Benchmarking methods integrating GWAS and single-cell transcriptomic data for mapping trait-cell type associations

Ang Li[1,2], Tian Lin[1], Alicia Walker[2], Xiao Tan[1,3], Ruolan Zhao[1], Shuyang Yao[4], Patrick F. Sullivan[4,5,6], Jens Hjerling-Leffler[5], Naomi R. Wray[1,2,7*], Jian Zeng[1,7*]

[1] Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD, Australia
[2] Department of Psychiatry, University of Oxford, Oxford, UK
[3] Infection and Inflammation Centre, Queensland Institute of Medical Research Berghofer, Australia
[4] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
[5] Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden
[6] Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, NC, USA
[7] These authors contributed equally to this work
* Correspondence: N.R.W. <naomi.wray@psych.ox.au.uk>, J.Z. <j.zeng@uq.edu.au>.

Abstract

Genome-wide association studies (GWAS) have discovered numerous trait-associated variants, but their biological context remains unclear. Integrating GWAS summary statistics with single-cell RNA-sequencing expression profiles can help identify the cell types in which these variants influence traits. Two main strategies have been developed to integrate these data types. The “single cell to GWAS” strategy (representing most methods) identifies gene sets with cell-type-specific expression and then follows with enrichment analyses applied to GWAS summary statistics. Conversely, the “GWAS to single cell” strategy begins with a list of trait-associated genes and calculates a cumulative disease score per cell based on gene expression count data. We systematically evaluated 19 approaches verses “ground truth” trait-cell type pairs to assess their statistical power and false positive rates. Based on these analyses, we draw seven key conclusions to guide future studies. We also propose a Cauchy approach to combine the two main strategies to maximize power for detecting trait-cell type associations.

Introduction

Genetic variants likely affect complex traits and diseases by acting at the level of specific cell types[1–5]. Identifying these cell types is essential for understanding disease etiology and for developing new preventative and therapeutic approaches for personalized medicine[6–8]. A powerful approach is to integrate two fundamental data modalities, summary statistics from GWAS, which capture SNP-trait associations, and gene transcript-by-cell matrices from single-cell or single-nucleus RNA sequencing (both abbreviated as scRNA-seq hereafter), which provide cell-type-specific gene expression profiles. In principle, a cell type is prioritized if GWAS signals are significantly associated with the gene expression specificity of the cell type. Methods that leverage these data to map cell types to complex traits, referred to as trait-cell type mapping methods, have gained increasing attention. As of Jan 2025, we identified over 15 computational methods[9–19] can integrate GWAS and scRNA-seq data for this purpose (**Supplementary Table 1**). However, which methods perform best and what factors affect performance have been difficult to assess.

Benchmark comparisons of these methods are challenging for several reasons. First, while GWAS datasets have grown rapidly in sample size over the past decade, statistical power varies widely across traits[20,21], which may lead to different methods performing optimally for different traits. Second, scRNA-seq datasets are increasingly available across different species, conditions (e.g., cases versus controls), longitudinal time points, organs, and tissue regions, presenting opportunities for discovery but also challenges in effectively integrating and analyzing multiple datasets[22]. Methods differ in how they identify gene sets associated with relative cell type specificity[23,24], and these gene sets can vary depending on experimental conditions, sequencing technology/depth, biological contexts, and data sources. Third, the causal relationships between cell types and traits remain largely unknown, making it difficult to compare methods using real data. Currently, no gold standard exists for evaluating the performance of trait-cell type mapping methods, and current studies have mainly relied on the increasing number of detected trait-cell type associations of new methods compared to existing methods, an approach that may be biased if false discovery is not appropriately controlled.

