An integrated molecular landscape implicates the regulation of dendritic spine formation through insulin-related signalling in obsessive–compulsive disorder

Ilse van de Vondervoort, MSc*; Geert Poelmans, MD, PhD*; Armaz Aschrafi, PhD; David L. Pauls, PhD; Jan K. Buitelaar, MD, PhD; Jeffrey C. Glennon, PhD†; Barbara Franke, PhD†

Introduction

Obsessive–compulsive disorder (OCD) is a neuropsychiatric disorder that affects an estimated 2.5% of the world's population1,2 and is characterized by obsessions (recurrent, intrusive, persistent thoughts, impulses and/or ideas that often cause anxiety or distress) and compulsions (ritualized and stereotypic behaviours or mental acts that are often performed to relieve anxiety or distress associated with obsessions). Although OCD is a heritable disorder, its complex molecular etiology is poorly understood. Methods: We combined enrichment analyses and an elaborate literature review of the top-ranked genes emerging from the 2 published genome-wide association studies of OCD and candidate genes implicated through other evidence in order to identify biological processes that, when dysregulated, increase the risk for OCD. Results: The resulting molecular protein landscape was enriched for proteins involved in regulating postsynaptic dendritic spine formation — and hence synaptic plasticity — through insulin-dependent molecular signalling cascades. Limitations: This study is a first attempt to integrate molecular information from different sources in order to identify biological mechanisms underlying OCD etiology. Our findings are constrained by the limited information from hypothesis-free studies and the incompleteness and existing limitations of the OCD literature and the gene function annotations of gene enrichment tools. As this study was solely based on in silico analyses, experimental validation of the provided hypotheses is warranted. Conclusion: Our work suggests a key role for insulin and insulin-related signalling in OCD etiology and — if confirmed by independent studies — could eventually pave the way for the development of novel OCD treatments.
serotonin,12 glutamate13 and dopamine14 neurotransmitter systems have been implicated in OCD etiology through candidate gene association studies. Furthermore, twin studies have revealed that OCD is influenced by genetic factors, with a heritability of approximately 40%15,16 and this was confirmed by a recent genomic heritability analysis based on common single nucleotide polymorphism (SNP) data that estimated the heritability of OCD at 37%.17 In addition, environmental factors and gene–environment interactions are thought to play a role in the etiology of the disorder.16

The genetic component of the disorder is likely to be highly complex, with multiple genetic variants of small individual effect size contributing to disease risk. Candidate gene association studies have implicated several genes in OCD, including genes encoding proteins involved in the serotonergic (e.g., SLC6A423), glutamatergic (e.g., SLC1A120), catecholaminergic (COMT and MAOA23) and neurotrophic (e.g., BDNF23) systems. In addition, an increasing body of evidence from animal models suggests that SLITRK5,23 SAPAP3,23 HOXB8,23 and FKBP1A25 may play an important role in the etiology of OCD. Recently, 2 independent genome-wide association studies (GWAS) of OCD were published,26,27 which provides the first opportunity to investigate the molecular mechanisms underlying OCD based on hypothesis-free, unbiased genetic data. In 2013, Stewart and colleagues26 published the first GWAS of OCD, which contained 2 independent discovery samples (i.e., a case–control sample and a trio sample). The authors generated genome-wide SNP genotyping data for 1465 individuals affected with OCD according to the DSM-IV criteria,3 5557 ancestry-matched controls and 400 complete trios. In the case–control analysis, the 2 SNPs with the smallest p values (p = 0.00000249 and p = 0.00000344) were located within DLGAP1, which codes for an important component of the postsynaptic scaffolding complex in neuronal cells.28 In the trio analysis, rs6131295, near BTBD3, exceeded the genome-wide significance threshold (p = 0.0000000384). BTBD3 regulates dendrite orientation toward axons in various mammalian cortical areas.29 However, in the meta-analyses of the 2 samples (case–control and trios) the significance of this locus was reduced considerably. More recently, Mattheisen and colleagues27 published a second GWAS of OCD, for which they used family-based association testing with additional population controls. The discovery sample contained 5061 individuals in total, including 1065 families comprising 1406 patients with OCD (based on DSM-IV criteria) and 1489 control individuals and an additional 1984 unrelated controls. Although Mattheisen and colleagues did not identify any genome-wide significant associations, they found nominally significant evidence of an association between OCD and several genes encoding proteins that interact with DLGAP1, the most significant finding from the case–control discovery sample in the first GWAS of OCD.

Aiming to increase our insights into the biological processes implicated in the disorder, we performed gene enrichment analyses and an elaborate literature review of the most significant findings from these 2 GWAS and combined this with information on candidate genes/proteins implicated in OCD through other evidence. From this, we generated an integrated molecular landscape to identify biological processes underlying the disorder and thus to take a first step toward the development of novel OCD treatments.

