin binding, and small GTPase signaling (19). This same study found that the global transcriptomic signatures for PTSD closely resemble schizophrenia, autism, and bipolar disorder, consistent with GWAS of PTSD (14, 20). Thus, there is at least some convergence in the pathology of PTSD using blood and brain tissues. Indeed, as the collection of post-mortem brain tissue involves ethical, logistic, and technical confounds (21, 22), blood is increasingly viewed as a viable and valid proxy tissue for gene expression data. Furthermore, several studies have pointed to a reasonable correlation of the total gene expression between different tissue, notably brain and blood (23–25). Finally, studying blood related gene expression is also advantageous as some processes underlying PTSD are reflected in peripheral blood gene expression [e.g., glucocorticoid sensitivity, (26)]. For all of these reasons, gene expression data extracted from whole blood is becoming increasingly commonplace.

Comparing gene expression profiles between subjects with PTSD and TE have suggested genes involved in the hypothalamic-pituitary-adrenal Axis (HPA) [e.g., *FKBP5*; (27–30)] and signal transduction processes [e.g., *STAT5B*; (16, 28)]. Most studies included individuals in the TE control group if individuals did not meet criteria for *current* PTSD, which is problematic, since differences may be obscured between those who previously were diagnosed with PTSD vs. those never diagnosed. A study of subjects with PTSD to those without TE found evidence for gene expression differences that are important for activating the adaptive cellular immune response (e.g., *IL-12* and *IL-18*) (31), but lack of a non-PTSD TE group limits conclusions about molecular genetic risk.

Gene Network Analysis

A limitation of studies utilizing single gene expression analyses is the focus on ranking a list of individual differentially expressed genes with the highest statistical significance. This selection approach is somewhat arbitrary, does not consider the potential interaction between genes in the dataset, and does not provide a broader, system-level view of expression. Thus, a network approach can be used to better understand the functional changes between the disease and normal transcriptomes. One study (32) using this approach identified over-expression of genes enriched for innate-immune response and interferon signaling (PTSD-TE), while another (33) found differences for genes enriched for signal transduction (PTSD-control).

Current Study

The few previously reported gene expression and gene network analyses are limited by the lack of use of a three-group design that differentiates between TE and PTSD. Thus, the first aim of this study was to extend prior expression studies of PTSD, using both TE and non-TE exposed controls. In so doing, we are able to understand what expression signatures are associated with trauma exposure (as seen in differences between TE and non-TE groups) and PTSD (as seen in differences between PTSD and TE groups). The second aim was to provide a system view of the expression signatures using weighted gene co-expression network analysis (WGCNA). While WGCNA identifies sets of genes showing related expression patterns and therefore having likely shared disease functions

(34–36), it does not provide information on the biological functions and process the significant modules are enriched for. Thus, all significantly correlated modules were assessed for the enrichment of cellular processes and biological functional categories using gene set enrichment analysis [GSEA; (37)]. These analyses allow us to better understand co-expression networks and potential biological functions that may differentiate those with and without trauma exposure, as well as-among those with trauma exposure-those with and without PTSD. Genome-wide expression data were generated using RNA isolated from Peripheral Blood Mononuclear Cell (PBMC) obtained from an ethnically homogeneous sample of female subjects, known to be at higher genetic risk for PTSD.

METHODS AND MATERIALS

Overview of Larger, Parent Study

Participants were sampled from a larger community sample of young adults. All subjects gave their informed consent before being included in the study. The study was conducted in accordance with the Declaration of Helsinki, and the research protocol was approved by the Institutional Review Board (IRB) at Medical University of South Carolina. These 281 young adults were 59.2% female; primarily Caucasian (84.7% Caucasian, 6.1% African-American, 9.2% Other), and between the ages of 21 and 30 ($M = 24.76$, $SD = 2.59$). Participants were recruited to one of three study groups: non-trauma exposed control (controls), trauma exposed without PTSD (TE), or trauma exposed with PTSD. See prior work for information for eligibility details (38). Of those, 72 controls, 72 TE, and 53 PTSD participants ($n = 197$) provided blood samples for genomic analyses.

Current Study Sub-sample Participants

For this study, 15 European Ancestry females from each group ($n = 45$) were included. PTSD group participants with the most PTSD symptoms were prioritized for inclusion. PTSD and TE groups were matched on trauma characteristics (i.e., number of traumas, time since traumas) and age. Control group participants were age-matched to the PTSD and TE groups.

There were no differences between groups on age ($F = 0.27$, NS). There were differences in number of experienced traumas between control ($M = 0.07$) and TE/PTSD groups, but not between TE and PTSD groups ($M = 0.93$, $M = 1.73$, respectively; $F = 11.90$, $p < 0.001$). The PTSD group ($M = 40.79$) reported more PTSD symptoms than TE group (19.29; $t = 15.842$, $p < 0.001$). The PTSD group ($M = 16.53$) reported more depressive symptoms than the TE and control groups ($M = 3.87$, $M = 1.80$, respectively; $F = 954.47$, $p < 0.001$).

Measures

Number of Traumatic Events

Using the Life Events Checklist (39), count scores for total endorsed witnessed or experienced traumas was used as a covariate in the differential expression analyses (see **Table 1**).

PTSD

Using the PTSD Checklist [PCL; (40)], individuals with scores of 30 or higher (41) and who met the minimum symptoms per cluster were given a diagnosis of probable PTSD (Cronbach's Alpha: 0.91). The Mini-International Neuropsychiatric Interview (MINI) PTSD scale was used to confirm PTSD diagnosis (42).

Depression Symptoms

a sum of 20 items assessing depressive symptoms using the Beck Depression Inventory [BDI; (43)] was entered as a covariate and additional phenotypic measure in the gene expression (Cronbach's Alpha: 0.92) and network analyses, respectively.

RNA Isolation

Total RNA containing the small RNA fraction was isolated from 9 ml of whole blood using the mirVana-PARIS kit (Life Technologies, Carlsbad, CA), following manufacturer's protocols. RNA concentration was measured using the Quant-iT Broad Range RNA Assay kit (Life Technologies), and the RNA Integrity Number (RIN) was measured on the Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA). All RNA samples had an excellent RIN scores (average RIN $\geq$ 9.0, s.d. $\pm$ 0.5).

mRNA Expression Microarrays

Genome-wide expression assays were ran using the Affymetrix (Santa Clara, CA) GeneChip® Human Genome U133A 2.0 (HG-U133A 2.0) following the Affymetrix® protocol (44). This array provides comprehensive coverage of the transcribed human genome using 22,214 probesets, and captures the expression of ≈18,400 human transcripts. Array quality was assessed by estimating the $3'/5'$ ratios of GAPDH, and the percentage of "Present" genes (%P) and array exhibiting GAPDH $3'/5' < 3.0$ and $\%P > 40\%$ were considered of good quality. Based on these metrics no arrays were excluded.

Microarray Normalization

Expression values were calculated following the pre-processing procedures: (1) Robust Multiarray Average adjusting for probe sequence (GCRMA) background correction, (2) $\log_2$ transformation, (3) quantile normalization, and (4) median-polish probeset summarization using **P**artek **G**enomics **S**uite v6.23 (PGS; Partek Inc., St. Louis, MO) (45, 46). The batch effect removal option in PGS was used to control for batch effects. Microarray quality was assessed by principal component analysis (PCA) and unsupervised hierarchical clustering to identify potential outliers. One sample was identified as outlier and removed from future analyses (**Supplementary Figure 1**).

Microarray reliability was assessed, by validating the expression levels of four genes, (DUS2, FRMD4B, CXCR6, and SRRT) via quantitative real-time PCR using a Taqman approach. We observed a high correlation between the two platforms, i.e., the Pearson correlation ranged from $r = 0.664$ (DUS2) to $r = 0.846$ (SRRT); see **Supplementary Table 1** and **Supplementary Figure 2**.

Adjusting for Cell Heterogeneity

We used the ImmQuant software (47), which takes as input transcription profiles, and provides the predicted quantities of various cell types within each sample. The deconvolution approach builds on the implementation of the digital cell quantification (DCQ) deconvolution algorithm (48), which is used for immune cell-type prediction in human samples, resulting in gene expression markers for 39 cell types being identified and accounted for to minimize the effect of cell based differential gene expression (**Supplementary Figure 3**).

Data Analytic Plan

Aim 1: Gene Expression Analyses

Detection of differentially expressed genes between groups was performed in the Number Cruncher Statistical Software (NCSS) v11, using a robust multiple regression model (49), with gene expression as the dependent variable and group as the explanatory variable. To reduce the number of potential covariates and account for potential collinearity between them, the 39 cell-based estimates together with RIN and Batch, were incorporated into a principle component analysis (PCA). The first 6 principle components (PCs) accounted for 13% of the variance in the data and were used as covariates, along with number of interpersonal traumatic events (50) and depressive symptoms (51). Our intention of conducting a PCA here was to reduce the number of covariates, increase our power, and use the higher order orthogonal relationships between the covariates to adjust for their effect on gene expression.

Aim 2: Network Analyses

The gene network analyses were conducted using WGCNA. The method for constructing scale-free networks by WGCNA has been described in previous studies (52, 53). The gene co-expression networks were constructed by using the WGCNA v1.36 package in R environment (v3.3.2) using a minimum module size of 35 genes and a minimum module merge height of 0.8. To build the networks, we used all transcripts differentially expressed between the three groups at $p \leq 0.05$ (i.e., for the PTSD/TE, PTSD/C, and TE/C comparisons a total of 1,701, 441, and 1,846 genes were included, respectively). This significance threshold was chosen to allow for: (**i**) the inclusion of true positive signals with a smaller effect size (which would otherwise be excluded from the more stringent statistical criteria applied in gene expression analyses), (**ii**) retain a sufficient number of genes with biological importance in PTSD and TE for the building of gene co-expression networks, and (**iii**) exclude genes with low variance, i.e., genes with limited importance for the disease trait (49, 54).

Following module definition, the module eigengene (ME) representing the first principle component (PC) of all summarized gene expression profiles within a given module was calculated. The ME was then correlated with the three diagnostic groups, PTSD, TE, and controls, and statistical significance assessed at Bonferroni-adjusted $p \leq 0.05$ between the groups (i.e., corrected for a number of modules tested). To account for biological or technical confounds, we fitted the regression model of each gene expression value on covariates and then used the residuals from the model as an input for the network analysis.

The rationale for conducting the network analyses within each diagnostic group separately is based on our ability to understand better group differences between subjects with PTSD, TE not reaching the threshold for PTSD, and controls. Our rationale for this approach was further indicated by our desire to test for network preservation between the three phenotypic groups, i.e., identify the precise network modules that differentiate between the PTSD, TE and controls. Thus, with the current analyses, we have increased confidence inferring the gene expression and gene network differences between groups are associated with trauma exposure (TE-C), PTSD (PTSD-TE), or potentially both (PTSD-C). Below are the results of these three comparisons.

Pathway Analyses

To identify known biological processes and pathways that our significant modules were enriched for, the gene set enrichment analysis [GSEA; (GSEAv2.0.14)] was used (37, 55). Individual gene lists for each of the gene modules significantly correlated with PTSD and TE status were generated by rank-ordering all differentially expressed genes by their module membership (MM) to each of the trait-associated modules. In GSEA, Affymetrix HG-U133A 2.0 probe IDs were converted to HUGO Gene Nomenclature Committee (HGNC) gene symbols, and in cases of multiple transcripts representing a single gene, the probeset with the highest MM was retained (56). A priori gene sets were obtained from the Molecular Signatures Database v4.0 from the Broad Institute. A total of 1320 gene sets from the Canonical KEGG Pathways subset of the C2: Curated Pathways collection of MSigDB were assessed. Default parameters were then applied to restrict the *a priori* gene set size between 15 and 500 genes, respectively. A total of 186 a priori gene sets were identified, of which in the PTSD/TE comparison, 165 were filtered out due to gene set size parameters (i.e., gene sets <15 and >500 genes) leading to 21 a priori gene sets tested in the final GSEA analysis. In the TE/Control comparison and the PTSD/Control comparison, 16 and 18 a priori gene sets were used (i.e. 170/186 and 168/186 being excluded).

Identification of Candidate Hub Genes

Hubs comprise the highly interconnected nodes within the gene network and have been shown to be functionally significant (57). In this study, hub genes were defined by (1) the strength of their intramodular connectivity (i.e., module membership (MM) (58, 59), which measures the strength of the Pearson correlation product moment ($r \geq 0.7$) between the individual gene expression and module eigengene (ME) and (2) their gene significant (GS) correlation with phenotype. The upper quartile of transcripts with the highest MM and GS (employed for the significant modules) was selected as candidate hub genes.

Network Preservation

We used the module preservation statistic $Z_{summary}$, outlined previously (60), to test for module preservation between the three comparisons (i.e., PTSD/TE, PTSD/C, and TE/C). One advantage $Z_{summary}$ statistics has over the more traditional cross-tabulation

approaches is that, in addition to overlap in module preservation, it also accounts for the module connectivity pattern and density. The module preservation also requires that a reference network is assigned which then is compared against the test network. The modules from the control groups (in the PTSD/C and TE/C comparisons) were used as reference modules and the modules from TE group were used as reference modules in the PTSD/TE comparisons. We followed previously outlined recommendations for network preservation, i.e., $Z_{summary} \leq 2$ implies no evidence for module preservation, $2 \leq Z_{summary} \leq 10$ implies weak to moderate evidence, and $Z_{summary} \geq 10$ implies strong evidence for module preservation. To ensure that module preservation statistics is not confounded by module size (larger module tend to be more preserved than smaller ones, solely due to their size), in addition to $Z_{summary}$ statistics to assess module preservation, we have also used the median rank statistics, that is invariant to module size.

Power

Power for the differential gene expression analysis was computed using the power analysis statistical software (PASS) version 19 (NCSS, LLC, Kaysville, Utah) assuming an FDR adjusted $p \leq 0.0005$, corresponding to an FDR of 10% in our analyses. Generally regarded as a rigorous cut-off (61, 62), in our power analysis, the values of the effect size ($R^2|C$) represents the amount by which the squared correlation coefficient (R^2) is increased when the test (T) variables are added to the regression model containing the covariates (C). The ($R^2|C$) was estimated from our gene expression analyses and had an effect sizes ranging between 0.01÷0.06.

Testing Whether Findings Hold in an Independent Sample

The gene overlap between the civilian and the unique Veteran samples (n=78) was assessed via the hypergeometric test. The test significance was based on the "representation factor (rf)", i.e., the number of overlapping genes divided by the expected number of overlapping genes drawn from two independent groups, with a value of ≥1 indicating a significant overlap between the gene datasets. The total number of genes was based on all unique 14,800 transcripts on the Affymetrix HG-U133_2.0 array. We limited our examination of effects across samples to all differentially expressed genes between PTSD/TE, TE/C, and PTSD/C comparisons at $p \leq 0.005$. Our rationale for choosing this p-value was based on the compromise between providing a sufficient number of genes for the analysis, while still controlling the Type I error (63).

RESULTS

Aim 1: Gene Expression Analyses

TE vs. PTSD. At a FDR of 10%, we identified 53 genes with evidence for differential expression between subjects with TE and PTSD (**Table 2**). Of these, 44 genes were upregulated, and 9 genes downregulated in the PTSD group compared to TE, respectively. The three top significant genes were the growth regulation by estrogen in breast cancer 1 (GREB1) gene (210562_at; $t = 11.19$, $p = 1.12E\text{-}08$, $q = 2.38E\text{-}04$), the abhydrolase domain containing 6 (ABHD6) gene (221552_at; $t = 8.41$, $p = 4.61E\text{-}07$, $q = 4.91E\text{-}03$), and the FERM domain containing 4B (FRMD4B) gene (216134_at; $t = 7.36$, $p = 2.38E\text{-}06$, $q = 1.69E\text{-}02$). Interestingly, compared to *PTSD vs. TE* comparison, we found a much lower number of differentially expressed genes in the *PTSD vs. Control* comparison, with only 5 differentially expressed genes at FDR of 10%. The top gene in this category was represented by the glutamate receptor, ionotropic, N-methyl D-aspartate 1 (GRIN1) gene (205915_x_at; $t = -6.04$, $p = 2.22E\text{-}06$, $q = 1.19E\text{-}02$) that was under expressed in subjects with PTSD. In the *TE vs. Cont*rol comparison at FDR of 10%, we observed only two genes: the keratin 4, type II (KRT4) gene (214399_s_at; $t = -6.64$, $p = 4.86E\text{-}07$, $q = 0.008178$) and the protocadherin gamma subfamily A, 11 (PCDHGA11) gene (211876_x_at; $t = -6.45$, $p = 7.84E\text{-}07$, $q = 0.008178$), with both genes under expressed in the TE group. The quantile-quantile (qq) plots of the p-value distribution for the gene expression between the three comparisons are shown in **Supplementary Figure 4**, and the entire list of differentially expressed genes is provided in **Supplementary Table 2**. The effect sizes for gene expression studies are generally larger than the effect sizes observed in GWAS, which may be one possible reason why the qq plots may appear inflated, without actually being so (64).

TABLE 1 | Percent of trauma-exposed (TE) and PTSD groups witnessing or experiencing different types of traumatic events.

	% TE group	% PTSD group
Natural disaster	42.9	28.6
Fire or explosion	14.3	7.1
Transportation accident	71.4	50.0
Serious accident	21.4	28.6
Exposure to toxic substance	7.1	0.0
Physical assault	35.7	64.3
Assault with a weapon	28.6	35.7
Sexual assault	21.4	50
Other unwanted or uncomfortable sexual experience	14.3	42.9
Combat or war zone exposure	7.1	0.0
Captivity (e.g., being kidnapped)	0.0	0.0
Life-threatening illness	35.7	28.6
Sudden violent death	21.4	14.3
Serious injury you caused to someone else	0.0	14.3
Other stressful event	21.4	21.4

Aim 2: Network Analysis

PTSD vs. TE Group Comparison

In our first set of network analysis, we identified 14 modules [including the *gray* module (M_{grey}), which contains genes unassigned to any of the other 13 modules (52)] between the PTSD and TE subjects.

The size of the 14 modules ranged from 312 transcripts in M_{blue} to 35 transcripts in M_{cyan} and $M_{midnightblue}$. In addition to

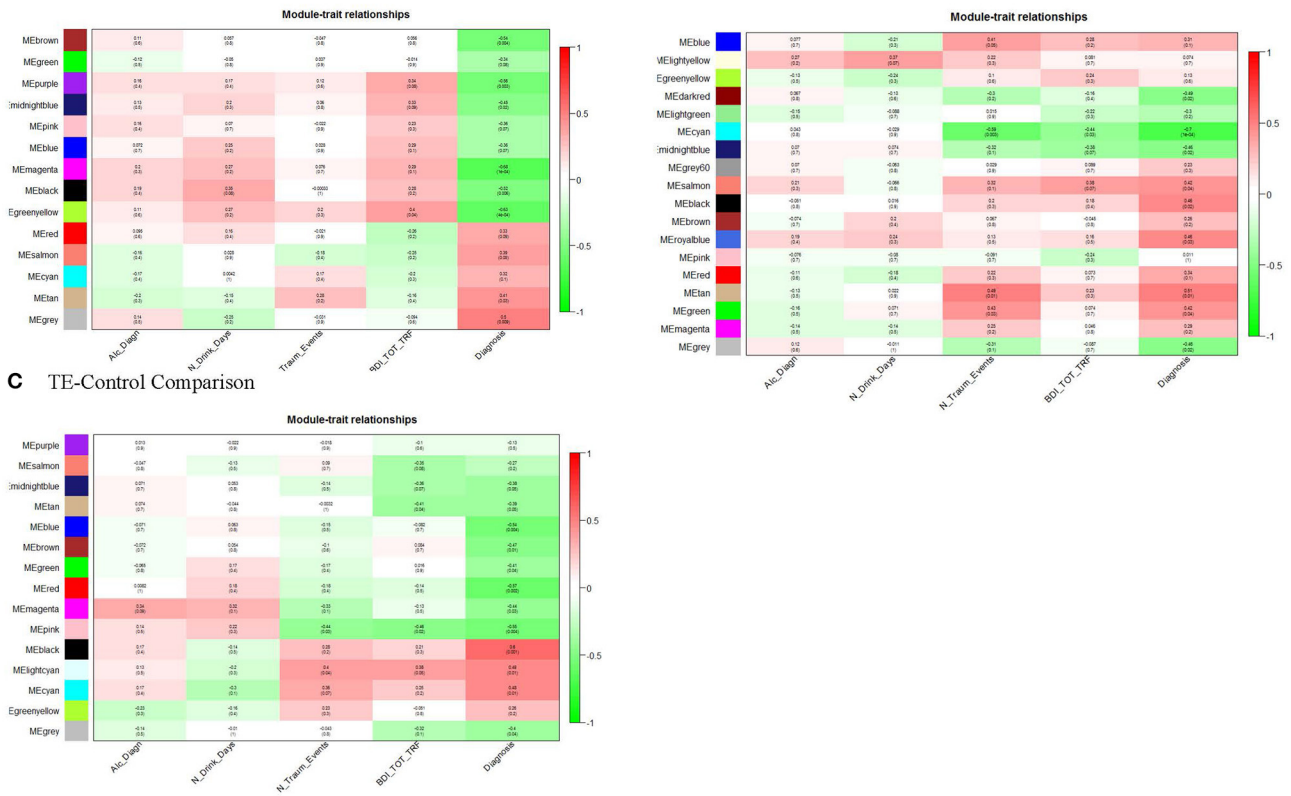


FIGURE 1 | Module-trait relationships. Module MEs are correlated (Pearson) to PTSD/TE case-status (Diagnosis), to identify disease-relevant modules, and to alcohol drinking (Alc_Diagn), number alcohol drinks per day (N_Drink_Days), the number of traumatic events (N_Traum_Events), and Beck depression scale (BDI_TOT_TRF) to assess for confounding of these variables. *P*-values shown are unadjusted for multiple testing. **(A)** After adjusting for number of modules tested, ME_{brown}, ME_{purple}, $ME_{magenta}$, and $ME_{greenyellow}$, are significantly correlated with PTSD/TE status (Diagnosis); **(B)** After adjusting *p*-values for number of modules tested, only ME_{cyan}, module is significantly correlated with PTSD/Cont status (Diagnosis). **(C)** Similarly, after adjusting *p*-values for number of modules tested, ME_{blue}, ME_{red}, ME_{pink}, and ME_{black} modules are significantly correlated with TE/Cont status (Diagnosis).

the main diagnosis phenotype (i.e., the PTSD/TE comparison) the module eigengenes (MEs) of the 14 modules, which represent the sum of gene expression profiles of each module, were correlated to two additional phenotypes: a number of trauma events (N_Traum_Events) and the Beck Depression Inventory scores (BDI_TOT_TRF). Of the 14 modules, the MEs for 4 modules (containing a total of 628 genes) were significantly and negatively correlated with PTSD at an adjusted Bonferroni $p \leq 0.004$ (**Figure 1A**).

PTSD vs. Control Comparison

In this comparison between the PTSD and Control groups, we identified 18 modules, of which after Bonferroni correction (adjusted $p = 0.0027$) only one module (M_{cyan}) consisting of 52 genes was significantly negatively correlated with the PTSD phenotype (**Figure 1B**). Interestingly, M_{cyan} appeared to be also significantly correlated with the number of interpersonal traumatic events as well (adj. $p \leq 0.003$). Thus, it may be that, to some extent, modules promoting PTSD development are shared with those that differentiate those with varying levels of interpersonal trauma (IPT) exposure.

TE vs. Control Comparison

Finally, in this third comparison, we identified 15 modules in total, of which two modules (M_{black} and M_{red}) remained significant at the Bonferroni adjusted significance threshold (adj. $p \leq 0.004$) and 2 additional modules (M_{pink} and M_{blue}) with a suggestive significance (**Figure 1C**). None of the identified modules appeared to be correlated with depressive symptoms. A table containing correlations and *p*-values of all modules across the three comparisons is provided in **Supplementary Table 3**.

Detection of Network Hub Genes

We observed a significant positive correlation between module membership (MM) and gene significance (GS) for all significant modules in the PTSD-TE (**Figures 2A–D**), in the TE-C (**Figures 2E–H**), and in the PTSD-C comparisons (**Figure 2I**),

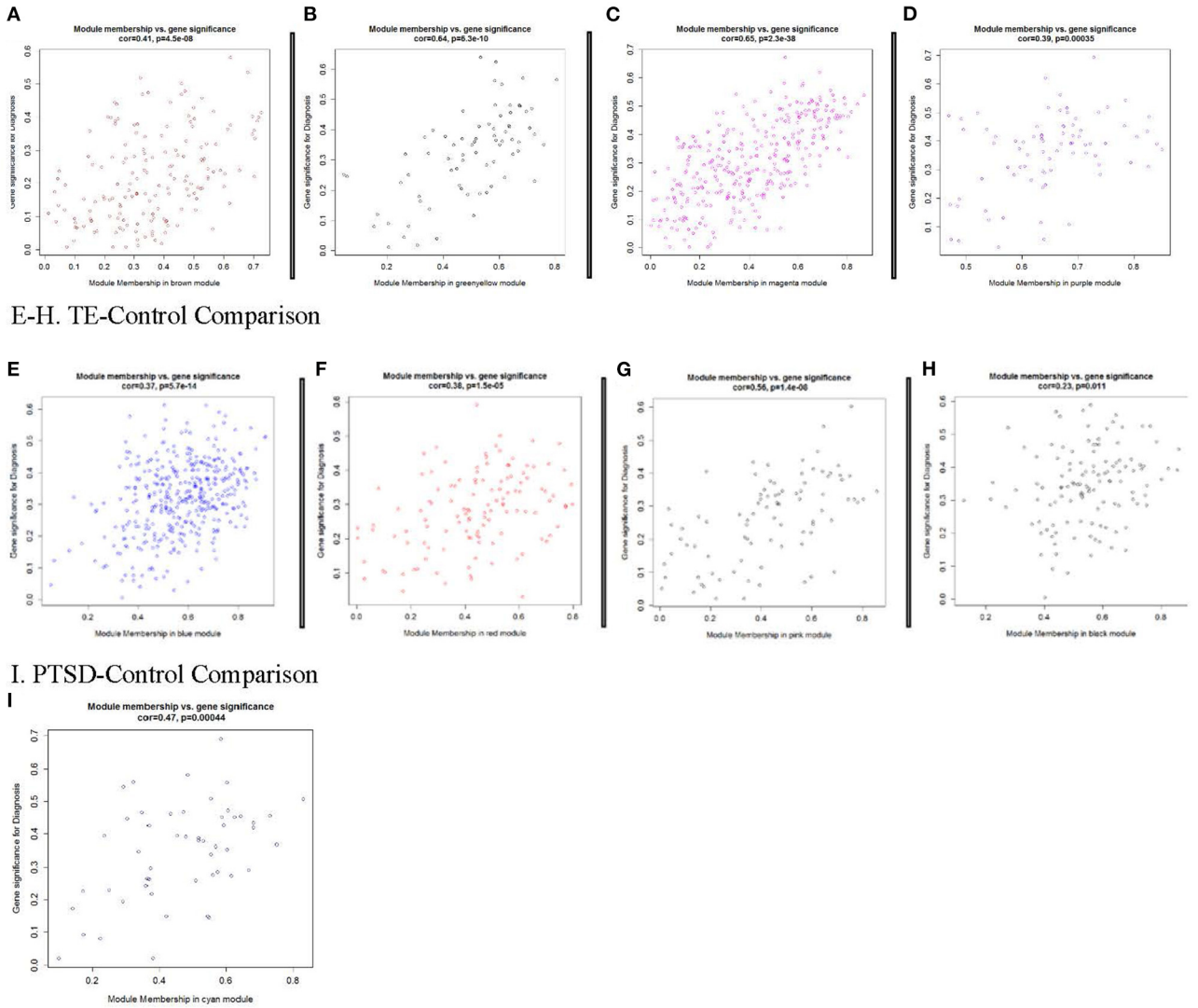


FIGURE 2 | Modules significantly correlated with Diagnosis. Each point represents an individual transcript within each module, which are plotted by the absolute value of their expression correlation to Diagnosis [Gene Significance (GS)] on the y-axis and module eigengene [Module Membership (MM)] on the x-axis. The correlation and corresponding *p*-values are shown for each plot, indicating that increases of GS for transcripts is reflected by the concomitant increase in intramodular connectivity (MM). **(A–D)** Modules significantly correlated with PTSD/TE; **(E-H)** Modules significantly correlated with TE/Cont; and **(I)** Module significantly correlated with PTSD/Cont.

supporting previous observations that genes significantly correlated with the disease are also the most important (or central) elements of the modules (65). Of the 916 genes clustered in the six modules that survive Bonferroni correction, 195 were located in the top quartile of MM and were selected as candidate hub transcripts (see Material and Methods). Interestingly, among these, peptidylprolyl isomerase A (*PPIA*) was shared as a hub between the M_{brown} and M_{pink} modules that were significantly correlated with trauma exposure in the *PTSD vs. TE* and *TE vs. Control* comparisons. Full transcript, GS, MM, and gene symbol annotation for candidate hub transcripts are available in **Supplementary Table 4**.

Assessing Module Preservation

An important consideration of the network analyses is our ability to test how well-preserved (stable) the modules between the different phenotypes are. To that end, we were interested to know whether any of the modules identified in the PTSD subjects, were also preserved in the Control and/or the TE groups; thus, we used 1,000 permutations in the preservation analysis to answer this question. We observed that except for the large modules (such as the turquoise module ≥1,000 genes) all other modules showed modest to no preservation between PTSD, TE and control subjects (**Figures 3A–F**) suggesting that the genes in these modules show distinct connectivity and/or density patterns.

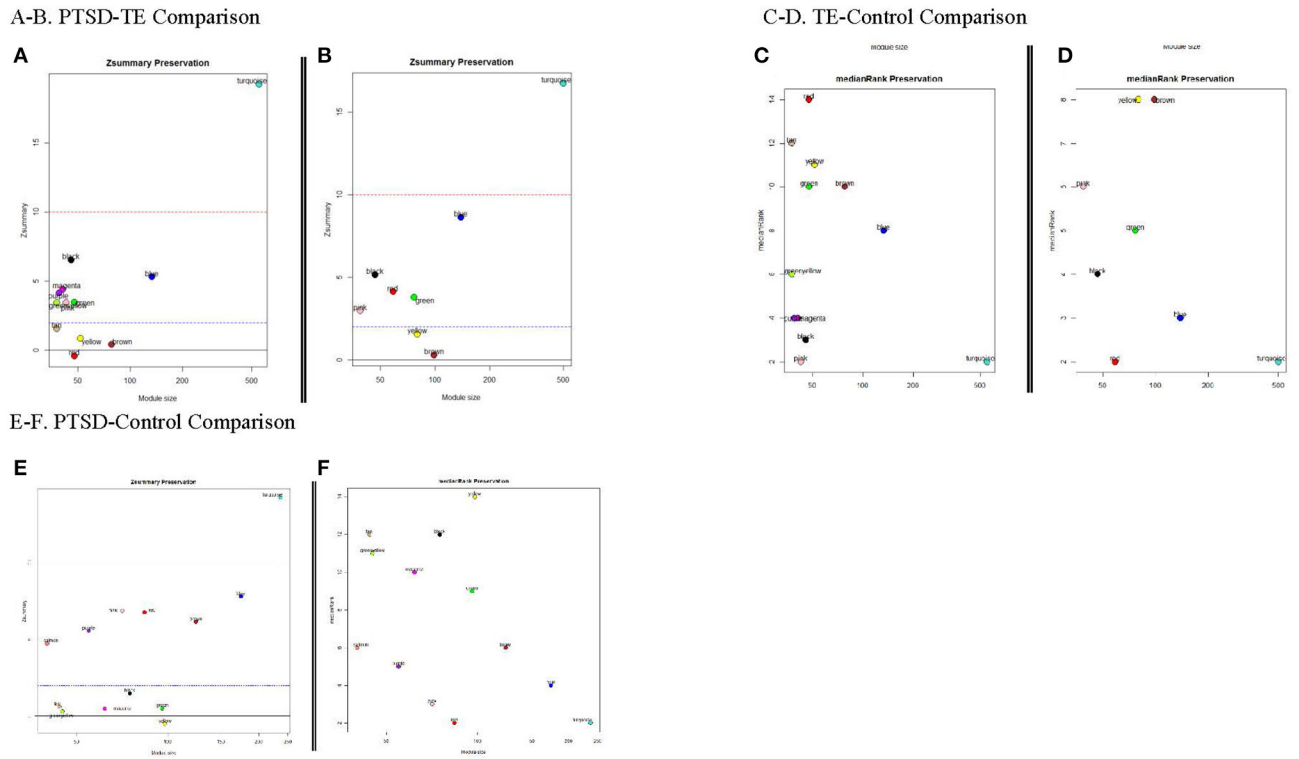


FIGURE 3 | Module Preservation Statistics. Zsummary and medianRank are two mutually complementing preservation statistics to assess module preservation, using either TE or the Cont modules as reference. Y-axis represents preservation statistics for the corresponding module(s) in the case set, and x-axis is the gene numbers in each module. The dashed blue and red lines indicate the thresholds $Z \leq 2$ and $Z \leq 10$, respectively (see text for more details). **(A,B)** Zsummary and medianRank preservations statistics for PTSD/TE comparison, respectively; **(C,D)** Zsummary and medianRank preservations statistics for TE/Cont comparison; **(E,F)** Zsummary and medianRank preservations statistics for PTSD/Cont comparison.

Pathway Enrichment

Using the default parameters in GSEA (see methods and material), at FDR $\leq$ 0.2, we identified a total of 16 *a priori* gene sets. We further observed some of these to be shared between more than one module. For example, *a priori* gene sets enriched for the neuroactive ligand receptor interaction, focal adhesion, and vascular smooth muscle contraction gene sets were predominantly shared between $M_{greenyellow}$, $M_{magenta}$, and M_{brown} modules in the *PTSD vs. TE* comparison and in M_{blue} in the *TE vs. Control*. Similarly, pathways enriched for genes involved in focal adhesion were shared between $M_{greenyellow}$, $M_{magenta}$, and M_{purple} in the *PTSD vs. TE* comparison. In addition to shared gene pathways, we also identified modules with unique enrichments. Among these M_{brown}, identified in the *PTSD vs. TE* comparison, was the only module enriched for genes involved in Alzheimer disease, long-term potentiation and chemokine signaling, while $M_{greenyellow}$ was uniquely enriched for gene sets in the cytokine receptor interaction pathway. Interestingly, while some of the molecular features of PTSD have been shown to be associated with neuroinflammation, in our study, the only module that was enriched for genes related to immune processes such as antigen processing and presentation, or natural killers cell mediated cytotoxicity pathways was in M_{pink} from the *TE vs. Control* comparison. In addition to genes involved in neuronal, cellular, and/or neurodegenerative processes we also identified pathways enriched for genes involved in MAPK, calcium, or chemokine signaling. A full list of the significant unique and shared enriched gene sets, as well as descriptions of PTSD biology, is given in **Table 3**.

Results From Independent Sample

We tested whether our initial findings held in an independent sample composed of military Veterans; the data from that sample are described elsewhere (66). Notably, the Veteran sample is primarily male and slightly older than the initial test sample (i.e., 94.5% male, age range 21–40). The extent to which differentially expressed genes were found across samples was assessed via hypergeometric test assaying the level of overlap between the two datasets and represented by a Venn diagram in **Supplementary Figure 5** and overlapping genes in **Supplementary Table 5**. The Venn diagram was drawn through an online source (https://bioinfogp.cnb.csic.es/tools/venny/index.html). Based on this comparison, we observed a significant overlap for the DEG (at unadjusted $p \leq 0.005$) between our civilian and Veteran samples in the PTSD/TE comparison (i.e., PTSD/TE_Civ ($n = 283$ at least nominally

TABLE 2 | Significant single gene expression analyses between the different comparisons at an FDR of 10%.

Probeset ID	Gene symbol	*T*-value	*p*-value	*q*-value
PTSD/TE				
210562_at	GREB1	11.19	1.12E-08	2.38E-04
221552_at	ABHD6	8.41	4.61E-07	4.91E-03
216134_at	FRMD4B	7.36	2.38E-06	1.69E-02
210049_at	SERPINC1	6.93	4.78E-06	2.54E-02
217144_at	–	−6.22	1.64E-05	4.83E-02
217065_at	–	6.13	1.91E-05	4.83E-02
208389_s_at	SLC1A2	6.13	1.92E-05	4.83E-02
222334_at	C1orf186	−6.12	1.97E-05	4.83E-02
220594_at	OGT	6.1	2.04E-05	4.83E-02
201679_at	SRRT	6.03	2.30E-05	4.90E-02
210782_x_at	GRIN1	−5.74	3.88E-05	6.87E-02
206224_at	CST1	5.63	4.83E-05	6.87E-02
201438_at	COL6A3	5.61	5.01E-05	6.87E-02
216847_at	–	5.6	5.04E-05	6.87E-02
208036_at	OPN1SW	5.52	5.84E-05	6.87E-02
206217_at	EDA	5.51	5.97E-05	6.87E-02
210354_at	IFNG	5.48	6.33E-05	6.87E-02
214382_at	UNC93A	5.48	6.40E-05	6.87E-02
209209_s_at	FERMT2	5.47	6.40E-05	6.87E-02
210431_at	ALPPL2	5.45	6.66E-05	6.87E-02
207506_at	GLRX3	5.43	6.99E-05	6.87E-02
201525_at	APOD	5.4	7.39E-05	6.87E-02
214399_s_at	KRT4	5.37	7.83E-05	6.87E-02
219486_at	DUS2	5.35	8.12E-05	6.87E-02
211469_s_at	CXCR6	−5.33	8.34E-05	6.87E-02
208194_s_at	STAM2	5.31	8.76E-05	6.87E-02
211570_s_at	RAPSN	5.3	8.92E-05	6.87E-02
205364_at	ACOX2	5.28	9.28E-05	6.87E-02
221429_x_at	TEX13A	5.27	9.35E-05	6.87E-02
218651_s_at	LARP6	5.2	0.000108	0.075568
214493_s_at	INADL	−5.18	0.000112	0.075568
208804_s_at	SRSF6	5.17	0.000114	0.075568
222354_at	F11R	5.11	0.000127	0.081852
213475_s_at	ITGAL	−5.09	0.000133	0.08331
218913_s_at	GMIP	−5.06	0.00014	0.08381
214933_at	CACNA1A	−5.05	0.000144	0.08381
205034_at	CCNE2	5.04	0.000146	0.08381
222325_at	–	5.02	0.000153	0.086012
215448_at	–	4.99	0.000161	0.086556
212993_at	NACC2	4.97	0.000167	0.086556
211550_at	EGFR	4.97	0.000167	0.086556
220580_at	BICC1	4.96	0.000171	0.086556
216946_at	HLA-DOA	4.93	0.000183	0.090485
207451_at	NKX2-8	4.9	0.000192	0.09288
218186_at	RAB25	4.86	0.000207	0.09304
205017_s_at	MBNL2	4.86	0.00021	0.09304
207772_s_at	PRMT8	4.85	0.000211	0.09304
215534_at	–	4.85	0.000213	0.09304
207885_at	S100G	4.84	0.000217	0.09304
219759_at	ERAP2	4.83	0.000218	0.09304
219768_at	VTCN1	4.82	0.000226	0.094414
216616_at	–	4.77	0.000249	0.100276
215366_at	SNX13	−4.77	0.00025	0.100276
TE/C				
214399_s_at	KRT4	−6.64	4.86E-07	0.008178
211876_x_at	PCDHGA11	−6.45	7.84E-07	0.008178
PTSD/C				
205915_x_at	GRIN1	−6.04	2.22E-06	1.19E-02
217144_at	–	−5	3.34E-05	6.93E-02
219838_at	TTC23	−4.69	7.58E-05	8.15E-02
200633_at	UBB	−4.61	9.45E-05	8.85E-02
213695_at	PON3	−4.51	0.000122	0.09248

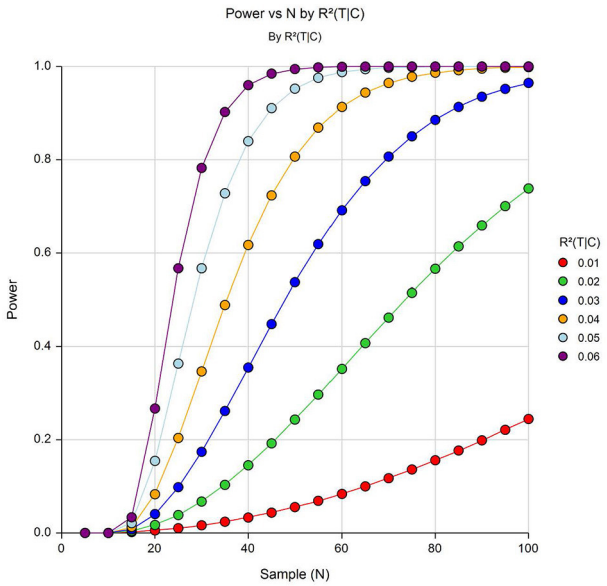


FIGURE 4 | Power to detect differentially expressed genes with varying effect sizes ($R^2$6C) at $\alpha = 5 \times 10^{-4}$. Y-axis shows power and the X-axis shows effective sample size.

different genes); PTSD/TE_Vet (n = 565 at least nominally different genes); gene overlap (n=19); total number of genes (n = 14,800; rf = 1.8; $p = 0.012$) only. There was no significant overlap between the other two comparisons (i.e., PTSD/C and TE/C).

Power Analysis Findings

As shown in **Figure 4**, we have a sufficient power to detect an R^2 increase of 6% attributed to one independent variable after adjusting for 9 covariates at an FDR adjusted $p \leq 0.0005$.

TABLE 3 | GSEA ontology enrichment results for the significant gene modules.

PTSD/TE_Brown_Module	**Size**	**ES**	**NES**	**NOM *p*-value**	**FDR *q*-value**	**FWER *p*-value**	**Rank at max**
KEGG_ALZHEIMERS_DISEASE	16	0.3808411	1.8343772	0.006012024	0.11456045	0.118	564
KEGG_CHEMOKINE_SIGNALING_PATHWAY	28	0.2661027	1.6540613	0.04158416	0.12534362	0.292	685
KEGG_VASCULAR_SMOOTH_MUSCLE_CONTRACTION	16	0.34704605	1.6623145	0.027027028	0.152568	0.283	520
KEGG_LONG_TERM_POTENTIATION	15	0.36446524	1.6956825	0.030592734	0.17543387	0.248	462
PTSD/TE_GreenYellow_Module	**Size**	**ES**	**NES**	**NOM *p*-value**	**FDR *q*-value**	**FWER *p*-value**	**Rank at max**
KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION	38	0.29289687	2.0918672	0	0.039426543	0.032	244
KEGG_FOCAL_ADHESION	25	0.31623906	1.9317259	0.003802281	0.054069344	0.083	921
KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	28	0.2855701	1.8023701	0.010141988	0.06624937	0.144	602
PTSD/TE_Magenta_Module	**Size**	**ES**	**NES**	**NOM *p*-value**	**FDR *q*-value**	**FWER *p*-value**	**Rank at max**
KEGG_FOCAL_ADHESION	25	0.46487573	2.8352916	0	0	0	518
KEGG_VASCULAR_SMOOTH_MUSCLE_CONTRACTION	16	0.3864319	1.9351841	0.004008016	0.04435187	0.067	461
KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION	38	0.25406504	1.81941	0.01026694	0.062306367	0.136	381
PTSD/TE_Purple_Module	**Size**	**ES**	**NES**	**NOM *p*-value**	**FDR *q*-value**	**FWER *p*-value**	**Rank at max**
KEGG_FOCAL_ADHESION	25	0.36741436	2.2619853	0	0.018515298	0.015	542
TE/Controls_Blue_Module	**Size**	**ES**	**NES**	**NOM *p*-value**	**FDR *q*-value**	**FWER *p*-value**	**Rank at max**
KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION	41	0.26657942	2.0054896	0.008032128	0.042644728	0.039	679
KEGG_CALCIUM_SIGNALING_PATHWAY	24	0.29648352	1.698969	0.02892562	0.11025008	0.188	601
KEGG_MAPK_SIGNALING_PATHWAY	31	0.23569264	1.5734301	0.052208837	0.13829629	0.316	768
TE/Controls_Pink_Module	**Size**	**ES**	**NES**	**NOM *p*-value**	**FDR *q*-value**	**FWER *p*-value**	**Rank at max**
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION	15	0.49288395	2.3176186	0	0.004643181	0.004	503
KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY	23	0.32229558	1.7953429	0.016129032	0.080599934	0.137	118

DISCUSSION

This study employs a three-group design that allowed disaggregation of expression differences among PTSD, TE, and C groups. The largest number of expression differences were observed between PTSD-TE groups (54), followed by PTSD-control (33) and TE-control (2). We observed overlap between those that differentiated the PTSD-TE groups, and PTSD-control groups (hypergeometric $p = 1.17E\text{-}33$). Specifically, 19 overlapped across the sets of analyses.

In terms of the network findings, there were 4 modules, 1 module, and 2 modules that differentiated the PTSD-TE, PTSD-C, and TE-C groups, respectively. The only overlapping module was the one that differentiated PTSD-C groups, also differentiating PTSD-TE groups. Supporting previous network findings (32, 33), we identified pathways enriched for genes involved in chemokine signaling and immune system functioning (i.e., natural killer cells; **Supplementary Table 4**), suggesting these modules captured genes sets involved in the inflammatory response to TE and/or PTSD.