Here, we present a benchmarking study of trait-cell type mapping methods that integrate GWAS results and scRNA-seq data. We used 33 GWAS datasets representing a wide range of complex traits and diseases (**Supplementary Table 2**), and 10 scRNA-seq datasets (**Supplementary Table 3**) from both single organs[25–32] and multi-organ atlases in human[33] and mouse[34]. For a fair comparison, we identify a set of putatively critical and control trait-cell type pairs as “ground truth”, based on established biological knowledge and empirical evidence from prior studies (**Figure 1a)**. Moreover, we classify existing methods into two primary strategies based on their conceptual distinctions (**Figure 1b**). (1) “From single cell transcriptomics to GWAS” (SC-to-GWAS) identifies a set of

specifically expressed genes (SEGs) for each cell type and then tests for GWAS enrichment. There are many approaches to identify SEGs[13,16,35,36]. These gene sets are then evaluated by stratified LD score regression[10,11] (evaluates SNP-based heritability enrichment in the genomic regions annotated by SEGs) or by MAGMA gene-set enrichment analysis which tests for overrepresentation of trait-associated genes in SEGs (MAGMA-GSEA)[12,37]. (2) "From GWAS to single cell transcriptomics" (GWAS-to-SC) starts with genes identified from GWAS and then computes single-cell disease relevance scores based on cumulative scRNA-seq expression of disease-associated genes, with scDRS[9] being a representative method. We systematically compare both approaches and propose a Cauchy p-value combination[38,39] to integrate results across different strategies. We investigate how factors such as GWAS sample size, scRNA-seq dataset selection, and marker gene specificity affect method performance. By using a comprehensive evaluation framework, our study aims to establish best practices for integration of genetic and single-cell data to prioritize cell types for complex traits.

Results

Establishing putative true-positive and true-negative trait-cell type pairs for methods evaluation

All methods comparisons in this study are based on the benchmark trait-cell type pairs (see **Methods**). Briefly, across 8 trait categories, we selected 33 complex traits (**Supplementary Table 2**), each with $\geq$ 10 independent genome-wide significant loci (median = 108) and SNP-based heritability > 0.05 (median = 0.24). For each trait, we identified a putative true-positive cell type that was most likely to be associated with the trait and a putative true-negative cell type that was the least likely association. These determinations were based on PubMed-supported evidence[16] and biological knowledge, facilitating method comparisons in power and false positive rate (FPR) (**Figure 1a**).

Effective trait-cell type mapping requires more than identification of cell-type markers

Both mapping strategies rely on lists of gene sets. In the SC-to-GWAS strategy, the gene lists consist of SEGs derived from single-cell transcriptomic features, whereas in the GWAS-to-SC strategy, the key gene list contains trait-associated genes identified from genetic data. Several metrics have been developed to define SEG, and we incorporated all of them in this study (**Supplementary Table 4)**.

To assess the impact of SEG selection on the SC-to-GWAS strategy, we first examined the overlap between gene sets prioritized by 8 different metrics (**Supplementary Table 4)**. For each metric, the top 10% of genes (a common threshold used in the literature[13,40]) were considered as the cell-type-specific gene set. We applied these metrics to three distinct cell types (microglial cells, B cells, and pancreatic β-cells), and found that only about half of the genes identified by each metric were shared with another metric. This suggests that different metrics capture different properties of the expression distribution of all genes within a cell type and of each gene across cell types (**Supplementary Figure**

1). The SEG gene sets comprise ~1,400 genes, which contrasts to the small number of gold-standard marker genes used to provide cell-type labels (at least 5 marker genes per cell type, with a median of 5, **Supplementary Table 7**). Hence, we evaluated how well different SEG metrics prioritized these gold-standard marker genes across 28 cell types. We found that the differential expression T-statistic (DET) consistently ranked gold-standard marker genes highest, followed by sc-linker[14] and Cepo[36] (**Figure 2a, Supplementary Figure 2**). However, DET showed no advantage in mapping trait-cell type associations, whereas Cepo outperformed others in mapping power and FPR control, regardless of the enrichment method used to link the cell-type specific gene sets with GWAS results (sLDSC or MAGMA-GSEA) (**Figure 2b, Supplementary Figure 3**). These results suggest that the best metric for trait-cell type mapping does not simply identify cell-type markers more precisely.

