# Comprehensive pathway analyses of schizophrenia risk loci point to dysfunctional postsynaptic signaling

Running title: Comprehensive pathway analyses in schizophrenia

Dick Schijven MSc [1,2,3,7], Daniel Kofink MSc [4], Vinicius Tragante PhD [4], Marloes Verkerke BSc [3], Sara L. Pulit PhD [2], René S. Kahn MD PhD [1,5], Jan H. Veldink MD PhD [2], Christiaan H. Vinkers MD PhD [1], Marco P. Boks MD PhD [1], Jurjen J. Luykx MD PhD [1,3,6]

1. Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht University, the Netherlands
2. Department of Neurology and Neurosurgery, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht University, the Netherlands
3. Department of Translational Neuroscience, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht University, the Netherlands
4. Department of Cardiology, Division Heart and Lungs, University Medical Center Utrecht, Utrecht University the Netherlands
5. Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA
6. Department of Psychiatry, ZNA hospitals, Antwerp, Belgium

[7] Corresponding author:

Dick Schijven

Brain Center Rudolf Magnus, University Medical Center Utrecht, Universiteitsweg 100, 3584CG Utrecht, the Netherlands

Email: D.Schijven@umcutrecht.nl

Telephone: +31 88 75 676 81

Fax: +31 88 75 518 14

**Abstract**

Large-scale genome-wide association studies (GWAS) have implicated many low-penetrance loci in schizophrenia. However, its pathological mechanisms are poorly understood, which in turn hampers the development of novel pharmacological treatments. Pathway and gene set analyses carry the potential to generate hypotheses about disease mechanisms and have provided biological context to genome-wide data of schizophrenia. We aimed to examine which biological processes are likely candidates to underlie schizophrenia by integrating novel and powerful pathway analysis tools using data from the largest published schizophrenia GWAS (N = 79,845). By applying a primary unbiased analysis (Multi-marker Analysis of GenoMic Annotation; MAGMA) to weigh the role of biological processes from the MSigDB database, we identified enrichment of common variants in synaptic plasticity and neuron differentiation gene sets. We supported these findings using MAGMA, Meta-Analysis Gene-set Enrichment of variaNT Associations (MAGENTA) and Interval Enrichment Analysis (INRICH) on detailed synaptic signaling pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) and found enrichment in mainly the dopaminergic synapse and long-term potentiation. Moreover, shared genes involved in these neurotransmitter systems had a large contribution to the observed enrichment, protein products of top genes in these pathways showed more direct and indirect interactions than expected by chance, and expression profiles of these genes were largely similar among brain tissues. In conclusion, we provide strong and consistent genetics and protein-interaction informed evidence for the role of postsynaptic signaling processes in schizophrenia, opening avenues for future psychopharmacological studies.

## Introduction

Antipsychotics target synaptic signaling by changing neurotransmission of the dopamine D2 receptor (DRD2) and the serotonin 5-$HT_2$ receptor [1,2]. These drugs are the mainstay treatment modality in schizophrenia and are fairly effective at reducing positive symptoms [3]. However, improving the cognitive and negative symptoms, which substantially affect quality of life, has proven challenging [2,4,5]. Although post-mortem studies, imaging and human genetic studies have contributed to theories about pathophysiological mechanisms in schizophrenia, the underlying molecular processes have not been fully elucidated.

Genetic studies provide a valuable resource to investigate the mechanisms that are likely at play in schizophrenia. Schizophrenia is highly heritable ($h^2$ estimates ranging from 45-80%) and polygenic [6,7]. Genome-wide association studies (GWAS) have identified 108 independent associated risk loci [8]. More recently, 142 associated loci were reported in a meta-analysis combining this GWAS with a British study population [9].

Pathway and gene set enrichment analysis methods are widely used to provide biological context to the results of genetic association studies by testing whether biologically relevant pathways or sets of genes are enriched for genetic variants associated with a phenotype [10]. These analyses have been widely applied to schizophrenia, providing evidence for the involvement of synaptic and immune-related processes [11,12]. Such findings are supported by pathway analyses in combined psychiatric disorders (schizophrenia, depression and bipolar disorder), revealing enrichment of genetic variants in neuronal, immune and histone pathways [13]. Involvement of calcium signaling and ion channels in schizophrenia has been reported in a gene set analysis paper combining GWAS data with post-mortem brain gene expression data [14]. Importantly, none of the abovementioned pathway analysis studies has used the full dataset reported in the latest (and largest) published schizophrenia GWAS [8]. Moreover, several novel or widely used pathway analysis tools have not yet been applied to

this schizophrenia GWAS. These tools constitute fast and powerful approaches to test gene set enrichment, despite variability in their correcting for confounding factors that may increase the type 1 error rate [10]. Additionally, tools and databases aimed at the integration of GWAS data with gene expression and protein-protein interaction data allow to further explore the biological impact of common variants associated with schizophrenia [15,16].

Aiming to comprehensively investigate the possible biological processes underlying schizophrenia, we set out to apply gene set and pathway enrichment analysis methods to the largest published GWAS in schizophrenia[8] and additionally integrate the results of these analyses with data on protein-protein interactions and tissue-specific gene expression (Figure 1). Our results elucidate the involvement of neuron differentiation and synaptic plasticity in schizophrenia, and reveal an accumulation of variants in post-synaptic signaling cascades. They moreover enable a more nuanced understanding of the several actionable classes of neurotransmitters implicated in the disease.