Notably, as none of our significant effects overlapped with genome-wide gene expression findings from prior studies, overlap between our findings and those from related literatures are discussed below. Some of our study findings are consistent with candidate gene literature on PTSD. Specifically, *CNR1,* a cannabinoid receptor gene, was part of a module differentially expressed between PTSD-TE, and prior work found evidence for this gene's association with PTSD (67). The Apolipoprotein E gene *APOE* was also part of a module differentially expressed between these two groups, with research finding evidence for its association with PTSD as well (68). Our findings do not provide support for genes whose SNPs have been implicated in PTSD genome-wide association studies [e.g., *ANKRD55, ZNF626*; (69)].

Some genes differentially expressed between groups were associated with functional processes underlying or related to stress response/PTSD. For example, the nicotinic and muscarinic receptor genes *CHRNA4* and *CHRM2*, respectively, within modules differentially expressed between PTSD-TE, impact immune function and inflammatory responses (70, 71), which are relevant to PTSD. *CHRNA4* also impacts arousal, anxiety, and memory (72), and *CHRM2*, is important in learning, memory and higher order brain functions (73). In addition, the gene *ABHD6*, also differentially expressed between PTSD-TE, and involved in cannabinoid system signaling, is implicated in symptoms of anxiety and depression (74), which are highly likely to co-occur with PTSD (75). The gene *GRIN1*, differentially expressed between PTSD-Controls, encodes a NMDA receptor involved in activation of the pre-frontal cortex and previously linked to Bipolar Disorder (76). Thus, the results from this study suggest that genes differentially expressed between the groups may be functionally involved in immune function and inflammation, stress response, arousal, learning and memory, as well as related psychopathology, including anxiety, depression, and Bipolar.

Recent whole blood, transcriptome-wide mega-analysis comparing those with PTSD to trauma-exposed controls was conducted and notably, while collapsing across trauma types (e.g., interpersonal combat) and sex, no case-control comparisons within the univariate analyses resulted in significant effects withstanding the FDR $p < 0.05$ correction (77). Of the

genes differentially expressed between those in the PTSD and TE groups in the current sample, none overlapped with the nominally significant genes differentially expressed in females exposed to IPT in Breen et al. (77). Although no genes overlapped, a number of genes in systems from which genes detected in the current study were found to be (at least nominally) important in differentiating groups. Specifically, the expression of genes involved in innate immune response, the cytokine receptor pathway, and interferon signaling differentiated those with PTSD and TE among females exposed to IPT in both samples.

Of the 283 genes at least nominally differentially expressed between the PTSD and TE groups in the current Civilian sample, 19 of these overlapped with those found to be differentially expressed between the PTSD and TE groups in the Veteran sample. Of these 19 genes that were found to be significant in both samples, some are involved in immune system functioning [e.g., *XIAP*; (78)] signaling processes [e.g., INPP4A; (79)] and learning and motivation [e.g., *EHMT1*; (80)], consistent with prior literature. Others, interestingly, were implicated in tumor growth [e.g., *YES1, PML*; (81, 82)] and a range of other health-related outcomes [e.g., *SNX13* and *CCHCR1*; heart failure, skin disease; (83, 84)]. Thus, over and above trauma type differences between the two samples, it appears that the expression of genes implicated in immune functioning, signal processing, learning/motivation, and broad health-related outcomes differentiate those with trauma exposure with and without PTSD. That is, the expression of genes implicated in these systems arose in two separate samples that differ on target trauma type (interpersonal Civilian vs. combat Veteran), sex (i.e., the civilian sample here was all adult and the Veteran sample was predominantly male), and age [the Veteran age range was higher (21–40) than the civilian (21–30)], which is quite notable. Interestingly, there were no overlapping genes differentially expressed between control and trauma-exposed groups, or between control and PTSD groups across the current study and Veteran samples. As noted above, these two samples differed in terms of the criterion trauma type required, as well as gender and age, so it is perhaps not surprising that two of the comparisons yielded no overlap.

The module preservation statistics further supports our differential expression and network analysis results suggesting separation between the profiles for not only subjects with PTSD and controls, but also for TE individuals. This separation further indicates that different gene modules identified in the PTSD and TE groups are in overall qualitatively different from those in the controls, as well as between each other.

Limitations, Future Directions, and Conclusions

This study is not without limitations. First, because of its cross-sectional, we are unable to determine the direction of effects. Second, given the stated sample characteristics, these findings may not generalize to males or individuals of other ancestral groups. Although the criterion trauma for the initial sample was interpersonal trauma and for the veteran sample was combat trauma, individuals tended to have other traumas in addition to the target trauma, consistent with prior literature (85). However, additional studies using other trauma types (e.g., disaster, bereavement) as the criterion trauma are needed. It is also important to note that the two samples on which we tested our research questions were quite different—that is, one was primarily female, civilian, and interpersonal trauma-exposed and the other was primarily male, veteran, and combat exposed. Thus, effects found across these two samples really point to gene expression differences that hold over and above sex and trauma type. Future work would benefit from examining these questions across two samples that are more similar on these characteristics. In addition, although we conducted power analyses which suggested we could detect effects explaining an R^2 increase of 6% variance, we, of course, would have missed effects that are smaller than this. Fourth, with any network approaches, our network can capture underlying technical biases rather than actual biological differences, though our investigation of all potential covariates that are present in any genome-wide expression data (i.e., batch effects, RIN, probe dependencies) argues against this possibility. Another limitation is the lack of specific significant thresholds, which we can use to declare whether our PTSD, TE, and control networks are specific to each of these phenotypic traits. We applied permutation-based statistics to account for the preservation significance, i.e., estimating the empirical *p*-value distribution, which is more powerful that the application of predefined *p*-values thresholds. However, it has been shown that this approach is dependent on the module size (86). To alleviate this potential shortcoming, we used an additional preservation measure such as *medianRank* measure that is invariant to the module size. Additionally, as a network is formed by using genes included in the network analysis, we could not examine whether genes outside of this network, i.e., not included in the network analysis (please see Materials and Methods, Network Analysis for our justification) were associated with group membership. Future research should attempt to answer this question. Finally, we chose to use a Microarray because we were interested in examining differentially expressed genes between our three groups, but in the future, RNA-sequencing may have future utility for identification of alternative variants.

Despite its limitations, this study makes several key contributions to the literature. It was an early study to utilize differential expression analyses, including a number of key covariates, as well as network analyses in the examination of expression differences within this three-group design. Broadly these findings suggest that there are more genes and networks differentiating those with and without PTSD, compared to the number that differentiate those with and without TE. This suggests that there is something clinically meaningful about PTSD, that differentiates it from exposure to trauma—in terms of gene expression signatures. Additionally, although there is some overlap, the genes differentially expressed following TE are not the same that are expressed among those with PTSD. These findings are a next step toward identifying potentially novel pharmacological targets for this debilitating condition.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because we as researchers do not have consent from participants to deposit de-identified data in a public repository. Requests to access the datasets should be directed to overall study PI, Carla Kmett Danielson, Ph.D., (danielso@musc.edu).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Medical University of South Carolina IRB. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KB was PI for this grant, and wrote introduction, part of method, and discussion. VV was Co-PI on this grant and analyzed data and wrote part of method for manuscript. GM and ZT helped with data analysis. GH, DC, CKD and AA were Co-Is on grant and provided critical scientific input on that grant. CKD and AA were also Co-PIs on larger parent grant. AA wrote part of the method section for this manuscript. ZA was Co-I on larger parent grant. All authors provided critical edits to this manuscript.

FUNDING

This project was supported by the South Carolina Clinical & Translational Research Institute with an academic home at the Medical University of South Carolina CTSA NIH/NCATS grant number UL1TR001450. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH or NCATS. Funding for this project was also provided by NIAAA P50 P50AA010761 (PI: Becker) and R21AA022749 to VV. KB is supported by 1K01AA028058. AA is supported by grants R01AA020179 and K02 AA023239. DC is supported by grant NIH/NIGMS R01 GM122078 and NIH/NCI R21-CA209848 and GH by funding from SC the Established Program to Stimulate Competitive Research (EPSCoR; PI: Hardiman, 2017). ZA is supported by grants K23DA038257.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2021.581093/full#supplementary-material

Supplementary Figure 1 | Microarray quality was assessed via a principal components analysis (PCA) on the expression values in which samples were plotted along the first three principal components (PCs) to identify potential microarray outliers. **(A)** Three samples did not load on two of the first three PCs and were removed from subsequent analysis. **(B)** The PCA results were further corroborated via unsupervised hierarchical clustering (using Euclidian metrics), with two of the same samples identified in the PCA analysis also presenting as outliers in the unsupervised hierarchical clustering.

Supplementary Figure 2 | Microarray expression data validation using quantitative (qPCR) approach. Expression levels of four genes measured by the expression array-based approach were validated by qPCR in all 45 RNA samples. The Y-axis represents the qPCR data and the X-axis the array data. The correlation coefficients were calculated using the Pearson product-moment.

Supplementary Figure 3 | Figure displays the cell deconvolution as a heatmap where each column is a sample, and each row represents a different immune cell type. Each entry (square) in the matrix shows the inferred quantity of a certain cell type in a specific input sample, i.e., the red/blue coloring scheme indicates increase or decrease in estimated cell proportion relative to the control sample.

Supplementary Figure 4 | Quantile–quantile (QQ) plots of the gene expression signals across the PTSD/TE, TE/Cont, and PTSD/Cont comparisons. The expected –log10 *p*-values under the null hypothesis are represented on the x-axis, while the observed values are represented on the y-axis.

Supplementary Figure 5 | Overlap between civilian and veteran samples.

Supplementary Table 1 | Microarray validation using quantitative (qPCR) approach.

Supplementary Table 2A | Single gene expression analysis statistics for PTSD/TE comparison. The highlighted rows represent significant DEG at FDR of 10%.

Supplementary Table 2B | Single gene expression analysis statistics for TE/Cont comparison. The highlighted rows represent significant DEG at FDR of 10%.

Supplementary Table 2C | Single gene expression analysis statistics for PTSD/Cont comparison. The highlighted rows represent significant DEG at FDR of 10%.

Supplementary Table 3A | The entire list of genes with their espective gene significance and module membership correlation values across all modules identified in PTSD/Cont comparison.

Supplementary Table 3B | The entire list of genes with their espective gene significance and module membership correlation values across all modules identified in PTSD/TE comparison.

Supplementary Table 3C | The entire list of genes with their espective gene significance and module membership correlation values across all modules identified in TE/Cont comparison.

Supplementary Table 4 | Identified hub genes for the three comparsions. Additional information such as correlation and *p*-values for gene significance (GS) and module membership (MM) are also presented.

Supplementary Table 5 | The set of genes overlapping between the civilian and veteran samples at single gene expression level of $p \leq 0.005$.

REFERENCES

1. Benjet C, Bromet E, Karam EG, Kessler RC, McLaughlin KA, Ruscio AM, et al. The epidemiology of traumatic event exposure worldwide: results from the world mental health survey consortium. *Psychol Med.* (2016) 46:327–43. doi: 10.1017/S0033291715001981
2. Breslau N, Davis G, Andreski P, Federman B, Anthony JC. Epidemiological findings on posttraumatic stress disorder and co-morbid disorders in the general population. In *Adversity, Stress, and Psychopathology*. London; Oxford: University Press (1998). p. 319–28
3. Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB. Posttraumatic stress disorder in the national comorbidity survey. *Arch Gen Psychiatry.* (1995) 52:1048–60. doi: 10.1001/archpsyc.1995.03950240066012
4. Boscarino JA, Forsberg CW, Goldberg J. A twin study of the association between PTSD symptoms and rheumatoid arthritis. *Psychosom Med.* (2010) 72:481–6. doi: 10.1097/PSY.0b013e3181d9a80c

5. Kessler RC. Posttraumatic stress disorder: the burden to the individual and to society. *J Clin Psychiatry.* (2000) 61:4–12. doi: 10.4088/JCP.v61n0713e
6. Neupane SP, Bramness JG, Lien L. Comorbid post-traumatic stress disorder in alcohol use disorder: relationships to demography, drinking and neuroimmune profile. *BMC Psychiatry.* (2017) 17:312. doi: 10.1186/s12888-017-1479-8
7. Sartor CE, McCutcheon VV, Pommer NE, Nelson EC, Grant JD, Duncan AE, et al. Common genetic and environmental contributions to posttraumatic stress disorder and alcohol dependence in young women. *Psychol Med.* (2011) 41:1497–505. doi: 10.1017/S0033291710002072
8. Stein MB, Jang KJ, Taylor S, Vernon PA, Livesley WJ. Genetic and environmental influences on trauma exposure and posttraumatic stress disorder: a twin study. *Am J Psychiatry*. (2002) 159:1675–81. doi: 10.1176/appi.ajp.159.10.1675
9. True WJ, Rice J, Eisen SA, Heath AC, Goldberg J, Lyons MJ, et al. A twin study of genetic and environmental contributions to liability for posttraumatic stress symptoms. *Arch Gen Psychiatry*. (1993) 50:257–64. doi: 10.1001/archpsyc.1993.01820160019002
10. Gelernter J, Sun N, Polimanti R, Pietrzak R, Levey DF, Bryois J, et al. Genome-wide association study of post-traumatic stress disorder reexperiencing symptoms in> 165,000 US veterans. *Nat Neurosci.* (2019) 22:1394–401. doi: 10.1038/s41593-019-0447-7
11. Nievergelt CM, Maihofer AX, Klengel T, Atkinson EG, Chen CY, Choi KW, et al. International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nat Commun*. 10:4558. doi: 10.1038/s41467-019-12576-w
12. Stein MB, Levey DF, Cheng Z, Wendt FR, Harrington K, Pathak GA, et al. Genome-wide association analyses of post-traumatic stress disorder and its symptom subdomains in the million veteran program. *Nat Genet*. 53:174–84. doi: 10.1038/s41588-020-00767-x
13. Daskalakis NP, Rijal CM, King C, Huckins LM, Ressler KJ. Recent genetics and epigenetics approaches to PTSD. *Curr Psychiatry Rep*. (2018) 20:30. doi: 10.1007/s11920-018-0898-7
14. Duncan LE, Ratanatharathorn A, Aiello AE, Almli LM, Amstadter AB, Ashley-Koch AE, et al. Largest GWAS of PTSD (N=20 070) yields genetic overlap with schizophrenia and sex differences in heritability. *Mol Psychiatry* (2017) 23:666–73. doi: 10.1038/mp.2017.77
15. Ressler KJ, Mercer KB, Bradley B, Jovanovic T, Mahan A, Kerley K, et al. Post-traumatic stress disorder is associated with PACAP and the PAC1 receptor. *Nature.* (2011) 470:492–7. doi: 10.1038/nature09856
16. Yehuda R, Cai G, Golier JA, Sarapas C, Galea S, Ising M, et al. Gene expression patterns associated with posttraumatic stress disorder following exposure to the world trade center attacks. *Biol Psychiatry.* (2009) 66:708–11. doi: 10.1016/j.biopsych.2009.02.034
17. Su YA, Wu J, Zhang L, Zhang Q, Su DM, He P, et al. Dysregulated mitochondrial genes and networks with drug targets in postmortem brain of patients with posttraumatic stress disorder (PTSD) revealed by human mitochondria-focused cDNA microarrays. *Int J Biol Sci.* (2008) 4:223–35. doi: 10.7150/ijbs.4.223
18. Lutz PE, Tanti A, Gasecka A, Barnett-Burns S, Kim JJ, Zhou Y, et al. Association of a history of child abuse with impaired myelination in the anterior cingulate cortex: convergent epigenetic, transcriptional, morphological evidence. *Am J Psychiatry.* (2017) 174:1185–94. doi: 10.1176/appi.ajp.2017.16111286
19. Girgenti MJ, Wang J, Ji D, Cruz DA, Stein MB, Gelernter J, et al. Transcriptomic organization of the human brain in post-traumatic stress disorder. *Nat Neurosci.* (2021) 24:24–33. doi: 10.1038/s41593-020-00748-7
20. Jovanovic T, Norrholm SD, Blanding NQ, Davis M, Duncan E, Bradley B, et al. Impaired fear inhibition is a biomarker of PTSD but not depression. *Depress Anxiety.* (2010) 27:244–51. doi: 10.1002/da.20663
21. Durrenberger PF, Fernando S, Kashefi SN, Ferrer I, Hauw JJ, Seilhean D, et al. Effects of antemortem and postmortem variables on human brain mRNA quality: a BrainNet Europe study. *J Neuropathol Exp Neurol.* (2010) 69:70–81. doi: 10.1097/NEN.0b013e3181c7e32f
22. Hynd MR, Lewohl JM, Scott HL, Dodd PR. Biochemical and molecular studies using human autopsy brain tissue. *J Neurochem*. (2003) 85:543–62. doi: 10.1046/j.1471-4159.2003.01747.x
23. Jasinska AJ, Service S, Choi OW, DeYoung J, Grujic O, Kong SY, et al. Identification of brain transcriptional variation reproduced in peripheral blood: an approach for mapping brain expression traits. *Hum Mol Genet.* (2009) 18:4415–27. doi: 10.1093/hmg/ddp397
24. Sullivan PF, Fan C, Perou CM. Evaluating the comparability of gene expression in blood and brain. *Am J Med Genet B Neuropsychiatr Genet*. (2006) 141B, 261–268. doi: 10.1002/ajmg.b.30272
25. Witt SH, Sommer WH, Hansson AC, Sticht C, Rietschel M, Witt CC. Comparison of gene expression profiles in the blood, hippocampus and prefrontal cortex of rats. *In Silico Pharmacol.* (2013) 1:15. doi: 10.1186/2193-9616-1-15
26. Hakonarson H, Bjornsdottir US, Halapi E, Bradfield J, Zink F, Mouy M, et al. Profiling of genes expressed in peripheral blood mononuclear cells predicts glucocorticoid sensitivity in asthma patients. *Proc Natl Acad Sci USA*. (2005) 102:14789–94. doi: 10.1073/pnas.0409904102
27. Levy-Gigi E, Szabo C, Kelemen O, Keri S. Association among clinical response, hippocampal volume, and FKBP5 gene expression in individuals with posttraumatic stress disorder receiving cognitive behavioral therapy. *Biol Psychiatry.* (2013) 74:793–800. doi: 10.1016/j.biopsych.2013.05.017
28. Sarapas C, Cai G, Bierer LM, Golier JA, Galea S, Ising M, et al. Genetic markers for PTSD risk and resilience among survivors of the World Trade Center attacks. *Dis Markers.* (2011) 30:101–10. doi: 10.1155/2011/328054
29. Segman RH, Shefi N, Goltser-Dubner T, Friedman N, Kaminski N, Shalev AY. Peripheral blood mononuclear cell gene expression profiles identify emergent post-traumatic stress disorder among trauma survivors. *Mol Psychiatry.* (2005) 10:500–13. doi: 10.1038/sj.mp.4001636
30. Yehuda R. Status of glucocorticoid alterations in post-traumatic stress disorder. *Ann N Y Acad Sci.* (2009) 1179:56–69. doi: 10.1111/j.1749-6632.2009.04979.x
31. Zieger J. Differential gene expression in peripheral blood of patients suffering from PTSD. *Mol Psychiatry.* (2007) 12:116–9. doi: 10.1038/sj.mp.4001905
32. Breen MS, Maihofer AX, Glatt SJ, Tylee DS, Chandler SD, Tsuang MT, et al. Gene networks specific for innate immunity define post-traumatic stress disorder. *Mol Psychiatry.* (2015) 20:1538–45. doi: 10.1038/mp.2015.9
33. Bam M, Yang X, Zhou J, Ginsberg JP, Leyden Q, Nagarkatti PS, et al. Evidence for epigenetic regulation of pro-inflammatory cytokines, interleukin-12 and interferon gamma, in peripheral blood mononuclear cells from PTSD patients. *J Neuroimmune Pharmacol.* (2016) 11:168–81. doi: 10.1007/s11481-015-9643-8
34. Mason MJ, Fan G, Plath K, Zhou Q, Horvath S. Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics.* (2009) 10:327. doi: 10.1186/1471-2164-10-327
35. Plaisier CL, Horvath S, Huertas-Vazquez A, Cruz-Bautista I, Herrera MF, Tusie-Luna T, et al. A systems genetics approach implicates USF1, FADS3, and other causal candidate genes for familial combined hyperlipidemia. *PLoS Genet.* (2009) 5:e1000642. doi: 10.1371/journal.pgen.1000642
36. Winden KD, Karsten SL, Bragin A, Kudo LC, Gehman L, Ruidera J, et al. A systems level, functional genomics analysis of chronic epilepsy. *PLoS ONE.* (2011) 6:e20763. doi: 10.1371/journal.pone.0020763
37. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. (2005) 102:15545–50. doi: 10.1073/pnas.0506580102
38. Walker J, Bountress KE, Calhoun CD, Metzger IW, Adams Z, Amstadter A, et al. Impulsivity and comorbid PTSD-Binge drinking. *J Dual Diag.* 14:89–95. doi: 10.1080/15504263.2018.1439212
39. Gray MJ, Litz BT, Hsu JL, Lombardo TW. Psychometric properties of the life events checklist. *Assessment*. (2004) 11:330–41. doi: 10.1177/1073191104269954
40. Weathers FW, Litz BT, Herman DS, Huska JA, Keane TM. The PTSD checklist (PCL): reliability, validity, diagnostic utility. In *Annual Convention of the International Society for Traumatic Stress Studies*. Vol. 462. San Antonio, TX (1993).
41. Yehuda R, Engel SM, Brand SR, Seckl J, Marcus SM, Berkowitz GS. Transgenerational effects of posttraumatic stress disorder in babies of mothers exposed to the world trade center attacks during pregnancy. *J Clin Endocrinol Metab.* (2005) 90:4115–8. doi: 10.1210/jc.2005-0550

42. Sheehan DV, Janavs J, Baker R, Harnett-Sheehan K, Knapp M, Sheehan M. *MINI Plus: Mini International Neuropsychiatric Interview (Version 1335)* (2006).
43. Beck AT, Steer RA, Brown GK. *Manual for the Beck Depression Inventory*. 2nd Edn. San Antonio, TX: The Psychological Corporation (1996). doi: 10.1037/t00742-000
44. Dumur CI, Ladd AC, Wright HV, Penberthy LT, Wilkinson DS, Powers CN, et al. Genes involved in radiation therapy response in head and neck cancers. *Laryngoscope.* (2009) 119:91–101. doi: 10.1002/lary.20005
45. Vardhanabhuti S, Blakemore SJ, Clark SM, Ghosh S, Stephens RJ, Rajagopalan D. A comparison of statistical tests for detecting differential expression using affymetrix oligonucleotide microarrays. *OMICS.* (2006) 10:555–66. doi: 10.1089/omi.2006.10.555
46. Wu Z, Irizarry RA, Gentleman R, Martinez-Murillo F, Spencer F. A model-based background adjustment for oligonucleotide expression arrays. *J Am Stat Assoc.* (2004) 99:909–17. doi: 10.1198/016214504000000683
47. Frishberg A, Brodt A, Steuerman Y, Gat-Viks I. ImmQuant: a user-friendly tool for inferring immune cell-type composition from gene-expression data. *Bioinformatics.* (2016) 32:3842–3. doi: 10.1093/bioinformatics/btw535
48. Altboum Z, Steuerman Y, David E, Barnett-Itzhaki Z, Valadarsky L, Keren-Shaul H, et al. Digital cell quantification identifies global immune cell dynamics during influenza infection. *Mol Syst Biol.* (2014) 10:720. doi: 10.1002/msb.134947
49. Mamdani M, Williamson V, McMichael GO, Blevins T, Aliev F, Adkins A, et al. Integrating mRNA miRNA weighted gene co-expression networks with eQTLs in the nucleus accumbens of subjects with alcohol dependence. *PLoS ONE.* (2015) 10:e0137671. doi: 10.1371/journal.pone.0137671
50. Perkonigg A, Kessler RC, Storz S, Wittchen HU. Traumatic events and post-traumatic stress disorder in the community: prevalence, risk factors and comorbidity. *Acta Psychiatr Scand.* (2000) 101:46–59. doi: 10.1034/j.1600-0447.2000.101001046.x
51. O'Donnell ML, Creamer M, Pattison P. Posttraumatic stress disorder and depression following trauma: understanding comorbidity. *Am J Psychiatry.* (2004) 161:1390–6. doi: 10.1176/appi.ajp.161.8.1390
52. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* (2008) 9:559. doi: 10.1186/1471-2105-9-559
53. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol.* (2005) 4*:* Article17. doi: 10.2202/1544-6115.1128
54. Ponomarev I, Wang S, Zhang L, Harris RA, Mayfield RD. Gene coexpression networks in human brain identify epigenetic modifications in alcohol dependence. *J Neurosci.* (2012) 32:1884–97. doi: 10.1523/JNEUROSCI.3136-11.2012
55. Hung JH, Yang TH, Hu Z, Weng Z, DeLisi C. Gene set enrichment analysis: performance evaluation and usage guidelines. *Brief Bioinformat.* (2012) 13:281–91. doi: 10.1093/bib/bbr049
56. Gray KA, Daugherty LC, Gordon SM, Seal RL, Wright MW, Bruford EA. Genenames.org: the HGNC resources in 2013. *Nucleic Acids Res.* (2013) 41 (Database issue):D545–52. doi: 10.1093/nar/gks1066
57. Albert R. Scale-free networks in cell biology. *J Cell Sci.* (2005) 118(Pt 21), 4947–57. doi: 10.1242/jcs.02714
58. Horvath S, Dong J. Geometric interpretation of gene coexpression network analysis. *PLoS Comput Biol.* (2008) 4:e1000117. doi: 10.1371/journal.pcbi.1000117
59. Presson AP, Sobel EM, Papp JC, Suarez CJ, Whistler T, Rajeevan MS, et al. Integrated weighted gene co-expression network analysis with an application to chronic fatigue syndrome. *BMC Syst Biol.* (2008) 2:95. doi: 10.1186/1752-0509-2-95
60. Langfelder P, Castellani LW, Zhou Z, Paul E, Davis R, Schadt EE, et al. A systems genetic analysis of high density lipoprotein metabolism and network preservation across mouse models. *Biochim Biophys Acta.* (2012) 1821:435–47. doi: 10.1016/j.bbalip.2011.07.014
61. Korthauer K, Kimes PK, Duvallet C, Reyes A, Subramanian A, Teng M, et al. A practical guide to methods controlling false discoveries in computational biology. *bioRxiv.* (2018) 20:458786. doi: 10.1101/458786
62. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA.* (2003) 100:9440–5. doi: 10.1073/pnas.1530509100
63. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, et al. Redefine statistical significance. *Nat Hum Behav.* (2018) 2:6–10. doi: 10.1038/s41562-017-0189-z
64. Battle A, Brown CD, Engelhardt BE, Montgomery SB. Genetic effects on gene expression across human tissues. *Nature.* (2017) 550:204–13. doi: 10.1038/nature24277
65. Zhang H, Wang F, Xu H, Liu Y, Liu J, Zhao H, et al. Differentially co-expressed genes in postmortem prefrontal cortex of individuals with alcohol use disorders: influence on alcohol metabolism-related pathways. *Hum Genet.* (2014) 133:1383–94. doi: 10.1007/s00439-014-1473-x
66. Lind MJ, Brown E, Farrell-Carnahan L, Brown RC, Hawn S, Berenz E, et al. Sleep disturbances in OEF/OIF/OND veterans: associations with PTSD, personality, and coping. *J Clin Sleep Med.* (2017) 13:291–9. doi: 10.5664/jcsm.6466
67. Smoller JW. The genetics of stress-related disorders: PTSD, depression, anxiety disorders. *Neuropsychopharmacology.* (2015) 41:297–319. doi: 10.1038/npp.2015.266
68. Freeman T, Roca V, Guggenheim F, Kimbrell T, Griffin WST. Neuropsychiatric association of apolipoprotein E alleles in subjects with combat-related posttraumatic stress disorder. *J Neuropsychiatry Clin Neurosci.* (2005) 17:541–3. doi: 10.1176/jnp.17.4.541
69. Stein MB, Chen C-Y, Ursano RJ, Cai T, Gelernter J, Heeringa SG, et al. Genome-wide association studies of posttraumatic stress disorder in 2 cohorts of US army soldiers. *JAMA Psychiatry.* (2016) 73:695–704. doi: 10.1001/jamapsychiatry.2016.0350
70. Hou D, Chen Y, Liu J, Xu L, Zhang Z, Zhang J, et al. Proteomics screen to reveal molecular changes mediated by C722G missense mutation in CHRM2 gene. *J Proteomics.* (2013) 89:39–50. doi: 10.1016/j.jprot.2013.05.027
71. Sfera A, Cummings M, Osorio C. Non-neuronal acetylcholine: the missing link between sepsis, cancer, and delirium? *Front Med.* (2015) 2:56. doi: 10.3389/fmed.2015.00056
72. CHRNA4 gene. (2020). *Genetics Home Reference.* Available online at: https://ghr.nlm.nih.gov/gene/CHRNA4#normalfunction
73. Gosso FM, de Geus EJ, Polderman TJ, Boomsma DI, Posthuma D, Heutink P. Exploring the functional role of the CHRM2 gene in human cognition: results from a dense genotyping and brain expression study. *BMC Med Genet.* (2007) 8:66. doi: 10.1186/1471-2350-8-66
74. Marrs WR, Horne EA, Ortega-Gutierrez S, Cisneros JA, Xu C, Lin YH, et al. Dual inhibition of alpha/beta-hydrolase domain 6 and fatty acid amide hydrolase increases endocannabinoid levels in neurons. *J Biol Chem.* (2011) 286:28723–8. doi: 10.1074/jbc.M110.202853
75. Spinhoven P, Penninx BW, van Hemert AM, de Rooij M, Elzinga BM. Comorbidity of PTSD in anxiety and depressive disorders: prevalence and shared risk factors. *Child Abuse Negl.* (2014) 38:1320–30. doi: 10.1016/j.chiabu.2014.01.017
76. Mundo E, Tharmalingham S, Neves-Pereira M, Dalton EJ, Macciardi F, Parikh SV, et al. Evidence that the N-methyl-D-aspartate subunit 1 receptor gene (GRIN1) confers susceptibility to bipolar disorder. *Mol Psychiatry.* (2003) 8:241–5. doi: 10.1038/sj.mp.4001218
77. Breen MS, Tylee DS, Maihofer AX, Neylan TC, Mehta D, Binder EB, et al. PTSD blood transcriptome mega-analysis: shared inflammatory pathways across biological sex and modes of trauma. *Neuropsychopharmacology.* (2018) 43:469–81. doi: 10.1038/npp.2017.220
78. Marsh RA, Madden L, Kitchen BJ, Mody R, McClimon B, Jordan MB, et al. XIAP deficiency: a unique primary immunodeficiency best classified as X-linked familial hemophagocytic lymphohistiocytosis and not as X-linked lymphoproliferative disease. *Blood.* (2010) 116:1079–82. doi: 10.1182/blood-2010-01-256099
79. Zhang M, Tian H, Li R, Yan W, Xu R. MicroRNA-4286 promotes esophageal carcinoma development by targeting INPP4A to evoke the JAK2/STAT3 pathway activation. *Pharmazie.* (2018) 73:342–8. doi: 10.1691/ph.2018.7339
80. Kramer JM, Kochinke K, Oortveld MA, Marks H, Kramer D, de Jong EK, et al. Epigenetic regulation of learning and memory by *Drosophila* EHMT/G9a. *PLoS Biol.* (2011) 9:e1000569. doi: 10.1371/journal.pbio.1000569
81. Hamanaka N, Nakanishi Y, Mizuno T, Horiguchi-Takei K, Akiyama N, Tanimura H, et al. YES1 is a targetable oncogene in cancers harboring YES1 gene amplification. *Cancer Res.* (2019) 79:5734–45. doi: 10.1158/0008-5472.CAN-18-3376

82. Mazza M, Pelicci PG. Is PML a tumor suppressor? *Front Oncol.* (2013) 3:174. doi: 10.3389/fonc.2013.00174
83. Li J, Li C, Zhang D, Shi D, Qi M, Feng J, et al. SNX13 reduction mediates heart failure through degradative sorting of apoptosis repressor with caspase recruitment domain. *Nat Commun.* (2014) 5:5177. doi: 10.1038/ncomms6177
84. Tervaniemi M, Siitonen HA, Soderhall C, Minhas G, Vuola J, Tiala I, et al. Centrosomal localization of the psoriasis candidate gene product, CCHCR1, supports a role in cytoskeletal organization. *PLoS Biol.* (2012) 7:e49920. doi: 10.1371/journal.pone.0049920
85. Kessler RC, Aguilar-Gaxiola S, Alonso J, Benjet C, Bromet EJ, Cardoso G, et al. Trauma and PTSD in the WHO world mental health surveys. *Eur J Psychotraumatol.* (2017) 8:1353383. doi: 10.1080/20008198.2017.1353383
86. Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? *PLoS Comput Biol.* (2011) 7:e1001057. doi: 10.1371/journal.pcbi.1001057

Disclaimer: The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH or NCATS.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychiatry

BRIEF RESEARCH REPORT
published: 29 April 2021
doi: 10.3389/fpsyt.2021.672241

Increased Prevalence of Minor Physical Anomalies Among the Healthy First-Degree Relatives of Bipolar I Patients – Results With the Méhes Scale

*Tímea Csulak[1], Györgyi Csábi[2], Róbert Herold[1], Viktor Vörös[1], Sára Jeges[3], András Hajnal[1], Márton Áron Kovács[1], Maria Simon[1], Márton Herold[1], Ákos Levente Tóth[4] and Tamás Tényi[1]**

[1] Department of Psychiatry and Psychotherapy, Medical School, University of Pécs, Pécs, Hungary, [2] Department of Pediatrics, Medical School, University of Pécs, Pécs, Hungary, [3] Faculty of Health Sciences, Institute of Nursing and Patients Care, University of Pécs, Pécs, Hungary, [4] Faculty of Sciences, Institute of Physical Education and Sport Science, University of Pécs, Pécs, Hungary

OPEN ACCESS

Edited by:
Patrick E. Williams, University of Kansas, United States

Reviewed by:
Ferihan Ahmed-Popova, Plovdiv Medical University, Bulgaria
Laiana Quagliato, Federal University of Rio de Janeiro, Brazil

***Correspondence:**
Tamás Tényi
tenyi.tamas@pte.hu

Specialty section:
This article was submitted to Mood and Anxiety Disorders, a section of the journal Frontiers in Psychiatry

Received: *25 February 2021*
Accepted: *07 April 2021*
Published: *29 April 2021*

Citation:
Csulak T, Csábi G, Herold R, Vörös V, Jeges S, Hajnal A, Kovács MÁ, Simon M, Herold M, Tóth ÁL and Tényi T (2021) Increased Prevalence of Minor Physical Anomalies Among the Healthy First-Degree Relatives of Bipolar I Patients – Results With the Méhes Scale. Front. Psychiatry 12:672241. doi: 10.3389/fpsyt.2021.672241

Minor physical anomalies are somatic markers of aberrant neurodevelopment, so the higher prevalence of these signs among the relatives of bipolar I patients can confirm minor physical anomalies as endophenotypes. The aim of the study was to evaluate the prevalence of minor physical anomalies in first-degree healthy relatives of patients with bipolar I disorder compared to normal control subjects. Using a list of 57 minor physical anomalies (the Méhes Scale), 20 first-degree unaffected relatives of patients with the diagnosis of bipolar I disorder and as a comparison 20 matched normal control subjects were examined. Minor physical anomalies were more common in the ear, head, mouth and trunk regions among the relatives of bipolar I patients compared to normal controls. By the differentiation of minor malformations and phenogenetic variants, we have found that both minor malformations and phenogenetic variants were more common among the relatives of bipolar I patients compared to the control group, while individual analyses showed, that one minor malformation (sole crease) and one phenogenetic variant (high arched palate) were more prevalent in the relative group. This is the first report in literature on the increased prevalence of minor physical anomalies among the first-degree unaffected relatives of bipolar I patients. The study support the concept, that minor physical anomalies can be endophenotypic markers of bipolar I affective disorder.

Keywords: psychotic disorders, somatic markers, endophenotype, bipolar disorder, neurodevelopment

INTRODUCTION

Minor physical anomalies (MPAs) are clinically nonsignificant errors of morphogenesis which have a prenatal origin and may bear important informational value. The presence of minor physical anomalies is a sensitive indicator of altered embryonic development. Since both the brain and the skin derived from the ectoderm, minor physical anomalies may be physical markers of aberrant neurodevelopment. Minor physical anomalies develop during the first and/or early second trimester of gestation (1–5) and represent potentially valuable indices of disturbances

in early brain development. "Once formed they persist into adult life and can be detected on visual examination of the particular body area" (9, 224.pp.). As our group (4, 6–10) and others (11, 12) have discussed earlier, differences and contradictions between studies on minor physical anomalies among adults and children with different neuropsychiatric disorders, may be associated, partly, with the problems in the use of the Waldrop-scale for the detection of these signs. The Waldrop-scale contains only 18 minor physical anomalies (13) while in recent genetic and pediatric literature more than 50 anomalies have been listed (1, 2, 8, 12). A major problem with the Waldrop-scale that it makes no distinction between minor malformations, which arise during organogenesis and phenogenetic variants, which appear after organogenesis (1, 2, 7). A clear distinction between morphogenetic events developing during and after organogenesis seems to be needed. Minor malformations are always abnormal and are qualitive defects of embryogenesis, which arise during organogenesis. All malformations are developmental field defects and usually they are all-or-none anomalies. In contrast phenogenetic variants are quantitative defects of final morphogenesis and arise after organogenesis" (9,224.pp.). Using a list of MPAs containing 57 minor signs evaluated by Méhes (2), previously we have studied the prevalence of minor physical anomalies in patients with schizophrenia, bipolar affective disorder, alcohol dependence, Tourette syndrome, major depression and among the healthy relatives of patients with schizophrenia (3, 6, 8–10, 14–16), and the list and definitions has become also acceptable for researchers (8).

The endophenotype concept of bipolar affective disorder represents an important approach in the exploration of the pathogenesis of the illness. Gottesman and Gould (17) described an endophenotype as an intermediate phenotype that fills the gap between genes and diseases. "Endophenotypes should be: (1) associated with the illness, (2) heritable, (3) state-independent, (4) found in unaffected relatives at a higher rate than in the general population, and (5) shown to co-segregate with the illness within families" (9,225.pp.). An important characteristic of an endophenotype, that it can be more commonly found among the healthy, first-degree relatives of patients with schizophrenia (for a review in schizophrenia, see 9). Minor physical anomalies are suggested as endophenotypes on account of the findings that in some (5, 10–12, 18, 19), but not all (20–22) studies MPAs were more common in bipolar patients than in healthy controls, MPAs are trait-markers, so they can be detected in remission too.

Only one single study (22) was reported with a very low number of participants (9 siblings of bipolar patients) on the prevalence of minor anomalies in bipolar relatives, where no difference was detected among the healthy first-degree relatives compared to normal controls, so we consider the field to be open for further research.

The aim of this study was to investigate the prevalence of minor physical anomalies - using the Méhes Scale to differentiate minor malformations and phenogenetic variants - in the relatives of patients with bipolar I disorder comparing them to normal control subjects. The following hypotheses have been tested: (1) MPAs are more prevalent in the relatives of bipolar I patients compared to normal controls, which supports the hypothesis, that MPAs can be endophenotypic markers of bipolar I disorder (2) a higher rate of MPAs is found predominantly in the craniofacial regions among the relatives of bipolar I patients, pointing at aberrant early (first and second trimester) neurodevelopment.

MATERIALS AND METHODS

Study Subjects

Using a list of 57 minor physical anomalies collected by Méhes (2), 20 first-degree healthy relatives of patients with the diagnosis of bipolar I disorder were examined. Eleven parents, 3 siblings and 6 children were included in the study, the mean age of the relatives was 57.20 years (standard deviation: 21.53, standard error mean: 4.81). The age of the 6 children were: 44, 47, 35, 42, 42 and 28 years. The sample size followed previous publication (9). As a comparison 20 normal control subjects matched for sex, age and ethnic origin were also observed for minor physical anomalies. Controls were excluded if they endorsed any personal or family history (in the first- or second-degree relatives) of psychotic disorders, mood disorders, personality disorders and anxiety disorders. First-degree relatives of bipolar I patients were excluded if they endorsed a personal history of psychotic disorders, mood disorders, personality disorders and anxiety disorders. For all participants, psychotic disorders, mood disorders, personality disorders and anxiety disorders were ruled out independently by two experienced psychiatrists according to the DSM-5 (23). All clinical information were obtained from structured clinical interviews.

Examination of Minor Physical Anomalies

We have used the Méhes Scale for evaluation of MPAs, which includes 57 minor signs (4, 6, 8–10). MPAs were connected to body regions for comparison. A clear differentiation between minor malformations and phenogenetic variants were introduced, the scale and detailed definitions were published earlier (8). All participants gave informed consent, the study was performed in accordance with the Declaration of Helsinki and was evaluated following institutional guidelines. The Commitee on Medical Ethics of University of Pécs accepted the proposal for the study (No.6416). Two examiners (T. Tényi, Gy. Csábi) investigated all the relatives and controls separately. "The raters were trained by Professor Károly Méhes, and they participated earlier in many minor anomaly studies, they have a long clinical experience in dysmorphology. The examination of minor physical anomalies was done qualitatively (present or absent) without scores being used, but where it was possible, measurements were taken with calipers and tape to improve the objectivity of examination" (9,226.pp).

Statistical Analysis

"Before the statistical analyses interrater reliability was tested and the kappa coefficient was >0.75 for all items. Statistical analyses were done by applying the Mann - Whitney *U*-test for and the chi-squared test for the comparison of the two groups with each other. Two-sided Fisher's exact tests were used to

TABLE 1 | The number of MPAs in the case of relatives and controls.

	0 MPA	1 MPA	2 MPAs	3 MPAs	4 MPAs	5 MPAs	6 MPAs	7 MPAs	8 MPAs
Controls (no. 20)	11	7	2	0	0	0	0	0	0
Relatives (no. 20)	0	2	3	8	3	2	1	0	1

TABLE 2 | Comparison of percentages of body regions in the two goups according to the dichotomization, where the cut point was: (1) at least 1 or more MPAs; (2) 0 MPA.

	Ear region	Head region	Mouth region	Eye region	Trunk region	Hand region	Foot region
Controls (no. 20)	0.00%	10.00%	0.00%	0.00%	25.00%	10.00%	10.00%
Relatives (no. 20)	45.00%	45.00%	70.00%	10.00%	65.00%	35.00%	35.00%
Fisher exact test, level of significance	$P = 0.001$ significant	$P = 0.038$ significant	$P < 0.001$ significant	$P = 0.487$ ns	$P = 0.025$ significant	$P = 0.127$ ns	$P = 0.127$ ns

Fisher-exact-test results.

compare the two groups with each other by body regions, the level of significance chosen was $p < 0.05$. For the analysis of the frequency of each individual minor physical anomaly the two-sided Fisher's exact probability test was used, the level of significance chosen was $p < 0.05$. All the statistical analyses were done by the use of SPSS Version 21" (9,226.pp).

Results

The comparison of two groups with the Mann-Whitney-*U*-test showed significant differences between them (relatives of bipolar I patients: mean rank: 29.8 vs. normal controls: mean rank: 11.2, $p < 0.001$). The differences of the MPA profiles between the two groups is shown on **Table 1**. As in our earlier publications (3, 8, 9) we did a dichotomization by establishing two groups: (1) none or only 1 MPA, (2) MPAs more than 1. While in the control group the number of individuals with none or only with 1 MPA was 18 (90%), this in the relative group was 2 (10%), the chi-squared test showed a statistically significant difference ($p < 0.001$). The significantly different between group data of the 2-sided Fisher's exact test comparisons of percentages according to body regions are shown on **Table 2**. Relatives of bipolar I patients showed a higher frequency of MPAs in the ear, head, mouth and trunk regions compared to normal control subjects. By the differentiation of minor malformations and phenogenetic variants, we have found that phenogenetic variants were more common among the relatives of bipolar I patients compared to the control group (relatives: mean rank: 26.15 vs. controls: mean rank: 14.85, $P = 0.002$), while minor malformations were also more prevalent in the relative group, (relatives: mean rank: 28.43, controls: mean rank: 12.57, $p < 0.001$). Comparing phenogenetic variants by body regions, between the two groups phenogenetic variants in the mouth region were more prevalent (Fisher's exact test, two-sided: 0.003) among the relatives of bipolar I patients. Comparing minor malformations by body regions: ear minor malformations (Fisher's exact test, two-sided: 0.008), trunk minor malformations (Fisher's exact test, two-sided: 0.025) and foot minor malformations (Fisher's exact test, two-sided: 0.047) were more prevalent in the relative group. The results of comparisons of individual MPAs among the two groups are shown on **Table 3**. Only one minor malformation (sole crease) and one phenogenetic variant (high arched palate) were more prevalent ($p = 0.047$) in the bipolar relatives group compared to the normal control group.

TABLE 3 | Individual analysis of MPAs between the two groups, significant differences.