Methods

SNP-based GWAS gene selection

The first step of our GWAS-based molecular landscape building approach30,31 was the selection of candidate genes based on SNPs and their corresponding p values. From the first published GWAS of OCD,26 we selected the SNPs that were associated with OCD at p < 0.0001 in either the trio-based or case–control analysis to compile a list of associated genes, and we did the same for the SNPs from the discovery sample of the second GWAS.27 The selected genes either contained a SNP that was located within an exonic, intronic or untranslated region of the gene or were found within 100 kb downstream and upstream of the SNP. We included genes implicated through 100 kb downstream and upstream flanking regions based on the fact that the vast majority of expression quantitative trait loci (eQTL) for a given gene are located within 100 kb downstream and/or upstream of a gene32–34 and because trait-associated SNPs are more likely eQTL.35 The chosen statistical cut-off for association of p < 0.0001 is often used to designate “suggestive” association and has been used previously in similar studies.30,31

Enrichment analyses

We conducted analyses aimed at identifying biological functions and themes that are enriched in the OCD GWAS genes. For this, we used 2 different tools based on different gene function annotations to find overlap between the enriched functions and themes. First, we performed a network analysis using the Ingenuity Pathway Analysis software package (http://www.ingenuity.com) with the default, preset parameters. To generate protein interaction networks, the Ingenuity software package uses the Ingenuity Knowledge Base, which is based on extensive information from the published literature as well as many other sources, including gene expression and gene function annotation databases. For each network, the Ingenuity software also generates an enrichment score that takes into account the number of eligible molecules/proteins in the network and its size, as well as the total number of network-eligible molecules analyzed and the total number of molecules in the Ingenuity Knowledge Base that could potentially be included in networks. The resulting score is calculated with the right-tailed Fisher exact test and displayed as the negative logarithm of the Fisher exact test result.

Second, we conducted a gene ontology (GO) term enrichment analysis of the OCD GWAS genes using GOstat (http://gostat.wehi.edu.au)30 with default, preset parameters. GOstat is a web-based tool that allows detection of statistically overrepresented GO terms within a group of genes. The program determines the annotated GO terms and calculates the number of appearances of each term for the group of genes as well as a reference group (i.e., the group of
all annotated GO terms in humans). Based on this, it calculates a \( p \) value for each GO term, using a \( \chi^2 \) test or Fisher exact test as appropriate, and corrects these \( p \) values for multiple testing using the Benjamini–Hochberg correction.

For the Ingenuity analysis, we report all significantly enriched protein interaction networks, whereas for the GOstat analysis, we report only the GO terms that contained at least 2 genes and had a corrected enrichment (\( p < 0.01 \)).

Other OCD candidate genes

We compiled a list of other OCD candidate genes that were not directly observed in the top-ranked GWAS findings, but have been implicated in other ways in OCD etiology. The literature on such genes is mixed, with varying support for most genes, and there have been few replicated findings. We critically evaluated the literature and selected only those genes that had received support through findings from (genetic) animal studies, gene mutations, and/or 2 or more independent candidate gene association studies (or at least nominal significance in meta-analysis) and/or mRNA/protein expression studies.

Literature analysis and molecular landscape building

Guided by the results of the enrichment analyses and as a next step of the molecular landscape building approach that we have previously reported for other neuropsychiatric disorders,\(^{26,27}\) we then searched the literature for the (putative) function and neuronal expression of all OCD GWAS candidate genes/proteins as well as the other OCD candidate genes/proteins from the list that we had compiled. For this literature analysis, we used the UniProt Protein Knowledge Base (UniProtKB, http://www.uniprot.org/uniprot)\(^{37}\) to gather basic information on the function of all genes and their encoded proteins. In addition, we searched PubMed for all genes and using the search terms “brain,” “neurosystem,” “neuron,” “neuronal growth,” “neurite,” “axon,” “synapse,” “insulin,” and “insulin receptor.” Further, we searched PubMed for functional interactions among all OCD candidate genes/encoded proteins.

Results

OCD GWAS genes

From the 2 GWAS,\(^{26,27}\) we compiled a list of SNPs for inclusion in the study (Appendix 1, Table S1, available at jpn.ca). These SNPs were located in (the vicinity of) 66 unique protein-coding genes in the first study\(^{26}\) and 23 in the second.\(^{27}\)

Enrichment analyses

The Ingenuity network analysis yielded 6 enriched networks (Appendix 1, Table S2). The network with the highest enrichment score by far (\( p = 1.00E-46 \)) and containing the highest number of proteins (encoded by 20 of the 89 candidate genes emerging from the 2 OCD GWAS) had (pro)insulin located at its centre, interacting with 21 other molecules/proteins, 8 of which are encoded by OCD GWAS candidate genes (Appendix 1, Fig. S1). The insulin receptor signalling pathway was also one of the GO terms that was significantly enriched in the GOstat enrichment analysis (Appendix 1, Table S3). Taken together, both gene enrichment tools that we used showed an enrichment of insulin and insulin-related signalling cascades among the OCD GWAS candidate genes. Additionally, significantly enriched GO terms were related to biological processes regulating nervous system development and function (i.e., “axon guidance,” “axonogenesis,” “generation of neurons,” “nervous system development,” “neurogenesis,” “neuron projection development” and “neuron projection morphogenesis”; Appendix 1, Table S3).