## Methods

### Input data and analysis overview

We used summary-level results from the largest and most recent GWAS in schizophrenia, made publically available by the Psychiatric Genomics Consortium (www.med.unc.edu/pgc/results-and-downloads; downloaded on 10 May, 2017)[8]. This GWAS was performed in 34,341 schizophrenia cases and 45,604 healthy controls and association results for 9.4 million single-nucleotide polymorphisms (SNPs) were reported in the summary-level data. We considered replication of our findings in independent datasets, e.g. the UK Biobank. However, low numbers of schizophrenia cases (170 ICD-10 and 337 self-reported) preclude any sensible pathway analyses. As detailed below (also see Figure 1), using Multi-marker Analysis of GenoMic Annotation (MAGMA)[17] we successively (A) mapped SNPs to genes, (B) calculated gene p-values based on GWAS SNP p-values, (C) performed a primary gene set enrichment analysis using gene ontology (GO) terms, and (D) tested the robustness of these findings in detailed molecular pathways derived from KEGG. Finally, we further investigated the results of the analysis on KEGG pathways using Meta-Analysis Gene-set Enrichment of variaNT Associations (MAGENTA) and Interval Enrichment Analysis (INRICH)[18,19], applied protein-protein interaction (PPI) analysis using Disease Association Protein-Protein Link Evaluator (DAPPLE)[15], and assessed tissue-specific expression using data from the Genotype-Tissue Expression (GTEx) project [16].

### Mapping SNPs to genes and assigning p-values to genes

SNPs present in the European subset of the 1000 Genomes Phase 3 dataset were extracted from the GWAS summary-level results [20]. Using MAGMA v1.06, we mapped SNPs to corresponding genes, extending gene footprints by an additional 20 kilobase (kb) up- and downstream, as a large proportion of regulatory elements involved in gene expression

regulation is likely to be captured by including this region [21]. We then applied a gene analysis to obtain a p-value for each gene to which at least one SNP was mapped. The p-value of a gene was based on the mean association statistic of SNPs contained in that gene. Linkage disequilibrium (LD) between SNPs included in the gene analysis was estimated from the European subset of 1000 Genomes Phase 3, which corresponds best with the ancestry of samples in the schizophrenia GWAS [8].

**Gene Ontology (GO) term enrichment analysis**

We downloaded 4,436 GO biological process gene sets from the Molecular Signatures Database (MSigDB, release 6.0, April 2017) [22,23]. MSigDB merged original GO terms with high similarity (Jaccard's coefficient > 0.85) into one gene set and the number of genes per gene set was limited to $10 < n_{genes} < 2000$. We applied a competitive gene set analysis -which tests whether the genes in a gene set are more strongly associated with the phenotype of interest than random sets of genes- using gene p-values from the previous step. The significance threshold (at $\alpha = 0.05$) was adjusted for multiple GO terms tested using a permutation procedure implemented in MAGMA (10,000 permutations, $p = 4.23\times10^{-6}$). A correction for confounding factors (gene size, gene density and minor allele count) was applied. We additionally tested the robustness of our analysis by implementing several sensitivity analyses: 1) excluding the MHC region on chromosome 6 (chr6:25500000-33500000, human genome assembly GRCh37/hg19), 2) excluding the X-chomosome from the analysis, and 3) applying strict filtering for gene set size ($10 \leq n_{genes} \leq 200$) as larger gene sets are sometimes perceived as too broad or uninterpretable.

**Follow-up using KEGG pathways**

To further test the results of our primary GO term analysis, we used different gene sets representing synaptic signaling processes from the Kyoto Encyclopedia of Genes and Genomes (KEGG, www.kegg.jp, downloaded on 3 January, 2017) [24] as input for MAGMA: hsa04724, Glutamatergic synapse (114 genes); hsa04725 Cholinergic synapse (111 genes); hsa04726 Serotonergic synapse (113 genes); hsa04727, GABAergic synapse (88 genes); hsa04728, Dopaminergic synapse (130 genes); hsa04720, Long-term potentiation (67 genes); hsa04730, Long-term depression (60 genes). We then ran a similar competitive gene set analysis on these pathways (multiple-testing corrected $p = 9.0\times10^{-3}$). For each significantly associated pathway, we mapped top associated genes (gene $p < 1.45\times10^{-4}$: 0.05/344 unique genes included in the enrichment analysis of the selected KEGG pathways) to pathway components using *pathview* in R version 3.3.3 (www.r-project.org) [25]. Because many genes overlapped between two or more of the tested KEGG pathways (Supplementary Figure S1), we conditioned the enrichment specific to each KEGG pathway on the association signal of genes shared with other tested KEGG pathways (built-in function in MAGMA). Additionally, we tested enrichment of sets of unique genes per KEGG pathway and a gene set containing all shared genes.

To support the results of the MAGMA analysis, we tested enrichment in the same KEGG pathways using MAGENTA and INRICH. MAGENTA calculated gene p-values based on the lowest SNP p-value in a gene (extended 20 kb up- and downstream) while correcting for confounding factors, such as gene length and gene overlap. It then assessed whether a gene set was enriched with low gene p-values at the 95$^{th}$ percentile cut-off (based on all gene p-values in the genome) compared to randomly sampled gene sets of similar size [18]. Enrichment p-values were corrected for multiple testing using false discovery rate (FDR < 0.05). INRICH tests whether sets of genes are enriched with association signals from independent genomic intervals [19]. Intervals were calculated around genome-wide significant SNPs ($p < 5\times10^{-8}$)

using the clump function in PLINK v1.90b3z [26]. Only SNPs with a LD $R^2 > 0.5$ and an association p-value < 0.05 were included in an interval, resulting in 114 non-overlapping intervals. Target gene regions were extended with 20 kb up- and downstream. Enrichment of KEGG pathways was tested using a permutation procedure where the enrichment of a gene set was tested against the normal distribution of enrichment in background gene sets. Enrichment p-values were corrected for multiple testing by a bootstrap method in INRICH.