	Relatives, number of individuals	Controls, number of individuals	Ficher exact-test, two-sided
High arched palate	5	0	$P = 0.047$ significant
Sole crease	5	0	$P = 0.047$ significant

DISCUSSION

This is the first report in literature on the increased prevalence of minor physical anomalies among the first-degree unaffected relatives of bipolar I patients. Our results on the overrepresentation of the examined anomalies in the relatives of bipolar I patients support the hypothesis, that MPAs may be endophenotypic markers of bipolar I affective disorder. We report here, that minor physical anomalies were more prevalent in the eye, head, mouth and trunk regions among the relatives of bipolar I patients compared to normal controls and the individual analyses showed, that one minor malformation (sole crease) and one phenogenetic variant (high arched palate) were also more prevalent in the relative group compared to the normal control group. First in literature, we report here on the analyses of MPAs among relatives of bipolar I affective patients by the differentiation of minor malformations and phenogenetic variants, and we emphasize that insults resulting aberrant neurodevelopment may appear both during and after the first and second trimester (both phenogenetic variants and minor malformations were more

prevalent in the relative group, while one minor malformation and one phenogenetic variant were significantly more common among the relatives of patients during the individual analysis). We see as an important result, that relatives of bipolar I patients showed a higher frequency of MPAs in the eye, head and the mouth regions and one phenogenetic variant (high arched palate) was more prevalent in this group of individuals. Previous findings suggested, that anomalies of the head and the mouth may have more relevance to the hypothetical neurodevelopmental failure in patients with several neuropsychiatric disorders (2, 8, 12, 15, 24–26). In our previous study on the higher prevalence of minor physical anomalies among healthy schizophrenia relatives (9), in harmony with the results of Tikka et al. (26), we reported that MPAs in the craniofacial region were significantly higher in the first-degree relative group than the healthy control group.

Limitations

A limitation of our study is the relatively smaller sample size, the replication of our findings on bigger samples seems important. Another caveat of the work, that the relative group is heterogeneous, as 11 parents, 3 siblings and 6 children of bipolar I patients were evaluated, 10% of the relatives (2 children) by chance can develop bipolar disorder later.

To conclude: considering the endophenotype concept (9, 17), it should be reminded that although minor physical anomalies are not specific to bipolar affective disorder and are reported in different neurodevelopmental disorders (2, 14, 27), however our first, pioneering results on the more prevalent appearence of these markers among the relatives of patients with bipolar I affective disorder, can suggest these anomalies as endophenotypic traits, emphasizing at aberrant brain development as an important etiological component in bipolar I disorder. Further studies on the MPA profile of healthy relatives of bipolar I patients, correlating with results from structural brain imaging, can clarify the endophenotypic nature of these somatic markers and the nature of genetic and/or enviromental insults on brain maldevelopment.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Commitee on Medical Ethics of University of Pécs. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

TT, GC, and RH: planning and writing the article. TC, TT, GC, VV, AH, MK, and MS: evaluating the research. MH, AT, and SJ: statistical analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Brain Research Program Grant No. NAP KTIA NAP-A-II/12(2017-2021) and The National Excellence programme (2019-2021, FIKP II.).

REFERENCES

1. Pinsky L. Informative morphogenetic variants. Minor congenital anomalies revisited. *Issues Rev Teratol.* (1985) 3:135–70. doi: 10.1007/978-1-4613-2495-9_5
2. Méhes K. *Informative Morphogenetic Variants in the Newborn*. Budapest: Budapest Akadémiai Kiadó (1988).
3. Tényi T, Trixler M, Csábi Gy, Jeges S. Minor physical anomalies in non-familial unipolar recurrent major depression. *J Aff Disord.* (2004) 79:259–62. doi: 10.1016/S0165-0327(02)00360-9
4. Tényi T, Trixler M, Csábi G. Minor physical anomalies in affective disorders: a review of the literature. *J Aff Disord*. (2009) 112:11–18. doi: 10.1016/j.jad.2008.04.025
5. Akabaliev V, Sivkov S, Mantarkov M. Minor physical anomalies in schizophrenia and bipolar I disorder and the neurodevelopmental continuum of psychosis. *Bipolar Disord.* (2014) 16:633–41. doi: 10.1111/bdi.12211
6. Trixler M, Tényi T, Csábi G, Szabó G, Méhes K. Informative morphogenetic variants in patients with schizophrenia and alcohol-dependent patients: beyond the waldrop scale. *Am J Psychiatry*. (1997) 154:691–3. doi: 10.1176/ajp.154.5.691
7. Trixler M, Tényi T. Problems with the waldrop scale. *Am J Psychiatry.* (2000) 157:486. doi: 10.1176/appi.ajp.157.3.486
8. Trixler M, Tényi T, Csábi G, Szabó R. Minor physical anomalies in schizophrenia and bipolar affective disorder. *Schizop Res.* (2001) 52:195–201. doi: 10.1016/S0920-9964(00)00182-1
9. Hajnal A, Csábi Gy, Herold R, Jeges S, Halmai T, Trixler D, et al. Minor physical anomalies are more common among the first-degree unaffected relatives of schizophrenia patients – results with the méhes scale. *Psychiatry Res.* (2016) 237:224–8. doi: 10.1016/j.psychres.2016.01.036
10. Berecz H, Csábi Gy, Jeges S, Herold R, Simon M, Halmai T, et al. Minor physical anomalies in bipolar I and bipolar II disorders- results with the méhes scale. *Psychiatry Res.* (2017) 249:120–4. doi: 10.1016/j.psychres.2017.01.014
11. Akabaliev V, Sivkov S, Mantarkov M, Ahmed-Popova F. Minor physical anomalies in patients with bipolar I disorder and normal controls. *J Affect Disord.* (2011) 135:193–200. doi: 10.1016/j.jad.2011.07.019
12. Ince B, Altinoz MA, Ayran AC, Cansiz A, Altinbas K, Guloksuz S, et al. Minor physical anomalies in bipolar disorder. *Compr Psychiatry.* (2020) 103:152206. doi: 10.1016/j.comppsych.2020.152206
13. Waldrop MF, Goering JD. Hyperactivity and minor physical anomalies in elementary school children. *Am J Orthop.* (1971) 41:602–7. doi: 10.1111/j.1939-0025.1971.tb03219.x
14. Csábi Gy, Gádoros J, Jeges S, Gyenge E, Trixler M, Tényi T. Minor physical anomalies in tourette syndrome. *Eur J Psychiatry.* (2008) 22:173–80. doi: 10.4321/S0213-61632008000300006
15. Tényi T, Halmai T, Antal A, Benke B, Jeges S, Tényi D, et al. Minor physical anomalies are more common in schizophrenia patients with the history of homicide. *Psychiatry Res.* (2015) 225:702–5. doi: 10.1016/j.psychres.2014.11.012
16. Trixler D, Tényi T. Iris structure and minor physical anomalies in schizophrenia. *Psychiatry Res.* (2017) 256:412–6. doi: 10.1016/j.psychres.2017.07.004

17. Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry*. (2003) 160:636–45. doi: 10.1176/appi.ajp.160.4.636
18. Goswami K, Ram D, Praharaj SK. Minor physical anomalies and neurological signs in bipolar disorders: a comparative study. *Bipolar Disord.* (2011) 13:50.
19. Kelly BD, Lane A, Waddintgon J, Larkin C, O'Callaghan E. Minor physical anomalies in schizophrenia and bipolar disorder. *Schizophr Res.* (2004) 67:139.
20. Alexander RC, Reddy R, Mukherjee S. Minor physical anomalies in schizophrenia. *Biol Psychiatry*. (1992) 31:209A. doi: 10.1016/0006-3223(92)90697-X
21. Alexander RC, Mukherjee S, Richter J, Kaufmann CA. Minor physical anomalies in schizophrenia. *J Nerv Ment Dis.* (1994) 182:639–44. doi: 10.1097/00005053-199411000-00007
22. Green MC, Satz P, Christenson C. Minor physical anomalies in schizophrenia patients, bipolar patients and their siblings. *Schizop Bull.* (1994) 20:433–40. doi: 10.1093/schbul/20.3.433
23. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. St. Washington, DC: American Psychiatric Association (2013).
24. Green MF, Satz P, Gaier DJ, Ganzell S, Kharabi F. Minor physical anomalies in schizophrenia. *Schizophrenia Bull.* (1989) 15:91–9. doi: 10.1093/schbul/15.1.91
25. Lane A, Kinsella A, Murphy P, Byrne M, Keenan J, Colgan K, et al. The antropometric assessment of dysmorphic features in schizophrenia as an index of its developmental origins. *Psychological Med.* (1997) 27:1155–64. doi: 10.1017/S0033291797005503
26. Tikka SK, Nizamie SH, Goyal N, Pradhan N, Tikka DL, Katshu MZUH. Evaluation of spontaneous dense array gamma oscillatory activity and minor physical anomalies as a composite neurodevelopmental endophenotype in schizophrenia. *Int J Dev Neurosci.* (2015) 40:43–51. doi: 10.1016/j.ijdevneu.2014.11.002
27. Compton MT, Chan RCK, Walker EF, Buckley PF. Minor physical anomalies: potentially informative vestiges of fetal developmental disruptions in schizophrenia. *Int J Dev Neurosci.* (2011) 29:245–50. doi: 10.1016/j.ijdevneu.2010.10.006

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 09 July 2021
doi: 10.3389/fgene.2021.647246

Relationships between Complement Protein Concentration and Genotype in First-episode Psychosis: Evidence from C3 and C4 in Peripheral Blood

Yu Chen[1,2], Zhenguo Zhao[1,2], Fen Lin[1,2], Lifang Wang[1,2], Zheng Lin[3] and Weihua Yue[1,2,4]**

[1] Institute of Mental Health, Peking University Sixth Hospital, Beijing, China, [2] Key Laboratory of Mental Health, National Clinical Research Center for Mental Disorders, Chinese Academy of Medical Sciences Research Unit (No. 2018RU006), Ministry of Health (Peking University), Beijing, China, [3] Second Hospital Zhejiang University School of Medicine (SAHZU), Hangzhou, China, [4] PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, China

OPEN ACCESS

Edited by:
Elena Martín-García, Pompeu Fabra University, Spain

Reviewed by:
Qiang Wang, Sichuan University, China
Hongsheng Gui, Henry Ford Health System, United States

***Correspondence:**
Weihua Yue
dryue@bjmu.edu.cn
Zheng Lin
linzzr@126.com

Specialty section:
This article was submitted to Behavioral and Psychiatric Genetics, a section of the journal Frontiers in Genetics

Received: *29 December 2020*
Accepted: *04 June 2021*
Published: *09 July 2021*

Citation:
Chen Y, Zhao Z, Lin F, Wang L, Lin Z and Yue W (2021) Relationships between Complement Protein Concentration and Genotype in First-episode Psychosis: Evidence from C3 and C4 in Peripheral Blood. Front. Genet. 12:647246. doi: 10.3389/fgene.2021.647246

Schizophrenia is a common neuropsychiatric disorder with complex pathophysiology. Recent reports suggested that complement system alterations contributed to pathological synapse elimination that was associated with psychiatric symptoms in schizophrenia. Complement component 3 (C3) and complement component 4 (C4) play central roles in complement cascades. In this study, we compared peripheral C3 and C4 protein levels between first-episode psychosis (FEP) and healthy control (HC). Then we explored whether single nucleotide polymorphisms (SNPs) at C3 or C4 genes affect peripheral C3 or C4 protein levels. In total, 181 FEPs and 204 HCs were recruited after providing written informed consent. We measured serum C3 and C4 protein levels using turbidimetric inhibition immunoassay and genotyped C3 and C4 polymorphisms using the Sequenom MassArray genotyping. Our results showed that three SNPs were nominally associated with schizophrenia (rs11569562/C3: A > G, $p = 0.048$; rs2277983/C3: A > G, $p = 0.040$; rs149898426/C4: G > A, $p = 0.012$); one haplotype was nominally associated with schizophrenia, constructed by rs11569562–rs2277983–rs1389623 (GGG, $p = 0.048$); FEP had higher serum C3 and C4 (both $p < 0.001$) levels than HC; rs1389623 polymorphisms were associated with elevated C3 levels in our meta-analysis (standard mean difference, 0.50; 95% confidence interval, 0.30 to 0.71); the FEP with CG genotype of rs149898426 had higher C4 levels than that with GG genotypes ($p = 0.005$). Overall, these findings indicated that complement system altered in FEP and rs149898426 of C4 gene represented a genetic risk marker for schizophrenia likely through mediating complement system. Further studies with larger sample sizes needs to be validated.

Keywords: first-episode psychosis, complement protein concentration, single nucleotide polymorphism, Chinese Han population, case-control studies

HIGHLIGHTS

- Peripheral C3 and C4 protein levels were increased in first-episode psychosis (FEP).
- The rs149898426 polymorphisms of C4 gene were associated with schizophrenia.
- The CG genotype of rs149898426 were associated with higher serum C4 level in FEP comparing with GG genotype.

INTRODUCTION

Schizophrenia is a severe and complicated neuropsychiatric disorder characterized by hallucinations, delusions and cognitive dysfunction (Kahn and Keefe, 2013; Kahn et al., 2015; Owen et al., 2016). While many studies attributed the causes of schizophrenia to biological or environmental factors, the etiology of schizophrenia are still unclear. Several lines of evidence suggested that dysregulation of the immune system contributed to the development of schizophrenia. Epidemiological studies found that infection and autoimmune disorders were associated with schizophrenia (Brown and Derkits, 2010; Benros et al., 2011, 2014; Arias et al., 2012; Khandaker et al., 2013). Besides, many cross-sectional studies reported that proinflammatory cytokines and C-reactive protein (CRP) were increased in schizophrenia compared with healthy control (HC) (Miller et al., 2011; Fineberg and Ellman, 2013; Fernandes et al., 2016). Lastly, antipsychotics has been demonstrated to have immunomodulatory effects, for example, clozapine (Hinze-Selch et al., 1998; Roge et al., 2012). Thus, immune dysregulation could represent a vulnerability factor for schizophrenia.

Complement system, an important part of the innate immune system, play important roles in clearing microbes and damaged cells from an organism, triggering inflammation and destroying foreign invaders (Merle et al., 2015a,b). Recent evidence identified that complement system orchestrated the balance for neurodevelopmental processes (Stephan et al., 2012; Ricklin et al., 2016). For example, overactivation of complement system contributed to neurotoxicity and pathological synapse loss through microglia, leading to progressive dysfunction in Alzheimer disease (AD) (Shi et al., 2017; Heneka et al., 2018).

Complement component 3 (C3) and complement component 4 (C4) play dominant roles in complement cascades. The C3 gene is located on chromosome 19, consisting of 41 exons and the mRNA has 5101 bp. C3 is a convergent point for activation both classical and alternative complement pathways (Merle et al., 2015a,b). Genetic evidence suggests that the polymorphisms of C3 were associated with risk of schizophrenia (Rudduck et al., 1985; Fañanás et al., 1992; Ni et al., 2015; Zhang et al., 2018). And patients with schizophrenia had increased C3 protein levels in peripheral blood (Santos Sória et al., 2012; Ali et al., 2017) and serum C3 concentrations were positively correlated with PANSS scores (Li et al., 2016). The C4 gene has two isoforms, C4A and C4B, located within the major histocompatibility complex (MHC) region on chromosome 6. C4A was proved to increase the risk for schizophrenia through mediating synapse pruning during postnatal development (Sekar et al., 2016).

The elucidation of pathophysiology in the early stages of schizophrenia may help understand disease etiology. First-episode psychosis (FEP) refers to the first time a person outwardly experiences symptoms of psychosis. Although some symptoms were unspecific in FEP, the underlying biological processes has changed, especially upregulating inflammatory status. Given that C3 and C4 are major plasma proteins of the complement pathway and are widely measured parameters during clinical practice, we investigated whether C3 or C4 protein levels are different between FEP and HC from our clinical data. Considering previous reports that C3 or C4 gene has multiple single nucleotide polymorphisms (SNPs) that were at the genome-wide level (Yang et al., 2012), we selected several SNPs at C3 (rs11569562, rs2277983, rs1389623) and C4 (rs406658, rs2746414, rs149898426) genes to explore the associations between genotype and complement protein concentration in FEP.

MATERIALS AND METHODS

Subjects

A total of 181 FEPs (94 males, 87 females; mean age: 29.9 ± 10.0 years) and 204 HCs (96 males, 108 females; mean age: 30.0 ± 10.2 years) were recruited from Peking University Sixth Hospital. The diagnosis was made according to the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) criteria using a Structured Clinical Interview and people with FEP did not take medication regularly. We excluded HC with a history of mental health or neurological diseases from communities through simple unstructured interviews conducted by psychiatrists. All participants were given written informed consent before the study began. The study was conducted under established ethical standards and was approved by the ethics committee of Peking University Sixth Hospital.

SNP Selection and Genotyping

We collected 5 mL venous blood from each subject, separated out serum by centrifugation, and stored it at −80°C until laboratory analysis. The genomic DNA was extracted from using the QIAamp DNA Blood Mini Kit. We selected 3 haplotype-tagging single nucleotide polymorphisms (tagSNPs) by using the UCSC (GRCh37/hg19) genome browser[1] for C3 and C4, respectively (C3: rs11569562, rs2277983, rs1389623; C4: rs406658, rs2746414, rs149898426). Genotyping was conducted for 6 tagSNPs by using the platform of the Sequenom MassArray system (Sequenom, San Diego, CA, United States). Locus-specific polymerase chain reaction (PCR) and detection primers were designed with the MassARRAY Assay Design Version 3.0 software (Sequenom, San Diego, CA, United States). The DNA samples were amplified by multiplex PCR reactions and the obtained products were used for locus-specific single-base extension reaction. The final products were desalted and transferred to 384-element SpectroCHIP arrays. Allele typing was performed using MALDI-TOF MS spectroscopy. The mass spectrum was analyzed by

[1] http://genome.ucsc.edu/

the MassARRAY TYPER software (Sequenom, San Diego, CA, United States). All DNA samples were analyzed in technical duplicates. We genotyped each sample thrice to minimize genotyping errors and only consensus genotypes were processed for further analysis. Genotyping primer sequences are indicated in **Table 1**.

TABLE 1 | SNP information and PCR and extension primer sequences from Sequenom SNP genotyping.

SNP	Position	Location	Second-PCRP	First-PCRP	UEP_SEQ
rs11569562	6678742	Intron	ACGTTGGATGTACTATGCAATGAGAGTGAC	ACGTTGGATGTGACTGTAGTTTTCCGTGGC	CCATGAGGCTACAGTATT
rs2277983	6679552	Intron	ACGTTGGATGGAAGATGAGAGGATAAGGGC	ACGTTGGATGCAGGTCTGAAGACTGAGAAC	CTCCCTCCAAAGACCA
rs1389623	6684186	Intron	ACGTTGGATGATGATTGTGTACTTTCCTCC	ACGTTGGATGTGTGAGACCACAGCAATGAC	AGCAATGACCACGTAAG
rs406658	32028747	Coding sequence	ACGTTGGATGTTGAAGGTCCTGAGTTTGGC	ACGTTGGATGGGACAGAAGCCAGTTAGATG	CAGTTTCTCAGGCGA
rs2746414	32029189	Coding sequence	ACGTTGGATGGTTTTCTTCCAGGAAGCCTC	ACGTTGGATGAGCCCAGCACTTGCTTTCTC	TCCATCTCAAAGGCAA
rs149898426	32029352	intron	ACGTTGGATGCAACAACCTCATGGCAATGG	ACGTTGGATGCACTCAGGGATCCTAAGGTC	GCTCACTGCCAGAGC

TABLE 2 | Comparison of genotype and allele distribution of 6 SNPs of C3 and C4 gene between FEP and HC.

Makers/gene	Genotype *N* (Freq.)				Chi-square (df = 2)	*p*-Value	HWE *P*	Allele *N* (freq.)		Chi-square (df = 1)	*p*-Value	OR (95% CI)
rs11569562/C3	A/A	A/G	G/G	NA				A	G			
FEP	54 (0.298)	89 (0.492)	38 (0.210)		4.325	0.114	0.905	197 (0.544)	165 (0.456)	3.886	0.048[a]	0.751 (0.566–0.999)
HC	43 (0.211)	107 (0.525)	54 (0.265)				0.457	193 (0.473)	215 (0.527)			
rs2277983/C3	A/A	A/G	G/G					A	G			
FEP	54 (0.298)	90 (0.497)	37 (0.204)		4.484	0.106	0.964	198 (0.547)	164 (0.453)	4.199	0.040[a]	0.743 (0.559–0.987)
HC	43 (0.212)	106 (0.522)	54 (0.266)	1			0.499	192 (0.473)	214 (0.527)			
rs1389623/C3	A/A	A/G	G/G					A	G			
FEP	3 (0.017)	41 (0.227)	137 (0.757)		0.178	0.915	0.973	47 (0.130)	315 (0.870)	0.040	0.841	1.044 (0.683–1.596)
HC	4 (0.020)	43 (0.211)	157 (0.767)				0.603	51 (0.125)	357 (0.875)			
rs406658/C4	A/A	A/C	C/C					A	C			
FEP	1 (0.006)	33 (0.189)	141 (0.806)	6	1.939	0.379	0.529	35 (0.100)	315 (0.900)	1.480	0.224	0.736 (0.468–1.158)
HC	3 (0.015)	47 (0.233)	152 (0.752)	2			0.769	53 (0.131)	351 (0.869)			
rs2746414/C4	A/A	A/G	G/G					A	G			
FEP	0 (0)	29 (0.160)	152 (0.840)		1.977	0.372	0.241	29 (0.080)	333 (0.920)	0.479	0.489	0.811 (0.490–1.341)
HC	2 (0.001)	35 (0.174)	164 (0.816)	3			0.931	39 (0.097)	363 (0.903)			
rs149898426/C4	C/C	C/G	G/G					C	G			
FEP	0 (0)	28 (0.156)	152 (0.844)	1	7.775	0.021[a]	0.258	28 (0.078)	332 (0.922)	6.344	0.012[b]	0.527 (0.325–0.853)
HC	3 (0.016)	47 (0.245)	142 (0.740)	12			0.690	53 (0.138)	331 (0.862)			

FEP, first-episode psychosis; HC, healthy control; Freq., frequencies; HWE P, The p-value for Hardy-Weinberg equilibrium tests; OR, odds ratio; CI, confidence interval; NA indicates a missing value.

[a]*Denotes a significant association ($P < 0.05$).*

[b]*Denotes a significant association after Bonferroni adjustment ($P < 0.05/3 = 0.017$).*

Measurement of Complement Protein

Serum C3 and C4 levels were measured using the turbidimetric inhibition immunoassay in an automatic chemistry analyzer, ROCHE (Roche Molecular Systems, Inc., Basel, Switzerland) module Cobas 8000 (C702). The assays were performed strictly by the manufacturer's recommended protocol. All samples were analyzed in duplicates and the mean of the two measurements was used in the analysis.

Statistical Analyses

Pairwise linkage disequilibrium (LD), haplotype construction, and genetic association analysis were performed by using Haploview 4.2 software[2]. Deviation of the genotype and allele frequency from the Hardy–Weinberg equilibrium were analyzed using a chi-square goodness-of-fit test. We conducted *t*-tests and a univariate analysis of covariance (ANCOVA) to examine the difference of serum complement protein levels (C3 and C4) between FEP and HC, considering that age and sex may affect serum complement protein levels. A two-way analysis of variance (ANOVA) was used to explore the interaction effects of group and genotype. *Post hoc* analyses were performed using a Tukey test. When the result from *F*-test was marginal significant ($0.05 < p < 0.1$), we would conduct a meta-analysis of the SNP for case group and control group to increase the statistical power. Statistical analyses were performed using R statistical software[3]. Data were expressed as the mean ± standard deviation (SD) or as number (proportion). Results were considered nominally significant at $p < 0.05$ (two-tailed). Multiple testing adjustment will be controlled via Bonferroni adjustment for comparison of genotype and allele distribution for C3 and C4 gene. $P < 0.05/3$ was considered as significant for the reason that the total number of independence tests (excluding LD) was 3 for all 6 SNPs across two genes.

RESULTS

Single Polymorphism Analysis

We observed no association of rs11569562, rs1389623, s406658, and rs2746414 polymorphism with schizophrenia for genotypes and alleles. There were two nominally significant associations between rs11569562 ($p = 0.048$) and rs2277983 ($p = 0.040$) alleles distributions and schizophrenia but no association between genotypes and schizophrenia. The genotypes ($p = 0.0205$) and alleles ($p = 0.012 < 0.05/3$) of rs149898426 were significantly different between FEP and HC. The genotypes and alleles frequencies of each polymorphism were presented in **Table 2**.

Haplotype-Based Association Analysis

The LD map and block structures of C3 and C4 polymorphisms were shown in **Figure 1**. A haplotype block of C3 was constructed by rs11569562, rs2277983, and rs1389623: GGG haplotype was significantly associated with schizophrenia ($\chi^2 = 3.922$, $p = 0.048$); while after 10,000 times of permutation tests, the difference of haplotype GGG frequencies between FEP and HC were not significant (empirical *p*-value was 0.104) (see details in **Table 3** and **Figure 1A**). No haplotype blocks were found in C4 gene (**Figure 1B**).

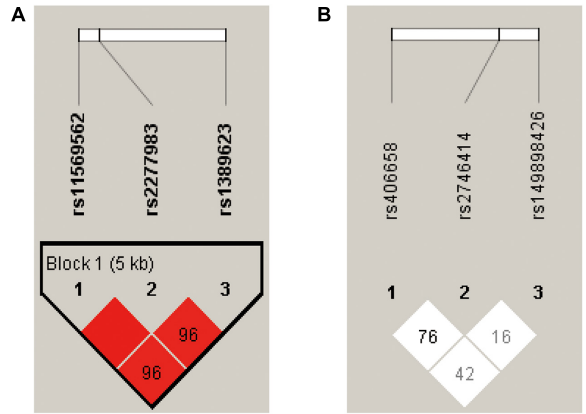


FIGURE 1 | The pairwise linkage disequilibrium (LD) analysis was applied to detect the inter-marker relationship of the three SNPs at the C3 and C4 gene receptively, by using the *D'* values. According to the *D'* values in this figure, **(A)** three SNPs constructed into one haplotype block in C3 gene. **(B)** No haplotype block was detected in C4 gene.

Complement Protein Levels

One sample testing C3 protein levels was excluded for being three SDs above the mean. The serum C3 protein levels (mean ± SD) in cases and controls were 1.18 ± 0.22 and 1.07 ± 0.20 g/L; the serum C4 protein levels (mean ± SD) in cases and controls were 0.28 ± 0.08 and 0.25 ± 0.08 g/L, respectively. *T*-tests showed that FEP had higher serum C3 ($t = 5.11$, $p < 0.001$, Cohen's $d = 0.52$) and C4 ($t = 4.21$, $p < 0.001$, Cohen's $d = 0.44$) protein levels than HC (**Figures 2A,B**). Although the differences of serum C3 and C4 protein levels were not significant considering age and sex in cases and controls, respectively (see details in **Table 4**), age- and sex-adjusted comparisons were performed with ANCOVA: the differences of serum C3 ($F = 26.55$, $p < 0.001$, $\eta^2 = 0.065$) and C4 ($F = 18.27$, $p < 0.001$, $\eta^2 = 0.046$) protein levels remained significant.

TABLE 3 | Comparison of haplotype frequencies at the C3 gene between FEP and HC.

Haplotypes	Freq. case	Freq. control	Chi-square	*p*-Value	Empirical *p*-Value
Block (rs11569562–rs2277983–rs1389623)					
GGG	0.524	0.452	3.922	0.048	0.104
AAG	0.351	0.415	3.320	0.069	0.158
AAA	0.125	0.129	0.092	0.761	1.000

FEP, first-episode psychosis; HC, healthy control; Freq., frequency. Haplotypes with frequency < 0.01 in FEP and HC were dropped. Empirical p-value: the p-value of the permutation tests (number of iterations, 10,000).

[2]http://www.broadinstitute.org/haploview/

[3]https://www.r-project.org

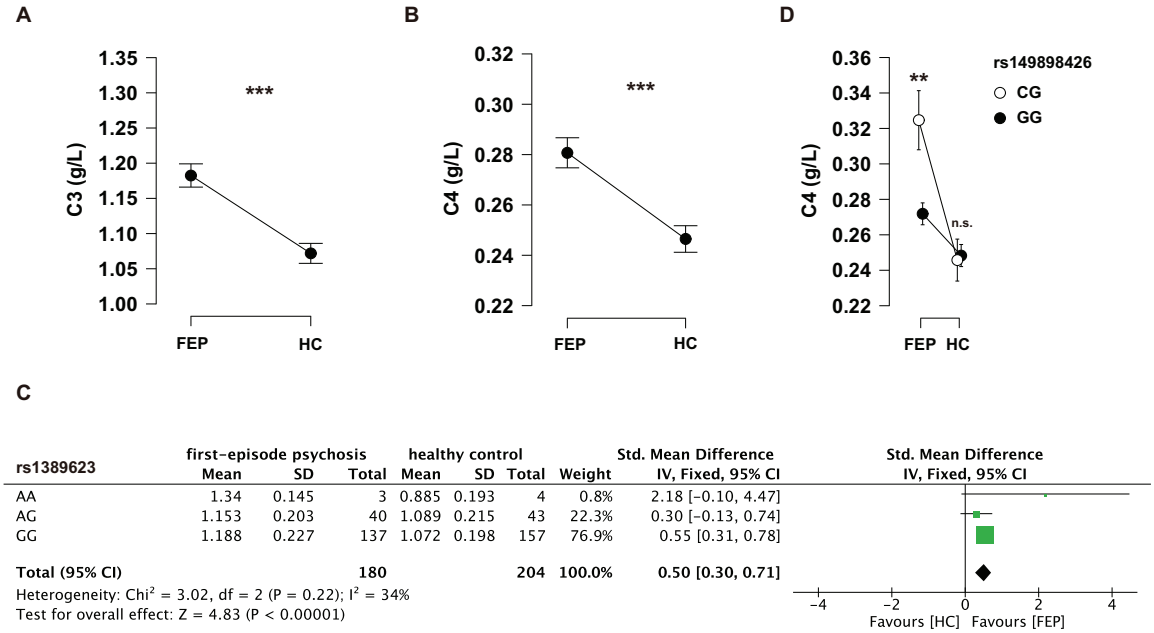


FIGURE 2 | (A) Serum C3 protein levels between FEP and HC. **(B)** Serum C4 protein levels between FEP and HC. Data were analyzed by Student's *t*-tests. **(C)** Serum C4 levels among the genotypes of rs149898426 (CG and GG) in FEP and HC receptively. Data were analyzed by a two-way ANOVA with Tukey *post hoc* tests. Data shown in the figure represent mean ± SEM. $^{**}P < 0.01$; $^{***}P < 0.001$; n.s.: non-significant. **(D)** A meta-analysis of the genotypes of rs1389623 on serum C3 levels.

TABLE 4 | Subgroups of serum C3 and C4 levels according to sex and clinical information.

	Group	Sex	N	Mean ± SD (g/L)	*P*
C3	FEP	Male	94	1.16 ± 0.24	0.44
		Female	86	1.20 ± 0.20	
	HC	Male	96	1.07 ± 0.22	0.77
		Female	108	1.08 ± 0.19	
C4	FEP	Male	94	0.28 ± 0.09	0.45
		Female	87	0.27 ± 0.07	
	HC	Male	96	0.24 ± 0.08	0.68
		Female	108	0.24 ± 0.07	

FEP, first-episode psychosis; HC, healthy control. One sample information (C3 for Female in cases) was lost for technical errors.

SNP-Group Interactions in Complement Proteins

Serum C3 and C4 protein levels of individual genotypes were evaluated between the case and control groups using two-way ANOVAs. We found that the interaction effects of group (FEP vs. HC) and rs1389623 were marginal significant ($F = 2.81$, $p = 0.062$, partial $\eta^2 = 0.015$) on serum C3 levels (see **Supplementary Tables 1–2**). A meta-analysis showed that rs1389623 polymorphism were associated with elevated serum C3 levels ($Z = 4.83$, $p < 0.001$) (see details in **Figure 2C**). There was no interaction effect between other genotypes and group on serum C3 protein levels (see **Supplementary Tables 3–6**). Besides, we excluded all CC genotypes in a two-way ANOVAs (group: FEP, HC × rs149898426 polymorphisms: CG, GG) on C4 protein levels, for the reason that there were three subjects with CC genotypes in HC and no subject in FEP. The interaction effects of group and rs149898426 were significant ($F = 7.33$, $p = 0.007$, partial $\eta^2 = 0.02$) (see **Supplementary Tables 7–8**). *Post hoc* tests found that FEP who carry the CG genotypes of rs149898426 had higher serum C4 protein levels than the patients who carry the GG genotypes ($t = 3.34$, $p_{tukey} = 0.005$) while HC not ($t = -0.198$, $p_{tukey} = 0.997$) (**Figure 2D**). No other interaction of genotypes and group was significant in serum C4 protein levels (see **Supplementary Tables 9–12**).

DISCUSSION

The present study confirmed that complement system overactivated in FEP, which were correlated with genetic factors. We observed that FEP had higher serum C3 and C4 protein levels than HC, suggesting that aberrant expression of complement proteins may be potential biomarkers for schizophrenia; C4 polymorphism affected serum C4 protein levels in FEP but not in HC, suggesting that genotype may be considered as an important risk factor for the development of schizophrenia.

The first finding was that FEP had increased serum C3 and C4 protein levels compared with HC. The results replicated many previous findings in schizophrenia (Maes et al., 1997) and extended to patients that were not FEP. But two other studies had no similar findings in FEP (Kopczynska et al., 2019;

Laskaris et al., 2019). One possible reason is that the sample size may not be sufficient to reach the statistical differences; both studies had lower sample sizes than our study. Another potential cause for this divergence of outcomes is race or ethnicity. In our study, study participants are limited to the Chinese Han population, while various racial groups were involved in other studies. However, one study pointed that individuals at ultra-high risk (UHR) for psychosis had elevated C3 and C4 protein levels (Laskaris et al., 2019), suggesting that alteration of complement system accompanied the development of schizophrenia. These results should be considered exploratory and further studies with larger cohorts would be required to confirm results.

To our knowledge, this is the first survey that compares serum C3 and C4 levels with their SNP in FEP of the China Han population. Two recent papers investigated the roles of C3 polymorphism in susceptibility to schizophrenia in the China Han population (Ni et al., 2015; Zhang et al., 2018). Although SNP rs11569562 in C3 was not associated with schizophrenia in the Chinese Han population (Zhang et al., 2018) consisting of 1086 patients with schizophrenia and 1154 HCs, patients with all genotypes of rs11569562 have higher serum C3 levels than the controls similar to our study. Other studies using expression quantitative trait locus (eQTL) analysis revealed that rs2277984, which is adjacent to rs2277983 in our study, regulates C3 expression in the liver (Zhang et al., 2017), suggesting that SNPs in the C3 gene affected the expression of C3 genes to further impact the C3 protein translation. Similarly, one study found that in the region near C4 of chromosome 6, the more strongly an SNP associated with schizophrenia, the more strongly it correlated with predicted C4A expression (Sekar et al., 2016). In short, the genotype of the complement gene had effects on its expression, which concurs with our results.

The present study has a major limitation that the sample size was small and may underpower the whole study in terms of C3 and C4 genotype distribution. Therefore, our results should be interpreted with caution. Further studies with larger sample sizes will be required to achieve sufficient statistical power and elucidate a potential link between complement system and schizophrenia in China Han population.

CONCLUSION

Our results showed that first-episode psychosis had higher serum C3 and C4 protein levels than healthy control. C4 gene polymorphisms affected C4 protein levels in peripheral blood: the FEP carrying CG genotypes of rs149898426 had higher serum C4 protein levels than that of GG genotypes. The available evidence that complement proteins were elevated in FEP and polymorphisms of complement genes played a contributing role in regulating peripheral complement proteins concentration.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the Peking University Open Research Data repository, accession number 10635 (http://opendata.pku.edu.cn/api/access/datafile/10635).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Peking University Sixth Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

WY provided the funds and designed the study. YC analyzed the data and wrote the draft of the manuscript. ZZ recruited schizophrenia patients and collected peripheral blood samples. Primers were designed by FL and LW. ZL supervised this study. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from the National Key R&D Program of China (2016YFC1307000), National Natural Science Foundation of China (81825009, 81901358, and 81221002), Peking University Clinical Scientist Program supported by "the Fundamental Research Funds for the Central Universities" (BMU2019LCKXJ012), and Academy of Medical Sciences Research Unit (2019-I2M-5-006). The findings had no further role in study design; in the collection, analysis, and interpretation of data; in the writing of the report, and in the decision to submit the manuscript for publication.

ACKNOWLEDGMENTS

We thank the participants who were involved in the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.647246/full#supplementary-material

REFERENCES

Ali, F. T., Abd El-Azeem, E. M., Hamed, M. A., Ali, M. A. M., Abd Al-Kader, N. M., and Hassan, E. A. (2017). Redox dysregulation, immuno-inflammatory alterations and genetic variants of BDNF and MMP-9 in schizophrenia: pathophysiological and phenotypic implications. *Schizophr. Res.* 188, 98–109. doi: 10.1016/j.schres.2017.01.016

Arias, I., Sorlozano, A., Villegas, E., de Dios Luna, J., McKenney, K., Cervilla, J., et al. (2012). Infectious agents associated with schizophrenia: a meta-analysis. *Schizophr. Res.* 136, 128–136. doi: 10.1016/j.schres.2011.10.026

Benros, M. E., Nielsen, P. R., Nordentoft, M., Eaton, W. W., Dalton, S. O., and Mortensen, P. B. (2011). Autoimmune diseases and severe infections as risk factors for Schizophrenia: a 30-year population-based register study. *Am. J. Psychiatry* 168, 1303–1310. doi: 10.1176/appi.ajp.2011.11030516

Benros, M. E., Pedersen, M. G., Rasmussen, H., Eaton, W. W., Nordentoft, M., and Mortensen, P. B. (2014). A Nationwide Study on the Risk of Autoimmune Diseases in Individuals With a Personal or a Family History of Schizophrenia and Related Psychosis. *Am. J. Psychiatry* 171, 218–226. doi: 10.1176/appi.ajp.2013.13010086

Brown, A. S., and Derkits, E. J. (2010). Prenatal infection and Schizophrenia: a review of epidemiologic and translational studies. *Am. J. Psychiatry* 167, 261–280. doi: 10.1176/appi.ajp.2009.09030361

Fañanás, L., Moral, P., Panadero, M. A., and Bertranpetit, J. (1992). Complement genetic markers in schizophrenia: C3, BF and C6 polymorphisms. *Hum. Hered.* 42, 162–167. doi: 10.1159/000154060

Fernandes, B. S., Steiner, J., Bernstein, H. G., Dodd, S., Pasco, J. A., Dean, O. M., et al. (2016). C-reactive protein is increased in schizophrenia but is not altered by antipsychotics: meta-analysis and implications. *Mol. Psychiatry* 21, 554–564. doi: 10.1038/mp.2015.87

Fineberg, A. M., and Ellman, L. M. (2013). Inflammatory cytokines and neurological and neurocognitive alterations in the course of Schizophrenia. *Biol. Psychiatry* 73, 951–966. doi: 10.1016/j.biopsych.2013.01.001

Heneka, M. T., McManus, R. M., and Latz, E. (2018). Inflammasome signalling in brain function and neurodegenerative disease. *Nat. Rev. Neurosci.* 19, 610–621.

Hinze-Selch, D., Becker, E. W., Stein, G. M., Berg, P. A., Mullington, J., Holsboer, F., et al. (1998). Effects of clozapine on in vitro immune parameters: a longitudinal study in clozapine-treated schizophrenic patients. *Neuropsychopharmacology* 19, 114–122.

Kahn, R. S., and Keefe, R. S. E. (2013). Schizophrenia is a cognitive illness time for a change in focus. *JAMA Psychiatry* 70, 1107–1112. doi: 10.1001/jamapsychiatry.2013.155

Kahn, R. S., Sommer, I. E., Murray, R. M., Meyer-Lindenberg, A., Weinberger, D. R., Cannon, T. D., et al. (2015). Schizophrenia. *Nat. Rev. Dis. Primers* 1:15067. doi: 10.1038/nrdp.2015.67

Khandaker, G. M., Zimbron, J., Lewis, G., and Jones, P. B. (2013). Prenatal maternal infection, neurodevelopment and adult schizophrenia: a systematic review of population-based studies. *Psychol. Med.* 43, 239–257. doi: 10.1017/s0033291712000736

Kopczynska, M., Zelek, W., Touchard, S., Gaughran, F., Di Forti, M., Mondelli, V., et al. (2019). Complement system biomarkers in first episode psychosis. *Schizophr. Res.* 204, 16–22. doi: 10.1016/j.schres.2017.12.012

Laskaris, L., Zalesky, A., Weickert, C. S., Di Biase, M. A., Chana, G., Baune, B. T., et al. (2019). Investigation of peripheral complement factors across stages of psychosis. *Schizophr. Res.* 204, 30–37. doi: 10.1016/j.schres.2018.11.035

Li, H., Zhang, Q., Li, N., Wang, F., Xiang, H., Zhang, Z., et al. (2016). Plasma levels of Th17-related cytokines and complement C3 correlated with aggressive behavior in patients with schizophrenia. *Psychiatry Res.* 246, 700–706. doi: 10.1016/j.psychres.2016.10.061

Maes, M., Delange, J., Ranjan, R., Meltzer, H. Y., Desnyder, R., Cooremans, W., et al. (1997). Acute phase proteins in schizophrenia, mania and major depression: modulation by psychotropic drugs. *Psychiatry Res.* 66, 1–11. doi: 10.1016/s0165-1781(96)02915-0

Merle, N. S., Church, S. E., Fremeaux-Bacchi, V., and Roumenina, L. T. (2015a). Complement system part I - molecular mechanisms of activation and regulation. *Front. Immunol.* 6:262. doi: 10.3389/fimmu.2015.00262

Merle, N. S., Noe, R., Halbwachs-Mecarelli, L., Fremeaux-Bacchi, V., and Roumenina, L. T. (2015b). Complement system part II: role in immunity. *Front. Immunol.* 6:257. doi: 10.3389/fimmu.2015.00257

Miller, B. J., Buckley, P., Seabolt, W., Mellor, A., and Kirkpatrick, B. (2011). Meta-analysis of cytokine alterations in Schizophrenia: clinical status and antipsychotic effects. *Biol. Psychiatry* 70, 663–671. doi: 10.1016/j.biopsych.2011.04.013

Ni, J., Hu, S., Zhang, J., Tang, W., Lu, W., and Zhang, C. (2015). A preliminary genetic analysis of complement 3 gene and Schizophrenia. *PLoS One* 10:e0136372. doi: 10.1371/journal.pone.0136372

Owen, M. J., Sawa, A., and Mortensen, P. B. (2016). Schizophrenia. *Lancet* 388, 86–97. doi: 10.1016/s0140-6736(15)01121-6

Ricklin, D., Reis, E. S., and Lambris, J. D. (2016). Complement in disease: a defence system turning offensive. *Nat. Rev. Nephrol.* 12, 383–401. doi: 10.1038/nrneph.2016.70

Roge, R., Moller, B. K., Andersen, C. R., Correll, C. U., and Nielsen, J. (2012). Immunomodulatory effects of clozapine and their clinical implications: What have we learned so far? *Schizophr. Res.* 140, 204–213. doi: 10.1016/j.schres.2012.06.020

Rudduck, C., Beckman, L., Franzén, G., and Lindström, L. (1985). C3 and C6 complement types in schizophrenia. *Hum. Hered.* 35, 255–258. doi: 10.1159/000153555

Santos Sória, L. D., Moura Gubert, C. D., Ceresér, K. M., Gama, C. S., and Kapczinski, F. (2012). Increased serum levels of C3 and C4 in patients with schizophrenia compared to Eutymic patients with bipolar disorder and healthy. *Rev. Bras. Psiquiatr.* 34, 119–120.

Sekar, A., Bialas, A. R., de Rivera, H., Davis, A., Hammond, T. R., Kamitaki, N., et al. (2016). Schizophrenia risk from complex variation of complement component 4. *Nature* 530, 177–183. doi: 10.1038/nature16549

Shi, Q., Chowdhury, S., Ma, R., Le, K. X., Hong, S., Caldarone, B. J., et al. (2017). Complement C3 deficiency protects against neurodegeneration in aged plaque-rich APP/PS1 mice. *Sci. Transl. Med.* 9:eaaf6295. doi: 10.1126/scitranslmed.aaf6295

Stephan, A. H., Barres, B. A., and Stevens, B. (2012). The complement system: an unexpected role in synaptic pruning during development and disease. *Annu. Rev. Neurosci.* 35, 369–389.

