Meta-analysis of transcriptional regulatory networks for lipid metabolism in neural cells from schizophrenia patients based on an open-source intelligence approach

Lisa Okamoto\textsuperscript{a,b,c,1}, Soyoka Watanabe\textsuperscript{a,b,1}, Senka Deno\textsuperscript{a,b,c,1}, Xiang Nie\textsuperscript{a}, Junichi Maruyama\textsuperscript{a}, Masaru Tomita\textsuperscript{b,c}, Atsushi Hatano\textsuperscript{a,d}, Katsuyuki Yugi\textsuperscript{a,b,c,e,f}\textsuperscript{*}

\textsuperscript{a} Laboratory for Integrated Cellular Systems, RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
\textsuperscript{b} Institute for Advanced Biosciences, Keio University, Fujisawa, 252-0882, Japan
\textsuperscript{c} Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa, 252-0882, Japan
\textsuperscript{d} Department of Omics and Systems Biology, Niigata University Graduate School of Medical and Dental Sciences, 757 Ichibancho, Asahimachi-dori, Chuo Ward, Niigata City, 951-8510, Japan
\textsuperscript{e} Department of Biological Sciences, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan
\textsuperscript{f} PRESTO, Japan Science and Technology Agency, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan

\textsuperscript{1} These authors contributed equally to this work.

**Keywords:** Schizophrenia, Transcriptional regulatory network, Enhancer, Lipid metabolism, Meta-analysis, Open-source intelligence

**A B S T R A C T**

There have been a number of reports about the transcriptional regulatory networks in schizophrenia. However, most of these studies were based on a specific transcription factor or a single dataset, an approach that is inadequate to understand the diverse etiology and underlying common characteristics of schizophrenia. Here we reconstructed and compared the transcriptional regulatory network for lipid metabolism enzymes using 15 public transcriptome datasets of neural cells from schizophrenia patients. Since many of the well-known schizophrenia-related SNPs are in enhancers, we reconstructed a network including enhancer-dependent regulation and found that 53.3\% of the total number of edges (7,577 pairs) involved regulation via enhancers. By examining multiple datasets, we found common and unique transcriptional modes of regulation. Furthermore, enrichment analysis of SNPs that were connected with genes in the transcriptional regulatory networks by eQTL suggested an association with hematological cell counts and some other traits/diseases, whose relationship to schizophrenia was either not or insufficiently reported in previous studies. Based on these results, we suggest that in future studies on schizophrenia, information on genotype, comorbidities and hematological cell counts should be included, along with the transcriptome, for a more detailed genetic stratification and mechanistic exploration of schizophrenia.

1. Introduction

Schizophrenia is known to have a high concordance rate between monozygotic twins, with an estimated heritability of 80\%, indicating that genetic factors may play a major role in its onset (Imamura et al., 2020; Sullivan et al., 2003). Recently, a large-scale genome-wide association study (GWAS) (n = 369,899 schizophrenia patients, n = 113,075 controls) conducted by the Schizophrenia Working Group of the Psychiatric Genomics Consortium reported evidence for 108 single nucleotide polymorphisms (SNPs) associated with the onset of schizophrenia (Ripke et al., 2014). These SNPs are found in various locations in the genome and have been linked to a variety of genes, suggesting that schizophrenia is a multifactorial disease.

Schizophrenia has clinical features such as neurological dysfunction (Ni and Chung, 2020; Roberts, 2021), inflammatory reactions (Müller, 2018; Upthegrove and Khandaker, 2020) and abnormal lipid metabolism (Cao et al., 2018; Ghosh et al., 2017; Liu et al., 2021; Tessler et al., 2016), and it has been hypothesized that mitochondrial dysfunction and the kynurenic pathway, which are responsible for tryptophan metabolism, are responsible for these traits.

https://doi.org/10.1016/j.neures.2021.12.006

* Corresponding author at: Laboratory for Integrated Cellular Systems, RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan.

E-mail address: katsuyuki.yugi@riken.jp (K. Yugi).

Received 24 August 2021; Received in revised form 23 December 2021; Accepted 23 December 2021
Available online 31 December 2021

0168-0102/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Lipid metabolism exhibits various characteristics among individuals with schizophrenia. Examples of such abnormalities include structural alterations of lipids in the brain and skin and changes in blood plasma and serum lipids (Shimamoto-Mitsuaya et al., 2021). For instance, multiple cases of derangements of membrane phospholipids and polyunsaturated fatty acids have been reported in the brains of schizophrenia patients (Ghosh et al., 2017; Tessier et al., 2016; Wood et al., 2014; Yao et al., 2000). Other studies have shown that the composition of skin ceramides can also be affected (Smesny et al., 2013). Furthermore, significant differences in phospholipids, such as phosphatidylethanolamine and phosphatidylcholine, and amino acids related to lipid metabolism, such as tryptophan and carnitine, have been found in plasma and serum of schizophrenia patients, making them biomarker candidates (Cao et al., 2019, 2018; Liu et al., 2021; Wang et al., 2019).

Since schizophrenia is considered a multifactorial disease with such a wide variety of symptoms, it is important to elucidate how the genes linked to SNPs are functionally linked together as a system. TCF4, a transcription factor (TF), has been suggested to function as a master regulator of gene network disruption during the early stages of neurodevelopment in schizophrenia (Tornhiel et al., 2019). The TCF4 gene network in neural progenitor cells was found to be enriched in pathways involved in neural activity and in schizophrenia-related risk genes. In a different study that integrated data from ChIP-seq to identify the binding sites of TCF4 in neuro-derived SH-SY5Y cells and findings on its association with schizophrenia, the TCF4 gene set was found to be enriched in Gene Ontology categories such as axon and neurogenesis, genes preferentially expressed in pyramidal cells of the brain and genes whose expression is reduced in brain tissue of schizophrenia patients (Xia et al., 2018). Also, direct regulation by TCF4 was suggested for 13 genes associated with the 108 SNPs. TCF4 was recognized as an important regulator of neural genes and the relationship between TCF4 and schizophrenia is being widely discussed.

Although informative, previous studies have mainly focused on promoter-mediated regulation and did not consider enhancer regulation. It has become clear in recent years that enhancer regulation plays an important role in psychiatric and neurological disorders. A representative previous study showed that the 108 SNPs that were proposed to be associated with schizophrenia were significantly enriched in enhancers activated in the brain and immune system (Ripke et al., 2014). In enhancers, RNAs known as enhancer RNAs (eRNAs) are transcribed bidirectionally and it has been reported that knocking down these eRNAs decreases the expression of target mRNAs (Kim et al., 2010; Murakawa et al., 2016). It is also speculated that eRNAs promote liquid-liquid phase separation of transcriptional complexes, which results in the up-regulation of gene transcription (Henninger et al., 2021). Other studies by Arner et al. (Arner et al., 2015) and Hirabayashi et al. (Hirabayashi et al., 2019) have also reported the relationship between eRNA and mRNA expression. Transcriptome analysis of 537 postmortem brain samples (schizophrenia patients n = 258, controls n = 279) identified 118 eRNAs with different transcription patterns in the schizophrenia group and 927 genetic variants affecting the expression of enhancers (Hauberg et al., 2019). Therefore, it is essential to focus on enhancer-mediated transcriptional regulation to elucidate the underlying mechanisms of gene dysregulation relevant to the pathogenesis of schizophrenia.

Furthermore, the fact that only certain TFs such as TCF4 were analyzed in these studies raises the concern that important factors may be ignored. This limitation can be prevented by considering transcriptional regulation via enhancers as a network to avoid missing important factors. In addition, most large-scale studies about schizophrenia discuss the system based on only one set of data and do not sufficiently compare the system in terms of similarities or differences among schizophrenia populations that display various characteristics such as neurological dysfunction, inflammatory response and lipid metabolism disorders (Myint, 2012; Suzuki et al., 2019).

In this study, we investigated all TFs, both promoters and enhancers, using existing information from GeneHancer and GeneALaCart (Fishilevich et al., 2017; Safran et al., 2010; Stelzer et al., 2016) and performed a meta-analysis on multiple public datasets. For the target genes, we focused on lipid metabolism, which seems to play a common and important role in a wide range of schizophrenia phenotypes, such as psychiatric and neurological disorders and inflammatory responses (Tufano and Pinna, 2020; Varga et al., 2011; Zolezzi et al., 2017). We did not examine networks of specific types of schizophrenia data but rather found networks shared between multiple samples (Barnes et al., 2011; Brennand et al., 2015, 2011; Hoffman et al., 2017; Lin et al., 2016; Narla et al., 2017; Ni et al., 2020a, b; Park et al., 2020; Pietersen et al., 2014; Shao et al., 2019; Topol et al., 2015, 2016). To do so, we followed the open-source intelligence (OSINT) approach (Benes, 2013; Carr, 2016; Kraus, 2008) and used a variety of publicly available information that could potentially provide clues to the missing links between transcriptional regulatory networks and phenotypes of schizophrenia. What we mean here by OSINT is to collect publicly available large-scale data, interpret the data using public databases and eventually characterize molecular mechanisms that underlie biological functions and diseases (Kraus, 2008). In this study, we used as many public datasets as possible to reconstruct differentially regulated networks (DRNs), which are transcriptional regulatory networks consisting of differentially expressed genes (DEGs), taking both enhancers and promoters into account. After the reconstruction of the DRNs, the genes of the DRNs were connected with SNPs by using GTEx eQTL data to address what types of genetic traits are associated with the schizophrenia DRNs. Then, we conducted a meta-analysis of the network structure to compare the underlying transcriptional networks in various schizophrenia populations.