**Protein-protein interaction analysis**

To increase insight into the biological impact of the pathways implicated in schizophrenia by our analyses, we performed PPI analysis using DAPPLE v0.17 [15]. We used DAPPLE to test whether proteins in a network comprised of 22 unique top genes from the KEGG dopaminergic, cholinergic or long-term potentiation pathways (gene $p < 1.45\times10^{-4}$) show more direct and indirect interactions with each other and with other proteins than expected by chance.

**Gene expression analysis**

Median relative transcript abundances (transcripts per milion, TPM) among 13 brain tissues and 40 non-brain tissues were obtained from the GTEx database (v7, RNASeQCv1.8.8, downloaded on 01-11-2017 from https://www.gtexportal.org/home/datasets) [16]. Per enriched KEGG pathway, we extracted TPM levels of top genes (gene $p < 1.45\times10^{-4}$) and visualized the median relative abundance of transcripts in heatmaps for each enriched pathway and for one set of genes that were shared between two or more of the enriched KEGG pathways using the R-package *gplots* [27]. Genes with high similarities in transcript abundance, as well as tissues with similar transcript abundance, were clustered.

**Results**

Primary gene set enrichment analysis in MAGMA (in 4,436 MSigDB GO biological processes) identified two gene sets that were enriched for schizophrenia-associated SNPs (Figure 2A and Supplementary Table S1): 'Regulation of synaptic plasticity' ($p = 6.52\times10^{-7}$) and 'Regulation of neuron differentiation' ($p = 1.00\times10^{-6}$), which are specialized GO terms of synaptic signaling and neuron differentiation (Supplementary Figure S2). Excluding the extended MHC region or the X-chromosome did not affect the outcome this analysis. When we applied strict filtering for gene set size, we found that the 'Regulation of neuron differentiation' GO term (containing 551 genes) was excluded from the analysis and only 'Regulation of synaptic plasticity' remained significant (Supplementary Figure S3).

To gain a more nuanced understanding of molecular synaptic pathways enriched for schizophrenia-associated variants, we tested enrichment of SNPs in KEGG pathways representing synaptic signaling. We found significant enrichment in pathways representing the dopaminergic synapse ($p = 1.62\times10^{-4}$), cholinergic synapse ($p = 1.51\times10^{-3}$) and long-term potentiation ($p = 4.00\times10^{-3}$) (Figure 2B, Supplementary Table S2). For each significant pathway, we mapped top genes (from the MAGMA gene analysis step, see Supplementary Table S3) on components within these KEGG pathways (Figure 3A, Supplementary Figure S4). SNP enrichment was mostly restricted to trans-membrane and postsynaptic components in the cholinergic and dopaminergic synapses. The long-term potentiation pathway only included post-synaptic components. We highlighted strong enrichment in signaling through extracellular signal-regulated kinase (ERK) and cAMP response element-binding protein (CREB), phospholipase C (PLC) and the inositol trisphosphate receptor ($IP_3R$), and protein kinase B (PKB/Akt). These cascades converge on mechanisms involved in synaptic growth regulation and synaptic plasticity. Furthermore, voltage-gated calcium channels,

glutamatergic NMDA and AMPA receptors, the dopamine D2 receptor (DRD2) and the muscarinic acetylcholine receptor $M_4$ were highly enriched.

Protein products of 22 unique top genes from enriched KEGG pathways showed more direct interactions than expected by chance ($p = 1.4\times10^{-2}$) and on more indirect connections to other proteins ($p = 4.0\times10^{-3}$), but not direct connections to other proteins ($p = 0.81$) (Figure 3B,C). Gene expression analysis revealed no large differences in transcript abundance for top genes among brain tissues, although relative expression seems higher in cerebellum and cerebellar hemisphere compared to other brain tissues (Supplementary Figure S5).

When we conditioned the enrichment of each KEGG pathways on shared genes with other tested pathways, all enrichment p-values increased to non-significance ($p > 9.0\times10^{-3}$), indicating a considerable contribution of shared genes to the enrichment signal (Supplementary Figure S6). Shared genes tested as a single gene set were however not enriched after multiple testing correction in a combined analysis with sets of unique genes per pathway ($p = 9.4\times10^{-3}$) (Supplementary Figure S7).

Using MAGENTA and INRICH with the same data and gene sets, we confirmed significant enrichment of the dopaminergic synapse (MAGENTA corrected FDR = $8.00\times10^{-4}$, INRICH corrected p-value = $2.82\times10^{-2}$) and long-term potentiation (MAGENTA corrected FDR = $1.14\times10^{-2}$, INRICH corrected p-value = $9.6\times10^{-3}$). Enrichment for the cholinergic synapse was only confirmed in MAGENTA (corrected FDR = $4.12\times10^{-2}$) (Figure 2B).

## Discussion

By implementing complementary gene set enrichment analysis tools (MAGMA, MAGENTA and INRICH), annotations from biological databases (MSigDB/GO and KEGG), protein-protein interaction data, and tissue-specific gene expression into a comprehensive analysis, we aimed to elucidate biological processes underlying schizophrenia. Thus, we first detected enrichment of schizophrenia-associated SNPs in synaptic plasticity and neuronal differentiation processes, which is in line with biological hypotheses that previous studies have made [11,13,14]. We did however not replicate significant enrichment of immune and histone pathways (Supplementary Figure S8), which has also been reported as highly associated to psychiatric disease [13]. Possible reasons for this are the larger sample size and the exclusion of other phenotypes than schizophrenia in the current study. We followed our first findings up in a targeted analysis on pathways representing synaptic signaling in all major neurotransmitter systems, and demonstrated enrichment of schizophrenia SNPs in the dopaminergic synapse, long-term potentiation through the glutamatergic system and the cholinergic synapse, although the latter could not be confirmed using INRICH. We then showed that protein products of significant schizophrenia-associated genes in our enriched KEGG pathways show on average a higher rate of interactions with other proteins than expected by chance. Finally, we highlight that expression patterns of top genes from the enriched KEGG pathways do not substantially differ between brain regions.

