A transcriptome-wide Mendelian randomization study in isolated human immune cells highlights *interferon regulatory transcription factor 3* as a schizophrenia hub gene.

David Stacey*, PhD [1,2,3]; Liam Gaziano, PhD [4]; Preethi Eldi, PhD [2]; Catherine Toben, PhD [5]; Beben Benyamin, PhD [1,3,6]; S Hong Lee, PhD [1,3,6]; Elina Hyppönen, PhD [1,2,3]

[1]Australian Centre for Precision Health, University of South Australia, Adelaide, South Australia, Australia. [2]UniSA Clinical and Health Sciences, University of South Australia, Adelaide, South Australia, Australia. [3]South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia. [4]Harvard Medical School, Boston, Massachusetts, USA. [5]University of Adelaide, Discipline of Psychiatry, Adelaide, South Australia, Australia. [6]UniSA Allied Health and Human Performance, University of South Australia, Adelaide, South Australia, Australia.

***Corresponding author**

Dr. David Stacey

Postal address: South Australian Health and Medical Research Institute (SAHMRI), level 8 south, North Terrace, Adelaide, SA 5000, Australia.

E-mail: david.stacey@unisa.edu.au

Tel: +61(0)8 8302 8339

ABSTRACT

Immune mechanisms are associated with schizophrenia, although the precise mechanisms involved are not fully understood. To identify immune genes associated with schizophrenia, we conducted a transcriptome-wide Mendelian randomization study using gene expression exposures from 29 *cis*-eQTL datasets encompassing 11 unique immune cell types, all publicly available from the eQTL catalogue. These analyses highlighted 196 genes, including 67 located within the human leukocyte antigen (HLA) region. Enrichment analyses indicated an over-representation of immune genes, which was driven primarily by the HLA genes. Stringent validation and replication steps retained 61 candidate genes, 16 of which were immune-related. As an exemplar, we performed follow-up analyses focused on *interferon regulatory transcription factor 3* (*IRF3*), which coordinates interferon responses to viral infections. We found evidence of shared genetic aetiology between schizophrenia and autoimmune diseases at the *IRF3* locus, and a significant enrichment of *IRF3* chromatin binding at schizophrenia risk loci. Together, our findings suggest the HLA region may be responsible for a wide range of immune perturbations not directly related to schizophrenia aetiology. They also provide support for *IRF3* as a schizophrenia hub gene responsible for mediating the impact of viral infections on schizophrenia risk, and as a potential novel therapeutic target.

INTRODUCTION

Schizophrenia (SCZ) is a devastating psychiatric illness affecting approximately 1 in 300 people worldwide[1], with onset generally occurring during late adolescence or early adulthood[2]. It is characterised by several debilitating symptoms including psychosis (i.e., hallucinations, delusions), disorganised thought and speech, and flattened affect[3]. Although SCZ is highly heritable (~80%)[4], its underlying genetic aetiology is complex and heterogeneous. Moreover, approximately 30% of patients are resistant to currently available antipsychotics[5], which underscores the need to identify novel drug targets. Importantly, recent studies have shown that anti-inflammatory medications can alleviate SCZ symptoms in some patients, including treatment-resistant patients[6, 7].

Immune hypotheses of SCZ have been in existence for decades, supported by several lines of evidence. For example, some people with SCZ have abnormalities in the levels of circulating immune markers[8, 9] and autoantibodies[10]. People with SCZ also have an increased risk of comorbid immune-mediated diseases, particularly autoimmune diseases[11]; while inflammation of the brain – or encephalitis – can often present with psychiatric symptoms such as psychosis[12]. Although the relationship between the immune system and SCZ is highly complex and likely confounded by several lifestyle factors (e.g., smoking) and medication effects (e.g., antipsychotics), findings from Mendelian randomization (MR) studies, which are robust to such confounds, suggest that several circulating immune markers may be causally associated with SCZ risk[13-15].

Of these immune markers, C-reactive protein (CRP) – a key modulator of the complement cascade – has received the most widespread attention, with several independent MR studies consistently showing that a genetic propensity for higher levels of circulating CRP is associated with a lower risk of SCZ[13, 15, 16]. The precise mechanisms underlying this association are currently unclear, though they may be mediated by protection against infection[17]. Indeed, early life or maternal viral/bacterial infection is a well-established risk factor not only for SCZ[18, 19] but also autoimmune diseases[20], which may explain a proportion of the observed SCZ-autoimmune comorbidities.

Over the past decade, GWASes have identified more than 200 independent risk loci for SCZ, of which the most highly significant (by many orders of magnitude) resides within the human leukocyte antigen (HLA) region[21]. Fine-mapping studies have highlighted *C4A*, which encodes complement component 4a, as a candidate effector gene at this locus, potentially mediated by perturbed synaptic pruning during adolescence[22]. Nevertheless, pinpointing the precise effector genes at complex disease risk loci is a non-trivial task, and so the extent to which other immune-related genes may mediate some of the known genetic risk is currently unclear. The continued integration of GWAS data with functional genomic data from immune tissues or cells may help to address this knowledge gap.

To this end, we performed a transcriptome-wide MR (TWMR) study of SCZ using publicly available expression quantitative trait locus (eQTL) data from 11 isolated immune cell types of both myeloid and lymphoid origin, including microglia. The TWMR approach utilises *cis*-eQTLs to instrument gene expression exposures at scale, and is a proven method for identifying causal genes for complex human diseases[23, 24]. We identified several immune-related genes as causal candidates for SCZ, potentially by modulating immune responses to neurotropic viral infection. We performed follow-up analyses of one of the immune gene candidates, *interferon regulatory factor 3* (*IRF3*), which encodes a transcription factor responsible for inducing interferon responses to viral infections. We show that SCZ shares genetic aetiology with autoimmune diseases at the *IRF3* locus, and that IRF3 chromatin binding sites are significantly enriched at SCZ risk loci. We propose *IRF3* to be a SCZ hub gene and a potentially druggable therapeutic target.

RESULTS

Transcriptome-wide Mendelian randomization (TWMR) and statistical colocalization of isolated human immune cell types highlight 61 schizophrenia-associated genes.

Fig. 1 describes the overall study design and main analyses conducted. To identify immune genes with potential causal roles in SCZ, we performed TWMR analyses using summary data from 29 *cis*-eQTL datasets selected (see **Methods**) from the eQTL catalogue generated by 15 different studies from 11 unique immune cell types (**Table 1**). To facilitate future TWMR efforts focused on exposures from immune cells, we have provided as an immunogenomic resource all the genetic variants used to instrument genes in each of the eQTL catalogue datasets we tested (**Table S1**). Overall, these analyses highlighted a total of 196 significant genes (see **Methods, Fig. 2a, Table S2**), of which 156 (79.59%) were protein-coding and 67 (34.18%) were located within the extended HLA region as defined by Horton et al. (2004) (chr6:25,726,063-33,400,644; hg38)[25]. Notably, this gene set did not include *CRP* or any genes encoding major cytokines. The absence of *CRP* was expected since it is specifically expressed in the liver, with only residual expression observed in immune cells[26].

Table 1. Summary of human immune cell *cis*-eQTL datasets downloaded from the eQTL catalogue for transcriptome-wide Mendelian randomization and statistical colocalization analyses.

Study	Cell types	Conditions	Samples	Gene expression platform
Alasoo et al. (2018)[27]	Macrophages	Naïve, IFN-γ, salmonella, IFN-γ & salmonella	84	RNA-seq
Blueprint[28]	Monocytes, Neutrophils, CD4+ T cells	Naïve	197	RNA-seq
Bossini-Castillo et al. (2019)[29]	Regulatory T cells	Naïve	119	RNA-seq
CAP[30]	LCLs	Naïve, statins	148	RNA-seq
CEDAR[31]	CD4+ T cells, CD8+ T cells, Monocytes, Neutrophils, B cells	Naïve	322	Microarray
Fairfax et al. (2012)[32]	B cells	Naïve	282	Microarray
Fairfax et al. (2014)[33]	Monocytes	Naïve, IFN-γ (24hrs), LPS (2hrs), LPS (24hrs)	424	Microarray
Gencord[34]	LCLs, T cells	Naïve	195	RNA-seq
GEUVADIS[35]	LCLs	Naïve	445	RNA-seq
Gilchrist et al. (2021)[36]	Natural Killer cells	Naïve	247	Microarray
Kasela et al. (2017)[37]	CD4+ T cells, CD8+ T cells	Naïve	297	Microarray
Naranbhai et al. (2015)[38]	Neutrophils	Naïve	93	Microarray
TwinsUK[39]	LCLs	Naïve	433	RNA-seq
Young et al. (2019)[40]	Microglia	Naïve	104	RNA-seq

Approximately 50% (97/196) of the genes were associated with SCZ in a single *cis*-eQTL dataset, ~15% (32/196) in two datasets, and a further ~11% (22/196) in three (**Fig. 2a**). There were examples of genes with ubiquitous associations across datasets. Indeed, three genes located within the HLA region – *BTN3A1*, *ZFP57*, and *HLA-DRB1* – were each associated with SCZ in more than 10 *cis*-eQTL datasets (**Fig. 2a**). There were also several examples of immune cell- or treatment-specific SCZ associations. For example, *MAP7D1*, which plays a key role in cytoskeleton organization, was only significant in the three neutrophil datasets tested (**Fig. 2a**). Furthermore, *C4A*, which has previously been highlighted as a potential effector gene at the SCZ risk locus within the human leukocyte antigen (HLA) region[22], was only highlighted in the two interferon-γ-treated macrophage datasets.

To characterise this list of 196 genes, we performed gene ontology (GO) term enrichment analyses using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (see **Methods**)[41]. An initial analysis including all 196 genes revealed significant (FDR<0.05) enrichment of 30 GO terms, most of which were immune-related (**Fig. 2b**, **Table S3**). However, none of these terms retained significance after removing the 67 HLA region genes from the test set (**Fig. 2c**, **Table S4**), demonstrating that the immune enrichment observed in our initial analysis was driven by the HLA region genes.

Since TWMR studies are susceptible to confounding by linkage disequilibrium (LD), we sought to validate our findings by performing statistical colocalization analyses (see **Methods**). Overall, we found robust evidence (PP_{H4}>0.8) of colocalization between SCZ and at least one *cis*-eQTL dataset for 61 of 196 genes (31.12%) (**Fig. 2a**, **Tables S5**,**6**) – herein, we will refer to these colocalising genes as the validated SCZ-associated gene set. Notably, most of the colocalizing genes resided outside of the extended HLA region, and the overall evidence for colocalization within the HLA region was sparse (**Fig. 2a**). Indeed, of the 67 HLA region genes, only six of them (*H4C13*, *APOM*, *RNF5*, *AGER*, *PBX2*, *GPSM3*) were in the validated SCZ-associated gene set, indicating that most of the associations observed in this region, including *C4A*, do not reflect causal effects on SCZ (at least not in the cell types tested here).