As a result, we found common DRNs and differences among multiple datasets. The SNP enrichment analysis, which included enhancers, suggested an association of the schizophrenia DRNs with terms in hematological cell counts, lung disease, eyes, etc. Thus, we suggest that when collecting samples from schizophrenia patients in the future, information about genotype, comorbidities, and hematological cell counts should be included, along with the transcriptome, for genetic stratification and mechanistic exploration of the disease.

2. Materials and methods

2.1. Step (i): downloading the schizophrenia transcriptome data

The data processing pipeline for this study is summarized in Fig. 1a, b. We obtained all public datasets used in this study from NCBI GEO (Gene Expression Omnibus). The search terms were set to Query: “schizophrenia”[MeSH Terms] OR “Schizophrenia”[All Fields]) AND “Homo sapiens”[porgn] AND “gse”[Filter], resulting in 213 hits. We manually excluded datasets whose study type was non-coding or DNA methylation and 65 datasets remained. From these, we manually selected 27 datasets of case/control studies using neuronal cells derived from iPSC cells of patients or postmortem brains that appeared to fit the purpose of this study. Of these 27 datasets, we excluded datasets that could not be processed, contained non-coding RNA (ncRNA), had redundant contents, lacked measurement of expression levels, had no comparisons between schizophrenia patients and healthy subjects, or included diseases other than schizophrenia.

2.2. Step (ii): identification of DEGs

To identify DEGs from the obtained public datasets, we first checked whether DEGs were already defined in each dataset. For the public datasets that had already set the criteria for DEGs in the related publications, we followed the original criteria in the corresponding publications. As a result, a list of DEGs was obtained from datasets 1–3, 6–8, 12, 13, and 15. For datasets 4, 5, 9–11, and 14, where no DEG lists were available or the number of DEGs was very small, we used DEGseq (Wang
Neuroscience Research 175 (2022) 82–97
84
L. Okamoto et al.

Fig. 1. Data processing pipeline.

(a) Reconstruction of DRNs.
All publicly available transcriptome data were obtained from NCBI GEO [step (i)]. The DEGs in each dataset were identified with the criteria as described in step (ii). The list of lipid metabolism genes was obtained from KEGG, and their TFs (layer 1 TFs) were estimated using GeneHancer via GeneAlaCart query [step (iii) and (iv)]. Layer 2 TFs, which are TFs of the layer 1 TF genes were estimated in the same way. The ‘List of GHIDs and TFBSs corresponding to each DEG’ of layer 2, shown at the bottom of the figure, follows to the top left of (b). Pathway enrichment analysis was performed for all DEGs [step (v)].

(b) Associating the DEGs in the DRNs with eQTL SNPs.
The reconstructions of DRNs were done using the results obtained from the TF estimation [step (vi)]. SNPs associated with TFs were identified via eQTLs from GTEx v8, and the traits/diseases were obtained by enrichment analysis of the SNPs using XGR [step (vii)]. The EFO terms were converted to MeSH terms and publications were searched in PubMed [step (viii)]. The ‘TF estimation results’ shown in the upper left corner of the figure, continues from (a).
et al., 2017) to determine the DEGs (see Data S1 for details). A multiple testing correction was conducted to adjust p-values for false discovery rate (FDR) using the method proposed by Storey and Tibshirani (FDR < 0.05, fold change > 3/2 or < 2/3) (Storey and Tibshirani, 2003).

2.3. Step (iii): conversion to KEGG IDs/identification of lipid metabolism genes

Pathway map information from the KEGG database was used to extract the lipid metabolism genes among the identified DEGs. The names and IDs of all lipid metabolism pathways were obtained from KEGG REST, along with the KEGG IDs of the genes in the pathways to create a lipid enzyme list. We also compiled a table of all gene symbols and their corresponding KEGG IDs from KEGG REST.

2.4. Step (iv): estimating TFs of DEGs related to lipid metabolism

GeneHancer, a database of human enhancers and their inferred target genes, provided in the GeneCards framework was used to identify TFs that potentially regulate each lipid metabolism gene (Fishilevich et al., 2017). The TFs were assessed in the following steps. First, arbitrary genes (lipid metabolism genes identified above) were entered into the Input field of GeneALaCart, a batch querying system for GeneCards, and their corresponding KEGG IDs from KEGG REST were retrieved. Next, the GeneHancer ID (GHID) and Transcription Factor Binding Site (TFBS) corresponding to each gene were extracted from the acquired data. Then, the TFs corresponding to these TFBSs, which are also DEGs, were extracted. After deleting the duplicates of the extracted TFBSs, the gene identifiers for the TFs were submitted to GeneALaCart Query and the same process was repeated again. This procedure was repeated multiple times for each public dataset. The purpose of this step was to determine how many layers of the schizophrenia transcriptional regulatory network should be examined. We expected that if we traced the network from the lipid metabolism genes to chains of upstream TFs involved in the regulation, the appearance of new TF species would saturate at some point. As a result of this procedure, we found that the number of TFBSs was saturated at the second iteration in all datasets, and thus we concluded that we could identify not only TFs that regulate lipid metabolism genes but also genes involved in the regulation of these TFs by repeating the TF estimation procedure twice. Thereafter, the TF estimation procedure was repeated twice for the remaining data, and each TF estimation result was referred to as layer 1 and layer 2, i.e., the layer 1 TFs directly regulate lipid metabolism gene expression, and the layer 2 TFs directly regulate the layer 1 genes. In the end, using the tissue information provided by GeneHancer for each GeneHancer ID, we filtered the TFBSs in layer 1 and layer 2 to include only the TFBSs associated with the brain and nerve tissues (see Data S1 for the list of tissues) for further analysis [step (vi)-(viii)].

2.5. Step (v): KEGG pathway enrichment analysis

We conducted an analysis to identify the KEGG pathways that are significantly enriched in the respective public datasets. First, we obtained non-disease pathway information (KEGG pathway IDs other than the 05000 s, 04900 s, 01500s, and 07000 s) of Homo sapiens from KEGG REST. Next, we determined the total number of DEGs in the pathways of interest. Subsequently, the probability of the number of genes with variable expression levels appearing by chance in the pathway of interest (p-value) was determined using Fisher’s exact test (p < 0.05). A multiple testing correction was performed to adjust the p-values for FDR using the Benjamini-Hochberg method (FDR < 0.1).

This analysis showed an overlap of the following three pathways: Hippo signaling pathway (path:hsa04390, path:hsa04392), Autophagy (path:hsa04140, path:hsa04136), and Drug metabolism (path: hsa00982, path:hsa00983). We chose path:hsa04390 for the Hippo pathway and path:hsa04140 for Autophagy out of the multiple pathways. In addition, Drug metabolism was distinguished by adding ‘cytochrome p450’ (path:hsa00982) and ‘other enzymes’ (path: hsa00983) to the pathway names.

2.6. Step (vi): comparing the differentially regulated networks

Frequencies of the TF-target gene pairs across all datasets were visualized in a heatmap and then subjected to hierarchical clustering based on Euclidean distance and Ward’s minimum variance method (“ward.D2” in R language). Classification of each TF was obtained from PANTHER Protein Class (http://www.pantherdb.org/). All genes in each cluster were subjected to enrichment analysis in terms of KEGG pathways or TF classes.

2.7. Step (vii): identification and enrichment analysis of eQTL SNPs

We next identified eQTL SNPs associated with the DRNs for lipid metabolism enzymes reconstructed in the previous section. The subsequent SNP identification and enrichment analyses were performed on the TFs identified in layers 1 and 2 but not on the target genes because it was expected that inclusion of the target genes might provide nothing other than an overrepresentation of lipid metabolism-related traits. The eQTLs were identified using the GTEx v8 cis-eQTL in GTEx Portal (https://gtexportal.org/home/). The SNPs were identified in 13 tissues beginning with ‘Brain’ (Brain Amygdala, Brain Anterior cingulate cortex BA24, Brain Caudate basal ganglia, Brain Cerebellum Hemispheres, Brain Cerebellum, Brain Cortex, Brain Frontal Cortex BA9, Brain Hippocampus, Brain Hypothalamus, Brain Nucleus accumbens basal ganglia, Brain Putamen basal ganglia, Brain Spinal cord cervical c-1, and Brain Substantia nigra). We used statistically significant eQTLs detected in the 13 tissues. Enrichment analysis of the trait/disease annotations for eQTL SNPs associated with the DRN was performed using XGR (Xploring Genomic Relations) (Fang et al., 2016) for each dataset. Conditions of XGR were set as follows: [ontology : Experimental Factor Ontology; LD r^2 > 0.8; the size of members of each term in consideration : min 3, max 2000; the minimum number of overlaps : 2; test : hypergeometric; the tail used to calculate p-values : one tail; adjust p-values : BH (Benjamini-Hochberg)]. The XGR uses EFO (Experimental Factor Ontology) developed by EMBL-EBI (Malone et al., 2010) as a controlled vocabulary for trait/disease names to handle notation variability. Subsequent functional annotations obtained from enrichment analysis of trait and disease names are based on EFO terms of FDR < 0.05.