Yang, X., Sun, J., Gao, Y., Tan, A., Zhang, H., Hu, Y., et al. (2012). Genome-wide association study for serum complement C3 and C4 levels in healthy Chinese subjects. *PLoS Genet.* 8:e1002916. doi: 10.1371/journal.pgen.1002916

Zhang, C., Zhang, Y., Cai, J., Chen, M., and Song, L. (2017). Complement 3 and metabolic syndrome induced by clozapine: a cross-sectional study and retrospective cohort analysis. *Pharmacogenomics J.* 17, 92–97. doi: 10.1038/tpj.2015.68

Zhang, S., Zhou, N., Liu, R., Rao, W., Yang, M., Cao, B., et al. (2018). Association between polymorphisms of the complement 3 gene and Schizophrenia in a Han Chinese population. *Cell. Physiol. Biochem.* 46, 2480–2486. doi: 10.1159/000489654

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer QW declared a past co-authorship with the authors LW and WY to the handling editor.

frontiers in Psychiatry

ORIGINAL RESEARCH
published: 29 July 2021
doi: 10.3389/fpsyt.2021.688464

Predicting Complex Traits and Exposures From Polygenic Scores and Blood and Buccal DNA Methylation Profiles

Veronika V. Odintsova[1†] Valerie Rebattu[1†], Fiona A. Hagenbeek[1], René Pool[1], Jeffrey J. Beck[2], Erik A. Ehli[2], Catharina E. M. van Beijsterveldt[1], Lannie Ligthart[1], Gonneke Willemsen[1], Eco J. C. de Geus[1], Jouke-Jan Hottenga[1], Dorret I. Boomsma[1] and Jenny van Dongen[1*] on behalf of BIOS consortium[‡]*

[1] Department of Biological Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, [2] Avera Institute for Human Genetics, Sioux Falls, SD, United States

We examined the performance of methylation scores (MS) and polygenic scores (PGS) for birth weight, BMI, prenatal maternal smoking exposure, and smoking status to assess the extent to which MS could predict these traits and exposures over and above the PGS in a multi-omics prediction model. MS may be seen as the epigenetic equivalent of PGS, but because of their dynamic nature and sensitivity of non-genetic exposures may add to complex trait prediction independently of PGS. MS and PGS were calculated based on genotype data and DNA-methylation data in blood samples from adults (Illumina 450 K; $N = 2{,}431$; mean age 35.6) and in buccal samples from children (Illumina EPIC; $N = 1{,}128$; mean age 9.6) from the Netherlands Twin Register. Weights to construct the scores were obtained from results of large epigenome-wide association studies (EWASs) based on whole blood or cord blood methylation data and genome-wide association studies (GWASs). In adults, MSs in blood predicted independently from PGSs, and outperformed PGSs for BMI, prenatal maternal smoking, and smoking status, but not for birth weight. The largest amount of variance explained by the multi-omics prediction model was for current vs. never smoking (54.6%) of which 54.4% was captured by the MS. The two predictors captured 16% of former vs. never smoking initiation variance (MS:15.5%, PGS: 0.5%), 17.7% of prenatal maternal smoking variance (MS:16.9%, PGS: 0.8%), 11.9% of BMI variance (MS: 6.4%, PGS 5.5%), and 1.9% of birth weight variance (MS: 0.4%, PGS: 1.5%). In children, MSs in buccal samples did not show independent predictive value. The largest amount of variance explained by the two predictors was for prenatal maternal smoking (2.6%), where the MSs contributed 1.5%. These results demonstrate that blood DNA MS in adults explain substantial variance in current smoking, large variance in former smoking, prenatal smoking, and BMI, but not in birth weight. Buccal cell DNA methylation scores have lower predictive value, which could be due to different tissues in the EWAS discovery studies and target sample, as well as to different ages. This study illustrates the value of combining polygenic scores with information from methylation data for complex traits and exposure prediction.

Keywords: DNA methylation, methylation scores, polygenic scores, multi-omics prediction, birth weight, maternal smoking, BMI, smoking

OPEN ACCESS

Edited by:
Noèlia Fernàndez-Castillo, Centre for Biomedical Network Research (CIBER), Spain

Reviewed by:
Marina Mitjans, University of Barcelona, Spain
Linn Maria Ellinor Gillberg, Copenhagen University Hospital, Denmark

****Correspondence:***
Veronika V. Odintsova
v.v.odintsova@vu.nl
Jenny van Dongen
j.van.dongen@vu.nl

[†] These authors have contributed equally to this work

[‡] The Biobank-based Integrative Omics Study (BIOS) consortium. For a complete list of authors, see the end of the article

Specialty section:
This article was submitted to Behavioral and Psychiatric Genetics, a section of the journal Frontiers in Psychiatry

Received: *30 March 2021*
Accepted: *15 June 2021*
Published: *29 July 2021*

Citation:
Odintsova VV, Rebattu V, Hagenbeek FA, Pool R, Beck JJ, Ehli EA, van Beijsterveldt CEM, Ligthart L, Willemsen G, de Geus EJC, Hottenga J-J, Boomsma DI and van Dongen J (2021) Predicting Complex Traits and Exposures From Polygenic Scores and Blood and Buccal DNA Methylation Profiles. Front. Psychiatry 12:688464. doi: 10.3389/fpsyt.2021.688464

INTRODUCTION

Nearly all complex traits in humans are a function of their genotype and of environmental exposures, as shown by family and twin studies (1–3). DNA-based predictors of complex traits can increasingly serve to improve prediction of health outcomes and disease and to optimize risk stratification (4) and are also considered for application in social sciences and education (5, 6). Whereas, DNA-based predictors are static and solely capture genomic information, other predictors such as those based on epigenome data are dynamic and may capture both genetic and environmental information.

Polygenic scores (PGS; sometimes referred to as Polygenic Risk Scores) are defined as the weighted sum of an individual's risk alleles, or increasing alleles for a continuous trait, of a pre-selected number of single nucleotide polymorphisms (SNPs). In some areas of medicine, polygenic risk scores are already beginning to be employed to predict individual risk of disease (7–9). The PGS of an individual for a trait is calculated by multiplying, for each SNP, the number of risk alleles by a weight and then summing over all SNPs. Weights are typically estimated in a regression analysis, from a genome-wide association study (GWAS) for the trait from an independent discovery sample (typically, a large GWAS meta-analysis), and are included in the GWAS summary statistics (i.e., the estimated effect sizes, the standard errors of the estimates and the corresponding *p*-values).

This polygenic type of approach can be generalized to other omics data, including epigenomics where it results in DNA methylation scores (MS) (10), which can be described as weighted sums of the individual's methylation levels of a selected number of CpG sites. The individual's methylation levels at each CpG in an independent study population are multiplied by their corresponding weights and summed over multiple sites. Here the weights are based on summary statistics from a single or a meta-analysis epigenome-wide association study (EWAS) of the trait. By combining the effects of multiple CpG sites into a MS, a larger proportion of variance in traits is likely be explained compared to the variance that is captured by individual CpG sites. In addition to their value for prediction of complex traits and disease risk, MSs could potentially be informative as biomarkers for environmental exposures (11) or to monitor disease progression, and might be considered in association analyses in which individual CpG sites do not achieve significance or as a dimension reduction approach in interaction and mediation analyses (12, 13).

The number of genetic variants and CpG sites associated with complex traits is growing based on findings from GWAS and EWAS meta-analyses. Birth weight was associated with 60 independent signals in a multi-ancestry GWA meta-analysis, capturing up to 4.9% of the variance in birth weight in different cohorts (14), and with 914 epigenome-wide Bonferroni-significant CpGs in an EWAS meta-analysis of multiple birth cohorts with cord blood DNA methylation data (15). Body mass index (BMI) was associated with 751 SNPs in adults in the currently largest European ancestry GWAS meta-analysis, capturing ∼6% of the BMI variance (16). The currently largest EWAS meta-analysis of BMI based on whole blood from adults identified association with 278 Bonferroni-significant CpGs (12). Smoking initiation was associated with 566 genetic variants in a GWAS of more than one million individuals, capturing 3.6 and 4.2% of the variance in the trait in prediction cohorts (17). A large EWAS meta-analysis of smoking identified 18,760 CpGs significantly differentially methylated in relation to current smoking in adults at a false discovery rate (FDR) of 5% from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium, and 2,623 FDR significant CpGs in association with former smoking (18). EWAS meta-analyses conducted in newborns using cord blood DNA methylation data identified 6,073 CpGs with FDR significance in association with prenatal maternal smoking (19).

Attempting to capture the DNA methylation differences, previous studies have developed polygenic methylation predictors. We extensively reviewed the literature on studies that report methylation predictors as single MS and studies that examined the combined predictive value of MS and PGS (see **Supplementary Table 1**). Taking the results from EWASs into independent target samples in which MSs are defined, has yielded promising results for birth weight (20), BMI (20–22), prenatal maternal smoking (23, 24), and smoking status (11, 23, 25–28). Reed et al. (20) computed MSs for birth weight based on the 135 CpGs from an adult BMI EWAS in the Framingham Heart Study and Lothian Birth cohorts ($N = 3{,}743$) (29). These scores captured 2% of birth weight variation in 823 ALSPAC newborns with DNA methylation data in cord blood, which was higher than the variance captured by a PGS (0.4%). Several studies created whole blood DNA MSs of BMI and made predictions in children and adults. MSs based on 78 probes from 2,377 adults of the Framingham Heart Study and weights (effect sizes) from 750 adults of the LifeLines DEEP study explained 11% of the variance in BMI in 1,366 adults from Lothian Birth cohorts and 5% of BMI variance in 403 adolescents from Brisbane Systems Genetic Study (BSGS) (21). MSs based on 400 CpGs from 2,562 Generation Scotland participants explained 10% of BMI variance in 892 adults from Lothian Birth cohort (22). MSs based on 135 probes from 3,742 adults from both Framingham Heart Study and Lothian Birth cohorts explained 10% of BMI variance in 726 ALSPAC women and up to 3% of BMI variance in children at different ages (20). It has been shown that MS for BMI perform better in adults compared to children and adolescents (20, 21). Attempts of cross-tissue performance testing were scarce (25, 30), however, it have been shown that some alterations persist across tissue types (31).

For prenatal maternal smoking, MS based on weights from cord blood DNA methylation EWASs of 1,057 newborns from Norwegian Mother and Child Cohort Study (MoBa) was tested on another MoBa subset of 221 newborns (24), and MS based on weights from cord blood DNA methylation EWAS meta-analysis of 6,685 newborns done by Joubert and colleagues (19) was tested on 754 ALSPAC women around 30 years old (23); the predictive accuracy (the amount of variation in the outcome explained by the score) was lower in women than in newborns. Smoking predictors have been described based on different numbers of probes from whole blood DNA methylation studies. Only 2 CpGs were included the smoking MS of Zhang

TABLE 1 | Discovery epigenome-wide and genome-wide association studies that provided the summary statistics to calculate DNA methylation scores and polygenic scores.

Trait/Exposure	References	Phenotype in Discovery Study—DNA methylation Tissue	*N* discovery cohort used to create scores	*N* CpGs/SNPs reported at significant level in reference
EWAS				
Birth weight	Küpers et al. (15)	Birth weight—Cord blood	8,825 newborns	914 (Bonferroni)
BMI	Wahl et al. (12)	BMI—Whole blood	5,387 adults	278 (Bonferroni)
Prenatal maternal smoking	Sidkar et al. (51)*	Prenatal maternal smoking—Cord blood	4,994 newborns (897 exposed to sustained prenatal smoking)	5,547 (FDR <0.05)
Smoking	Sidkar et al. (51)*	Current vs. never smoked—Whole blood	9,389 adults (2,433 current smokers)	34,541 (FDR <0.05)
	Joehanes et al. (18)	Former vs. never smoked—Whole blood	13,474 adults (6,518 former smokers)	2,623 (FDR <0.05)
GWAS				
Birth weight	http://www.nealelab.is/uk-biobank/	Birth weight	280,250 (UK biobank)	not published
BMI	Yengo et al. (16)	BMI	∼700,000 individuals of different ancestry 652,099** were used in our study for calculation of PGS	941
Prenatal maternal smoking	UK Biobank http://www.nealelab.is/uk-biobank/	Maternal smoking around birth	331,862 (UK Biobank)	Not published
Smoking	Liu et al. (17)	Smoking initiation (ever/never smoked)	Up to 1.2 million individuals in discovery study 625,536** used in our study for PGS	566

EWAS, Epigenome-wide association study; GWAS, Genome-wide association study; FDR, False Discovery Rate.
**Sikdar et al. (51) repeated the meta-analysis by Joubert et al. (2016) (EWAS in newborns) and Joehanes et al. (18) (EWAS in never vs. current smokers), and provided full genome-wide summary statistics for a fixed-effects meta-analysis of maternal smoking in newborns (cord blood) and for current vs. never smokers (whole blood).*
NTR and 23andMe are excluded. For additional details, For additional details, see* ***Supplementary Tables 2, ***3***.

et al. that predicted smoking status in 9,949 older adults (28). The largest smoking MSs included 2,623 Bonferroni significant CpGs from EWAS meta-analysis of 15,907 individuals (18) and predicted smoking status during pregnancy in 754 women by Richmond et al. (23). The same CpGs were used by Sugden et al. (11) to predict smoking status in 1,037 adults from the Dunedin Longitudinal Study and 2,232 twins from the Environmental Risk Longitudinal Study.

Despite the growing number of cohorts that have both genomic and methylation data, few attempts have been made to combine PGS and MS in a multi-omics model. To the best of our knowledge, BMI, and height are currently the only traits for which the prediction by PGS and MS combined has been investigated (21, 22). In a combined model, the PGS and MS together explained 17% of the variance in BMI in 1,366 adults (21) and 18% in 889 adults (22), both from the Lothian Birth cohorts, 13–16% in 750 adults from Lifelines and 8% in adolescents from the Brisbane Systems Genetic Study (21), corresponding to an added ∼4–9% extra variance explained compared to the PGS alone.

We expand on the previous work by addressing several points. First, it is largely unknown to what extent MS based on EWAS weights derived in adults predict trait variation in children and vice versa. Second, previous studies of MS were based on cord blood or whole blood, and it is unknown if these scores translate to other tissues. Third, for all traits, except BMI and height (20, 21), it is unknown whether MS add to prediction independently of PGS.

In the current biomarker study, we analyze the predictive accuracy of PGS and MS (both individually and combined). The goal of our study is to examine if the MSs add predictive value above the PGSs. The weights required for DNA methylation data were obtained large EWAS and applied to methylation levels from two different tissues (blood and buccal). We analyze data from large groups of adults with DNA methylation in blood (N = 2,431, mean age = 35.6) and children with DNA methylation in buccal cells (N = 1,128, mean age = 9.6) who participate in research projects of the Netherlands Twin Register and consider multiple traits. For an early-life trait we analyze birth weight, and for a trait that is dynamic in childhood and adulthood, we analyze BMI. As early and later life exposures we examine prenatal maternal smoking during pregnancy and own smoking. These four phenotypes represent complex traits and exposures with different

relative contributions of genetics and environment to inter-individual variance.

MATERIALS AND METHODS

Overview

This study included adults and children who participated in studies from the Netherlands Twin Register (NTR). DNA samples in adult twins and family members were isolated from whole blood DNA data and in twin children from buccal cells. Adults took part in the NTR-Biobank (32) and children in the FP7-Action project (33–36). The study was approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre, Amsterdam, an Institutional Review Board certified by the U.S. Office of Human Research Protections (IRB number IRB00002991 under Federal-wide Assurance FWA00017598; IRB/institute codes, NTR 03-180). Adults provided written informed consent, for children consent was given by their parents.

Adults

Study Population and Samples

After quality control, genome-wide DNA methylation profiles in whole blood and genotype data were available for 2,431 NTR adults (37). This dataset included 2,426 individuals from twin pairs, and 5 family members (mothers and spouses). The mean age at DNA collection was 35.6 years (range = 17.6–79.2 years) and 32.7% of subjects were males. For 20 participants, longitudinal methylation data (methylation data at two time points) were available. Individuals with missing data on phenotypes or covariates, and phenotype outliers were excluded from analysis, resulting in a sample size of 2,040 for birth weight, 2,410 for BMI, 1,914 for current vs. never smoking, and 1,938 for former vs. never smoking. Because prenatal maternal smoking exposure is equal for co-twins, one twin from each pair was randomly included in the analysis, resulting in a sample size of 720. The blood sampling procedure has been described by Willemsen et al. (32).

DNA Methylation

DNA methylation in blood was assessed with the Infinium HumanMethylation450 BeadChip Kit (Illumina, San Diego, CA, USA) by the Human Genotyping facility (HugeF) of ErasmusMC, the Netherlands (http://www.glimdna.org/) as part of the Biobank-based Integrative Omics Study (BIOS) consortium (38). DNA methylation measurements have been described previously (37, 38). Genomic DNA (500 ng) from whole blood was bisulfite treated using the Zymo EZ DNA Methylation kit (Zymo Research Corp, Irvine, CA, USA), 12 μl of buffer was utilized to elute the converted DNA off the column after conversion, and 4 μl (∼33 ng/μl) of bisulfite-converted DNA was measured on the Illumina 450 K array following the manufacturer's protocol. A number of sample- and probe-level quality checks and sample identity checks were performed, as described in detail previously (37). In short, sample-level QC was performed using MethylAid (39). Probes were set to missing in a sample if they had an intensity value of exactly zero, or a detection $p > 0.01$, or a bead count of <3. After these steps, probes that failed based on the above criteria in >5% of the samples were excluded from all samples (only probes with a success rate ≥0.95 were retained). For all samples, ambiguously mapped probes were excluded, based on the definition of an overlap of at least 47 bases per probe from Chen et al. (40), and all probes containing a SNP, identified in the Dutch population (41), within the CpG site (at the C or G position) were excluded, irrespective of minor allele frequency. Only autosomal sites were kept in the current analyses ($N = 411{,}169$). The methylation data were normalized with functional normalization (42). Probes with missing values (probes with missing values in more than 5% of the sample were removed) were imputed with the function imputePCA from the package missMDA as implemented in the pipeline for DNA methylation array analysis developed by the Biobank-based Integrative Omics Study (BIOS) consortium (43).

Phenotyping

Data on birth weight were obtained from self-report or by parental report. If data were available from multiple surveys by Adult Netherlands Twin Register (ANTR) and/or informants, they were checked for consistency (44). When multiple data points differed by <200 g, the average was taken, and in the cases of larger differences, data were excluded. Information on maternal smoking during pregnancy was obtained in ANTR Survey 10 (data collection in 2013) with the following question: "Did your mother ever smoke during pregnancy?" Answer categories were "no," "yes," and "I don't know." For twin pairs, the answers were checked for consistency and missing data for one twin were supplemented with data from the co-twin where possible. In the case of inconsistent answers, the data from both co-twins were set to missing. If both twins answered "I don't know," the variable was coded as missing. Data on body mass index (BMI) and smoking status were collected at blood draw (32). We analyzed two smoking phenotypes: current smokers (1) vs. never smokers (0), and former smokers (1) vs. never smokers (0). The percentage of white blood cell was obtained in fresh blood samples collected in EDTA (Ethylene Diamine Tetra Acetic acid) tubes (45). For birth weight and BMI, we removed outliers using a cut-off of 3 standard deviations from the mean. For birth weight, 6 outliers were removed; for BMI, 27 outliers were removed.

Children

Study Population and Samples

Genotype data and genome-wide DNA methylation profiles in buccal cells were collected in a children that participated in a larger project on childhood aggression "Aggression in Children: Unraveling gene-environment interplay to inform Treatment and InterventiON strategies" (ACTION; http://www.action-euproject.eu/) and consists of twins who score high or low on aggression (33–36). After quality control, genome-wide DNA methylation data and genotype data were available for 1,128 children from twin pairs (mainly monozygotic twins). The mean age at DNA collection was 9.6 years (range = 5.6–12.9 years) and 52.8% were males. For 2 participants, a technical replicate measure on with the Infinium MethylationEPIC BeadChip

Kit was included (36). Individuals without missing data on phenotypes or covariates were included in the analyses, and phenotype outliers were excluded, resulting in a sample size of 1,070 children for birth weight and 1,072 for BMI. Because prenatal maternal smoking exposure is equal for co-twins, one twin from each pair was randomly included in the analysis, resulting in a sample size of 547. The sample collection protocol is available at: http://www.action-euproject.eu/content/data-protocols. DNA was collected from buccal swabs at home: 16 cotton sticks were individually rubbed against the inside of the cheek in the morning and evening on 2 days by the participants and placed in buffer. Individuals were asked to refrain from eating or drinking 1 h prior to sampling. High molecular weight genomic DNA was extracted from the swabs by standard DNA extraction techniques and visualized using agarose gel electrophoresis. The DNA samples were quantified using the Quant-iT PicoGreen dsDNA Assay Kit (ThermoFisher Scientific, Waltham, MA, USA).

DNA Methylation

DNA methylation was assessed with the Infinium MethylationEPIC BeadChip Kit (Illumina, San Diego, CA, USA) by the Human Genotyping facility (HugeF) of ErasmusMC, the Netherlands (http://www.glimdna.org/) [see van Dongen et al. (36)]. Quality control (QC) and normalization of the methylation data were performed using a pipeline developed by the Biobank-based Integrative Omics Study (BIOS) consortium (43), which includes sample quality control using the R package MethylAid (39) and probe filtering and functional normalization (42) as implemented in the R package DNAmArray. The following probe filters were applied: probes were set to missing (NA) in a sample if they had an intensity value of exactly zero, detection $P > 0.01$, or bead count < 3; probes were excluded from all samples if they mapped to multiple locations in the genome, if they overlapped with a SNP or Insertion/Deletion (INDEL), or if they had a success rate < 0.95 across samples. Annotations of ambiguous mapping probes (based on an overlap of at least 47 bases per probe) and probes where genetic variants (SNPs or INDELS) with a minor allele frequency > 0.01 in Europeans overlap with the targeted CpG or single base extension site (SBE) were obtained from Pidsley et al. (46). For two twins, a technical replicate measure on EPIC was obtained (on different BeadChip Arrays). Probes with missing values (probes with missing values in more than 5% of the sample were removed) were imputed with the function imputePCA from the package missMDA as implemented in the pipeline for DNA methylation array analysis developed by the BIOS consortium (43).

Phenotyping

Data on birth weight of the young twins came from surveys sent to mothers shortly after the registration of the newborn twins (47). Data on BMI were collected from surveys filled out by mothers and fathers in the Young Netherlands Twin Register (YNTR) when children were around 5, 7, 10, and 12 years of age. If both parents completed the survey, preference was given to data provided by the mother. BMI closest to the date of DNA collection was selected. The average time between DNA collection and BMI assessment was 1.9 years before the survey (median $= -0.9$, range: from buccal sample collection 10.3 years before survey to buccal sample collection 2.1 years after survey). Information on maternal smoking during pregnancy was reported by mothers after registration for three trimesters of pregnancy and was coded as "non-smoking" if the mother did not smoke during the entire pregnancy and "smoking" if the mother smoked at least during one trimester (48). For birth weight and BMI, we removed outliers using a cut-off of 3 standard deviations from the mean. For birth weight, 1 outlier was removed; for BMI, 12 outliers were removed.

Cellular proportions were predicted with hierarchical epigenetic dissection of intra-sample-heterogeneity (HepiDISH) with the RPC method (reduced partial correlation), as described by Zheng et al. (49) and implemented in the R package HepiDISH. HepiDISH is a cell-type deconvolution algorithm developed for estimating cellular proportions in epithelial tissues based on genome-wide methylation profiles and makes use of reference DNA methylation data from epithelial cells, fibroblasts, and seven leukocyte subtypes. This method was applied to the data after data QC and normalization.

Genotyping

Genotyping in children (YNTR) and adults (ANTR) was done on multiple platforms over time including Perlegen-Affymetrix, Affymetrix 6.0, Affymetrix Axiom, Illumina Human Quad Bead 660, Illumina Omni 1M and Illumina GSA. Quality control and processing of the genotype data was performed on the complete dataset of all genotyped participants from the NTR. Quality control was carried out and haplotypes were estimated in PLINK. CEU population outliers, based on per platform 1000 Genomes PC projection with the Smartpca software (50), were excluded. Data were phased per platform using Eagle, and then imputed to 1000 Genomes using Minimac, following the Michigan imputation server protocols. For the polygenic scoring imputed data were converted to best guess genotypes, and filtered to include only ACGT SNPs, SNPs with MAF > 0.01, HWE $p > 10^{-5}$ and genotype call rate > 0.98, and exclude SNPs with more than 2 alleles. All Mendelian errors were set to missing. Principal components (PCs) were calculated with Smartpca using linkage-disequilibrium-pruned (LD-pruned) 1000 Genomes–imputed SNPs that were also genotyped on at least one platform, had MAF > 0.05 and were not present in the long-range LD regions.

Statistical Methods

EWAS and GWAS Summary Statistics

MSs and PGSs were created using weights based on large epigenome-wide association study (EWAS) and genome-wide association study (GWAS) meta-analyses. These studies are summarized in **Table 1**. Additional information on the studies and derived scores is provided in **Supplementary Tables 2, 3**.

DNA Methylation Scores

The effect sizes obtained from the summary statistics from previously published EWAS meta-analyses (**Table 1**, **Supplementary Table 2**) were used to calculate weighted MSs

in NTR participants as previously done by Sugden et al. (11), Wahl et al. (12), Shah et al. (21), Richmond et al. (23), and Elliott et al. (26). For each trait and for each individual, a score was calculated by multiplying the methylation level at a given CpG by the previously reported effect size of the CpG (β), and then summing these values over all CpGs:

$$DNA\ methylation\ score = \beta_1^* CpG_1 + \beta_2^* CpG_2 \ldots . + \beta_i^* CpG_i$$

where CpG_i is the methylation level at CpG site *i*, which ranges between 0 and 1, and β_i is the effect size (regression coefficient) at the CpG_i obtained from summary statistics of EWAS meta-analyses that did not include participants from the NTR.

For each phenotype, except for former smoking, we calculated multiple MSs based on different subsets of CpGs according to their significance level. Subsets of CpGs were selected based on *p*-value thresholds of $<1 \times 10^{-1}$, $<1 \times 10^{-5}$, and $<1 \times 10^{-7}$. For former vs. never smoking, genome-wide summary statistics were not available, and we calculated MSs for former smoking based on CpGs that were significant in the EWAS of former vs. never smokers at a False Discovery Rate of 5%. Additionally, we tested prediction of former smoking based on the MSs derived from the genome-wide EWAS summary statistics of current vs. never smoking. To examine if removal of CpGs with correlated DNA methylation levels affects trait prediction, we also calculated pruned scores by step-wise selection of the most significant CpG site and excluding CpG sites with a correlation of 0.1 or higher [threshold chosen based on Shah et al. (21)] in order to keep an independent set of CpGs.

Polygenic Scores

Polygenic scores (PGSs) were calculated based on weighting of genotypes by effect sizes as made available from GWAS summary statistics (see **Table 1**, **Supplementary Table 3**) in discovery samples without NTR. Before calculating the PGSs, linkage disequilibrium (LD) weighted β's were calculated from these summary statistics by the LDpred package to correct for the effects of LD and to maximize predictive accuracy of the PGSs (52). QC has been applied: MAF > 0.01, duplicated SNPs, mismatching alleles, ambiguous SNPs were excluded. We randomly selected 2,500 unrelated individuals from NTR as a reference population to calculate the LD patterns. The adjusted β's were calculated from an LD pruning window of 250 KB, with the fraction of causal SNPs set at 0.01 for birth weight, because this fraction was previously shown to perform optimally for birth weight in the NTR population (53) and at 0.50 for other phenotypes. The PGSs were obtained for all NTR participants with genotyping data with the PLINK 1.9 software.

Statistical Analysis

Continuous traits, MSs and PGSs were z-score transformed [trait value—trait mean/trait standard deviation] before analysis. Pairwise Pearson correlations between each trait, MSs, PGSs, and covariates were computed in NTR adults and children EWAS datasets for each phenotype and visualized in correlation plots. For each trait, we fitted a series of regression models to examine: (1) the predictive value of MSs; (2) the predictive value of a PGS; and (3) whether MS and PGS contributed independently to trait prediction in a combined predictor. First, for each trait, we evaluated the performance of multiple different MSs based on different *p*-value thresholds, pruned, and unpruned. We took the score that explained the largest amount of variance forward to the combined model. Second, we evaluated the performance of PGS in prediction of each trait. Finally, we examined if MSs predict these traits over and above the PGSs and estimated how much variance in each trait was explained by multi-omics predictor, e.g., by MSs and PGSs together. Sex and age at DNA collection were included as covariates in all three models. In the prediction models with whole blood DNA MSs, we corrected for percentages of neutrophils, monocytes, and eosinophils. In the prediction models with buccal DNA MSs, we corrected for epithelial cell and natural killer cell proportions. To adjust for technical variation, array row and bisulfite plate (dummy-coding) were included as covariates in all models with EWAS covariates. In models including PGSs, we corrected for genotype data-specific covariates: the first ten genetic principal components and genotype platform dummy variables (GWAS covariates).

TABLE 2 | Characteristics: NTR adults and children.

	Adults	Children
DNA methylation tissue	**Whole blood**	**Buccal cells**
DNA methylation array	Illumina 450 k	Illumina EPIC
Sample size	2,431	1,128
Age, mean (sd)	35.6 (11.9)	9.6 (1.9)
Males, *n* (%)	794 (32.7%)	596 (52.8%)
Females, *n* (%)	1,637 (67.3%)	532 (47.2%)
Birth weight, mean (sd)	2,507.8 (573.7)	2,395.9 (544.1)
Missing	404	59
BMI, mean (sd)	24.1 (3.9)	16 (2.1)
Missing	14	46
Prenatal maternal smoking (no), *n* (%)	1,169 (90.1%)	921 (91.1%)
Prenatal maternal smoking (yes), *n* (%)	129 (9.9%)	90 (8.9%)
Missing	1,133	117
Never smokers, *n* (%)	1,395 (57.5%)	NA
Former smokers, *n* (%)	527 (21.7%)	NA
Current smokers, *n* (%)	506 (20.8%)	NA
Missing	3	NA
Cell counts, mean (sd)		
Neutrophil percentage	52.5 (9.1)	NA
Eosinophil percentage	3.1 (2.3)	NA
Monocyte percentage	8.4 (2.4)	NA
Epithelial cell proportion	NA	0.806 (0.116)
NK cell proportion	NA	0.03 (0.013)

NTR, Netherlands Twin Register; sd, standard deviation; NA, not available; NK, natural killer cells.

Descriptives are provided for covariates that were included as covariates in the prediction models.

The number of samples with complete data on DNA methylation, cell counts, and genotyping data are presented. For a small number of individuals, two samples are included in the analysis in adults (n = 20) and children (n = 2).

Continuous Traits

For birth weight and BMI, the following models were fitted in each of the two datasets (whole blood methylation data from adults and buccal methylation data from children):

Model 1: Trait ~ MS + sex + age + EWAS covariates
Model 2: Trait ~ PGS + sex + age + GWAS covariates
Model 3: Trait ~ MS + PGS + sex + age + EWAS covariates + GWAS covariates

Analyses were carried out with generalized estimation equation (GEE) models accounting for familial relatedness, fitted with the R package "gee" with the following settings: Gaussian link function for continuous data, 100 iterations, and the "exchangeable" option to account for the correlation structure within families. To calculate the variance explained by the MS and the PGS, we squared the regression coefficient of each score obtained in GEE. This value was multiplied by 100 to obtain the percentage of variance explained.

Dichotomous Traits

For dichotomous traits, i.e., prenatal maternal smoking, current vs. never smoking, and former vs. never smoking, the following models were fitted in two datasets (whole blood methylation data from adults and buccal methylation data from children):

Model 1a: Trait ~ MS + sex + age + EWAS covariates
Model 1b: Trait ~ sex + age + EWAS covariates
Model 2a: Trait ~ PGS + sex + age + GWAS covariates
Model 2b: Trait ~ sex + age + GWAS covariates
Model 3a: Trait ~ MS + PGS + sex + age + EWAS covariates + GWAS covariates
Model 3b: Trait ~ PGS + sex + age + EWAS covariates + GWAS covariates
Model 3c: Trait ~ MS + sex + age + EWAS covariates + GWAS covariates

To obtain the variance explained, models were fitted with logistic regression with binomial family setting (link = "logit"). Estimation of the variance explained by the MS and PGS, was based on the approach proposed by Lee et al., where coefficients of determination (R^2) for binary responses are calculated on the liability scale (54). R^2 is equal to the explained variance divided by the total variance; that is the sum of explained variance and residual (homoscedastic) variance. We first regressed the trait on the MS, sex, age and EWAS covariates (model 1a), and then on sex, age, and EWAS covariates only (model 1b). We calculated variance explained by all predictors in each model. We calculated the predictive value of the MS by subtracting the difference between the variance explained by the model 1a and 1b. The same was done for models with PGS with sex, age, and GWAS covariates (model 2a and 2b), and then for combined model with both MS and PGS scores (models 3a-c). In the last case, the difference between explained variance in model 3a and model 3b gave us an estimate explained by MS, and the difference between explained variance in model 3a and model 3c resulted in estimate explained by PGS.

To correct for relatedness in smoking prediction, *p*-values were obtained from GEE models, fitted with the R package "gee," with the binomial link function for dichotomous data, 100 iterations, and the "exchangeable" option to account for the correlation structure within families. For prenatal smoking exposure (yes/no) we randomly chose one of the twins from the pair, and *p*-values were obtained from logistic regression models.

Sensitivity Analysis

We carried out a sensitivity analysis in which we repeated the models for BMI prediction in children from MSs after removal of children for whom information on BMI was collected more than 3 years before or after DNA collection ($N = 324$ children removed; new $N = 748$).

Multiple Testing Correction

Statistical significance was assessed following Bonferroni correction for multiple testing (six tests in model 1 for birth weight, BMI, prenatal maternal smoking and current smoking, seven tests in model 1 for former smoking, one test in models 2 and 3 for each trait in adults; the same number of tests in children except smoking status and plus eight tests in sensitivity analysis for BMI). This resulted in a significance level of 0.0012 ($\alpha = 0.05/42$) for adults and 0.0016 ($\alpha = 0.05/32$) for children.

RESULTS

Characteristics of the NTR adult and children are presented in **Table 2**. Distribution of the traits/exposures as main outcomes and MSs and PGs as predictors are presented in **Supplementary Figures 1–3**. Correlations between trait/expoosure, PGS, MSs, sex, age, and cellular compositions of the samples are shown in **Supplementary Figures 4, 5**. The correlations between PGS and MSs for the same trait were weak in adults ($r = [0.01–0.15]$) and children ($r = [0.01–0.05]$). Further, we report the correlation between the PGS and the MS that captured the largest amount of variation in the trait. We examined prediction of each phenotype by its MS and PGS separately. The explained variance and corresponding *p*-values for unpruned and pruned MSs with different thresholds for inclusion of CpGs are presented in **Table 3**, and for PGS in **Table 4**. To examine to what extent the PGS and the MS capture independent information, we fitted the model in which the outcome was regressed on both scores as multi-omics prediction presented in **Table 5**. **Figure 1** shows the variance explained by the MSs and PGSs separately and together as multi-omics predictor in previous and our studies.

Birth Weight

The birthweight MSs were calculated based on the birthweight EWAS of cord blood samples from neonates (15). The results of GEE showed that none of the MSs was strongly associated with birth weight in adults ($p < 0.0012$) and children ($p < 0.0016$). The pruned blood MS based on 934 CpGs with a *p*-value lower than 1×10^{-1} performed better in prediction of birth weight in adults compared with unpruned and other threshold pruned scores, accounting for 0.39% of the variance ($p = 0.004$) (**Table 3**, **Figure 1A**). The PGS significantly predicted birth weight in adults (variance explained by PGS = 1.52%, $p = 1.56 \times 10^{-7}$)

TABLE 3 | Results of the methylation score prediction of birth weight, BMI, prenatal maternal smoking and current and former smoking.

Trait	Group—DNA methylation tissue	CpGs included in MS	N CpGs	β_{MS}	SE_{MS}	P_{MS}	MS R^2 (%)
Birth weight	Adults—Whole blood ($N = 2,040$)	$p < 10^{-1}$	72,570	0.044	0.029	0.127	0.192
		$p < 10^{-5}$	2,274	0.055	0.029	0.057	0.304
		$p < 10^{-7}$	963	0.049	0.027	0.074	0.238
		$p < 10^{-1}$ pruned**	934	0.062	0.022	0.004	0.386
		$p < 10^{-5}$ pruned	30	0.015	0.023	0.513	0.023
		$p < 10^{-7}$ pruned	18	0.025	0.023	0.277	0.064
	Children—Buccal cells ($N = 1,070$)	$p < 10^{-1}$	72,205	−0.013	0.031	0.679	0.016
		$p < 10^{-5}$	2,249	0.031	0.032	0.344	0.094
		$p < 10^{-7}$**	958	0.038	0.032	0.237	0.141
		$p < 10^{-1}$ pruned	184	0.015	0.029	0.613	0.022
		$p < 10^{-5}$ pruned	13	0.016	0.032	0.613	0.025
		$p < 10^{-7}$ pruned	9	−0.005	0.033	0.882	0.002
BMI	Adults—Whole blood ($N = 2,410$)	$p < 10^{-1}$	55,653	0.134	0.025	$6.98 \times 10^{-08*}$	1.786
		$p < 10^{-5}$	1,067	0.261	0.026	$5.07 \times 10^{-24*}$	6.822
		$p < 10^{-7}$**	412	0.277	0.026	$9.79 \times 10^{-27*}$	7.673
		$p < 10^{-1}$ pruned	671	0.124	0.021	$3.14 \times 10^{-09*}$	1.538
		$p < 10^{-5}$ pruned	13	0.220	0.023	$1.70 \times 10^{-22*}$	4.827
		$p < 10^{-7}$ pruned	6	0.206	0.023	$7.50 \times 10^{-20*}$	4.258
	Children—Buccal cells ($N = 1,072$)	$p < 10^{-1}$	55,279	0.003	0.017	0.878	0.001
		$p < 10^{-5}$	1,079	−0.006	0.017	0.733	0.003
		$p < 10^{-7}$	422	−0.008	0.018	0.676	0.006
		$p < 10^{-1}$ pruned**	183	0.021	0.019	0.276	0.042
		$p < 10^{-5}$ pruned	13	0.005	0.019	0.772	0.003
		$p < 10^{-7}$ pruned	6	0.007	0.018	0.720	0.004
Prenatal maternal smoking	Adults—Whole blood ($N = 720$)	$p < 10^{-1}$	76,531	0.761	0.168	$5.83 \times 10^{-06*}$	8.512
		$p < 10^{-5}$	1,581	0.946	0.140	$1.51 \times 10^{-11*}$	15.115
		$p < 10^{-7}$**	607	1.009	0.136	$1.28 \times 10^{-13*}$	17.277
		$p < 10^{-1}$ pruned	962	0.580	0.136	$2.01 \times 10^{-05*}$	7.076
		$p < 10^{-5}$ pruned	33	0.661	0.123	$7.12 \times 10^{-08*}$	6.820
		$p < 10^{-7}$ pruned	16	0.620	0.119	$1.94 \times 10^{-07*}$	5.963
	Children—Buccal cells ($N = 547$)	$p < 10^{-1}$	76,146	−0.256	0.158	0.105	0.930
		$p < 10^{-5}$	1,571	−0.285	0.187	0.127	1.683
		$p < 10^{-7}$**	606	−0.304	0.183	0.096	2.223
		$p < 10^{-1}$ pruned	187	−0.125	0.153	0.416	0.386
		$p < 10^{-5}$ pruned	16	−0.096	0.164	0.559	0.149
		$p < 10^{-7}$ pruned	9	−0.120	0.161	0.456	0.274
Smoking current vs. never	Adults—Whole blood ($N = 1,914$)	$p < 10^{-1}$	98,972	1.420	0.095	$4.65 \times 10^{-43*}$	20.974
		$p < 10^{-5}$	11,433	1.923	0.104	$1.93 \times 10^{-66*}$	40.086
		$p < 10^{-7}$	6,938	1.999	0.105	$6.89 \times 10^{-69*}$	44.449
		$p < 10^{-1}$ pruned	913	1.170	0.070	$2.70 \times 10^{-47*}$	27.626
		$p < 10^{-5}$ pruned	37	2.194	0.108	$5.21 \times 10^{-75*}$	56.237
		$p < 10^{-7}$ pruned**	24	2.245	0.111	$1.25 \times 10^{-73*}$	57.461
Smoking former vs. never	Adults—Whole blood ($N = 1,938$)	$p < 10^{-1}$ from current vs. never smoking EWAS	98,972	0.429	0.085	$6.63 \times 10^{-07*}$	2.004

(Continued)

TABLE 3 | Continued

Trait	Group—DNA methylation tissue	CpGs included in MS	N CpGs	β_{MS}	SE_{MS}	P_{MS}	MS R^2 (%)
		$p < 10^{-5}$ from current vs. never smoking EWAS	11,433	0.689	0.086	$2.73 \times 10^{-15*}$	5.127
		$p < 10^{-7}$ from current vs. never smoking EWAS	6,938	0.758	0.086	$3.37 \times 10^{-18*}$	6.217
		FDR significant from Former vs. Never Smoking EWAS	2,568	0.935	0.088	$3.56 \times 10^{-24*}$	9.340
		$p < 10^{-1}$ pruned from current vs. never smoking EWAS	913	0.499	0.066	$3.38 \times 10^{-13*}$	4.399
		$p < 10^{-5}$ pruned from current vs. never smoking EWAS	37	1.186	0.087	$6.64 \times 10^{-36*}$	15.625
		$p < 10^{-7}$ pruned from current vs. never smoking EWAS**	24	1.226	0.088	$5.44 \times 10^{-36*}$	16.316
		FDR significant from Former vs. Never Smoking EWAS, pruned	2,330	0.692	0.071	$1.96 \times 10^{-18*}$	7.569

β is the regression coefficient for each methylation score (MS) with standard error (SE) and p-value (P). MS R^2 is the phenotypic variance explained by the MS.
** indicates p < 0.0012 in adults and <0.0016 in children. **indicate methylation score with lowest p-value for a trait/exposure.*

TABLE 4 | Results of the polygenic score prediction of birth weight, BMI, prenatal maternal smoking, and current and former smoking.

Trait	Group	β_{PGS}	SE_{PGS}	P_{PGS}	PGS R^2 (%)
Birth weight	Adults ($N = 2,040$)	0.123	0.024	$1.56 \times 10^{-07*}$	1.520
	Children ($N = 1,070$)	0.118	0.030	$9.67 \times 10^{-05*}$	1.387
BMI	Adults ($N = 2,410$)	0.259	0.022	$3.43 \times 10^{-32*}$	6.725
	Children ($N = 1,072$)	0.173	0.041	$2.20 \times 10^{-05*}$	3.003
Prenatal maternal smoking	Adults ($N = 720$)	0.259	0.132	0.049	1.797
	Children ($N = 547$)	0.282	0.165	0.086	1.622
Smoking current vs. never	Adults ($N = 1,914$)	0.330	0.056	$2.24 \times 10^{-07*}$	2.794
Smoking former vs. never	Adults ($N = 1,938$)	0.197	0.058	0.001	0.909

β is the regression coefficient for each polygenic score (PGS) with standard error (SE) and p-value (P). PGS R^2 is the phenotypic variance explained by the PGS.
PGS to predict current vs. never smoking and former vs. never smoking were created based on GWAS on smoking initiation (17).
**indicates p < 0.0012 in adults and <0.0016 in children.*

(**Table 4**). The correlation between the whole blood MS and PGS in adults was −0.03 ($p = 0.182$; **Supplementary Figure 4A**). In the model combining MS and PGS to predict birth weight, the PGS, and blood MS in adults both significantly explained variation in birth weight (variance explained by MS in combined model: 0.39%, $p = 0.003$; by PGS: 1.53%, $p = 1.96 \times 10^{-7}$ and MS+PGS: 1.92%) (**Table 5**, **Figure 1E**).

In children, the best performing score was based on 958 CpGs with a *p*-value lower than 1×10^{-7}, explaining 0.14% of the variance ($p = 0.263$) (**Table 3**). The PGS predicted birth weight in children (variance explained by PGS = 1.39%, $p = 9.67 \times 10^{-5}$) (**Table 4**). MSs did not add predictive value to PGS in the combined model (**Table 5**, **Figure 1F**).

BMI

Blood MSs for BMI were based on the EWAS by Wahl et al. (12) in blood DNA in adults. These account for a

TABLE 5 | Results of the multi-omics prediction of birth weight, BMI, prenatal maternal smoking, and current and former smoking.

