

SC-VAR: a computational tool for interpreting polygenic disease risks using single-cell epigenomic data

Gefei Zhao^{1,2} and Binbin Lai^{1,2,3,4,*}

¹Institute of Medical Technology, Peking University Health Science Center, 38 Xueyuan Rd, Hai Dian Qu, Beijing 100191, China

²Biomedical Engineering Department, Institute of Advanced Clinical Medicine, Peking University, 5 Yiheyuan Rd, Haidian District, Beijing 100191, China

³Department of Dermatology and Venerology, Peking University First Hospital, 8 Xishiku Ave, Xicheng District, Beijing 100191, China

⁴State Key Laboratory of Molecular Oncology, Peking University International Cancer Institute, 5 Yiheyuan Rd, Haidian District, Beijing 100191, China

*Corresponding author: Binbin Lai, Biomedical Engineering Department, Institute of Advanced Clinical Medicine, Peking University, 5 Yiheyuan Rd, Haidian District, Beijing 100191, China. E-mail: laib@bjmu.edu.cn

Abstract

Motivation: One major challenge of interpreting variants from genome-wide association studies (GWAS) of complex traits or diseases is how to efficiently annotate noncoding variants. These variants influence gene expression by disrupting cis-regulatory elements (CREs), whose spatial and cell-type specificity are not adequately captured by conventional tools like multi-marker analysis of genomic annotation. Current methods either rely on linear proximity to genes or quantitative trait locus (QTL) data yet fail to integrate single-cell epigenomic information for a comprehensive annotation. **Results:** We present SC-VAR, a novel computational tool designed to enhance the interpretation of disease-associated risks from GWAS using single-cell epigenomic data. SC-VAR leverages single-cell epigenomic data to predict functional outcomes including risk genes, pathways, and cell types for both coding and noncoding disease-associated variants. We demonstrate that SC-VAR outperforms state-of-the-art methods by predicting more validated disease-related genes and pathways for multiple diseases. Additionally, SC-VAR identifies cell types that are susceptible to disease, along with their specific CREs and target genes linked to risk. By capturing a broad range of disease risks across human tissues at distinct developmental stages, SC-VAR could enhance our understanding of disease mechanisms in complex tissues across different life stages.

Keywords: single-cell epigenomics; polygenic disease risks; genome-wide association studies (GWAS); noncoding variant annotation; disease relevance score

Introduction

Genome-wide association studies (GWAS) have identified thousands of variants associated with complex human traits [1–3]. However, translating this knowledge into clinical insights or disease-relevant biology remains challenging, primarily because more than 90% of GWAS variants fall in noncoding regions of the genome [4, 5]. These variants might affect gene functions by accumulating in cis-regulatory elements (CREs) disrupting binding of transcription factors (TFs) and then affecting gene expression levels [6].

Several methods have been developed to annotate CREs loci, such as GREAT [7] and loci2path [8]. GREAT [7] annotates CREs by analyzing their functional significance related to a gene ontology or a pathway based on the proximity of genes in the genome. The loci2path [8] links loci to genes by incorporating eQTL information. In addition, some models and methods have been developed to convert variant-level risk significance from GWAS to gene- and gene-set-level risk significance, such as multi-marker analysis of genomic annotation (MAGMA) [9], which performs both gene-level association testing and pathway enrichment analysis to uncover potential biological mechanisms underlying complex traits. However, MAGMA employs a window-based strategy to assign variants to nearby genes. Although MAGMA can extend the region to include additional variants, it has obvious limitations in dealing with variants within noncoding regions, owing to the

arbitrary setting of the window size: a large genomic window may include many regions without functional connectivity while a small window would miss true distal CREs that regulate genes via chromatin looping but are far away from the gene in the linear distance of genome. Some extensions have been proposed to extend gene variant annotation to provide more powerful functionality using transcriptomics networks (e.g. nMAGMA [10]) or eQTL data (e.g. E-MAGMA [11]). These methods have been widely used to identify novel gene loci and functional pathways that affect complex traits and diseases [12–14].

Besides gene variant annotation approaches based on genomic loci, gene network, and eQTL, epigenomic-based approach is another promising one. It has at least two advantages in explaining the variants in noncoding regions by exploring their regulatory functions. Firstly, epigenomic assays help link distal CREs to target genes reliably and consequently improve the capability to annotate distal disease-related SNPs. For example, one extension of MAGMA based on Hi-C assays (H-MAGMA) utilizes distal SNPs to explore disease risks by calling SNP-gene links via predicted chromatin loopings of CREs and genes [15]. Secondly, advanced epigenomic techniques such as single-cell genomic assays depict CREs in cell-type-specific manner associated with different tissues and different stages of life time, providing great opportunities to do spatial-temporal interpretation of disease-related SNPs [16, 17] in tissues with

Received: December 9, 2024. **Revised:** February 11, 2025. **Accepted:** March 1, 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

high level of cellular heterogeneity, which cannot be achieved by gene network-based or eQTL-based approaches. Recent single-cell epigenomic atlas studies employed methods to assign risk scores to different cell types by estimating the heritability of SNPs residing within cell-type-specific CREs [18–20]. However, the method is still limited to use single-cell epigenomic or single-cell multiome data to annotate noncoding variants in terms of exploring the affected genes and pathways, which is crucial for understanding the molecular mechanisms underlying the genetic basis of polygenetic diseases.

In addition to gene and pathway insights, identifying cell-type-specific contributions is also critical for understanding disease etiology and can provide valuable information for new therapeutic strategies. Methods have been developed to prioritize cell-type risks related to diseases based on single-cell RNA sequencing, such as scRDS [21], which mainly account for variants within gene bodies, but the methods for prioritizing cell-type risks based on noncoding variants are not yet available.

Here we report a human variant interpreter, named SC-VAR, based on single-cell epigenomic data. This tool aims to interpret disease-related risks on three layers: risk genes (individual genes and gene sets), CREs, and cells (individual cells and cell type). SC-VAR extends conventional SNP annotation by linking noncoding GWAS variants to target genes via single-cell ATAC-seq data and enabling the identification of risk SNPs located in cell-type-specific CREs. Unlike previous methods that primarily annotate variants within gene bodies or rely on arbitrary genomic window settings, SC-VAR leverages cell-type and stage-specific SNP-CRE-Gene interactions, offering a refined, biologically informed approach to genome annotation that more accurately reflects the complexity of gene regulation in human disease. When applied with single-cell multi-omics data, SC-VAR further strengthens these linkages by correlating CRE activity with gene expression.

We applied SC-VAR on complex disorders with 12 single-cell epigenomic and multiome datasets from human tissues to examine several complex disorders, including skin diseases, gastrointestinal tract diseases, heart diseases, psychiatric, and neurodegenerative diseases. By correctly assigning more SNPs in CREs, SC-VAR was able to identify more reliable SNP-gene links and detected more risk genes than MAGMA. Several risk genes detected by SC-VAR have been verified to be functionally associated with disease in previous research [22–24]. SC-VAR further highlighted developmental stage-specific risk genes and pathways relevant to psychiatric disorders. By analyzing risk genes and CREs, SC-VAR identifies high-risk cell types within complex tissues which are unexplored by previous GWAS-based methods.

In summary, SC-VAR advances the field by linking single-cell epigenomic data with genetic risk, delivering a robust framework for interpreting the impact of noncoding variants on disease biology across different tissues and developmental stages. This method improves understanding of genetic influences on diseases and provides valuable insights into complex genetic diseases from both gene and cell type perspectives. In summary, we provide a powerful tool to interpret genome-wide risks of SNPs for complex genetic traits or diseases with single-cell epigenomic data.

Materials and methods

Genome-wide association studies dataset

The complete set of summary statistics from the GWAS for schizophrenia (SCZ) was downloaded from the fastGWA data

portal [25] (<https://yanglab.westlake.edu.cn/data.html>), which was generated by applying genome-wide association analysis to traits on individuals of European ancestry in the UK Biobank. The dataset contains a total number of 11 842 642 SNPs associated with traits for SCZ. To expand the application scenarios of our method, we additionally used publicly available GWAS summary datasets from FINNGEN R9 version [26]. These datasets include the phenotypes: Attention Deficit Hyperactivity Disorder (ADHD), Bipolar Disorder (BD), Obsessive-Compulsive Disorder (OCD), Alzheimer's Disease (AD), Parkinson's Disease (PKS), Autism Spectrum Disorder (ASD), Inflammatory Bowel Disease (IBD), Eczema, and Atrial Fibrillation (AF). Detailed information about these datasets is provided in [Supplementary Table S1](#).

As publicly available GWAS summary statistics were used, no data points were excluded from analysis, and no statistical methods were employed to predetermine the sample size. All data are aligned and annotated or convert to hg38 reference genome.

Single-cell epigenomic dataset

The scATAC-seq of adult brain data and the multi-omics data of the human neocortex are available in NIH GEO under accession numbers GSE147672 [27] and GSE204684 [28]. The scATAC-seq of fetal brain and heart data can be downloaded from Descartes database (<https://descartes.brotmanbaty.org/>). The scATAC-seq dataset of human peripheral blood mononuclear cells (PBMCs) is freely available from 10X Genomics (<https://www.10xgenomics.com/datasets>). The scATAC-seq dataset of human sigmoid colon, small intestine, transverse colon, skin (sun-exposed and normal), and heart data is available under accession number GSE184462 [19].

Disease risk genes

To validate the accuracy and interpretability of our method for identifying risk genes for SCZ, we collected a list of risk genes from various public resources including CTD, dbGAP, GWASdb, GAD, DisGeNET, PheGenI, GeneRIF, and HuGE. Genes from CTD were selected based on manually curated disease–gene associations from literatures while genes from dbGAP, GWASdb, PheGenI, and GAD were selected based on GWAS or GWAS Catalog SNP-Phenotype associations. The genes selected from GeneRIF and HuGE navigator were obtained based on the genetic association by text-mining. DisGeNET integrates data from expert curated repositories, GWAS catalogues, animal models, and the scientific literature. In total, we obtained a list of 2017 candidate risk genes associated with SCZ after removing duplicates. For IBD, we collected a list of risk genes from IBDDb which is a manually curated and text-mining-enhanced database. The database provides a unique combination of 289 IBD-related genes.

We collected a list of AF risk genes from public resources, similar to those used for SCZ, including CTD, dbGAP, GWASdb, GAD, GeneRIF, and HuGE. After removing duplicates, we compiled a list of 782 candidate risk genes associated with AF after removing duplicates. The same for eczema risk genes from public resources. And we compiled a list of 683 candidate risk genes after removing duplicates ([Supplementary Table S2](#)).

