

METHOD

Open Access



scGRNom: a computational pipeline of integrative multi-omics analyses for predicting cell-type disease genes and regulatory networks

Ting Jin^{1,2†}, Peter Rehani^{2,3,4†}, Mufang Ying^{5,6†}, Jiawei Huang⁵, Shuang Liu², Panagiotis Roussos^{7,8} and Daifeng Wang^{1,2,9*} 

Abstract

Understanding cell-type-specific gene regulatory mechanisms from genetic variants to diseases remains challenging. To address this, we developed a computational pipeline, scGRNom (single-cell Gene Regulatory Network prediction from multi-omics), to predict cell-type disease genes and regulatory networks including transcription factors and regulatory elements. With applications to schizophrenia and Alzheimer's disease, we predicted disease genes and regulatory networks for excitatory and inhibitory neurons, microglia, and oligodendrocytes. Further enrichment analyses revealed cross-disease and disease-specific functions and pathways at the cell-type level. Our machine learning analysis also found that cell-type disease genes improved clinical phenotype predictions. scGRNom is a general-purpose tool available at <https://github.com/daifengwanglab/scGRNom>.

Keywords: Single-cell genomics, Single-cell multi-omics integration, Cell-type gene regulatory network, Cell-type disease risk genes, Cross-disease functional genomics, Schizophrenia, Alzheimer's disease

Background

Recent genome-wide association studies (GWAS) studies have identified a variety of genetic risk variants associated with multiple brain diseases. For example, a recent study found 109 pleiotropic loci significantly associated with at least two brain disorders [1]. Many cross-disease common genetic risk factors have revealed many shared functional consequences in clinical presentations [2]. Recent studies have also revealed shared symptoms at both psychiatric and physical levels between neurodegenerative and neuropsychiatric diseases [3]. For instance, 97% of Alzheimer's

disease patients develop neuropsychiatric symptoms throughout the disease [4]. Besides, additional insights into each disease's progression and causes have further demonstrated the highly interlinked nature of both disease types [5]. However, our understanding of the molecular mechanisms of genetic variants between diseases remains elusive, particularly at the cell-type levels.

Alzheimer's disease (AD) and schizophrenia (SCZ) are neurodegenerative and neuropsychiatric diseases, respectively. Both are significantly associated with genetic variants and have complex underlying cellular and molecular mechanisms from genotype to phenotype [6, 7]. Notably, AD is physiologically characterized by accumulations of amyloid beta plaques and neurofibrillary tau protein tangles in the brain [8]. Amyloid beta plaques primarily originate from the apolipoprotein E-encoding gene APOE and its multiple variants. The APOE gene is

* Correspondence: daifeng.wang@wisc.edu

[†]Ting Jin, Peter Rehani and Mufang Ying contributed equally to this work.

¹Department of Biostatistics and Medical Informatics, University of Wisconsin – Madison, Madison, WI 53706, USA

²Waisman Center, University of Wisconsin – Madison, Madison, WI 53705, USA

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

a single step in the broader amyloidogenic processing pathway (APP), and additional genes involved in the process contribute to the regulation of amyloid beta production [6]. Much work has identified major genes of interest involved in the APP [6]. However, a distinct need still exists to further explore these disease loci to understand better the interplay between their regulatory elements and eventual amyloid beta creation and accumulation. Similarly, neurofibrillary tau tangles are associated with many genetic loci and require a study of the highly complex molecular mechanisms required to achieve disease pathology [8]. Further, the downstream effects from both amyloid beta and neurofibrillary tangles within and between various cell types add additional complexity toward linking specific regulatory events and elements with clinical pathology [9, 10].

Also, SCZ is a neuropsychiatric disorder characterized by disruptions in dopamine, glutamate, and GABA-based receptor signaling pathways [11]. Pathologically, the direct connection is less clear between known molecular abnormalities and observed physical changes through neurological imaging studies [12]. Thus, attempting to understand the interactions between activation of known risk genes and higher-level pathway disruptions may help elucidate the causes for structural shifts in SCZ patients [10]. At a psychiatric level, the alterations to various cortical structures create multiple forms of symptoms, including positive (e.g., hallucinogenic episodes) and negative (e.g., anti-social tendencies) [13]. Finally, GWAS for AD cohorts has revealed multiple conserved genetic loci that could encode shared risk factors between the two diseases [14]. Clinical presentations, specifically psychiatric effects, create a crucial point of intersection to be explored where general psychosis was found in up to 60% of AD patients, including hallucination events as well as other effects mirroring those of the positive symptoms found in SCZ patients [15]. Thus, studying shared risk variants and genes between both diseases may help elucidate functional genomics of interest in both diseases and can further uncover cross-disease and disease-specific mechanisms between neurodegenerative and neuropsychiatric diseases.

Recently, advances in single-cell sequencing technologies have generated a great deal of excitement and interest in studying functional genomics at cellular resolution. For example, scRNA-seq and scATAC-seq techniques have measured the transcriptomics and epigenomics of individual cells in the human brain [16, 17]. Further computational analyses have clustered cells into many cell types [16]. The cells in the same type share similar transcriptional activities such as gene expression and genomic functions. Differential gene expression across cell types is a complex, multi-gene dynamic process that tightly regulates and controls functions and is governed by gene regulatory factors such as transcriptional factors (TFs) and

non-coding regulatory elements. These factors cooperate as a gene regulatory network (GRN) to facilitate the correct cellular and molecular functions on the genome scale. Disrupted cooperation can give rise to abnormal gene expression, such as those present in diseases. Therefore, GRN has been used as a robust system to infer genomic functions and molecular mechanisms, especially for human diseases [18].

Recent analyses have also revealed that brain disease risk variants are located in non-coding regulatory elements (e.g., enhancers). The risk genes likely have cell-type-specific effects for both neuronal and non-neuronal cell types [19, 20]. Besides, recent single-cell studies suggest changes to cell-type-specific gene expression in brain diseases [9, 10]. However, our understanding of the underlying gene regulatory mechanisms driving cell-type- and disease-specific gene expression, especially across diseases, remains elusive. To better understand cell-type gene regulatory mechanisms, several computational methods have recently been developed to predict cell-type GRNs [21], such as PIDC [22], GENIE3 [23], and GRNBoost2 [24], aiming to provide deeper mechanistic insights on how transcription factors regulate target gene expression at the cell-type level. However, these methods typically only use single omics (e.g., transcriptomics) and predict networks based on statistical associations (e.g., co-expression), providing insufficient mechanistic insights into gene regulation at the cellular resolution. For instance, how the disease variants affect the transcription factor binding sites (TFBSs) on the distal regulatory elements (e.g., enhancers) that control disease genes is still unclear, especially at the cell-type level. Thus, it is essential to integrate emerging multi-omics data to understand cell-type gene regulation, especially involving non-coding regulatory elements. Recent studies have shown that integrating multi-omics data can reduce the impact of noise from a single omics data and achieve better prediction accuracy [25].

To explore these ideas, we developed a computational pipeline, scGRNom (single-cell Gene Regulatory Network prediction from multi-omics), to integrate multi-omics data and predict cell-type GRNs linking TFs, regulatory elements (e.g., enhancers and promoters), and target genes. scGRNom is a general-purpose tool open-source available at <https://github.com/daifengwanglab/scGRNom> [26]. In particular, we applied scGRNom to the multi-omics data at the cellular resolution, such as chromatin interactions, epigenomics, and single-cell transcriptomics of primary cell types in the human brain, including different excitatory and inhibitory neuronal types, microglia, and oligodendrocyte. Our predictions have high overlapping with state-of-the-art methods for revealing TFs and target genes [21], but they provide additional information on cell-type gene

regulation, such as linking the regulatory elements to the genes. We also found that the enhancers in our cell-type GRNs are enriched with GWAS SNPs in human brain diseases, including psychiatric disorders and AD. Thus, we further linked the GWAS SNPs that interrupt TFBSs to cell-type disease genes based on the cell-type GRNs for SCZ and AD and found cross-disease and disease-specific genomic functions at the cell-type level. Finally, we found that the cell-type disease genes shared by AD and SCZ have improved predicting clinical phenotypes in AD, like disease staging and cognitive impairment.

Methods

Predicting gene regulatory networks from multi-omics data

scGRNom is a computational pipeline in R as a general-purpose tool [26] to (I) integrate multi-omics datasets for predicting gene regulatory networks linking transcription factors, non-coding regulatory elements, and target genes and (II) identify disease genes and regulatory elements. scGRNom can be applied in general to predict either bulk or cell-type disease genes and regulatory networks. First, for predicting gene regulatory networks from multi-omics, scGRNom has three steps (Fig. 1), each of which is available as an R function:

Step 1: Finding chromatin interactions. The function, *scGRNom_interaction*, inputs the chromatin interaction data (e.g., Hi-C) and predicts all possible interactions between enhancers and promoters in the data or the user-provided list—for example, those from topologically associating domains (TADs) in Hi-C data. In addition, the function uses an R package, *GenomicInteractions* [27], to annotate interacting regions and link them to genes.

Step 2: Inferring the transcription factor binding sites on interacting regions. The function, *scGRNom_getTF*, infers the transcription factor binding sites (TFBSs) based on consensus binding site sequences in the enhancers and promoters that potentially interact from the previous step *scGRNom_interaction*. It outputs a reference gene regulatory network linking these TF, enhancers, and/or promoters of genes. In particular, this function uses *TFBSTools* [28] to obtain the position weight matrices of the TFBS motifs from the JASPAR database [29] and predicts the TFBS locations on the enhancers and promoters via mapping TF motifs. The function further links TFs with binding sites on all possible interacting enhancers and promoters and outputs the reference regulatory network. Furthermore, this function can run on a parallel computing version via an R package, *motifmatchr* [30] for computational speed-up.

Step 3: Predicting the gene regulatory network. The function, *scGRNom_getNt*, predicts the final gene

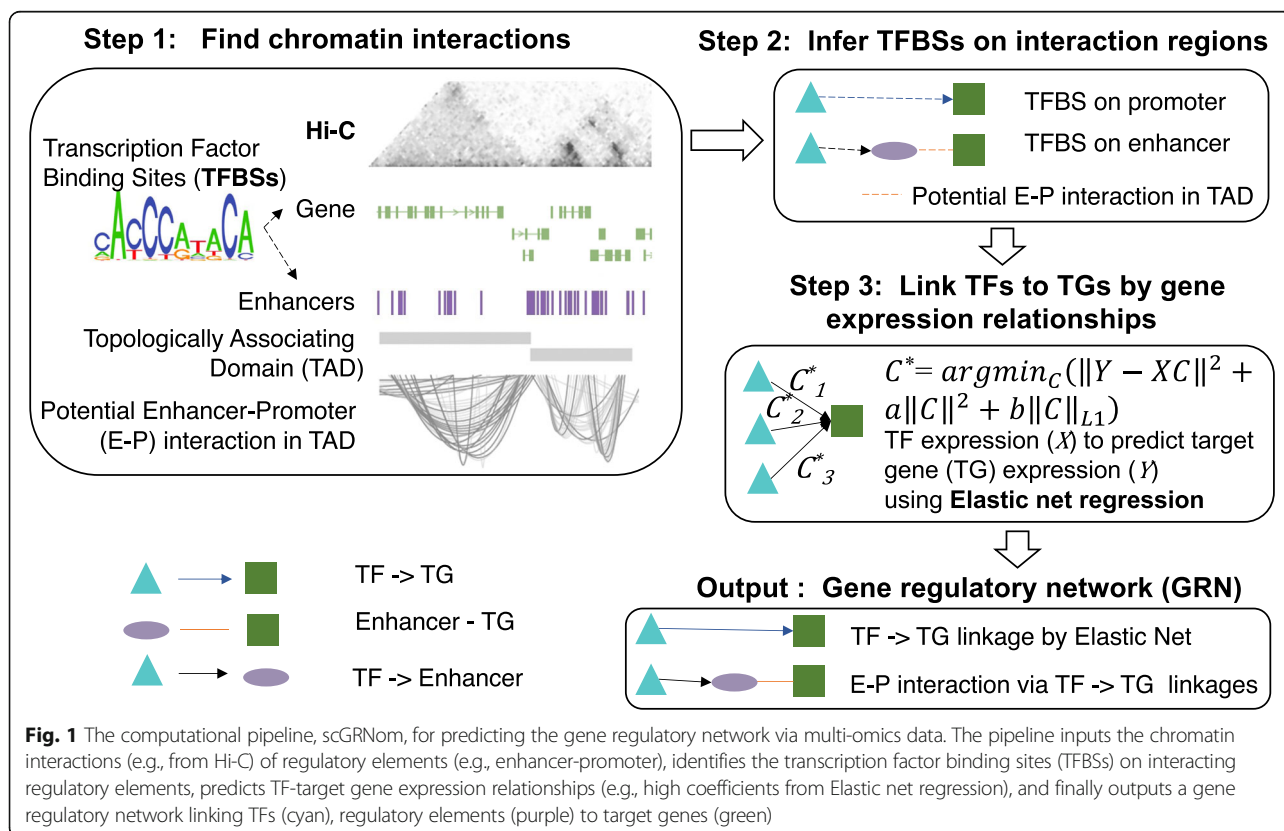


Fig. 1 The computational pipeline, scGRNom, for predicting the gene regulatory network via multi-omics data. The pipeline inputs the chromatin interactions (e.g., from Hi-C) of regulatory elements (e.g., enhancer-promoter), identifies the transcription factor binding sites (TFBSs) on interacting regulatory elements, predicts TF-target gene expression relationships (e.g., high coefficients from Elastic net regression), and finally outputs a gene regulatory network linking TFs (cyan), regulatory elements (purple) to target genes (green)

regulatory network based on the TF-target gene expression relationships in the reference network. The reference gene regulatory network from the previous step provides all possible regulatory relationships (wires) between TF, enhancers, and target genes. However, the chromatin interacting regions are broad, so that many TFs likely have binding sites on them. Also, changes in gene expression may trigger different regulatory wires in the reference network. To refine our maps and determine the activity status of regulatory wires, we apply elastic net regression, a machine learning method that has successfully modeled TF-target gene expression relationships in the gene regulatory networks by our previous work [10]. Further, suppose the chromatin accessibility information is available (e.g., from scATAC-seq data for a cell type). In that case, the function can also filter the enhancers based on their chromatin accessibilities and then output the network links only having the enhancers with high accessibility (e.g., overlapped with scATAC-seq peaks). The parameter “open_chrom” inputs a list of user-defined chromatin accessible regions.

Mathematically, given a gene expression dataset and a reference network (e.g., from *scGRNom_getTF*), the function uses the TF expression to predict each target gene expression and finds the TF with high regression coefficients. Given a target gene, let $y \in \mathcal{R}^n$ be a gene expression vector modeling its expression values across n samples (e.g., n cells from single-cell data) and $X \in \mathcal{R}^{n \times m}$ be the gene expression matrix of m TFs across n samples. Those m TFs should link to the target gene from the reference network, implying possible regulatory relationships to the gene. The elastic net regression model then aims to find the optimal coefficients for m TFs $c^* \in \mathcal{R}^m$ to solve the following optimization problem:

$$c^* = \operatorname{argmin}_c (\|y - Xc\|^2 + \alpha \|c\|^2 + \beta \|c\|_1),$$

where α and β are parameters to adjust the contributions from L_2 and L_1 regularizations of $c \in \mathcal{R}^m$. The samples are randomly divided into the training and testing sets by the parameter, *train_ratio* (e.g., if *train_ratio* = 0.7, then 70% training and 30% testing data). The optimal TF coefficients c^* are estimated by the training data. Also, for the model evaluation, the mean square error (MSE) of the regression is calculated and reported by $\|y_{\text{test}} - X_{\text{test}}c^*\|^2$ using the testing data. Further, the top TFs with high coefficients can be either selected by absolute coefficient values (the parameter, *cutoff_absolute*) or a percentage from all m TFs (the parameter, *cutoff_percentage*). Finally, the function outputs a final gene regulatory network linking the top TFs as well as their linked enhancers (from the reference network, if any) to all possible target genes.

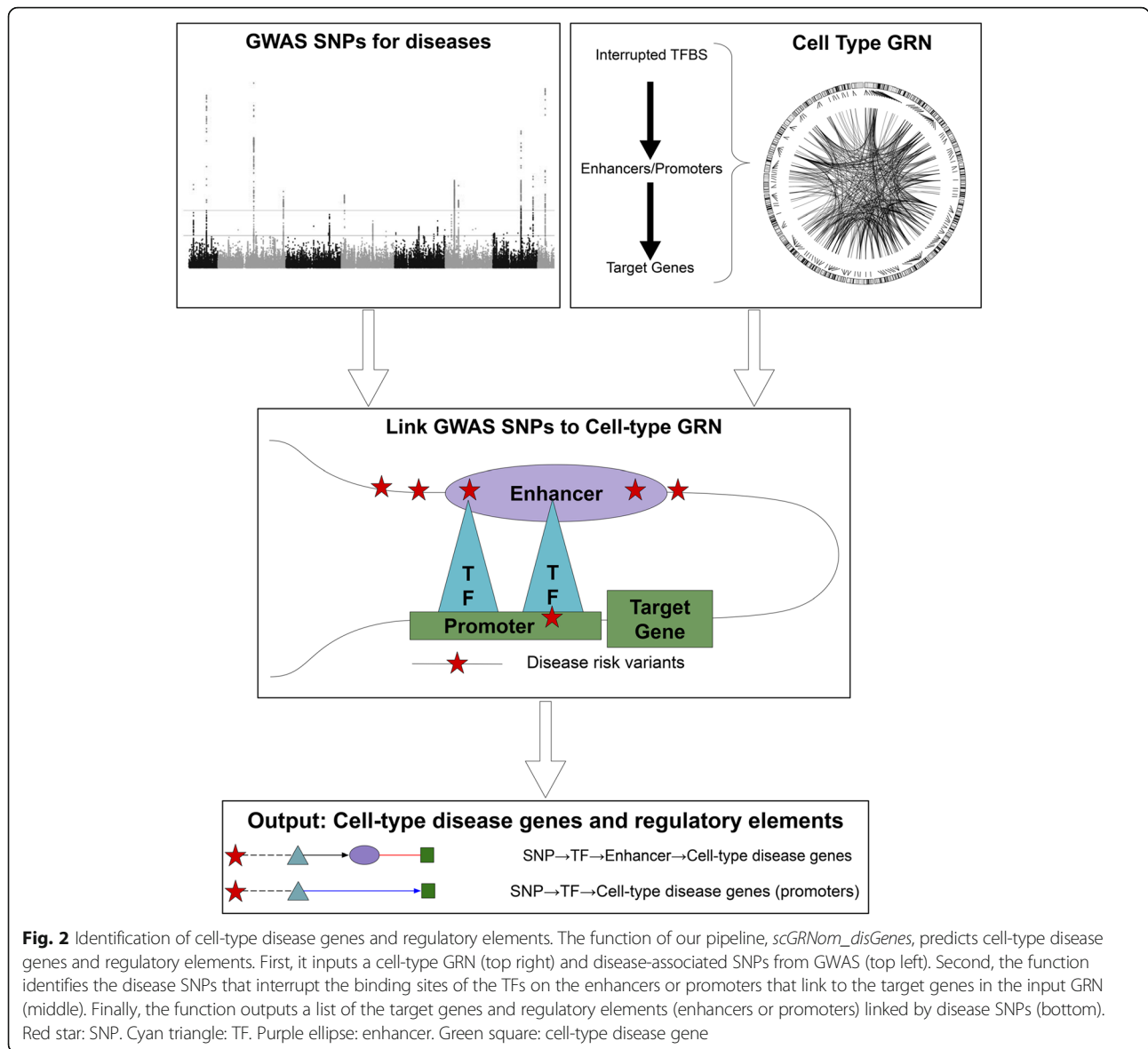
Identifying cell-type disease genes and regulatory elements

In addition to predicting gene regulatory networks, the pipeline also provides another function, *scGRNom_disGenes*, for identifying cell-type disease genes and regulatory elements (e.g., enhancers, promoters). This function's input includes a cell-type gene regulatory network and a list of GWAS SNPs associated with a disease (Fig. 2). The function uses an R package, *GenomicRanges* [31], to overlap these disease SNPs with the enhancers and promoters of the input cell-type gene regulatory network, and then find the ones that interrupt the binding sites of all possible TFs (TFBSs) on the enhancers and promoters by *motif-breakR* [32]. It finally maps the overlapped enhancers or promoters and TFs with interrupted TFBSs onto the input network to find the linked genes and enhancers/promoters as the output cell-type disease genes and regulatory elements.

Application to multi-omics data and data processing for predicting cell-type gene regulatory networks in the human brain

We applied the *scGRNom* pipeline to the multi-omics data for the human brain, including cell-type chromatin interactions [19], transcription factor binding sites [28], single-cell transcriptomics [16], and recent cell-type open chromatin regions [17] for predicting cell-type GRNs in the human brain. In particular, we predicted cell-type gene regulatory networks for major cell types: excitatory and inhibitory neurons, microglia, and oligodendrocyte for the human brain [16]. The excitatory neuronal types include Ex1, Ex2, Ex3e, Ex4, Ex5b, Ex6a, Ex6b, Ex8, and Ex9. The inhibitory neuronal types include In1a, In1b, In1c, In3, In4a, In4b, In6a, In6b, In7, and In8. We first input recently published cell-type chromatin interactome data in the human brain [19] to *scGRNom_interaction* to reveal all possible interactions from enhancers to gene promoters in the neuronal, microglia, and oligodendrocyte types. The genome annotation was from TxDb.Hsapiens.UCSC.hg19.knownGene [33]. We then predicted a reference regulatory network for each of these cell types using *scGRNom_getTF*. Finally, given a cell type, we input the single-cell gene expression data for the type and the reference network from *scGRNom_getTF* to *scGRNom_getNt* for predicting the cell-type gene regulatory network (GRN).

Specifically, to make our predicted networks comparable across cell types, we made the following data preparation and processing steps. First, although different studies have generated increasing numbers of single-cell data (e.g., for microglia [9]), we used the data from one study (GSE97942) [16] that includes the gene expression data (UMI) of individual cells of primary cell types, all from one postmortem tissue of human frontal cortex,



aiming not to introduce additional batches from different studies, which helps our further comparative analyses across cell types. In addition, we filtered the genes that express in less than 100 cells. We then normalized gene expression by *Seurat 4.0* [34] for further removing noises and batches across cell types. We also applied the method *MAGIC* [35] to impute the single-cell gene expression of all cells to address potential dropout issues. Then, for each cell type, we removed the lowly expressed genes with $\log_{10}(\text{sum of imputed gene expression levels of the cells of the cell type}+1) < 1$. The numbers of genes and cells for each cell type used for prediction are available in Additional file 1.

For predicting a cell-type GRN using the Elastic net regression model, we randomly split the cells into 70% training and 30% testing sets. We then selected

the best Elastic net model that minimized the mean square error (MSE). We then filtered the target genes based on the goodness of fit by MSEs. In particular, we removed the target genes predicted by Elastic net regression with $\text{MSE} > 0.1$ and also TF-target gene with absolute Elastic net coefficient < 0.01 . The output cell-type GRN consists of the network edges that link TFs, enhancers (if any), and target genes (TGs), as well as the Elastic net coefficient of TF-TGs for each edge. Finally, we provided two versions of each cell-type GRN (Additional file 2):

- (I) The edges that include the enhancers that overlap the cell-type open chromatin regions predicted by recent scATAC-seq data (broad excitatory and inhibitory neurons, microglia, and oligodendrocyte) [17]

- (II) The edges that only include the top 10% TFs with absolute coefficients for each target gene without considering cell-type open chromatin regions (for scATAC-seq data might be likely noisy and no open chromatin regions available for neuronal subtypes)

Comparison with state-of-the-art methods

We compared our scGRNom predictions with existing state-of-the-art methods. In particular, we input the single-cell gene expression data for each cell type to a recently published benchmark framework, BEELINE [21], and predicted the cell-type regulatory networks using three of the most consistent and highly accurate methods, PIDC [22], GENIE3 [23], and GRNBoost2 [24]. These methods only input gene expression data to predict all possible TF-target gene (TG) regulatory links based on their expression relationships, without considering the regulatory elements or open chromatin regions at the cell-type level. Thus, after applying to our processed single-cell gene expression data, they generated more network edges than us because scGRNom only keeps the TF-TG links in which TFs have binding sites on the regulatory elements (e.g., enhancers and promoters). Therefore, to make these networks comparable with ours, we extended our networks by selecting up to the top 30% TFs for each TG and then checked if our TF-TG links predicted by the state-of-the-art method (also up to top 30% TFs for each TG selected for each method). In particular, given a cell-type GRN by scGRNom, we selected top K TFs per target gene (TG) to see if the TF-TG pairs were also predicted by PIDC, GENIE3, or GRNBoost2 (again picking top K TFs per TG). We varied K values from 0 to 30% and then calculated the percentages of the scGRNom's TF-TG pairs that can be predicted by one of those methods (Additional file 3: Figure S1).

GWAS SNPs and heritability enrichment analyses for the enhancers in the cell-type gene regulatory networks in the human brain

Genome-wide association studies (GWAS) have identified a variety of genetic risk variants, including single nucleotide polymorphisms (SNPs) that are significantly associated with diseases and phenotypes (i.e., the traits). For example, recent GWAS studies have identified many SNPs associated with AD (2357 credible SNPs) [6] and SCZ (6105 credible SNPs) [7]. In addition to the credible SNPs, we also included additional SNPs with $p < 5e-5$ from the AD and SCZ GWAS summary statistics for linking potential additional cell-type disease genes [36]. We applied the *partitioned linkage disequilibrium score regression* (LDSC) [37] to evaluate the heritability explained by the enhancers of our cell-type gene regulatory

networks for GWAS SNPs. In particular, our heritability enrichment analyses used the GWAS summary statistics for the diseases or traits: SCZ [7], AD [6], autism spectrum disorder (ASD) [38], bipolar disorder (BPD) [39], amyotrophic lateral sclerosis (ALS) [40], major depressive disorder (MDD) [41], intelligence [42], multiple sclerosis (MS) [43], Parkinson's disease (PD) [44], attention-deficit hyperactivity disorder (ADHD) [45], education [46], type 2 diabetes (T2D) [47], inflammatory bowel disease (IBD) [48], and coronary artery disease (CAD) [49]. Also, we provided the numbers of GWAS SNPs for AD and SCZ that interrupt the binding sites of at least one of all possible TFBSs and the binding sites of the regulatory TFs in each cell-type GRN (Additional file 3: Table S1).

Identification and enrichment analyses of cell-type disease genes

Given a cell type and a disease type (AD or SCZ), we input the cell-type GRNs (both versions) and the GWAS SNPs for the disease to the *scGRNom_disGenes* function for identifying the cell-type disease genes. We identified the cell-type disease genes (AD and SCZ) for all the cell types as above, including excitatory and inhibitory neuronal subtypes, microglia, and oligodendrocyte (Additional file 4). We also merged the disease genes for excitatory/inhibitory neuronal subtypes as the broad excitatory/inhibitory neuronal disease genes. We used the web app, *Metascape* [50], to find the enrichments of cell-type disease genes such as KEGG pathways, GO terms, protein-protein interactions, and diseases (via DisGeNET). Enrichment p -values shown in this paper were adjusted using the Benjamin-Hochberg (B-H) correction. Also, we looked at the expression levels of our cell-type disease genes in the disease samples. In particular, we compared the published population gene expression data in AD for single-cell expression [9].

Machine learning prediction of clinical phenotypes from cell-type disease genes

Finally, we used the machine learning approach to predict clinical phenotypes from our cell-type disease genes using the population data of the ROSMAP project, an independent AD cohort [51]. Given a clinical phenotype, we assume that $X_i \in R^d$ represents the expression data of d disease genes for the i th individual in the cohort and $Y_i \in \{0, 1\}$ represents the binarized class of the i th individual's phenotype with $i \in 1, \dots, n$ individuals for training. We then found the optimal logistic regression model with the parameters $\{\beta_0^*, \beta_1^* \in R^d\}$ to classify the phenotype from the disease gene expression data via minimizing the following loss function:

$$\{\beta_0^*, \beta_1^*\} = \underset{\beta_0, \beta_1}{\operatorname{argmax}} \sum_{i=1}^n -Y_i \log(\beta_0 + \beta_1^T X_i) - (1 - Y_i) \log(1 - \beta_0 - \beta_1^T X_i)$$

where $(\cdot)^T$ is the transpose operation. Also, we performed cross-validation ($K = 5$) for the individual samples with 80% training and 20% testing sets. We also balanced the class sample size in each training set by the weighting method [52] so that the baseline of the classification accuracy is 50% for two classes. We used the individuals from the training sets to train the classification model. We then used the individuals from the testing sets to evaluate the classification performance, i.e., accuracy. In particular, we predicted four clinical phenotypes in ROSMAP: Braak stages that measure the severity of neurofibrillary tangle (NFT) pathology (Braak early stages (0, 1, 2, 3) vs. late stages (4, 5, 6)), CERAD scores that measure neuritic plaques (no AD vs. AD), the diagnosis of cognitive status (DCFDX, no or mild cognitive impairments (1, 2, 3) vs. Alzheimer's dementia (4, 5)), and the cognitive status at the time of death (COGDX, no or mild cognitive impairments (1, 2, 3) vs. Alzheimer's dementia (4, 5)).

Results

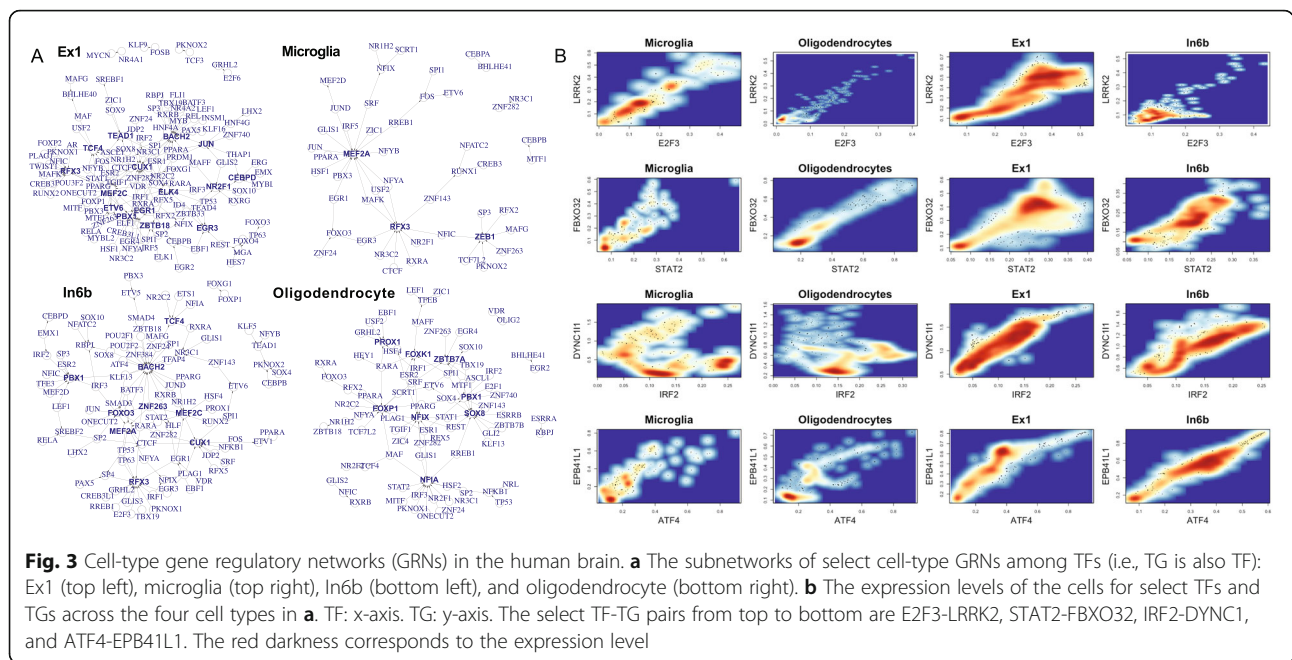
Predicting cell-type gene regulatory networks in the human brain

We applied the scGRNom pipeline to the multi-omics data for the human brain, including cell-type chromatin interactions [19], transcription factor binding sites [28], single-cell transcriptomics [16], and cell-type open chromatin regions [17] (“Methods” section). We predicted the cell-type gene regulatory networks (GRNs) for both glial and neuronal cell types in the human brain, including microglia, oligodendrocyte, excitatory neuronal subtypes (Ex1, Ex2, Ex3e, Ex4, Ex5b, Ex6a, Ex6b, Ex8, and Ex9), and inhibitory neuronal subtypes (In1a, In1b, In1c, In3, In4a, In4b, In6a, In6b, In7, and In8). Each cell-type GRN links TFs to enhancers to target genes (TGs) and has two versions that (I) only include the network edges with open cell-type enhancers predicted by recent scATAC-seq data [17] and (II) only include the edges with top 10% TFs with highest absolute coefficients for each target gene without using cell-type open chromatin regions. The reason why we included the second version is that the open chromatin regions predicted by scATAC-seq may not be highly accurate and cell-type-specific, given that the scATAC-seq data is noisy and also currently unavailable for neuronal subtypes (e.g., Ex1-9, In1-8). The network statistics such as numbers of cells, edges, TFs, enhancers, and TGs for all cell-type GRNs in the human brain as above are available in Additional file 1. For instance, we found that the microglia

GRN with open chromatin regions consists of 47,353 edges linking 180 TFs, 1893 microglia open enhancers, and 1236 TGs. The edge lists of cell-type GRNs are available in Additional file 2.

Our GRNs reveal many known cell-type-specific regulations. Figure 3A visualizes the subnetworks among TFs for select cell types (i.e., TGs are also TFs). For example, two known TFs, MEF2A and RFX3, that control microglia phenotypes play hub roles in the microglia network [53, 54]. The nuclear factors NFIA, NFIX, and FOXP1 controlling neural differentiation and gliogenesis are also hub genes in the oligodendrocyte network [55–57]. MEF2C regulating inhibitory and excitatory synapses is a central node in both excitatory and inhibitory networks (e.g., Ex1 and In6b) [58]. In addition to cell-type TFs, we also observed the cell-type-specific expression relationships between TFs and TGs (high correlation). For instance, in Fig. 3B, E2F3-LRRK2, STAT2-FBXO32, IRF2-DYNC1, and ATF4-EPB41L1 show cell-type-specific expression relationships across the cells of microglia, oligodendrocyte, Ex1, and In6b types.

In addition, we compared our predicted cell-type gene regulatory networks with existing state-of-the-art methods for predicting cell-type gene regulatory networks, particularly those that are consistent and highly accurate PIDC, GENIE3, and GRNBoost2 benchmarked by BEELINE [21] (“Methods” section). The percentages of the overlapped TF-TG links of the cell-type network between scGRNom (both versions) and the state-of-the-art methods are over 50% (“Methods” section, Additional file 3: Figure S1). This suggests a high consistency between scGRNom and these methods. However, these methods predicted TF-TG links without providing information on regulatory elements like enhancers. Thus, we looked further at the enhancers in our cell-type GRNs and found that they have significantly high heritability enrichments for GWAS SNPs of multiple brain diseases and traits ($p < 0.05$, “Methods” section). For example, the enhancers in our excitatory-neuron and inhibitory-neuron GRNs have high enrichments for AD, SCZ, major depressive disorder, intelligence, and education (Fig. 4A, Additional file 3: Figure S2). Besides, these brain-cell-type enhancers do not have significant enrichment of GWAS SNPs for non-brain diseases. Therefore, the heritability enrichment analysis suggests that the enhancers in our cell-type GRNs have potential pleiotropic roles associated with multiple brain diseases or traits. Finally, we also compared our cell-type networks with public GRN databases such as TRRUST [59], Dorothea [60], and RegNetwork [61]. We found that the overlaps are not significant (hypergeometric test p -value > 0.999 , Additional file 3: Table S2). In fact, those public GRNs were primarily inferred by integrating different studies from the literature (e.g., via physical interactions, co-



expressed genes at the bulk tissue level) and thus may not be specific for the neuronal and glial cell types in the human brain. However, the overlapped network edges suggest the potential associations of those public GRNs with the human brain’s cell types. For example, we found that 319 TRRUST edges, 5935 Dorothea edges, and 106 RegNetwork edges overlap with at least one of our cell-type GRNs.

Identifying cell-type disease genes in AD and SCZ for neuronal and glial cell types

We used these cell-type GRNs to link GWAS SNPs with disease risk genes for each cell type, advancing knowledge on cross-disease and disease-specific interplays among genetic, transcriptional, and epigenetic risks at cellular resolution. In particular, we chose SCZ and AD, two majorly represented neuropsychiatric and neurodegenerative diseases with potential convergent underlying mechanisms [4], and we linked a number of cell-type disease genes (“Methods” section, Additional file 4) and performed their enrichments (Additional file 5). We found that many disease genes present in one or a few cell types only, suggesting potential cell-type-specific contributions to AD and SCZ (Additional file 3: Figure S3). As shown in Fig. 4B, the cell-type AD genes are also significantly enriched with known disease genes for AD and other dementia diseases such as brain atrophy and cerebral amyloid angiopathy ($p < 0.01$). For example, the microglia AD genes, including APP, CLU, BACE2, and BIN1, are specifically enriched for other diseases such as neurofibrillary degeneration, Parkinson’s dementia, and aging memory impairment. Also, the cell-type SCZ

genes are enriched for various disorders such as neurodevelopmental, mental, and mood disorders and depression ($p < 0.01$, Fig. 4C). Furthermore, many cell-type disease genes have corresponding expression activities in the disease samples. For example, 72 excitatory AD genes (86%) and 65 inhibitory AD genes (80%) (plus 22 oligodendrocyte AD genes and three microglia AD genes) are significantly differentially expressed in the corresponding cell types in AD individuals, respectively ($p < 0.05$) [9].

The functional enrichment analyses for our cell-type disease genes also uncover known genomic functions and pathways at the cell-type level. For instance, the microglia AD genes are enriched with amyloid beta formation and clearance [62], MAPK signaling [63], and neuron death [64] (Fig. 4D), and the oligodendrocyte AD genes are enriched with Tau protein binding [65]. This is vital to understanding a multitude of diseases that commonly demonstrate atrophy of cortical tissue as a hallmark feature. In SCZ genes, we also found that multiple key hallmark pathways were enriched, such as dopaminergic synapse [66], trans-synaptic signaling [67], and synapse organization [68] for excitatory SCZ genes (Fig. 4E). For inhibitory SCZ genes, we observed that MAPK family signaling [69], regulation of NMDA receptor activity [70], dopaminergic synapses [66], and neurotransmission [70] are enriched.

Comparative analyses reveal the interplays between genomic functions, pathways, cell types, and diseases

In addition to cell-type-specific pathways in these diseases, we also identified those involving multiple cell

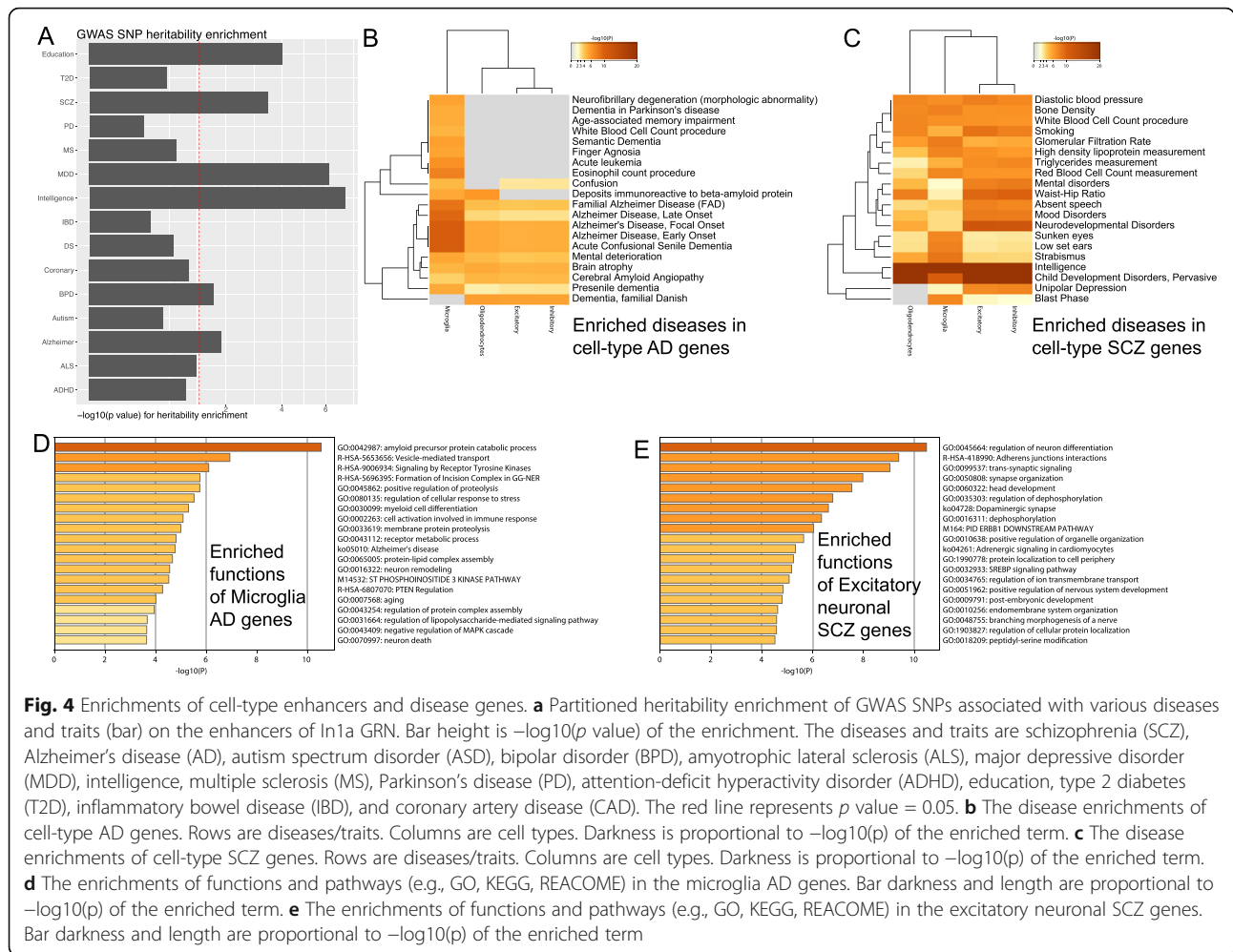


Fig. 4 Enrichments of cell-type enhancers and disease genes. **a** Partitioned heritability enrichment of GWAS SNPs associated with various diseases and traits (bar) on the enhancers of In1a GRN. Bar height is $-\log_{10}(p)$ value of the enrichment. The diseases and traits are schizophrenia (SCZ), Alzheimer’s disease (AD), autism spectrum disorder (ASD), bipolar disorder (BPD), amyotrophic lateral sclerosis (ALS), major depressive disorder (MDD), intelligence, multiple sclerosis (MS), Parkinson’s disease (PD), attention-deficit hyperactivity disorder (ADHD), education, type 2 diabetes (T2D), inflammatory bowel disease (IBD), and coronary artery disease (CAD). The red line represents p value = 0.05. **b** The disease enrichments of cell-type AD genes. Rows are diseases/traits. Columns are cell types. Darkness is proportional to $-\log_{10}(p)$ of the enriched term. **c** The disease enrichments of cell-type SCZ genes. Rows are diseases/traits. Columns are cell types. Darkness is proportional to $-\log_{10}(p)$ of the enriched term. **d** The enrichments of functions and pathways (e.g., GO, KEGG, REACOME) in the microglia AD genes. Bar darkness and length are proportional to $-\log_{10}(p)$ of the enriched term. **e** The enrichments of functions and pathways (e.g., GO, KEGG, REACOME) in the excitatory neuronal SCZ genes. Bar darkness and length are proportional to $-\log_{10}(p)$ of the enriched term

types in each disease, implying that potential cell-type interactions drive the disease pathology. For example, the enrichment of SCZ primarily includes changes in synapse structures and cell shaping and differentiation (Fig. 5A). Clinically, this is consistent with the consensus that SCZ is strictly neuropsychiatric as opposed to degenerative. In particular, cell morphogenesis and regulation of neuron differentiation are enriched in all four major cell type SCZ genes ($p < 0.01$). Early life neurodevelopmental genetic markers may suggest causal links with alterations in hippocampal cell differentiation points on the front of cell morphogenesis, leading to cascades of downstream effects [71]. This has primarily been studied and modeled within the scope of iPSC-based analyses, which make correlations and connections to the clinical presentation more difficult due to the additional abstraction from the standard pathology-based analysis. Also, the BDNF signaling pathway that potentially relates to intercellular communications is enriched as well in multiple cell types ($p < 0.01$) [72]. Finally, we also observed that protein-protein interactions

(PPIs) are enriched among the disease genes of the cell types at a higher level. As shown in Fig. 5B, the SCZ genes for dopaminergic synapse, NMDA receptors, glutamate binding, and activation are shared by multiple cell types and have strong PPIs, implying protein-level cross-type coordination [73]. In AD, multiple pathways were significantly enriched across various cell types (Fig. 5C). For instance, the catabolic process for the AD key player, amyloid precursor protein (APP), is enriched with both glial and neuronal types ($p < 0.01$) [74].

Furthermore, we found that cross-disease conserved functions/pathways are involved in one or multiple cell types, revealing potential novel functional interplays across cell types and diseases (Fig. 5D). For example, the cell morphogenesis involved in differentiation is enriched in both AD and SCZ neuronal genes. Another example is that the vesicle-mediated transport is enriched for both AD and SCZ microglia genes. In total, we found 11 microglia genes shared by AD and SCZ. In particular, the VEGF signaling pathway is enriched in the SCZ microglia genes. In the general theme of AD

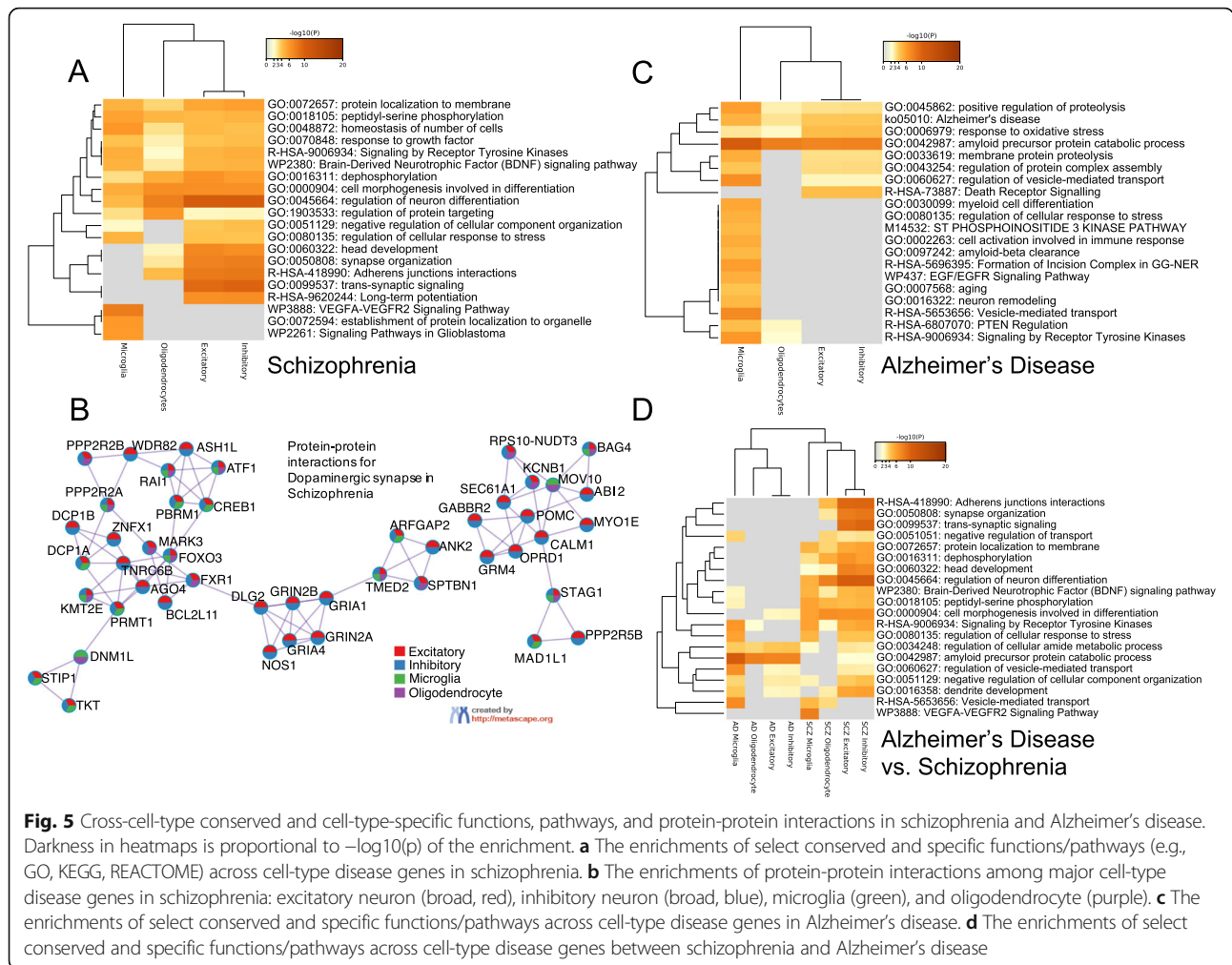


Fig. 5 Cross-cell-type conserved and cell-type-specific functions, pathways, and protein-protein interactions in schizophrenia and Alzheimer's disease. Darkness in heatmaps is proportional to $-\log_{10}(p)$ of the enrichment. **a** The enrichments of select conserved and specific functions/pathways (e.g., GO, KEGG, REACTOME) across cell-type disease genes in schizophrenia. **b** The enrichments of protein-protein interactions among major cell-type disease genes in schizophrenia: excitatory neuron (broad, red), inhibitory neuron (broad, blue), microglia (green), and oligodendrocyte (purple). **c** The enrichments of select conserved and specific functions/pathways across cell-type disease genes in Alzheimer's disease. **d** The enrichments of select conserved and specific functions/pathways across cell-type disease genes between schizophrenia and Alzheimer's disease

pathology, increased VEGF expression has been linked to worse cognitive outcomes in postmortem analysis [75]. Similarly, multiple meta-analyses have revealed differential expression levels between healthy controls and SCZ patients [76]. However, little has been done to link potential gene function to cell-type-level interactions and pathways. Here, these SCZ-AD shared microglia genes may help explain shared higher-level dysfunction between both diseases as evidenced by higher expression. Also, we found that some functions involve different cell types across diseases. The dendrite development has been found in both SCZ and AD pathology [77, 78]. We found that it is mainly enriched with microglia in AD but neuronal types in SCZ.

More interestingly, when exploring the interactions between cell types that change between diseases, the disease-specific pathologies enter to explain the cause of discrepancies. In particular, for AD, it is shown that phagocytic microglia are activated during the early stages of synaptic decline, leading to eventual neuroinflammation and programmed cell death [79]. For SCZ, the

oligodendrocyte enrichment reveals similar intercellular mechanisms between excitatory and inhibitory neurons, specifically those regulating neuron differentiation (Fig. 5A), providing potential direction for future exploration and validation of the communication role of oligodendrocytes [80].

Prediction of clinical phenotypes from cell-type disease genes

Finally, we want to investigate the clinical applications of our cell-type disease genes. To this end, we looked at the population-level gene expression data for AD in the ROSMAP cohort [51]. In particular, we first found that many cell-type AD genes have significantly associated expression levels with clinical phenotypes across individuals in ROSMAP. For example, out of 195 cell-type AD genes, we found that 72 genes are significantly associated with the Braak stages that measure the severity of neurofibrillary tangle (NFT) pathology (ANOVA $p < 0.05$), 78 genes with the CERAD scores that measure neuritic plaques (ANOVA $p < 0.05$), 89 genes with the diagnosis of

cognitive status (DCFDX) (ANOVA $p < 0.05$), 73 genes with the cognitive status at the time of death (COGDX) (ANOVA $p < 0.05$), and 92 genes with the Mini-Mental State Examination (MMSE) scores (Pearson correlation $p < 0.05$). In total, 135 cell-type AD genes were significantly associated with at least one clinical phenotype in ROSMAP (Additional file 6).

In addition to statistically significant associations between our cell-type disease genes and clinical phenotypes, we also applied the machine learning approach to predict clinical phenotypes from these cell-type disease genes using ROSMAP data (“Methods” section). Specifically, for each clinical phenotype, we used the logistic regression model to classify individual states of the phenotype (as classes) from their expression data of 53 AD-SCZ shared cell-type genes. We performed the cross-validation ($K = 5$) for the individual samples with 80% training and 20% testing sets. As shown in Fig. 6, our average accuracy values for classifying all four major clinical phenotypes: Braak stages, CERAD scores, DCFDX status, and COGDX status, are much higher than the baselines (50%), the random select genes, and AD genes from the latest GWAS study [6]. This suggests that using cell-type disease genes shared by AD and SCZ has improved predicting those clinical phenotypes, especially for cognitive-related ones in AD.

Discussion

This paper focuses on the scGRNom’s application to single-cell data for AD and SCZ. However, scGRNom is

general-purpose for understanding functional genomics and gene regulation in other diseases. Besides the cell types, the pipeline also predicts the gene regulatory networks for cell clusters (unknown cell types) and bulk tissue types. Furthermore, recent eQTL studies have identified various SNPs associating with gene expression in multiple brain tissue types using the population data such as GTEx [81] and PsychENCODE [10]. Although those brain eQTLs suggest the SNP-gene association at the bulk tissue level, we still found that several eQTLs in PsychENCODE match our linked GWAS SNPs and cell-type disease genes, e.g., SNPs chr2:127846321 for and chr20:43598154 for STK4, two microglia disease genes for AD and SCZ, respectively. This suggests potential cell-type effects of these human brain eQTLs. Thus, increasing single-cell data at the population level allows us to predict the cell-type eQTLs [82], which will likely help understand cell-type gene regulation and refine linking disease genes at the cell-type level.

For linking disease genes, we primarily used the interrupted TFBSs by GWAS SNPs. Future studies utilizing scGRNom would be able to take advantage of the ever-growing number of GWAS for a wide variety of diseases as well as single-cell data. However, additional information can also help link GWAS SNPs to disease genes. For example, existing tools such as FUMA [83] have linked GWAS loci to genes by integrating information from multiple resources, providing more functional linkages from genotype to genes to phenotype in human diseases. Thus, incorporating multiple GWAS data (e.g.,

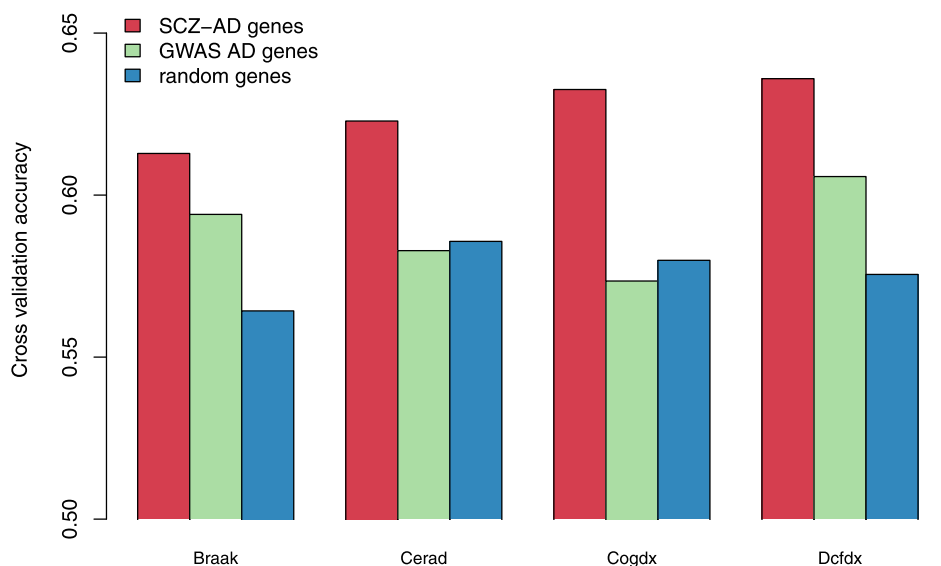


Fig. 6 Prediction accuracy of AD clinical phenotypes from disease genes. The AD population data for prediction was from the ROSMAP cohort [51]. AD clinical phenotypes include Braak—stages that measure the severity of neurofibrillary tangle (NFT) pathology; Cerad—scores that measure neuritic plaques; Cogdx—cognitive status at the time of death; and Dcfdx—the diagnosis of cognitive status. The bar height represents the average accuracy of cross-validation ($K = 5$) from the prediction using logistic regression (“Methods” section). Red: scGRNom’s cell-type disease genes shared by AD and SCZ (SCZ-AD genes). Green: AD genes from GWAS [6]. Blue: randomly select genes (same number as SCZ-AD genes)

various brain regions) exposes key areas of observed phenotypes. Previous studies have demonstrated the caution that must be exercised when attempting to correlate GWAS data with clinical phenotypes, and methods such as our analysis mitigate these effects [20]. A similar methodology as outlined could be used where common loci within each set of summary statistics are incorporated and established before integration into the cell-type GRNs, thus linking neuronal spatial information with known mutation sites in patients along with potentially cell-type-specific functionality. Expanding past the cell types examined here into additional multi-cellular analyses is also possible given expanded interactome data. Notably, this would allow for further investigation of complex neuropsychological diseases and cases where the line between different clinical classifications becomes blurred and leads to additional complications about clinically relevant genetic therapies. One such example includes autism spectrum disorder (ASD), where clinical presentations can vary in multiple axes of severity, creating a broad spectrum of phenotypes. In such cases, potentially linking specific symptoms or aspects of a particular subset of ASD to particular brain regions and cell types allows for a better-informed picture of functional consequences associated with genetic mutation sites. Such connections could aid in determining genetic risk factors associated with variations in edge case patients; they also create the opportunity to take advantage of induced pluripotent stem cell (iPSC) technology using genetic engineering technologies to create point mutations matching computationally identified genes. Moreover, because the single-cell multi-omics data we used for predicting cell-type GRNs are not specific for particular diseases (AD or SCZ), we used the SNPs disrupting all possible TFs on the enhancers and promoters from our cell-type GRNs to link at large cell-type disease genes. However, the scGRNom pipeline is general-purpose and able to work for incoming disease-specific single-cell multi-omics and link to cell-type disease genes via interrupted regulatory TFs in the diseases.

Machine learning has also been widely used to analyze multi-omics, such as multiview learning and deep learning [10, 84]. Multiview learning has great potential for understanding functional multi-omics and revealing nonlinear interactions across omics. Therefore, integrating such emerging machine learning approaches will enable identifying different cross-omic patterns, especially for increasing single-cell multi-omics data and providing more comprehensive mechanistic insights in cell-type gene regulation and linking to disease genes. For example, this means adding more omics such as methylation data that reflect epigenetic changes that may occur due to wide variations of inherited and environmental

factors [85]. At a deeper functional level, variations in methylation have been attributed to alterations in splicing activity, ultimately impacting the regulation and expression of key genes [86]. Additionally, integrating proteomic data at a single-cell level enhances the broader picture formed through additional data sources even further [87]. Lastly, expanding past simple methylation and proteomics allows for the ability to include all forms of data incorporated through the single-cell cytometry [88].

Conclusions

We developed a computational pipeline, scGRNom, to integrate multi-omics data and predict gene regulatory networks (GRNs), which link TFs, non-coding regulatory elements (e.g., enhancers), and target genes. With applications to the data from single-cell multi-omics of the human brain, we predicted cell-type GRNs for both neuronal (e.g., excitatory, inhibitory) and glial cell types (e.g., microglia, oligodendrocyte). Further, scGRNom can input cell-type GRNs and disease risk variants to link disease genes at the cell-type level, such as brain diseases like AD and SCZ. These disease genes revealed conserved and specific genomic functions across neuropsychiatric and neurodegenerative diseases, providing cross-disease regulatory mechanistic insights at the cellular resolution. Although this paper focuses on AD and SCZ, scGRNom is a general-purpose tool for understanding functional genomics and gene regulation in other diseases, at either bulk tissue or cell-type levels. Finally, scGRNom is open-source available at <https://github.com/daifengwanglab/scGRNom> [26].

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-021-00908-9>.

Additional file 1. Statistics of cell-type gene regulatory networks in the human brain.

Additional file 2. Human brain cell-type gene regulatory networks (TF, enhancer, target gene).

Additional file 3. Supplementary figures (Figures S1-S3) and tables (Tables S1-S2).

Additional file 4. Human brain cell-type disease genes for AD and SCZ.

Additional file 5. Enrichments of human brain cell-type disease genes for AD and SCZ.

Additional file 6. Human brain cell-type disease genes associated with AD clinical phenotypes.

Acknowledgements

We thank the reviewers and editors for their careful work and insightful comments and suggestions.

Authors' contributions

D.W. conceived and designed the study. T.J. and M.Y. implemented the software. T.J., P.R., M.Y., J.H., S.L., and D.W. analyzed the data. T.J., P.R., M.Y.,

Pa.R., and D.W. wrote the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the grants of the National Institutes of Health, R01AG067025, R21CA237955, and U01MH116492 to D.W., U54HD090256 to Waisman Center, and the start-up funding for D.W. from the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison.

Availability of data and materials

Our pipeline for predicting gene regulatory networks via multi-omics with a tutorial and codes are open-source available at <https://github.com/daifengwanglab/scGRNom> [26]. All data supporting this study are included in this paper and its additional files.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Biostatistics and Medical Informatics, University of Wisconsin – Madison, Madison, WI 53706, USA. ²Waisman Center, University of Wisconsin – Madison, Madison, WI 53705, USA. ³Department of Integrative Biology, University of Wisconsin – Madison, Madison, WI 53706, USA. ⁴Present address: Morgridge Institute for Research, Madison, WI 53715, USA. ⁵Department of Statistics, University of Wisconsin – Madison, Madison, WI 53706, USA. ⁶Present address: Department of Statistics, Rutgers University, Piscataway, NJ 08854, USA. ⁷Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. ⁸Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. ⁹Department of Computer Sciences, University of Wisconsin – Madison, Madison, WI 53706, USA.

Received: 4 October 2020 Accepted: 13 May 2021

Published online: 27 May 2021

References

- Cross-Disorder Group of the Psychiatric Genomics Consortium. Electronic address: plee0@mgh.harvard.edu, Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell*. 2019;179:1469–1482.e11.
- Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. Analysis of shared heritability in common disorders of the brain. *Science*. 2018;360(6395):eaap8757.
- Ciccocioppo F, Bologna G, Ercolino E, Pierdomenico L, Simeone P, Lanuti P, et al. Neurodegenerative diseases as proteiopathies-driven immune disorders. *Neural Regen Res*. 2020;15:850–6.
- Steinberg M, Shao H, Zandi P, Lyketsos CG, Welsh-Bohmer KA, Norton MC, et al. Point and 5-year period prevalence of neuropsychiatric symptoms in dementia: the Cache County Study. *Int J Geriatr Psychiatry*. 2008;23:170–7.
- Cummings J, Ritter A, Rothenberg K. Advances in management of neuropsychiatric syndromes in neurodegenerative diseases. *Curr Psychiatry Rep*. 2019;21:79.
- Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease. *Int J Geriatr Psychiatry*. 2019;51:404–51.
- Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet*. 2018;50:381–9.
- Shoghi-Jadid K, Small GW, Agdeppa ED, Kepe V, Ercoli LM, Siddarth P, et al. Localization of neurofibrillary tangles and beta-amyloid plaques in the brains of living patients with Alzheimer disease. *Am J Geriatr Psychiatry*. 2002;10:24–35.
- Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*. 2019;570:332–7.
- Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP, et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science*. 2018;362(6420):eaat8464.
- Moghaddam B, Javitt D. From revolution to evolution: the glutamate hypothesis of schizophrenia and its implication for treatment. *Neuropsychopharmacology*. 2012;37:4–15.
- Akbarian S, Liu C, Knowles JA, Vaccarino FM, Farnham PJ, Crawford GE, et al. The PsychENCODE project. *Nat Neurosci*. United States. 2015;18:1707–12.
- Karlsgodt KH, Sun D, Cannon TD. Structural and functional brain abnormalities in schizophrenia. *Curr Dir Psychol Sci*. 2010;19:226–31.
- DeMichele-Sweet MAA, Weamer EA, Klei L, Vrana DT, Hollingshead DJ, Seltman HJ, et al. Genetic risk for schizophrenia and psychosis in Alzheimer disease. *Mol Psychiatry*. 2018;23:963–72.
- Murray PS, Kumar S, Demichele-Sweet MAA, Sweet RA. Psychosis in Alzheimer's disease. *Biol Psychiatry*. 2014;75:542–52.
- Lake BB, Chen S, Sos BC, Fan J, Kaeser GE, Yung YC, et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol*. 2018;36:70–80.
- Corces MR, Shcherbina A, Kundu S, Gloudemans MJ, Frésard L, Granja JM, 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.
- Marbach D, Lamarter D, Quon G, Kellis M, Kutalik Z, Bergmann S. Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nat Methods*. 2016;13:366–70.
- Nott A, Holtman IR, Coufal NG, Schlachetzki JCM, Yu M, Hu R, et al. Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science*. 2019;366:1134–9.
- Sey NYA, Hu B, Mah W, Fauni H, McAfee JC, Rajarajan P, 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(4):583–93.
- Pratapa A, Jalihal AP, Law JN, Bharadwaj A, Murali TM. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat Methods*. 2020;17:147–54.
- Chan TE, Stumpf MPH, Babbie AC. Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst*. 2017;5:251–267.e3.
- Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. *PLoS One*. 2010;5(9):e12776.
- Moerman T, Aibar Santos S, Bravo González-Blas C, Simm J, Moreau Y, Aerts J, et al. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinform Oxf Engl*. 2019;35:2159–61.
- Li Y, Wu F-X, Ngom A. A review on machine learning principles for multi-view biological data integration. *Brief Bioinform*. 2018;19:325–40.
- Ting J, Ying M, Wang D. scGRNom (single-cell gene regulatory network prediction from multi-omics), <https://github.com/daifengwanglab/scGRNom>. Github; 2021.
- Harmston N, Ing-Simmons E, Perry M, Baresic A, Lenhard B. GenomicInteractions: R package for handling genomic interaction data [Internet]. 2020. Available from: <https://github.com/ComputationalRegulatoryGenomics/GenomicInteractions/>
- Tan G, Lenhard B. TFBSTools: an R/bioconductor package for transcription factor binding site analysis. *Bioinform Oxf Engl*. 2016;32:1555–6.
- Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2020;48:D87–92.
- Schep, Alicia. motifmatchr: fast motif matching in R [Internet]. 2019. Available from: <https://www.bioconductor.org/packages/release/bioc/html/motifmatchr.html>
- Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol*. 2013;9:e1003118.
- Coetzee SG, Coetzee GA, Hazelett DJ. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinform Oxf Engl*. 2015;31:3847–9.

33. Carlson M. TxDb.Hsapiens.UCSC.hg19.knownGene: annotation package for TxDb object(s) [Internet]: Bioconductor; 2015. Available from: <https://bioconductor.org/packages/release/data/annotation/html/TxDb.Hsapiens.UCSC.hg19.knownGene.html>
34. Stuart T, Butler A, Hoffman B, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive integration of single-cell data. *Cell*. 2019;177:1888–1902.e21.
35. van Dijk D, Sharma R, Nainys J, Yin K, Kathail P, Carr AJ, et al. Recovering gene interactions from single-cell data using data diffusion. *Cell*. 2018;174:716–729.e27.
36. Panagiotou OA, Ioannidis JPA. for the Genome-Wide Significance Project. What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. *Int J Epidemiol*. 2012;41:273–86.
37. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47:1228–35.
38. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet Lond Engl*. 2013;381:1371–9.
39. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet*. 2011;43:977–83.
40. van Rheenen W, Shatunov A, Dekker AM, McLaughlin RL, Diekstra FP, Pulit SL, et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat Genet*. 2016;48:1043–8.
41. Howard DM, Adams MJ, Clarke T-K, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci*. 2019;22:343–52.
42. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet*. 2018;50:912–9.
43. International Multiple Sclerosis Genetics Consortium, Wellcome Trust Case Control Consortium 2, Sawcer S, Hellenthal G, Pirinen M, Spencer CCA, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*. 2011;476:214–9.
44. Blauwendraat C, Heilbron K, Vallerga CL, Bandres-Ciga S, von Coelln R, Pihlström L, et al. Parkinson's disease age at onset genome-wide association study: Defining heritability, genetic loci, and α -synuclein mechanisms. *Mov Disord Off J Mov Disord Soc*. 2019;34:866–75.
45. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet*. 2019;51:63–75.
46. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*. 2016;533:539–42.
47. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*. 2012;44:981–90.
48. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491:119–24.
49. Schunkert H, Kke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic13 new susceptibility loci for coronary artery disease. *Nat Genet*. 2011;43:333–8.
50. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10:1523.
51. De Jager PL, Ma Y, McCabe C, Xu J, Vardarajan BN, Felsky D, et al. A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci Data*. 2018;5:180142.
52. King G, Zeng L. Logistic regression in rare events data. *Polit Anal*. 2001;9:137–63.
53. Holtman IR, Skola D, Glass CK. Transcriptional control of microglia phenotypes in health and disease. *J Clin Invest*. 2017;127:3220–9.
54. Zeisel A, Munoz-Manchado AB, Codeluppi S, Lonnerberg P, La Manno G, Jureus A, et al. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science*. 2015;347:1138–42.
55. Deneen B, Ho R, Lukaszewicz A, Hochstim CJ, Gronostajski RM, Anderson DJ. The transcription factor NFIA controls the onset of gliogenesis in the developing spinal cord. *Neuron*. 2006;52:953–68.
56. Zhou B, Osinski JM, Mateo JL, Martynoga B, Sim FJ, Campbell CE, et al. Loss of NFIX transcription factor biases postnatal neural stem/progenitor cells toward oligodendrogenesis. *Stem Cells Dev*. 2015;24:2114–26.
57. Pearson CA, Moore DM, Tucker HO, Dekker JD, Hu H, Miquelajl CE, et al. Loss of NFIREGULATES neural stem cell self-renewal and bias toward deep layer cortical fates. *Cell Rep*. 2020;30:1964–1981.e3.
58. Harrington AJ, Raissi A, Rajkovich K, Berto S, Kumar J, Molinaro G, et al. MEF2C regulates cortical inhibitory and excitatory synapses and behaviors relevant to neurodevelopmental disorders. *eLife*. 2016;5:e20059.
59. Han H, Cho J-W, Lee S, Yun A, Kim H, Bae D, et al. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res*. 2018;46:D380–6.
60. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res*. 2019;29:1363–75.
61. Liu Z-P, Wu C, Miao H, Wu H. RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database*. 2015;2015:bav095.
62. Baik SH, Kang S, Son SM, Mook-Jung I. Microglia contributes to plaque growth by cell death due to uptake of amyloid β in the brain of Alzheimer's disease mouse model. *Glia*. 2016;64:2274–90.
63. Kheiri G, Dolatshahi M, Rahmani F, Rezaei N. Role of p38/MAPKs in Alzheimer's growth by cell death due to uptake of amyloid β in the brain of A. *Rev Neurosci*. 2018;30:9–30.
64. Marín-Teva JL, Cuadros MA, Martín-Oliva D, Navascués J. Microglia and neuronal cell death. *Neuron Glia Biol*. 2011;7:25–40.
65. Barbier P, Zejneli O, Martinho M, Lasorsa A, Belle V, Smet-Nocca C, et al. Role of tau as a microtubule-associated protein: structural and functional aspects. *Front Aging Neurosci*. 2019;11:204.
66. McCutcheon RA, Krystal JH, Howes OD. Dopamine and glutamate in schizophrenia: biology, symptoms and treatment. *World Psychiatry Off J World Psychiatr Assoc WPA*. 2020;19:15–33.
67. Berdenis van Berlekom A, Muflihah CH, Snijders GJJ, MacGillavry HD, Middeldorp J, Hol EM, et al. Synapse pathology in schizophrenia: a meta-analysis of postsynaptic elements in postmortem brain studies. *Schizophr Bull*. 2020;46:374–86.
68. Osimo EF, Beck K, Reis Marques T, Howes OD. Synaptic loss in schizophrenia: a meta-analysis and systematic review of synaptic protein and mRNA measures. *Mol Psychiatry*. 2019;24:549–61.
69. McGuire JL, Depasquale EA, Funk AJ, O'Donovan SM, Hasselfeld K, Marwaha S, et al. Abnormalities of signal transduction networks in chronic schizophrenia. *NPJ Schizophr*. 2017;3:30.
70. Kehrre C, Maziashvili N, Dugladze T, Gloveli T. Altered excitatory-inhibitory balance in the NMDA-hypofunction model of schizophrenia. *Front Mol Neurosci*. 2008;1:6.
71. Ahmad R, Sportelli V, Ziller M, Spengler D, Hoffmann A. Tracing early neurodevelopment in schizophrenia with induced pluripotent stem cells. *Cells*. 2018;7(9):140.
72. Sasi M, Vignoli B, Canossa M, Blum R. Neurobiology of local and intercellular BDNF signaling. *Pflug Arch Eur J Physiol*. 2017;469:593–610.
73. Fabiani C, Antollini SS. Alzheimer's disease as a membrane disorder: spatial cross-talk among beta-amyloid peptides, nicotinic acetylcholine receptors and lipid rafts. *Front Cell Neurosci*. 2019;13:309.
74. O'Brien RJ, Wong PC. Amyloid precursor protein processing and Alzheimer's disease. *Annu Rev Neurosci*. 2011;34:185–204.
75. Mahoney ER, Dumitrescu L, Moore AM, Cambroner FE, De Jager PL, Koran MEI, et al. Brain expression of the vascular endothelial growth factor gene family in cognitive aging and Alzheimer's disease. *Mol Psychiatry*. 2021;26(3):888–96.
76. Misiak B, Stramecki F, Stańczykiewicz B, Frydecka D, Lubeiro A. Vascular endothelial growth factor in patients with schizophrenia: A systematic review and meta-analysis. *Prog Neuropsychopharmacol Biol Psychiatry*. 2018;86:24–9.
77. Glausier JR, Lewis DA. Dendritic spine pathology in schizophrenia. *Neuroscience*. 2013;251:90–107.
78. Cochran JN, Hall AM, Roberson ED. The dendritic hypothesis for Alzheimer's disease pathophysiology. *Brain Res Bull*. 2014;103:18–28.
79. Fakhoury M. Microglia and Astrocytes in Alzheimer's Disease: Implications for Therapy. *Curr Neuropharmacol*. 2018;16:508–18.
80. Raabe FJ, Slapakova L, Rossner MJ, Cantuti-Castelvetri L, Simons M, Falkai PG, et al. Oligodendrocytes as a new therapeutic target in schizophrenia:

- from histopathological findings to neuron-oligodendrocyte interaction. *Cells*. 2019;8(12):1496.
81. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45:580–5.
 82. Kim-Hellmuth S, Aguet F, Oliva M, Muñoz-Aguirre M, Kasela S, Wucher V, et al. Cell type-specific genetic regulation of gene expression across human tissues. *Science*. 2020;369:eaa28528.
 83. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8:1826.
 84. Nguyen ND, Wang D. Multiview learning for understanding functional multiomics. *PLoS Comput Biol*. United States. 2020;16:e1007677.
 85. Lacal I, Ventura R. Epigenetic inheritance: concepts, mechanisms and perspectives. *Front Mol Neurosci*. 2018;11:292.
 86. Linker SM, Urban L, Clark SJ, Chhatriwala M, Amatya S, McCarthy DJ, et al. Combined single-cell profiling of expression and DNA methylation reveals splicing regulation and heterogeneity. *Genome Biol*. 2019;20:30.
 87. Marx V. A dream of single-cell proteomics. *Nat Methods*. 2019;16:809–12.
 88. Brummelman J, Haftmann C, Núñez NG, Alvisi G, Mazza EMC, Becher B, et al. Development, application and computational analysis of high-dimensional fluorescent antibody panels for single-cell flow cytometry. *Nat Protoc*. 2019;14:1946–69.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