Next, we explored the impact of SEG-derived genomic annotations when incorporated with sLDSC and MAGMA-GSEA methods under different settings of annotations **(Supplementary Tables 1,4**). The SEG methods yield continuous per-gene metrics that can be incorporated with GWAS results as continuous or binarized annotations (**Methods**) while conditioning on a set of functional genomic annotations included in the baseline sLDSC model (v1.1 contained 53 annotations[10] and v2.3 contains 111 annotations[41]). Our results indicate that continuous annotations with the minimal baseline produced the most robust results (highest power and lowest FPR on average across traits, **Supplementary Figures 3-5**), whereas binarized annotations can help control inflation of FPR in MAGMA-GSEA (**Supplementary Figures 3,4**).

The GWAS-to-SC strategy does not require SEG sets because the cumulative gene expression of trait-associated genes is calculated per cell which can be evaluated across cells of the same cell-type label. We found that using the gene-based test mBAT-combo[42] to identify a list of trait-associated genes from GWAS results gave slightly more robust results in the scDRS method than using MAGMA-GBT[37] genes, especially in FPR control (**Supplementary Figure 6**). This improvement is likely due to the superior ability of mBAT-combo to identify genes with masking effects caused by LD (trait-associated variants of opposite effect that are on the same haplotype).

Based on these results, in subsequent sections of main results we limit comparisons to the SC-to-GWAS methods of Cepo–›sLDSC and Cepo–›MAGMA-GSEA (i.e., Cepo as the metric to identify the cell-type-specific gene lists, followed by either sLDSC or MAGMA-GSEA using GWAS summary statistics). We also continue to present results for EP–›binary-sLDSC because this is the method most commonly implemented in published papers. For the GWAS-to-SC method, we take mBAT-combo–›scDRS forward into subsequent sections.

Consistent trait-cell type mapping between mouse and human scRNA-seq data

Mouse models are widely used in biomedical research because of their cost-effectiveness, ethical feasibility, and tractability in accessing all tissue types and exploring impact of treatment exposures, making them an accessible starting point for studying disease mechanisms before translation to human studies. However, findings from mice must be validated in human data to ensure biological relevance. Given the broader availability of mouse-based scRNA-seq data, we investigated whether the choice of species influences trait-cell type mapping performance.

To assess this, we focused on putative true-positive and true-negative trait-cell type pairs in common between a human atlas[33] and a mouse atlas[34] dataset generated from the same lab. As shown in **Figure 3a**, among the cell types represented in both atlases, the human atlas contains markedly more cells than the mouse atlas. Despite this difference, all methods demonstrated significant differentiation between true-positive and true-negative groups (t-test $P < 0.01$, **Figure 3b**). While some traits appeared to have stronger signals in MAGMA-GSEA and scDRS analyses for human data compared to mouse data, no significant differences were observed between the two datasets on average across traits (t-test $P > 0.05$, **Figure 3b, Supplementary Figure 6**). Interestingly, the hepatocyte data in humans tended to yield stronger association signals, whereas immune cells in mice consistently gave robust results across traits (**Figure 3c**). Overall, the mouse vs human origin of the scRNA-seq data had minimal impact on the performance of mapping methods, consistent with previous findings[24]. These results suggest that both human and mouse scRNA-seq data can be used effectively in trait-cell type mapping, supporting the feasibility of leveraging datasets from model species in initial analyses.

Atlas scRNA-seq data enhances trait-cell type mapping compared to single-organ data

Although scRNA-seq datasets are increasingly available, most have been collected from a single organ due to cost, complexity, and specific research interests. To assess whether using a single-organ dataset affects the conclusions about method performance, we repeated our analyses across datasets from different organs, while ensuring comparable pairwise evaluations to those based on atlas datasets. Notably, atlas-based analyses consistently yielded stronger association signals for true-positive trait-cell type pairs, regardless of the strategy used. Specifically, a higher proportion of true-positive pairs showed significant associations in atlas-based analyses compared to using single-organ datasets – of 27 common pairs among datasets, 24 were detected by Cepo–›sLDSC, 24 by Cepo–›MAGMA-GSEA, and 25 by mBAT-combo–›scDRS when using atlas datasets, in contrast to 19 were detected by Cepo–›sLDSC, 22 by Cepo–›MAGMA-GSEA, and 18 by mBAT-combo–›scDRS, respectively when using single-organ datasets (**Figure 4a**). Meanwhile, FPR were effectively controlled in the true-negative pairs (**Supplementary Figure 7**).