OCD candidate genes

In addition to the top-ranked GWAS findings, an elaborate literature review yielded 26 candidate genes that are implicated in OCD etiology through various types of (genetic) evidence (Appendix 1, Table S4).

Literature analysis and molecular landscape building

On the basis of the findings from the gene enrichment and elaborate literature analyses, we found that the proteins encoded by 40 of the 89 GWAS OCD candidate genes (Appendix 1, Table S1; 45%) and 11 of the 26 other OCD candidate genes (Appendix 1, Table S4; 42%) functionally interact within a distinct molecular landscape. This molecular OCD landscape is shown in Figure 1 and contains a number of signalling cascades that are involved in the endogenous synthesis, secretion and extracellular signalling of insulin in and around postsynaptic dendritic spines and that regulate the formation of these spines, which in turn affects and regulates synaptic plasticity. The insulin-related molecular cascades interact with and regulate the activity of important landscape proteins, including PI3K, RAC1, AKT and MTOR. Two additional signalling cascades in the landscape centre around receptors for the neurotransmitters glutamate and serotonin, which have both been linked to OCD, and both cascades regulate the secretion of insulin that can have autocrine or paracrine effects on (neighbouring) neurons, leading to, among other processes, the modulation of dendritic spine formation. The function and/or expression of several of the key proteins/molecules in the landscape is regulated by nitric oxide (NO) synthesized by the NOS1 enzyme, and by ELAVL1, an RNA-binding protein that upregulates the expression of its target genes by binding to and stabilizing the mRNAs that are produced by these genes. A more detailed description of the evidence linking all the genes and encoded proteins in the landscape is provided in Appendix 1.

Discussion

In this study, we used both gene enrichment and extensive manual literature mining to integrate the top-ranked findings of 2 published GWAS of OCD with other OCD candidate genes in order to build a molecular landscape of biological processes involved in OCD etiology. Our landscape approach
provides compelling evidence for a molecular and functional link between OCD, synaptic plasticity and insulin-related signalling. The finding of altered synaptic plasticity is in keeping with previous literature suggesting an involvement of (dysregulated) synaptic plasticity in OCD pathogenesis. For instance, the differential expression of several synaptic proteins in the landscape, including the glutamate transporter SLC1A1,19 SLITRK522 and SAPAP3 (also known as DLGAP3),23 has been associated with OCD and OCD-like behaviours. Slitrk5 and Sapap3 knockout mice also show alterations in postsynaptic glutamate receptor composition,22,38 and GRIN2B, which has been found to be associated with OCD,39 encodes an N-methyl-D-aspartate (NMDA) glutamate receptor subunit that has an important role in regulating synaptic plasticity.40 Taken together, these data suggest that OCD could indeed have its biological substrate at the level of the synapse and, more specifically, the postsynaptic density and dendritic spines.

To our knowledge, this is the first study to link insulin signalling in the central nervous system (CNS) with OCD. Insulin has been reported to promote both dendritic spine formation and excitatory synaptogenesis through activating the signalling pathways involving PI3K, RAC1, AKT and MTOR,41 as shown in our molecular landscape (Fig. 1). It is worth noting that a link between OCD and dysregulated peripheral insulin signalling (i.e., diabetes mellitus), has also been reported. Winocour and colleagues42 observed increased obsessive symptoms in men with type 1 diabetes, which is

![Molecular landscape](image-url)
caused by an absolute lack of insulin due to breakdown of islet cells in the pancreas. In addition, a recent study found a positive correlation between OCD symptoms and the level of glycosylated hemoglobin, a measure of the severity of type 2 diabetes, which is caused by insulin resistance and hence a relative lack of insulin (signalling).