Trait	Group—DNA methylation tissue	Methylation score					Polygenic score				Total R^2 for MS+PGS (%)
		CpGs included in MS	β_{MS}	SE_{MS}	P_{MS}	MS R^2 (%)	β_{PGS}	SE_{PGS}	P_{PGS}	PGS R^2 (%)	
Birth weight	Adults—Whole blood ($N = 2,040$)	$p < 10^{-1}$ pruned	0.063	0.021	0.003	0.394	0.124	0.024	$1.96 \times 10^{-07*}$	1.525	1.92
	Children—Buccal cells ($N = 1,070$)	$p < 10^{-7}$	0.043	0.032	0.172	0.186	0.110	0.030	$2.61 \times 10^{-04*}$	1.219	1.41
BMI	Adults—Whole blood ($N = 2,410$)	$p < 10^{-7}$	0.253	0.026	$5.5 \times 10^{-23*}$	6.40	0.235	0.022	$7.02 \times 10^{-28*}$	5.54	11.95
	Children—Buccal cells ($N = 1,072$)	$p < 10^{-1}$ pruned	0.017	0.019	0.347	0.030	0.157	0.042	$1.7 \times 10^{-04*}$	2.451	2.48
Prenatal maternal smoking	Adults—Whole blood ($N = 720$)	$p < 10^{-7}$	1.031	0.140	$2.4 \times 10^{-13*}$	16.886	0.207	0.140	0.139	0.837	17.72
	Children - Buccal cells (N = 547)	$p < 10^{-7}$	−0.291	0.187	0.120	1.529	0.283	0.168	0.091	1.093	2.62
Smoking current vs. never	Adults—whole blood ($N = 1,914$)	$p < 10^{-7}$ pruned	2.251	0.113	$1.3 \times 10^{-70*}$	54.41	0.165	0.080	0.042	0.16	54.57
Smoking former vs. never	Adults—whole blood ($N = 1,938$)	$p < 10^{-7}$ pruned from current vs. never smoking EWAS	1.216	0.089	$1.5 \times 10^{-34*}$	15.52	0.159	0.064	0.018	0.51	16.03

*β is the regression coefficient for each term with standard error (SE) and p-value (P). MS R^2 is the phenotypic variance explained by the MS. PGS R^2 is the phenotypic variance explained by the PGS. The combined model included the PGS and the best performing MS. *indicates $p < 0.0012$ in adults and <0.0016 in children.*

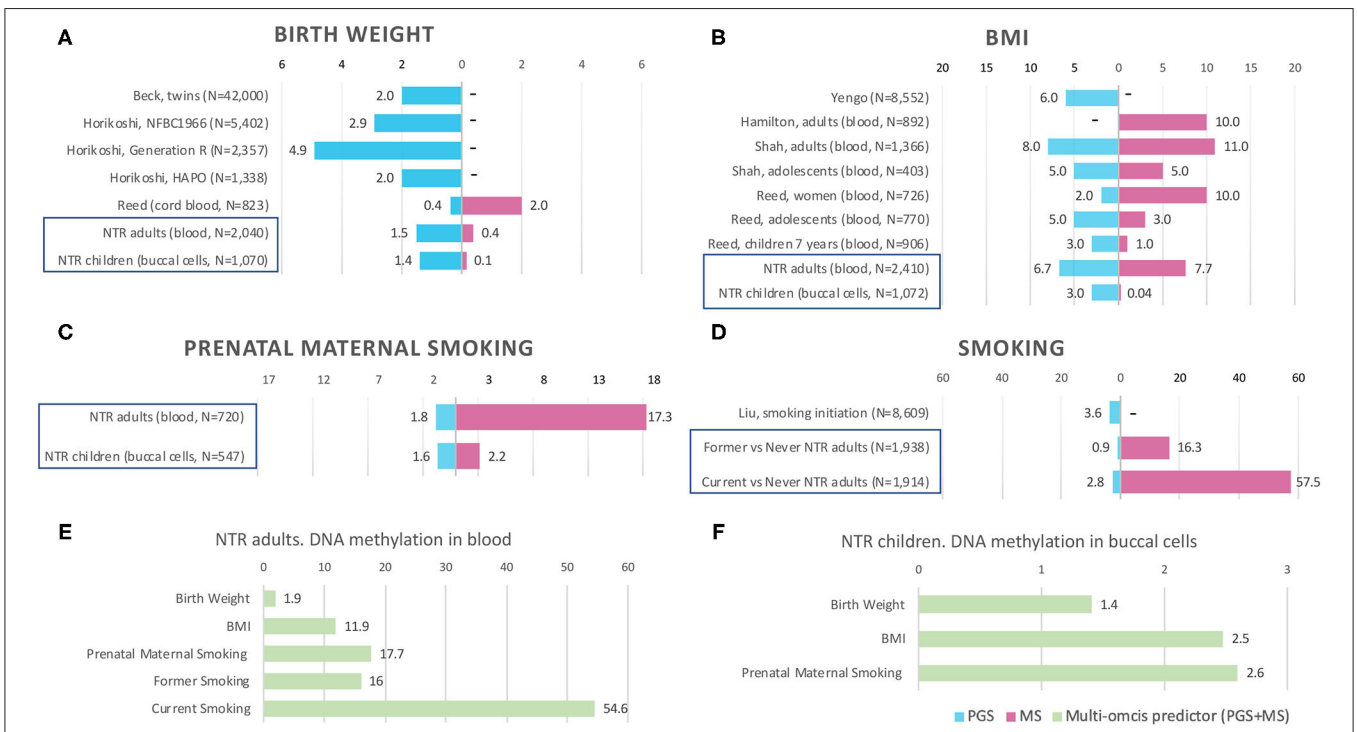


FIGURE 1 | Prediction by methylation and polygenic scores in previous studies and NTR. Bars are the phenotypic variance explained by the score (R^2), x-axis shows R^2 in %. MS, methylation score. PGS, polygenic score. PGS+MS, polygenic and methylation scores in combined model (multi-omics predictor). "–" indicates that the score is not available in the study. Prediction by PGS and MS separately in NTR cohorts is indicated by blue frames in **(A–D)**. Multi-omics prediction in NTR is presented in **(E–F)**. Full references on previous studies in **(A,B,D)** can be found by first author in References. For more details on previous EWASs included in **(A,B)**, see **Supplementary Table 1**.

moderate proportion of the variance in BMI in adults (1.5–7.7%). The best performing score explained 7.7% of the variance in BMI ($p = 9.79 \times 10^{-27}$) and was based on 412 Bonferroni significant CpG sites (**Table 3**, **Figure 1B**). The pruned MSs explained less variation in BMI (1.5–4.8% explained variance). The PGS for BMI explained 6.7% of the variance in adults ($p = 3.43 \times 10^{-6}$; **Table 4**). The correlation between whole blood MS and PGS was 0.1 in adults ($p = 1.31 \times 10^{-6}$; **Supplementary Figure 4B**). In a combined regression model in adults the MS and PGS contributed independently to the prediction of BMI (variance explained by MS in combined model: 6.4%, $p = 5.46 \times 10^{-23}$, by PGS: 5.5%, $p = 7.02 \times 10^{-28}$, and MS+PGS: 11.9%) (**Table 5**, **Figure 1E**).

In children, the BMI MSs based on buccal methylation data had a considerably lower predictive performance, and none of the scores significantly predicted BMI: the best score in children explained 0.04% of the variance ($p = 0.276$), and was based on 183 pruned CpG sites with a p-value lower than 1×10^{-1}. The PGS explained 3% of the BMI variance ($p = 2.2 \times 10^{-5}$). MSs did not outperform PGSs in the combined model (**Table 5**, **Figure 1F**). Furthermore, removal of children for whom information on BMI was collected more than 3 years before or after DNA collection did not lead to an increase in explained variance (**Supplementary Table 4**).

Prenatal Maternal Smoking

In adults, the MSs were based on the EWAS in cord blood from 4,994 newborns (19, 51) and significantly predicted prenatal maternal smoking exposure (**Table 3**, **Figure 1C**). The score based on weights of 607 unpruned CpGs at $p < 1 \times 10^{-7}$ accounted for largest variance of 17.3% of prenatal maternal exposure ($p = 1.28 \times 10^{-13}$). The pruned MSs performed worse (5.9–7% explained variance). The PGS for maternal smoking around birth did not significantly predict prenatal maternal smoking (variance explained by PGS 1.8%, $p = 0.05$) (**Table 4**). The correlation between the best performing MS and PGS was 0.06 ($p = 0.003$; **Supplementary Figure 4C**). The variance explained by MS and PGS in the combined model was slightly lower than predicted by MS alone (variance explained by MS in combined model: 16.9%, $p = 2.4 \times 10^{-13}$, by PGS: 0.84%, $p = 0.139$ and by MS+PGS: 17.7%) (**Table 5**, **Figure 1E**). Maternal smoking scores in buccal methylation data from children, based on the same cord blood discovery EWAS, were not significantly predictive (**Table 5**, **Figure 1F**).

Smoking

The smoking MS in adults were based on the EWAS for never vs. current smokers (18, 51) and were strongly predictive for smoking status. The pruned MSs had a considerably better predictive performance (28–57.5 vs. 21–44% explained variance). The best performing MSs was based on 24 pruned CpGs at $p < 1 \times 10^{-7}$, and explained 57.5% of variance for current smoking ($p = 1.25 \times 10^{-73}$) and 16.3% of the variance for former smoking ($p = 5.44 \times 10^{-36}$) (**Table 3**, **Figure 1D**). The PGS for smoking initiation explained 2.8% of the variance in current smoking ($p = 2.24 \times 10^{-7}$) and 0.9% of the variance of former smoking ($p = 0.001$) (**Table 4**). The correlation between the PGS of smoking initiation and the best performing MS was 0.14 ($p = 7.71 \times 10^{-13}$; **Supplementary Figure 4D**). In the combined prediction model, the MSs outperformed PGSs in the prediction of smoking status, and the PGSs were no longer significant (variance explained by MS in combined model for current vs. never smoking: 54.4%, $p = 1.3 \times 10^{-70}$, by PGS: 0.16%, $p = 0.042$, and by MS+PGS: 54.6%), indicating that the PGS and MS for smoking do not independently add to prediction of this trait (**Table 5**, **Figure 1E**).

DISCUSSION

We examined if a combined model that includes methylation scores (MS) and polygenic scores (PGS) captures more variance in body size, i.e., birth weight and BMI, and in two exposures, i.e., prenatal maternal smoking exposure and smoking in adulthood, in comparison to the PGS alone. Our results showed that MSs in adults, from blood DNA, predicted BMI, prenatal maternal smoking, and smoking status independent of PGSs, and outperformed PGSs for BMI, prenatal maternal smoking, and smoking status, but not for birth weight. In children, MSs from buccal-cell DNA did not show predictive value in children, but here the tissue in the discovery studies derived from EWASs of cord blood and whole blood DNA methylation profiles.

The most successful MS predictor in our study is for smoking. Blood DNA MS explained up to 57.5% of the variance in current smoking status and 16.3% of the variance in former smoking status. This was substantially better compared to the performance of PGS. Tobacco exposure, both prenatal and current, is a potential environmental exposure that modifies DNA methylation. Several previous studies reported successful application of blood DNA MS created based on weights from an independent discovery EWAS, as we did in the current study (11, 23, 26), based on calculation of indexes (27, 28) or based on machine learning algorithms (25). In line with previous studies, MSs performed better for predicting current vs. never smoking than for former vs. never smoking (25). Most studies of smoking were done on blood DNA methylation. It has been suggested that buccal cell DNA methylation predictors should perform even better (25). Currently, our participants with buccal cell DNA methylation data are too young (methylation data in buccal cells was available for children around 9 years old) to have initiated smoking.

The blood DNA MSs for prenatal maternal smoking, based on cord blood-derived weights from newborns, significantly explained 17.3% of variance in adults. Earlier reports demonstrated that maternal smoking during pregnancy is associated with alterations in offspring blood DNA methylation in newborns (19, 55, 56), children and adolescents (57, 58), and adults several decades after exposure (23, 59). However, the effects of sustained maternal smoking during pregnancy fade

away with time, and the predictive accuracy in blood samples from adults is much lower than the accuracy obtained with cord blood samples from newborns (56–58, 60).

We showed that MSs perform better than PGSs in adults for both exposures capturing the effect of smoking, and add value to prediction in combined models. The effects of individual SNPs on behavioral traits such as smoking is small, hence, larger GWAS meta-analyses are required for smoking and maternal smoking to obtain better PGSs. In contrast to PGSs, which capture an individual's genetic predisposition for smoking behavior, MSs capture the effect of exposure to smoking on the methylome. Smoking as an exposure is strongly associated with DNA methylation, and EWAS meta-analyses of smoking and maternal smoking have identified very large numbers of CpGs associated with these traits, allowing for the calculation of fairly reliable MSs, i.e., the EWAS meta-analysis identified over 2,000 significant loci associated with smoking (18), while the currently largest GWAS detected 566 loci associated with smoking initiation (17).

The BMI MS derived in blood samples from adults explained up to 7.7% of the variation in BMI, thereby outperforming the PGS, which explained 6.7% of the variance. Both scores contributed independently to the prediction of BMI in the combined model (12% of explained variance). The performance of adult NTR MSs for BMI was in line with other studies that reported around 10–11% variance explained by MS only in adults (see **Figure 1B**) and larger variance explained by combined MS and PGS predictors (21, 22). In children with buccal methylation data, a considerably smaller proportion of 2.5% of variation in BMI was explained by the MS and PGS in combined predictor. The lower predictive performance of BMI MSs in children than in adults was also observed in other studies (20, 21), and could be explained by increase of environmental contribution to the trait with age (61) as BMI tends to increase during most of adult life (62). Shah et al. (21) reported that BMI MSs based on an EWAS in adults from the Lothian Birth Cohorts explained 4.9% of variation in adults from the Lifelines DEEP study, but did not account for any BMI variation in adolescents (mean age 14 years) from the Brisbane Systems Genetic Study. Reed et al. (20) observed 10% of BMI variance explained in women in comparison with 1% in children age 7 years and 3% in adolescents age 15 years by MSs calculated on the same set of CpGs from an EWAS in adults.

Birthweight MSs were not strongly predictive in our study, with 0.4 and 0.1% of explained variance in adults and children, respectively, while PGS were significant with 1.5 and 1.4% of explained variance in adults and children, respectively. The PGS for birth weight in NTR was in between the variance explained in previous studies (see **Figure 1A**): 0.4% in an ALSPAC cohort of 823 newborns (20), 2% in multi-cohort study of 42 thousand twins (53) and in Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study of 1,338 individuals (14), 2.9% in the Northern Finland Birth Cohort (NFBC) of 5,402 individuals and 4.9% in Generation R cohort of 2,357 individuals (14). The discovery EWAS of birth weight ($N = 8{,}825$ newborns) detected 914 Bonferroni significant CpGs in cord blood (15), suggesting that birth weight does have a large epigenetic signal in cord blood at birth. The cord-blood DNA MS base on weights from adult BMI EWAS accounted for 2% of variance in birth weight (20). According to our knowledge, the performance of birthweight MS based on weights from newborn EWAS has not been previously examined. The low predictive accuracy of the MSs can be caused by the fact the birth weight scores in whole blood and buccal cells were based on an EWAS in cord blood of neonates. Across different tissues and ages, different CpG sites may be associated with birth weight. Another explanation is, that the association between birth weight and DNA methylation fades away with age. Küpers et al. (15) took the 914 significant neonatal blood CpG sites and examined their associations with birth weight in blood samples of adults. No CpG site reached Bonferroni significance in the adults. At present, there are no published EWASs of birth weight based on buccal DNA methylation and no large EWASs of birth weight on blood samples from adults.

The lower predictive accuracy of the buccal cell DNA MSs for all four phenotypes may have several reasons: (1) for some traits there is evidence that DNA methylation signatures increases with age, e.g., BMI, and thus can be not evident at age of 9 years old; (2) unreliability in the phenotype, e.g., prenatal maternal smoking reported participants on their mother's smoking behavior; (3) use of effect sizes from EWASs in cord blood and whole blood methylation data to calculate the scores in buccal cell DNA methylation data. The CpG sites that are predictive for trait/exposure in blood may not be the same CpG sites that are predictive in buccal cells, or the strength of the association may differ across tissues. This last explanation can be tested once more EWASs in buccal cells become available.

The best performing score for each trait will depend on the true number of CpGs associated with the trait and their effect size, the correlation among CpGs, and the power of the discovery EWAS analysis. If the discovery EWAS had full power to detect all CpGs associated with the trait, and there is no large heterogeneity in the effects across cohort, scores created within the same tissue as the discovery EWAS with CpGs based on the most stringent p-value threshold (i.e., $<1 \times 10^{-7}$) are expected to perform best. More likely, EWASs for these traits did not yet detect all CpGs that are truly associated and larger discovery samples are required to detect CpGs with smaller effects. Therefore, we also examined the performance of scores created based on more lenient p-value thresholds. More lenient p-value thresholds will potentially add more CpGs to the MS that are truly associated with the phenotype, but which did not yet reach epigenome-wide significance in the discovery EWAS meta-analysis, thereby improving the score. At the same time, inclusion of more CpGs that are not truly associated with the trait and less accurate weights at more lenient thresholds, add more noise to the MS. Pruning was performed to remove correlated CpGs that are redundant (and potentially add noise to scores). The expectation is that if the set of CpGs associated with a trait is correlated (and especially if correlations are strong or abundant), pruning will improve performance of the MS. We found this to be the case, for instance, for blood DNA MS for smoking in adults. For simplicity, we compared two options that have been previously applied in

the literature: (1) no pruning at all and (2) a correlation cut-off of 0.1 to select an approximate independent set of CpGs, but we note that the optimal correlation cut-off for pruning may also vary across traits. In adults, pruning reduced the performance of some scores, namely BMI and prenatal maternal smoking, while it improved the performance for birth weight and smoking. Sophisticated methods for MS calculation that model the exact correlation structure between CpGs, as are available for PGS (52), are yet to be developed. In our study, we have selected the best performing score for each trait based on the currently available largest EWAS. With larger discovery EWASs, the optimal selection approach for CpGs is also expected to change.

Birth weight and BMI are physical characteristics, whereas prenatal maternal smoking and own smoking are commonly labeled as exposures and behavioral traits. Birth weight is the least heritable of these traits, while mother's behavior in prenatal maternal smoking consists of an exposure whose genetic contribution is genetically transmittable to offspring, who inherit 50% of mothers' genes. All four complex traits are influenced by genetic variants and environmental factors, although some have argued that behavioral traits are more distal and less directly under biological control than physical traits. Polygenic signals from PGS and MS are composites of signals from different sources that are a result of different combinations of underlying biological processes. Notwithstanding the gap in our understanding about biological processes between the polygenic signals and phenotypes and exposure outcomes, the hypothesis-free approaches from GWAS and EWAS allow for construction of polygenic and methylation scores that have certain predictive accuracy, as demonstrated in research and that have potential for clinical use (10, 63).

The pathways between genome and complex physical and behavioral traits may pass over many different cascades of biological processes in interplay and interaction with environmental factors. PGS and MS can capture different sources of information, from GWASs and EWASs. PGS will capture only genetic vulnerability for a trait, while MSs may capture, in addition to genetic influences on the trait, environmental and stochastic influences and the effect of the trait on the MS.

Pathway analyses indicate that protein products of genes within birthweight-associated regions in GWAS are enriched for diverse processes including insulin signaling, glucose homeostasis, glycogen biosynthesis and chromatin remodeling (14). Birthweight-associates CpGs are among sites that have previously been linked to prenatal maternal smoking and mother's BMI before pregnancy (15). Genes annotated to BMI-associated SNPs are mostly enriched among genes involved in neurogenesis and more generally in the development of the central nervous system (16). Cell type-specific gene expression analysis identified enrichment of brain cell types in BMI (64). These findings suggest that BMI could be considered as a behavioral trait and not only metabolic one. Genes annotated to BMI-associated CpGs play role in adipose tissue biology, insulin resistance, inflammation, as well as metabolic, cardiovascular, respiratory and neoplastic disease (12).

Smoking-initiation-associated genes are involved in dopaminergic and glutamatergic neurotransmission among several regions in the central nervous system related to addictive behavior (17). Many CpGs overlap in both in newborns exposed to prenatal maternal smoking and smoking adults (including cg05575921 (*AHRR*) indicative to smoking exposure in many studies) are implicated in numerous neurological pathways, embryogenesis, and various developmental pathways (51). Unique pathways observed in newborns include xenobiotic-related pathways, cytochrome P450 and uridine-glucoronosyltansferases involved in metabolism of nicotine and other compounds of tobacco smoke (51), and pathways associated to susceptibility to orofacial clefts (19). Unique pathways observed in adults EWAS are enriched for variants associated in GWAS with smoking-related disease, including osteoporosis, colorectal cancers, and chronic obstructive pulmonary disease.

The limitations of our study relate to measurement reliability and the missing data for some phenotypes, which reduced the study power. The largest number of missing data was for prenatal maternal smoking in adults (47%, $N = 1{,}133$). Women may underreport smoking during pregnancy (24), although in NTR the prevalence of maternal SDP was 19.5 % for the mothers of young twins, which is in line with the prevalence reported in the general Dutch population (48). Birth weight data were missing for many participants who joined the NTR as adults (17%, $N = 404$). In children, there were varying time differences between DNA methylation and BMI measurement. Our sensitivity analysis showed that MS prediction accuracy was not affected by this difference. In the prediction models, we adjusted for age, sex, and cellular composition of samples, hence the predictive performance of MSs reported in this paper is over and above the effects of age, sex, and cellular composition. We recognize that the MSs, and their ability to predict the phenotypes that we study are likely to be impacted by other factors, such as gestational age for birth weight and prenatal maternal smoking, BMI and amount of cigarettes in smoking exposure and vice versa smoking in BMI. Further explorations of potential confounders and mediators will be valuable.

The combination of PGS and MS is a tool to address research questions, such as mediation by DNA methylation of the effect of certain exposures on a trait of interest, where a score based on multiple CpGs may increase the power of such studies compared to a single CpG site.

Lifestyle variables, such as smoking behavior, are often assessed in epidemiologic studies by interviews or questionnaires, and individuals may hide their smoking status or adults may not know if their mother smoked during pregnancy. In such cases, the use of epigenetic profiles can serve as biomarkers and be applied an alternative of survey data (11). Further, the MSs also have potential to be used in risk stratification and disease risk prediction. For example, BMI MSs were shown to predict type 2 diabetes beyond traditional risk factors including BMI and waist–hip ratio (12).

In conclusion, this study illustrates the value of combining PGS with MS for complex trait and exposure prediction. The results of our study provide new insights into the predictive

performance of PGS and MS for different traits, across different tissues and ages. Because we analyzed buccal data in NTR children and blood data in NTR adults, the current study could not distinguish between age and tissue as cause for the differences in predictive performance of the scores in the two groups. To make a better distinction between differences caused by age or tissue type, future studies that can create PGS and MS based on both blood and buccal data in children and adults are warranted and ideally both tissues are available for the same individuals. Furthermore, the predictive performance of MSs in blood and buccal methylation data may improve if MSs will be created based on EWA studies performed in the same type of tissue collected at the same age with larger sample size, with other approaches rather than weighted score approach (e.g., machine learning), and the prediction of traits may be further improved by adding information from additional omics levels. Future follow-up studies should investigate relationships between the DNA sequence and DNA methylation in complex traits and exposure outcomes.

DATA AVAILABILITY STATEMENT

The datasets analyzed in the current study are available from the Netherlands Twin Register on reasonable request. The HumanMethylation450 BeadChip data from the NTR are available as part of the Biobank-based Integrative Omics Studies (BIOS) Consortium in the European Genome-phenome Archive (EGA), under the accession code EGAD00010000887. Analysis code is available upon request from the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Center, Amsterdam, an Institutional Review Board certified by the U.S. Office of Human Research Protections (IRB number IRB00002991 under Federal-wide Assurance FWA00017598; IRB/institute codes, NTR 03-180). Adult participants or participants' parents/legal guardians provided informed consent.

BIOS CONSORTIUM (BIOBANK-BASED INTEGRATIVE OMICS STUDY)

Management Team

Bastiaan T. Heijmans (chair)[1], Peter A.C. 't Hoen[2], Joyce van Meurs[3], Aaron Isaacs[4], Rick Jansen[5], Lude Franke[6].

Cohort Collection

Dorret I. Boomsma[7], René Pool[7], Jenny van Dongen[7], Jouke J. Hottenga[7] (Netherlands Twin Register); Marleen MJ van Greevenbroek[8], Coen D.A. Stehouwer[8], Carla J.H. van der Kallen[8], Casper G. Schalkwijk[8] (Cohort study on Diabetes and Atherosclerosis Maastricht); Cisca Wijmenga[6], Lude Franke[6], Sasha Zhernakova[6], Ettje F. Tigchelaar[6] (LifeLines Deep); P. Eline Slagboom[1], Marian Beekman[1], Joris Deelen[1], Diana van Heemst[9] (Leiden Longevity Study); Jan H. Veldink[10], Leonard H. van den Berg[10] (Prospective ALS Study Netherlands); Cornelia M. van Duijn[4], Bert A. Hofman[11], Aaron Isaacs[4], André G. Uitterlinden[3] (Rotterdam Study).

Data Generation

Joyce van Meurs (Chair)[3], P. Mila Jhamai[3], Michael Verbiest[3], H. Eka D. Suchiman[1], Marijn Verkerk[3], Ruud van der Breggen[1], Jeroen van Rooij[3], Nico Lakenberg[1].

Data Management and Computational Infrastructure

Hailiang Mei (Chair)[12], Maarten van Iterson[1], Michiel van Galen[2], Jan Bot[13], Dasha V. Zhernakova[6], Rick Jansen[5], Peter van 't Hof[12], Patrick Deelen[6], Irene Nooren[13], Peter A.C. 't Hoen[2], Bastiaan T. Heijmans[1], Matthijs Moed[1].

Data Analysis Group

Lude Franke (Co-Chair)[6], Martijn Vermaat[2], Dasha V. Zhernakova[6], René Luijk[1], Marc Jan Bonder[6], Maarten van Iterson[1], Patrick Deelen[6], Freerk van Dijk[14], Michiel van Galen[2], Wibowo Arindrarto[12], Szymon M. Kielbasa[15], Morris A. Swertz[14], Erik. W van Zwet[15], Rick Jansen[5], Peter-Bram 't Hoen (Co-Chair)[2], Bastiaan T. Heijmans (Co-Chair)[1].

1. Molecular Epidemiology Section, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, Netherlands
2. Department of Human Genetics, Leiden University Medical Center, Leiden, Netherlands
3. Department of Internal Medicine, ErasmusMC, Rotterdam, Netherlands
4. Department of Genetic Epidemiology, ErasmusMC, Rotterdam, Netherlands
5. Department of Psychiatry, VU University Medical Center, Neuroscience Campus Amsterdam, Amsterdam, Netherlands
6. Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, Netherlands
7. Department of Biological Psychology, VU University Amsterdam, Neuroscience Campus Amsterdam, Amsterdam, Netherlands
8. Department of Internal Medicine and School for Cardiovascular Diseases (CARIM), Maastricht University Medical Center, Maastricht, Netherlands
9. Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, Netherlands
10. Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands
11. Department of Epidemiology, ErasmusMC, Rotterdam, Netherlands
12. Sequence Analysis Support Core, Leiden University Medical Center, Leiden, Netherlands
13. SURFsara, Amsterdam, Netherlands
14. Genomics Coordination Center, University Medical Center Groningen, University of Groningen, Groningen, Netherlands
15. Medical Statistics Section, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, Netherlands

AUTHOR CONTRIBUTIONS

VO, VR, JD, and DB: conceptualization and writing—original draft. VO, VR, and JD: prediction analysis. GW, EG, and DB: funding acquisition. FH: sample collection. CB, LL, and GW: phenotype data management. VO and VR: calculation MS. EE, JB, and J-JH: genotyping and genotype data processing. RP and J-JH: calculation PGS. DB and JD: supervision. BIOS consortium: Quality control and analysis pipelines for DNA methylation arrays. All authors: writing—review and editing.

FUNDING

We acknowledge funding from the Netherlands Organization for Scientific Research (NWO): Biobanking and Biomolecular Research Infrastructure (BBMRI–NL, 184.033.111) and the BBMRI-NL funded BIOS Consortium (NWO184.021.007); and Aggression in Children: Unraveling gene-environment interplay to inform Treatment and InterventiON strategies project (ACTION). ACTION received funding from the European Union Seventh Framework Program (FP7/2007-2013) under grant agreement no 602768. The Netherlands Twin Register is supported by grant NWO 480-15-001/674: Netherlands Twin Registry Repository: researching the interplay between genome and environment, the Avera Institute for Human Genetics and by multiple grants from the Netherlands Organization for Scientific Research (NWO). Genotyping was made possible by grants from NWO/SPI 56-464-14192, Genetic Association Information Network (GAIN) of the Foundation for the National Institutes of Health, Rutgers University Cell and DNA Repository, the Avera Institute, Sioux Falls (USA) and the National Institutes of Health (NIH R01 HD042157-01A1, MH081802, Grant Opportunity grants 1RC2 MH089951 and 1RC2 MH089995) and European Research Council (ERC-230374). Epigenetic data were generated at the HUMAN GENOMICS FACILITY (HUGE-F) at EUR. JD was supported by NWO Large Scale infrastructures, X-Omics (184.034.019). DB acknowledges the Royal Netherlands Academy of Science Professor Award (PAH/6635).

ACKNOWLEDGMENTS

We would like to thank the twins and their family members for their participation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2021.688464/full#supplementary-material

REFERENCES

1. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. *Nat Rev Genet.* (2002) 3:872–82. doi: 10.1038/nrg932
2. Polderman TJC, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet.* (2015) 47:702–9. doi: 10.1038/ng.3285
3. van Dongen J, Slagboom PE, Draisma HHM, Martin NG, Boomsma DI. The continuing value of twin studies in the omics era. *Nat Rev Genet.* (2012) 13:640–53. doi: 10.1038/nrg3243
4. Yong SY, Raben TG, Lello L, Hsu SDH. Genetic architecture of complex traits and disease risk predictors. *Sci Rep.* (2020) 10:12055. doi: 10.1038/s41598-020-68881-8
5. Harden KP, Koellinger PD. Using genetics for social science. *Nat Hum Behav.* (2020) 4:567–76. doi: 10.1038/s41562-020-0862-5
6. Smith-Woolley E, Selzam S, Plomin R. Polygenic score for educational attainment captures DNA variants shared between personality traits and educational achievement. *J Pers Soc Psychol.* (2019) 117:1145–63. doi: 10.1037/pspp0000241
7. Khera A, Chaffin M, Wade K, Zahid S, Brancale J, Xia R, et al. Polygenic prediction of weight and obesity trajectories from birth to adulthood. *Cell.* (2019) 177:587–96.e9. doi: 10.1016/j.cell.2019.03.028
8. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am J Hum Genet.* (2019) 104:21–34. doi: 10.1016/j.ajhg.2018.11.002
9. Toulopoulou T, Zhang X, Cherny S, Dickinson D, Berman KF, Straub RE, et al. Polygenic risk score increases schizophrenia liability through cognition-relevant pathways. *Brain.* (2018) 142:471–85. doi: 10.1093/brain/awy279
10. Hüls A, Czamara D. Methodological challenges in constructing DNA methylation risk scores. *Epigenetics.* (2020) 15:1–11. doi: 10.1080/15592294.2019.1644879
11. Sugden K, Hannon EJ, Arseneault L, Belsky DW, Broadbent JM, Corcoran DL, et al. Establishing a generalized polyepigenetic biomarker for tobacco smoking. *Transl Psychiatry.* (2019) 9:92. doi: 10.1038/s41398-019-0430-9
12. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature.* (2017) 541:81–86. doi: 10.1038/nature20784
13. Rask-Andersen M, Karlsson T, Ek WE, Johansson Å. Gene-environment interaction study for BMI reveals interactions between genetic factors and physical activity, alcohol consumption and socioeconomic status. *PLOS Genet.* (2017) 13:e1006977. doi: 10.1371/journal.pgen.1006977
14. Horikoshi M, Beaumont RN, Day FR, Warrington NM, Kooijman MN, Fernandez-Tajes J, et al. Genome-wide associations for birth weight and correlations with adult disease. *Nature.* (2016) 538:248–52. doi: 10.1038/nature19806
15. Küpers LK, Monnereau C, Sharp GC, Yousefi P, Salas LA, Ghantous A, et al. Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight. *Nat Commun.* (2019) 10:1893. doi: 10.1038/s41467-019-09671-3
16. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. *Hum Mol Genet.* (2018) 27:3641–49. doi: 10.1093/hmg/ddy271
17. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet.* (2019) 51:237–44. doi: 10.1038/s41588-018-0307-5
18. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic signatures of cigarette smoking. *Circ Cardiovasc Genet.* (2016) 9:436–47. doi: 10.1161/CIRCGENETICS.116.001506
19. Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, et al. DNA methylation in newborns and maternal smoking in pregnancy: genome-wide consortium meta-analysis. *Am J Hum Genet.* (2016) 98:680–96. doi: 10.1016/j.ajhg.2016.02.019
20. Reed ZE, Suderman MJ, Relton CL, Davis OSP, Hemani G. The association of DNA methylation with body mass index: distinguishing

between predictors and biomarkers. *Clin Epigenetics.* (2020) 12:50. doi: 10.1186/s13148-020-00841-5
21. Shah S, Bonder MJ, Marioni RE, Zhu Z, McRae AF, Zhernakova A, et al. Improving phenotypic prediction by combining genetic and epigenetic associations. *Am J Hum Genet.* (2015) 97:75–85. doi: 10.1016/j.ajhg.2015.05.014
22. Hamilton OKL, Zhang Q, McRae AF, Walker RM, Morris SW, Redmond P, et al. An epigenetic score for BMI based on DNA methylation correlates with poor physical health and major disease in the Lothian Birth Cohort. *Int J Obes.* (2019) 43:1795–802. doi: 10.1038/s41366-018-0262-3
23. Richmond RC, Suderman M, Langdon R, Relton CL, Smith GD. DNA methylation as a marker for prenatal smoke exposure in adults. *Int J Epidemiol.* (2018) 47:1120–30. doi: 10.1093/ije/dyy091
24. Reese SE, Zhao S, Wu MC, Joubert BR, Parr CL, Håberg SE, et al. DNA methylation score as a biomarker in newborns for sustained maternal smoking during pregnancy. (2016) 125:760–6. doi: 10.1289/EHP333
25. Bollepalli S, Korhonen T, Kaprio J, Anders S, Ollikainen M. EpiSmokEr: a robust classifier to determine smoking status from DNA methylation data. *Epigenomics.* (2019) 11:1469–86. doi: 10.2217/epi-2019-0206
26. Elliott HR, Tillin T, McArdle WL, Ho K, Duggirala A, Frayling TM, et al. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clin Epigenetics.* (2014) 6:4. doi: 10.1186/1868-7083-6-4
27. Nwanaji-Enwerem JC, Cardenas A, Chai PR, Weisskopf MG, Baccarelli AA, Boyer EW. Relationships of long-term smoking and moist snuff consumption with a DNA methylation age relevant smoking index: an analysis in buccal cells. *Nicotine Tob Res Off J Soc Res Nicotine Tob.* (2019) 21:1267–73. doi: 10.1093/ntr/nty156
28. Zhang Y, Schöttker B, Florath I, Stock C, Butterbach K, Holleczek B, et al. Smoking-associated DNA methylation biomarkers and their predictive value for all-cause and cardiovascular mortality. *Environ Health Perspect.* (2016) 124:67–74. doi: 10.1289/ehp.1409020
29. Mendelson MM, Marioni RE, Joehanes R, Liu C, Hedman ÅK, Aslibekyan S, et al. Association of body mass index with DNA methylation and gene expression in blood cells and relations to cardiometabolic disease: a mendelian randomization Approach. *PLoS Med.* (2017) 14:e1002215. doi: 10.1371/journal.pmed.1002215
30. Gao X, Zhang Y, Breitling LP, Brenner H. Relationship of tobacco smoking and smoking-related DNA methylation with epigenetic age acceleration. *Oncotarget.* (2016) 7:46878–89. doi: 10.18632/oncotarget.9795
31. Nwanaji-Enwerem JC, Colicino E. DNA methylation–based biomarkers of environmental exposures for human population studies. *Curr Environ Heal Rep.* (2020) 7:121–8. doi: 10.1007/s40572-020-00269-2
32. Willemsen G, de Geus EJC, Bartels M, van Beijsterveldt, CEMT, Brooks AI, et al. The Netherlands twin register biobank: a resource for genetic epidemiological studies. *Twin Res Hum Genet.* (2010) 13:231–45. doi: 10.1375/twin.13.3.231
33. Boomsma DI. Aggression in children: unravelling the interplay of genes and environment through (epi)genetics and metabolomics. *J Pediatr Neonatal Individ Med.* (2015) 4:e040251. https://doi.org/10.7363/040251
34. Bartels M, Hendriks A, Mauri M, Krapohl E, Whipp A, Bolhuis K, et al. Childhood aggression and the co-occurrence of behavioural and emotional problems: results across ages 3–16 years from multiple raters in six cohorts in the EU-ACTION project. *Eur Child Adolesc Psychiatry.* (2018) 27:1105–21. doi: 10.1007/s00787-018-1169-1
35. Hagenbeek FA, Roetman PJ, Pool R, Kluft C, Harms AC, van Dongen J, et al. Urinary amine and organic acid metabolites evaluated as markers for childhood aggression: the ACTION Biomarker study. *Front Psychiatry.* (2020) 11:165. doi: 10.3389/fpsyt.2020.00165
36. van Dongen J, Hagenbeek FA, Suderman M, Roetman PJ, Sugden K, Chiocchetti AG, et al. DNA methylation signatures of aggression and closely related constructs: a meta-analysis of epigenome-wide studies across the lifespan. *Mol Psychiatry.* (2021) 26(6):1–15. doi: 10.1038/s41380-020-00987-x
37. van Dongen J, Nivard MG, Willemsen G, Hottenga, J.-J., Helmer Q, et al. Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun.* (2016) 7:11115. doi: 10.1038/ncomms11115
38. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet.* (2017) 49:131–38. doi: 10.1038/ng.3721
39. van Iterson M, Tobi EW, Slieker RC, den Hollander W, Luijk R, Slagboom PE, et al. MethylAid: visual and interactive quality control of large Illumina 450k datasets. *Bioinformatics.* (2014) 30:3435–7. doi: 10.1093/bioinformatics/btu566
40. Chen Y, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics.* (2013) 8:203–9. doi: 10.4161/epi.23470
41. Francioli LC, Menelaou A, Pulit SL, van Dijk F, Palamara PF, Elbers CC, et al. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet.* (2014) 46:818–25. doi: 10.1038/ng.3021
42. Fortin, J.-P., Labbe A, Lemire M, Zanke BW, Hudson TJ, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* (2014) 15:503. doi: 10.1186/s13059-014-0503-2
43. Sinke L, van Iterson M, Cats D, Slieker R, Heijmans B. (2019). *DNAmArray: Streamlined Workflow for the Quality Control, Normalization, and Analysis of Illumina Methylation Array Data*. Available online at: https://zenodo.org/record/3355292#.XYj9KC2ZPx4 (accessed July 4, 2021).
44. Ligthart L, van Beijsterveldt CEM, Kevenaar ST, de Zeeuw E, van Bergen E, Bruins S, et al. The Netherlands twin register: longitudinal research based on twin and twin-family designs. *Twin Res Hum Genet.* (2019) 22:623–36. doi: 10.1017/thg.2019.93
45. van Dongen J, Nivard MG, Baselmans BML, Zilhão NR, Ligthart L, Heijmans BT, et al. Epigenome-wide association study of aggressive behavior. *Twin Res Hum Genet.* (2015) 18:686–98. doi: 10.1017/thg.2015.74
46. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.* (2016) 17:208. doi: 10.1186/s13059-016-1066-1
47. van Beijsterveldt CEM, Groen-Blokhuis M, Hottenga JJ, Franić S, Hudziak JJ, Lamb D, et al. The Young Netherlands Twin Register (YNTR): longitudinal twin and family studies in over 70,000 children. *Twin Res Hum Genet.* (2013) 16:252–67. doi: 10.1017/thg.2012.118
48. Dolan C, Geels L, Vink J, van beijsterveldt T, Neale M, Bartels M, et al. Testing causal effects of maternal smoking during pregnancy on offspring's externalizing and internalizing behavior. *Behav Genet.* (2016) 46:378–88. doi: 10.1007/s10519-015-9738-2
49. Zheng SC, Webster AP, Dong D, Feber A, Graham DG, Sullivan R, et al. A novel cell-type deconvolution algorithm reveals substantial contamination by immune cells in saliva, buccal and cervix. *Epigenomics.* (2018) 10:925–40. doi: 10.2217/epi-2018-0037
50. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet.* (2006) 2:e190. doi: 10.1371/journal.pgen.0020190
51. Sikdar S, Joehanes R, Joubert BR, Xu C-J, Vives-Usano M, et al. Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking. *Epigenomics.* (2019) 11:1487–500. doi: 10.2217/epi-2019-0066
52. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am J Hum Genet.* (2015) 97:576–92. doi: 10.1101/015859.
53. Beck JJ, Pool R, van de Weijer M, Chen X, Krapohl E, Gordon SD, et al. Genetic meta-analysis of twin birth weight shows high genetic correlation with singleton birth weight. *Hum Mol Genet.* (2021). doi: 10.1093/hmg/ddab121. [Epub ahead of print].
54. Lee SH, Goddard ME, Wray NR, Visscher PM. A better coefficient of determination for genetic profile analysis. *Genet Epidemiol.* (2012) 36:214–24. doi: 10.1002/gepi.21614
55. Markunas CA, Xu Z, Harlid S, Wade PA, Lie RT, Taylor JA, et al. Identification of DNA methylation changes in newborns related to maternal smoking during pregnancy. *Environ Health Perspect.* (2014) 122:1147–53. doi: 10.1289/ehp.1307892

56. Küpers LK, Xu X, Jankipersadsing SA, Vaez A, la Bastide-van Gemert S, Scholtens S, et al. DNA methylation mediates the effect of maternal smoking during pregnancy on birthweight of the offspring. *Int J Epidemiol.* (2015) 44:1224–37. doi: 10.1093/ije/dyv048
57. Richmond RC, Simpkin AJ, Woodward G, Gaunt TR, Lyttleton O, McArdle WL, et al. Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Hum Mol Genet.* (2015) 24:2201–17. doi: 10.1093/hmg/ddu739
58. Lee KWK, Richmond R, Hu P, French L, Shin J, Bourdon C, et al. Prenatal Exposure to maternal cigarette smoking and dna methylation: epigenome-wide association in a discovery sample of adolescents and replication in an independent cohort at birth through 17 years of age. *Environ Health Perspect.* (2015) 123:193–9. doi: 10.1289/ehp.1408614
59. Wiklund P, Karhunen V, Richmond RC, Parmar P, Rodriguez A, De Silva M, et al. DNA methylation links prenatal smoking exposure to later life health outcomes in offspring. *Clin Epigenetics.* (2019) 11:97. doi: 10.1186/s13148-019-0683-4
60. Ladd-Acosta C, Shu C, Lee BK, Gidaya N, Singer A, Schieve LA, et al. Presence of an epigenetic signature of prenatal cigarette smoke exposure in childhood. *Environ Res.* (2016) 144:139–48. doi: 10.1016/j.envres.2015.11.014
61. Elks CE, den Hoed M, Zhao JH, Sharp SJ, Wareham NJ, Loos RJF, et al. Variability in the heritability of body mass index: a systematic review and meta-regression. *Front Endocrinol.* (2012) 3:29. doi: 10.3389/fendo.2012.00029
62. Villareal DT, Apovian CM, Kushner RF, Klein S. Obesity in older adults: technical review and position statement of the American Society for Nutrition and NAASO. the obesity society. *Am J Clin Nutr.* (2005) 82:923–34. doi: 10.1093/ajcn/82.5.923
63. Abdellaoui A, Verweij KJH. Dissecting polygenic signals from genome-wide association studies on human behaviour. *Nat Hum Behav.* (2021) 13:1–9. doi: 10.1038/s41562-021-01110-y
64. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh R, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet.* (2015) 47:1228–35. doi: 10.1038/ng.3404

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 September 2021
doi: 10.3389/fgene.2021.688488

Haploinsufficiency of the Attention-Deficit/Hyperactivity Disorder Risk Gene *St3gal3* in Mice Causes Alterations in Cognition and Expression of Genes Involved in Myelination and Sialylation

Olga Rivero[1,2†‡], Judit Alhama-Riba[1†‡], Hsing-Ping Ku[1], Matthias Fischer[3‡], Gabriela Ortega[1], Péter Álmos[1‡], David Diouf[1], Daniel van den Hove[4] and Klaus-Peter Lesch[1,4,5]*

[1] Division of Molecular Psychiatry, Center of Mental Health, University of Würzburg, Würzburg, Germany, [2] Biomedical Network Research Centre on Mental Health (CIBERSAM), Valencia, Spain, [3] Department of Psychiatry, Psychosomatics and Psychotherapy, University of Würzburg, Würzburg, Germany, [4] Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University, Maastricht, Netherlands, [5] Laboratory of Psychiatric Neurobiology, Institute of Molecular Medicine, I.M Sechenov First Moscow State Medical University, Moscow, Russia

Genome wide association meta-analysis identified *ST3GAL3,* a gene encoding the beta-galactosidase-alpha-2,3-sialyltransferase-III, as a risk gene for attention-deficit/hyperactivity disorder (ADHD). Although loss-of-function mutations in *ST3GAL3* are implicated in non-syndromic autosomal recessive intellectual disability (NSARID) and West syndrome, the impact of *ST3GAL3* haploinsufficiency on brain function and the pathophysiology of neurodevelopmental disorders (NDDs), such as ADHD, is unknown. Since *St3gal3* null mutant mice display severe developmental delay and neurological deficits, we investigated the effects of partial inactivation of *St3gal3* in heterozygous (HET) knockout (*St3gal3*$^{\pm}$) mice on behavior as well as expression of markers linked to myelination processes and sialylation pathways. Our results reveal that male *St3gal3* HET mice display cognitive deficits, while female HET animals show increased activity, as well as increased cognitive control, compared to their wildtype littermates. In addition, we observed subtle alterations in the expression of several markers implicated in oligodendrogenesis, myelin formation, and protein sialylation as well as cell adhesion/synaptic target glycoproteins of ST3GAL3 in a brain region- and/or sex-specific manner. Taken together, our findings indicate that haploinsufficiency of ST3GAL3 results in a sex-dependent alteration of cognition, behavior and markers of brain plasticity.

Keywords: *St3gal3*, sialyltransferase, sialic acid, psychiatric disorders, attention-deficit/hyperactivity disorder (ADHD), prefrontal cortex, hippocampus, mouse model

OPEN ACCESS

Edited by:
Elena Martín-García,
Pompeu Fabra University, Spain

Reviewed by:
Simon Trent,
Keele University, United Kingdom
Nathalie Berube,
Western University (Canada), Canada
Lay Khoon Too,
The University of Sydney, Australia

***Correspondence:**
Olga Rivero
olga.rivero@uv.es;
olgarivero1980@gmail.com

† These authors have contributed equally to this work

‡ Present address:
Olga Rivero,
INCLIVA Biomedical Research Institute, Fundación Investigación Hospital Clínico de Valencia, Valencia, Spain
Judit Alhama-Riba,
Department for Developmental Origins of Disease, University Medical Center Utrecht Brain Center, Utrecht, Netherlands
Matthias Fischer,
Department for Psychiatry and Psychotherapy, University of Rostock, Rostock, Germany
Péter Álmos,
Department of Psychiatry, Faculty of Medicine, University of Szeged, Szeged, Hungary

Specialty section:
This article was submitted to Behavioral and Psychiatric Genetics, a section of the journal Frontiers in Genetics

Received: *30 March 2021*
Accepted: *31 August 2021*
Published: *28 September 2021*

INTRODUCTION

Neurodevelopmental disorders (NDDs) constitute frequent causes of disability among children and adolescents world-wide. Being the most common NDD, attention-deficit/hyperactivity disorder (ADHD, MIM #143465), is defined as a clinically heterogeneous syndrome comprising developmentally inappropriate inattention, hyperactivity, increased impulsivity, and motivational/emotional dysregulation. The syndromal dimensions of inattention and impulsivity are increasingly being recognized as highly persistent into adulthood and are associated with a considerable risk for failure in psychosocial adaptation as well as psychiatric co-morbidity, such as depression and substance use disorders (Bush, 2010; Faraone and Larsson, 2019).