For the SC-to-GWAS strategy, we hypothesized that the improved performance of atlas-based over single-organ-based analyses is likely due to the increased number of cell types, which improves the

precision of cell-type specificity measurement. This specificity quantifies how the mean expression of each gene varies across cell types rather than within a single cell type. This means that as the number of cell types increases, the sampling variance would decrease, leading to improved precision of specificity score estimates. We performed simulations by random cell sampling and calculated the expression proportion (EP, the most commonly used cell-type specificity metric in genetic studies, **Supplementary Tables 1,4**), in each subsample of cells for six example genes. Across datasets (atlas, independent single-organ, or single-organ subset extracted from the atlas), we observed distinct distributions of the number of cell types with nonzero gene expression and the sampling distributions of cell-type specificity scores (**Supplementary Figure 8**). Compared to the atlas dataset, single-organ datasets produced higher mean specificity scores but substantially greater sampling variance (the mean-variance ratio in atlas was substantially higher, with a value of 200.0 in atlas, 99.9 in atlas-lung, and only 15.6 in lung dataset), indicating noisier estimates of cell-type specificity in single-organ datasets (result details are provided in **Supplementary Table 8**).

For the GWAS-to-SC strategy, we investigated two trait-cell type pairs: atrial fibrillation with atrial cardiomyocyte and celiac disease with CD4+ αβ T cells. These examples were chosen because atlas and single-organ datasets produced notably different results. In this scDRS method[9], disease scores are made per cell by summing the expression of the top 1,000 genes associated with the disease. These scores are compared to those generated from 1,000 sets of 1,000 control genes that matched to the disease genes in expression mean and variance across cells. Using potential hypothetical examples (**Supplementary Figure 9**), we showed that this matching process implicitly accounts for background cell types, leading to higher power in detecting differences between disease and control scores in the target cell type. Empirically, we observed that both raw disease and control scores (before standardization by control score mean and variance, as part of the scDRS method) for atrial fibrillation across cells were larger in variance in the atlas dataset than in the single-organ (heart) dataset (**Supplementary Figures 10a** and **11a**), likely due to the broader cell-type diversity in the atlas (174 cell types vs only 12 in the heart dataset). This case was consistent with our hypothetical example (**Supplementary Figure 9**). Notably, the difference between disease and control scores in the atrial cardiomyocytes was larger in the atlas than in the heart dataset (**Supplementary Figure 11b**), resulting in a strong association signal in scDRS using the atlas (P_{scDRS} = 1.52e-21 vs 0.81 using heart data, **Figure 4a**). While the atlas generally outperformed single-organ datasets, counterexamples existed. In the analysis of CD4+ αβ T cells and celiac disease, the lung dataset produced the strongest association (P_{scDRS} = 1.52e-07, 49 cell types, 13,485 CD4+ αβ T cells), outperforming the atlas and other organs where this cell type was also present. However, this case that both raw disease and control for celiac disease across cells in the lung dataset showed smaller variance than all other datasets (**Supplementary Figures 12** and **13**), coincided with our second hypothetical example (**Supplementary Figure 9**). It is important to note that the lung dataset itself is an atlas integrating 49

datasets with 2.4 million cells from 486 individuals. These results imply that neither high cell-type diversity, a large number of cells, nor genes with high expression (low variance) within a cell type alone guarantee statistical significance, but rather that all these factors contribute, highlighting the value of single-cell dataset integration for trait-cell type mapping.