2.8. Step (viii): bibliographic analysis

We systematically searched the PubMed database to determine if overrepresented traits/diseases were studied in association with schizophrenia. EFO terms that appeared in the public datasets were replaced with MeSH terms by searching the EMBL-EBI Ontology Lookup Service and the NCBI MeSH database. When the original EFO term did not appear, it was paraphrased to an analogical term and then replaced with a MeSH term. The MeSH terms were searched in PubMed on 10th November 2021 by: (MeSH term + “ANd (schizophrenia[MESH] OR schizophrenia[WORD]) AND English[FILT] AND English journal article[FILT] NOT review[FILT]”).

2.9. Source code availability

Source codes for the data processing pipeline are available online (https://github.com/YugiLab/Publications/tree/main/Okamoto2021).

3. Results

3.1. Statistics of the schizophrenia transcriptome data

We selected 15 publicly available transcriptome datasets that match...
the subjects of this study from the downloaded data. See Table 1 for the number designations for each dataset included in this study (e.g. dataset 1), GEO accession numbers, references, and phenotypes of the selected datasets. The other downloaded datasets are presented in Table 2, along with reasons for their exclusion from this study.

| Table 1 | Datasets used in the analysis. |
|---------|--------------------------------|
| No.     | Accession No. | References | Sample origin | Method |
| 1       | GSE125999     | Ni et al. (2020a, b) | Caucasian male patients requiring treatment by clozapine | RNA-seq |
| 2       | GSE92874      | Narla et al. (2017) | Caucasian men and women with a history of symptoms and/or hospitalization. Including some with family history. | RNA-seq |
| 3       | GSE40102      | Brennand et al. (2015) | Caucasian men and women with description of symptoms. | Microarray |
| 4       | GSE125805     | Ni et al. (2020a, b) | Caucasian male patients requiring treatment by clozapine. | RNA-seq |
| 5       | GSE121376     | Shao et al. (2019) | Caucasian male patients requiring treatment by clozapine. | RNA-seq |
| 6       | GSE46562      | Lin et al. (2016) | SZ and SAD patients with 22q11.2 deletion | RNA-seq |
| 7       | GSE53738      | Topol et al. (2015, 2016) | Caucasian men and women with descriptions of hospitalization and family history. | RNA-seq |
| 8       | GSE63734      | Topol et al. (2016) | Caucasian males and females, mainly COS patients. Includes patients with family history and gene deletions. | RNA-seq |
| 9       | GSE106589     | Hoffman et al. (2017) | COS patients with a description of family history and gene deletions. | RNA-seq |
| 10      | GSE152438     | Park et al. (2020) | Caucasian male patients requiring treatment by clozapine. | RNA-seq |
| 11      | GSE152438     | Park et al. (2020) | Caucasian male patients requiring treatment by clozapine. | RNA-seq |
| 12      | GSE46509      | Pietersen et al. (2014) | Postmortem brains of men and women who had been treated by drugs. | Microarray |
| 13      | GSE21935      | Barnes et al. (2011) | Postmortem brain samples collected from elderly patients with significant symptoms diagnosed according to DSM-III criteria. | Microarray |
| 14      | GSE25673      | Brennand et al. (2011) | Four Caucasian men and women with a high probability of having heritable factors behind the disease. | Microarray |
| 15      | GSE118313     | Shao et al. (2019) | Caucasian male patients requiring treatment by clozapine. | RNA-seq |

Out of the 27 datasets obtained from NCBI GEO, 15 were selected for analyses in this study. In all cases, we selected studies that compared cases between schizophrenia patients and healthy individuals. The accession No., reference and sample origin corresponding to each data are shown in the table. Datasets 10 and 11 are obtained from different sample populations in the same study. Specifically, the population of dataset 10 is healthy control vs. schizophrenia without inflammatory stimuli [GSE152438 (GSM4615614-GSM4615621)], and the population of dataset 11 is healthy control vs. schizophrenia without IL-17 stimulus [GSE152943 (GSM4615631-GSM4615637)].

Table 2 | Datasets excluded from the analysis. |
|---------|-------------------------------------|
| No.     | Accession No. | References | Reasons for exclusion |
| 16      | GSE143666     | Hauberg et al. (2020) | This study does not contain comparisons between SZ patients and healthy controls. |
| 17      | GSE48827      | Kiltchevskiy et al. (2020) | This study does not contain comparisons between SZ patients and healthy controls. |
| 18      | GSE76922      | Zhang et al. (2018) | This data includes diseases other than SZ. |
| 19      | GSE92876      | Narla et al. (2017) | mRNA expression levels were not measured. |
| 20      | GSE92873      | Narla et al. (2017) | mRNA expression levels were not measured. |
| 21      | GSE59978      | Chen et al. (2013, 2018) | Includes datasets 22 and 23 |
| 22      | GSE5977       | Chen et al. (2013, 2018) | This data includes diseases other than SZ. |
| 23      | GSE5974       | Chen et al. (2013) | This data includes diseases other than SZ. |
| 24      | GSE30737      | mRNA expression levels were not measured. |
| 25      | GSE27356      | Narayan et al. (2008) | Unable to process data. |
| 26      | GSE6629       | Pedrosa et al. (2011) | This data does not contain detailed numerical data about mRNA expression levels. |
| 27      | GSE410843     | Citation missing | This study does not contain comparisons between SZ patients and healthy controls. |
| 28      | GSE132698     | Park et al. (2020) | This study uses cortical interneurons under inflammation conditions. |

Out of the 27 datasets obtained from NCBI GEO, the following 12 datasets were excluded from the analysis. The reasons for these exclusions were, inability to process the files, inclusion of ncRNAs, overlap of the contents, absence of expression assays, lack of comparison between schizophrenia patients and healthy subjects, and inclusion of diseases other than schizophrenia in the study. The corresponding accession No., reference and reason for exclusion of each dataset are shown in the table.

data points were considered in subsequent data analysis. First, we identified 9,062 DEGs (counted by gene symbols) that exhibited significant changes in schizophrenia patients (see Data S1 for details) out of 491,917 genes (counted by gene symbols) without missing data points in total (Table 3). Lipid metabolism genes were then extracted from these DEGs (see Table 3 for KEGG IDs of these enzymes). Lists of all identified DEGs are available in Data S1.

3.2. Estimating TFs of DEGs related to lipid metabolism

The numbers of target genes (i.e. lipid metabolism genes) in the DRNs differed across the datasets. The fraction of the number of lipid metabolism genes (= target genes) in all the DEGs was 1.2 % on average (Fig. 2a). Reconstruction of the DRNs provided TFs and their modes of regulation (Promoter, Enhancer or Promoter/Enhancer) of the target genes. For all the datasets, any form of transcriptional regulation via enhancers was in the majority (Fig. 2b, c). Note that the number of TF was zero in datasets 8, 13, and 15 since no lipid metabolism genes were identified in their DEGs.

3.3. KEGG pathway enrichment analysis

Functional enrichment analysis of non-disease pathways in each public dataset detected enrichments in pathways associated with the nervous system, cell-cell adhesion, protein and amino acid, immunity and inflammation, lipid metabolism, and hematolgy (Fig. S1).

Neurological pathways were significantly enriched in datasets 2, 3, 5, 7, 9–11, and 14 at p < 0.05 and in datasets 2, 3, 11, and 14 at FDR < 0.1. In these data sets, at least one of the following pathways was found...
to be enriched: ‘Glutamatergic synapse’, ‘Dopaminergic synapse’, ‘GABAergic synapse’, ‘Cholinergic synapse’, ‘Synaptic vesicle cycle’, ‘Axon guidance’, and ‘Gap junction’. For pathways related to cell-cell adhesion relationships, enrichments of ‘Gap junction’, ‘Focal adhesion’, ‘Adherens junction’, ‘Tight junction’, ‘Glycosylphosphatidylinositol (GPI)-anchor biosynthesis’ and ‘Cell adhesion molecules’ were observed in datasets 1–7, 9–11, and 14 at p < 0.05 and in datasets 2–5, 9–11, and 14 at FDR < 0.1. The ‘ECM-receptor interaction’ pathway was found to be enriched at a significantly high frequency in datasets 2–5, 7, 9–11, 14 at p < 0.05 and in datasets 2, 4, 5, 9, 11, 14 at FDR < 0.05. The extracellular matrix (ECM) receptor is known to be associated with various functions in the brain, such as neuronal differentiation and migration, axon elongation, myelination, and regulation of synaptogenesis and synaptic plasticity.