In addition to validating our TWMR findings using statistical colocalization, we sought to replicate them using the whole blood *cis*-eQTL data from the eQTLgen consortium. We selected whole blood as a proxy tissue because most of the immune cells investigated in our discovery analyses should be represented within blood samples. We restricted replication analyses to the set of validated SCZ-associated genes. Overall, 51 out of 61 genes were present in the eQTLgen data and had at least one suitable instrument (see **Methods**), of which 41 replicated with p_{adj}< 0.05/61 genes (**Fig. S1**, **Table S6**). Furthermore, of the 41 genes that replicated, 38 had directionally consistent estimates across both the immune cell and whole blood *cis*-eQTL datasets. These findings support the broad replicability of the validated SCZ-associated gene set.

Immune cell and brain tissue gene candidates partially overlap, and a high ratio of colocalising to non-colocalising TWMR signals implicates Natural Killer cells in schizophrenia aetiology.

To contextualise our immune cell type findings, we compared them with analogous findings from human brain tissue. To this end, we applied our TWMR-colocalization pipeline to *cis*-eQTL data from the 13 brain tissues from the Genotype-Tissue Expression (GTEx) project[42]. To ensure a direct comparison, we utilised GTEx summary data generated by a re-analysis of the raw data conducted by the eQTL catalogue team using the same methods applied to the immune cell type data utilised here.

Collectively, we found a total of 81 significant genes from the 13 GTEx brain tissues, 21 (~25%) of which colocalised with SCZ (**Fig. 3a**, **Tables S7**,**8**). Of these 81 genes, 44 (54.32%) were also present in the set of 196 genes identified in the immune cell type analyses, while 10 (47.62%) of the 21 colocalising genes were present in the validated SCZ-associated set of 61 genes (**Figs. 3a, b**). Thus, the observed overlap indicates appreciable concordance between the immune cell and brain tissue datasets, with the unique contributions from each likely reflecting a mixture of sub-threshold associations and distinct aetiological mechanisms.

To directly compare the global TWMR estimate profiles for each of the brain tissue and immune cell datasets, we correlated the estimates (regardless of *p*-value and colocalization status) across all shared genes for each pair of datasets (**Fig. 3c**). Overall, the observed coefficients skewed positive, with only a few very small negative correlations. Hierarchical clustering assigned the GTEx brain tissues into two brain-specific clusters, one containing the two cerebellar tissues, and the other containing the remaining 11 brain tissues (**Fig. 3c**). In general, clustering of the immune cell datasets recapitulated their corresponding immune cell types. For example, the three neutrophil datasets clustered together, as did the three untreated monocyte datasets (**Fig. 3c**). The microglial dataset formed its own cluster between the brain tissue and macrophage datasets (**Fig. 3c**), which is consistent with microglia being the resident macrophages of the central nervous system (CNS). These findings are consistent with brain tissue and immune cell types reflecting at least partially distinct SCZ-related aetiology.

To prioritise specific immune cell types with potential relevance for the aetiology of SCZ, we compared the number of significant TWMR signals contributed by each of the immune cell datasets tested. Overall, the largest number of signals were identified in the monocyte, neutrophil, and T cell datasets generated by the BLUEPRINT epigenome project[28], as well as the lymphoblastoid cell dataset from TWINS UK[39] (**Fig. 3d**). For context, the signal counts for these four datasets were comparable to those for the cerebellum and cerebellar hemisphere datasets, which contributed most of the GTEx brain signals (**Fig. 3e**). Since three of the four immune cell datasets with the most signals were from a single study, we checked whether TWMR signal counts were associated with study-specific factors. We found significant correlations with the number of eGenes (**Figs. 3f, g**), but not eQTL sample size (**Figs. 3h,i**) or gene expression platform (**Figs. 3j, k**). Together, these findings suggest that signal counts may have been driven more by the number of eGenes than immune cell type.

As an alternative metric to prioritise immune cell types in SCZ aetiology, we compared the signal-to-noise (i.e., colocalising vs. non-colocalising signals) ratio across the immune datasets. Notably, we found that almost 50% of the TWMR signals identified in the Natural Killer (NK) cell dataset from Gilchrist et al. (2021)[36] colocalised with SCZ (**Fig. 3d**). Moreover, when we recalculated the signal-to-noise ratio after removing the extended HLA region signals, the ratio increased to more than 70%. Since NK cells represent the immune system's first line of defence against virally infected cells, this cell type may warrant further investigation as a potential mediator of the association between prenatal/early life viral infections and SCZ risk. Conversely, none of the seven TWMR signals identified in the microglia dataset from Young et al. (2019)[40] colocalised with SCZ. This was unexpected given that microglia play a key role in synaptic pruning during adolescence, a process that when perturbed is thought to increase the risk of SCZ. This lack of colocalising signals is therefore inconsistent with a major causal role for microglia.

Immune-related genes are candidate effectors at several genome-wide significant schizophrenia risk loci.

Although there was no evidence of immune enrichment in our initial list of SCZ-associated genes after removing the HLA region, this does not preclude the possibility that some of the remaining genes may still modulate SCZ risk through the regulation of immune mechanisms. To explore this possibility, we focused on the validated SCZ-associated set of 61 genes. We first annotated them with ontological terms and pathways using the functional annotation tool from DAVID, and then manually searched for immune-related annotations. Overall, we found 16 potential immune-related genes within this gene set, four of which reside within the extended HLA region (**Table 2**, **Table S9**).

To determine whether these immune-related genes are candidate effector genes at known SCZ risk loci, we extracted their genomic coordinates and intersected them with the independent genome-wide significant ($p<5x10^{-8}$) loci identified in Trubetskoy et al. (2002)[21] (see **Methods**). Overall, 12 of the 16 genes reside within 1Mb of a SCZ risk locus, six of which also reside within 100Kb (**Table 2**). Moreover, four of them contain a SCZ risk sentinel variant from Trubestkoy et al. within their gene bodies: *FURIN* (rs4702, 3' untranslated region [UTR]), *IRF3* (rs7251, missense), *MAD1L1* (rs58120505, intronic), and *ZNF823* (rs72986630, 5' UTR) (**Table 2**). Thus, these four immune-related genes are strong candidate SCZ effector genes.

To assess whether the SCZ risk variants from **Table 2** are also associated with other immune-related phenotypes, we performed phenome-wide association studies (PheWASes) using the Open Targets Genetics platform (see **Methods**) (**Tables S10-S25**). We found several associations with immune-related blood cell traits (e.g., lymphocyte, neutrophil, and eosinophil counts) and immune-mediated diseases. Most notably, two of the 16 SCZ risk variants were associated with autoimmune diseases: rs140365013 at the HLA region was associated with type 1 diabetes and systemic lupus erythematosus (**Table S15**), while rs7251 at the *IRF3* locus on chromosome 19 was associated with autoimmune thyroid disease and several other thyroid conditions likely driven by autoimmunity (e.g., hypothyroidism) (**Table S16**). These two risk loci may therefore be key drivers of the observed comorbidities between SCZ and autoimmune disease.

Finally, effector genes at complex disease-associated risk loci often contain rare protein-altering variants that cause monogenic or oligogenic forms of related diseases[43]. We therefore performed lookups of the 16 immune-related genes in the Online Mendelian Inheritance in Man (OMIM) database[44] and Orphanet, which revealed rare disease annotations for four genes: *MAD1L1*, *NT5C2*, *PIK3C2A*, and *IRF3* (**Table S9**). Notably, *MAD1L1* was associated with B cell lymphoma and mosaic variegated aneuploidy syndrome 7 with inflammation and tumour predisposition, both of which are characterized by immune dysfunction. Furthermore, *IRF3* was associated with herpes simplex virus-1 (HSV-1)-induced encephalitis[45, 46], which is characterized by acute inflammation in the brain and can present with psychotic symptoms[12, 47]. Not only does this indicate a clear link between *IRF3* and inflammation, but it also hints at a mechanism by which *IRF3* may contribute towards SCZ-related symptoms.

Table 2. List of immune-related genes from the validated schizophrenia-associated genes. Genomic coordinates are in human genome build 38 (hg39). Genome-wide significant ($p<5x10^{-8}$) schizophrenia risk loci are represented by the sentinel variants at each locus from Trubetskoy et al. (2022). Hyphens indicate there are no genome-wide significant schizophrenia risk loci within 1Mb. SCZ: schizophrenia, Mb: megabases, Kb: kilobases.

Ensembl ID	Gene symbol	Gene name	Gene chr:start-end	SCZ risk loci within 1Mb	Position	Distance (Kb)
ENSG00000150967	*ABCB9*	ATP binding cassette subfamily B member 9	12:122920951-122981649	rs1790135 rs61952922	123184688 122082669	203.04 838.28
ENSG00000115073	*ACTR1B*	Actin related protein 1B	2:97655939-97664044	rs7575796	97080789	575.15
ENSG00000204305	*AGER*	Advanced glycosylation end-product specific receptor	6:32180968-32184322	-	-	-
ENSG00000143578	*CREB3L4*	cAMP responsive element binding protein 3 like 4	1:153967534-153974361	rs10559	153948088	19.45
ENSG00000140564	*FURIN*	Furin, paired basic amino acid cleaving enzyme	15:90868588-90883564	rs4702	90883330	0
ENSG00000213654	*GPSM3*	G protein signaling modulator 3	6:32190766-32195523	-	-	-
ENSG00000275126	*H4C13*	H4 clustered histone 13	6:27873148-27873534	rs140365013	27556090	317.06
ENSG00000126456	*IRF3*	Interferon regulatory factor 3	19:49659569-49665875	rs7251 rs2387414	49659652 50530986	0 865.11
ENSG00000002822	*MAD1L1*	Mitotic arrest deficient 1 like 1	7:1815793-2233243	rs58120505 rs7811528	1990232 2663182	0 429.94
ENSG00000102882	*MAPK3*	Mitogen-activated protein kinase 3	16:30114105-30123506	rs9925915	29982365	131.74
ENSG00000197696	*NMB*	Neuromedin B	15:84655129-84658563	rs11638554	84605000	50.13
ENSG00000076685	*NT5C2*	5'-nucleotidase, cytosolic II	10:103087185-103277605	rs12571643	102760520	326.67
ENSG00000167004	*PDIA3*	Protein disulfide isomerase family A member 3	15:43746394-43773279	rs2929278	43958115	184.84
ENSG00000011405	*PIK3C2A*	Phosphatidylinositol-4-phosphate 3-kinase catalytic subunit type 2 alpha	11:17077730-17207986	-	-	-
ENSG00000204308	*RNF5*	Ring finger protein 5	6:32178405-32180793	-	-	-
ENSG00000197933	*ZNF823*	Zinc finger protein 823	19:11721265-11739009	rs72986630 rs322124	11738921 11288696	0 432.57

Interferon regulatory factor 3 is expressed in the brain and may contribute towards schizophrenia-autoimmune comorbidities.