Protein coding genes are sufficient for trait-cell type mapping

In addition to cell types, the selection of background genes may also affect the performance of trait-cell type mapping analyses. Most studies focus on protein-coding genes for their more direct functional relevance and high expression stability relative to other gene types (e.g., long noncoding (lnc) RNAs, short ncRNAs and microRNAs[43]) but recent studies suggested that non-coding genes contribute to complex trait variation[44,45]. Using both atlas and single-organ scRNA-seq datasets, we tested whether including all expressed genes or only protein-coding genes affects mapping performance. Despite variations in the number of genes expressed across organs, our results showed that restricting analyses to protein-coding genes (as used in our analyses so far) achieved similar or higher power in detecting true-positives (**Figure 4e, Supplementary Figures 14** and **15**), whereas including all expressed genes led to an inflation with binary annotations sLDSC (**Supplementary Figure 15a, b**). Using all expressed genes led to significant inflations in MAGMA-GSEA regardless metrics for cell-type specificity **(Supplementary Figure 15c, d**). In scDRS, similar results were observed between using all expressed genes and protein-coding genes **(Supplementary Figure 15e-h)**.

Impact of GWAS sample size and number of cells in scRNA-seq data

To understand power constraints in discovery of trait-cell type associations, we evaluated the impact of GWAS sample size and the number of cells in scRNA-seq datasets on trait-cell type mapping. Here, we focused on height, type 2 diabetes (T2D), and coronary artery disease (CAD), for which GWAS with different samples sizes are available. As expected, increasing the effective GWAS sample sizes led to stronger signals of association for all methods in detecting true-positive pairs (**Figure 5a**), while effectively controlling for FPR. Consistently, atlas-based analyses demonstrated higher power than single-organ-based analyses, with a higher rate of power gain (larger slope) as the GWAS sample size increased. Of note, increasing the GWAS sample size had a limited effect on the power for CAD when using scRNA-seq data from heart.

To investigate the impact of scRNA-seq cell numbers, we systematically down-sampled cells within corresponding cell types for both single-organ and atlas datasets while keeping the GWAS sample size at its maximum. Our results showed that as the number of cells in the cell type increased, power also improved (**Figure 5b**). Given a large GWAS sample size, a minimum number of 20 cells per cell type after quality control (QC) in atlas datasets was sufficient to identify the true-positive trait-cell

type pairs, such as height-chondrocytes and T2D-pancreatic β cells. However, despite an increase in the number of cells, power eventually plateaued. Moreover, the higher the GWAS power, the sooner this plateau was reached. For instance, in height-chondrocytes (SNP-based heritability, $h^2_{SNP} = 0.54$, number of independent GWAS loci N_{loci} = 4,114), the power plateaued at around 80–100 cells, whereas for T2D-pancreatic β cells ($h^2_{SNP} = 0.20$, N_{loci} = 143), power plateaued at 150 cells. When the number of chondrocytes reached its maximum in the data, reducing the cell number in other cell types (vascular smooth muscles and pancreatic B cells) led to a decrease in power even though the numbers of the relevant cell type remained the same, demonstrating the importance of comparator cell types in the dataset. Similarly, when the GWAS sample size was limited, increasing the number of cells improved association detection, as seen in CAD-smooth muscle cells ($h^2_{SNP} = 0.06$, N_{loci} = 243) from the heart (**Figure 5b**). With the same number of cells for all cell types, the change of the number of cells in irrelevant cell types have limited impact on the detection of target trait-cell type pair associations (**Supplementary Figure 16**). In conclusion, a large effective GWAS sample size, combined with a minimal number of QC-passed cells (n = 20) across diverse cell types in the dataset, plays a critical role in the performance of the trait-cell type mapping analyses. The majority of cell types—82% in the human atlas and 81% in the mouse atlas, defined as those with ≥100 cells—have already reached the power plateau corresponding to height GWAS. This highlights that enhancing GWAS power is now the critical factor for improvement, as most cell types in TS and TMS have achieved sufficient cell numbers. Achieving GWAS power comparable to height GWAS would yield the greatest benefits.