Dysfunctional synaptic transmission impacts synaptic plasticity and brain development, mediated through long-term potentiation (LTP) and long-term depression (LTD) [28]. Although all five major neurotransmitter systems (dopamine, gamma-aminobutyric acid, glutamate, serotonin, and acetylcholine) have been implicated in schizophrenia, the extent to which each of them is involved had remained elusive [2,29]. Our results strongly support the involvement of the dopaminergic system, which has been extensively examined

in schizophrenia. Previous studies have reported increased dopamine synthesis and release, and increased dopamine receptor expression in schizophrenia [2,30]. *DRD2* genetic variants are also implicated in schizophrenia and several of its intermediate phenotypes [31,32]. We here confirm an accumulation of *DRD2* genetic variants in schizophrenia and in signaling cascades downstream of this receptor. Enrichment of the glutamate-induced LTP pathway was another finding that could be verified using all enrichment analysis tools. Mediation of LTP is, however, not limited to the glutamatergic system as post-synaptic signaling molecules such as the above mentioned CREB, $IP_3R$ and PKB mediate synaptic plasticity in other neurotransmitter systems (e.g. the dopaminergic system). Multiple lines of evidence link LTP to cognitive deficits in schizophrenia [33]. Cholinergic transmission may also be relevant to symptomatology of schizophrenia, especially in light of the high rates of nicotine abuse and a range of cognitive symptoms [34,35]. The implication of acetylcholine in schizophrenia is further supported by a landmark study investigating chromatin interactions between enhancer regions containing schizophrenia-associated loci and promoter regions of target genes [36]. A recent GWAS meta-analysis on schizophrenia, that extended the data used in this study with a British cohort, also points towards synaptic transmission [9]. In contrast to our study, it also identified significant association at two genes involved in GABA signaling (*GABBR2* and *SCL6A11*). Although these genes had assigned gene p-values below 0.05 in our analysis, we did not find an enrichment of the GABAergic synapse after multiple-testing correction. Possible explanations could be phenotypic differences between the 2014 schizophrenia GWAS and the recent meta-analysis (e.g. enrichment of relatively treatment-resistant patients in the latter), or the fact that the contribution of these genes to the enrichment of the GABAergic synapse pathway is attenuated by non-associated genes present in this gene set. In a broader perspective, however, our findings are in line with rare variant

studies in schizophrenia that reported mutations in genes coding for postsynaptic signaling components [37-40].

Our detailed analysis of downstream signaling cascades in all major neurotransmitter system gene sets revealed several of these cascades to be highly enriched for schizophrenia-associated variants: the phospholipase pathway, CREB signaling and the PKB/Akt signaling cascade. All of these cascades may be linked to schizophrenia by numerous lines of neurobiological evidence, as outlined below. First, the phospholipase pathway (particularly PLC) controls neuronal activity and thereby maintains synaptic functioning and development. Gene deletions in *PLC* are associated with schizophrenia and altered expression of *Plc* and schizophrenia-like behavior have been reported in *Plc* knock-out mice [41,42]. Second, signaling through CREB modulates synaptic plasticity. A recent study focusing on the cyclic adenosine monophosphate (cAMP)/PKA/CREB pathway shows a significant association of a SNP in this system with schizophrenia [43]. Additionally, ERK is part of the CREB signaling cascade and was also found to be enriched in our analyses. Impairment of signaling through ERK is hypothesized as a disease mechanism in schizophrenia [44,45]. Third, we found a significant enrichment of schizophrenia SNPs in postsynaptic protein kinase B (PKB or Akt). *AKT1* messenger RNA levels are higher in blood of schizophrenia patients compared to healthy controls and interactions between genetic variation in *AKT1* and cannabis use are associated with schizophrenia, possibly mediated through AKT signaling downstream of DRD2 [46,47]. Interestingly, phosphorylation of glycogen synthase kinase 3 beta (Gsk3β) by the antipsychotic aripiprazole is mediated by Akt [48]. Finally, we detected an accumulation of SNPs in protein phosphatase 1 (PP1) and protein phosphatase 2A (PP2A). PP2A is one of the mediators of sensorimotor gating, an intermediate phenotype for schizophrenia [49].

Conditional analysis correcting for shared genes between pathways resulted in a non-significant enrichment of all KEGG pathways, supporting the hypothesis that shared

signaling mechanisms across the implicated pathways are likely to be a strong underlying factor in schizophrenia in addition to neurotransmitter-specific pathways. Additionally, the finding of higher interaction rate of proteins coded by SNP-enriched genes in the dopaminergic, LTP and cholinergic pathways might highlight an important biological role of these genes in various biological processes, including synaptic signaling. Although this interaction is expected given that these proteins all operate in the same cellular compartment (the synapse), these genes might be interesting targets in light of a novel hypothesis recently outlined in a hallmark paper about core disease genes that are connected to many other gene networks [50]. Expression profiles of top genes from enriched KEGG pathways among brain tissues did not show major differences, similarly to brain tissues versus other tissues (Supplementary Figure S9). This could support the widespread role of these genes in biological processes. Seemingly higher relative transcript abundance of these genes in cerebellum and cerebellar hemisphere might be explained by the fact that these genes were selected from neuronal gene sets and that neuron density is higher in these tissues compared to the rest of the brain [51]. It has to be emphasized that the expression data of these genes was generated using healthy tissue.