Lipid metabolism-related pathways appeared significantly more frequently in datasets 2 and 12 at p < 0.05, but not in any datasets at FDR < 0.1. In dataset 2, ‘Steroid biosynthesis’, and in dataset 12, ‘Steroid hormone biosynthesis’ were found to be enriched. For immune responses and inflammation-related pathways, ‘Th1 and Th2 cell differentiation’, ‘Antigen processing and presentation’, ‘TGF-β signaling pathway’ and ‘JAK-STAT signaling pathway’ were found to be enriched in datasets 5–7, 11, 12 at p < 0.05 and datasets 7 and 11 at FDR < 0.1. Other blood-related pathways such as, ‘Porphyria metabolism’, ‘Hematopoietic cell lineage’, ‘Vascular smooth muscle contraction’, and ‘Complement and coagulation cascades’ also stood out. Pathways related to protein and amino acid metabolism were also found enriched at a significantly higher frequency, including ‘Tyrosine metabolism’, ‘Protein digestion and absorption’, ‘Tryptophan metabolism’, ‘Alanine, aspartate and glutamate metabolism’, ‘Arginine and proline metabolism’, and ‘D-Glutamine and d-glutamate metabolism’.

3.4. Comparing the differentially regulated networks

In all datasets, the number of TF species in the DRNs were saturated at layer 2, which is the layer of TFs directly regulating layer 1 genes, i.e. the TFs that directly regulate the lipid metabolism enzymes (Fig. 5a). Based on this result, we reconstructed the DRNs up to layer 2 in the subsequent analysis because the inclusion of TFs that belong to layers over layer 3 elaborates the DRNs just by repeating TF autoregulatory loops. Connectivity of the DRNs corresponding to each dataset is available in Fig. S2 and Data S2.

There was a total of 14,215 edges representing physical interactions between TFs and cis-elements of target genes, of which 7,577 were interactions via enhancers and 6,638 were via promoters. Of the total number of edges, 53.3 % were found to be interactions via enhancers.

Hierarchical clustering based on frequencies of the TF-target gene pairs across all datasets provided six clusters for the TFs (a–f, Fig. 2b) and seven clusters for the target genes (1–7, Fig. 3b). Each cluster had one or more unique datasets that characterized the cluster itself by providing TFs and target genes that were not observed in other clusters. The intersection of these clusters revealed DRNs that commonly appeared in multiple datasets. The DRNs of blocks a1, a2, b1, and b2 in Fig. 2b consisted mainly of dataset 2. Block a1 was overlapping with dataset 11. TFs of block b2 were seen partly in datasets 1, 6, and 12. Target genes of block b2 were also seen partly in dataset 7. These findings imply the existence of a transcriptional regulatory module composed of common TFs and target genes that are shared in datasets 1, 2, 6, 7, 11, and 12. Block a3, c1, and c3 were mainly composed of dataset 11. Block c3 was partly overlapping with those of dataset 10. Block d4 was mainly composed of dataset 1. TFs of block d4 were partly seen in dataset 6. Block e6 was mainly composed of dataset 12. Block f5 includes TFs and target genes from dataset 6. Block f7 includes TFs and target genes from datasets 3, 7, 10, and 14. Dataset 2 included more connections between TFs than the other datasets, such as dataset 11, and occupied a larger area in the heat map.

The most frequently found TF-target gene pair was autoregulation of TFs, such as LRF1 (found in six datasets), FOXC2 and ZIP2 (found in five datasets), AFTF, EMT2, FOS, PBX2, and SCRT2 (found in four datasets), and EBF1, GATA2, KLF6, NFIA/B, REXPR, SOX6, and TCF7L2 (found in three datasets). These TFs regulate target genes via both enhancers and promoters in the DRNs.

Although there were no clusters whose TF classes are significantly biased, some TF classes exhibited raw p-values less than 0.05 in enrichment analysis; Cluster c: ‘DNA-binding transcription factor’ (PC00218) and ‘Basic helix-loop-helix transcription factor’ (PC00055), Cluster d: ‘protein-binding activity modulator’ (PC00095), and Cluster f: ‘general transcription factor’ (PC00259) (Fig. 3c).

Similarly, there were no clusters with significantly overrepresented target gene pathways, however, some pathways had raw p-values of less than 0.05 in the enrichment analysis; four pathways in Cluster 1, 10 pathways in Cluster 2, four cancer-related pathways in Cluster 3, one pathway in Cluster 4, two pathways in Cluster 5, two pathways in Cluster 6, and three pathways in Cluster 7. These included pathways related to spinocerebellar ataxia, cancer, infectious disease, amino acid, and central carbon metabolism (Data S3).
3.5. Identification and enrichment analysis of eQTL SNPs

The number of SNPs associated with GTEx v8 eQTLs was approximately 1.8 times larger than the number of DEGs identified from the public datasets, though the number differed between each dataset (Fig. 4a and Data S4). The enrichment analysis of the eQTL SNPs associated with the DRN of each dataset revealed significant over-representation (FDR < 0.05) of the SNPs annotated with EFO terms related to the immune system, nervous system, eyes, skin, etc. (Figs. 4b and S3). In datasets 3, 7, 9, and 14, no EFO terms were significantly enriched. The immune response-related terms were enriched in datasets 1, 2, 4, 5, 10–12 such as ‘immune system disease’ (in datasets 1, 2, 4, 5, 10, and 11) and ‘hematological measurement’ (datasets 4, 5, 10–12). Among the terms under ‘nervous system disease’, the term ‘schizophrenia’ appeared in datasets 1 and 6. The other terms that appeared frequently under ‘nervous system disease’ were ‘mood disorder’ (datasets 1 and 6), ‘autism spectrum disorder’ (datasets 1, 4–6, and 10), and ‘unipolar depression’ (datasets 1 and 6). It was also found that some datasets were enriched with eye-related terms such as ‘eye disease’ (datasets 4, 5, 10, and 11) and ‘eye measurement’ (datasets 4–6, 10, and 11). For other diseases and traits, we observed significant enrichment of EFO terms related to the respiratory system, body size, coronary artery disease, etc.: ‘lung disease’ (datasets 4, 5, 10, and 11), ‘smoking behaviour (status) measurement’ (datasets 1, 4, 5, and 10), ‘BMI-adjusted waist-hip ratio’ (datasets 1, 4, 5, 10, and 11), ‘waist-hip ratio’ (datasets 4, 5, 10, and 11), ‘body ratio measurement’ (datasets 4, 5, 10, and 11), ‘coronary heart disease’ (datasets 4, 5, and
Fig. 3. Differentially regulated networks.

(a) The number of TFBSs and their layers resulting from TF estimation.

Transition in the number of TFBSs when TF estimation was repeated multiple times. The horizontal axis shows the number of the TF layers (e.g., layer 1, layer 2, …), and the vertical axis shows the number of TFBSs that were estimated. The dotted line drawn in the y-axis direction indicates the point where the numbers of TFBSs are saturated. The number of TFBSs and TF species in the DRNs were saturated in layer 2 in all datasets. Considering this result, we decided to examine the DRNs up to layer 2 in the subsequent analysis.

(b) DRNs reconstructed from all public datasets.

Heatmap representing the presence or absence of binding of each TF and target gene combination. Hierarchical clustering based on the frequency of TF-target gene pairs in all datasets provided seven clusters for target genes and six clusters for TFs. The areas delimited by black lines represent each cluster. Red labels for the target genes indicate lipid metabolism genes. The points where the clusters were divided are indicated by the green lines. Six clusters (a–f) of TFs were formed and labeled on the x-axis in black from left to right. Seven clusters (1–7) of target genes were formed and numbered from the bottom right y-axis.

(c) Classes of the TFs.

Functional classes of the TF-target gene pairs in each cluster are presented in different colors according to the class to which the TF belongs. The symbol of each cluster (a–f) corresponds to the six clusters (a–f) for the TFs in (b). C2H2 zinc finger transcription factors were dominant in most of the clusters.

10), and lipid-related EFO terms (dataset 11) such as ‘triglyceride measurement’, ‘lipoprotein measurement’, and ‘lipid or lipoprotein measurement’. See Data S5 for lists of the DEGs associated with the EFO terms.

3.6. Bibliographic analysis

The PubMed searches with schizophrenia and MeSH terms, which correspond to the overrepresented EFO terms, provided small numbers (<100) of original research articles related to blood cell counts such as eosinophil and basophil counts, whereas these terms were frequently associated with the SNPs connected with the DRNs (Fig. 4c). Also, diseases such as coronary artery disease, systemic lupus erythematosus (SLE) and Sjogren syndrome appeared in most datasets with fewer hits by MeSH term in PubMed. Terms related to the clinical picture of schizophrenia were associated with most of the datasets, and more than hundreds of papers were found: smoking-related terms, 1878, ‘drug-induced agranulocytosis’, 341, and ‘lung disease’, 346. Also, terms related to brain disorders were associated with most of the datasets, and hundreds of papers were found: ‘autism spectrum disorder’, 1517 and ‘multiple sclerosis’, 313. The ocular terms appeared in most of the datasets but the numbers of papers related to schizophrenia were variable, ‘eye disease’, 433, ‘eye measurement’, 27 and ‘intraocular pressure measurement’, five.

