



# The Stanley Neuropathology Consortium Integrative Database (SNCID) for Psychiatric Disorders

Sanghyeon Kim<sup>1</sup> · Maree J. Webster<sup>1</sup>

Received: 20 July 2018 / Accepted: 17 August 2018 / Published online: 27 November 2018  
© Shanghai Institutes for Biological Sciences, CAS and Springer Nature Singapore Pte Ltd. 2018

The Stanley Medical Research Institute (SMRI) is a non-profit organization with a primary mission to fund research on the cause and treatment of severe mental illnesses. The SMRI also supports a brain bank as part of the mission to facilitate research into mental illness. The SMRI brain bank distributes postmortem samples from individuals with serious mental illness, free of charge, to scientists around the world. The SMRI brain bank is recognized for the unique way it is set up, organized, and administered. Cohorts of demographically-matched groups of patients with schizophrenia, bipolar disorder (BP), or major depression (DEP) and unaffected controls are organized such that all researchers applying for tissue received samples from the same cohort. The SMRI was the first to include multiple diagnostic categories in the cohorts as well as the first to include a large number (N) of cases in each group. The Stanley Neuropathology Consortium (SNC) was the first cohort established and contains 60 brains (15 in each of 4 groups); it was the largest cohort of subjects with major mental illness that was available when the cohort was established in 1997. The Array Collection (AC) was a subsequent cohort with an N of 35 in each of 3 groups and was established in 2002. Pathologists from the offices of designated medical examiners in the USA were trained in standardized collection techniques. All medical and

psychiatric records were obtained and reviewed by two senior psychiatrists. A psychiatrist also contacted one or more family members by telephone to clarify and verify before making the final DSM-IV diagnosis. For the normal controls, a psychiatrist conducted a structured telephone interview with first-degree family members to obtain all pertinent psychiatric and medical history as described in Torrey *et al.*, 2000 [1]. A pathologist performed a standard neuropathological examination for every case. While cases were relatively young (average age 45 years), all brains were screened for the presence of cardiovascular disease, hemorrhage, trauma, tumors, or other pathology and confirmed by the examination of appropriate sections from the suspect area. Cases were also screened for Alzheimer's disease, Parkinson's disease, ethanol-induced changes, and anoxic/hypoxic-related alterations and also for RNA integrity. Only brains that were free from pathology and that had intact RNA were included in the final research cohorts. Controls have no history of any neurologic or psychiatric disorder. Within each research cohort the diagnostic groups and controls were matched for age, sex, race, postmortem interval (PMI), pH, and side of brain frozen.

The SMRI brain bank was involved in the initial efforts to identify and develop methodology for optimizing anatomical and tissue integrity for postmortem tissue necessary for the rapidly-evolving molecular biology techniques [1, 2]. The SMRI maximized the use of the brains by distributing aliquots of DNA, RNA, and protein as well as slide-based sections. In addition to the post-mortem brain tissue, liver, spleen, serum, and cerebrospinal fluid (CSF) were also collected from each subject. CSF can be important for the cross-validation of abnormal protein expression in the brain, and abnormal CSF proteins may be useful in the discovery of potential serum biomarkers [3].

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s12264-018-0314-7>) contains supplementary material, which is available to authorized users.

✉ Maree J. Webster  
websterm@stanleyresearch.org

<sup>1</sup> Stanley Brain Research Laboratory, Stanley Medical Research Institute, 9800 Medical Center Drive, Rockville, MD 20850, USA