Several limitations should be considered when interpreting our results. First, no significant SNP enrichment was found in other systems hypothesized to be dysregulated in schizophrenia, such as glutamatergic and GABAergic neurotransmission [2,52,53]. As our analyses are dependent on the power of GWAS, we cannot rule out the possibility that increased sample sizes in future studies may flag such systems, as illustrated by the identification of genes involved in GABA signaling in novel GWAS (doi.org/10.1101/068593). Second, we can only test for enrichment in gene sets and pathways that are annotated based on the knowledge currently available. Third, only protein-coding regions of the genome and up- and downstream regions in close proximity to genes were

considered in our analyses. Non-coding stretches of the genome account for a major part of disease heritability and transcription regulation [36]. As the current analyses do not allow us to probe non-coding regions, we cannot take into account the effects that such genomic areas may have on neurotransmitter systems. Future integration of expression quantitative trait locus (eQTL) data and genomic interactions in pathway analysis tools that may be under development has the potential to further deepen our understanding of the molecular mechanisms involved in the pathways underlying schizophrenia. Despite these limitations and slight technical differences between MAGMA, MAGENTA and INRICH [10,17-19], we found that our findings are fairly robust to analytic approach.

In conclusion, using complementary enrichment analysis approaches, we highlight downstream signaling cascades as the most likely part of the dopaminergic, cholinergic and LTP systems to have a role in schizophrenia. Our results open avenues for further research aimed at elucidating signaling pathways in schizophrenia, e.g. through more comprehensive integration with genotype, expression and protein-interaction data. Finally, our findings may aid the discovery of novel drug targets to hopefully reduce the burden imposed on quality of life in patients suffering from this disabling disorder.

**Acknowledgments**

The authors thank the Psychiatric Genomics Consortium for sharing summary-level genetic data on which this research is based.

The GTEx Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. Gene expression data were obtained from the GTEx Portal.

**Conflict of interest**

The authors declare no conflict of interest.

## Figure legends

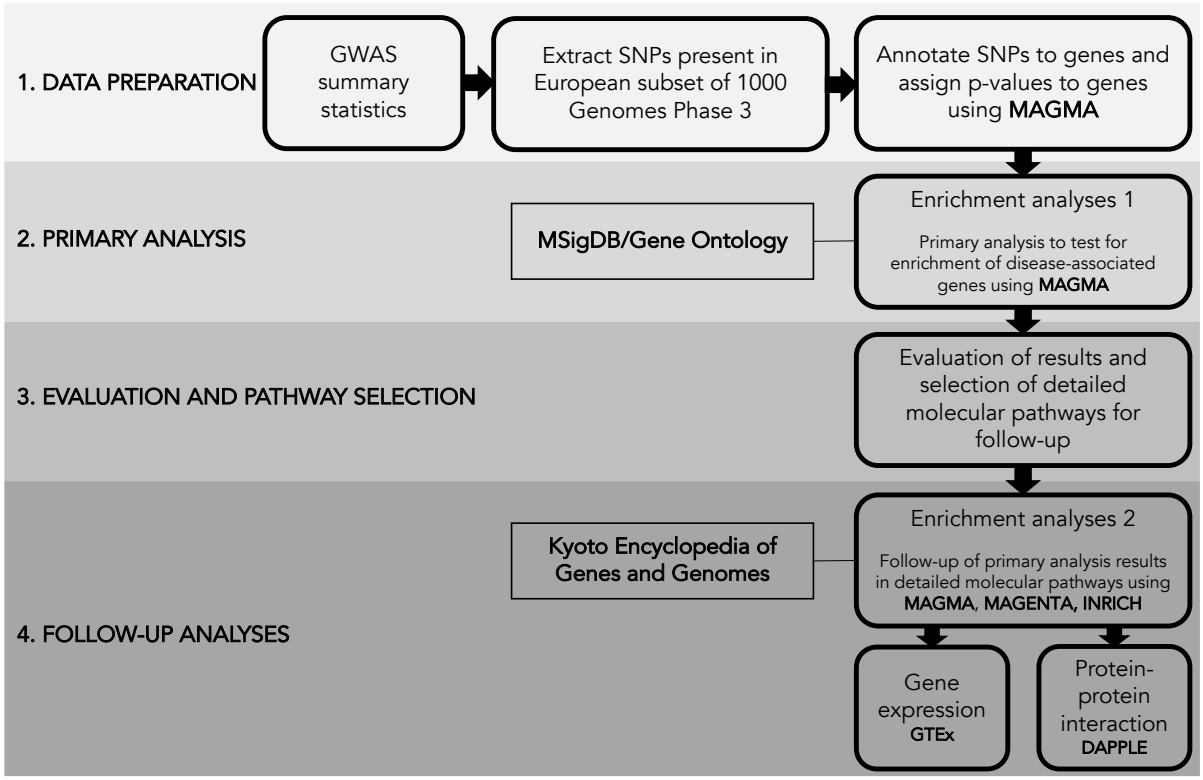


**Figure 1 Overview of pathway analysis pipeline.** Our analysis pipeline consisted of four stages: **(1)** Data preparation; **(2)** Primary gene set enrichment analysis on MSigDB gene ontology (GO) biological processes; **(3)** Evaluation of primary results and selection of detailed pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG); **(4)** Follow-up analyses on these detailed molecular pathways to further investigate involvement of biological processes found in the primary analysis.

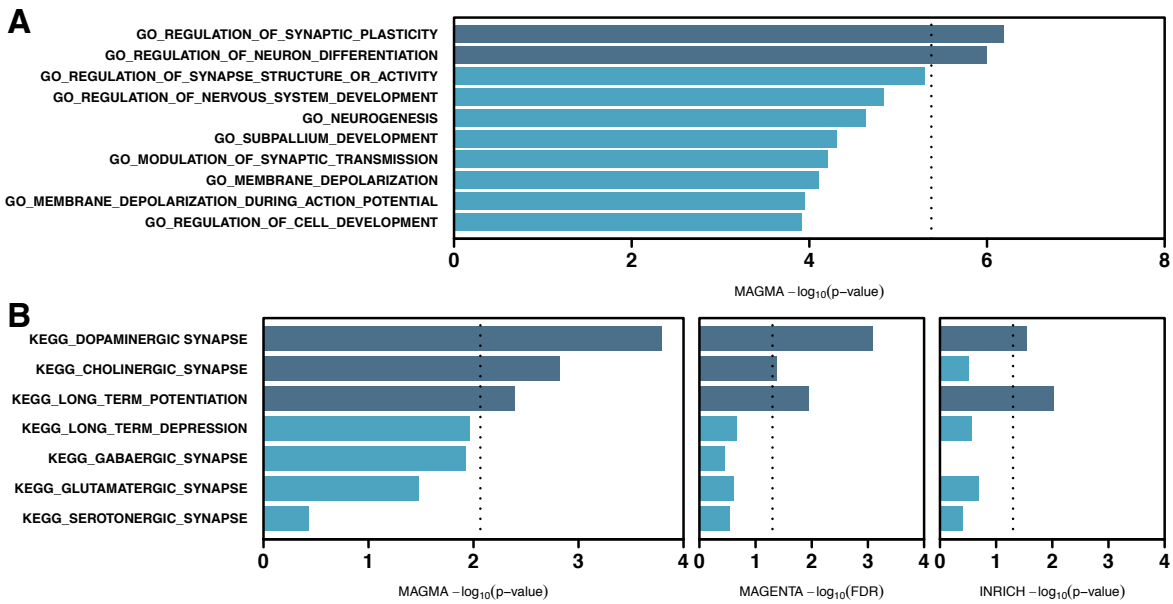


**Figure 2 Results of enrichment analysis on MSigDB biological processes and KEGG pathways.** Reported p-values or FDR are $-\log_{10}$ converted. Significance thresholds are different per analysis depending on the used method. Significant enrichment is indicated with dark blue. **(A)** Top ten enriched MSigDB gene sets (Significance threshold $p = 4.23\times10^{-6}$). **(B)** Enrichment in KEGG synaptic signaling pathways tested using MAGMA (left panel, significance threshold $p = 9.0\times10^{-3}$), MAGENTA (middle panel, significance threshold FDR = 0.05) and INRICH (right panel, significance threshold p = 0.05).

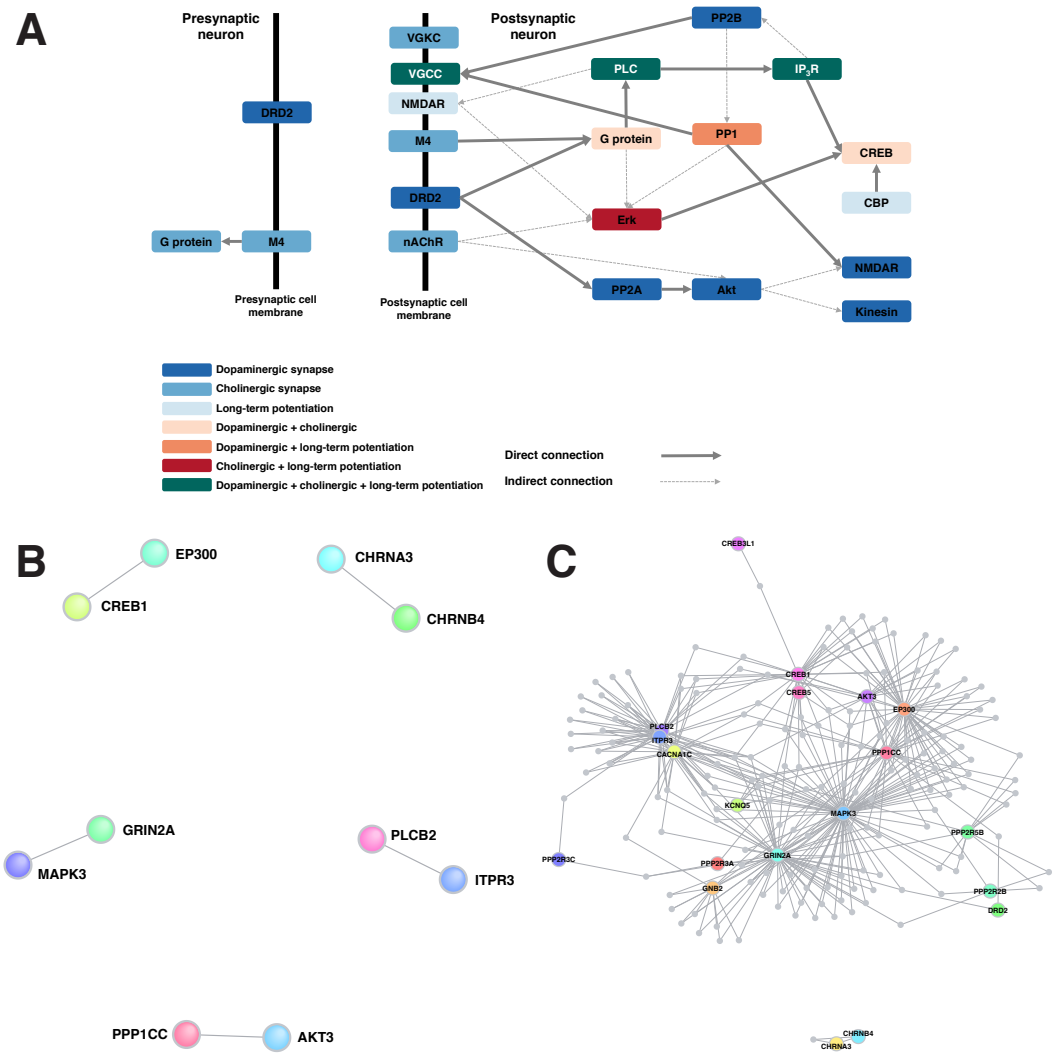


**Figure 3 Enrichment and protein-protein interactions in KEGG synaptic signaling pathways. (A)** Significant pathway components (gene p-value < $1.5\times10^{-4}$) for dopaminergic synapse, cholinergic synapse and long-term potentiation. Colors indicate specificity to one or more pathways. Arrows indicate connections between signaling components (solid, direct interaction; dashed, indirect interaction), based on KEGG.[24] **(B)** Direct (interaction p = $1.4\times10^{-2}$) and **(C)** indirect (interaction p = $4.0\times10^{-3}$) protein-protein interaction networks generated with DAPPLE [15] for 22 top genes from the dopaminergic synapse, long-term potentiation and cholinergic synapse gene sets.

Abbreviations: DRD2, dopamine receptor D2; M4, muscarinic acetylcholine receptor M4; G protein, guanine nucleotide-binding protein; VGKC, voltage-gated potassium channel; VGCC, voltage-gated calcium channel; NMDAR, N-methyl-D-aspartate receptor; nAChR, nicotinic acetylcholine receptor; PP2A, protein phosphatase 2A; Akt, protein kinase B; Erk, Extracelluar signal-regulated kinase; CREB, cAMP response element binding protein; CBP, CREB-binding protein; PP1, protein-phosphatase 1; PLC, phospholipase C; $IP_3R$, inositol trisphosphate receptor; PP2B, protein-phosphatase 2B.

## References


1 Celada P, Bortolozzi A, Artigas F. Serotonin 5-HT1A receptors as targets for agents to treat psychiatric disorders: rationale and current status of research. *CNS drugs* 2013; **27**: 703–716.

2 Kahn RS, Sommer IE, Murray RM, Meyer-Lindenberg A, Weinberger DR, Cannon TD *et al.* Schizophrenia. *Nature Reviews Disease Primers* 2015; : 15067.

3 Kane JM, Correll CU. Past and present progress in the pharmacologic treatment of schizophrenia. *The Journal of clinical psychiatry* 2010; **71**: 1115–1124.

4 Köster L-S, Carbon M, Correll CU. Emerging drugs for schizophrenia: an update. *Expert opinion on emerging drugs* 2014; **19**: 511–531.

5 Karow A, Wittmann L, Schöttle D, Schäfer I, Lambert M. The assessment of quality of life in clinical practice in patients with schizophrenia. *Dialogues Clin Neurosci* 2014; **16**: 185–195.

6 Lichtenstein P, Björk C, Hultman CM, Scolnick E, Sklar P, Sullivan PF. Recurrence risks for schizophrenia in a Swedish national cohort. *Psychological medicine* 2006; **36**: 1417–1425.

7 Wang K, Gaitsch H, Poon H, Cox NJ, Rzhetsky A. Classification of common human diseases derived from shared genetic and environmental determinants. *Nat Genet* 2017; **6**: 111.

8 Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421–427.

9 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 maintained by background selection. *bioRxiv* 2016; : 068593.

10 de Leeuw CA, Neale BM, Heskes T, Posthuma D. The statistical properties of gene-set analysis. *Nat Rev Genet* 2016; **17**: 353–364.

11 Lips ES, Cornelisse LN, Toonen RF, Min JL, Hultman CM, Holmans PA *et al.* Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. *Molecular psychiatry* 2012; **17**: 996–1006.

12 Duncan LE, Holmans PA, Lee PH, O'Dushlaine CT, Kirby AW, Smoller JW *et al.* Pathway analyses implicate glial cells in schizophrenia. *PloS one* 2014; **9**: e89441.

13 Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nature neuroscience* 2015; **18**: 199–209.

14 Hertzberg L, Katsel P, Roussos P, Haroutunian V, Domany E. Integration of gene expression and GWAS results supports involvement of calcium signaling in Schizophrenia. *Schizophrenia research* 2015; **164**: 92–99.

15 Rossin EJ, Lage K, Raychaudhuri S, Xavier RJ, Tatar D, Benita Y *et al.* Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet* 2011; **7**: e1001273.

16 Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S *et al.* The Genotype-Tissue Expression (GTEx) project. *Nature genetics* 2013; **45**: 580–585.

17 de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput Biol* 2015; **11**: e1004219.

18 Segrè AV, Groop L, Mootha VK, Daly MJ, Altshuler D, DIAGRAM Consortium *et al.* Common Inherited Variation in Mitochondrial Genes Is Not Enriched for Associations with Type 2 Diabetes or Related Glycemic Traits. *PLoS Genet* 2010; **6**: e1001058.

19 Lee PH, O'Dushlaine C, Thomas B, Purcell SM. INRICH: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics (Oxford, England)* 2012; **28**: 1797–1799.

20 The 1000 Genomes Project Consortium C. A global reference for human genetic variation. *Nature* 2015; **526**: 68–74.

21 Veyrieras JB, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M *et al.* High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS genetics* 2008; **4**: e1000214.