Despite evidence for high heritability, the multifactorial and polygenic nature of ADHD has rendered it challenging to unravel the genetic contribution to each syndromal dimension. Recently, a collaborative effort which assessed a sample of 20,183 individuals with ADHD and 35,191 controls identified ADHD risk variants surpassing genome-wide significance in 12 independent loci (Demontis et al., 2019). This meta-analysis shed light on a locus located on chromosome 1, whose most significant single nucleotide polymorphism (SNP), rs11420276 ($p = 2.14 \times 10^{-13}$), is located in *ST3GAL3,* a gene encoding the Golgi membrane enzyme beta-galactosidase-alpha-2,3-sialyltransferase-III (ST3GAL3). Of note, a recent genome-wide methylome screening by Walton et al. (2017) also identified *ST3GAL3* as differentially methylated during the course of ADHD. It was pointed out that gene-by-environment (GxE) interaction may moderate *ST3GAL3* expression, impacting epigenetic programming of brain development and maturation. Remarkably, *ST3GAL3* was implicated in general cognitive function in a GWAS meta-analysis (Davies et al., 2018). Together, these findings identify *ST3GAL3* as a putative risk gene for ADHD, specifically for associated impairment of cognitive performance.

In support of this notion, loss-of-function mutations in *ST3GAL3* were previously reported to result in severe non-syndromic autosomal recessive intellectual disability (NSARID), a familial disorder characterized by low cognitive abilities associated with difficulty in learning basic language and motor skills, as well as West syndrome (WS), a severe infantile seizure disorder associated with developmental regression and intellectual disability that can occur as an autosomal recessive familial disorder (Hu et al., 2011; Edvardson et al., 2013; Schnaar et al., 2014; Indellicato et al., 2020), compellingly suggesting that ST3GAL3 function is critical for normal development and functioning of the brain.

ST3GAL3 is mainly involved in the formation of the sialyl epitope on glycoproteins, a process with critical impact on brain function which also determines the functional specificity of glycans (Schnaar et al., 2014). Sialic acids (Sia) are 9-carbon backbone sugars with chemical properties that are well suited for molecular recognition and communication (Yoo et al., 2015). Brain sialoglycans are dominated by gangliosides, sialic acid-bearing glycosphingolipids that constitute approximately 75% of the brain's sialic acid, whereas sialoglycoproteins constitute the remaining 25% (Yoo et al., 2015).

Sialoglycans are biosynthesized by sialyltransferases (STs), which catalyze the transfer of Sia from its activated nucleotide sugar donor cytidine monophosphate (CMP-Sia) to the terminus of an oligosaccharide chain of a glycoprotein or glycolipid (Wang, 2009; Schnaar et al., 2014). Based on their acceptor substrates, the type of glycosidic linkage they generate (α2-3, α2-6, and α2-8) and their primary saccharide receptor [galactose (Gal), N-Acetyl-D-galactosamine (GalNAc) or Sia] (Wang, 2009; Schnaar et al., 2014), 20 mammalian STs have been described and categorized into four families: β-galactoside α2,3-sialyltransferases (ST3Gal), β-galactoside α2,6-sialyltransferases (ST6Gal), alpha-N-acetylgalactosaminide α2,6-sialyltransferases (ST6GalNAc), and α2,8-sialyltransferases or polysialyltransferases (ST8Sia). Particularly, ST3Gal3 utilizes both Galß1,3GlcNAc and Galß1,4GlcNAc structures (with preference for Galß1,3GlcNAc structure), mainly on glycoproteins as acceptor substrates (Yoo et al., 2015). Additionally, polysialyltransferases (in vertebrates ST8Sia2 and ST8Sia4) attach polysialic acid (PSA) to complex glycans with the first sialic acid.

Among several other proteins, PSA is added to the neural cell adhesion molecule-1 (NCAM1). Attachment of PSA to NCAM1 influences its properties, altering cell motility, neurogenesis, axon outgrowth as well as structural and synaptic plasticity (Galuska et al., 2008; Wang, 2009; Nacher et al., 2013; Schnaar et al., 2014). PSA is also added to other molecules, such as synaptic cell adhesion molecule-1 (SynCAM1, CADM1), Neuropilin-2 or voltage-gated ion channels (Ednie and Bennett, 2012; Mühlenhoff et al., 2013), in specific cell types. The effects of sialylation on these glycoproteins are yet unknown.

Although ST3GAL3 is well characterized in the context of cancer or lung inflammation, little is known about its function in the brain. Yoo et al. (2015) assessed neuroanatomy, brain histology, biochemistry as well as behavior in *St3gal2* and *St3gal3* single and double-null mutant mice. Notably, the authors demonstrated that sialylation of sialoglycoproteins is decreased approximately 50% in *St3gal3*$^{-/-}$ mice. Moreover, the authors reported alterations in different markers linked to oligodendrogenesis and the process of myelination (MBP, Olig2, and MAG). Additionally, they reported reduced motor coordination, marked cognitive deficits and hyperactive behavior in *St3gal3* single-null mice.

The notion that *ST3GAL3* inactivation causes severe intellectual disability with distinguishable developmental, neurological and neuropsychiatric phenotypes led us to predict that more subtle allelic variation of *ST3GAL3* expression may increase the vulnerability to ADHD and related conditions, or predispose individuals to the effects of other genetic and/or environmental risk factors. Since *St3gal3* null mutant mice display severe developmental delay, neurological deficits, including spontaneous seizures, and overall reduced viability, thus limiting behavioral and neurobiological assessment, we here investigated the effects of partial inactivation of *St3gal3* in male and female heterozygous knockout (*St3gal3*$^{\pm}$) mice with focus

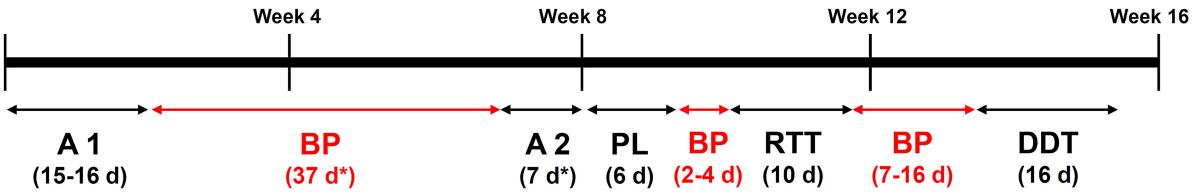


FIGURE 1 | Timeline of the different behavioral experiments performed in this study by the male and female cohort. A, adaptation phase; BP, break phase; PL, place learning; RTT, reaction time task; DDT, delayed discounting task; d, days. *Males only.

on behavior as well as expression of markers linked to sialylation and myelination pathways.

MATERIALS AND METHODS

Generation of *St3gal3*-Deficient Mice

Experiments were performed using a constitutive *St3gal3*-deficient mouse line (B6.129-*St3gal3*^{TM1Jxm}/J) that had been previously generated by Cre-loxP gene targeting to produce the deletion of exon 2 that encodes the ST3GAL3 transmembrane domain (Ellies et al., 2002). Heterozygous *St3gal3*-deficient (*St3gal3*$^{\pm}$, in the following referred to as HET) were obtained from the Jackson Laboratory (JAX, Bar Harbor, ME, United States) and re-derived by *in vitro* fertilization and implantation in C57BL/6J dams at the Center for Experimental Molecular Medicine (ZEMM), University of Würzburg. After the rederivation, the mouse line was maintained by breeding HET males with *St3gal3*$^{+/+}$ wildtype (WT) females. The genotyping of the *St3gal3* locus was done by polymerase chain reaction (PCR) as described on the Jackson Lab website.[1] In order to confirm the deletion and deficiency of ST3GAL3, we measured mRNA and protein levels in the mouse brain by quantitative real time PCR (qRT-PCR), as described in 2.5.

Animal Husbandry

All mice used for this study were housed in standard Makrolon cages at the animal facility of the ZEMM (University Hospital of Würzburg), under a 12 h light/dark cycle with food and water *ad libitum* unless specified otherwise. After weaning, male mice were kept in type IV Makrolon cages (rat type) in groups of 9–10 animals of mixed genotypes, in order to follow the group distribution needed for the behavioral experiments in the Intellicage System (see section "Behavioral Assessment"). This was done to avoid future aggressive reactions between adult male animals coming from different litters due to the territorial behavior that naturally occurs in male mice (Koolhaas et al., 2013), which can be avoided if animals from different litters are mixed at an early age. In the case of female mice, they were kept in type III Makrolon cages in genotype-mixed groups of 3–5 per cage until the start of the behavioral experiments. Animals used for the molecular experiments were housed following the same strategy described above. Experimental procedures of the IntelliCage System were approved by the local authorities of the Government of Lower Franconia and performed in accordance with the guidelines for animal care and use provided by the European Parliament and Council Directive (2010/63/EU).

[1]https://www.jax.org/

Behavioral Assessment

Behavioral tests were performed to assess activity, memory and learning abilities as well as cognitive flexibility and impulsivity in two separate cohorts of adult male (n = 18; 9 WT and 9 HET) and female (n = 18; 8 WT and 10 HET) mice. Due to the impossibility to test females and males simultaneously, both sexes had to be tested at different time points: the male cohort was 3–4 months of age at the beginning of the testing; the female cohort was 5–7 months old. However, we consider that the age difference between males and females, which is around 2–3 months, is not as high as being considered a main confounding factor (Benice et al., 2006). In addition, all the animals used in the behavioral experiment are full adults that have not started with any aging process which may have underlid strong differences in their performance in the Intellicage (Yanai and Endo, 2021).

Behavioral assessment was carried out using the IntelliCage System (New Behavior, Zurich, Switzerland), an automated home cage system (Krackow et al., 2010; Kobayashi et al., 2013; Fischer et al., 2017). For that purpose, mice were anesthetized with isoflurane and radio frequency identification transponders (Planet ID, Essen, Germany) were implanted subcutaneously in the dorso-cervical region. Animals were returned to their home cages and stayed in observation for recovery for a few days. Afterward, mice were transferred to the Intellicage System, where the following phases/behavioral tasks took place (**Figure 1**).

Adaptation Phase

The adaptation phase allowed us to investigate exploratory behavior in a novel environment. During the first 7 days (spontaneous activity subphase, SA), all IntelliCage doors were open, providing free access to all eight water bottles (two in each corner). Animals that were unable to drink from the bottles 48 h after being introduced in the Intellicage were excluded from the experiment. Afterward, the nosepoke (np) phase took place; during np phase, the mice had to make a np to be able to drink from the water bottles. Likewise, those animals that were unable to learn the task and drink from the bottles after 48 h were excluded from the experiment. In this module, the total number of visits/h, visits without np or licks/h, visits with np/h, visits with licks/h, median number of np and licks per visit, median duration of nosepokes as well as median duration of licking were measured.

Due to technical reasons (forced experimental discontinuity in the animal facility), male mice had to be removed from the Intellicage after the Adaptation Phase and returned to normal cages. Once again in the Intellicage and before the start of the place learning phase, they were subjected to a second Adaptation period, which consisted only in np phase (**Figure 1**). Results concerning this second Adaptation period are shown in **Supplementary Figure 1**.

Place-Learning Phase

This phase allowed us to evaluate learning and memory abilities as well as cognitive flexibility. In this phase, water was only available for each mouse in one specific corner (termed as "correct"), while the rest of the corners (termed as "incorrect") did not open the doors even if the mouse had made a nosepoke. Correct corners were assigned depending on the corner preferences of each mouse during the adaptation phase; the least visited corner was assigned as the correct one. To avoid learning by imitation, balance between corner preferences was also taken into account. The first place learning module lasted for three days. This was followed by 3 days of place reversal, in which the rewarded corner was swapped to the opposite one, allowing us to evaluate cognitive flexibility. In this module, correct and incorrect responses were measured and the proportion of correct responses was calculated. The learning criterion was set at 35% of correct responses, that is, random expectation plus 10%.

Reaction Time-Task (RTT)

The reaction time-task (RTT) evaluates motor impulsivity and was divided into training and testing phases, which lasted 3 and 5 days, respectively. During the training part, the first np on each visit was followed by a delay period of variable (random) duration (0.5, 1.5, or 2.5 s), after which three green LEDs were switched on exactly on the same side where the mouse had made that first np. Afterward, the door opened for 5 s, allowing the mouse to drink as a response to a np. Any nosepoke during the delay period was considered as a premature or anticipatory response. During the training, premature responses had no consequence. However, during the testing phase, premature responses stopped the trial, requiring the mouse to leave the corner and start again. Therefore, animals had to learn to wait until the LED was on before making a second np.

Impulsivity was measured by the proportion of premature and correct responses (visits without premature responses).

Delay Discounting Task (DDT)

The delay discounting task (DDT) evaluates cognitive impulsivity. During this task, one of the bottles at each corner was filled with 2% sucrose water, while the other contained only water. In addition, while the doors leading to the water bottle could be opened immediately upon a np, there was a delay between the np and the opening of the sucrose doors; this delay started at 0.5 s and increased by 0.5 s every 24 h, reaching 8 s of delay at the end. Additional np at the closed sucrose door had no effect. Preference for sucrose and water bottles was measured. Increasing delays are expected to reduce the preference for sucrose; however, an earlier breakdown of sucrose preference in one of the groups would be taken as evidence for increased cognitive impulsivity.

Behavioral Data Analysis

The FlowR software (XBehavior, Bänk, Switzerland) was used to analyze the raw data obtained directly from the IntelliCage System software. To analyze the data from the adaptation phase, a saturated linear model was implemented to ANOVA, followed by linear model multiple comparisons as previously described in Fischer et al. (2017). With regard to the data from the place learning phase, RTT and DDT, the proportions of correct responses and premature responses were compared via generalized linear mixed models using a binomial link function on the logits, as described in Fournier et al. (2012) and Fischer et al. (2017). *F*-tests were taken for significance testing to take individual variation into account for error variance estimation, as proportions tended to be heavily overdispersed (Krackow et al., 2010; Fischer et al., 2017). For designs measuring individuals repeatedly, subject identity was added as random effect to the within-subject and between-subject fixed effects (giving essentially repeated measures ANOVA). Additionally, the raw data obtained by FlowR software were also used to analyze and plot the proportion of visits with nosepokes and proportion of visits with licks in the adaptation phase, using GraphPad Prism version 6.0. The normality of the data sets was verified using the Shapiro–Wilk normality test. Once a normal distribution was confirmed, genotype effects were analyzed by a two-tailed unpaired *t*-test; the level of significance was set to 0.05 (two-tailed) in all the statistical analyses. GraphPad Prism was also used to plot the sucrose preference in a timeline (delay-dependent) manner.

Brain Tissue Preparation for Molecular Analysis

A cohort of *St3gal3* littermate HET (19 females, 7 males) and WT (13 females, 8 males) adult mice that were behaviorally naïve (age between 3 and 5 months old) were anesthetized with a lethal dose of isoflurane and sacrificed by cervical dislocation. Brains were dissected, frozen immediately in dry ice-cold isopentane and subsequently stored at −80°C until dissection of the areas of interest. Afterward, the left and right parts of the prefrontal cortex, hippocampus and striatum were separated on a pre-cooled plate, collected in separate vials and stored at −80°C for later use (see below).

RNA Isolation and Quantitative RT-PCR

This analysis was done in the male and female sample sets. RNA was isolated from the left hemispheres of each brain region with the miRNeasy Mini Kit (Qiagen, Hilden, Germany). RNA integrity was verified by agarose gel electrophoresis. Complementary DNA (cDNA) was synthesized using the QuantiTect Reverse Transcription Kit (Qiagen, Hilden, Germany) using 600ng total RNA per reaction. We next proceeded with the qRT-PCR analysis to check the expression of the target genes *Cadm1*, *Cnp*, *Ncam1*, *St3gal3*, *St8Sia2*, *St8Sia4*, *Mag*, *Mog*, *Mbp*, *Cspg4*, *Olig2*, *Sox10*, *Plp1*, and *Plp2* (primer sequences for target genes can be consulted

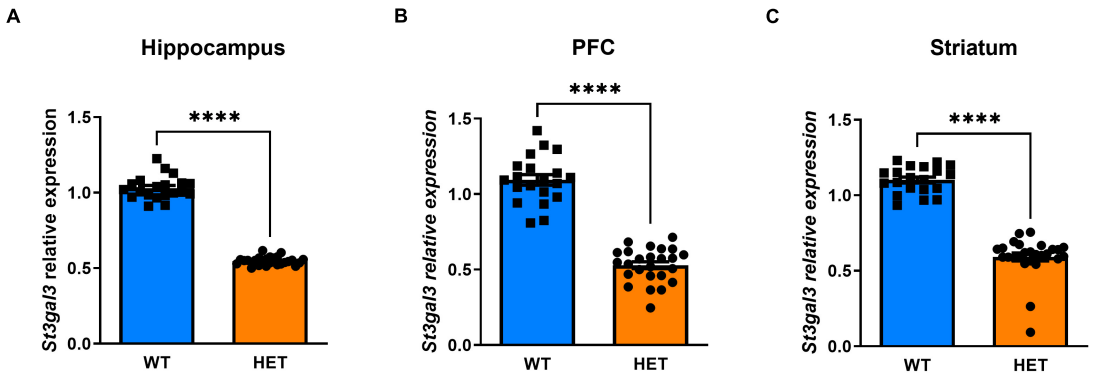


FIGURE 2 | Relative expression of *St3gal3* in hippocampus, PFC and striatum of HET and WT mice. **(A)** Relative levels of *St3gal3* mRNA in the hippocampus (HET: $n = 21$, WT: $n = 13$). **(B)** Levels of *St3gal3* mRNA expression in the PFC (HET: $n = 19$, WT: $n = 14$). **(C)** Levels of *St3gal3* mRNA expression in the striatum (HET: $n = 20$, WT: $n = 14$). In all analyzed brain regions, a strongly significant decrease of around 50% in *St3gal3* expression was observed in heterozygote animals compared to their wildtype littermates (hippocampus: $t = 29.59$, $p < 0.0001$; PFC: $t = 13.97$, $p < 0.0001$; striatum: $t = 14.62$, $p < 0.0001$). Data were analyzed by Student's *t*-test and presented as mean ± SEM. ****$p < 0.0001$. WT, wildtype; HET, heterozygote.

in **Supplementary Table 1**). *Rplp0*, *Ubc*, *B2m*, *Gapdh*, and *Actb2* served as reference genes for normalization of the gene expression. Reactions were run as previously described (Forero et al., 2020). PCR efficiencies for each primer pair were determined with LinRegPCR (Ruijter et al., 2009), using fluorescence raw data extracted from the CFX software. qbase + qPCR analysis software (Biogazelle, Zwijnaarde, Belgium) was used for calculation of relative expression values of all the target genes, using the threshold cycle values (Cq) from Bio-Rad CFX Maestro software (Bio-Rad Laboratories, Feldkirchen, Germany) and the primer efficiency values from LinReqPCR (Academic Medical Center, Amsterdam, Netherlands). These values were normalized by the most stable reference genes (Gapdh, *Actb* and *Rplp0* in hippocampus; *Ubc*, *B2m*, and *Rplp0* in PFC and striatum).

Western Blot

Prefrontal cortex, hippocampal and striatal regions of the right hemispheres of the male sample set (WT = 8; HET = 7) were homogenized by sonication in Radio-Immunoprecipitation Assay (RIPA) buffer (Sigma-Aldrich, Darmstadt, Germany), supplemented with protease and phosphatase inhibitor tablets (Roche, Basel, Switzerland). A total of 20 μg of protein were loaded on 4–12% Bis-Tris protein gels (Thermo Fisher Scientific, Dreieich, Germany), electroblotted to a 0.45 μm pore nitrocellulose membrane (Thermo Fisher Scientific, Dreieich, Germany) and probed using one of the following primary antibodies, always in combination with either rabbit anti-α-tubulin antibody or goat anti- beta-tubulin (Abcam, Cambridge, United Kingdom) as loading controls: goat polyclonal anti-NCAM1/CD56 antibody (1:2,000, R&D Systems, Minneapolis, MA, United States), mouse monoclonal anti-PSA antibody (1:500, #NBP2-52710, Novus, Lingen, Germany), rabbit polyclonal anti-SynCAM1/2/3 (1:2000, Synaptic Systems, Göttingen, Germany), rabbit polyclonal anti-Myelin PLP (1:5,000, #ab28486, Abcam, Cambridge, United Kingdom), rabbit polyclonal anti-MBP (1:1,000, #PD004, MBL, Woburn, MA, United Kingdom). Finally, membranes were incubated with appropriate secondary fluorescent antibodies (all at 1:10,000; Li-Cor Biotechnology, Bad Homburg, Germany). Fluorescent band detection was conducted using the Fusion FX imaging system (Vilber Lourmat, Eberhardzell, Germany) and quantified using the Image Studio Lite software (Li-Cor Biotechnology, Bad Homburg, Germany). Target protein levels were normalized by the internal loading control and referred to average WT levels.

Molecular Data Analysis

All statistical analyses as well as data graphs were performed using Prism version 9.2.0 (GraphPad Software, San Diego, CA, United States). Normal distribution was checked using the Shapiro–Wilk normality test and data were analyzed by two-way ANOVA or Student t-test with a level of significance of 0.05 (two-tailed).

RESULTS

Heterozygous *St3gal3*-Deficient Mice Display Intermediate Levels of *St3gal3* mRNA

We first determined the expression of *St3gal3* in several brain regions of HET mice compared to WT littermates. By using PCR primers located in exons 2 (forward) and 3 (reverse) targeting specifically the *St3gal3* WT allele, we observed a gene dose dependent reduction of ∼50% in *St3gal3* levels in HET mice when compared to WT littermates, in all analyzed brain regions ($p < 0.0001$) (**Figure 2**).

Heterozygous *St3gal3*-Deficient Mice Present With Sex-Dependent Alterations in Cognitive and Locomotor Activity

We evaluated the effects of *St3gal3* haploinsufficiency using the Intellicage System, which allows the assessment of different

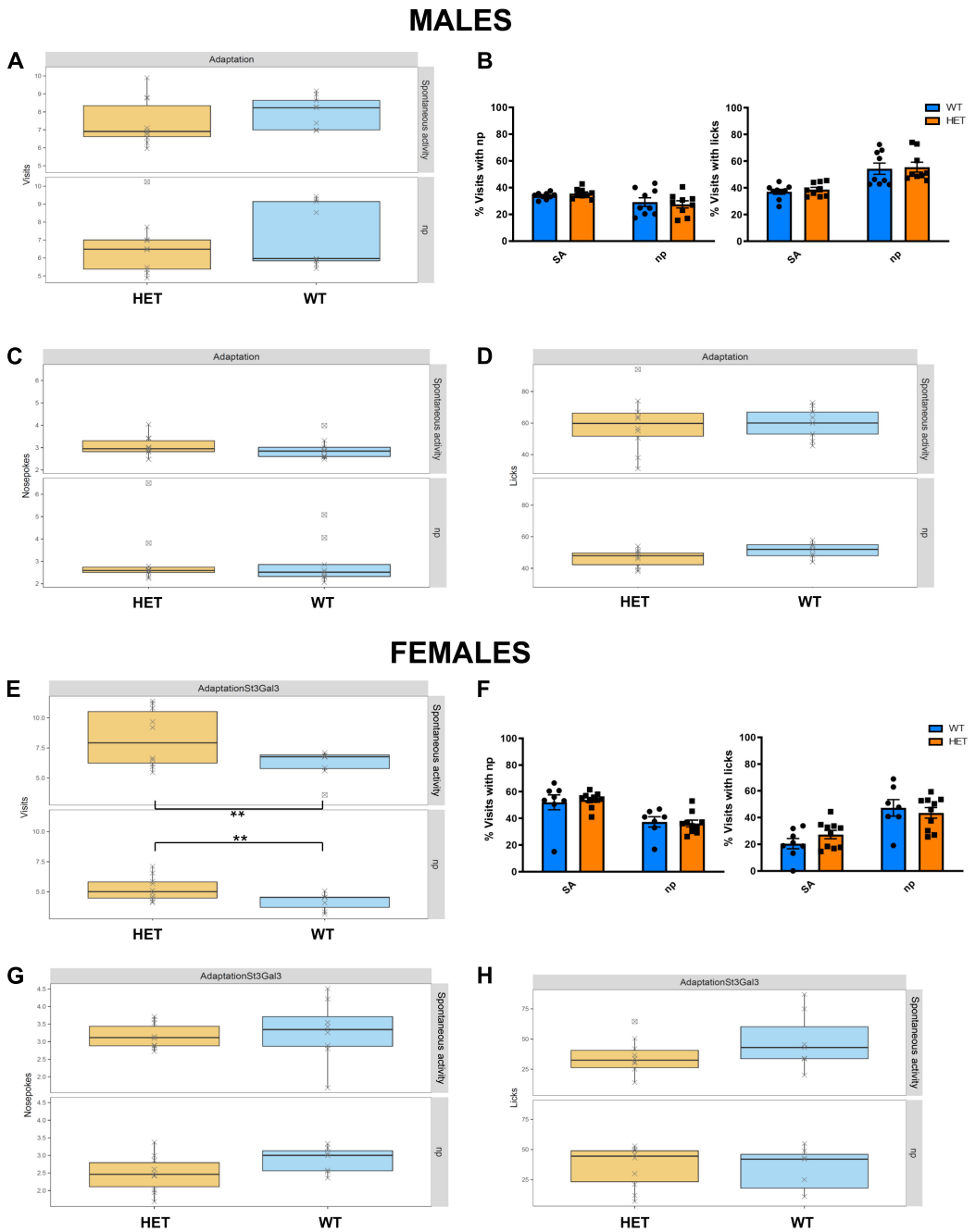


FIGURE 3 | Assessment of exploratory behavior and association between nose-poking and water reward in HET mice. Adaptation phase was divided in spontaneous activity (SA) and nose-poke (Np) subphases, where the following parameters were assessed in the male cohort (n = 9/genotype) and the female cohort (n = 8 WT and 10 HET): **(A,E)** Number of visits/h; **(B,F)** proportion of visits with nosepokes and visits with licks; **(C,G)** mean number of nosepokes during visits with nosepokes and without licks; **(D,H)** median number of licks per visit. No genotype effects were observed in the male dataset. By contrast, female HET animals made significantly more visits compared to their WT littermates ($F_{1,31}$ = 9.35, p = 0.0045) Data in panels **(B,F)** were analyzed by two-way ANOVA. Remaining data were analyzed and plotted with FlowR software. All data are presented as mean ± SEM.

behavioral domains that are related to ADHD pathophysiology. Profound differences between the male and the female *St3gal3*-deficient mice were observed. In HET males no significant genotype effects during the adaptation phase were detected. As shown in **Figures 3A–D**, the genotype had no effects on the number of visits, percentage of visits with either np or lick, number of np or number of licks, neither in the free adaptation subphase nor in the np subphase, suggesting unaltered activity and exploratory behavior in HET male mice. In contrast, female HET mice (**Figures 3E–H**) made significantly more visits

per hour compared to their WT littermates during the entire adaptation phase ($p = 0.0046$), a finding that indicates enhanced exploratory behavior. However, the percentage of visits either with np or with np + licks remained unaltered.

While ADHD is frequently associated with reduced cognitive performance, including impaired working memory that results in learning disabilities (Andersen et al., 2013; Tistarelli et al., 2020), *ST3GAL3* was also implicated in by a GWAS meta-analysis on general cognitive function (Davies et al., 2018). This deficit is also observed in null mutant mice (Yoo et al., 2015) and patients with loss-of-function mutations in *ST3GAL3* (Hu et al., 2011; Edvardson et al., 2013; Indellicato et al., 2020). These fact led us to evaluate learning and memory as well as cognitive flexibility in *St3gal3*-deficient mice in the place learning task. This module was divided into two phases, acquisition and reversal, in which the percentage of correct responses (visits to correct corner) was used to evaluate learning performance of the male and the female cohorts. Although all groups were able to learn the task and reached the learning criterion (set at 35% of correct responses, that is, random expectation plus 10%), HET male mice displayed a reduced averaged proportion of correct responses compared to their WT littermates in acquisition (**Figures 4A,C**; $p = 0.0031$) and reversal (**Figures 4B,D**; $p = 0.0035$); this effect was not observed in the female group, however (**Figures 4E,F**).

In the RTT, which evaluates motor impulsivity (**Figure 5**), a strongly significant effect of the delay on premature and correct responses was observed in both cohorts ($p < 1e\text{-}08$), as expected, with increased premature responses and reduced correct responses when the delay was higher. In addition, a genotype effect was observed in the male cohort in the proportion of correct responses (**Figure 5A**) in the test ($p = 0.0277$) as well as the training ($p = 0.0035$) phases. However, no effects in the number of premature responses were seen. On the other hand, in females, we observed a significant genotype x delay interaction in the proportion of correct responses (**Figure 5C**; $p = 0.0150$), with HET females having an increased number of correct responses only in shorter delays. Finally, significant genotype effect in the number of premature responses during the test phase was also detected in the female cohort (**Figure 5D**; $p = 0.0370$); in this case HET females made less premature responses in the test phase compared to their WT littermates; this finding was delay-independent and suggests an increased inhibitory control in the HET females.

The behavioral testing ended with the DDT, which assesses cognitive impulsivity (**Supplementary Figure 2**), however, no genotype effects were found.

Heterozygous *St3gal3*-Deficient Mice Display Dysregulation of Genes Involved in Oligodendrogenesis, Myelin Formation and (Poly)sialylation

We next assessed the effect of partial *St3gal3* deficiency on the expression of genes involved in oligodendrogenesis and myelin formation in three key brain regions involved in cognition and ADHD pathophysiology (PFC, hippocampus and striatum). We only found a genotype effect on *Plp2* levels in the PFC ($p = 0.0286$; **Figure 6C**), with HET animals having a reduced expression (average reduction of ∼10%) of this gene compared to WT mice. However, no genotype effects were observed in the hippocampus (**Figure 6A**) and striatum (**Supplementary Figure 3**). In addition, the expression of the oligodendrocyte marker *Cspg4/Ng2* was significantly reduced in the PFC of HET mice ($p = 0.073$; **Figure 7B**, average reduction in HET mice was 14%), an effect that was not observed in the hippocampus (**Figure 7A**). In both cases effects were mainly driven by the HET male dataset (16 and 25% reductions in the expression of *Plp2* and *Cspg4* in HET males compared to WT males, respectively). Trends for a sex effect on hippocampal *Mbp* ($p = 0.062$; **Figure 6A**) as well as PFC *Plp1* ($p = 0.0537$; **Figure 6C**) values were also detected.

Notably, Western blot analysis of MBP and myelin PLP as representative protein markers of myelination (**Figures 6B,D**) revealed no differences in PLP levels, but a significant decrease of ∼50% in PFC MBP levels in HET animals when compared to WT littermates ($p = 0.0409$; **Figure 6D**).

In addition, *St3gal3* deficiency was associated with expression changes of several genes involved in the process of (poly)sialylation. In the hippocampus (**Figure 8A**), we detected a significant genotype effect on *Cadm1* levels, with HET animals displaying higher levels (∼10% increase) compared to their WT littermates ($p = 0.0485$). Moreover, ANOVA detected a genotype main effect ($F_{1,43} = 5.624$, $p = 0.0223$) as well as sex × genotype interaction (s × g: $F_{1,43} = 4.736$, $p = 0.0351$) that affected hippocampal *St8sia2* expression. While *St8sia2* levels were generally higher in HET animals compared to WT, *post hoc* analysis revealed that *St8sia2* expression was particularly upregulated in HET males when compared to HET females (∼14% increase, $p = 0.0481$) and WT males (∼23% increase, $p = 0.0083$). Finally, the hippocampal expression of the other polysialyltransferase included in the study, *St8sia4,* was also significantly increased in HET mice compared to their WT littermates (∼16% increase, $p = 0.0088$).

In the PFC (**Figure 8C**), effects were generally modest, only sex main effects in the expression of *Cadm1* ($p = 0.0239$) and *St8sia4* ($p = 0.0231$) were observed. Trends toward significant effects of genotype and sex on *Ncam1* expression were also seen (sex: $p = 0.0853$; genotype: $p = 0.0916$). As for striatum, a trend toward a sex effect was observed for *St8sia4* (**Supplementary Figure 2**; $p = 0.0657$).

The evidence that HET animals display an upregulation of genes involved or affected by the process of polysialylation led to us to investigate whether the levels of the main targets of polysialylation (NCAM1, SynCAM1) were altered in those mice (**Figure 8**). We also analyzed global PSA levels in polysialylated proteins, but Western blot analyses revealed no differences, neither in the levels of total NCAM1, total SynCAM1 nor in PSA-glycoproteins in the PFC, hippocampus and striatum of HET mice, compared to their WT littermates (**Figure 9**).

DISCUSSION

Our results demonstrate how partial reduction in *St3gal3* expression produces critical changes at the molecular

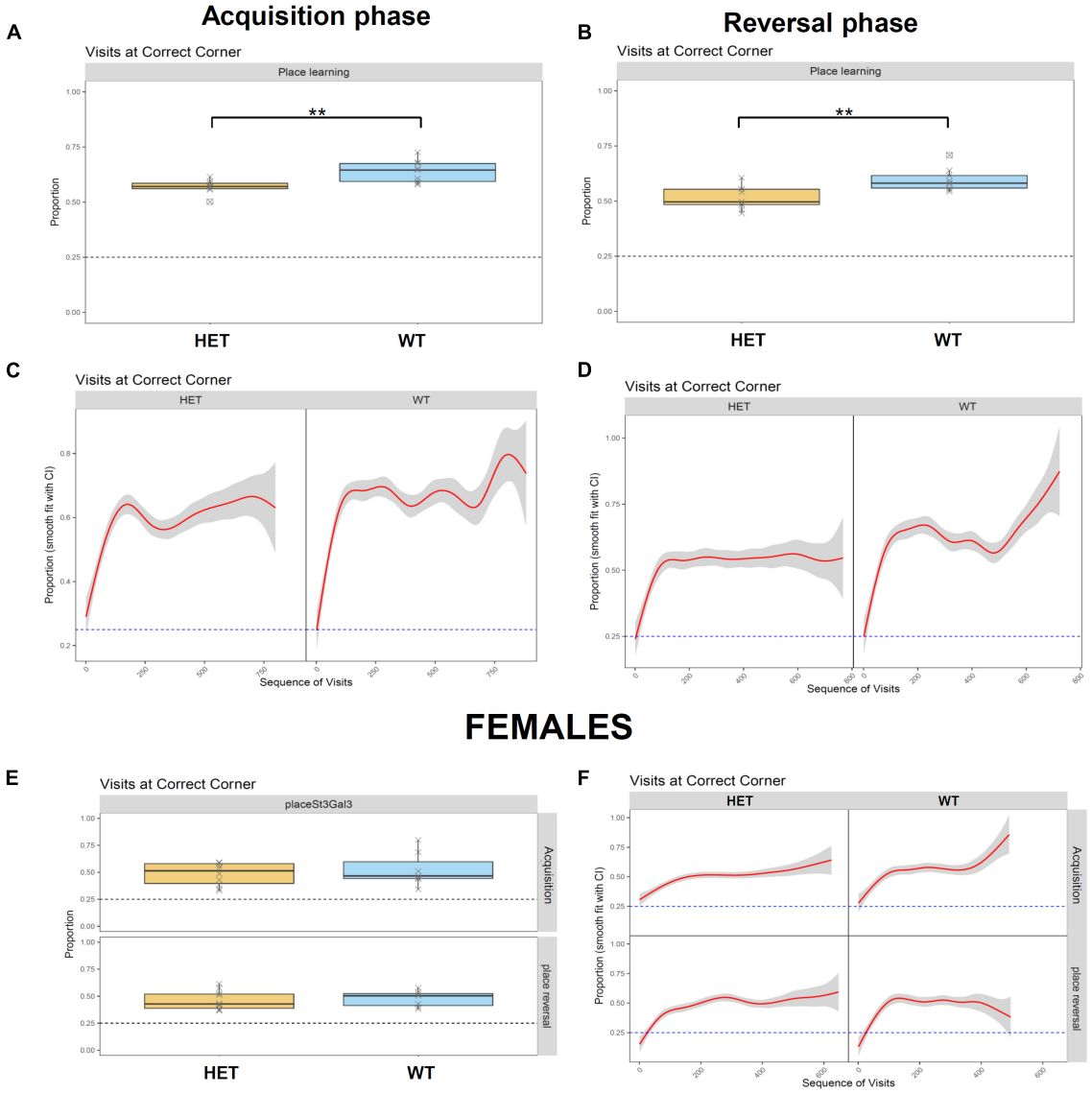


FIGURE 4 | Assessment of learning performance and cognitive flexibility in HET mice. Place learning module was divided in two phases, Acquisition and Reversal, where the percentage of correct responses (visits to the correct corner) was used to evaluate learning performance of the male (n = 9/genotype) and the female (n = 8 WT and 10 HET) cohorts. **(A,B)** HET male mice displayed a reduced proportion of correct responses compared their WT littermates in Acquisition **(A)** ($F_{1,15}$ = 12.35, p = 0,003128) and Reversal **(B)** ($F_{1,15}$ = 11.97, p = 0,003496). **(C,D)** Cumulative visits at the correct corner in Acquisition **(C)** and Reversal **(D)**. **(E)** Averaged proportion of correct responses by the female cohort in Acquisition and Reversal phases, where no genotype effects were observed. **(F)** Cumulative visits at the correct corner by the female cohort in Acquisition and Reversal. Data were analyzed and plotted with FlowR software. Data in panels **(A,B,E)** presented as mean ± SEM. Dotted line on each graph indicates random expectation.

and behavioral level, with a specific involvement of genes central to the myelination process and the formation of PSA-glycoproteins.

Despite representing the most common NDD, the search for genetic risk factors for ADHD has been a considerable challenge. Therefore, the identification of risk variants surpassing genome-wide significance has provided new sources of hypothesis-generating research (Demontis et al., 2019). Among several other ADHD risk genes, *ST3GAL3*, previously related to several severe intellectual disability (ID)-presenting conditions (Hu et al., 2011; Edvardson et al., 2013; Indellicato et al., 2020), encodes an enzyme that is crucial for sialoglycan formation and therefore contributes to critical processes in the mammalian forebrain, such as sialylation of glycoproteins, oligodendrogenesis and ultimately myelin formation, with numerous roles in cell adhesion, synaptic plasticity and neuronal communication (Kolter et al., 2002; Wang, 2009; Audry et al., 2011; Yoo et al., 2015).

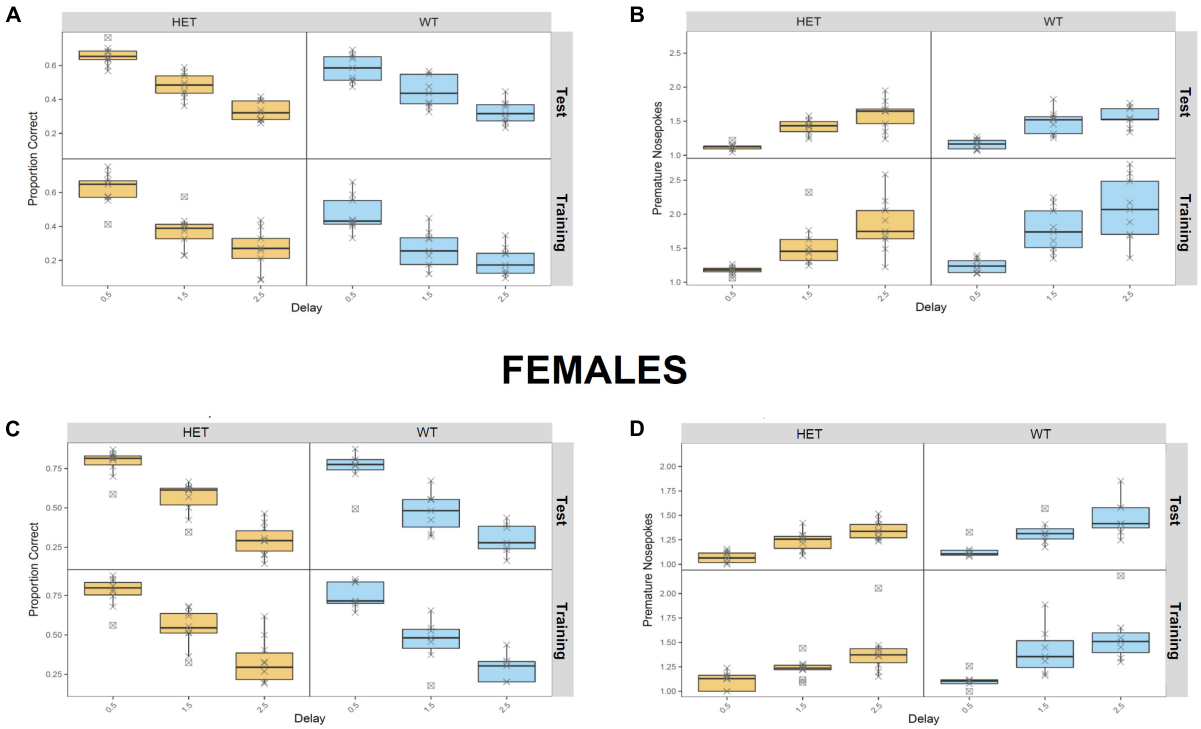


FIGURE 5 | Assessment of motor impulsivity in *St3gal3* HET mice with the reaction time task. **(A,C)** Mean proportion of correct responses (depending on the delay between first and second nosepoke) made by males **(A)** and females **(C)** in the training and test phases. **(B,D)** Average number of premature nosepokes made by males **(B)** and females **(D)** in the training and test phases, depending on the delay between first and second nosepoke. A strongly significant effect of the delay on premature responses and correct responses was observed in both cohorts (p < 1e-08). In males, a genotype effect in the proportion of correct responses **(A)** was also detected in the test ($F_{1,47}$ = 5.16, p = 0.02774) and training ($F_{1,47}$ = 9.475, p = 0.003471) phases, but no effects in the number of premature responses were seen. In females, we observed a significant effect of the genotype x delay interaction in the proportion of correct responses ($F_{2,44}$ = 4.624, p = 0.01504), as well as a significant genotype effect in the number of premature responses during the test phase ($F_{1,15}$ = 5.236, p = 0.03706). Data were analyzed with FlowR using a generalised linear binomial modelling with random error estimation (F-test) to prevent significance overestimation due to overdispersion. Data are shown as mean ± SEM.

In concordance with the human phenotypes caused by loss-of-function mutations, *St3gal3* null mutant mice are characterized by marked cognitive deficits, reduced motor coordination and hyperactive behavior (Yoo et al., 2015) as well as spontaneous non-lethal epileptic seizures (Jamey Marth, personal communication; also see: www.jax.org/strain/006899). All these profound abnormalities might be the consequence of a disrupted myelination process. The question remains how severe neurological and cognitive alterations resulting from ST3GAL3 homozygous inactivation help to understand ADHD-related pathophysiology? Indeed, deficient ST3GAL3 activity may be a key factor in ADHD pathogenesis for several reasons. First, dysregulated myelin formation may compromise the developmental trajectory leading to ADHD (Lesch, 2019). Several diffusion tensor imaging (DTI) studies and meta-analyses revealed alterations in white matter integrity assessed by fractional anisotropy in cortico-striatal tracts and commissural tracts of the corpus callosum, all of them relevant for ADHD pathophysiology as well as to ADHD-associated impairment of cognitive performance (van Ewijk et al., 2012; Chen et al., 2016). Thus, our findings support the notion that a more subtle alteration in ST3GAL3 function, e.g., through common risk variants, may underlie distinct dimensions of ADHD symptomatology, such as inattention or working memory impairment, whereas loss-of-function mutations lead to severe phenotypes. In agreement with this notion, a recent study found that a large set of ID-causing genes, including *ST3GAL3*, contributes to ADHD risk via common variants (Klein et al., 2020).