Cauchy combination across strategies improves trait-cell type mapping

We found that no single method could successfully detect all predefined true-positive pairs (**Figure 4a**). To improve power, we applied a Cauchy combination approach[38,46] to combine P-values of the optimal methods of both SC-to-GWAS (P-values from Cepo-continuous-LDSC and Cepo-binary-MAGMA-GSEA) and GWAS-to-SC strategies (P-values from mBAT-combo-scDRS). Results show that this combined approach successfully detected all true-positive pairs and appropriately controlled inflation in true-negative pairs using atlas dataset (**Figure 6a, b**). Similar results were observed in the single organ-based analyses (**Figure 6c, d**). Still, the overall power (average -log10(P)) in the atlas dataset was larger than single organ-based datasets across all methods, consistent with the results in **Figure 4**.

Discussion

In this study, we evaluated methods for identification of trait-relevant cell types under the hypothesis that trait-associated genes will show gene expression specificity in trait-relevant cell types. To benchmark methods, we assessed power of detection through 33 true-positive trait-cell type pairs and

assessed FPR through 33 true-negative trait-cell type pairs (**Figure 1b**). Based on analyses across a variety of settings, we draw the following conclusions: 1) the method Cepo[36] is the superior method for defining cell-type specificity genes used to identify trait-cell type associations in the SC-to-GWAS approach, even though the differential expression test statistics, performed best in ranking gold-standard marker genes used to label cells to cell types (**Figure 2**). 2) Performance is comparable when using transcriptomic data derived from either mouse or human origins (**Figure 3**). 3) Single-cell atlases of multiple tissues or single-organ datasets with high cellular diversity provide greater power than single-organ datasets with limited cell-type representation (**Figure 4a**). 4) Focusing on protein-coding genes performs equivalently or slightly better than using all expressed genes in the dataset (**Figure 4b**). 5) The power of these methods is positively correlated with GWAS sample size and the number of cells per cell type, where power of GWAS plays a more important role than the sample size of scRNA-seq datasets (**Figure 5**). The primary factors influencing performance include a sufficient GWAS sample size (i.e., power plateaus after this sufficient sample size), the diversity of cell types in scRNA-seq, and a sufficient number of cells per cell type (which can be as low as 20 if the GWAS sample size is large). 6) Cauchy p-value combinations of the best performing SC-to-GWAS and GWAS-to-SC maximizes power while avoiding false positives (**Figure 6**). Since each method is well controlled for false-positive discovery but exploits different properties of SC gene expression data, the Cauchy combination gains power for discovery of true positive associations while still controlling false discovery. 7) Our benchmark pipeline (SC-GWAS:GWAS-SC) provide reproducible and extensible evaluation for future trait-cell type mapping methods development.

We demonstrated that the optimal method for identifying cell type marker genes differs from the optimal method for identifying genes important for disease risk. Marker gene detection focuses on discriminative power, i.e., finding genes that best differentiate one cell type from another, which has been well summarized and benchmarked elsewhere[24]. In contrast, risk genes likely reflect the functional importance rather than cell type identity. A gene can be a strong marker but have no direct role in disease, and conversely, a gene driving disease risk may be expressed across multiple cell types, making it a poor marker. Though the SC-to-GWAS and GWAS-to-SC strategies rely on distinct gene features, they had comparable power in identifying associations between cell types and traits, consistent with the original findings when each method was first proposed. However, neither strategy detected all putatively causal pairs in our benchmark, suggesting the two strategies may be complementary to each other and the combination of both could be beneficial. In line with the conclusion of the CELLECT study[13], our results found that combining a set of four SEG metrics, averaged as ESμ had improved performance in detecting true-positive pairs over performance of each method across traits. While most of the metrics focus on detecting the mean differences in gene expression between cell types, Cepo first filters genes to have both high average level and low variance of expression amongst cells of the same cell type relative to other cell types. When we

incorporated Cepo in the cell type identification methods, it achieved comparable and more stable performance to ESμ in both sLDSC and MAGMA-GSEA frameworks.