4. Discussion

We conducted a meta-analysis of 15 publicly available transcriptome datasets to characterize diverse etiologies and underlying common mechanisms of schizophrenia that cannot be revealed by studies based on a single dataset. Reconstruction of the DRNs for lipid metabolism enzymes revealed that promoters and enhancers mediate distinct TF-target gene pairs and that enhancer-mediated regulation covers 7,577 enzymes revealed that promoters and enhancers mediate distinct TF-target gene pairs in all datasets provided seven clusters for target genes and six clusters for TFs. The areas delimited by black lines represent each cluster. Red labels for the target genes mediate lipid metabolism genes. The points where the clusters were divided are indicated by the green lines. Six clusters (a–f) of TFs were formed and labeled on the x-axis in black from left to right. Seven clusters (1–7) of target genes were formed and numbered from the bottom right y-axis.

(c) Classes of the TFs.

Functional classes of the TF-target gene pairs in each cluster are presented in different colors according to the class to which the TF belongs. The symbol of each cluster (a–f) corresponds to the six clusters (a–f) for the TFs in (b). C2H2 zinc finger transcription factors were dominant in most of the clusters.

10), and lipid-related EFO terms (dataset 11) such as ‘triglyceride measurement’, ‘lipoprotein measurement’, and ‘lipid or lipoprotein measurement’. See Data S5 for lists of the DEGs associated with the EFO terms.

3.6. Bibliographic analysis

The PubMed searches with schizophrenia and MeSH terms, which correspond to the overrepresented EFO terms, provided small numbers (<100) of original research articles related to blood cell counts such as eosinophil and basophil counts, whereas these terms were frequently associated with the SNPs connected with the DRNs (Fig. 4c). Also, diseases such as coronary artery disease, systemic lupus erythematosus (SLE) and Sjogren syndrome appeared in most datasets with fewer hits by MeSH term in PubMed. Terms related to the clinical picture of schizophrenia were associated with most of the datasets, and more than hundreds of papers were found: smoking-related terms, 1878, ‘drug-induced agranulocytosis’, 341, and ‘lung disease’, 346. Also, terms related to brain disorders were associated with most of the datasets, and hundreds of papers were found: ‘autism spectrum disorder’, 1517 and ‘multiple sclerosis’, 313. The ocular terms appeared in most of the datasets but the numbers of papers related to schizophrenia were variable, ‘eye disease’, 433, ‘eye measurement’, 27 and ‘intraocular pressure measurement’, five.

4. Discussion

We conducted a meta-analysis of 15 publicly available transcriptome datasets to characterize diverse etiologies and underlying common mechanisms of schizophrenia that cannot be revealed by studies based on a single dataset. Reconstruction of the DRNs for lipid metabolism enzymes revealed that promoters and enhancers mediate distinct TF-target gene pairs and that enhancer-mediated regulation covers 7,577 edges of the DRNs out of 14,215 (53.3 %). Comparison of the DRNs provided networks that are commonly utilized among the datasets, as well as differences among them. Clustering the frequencies of the TF-target pairs elucidated biases in the common DRNs, which might be caused by sample preparation, measurement methods and environment.

4.1. KEGG pathway enrichment analysis

Enrichment analysis using the KEGG pathway map showed that pathways related to the nervous system, cell-cell adhesion, protein and amino-acid, immunology and inflammation, lipid metabolism, and hematology were significantly overrepresented. Out of the 15 public datasets analyzed in this study, two of them showed enrichment of lipid metabolism-related pathways. This implies that abnormal lipid metabolism is not a common feature, at least at the level of gene expression in brain tissue and iPS-derived neurons.

The ‘Gap junction’, which showed enrichment in datasets 2, 5, 9, 10, and 11, is a pathway relevant to the blood-brain barrier (BBB), which has recently been suggested to be involved with psychiatric disorders. Studies showing a relationship between schizophrenia and the BBB include research using the 22q11.2 deletion mouse model (Crockett et al., 2021). In this model, the BBB was disrupted and there was hyper inflammation and microglia/astrocyte activation. The ‘ECM-receptor interaction’ pathway, which is enriched at a significantly high frequency in datasets 2–5, 7, 9–11, and 14, is also expected to affect the BBB. ECM proteins and receptors are important molecules that support cell adaption to environmental changes and regulate the growth and motility of cells (Lukashev and Werb, 1998). The composition of the ECM alters during disruption of the BBB and has been shown to directly influence the progression of neurological diseases (Baeten and Akassoglou, 2011). Furthermore, ‘Adherens junction’ has been suggested to be a factor related to both neurological disorders and intellectual disability. According to a previous study, human iPSC-derived neural progenitor cells with a 15q11.2 microdeletion, which is a risk factor for schizophrenia and autism, show deficiencies in the adherens junction and apical polarity (Yoon et al., 2014). Since inhibiting the function of cadherin and catenin in the developing nervous system causes disruption of the adherens junction and impairs the structural formation of the cerebral cortex (Kazdowski et al., 2007), it is possible that abnormal cell adhesion can lead to the decline of brain functions in schizophrenia patients.

The metabolic pathway map for tryptophan was observed among the amino acid metabolic pathways for which enrichment was observed in the public datasets. Tryptophan is a precursor of serotonin that is degraded by the kynurenine metabolic pathway. Significant increases in kynurenine aminotransferase mRNA and kynurenine pathway metabolites have been reported in the brain and plasma of schizophrenic disease groups, particularly those with higher cytokine levels (Kindler et al., 2020). Also, metabolic pathway maps related to candidate biomarkers of schizophrenia, such as s-glutamate, were observed (Cao et al., 2019). In particular, the degradation of D-type amino acids has been suggested to contribute to the degradation of NMDA receptor function in schizophrenia (Verrall et al., 2010). Research on D-amino acid oxidase (DAAO) has shown that the DAAO activity is doubled in patients with schizophrenia compared to controls (Madeira et al., 2008). Increased D-amino acid degradation due to DAAO activity leads to reduced D-serine levels, which serves as an intrinsic co-agonist of NMDA receptors in the nervous system (Madeira et al., 2008). Based on this evidence, we can speculate that the abnormalities in the nervous system associated with schizophrenia are reflected as alterations in amino acid metabolism. Enrichment of the ‘Protein digestion and absorption’ pathway was also found in nine datasets out of 15 (datasets 1-5, 7, 9-11). There is a report that protein fermentation in the gut correlates with the severity of psychiatric symptoms in schizophrenia (Kang et al., 2021). There have been other studies that found changes in the gut microbiota in schizophrenia patients, which could indicate that the state of proteins in the intestinal tract may affect the gut microbiota (Dinan et al., 2014). Accordingly, the enrichment of the ‘Protein digestion and absorption’ pathway observed in many datasets suggests the involvement of brain-gut correlation in schizophrenia.

L. Okamoto et al.

Neuroscience Research 175 (2022) 82–97
Fig. 4. SNPs associated with the DEGs.
(a) The number of DEGs, and SNPs associated with DEGs via GTEx v8 eQTL.
The number of DEGs identified from public datasets and the number of SNPs associated with the DEGs are shown. The numbers on the vertical axis correspond to the public dataset numbering shown in Table 1. Cyan: DEG and red: SNP. There were on average approximately three times more SNPs than DEGs, although the numbers differed greatly among each dataset.
(b) Results of the SNP Enrichment analysis.
The results of the SNP enrichment analysis identified from DEGs that belong to the DRNs are shown in the heat map. The vertical axis shows the representative upper level EFO terms enriched as the result of the analysis, and the numbers on the horizontal axis correspond to the numbering of the public datasets shown in Table 1. Enrichment of EFO terms related to psychiatric and neurological diseases was observed for datasets 1, 4–6, and 10. In addition, terms related to immunity, respiratory, eye, and skin were overrepresented in multiple datasets.
(c) Overrepresented EFO terms and their appearance in the literature.
The number of times the EFO terms appeared in the public datasets (horizontal axis) and the number of original papers (vertical axis) that were found in PubMed when the EFO terms or their related terms were used for MeSH terms and searched with schizophrenia. Blue dots indicate EFO terms that appeared in three or more databases and had more than 100 original papers. Blue dots include EFO terms related to brain disorders such as ‘autism spectrum disorder’ and ‘multiple sclerosis’. In addition, inflammation-related terms such as ‘inflammatory biomarker measurement’ and ‘abnormality of the immune system’ were also included. Red dots indicate EFO terms that appeared in three or more datasets with fewer than 100 original papers. The red dots included terms related to blood cells, such as ‘basophil count’ and ‘eosinophil count’, and diseases such as ‘coronary artery disease’, ‘systemic lupus erythematosus’ and ‘Sjogren syndrome’.
In dataset 12, we found enrichment of the ‘Th1 and Th2 cell differentiation pathway’. The levels of proinflammatory markers, such as cytokines, are found to increase in both blood and the brain of patients with schizophrenia, while some individuals display mild systemic inflammation (Müller, 2018; Upthagrove and Khandaker, 2020). Moreover, several studies on the Th1/Th2 cytokine balance in psychiatric disorders have been conducted over the years. Based on this, it has been hypothesized that schizophrenia may be associated with an imbalance in these cytokines, resulting in a relative dominance of Th2 immunity in schizophrenia (Augustin et al., 2005; Schwarz et al., 2001a, 2001b; Teixeira et al., 2008). Also, the TGF-β signaling pathway, which inhibits Th1 differentiation and activity (Gorelik et al., 2002), showed enrichment in datasets 5, 7, and 11. TGF-β is known to be increased in the blood and in blood cells of schizophrenia patients at both protein and mRNA levels, respectively, which may also be linked to disruption of the Th1/Th2 cytokine balance (Amoli et al., 2019; Borovcanin et al., 2012; Kim et al., 2004). Note, however that there have been results that do not support this hypothesis, and no consistent conclusion has yet been reached (Potvin et al., 2005).

The Hippo signaling pathway was also found enriched in multiple datasets. In recent years, this signaling pathway has been shown to play an important role in development of the nervous system, such as proliferation of neural stem cells, differentiation of NSCs into neurons, synapse formation, and brain development, as well as controlling immunity (Cheng et al., 2020; Sahu and Mondal, 2021). It has been found that the Hippo pathway is involved in neurodegenerative disorders and it has been suggested that it may also contribute to other neurodegenerative diseases and neuropsychiatric disorders (Cheng et al., 2020; Sahu and Mondal, 2021; Stepan et al., 2018). A study reported that genes in the Hippo signaling pathway were included amongst the genes that appeared commonly linked to schizophrenia, autism, bipolar disorder, and obsessive-compulsive disorder (O’Connell et al., 2018).

To summarize, pathway analysis based only on DEGs of schizophrenia patients, without including transcription factors, showed significant enrichment of pathways (e.g. nervous system, cell-cell adhesion, protein and amino-acid, etc.) that support the characteristics of schizophrenia. In addition, several pathways were found to be associated with the BBB. This result provides a rationale to explore the involvement of the BBB in understanding the pathogenesis of schizophrenia.

4.2. Comparing the differentially regulated networks

Counting the frequency of TF-target gene pairs across all datasets revealed that there are DRNs that commonly appear in multiple datasets and, at the same time, there are DRNs specifically preferred by particular datasets (Fig. 3b). We also found that more than half of the edges in the DRNs (53.3%, 7,577 out of 14,215) involve regulation via enhancers (datasets (Fig. 3b). We also found that more than half of the edges in the DRNs (53.3%, 7,577 out of 14,215) involve regulation via enhancers (Fig. 3c). The bias could be caused by the variation of numbers of DEGs among the datasets that are possibly rooted in differences in sample preparation protocols, measurement methods, environmental factors around the individuals, etc. For example, the DRNs based on datasets from similar sample origins (e.g. datasets 3, 10, and 14 in block f7) tended to be gathered in the same blocks or overlapped with each other. To explore the origin of such dataset-specific biases in future studies, measuring the genotypes of patients together with their transcriptional networks may be helpful to discern whether the biases are genetic or environmental.

4.3. Identification and enrichment analysis of eQTL SNPs/bibliographic analysis

In datasets 4, 5, 10–12, the term ‘hematological measurement’ was significantly enriched, including ‘erythrocyte measurement’ and its subterms, ‘eosinophil count’, ‘basophil count’, and ‘granulocytosis’. There is a report of a patient with schizophrenia who developed acute myelogenous leukemia (AML) and was treated by bone marrow transplantation, resulting in remission of the schizophrenia as well (Miyakoa et al., 2017). Furthermore, several GWAS SNPs that are significantly associated with leukocyte counts, and monocytes were found to be significantly increased in the blood of schizophrenia patients compared to controls (Jackson and Miller, 2020; Steiner et al., 2020a). However, no consistent pattern has yet been identified for the changes in eosinophils and basophils. Since eosinophils release Th1-type pro-inflammatory cytokines (IL-6, IFN-gamma, TNF-alpha, etc.), this may suggest a relationship to the inflammation observed in schizophrenia (Spencer et al., 2009). High levels of eosinophils are also detected in autoimmune diseases such as atopic dermatitis and asthma, and it is known that the heart, lungs, skin, and nervous system can be damaged by the increased eosinophils (Diny et al., 2017). Asthma due to increased eosinophils (Lee et al., 2021) and pneumonia (eosinophilic pneumonia) are examples of such diseases. Based on this evidence, we can hypothesize that schizophrenia may be a form of autoimmune disease. As a matter of fact, several terms indicating autoimmune diseases appeared in the enrichment analysis of SNPs; i.e. ‘systemic lupus erythematosus’ (datasets 4, 5, 10, and 11) and ‘ulcerative colitis’ (datasets 2 and 4). It has been argued that exposure to autoimmune diseases raises the risk of developing schizophrenia (Bennros et al., 2011). It has also been reported that patients with autoimmune encephalitis exhibit negative symptoms and cognitive dysfunction similar to what is seen in schizophrenia (Hao et al., 2017; Steiner et al., 2020b). These observations support the possibility that inflammation, such as that caused by autoimmune diseases, may be one of the causes of schizophrenia.

In datasets 1, 4, and 10, the terms ‘lung disease’ and ‘smoking behaviour (status) measurement’ were enriched, respectively. It is known that people with schizophrenia have a higher risk of lung cancer morbidity and mortality than those without the disease (Ni et al., 2019;
In a GWAS involving more than 220,000 individuals, three loci were found to be shared by schizophrenia and lung cancer (Zuber et al., 2018). The locus with the strongest association included the nicotinic acetylcholine receptor, which has been established as a multifaceted locus shared between lung cancer and smoking behavior. The other two remaining loci showed no genetic association with smoking. Moreover, studies have shown that impaired cardio-respiratory fitness is associated with cognitive dysfunction (Filik et al., 2006; Holmen et al., 2018) and, in fact, patients with schizophrenia have a higher prevalence of angina and respiratory symptoms, and often have impaired lung function. It is also believed that there are genetic factors that increase the risk of both becoming a smoker and developing schizophrenia (Hartz et al., 2018).

Interestingly, the enrichment of terms related to the eyes were conspicuous in the results, including ‘eye disease’ in datasets 4, 5, 10, and 11, and ‘eye measurement’ in datasets 4-6, 10, and 11. Studies have shown that there are multiple structural and physiological disorders in the eyes of individuals with schizophrenia. More specifically, patients can exhibit retinopathy, cataracts, decreased visual acuity, strabismus, increased retinal vein width, loss of ganglion cell axons, abnormal dopaminergic activity in photoreceptors, and decreased GABA- and glycine-related lateral inhibition (Jurisic et al., 2020; Silverstein and Rosen, 2015). These symptoms could be related to schizophrenia or they may be caused by comorbidities such as diabetes (e.g. diabetic retinopathy) or psychotropic medications. Other follow-up studies of ocular imaging (e.g. Optical Coherence Tomography and Optical Coherence Tomography Angiography) in patients with type 1 diabetes reported that retinal nerve and vascular structures were associated with a decline in cognitive function (Fickweiler et al., 2021).

Collectively, we found that terms related to immunity and inflammation tend to appear more frequently in the SNP enrichment analysis, including transcription factors, compared to the pathway enrichment analysis of DEGs obtained from the public datasets. Thus, the immune system appears to serve a major role in the pathogenesis of schizophrenia, as pointed out in previous studies. The results also suggested an association with blood cell counts, such as eosinophils and basophils, and the ocular system. However, there are very few publications that have investigated the relationship between schizophrenia and these traits (Fig. 4c). Considering their frequent occurrence in the SNP enrichment analysis, we suggest that further studies are needed on the relationship between schizophrenia and these traits in the future.

4.4. Extending the research strategy to other target genes and other omic layers

Although this meta-analysis examined only transcriptome and SNP data, it is possible to explore far more mechanistic details if other omics data are measured. Gene expression regulation by ncRNA is another possible future addition to this work. In addition to miRNAs, which were introduced in this paper, other ncRNAs that have been reported to be altered in schizophrenia are long noncoding RNA (lncRNA) and Piwi-interacting RNA (piRNA) (Bundo et al., 2014; Mishra and Kumar, 2021; Ni et al., 2020a, b). piRNA exerts epigenetic functions such as silencing retrotransposons in germ cells and post-transcriptional regulation that contribute to gene expression in nerve cells. The copy number of LINE-1, a piRNA, has been confirmed to increase in neurons differentiated from patient-derived iPS cells including those with the 22q11 deletion (Bundo et al., 2014).

ncRNA data were not measured in all the public datasets used in this study, so it was not possible to compare the DRNs that we reconstructed with the network that includes regulation by miRNAs (Guo et al., 2010). TargetScan (http://www.targetscan.org/vert_72/) facilitates predicting of miRNA target genes. For analysis of lncRNAs, NONCODE (http://www.noncode.org) and LNCipedia (https://lncipedia.org/) are available. piRBase provides manually curated piRNA data (http://www.regulatoryrna.org/database/piRNA/). In our next work, we plan to reconstruct gene expression regulatory networks in which ncRNA is also taken into consideration by using these resources based on the OSINT approach.