Experimental animal models represent an invaluable source of information to explore the contribution of individual genes to human disorders (Homberg et al., 2021). However, null mutant models exhibiting a complete inactivation of a given gene may be inadequately informative in the case of genetically complex diseases, in which each risk gene contributes to vulnerability only modestly. We consequently focused our study on *St3gal3* HET mice. Our results show that HET mice display a gene dose-dependent 50% reduction in *St3gal3* expression. Interestingly, such a reduction was sufficient to give rise to expression changes of several genes and a myelination-related protein. In the

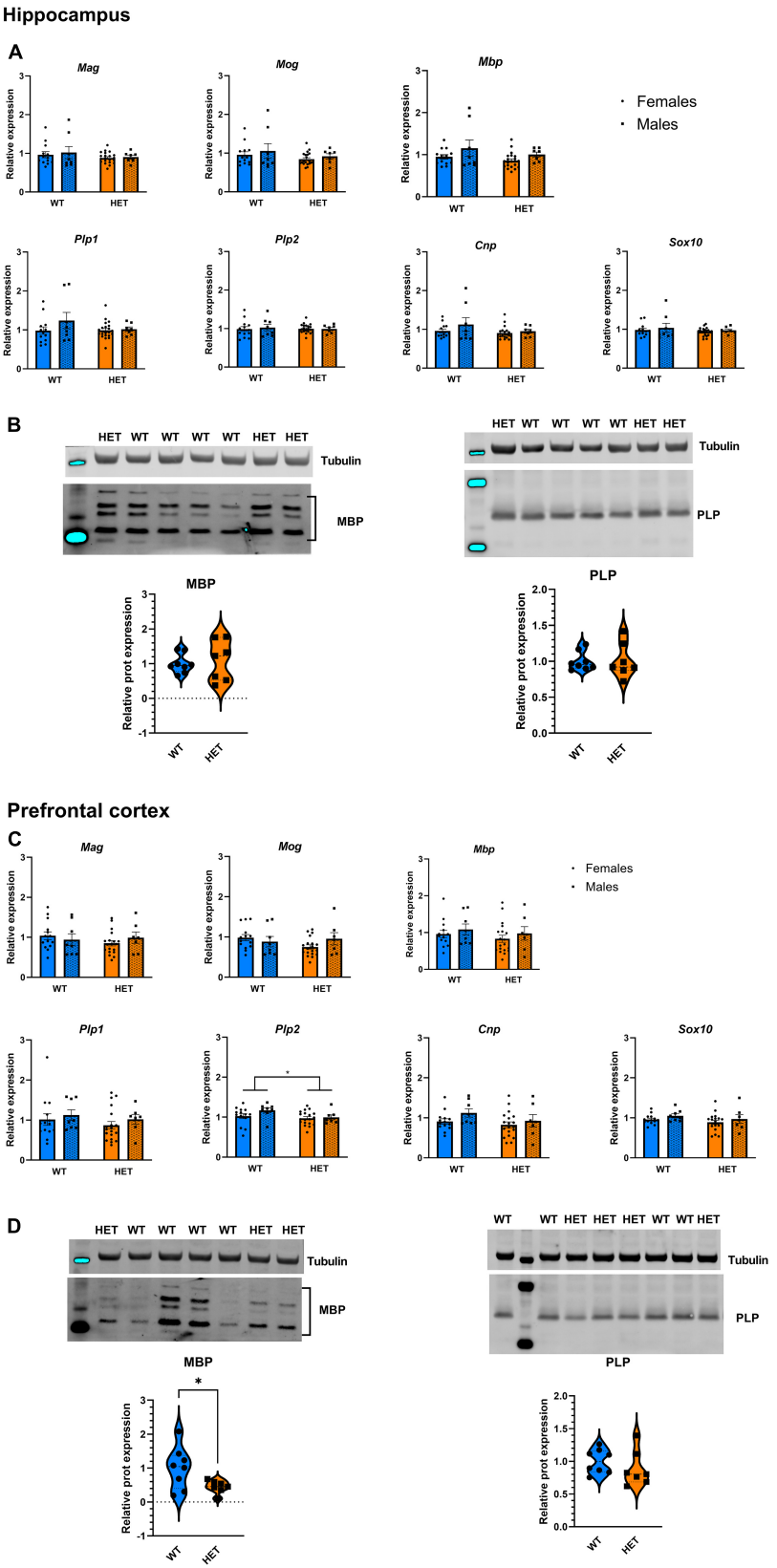


FIGURE 6 | (Continued)

FIGURE 6 | Expression analysis of genes involved in the myelination process. **(A)** Relative mRNA levels of 7 myelination biomarkers in the hippocampus of *St3gal3* mice (HET: $n = 21$, WT: $n = 13$). A trend toward a reduction of *Mbp* expression levels in females was observed (sex: $F_{1,43} = 3.671$, $p = 0.062$). **(B)** Protein levels of the myelination markers MBP and PLP in the hippocampus of *St3gal3* male mice (HET: $n = 7$, WT: $n = 8$). Representative blots are shown in each case. No differences were observed for any of the analyzed proteins. **(C)** Relative mRNA levels of 7 myelin biomarkers in the PFC of *St3gal3* mice (HET: $n = 19$, WT: $n = 14$). A significant genotype effect on *Plp2* levels was found (genotype: $F_{1,42} = 5.142$, $p = 0.0286$). A trend toward a reduction of *Plp1* expression levels in females (regardless from genotype) was also observed (sex: $F_{1,43} = 3.936$, $p = 0.0537$). **(D)** Protein levels of the myelination markers MBP and PLP in the PFC of *St3gal3* male mice (HET: $n = 7$, WT: $n = 8$). Representative blots are shown in each case. A significant reduction in MBP protein levels was observed in HET mice ($t = 2.411$, $p = 0.0409$). qPCR data were analyzed by two-way ANOVA and presented as mean ± SEM. Western blot data were analyzed by Student's *t*-test with Welch's correction and presented as violin plots with median + quartiles. $^{*}p < 0.05$. For simplicity, only genotype effects are indicated in the figures.

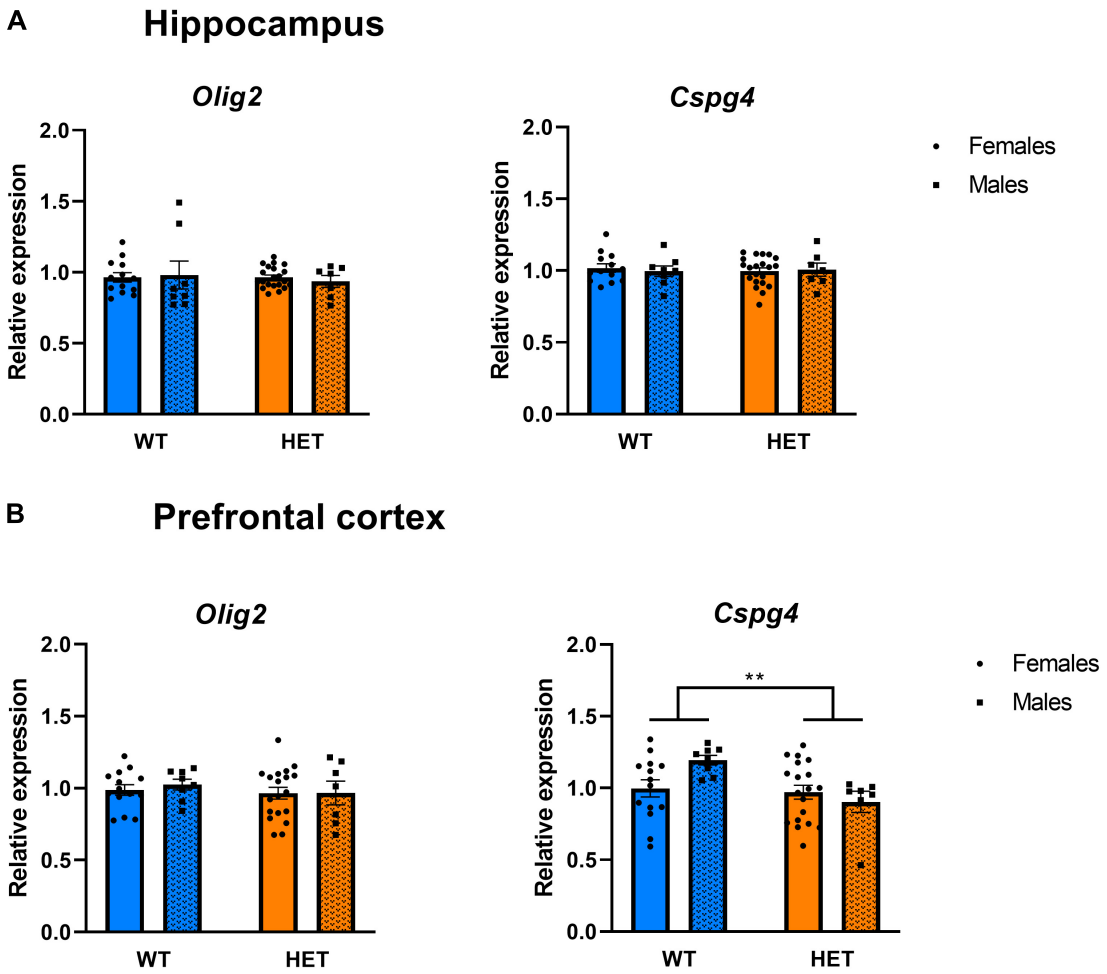


FIGURE 7 | Expression analysis of genes involved in the generation and differentiation of oligodendrocytes. **(A)** Relative mRNA expression of the oligodendrocyte progenitor markers *Olig2* and *Cspg4* in the hippocampus of *St3gal3* mice (HET: $n = 21$, WT: $n = 13$). No significant differences were found. **(B)** Relative mRNA levelsexpression of the oligodendrocyte progenitor markers *Olig2* and *Cspg4* in the PFC (HET: $n = 19$, WT: $n = 14$). A significant effect of genotype on *Cspg4* expression levels was seen (genotype: $F_{1,42} = 7.961$, $p = 0.0073$). Data were analyzed by two-way ANOVA followed by Sidak *post hoc* test and presented as mean ± SEM. $^{**}p < 0.01$.

PFC, the effects of *St3gal3* deficiency were largely restricted to myelination markers, with downregulation of *Plp2* and *Cspg4* as well as a 50% reduction in MBP, a major constituent of the myelin sheath in brain. Consistent with our observation, a previous study revealed that the expression level of MBP, was also reduced in the brain of *St3gal3* null mutant mice, along with a significant reduction in the number of myelinated axons in the corpus callosum (Yoo et al., 2015). The fact that partial loss of St3gal3 resembles the molecular signatures of complete depletion further highlights the critical role of ST3GAL3 function in brain plasticity.

In the hippocampus of *St3gal3*-deficient mice, the strongest expression differences were observed for several genes related to glycoprotein sialylation. Specifically, the two polysialyltransferases, *St8sia4* and *St8sia2* (responsible for the attachment of polySia to several adhesion molecules) were

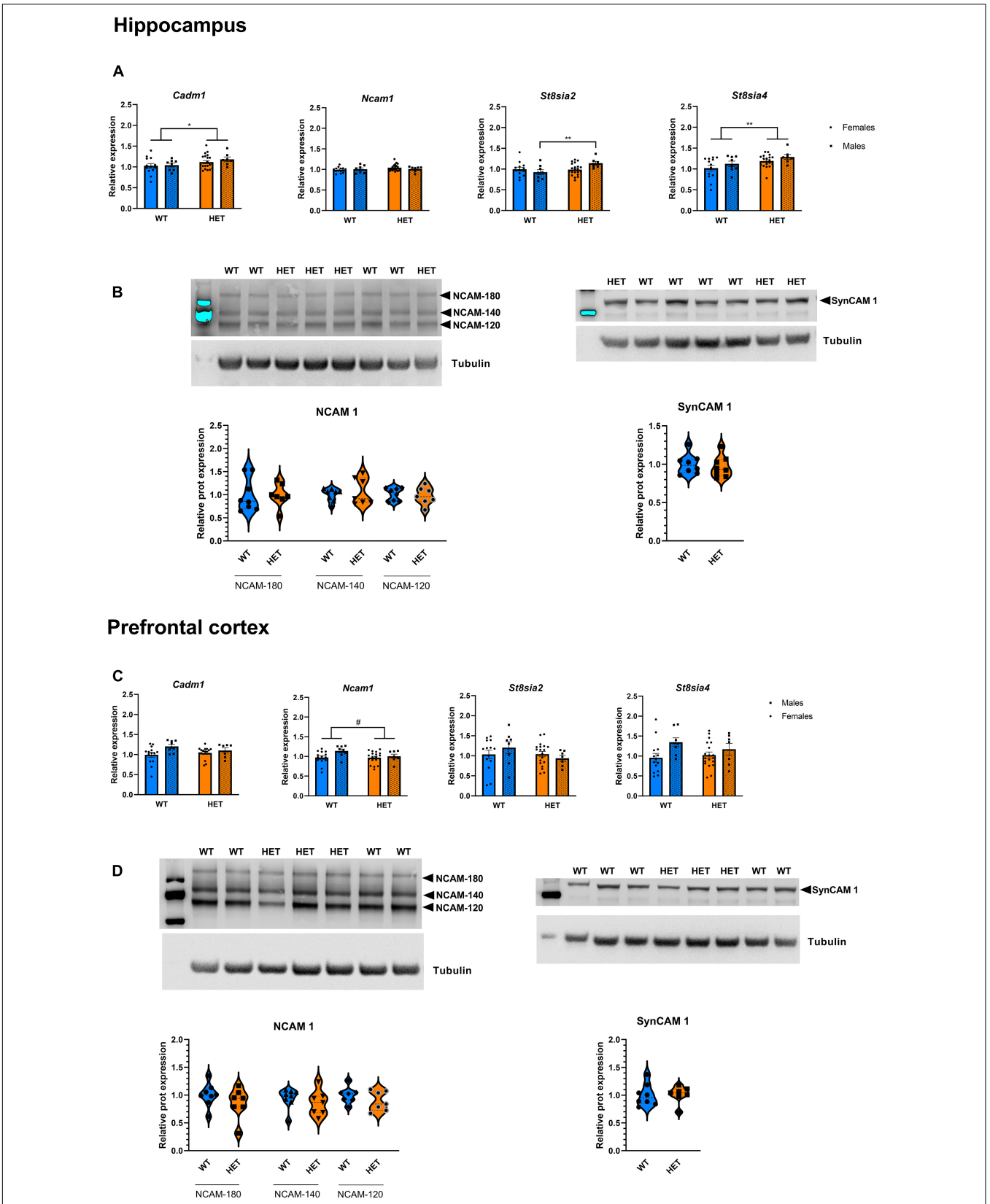


FIGURE 8 | (Continued)

FIGURE 8 | Expression analysis of genes involved in the polysialylation process. **(A)** Relative mRNA level of genes involved in the (poly)sialylation process in the hippocampus of *St3gal3* mice (HET: $n = 21$, WT: $n = 13$). A significant genotype effect on *Cadm1* levels was observed (genotype: $F_{1,43} = 4.124$, $p = 0.0485$). We detected a genotype main effect (genotype: $F_{1,43} = 5.624$, $p = 0.0223$) as well as a significant genotype × sex interaction ($F_{1,43} = 4.736$, $p = 0.0351$) on the levels of *St8sia2* expression; *post hoc* analysis revealed an upregulation in HET male mice compared to WT male mice (*post hoc* $p = 0.0083$) and HET females (*post hoc* $p = 0.0481$). A genotype effect was also observed for *St8sia4* (genotype: $F_{1,43} = 7.538$, $p = 0.0088$). **(B)** Protein levels of the adhesion molecules NCAM (isoforms NCAM-180, NCAM-140, and NCAM-120) and SynCAM 1 in the hippocampus of *St3gal3* male mice (HET: $n = 7$, WT: $n = 8$). Representative blots are shown in each case. No differences were observed for any of the analyzed proteins. **(C)** Relative mRNA level of genes involved in the polysialylation process in the PFC of *St3gal3* mice (HET: $n = 19$, WT: $n = 14$). A significant sex effect was observed in *Cadm1* (sex: $F_{1,42} = 5.495$, $p = 0.0239$). Trends toward sex and genotype effects were found for *Ncam1* (sex: $F_{1,42} = 3.105$, $p = 0.0853$; genotype: $F_{1,42} = 2.981$, $p = 0.0916$). A significant sex effect on *St8sia4* expression was also detected ($F_{1,42} = 5.558$, $p = 0.0231$). **(D)** Protein levels of the adhesion molecules NCAM (isoforms NCAM-180, NCAM-140, and NCAM-120) and SynCAM 1 in the hippocampus of *St3gal3* male mice (HET: $n = 7$, WT: $n = 8$). Representative blots are shown in each case. No differences were observed for any of the analyzed proteins. qPCR data were analyzed by two-way ANOVA followed by Sidak *post hoc* test (when necessary) and presented as mean ± SEM. Western blot data were analyzed by Student's *t*-test with Welch's correction and presented as violin plots with median + quartiles. $^{\#}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$. For simplicity, only genotype effects are indicated in the figures.

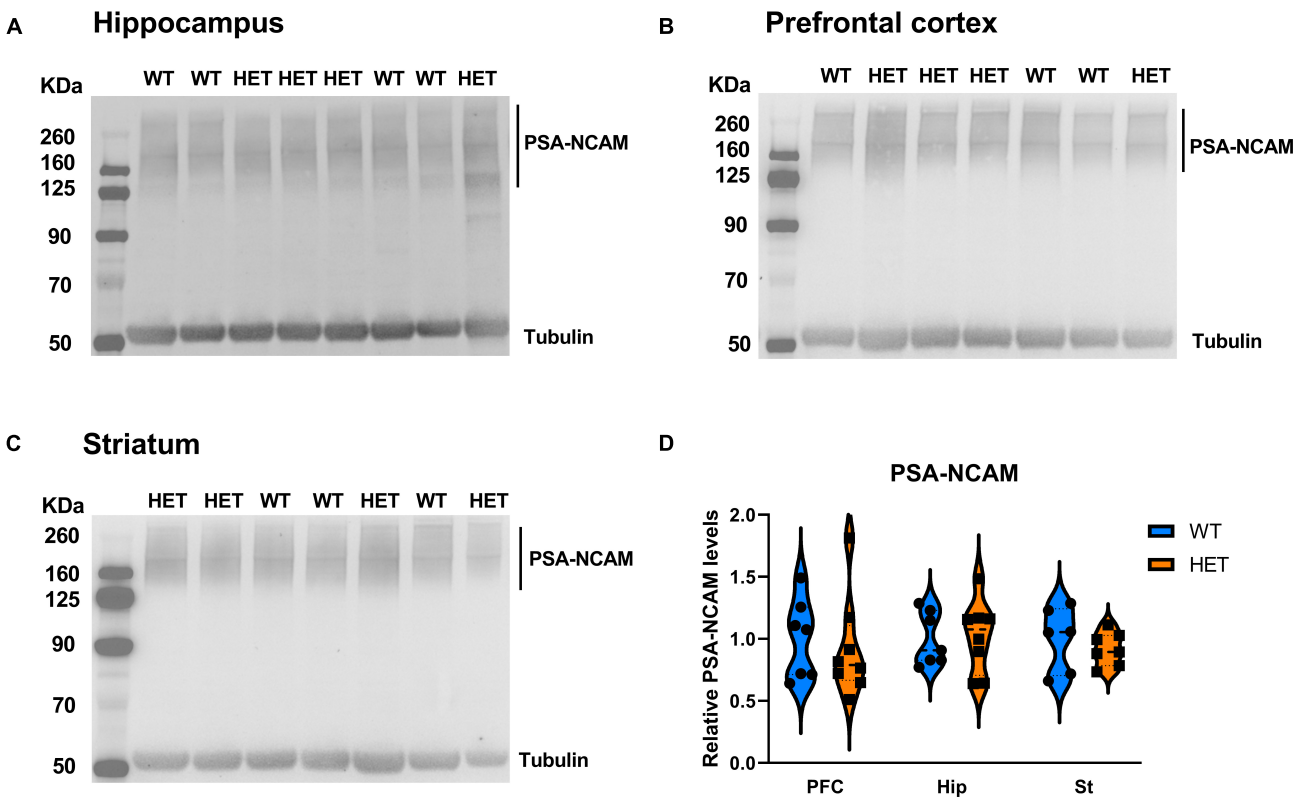


FIGURE 9 | Analysis of polysialylated proteins in different brain regions of the mouse brain. **(A–C)** Representative images of immunoblots of PSA in hippocampus **(A)**, PFC **(B)**, and Striatum **(C)**. **(D)** Quantification of PSA levels, normalized to α-tubulin (HET: $n = 8$, WT: $n = 7$). Data were analyzed by two-way ANOVA and presented as mean ± SEM.

upregulated. Interestingly, several association studies have identified *ST8SIA2* (also known as *SIAT8B*) as a risk gene for several neurodevelopmental and psychiatric disorders, including autism spectrum disorder (ASD) (Anney et al., 2010), bipolar disorder (McAuley et al., 2012) and schizophrenia (Arai et al., 2006; Tao et al., 2007; McAuley et al., 2012). These findings are further supported by the identification of loss-of-function mutations implicating *ST8SIA2* in patients with schizophrenia (Isomura et al., 2011) and ASD with epilepsy (Kamien et al., 2014). A *ST8SIA2* haplotype has also been shown to have differential effects in patients with psychotic disorders and healthy subjects on white matter integrity of the parietal lobe (Fullerton et al., 2018). In addition, *St8sia2* null mutant mice display behavioral symptoms that resemble those from patients with psychoses, such as cognitive dysfunction, deficits in prepulse inhibition, and increased sensitivity to amphetamine-induced locomotion (Kröcher et al., 2015). The upregulation of *St8sia4* and *St8sia2* observed in the brain of *St3gal3* HET mice may be interpreted as a compensatory mechanism to the reduction in *St3gal3* expression, in order to maintain the levels of PSA glycoproteins such as PSA-NCAM1 or PSA-SynCAM1 (PSA-CADM1). Indeed,

in our study *Cadm1* expression was also upregulated in the hippocampal region of *St3gal3* HET mice, possibly as a result of the same compensatory mechanism. CADM1 is essential for many neurodevelopmental processes including myelination (Niederkofler et al., 2010). Additionally, a study with a mouse model harboring a dominant negative mutation in *Cadm1* underscores a relationship between CADM1 function and ADHD-like behaviors (Sandau et al., 2012). Of note, CADM1 shows broadly overlapping expression with NCAM1 during brain development, but its protein product SynCAM1 is specifically polysialylated in the NG2 cells (Schnaar et al., 2014), which are considered progenitor cells types for oligodendrocytes, astrocytes, and neurons (Dimou and Gallo, 2015). Probably in cojunction with the upregulation of *Cadm1* and the polysialyltransferases *St8sia4* and *St8sia2*, *St3gal3* HET mice present with a downregulation of *Cspg4*, which encodes the proteoglican NG2, the only common marker of all NG2-specific cells (Dimou and Gallo, 2015).

Nevertheless, despite the robust alterations in the mRNA expression of several genes related to polysialylation, we did not detect significant changes at the protein level, neither for PSA-glycoproteins nor for total levels of SynCAM1 or NCAM1. This is in contrast with the increase in PSA glycoproteins observed in *St3gal3* full knockout mice (Yoo et al., 2015). There are several explanations to our negative finding. On the one hand, the upregulation of *St8sia4* and *St8sia2* could counteract the effects of *St3gal3* partial deficiency. On the other hand, the fact that *St3gal3* deficiency has a greater impact on *Cadm1* expression rather than on *Ncam1* (which encodes the main PSA acceptor) could explain why total PSA levels are not modified in those mice.

The stronger effects on the sialylation-related genes in the hippocampus compared to PFC might be linked to the higher rates of adult neurogenesis described in the hippocampus in comparison to other brain regions. In agreement, PSA is restricted in adulthood to regions with ongoing neurogenesis and synaptic plasticity (Galuska et al., 2010; Schnaar et al., 2014).

Strikingly, *St3gal3*-deficient mice presented with various behavioral alterations resembling deficits observed in ADHD patients such as locomotor activity and cognitive performance. In this regard, male HET mice showed impaired memory in the place learning task and, it is well known that learning disabilities are frequently observed in ADHD patients (Andersen et al., 2013; Tistarelli et al., 2020) and also supported by a GWAS meta-analysis implicating *ST3GAL3* in general cognitive function (Davies et al., 2018). This deficit is also observed in null mutant mice (Yoo et al., 2015) and patients with loss-of-function mutations in *ST3GAL3* (Hu et al., 2011; Edvardson et al., 2013; Indellicato et al., 2020). On the other hand, HET females demonstrated increased locomotor activity. As hyperactivity is one of the core features of ADHD, together with inattention and increased impulsivity (Tistarelli et al., 2020), moderate deficiency in *St3gal3* expression in female HET mice is able to capture some of the most representative features of ADHD. Unexpectedly, HET mice appeared to display increased inhibitory control (increased percentage of correct responses in the RTT). In agreement with such an increased inhibitory control, we also observed that HET females made a reduced number of premature nosepokes per visit during the test phase of the RTT. Our behavioral findings suggest that the contribution of *ST3GAL3* to ADHD may be restricted to distinct behavioral domains, such as cognitive performance, learning abilities and locomotor activity, whereas other traits (attention, impulsivity) are affected to a lesser extent. However, we cannot rule out that the effects of *St3gal3* deficiency on impulsivity or attention might be different in younger animals or under stressful conditions that may act as a second-hit trigger for other dysfunctional behavioral phenotypes.

Another feature of our findings is the sexual dimorphism. The locomotor hyperactivity in *St3gal3* HET females is in contrast to the clinical observation of female patients with ADHD, which tend to show less hyperactivity and more inattention, when compared to males diagnosed with ADHD. In addition, males normally suffer more often from comorbid disorders implicating impaired cognitive control, such as oppositional defiant, conduct or drug use disorders (Mansouri et al., 2016). However, because male patients with ADHD tend to display more externalizing behaviors, ADHD among females remains underdiagnosed (Slobodin and Davidovitch, 2019). Together, these behavioral characteristics may be explained by sex-specific differences at the molecular level that impact brain (cyto)architecture and synaptic plasticity (Manoli and Tollkuhn, 2018). Accordingly, we identified several genes that were differentially expressed in males and females, such as *Cadm1*, *St8sia4* and *Ncam1* in the PFC, among others.

In conclusion, our findings indicate that haploinsufficiency of ST3GAL3 results in a sex-dependent alteration of cognition, behavior and markers of brain plasticity. The *St3gal3*-haplodeficient mouse proved to represent a remarkably informative model to better understand the contribution of *ST3GAL3* variation to ADHD risk. Our results thus support ST3GAL3 as a key hub in the pathway toward innovative strategies to prevent or treat cognitive and behavioral deficits related to ADHD pathophysiology.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by Government of Lower Franconia.

AUTHOR CONTRIBUTIONS

OR, MF, and K-PL developed the conception and design of the study. OR, JA-R, PA, and H-PK performed the behavioral experiments. JA-R, GO, H-PK, DD, and DH performed the

molecular experiments. JA-R, MF, OR, and H-PK performed the statistical analysis of behavioral data. JA-R and OR performed the statistical analysis of the molecular data. OR led the manuscript writing. OR and JA-R wrote the manuscript. All authors contributed to manuscript revision, critically read and approved the submitted version.

FUNDING

This study received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant No. 953327 (Serotonin & BEYOND) and No. 728018 (Eat2beNICE), ERA-Net NEURON/DECODE! Grant No. FKZ01EW1902, and the 5-100 Russian Academic Excellence Project. This publication was supported by the Open Access Publication Fund of the University of Würzburg.

ACKNOWLEDGMENTS

We are grateful to N. Steigerwald, J. Zöller, M. Vitale, and A. Schmitt-Böhrer for valuable assistance in mouse genotyping and sample preparation, and personnel at the ZEMM for excellent animal caretaking. We also thank Juan A. Navarro for critical reading of the manuscript, and Sven Krackow for providing advice in the performance and statistical analysis of the Intellicage experiments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.688488/full#supplementary-material

REFERENCES

Andersen, P. N., Egeland, J., and Øie, M. (2013). Learning and memory impairments in children and adolescents with attention-deficit/hyperactivity disorder. *J. Lear. Disabil.* 46, 453–460. doi: 10.1177/0022219412437040

Anney, R., Klei, L., Pinto, D., Regan, R., Conroy, J., Magalhaes, T. R., et al. (2010). A genome-wide scan for common alleles affecting risk for autism. *Hum. Mol. Genet.* 19, 4072–4082. doi: 10.1093/hmg/ddq307

Arai, M., Yamada, K., Toyota, T., Obata, N., Haga, S., Yoshida, Y., et al. (2006). Association between polymorphisms in the promoter region of the sialyltransferase 8B (SIAT8B) gene and schizophrenia. *Biol. Psychiatry* 59, 652–659. doi: 10.1016/j.biopsych.2005.08.016

Audry, M., Jeanneau, C., Imberty, A., Harduin-Lepers, A., Delannoy, P., and Breton, C. (2011). Current trends in the structure-activity relationships of sialyltransferases. *Glycobiology* 21, 716–726. doi: 10.1093/glycob/cwq189

Benice, T. S., Rizk, A., Kohama, S., Pfankuch, T., and Raber, J. (2006). Sex-differences in age-related cognitive decline in C57BL/6J mice associated with increased brain microtubule-associated protein 2 and synaptophysin immunoreactivity. *Neuroscience* 137, 413–423. doi: 10.1016/j.neuroscience.2005.08.029

Bush, G. (2010). Attention-deficit/hyperactivity disorder and attention networks. *Neuropsychopharmacology* 35, 278–300. doi: 10.1038/npp.2009.120

Chen, L., Hu, X., Ouyang, L., He, N., Liao, Y., Liu, Q., et al. (2016). A systematic review and meta-analysis of tract-based spatial statistics studies regarding attention-deficit/hyperactivity disorder. *Neurosci. Biobehav. Rev.* 68, 838–847. doi: 10.1016/j.neubiorev.2016.07.022

Davies, G., Lam, M., Harris, S. E., Trampush, J. W., Luciano, M., Hill, W. D., et al. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat. Commun.* 9:2098. doi: 10.1038/s41467-018-04362-x

Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., et al. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* 51, 63–75. doi: 10.1038/s41588-018-0269-7

Dimou, L., and Gallo, V. (2015). NG2-glia and their functions in the central nervous system. *Glia* 63, 1429–1451. doi: 10.1002/glia.22859

Ednie, A. R., and Bennett, E. S. (2012). Modulation of voltage-gated ion channels by sialylation. *Compr. Physiol.* 2, 1269–1301. doi: 10.1002/cphy.c110044

Edvardson, S., Baumann, A. M., Mühlenhoff, M., Stephan, O., Kuss, A. W., Shaag, A., et al. (2013). West syndrome caused by ST3Gal-III deficiency. *Epilepsia* 54, e24–7. doi: 10.1111/epi.12050

Ellies, L. G., Ditto, D., Levy, G. G., Wahrenbrock, M., Ginsburg, D., Varki, A., et al. (2002). Sialyltransferase ST3Gal-IV operates as a dominant modifier of hemostasis by concealing asialoglycoprotein receptor ligands. *Proc. Natl. Acad. Sci. U. S. A.* 99, 10042–10047. doi: 10.1073/pnas.142005099

Faraone, S. V., and Larsson, H. (2019). Genetics of attention deficit hyperactivity disorder. *Mol. Psychiatry* 24, 562–575. doi: 10.1038/s41380-018-0070-0

Fischer, M., Cabello, V., Popp, S., Krackow, S., Hommers, L., Deckert, J., et al. (2017). Rsk2 knockout affects emotional behavior in the IntelliCage. *Behav. Genet.* 47, 434–448. doi: 10.1007/s10519-017-9853-3

Forero, A., Ku, H. P., Malpartida, A. B., Wäldchen, S., Alhama-Riba, J., Kulka, C., et al. (2020). Serotonin (5-HT) neuron-specific inactivation of Cadherin-13 impacts 5-HT system formation and cognitive function. *Neuropharmacology* 168:108018. doi: 10.1016/j.neuropharm.2020.108018

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M., et al. (2012). AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optim. Methods Softw.* 27, 233–249. doi: 10.1080/10556788.2011.597854

Fullerton, J. M., Klauser, P., Lenroot, R. K., Shaw, A. D., Overs, B., Heath, A., et al. (2018). Differential effect of disease-associated ST8SIA2 haplotype on cerebral white matter diffusion properties in schizophrenia and healthy controls. *Transl. Psychiatry* 8:21. doi: 10.1038/s41398-017-0052-z

Galuska, S. P., Geyer, R., Gerardy-Schahn, R., Mühlenhoff, M., and Geyer, H. (2008). Enzyme-dependent variations in the polysialylation of the neural cell adhesion molecule (NCAM) in vivo. *J. Biol. Chem.* 283, 17–28. doi: 10.1074/jbc.M707024200

Galuska, S. P., Rollenhagen, M., Kaup, M., Eggers, K., Oltmann-Norden, I., Schiff, M., et al. (2010). Synaptic cell adhesion molecule SynCAM 1 is a target for polysialylation in postnatal mouse brain. *Proc. Natl. Acad. Sci. U. S. A.* 107, 10250–10255. doi: 10.1073/pnas.0912103107

Homberg, J. R., Adan, R. A. H., Alenina, N., Asiminas, A., Bader, M., Beckers, T., et al. (2021). The continued need for animals to advance brain research. *Neuron* 109, 2374–2379. doi: 10.1016/j.neuron.2021.07.015

Hu, H., Eggers, K., Chen, W., Garshasbi, M., Motazacker, M. M., Wrogemann, K., et al. (2011). ST3GAL3 mutations impair the development of higher cognitive functions. *Am. J. Hum. Genet.* 89, 407–414. doi: 10.1016/j.ajhg.2011.08.008

Indellicato, R., Domenighini, R., Malagolini, N., Cereda, A., Mamoli, D., Pezzani, L., et al. (2020). A novel nonsense and inactivating variant of ST3GAL3 in two infant siblings suffering severe epilepsy and expressing circulating CA19.9. *Glycobiology* 30, 95–104. doi: 10.1093/glycob/cwz079

Isomura, R., Kitajima, K., and Sato, C. (2011). Structural and functional impairments of polysialic acid by a mutated polysialyltransferase found in schizophrenia. *J. Biol. Chem.* 286, 21535–21545. doi: 10.1074/jbc.M111.221143

Kamien, B., Harraway, J., Lundie, B., Smallhorne, L., Gibbs, V., Heath, A., et al. (2014). Characterization of a 520 kb deletion on chromosome 15q26.1 including ST8SIA2 in a patient with behavioral disturbance, autism spectrum disorder, and epilepsy. *Am. J. Med. Genet. Part A* 164A, 782–788. doi: 10.1002/ajmg.a.36345

Klein, M., Singgih, E. L., van Rens, A., Demontis, D., Børglum, A. D., Mota, N. R., et al. (2020). Contribution of Intellectual Disability-Related Genes to ADHD

Risk and to Locomotor Activity in *Drosophila*. *Am. J. Psychiatry* 177, 526–536. doi: 10.1176/appi.ajp.2019.18050599

Kobayashi, Y., Sano, Y., Vannoni, E., Goto, H., Suzuki, H., Oba, A., et al. (2013). Genetic dissection of medial habenula-interpeduncular nucleus pathway function in mice. *Front. Behav. Neurosci.* 7:17. doi: 10.3389/fnbeh.2013.00017

Kolter, T., Proia, R. L., and Sandhoff, K. (2002). Combinatorial ganglioside biosynthesis. *J. Biol. Chem.* 277, 25859–25862. doi: 10.1074/jbc.R200001200

Koolhaas, J. M., Coppens, C. M., de Boer, S. F., Buwalda, B., Meerlo, P., and Timmermans, P. J. (2013). The resident-intruder paradigm: a standardized test for aggression, violence and social stress. *J. Visual. Exp. JoVE* e4367. doi: 10.3791/4367

Krackow, S., Vannoni, E., Codita, A., Mohammed, A. H., Cirulli, F., Branchi, I., et al. (2010). Consistent behavioral phenotype differences between inbred mouse strains in the IntelliCage. *Genes Brain Behav.* 9, 722–731. doi: 10.1111/j.1601-183X.2010.00606.x

Kröcher, T., Malinovskaja, K., Jürgenson, M., Aonurm-Helm, A., Zharkovskaya, T., Kalda, A., et al. (2015). Schizophrenia-like phenotype of polysialyltransferase ST8SIA2-deficient mice. *Brain Struct. Funct.* 220, 71–83. doi: 10.1007/s00429-013-0638-z

Lesch, K. P. (2019). Editorial: can dysregulated myelination be linked to ADHD pathogenesis and persistence? *J. Child Psychol. Psychiatry and Allied Discipl.* 60, 229–231. doi: 10.1111/jcpp.13031

Manoli, D. S., and Tollkuhn, J. (2018). Gene regulatory mechanisms underlying sex differences in brain development and psychiatric disease. *Ann. N. Y. Acad. Sci.* 1420, 26–45. doi: 10.1111/nyas.13564

Mansouri, F. A., Fehring, D. J., Gaillard, A., Jaberzadeh, S., and Parkington, H. (2016). Sex dependency of inhibitory control functions. *Biol. Sex Diff.* 7:11. doi: 10.1186/s13293-016-0065-y

McAuley, E. Z., Scimone, A., Tiwari, Y., Agahi, G., Mowry, B. J., Holliday, E. G., et al. (2012). Identification of sialyltransferase 8B as a generalized susceptibility gene for psychotic and mood disorders on chromosome 15q25-26. *PLoS One* 7:e38172. doi: 10.1371/journal.pone.0038172

Mühlenhoff, M., Rollenhagen, M., Werneburg, S., Gerardy-Schahn, R., and Hildebrandt, H. (2013). Polysialic acid: versatile modification of NCAM, SynCAM 1 and neuropilin-2. *Neurochemical Res.* 38, 1134–1143. doi: 10.1007/s11064-013-0979-2

Nacher, J., Guirado, R., and Castillo-Gómez, E. (2013). Structural plasticity of interneurons in the adult brain: role of PSA-NCAM and implications for psychiatric disorders. *Neurochem. Res.* 38, 1122–1133. doi: 10.1007/s11064-013-0977-4

Niederkofler, V., Baeriswyl, T., Ott, R., and Stoeckli, E. T. (2010). Nectin-like molecules/SynCAMs are required for post-crossing commissural axon guidance. *Development* 137, 427–435. doi: 10.1242/dev.042515

Ruijter, J. M., Ramakers, C., Hoogaars, W. M., Karlen, Y., Bakker, O., van den Hoff, M. J., et al. (2009). Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res.* 37:e45. doi: 10.1093/nar/gkp045

Sandau, U. S., Alderman, Z., Corfas, G., Ojeda, S. R., and Raber, J. (2012). Astrocyte-specific disruption of SynCAM1 signaling results in ADHD-like behavioral manifestations. *PLoS One* 7:e36424. doi: 10.1371/journal.pone.0036424

Schnaar, R. L., Gerardy-Schahn, R., and Hildebrandt, H. (2014). Sialic acids in the brain: gangliosides and polysialic acid in nervous system development, stability, disease, and regeneration. *Physiol. Rev.* 94, 461–518. doi: 10.1152/physrev.00033.2013

Slobodin, O., and Davidovitch, M. (2019). Gender Differences in Objective and Subjective Measures of ADHD Among Clinic-Referred Children. *Front. Hum. Neurosci.* 13:441. doi: 10.3389/fnhum.2019.00441

Tao, R., Li, C., Zheng, Y., Qin, W., Zhang, J., Li, X., et al. (2007). Positive association between SIAT8B and schizophrenia in the Chinese Han population. *Schizophr. Res.* 90, 108–114. doi: 10.1016/j.schres.2006.09.029

Tistarelli, N., Fagnani, C., Troianiello, M., Stazi, M. A., and Adriani, W. (2020). The nature and nurture of ADHD and its comorbidities: a narrative review on twin studies. *Neurosci. Biobehav. Rev.* 109, 63–77. doi: 10.1016/j.neubiorev.2019.12.017

van Ewijk, H., Heslenfeld, D. J., Zwiers, M. P., Buitelaar, J. K., and Oosterlaan, J. (2012). Diffusion tensor imaging in attention deficit/hyperactivity disorder: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* 36, 1093–1106. doi: 10.1016/j.neubiorev.2012.01.003

Walton, E., Pingault, J. B., Cecil, C. A., Gaunt, T. R., Relton, C. L., Mill, J., et al. (2017). Epigenetic profiling of ADHD symptoms trajectories: a prospective, methylome-wide study. *Mol. Psychiatry* 22, 250–256. doi: 10.1038/mp.2016.85

Wang, B. (2009). Sialic acid is an essential nutrient for brain development and cognition. *Ann. Rev. Nutr.* 29, 177–222. doi: 10.1146/annurev.nutr.28.061807.155515

Yanai, S., and Endo, S. (2021). Functional Aging in Male C57BL/6J Mice Across the Life-Span: a Systematic Behavioral Analysis of Motor, Emotional, and Memory Function to Define an Aging Phenotype. *Front. Aging Neurosci.* 13:697621. doi: 10.3389/fnagi.2021.697621

Yoo, S. W., Motari, M. G., Susuki, K., Prendergast, J., Mountney, A., Hurtado, A., et al. (2015). Sialylation regulates brain structure and function. *FASEB J.* 29, 3040–3053. doi: 10.1096/fj.15-270983

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Citation: Rivero O, Alhama-Riba J, Ku H-P, Fischer M, Ortega G, Álmos P, Diouf D, van den Hove D and Lesch K-P (2021) Haploinsufficiency of the Attention-Deficit/Hyperactivity Disorder Risk Gene St3gal3 in Mice Causes Alterations in Cognition and Expression of Genes Involved in Myelination and Sialylation. Front. Genet. 12:688488. doi: 10.3389/fgene.2021.688488

frontiers in Psychiatry

SYSTEMATIC REVIEW
published: 12 January 2022
doi: 10.3389/fpsyt.2021.685041

The Brain-Derived Neurotrophic Factor Val66Met Polymorphism Is Associated With Female Obsessive-Compulsive Disorder: An Updated Meta-Analysis of 2765 Obsessive-Compulsive Disorder Cases and 5558 Controls

Yanan Shang[1†], *Na Wang*[2†], *En Zhang*[3†], *Qian Liu*[1], *Hengfen Li*[1] *and Xiaofeng Zhao*[1*]

[1] *Department of Psychiatry, First Affiliated Hospital of Zhengzhou University, Zhengzhou, China,* [2] *Department of Neurological Rehabilitation, Second Affiliated Hospital of Zhengzhou University, Zhengzhou, China,* [3] *Department of Psychiatry, The Fourth Hospital of Wuhu City, Wuhu, China*

Objective: Accumulated evidence has implicated that brain-derived neurotrophic factor (BDNF) gene polymorphisms play a role in the etiology of obsessive-compulsive disorder (OCD). A single nucleotide polymorphism in the coding exon of the BDNF gene at position 66, Val66Met (rs6265), is found to be associated with OCD in different populations, but results linking Val66Met with OCD have been inconsistent and inconclusive. In our study we performed a meta-analysis to further examine whether rs6265 genetic variants are involved in the etiology of OCD.

Methods: By searching databases, relevant case-control studies were retrieved; using established inclusion criteria, we selected eligible studies for analysis.

Results: Thirteen studies were identified that examined the association between the rs6265 polymorphism and OCD. After statistical analyses, no significant association was found between the rs6265 polymorphism and OCD (OR = 1.07, 95% CI = 1.00–1.15, $P = 0.06$ for genotype; OR = 1.06, 95% CI = 0.98–1.15, $P = 0.15$ for allele). However, in gender-specific analysis, female Val carriers might be a risk factor for OCD (OR = 1.36, 95% CI = 1.03–1.80, $P = 0.03$ for genotype; OR = 1.15, 95% CI = 1.01–1.32, $P = 0.04$ for allele).

Conclusion: Our updated meta-analysis suggests that female carriers of the Val66Met BDNF polymorphism might be more suspectable to develop OCD.

Keywords: Val66Met polymorphism, brain-derived neurotrophic factor (BDNF) gene, obsessive-compulsive disorder (OCD), meta-analysis, female

OPEN ACCESS

Edited by:
Elena Martín-García, Pompeu Fabra University, Spain

Reviewed by:
Reiji Yoshimura, University of Occupational and Environmental Health Japan, Japan
Catia Scassellati, San Giovanni di Dio Fatebenefratelli Center (IRCCS), Italy

***Correspondence:**
Xiaofeng Zhao
zxf0605004@163.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Mood Disorders, a section of the journal Frontiers in Psychiatry

Received: *24 March 2021*
Accepted: *17 December 2021*
Published: *12 January 2022*

Citation:
Shang Y, Wang N, Zhang E, Liu Q, Li H and Zhao X (2022) The Brain-Derived Neurotrophic Factor Val66Met Polymorphism Is Associated With Female Obsessive-Compulsive Disorder: An Updated Meta-Analysis of 2765 Obsessive-Compulsive Disorder Cases and 5558 Controls. Front. Psychiatry 12:685041. doi: 10.3389/fpsyt.2021.685041

INTRODUCTION

It has been demonstrated by twin studies that genetic factors play a key role in the etiology of obsessive-compulsive disorder (OCD) (1). While these genetic factors account for 27–65% of the variance in this disorder (1), its etiology has not been clearly elucidated in the past decades.

The brain-derived neurotrophic factor (BDNF) gene is considered to be one of the candidate genes of psychiatric disorders, as it is widely expressed in the brain and plays a major role in neurogenesis, neuronal growth, maturation, survival, synaptic plasticity, and microarchitectural integrity (2, 3). BDNF is also considered to play a role in regulating mood and behavior, and abnormalities in the expression of this gene have been involved in major anxiety disorder, depressive disorder, and OCD. Preclinical studies have shown that levels of BDNF are drastically decreased in the serum and brain tissue of rats with obsessive-compulsive disorder compared to a control group (4). In a clinical study, Maina was the first to demonstrate that levels of serum BDNF were significantly decreased in OCD patients compared to normal controls (5). Following that study, Wang et al. (6) suggest that BDNF is involved in the etiology of OCD, and may be a peripheral marker indicating neurotrophic impairment in OCD similarly, Suliman et al. demonstrated that peripheral BDNF levels in OCD patients were significantly lower compared with that of normal controls with this data, it can be speculated that the BDNF gene is involved in the etiology of OCD.

Researchers have extensively focused on the BDNF polymorphism, Val66Met, also known as G196A or rs6265 (7). This polymorphism influences the activity of BDNF and impairs BDNF signal transduction, and therefore may be a possible candidate gene involved in the etiology of OCD (7). While previous evidence suggests that the rs6265 polymorphism is involved in the etiology of OCD, the outcomes of such studies have been inconsistent and even contradictory. Findings from the latest meta-analytic investigation by Wang et al. in 2019 provide little support for the Val66Met variant as a predictor of OCD (8). In order to derive a more putative conclusion, an updated meta-analysis was performed to study the connection between the rs6265 polymorphism of the BDNF gene and the risk of OCD.

METHODS

Inclusion Criteria

(1) All studies investigated the association between OCD and Val66Met, G196A, or rs6265 polymorphisms. (2) OCD was diagnosed according to the Diagnostic and Statistical Manual of Mental Disorders (DSM) or the International Classification of Diseases (ICD) and included DSM-III, DSM-III-R, DSM-IV, DSM-IV-R, and ICD-10. (3) The study had a patient group with OCD and a control group. (4) The study included samples from varied ethnic backgrounds. (5) The frequencies of each allele and genotype met the criteria for Hardy-Weinberg disequilibrium in both case and control groups.