The SMRI was at the forefront for requiring data-sharing by distributing all tissue coded and researchers only obtained the code, demographic, and clinical information for each case after they sent the data back to the SMRI. The demographic and clinical information includes a wide range of variables such as age, sex, race, PMI, cause of death, education, marital status, age at onset, family history, total duration of hospitalization, estimate of lifetime antipsychotic medication (fluphenazine milligram equivalents), and history of drug and alcohol use, including smoking. To date, the SMRI has obtained more than 6000 individual neuropathology datasets, as well as multiple microarray gene expression datasets, microRNA array data, single nucleotide polymorphism (SNP) array data, proteomic data, epigenetic data, and RNA-seq data from multiple brain regions, all from the same set of brains. Consequently, the SMRI has the most well-characterized set of brains ever established and has developed a web-based interactive tool to integrate all this data. The integrative database that houses all this data has been described in detail [4, 5], however it is constantly expanding and evolving. Thus, this brief report provides an update of the data and the tools recently added to the functions of the database. Using the Stanley Neuropathology Consortium Integrative Database (SNCID) data we investigate whether any of the SNPs recently associated with the risk for schizophrenia can be identified in the SNP array data and then determine if they are associated with any of the neuropathological traits that have been measured in these same brains.

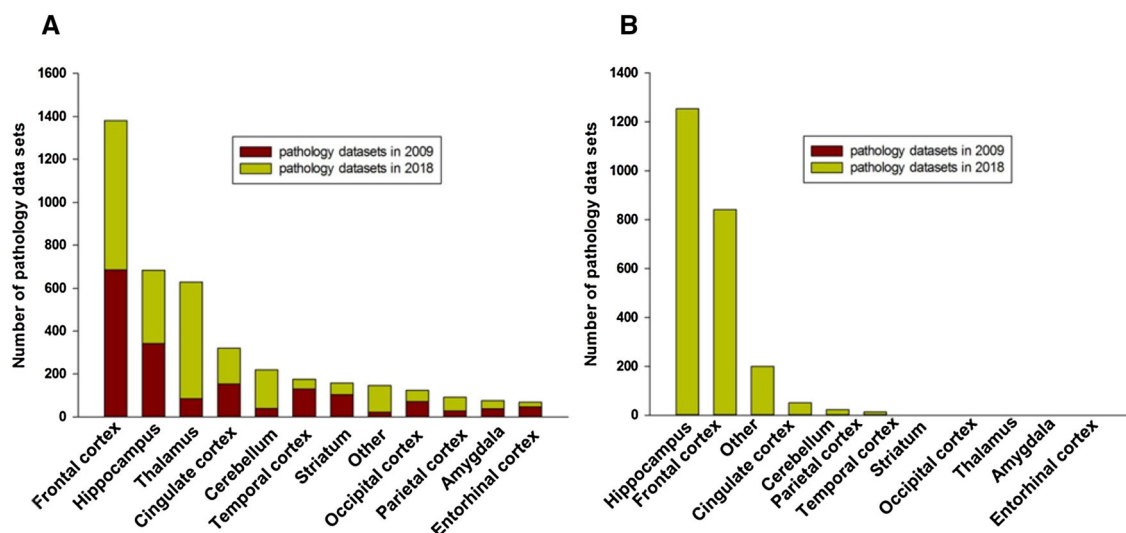
A total of 4712 individual neuropathology datasets have been added to the SNCID [4] since the integrative database was first made public in March, 2009. As of June 2018, the SNCID contains 4078 neuropathology markers measured in 12 different brain regions of the 60 cases in the SNC. This is a 136% increase compared to when the database was first made public (Fig. 1A). There are also 2383 markers measured in 7 brain regions in the 105 cases of the AC. These datasets are particularly useful for validation studies (Fig. 1B). These neuropathology datasets include traditional cytoarchitectural studies that measure the size and density of neurons and glia, as well as protein studies (western blots, ELISAs, and immunohistochemistry), RNA studies (qPCR and *in-situ* hybridization) and biochemical studies (receptor binding). In addition, RNA-Seq data from the hippocampus of the SNC have been integrated into the correlation module of the SNCID and the raw RNA-Seq data from seven brain regions of three Stanley brain collections are available to download.

A variety of statistical analysis tools such as variance analysis, correlation analysis, and functional annotation tools are integrated into the database, but new tools continue to be developed and added to the database [4, 6].

The gene co-expression network analysis algorithm is widely used to identify modules of co-expressed genes that show a similar expression pattern across samples [7, 8]. A quantitative comparison of two co-expression modules could enable researchers to compute the similarity between two networks and thereby lead to the possible identification of the mechanisms underlying complex traits and diseases. However, the availability of such statistical algorithms to analyze co-expression modules is limited. We therefore developed a new web tool that enables researchers to compare two co-expression modules [8]. This feature has also been integrated into the SNCID (Fig. S1).

A SNP-based analysis tool has also been developed to perform association analyses between the SNP genotyping data and the neuropathology data or gene expression data and has been added to the SNCID. The genome-wide SNP genotyping data were generated by Dr. Chunyu Liu and colleagues using the Human SNP Array 5.0 chips (Affymetrix). Quality-control of the SNP data was performed as previously described [9]. For the tool, the SNPs with a call rate of < 90% and minor allele frequency < 5% were filtered out. A total of 308,560 SNPs passed this filter and were included in the analysis tool. Our previous study identified four ethnic outliers from the SNC and three outliers from the AC [9], and so these cases were excluded from the SNP-based analysis tool. The new SNP-based analysis tool enables users to explore whether a particular SNP of interest is significantly associated with a neuropathology marker (Fig. S2). The tool accepts a gene symbol, an Entrez ID, dbSNP ID, or chromosomal location as an input (Fig. S2A). Using correlation analysis, the tool yields a list of markers in the SNCID that are significantly associated with that SNP of interest (e.g. Fig. S2B).

A previous large-scale genome-wide association study (GWAS) identified 108 risk loci for schizophrenia [10], however the causal genetic variations and the underlying mechanisms leading to abnormal brain function remain to be determined. Most GWAS SNPs associated with schizophrenia risk are located in noncoding regions of the chromosome, and thus probably mediate their influence through gene regulation in the brain rather than causing changes in protein function [10, 11]. The SNCID includes several different types of quantitative data for various markers such as RNA, protein, metabolites, and the density of cells expressing specific markers that are directly related to the schizophrenia risk SNPs. The SNCID tool allows users to identify the neuropathology markers that are correlated with the GWAS SNPs and therefore provides a unique opportunity to determine if the SNP could actually cause a change in the levels of the corresponding RNA or protein. However, among the 108 risk loci for schizophrenia, only 10 could be identified on the SNP array in the SNCID. We identified 18 neuropathology datasets that



**Fig. 1** Total numbers of neuropathology markers in the Stanley Neuropathology Consortium Integrative Database. The number of markers from each brain region from the Stanley Neuropathology Consortium (A) and the Array Collection (B).

measured the RNA levels of the genes that are located in the vicinity of these 10 SNPs. We used the SNP-based analysis tool to determine if any of the 10 SNPs was correlated with these datasets or with any of the other unrelated neuropathology datasets ( $P < 0.001$  was considered significant).

Using the SNP-based analysis tool, we determined whether the 10 schizophrenia risk SNPs that are currently integrated into the database were significantly correlated with the neuropathology markers (Table 1). While we found 9 of the 10 SNPs to be correlated with at least 2 neuropathology markers, none showed a significant correlation with the RNA levels of their nearby genes (Table 1). For example, the SNP rs2514218 is located near the dopamine receptor gene *DRD2* (a schizophrenia risk locus) in chromosome 11 and the SNCID has 12 and 2 gene expression datasets that measure RNA expression for *DRD2* in different brain regions of the SNC and the AC, respectively. However none of these datasets were significantly correlated with the *DRD2* SNP. Similarly, rs9636017 is located in the intron of the *TCF4* gene. There are three *TCF4* RNA expression datasets that were measured independently in the frontal cortex of the AC and none of them were significantly associated with the *TCF4* SNP.

However, there was a significant correlation between the SNP rs11210892, located in the vicinity of the *KDM4A* gene in chromosome 1, and a possible functionally-related marker. *KDM4A* encodes the lysine-specific demethylase 4A enzyme and plays an important role in neuronal differentiation and the survival of neural stem cells [12]. We found a significant correlation between this SNP and the density of NeuN-containing neurons in the dorsal raphe

( $P < 0.0001$ ) (Table 1), indicating that the risk allele may affect the neuronal density of the dorsal raphe by altering expression of the *KDM4A* gene.

There are also  $> 352$  genes located within or near the 108 risk loci for schizophrenia [10]. Of these 352 genes, we found 13 had corresponding RNA or protein (neuropathology markers) measured in the SNC, 22 were measured in the AC, and all were available in the SNCID. We performed a non-parametric variance analysis using the SNCID to determine if there is a statistical difference between the diagnostic groups and unaffected controls in the levels of these markers ( $P < 0.05$  considered significant). Using the group comparison tool we analyzed the various markers that are directly associated with these 352 genes and that have been measured in the SNCID. The results were visualized using heat maps (Fig. 2). The *DRD2* RNA levels in frontal cortex of the SNC were significantly decreased in schizophrenia and major depression, whereas the levels in the hippocampus were not significantly altered (Fig. 2A). This result was replicated in the frontal cortex of the AC (Fig. 2B). Dopamine is involved in the pathophysiology of schizophrenia [13] and *DRD2* is the main target of antipsychotic drugs [14] so the change in *DRD2* RNA expression in the frontal cortex in schizophrenia supports the dopamine hypothesis. However, as mentioned above, we found no significant association between *DRD2* expression in the frontal cortex and the rs2514218 SNP that is located near the *DRD2* gene. Thus there may be other risk variants that affect expression of the gene or there may be unidentified risk variants that regulate the expression of *DRD2* in the frontal cortex. However, it is also possible that the results are underpowered because of the small sample size in the SNCID.

**Table 1** Genome-wide risk loci for schizophrenia and the neuropathology markers that were significantly correlated with the loci.

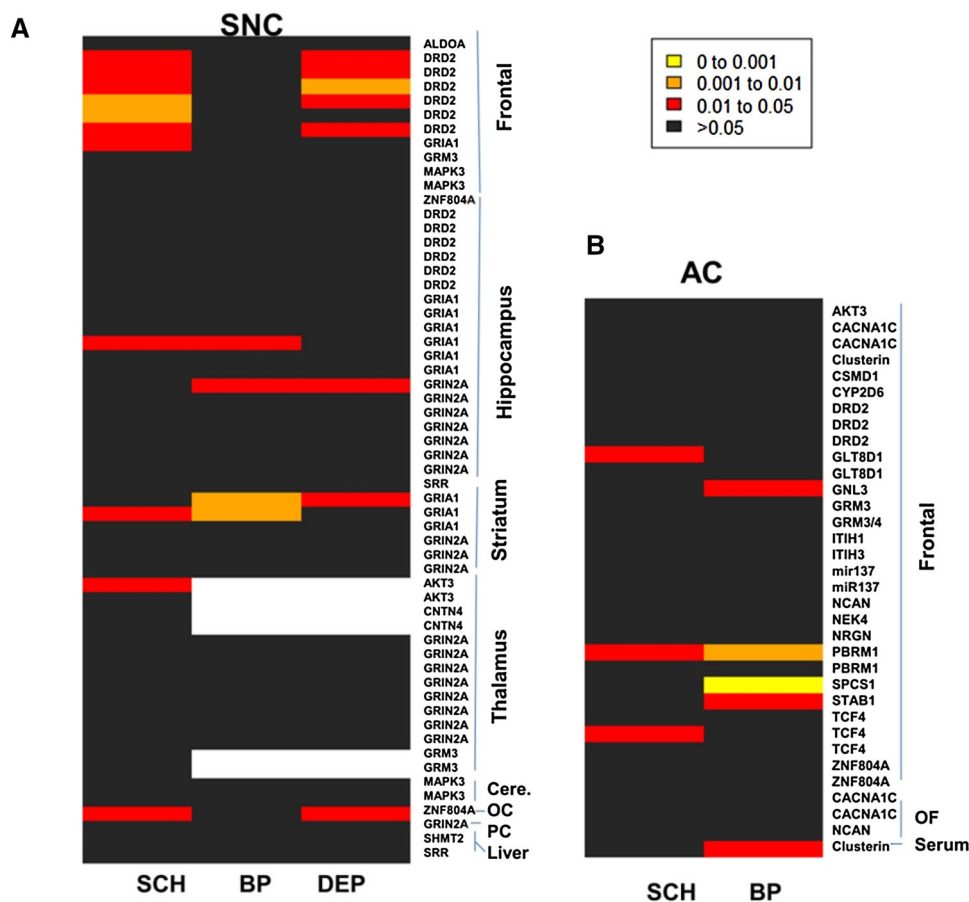
dbSNP_ID	Position	Genes	Tissue collection	Marker	Brain/Region	Marker type	cor <sup>a</sup>	P-value
rs10803138	1:243555219	<i>AKT3, SDCCAG8</i>	SNC	Inositol monophosphatase	Frontal cortex	Other	0.52	0.0001
rs11210892	1:44100084	<i>KDM4A, PTPRF</i>	SNC	N-Acetylputrescine	Hippocampus	Other	−0.53	0.0009
			SNC	NeuN density (cells/mm <sup>2</sup> )	Dorsal raphe	Cell	−0.66	0.0000
			SNC	NeuN density (cells/mm <sup>2</sup> )	Dorsal raphe	Cell	−0.61	0.0001
			SNC	PSA-NCAM	Hippocampus	Protein	0.58	0.0005
			AC	GRO-alpha	Frontal cortex	Protein	0.35	0.0008
			AC	Amphiregulin	serum	Protein	0.37	0.0005
			SNC	MBP	Frontal cortex	RNA	0.62	0.0005
rs6704641	2:200164252	<i>SATB2</i>	n/a	n/a	n/a	n/a	n/a	n/a
			rs6704768	2:233592501	<i>C2orf82 EFHD1 GIGYF2 KCNJ13 NGEF</i>	SNC	NOX2 (microglia)	Nucleus accumbens
rs1106568	4:176861301	<i>GPM6A</i>	SNC	GRIK5	Hippocampus	RNA	0.47	0.0006
			SNC	SLC16A1*	Thalamus	RNA	−0.60	0.0008
			SNC	SLC16A1 M NM**	Thalamus	RNA	−0.60	0.0008
			SNC	GFAP (astrocytes)	Orbitofrontal cortex	Cell	0.65	0.0006
rs215411	4:23423603	<i>MIR548AJ2</i>	SNC	GFAP (astrocytes)	Orbitofrontal cortex	Cell	0.64	0.0007
			SNC	NTRK2	Frontal cortex	RNA	−0.47	0.0004
			SNC	SOX11	Thalamus	RNA	−0.46	0.0009
			SNC	[3H]MK801	Frontal cortex	Other	−0.47	0.0006
			SNC	CDP-choline	Hippocampus	Other	0.47	0.0005
			SNC	Succinic acid	Hippocampus	Other	0.44	0.0010
			SNC	MAPK3	Frontal cortex	Protein	−0.47	0.0004
			SNC	Normalized AVP-NP11 intensity	Pituitary	Protein	0.46	0.0009
			SNC	FYN	Thalamus	RNA	0.47	0.0004
			SNC	PIP4K2A*	Thalamus	RNA	0.51	0.0010
rs4388249	5:109036066	<i>MAN2A1</i>	SNC	PIP4K2A M NM**	Thalamus	RNA	0.51	0.0010
			SNC	COMPLEX 1 activity	Frontal cortex	Other	0.48	0.0002
			SNC	NMDA receptor stimulation/unstimulated_NR1	Cingulate cortex	Other	0.45	0.0007
			SNC	Trigonelline	Hippocampus	Other	−0.70	0.0006
			SNC	PHF1	Frontal cortex	Protein	−0.47	0.0004
			AC	Alpha-1 antitrypsin	Frontal cortex	Protein	−0.34	0.0007
			SNC	GRIA3	Frontal cortex	RNA	0.45	0.0007
			SNC	MKP1	Striatum	RNA	−0.44	0.0009
			SNC	24 kDa NDUFV1	Cerebellum	RNA	−0.46	0.0005
			AC	TNFRSF6	Orbitofrontal	RNA	−0.36	0.0004