Since *IRF3* is a strong candidate effector gene at the rs7251 SCZ risk locus (**Table 2**) and a well-established immune gene, we sought to further elucidate the immune component of SCZ by performing follow-up analyses focused on the *IRF3* locus. *IRF3* encodes interferon regulatory transcription factor 3, which is responsible for inducing interferon responses to viral infection[48]. In

our TWMR analyses, we found an association between lower *IRF3* mRNA expression in natural killer (NK) cells and a higher risk of SCZ (bxy= -0.32, se=0.07, p=5.96x10^{-6}) (**Figs. 4a, b**). This association replicated in the eQTLgen whole blood dataset, and we also observed directionally consistent sub-threshold associations in both B cells and neutrophils (**Fig. 4a**). Moreover, statistical colocalization analyses revealed robust evidence of colocalization at the *IRF3* locus between SCZ and four of the six phenotypes highlighted in the PheWAS, including autoimmune thyroid disease (**Fig. 4c**).

To further assess the functional impact of the *IRF3* SCZ risk locus, we performed a look-up of the sentinel variant (rs7251) (**Table 2**) in *cis*-eQTL data from the Genotype-Tissue Expression (GTEx) project. Of the 49 GTEx tissues, rs7251 was significantly associated with *IRF3* mRNA expression in 16 of them, including two brain regions (cerebellum, cerebellar hemisphere) and the pituitary gland (**Fig. S2**). Notably, the lowest p-value was observed in thyroid tissue (p=2.5x10^{-14}). We also performed pairwise statistical colocalization analyses between SCZ GWAS and *IRF3 cis*-eQTL data from all 49 tissues, regardless of p-value. We found robust evidence of colocalization (PP_{H4}>0.8) for all 16 significant tissues and several other tissues including the hypothalamus (**Fig. 4d**). We also found suggestive colocalization (PP_{H4}>0.5) in two additional brain tissues: the frontal cortex (BA9) (PP_{H4}=0.73) and the nucleus accumbens (PP_{H4}=0.74) (**Fig. 4d**). Thus, since rs7251 is associated with regulatory function in the brain, *IRF3* may mediate SCZ risk independently of peripheral immune pathways.

Interferon regulatory factor 3 may represent a schizophrenia hub gene

Transcription factors are often hub genes responsible for orchestrating complex transcriptional networks that regulate important biological processes. Since *IRF3* is a transcription factor, we sought to assess the extent to which *IRF3* might regulate the expression of other SCZ risk genes. To explore this, we performed enrichment analyses using GARFIELD[49] (GWAS analysis of regulatory or functional enrichment with LD correction) by leveraging (i) full SCZ GWAS summary data from Trubetskoy et al. (2022)[21] and (ii) all available chromatin immunoprecipitation sequence (ChIP-seq) data for IRF3 from the ENCODE project[50] (see **Methods**). We found significant enrichment of SCZ risk loci (defined using genome-wide [p<5x10^{-8}] and suggestive [p<1x10^{-5}] significance thresholds) at IRF3 peaks from lymphoblastoid, epithelial, and neuroblastoma cells (**Fig. 5a**).

Overall, IRF3 peaks overlapped with 15 independent genome-wide significant (p<5x10^{-8}) and 11 suggestive (p<1x10^{-5}) SCZ risk loci in at least one of the three significant cell types (**Fig. 5b**). We found that 18 of these loci (11 genome-wide significant, 7 suggestive) overlapped with IRF3 peaks at one or more gene promoters (+/-1Kb of the transcription start site). Several of these promoters belonged to genes from our set of validated SCZ-associated genes, including *ZNF823* and *CREB3L4*, suggesting *IRF3* may regulate the expression of these genes. Other notable promoters belonged to *LETM2* and *TRIM38*. Rare variants affecting *LETM2* cause acute haemorrhagic leukoencephalitis (AHL), which is consistent with a role for *IRF3* in modulating maladaptive brain inflammation, whereas *TRIM38*, which encodes tripartite motif containing 38, resides at the primary SCZ signal within the HLA region. Intriguingly, the IRF3 peak-overlapping variant (rs17587226) at the *TRIM38* promoter was previously identified as a plasma *trans*-pQTL for the protein encoded by *AGER*[51], which was also in our validated SCZ-associated gene set. This suggests *IRF3* might modulate plasma abundance of the protein encoded by *AGER* with *TRIM38* as the mediator.

To confirm that SCZ-associated variants overlapping IRF3 peaks exert *cis*-effects on gene expression, we conducted systematic pairwise statistical colocalization analyses using SCZ GWAS summary

data[21] from the 26 IRF3 peak-overlapping regions (sentinel variant +/-1Mb) and *cis*-eQTL data for overlapping genes from the eQTLgen (whole blood) consortium. We found robust evidence of colocalization (PP_{H4}>0.8) between the SCZ signals at these 26 regions with *cis*-eQTL signals for a total of 49 genes (**Table S26**). These analyses confirmed that the SCZ signals overlapping IRF3 peaks at gene promoters (e.g., *ZNF823*, *CREB3L4*) do indeed colocalize with cognate *cis*-eQTL signals. Moreover, these analyses also highlighted several additional genes, suggesting that IRF3 may be binding other regulatory elements (e.g., enhancers) at SCZ risk loci. Taken together, these findings are consistent with a model whereby SCZ-associated variants at these 26 risk loci (15 genome-wide significant, 11 suggestive) exert regulatory effects on local gene expression either by strengthening or weakening IRF3 binding sites.

DISCUSSION

Here, we sought to identify immune genes and elucidate immune mechanisms involved in SCZ by leveraging 29 *cis*-eQTL datasets generated by 15 different studies from 11 unique human immune cell types, all publicly available from the eQTL catalogue[52, 53]. Overall, TWMR analyses highlighted 196 unique candidate genes, 67 of which resided within the extended HLA region. We applied a series of validation, replication, and filtering steps culminating in the prioritization of 16 immune-related genes that may play a role in SCZ aetiology. We performed in-depth follow-up analyses of one of them, *IRF3*, which we showed to be a SCZ hub gene and a novel drug target candidate that may impact SCZ risk by modulating host responses to viral infections.

As noted above, more than a third of the candidate SCZ-associated genes identified in our initial TWMR analyses resided at the HLA region. The extended HLA region has a complex LD structure[25] and contains the most highly significant SCZ risk locus (by many orders of magnitude)[21]. Given these factors, an overrepresentation of HLA region genes in our initial candidate set is unsurprising. Indeed, statistical colocalization analyses uncovered robust evidence of colocalization with SCZ for just six of the 67 HLA region candidates, which is consistent with widespread confounding by LD. These findings suggest that a significant proportion of the immune perturbations observed in SCZ patients (e.g., elevated cytokine levels, immune-mediated disease comorbidities) are not directly related to SCZ pathophysiology but may instead be a consequence of knock-on effects in those carrying SCZ risk alleles at the HLA region.

Given the complex LD structure at the HLA region, identifying the precise mechanisms driving the genetic association between this region and SCZ has proven challenging. However, previous fine-mapping and functional studies have highlighted *C4A* as a strong candidate effector gene at this locus, potentially impacting SCZ risk by perturbing synaptic pruning during adolescence[22]. Our TWMR analyses highlighted *C4A* as a candidate, but only in the two IFN-γ-treated macrophage datasets from Alasoo et al. (2018)[27] (one treated with IFN-γ alone, the other treated with IFN-γ in conjunction with salmonella), where IFN-γ appeared to induce *C4A* expression. IFN-γ treatment of cells may therefore provide a useful model to explore the impact of *C4A* on SCZ risk in future studies. Nevertheless, we found no evidence of colocalization between signals for SCZ and *C4A* mRNA expression at the HLA region in any of the immune cell types or GTEx brain tissues tested here.

We did however find robust evidence of colocalization with six other genes at the HLA region (*H4C13*, *APOM*, *RNF5*, *AGER*, *PBX2*, *GPSM3*), three of which we annotated as being potentially immune-related: *RNF5*, *AGER*, and *GPSM3* (**Table 2**). Of these genes, *AGER* has previously been linked to SCZ, with several observational studies showing lower levels of a soluble form of RAGE in serum from SCZ patients relative to healthy controls[54-56]. *AGER* encodes Receptor for Advanced Glycation Endproducts (RAGE), which plays an important role in the innate immune system by inducing inflammatory cytokines such as interleukin 6 (IL-6), IL-8, and tumour necrosis factor-α (TNF-α). Our findings therefore highlight other plausible candidate effector genes at the HLA region, including *AGER*, that may warrant further attention.

A comparison of the different immune cell datasets tested indicated that the number of significant TWMR signals was dictated more by study-specific factors (particularly number of eGenes) than type of immune cell. However, the NK cell dataset from Gilchrist et al. (2021) exhibited a high signal-to-noise ratio (i.e., colocalizing vs. non-colocalizing TWMR signals) compared to the other datasets. NK cells are a critical component of the innate immune system and represent the body's first line of defence against viral infections by recognizing and killing infected cells[57]. Indeed, people with functionally impaired NK cells, or those with low NK cell abundance, often suffer particularly severe

and recurrent viral infections that can prove fatal in extreme cases[57]. Since viral infection has long been considered a causal risk factor not only for SCZ[19, 20, 58] but also several other diseases including autoimmune diseases[20], NK cells may warrant further investigation as a potential mediator of these relationships.

Several different viruses have been implicated in SCZ aetiology including influenza and more recently SARS-CoV-2 (otherwise known as COVID-19), though neurotropic viruses (i.e., those with a propensity to infect the brain) like herpes simplex virus-1 (HSV-1), cytomegalovirus (CMV), and Epstein-Barr virus (EBV) are primary candidates[59]. As an example, HSV-1 initially infects epithelial cells at orofacial surfaces causing vesicular lesions known colloquially as cold sores[60]. However, following initial infection, HSV-1 is capable of travelling afferently along sensory neurons from the skin into the brain and has been shown to specifically infect parts of the limbic system[61]. This results in central inflammation to maintain HSV-1 latency[62] and in extreme cases can cause encephalitis, which can present with psychotic symptoms[12, 47]. Thus, chronic low-grade or acute intermittent bouts of inflammation in response to infection within the limbic system, for example, may play a role in triggering or aggravating SCZ symptoms in some patients.