In the publicly available transcriptome datasets used in this study, only three (datasets 6, 8, and 14) out of 15 were published with the genotype of the patients. There were two more datasets (datasets 2 and 7) that showed family histories but did not show specific genotypes. As mentioned in the discussion corresponding to the comparison of the DRNs, if the genotype of the patient is known, differences among the DRNs could be interpreted for each genotype. In addition to the limited genotype information, it was not possible to access other data, such as descriptions of comorbidities of the patients. Based on the results of this study, we propose to collect information such as gene deletions, descriptions of comorbidities and blood cell counts in comprehensive studies of schizophrenia in the future.

5. Conclusion

We conducted a meta-analysis of 15 publicly available transcriptome datasets of neural cells from schizophrenia patients to reconstruct and compare transcriptional regulatory networks for lipid metabolism enzymes. Comparison of the DRNs revealed commonly utilized network modules among the datasets, as well as differences between them. Since promoters and enhancers may cover different networks, it is essential to include not only TF-target gene pairs via promoters but also those via enhancers. In fact, 53.3% of the total 14,215 edge pairs were involved in enhancer-mediated regulation in our study. However, by clustering the TF-target gene pairs based on their frequency of occurrence, we also found that there was a bias in the frequent DRNs among the datasets. Possible reasons for this bias are differences in sample preparation protocols, measurement methods and environmental factors surrounding individuals among the datasets. Moreover, the enrichment analysis of the SNPs associated with the genes in the DRNs led to the suggestion that terms involved with hematological and eye systems were associated with schizophrenia. Although these terms appeared in many datasets, the number of publications studying their association with schizophrenia was small. Based on these results, we suggest that when collecting samples from schizophrenia patients in the future, information such as genotype, description of comorbidities and hematological cell counts should be included, as well as the transcriptome, for more detailed genetic stratification and mechanistic exploration of schizophrenia.

Acknowledgements

This work is supported by funding from JSPS KAKENHI grants JP15H055882, JP18H05431 and “Creation of Innovative Technology for Medical Applications Based on the Global Analyses and Regulation of Disease-Related Metabolites”, PRESTO (JPMJPR1538) from JST. L.O. and S.D. were supported by research funds from the Yamagata prefectural government and the City of Tsuruoka. A.H. was supported by a Grant-in-Aid for Scientific Research on Innovative Areas “Transomic Analysis of Metabolic Adaptation” (JP17H06300) from JSPS KAKENHI. X.N. is supported by a JSPS KAKENHI grant JP21J01116. L.O., A.H. and K.Y. conceived the project. L.O. reconstructed differentially regulated networks and S.D. visualized and analyzed the reconstructed networks. S.W. performed identification of DEGs from the transcriptome data, integration of GTex eQTL data with the reconstructed networks and enrichment analysis of the SNPs associated with the DEGs. X.N. conducted bibliometric analysis based on the enriched EFO terms. J.M. compiled the data processing pipeline. L.O., S.W., S.D., X.N., J.M., M.T., A.H., and K.Y. wrote the manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the
online version, at doi:https://doi.org/10.1101/2020.12.06.102318

References

Amoli, M.M., Khatami, F., Arzaghi, S.M., Enayati, S., Nejatisafa, A.-A., 2019. Over-
expression of TGF-beta and its role in fibrosis formation in the brain barrier in 22q11.2
deletion syndrome. Brain. https://doi.org/10.1093/brain/1149.8376.

Brenna, K.J., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., Sangar, S., Li, Y.,
Borovcanin, M., Jovanovic, I., Radosavljevic, G., Djukic Dejanovic, S., Bankovic, D.,
Benros, M.E., Nielsen, P.R., Nordentoft, M., Eaton, W.W., Dalton, S.O., Mortensen, P.B.,
Avgustin, B., Wraber, B., Tavcar, R., 2005. Increased Th1 and Th2 immune reactivity
in multiple risk loci for chronic lymphocytic leukemia. Nat. Genet. 45, 868–76.

Cao, B., Wang, D., Brietzke, E., McIntyre, R.S., Fun, Z., Cha, D., Rosenblat, J.D.,
Eckert, H., Lui, Y. 2018. Characterization of the immune signatures among individuals with schizophrenia: a case-control study. Amino
Acids 50, 1013–1023. https://doi.org/10.1007/s00726-017-2597-9.

Cao, B., Jin, M., Brietzke, E., McIntyre, R.S., Wang, D., Rosenblat, J.D., Raggett, R.-M.,
Zhang, G., Sun, X., Rong, C., Wang, J., 2019. Serum metabolic profiling using small
molecular water-soluble metabolites in individuals with schizophrenia: a longitudinal study using a pre-post-treatment design. Psychiatry Clin. Neurosci. 73, 1008–1016. https://doi.org/10.1111/pcn.13104.

Carr, E., 2016. Shaken and stirred. The Economist, 12 November. London.

Chen, C., Meng, Q., Xiang, Y., Ding, C., Wang, L., Dai, R., Cheng, L., Gunaratne, P., Gibbs, R.
A., Min, S., Cooira, C., Reid, J.G., Zhang, C., Jiao, C., Jiang, Y., Glase, G., Thomas, A.,
Fitzgerald, D., Brunetti, T., Shi, H., Aix, C., Wang, Y., Wang, Y., Badner, J.A., Gershon, E.S., White, K.P., Liu, C., 2018. The transcription factor FOXP2 regulates a gene coexpression network in brain tissue from patients with psychiatric disorders. Sci. Transl. Med. 10. https://doi.org/10.1126/scitranslmed.aat1879.

Chen, J., Wang, S., Dong, Y., Yuan, Z., 2020. The role and regulatory mechanism of
hippoc signaling components in the neuronal component. Front. Neurosci. 14. https://doi.org/10.3389/fnins.2020.00384.

Clin. Res. 95, 135–142. https://doi.org/10.1016/j.jpsychires.2017.08.009.

Dinan, T.G., Sayer, W., Walford, R.L., Beasley, R., 1996. The effects of calorie restriction on
longevity. J. Gerontol. A Biol. Sci. Med. Sci. 51, B127–B130. https://doi.org/10.1093/gerona/51.2.B127.

Dobrovolski, A.D., Salazar, S., Paredes, L., Al-Zahabi, A., Jezova, D., 2012. Elevated serum level of type-2 cytokine and low
molecular water-soluble metabolites in individuals with schizophrenia: a longitudinal study using a pre-post-treatment design. Psychiatry Clin. Neurosci. 73, 1008–1016. https://doi.org/10.1111/pcn.13104.

Fickweiler, W., Wolfson, E.A., Paniagua, S.M., Yu, M.G., Adam, A., Bahnam, V.,
Fang, H., Knezevic, B., Burnham, K.L., Knight, J.C., 2016. XGR software for enhanced
genome-wide integration of enhancers and target genes in GenCards. Database 2017, 1–17. https://doi.org/10.1093/database/bax028.

Gershon, E.S., Brinkman, D., Zhang, E., Wischhusen, J., Zanis, M., 2012. GeneHancer: genome-
wide integration of enhancers and target genes in GenCards. Database 2017, 1–17. https://doi.org/10.1093/database/bax028.

Ghosh, S., Dyer, R.A., Beasley, C.L., 2017. Evidence for altered cell membrane lipid
composition in postmortem prefrontal white matter in bipolar disorder and
schizophrenia. J. Psychiatr. Res. 95, 135–142. https://doi.org/10.1016/j.
psychiatriy.2017.06.009.

Gorelik, L., Constant, F., Flavell, R.A., 2002. Mechanism of transforming growth factor-
beta-induced inhibition of T help type 1 differentiation. J. Exp. Med. 195, 1499–1505. https://doi.org/10.1084/jem.20012076.

Gotlib, I., 2017. World Health Organization-defined eosinophilic disorders: 2017 update
and management. Am. J. Gastroenterol. 112, 789–802. https://doi.org/10.1038/jn.2017.149.

Hartz, S.M., Horton, A.C., Hancock, D.B., Baker, T.B., Caporaso, N.E., Chen, L.-S.,
Chin, C., 2021. Disruption of the blood–brain barrier in 22q11.2 deletion syndrome. Brain. https://doi.org/10.1093/brain.
awab174.

Henein, M., Carter, T., Huennekens, F., Petrini, J.O., Tcheng, J.E., 1981. Evolution of calr-positive primary myelofibrosis toward acute myeloid leukemia. Blood 58, 274–83. https://doi.org/10.1182/jn.1981.58.1.274.

Immunol. 8, 484. https://doi.org/10.3389/fimmu.2017.00484.

Kohn, A., Twik, M., Safran, M., Lancet, D., Cohen, D., 2017. GeneHancer: genome-
wide integration of enhancers and target genes in GenCards. Database 2017, 1–17. https://doi.org/10.1093/database/bax028.

Koos, Z., 2018. Neuropharmacology of schizophrenia: the role of the brake and
accelerator. Neuropharmacology 136, 1–13. https://doi.org/10.1016/j.
neuropharm.2018.05.004.

Kotzbue, S., Andrasik, F., 2012. The role of transforming growth factor-beta in
the development and progression of idiopathic pulmonary fibrosis. Respir. Res. 13, 1–9. https://doi.org/10.1186/respirology.2012.08.016.