Design of SC-VAR method

SC-VAR takes GWAS data and single-cell epigenomic data (scATAC-seq or single-cell multi-omic ATAC + RNA sequencing data) as input and includes two modules to annotate the GWAS variants. The first module is designed to perform single-cell epigenome-coupled multi-marker analysis of genomic annotation (sce-MAGMA) to predict risk genes and gene sets associated

Table 1. Selected risk biological processes detected by sce-MAGMA

| Stage | GOBP | No. of genes | P-value | Reference ^a |
|-------|---|--------------|---------|--|
| Adult | Regulation of synaptic vesicle exocytosis | 50 | 6.8E-4 | Egbujo et al. [46] |
| | Regulation of DNA metabolic process | 378 | 1.9E-3 | Parellada et al. [73] |
| | Regulation of anoikis | 24 | 3.4E-3 | Jarskog et al. [74] |
| | Response to amyloid beta | 51 | 4.3E-3 | Albertini et al. [44] |
| Fetal | Regulation of metal ion transport | 365 | 1.4E-3 | Schoonover et al. [75]; Mealer et al. [76] |
| | Macromolecule catabolic process | 1226 | 1.7E-3 | Trubetskoy et al. [2] |
| | sno(s)RNA metabolic process | 14 | 1.8E-3 | Gibbons et al. [45] |
| | RNA phosphodiester bond hydrolysis | 136 | 2.0E-3 | Ermakov et al. [77]; Parshukova et al. [78] |
| | Regulation of synaptic vesicle exocytosis | 49 | 2.5E-3 | Egbujo et al. [46]; Osimo et al. [47]; Radhakrishnan et al. [48] |

^aReferences with evidence related to SCZ.

with disease or phenotype of interest. This module consists of two steps. The first step establishes the CRE peak-gene connections via the single-cell data and then generates the SNP-gene annotation by aggregating SNPs within the CREs and gene bodies. This step provides a whole genome multi-marker annotation. In the second step, sce-MAGMA applies MAGMA on the SNP-gene annotation to prioritize the disease risks for genes and gene sets. In Module 2, SC-VAR employs an algorithm named sce-DRS (single-cell epigenome-based disease relevance score) to integrate the activity of risk CREs of single cells with the significance of genetic variants from GWAS to prioritize disease-associated cells and cell types within complex tissues at single-cell resolution, independent of annotated cell types. The whole workflow is illustrated in Fig. 1.

Module 1. Single-cell epigenome-coupled multi-marker analysis of genomic annotation

Step 1. Linking risk variants to target genes: To determine which SNPs are located within the CREs linked to the genes, we predict the CRE-gene links from single-cell ATAC-seq or multi-omics (single-cell ATAC and RNA sequencing) data. Specifically, for scATAC-seq data we employ cicero [29] method to calculate co-accessibility of the peaks and construct cis-regulatory networks. For a given gene, the peaks linking to the peak at the promoter region of the gene within the cis-regulatory network are inferred as linked to the gene. For single-cell ATAC and RNA data, we use LinkPeaks function in Signac package, which implements the method based on the correlation between peak accessibility and gene expression described in SHARE-seq [30], to infer peak-gene associations.

We assign two parts of SNPs to genes. The first part of SNPs is those located within the gene body region, which are exactly what traditional MAGMA method include. The second part of SNPs is located within the CREs linked to the genes. Both parts of SNPs are assigned to the gene, and the SNP-gene annotation file is constructed after removing redundant SNPs for downstream analysis. This SNP-gene annotation includes both coding and noncoding variants linked to genes.

Step 2. Running MAGMA gene and gene set analysis: With the SNP-gene annotation containing both coding and noncoding variants, we perform MAGMA for gene and gene-set layers to predict risk genes and risk biological processes or pathways. Briefly, the gene analysis employs a multiple linear principal components regression model on gene-SNP matrix and uses an F-test to compute the gene P-value. The gene-set analysis tests whether the genes in a gene set are strongly associated with the disease or phenotype of interest within a regression framework. The detailed methods for MAGMA are described previously [9].

Module 2. Prioritizing disease association of cells and cell types based on single-cell epigenome-based disease relevance score

To identify cell or cell types relevant to trait or disease of interest, we developed the algorithm, called sce-DRS. The notion and definition of sce-DRS is modified from scDRS [21] which was proposed to measure disease relevance of single cells using single-cell RNA-sequencing data. Briefly, the sce-DRS method tests whether GWAS-weighted accessibility of risk peaks for cells is stronger than other cells, at both individual cell and cell type (a cluster of cells) level.

Specifically, sce-DRS creates a collection of risk peaks those were identified as being associated with risk genes from Peak-SNP-Gene annotation, as output from the first module sce-MAGMA. To maximize power, each risk peak was weighted by its GWAS Z score and inversely weighted by its technical noise level. We took the largest SNPs Z score under the peaks as the weight. Therefore the putative disease risk peak set was $P = \{1, 2, \dots, n_{\text{peaks}}\}$ with Z_p as the weight of each peak.

Then, we calculated the disease association score for each cell $c = 1, \dots, n_{\text{cell}}$ as follows:

$$\text{Disease score} : S_c = \frac{\sum_{p \in P} Z_p \sigma_{\text{tech},p}^{-1} X_{cp}}{\sum_{p \in P} Z_p \sigma_{\text{tech},p}^{-1}}$$

where X_{cp} denotes the cell-peak matrix of single-cell ATAC-seq data. The weights vector of the risk peaks is $\{Z_p\}_{p \in P}$ and Z_p is defined as the largest Z score of the SNPs located within the peak p . $\sigma_{\text{tech},p}^2$ is a vector that represents the technical variance of each CRE across the samples. We estimate the technical variance similarly to previous works [31, 32]. Specifically, we compute the mean and variance for each CRE across all cells in the original non-log-transformed space. Next, we fit a nonlinear trend to the log10-scale variance/mean relationship using local polynomial regression. The estimated trend models the expected technical variance based on the mean accessibility while the observed variance values above this expected trend reflect biological variance. Given this trend, the proportion of technical variance is computed as the ratio of predicted technical variance to observed variance (in the original space). Finally, the technical variance of the log-transformed data $\sigma_{\text{tech},p}^2$ is computed as the product of the variance of the CRE in log-transformed data and the proportion of technical variance as estimated above.

After obtaining epigenomic-based cell disease scores, sce-DRS tests whether a cell or cell type is significantly associated with disease using the Monte Carlo (MC) test. sce-DRS constructs a control set by randomly selecting peaks matching the mean accessible level and variance of the risk peaks calculated across all cells and

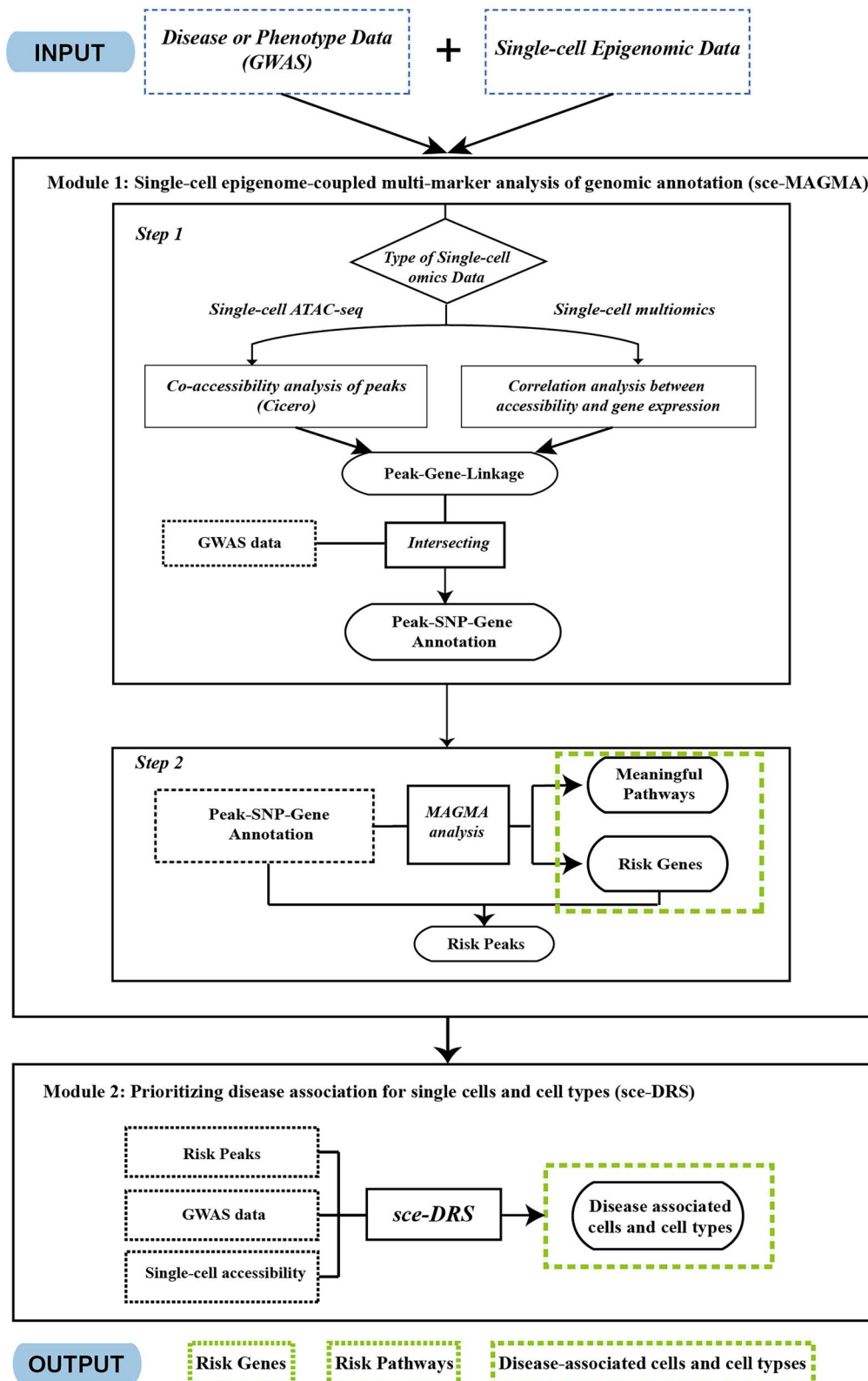


Figure 1. Flowchart of the SC-VAR analytical framework. The process starts with the input of single-cell omics data along with GWAS information related to diseases or phenotypes. Then, SC-VAR employs two modules. The first module sce-MAGMA predicts risk genes, pathways, and CREs. The second module sce-DRS prioritizes disease association for single cells and cell types. The major output of SC-VAR includes risk genes, pathways, and disease-associated cells and cell types.