Search Strategy

This meta-analysis was conducted according to PRISMA guidelines. The English databases used for retrieval were PubMed, Cochrane, Web of Science, Embase, PsycINFO,

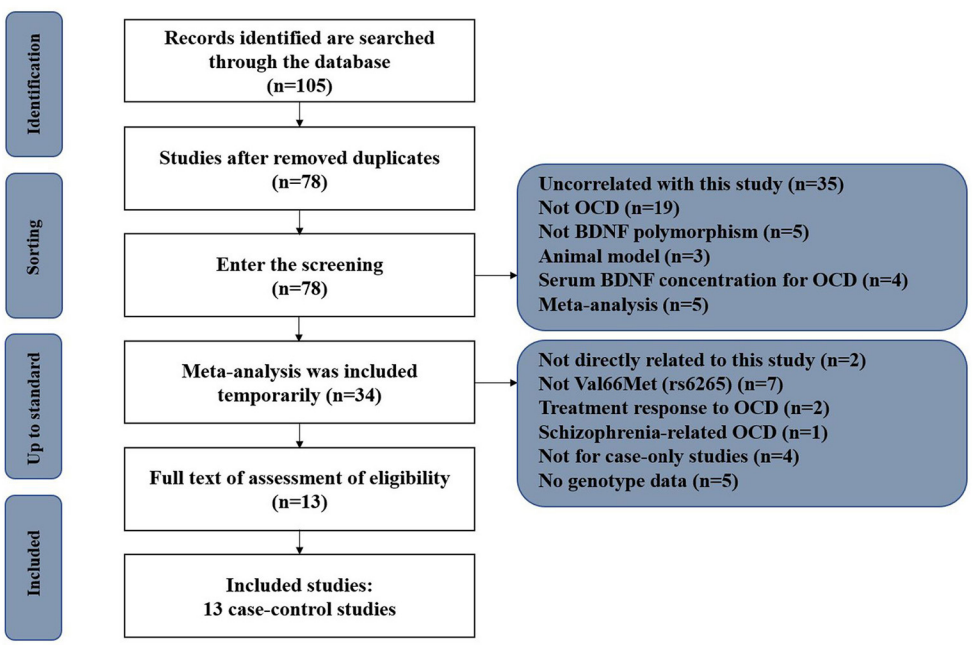


FIGURE 1 | The process of independent studies for inclusion in this meta-analysis.

TABLE 1 | Summary of studies examining the relationship between the BDNF val66met polymorphism and OCD.

References	Race	Results	Genotype						Allele			
			Case (n)			Control (n)			Case (n)		Control (n)	
			Val/Val	Val/Met	Met/Met	Val/Val	Val/Met	Met/Met	Val	Met	Val	Met
Wendland et al. (10)	Caucasian	No association	192	92	11	428	206	23	476	114	1,062	252
Hemmings et al. (11)	Caucasian	val66met variant associated with male EO OCD	73	33	6	95	43	2	179	45	233	47
Katerberg et al. I (12)	Caucasian	Val associated with sexual/religious obsessions	128	62	9	76	36	3	318	80	188	42
Katerberg et al. II (12)	Caucasian	Val associated with sexual/religious obsessions	132	75	13	352	166	17	339	101	870	200
Wang et al. (13)	Chinese Han	No association	31	76	41	24	51	19	138	158	99	89
Da Rocha et al. (14)	Caucasian	Met allele associated with OCD	83	33	11	98	21	8	199	55	217	37
Tükel et al. (15)	Caucasian	No association	23	54	23	29	51	30	100	100	109	111
Hemmings et al. (18)	Caucasian	No association	85	33	3	131	50	6	203	39	312	62
Liu et al. (17)	Chinese Han	No association	40	107	43	60	167	82	187	193	287	331
Márquez et al. (16)	NS	Val/Val associated with OCD	179	47	6	162	99	22	405	59	426	143
Liu et al. (19)	Chinese Han	val66met variant associated with OCD	101	166	54	110	221	105	368	274	431	421
Wang et al. (20)	Chinese Han	No association	31	76	41	24	54	21	138	158	102	96
Umehara et al. (21)	Japanese	No association	59	83	33	686	1,002	339	201	149	2,374	1,680
Taj et al. (22)	Indian	Met allele play a protective role in OCD	277	91	9	299	133	17	645	109	731	167

BDNF, brain-derived neurotrophic factor; OCD, obsessive-compulsive disorder.

PsycARTICLE, and the Chinese database was CNKI, CBM and WANFANG. The deadline is September 2020. In order to search OCD and BDNF rs6265 articles in the database more widely, no further restrictions were added. All published papers examined the association between OCD and the BDNF Val66Met polymorphism. The keywords "Brain-derived neurotrophic factor (BDNF)," "Brain Derived Neurotrophic Factor," "Factor, Brain-Derived Neurotrophic," "Neurotrophic Factor, Brain-Derived," "BDNF," "obsessive-compulsive disorder (OCD)," "Obsessive-Compulsive Neuroses," "Val66Met," "rs6265," "G196A," and "196G/A" were used to search the articles. We also examined all the references cited in the identified articles, as well as those cited in relevant review articles. The quality of the included studies was evaluated, and publication bias was evaluated.

Data Extraction

From each enrolled article, data were extracted; including the name of the first author, year of publication, ethnicity of samples, diagnostic criteria of OCD, exclusion criteria, sample size (number of cases and controls), sample age, sex ratio, experimental design, severity of symptoms, and the available genotype information of the BDNF Val66Met polymorphism. To prevent potential errors, all data were extracted independently by the three authors, and a discussion within the research group resolved any disagreement.

Data Analysis and Statistical Methods

The data analysis and statistical methods were based on previous research methods (9).

Description of Studies Identified in Meta-Analysis

Through our search strategy, we identified 105 linked research papers with potential. After reading the abstracts and full compilations of the studies, 65 of them did not meet the inclusion criteria and were not able to be included (**Figure 1**). At the end, 13 available case-control studies were selected; including Wendland et al. (10), Hemmings et al. (11), Katerberg et al. (12), Wang et al. (13), Da Rocha et al. (14), Tükel et al. (15), Marquez et al. (16), Liu et al. (17), Hemmings et al. (18), Liu et al. (19), Wang et al. (20), Umehara et al. (21), and Taj et al. (22). One of the papers included separate data on two different groups of patients (12). After discussion within the research group, it was decided that two dependent sample data was still available, and therefore there were 13 articles in total included. Finally, all together there are 2765 OCD patients and 5558 psychiatrically healthy controls included. **Table 1** lists the detailed characteristics of each study.

RESULTS

Studies selected by the database searches examined the relationship between patients with OCD and Val66Met polymorphism distributions in the following populations: Chinese Han ($n = 4$), Caucasian ($n = 7$), Japanese ($n = 1$), Indian ($n = 1$) (**Table 1**). Polymerase chain reaction-based methods were used to identify the genotype and allele frequencies of both OCD patients and control subjects in all studies. The modified Newcastle-Ottawa scale (NOS) was used for quality assessment of the included studies, and the evaluation results are

TABLE 2 | The clinical characteristics of the included studies.

Characteristics	Sample size	Sex (male/female)	Age	Design	Exclusion criteria	Diagnosis	Measure to assess symptom severity
Wendland et al. (10)	925	Not mentioned	≥18	Case-control study	Active schizophrenia or psychosis, severe mental retardation that does not permit an evaluation to characterize OCD, or OCD symptoms that occur exclusively in the context of depression were excluded	DSM-IV criteria	Y-BOCS
Hemmings et al. (11)	252	90/162	9–65	Case-control study	Not meet the DSM-IV criteria (APA 1994) for a primary diagnosis of OCD on the Structured Clinical Interview for Axis I disorders Patient Version were excluded.	DSM-IV criteria	Y-BOCS-SS
Katerberg et al. (12)	755	345/410	Not mentioned	Case-control study	Not meet the DSM-IV criteria were excluded.	DSM-IV criteria	Y-BOCS-SS
Katerberg et al. (12)	314	133/181	Not mentioned	Case-control study	Not meet the DSM-IV criteria were excluded.	DSM-IV criteria	Y-BOCS-SS
Wang et al. (13)	242	139/103	18–64	Case-control study	Severe physical illness, psychoactive substance abusers, serious suicide attempts, pregnant or lactating women were excluded.	DSM-IV criteria	Y-BOCS
Da Rocha et al. (14)	254	Not mentioned	Not mentioned	Case-control study	Not mentioned	DSM-IV criteria	Not mentioned
Tükel et al. (15)	210	80/130	17–50	Case-control study	Current psychiatric disorder other than OCD diagnosed with the SCID-I/CV, history of alcohol and/or drug abuse/dependence, neurological disease, pregnancy or lactation, medical disorders that may have a causal relationship with OCD	DSM-IV criteria	Y-BOCS
Márquez et al. (16)	515	260/255	Not mentioned	Case-control study	Not mentioned	DSM-IV-TR criteria	Y-BOCS
Liu et al. (17)	499	223/276	13–75	Case-control study	Severe physical disorder, abuse of psychoactive substances, pregnancy or lactation	DSM-IV criteria	Y-BOCS
Hemmings et al. (18)	322	134/188	12–72	Case-control study	History of neurological disease, schizophrenia, schizo-affective disorder, other psychotic conditions or a history of substance dependence, as determined from the interviews or previous medical records	DSM-IV criteria	Y-BOCS
Liu et al. (19)	747	414/333	≥18	Case-control study	Serious psychiatric diseases other than OCD or related family history, history of alcohol and/or drug abuse/dependence; any serious concomitant general medical condition or neurological disease except depression, phobia, anxiety or TS; history of medical disorders that may have a causal relationship with OCD; and pregnancy or lactation	DSM-IV criteria	Y-BOCS
Wang et al. (20)	247	143/104	16–64	Case-control study	Met any other DSM-IV axis I diagnosis; had any prior or current suicide attempts; were pregnant or lactating; or were in physical health such that they could not complete the study were excluded.	DSM-IV criteria	Y-BOCS
Umehara et al. (21)	2206	937/1265	Not mentioned	Case-control study	Patients comorbid with other axis I disorders were excluded.	DSM-IV criteria	Y-BOCS
Taj et al. (22)	826	Not mentioned	Not mentioned	Case-control study	Patients with co-morbid psychosis, bipolar disorder and mental retardation were excluded	DSM-IV criteria	Y-BOCS

Axis I disorders (including dementia, substance abuse, dysthymia, panic disorder, obsessive-compulsive disorder, and generalized anxiety disorder); DSM-IV; the 4th edition of the diagnostic and statistical manual of mental disorder; DSM-IV-TR, DSM-IV-text version; Y-BOCS, Yale-Brown Obsessive Compulsive Scale; APA, American Psychological Association; Y-BOCS-SS, the Y-BOCS severity scale; SCID-I/CV, DSM-IV-TR Axis I disorders were examined using clinical protocols.

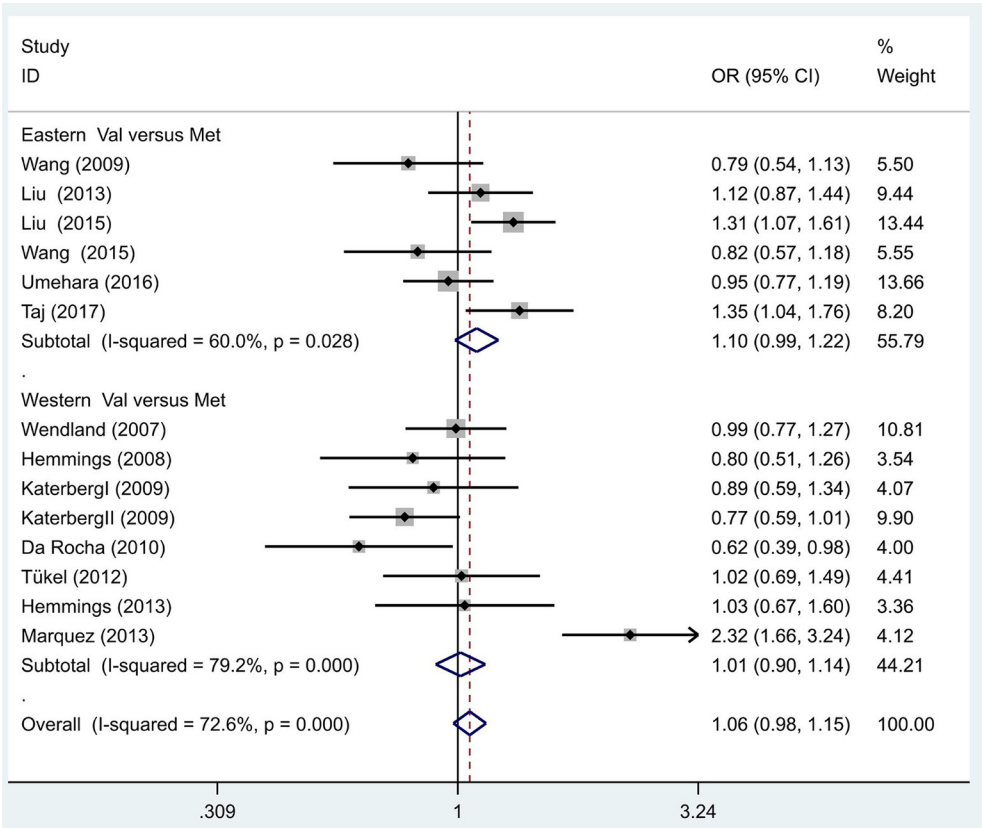


FIGURE 2 | Results of the random-effects meta-analysis for the BDNF Val66Met allele (Val vs. Met) in OCD and control groups.

shown in **Supplementary Table 1**. The exclusion of articles with low scores (14) did not affect the results after statistical analysis, and detailed data can be seen in the **Supplementary Figure 1** and **Supplementary Table 2**.

The studies included in **Table 2** permitted our team to make planned comparisons of allele and genotype frequencies between OCD cases and controls. Prior studies have shown that the Val allele is linked with higher activity of the BDNF system when compared with the Met allele; therefore, we grouped the data for this meta-analysis according to the law of dominant or recessive modeling. We hypothesized that the Val genotype is required to confer susceptibility to OCD. The Stata software (Version15.0) was used to analyze the data.

We also hypothesized that the Met allele is required to see a susceptibility to OCD. The between-study heterogeneity results hinted that the effect each individual study size had was statistically significant (I^2 = 72.6%, P = 0.00; **Figure 2**). The pooled OR from these studies was 1.06 (95% CI = 0.98–1.15, z = 1.43, P = 0.15), denoting that there was no link between the Met allele and OCD.

Additionally, we hypothesized that in different populaces, the Met allele is linked with OCD, and proceeded to execute separate analyses in Eastern and Western populations. The between-study heterogeneity results hinted that in the Eastern population, the effect of study size was statistically significant (I^2 = 60.0%, P = 0.03; **Figure 2**). The pooled OR from these studies was 1.10 (95% CI = 0.99–1.22, z =1.71, P = 0.09), hinting there was no link between the Met allele and OCD in the Eastern population. Similarly, no association was identified in the Western populations (OR = 1.01, 95% CI = 0.90–1.14, z = 0.22, P = 0.83; **Figure 2**).

Following our primary analysis, we hypothesized that the Met/Met genotype is a risk factor for OCD. It was suggested by the between-study heterogeneity results that the effect of sample size of each study was insignificant (I^2 = 61.1%, P = 0.00). The pooled OR from these studies was 1.07 (95% CI = 1.00–1.14, z = 1.86, P = 0.06; **Figure 3**), indicating no association between the Met/Met genotype and OCD. Similarly, we also compared the Met/Met genotype vs. Val/Met +Val/Val alleles in the Eastern and Western populations, in order to explore their relationship with OCD further. We found that the results did not change in either the Eastern or Western population (OR = 1.10, 95% CI = 0.99–1.22, z = 1.71, P = 0.09; OR = 1.01, 95% CI = 0.90–1.14, z = 0.22, P = 0.83, respectively; **Figure 3**). Previous studies have demonstrated that sex-specific genetic architecture is a risk factor for OCD; thus, we compared Val *vs.* Met carriers in male and female groups. In gender-specific analysis, the female Val carriers pose as a risk factor for OCD (OR = 1.15, 95% CI = 1.01–1.32,

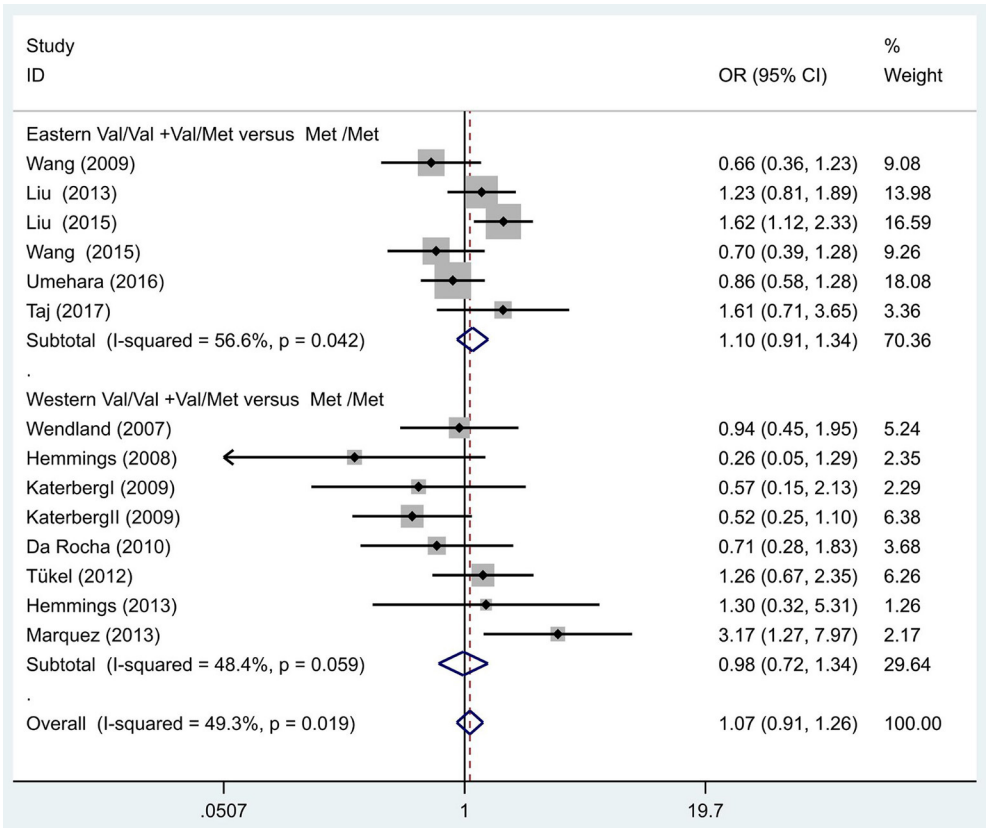


FIGURE 3 | Results of the random-effects meta-analysis for the BDNF Val66Met genotype (Val/Val+Val/Met vs. Met/Met) in OCD and control groups.

$P = 0.04$ for allele; OR = 1.36, 95% CI = 1.03–1.80, $P = 0.03$ for genotype; **Figure 4**). No associations were found between male Val carriers and OCD (OR = 1.09, 95% CI = 0.94–1.26, $P = 0.25$ for allele; OR = 1.13, 95% CI = 0.84–1.56, $P = 0.46$ for genotype; **Figure 5**).

Sensitivity Analyses

The 13 studies we screened evaluated Val66Met polymorphisms using small sample sizes, indicating possible differences in genotype frequency and/or population stratification. We then performed a sensitivity analysis that excluded these studies one by one. However, the outcomes of the meta-analysis were not changed by sensitivity analysis (**Figure 6**).

Publication Bias

The Begg-Mazumdar test for the BDNF Val66Met polymorphism demonstrated a low possibility of publication bias ($z = -0.91$, $P = 0.363$), and the Egger's test showed no significant results ($t = -1.02$, $P = 0.86$). Thus, we deduced there was no publication bias.

DISCUSSION

The currently updated meta-analysis is the most recent quantitative collection of available results from gene-psychiatric disorder association studies. The main findings of our studies indicate that the Val66Met polymorphism is not associated with OCD. Two prior meta-analyses have indicated that the Val66Met polymorphism is not related to OCD, which was in agreement with our findings (8, 20).

We first compared the gene gender-specific interaction effects in the etiology of OCD using a meta-analysis method. The main finding of this study is that in gender-specific analysis, the female Val carriers might be a risk factor for OCD (OR = 1.36, 95% CI = 1.03–1.80, $P = 0.03$ for genotype; OR = 1.21, 95% CI = 1.04–1.40, $P = 0.01$ for allele).

Sex differences in OCD etiology indicate a protective role of estrogens. It has been reported that male OCD patients display symptoms at a younger age and with more severe outcomes when compared to female patients (23), suggesting that sex differences may play a role in the progression of OCD. Furthermore, during the menstrual cycle, there are significant alternations in peripheral levels of BDNF, and BDNF levels have been strongly correlated with estradiol levels in females (24). In addition, Sohrabji et al. (25), demonstrated that through the estrogen response element's influence on the BDNF gene, the expression of BDNF could be regulated by estrogen. Even more, estrogen and some neurotrophins share certain signaling pathways (26), which explains how the BDNF gene is involved in the etiology of OCD. It also

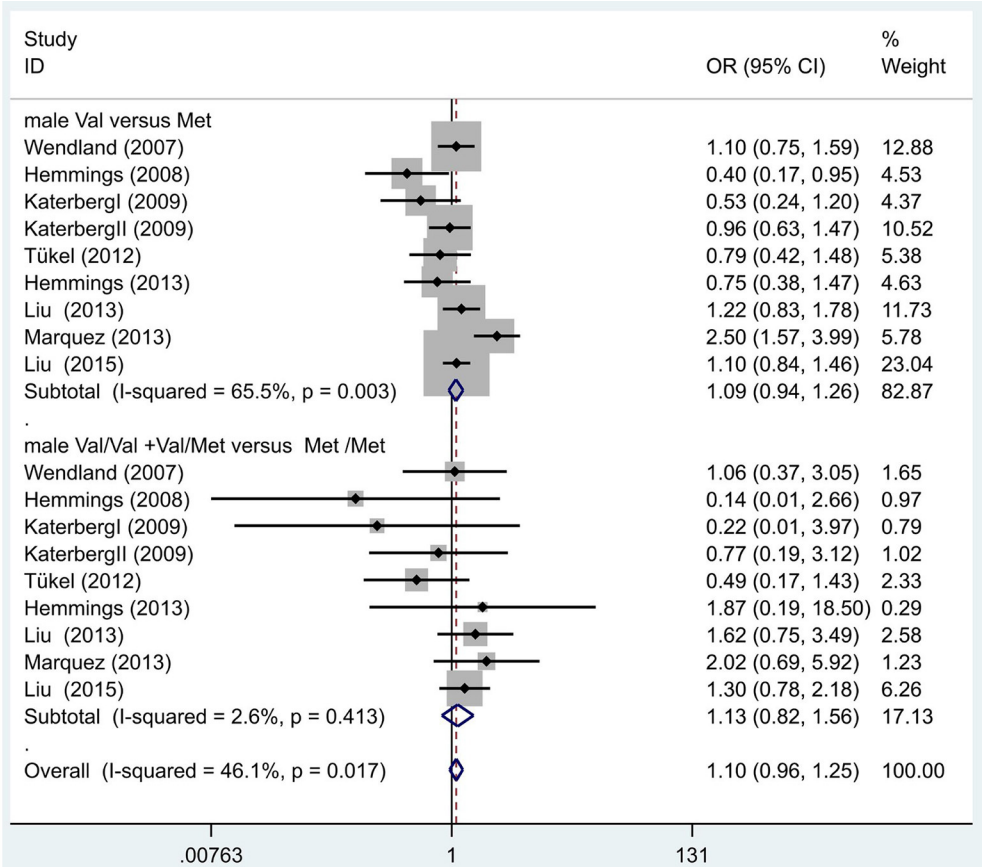


FIGURE 4 | Results of the random-effects meta-analysis for the BDNF Val66Met Val carriers in male OCD and control groups, respectively.

supports the notion that estrogens play a protective role in OCD development.

The results from Atmaca and Fouch (27, 28) demonstrated that subjects with OCD had smaller hippocampus volumes than controls. According to Egan et al. (29) and Hariri et al. (30), the Met allele is associated with abnormal hippocampal activation. Similarly, Egan has demonstrated that the BDNF Met allele significantly reduces the activity-dependent secretion of BDNF when compared with the Val allele (31). Therefore, the Met/Met genotype results in low BDNF protein activity and contributes to decreased hippocampus volumes, as seen in patients with OCD. Based on these findings and our results, we can speculate that the BDNF Val66Met polymorphism may confer susceptibility of hippocampus volumes in patients with OCD, and that is involved in the etiology of OCD.

Data that suggests there is no association of the Val66Met polymorphism with OCD is that the Met allele distributions range from 0 to 72% across different populations. This might explain the inconsistent results in different ethnic-backgrounds (32). The other possible explanation for the lack of association between the BDNF Val66Met polymorphism and OCD is the phenotypic heterogeneity of patients in the OCD sample. Selecting cases according to diagnostic criteria cannot assure an etiologically homogenous sample, because the diagnosis of OCD focuses on symptom-based criteria. Additionally, in the etiological mechanism behind OCD, the BDNF Val66Met polymorphism might not be directly involved, and may function instead to modulate disease development, such as symptom dimensions, age of onset, symptom severity, and family history of symptoms (10, 33–35). It is worth mentioning that the involvement of the BDNF gene may vary among distinct subtypes of OCD. Therefore, it is also possible that SNPs in the BDNF gene may only affect some of the symptoms of OCD (11, 12). Moreover, the Val66Met polymorphism may be in linkage disequilibrium with other functional polymorphisms of the BDNF gene, or in regulatory regions, such as BDNF C-270T, rs11030104, and rs10501087 (36–38). Combine these all together, it could be theorized that the rs6265 variant probably does not have a direct effect, but is in linkage disequilibrium with another causal variant. It can also be assumed that the linkage disequilibrium of rs6265, C-270T, rs11030104, and rs10501087 may vary among ethnic populations, thus contributing to the susceptibility to OCD. This may answer

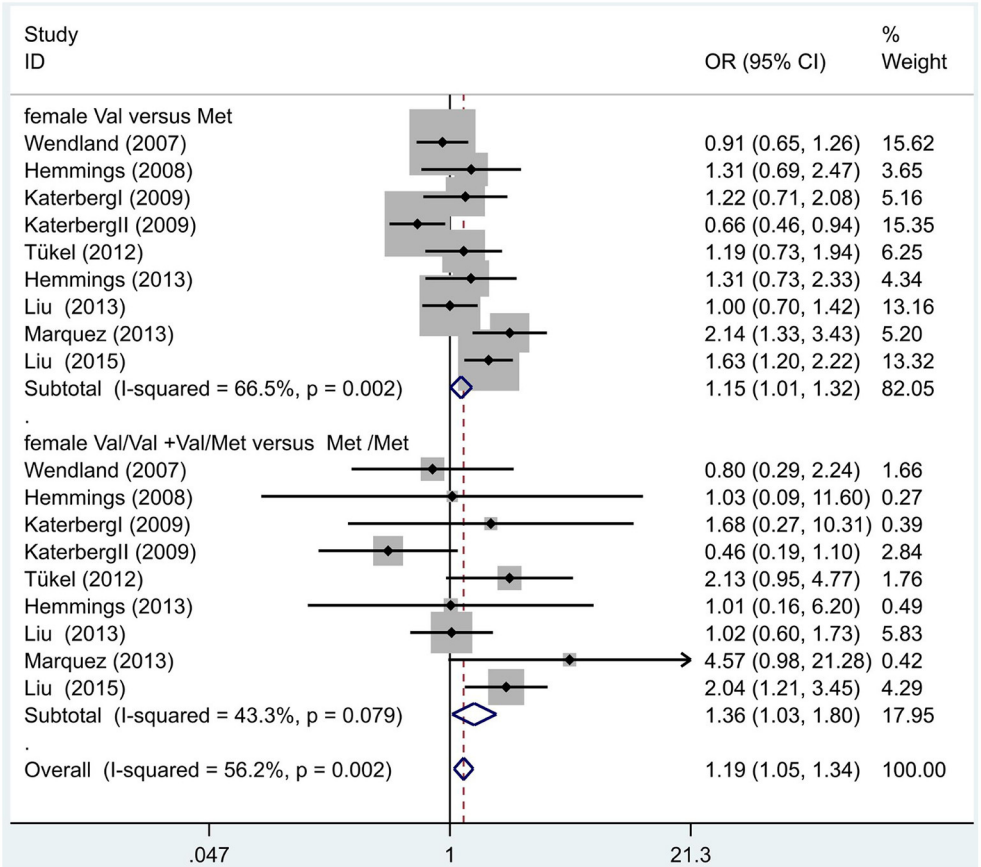


FIGURE 5 | Results of the random-effects meta-analysis for the BDNF Val66Met Val carriers in female OCD and control groups.

the inconsistency seen in the results from studies evaluating the association between the rs6265 polymorphism and OCD in different populaces.

A recent investigation highlighted that those individuals with obsessive–compulsive disorder (OCD) has comorbid lifetime diagnoses of major depressive disorder (MDD; rates <70%). Overlap exists in some common genetic variants (BDNF gene, for example) found across the affective/anxiety disorder spectrums (38). However, according to several meta-analysis of Val66Met polymorphism in the BDNF gene and MDD, the majority of them concluded that Val66Met polymorphism was not associated with MDD (39–41). Verhagen et al. suggested that the BDNF Val66Met polymorphism is of greater importance in the development of MDD in men than in women, while we reported that Val66Met polymorphism is of greater importance in the development of OCD in woman than in men. Future research into gender issues of these two diseases will be of interest (40).

Finally, in order to avoid the shortcomings of previous case-control and family-based studies in elucidating genes that contribute to psychiatric disorders, an endophenotype-based analysis has been recently adopted. This approach allows for a biological underpinning for diagnosis and classification, a net outcome to be established, and may help to identify the susceptibility of genes in the etiology of OCD. Future studies should adopt the endophenotype approach to improve ability to detect susceptibility loci in OCD in order to elucidate its etiology fully.

One major limitation for this study is that except for different genetic background, negative life events, personality, gender and environmental factors may also involve in the etiology of OCD. As it has been demonstrated that exposure to stressful life events before the onset of the illness seems to confer an increased risk for OCD. OCD probands have a significantly greater prevalence of 'anxious' personality disorders in general and avoidant and obsessive-compulsive personality disorders as compared to controls (39–41), while environmental influences have shown a modest but significant impact on the OCD (42). Thus, in future study we combine BDNF val66met polymorphism, subclinical and environment factors interaction in the etiology of OCD. It has been demonstrated that gene–environment interactions are more likely to be involved in the etiology of OCD.

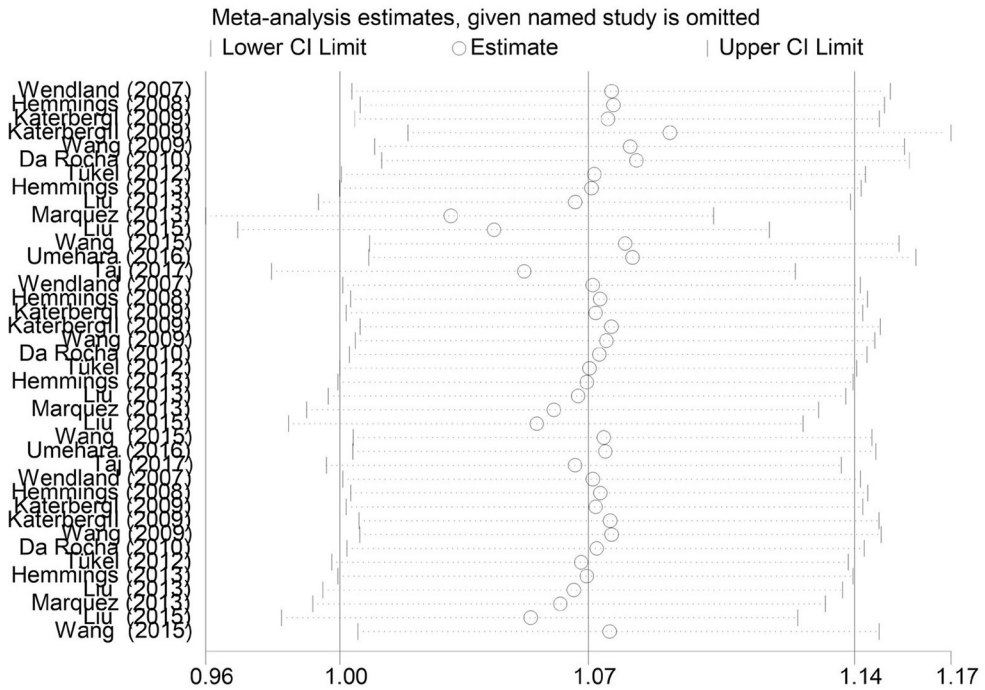


FIGURE 6 | Results of sensitivity analysis.

To conclude, our updated study suggests that the female BDNF Val66Met Val carriers may have an increased susceptibility to OCD.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

XZ and HL conceived and designed the experiments. YS, NW, and EZ performed the experiments. QL and NW analyzed the data. YS wrote the article. All authors contributed to the article and approved the submitted version.

FUNDING

A National Natural Science Foundation of China (No. 81371494) supported this study.

ACKNOWLEDGMENTS

We would like to thank all the subjects who served as research participants.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2021.685041/full#supplementary-material

REFERENCES

1. van Grootheest DS, Cath DC, Beekman AT, Boomsma DI. Twin studies on obsessive-compulsive disorder: a review. *Twin Res Hum Genet.* (2005) 8:450–8. doi: 10.1375/twin.8.5.450
2. Park H, Poo MM. Neurotrophin regulation of neural circuit development and function. *Nat Rev Neurosci.* (2013) 14:7–23. doi: 10.1038/nrn3379
3. Bathina S, Das UN. Brain-derived neurotrophic factor and its clinical implications. *Arch Med Sci.* (2015) 11:1164–78. doi: 10.5114/aoms.2015.56342
4. Guo HR, Huang BL, Wang YL, Zhang YY, Ma QG, Lv PP, et al. Effect of escitalopram on serum GDNF and BDNF levels and 5-HT level of brain tissue of obsessive-compulsive disorder rats. *Cell Mol Neurobiol.* (2020) 40:991–7. doi: 10.1007/s10571-020-00788-4
5. Maina G, Rosso G, Zanardini R, Bogetto F, Gennarelli M, Bocchio-Chiavetto L. Serum levels of brain-derived neurotrophic factor in drug-na?ve obsessive-compulsive patients: a case -control study. *J Affect Disord.* (2010) 122:174–8. doi: 10.1016/j.jad.2009.07.009
6. Wang Y, Mathews CA, Li Y, Lin Z, Xiao Z. Brain-derived neurotrophic factor (BDNF) plasma levels in drug-na?ve OCD patients are lower than those in

healthy people, but are not lower than those in drug-treated OCD patients. *J Affect Disord.* (2011) 133:305–10. doi: 10.1016/j.jad.2011.04.002
7. Chen ZY, Jing D, Bath KG, Ieraci A, Khan T, Siao CJ, et al. Genetic variant BDNF (Val66Met) polymorphism alters anxiety-related behavior. *Science.* (2006) 314:140–3. doi: 10.1126/science.1129663
8. Wang S, Xu X, Yan P, Song M, Li J, Wang S. Is Brain-Derived Neurotrophic Factor (BDNF) Val66Met polymorphism associated with obsessive-compulsive disorder? A meta-analysis. *Psychiatria Danubina.* (2019) 31:141–7. doi: 10.24869/psyd.2019.141
9. Zhao X, Huang Y, Ma H, Jin Q, Wang Y, Zhu G. Association between major depressive disorder and the norepinephrine transporter polymorphisms T-182C and G1287A: a meta-analysis. *J Affect Disord.* (2013) 150:23–8. doi: 10.1016/j.jad.2013.03.016
10. Wendland JR, Kruse MR, Cromer KR, Cromer KC, Murphy DL. A large case-control study of common functional SLC6A4 and BDNF variants in obsessive-compulsive disorder. *Neuropsychopharmacology.* (2007) 32:2543–51. doi: 10.1038/sj.npp.1301394
11. Hemmings SM, Kinnear CJ, Van der Merwe L, Lochner C, Corfield VA, Moolman-Smook JC, et al. Investigating the role of the brain-derived neurotrophic factor (BDNF) val66met variant in obsessive-compulsive disorder (OCD). *World J Biol Psychiatry.* (2008) 9:126–34. doi: 10.1080/15622970701245003
12. Katerberg H, Lochner C, Cath DC, de Jonge P, Bochdanovits Z, Moolman-Smook JC, et al. The role of the brain-derived neurotrophic factor (BDNF) val66met variant in the phenotypic expression of obsessive-compulsive disorder (OCD). *Am J Med Genet B Neuropsychiatr Genet.* (2009) 150b:1050–62. doi: 10.1002/ajmg.b.30930
13. Wang Y, Xiao ZP, Yu SY, Li Y, Wang DX. Association study of brain-derived neurotrophic factor gene polymorphism with generalized anxiety disorder and obsessive-compulsive disorder. *Chin J Neuropsychiatr Dis.* (2009) 10:623–5. doi: 10.3969/j.issn.1002-0152.2009.10.013
14. Da Rocha FF, Malloy-Diniz L, Lage NV, Corrêa H. Positive association between MET allele (BDNF Val66Met polymorphism) and obsessive-compulsive disorder. *Rev Bras Psiquiatr.* (2010) 32:323–4. doi: 10.1590/S1516-44462010000300023
15. Tükel R, Gürvit H, Ozata B, Oztürk N, Ertekin BA, Ertekin E, et al. Brain-derived neurotrophic factor gene Val66Met polymorphism and cognitive function in obsessive-compulsive disorder. *Am J Med Genet B Neuropsychiatr Genet.* (2012) 159b:850–8. doi: 10.1002/ajmg.b.32092
16. Márquez L, Camarena B, Hernández S, Lóyzaga C, Vargas L, Nicolini H. Association study between BDNF gene variants and Mexican patients with obsessive-compulsive disorder. *Eur Neuropsychopharmacol.* (2013) 23:1600–5. doi: 10.1016/j.euroneuro.2013.08.001
17. Liu YH, Liu SG, Kou HY, Zhang XH. Association study of brain-derived neurotrophic factor (BDNF) genetic polymorphism and obsessivecompulsive disorder. *J Clin Psychiatry.* (2013) 1:29–31.
18. Hemmings SM, Lochner C, van der Merwe L, Cath DC, Seedat S, Stein DJ. BDNF Val66Met modifies the risk of childhood trauma on obsessive-compulsive disorder. *J Psychiatr Res.* (2013) 47:1857–63. doi: 10.1016/j.jpsychires.2013.08.012
19. Liu S, Cui J, Niu Z, Yi M, Zhang X, Che F, et al. Do obsessive-compulsive disorder and Tourette syndrome share a common susceptibility gene? An association study of the BDNF Val66Met polymorphism in the Chinese Han population. *World J Biol Psychiatry.* (2015) 16:602–9. doi: 10.3109/15622975.2015.1012226
20. Wang J, Zhang F, Zhu W, Liu Y, Zhou Z. Meta-analysis of the association of brain-derived neurotrophic factor Val66Met polymorphism with obsessive-compulsive disorder. *Acta Neuropsychiatr.* (2015) 27:327–35. doi: 10.1017/neu.2015.38
21. Umehara H, Numata S, Kinoshita M, Watanabe S, Nakaaki S, Sumitani S, et al. No association between BDNF Val66Met polymorphism and treatment response in obsessive-compulsive disorder in the Japanese population. *Neuropsychiatr Dis Treat.* (2016) 12:611–5. doi: 10.2147/NDT.S102100
22. Taj MJ RJ, Ganesh S, Shukla T, Deolankar S, Nadella RK, Sen S, et al. BDNF gene and obsessive compulsive disorder risk, symptom dimensions and treatment response. *Asian J Psychiatr.* (2018) 38:65–9. doi: 10.1016/j.ajp.2017.10.014
23. Castle DJ, Deale A, Marks IM. Gender differences in obsessive compulsive disorder. *Aust N Z J Psychiatry.* (1995) 29:114–7. doi: 10.3109/00048679509075899
24. Matsuki C, To M, Kondo Y, Sugiyama H, Yamamoto Y, Shimizu T, et al. Associations between brain-derived neurotrophic factor and estradiol in women's saliva. *Neuro Endocrinol Lett.* (2014) 35:236–41.
25. Sohrabji F, Miranda RC, Toran-Allerand CD. Identification of a putative estrogen response element in the gene encoding brain-derived neurotrophic factor. *Proc Natl Acad Sci USA.* (1995) 92:11110–4. doi: 10.1073/pnas.92.24.11110
26. Singh M, S|ťcl® G, Guan X, Warren M, Toran-Allerand CD. Estrogen-induced activation of mitogen-activated protein kinase in cerebral cortical explants: convergence of estrogen and neurotrophin signaling pathways. *J Neurosci.* (1999) 19:1179–88. doi: 10.1523/JNEUROSCI.19-04-01179.1999
27. Atmaca M, Yildirim H, Ozdemir H, Ozler S, Kara B, Ozler Z, et al. Hippocampus and amygdalar volumes in patients with refractory obsessive-compulsive disorder. *Prog Neuropsychopharmacol Biol Psychiatry.* (2008) 32:1283–6. doi: 10.1016/j.pnpbp.2008.04.002
28. Fouche JP, du Plessis S, Hattingh C, Roos A, Lochner C, Soriano-Mas C, et al. Cortical thickness in obsessive-compulsive disorder: multisite mega-analysis of 780 brain scans from six centres. *Br J Psychiatry.* (2017) 210:67–74. doi: 10.1192/bjp.bp.115.164020
29. Egan MF, Weinberger DR, Lu B. Schizophrenia, III: brain-derived neurotropic factor and genetic risk. *Am J Psychiatry.* (2003) 160:1242. doi: 10.1176/appi.ajp.160.7.1242
30. Hariri AR, Goldberg TE, Mattay VS, Kolachana BS, Callicott JH, Egan MF, et al. Brain-derived neurotrophic factor val66met polymorphism affects human memory-related hippocampal activity and predicts memory performance. *J Neurosci.* (2003) 23:6690–4. doi: 10.1523/JNEUROSCI.23-17-06690.2003
31. Egan MF, Kojima M, Callicott JH, Goldberg TE, Kolachana BS, Bertolino A, et al. The BDNF val66met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function. *Cell.* (2003) 112:257–69. doi: 10.1016/S0092-8674(03)00035-7
32. Petryshen TL, Sabeti PC, Aldinger KA, Fry B, Fan JB, Schaffner SF, et al. Population genetic study of the brain-derived neurotrophic factor (BDNF) gene. *Mol Psychiatry.* (2010) 15:810–5. doi: 10.1038/mp.2009.24
33. Hall D, Dhilla A, Charalambous A, Gogos JA, Karayiorgou M. Sequence variants of the brain-derived neurotrophic factor (BDNF) gene are strongly associated with obsessive-compulsive disorder. *Am J Hum Genet.* (2003) 73:370–6. doi: 10.1086/377003
34. Zai G, Arnold P, Strauss J, King N, Burroughs E, Richter MA, et al. No association between brain-derived neurotrophic factor gene and obsessive-compulsive disorder. *Psychiatr Genet.* (2005) 15:235. doi: 10.1097/00041444-200512000-00002
35. Alonso P, Gratacòs M, Menchón JM, Saiz-Ruiz J, Segalàs C, Baca-García E, et al. Extensive genotyping of the BDNF and NTRK2 genes define protective haplotypes against obsessive-compulsive disorder. *Biol Psychiatry.* (2008) 63:619–28. doi: 10.1016/j.biopsych.2007.06.020
36. Zhang JP, Lencz T, Geisler S, DeRosse P, Bromet EJ, Malhotra AK. Genetic variation in BDNF is associated with antipsychotic treatment resistance in patients with schizophrenia. *Schizophr Res.* (2013) 146:285–8. doi: 10.1016/j.schres.2013.01.020
37. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. *Science.* (2002) 296:2225–9. doi: 10.1126/science.1069424
38. Kang JI, Song DH, Namkoong K, Kim SJ. Interaction effects between COMT and BDNF polymorphisms on boredom susceptibility of sensation seeking traits. *Psychiatry Res.* (2010) 178:132–6. doi: 10.1016/j.psychres.2010.04.001
39. Gothelf D, Aharonovsky O, Horesh N, Carty T, Apter A. Life events and personality factors in children and adolescents with obsessive-compulsive disorder and other anxiety disorders. *Comprehen psychiatry.* (2004) 45:192–8. doi: 10.1016/j.comppsych.2004.02.010
40. Naguy A, Moodliar-Rensburg S, Alamiri B. Obsessive-compulsive disorder in women: any gender-based differences? *J Nerv Ment Dis.* (2021) 209:150. doi: 10.1097/NMD.0000000000001260
41. Samuels J, Nestadt G, Bienvenu OJ, Costa PT, Jr., Riddle MA, et al. Personality disorders and normal personality dimensions in obsessive-compulsive disorder. *Brit J Psychiatry.* (2000) 177:457–62. doi: 10.1192/bjp.177.5.457

42. Vidal-Ribas P, Stringaris A, Rück C, Serlachius E, Lichtenstein P, Mataix-Cols D. Are stressful life events causally related to the severity of obsessive-compulsive symptoms? A monozygotic twin difference study. *Eur Psychiatry*. (2015) 30:309–16. doi: 10.1016/j.eurpsy.2014.11.008

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership