

Functional Network Composed of 1,219 Genes for Schizophrenia-a Literature Data Mining and Enrichment Analysis

Shunan Li¹, Benjamin H Lehrman², Hongbao Cao³, Lydia C Manor^{4*}

1. Vanderhouwen & Associates, 6342 SW Macadam Ave, Portland 97239, OR;

2. Rush Medical College, 600 S Paulina St, Chicago 60612, IL ;

3. Elsevier Inc., 5635 Fishers Ln Rockville 20852, MD;

4. American Informatics Consultant LLC, Rockville 20852, MD.

***Correspondence:** Dr. Manor, American Informatics Consultant LLC, Rockville, MD, 20852, USA. Email: l.manor@gousinfo.com

ABSTRACT

Background: Over the past decade, numerous studies have focused on identifying genetic factors associated with schizophrenia (SCZ). Sample variations, such as size, population, race, disease status, and data processing methods resulted in selection differences. Nevertheless, no systemic study has been completed to summarize these reports and provide an objective full list of genes with a reported association to SCZ.

Methods: We conducted a literature data mining (LDM) of over 13,515 articles covering publications from 1958 to Feb. 2016. These articles reported multiple types of marker-disease associations between 1,219 genes and SCZ. Then we conducted a gene set enrichment analysis (GSEA) and a sub-network enrichment analysis (SNEA) to study the functional profile and validate the pathogenic significance of these genes to SCZ. Finally, we presented additional results from the systemic review, including publication date, quality scores, and author affiliations.

Results: All of these genes have been demonstrated to present multiple mutations associating with SCZ, some of which were supported by a large number of high quality articles. Enrichment analyses showed that many psychiatric and neuropathic pathways/groups related to SCZ have been significantly enriched by these genes and that they are functionally associated with each other.

Conclusion: Our results indicate that these genes may operate as a functional biomarker network influencing the development of SCZ, and that LDM together with GSEA and SNEA could serve as an effective approach in finding these potential target genes.

INTRODUCTION

Schizophrenia (SCZ) is a mental disorder characterized by abnormal social behavior and failure to understand reality. The cause of SCZ is believed to be a combination of environmental and genetic factors (Owen et al., 2016). Although the estimates of heritability vary, due to the difficulty in separating the genetic and environmental effects, the averaged number of 0.80 has been used



OPEN ACCESS

DOI: 10.20900/jpbs.20160004

Received: January 19, 2016

Accepted: February 15, 2016

Published: April 25, 2016

website: <http://jpbs.qingres.com>

Copyright: ©2016 Cain et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors received no specific funding for this work.

Competing Interests: The authors have declared that no competing interests exist.

(Herson M, 2011). Over time, we have noticed an increased number of articles reporting over a thousand of genes/proteins that may potentially be related to SCZ (Sacchi et al., 2013; Winchester et al., 2014; Liao et al., 2015). Among those reported marker-disease relations, many were supported by high quality articles with thousands of citations (Stefansson et al., 2002; Egan et al., 2003), while some of those genes were newly reported in recent years, with a limited number of publication supports (Schmidt et al., 2015). Moreover, most of the genes have been reported in different genomic studies to have different types of relations with SCZ.

According to how the gene-SCZ relations were reported, those articles can generally be classified into several different categories: 1) Biomarker 2) Clinical Trial 3) Genetic Change; 4) Quantitative Change ; 5) Regulation ; 6) State Change.

A relatively small amount of articles suggested that the genes reported in their study could serve as biomarkers for the disease (Leonard et al., 2006; Boneberg et al., 2006; Kanazawa et al. 2007; Nohesara et al., 2011; Tenback et al., 2011; Fatemi et al., 2013; Abdolmaleky et al., 2014). Compared to the total number of articles reporting marker-disease relationships, biomarker-type articles are in the minority. It also should be noted that most articles in this group are using a weak modal for their conclusion. For example, AKT1 has been reported to be potentially related to SCZ, because of the weak p-values of the study (Thiselton et al., 2008).

An even smaller number of articles reported SCZ clinical trial related genes ((Strickett et al., 2011; Lindenmayer et al. 2015). However, clinical study results are effective in evaluating the functional relation between genes and SCZ in a more direct way.

There are large number of articles claiming genetic changes of the genes in SCZ in terms of gene deletions, amplifications, mutations, or epigenetic changes. This is the largest group of all article categories. Many researchers working in this group studied the direct relation between genetic change of marker genes and the development of the disease using GWAS or sequencing data from both human and animals (Sacchi et al., 2013; Winchester et al., 2014; Porteous et al., 2014; Agha et al., 2014). Alternatively, many observed marker-disease connections are actually associations between markers and one or more symptoms of SCZ (Deng et al., 2013; Pearson-Fuhrhop et al., 2013; Cheng et al., 2014). It should also be noted that many studies focused on gene polymorphisms (Houston et al., 2013; Liao et al., 2015; Hu et al., 2015). Interestingly, the molecular changes of some genes have been identified to be case-sensitive, and thereby have been suggested as markers to differentiate subgroups of SCZ (Villar-Menéndez et al., 2014). Additionally, the population of the study has been identified to influence the association between the markers and the disease (Gu et al., 2013). Nevertheless, the results are not always consistent. For example, both positive and negative associations of common SNPs in CHRNA7 have been reported in schizophrenia GWAS studies (Winchester et al., 2014).

Quantitative Change is another big group. Articles in this group reported changes in abundance/activity/ expression of a gene/protein in the disease state, and most were using gene expression studies (Yanagi et al., 2014; Chen et al., 2014; Yin et al., 2013; Le Magueresse et al., 2013; Du et al., 2014; Vauquelin et al., 2012; Penzes et al., 2011). Among those reports, some genes were observed to have increased expression levels in the case of SCZ (Zvara et al., 2005; Graziane et al., 2009; Fuxe et al., 2010; Sacchi et al., 2013; Yin et al., 2013; Chen et al., 2014; Joshi et al., 2014; Du et al., 2014). These genes include: BDNF, NRG1, DRD4, ERBB4, COMT, HTR2A, DRD3, and DAO. Alternatively, many other gene were reported to present decreased gene expression levels in SCZ studies: CHRNA7 (Zheng et al., 2012), GAD1 (Yanagi et al., 2014), DTNBP1 (Le Magueresse et al., 2013), RELN (Folsom et al., 2012), AKT1 (Chakraborty et al., 2014), DRD2 (Vauquelin et al., 2012), DISC1 (Penzes et al., 2011), MTHFR (Hill et al., 2011), and SLC6A3 (El Hage et al., 2015).

Many articles reported changing activity of a target gene by an unknown mechanism, which are clustered as a regulation group. The quantity of those types of article reports are similar to Quantitative Change. This is a less specific relation type than the others provided. Although reports of this type of relation were equivocal when describing the mechanism of association, they are affirmative in stating that the genes/proteins were related to the disease (Nikolaus et al., 2014; Yu et al., 2013; Ivleva et al. 2008; Kret et al., 2015; Luoni et al., 2014; Schroeder et al., 2015; Lee et al., 2012; Roffeei et al., 2014). It should be noted that, many genes were suggested to be important for new SCZ drug development (Winchester et al., 2014; Khoddami et al., 2015; Sacchi et al., 2013). Although some conflicting outcomes among different methodologies led to continuous debate, the vast amount of regulation-type reports did provide more information regarding the association between the markers and SCZ.

The last group is state change, which is about the same size as the Biomarker group. Articles in this group report changes in a protein/gene post-translational modification status or alternative splicing events

associated with SCZ (Guest et al., 2000; Kroeger et al., 2003; Neddens et al., 2011; Sequeira et al., 2012; Wilkinson et al., 2013). Although state change is a relatively smaller group, those studies reported specific protein/gene state changes that may be related to SCZ, and therefore are important for the understanding of the mechanism of the disease.

Although it is possible that false positives can be reported in some articles, those genes that have been frequently identified by different studies associating with SCZ should be brought up for further analysis and could be used as potential biomarker targets. Moreover, the ones that are newly reported in recent years may attract extra attention. Nevertheless, by far, no systematic analysis has evaluated the quality and strength of these reports to provide a full list of genes associating with SCZ. This is the aim of this study. Instead of focusing on one specific marker or function as were done in other reviews (Ando et al., 2011; Bliksted2016), we attempt to provide a full view of the genetic-map that is related to SCZ.

METHODS AND MATERIALS

Here is the workflow of this study: 1.) Literature data mining (LDM) to discover gene-SCZ relations; 2.) Enrichment analysis on the genes identified to validate their pathogenic significance to SCZ. 3.) Results of systematic review of the articles underlie these relations, including publication date, quality metric score, and author-affiliation analysis.

Literature Data Mining

In this study, we performed a literature data mining (LDM) of all articles available in the Pathway Studio database (www.pathwaystudio.com), which covers over 40 million scientific articles, seeking the ones that reported gene-SCZ relations. The LDM was conducted by employing the Natural Language Processing (NLP) text mining module of the Pathway Studio software.

Gene Set Enrichment Analysis

To better understand the underlying functional profile and validate the pathogenic significance of the reported genes, we performed a gene set/pathway enrichment analysis (GSEA) and a sub-network enrichment analysis (SNEA) on five groups: 1) Whole gene list (1,219 genes); 2) 4-subgroups selected using the highest quality matrix scores (61 genes in each group). In addition, we conducted a network connectivity analysis on a subset of genes using Pathway Studio(www.pathwaystudio.com).

GSEA (also known as functional enrichment analysis) is a method for analyzing biological high throughput experiments, which identify classes of genes or proteins that are over-represented in a large set of genes or proteins. These gene sets could be known biochemical pathways or otherwise functionally related genes. The method uses statistical approaches to identify significantly enriched or depleted groups of genes to retrieve a functional profile of the input gene set, in order to better understand the underlying biological processes. With this method, one does not consider the perturbation of single genes but of whole (functionally related) gene sets. The advantage of this approach is that it is more robust. It is more likely that a single gene is found to be falsely perturbed than it is that a whole pathway is falsely perturbed.

In addition to GSEA, we performed a sub-network enrichment analysis (SNEA), which is implemented in Pathway Studio using master casual networks (database) containing more than 6.5 million relationships derived from more than 4 million full text articles and 25 million PubMed abstracts. These networks are generated by a finely-tuned Natural Language Processing (NLP) text mining system to extract relationship data from the scientific literature, rather than the manual curation process used by IPA (<http://www.ingenuity.com/products/ipa>). The ability to quickly update the terminologies and linguistics rules used by NLP systems ensures that new terms can be captured soon after entering regular use in the literature. This extensive database of interaction data provides high levels of confidence when interpreting experimentally-derived genetic data against the background of previously published results(http://help.pathwaystudio.com/fileadmin/standalone/pathway_studio/help_ps_10.0/index.html?analyze_experiment.htm).

Quality Metric Analysis

In this work, we perform a quality metrics analysis on all marker-disease relations. Output of the analysis includes quality score (QScore), citation score (CScore), novelty score (NScore) and report frequency score (RScore) at article level as well as marker level. These quality measures could be used to sort the marker list.

Using the RScore one can identify the most frequently reported markers. At article level, RScore=1, in dictating a marker-disease relation has been reported; otherwise RScore=0. At marker level, RScore is the

sum of article level RScores, representing the report frequency of the marker.

Using the NScore one can identify the newly reported markers. Here we define the publication age as the Current year - publication year +1. According to different publication age threshold, we differentiate NScores into, where (years) =1, 2...; at article level, =0 when the publication age of the article is older than; otherwise >0. At marker level, =0 means the marker-disease relation has been reported more than years ago.

Using the CScore one can identify the marker-disease relations that are highly cited. The CScore of an article is defined as its number of citations, and the marker level CScore of a relation is the sum of the total citations of all the articles supporting the relation between the marker and the disease.

The QScore is a composite index considering three factors of an article-reported relation: 1) the citation number; 2) the publication age, and 3) the RScore. The QScore of an article is in the range of (0,1), and it is inversely related to publication age and positively related to its citation number. If an article is recently published with a high citation number, its QScore will be close to 1; otherwise, if the article is old with poor citations, its QScore will be close to 0. The marker level QScore is the sum of the QScores of all the articles supporting the marker. It should be noted that both article level and marker level QScores are designed on the relation level to evaluate the significance of the article(s) to the relation. If multiple marker-disease relations have been reported by one article, this articles will have QScores for each of those relations, although they are equal in value.

Article Author-affiliation Analysis

We summarized author-affiliation info of all the articles that reported a gene-SCZ relationship. Our intent is to provide the affiliations of the authors who are most active in the field, which is especially helpful for those who are seeking collaborators for SCZ related genetic/genomic studies.

RESULTS

Results Summary of LDM

Through the LDM approach, we discovered 13,515 articles (1958~ Feb. 2016) reporting 1,219 genes associated with SCZ. According to the reported category of gene-SCZ relations, the articles can generally be clustered into 6 different groups: 1) Biomarker (0.62%); 2) Clinical Trial (0.16%); 3) Genetic Change (53.91%); 4) Quantitative Change (22.51%); 5) Regulation (21.75%); 6) State Change (1.05%).

Among the reported genes, some are frequently reported with huge number of article support, such as COMT (430 articles), DRD2 (394 articles) and BDNF (341articles). These genes are the ones with highest RScores; Some are recently reported (e.g., reported within the last two years) to have a high NScore, such as FKBP5 (NScore:6.5), AGO2 (NScore: 4) and TREM2 (NScore:4). Some are frequently cited to have a high CScore: COMT (CScore: 20369), NRG1 (CScore:18039), GAD1 (CScore:18000). It should also be noted that genes with high report frequencies (RScore), do not necessarily have a higher number of citations (CScore), which may be caused by many factors such as the total number of underlying articles and their publication age. To balance these factors, we can use the QScore. We present in Table 1 the top 61 genes ranked by different scores, representing genes with different significance. The full results are provided in Supplementary Material 1.

Table 1 Top 61 Genes Reported Associations with SCZ ranked by Different Scores

Genes By RScore	COMT;DRD2;BDNF;NRG1;DISC1;DTNBP1;GAD1;HTR2A;RELN;ERBB4;DRD3;SLC6A4;DRD4;RGS4;MTHFR;NRXN1;AKT1;CHRNA7;SLC6A3;DAO;OXT;MAOA;S100B;IL1B;PRODH;IL6;TNF;GSK3B;GRIN2B;PVALB;CNR1;NOS1;EGF;HTR1A;SHANK3;PPP1R1B;HTR2C;TP53;ACE;HTR6;DRD1;NOTCH4;IL10;IL2;ZNF804A;GRM3;NCAM1;SOD2;ESR1;NPY;IL1RN;EP300;MIR137;CHRM1;CYP2D6;SYP;SRR;SYN2;NRG3;NTF3;APOD
Genes By NScore	FKBP5;AGO2;TREM2;C4BPA;S100A8;FRMPD4;MIR219-1;GRIN3B;MNG1;ACAN;CXCL10;GPR52;IL1R1;MIR1271;MIR134;MIR200C;MIR664;NDE1;NPAS4;PSMD9;REST;TLR3;TLR4;ARHGDI1A;LIMK1;AADAC;BCL11A;EPOR;YWHAG;AKAP9;C4BPB;RANBP9;C3;ACSM1;ANKS1B;ARNTL;ASMT;BAIAP2;BLOC1S6;CCL22;CDH13;CHD8;FXR1;FZD1;FZD5;FZD8;HLA-C;KCNB1;LHX5;MAP1A;MIR193A;NDUFS2;NEDD4;NLGN3;PLAT;POGZ;RAI14;SMARCC2;SREBF1;SYNGAP1;TNFRSF11B

Genes By CScore	COMT;NRG1;GAD1;DISC1;RELN;DRD2;BDNF;DTNBP1;HTR2A;RGS4;DRD3;DRD4;ERBB4;AKT1;DAO;CHRNA7;MTHFR;PRODH;GSK3B;HTR6;NRXN1;PVALB;PPP1R1B;S100B;IL6;SLC6A4;SLC6A3;SYP;HTR2C;HTR7;SRR;IL1B;SHANK3;OXT;TNF;MIR137;DNMT1;IL1RN;GRM3;EGF;GRIN2B;CNR1;SOX10;NOS1;CHRM1;HTR1A;GAP43;IL2;CACNA1C;MAOA;NCS1;IL2RA;APOD;NTRK2;NTF3;KALRN;FOLH1;NCAM1;DLG4;SLC6A1;DRD1
Genes By QScore	COMT;DRD2;BDNF;NRG1;DISC1;DTNBP1;GAD1;ERBB4;RELN;HTR2A;NRXN1;SLC6A4;RGS4;DRD3;AKT1;MTHFR;DRD4;CHRNA7;DAO;SLC6A3;S100B;IL1B;OXT;PRODH;IL6;TNF;SHANK3;GSK3B;PVALB;MAOA;GRIN2B;CNR1;NOS1;PPP1R1B;HTR2C;ZNF804A;DRD1;HTR1A;EGF;TP53;MIR137;ACE;IL10;GRM3;HTR6;CHRM1;NRG3;ESR1;EP300;SRR;IL2;NOTCH4;TCF7L2;RTN4R;NCAM1;IL1RN;RTN4;SOD2;CACNA1C;SYN2;HTR7

Note: NScore here is NScore₂; Any marker have been reported 3 years ago will have an NScore of 0.

We note that the top genes selected by QScore, CScore, RScore have large overlaps with each other; however, they show no overlap with the gene group selected by NScore, which contains all the genes that are newly reported within the last two years.

In addition, we demonstrated that 53 out of the 1,219 genes identified in this study are encoded by the 108 loci reported by Ripke et al., 2014. Some of those genes have been frequently reported with many citations, such as DRD2, ZNF804A and MIR137, while many others are relatively new with fewer supporting articles, such as MAD1L1 and AMBRA1. More information regarding these 53 genes are presented in Supplementary Material 1.

Enrichment Analysis

In this section, we present GSEA and SNEA results for 5 different groups: 1) All 1,219 genes; 2) The 4 gene groups listed in Table 1.

1. Enrichment Analysis on All 1,219 Genes

The full list of 819 pathways/gene sets that has been enriched with $p\text{-value} < 1E-5$ has been listed in Supplementary Material 2, where 605 pathways/genes were enriched with $p\text{-values} < 1E-6$, and 279 pathways/genes were enriched with $p\text{-values} < 1E-10$, 87 pathways/genes were enriched with $p\text{-values} < 1E-20$, and 22 pathways/genes were enriched with $p\text{-values} < 1E-40$ (Jaccard similarity > 0.04), as shown in Table 2.

Table 2 Molecular function pathways/ groups enriched by 1,219 genes reported

Pathway/gene set name	Hit type	GO ID	# of Entities	Overlap	p-value	Jaccard similarity
synaptic transmission	biological_process	0007268	472	165	9.72E-92	0.11
response to drug	biological_process	0042493	509	158	5.03E-79	0.10
neuronal cell body	cellular_component	0043025	466	149	1.16E-78	0.09
Synapse	cellular_component	0045202	466	141	8.24E-71	0.09
signal transduction	biological_process	0007165	1843	287	6.53E-68	0.10
Axon	cellular_component	0030424	318	115	3.75E-67	0.08
Dendrite	cellular_component	0030425	396	126	4.75E-66	0.08
postsynaptic membrane	cellular_component	0045211	227	94	3.47E-61	0.07
neuron projection	cellular_component	0043005	378	114	6.02E-57	0.08
plasma membrane	cellular_component	0005904	5390	509	4.1E-56	0.08
postsynaptic density	cellular_component	0014069	168	78	1.64E-55	0.06
response to ethanol	biological_process	0045471	161	75	2.35E-52	0.05
Membrane	cellular_component	0016020	7012	593	2.15E-51	0.08

integral component of plasma membrane	cellular_component	0005887	1360	210	2.65E-50	0.09
Cytosol	cellular_component	0005829	3173	348	7.52E-49	0.09
Aging	biological_process	0016280	254	88	2.79E-48	0.07
nervous system development	biological_process	0007399	523	119	1.53E-43	0.08
response to estradiol	biological_process	0032355	175	69	2.58E-42	0.05
Cytoplasm	cellular_component	0005737	6831	556	5.02E-41	0.07
Memory	biological_process	0007613	76	47	6.77E-41	0.04
enzyme binding	molecular_function	0019899	369	98	2.39E-40	0.07
cell junction	cellular_component	0030054	705	132	3.32E-40	0.08

From Table 2 we show that most of the extremely enriched pathways/gene sets were related to neuron system development or neuron signal transition. For example, synapse, neuronal cell body, axon, dendrite, postsynaptic density, signal transduction, neuron projection, synaptic transmission, and nervous system development. Enrichment in these pathways suggests that the genes are related to the neurological basis of SCZ (Danielyan and Nasrallah, 2009).

Some pathways/gene sets were related to drug effects, including response to drug (GO: 0042493, $P=5.03E-79$), response to ethanol (GO: 0045471; $P=2.35E-52$), and response to estradiol (GO: 0032355; $p\text{-value}=2.58E-42$). The genes in these pathways play important roles in the change in state or activity of a cell or an organism in terms of movement, secretion, enzyme production, gene expression, etc. as a result of a drug stimulus. A drug is a substance used in the diagnosis, treatment or prevention of a disease. Enrichment of the groups suggests that the genes may serve as drug targets for SCZ. Coincidentally, many of them, such as COMT, DRD2, DRD3, DRD3 and HTR2A (Stricket al., 2011; Winchester 2014), have already been suggested as antagonist genes for SCZ.

Other interesting pathways/gene sets include aging (GO: 0016280; $p\text{-value}=2.79E-48$; 88 genes) and memory (GO: 000761; $p\text{-value}=6.77E-41$; 47 Genes). It has been well established that SCZ is related to learning and memory impairments, and that age is an important factor influencing the pathogenesis of SCZ (Paulsen et al., 1995).

More significantly enriched pathways have been identified and presented in Supplementary Material 2. The results provide insights for understanding the functional processing through which the genes affect the development of SCZ, and validate the pathogenic significance of these 1,219 genes to SCZ.

In addition to GSEA, we performed SNEA using Pathway Studio with the purpose of identifying the pathogenic significance of reported genes with other disorders, especially psychiatric disorders. We provide the full list of results in Supplementary Material 3. We list in Table 3 the disease related sub-networks enriched with a $p\text{-value}<1.24E-168$.

Table 3 Sub-networks Enriched by the by 1,219 Genes Reported

Gene Set Seed	Total # of Neighbors	Overlap	p-value	Jaccard similarity
Bipolar Disorder	484	340	$<1E-308$	0.258949
Alzheimer Disease	1304	442	$5.90E-274$	0.217627
Depressive Disorder, Major	436	281	$1.50E-266$	0.212236
Seizures	824	323	$1.99E-216$	0.193413
Autistic Disorder	537	253	$4.22E-191$	0.174123
Anxiety	375	218	$5.14E-190$	0.164404
Parkinson Disease	720	283	$4.26E-188$	0.176214
Cognitive impairment	406	211	$1.29E-169$	0.154692

From Table 3 we see that many of these reported SCZ related genes were also identified in other mental

health disorders, with a large percentage of overlaps (Jaccard similarity > 0.15).

2. Enrichment Analysis on Top 61 Genes with Highest Scores

From Table 1 we see that the gene set with the highest NScores has few overlaps with the top gene sets ranked by the other scores, while the top genes selected by QScore, CScore, RScore have large overlaps with each other. This is because QScore, CScore, RScore are strongly related (See Methods) while NScores represents relatively different meanings. Here we compared their difference in terms of GSEA and SNEA results. Considering the similarity of the groups selected by QScore, CScore, RScore, we only present the results for the NScore group and the RScore group here (Table 4), and report the full results for QScore and CScore groups in Supplementary Material 2 and 3.

Table 4 Pathways/groups Enriched by 61 Genes with the Highest NScore and RScore

	Pathway/gene set Name	GO ID	p-value
The first 10 pathways/gene sets enriched by top 61 genes with highest NScores	signal transduction	0007165	4.98E-08
	positive regulation of transcription from RNA polymerase II promoter	0010552	1.96E-07
	TIR domain		2.56E-06
	innate immune response	0002226	3.69E-06
	regulation of synaptic plasticity	0048167	5.47E-06
	regulation of opsonization	1903027	6.26E-06
	PAZ domain		1.1E-05
	Rho protein signal transduction	0007266	1.7E-05
	Proteins Involved in Pathogenesis of Age-Related Macular Degeneration		1.95E-05
	transmembrane signaling receptor activity	0004926	2.14E-05
The first 10 pathways/gene sets enriched by top 61 genes with highest RScore	response to drug	0017035	1.93E-25
	Biofluids assayable substances		3.43E-22
	memory	0007613	1.73E-20
	drug binding	0008144	1.54E-19
	axon	0030424	4.36E-18
	Proteins Involved in Pathogenesis of Obesity		6.89E-18
	synaptic transmission	0007268	6.12E-15
	social behavior	0035176	1.65E-14
	Proteins Involved in Pathogenesis of Amyotrophic Lateral Sclerosis		4.22E-14
	dendritic spine	0043197	5.82E-14

Note: 1) NScore is used as NScore_2, non-zero-value of which represents the gene is reported within the last two years; there are a total of 61 genes reported to have non-zero NScore_2 values.

From Table 4 we see that, the genes with the top NScores and those with the top RScores are enriching different groups of pathways, with different p-values (NScore group: 2.14E-05~4.98E-08; RScore group: 5.82E-14~1.93E-25), indicating that the newly reported genes are functionally different from the frequently reported ones.

Moreover, we observe that the top 10 RScore group enriched pathways/gene sets (Table 4) have several overlaps with the 22 pathways/gene sets enriched by the overall 1,219 genes (Table 2), including: response to drug (GO: 0017035), memory (GO: 0007613), axon (GO: 0030424), and synaptic transmission (GO: 0007268), though with relatively bigger but still strong p-values. Similarly, we see that the signal transduction group (GO: 0007165) is enriched by both overall genes and the NScore group alone, although with much weaker significance (6.53E-68 vs. 4.98E-08), indicating that many more genes with similar function have already been discovered.

For the SNEA analysis, we only test the disease sub-networks that have been enriched by the two groups

of genes. We provide the full list of results in Supplementary Material 3. We present in Table 5 the top 10 disease related sub-networks enriched by the two groups of genes.

Table 5 SNEA Results by 61 Genes with the Highest NScore and RScore

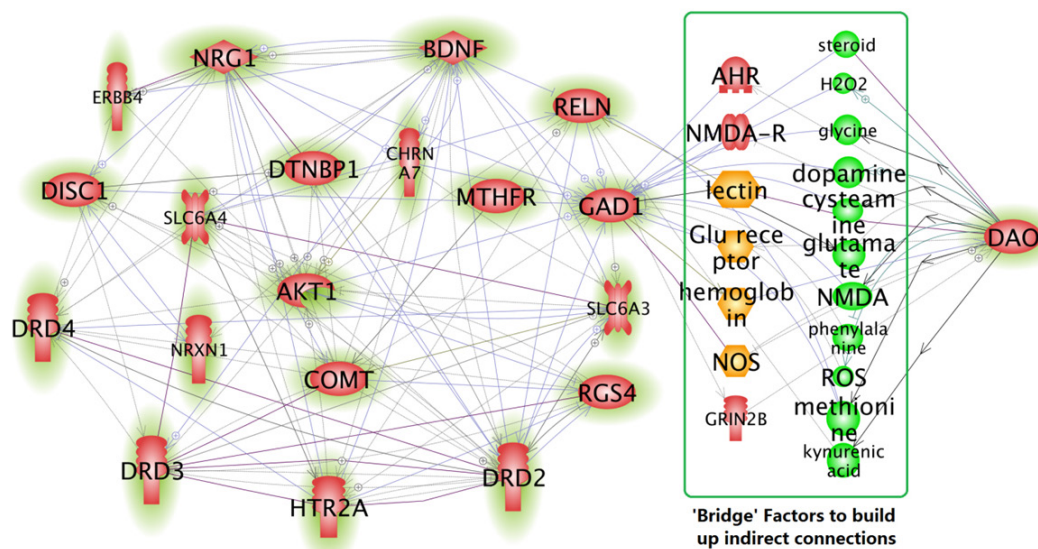
	Gene Set Seed	Overlap	p-value	Jaccard similarity
The first 10 pathways/ gene sets enriched by top 61 genes with highest NScores	Injuries	21	3.94E-10	0.011858
	Insulin Resistance	16	4.97E-10	0.016598
	Autism Spectrum Disorders	12	5.45E-10	0.025052
	Arthritis, Psoriatic	8	1.58E-09	0.042553
	Breast Cancer	26	4.3E-09	0.008309
	Autistic Disorder	12	6.6E-09	0.020478
	Neoplasms	34	7.69E-09	0.006267
	Alzheimer Disease	17	1.01E-08	0.012611
	Brain ischemia	12	2.95E-08	0.018072
The first 10 pathways/ gene sets enriched by top 61 genes with highest RScore	Psychotic Disorders	50	1.49E-87	0.163934
	Mental Disorders	50	1.1E-81	0.128535
	Bipolar Disorder	51	1.85E-78	0.103239
	Depressive Disorder, Major	47	3.36E-71	0.104444
	Mood Disorders	38	7.02E-67	0.183575
	Alzheimer Disease	52	6.09E-58	0.039604
	cognitive impairment	40	4.19E-57	0.093677
	negative syndrome	25	1.17E-54	0.297619
	Autistic Disorder	41	4.44E-54	0.073609

From Table 4 we see that both groups enriched some psychiatric disorder related sub-networks. However, the enrichment p-values by the RScore group are much more smaller than those by the NScore group. Moreover, more disease sub-networks were enriched by both RScore group and over-all group.

3. Connectivity Analysis

In addition to GSEA and SNEA, we performed a network connectivity analysis on the top 20 genes with the highest RScore (from Table 1) to generate a functional network, as illustrated in Fig. 1.

Fig. 1 Connectivity Network of 20 Genes with Highest RScore. The network is generated using Pathway Studio. The 20 genes discussed in the paper are highlighted using a green halo. All, but one gene DAO, have functional associations with each other, supported by over 2,000 articles. DAO has a strong indirect connection with other genes, and here we only present its indirect connection with RELN and GAD1 as an example.



From Fig. 1 we see that the genes, especially the 'hot' ones in terms of reported frequency, demonstrate strong connectivity between each other.

Quantitative Analysis of Literature Support

Fig. 2 presents the histogram of the number of articles against publication date. As shown in Fig. 2, the LDM of this study covers articles within the past 59 years (1958 to 2016). However, they have an averaged publication age of 7 years, suggesting that most of the articles are published in recent years. Here we define the publication age = current year - publication date + 1. It should be noted that, recent years saw an increased number of publications, especially after 2009. Although we observe the highest number of articles published in 2012, we cannot neglect that overall there are similar and higher volume of publications in recent years. The frequent and continuous exposure in terms of scientific reports in recent years suggest that these genes have been, and are still attracting significant attention in the genetic studies of SCZ.

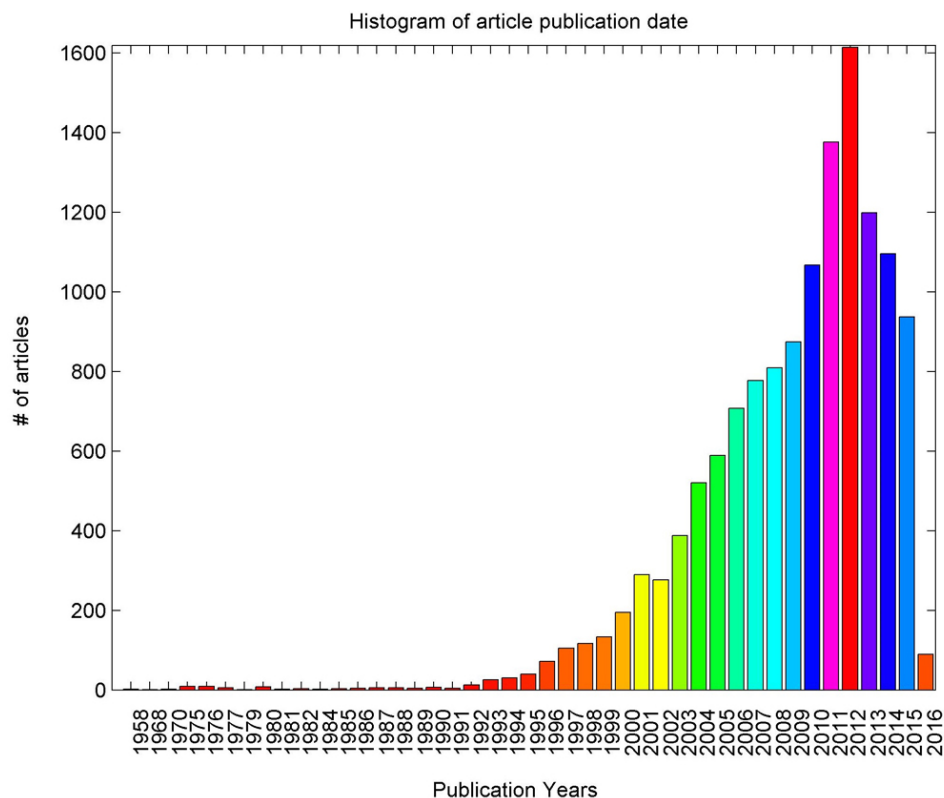


Fig. 2 Histogram of the Publications Reporting Marker-disease Relationships between SCZ and 1,219 Genes

In addition, we performed a marker-wise analysis of the publication date distribution of the supporting articles and provided the error-bar plot as shown in Fig. 3. The X-axis contains the names of the markers ranked by QScore. The Y-axis represents the mean of the publication date plus/minus the corresponding standard deviation. From the figure, we can see that the publication date distribution of most articles associated with each marker is similar, which are all centered around 2010.

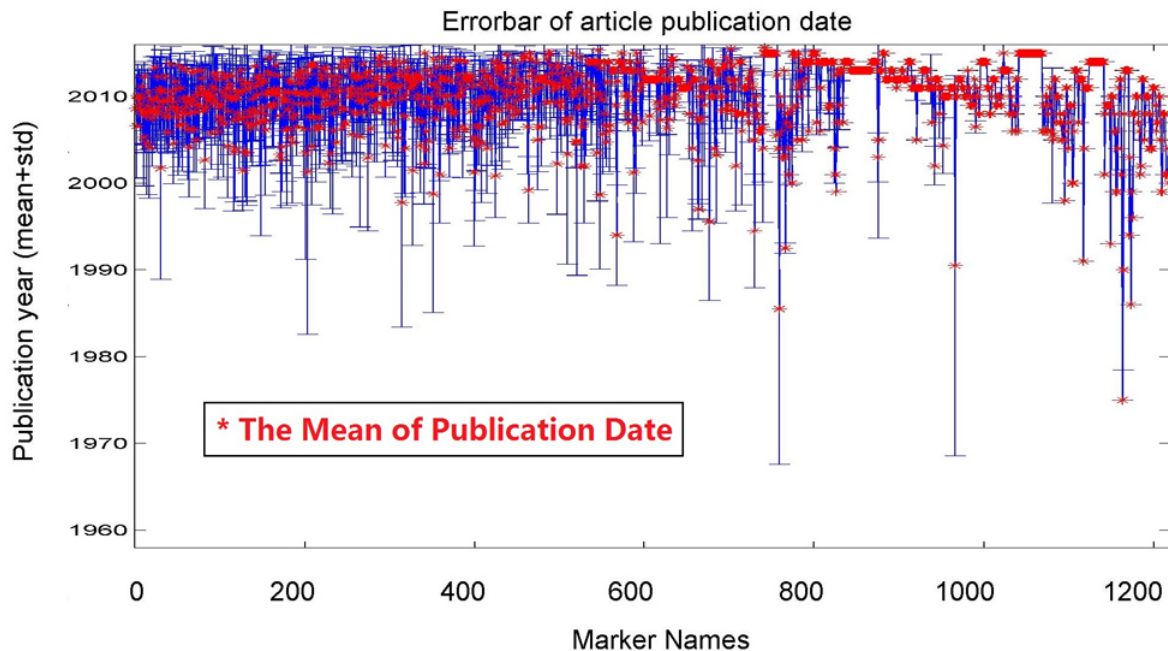


Fig. 3 Error-bar Plot of the Publication Date of the Articles for Each Gene

Quality Metrics Analysis

Fig. 4 presents the marker-wise CScore, QScore and NScore, along with the number of references (RScore). The X-axis represents the index of markers ranked by QScore; The Y-axis contains the CScore, QScores, NScore, and RScore normalized by their maximum values, respectively. As shown in Fig. 4, the QScores are roughly proportional to the RScore and CScore. Similarly Pathway Studio documents also demonstrate that QScore is inversely related to publication age. Nevertheless, the markers we discussed have similar publication ages (Fig. 3). Therefore, publication age contributes less in the ranking of these markers. In this study, the genes we identified have QScores in the range of [0.01, 172.16].

It also should be noted that some genes with low QScores present relatively high NScores (Fig. 3). This signifies that these genes have been newly identified during the past two years and it also signifies that the higher the NScore, the 'hotter' the newly identified genes for SCZ.

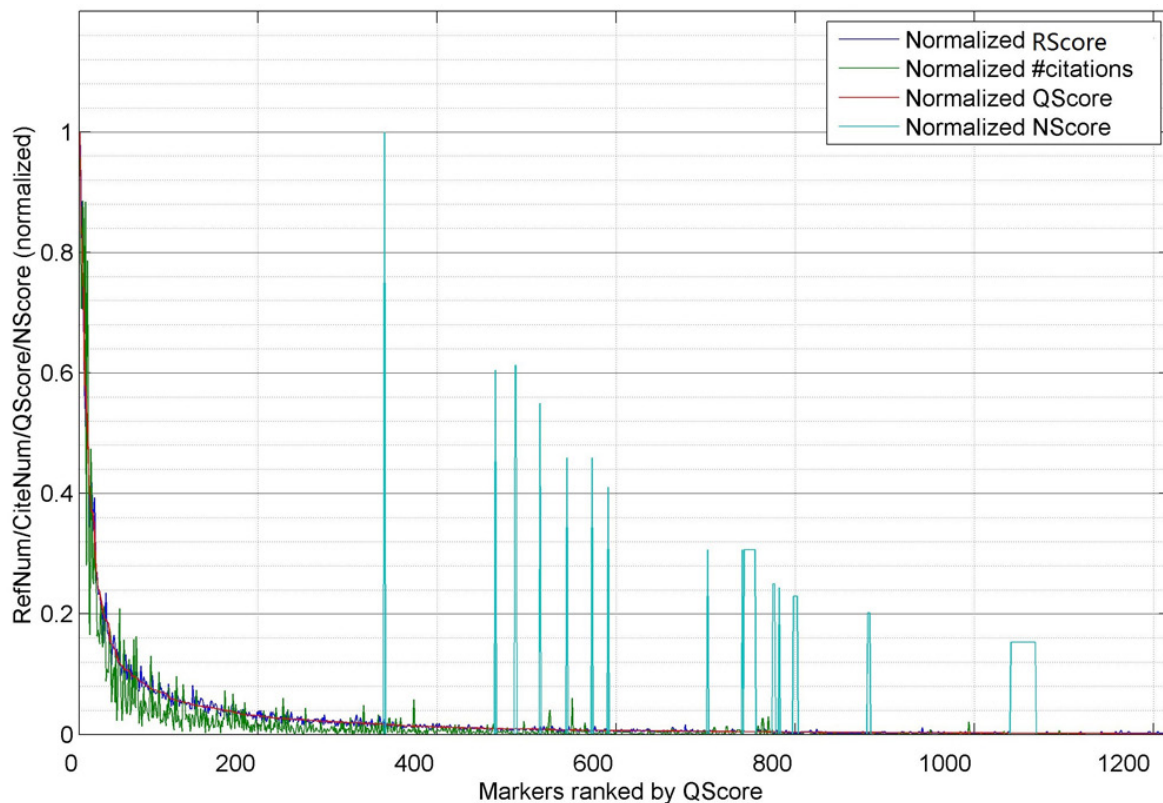


Fig.4 Plot of CScore, QScore, NScore and RScore. Each of these measures were normalized using the corresponding maximum value. The NScore presented in this figure are the NScore₂, so that the NScore will be zero if corresponding gene gets supports from articles with publication age older than 2 years.

Interestingly, we found that the majority of the QScores of the articles underlying associations between particular genes and the disease follow a normal distribution, as shown in the QQPlot provided by Fig. 5. The articles with QScores following normal distributions are roughly corresponding to the ones that lie within a band around the mean with a width of one standard deviation, representing about 75.29% of the total articles, while for an normal distribution, this percentage is 68.27%.

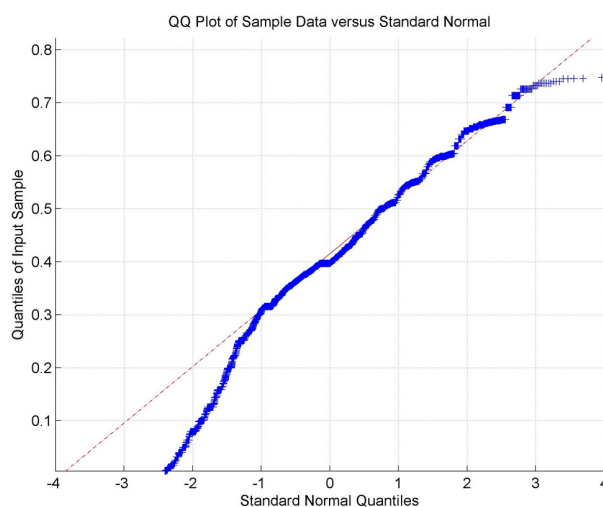


Fig.5 QQPlot of All QScores of 13,515 Articles

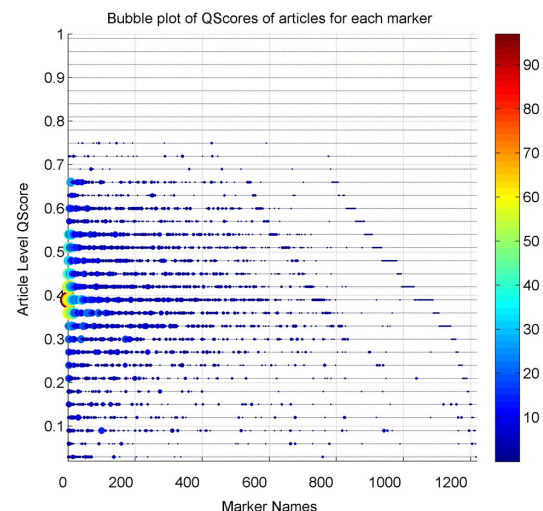


Fig.6 The Gene-wise Bubble Plot of QScores

Fig. 6 presents the marker-wise bubble plot of the articles, where the number of articles are represented by both the size and the color of the bubbles. As shown in Fig. 6, most of the articles are supporting the genes with higher QScore.

Furthermore, there are approximately 11.93% articles above the 1-standard-deviation band, and about

12.78% articles below the 1-standard-deviation band, as shown in Fig. 5. Although all the articles designating a marker are important in supporting the marker-disease relation, the quantity metric scores, namely RScore, NScore, CScore and QScore, provide several tentative methods to rank the markers related to the disease, as well as provide an approach to rank the articles behind a marker if there are numerous.

Author-affiliation Analysis of Literature Support

In this section, we analyzed the affiliations that publish these articles, and present the analyzing results including: 1) The top affiliations ranked by number of publications; 2) The collaboration network among these top affiliations; 3) The classification of these top affiliations; and 4) The top companies that studied the relations between the markers and the disease.

Fig. 7 presents the top 15 affiliations with the number of articles they published. There is a total of 13,515 articles from 3,601 affiliations supporting the relations between the markers and SCZ, out of which 3,858 articles are from the top 15 affiliations (0.4%), accounting for 28.5% of all articles. As shown in Fig. 7, most of these top affiliations are famous academic institutes, including National Institute of Mental Health, Shanghai Jiaotong University, VA Medical Center, University of Toronto, Shanghai Mental Health Center, Japan Science and Technology Agency, King's College London, The Johns Hopkins School of Medicine, Icahn School of Medicine at Mount Sinai, Peking University, Chinese Academy of Sciences, Ludwig-Maximilians-University at Munchen, Karolinska Institute, University of Pittsburgh, and Centre for Addiction and Mental Health. Although an article by a famous affiliation is not necessarily a guarantee of a good publication, it's still good to know that these organizations with high reputation were involved in the studies of the marker-disease relations between the genes reported and SCZ.

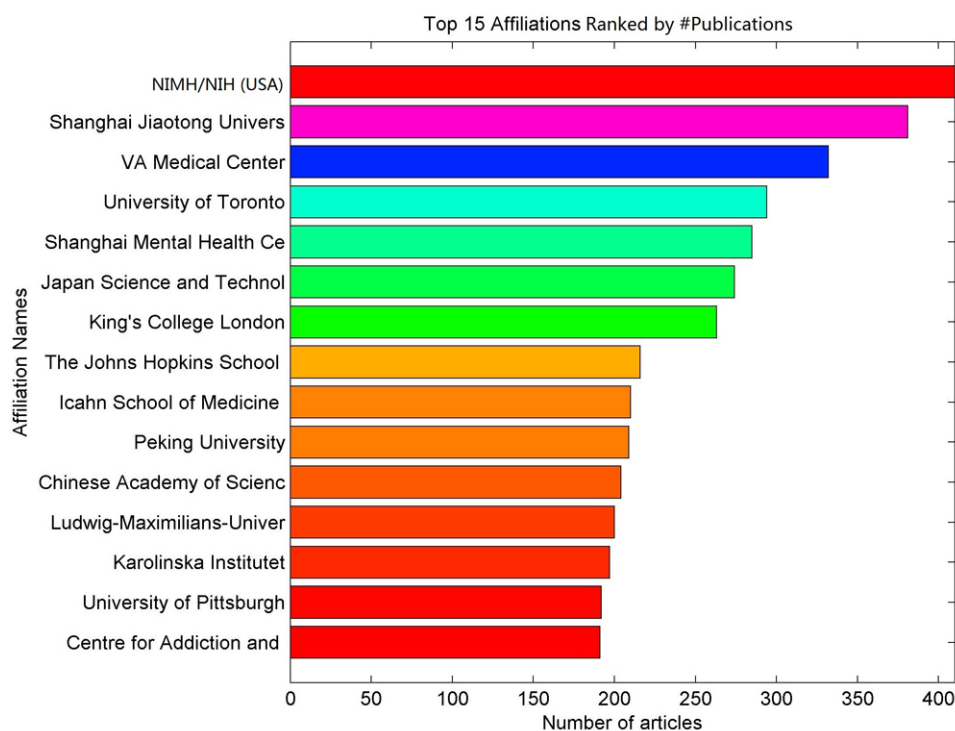


Fig.7 The Top 15 Affiliations Ranked by Article Numbers They Published Supporting the Marker-disease Relations Between the 1218 Genes and SCZ

During the affiliation analysis of these articles, we noticed that many of the studies reported in the papers were developed under the joint efforts from two or more affiliations, as shown in Fig. 8.

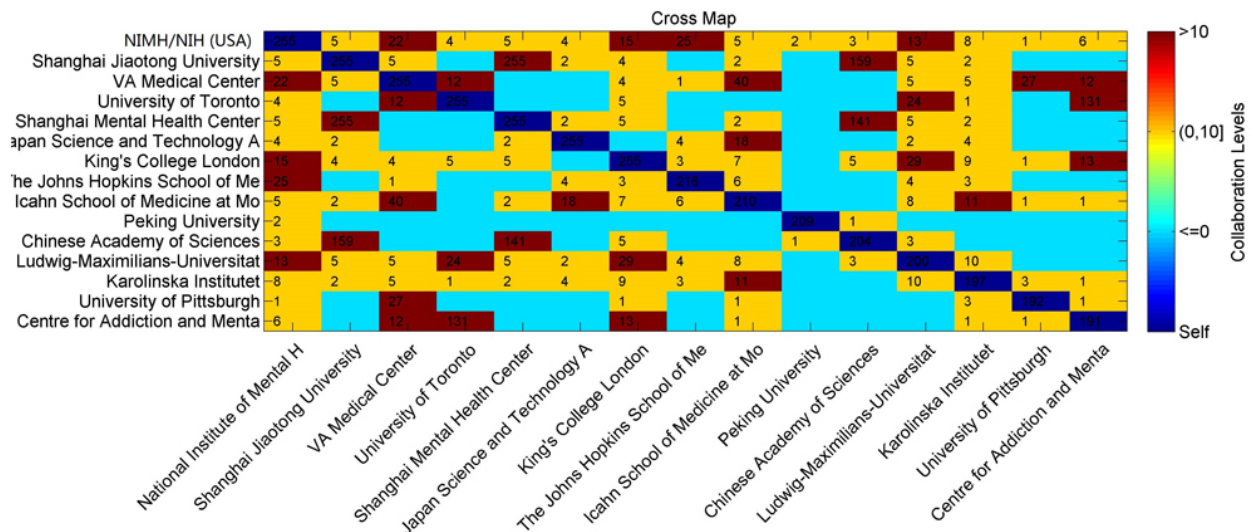


Fig.8 Collaboration Map of the Top Organizations Ranked by Publication Numbers; The number in each cell is the number of articles coauthored by both corresponding affiliations. If the number is equal to or greater than 255, it will appear as 255.

Fig.8 suggests that the top affiliations have strong collaborations with each other in terms of paper publications. For example, Shanghai Mental Health Center and Shanghai Jiaotong University have been published over 255 articles together studying the relations between the disease and the genes discovered. These collaborations indicated the joint efforts in the field. Generally, collaborations would lead to more solid and reliable conclusions.

It should also be noted that, besides schools and universities, there are also many other types of affiliations, such as hospitals, medical centers and companies. In this work, we separated the affiliations into four different classes (Fig. 9): 1) Academic institutes, including schools, universities, research centers and government academic institutes; 2) Hospitals and medical centers; 3) Companies; 4) Others.

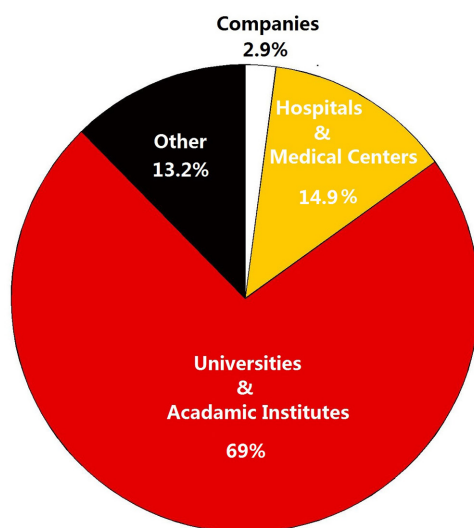


Fig.9 Pie Plot of the Author Affiliation Distribution

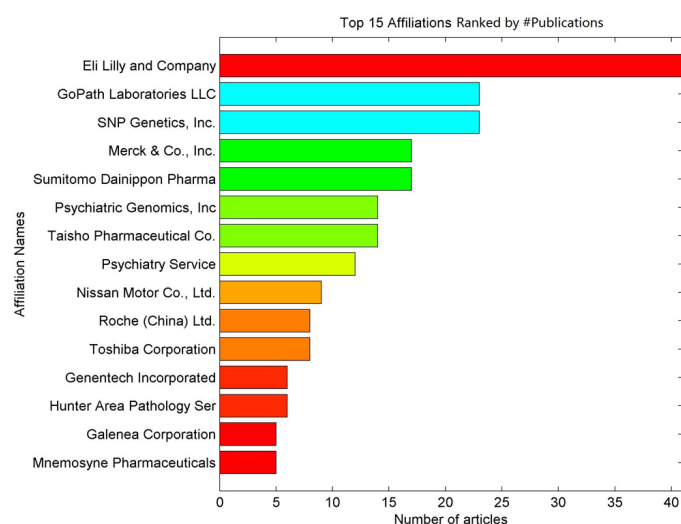


Fig.10 The 15 Companies Reporting Associations between the Genes and SCZ

From Fig. 9 we can see that most of the papers (69.0%) were developed by academic institutes, while the number for hospitals and medical center and companies were 14.9% and 2.9%, respectively. About 13.2% of articles have affiliations that could not be classified to any of these three classes. This was partially due to the incomplete information extraction. We extracted the affiliation information using Elsevier Scopus Search API (<http://dev.elsevier.com>), which is publicly available, but is not 100% percent responsive to quotes. Although the articles by companies only contribute a small part to the total publications (Fig. 10), the companies with

a relatively high number of publications should be considered with higher regard. Different from academic institutes whose research studies are mostly in the field of basic science, companies are more focused on applied research. The research they sponsor or directly perform is more likely to be linked to immediate applications, such as drug development. Fig.10 depicts the top 15 companies reporting associations between the markers we discussed and SCZ.

DISCUSSION AND CONCLUSION

In this work, we performed a LDM and a systematic summarization of over 13,515 articles (from 1958 to Feb. 2016) reporting 1,219 genes associated with SCZ. We provided in Supplementary Materials 1 the full analysis results. In addition, we conducted GSEA to study functional profile and the pathogenic significance of the reported genes with SCZ, with full results provided in Supplementary Materials 2. In addition, we conducted SNEA to further study the significance of these genes to potential SCZ related diseases and present the full analysis results in Supplementary Materials 3. Different from the genetic studies using raw data to report novel discoveries, this is a literature-based summarization and validation of already reported marker-diseases relations.

As an automatic data mining approach, the Natural Language Processing (NLP) technique used for LDM is effective and necessary in dealing with millions of articles. However, the automatic LDM method may produce some false positives. Therefore, the results of this study are intended to lay the groundwork for further studies in the area. Towards this purpose, we provided in Supplementary material 1 the detailed information of all the 13,515 articles studied for further investigation, including the sentences where a specific relation has been located.

This study has several limitations that need further work. 1) The literature data of 13,515 articles studied were extracted from Pathway Studio database. Although Pathway Studio database is composed of over 40 million articles, it is still possible that some articles studying gene-SCZ associations were out of their coverage. 2) The 4 quality scores, RScore, NScore, CScore and QScore are proposed as quality measures of LDR identified marker-disease relations, feasible to rank the markers/relations according to different needs/significance. Even though related to, however, they are not biological significance measures of the markers to the disease, which is different from other statistical studies like GWAS, meta-analysis and enrichment analysis.

Nevertheless, results from this up-to-date LDM suggest that these 1,219 genes have multiple types of association with SCZ, and enrichment analysis suggests that these genes play significant roles in the pathogenesis of SCZ, as well as in the pathogenesis of many other SCZ related psychiatric disorders. Together with the network connectivity analysis, we conclude that SCZ is a complex disease whose genetic causes are linked to a network composed of a large group of genes. LDM together with GSEA and SNEA could serve as an effective approach in finding these potential target genes.

DECLARATION OF INTERESTS

Dr. Hongbao Cao is from Elsevier Inc. The other authors declare no conflict of interests.

REFERENCES

1. Abdolmaleky HM, Nohesara S, Ghadirivasfi M, Lambert AW, Ahmadkhaniha H, Ozturk S, Wong CK, Shafa R, Mostafavi A, Thiagalingam S (2014). DNA hypermethylation of serotonin transporter gene promoter in drug naïve patients with schizophrenia. *Schizophrenia Research*, 152:(2-3):373-380.
2. Agha Z, Iqbal Z, Azam M, Siddique M, Willemsen MH, Kleefstra T, Zweier C, de Leeuw N, Qamar R, van Bokhoven H (2014). A complex microcephaly syndrome in a Pakistani family associated with a novel missense mutation in RBBP8 and a heterozygous deletion in NRXN1. *Gene*, 538(1):30-35.
3. Ando S, Clement S, Barley EA, Thornicroft G (2011). The simulation of hallucinations to reduce the stigma of schizophrenia: a systematic review. *Schizophr Res*. 133(1-3):8-16.
4. Balu DT, Li Y, Puhl MD, Benneyworth MA, Basu AC, Bolshakov VY, Coyle JT, Balu DT, Puhl MD, Benneyworth MA, Basu AC, Coyle JT, Takagi S (2013). Multiple risk pathways for schizophrenia converge in serine racemase knockout mice, a mouse model of NMDA receptor hypofunction. *Proceedings of the*

National Academy of Sciences of the United States of America, 110(26):E2400-E2409.

5. Barry G, Vanichkina DP, Taft RJ, Mattick JS, Briggs JA, Nayler SP, Wolvetang EJ, Poth EM, Blackshaw S, Beveridge NJ, Cairns MJ, Beveridge NJ, Cairns MJ, Ratnu VS, Bredy TW, Nones K, Hu J, Nakagawa S, Rigo F, Mattick JS, Mattick JS (2014). The long non-coding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing. *Molecular Psychiatry*, 19(4):486-494.
6. BeveridgeA(2002). The Creation of Psychopharmacology. *J R Soc Med*, 95(11): 568–569.
7. Bliksted V, Ubukata S, Koelkebeck K (2016). Discriminating autism spectrum disorders from schizophrenia by investigation of mental state attribution on an on-line mentalizing task: A review and meta-analysis. *Schizophr Res*. 171(1-3):16-26.
8. Boneberg EM, von Seydlitz E, Pröpster K, Watzl H, Rockstroh B, Illges H (2006).D3 dopamine receptor mRNA is elevated in T cells of schizophrenic patients whereas D4 dopamine receptor mRNA is reduced in CD4 +-T cells. *Journal of Neuroimmunology*, 173(1-2):180-187.
9. Braff DL, Geyer MA, Swerdlow NR (2001). Human studies of prepulse inhibition of startle: normal subjects, patient groups, and pharmacological studies. *Psychopharmacology (Berl)*. 156(2-3):234-58.
10. Chakraborty A, Xu J, Snyder SH, Chakraborty A, Latapy C, Beaulieu J-M, Snyder SH, Snyder SH, Chakraborty A (2014). Inositol hexakisphosphate kinase-1 regulates behavioral responses via GSK3 signaling pathways. *Molecular Psychiatry*, 19(3):284-293.
11. Chen SL, Lee SY, Chang YH, Chen SH, Chu CH, Wang TY, Chen PS, Lee IH, Yang YK, Hong JS, Lu RB (2014). The BDNF Val66Met polymorphism and plasma brain-derived neurotrophic factor levels in Han Chinese patients with bipolar disorder and schizophrenia. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 51(0):99-104.
12. Cheng J, Wang Y, Zhou K, Wang L, Li J, Zhuang Q, Xu X, Xu L, Zhang K, Dai D, Zheng R, Li G, Zhang A, Gao S, Duan S (2014). Male-specific association between dopamine receptor D4 gene methylation and schizophrenia. *PLoS ONE*,9(2). DOI: 10.1371/journal.pone.0089128
13. Danielyan A, Nasrallah HA (2009). Neurological disorders in schizophrenia, *PsychiatrClin North Am*. 32(4):719-57.
14. Debanne D, Campanac E, Bialowas A, Carlier E, Alcaraz G (2009). Axon physiology.*Physiol Rev*, 91(2): 555-602.
15. Deng C (2013).Effects of Antipsychotic Medications on Appetite, Weight, and Insulin Resistance. *Endocrinology and Metabolism Clinics of North America*, 42(3):545-563.
16. Du L, Merali Z, Poulter MO, Palkovits M, Faludi G, Anisman H (2014). Catechol-O-methyltransferase Val158Met polymorphism and altered COMT gene expression in the prefrontal cortex of suicide brains. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 50:178-183.
17. Egan MF, Kojima M, Callicott JH, Goldberg TE, Kolachana BS, Bertolino A, Zaitsev E, Gold B, Goldman D, Dean M, Lu B, Weinberger DR (2003). The BDNF val66met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function. *Cell*. 112(2):257-69.
18. El Hage C, Bédard AM, Samaha AN (2015). Antipsychotic treatment leading to dopamine supersensitivity persistently alters nucleus accumbens function. *Neuropharmacology*. 99:715-25. doi: 10.1016/j.neuropharm.2015.03.012.
19. Elena Ivleva,2 Guntav Thaker,3 and Carol A. Tamminga (2008).Comparing Genes and Phenomenology in the Major Psychoses: Schizophrenia and Bipolar 1 Disorder, *Schizophr Bull*. 34(4): 734–742.
20. Fatemi SH, Fatemi SH, Fatemi SH (2013). Cerebellum and autism. *Cerebellum*, 12(5):778-779
21. Folsom TD, HosseinFatemi S (2012). The involvement of Reelin in neurodevelopmental disorders. *Neuropharmacology*, 68(0):122-135.
22. Fuxe K, Marcellino D, Leo G, Agnati LF (2010). Molecular integration via allosteric interactions in receptor heteromers. A working hypothesis. *Current Opinion in Pharmacology*, 10(1):14-22.
23. Graziane NM, Yuen EY, Yan Z (2009). Dopamine D4 Receptors Regulate GABAA Receptor Trafficking via an Actin/Cofilin/Myosin-dependent Mechanism. *Journal of Biological Chemistry*, 284(13):8329-8336.
24. Gu L, Long J, Yan Y, Chen Q, Pan R, Xie X, Mao X, Hu X, Wei B, Su L (2013). HTR2A-1438A/G

- polymorphism influences the risk of schizophrenia but not bipolar disorder or major depressive disorder: a meta-analysis. *Journal of Neuroscience Research*, 91(5):623-633.
25. Guest PC, Salim K, Skynner HA, George SE, Bresnick JN, McAllister G (2000). Identification and characterization of a truncated variant of the 5-hydroxytryptamine 2A receptor produced by alternative splicing. *Brain Research*, 876(1-2):238-244.
 26. Hayashi-Takagi A, Sawa A, Hayashi-Takagi A, Nakamura M, Kasai H, Hayashi-Takagi A, Araki Y, Haganir RL, Sawa A, Vollrath B, Duron SG, Campbell DA, Yan Z (2014). PAKs inhibitors ameliorate schizophrenia-associated dendritic spine deterioration in vitro and in vivo during late adolescence. *Proceedings of the National Academy of Sciences of the United States of America*, 111(17):6461-6466.
 27. Herson M (2011). *Etiological considerations. Adult psychopathology and diagnosis*. John Wiley & Sons. ISBN 9781118138847
 28. Hill M, Shannahan K, Jasinski S, Macklin EA, Raeke L, Roffman JL, Goff DC (2011). Folate supplementation in schizophrenia: A possible role for MTHFR genotype. *Schizophrenia Research*, 127(1-3):41-45.
 29. Houston I, Peter CJ, Mitchell A, Rogaev E, Akbarian S, Straubhaar J (2013). Epigenetics in the human brain. *Neuropsychopharmacology*, 38(1):183-197.
 30. Howes OD, Kapur S (2009). The dopamine hypothesis of schizophrenia: version III--the final common pathway. *Schizophr Bull*, 35(3):549-62.
 31. Hu CY, Qian ZZ, Gong FF, Lu SS, Feng F, Wu YL, Yang HY, Sun YH (2015). Methylenetetrahydrofolate reductase (MTHFR) polymorphism susceptibility to schizophrenia and bipolar disorder: an updated meta-analysis. *Journal of Neural Transmission*, 122(2):307-320.
 32. Ito C (2004). The role of the central histaminergic system on schizophrenia. *Drug News & Perspectives* 17 (6): 383-7.
 33. Jääskeläinen E, Haapea M, Rautio N, Juola P, Penttilä M, Nordström T, Rissanen I1, Husa A, Keskinen E, Marttila R, Filatova S, Paaso TM, Koivukangas J, Moilanen K, Isohanni M, Miettunen J (2015). Twenty Years of Schizophrenia Research in the Northern Finland Birth Cohort 1966: A Systematic Review. *Schizophr Res Treatment*, 2015:524875.
 34. Joshi D, Fullerton JM, Weickert CS (2014). Elevated ErbB4 mRNA is related to interneuron deficit in prefrontal cortex in schizophrenia. *Journal of Psychiatric Research*, 53(1):125-132.
 35. Kanazawa T, Glatt SJ, Kia-Keating B, Yoneda H, Tsuang MT (2007). Meta-analysis reveals no association of the Val66Met polymorphism of brain-derived neurotrophic factor with either schizophrenia or bipolar disorder. *Psychiatric Genetics*, 17(3):165-170.
 36. Khoddami M, Nadri H, Moradi A, Sakhteman A (2015). Homology modeling, molecular dynamic simulation, and docking based binding site analysis of human dopamine (D4) receptor. *Journal of Molecular Modeling*, 21(2):36
 37. Kret ME, Ploeger A (2015). Emotion processing deficits: A liability spectrum providing insight into comorbidity of mental disorders. *Neuroscience and Biobehavioral Reviews*, 52:153-171.
 38. Kroeger KM, Pflieger KDG, Eidne KA (2003). G-protein coupled receptor oligomerization in neuroendocrine pathways. *Frontiers in Neuroendocrinology*, 24(4):254-278.
 39. Lee KW, Woon PS, Teo YY, Sim K (2012). Genome wide association studies (GWAS) and copy number variation (CNV) studies of the major psychoses: What have we learnt?. *Neuroscience and Biobehavioral Reviews*, 36(1):556-571.
 40. Leonard S, Freedman R (2006). Genetics of Chromosome 15q13-q14 in Schizophrenia. *Biological Psychiatry*, 60(2):115-122.
 41. Liao Q, Wang Y, Cheng J, Dai D, Zhou X, Zhang Y, Gao S, Duan S (2015). DNA methylation patterns of protein coding genes and long noncoding RNAs in female schizophrenic patients. *European Journal of Medical Genetics*, 58(2):95-104.
 42. Lindenmayer JP, Khan A, Lachman H, McGurk SR, Goldring A, Thanju A, Kaushik S (2015). COMT genotype and response to cognitive remediation in schizophrenia. *Schizophr Res*. Oct; 1-2(168):279-84. doi: 10.1016/j.schres.2015.07.037.
 43. Luoni A, Berry A, Calabrese F, Capoccia S, Bellisario V, Gass P, Cirulli F, Riva MA (2014). Delayed BDNF alterations in the prefrontal cortex of rats exposed to prenatal stress: Preventive effect of lurasidone treatment during adolescence. *European Neuropsychopharmacology*, 24(6):986-995.

44. Neddens J, Fish KN, Tricoire L, Vullhorst D, Shamir A, Chung W, Lewis DA, McBain CJ, Buonanno A (2011). Conserved Interneuron-Specific ErbB4 Expression in Frontal Cortex of Rodents, Monkeys, and Humans: Implications for Schizophrenia. *Biological Psychiatry*, 70(7):636-645.
45. Nohesara S, Ghadirivasfi M, Mostafavi S, Eskandari MR, Ahmadkhaniha HR, Thiagalingam S, Abdolmaleky HM (2011). DNA hypomethylation of MB-COMT promoter in the DNA derived from saliva in schizophrenia and bipolar disorder. *Journal of Psychiatric Research*, 45(11):1432-1438.
46. Owen MJ, Sawa A, Mortensen PB (2016). Schizophrenia. *Lancet*. Jan 14. pii: S0140-6736(15)01121-6. doi: 10.1016/S0140-6736(15)01121-6.
47. Paulsen JS, Heaton RK, Sadek JR, Perry W, Delis DC, Braff D, Kuck J, Zisook S, Jeste DV (1995). The nature of learning and memory impairments in schizophrenia. *J Int Neuropsychol Soc*. 1(1):88-99.
48. Pearson-Fuhrhop KM, Minton B, Acevedo D, Cramer SC, Shahbaba B, Cramer SC (2013). Genetic Variation in the Human Brain Dopamine System Influences Motor Learning and Its Modulation by L-Dopa. *PLoS ONE*, 4(8):e61197.
49. Penzes P, Cahill ME, Jones KA, Vanleeuwen J-E, Woolfrey KM, Penzes P (2011). Dendritic spine pathology in neuropsychiatric disorders. *Nature Neuroscience*, 14(3):285-293.
50. Pich EM, Collo G (2015). Pharmacological targeting of dopamine D3 receptors: Possible clinical applications of selective agents. *European Neuropsychopharmacology*, 25(9):1437-1447.
51. Porteous DJ, Thomson PA, Millar JK, Evans KL, Soares DC, Hennah W, McCarthy S, McCombie WR, Clapcote SJ, Korth C, Brandon NJ, Sawa A, Kamiya A, Roder JC, Lawrie SM, McIntosh AM, St Clair D, Blackwood DH (2014). DISC1 as a genetic risk factor for schizophrenia and related major mental illness: Response to Sullivan. *Molecular Psychiatry*, 19(2):141-143.
52. Roffeei SN, Mohamed Z, Reynolds GP, Said MA, Hatim A, Mohamed EH, Aida SA, Zainal NZ (2014). Association of FTO, LEPR and MTHFR gene polymorphisms with metabolic syndrome in schizophrenia patients receiving antipsychotics. *Pharmacogenomics*, 15(4):477-485.
53. Sacchi S, Rosini E, Pollegioni L, Molla G, Sacchi S, Rosini E, Pollegioni L, Molla G (2013). D-Amino acid oxidase inhibitors as a novel class of drugs for schizophrenia therapy. *Current Pharmaceutical Design*, 19(14):2499-2511.
54. Schmidt U, Buell DR2, Ionescu IA2, Gassen NC3, Holsboer F4, Cox MB5, Novak B2, Huber C2, Hartmann J6, Schmidt MV6, Touma C6, Rein T3, Herrmann L2 (2015). A role for synapsin in FKBP51 modulation of stress responsiveness: Convergent evidence from animal and human studies. *Psychoneuroendocrinology*. 52:43-58.
55. Schroeder A, Buret L, Hill RA, van den Buuse M (2015). Gene-environment interaction of reelin and stress in cognitive behaviours in mice: Implications for schizophrenia. *Behavioural Brain Research*, 287:304-314.
56. Sequeira PA, Martin MV, Vawter MP (2012). The first decade and beyond of transcriptional profiling in schizophrenia. *Neurobiology of Disease*, 45(1):23-36.
57. Serper MR1, Alpert M, Richardson NA, Dickson S, Allen MH, Werner A (1995). Clinical effects of recent cocaine use on patients with acute schizophrenia. *Am J Psychiatry*. 152(10):1464-9.
58. Serper MR1, Chou JC, Allen MH, Czobor P, Cancro R (1999). Symptomatic overlap of cocaine intoxication and acute schizophrenia at emergency presentation. *Schizophr Bull*. 25(2):387-94.
59. Stefansson H, Sigurdsson E, Steinthorsdottir V, Bjornsdottir S, Sigmundsson T, Ghosh S, Brynjolfsson J, Gunnarsdottir S, Ivarsson O, Chou TT, Hjaltason O, Birgisdottir B, Jonsson H, Gudnadottir VG, Gudmundsdottir E, Bjornsson A, Ingvarsson B, Ingason A, Sigfusson S, Hardardottir H, Harvey RP, Lai D, Zhou M, Brunner D, Mutel V, Gonzalo A, Lemke G, Sainz J, Johannesson G, Andresson T, Gudbjartsson D, Manolescu A, Frigge ML, Gurney ME, Kong A, Gulcher JR, Petursson H, Stefansson K (2002). Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet*. 71(4):877-92.
60. Strick CA1, Li C, Scott L, Harvey B, Hajós M, Steyn SJ, Piotrowski MA, James LC, Downs JT, Rago B, Becker SL, El-Kattan A, Xu Y, Ganong AH, Tingley FD 3rd, Ramirez AD, Seymour PA, Guanowsky V, Majchrzak MJ, Fox CB, Schmidt CJ, Duplantier AJ (2011). Modulation of NMDA receptor function by inhibition of D-amino acid oxidase in rodent brain. *Neuropharmacology*. 2011 Oct-Nov; 61(5-6):1001-15.
61. Susanne Nikolaus, Hubertus Hautzel1, Hans-Wilhelm Müller (2014). Neurochemical dysfunction in

- treated and nontreated schizophrenia-A retrospective analysis of in vivo imaging studies. *Reviews in the Neurosciences*, 25(1):25-96.
62. Swerdlow NR1, Weber M, Qu Y, Light GA, Braff DL. Realistic expectations of prepulse inhibition in translational models for schizophrenia research. *Psychopharmacology (Berl)*. 2008 Aug;199(3):331-88.
 63. Tenback DE, van Harten PN (2011). Epidemiology and Risk Factors for (Tardive) Dyskinesia. *International Review of Neurobiology*, 98: 211-230.
 64. Thiselton DL, Vladimirov VI, Kuo PH, McClay J, Wormley B, Fanous A, Oeill FA, Walsh D, Van den Oord EJCG, Kendler KS, Riley BP (2008). AKT1 Is Associated with Schizophrenia Across Multiple Symptom Dimensions in the Irish Study of High Density Schizophrenia Families. *Biological Psychiatry*, 63(5):449-457.
 65. Vauquelin G, Bostoen S, Vanderheyden P, Seeman P (2012). Clozapine, atypical antipsychotics, and the benefits of fast-off D 2 dopamine receptor antagonism. *Naunyn-Schmiedeberg's Archives of Pharmacology*, 385(4):337-372.
 66. Villar-Menéndez I, Díaz-Sánchez S, Blanch M, Albasanz JL, Pereira-Veiga T, Monje A, Planchat LM, Ferrer I4, Martín M, Barrachina M (2014). Reduced striatal adenosine A2A receptor levels define a molecular subgroup in schizophrenia. *Journal of Psychiatric Research*, 51(1):49-59.
 67. Wilkinson B, Campbell DB (2013). Contribution of Long Noncoding RNAs to Autism Spectrum Disorder Risk. *International Review of Neurobiology*, 113(0):35-59.
 68. Winchester CL, Pratt JA, Morris BJ (2014). Risk genes for schizophrenia: Translational opportunities for drug discovery. *Pharmacology and Therapeutics*, 143(1):34-50.
 69. Yanagi M, Southcott SA, Shukla AA, Ghose S, Tamminga CA, Joho RH (2014). Kv3.1-containing K⁺ channels are reduced in untreated schizophrenia and normalized with antipsychotic drugs. *Molecular Psychiatry*, 19(5):573-579.
 70. Yin DM, Chen YJ, Lu YS, Bean JC, Sathyamurthy A, Shen C, Liu X, Lin TW, Smith CA, Xiong WC, Mei L (2013). Reversal of Behavioral Deficits and Synaptic Dysfunction in Mice Overexpressing Neuregulin 1. *Neuron*, 78(4):644-657.
 71. Yu Z, Fang Q, Xiao X, Cao H, Gong N, Yu Z, Xiao X, Wang Y-Z, Xu T-L, Fang Q, Chen Z, Cai Y-Q, Fei J, Hu G (2013). GABA Transporter-1 Deficiency Confers Schizophrenia-Like Behavioral Phenotypes. *PLoS ONE*, 8(7). DOI: 10.1371/journal.pone.0069883
 72. Zhang HX, Zhao JP, Lv LX, Li WQ, Xu L, Ouyang X, Yuan ZQ, Huang JS (2008). Explorative study on the expression of neuregulin-1 gene in peripheral blood of schizophrenia. *Neuroscience Letters*, 438(1):1-5.
 73. Zheng W, Wang H, Zeng Z, Lin J, Little PJ, Srivastava LK, Quirion R (2012). The possible role of the Akt signaling pathway in schizophrenia. *Brain Research*, 1470(0):145-158.
 74. Zvara A, Szekeres G, Janka Z, Kelemen JZ, Cimmer C, Sántha M, Puskás LG. (2005). Over-expression of dopamine D2 receptor and inwardly rectifying potassium channel genes in drug-naïve schizophrenic peripheral blood lymphocytes as potential diagnostic markers. *Disease Markers*, 21(2):61-69.