Although nothing is known about the possible role of insulin in OCD etiology and of currently unknown relevance to CNS insulin signalling, an increasing body of evidence suggests that several neuropsychiatric disorders may have a link with insulin. In 2011 Stern hypothesized that insulin signalling may contribute to autism-spectrum disorder (ASD) etiology in genetically susceptible individuals by regulating PKB and MTOR pathway activation in CNS neurons. Moreover, insulin is known to regulate the function of cocaine-sensitive monoamine transporters in the nucleus accumbens, which provides a link to impulsivity associated with cocaine addiction. Interestingly, these disorders also feature compulsivity-like symptoms, including stereotypical behaviour in patients with ASDs and impulsive/compulsive drug-seeking behaviour in those with addiction disorders. In addition to these effects of insulin and insulin-related signalling on neuropsychiatric disorders at the CNS level, a peripheral effect may also be present. A case–control study found that patients with ADHD had a higher prevalence of type 2 diabetes than controls (0.9% v. 0.4%, respectively). Moreover, plasma insulin concentrations were found to be positively correlated with obsessional craving scores in alcohol-dependent patients. Bozdagi and colleagues recently reported that intraperitoneal injections of human insulin-like growth factor 1 (IGF1), an extracellular molecule present in our OCD landscape (Fig. 1), reversed deficits in hippocampal AMPA glutamate receptor signalling, long-term potentiation and motor performance in the Shank3-deficient ASD mouse model. Nevertheless, one cannot exclude the possibility that the effective site of action of these seemingly peripheral effects of insulin signalling is still within the brain, as insulin can enter the CNS from the periphery by active transport across the blood–brain barrier. Therefore, changes in the peripheral levels of insulin and insulin signalling–related proteins could also have a direct influence on insulin levels and signalling in the brain, which would in turn result in aberrant behaviours.

Limitations

This study is a first attempt to integrate molecular information from different sources in order to identify biological mechanisms underlying OCD etiology. Our findings implicate insulin signalling, but it is important to note that this is unlikely to be the only biological process of importance to OCD. Our study was strongly constrained by the incompleteness of gene function annotations of the gene enrichment tools that we used as well as by the existing limitations of the published OCD literature, gearing toward studies on neurotransmitter-related candidate genes. An additional potential bias comes from the fact that brain-expressed genes tend to be large; therefore, some of the top SNPs from the GWAS of OCD could have been found by chance. At the same time, however, genes that have been replicated for neuropsychiatric disorders are often large, so correcting for this gene size bias may result in the loss of valid candidates. In conjunction with this, it is important to note that this study was solely based on enrichment and literature analyses of published GWAS of OCD and other OCD candidate gene data, without experimental validation. Therefore, future studies should be aimed at replicating/validating our findings in well-powered samples as well as at functionally validating specific proteins and molecules from our OCD landscape in animal models with compulsive behavioural readouts and, at a later stage, in patients with OCD.

Conclusion

We have built a molecular landscape for OCD in which insulin-related signalling regulates and ultimately alters synaptic plasticity, integrating all currently available molecular genetic information. Therefore, when additional OCD GWAS data become available, we propose to conduct genetic validation studies of the insulin-related findings. Furthermore, the fact that insulin-related signalling contributes to the pathogenesis of OCD could lead to the development of novel treatments for OCD patients aimed at modulating (CNS) insulin levels and/or insulin-related signalling.

Acknowledgments: The authors thank the families who made all the genetic studies of OCD possible and the many investigators whose work drives the OCD genetics field forward. The authors also thank the anonymous reviewers for their useful comments and suggestions regarding text revision. The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007–2013) under grant agreement n°278948 (TACTICS). This funding organization has had no involvement with the conception, design, data analysis and interpretation, review and/or any other aspects relating to this paper.

Affiliations: From the Department of Cognitive Neurosciences, Donders Institute for Brain, Cognition and Behaviour, Radboud university medical center, Nijmegen, The Netherlands (van de Vondervoort, Poelmans, Buitelaar, Glennon); the Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands (Poelmans, Franke); the Department of Molecular Animal Physiology, Donders Institute for Brain, Cognition and Behaviour, Radboud University for Molecular Life Sciences (RIMLS), Radboud University, Nijmegen, The Netherlands (Poelmans); the Department of Neuroinformatics, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands (Aschrafi); the Psychiatric and Neurodevelopmental Genetics Unit, Centre for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA (Pauls); the Department of Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud university medical center, Nijmegen, The Netherlands (Franke); and the Karakter Child and Adolescent Psychiatry University Centre, Nijmegen, The Netherlands (Buitelaar).

Competing interests: J. Buitelaar declares lecture fees from Janssen Cilag, Lilly, Shire and Medice; personal fees from Shire and Lundbeck; and grants from Lundbeck and Vifor outside the submitted work. B. Franke declares speaker fees from Merz. No other competing interests declared.

Contributors: I. van de Vondervoort, G. Poelmans, J. Glennon and B. Franke designed the study. I. van de Vondervoort and G. Poelmans acquired the data, which all authors analyzed. I. van de Vondervoort, G. Poelmans and B. Franke wrote the article, which all authors reviewed and approved for publication.
References

22. Lee CC, Huang CC, Hsu KS. Insulin promotes dendritic spine and synapse formation and synapse formation by the PI3K/Akt/mTOR and Rac1 signaling pathways. Neuropharmacology 2011;61:867-79.