let $\pi(p)$ be the corresponding matched control peaks. The weights vector of the control risk peak is $\{Z_{\pi(p)}\}_{p \in P}$. Let B be the number of MC samples of control sets (default 1000). Then for each cell,

$$\text{Control score : } S_{cb}^{\text{ctrl}} = \frac{\sum_{p \in P} Z_{\pi(p)} \sigma_{\text{tech},p}^{-1} X_{cp}}{\sum_{p \in P} Z_{\pi(p)} \sigma_{\text{tech},p}^{-1}}, \quad \forall b \in \{1, \dots, B\}$$

For each cell c compute cell-level P -values using a MC test based on the empirical distribution of the pooled normalized control scores.

$$p_c = \frac{1 + \sum_{c=1}^{n_{\text{cell}}} \sum_{b=1}^B \mathbb{I}(S_c \leq S_{cb}^{\text{ctrl}})}{1 + n_{\text{cell}} B}$$

For each cell types, let t be the top 5% quantile of the disease score of cells from the given cell type as the test statistic, and $t_1^{\text{ctrl}}, \dots, t_B^{\text{ctrl}}$ be the same test statistics from the control score of the same cells. The MC P value for cell type level can be written as:

$$p_c^{\text{mc}} = \frac{1 + \sum_{b=1}^B \mathbb{I}(t \leq t_b^{\text{ctrl}})}{1 + B}$$

The main feature of sce-MAGMA is to map noncoding variants to CREs and allocate them to target genes, which efficiently expands the range of the annotated variants than traditional “window-based” strategy. sce-MAGMA predicts CRE-gene links via either co-accessibility among enhancer and promoter peaks from scATAC-seq or the correlation between peak accessibility and gene expression from single-cell multiome data. With more meaningful SNPs assigned to genes, sce-MAGMA is expected to identify more candidate risk genes and gene sets compared to “window-based” strategy. Since some risk peaks and risk genes might be cell-type-specific, another feature of SC-VAR is to identify which cells or cell types are highly associated with the disease. It calculates the weighted accessibility score of risk peaks for individual cells to evaluate their association with the disease. More details about the model and algorithm of SC-VAR can be found in the Materials and methods section. The SC-VAR package has been released on the Pypi website (<https://pypi.org/project/sc-var/>) and a manual detailing how to use SC-VAR is available on GitHub (https://github.com/gefeiz/sc_var/).

Integration of single-cell multiome datasets

Integration of single-cell multiome datasets from six developmental stages was performed using the merge function in Seurat [31] and Signac [33] packages.

Pseudotime analysis

Pseudotime analysis was performed using scFates [34] package, which combines graph-based algorithms with probabilistic models to assign pseudotime values to each cell. Starting from a root population, the pseudotime analysis revealed distinct branching events corresponding to key developmental milestones.

Gene ontology analysis

Gene ontology analysis was performed using g:Profiler [35]. Input gene lists were derived from differential risk gene analyses between key cell types, while an adjusted P -value threshold (FDR < 0.05) was applied to filter significant terms.

Estimating type 1 error rate for single-cell epigenome-coupled multi-marker analysis of genomic annotation

We followed the approach described in MAGMA literature [9] to estimate the type 1 error rate for sce-MAGMA and compared with

MAGMA. Specifically, random perturbation of SCZ SNPs was performed and sce-MAGMA was applied on perturbed SNPs. Under each perturbation, type 1 error rate was defined as percentage of predicted genes.

Results

SC-VAR efficiently assigns SNPs within distal cis-regulatory elements to target genes

To evaluate the capability of SC-VAR, we applied it to GWAS data for several different complex diseases (see Materials and methods section), including SCZ, AF, eczema, and IBD. For each disease, we downloaded single-cell ATAC-seq data with profiled single-cell chromatin accessibility landscapes for different disease-related human tissues: brain cerebrum for SCZ, heart for AF, skin for eczema, and digestive system (PBMC, colon, and intestine) for IBD. The data sources are outlined in the Materials and methods section, and a summary of the data is presented in [Supplementary Fig. S1A](#).

We first examined SC-VAR’s ability to link SNPs in CREs outside gene body regions to their target genes. The differences between the sce-MAGMA and MAGMA strategies are illustrated in [Fig. 2A](#). In the adult cerebrum dataset for SCZ, for instance, sce-MAGMA assigned 63 (23.5%) SNPs located in the distal peaks to the target gene SLC6A1 based on co-accessibility between the distal peaks and the peak at the promoter region ([Fig. 2B](#)). In comparison, we also used traditional MAGMA method to assign SNPs with two commonly used parameter settings [9]. Traditional MAGMA assigned SNPs using only gene body regions or a 10 kb window around each gene ([Fig. 2A](#)). Take the result of SCZ data as an example, we found that sce-MAGMA could annotate additional distal SNPs in a high fraction of annotated genes. Specifically, in the analysis of adult cerebrum data, sce-MAGMA assigned additional distal SNPs for 13 793 genes (61.8%) compared to MAGMA, of which 2532 genes (11.3%) were exclusively annotated by sce-MAGMA ([Fig. 2C](#)). In fetal brain cerebrum dataset, sce-MAGMA assigned additional distal SNPs for 19 436 (94.9%) genes, of which 689 (11.3%) genes were exclusively annotated by sce-MAGMA ([Fig. 2D](#)). SNPs shared by both methods were generally within gene bodies, while the MAGMA-derived unique SNPs were located within extended 10 kb region upstream of or downstream from the gene body, but these SNPs were not located within the CRE peaks linked to genes. ([Fig. 2A](#)).

We further tested SC-VAR on other studied diseases. Results showed that SC-VAR significantly expanded the range of assigned SNPs for a large proportion of genes across most datasets (type 2 and type 3 genes in [Fig. 2E](#)). We also examined the significance of these SNPs, especially for those in distal CREs. The results showed that sce-MAGMA annotations for distal SNPs had a consistently higher density in regions with larger logarithm P -values for all datasets compared to MAGMA 10 kb extension annotations, suggesting that our method captured more distal SNPs with significant association with diseases.

Next, we assessed an alternative SC-VAR strategy that leverages correlation between CRE accessibility and gene expression, specifically for single-cell multiome ATAC and RNA data (see Materials and methods section). We downloaded six single-cell multiome data of human neocortex from different developmental stages [28] and applied sce-MAGMA analysis to SCZ-related GWAS data. Results showed that sce-MAGMA consistently assigned additional SNPs in distal CREs for the majority of genes (type 2 and type 3 genes in [Fig. 2G](#)), uniquely annotating 1.5%–2.1% of genes

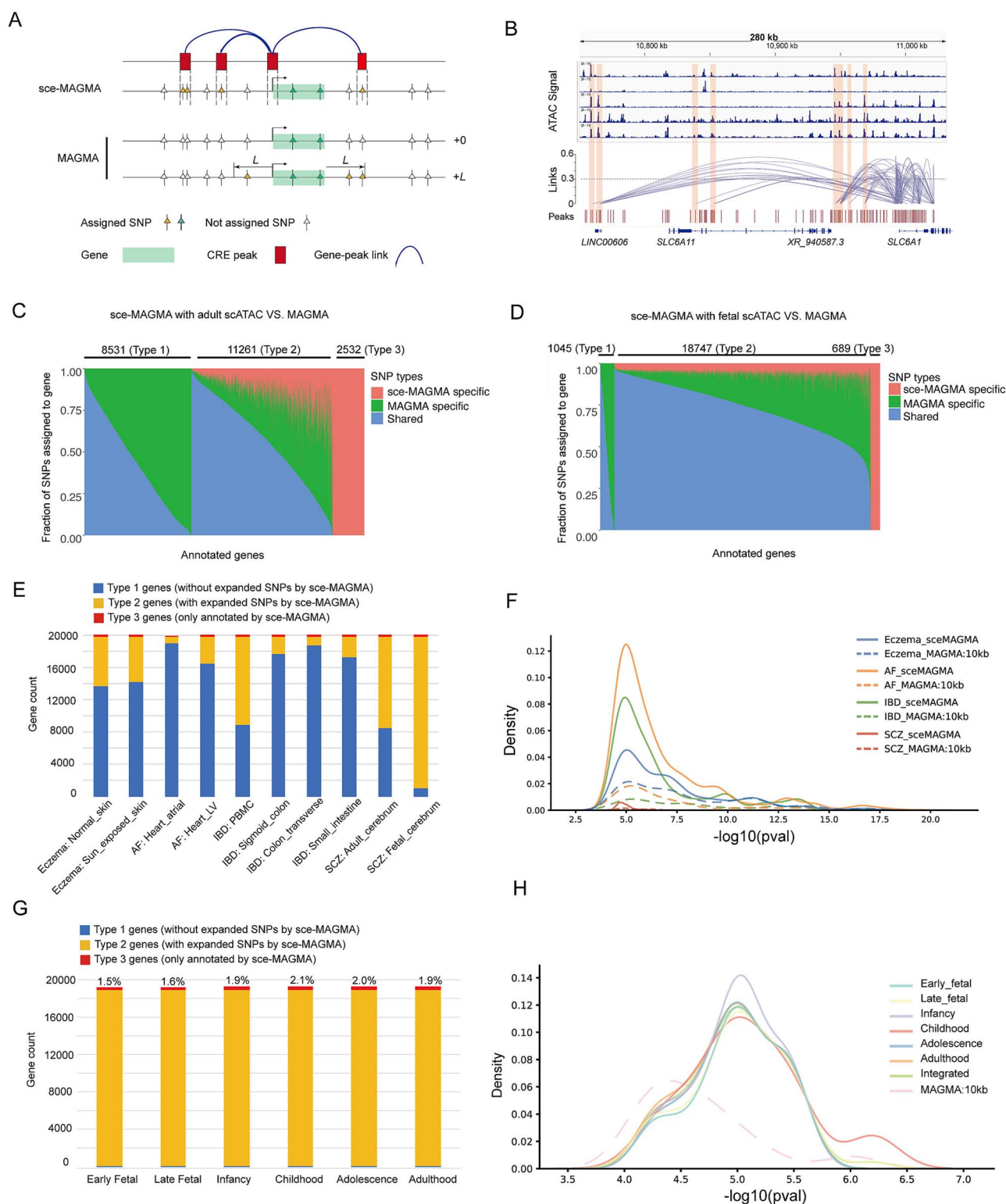


Figure 2. sce-MAGMA efficiently assigns distal noncoding variants to target genes. (A) Schema illustrates strategy of assigning SNPs to target genes for sce-MAGMA and MAGMA. (B) Significant co-accessibility peak links related to gene *SLC6A1*. Accessibility of different cell clusters in scATAC-seq (upper panel) and significant peak links (middle panel) were shown. Distal peaks upstream of gene body region of *SLC6A1* are highlighted. Distribution of different types of SNPs (sce-MAGMA specific, MAGMA specific, and shared SNPs) assigned to genes for analysis with adult (C) and fetal (D) datasets. Three types of genes are defined and indicated. Type 1 genes are those without sce-MAGMA-specific SNPs; type 2 genes are those having sce-MAGMA-specific SNPs and MAGMA SNPs; type 3 genes are those only having sce-MAGMA-specific SNPs. (E) Distribution of three types of annotated genes defined in (C) and (D) derived using multiple human tissue scATAC-seq datasets. (F) The KDE plot illustrates the distribution of SNPs outside gene body annotated to target genes across different diseases. The solid lines indicate annotations made by sce-MAGMA, while the dashed line denotes those by MAGMA. (G) Distribution of three types of annotated genes defined in (C) and (D) derived using multiome ATAC + RNA sequencing datasets from six different developmental stages. (H) The KDE plot illustrates the distribution of SNPs outside gene body annotated to target genes across different developmental stages.

(type 3 genes) in each dataset. Moreover, examining the significance of these sce-MAGMA-specific distal SNPs across developmental stages revealed similar patterns to those in other tissues (Fig. 2H). These findings suggest that SC-VAR efficiently annotates more significant SNPs distal from genes, resulting in a greater number of candidate genes for downstream analysis compared to MAGMA.

SC-VAR improved the ability to identify stage-specific or site-specific risk genes through cis-regulatory elements information

We then applied SC-VAR to predict disease-related genes using scATAC-seq data from specific tissues and compared the results with MAGMA. Although we cannot directly determine the true disease-related genes, we curated a list of validated genes from external public database that are either directly or indirectly associated with the phenotypes of interest for evaluation (Supplementary Table S2, see Materials and methods section).

As shown in Fig. 3A and B, sce-MAGMA identified more significant genes and verified genes across all disease contexts than the classical MAGMA approach, highlighting its enhanced sensitivity in detecting relevant genes for complex diseases. Further, the Venn-diagrams of predicted genes by different approaches revealed that sce-MAGMA almost covered all the predicted genes by MAGMA for different datasets (Fig. 3C–F). This improvement is largely due to sce-MAGMA's ability to include SNPs in noncoding CRE regions, which are not considered by classical MAGMA.

Since MAGMA can extend the “window size” to include SNPs beyond the gene body and account for noncoding variants, we compared sce-MAGMA with MAGMA using a 10 kb extension around the gene body [9–11] on SCZ GWAS data. Our results showed that sce-MAGMA, for both adult and fetal datasets, predicted more significant disease-risk genes and validated genes compared to MAGMA with a 10 kb extension did (Fig. 4A). Next, we estimated the type 1 error rate by randomly perturbing the positions of the SNPs related to SCZ. Results show that sce-MAGMA had similar mean type 1 error rate (0.064) as MAGMA did (0.069) (Fig. 4B).

It is of note that the prediction of sce-MAGMA yielded tissue-specific disease-risk genes, which were highly dependent on varied activity of the CREs in different tissues (Fig. 3C–F). For example, in the result of SCZ datasets, genes uniquely predicted by sce-MAGMA had a higher fraction of linked CRE peaks outside gene body regions compared to genes shared with MAGMA predictions (Fig. 4C). This suggests that these sce-MAGMA-specific genes are mainly influenced by SNPs in distal CREs. When comparing adult and fetal datasets, we found that 19.0% of predicted genes and 10.0% of validated genes were specific to the adult stage, while 15.2% of predicted genes and 9.4% of validated genes were specific to the fetal stage (Fig. 3E). Moreover, adult-stage-specific genes have higher fraction of adult-specific peaks linked to the genes than common genes in adult stage, and fetal-specific genes have higher fraction of fetal-specific peaks linked to the genes than common genes in fetal stage (Fig. 4D), suggesting the risk of stage-specific genes is mainly contributed from SNPs located in stage-specific CREs. For example, the gene *HTR1A* and *TAGLN* are adult-specific validated genes that have adult-specific peak links (Fig. 4E and F) while the gene *VAMP2* and *TAP2* are fetal-specific validated genes that have fetal-specific peak links (Fig. 4G and H). The 5-HT dysfunction has been demonstrated to contribute to not only emotional difficulty but also cognition deficit [36]. *HTR1A* is the most abundantly expressed 5-HT receptor subtype in the mammalian brain and have been found to play important roles

in the development and treatment of SCZ [37, 38]. The *VAMP2* gene also known as Synaptobrevin-2 is primarily involved in the fusion of neuronal synaptic vesicles with the cell membrane, and its aberrant function can disrupt human neural development [39–41]. *VAMP2* has also been found to be involved in the regulation of synaptic plasticity [42, 43], suggesting that it may play an important role in the development of SCZ. Our analysis provides the clues implying that the variants on the distal CREs might influence the genes in SCZ in a developmental-stage-dependent manner.

Furthermore, applying SC-VAR to single-cell multi-omics data from the human neocortex at different developmental stages detected SCZ-risk genes that covered most of the genes derived from scATAC-seq data of the adult and fetal samples. As expected, we observed some stage-specific risk genes. Meanwhile, 1238 (69.7%) of risk genes are common across stages (Fig. 4I), such as *FOXP2*, *ERBB4*, and *NRXN1*, consistently implicated in SCZ susceptibility across all stages.

In conclusion, SC-VAR enhances the identification of risk genes by incorporating CREs specific to developmental stages, demonstrating its superiority over the traditional MAGMA approach.

SC-VAR identifies biological processes or pathways related to disease risks

Since the risk of an individual for a particular GWAS disease or trait is the output of combinatorial effects of multiple SNPs and genes, we extended the evaluation of risk of single genes to meta-genes or gene sets. SC-VAR follows the similar algorithm as MAGMA to calculate the significance of a gene set for disease association, differing only in the gene-SNP annotation (Materials and methods section). Based on the results of detailed gene analysis above, we further explored the gene-sets enriched in those genes to see how the genes are involved in SCZ. We tested the association of gene sets from the MSig7.5 database with the SCZ by both SC-VAR and MAGMA. Many gene sets are detected by both methods as having significant connection with SCZ. However, while MAGMA exclusively detected some gene sets with general biological functions or pathways such as protein import, metabolic process, and receptor signaling, SC-VAR uniquely identified pathways more specifically linked to SCZ (Table 1 and Supplementary Table S3). For example, the response to amyloid beta process is detected as a risk pathway by SC-VAR at the adult stage. In studies of SCZ, amyloids are indicators of cognitive impairment [44]. Notably, SC-VAR also pinpointed the process of snoRNAs as the risk gene set in fetal sample. Previous studies have reported the change of the snoRNA level in schizophrenic patients, which is thought to be related to the dysregulation of alternative splicing in SCZ [45]. Another biological process of note detected by SC-VAR is the regulation of synaptic vesicle exocytosis. Synaptic proteins regulate a number of vesicles that can undergo cytokinesis, and it has been reported that there is a significant reduction in synaptic proteins in patients with SCZ [46–48]. Our data support that neurotransmitter transmission failure at the synapse may lead to SCZ. These results suggest that SC-VAR can uncover putative pathways and biological processes potentially implicated in variant-driven diseases.

Linking high-risk cell types from diverse tissues to disease

Within a complex tissue consisting of distinct cell types, the impacts from disease-relevant variants are different across cell types and cell states due to their specific transcriptomic and epigenomic profiles. The module sce-DRS of SC-VAR is specifically

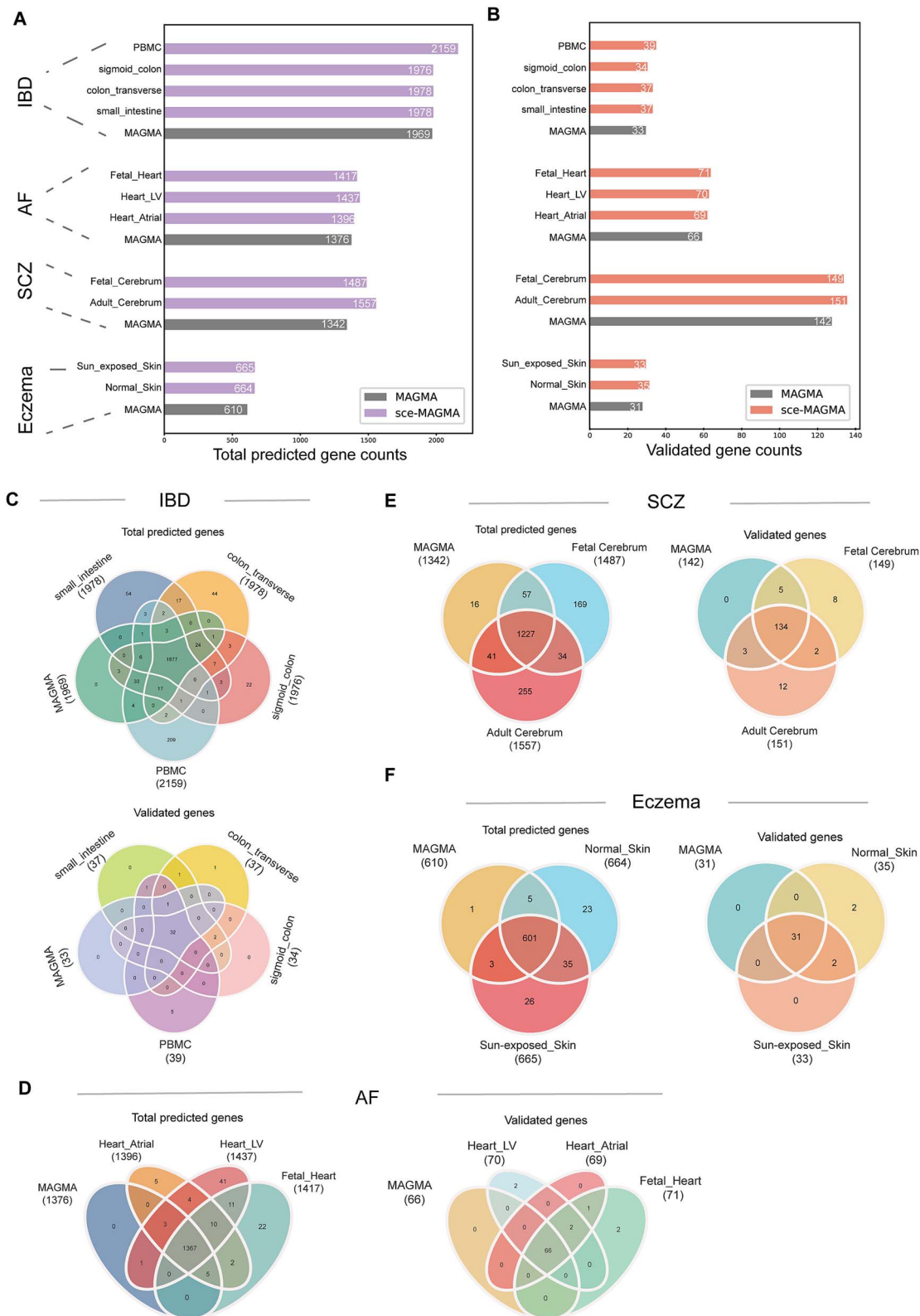


Figure 3. SC-VAR improved the ability to identify stage-specific disease-risk genes. (A) and (B) Bar plot of total predicted genes (A) and validated genes for different diseases (B). (C) Venn diagrams of predicted genes (top) and validated genes (bottom) for IBD. (D)–(F) Venn diagrams of predicted genes (left) and validated genes (right) for AF, SCZ, and eczema.

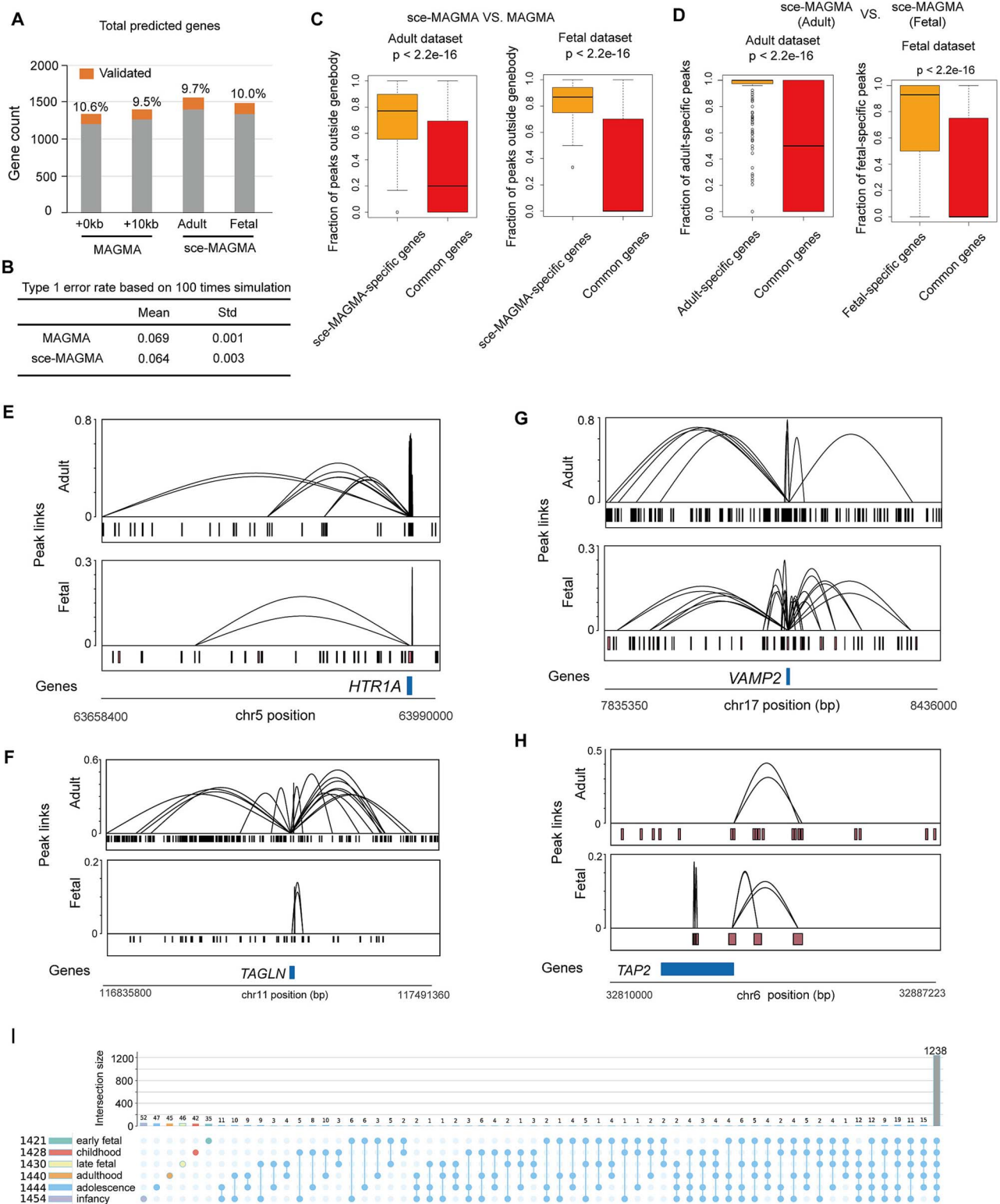


Figure 4. Comparison of gene prediction and validation across different annotation methods. (A) Counts of predicted genes for four approaches, validated genes are highlighted, and fractions of validated genes are labeled. (B) Type 1 error rate comparison between sce-MAGMA and MAGMA. (C) Proportion of CRE-linked peaks outside gene bodies in sce-MAGMA stage-specific genes versus in common genes shared between sce-MAGMA and MAGMA. P values were calculated using K-S test. (D) Higher fractions of stage-specific peaks linked to stage-specific genes. P values were calculated using K-S test. (E)–(H) Examples of adult-specific gene *HTR1A* (E) and *TAGLN* (F) with stage-specific peak links. Fetal specific gene *VAMP2* (G) and *TAP2* (H) with fetal stage-specific peak links. (I) Upset plot of unique and shared risk genes at six developmental stages for cortex multiome data.

designed to prioritize disease risks at single-cell and cell-type level (see Materials and methods section). Using GWAS data from IBD, eczema, SCZ, and AF, we evaluated the association of various cell types with these diseases (Fig. 5A–D).

For IBD, significant enrichment was observed in multiple cell types across gut regions (e.g. transverse colon, sigmoid colon, small intestine) and PBMC (Fig. 5A). Plasma cells, smooth muscle cells, and T lymphocytes showed the strongest associations, reflecting the autoimmune nature of IBD [49]. T lymphocytes and plasma cells are considered key players in the intestinal immune system and their dysregulation, leading to excessive immune responses, is a major contributor to the chronic inflammation observed in IBD [50]. In eczema, skin cell types such as basal epidermal, eccrine sweat epidermal, and granular epidermal cells, along with T lymphocytes, were notably enriched (Fig. 5B). Eccrine sweat epidermal and T lymphocytes were specifically enriched in sun-exposed skin, which is interesting. Sweat secretion, as an integral part of the skin's protective barrier, plays a crucial role in maintaining barrier homeostasis [51]. Previous studies have suggested that sweat management could serve as a potential therapeutic approach for skin dermatitis [52]. Our findings reveal the specific role of eccrine sweat epidermal cells in sun-exposed skin, demonstrating the advantage of using tissue-specific data to uncover novel insights into disease pathophysiology.

For AF, significant associations were identified in heart-related cell types (Fig. 5C). The strong link between ventricular cardiomyocytes and AF is consistent with the electrical dysregulation characteristic of the condition [53]. And the identification of AF-associated cell types aligns with findings from chromatin accessibility profiles, which demonstrated strong enrichment of AF-associated SNPs in cardiomyocyte-specific open chromatin regions [19, 54, 55]. Both our results and these studies highlight the critical role of noncoding SNPs in mediating disease risk. In SCZ, neuronal cell types from both fetal and adult cerebrum were significantly enriched in high risks (Fig. 5D). Astrocytes and microglia were also implicated, underscoring the importance of glial and neuronal interactions in the brain's pathology [56]. Notably, SCZ displayed substantial within-cell-type heterogeneity, particularly among neuronal and glial cells, suggesting a diverse set of mechanisms involved in the disease's development.

SC-VAR reveals dynamic genetic susceptibility across high-risk cell types for brain disorders during brain development

Next, we conducted a detailed analysis of finely divided multi-omics brain single-cell data at different developmental stages. We applied sce-DRS on single-cell multiome ATAC + gene expression datasets for neocortex samples derived from different developmental stages (Fig. 6A). By integrating data across developmental stages, we observed dynamic changes in disease risk relevance among neocortical cell types (Fig. 6B and C). In early and late fetal stages, radial glia (RG), as neural progenitor cells, exhibited the highest risk relevance, reflecting their critical role in early brain development and susceptibility to disruptions during this period [57]. As the neocortex matured, microglia displayed the highest risk scores across all cell types, highlighting their involvement in maintaining neuroimmune homeostasis and synaptic pruning [58, 59]. Mature neuronal cell types, including excitatory (EN) and inhibitory neurons (IN), also showed significantly higher risk levels, particularly after the childhood stage, underscoring the importance of excitatory/inhibitory (E/I) balance and neuronal connectivity in SCZ pathogenesis [60, 61].

We then aligned risk cell types with cell-type-specific risk genes and CRE peaks to elucidate the genes and regulatory foundations relevant to high-risk cell types. Our analysis focused on four major high-risk cell types: excitatory neurons (EN), IN-MGE, IN-CGE, and microglia (Supplementary Fig. S1A). EN-specific risk genes (e.g. RYR2, CCKBR, and CACNA1C) are enriched in the Ras-mitogen activated protein kinase (MAPK/ERK) pathway, which facilitates Ca^{2+} flux through NMDA receptors and is crucial for the expression of numerous key downstream genes in neurons [62–64]. This result is in line with the report that NMDA receptor dysfunction plays a significant role in the pathogenesis of SCZ [65, 66]. IN-MGE-specific risk genes (e.g. FGF14, GABRA5, and ZNF2) and IN-CGE-specific risk genes (e.g. GRIK3, PTPN5, and BSN) are enriched in pathways related to GABAergic synapse. Microglia-specific risk genes (e.g. RGS10, TNFRSF1A, and MYCBP2) were enriched in pathways associated with neurotrophic factors, such as Fcγ receptors, which play important roles in the generation, maturation, and integration of new neurons into neural circuits during neurogenesis [67–69]. These results highlight the pathways contributing to cell-type-specific risk in SCZ.

As a proof of principle and to assess the robustness of SC-VAR in elucidating cell types involved in the pathogenic mechanisms of brain diseases, we applied SC-VAR to GWAS data from six brain disorders in addition to SCZ: ADHD, BD, AD, ASD, OCD, and PKS, to evaluate cell-type-specific risk. Our analysis identified 20 significantly associated disease-cell type pairs (Fig. 6D and E), including known associations such as high risks of multiple neuronal subtypes for BD and ADHD [20], and high risks of glial cells for AD and PKS [70, 71]. Additionally, we uncovered several associations that were less clear in existing genetic data but are biologically plausible, like glial cells for compulsive disorder (OCD) and ASD, and neurons for AD.

The association of neurons in AD is consistent with recent studies by Mathys *et al.* [72] using single-cell transcriptomics datasets from Alzheimer's patients and controls. In these studies, neurons, which are early targets of tau pathology and amyloid-beta accumulation, experience degeneration leading to synaptic dysfunction and compromised neural circuit integrity. Disrupted inhibitory signaling further destabilizes neural circuits, potentially accelerating disease progression. Notably, microglia show high risk score in all studied brain disorders, highlighting the critical roles of neuroinflammation and glial dysfunction in brain disease pathology.

We also traced the developmental trajectory of neuron cells from RG to all mature neurons using pseudotime analysis, exploring the dynamics of risk scores for each brain disorder along the developmental pseudotime (Fig. 6F). The results revealed strong associations with all these disorders ($P < .005$, MC test), indicating a consistent pattern of increased risk scores along the pseudotime and suggesting that mature neurons are more susceptible to polygenic factors.

Discussion

In this study, we introduce SC-VAR, a software designed to enhance our understanding of the genetic basis of complex diseases. SC-VAR integrates single-cell data to offer a comprehensive genomic analysis that encompasses not only coding regions but also the extensive noncoding regions of the genome. Through extensive evaluation on multiple diseases, we demonstrate that SC-VAR is well-calibrated and robust under realistic scenarios. Leveraging single-cell and multi-omics sequencing technologies, SC-VAR effectively identifies disease-associated genes, pathways,

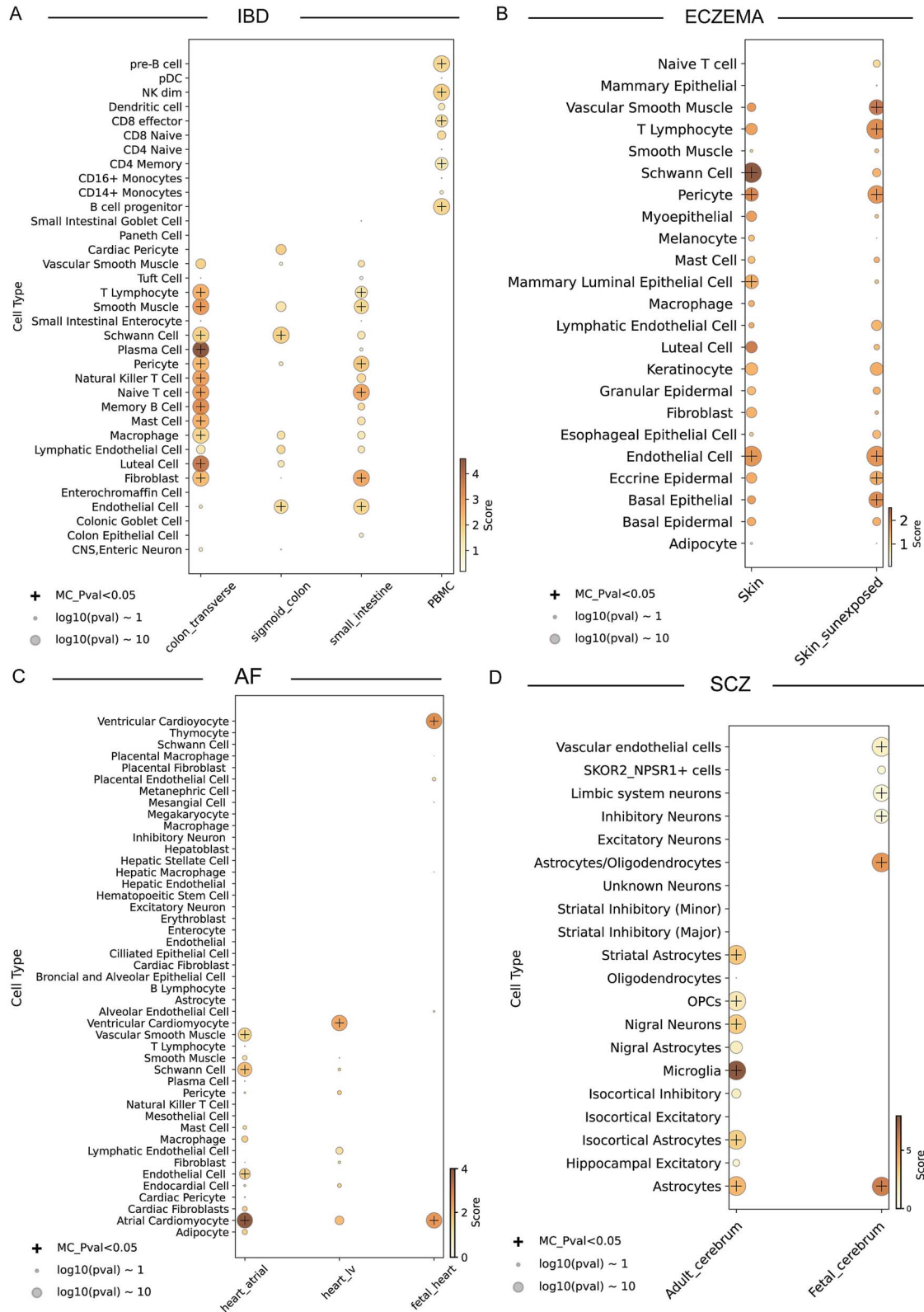


Figure 5. Cell type-specific associations for disease-associated variants across tissues and developmental stages. Bubble plots showing disease association scores of cell types across different tissues and developmental stages for four complex diseases: (A) IBD, (B) eczema, (C) AF, and (D) SCZ. Each panel represents one disease, displaying specific cell types along the y-axis and tissue or body part on the x-axis. Only cell types with MC P-values less than .05 are labeled “+” on the plots.

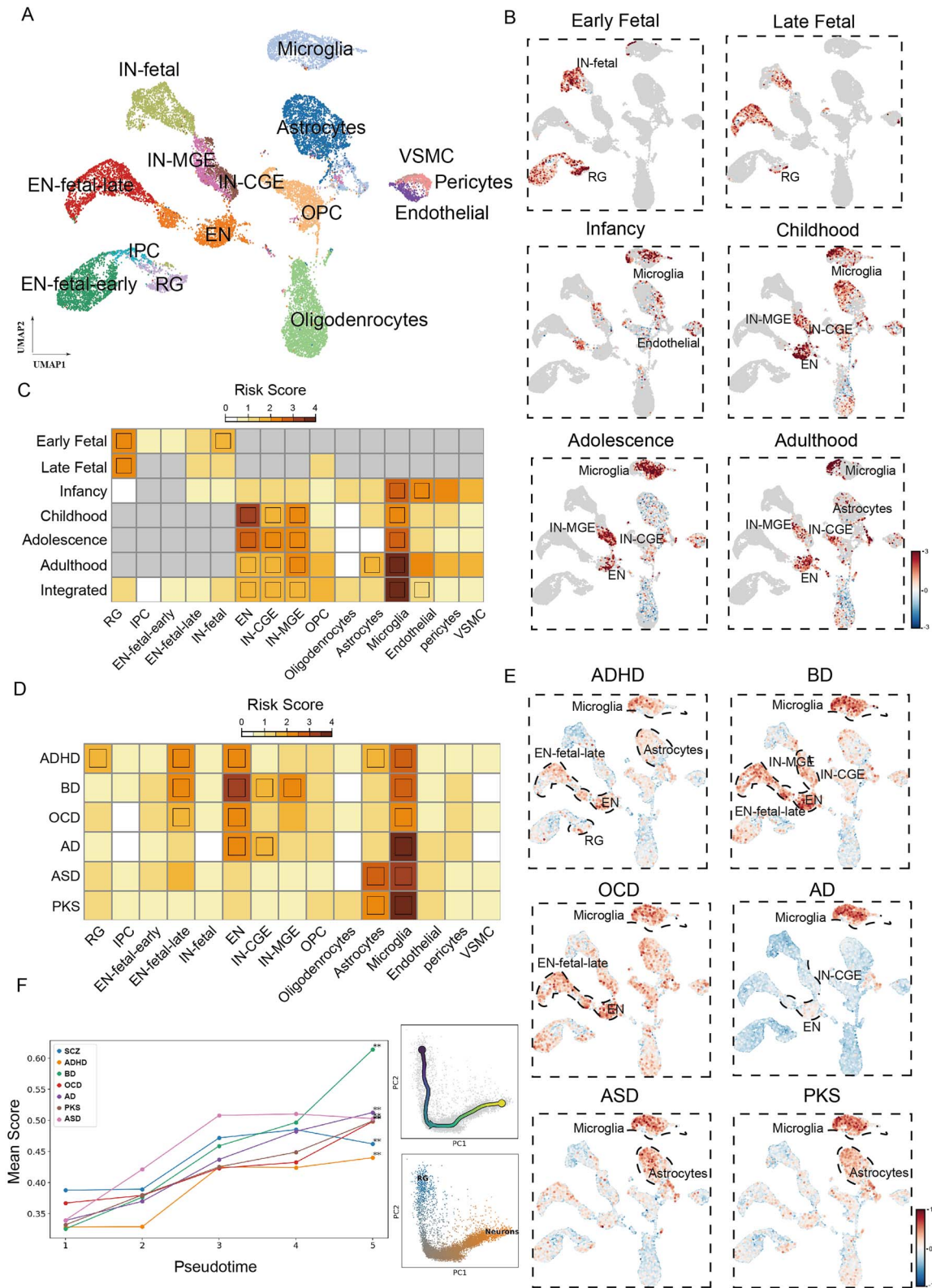


Figure 6. Various disease risk scores of single cells for neocortex samples. (A) UMAP of multi-omics data, showing distinct groups labeled by cell types. (B) Risk scores of single cells at individual developmental stages, significantly associated cell types are labelled on the graph. (C) Heat map shows statistical significance of cell-type associations with SCZ for individual brain development stages and integrated landscape. (D) Heat map shows statistical significance of cell-type associations with different brain disorders for integrated landscape. (E) Risk scores of single cells at integrated landscape of neocortex samples for six different brain disorders. The boundaries of the significant associated cell types are labeled in dash lines. (F) Line plot (left panel) shows increased association score for different brain disorders along pseudo-time trajectory indicating development from RG to mature neurons (right panel). The x-axis denotes time quintile bins, and the y-axis denotes the average sce-DRS disease score in each bin for each disease. * $P < .05$ and, ** $P < .005$ (one-sided MC test).

and cell types, offering significant advancements over traditional methods.

We applied SC-VAR to various diseases and traits using 12 scATAC-seq datasets. Our findings emphasize the critical role of epigenomic annotation in interpreting the regulatory functions of variants within noncoding regions. SC-VAR enabled us to uncover associations between genetic variants and genes across different tissues and developmental stages. For example, we found that SCZ risk genes are significantly enriched in neurodevelopmental processes in fetal brain cells, supporting the neurodevelopmental hypothesis of psychiatric disorders. This insight has potential implications for therapeutic strategies, particularly by highlighting the importance of early intervention. Additionally, SC-VAR identified disease-associated cell types and corresponding genes, which may guide targeted *in vitro* experiments to explore the relevant cellular mechanisms.

Our results have demonstrated SC-VAR's effectiveness in identifying risk genes and cell types. SC-VAR outperforms traditional MAGMA in identifying more verified disease-related risk genes while having the similar type 1 error rates. The main advance of SC-VAR compared to MAGMA is that SC-VAR could identify more reliable regulatory element-related SNPs which are distal from the gene bodies. It is of note that H-MAGMA, which depends on three-dimensional chromatin interaction data (HiC) [15], is also designed to leverage enhancer-promoter links to interpret association of SNPs in regulatory region with disease. However, the HiC data are often not available for tissue samples and could not dissect cell types in heterogeneous tissue samples, while single-cell epigenomic data, especially scATAC-seq data, is currently available for many human tissues and organs, or can be much more easily obtained for most laboratories with low-cost commercial supports. Therefore, SC-VAR could be widely used to interpret genetic variants in clinical studies where single-cell epigenomic data can be easily obtained.

While SC-VAR offers a comprehensive framework for analyzing single-cell genomic data, it also faces challenges, particularly regarding computational demands and the need for tissue datasets that are not always available. In SC-VAR, the computation of cell-type associations is highly tissue-specific, indicating that the identified risk CREs are not readily transferable to data from similar tissues or sites. This tissue specificity underscores the importance of context when interpreting risk genes and regulatory elements, as the functional relevance of these associations may vary significantly across different biological environments. Additionally, the interpretation of complex multi-omics data requires expertise, which could limit its accessibility for some researchers. Although computational methods like SC-VAR offer valuable insights into disease-associated cell types and regulatory elements, they have inherent limitations in fully uncovering the intricate pathophysiology and mechanisms of diseases. Complementary biological experiments are crucial to validate these findings and provide deeper understanding. In the future, SC-VAR could be expanded to include more diverse types of single-cell data and further refined to increase its computational efficiency.

Key Points

- SC-VAR is a novel computational tool for predicting stage-specific, disease-associated risk genes, pathways, and cell types, using genome-wide association studies data and single-cell epigenomic data.

- SC-VAR outperforms state-of-the-art methods in terms of annotating more noncoding single-nucleotide polymorphisms (SNPs) located within distal cis-regulatory elements and predicting more risk genes and meaningful biological pathway.
- We have benchmarked SC-VAR on multiple complex diseases, demonstrating its capability in interpreting polygenic disease risks in a broad range of diseases.

Acknowledgments

This work was supported by the Science Fund Program for Distinguished Young Scholars of the National Natural Science Foundation of China.

Author contributions

Binbin Lai (designed this research, interpreted the data, wrote the manuscript) and Gefei Zhao (responsible for the method implementation and software maintenance, analyzed the data, interpreted the data, and wrote the manuscript)

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Conflict of interest

The authors declare no competing interests.

Funding

None declared.

Data availability

SC-VAR software and source code are under MIT license and is freely available at Pypi (<https://pypi.org/project/sc-var/>). All codes used in this study and the manual are also provided in the GitHub repository (https://github.com/gefeiz/sc_var/). Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

References

1. Bycroft C, Freeman C, Petkova D. et al. The UK biobank resource with deep phenotyping and genomic data. *Nature* 2018;**562**: 203–9. <https://doi.org/10.1038/s41586-018-0579-z>
2. Trubetskoy V, Pardiñas AF, Qi T. et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* 2022;**604**:502–8. <https://doi.org/10.1038/s41586-022-04434-5>
3. Mallard TT, Linnér RK, Grotzinger AD. et al. Multivariate GWAS of psychiatric disorders and their cardinal symptoms reveal two dimensions of cross-cutting genetic liabilities. *Cell Genomics* 2022;**2**:100140. <https://doi.org/10.1016/j.xgen.2022.100140>
4. Buniello A, MacArthur JAL, Cerezo M. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 2018;**47**: 1005–12. <https://doi.org/10.1093/nar/gky1120>

5. Boix C, James B, Park Y. et al. Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature* 2021;**590**: 300–7. <https://doi.org/10.1038/s41586-020-03145-z>
6. Cano-Gamez E, Trynka G. From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. *Front Genet* 2020;**11**:424. <https://doi.org/10.3389/fgene.2020.00424>
7. McLean CY, Bristol D, Hiller M. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 2010;**28**: 495–501. <https://doi.org/10.1038/nbt.1630>
8. Xu T, Jin P, Qin ZS. Regulatory annotation of genomic intervals based on tissue-specific expression QTLs. *Bioinformatics* 2019;**36**: 690–7. <https://doi.org/10.1093/bioinformatics/btz669>
9. De Leeuw C, Mooij JM, Heskes T. et al. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015;**11**:e1004219. <https://doi.org/10.1371/journal.pcbi.1004219>
10. Yang A, Chen J, Zhao X. nMAGMA: a network-enhanced method for inferring risk genes from GWAS summary statistics and its application to schizophrenia. *Brief Bioinform* 2020;**22**. <https://doi.org/10.1093/bib/bbaa298>
11. Gerring ZF, Mina-Vargas A, Gamazon ER. et al. E-MAGMA: an eQTL-informed method to identify risk genes using genome-wide association study summary statistics. *Bioinformatics* 2021;**37**:2245–9. <https://doi.org/10.1093/bioinformatics/btab115>
12. Wray NR, Ripke S, Mattheisen M. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* 2018;**50**:668–81. <https://doi.org/10.1038/s41588-018-0090-3>
13. Lee JJ, Wedow R, Okbay A. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* 2018;**50**:1112–21. <https://doi.org/10.1038/s41588-018-0147-3>
14. Watanabe K, Stringer S, Frei O. et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* 2019;**51**:1339–48. <https://doi.org/10.1038/s41588-019-0481-0>
15. Sey NYA, Hu B, Mah W. et al. A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nat Neurosci* 2020;**23**: 583–93. <https://doi.org/10.1038/s41593-020-0603-0>
16. Preissl S, Gaulton KJ, Ren B. Characterizing cis-regulatory elements using single-cell epigenomics. *Nat Rev Genet* 2022;**24**: 21–43. <https://doi.org/10.1038/s41576-022-00509-1>
17. Gaulton KJ, Preissl S, Ren B. Interpreting non-coding disease-associated human variants using single-cell epigenomics. *Nat Rev Genet* 2023;**24**:516–34. <https://doi.org/10.1038/s41576-023-00598-6>
18. Domcke S, Hill AJ, Daza RM. et al. A human cell atlas of fetal chromatin accessibility. *Science* 2020;**370**:eaba7612. <https://doi.org/10.1126/science.aba7612>
19. Zhang K, Hocker JD, Miller M. et al. A single-cell atlas of chromatin accessibility in the human genome. *Cell* 2021;**184**:5985–6001.e19. <https://doi.org/10.1016/j.cell.2021.10.024>
20. Li YE, Preissl S, Miller M. et al. A comparative atlas of single-cell chromatin accessibility in the human brain. *Science* 2023;**382**:eadf7044. <https://doi.org/10.1126/science.adf7044>
21. Zhang MJ, Hou K, Dey KK. et al. Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *Nat Genet* 2020;**54**:1572–80. <https://doi.org/10.1038/s41588-022-01167-z>
22. Fan J-C, Lu Y, Gan J-H. et al. Identification of potential novel targets for treating inflammatory bowel disease using Mendelian randomization analysis. *Int J Colorectal Dis* 2024;**39**:165. <https://doi.org/10.1007/s00384-024-04744-2>
23. Sebastian R, Jin K, Pavon N. et al. Schizophrenia-associated NRXN1 deletions induce developmental-timing- and cell-type-specific vulnerabilities in human brain organoids. *Nat Commun* 2023;**14**:3770. <https://doi.org/10.1038/s41467-023-39420-6>
24. Rees E, Han J, Morgan J. et al. De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in schizophrenia. *Nat Neurosci* 2020;**23**:179–84. <https://doi.org/10.1038/s41593-019-0565-2>
25. Jiang L, Zheng Z, Fang H. et al. A generalized linear mixed model association tool for biobank-scale data. *Nat Genet* 2021;**53**: 1616–21. <https://doi.org/10.1038/s41588-021-00954-4>
26. Kurki M, Karjalainen J, Palta P. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* 2023;**613**:508–18. <https://doi.org/10.1038/s41586-022-05473-8>
27. Corces MR, Shcherbina A, Kundu S. et al. Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat Genet* 2020;**52**: 1158–68. <https://doi.org/10.1038/s41588-020-00721-x>
28. Zhu K, Bendl J, Rahman S. et al. Multi-omic profiling of the developing human cerebral cortex at the single-cell level. *Science. Advances* 2023;**9**:eadg3754. <https://doi.org/10.1126/sciadv.adg3754>
29. Pliner HA, Packer JS, McFaline-Figueroa JL. et al. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol Cell* 2018;**71**:858–71.e8. <https://doi.org/10.1016/j.molcel.2018.06.044>
30. Ma S, Zhang B, LaFave LM. et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* 2020;**183**:1103–1116.e20. <https://doi.org/10.1016/j.cell.2020.09.056>
31. Hao Y, Hao S, Andersen-Nissen E. et al. Integrated analysis of multimodal single-cell data. *Cell* 2021;**184**:3573–87.e29. <https://doi.org/10.1016/j.cell.2021.04.048>
32. Frost HR. Variance-adjusted Mahalanobis (VAM): a fast and accurate method for cell-specific gene set scoring. *Nucleic Acids Res* 2020;**48**:e94. <https://doi.org/10.1093/nar/gkaa582>
33. Stuart T, Srivastava A, Madad S. et al. Single-cell chromatin state analysis with Signac. *Nat Methods* 2021;**18**:1333–41. <https://doi.org/10.1038/s41592-021-01282-5>
34. Faure L, Soldatov R, Kharchenko PV. et al. scFates: a scalable python package for advanced pseudotime and bifurcation analysis from single-cell data. *Bioinformatics* 2022;**39**. <https://doi.org/10.1093/bioinformatics/btac746>
35. Raudvere U, Kolberg L, Kuzmin I. et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* 2019;**47**:W191–8. <https://doi.org/10.1093/nar/gkz369>
36. Filip M, Bader M. Overview on 5-HT receptors and their role in physiology and pathology of the central nervous system. *Pharmacol Rep* 2009;**61**:761–77. [https://doi.org/10.1016/S1734-1140\(09\)70132-X](https://doi.org/10.1016/S1734-1140(09)70132-X)
37. Guan F, Lin H, Chen G. et al. Evaluation of association of common variants in HTR1A and HTR5A with schizophrenia and executive function. *Sci Rep* 2016;**6**:38048. <https://doi.org/10.1038/srep38048>
38. Takekita Y, Fabbri C, Kato M. et al. HTR1A gene polymorphisms and 5-HT1A receptor partial agonist antipsychotics efficacy in schizophrenia. *J Clin Psychopharmacol* 2015;**35**:220–7. <https://doi.org/10.1097/JCP.0000000000000304>
39. Gao Y, Zorman S, Gundersen G. et al. Single reconstituted neuronal SNARE complexes zipper in three distinct stages. *Science* 2012;**337**:1340–3. <https://doi.org/10.1126/science.1224492>

40. Rothman JE, Söhlner TH. Throttles and dampers: controlling the engine of membrane fusion. *Science* 1997;**276**:1212. <https://doi.org/10.1126/science.276.5316.1212>
41. Salpietro V, Malintan NT, Llano-Rivas I. et al. Mutations in the neuronal vesicular SNARE VAMP2 affect synaptic membrane fusion and impair human neurodevelopment. *Am J Hum Genet* 2019;**104**:721–30. <https://doi.org/10.1016/j.ajhg.2019.02.016>
42. Deák F, Schoch S, Liu X. et al. Synaptobrevin is essential for fast synaptic-vesicle endocytosis. *Nat Cell Biol* 2004;**6**:1102–8. <https://doi.org/10.1038/ncb1185>
43. Yan C, Jiang J, Yang Y. et al. The function of VAMP2 in mediating membrane fusion: an overview. *Front Mol Neurosci* 2022;**15**:948160. <https://doi.org/10.3389/fnmol.2022.948160>
44. Albertini V, Benussi L, Paterlini A. et al. Distinct cerebrospinal fluid amyloid-beta peptide signatures in cognitive decline associated with Alzheimer's disease and schizophrenia. *Electrophoresis* 2012;**33**:3738–44. <https://doi.org/10.1002/elps.201200307>
45. Gibbons A, Udawela M, Dean B. Non-coding RNA as novel players in the pathophysiology of schizophrenia. *Non-Coding RNA* 2018;**4**:11. <https://doi.org/10.3390/ncrna4020011>
46. Egbujo CN, Sinclair D, Hahn C-G. Dysregulations of synaptic vesicle trafficking in schizophrenia. *Curr Psychiatry Rep* 2016;**18**:77. <https://doi.org/10.1007/s11920-016-0710-5>
47. Osimo EF, Beck K, Marques TR. et al. Synaptic loss in schizophrenia: a meta-analysis and systematic review of synaptic protein and mRNA measures. *Mol Psychiatry* 2018;**24**:549–61. <https://doi.org/10.1038/s41380-018-0041-5>
48. Radhakrishnan R, Skosnik PD, Ranganathan M. et al. In vivo evidence of lower synaptic vesicle density in schizophrenia. *Mol Psychiatry* 2021;**26**:7690–8. <https://doi.org/10.1038/s41380-021-01184-0>
49. De Souza HSP, Fiocchi C. Immunopathogenesis of IBD: current state of the art. *Nat Rev Gastroenterol Hepatol* 2015;**13**:13–27. <https://doi.org/10.1038/nrgastro.2015.186>
50. Petagna L, Antonelli A, Ganini C. et al. Pathophysiology of Crohn's disease inflammation and recurrence. *Biol Direct* 2020;**15**:23. <https://doi.org/10.1186/s13062-020-00280-5>
51. Proksch E, Fölster-Holst R, Jensen J-M. Skin barrier function, epidermal proliferation and differentiation in eczema. *J Dermatol Sci* 2006;**43**:159–69. <https://doi.org/10.1016/j.jdermsci.2006.06.003>
52. Hendricks AJ, Vaughn AR, Clark AK. et al. Sweat mechanisms and dysfunctions in atopic dermatitis. *J Dermatol Sci* 2017;**89**:105–11. <https://doi.org/10.1016/j.jdermsci.2017.11.005>
53. Colman MA, Aslanidi OV, Kharche S. et al. Pro-arrhythmogenic effects of atrial fibrillation-induced electrical remodelling: insights from the three-dimensional virtual human atria. *J Physiol* 2013;**591**:4249–72. <https://doi.org/10.1113/jphysiol.2013.254987>
54. Hocker JD, Poirion OB, Zhu F. et al. Cardiac cell type-specific gene regulatory programs and disease risk association. *Sci Adv* 2021;**7**:eabf1444. <https://doi.org/10.1126/sciadv.abf1444>
55. Selewa A, Luo K, Wasney M. et al. Single-cell genomics improves the discovery of risk variants and genes of atrial fibrillation. *Nat Commun* 2023;**14**:4999. <https://doi.org/10.1038/s41467-023-40505-5>
56. Pietiläinen O, Trehan A, Meyer D. et al. Astrocytic cell adhesion genes linked to schizophrenia correlate with synaptic programs in neurons. *Cell Rep* 2023;**42**:111988. <https://doi.org/10.1016/j.celrep.2022.111988>
57. Casingal CR, Descant KD, Anton ES. Coordinating cerebral cortical construction and connectivity: unifying influence of radial progenitors. *Neuron* 2022;**110**:1100–15. <https://doi.org/10.1016/j.neuron.2022.01.034>
58. Hartmann S-M, Heider J, Wüst R. et al. Microglia-neuron interactions in schizophrenia. *Front Cell Neurosci* 2024;**18**:1345349. <https://doi.org/10.3389/fncel.2024.1345349>
59. Sellgren CM, Gracias J, Watmuff B. et al. Increased synapse elimination by microglia in schizophrenia patient-derived models of synaptic pruning. *Nat Neurosci* 2019;**22**:374–85. <https://doi.org/10.1038/s41593-018-0334-7>
60. Sohal VS, Rubenstein JLR. Excitation-inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders. *Mol Psychiatry* 2019;**24**:1248–57. <https://doi.org/10.1038/s41380-019-0426-0>
61. Caballero A, Orozco A, Tseng KY. Developmental regulation of excitatory-inhibitory synaptic balance in the prefrontal cortex during adolescence. *Semin Cell Dev Biol* 2021;**118**:60–3. <https://doi.org/10.1016/j.semcdb.2021.02.008>
62. Berridge MJ, Bootman MD, Roderick HL. Calcium signalling: dynamics, homeostasis and remodelling. *Nat Rev Mol Cell Biol* 2003;**4**:517–29. <https://doi.org/10.1038/nrm1155>
63. Mizoguchi Y, Kato TA, Horikawa H. et al. Microglial intracellular Ca²⁺ signaling as a target of antipsychotic actions for the treatment of schizophrenia. *Front Cell Neurosci* 2014;**8**:370. <https://doi.org/10.3389/fncel.2014.00370>
64. Espana A, Seth H, Jézéquel J. et al. Alteration of NMDA receptor trafficking as a cellular hallmark of psychosis. *Transl Psychiatry* 2021;**11**:444. <https://doi.org/10.1038/s41398-021-01549-7>
65. Cohen SM, Tsien RW, Goff DC. et al. The impact of NMDA receptor hypofunction on GABAergic neurons in the pathophysiology of schizophrenia. *Schizophr Res* 2015;**167**:98–107. <https://doi.org/10.1016/j.schres.2014.12.026>
66. Koshiyama D, Kiriara K, Tada M. et al. Electrophysiological evidence for abnormal glutamate-GABA association following psychosis onset. *Translational Psychiatry* 2018;**8**:211. <https://doi.org/10.1038/s41398-018-0261-0>
67. Koskivi M, Pörsti E, Hewitt T. et al. Genetic contribution to microglial activation in schizophrenia. *Mol Psychiatry* 2024;**29**:2622–33. <https://doi.org/10.1038/s41380-024-02529-1>
68. Hu X, Liou AKF, Leak RK. et al. Neurobiology of microglial action in CNS injuries: receptor-mediated signaling mechanisms and functional roles. *Prog Neurobiol* 2014;**119–120**:60–84. <https://doi.org/10.1016/j.pneurobio.2014.06.002>
69. De Picker LJ, Victoriano GM, Richards R. et al. Immune environment of the brain in schizophrenia and during the psychotic episode: a human post-mortem study. *Brain Behav Immun* 2021;**97**:319–27. <https://doi.org/10.1016/j.bbi.2021.07.017>
70. Xie A, Gao J, Xu L. et al. Shared mechanisms of neurodegeneration in Alzheimer's disease and Parkinson's disease. *Biomed Res Int* 2014;**2014**:1–8. <https://doi.org/10.1155/2014/648740>
71. Zhang X, Zhang R, Awan MUN. et al. The mechanism and function of glia in Parkinson's disease. *Front Cell Neurosci* 2022;**16**:903469. <https://doi.org/10.3389/fncel.2022.903469>
72. Mathys H, Davila-Velderrain J, Peng Z. et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* 2019;**570**:332–7. <https://doi.org/10.1038/s41586-019-1195-2>
73. Parellada E, Gassó P. Glutamate and microglia activation as a driver of dendritic apoptosis: a core pathophysiological mechanism to understand schizophrenia. *Transl Psychiatry* 2021;**11**:271. <https://doi.org/10.1038/s41398-021-01385-9>
74. Jarskog LF, Glantz LA, Gilmore JH. et al. Apoptotic mechanisms in the pathophysiology of schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry* 2005;**29**:846–58. <https://doi.org/10.1016/j.pnpbp.2005.03.010>
75. Schoonover KE, Queern SL, Lapi SE. et al. Impaired copper transport in schizophrenia results in a copper-deficient

- brain state: a new side to the dysbindin story. *World J Biol Psychiatry* 2018;**21**:13–28. <https://doi.org/10.1080/15622975.2018.1523562>
76. Mealer RG, Jenkins BG, Chen C-Y. et al. The schizophrenia risk locus in SLC39A8 alters brain metal transport and plasma glycosylation. *Sci Rep* 2020;**10**:13162. <https://doi.org/10.1038/s41598-020-70108-9>
77. Ermakov EA, Ivanova SA, Buneva VN. et al. Hydrolysis by catalytic IgGs of microRNA specific for patients with schizophrenia. *IUBMB Life* 2018;**70**:153–64. <https://doi.org/10.1002/iub.1712>
78. Parshukova DA, Smirnova LP, Kornetova EG. et al. IgG-dependent hydrolysis of myelin basic protein of patients with different courses of schizophrenia. *J Immunol Res* 2020;**2020**:1–12. <https://doi.org/10.1155/2020/8986521>