We performed follow-up analyses of *IRF3*, which encodes an interferon regulatory transcription factor (IRF) that is a downstream effector of the cGAS-STING signalling pathway responsible for triggering interferon responses to viral infections[48]. *IRF3* resides at a SCZ risk locus on chromosome 19, where the missense sentinel variant (rs7251) results in an amino acid substitution at the 3' end of the IRF3 protein. This, combined with findings implicating rare loss-of-function variants affecting *IRF3* in HSV-1-induced encephalitis[45, 46], indicates that *IRF3* is a compelling causal risk gene for SCZ. In our TWMR analyses, *IRF3* was only significant in NK cells, where lower *IRF3* expression was associated with higher SCZ risk. We therefore propose a model whereby *IRF3* activity in NK cells might modulate SCZ risk by acting as a buffer against the spread of neurotropic viral infections from the periphery into the brain.

Alternatively, since *IRF3* is also expressed centrally, it may regulate immune mechanisms directly in the brain. Indeed, a recent animal study found increased depression- and anxiety-like behaviours in *Irf3*$^{-/-}$ mice accompanied by perturbed glutamatergic and GABAergic transmission in hippocampal and prefrontal cortex neurons relative to wild-type controls[63]. Importantly, the authors also showed that both the behavioural and neurotransmission phenotypes could be rescued – at least partially – in *Irf3*$^{-/-}$ mice by over-expressing *Irf3* specifically in the brain using an adeno-associated virus (AAV) construct[63]. Thus, based on these findings, a direct brain-centric role for *IRF3* in modulating SCZ risk cannot be ruled out.

In addition to the relevant tissue or cell type, another key question relates to the timepoint at which *IRF3* impacts SCZ risk or symptoms. A recent study found that *IRF3* mRNA expression, along with several other putative SCZ risk genes, was upregulated in placentas from pregnancies with SARS-CoV-2 (COVID-19) maternal infection[23]. This is consistent with an important role for *IRF3* in maternal and early life viral infections and stresses. Furthermore, since HSV-1-induced encephalitis can induce psychotic symptoms in adults with no clinical history of SCZ, this suggests *IRF3* can also impact SCZ-like symptoms outside of important neurodevelopmental timepoints[45, 46]. The latter findings indicate that *IRF3* may constitute a viable therapeutic target in those already presenting with SCZ, while the former suggest that population-level preventative strategies (e.g., vaccination) focused on pre-pregnant mothers and infants may also prove effective. Moreover, future studies should investigate whether vaccination is more or less efficacious in mothers and infants carrying SCZ risk alleles in *IRF3* and related genes.

A GARFIELD analysis revealed a significant enrichment of ChIP-seq *IRF3* peaks at known SCZ risk loci, thereby highlighting *IRF3* as a potential hub gene for SCZ risk and further strengthening its credentials as a putative drug target. Several of these peaks resided at proximal promoters or at *cis*-eQTLs for genes highlighted in our TWMR analyses. These genes included *ZNF823*, *MAPK3*, *AGER*, and *RNF5*, all of which belong to our prioritised set of 16 immune-related candidates (**Table 2**). Intriguingly, *ZNF823* and *MAPK3* both belong to the HSV-1 infection (hsa05168) KEGG pathway, while *MAPK3* was also found to be up-regulated in placentas from pregnancies with SARS-CoV-2 (COVID-19) maternal infection[23]. Furthermore, *RNF5* has been shown to inhibit type I interferon responses in HSV keratitis through the cGAS-STING/IRF3 signalling pathway[64]. Taken together, this is strong evidence consistent with a role for *IRF3* in coordinating the expression of a sub-network of putative SCZ risk genes, many of which are already known to be involved in modulating responses to viral infections.

Although we focused our follow-up efforts primarily on *IRF3*, future studies focussing on other genes from the set of 16 prioritised immune-related genes may yield valuable further insights. As mentioned above, *ZNF823*, *MAPK3*, and *RNF5* have established roles in viral infections, so deep-dive follow-on studies focused on these genes may help to better elucidate the mechanisms mediating the relationship between viral infections and SCZ symptoms. Furthermore, *FURIN* is another potential viral response-related candidate given it encodes a protease responsible for cleaving proteins produced by several viruses, including herpesvirus glycoprotein B produced by HSV, as well as the SARS-CoV-2 (COVID-19) spike protein[65]. However, it should be noted that FURIN also cleaves many endogenous substrates as well, some of which may plausibly be involved in SCZ pathophysiology. For example, cleavage by FURIN converts pro-brain derived neurotrophic factor (BDNF) into its mature, bioactive form, which has long been thought to play a role in SCZ[66]. Thus, we should be wary of the possibility of pleiotropy when further exploring immune-related SCZ candidates.

There are several potential limitations to this study. First, our TWMR analysis was conducted with exposures from nearly 30 *cis*-eQTL datasets, posing an increased risk of type I errors. To minimise this risk, we applied a conservative Bonferroni correction to each dataset, (see **Methods**) complemented by stringent validation and replication analyses. Second, most of the *cis*-eQTL datasets from isolated immune cells were derived from relatively small sample sizes, which limits discovery power. Access to larger datasets in the future will facilitate more robust discoveries. Third, certain immune cell types tested were represented by a single *cis*-eQTL study (e.g., NK cells, microglia). Independent *cis*-eQTL datasets for these cell types are necessary for replication and to further explore their role in SCZ.

In conclusion, we highlight several immune-related genes as potential candidates for involvement in SCZ aetiology. Of these genes, we have demonstrated that *IRF3*, which encodes a transcription factor, is a potential hub gene that likely regulates the expression of effector genes at multiple known SCZ risk loci. Since *IRF3* is activated in response to viral infections[48] and has been shown to mediate HSV-1-induced encephalitis[45, 46], *IRF3* is a strong candidate mediator of the link between prenatal/early life viral infections and SCZ risk and a potential therapeutic target for SCZ, particularly in people presenting with an underlying immune component. Thus, addressing unmet needs in vaccination (e.g., HSV-1) may not only serve as a preventative measure against the viruses themselves, but may also help to mitigate SCZ risk. Given the ample evidence for neurotropic viruses as causal risk factors for SCZ, pre-pregnant mothers, infants, and young children should be amongst those prioritized to receive corresponding vaccines.

METHODS

Datasets. For this study we utilized publicly available summary data from: (i) the largest and most recent GWAS of SCZ available from the psychiatric genetics consortium (PGC) (https://pgc.unc.edu/for-researchers/download-results/) in up to 130,644 Europeans, including 53,386 cases; (ii) *cis*-eQTL studies of isolated human immune cell types available through the eQTL catalogue[52, 53] (https://www.ebi.ac.uk/eqtl/) (**Table 1**), as well as the genotype tissue expression (GTEx) project[42] (https://gtexportal.org/home/) and the eQTLgen (phase I) consortium[67] (https://www.eqtlgen.org/phase1.html); (iii) GWASes of autoimmune thyroid disease[68] (European; up to 30,234 cases, 725,172 controls), cognitive performance[69] (European; up to 257,841 participants; https://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST006001-GCST007000/GCST006572/), diabetes or endocrine disease[70] (mixed, mostly European; up to 51,949 cases, 432,649 controls; https://www.ebi.ac.uk/gwas/publications/33959723), hypothyroidism or myxedema[70] (mixed, mostly European; up to 23,497 cases, 461,101 controls; https://www.ebi.ac.uk/gwas/publications/33959723), thyroid problem (not cancer)[70] (mixed, mostly European; up to 28,254 cases, 456,344 controls; https://www.ebi.ac.uk/gwas/publications/33959723), and vitiligo[71] (European; up to 4,680 cases, 39,586 controls; https://www.ebi.ac.uk/gwas/publications/27723757).

To select datasets from the eQTL catalogue, we first identified those derived from isolated human immune cell types. After running TWMR analyses, we then removed datasets flagged by the generalised summary-data-based Mendelian randomization (GSMR)[24] software tool as having too many variants with more than a 20% difference (--diff-freq 0.2) in allele frequency compared to the LD reference panel used. This filtering step protected against large ancestral differences between the exposure and outcome samples, and left us with 29 *cis*-eQTL datasets generated by 15 studies from 11 unique immune cell types for reporting and subsequent analyses.

Transcriptome-wide Mendelian randomization. We performed TWMR analyses using GSMR[24], as implemented in the Genome-Wide Complex Trait Analysis (GCTA) suite of tools (v1.94.1)[72]. Prior to running GSMR, we first extracted all *cis*-eQTLs from the eQTL catalogue[52, 53] datasets with *p*-values lower than the estimated empirical *p*-values (calculated from 1,000 permutations) for the corresponding eGene based on the beta distribution. We obtained all pre-computed empirical *p*-values from the eQTL catalogue[52, 53]. We supplied the pre-filtered exposure input files to GSMR and artificially set the GSMR significance threshold to $p<0.05$ (--gwas-thresh 0.05) to ensure all extracted *cis*-eQTLs were available for LD clumping within the GSMR workflow. Because *cis*-eQTL data generally yields fewer independent genetic instruments than well-powered GWASes, we set the minimum number of significant variants required for GSMR analysis to 1 (--gsmr-snp-min 1) and we applied a lenient LD (r^2) threshold of 0.1 (--clump-r2 0.1) (GSMR natively adjusts estimates for correlated instruments). We set all other GSMR parameters to default.

To estimate variant correlations, we utilized an individual-level reference genotype panel comprising 4,994 participants from the INTERVAL population study[73]. To minimise type I errors, we adjusted significance thresholds by applying a Bonferroni correction per eQTL dataset for the total number of eGenes tested (i.e., 0.05 / *n* eGenes). We plotted the GSMR estimates using the forestplot (v3.1.1) (https://CRAN.R-project.org/package=forestplot) and corrplot (v0.92)[74] R packages.

Statistical colocalization. We performed statistical colocalization analyses using two software tools: (i) the Coloc R package (v5.1.0.1)[75]; and (ii) a C++ implementation of pair-wise conditional and colocalization (PWCoCo) analysis[76]. In Coloc, we utilized the coloc.abf() function with priors and parameters set to default. The Coloc method has been described in detail previously, but briefly, it is

a suite of Bayesian tools that takes locus-specific summary data for a pair of traits as input. The coloc.abf function returns posterior probabilities (PPs) to estimate the likelihood that five distinct scenarios – or hypotheses – are true, assuming a single causal variant for each trait. These five hypotheses are: H_0, there is no causal variant for either trait at the locus; H_1/H_2, there is a causal variant for the first/second trait only; H_3, there are two distinct causal variants at the locus, one for each trait; H_4, both traits share the same causal variant. PWCoCo integrates Coloc with GCTA-COJO to enable colocalization after conditioning on independent regional signals. In PWCoCo, we set the *p*-value threshold for identifying significant signals in eQTL data from the eQTL catalogue and the GTEx project to the estimated empirical *p*-value for the corresponding eGene and eQTL dataset, as described above (***Transcriptome-wide Mendelian randomization***). Conversely, we set the *p*-value thresholds for SCZ and eQTLgen to $p<5x10^{-8}$. For both Coloc and PWCoCo, we considered a H_4 posterior probability (PP_{H4}) >0.8 to be robust evidence of colocalization between two traits, and a PP_{H4}>0.5 but ≤0.8 to be suggestive evidence though with potential for confounding by LD, as per Sun et al. (2018)[77].

Gene ontology term enrichment analyses. We conducted gene ontology (GO) term enrichment analyses using the functional annotation chart webtool available at the Database for Annotation, Visualization, and Integrated Discovery (DAVID)[41]. To facilitate interpretation, we performed enrichment analyses focused solely on the gene ontology (GO) direct terms for biological process, molecular function, and cellular component. We unselected all other functional annotation sources before beginning analyses. For the background gene set, we extracted all unique eGenes across all eQTL catalogue datasets (**Table 1**) used in TWMR analyses.

Variant and gene annotation. We annotated genetic variants using the Variant Effect Predictor (VEP) (GRCh37, release 109) webtool[78]. To identify potential immune genes within our set of candidates SCZ-associated genes, we first annotated them with GO terms and Kyoto encyclopedia of genes and genomes (KEGG) pathways using the functional annotation table function at DAVID[41]. We then manually searched for those genes assigned to immune-related terms or pathways. We annotated the 16 immune genes with rare disease associations by performing gene-based look ups at the online Mendelian inheritance in man (OMIM) and Orphanet websites. We also intersected the genomic coordinates of these same 16 immune genes with known SCZ risk loci using the findOverlaps() function from the GenomicRanges R package (v1.50.2). We set the 'maxgap' parameter to 1,000,000 and 100,000 to identify genes within 1Mb and 100Kb from a SCZ risk locus, respectively.

Phenome-wide association study. We conducted phenome-wide association studies (PheWASes) using the Open Targets Genetics platform. For each variant of interest, we extracted all phenotypes from the 'GWAS lead variants' label, which contained phenotypes/studies with a lead variant linked with the variant of interest either through linkage disequilibrium (r^2>0.5) or fine-mapping (i.e., appears in the 95% credible causal set).

Enrichment analysis of regulatory data. To determine whether IRF3 ChIP-seq peaks were enriched at SCZ risk loci, we utilized the GWAS analysis of regulatory or functional information enrichment with LD correction (GARFIELD) v2 software tool[49], available at https://www.ebi.ac.uk/birney-srv/GARFIELD/. We downloaded all available human IRF3 ChIP-seq data from the encyclopedia of DNA elements (ENCODE) project, which originated from five distinct experiments: two using GM12878 cells (ENCSR000DZX, ENCSR408JQO) and one each using HeLa-S3 (ENCSR000EDF), HepG2 (ENCSR000EEJ), and SK-N-SH (ENCSR422ZAO) cells. To ensure we were inputting high quality and reproducible regulatory data, for each experiment we downloaded .bed files containing only the irreproducible discovery rate (IDR) thresholded peaks. We excluded the data for HepG2 cells due to

an insufficient number of peaks relative to the other four datasets. We utilized the SCZ GWAS summary data from Trubestkoy et al. (2022)[21] and tested for enrichment using two different *p*-value thresholds to define SCZ risk loci: one set at genome-wide significance ($p<5x10^{-8}$) and the other using a less stringent threshold of $p<1x10^{-5}$. We used in-house R scripts (available on request) to format the IRF3 ChIP-seq peak and SCZ GWAS summary data ready for input to GARFIELD. We considered an annotation to be significantly enriched at SCZ risk loci by applying an adjusted *p*-value threshold ($p_{adj}<0.012$) corrected for the number of effective IRF3 ChIP-seq annotations using the garfield-Meff-Padj.R script provided as part of the GARFIELD software. We extracted all variants overlapping an enriched IRF3 ChIP-seq annotations using the 'garfield_extract_variants_overlapping_enriched_annotations.sh' bash script. Finally, we generated a network graph depicting the overlapping loci and variants using Cytoscape v3.10.1.

ACKNOWLEDGEMENTS

We would like to thank Dr. Kaur Alasoo and colleagues for their excellent eQTL catalogue resource.

DATA AVAILABILITY

All data generated by this study are available as Supplementary Tables or will be provided upon request from the corresponding author. All transcriptomic exposure data used in this study are publicly available either from the eQTL catalogue (https://www.ebi.ac.uk/eqtl/Data_access/) or the Genotype-Tissue Expression project (https://gtexportal.org/home/downloads/adult-gtex/overview). Full GWAS summary data for schizophrenia from Trubetskoy et al. (2022)[21] are publicly available from the Psychiatric Genomics Consortium (PGC) website (https://pgc.unc.edu/for-researchers/download-results/).

CONFLICTS OF INTEREST

All authors declare no conflicts of interest.

REFERENCES

1. van Os J, Kapur S. Schizophrenia. *Lancet* 2009; **374**(9690)**:** 635-645.

2. Häfner H, Maurer K, Löffler W, Fätkenheuer B, an der Heiden W, Riecher-Rössler A *et al.* The epidemiology of early schizophrenia. Influence of age and gender on onset and early course. *Br J Psychiatry Suppl* 1994; (23)**:** 29-38.

3. Patel KR, Cherian J, Gohil K, Atkinson D. Schizophrenia: overview and treatment options. *P t* 2014; **39**(9)**:** 638-645.

4. Hilker R, Helenius D, Fagerlund B, Skytthe A, Christensen K, Werge TM *et al.* Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register. *Biol Psychiatry* 2018; **83**(6)**:** 492-498.

5. Potkin SG, Kane JM, Correll CU, Lindenmayer JP, Agid O, Marder SR *et al.* The neurobiology of treatment-resistant schizophrenia: paths to antipsychotic resistance and a roadmap for future research. *NPJ Schizophr* 2020; **6**(1)**:** 1.

6. Laan W, Grobbee DE, Selten JP, Heijnen CJ, Kahn RS, Burger H. Adjuvant aspirin therapy reduces symptoms of schizophrenia spectrum disorders: results from a randomized, double-blind, placebo-controlled trial. *J Clin Psychiatry* 2010; **71**(5)**:** 520-527.

7. Nitta M, Kishimoto T, Müller N, Weiser M, Davidson M, Kane JM *et al.* Adjunctive use of nonsteroidal anti-inflammatory drugs for schizophrenia: a meta-analytic investigation of randomized controlled trials. *Schizophr Bull* 2013; **39**(6)**:** 1230-1241.

8. Miller BJ, Culpepper N, Rapaport MH. C-reactive protein levels in schizophrenia: a review and meta-analysis. *Clin Schizophr Relat Psychoses* 2014; **7**(4)**:** 223-230.

9. Wang AK, Miller BJ. Meta-analysis of Cerebrospinal Fluid Cytokine and Tryptophan Catabolite Alterations in Psychiatric Patients: Comparisons Between Schizophrenia, Bipolar Disorder, and Depression. *Schizophr Bull* 2018; **44**(1)**:** 75-83.

10. Jernbom Falk A, Galletly C, Just D, Toben C, Baune BT, Clark SR *et al.* Autoantibody profiles associated with clinical features in psychotic disorders. *Transl Psychiatry* 2021; **11**(1)**:** 474.

11. Benros ME, Pedersen MG, Rasmussen H, Eaton WW, Nordentoft M, Mortensen PB. A nationwide study on the risk of autoimmune diseases in individuals with a personal or a family history of schizophrenia and related psychosis. *Am J Psychiatry* 2014; **171**(2)**:** 218-226.

12. Ford H, Griffith S, Warren N, Swayne A, Blum S, Butzkueven H *et al.* Psychiatric manifestations of autoimmune encephalitis. *Autoimmun Rev* 2022; **21**(9)**:** 103145.

13. Hartwig FP, Borges MC, Horta BL, Bowden J, Davey Smith G. Inflammatory Biomarkers and Risk of Schizophrenia: A 2-Sample Mendelian Randomization Study. *JAMA Psychiatry* 2017; **74**(12)**:** 1226-1233.

14. Perry BI, Upthegrove R, Kappelmann N, Jones PB, Burgess S, Khandaker GM. Associations of immunological proteins/traits with schizophrenia, major depression and bipolar disorder: A bi-directional two-sample mendelian randomization study. *Brain Behav Immun* 2021; **97:** 176-185.

15. Prins BP, Abbasi A, Wong A, Vaez A, Nolte I, Franceschini N *et al.* Investigating the Causal Relationship of C-Reactive Protein with 32 Complex Somatic and Psychiatric Outcomes: A Large-Scale Cross-Consortium Mendelian Randomization Study. *PLoS Med* 2016; **13**(6)**:** e1001976.

16. Ligthart S, Vaez A, Võsa U, Stathopoulou MG, de Vries PS, Prins BP *et al.* Genome Analyses of >200,000 Individuals Identify 58 Loci for Chronic Inflammation and Highlight Pathways that Link Inflammation and Complex Disorders. *Am J Hum Genet* 2018; **103**(5)**:** 691-706.

17. Ligthart S. Commentary: CRP and schizophrenia: cause, consequence or confounding? *Int J Epidemiol* 2019; **48**(5)**:** 1514-1515.

18. Brown AS, Derkits EJ. Prenatal infection and schizophrenia: a review of epidemiologic and translational studies. *Am J Psychiatry* 2010; **167**(3)**:** 261-280.

19. Nielsen PR, Benros ME, Mortensen PB. Hospital contacts with infection and risk of schizophrenia: a population-based cohort study with linkage of Danish national registers. *Schizophr Bull* 2014; **40**(6)**:** 1526-1532.

20. Smatti MK, Cyprian FS, Nasrallah GK, Al Thani AA, Almishal RO, Yassine HM. Viruses and Autoimmunity: A Review on the Potential Interaction and Molecular Mechanisms. *Viruses* 2019; **11**(8).

21. Trubetskoy V, Pardiñas AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* 2022; **604**(7906)**:** 502-508.

22. Sekar A, Bialas AR, de Rivera H, Davis A, Hammond TR, Kamitaki N *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* 2016; **530**(7589)**:** 177-183.

23. Ursini G, Di Carlo P, Mukherjee S, Chen Q, Han S, Kim J *et al.* Prioritization of potential causative genes for schizophrenia in placenta. *Nat Commun* 2023; **14**(1)**:** 2613.

24. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun* 2018; **9**(1)**:** 224.

25. Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK *et al.* Gene map of the extended human MHC. *Nat Rev Genet* 2004; **5**(12)**:** 889-899.

26. Pepys MB, Hirschfield GM. C-reactive protein: a critical update. *J Clin Invest* 2003; **111**(12)**:** 1805-1812.

27. Alasoo K, Rodrigues J, Mukhopadhyay S, Knights AJ, Mann AL, Kundu K *et al.* Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nat Genet* 2018; **50**(3)**:** 424-431.

28. Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martín D *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* 2016; **167**(5)**:** 1398-1414.e1324.

29. Bossini-Castillo L, Glinos DA, Kunowska N, Golda G, Lamikanra AA, Spitzer M *et al.* Immune disease variants modulate gene expression in regulatory CD4(+) T cells. *Cell Genom* 2022; **2**(4)**:** None.

30. Theusch E, Chen YI, Rotter JI, Krauss RM, Medina MW. Genetic variants modulate gene expression statin response in human lymphoblastoid cell lines. *BMC Genomics* 2020; **21**(1)**:** 555.

31. Momozawa Y, Dmitrieva J, Théâtre E, Deffontaine V, Rahmouni S, Charloteaux B *et al.* IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat Commun* 2018; **9**(1)**:** 2427.

32. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A *et al.* Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat Genet* 2012; **44**(5)**:** 502-510.

33. Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, Lau E *et al.* Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* 2014; **343**(6175)**:** 1246949.

34. Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A *et al.* Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife* 2013; **2:** e00523.

35. Lappalainen T, Sammeth M, Friedländer MR, t Hoen PA, Monlong J, Rivas MA *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 2013; **501**(7468)**:** 506-511.

36. Gilchrist JJ, Makino S, Naranbhai V, Sharma PK, Koturan S, Tong O *et al.* Natural Killer cells demonstrate distinct eQTL and transcriptome-wide disease associations, highlighting their role in autoimmunity. *Nat Commun* 2022; **13**(1)**:** 4073.

37. Kasela S, Kisand K, Tserel L, Kaleviste E, Remm A, Fischer K *et al.* Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4+ versus CD8+ T cells. *PLoS Genet* 2017; **13**(3)**:** e1006643.

38. Naranbhai V, Fairfax BP, Makino S, Humburg P, Wong D, Ng E *et al.* Genomic modulators of gene expression in human neutrophils. *Nat Commun* 2015; **6:** 7545.

39. Buil A, Brown AA, Lappalainen T, Viñuela A, Davies MN, Zheng HF *et al.* Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat Genet* 2015; **47**(1)**:** 88-91.

40. Young AMH, Kumasaka N, Calvert F, Hammond TR, Knights A, Panousis N *et al.* A map of transcriptional heterogeneity and regulatory variation in human microglia. *Nat Genet* 2021; **53**(6)**:** 861-868.

41. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC *et al.* DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res* 2022; **50**(W1)**:** W216-w221.

42. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013; **45**(6)**:** 580-585.

43. Aragam KG, Jiang T, Goel A, Kanoni S, Wolford BN, Atri DS *et al.* Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat Genet* 2022; **54**(12)**:** 1803-1815.

44. Amberger JS, Bocchini CA, Scott AF, Hamosh A. OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res* 2019; **47**(D1)**:** D1038-d1043.

45. Andersen LL, Mørk N, Reinert LS, Kofod-Olsen E, Narita R, Jørgensen SE *et al.* Functional IRF3 deficiency in a patient with herpes simplex encephalitis. *J Exp Med* 2015; **212**(9)**:** 1371-1379.

46. Mørk N, Kofod-Olsen E, Sørensen KB, Bach E, Ørntoft TF, Østergaard L *et al.* Mutations in the TLR3 signaling pathway and beyond in adult patients with herpes simplex encephalitis. *Genes Immun* 2015; **16**(8)**:** 552-566.

47. Steiner J, Prüss H, Köhler S, Frodl T, Hasan A, Falkai P. Autoimmune encephalitis with psychosis: Warning signs, step-by-step diagnostics and treatment. *World J Biol Psychiatry* 2020; **21**(4)**:** 241-254.

48. Hiscott J, Pitha P, Genin P, Nguyen H, Heylbroeck C, Mamane Y *et al.* Triggering the interferon response: the role of IRF-3 transcription factor. *J Interferon Cytokine Res* 1999; **19**(1)**:** 1-13.

49. Iotchkova V, Ritchie GRS, Geihs M, Morganella S, Min JL, Walter K *et al.* GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat Genet* 2019; **51**(2)**:** 343-353.

50. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; **489**(7414)**:** 57-74.

51. Folkersen L, Gustafsson S, Wang Q, Hansen DH, Hedman Å K, Schork A *et al.* Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat Metab* 2020; **2**(10)**:** 1135-1148.

52. Kerimov N, Hayhurst JD, Peikova K, Manning JR, Walter P, Kolberg L *et al.* A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat Genet* 2021; **53**(9)**:** 1290-1299.

53. Kerimov N, Tambets R, Hayhurst JD, Rahu I, Kolberg P, Raudvere U *et al.* eQTL Catalogue 2023: New datasets, X chromosome QTLs, and improved detection and visualisation of transcript-level QTLs. *PLoS Genet* 2023; **19**(9)**:** e1010932.

54. Emanuele E, Martinelli V, Carlin MV, Fugazza E, Barale F, Politi P. Serum levels of soluble receptor for advanced glycation endproducts (sRAGE) in patients with different psychiatric disorders. *Neurosci Lett* 2011; **487**(1)**:** 99-102.

55. Miyashita M, Watanabe T, Ichikawa T, Toriumi K, Horiuchi Y, Kobori A *et al.* The regulation of soluble receptor for AGEs contributes to carbonyl stress in schizophrenia. *Biochem Biophys Res Commun* 2016; **479**(3)**:** 447-452.

56. Takeda M, Ohnuma T, Takeuchi M, Katsuta N, Maeshima H, Takebayashi Y *et al.* Altered serum glyceraldehyde-derived advanced glycation end product (AGE) and soluble AGE receptor levels indicate carbonyl stress in patients with schizophrenia. *Neurosci Lett* 2015; **593:** 51-55.

57. Dai HS, Caligiuri MA. Molecular Basis for the Recognition of Herpes Simplex Virus Type 1 Infection by Human Natural Killer Cells. *Front Immunol* 2018; **9:** 183.

58. Yolken R. Viruses and schizophrenia: a focus on herpes simplex virus. *Herpes* 2004; **11 Suppl 2:** 83a-88a.

59. Kotsiri I, Resta P, Spyrantis A, Panotopoulos C, Chaniotis D, Beloukas A *et al.* Viral Infections and Schizophrenia: A Comprehensive Review. *Viruses* 2023; **15**(6).

60. Arduino PG, Porter SR. Herpes Simplex Virus Type 1 infection: overview on relevant clinico-pathological features. *J Oral Pathol Med* 2008; **37**(2)**:** 107-121.

61. Klein HC. Silencing of Viral Elements: An Available Cure for Schizophrenia? *Front Psychiatry* 2017; **8:** 284.

62. Marcocci ME, Napoletani G, Protto V, Kolesova O, Piacentini R, Li Puma DD *et al.* Herpes Simplex Virus-1 in the Brain: The Dark Side of a Sneaky Infection. *Trends Microbiol* 2020; **28**(10)**:** 808-820.

63. Li J, Pang Y, Du Y, Xia L, Chen M, Fan Y *et al.* Lack of interferon regulatory factor 3 leads to anxiety/depression-like behaviors through disrupting the balance of neuronal excitation and inhibition in mice. *Genes Dis* 2023; **10**(3)**:** 1062-1074.

64. Liu Z, Xia L. E3 ligase RNF5 inhibits type I interferon response in herpes simplex virus keratitis through the STING/IRF3 signaling pathway. *Front Microbiol* 2022; **13:** 944101.

65. Braun E, Sauter D. Furin-mediated protein processing in infectious diseases and cancer. *Clin Transl Immunology* 2019; **8**(8)**:** e1073.

66. Di Carlo P, Punzi G, Ursini G. Brain-derived neurotrophic factor and schizophrenia. *Psychiatr Genet* 2019; **29**(5)**:** 200-210.

67. Võsa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet* 2021; **53**(9)**:** 1300-1310.

68. Saevarsdottir S, Olafsdottir TA, Ivarsdottir EV, Halldorsson GH, Gunnarsdottir K, Sigurdsson A *et al.* FLT3 stop mutation increases FLT3 ligand level and risk of autoimmune thyroid disease. *Nature* 2020; **584**(7822)**:** 619-623.

69. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics*, vol. 502018, pp 1112-1121.

70. Dönertaş HM, Fabian DK, Valenzuela MF, Partridge L, Thornton JM. Common genetic associations between age-related diseases. *Nat Aging* 2021; **1**(4)**:** 400-412.

71. Jin Y, Andersen G, Yorgov D, Ferrara TM, Ben S, Brownson KM *et al.* Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. *Nature genetics* 2016; **48**(11)**:** 1418-1424.

72. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; **88**(1)**:** 76-82.

73. Di Angelantonio E, Thompson SG, Kaptoge S, Moore C, Walker M, Armitage J *et al.* Efficiency and safety of varying the frequency of whole blood donation (INTERVAL): a randomised trial of 45?000 donors. *Lancet* 2017; **390**(10110)**:** 2360-2371.

74. Wei T, Simko V. R package 'corrplot': Visualization of a Correlation Matrix (Version 0.92). 2021.

75. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 2014; **10**(5)**:** e1004383.

76. Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurle MR *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet* 2020; **52**(10)**:** 1122-1131.

77. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J *et al.* Genomic atlas of the human plasma proteome. *Nature* 2018; **558**(7708)**:** 73-79.

78. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* 2016; **17**(1)**:** 122.

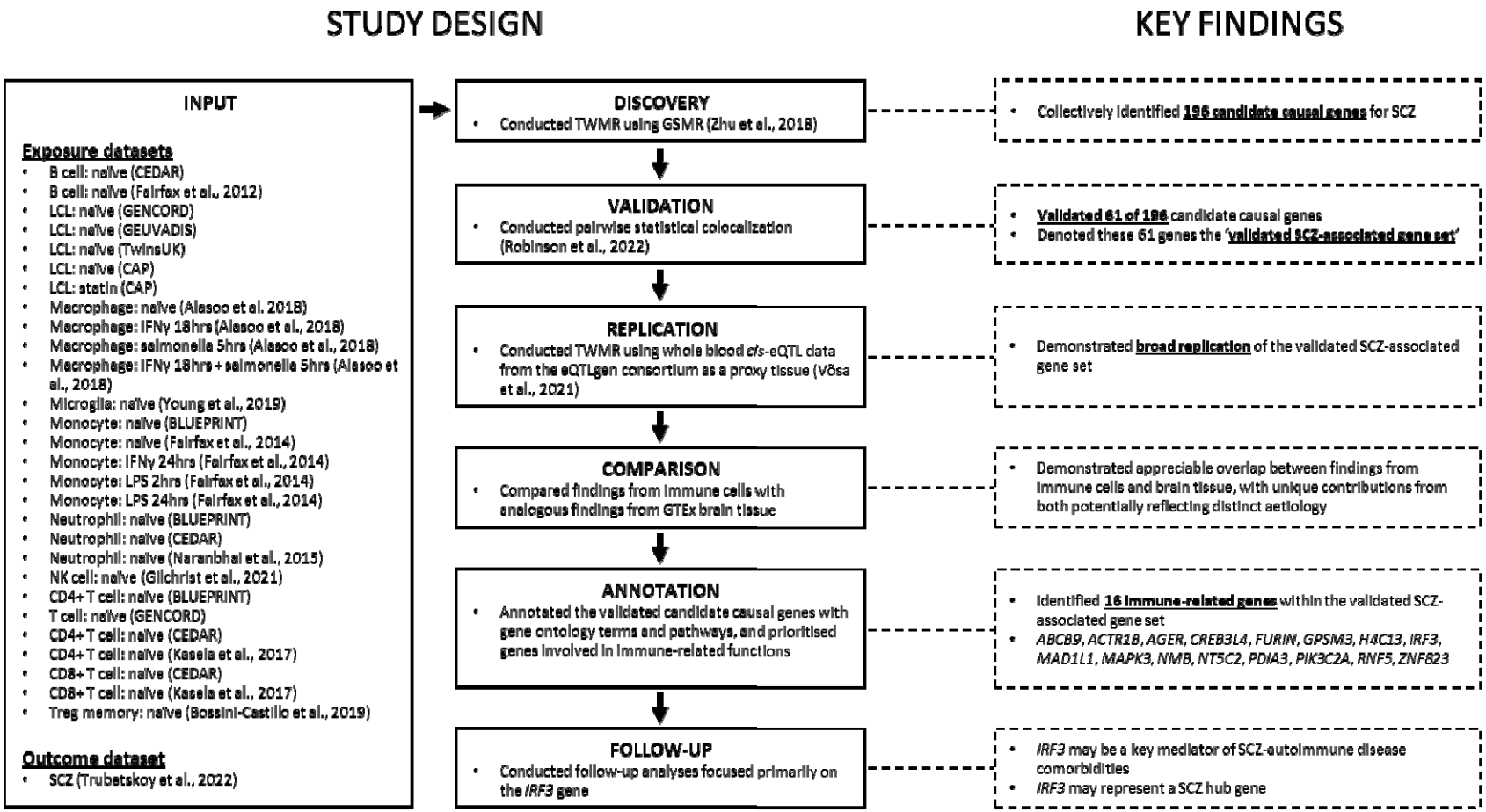

Figure 1. Flow diagram summarising the overall study design and key findings from each analysis step. GSMR: generalised summary-data-based Mendelian randomization, GTEx: genotype-tissue expression project, SCZ: schizophrenia, TWMR: transcriptome-wide Mendelian randomization.

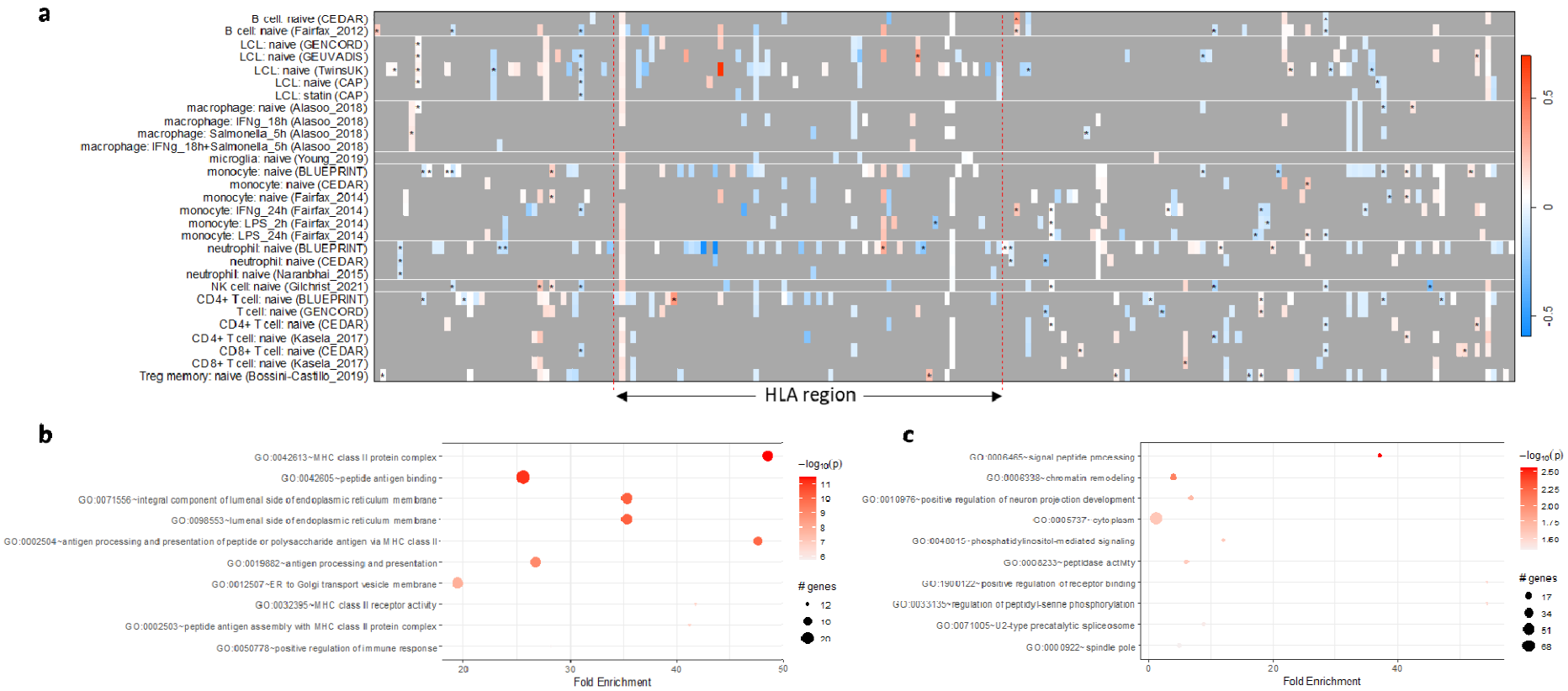

Figure 2. Transcriptome-wide Mendelian randomization analyses identifies schizophrenia-associated candidate genes. (a) Heatmap plotting the bxy (logOR) estimates of the causal effect of isolated immune cell gene expression exposures on schizophrenia risk (y-axis: *cis*-eQTL datasets; x-axis: genes). Gene expression exposures with robust evidence of colocalization (PP_{H4}>0.8) with schizophrenia risk are shown with asterisks. Vertical red dashed lines indicate the location of the human leukocyte antigen region (HLA). **(b,c)** Gene ontology term enrichment analyses with **(b)** and without **(c)** the HLA region genes included.

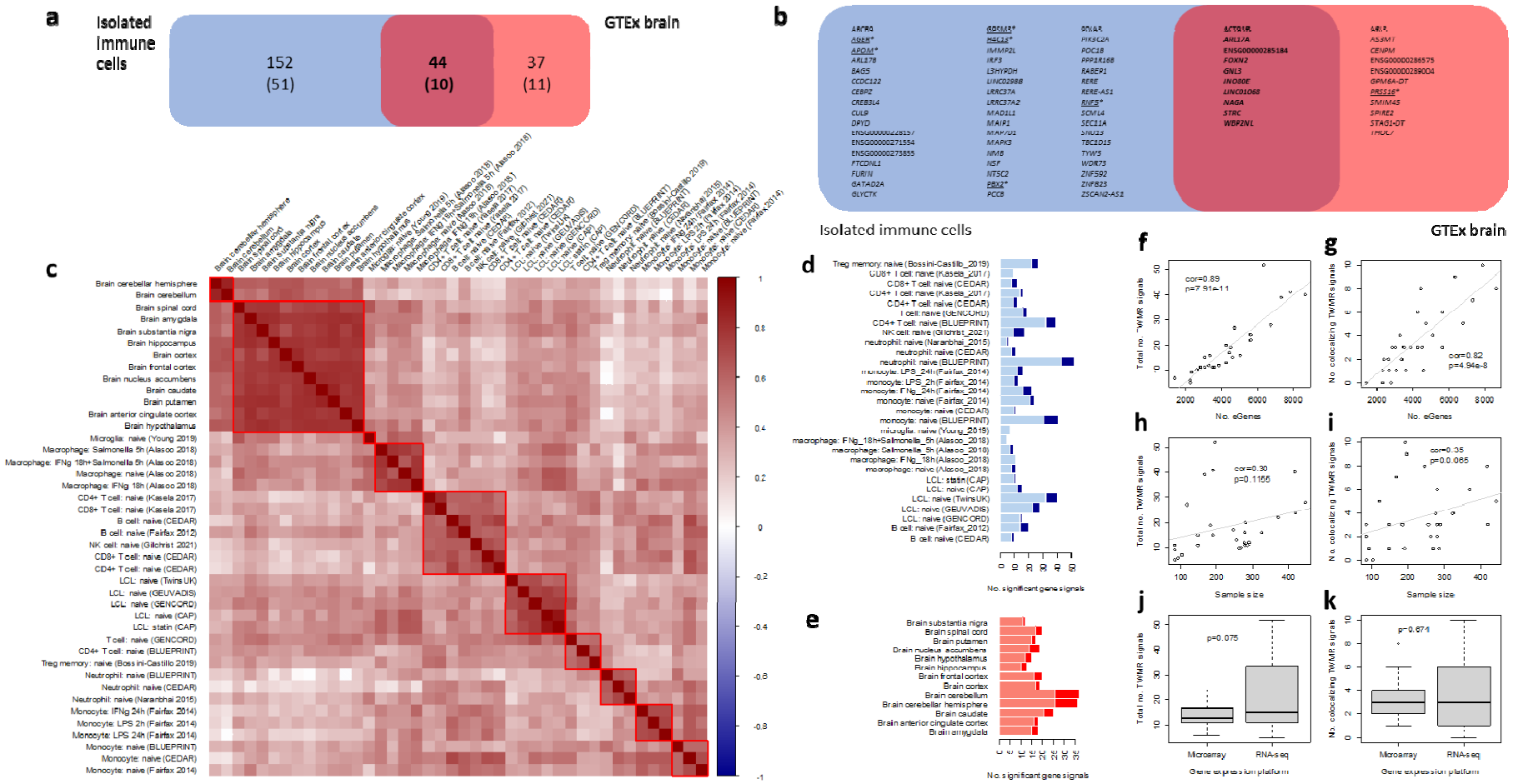

Figure 3. A comparison of transcriptome-wide MR findings from immune cell types and brain tissues demonstrates partial overlap. (**a**, **b**) Venn diagrams illustrating the unique and overlapping contributions of immune cell types and brain tissues to (**a**) the overall number of significant transcriptome-wide MR (TWMR) signals and (**b**) the colocalizing gene signals (genes from the validated SCZ-associated set are underlined and asterisked). (**c**) A symmetrical correlation (Pearson's *r*) matrix comparing GSMR estimates for shared genes for each pair of *cis*-eQTL datasets. Hierarchical clusters are indicated by red squares. (**d**, **e**) Stacked bar plots illustrating the number of significant TWMR signals identified in the (**d**) immune cell datasets and (**e**) brain tissues. The darker coloured segment of each stack indicates the number of colocalizing signals. (**f-k**) Plots illustrating the relationships between the overall number of TWMR signals and the number of colocalizing signals for each *cis*-eQTL dataset in relation to the (**f**, **g**) number of eGenes, (**h**, **i**) sample size, and (**j**, **k**) gene expression platform. Pearson's correlation coefficients and associated *p*-values are indicated, along with the corresponding lines of best fit.

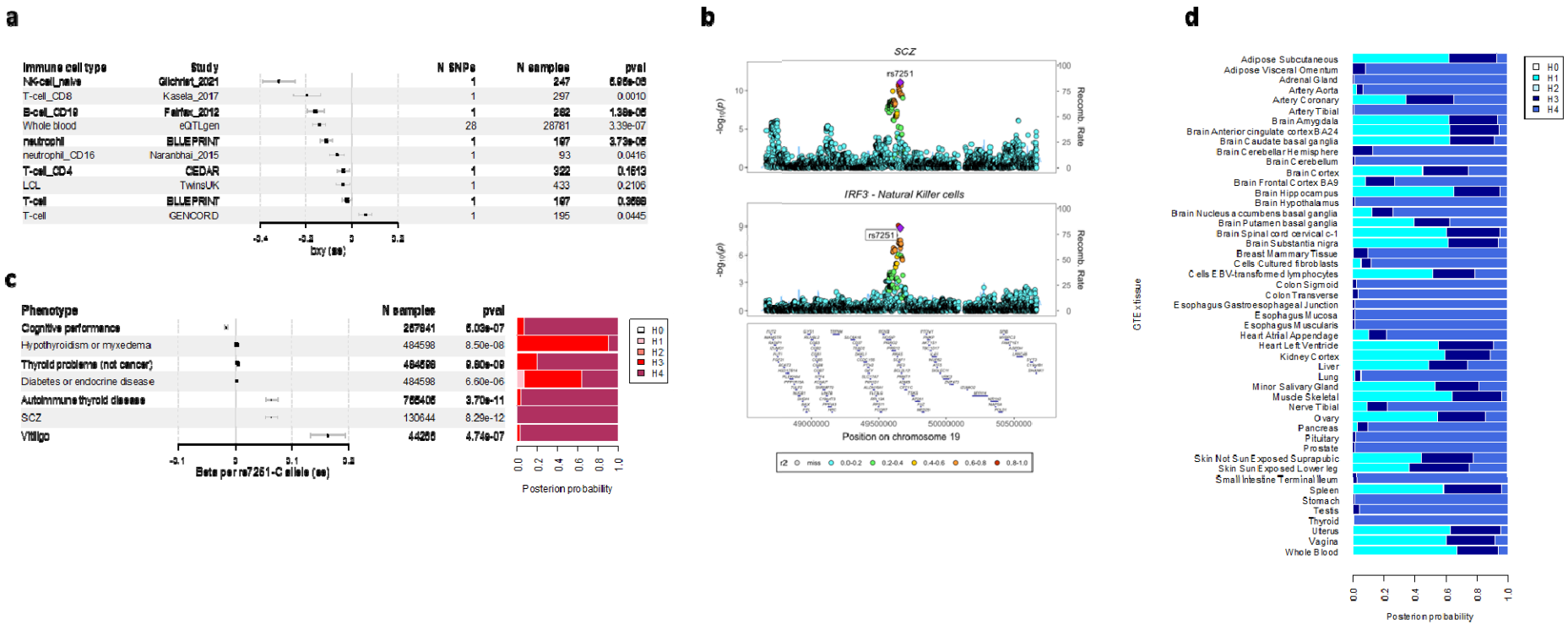

Figure 4. The rs7251 schizophrenia risk locus is associated with peripheral and central *IRF3* mRNA expression as well as risk of autoimmune diseases. (a) Forest plot depicting the bxy (logOR) estimates of the causal effect of immune cell and whole blood *IRF3* mRNA expression on schizophrenia risk. **(b)** Stacked regional association plot showing the genetic association signals at the *IRF3* locus for schizophrenia (top) and *IRF3* mRNA expression in natural killer cells (bottom). **(c)** Forest plot showing the results from a PheWAS of rs7251 and a stacked bar plot depicting PPs of colocalization between schizophrenia and each rs7251-associated phenotype. **(d)** Stacked bar plot depicting posterior probabilities (PP) of statistical colocalization between *IRF3* mRNA expression in 49 GTEx tissues and schizophrenia risk. We used coloc, which returns the following PPs: PP_{H0} (no causal variant), PP_{H1} (causal variant for *IRF3* mRNA expression only), PP_{H2} (causal variant for schizophrenia only), PP_{H3} (two distinct causal variants), and PP_{H4} (one shared causal variant). We consider PP_{H4}>0.8 as robust evidence of colocalization.

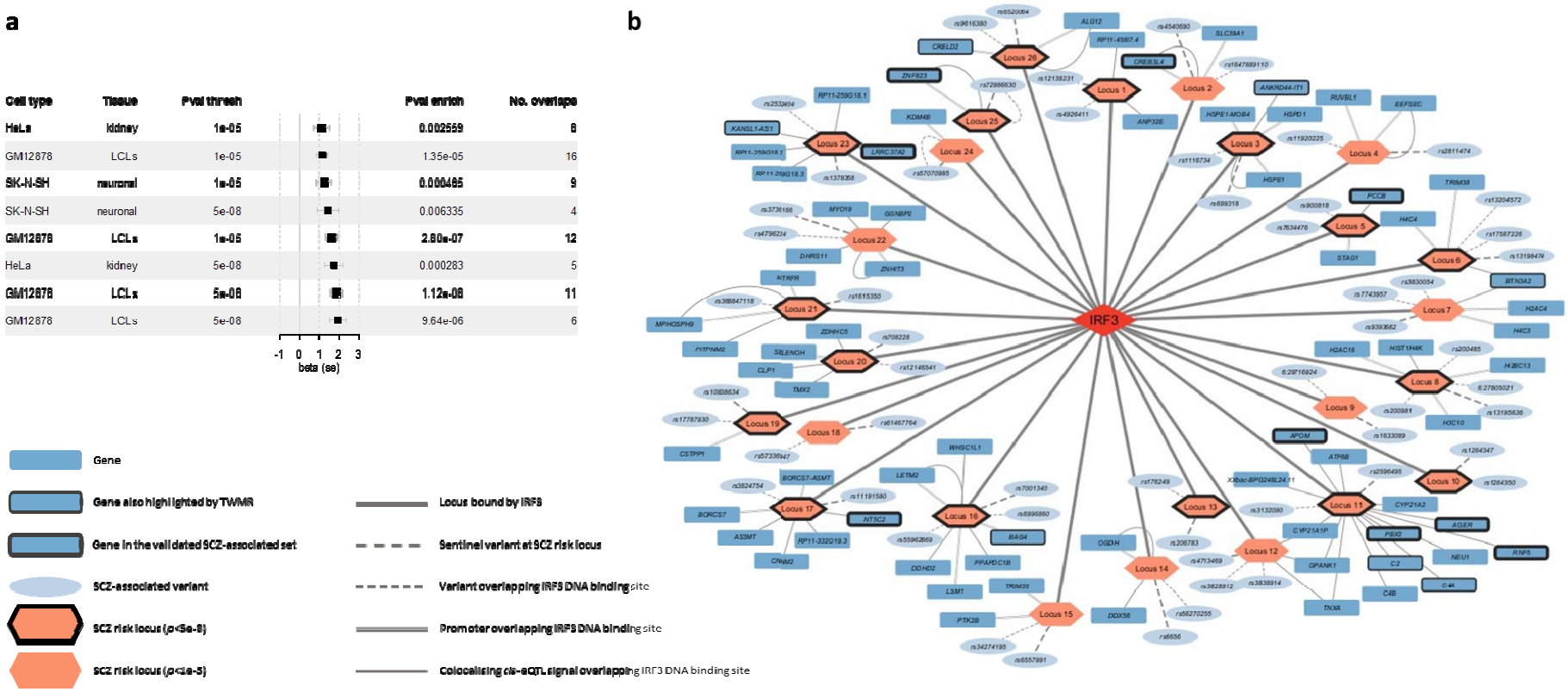

Figure 5. Interferon regulatory transcription factor 3 (IRF3) DNA binding sites are enriched at known schizophrenia risk loci. (a) Forest plot showing beta estimates (se) from GARFIELD enrichment analyses. **(b)** Graph showing the 26 schizophrenia risk loci with an overlapping IRF3 DNA-binding site as evidenced by chromatin immunoprecipitation-sequence (ChIP-seq) data. Colocalising (PP_{H4}>0.8) *cis*-eQTL signals were from whole blood (eQTLgen consortium). SCZ: schizophrenia.