Kotzbue, S., Andrasik, F., 2012. The role of transforming growth factor-
beta in the development and progression of idiopathic pulmonary fibrosis. Respir. Res. 13, 1–9. https://doi.org/10.1186/respirology.2012.08.016.

Kotzbue, S., Andrasik, F., 2012. The role of transforming growth factor-
beta in the development and progression of idiopathic pulmonary fibrosis. Respir. Res. 13, 1–9. https://doi.org/10.1186/respirology.2012.08.016.

Kotzbue, S., Andrasik, F., 2012. The role of transforming growth factor-
beta in the development and progression of idiopathic pulmonary fibrosis. Respir. Res. 13, 1–9. https://doi.org/10.1186/respirology.2012.08.016.

Kotzbue, S., Andrasik, F., 2012. The role of transforming growth factor-
beta in the development and progression of idiopathic pulmonary fibrosis. Respir. Res. 13, 1–9. https://doi.org/10.1186/respirology.2012.08.016.

Kotzbue, S., Andrasik, F., 2012. The role of transforming growth factor-
beta in the development and progression of idiopathic pulmonary fibrosis. Respir. Res. 13, 1–9. https://doi.org/10.1186/respirology.2012.08.016.

Kotzbue, S., Andrasik, F., 2012. The role of transforming growth factor-
beta in the development and progression of idiopathic pulmonary fibrosis. Respir. Res. 13, 1–9. https://doi.org/10.1186/respirology.2012.08.016.
Topol, A., Zhu, S., Hartley, B.J., English, J., Hauberg, M.E., Tran, N., Rittenhouse, C.A., Simone, A., Ruderfer, D.M., Johnson, J., Rendhead, B., Hadan, Y., Gochnach, P.A., Wang, Y.-C., Shah, H., Caprey, G., Rapoport, J., Gage, F.H., Dudley, J.T., Sklar, P., Matthiesen, M., Cotter, D., Fang, G., Brennand, K.J., 2016. Disregulation of mRNAs 9 in a subset of schizophrenia patient-derived neural progenitor cells. Cell Rep. 15, 1024–1036. https://doi.org/10.1016/j.celrep.2016.03.090.

Topol, A., Zhu, S., Tran, N., Simone, A., Fang, G., Brennand, K.J., 2015. Altered WNT signaling in human induced pluripotent stem cell neural progenitor cells derived from four schizophrenia patients. Biol. Psychiatry 78, e29–34. https://doi.org/10.1016/j.biopsych.2014.12.028.

Torshizi, A.D., Armaksu, C., Zhang, H., Forrest, M.P., Zhang, S., Souaiaia, T., Evgrafov, O.V., Knowles, J.A., Duan, J., Wang, K., 2019. Deconvolution of transcriptional networks identifies TCF4 as a master regulator in schizophrenia. Sci. Adv. 5, eaau4139. https://doi.org/10.1126/sciadv.aau4139.

Tufano, M., Pinna, G., 2020. Is there a future for PPARs in the treatment of neuropsychiatric disorders? Molecules 25. https://doi.org/10.3390/molecules25051002.

Upthegrove, R., Khandaker, G.M., 2020. Cytokines, oxidative stress and cellular markers of inflammation in schizophrenia. In: Khandaker, G.M., Meyer, U., Jones, P.B. (Eds.), Neuroinflammation and Schizophrenia. Springer International Publishing, Cham, pp. 49–66. https://doi.org/10.1007/978-3-030-14317-8.

Varga, T., Czirmerer, Z., Nagy, L., 2011. PPARs are a unique set of fatty acid regulated transcription factors controlling both lipid metabolism and inflammation. Biochim. Biophys. Acta 1812, 1007–1022. https://doi.org/10.1016/j.bbadis.2011.02.014.

Verrall, L., Burnet, P.W.J., Betts, J.F., Harrison, P.J., 2010. The neurobiology of D-amino acid oxidase and its involvement in schizophrenia. Mol. Psychiatry 15, 122–137. https://doi.org/10.1038/mp.2009.99.

Wang, L., Feng, Z., Wang, X., Wang, X., Zhang, X., 2010. DEGseq: an R package for differential expression in sequencing data. Bioinformatics 26, 136–138. https://doi.org/10.1093/bioinformatics/btp612.

Wang, D., Cheng, S.L., Fei, Q., Gu, H., Raftery, D., Cao, B., Sun, X., Yan, J., Zhang, C., 2014. Lipidomics reveals Xie, L., Shabalin, A.A., Bryois, J., Sweet, D.H., Kronfol, M.M., Palasuubenniam, P., McRae, M.P., Riley, B.P., Sullivan, P.F., Van Den Oord, E.J., McClay, J.L., 2018. Building a schizophrenia genetic network: transcription factor 4 regulates genes involved in neuronal development and schizophrenia risk. Hum. Mol. Genet. 27, 3246–3256. https://doi.org/10.1093/hmg/ddy222.

Yao, J.K., Leonard, S., Reddy, R.D., 2000. Membrane phospholipid abnormalities in postmortem brains from schizophrenic patients. Schizophr. Res. 42, 7–17. https://doi.org/10.1016/s0920-9964(99)00095-x.

Yin, X., Kim, K., Suetsugu, H., Bang, S.-Y., Wen, L., Koido, M., Ha, E., Liu, L., Sakamoto, Y., Jo, S., Seng, E.R., Otomo, N., Laurynenka, V., Kwon, Y.-C., Sheng, Y., Sugano, N., Hwang, M.Y., Li, W., Muki, M., Yoon, K., Cai, M., Ishigaki, K., Chung, W.T., Huang, H., Takahashi, D., Lee, S.-S., Wang, M., Karino, K., Shim, S.-C., Zheng, X., Miyamura, T., Kang, Y.M., Ye, D., Nakamura, J., Sub, C.-H., Tang, Y., Motomura, G., Park, Y.-B., Ding, H., Kuroda, T., Choe, J.-Y., Li, C., Niño, H., Park, Y., Shen, C., Miyamoto, T., Ahn, G.-Y., Fei, W., Takeuchi, T., Shin, J.-M., Li, K., Kawaguchi, Y., Lee, Y.-K., Wang, Y., Amano, K., Park, D.J., Yang, W., Tada, Y., Yamaji, K., Shimizu, M., Arai, T., Sumiki, A., Sumida, T., Okada, Y., Matsuoka, K., Kochi, Y., Japanese Research Committee on Idiopathic Osteonecrosis of the Femoral Head, Kottyan, L.C., Weirach, M.T., Parameswaran, S., Esvar, S., Salim, H., Chen, X., Yamamoto, K., Harley, J.B., Ohmura, K., Kim, T.-H., Yang, S., Yamamoto, T., Kim, B.-J., Shen, N., Ikegawa, S., Lee, H.-S., Zhang, X., Terao, C., Cui, Y., Bae, S.-C., 2021. Meta-analysis of 208370 East Asians identifies 113 susceptibility loci for systemic lupus erythematosus. Ann. Rheum. Dis. 80, 630–640. https://doi.org/10.1136/annrheumdis-2020-219209.

Yoon, K.-J., Nguyen, H.N., Utrinsi, G., Zhang, F., Kim, N.-S., Wen, Z., Makri, G., Nauen, D., Shin, J.H., Park, Y., Chung, B., Pekle, E., Zhang, C., Towe, M., Hussaini, S.M.Q., Lee, Y., Rujescu, D., St Clair, D., Kleinman, J.E., Hyde, T.M., Kraus, G., Christians, K.M., Rapoport, J.L., Weinberger, D.R., Song, H., Ming, G.-L., 2014. Modeling a genetic risk for schizophrenia in iPSCs and mice reveals neural stem cell deficits associated with adherens junctions and polarity. Cell Stem Cell 15, 79–91. https://doi.org/10.1016/j.stem.2014.05.003.

Zhang, X., Zhang, Y., Zhu, X., Purman, C., Haney, M.S., Ward, T., Khechaduri, A., Yao, J., Weissman, S.M., Urban, A.E., 2018. Local and global chromatin interactions are altered by large genomic deletions associated with human brain development. Nat. Commun. 9, 5356. https://doi.org/10.1038/s41467-018-07766-x.

Zolezzi, J.M., Santos, M.J., Bastias-Candia, S., Pinto, C., Godoy, J.A., Inestrosa, N.C., 2017. PPARs in the central nervous system: roles in neurodegeneration and neuroinflammation. Biol. Rev. Camb. Philos. Soc. 92, 2046–2069. https://doi.org/10.1111/brv.12320.

Zuber, V., Jonsson, E.G., Frei, O., Witteolar, A., Thompson, W.K., Schork, A.J., Bettella, F., Wang, Y., Djurovic, S., Smeland, O.B., Dietert, S., Fanous, A.H., Desikan, R.S., Koury, S., Beizai, S., Dale, A.M., Mills, I.G., Andreassen, O.A., 2018. Identification of shared genetic variants between schizophrenia and lung cancer. Sci. Rep. 8, 674. https://doi.org/10.1038/s41598-017-16481-4.