**Table 1** continued

dbSNP_ID	Position	Genes	Tissue collection	Marker	Brain/Region	Marker type	cor <sup>a</sup>	P-value
rs11027857	11:24403620	<i>LUZP2</i>	SNC	Homocarnosine	Hippocampus	Other	-0.48	0.0003
rs2514218	11:113392994	<i>DRD2</i>	AC	Alpha-fetoprotein	Serum	Protein	-0.36	0.0006
rs9636107	18:53200117	<i>TCF4</i>	SNC	4-Guanidinobutyric acid	Frontal cortex	Other	0.46	0.0005
			SNC	4-Guanidinobutyric acid	Hippocampus	Other	0.52	0.0001
			SNC	Azelaic acid	Hippocampus	Other	0.45	0.0010
			SNC	ADAR1	Cingulate cortex	RNA	0.44	0.0009
			SNC	CPLX1	Hippocampus	RNA	0.45	0.0008

<sup>a</sup>Correlation coefficients between modules associated with schizophrenia and cellular markers. Correlation analysis was performed using the SNCID tool. P-values < 0.001 were considered significant. Correlation coefficients represent only significant correlation between the module and cellular markers. \* Data normalized to geometric mean of 3 housekeeping genes; \*\* Data normalized to mean of control group. ns, not significant. SNC; Stanley Neuropathology Consortium, AC; Array Collection.

**Fig. 2** Heat maps of the neuropathology markers corresponding to genome-wide risk loci for schizophrenia. The color code represents P-values in non-parametric variance tests of the neuropathology markers in the brains of the Stanley Neuropathology Consortium (A) and the Array Collection (B). SCH, schizophrenia; BP, bipolar disorder; DEP, major depression; cere, cerebellum; OC, occipital cortex; PC, parietal cortex; OF, orbitofrontal cortex.



The previous GWASs also identified SNPs significantly associated with schizophrenia and located within or in the vicinity of several glutamate receptor genes: *GRIN2A*, *GRIA1*, and *GRM3*. *GRIN2A* RNA expression has been

measured in four different brain regions of the SNC and is available in the SNCID (Fig. 2A). Of the 19 datasets, we found one qPCR dataset for *GRIN2A* expression that was significantly lower in BP and DEP than in controls



(Fig. 2A). *GRIA1* RNA expression was measured in three brain regions of the SNC and is also available in the database (Fig. 2A). In frontal cortex, *GRIA1* expression was significantly decreased in schizophrenia, whereas in the striatum it was decreased in all three disorder groups (Fig. 2A). However, in the thalamus there was no difference in *GRIA1* RNA expression in any disorder group. There was also no significant difference in *GRM3* RNA expression in frontal cortex of the SNC or AC in any disorder group (Fig. 2A, B). The glutamatergic system and dysfunction of N-methyl-D-aspartate (NMDA) receptors have consistently been implicated in the pathophysiology of schizophrenia [15]. Our results suggest that  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid and NMDA receptors are more likely to be involved in the pathophysiology of schizophrenia than metabotropic glutamate receptors. Unfortunately, none of these glutamate-related SNPs were represented in our SNP array data and therefore an association analysis between SNPs and related RNA levels could not be performed.