22 Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic acids research* 2015; **43**: D1049–56.

23 Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics (Oxford, England)* 2011; **27**: 1739–1740.

24 Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* 2000; **28**: 27–30.

25 Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* 2013; **29**: 1830–1831.

26 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 2015; **4**: 7.

27 Warnes GR, Ben Bolker, Bonebakker L, Gentleman R, Huber W, Liaw A *et al.* gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1. https://CRAN.R-project.org/package=gplots (accessed 6 Nov2017).

28 Bernardinelli Y, Nikonenko I, Muller D. Structural plasticity: mechanisms and contribution to developmental psychiatric disorders. *Frontiers in Neuroanatomy* 2014; **8**. doi:10.3389/fnana.2014.00123.

29 Pocklington AJ, O'Donovan M, Owen MJ. The synapse in schizophrenia. *Eur J Neurosci* 2014; **39**: 1059–1067.

30 Kaalund SS, Newburn EN, Ye T, Tao R, Li C, Deep-Soboslay A *et al.* Contrasting changes in DRD1 and DRD2 splice variant expression in schizophrenia and affective disorders, and associations with SNPs in postmortem brain. *Molecular psychiatry* 2014; **19**: 1258–1266.

31 Luykx JJ, Broersen JL, de Leeuw M. The DRD2 rs1076560 polymorphism and schizophrenia-related intermediate phenotypes: A systematic review and meta-analysis. *Neurosci Biobehav Rev* 2017; **74, Part A**: 214–224.

32 Vink M, de Leeuw M, Luykx JJ, van Eijk KR, van den Munkhof HE, van Buuren M *et al.* DRD2 Schizophrenia-Risk Allele Is Associated With Impaired Striatal Functioning in Unaffected Siblings of Schizophrenia Patients. *Schizophrenia bulletin* 2016; **42**: 843–850.

33 Salavati B, Rajji TK, Price R, Sun Y, Graff-Guerrero A, Daskalakis ZJ. Imaging-based neurochemistry in schizophrenia: a systematic review and implications for dysfunctional long-term potentiation. *Schizophrenia bulletin* 2015; **41**: 44–56.

34 Carruthers SP, Gurvich CT, Rossell SL. The muscarinic system, cognition and schizophrenia. *Neurosci Biobehav Rev* 2015; **55**: 393–402.

35 Parikh V, Kutlu MG, Gould TJ. nAChR dysfunction as a common substrate for schizophrenia and comorbid nicotine addiction: Current trends and perspectives. *Schizophrenia research* 2016; **171**: 1–15.

36 Won H, la Torre-Ubieta de L, Stein JL, Parikshak NN, Huang J, Opland CK *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* 2016; **538**: 523–527.

37 Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D *et al.* De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry* 2012; **17**: 142–153.

38 CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nature genetics* 2017; **49**: 27–35.

39 Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* 2014; **506**: 179–184.

40 Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 2014; **506**: 185–190.

41 Koh HY. Phospholipase C-beta1 and schizophrenia-related behaviors. *Advances in biological regulation* 2013; **53**: 242–248.

42 Vasco Lo VR, Cardinale G, Polonia P. Deletion of PLCB1 gene in schizophrenia-affected patients. *Journal of cellular and molecular medicine* 2012; **16**: 844–851.

43 Forero DA, Herteleer L, De Zutter S, Norrback KF, Nilsson LG, Adolfsson R *et al.* A

network of synaptic genes associated with schizophrenia and bipolar disorder. *Schizophrenia research* 2016; **172**: 68–74.

44 Kyosseva SV. The role of the extracellular signal-regulated kinase pathway in cerebellar abnormalities in schizophrenia. *Cerebellum* 2004; **3**: 94–99.

45 Yuan P, Zhou R, Wang Y, Li X, Li J, Chen G *et al.* Altered levels of extracellular signal-regulated kinase signaling proteins in postmortem frontal cortex of individuals with mood disorders and schizophrenia. *J Affect Disord* 2010; **124**: 164–169.

46 Liu L, Luo Y, Zhang G, Jin C, Zhou Z, Cheng Z *et al.* The mRNA expression of DRD2, PI3KCB, and AKT1 in the blood of acute schizophrenia patients. *Psychiatry research* 2016; **243**: 397–402.

47 van Winkel R. Family-based analysis of genetic variation underlying psychosis-inducing effects of cannabis: sibling analysis and proband follow-up. *Archives of general psychiatry* 2011; **68**: 148–157.

48 Pan B, Chen J, Lian J, Huang X-F, Deng C. Unique Effects of Acute Aripiprazole Treatment on the Dopamine D2 Receptor Downstream cAMP-PKA and Akt-GSK3β Signalling Pathways in Rats. *PLoS One* 2015; **10**: e0132722.

49 Kapfhamer D, Berger KH, Hopf FW, Seif T, Kharazia V, Bonci A *et al.* Protein Phosphatase 2a and glycogen synthase kinase 3 signaling modulate prepulse inhibition of the acoustic startle response by altering cortical M-Type potassium channel activity. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2010; **30**: 8830–8840.

50 Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 2017; **169**: 1177–1186.

51 Azevedo FAC, Carvalho LRB, Grinberg LT, Farfel JM, Ferretti REL, Leite REP *et al.* Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J Comp Neurol* 2009; **513**: 532–541.

52 Yin DM, Chen YJ, Sathyamurthy A, Xiong WC, Mei L. Synaptic dysfunction in schizophrenia. *Advances in experimental medicine and biology* 2012; **970**: 493–516.

53 Schmidt MJ, Mirnics K. Neurodevelopment, GABA system dysfunction, and schizophrenia. *Neuropsychopharmacology* 2015; **40**: 190–206.