The SNCID enables researchers to efficiently reanalyze numerous neuropathology datasets and to test hypotheses without redoing experiments. However, the results from these analyses, using tools in the database, should be interpreted cautiously. The sample size for the neuropathology datasets of the SNC is relatively small ( $N = 60$ ) so the results may not be sufficiently powered to detect differences in the neuropathology markers with a small effect size between diagnostic groups. Thus, we have integrated the additional datasets from the AC ( $N = 105$ ) in order to replicate the results from the SNC data analyses. Replicating the results in the AC data may increase the reliability of the results derived from SNC data. Furthermore, increasing sample size by pooling data from both tissue collections can reduce type II errors. The SNCID provides several statistical analysis tools such as an omnibus ANOVA and simple correlation analysis. However, neuropathology markers are often confounded by demographic and clinical variables. Therefore, we strongly recommended that users further investigate any interesting finding by downloading the raw datasets from the repository and examining them with more sophisticated statistical models. We believe this integrative database will give researchers a unique opportunity to explore the abnormal neuropathological markers that occur in the major psychiatric disorders and will provide the data and tools necessary to explore the genes and biological processes associated with those abnormal markers.

**Acknowledgements** We thank all the investigators who generated the data in the SNCID (<http://sncid.stanleyresearch.org>). We also thank Jonathan Cohen for technical support.

**Conflicts of interest** The authors declare that they have no conflict of interest.

## References

1. Torrey EF, Webster M, Knable M, Johnston N, Yolken RH. The stanley foundation brain collection and neuropathology consortium. *Schizophr Res* 2000, 44: 151–155.
2. Webster MJ. Tissue preparation and banking. *Prog Brain Res* 2006, 158: 3–14.
3. Yan X, Mai L, Lin C, He W, Yin G, Yu J, *et al.* CSF-based analysis for identification of potential serum biomarkers of neural tube defects. *Neurosci Bull* 2017, 33: 436–444.
4. Kim S, Webster MJ. The stanley neuropathology consortium integrative database: a novel, web-based tool for exploring neuropathological markers in psychiatric disorders and the biological processes associated with abnormalities of those markers. *Neuropsychopharmacology* 2010, 35: 473–482.
5. Huitinga I, Webster MJ. *Brain Banking*. Elsevier, 2018.
6. Kim S, Webster MJ. Integrative genome-wide association analysis of cytoarchitectural abnormalities in the prefrontal cortex of psychiatric disorders. *Mol Psychiatry* 2011, 16: 452–461.
7. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008, 9: 559.
8. Jo Y, Kim S, Lee D. Identification of common coexpression modules based on quantitative network comparison. *BMC Bioinformatics* 2018, 19: 213.
9. Kim S, Cho H, Lee D, Webster MJ. Association between SNPs and gene expression in multiple regions of the human brain. *Transl Psychiatry* 2012, 2: e113.
10. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014, 511: 421–427.
11. Harrison PJ. Recent genetic findings in schizophrenia and their therapeutic relevance. *J Psychopharmacol* 2015, 29: 85–96.
12. Cascante A, Klum S, Biswas M, Antolin-Fontes B, Barnabe-Heider F, Hermanson O. Gene-specific methylation control of H3K9 and H3K36 on neurotrophic BDNF versus astroglial GFAP genes by KDM4A/C regulates neural stem cell differentiation. *J Mol Biol* 2014, 426: 3467–3477.
13. Howes OD, Kapur S. The dopamine hypothesis of schizophrenia: version III—the final common pathway. *Schizophr Bull* 2009, 35: 549–562.
14. Seeman P. Targeting the dopamine D2 receptor in schizophrenia. *Expert Opin Ther Targets* 2006, 10: 515–531.
15. Coyle JT. The glutamatergic dysfunction hypothesis for schizophrenia. *Harv Rev Psychiatry* 1996, 3: 241–253.