Despite providing a comprehensive assessment of state-of-the-art methods for the first time, our studies have limitations. First, our analyses were conducted at the level of broad cell types rather than fine-grained subtypes, which may overlook the capabilities of different methods in capturing trait-associations at the cell subclass level. It would be valuable for future studies to assess whether our conclusions hold at deeper levels of cellular hierarchy. Second, our "ground truth" selection is certainly not perfect and is limited by the availability of traits with sufficient GWAS power and cell type availability in scRNA-seq datasets. This selection of trait-cell type pairs could impact our conclusion that single cell data from mouse is equally good as human in identifying trait cell-type associations. Nonetheless, the robustness of results using single cell data from model species is encouraging given that some environmental exposures relevant for diseases are much more easily explored using non-human species. Third, we only focused on samples of European ancestry and whether our findings generalized to other ancestries warrants further research considering the rapidly increasing number of scRNA seq datasets. Given that common variant associations are shared across ancestries[47] and given the robustness of results across species, we do not expect our results to be impacted by human ancestry. Fourth, differences between scRNA-seq and snRNA-seq are not considered. snRNA-seq may miss important gene expression signals[12], such as microglial activations shown by previous research[48]. However, our study was limited to available data, our technical limitations mean that only snRNA-seq data can be generated for some organs, such as brain. Fifth, part of the differences between single-organ and atlas datasets may be due to technical aspects (data generated by different labs) which is out of our control. A completely fair comparison is difficult, although we did subset atlas data to make single organ data sets. Sixth, we focused only on the methods using scRNA-seq and GWAS data with common SNPs. Methods that utilized additional information, such as EPIC[17] with rare variants, ECLIPSER[49] incorporating QTLs, and the optional program in sc-linker[14] that integrating tissue enhancers, were not considered. Although extra functional annotations are generally expected to enhance power[2,50,51], our investigation did not consider such an effect.

Despite these limitations, our research provides valuable insights for method developers and researchers focusing on specific diseases or data applications, offering guidance and recommendations for future investigations in trait-cell type mapping.

Acknowledgements

We gratefully acknowledge DPhil Malte D. Luecken, A/Prof. Nathan Palphant and Dr. Sonia Shah for consultation on true-positive pairs selection in relative fields. This research was funded by the Australian National Health and Medical Research Council (1173790, 1177268), the National Institute for Mental Health (1RO1MH124871-01), and the University of Oxford Michael Davys Trust. J.Z. acknowledges support from Australian Research Council grant FL180100072 (awarded to P.M. Visscher).

Contributions

Conceptualization: A.L., J.Z., N.R.W., Methodology: A.L., J.Z., N.R.W., Formal analysis and investigation: A.L., GWAS data QC and format, SNP-heritability on liability scale estimates: T.L., Independent number of SNPs estimates: R.Zh., Cell ontology terms match: A.L., R.Zh., Software support: A.W., X.T., Consultation: S.Y, P.F.S, J.H.L, Writing, original draft: A.L., J.Z., N.R.W. Writing, review and editing: A.L., J.Z., N.R.W., P.F.S. Visualization, A.L., J.Z., N.R.W. Supervision: J.Z., N.R.W.

Competing interests

The authors declare that they have no competing interests.

Data and materials availability

The publicly available data used in this study are available through the references provided in this manuscript or requested through their link. Pipeline of SC-GWAS:GWAS-SC are deposited in GitHub, available through https://github.com/Share-AL-work/SC-GWAS_GWAS-SC.

References:

1. Morris, J. A. *et al.* Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens. *Science (1979)* **380**, (2023).
2. Yazar, S. *et al.* Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* **376**, (2022).
3. Werder, R. B. *et al.* CRISPR interference interrogation of COPD GWAS genes reveals the functional significance of desmoplakin in iPSC-derived alveolar epithelial cells. *Sci Adv* **8**, (2022).
4. Onuora, S. Single-cell RNA sequencing sheds light on cell-type specific gene expression in immune cells. *Nat Rev Rheumatol* **18**, 363 (2022).
5. Cuomo, A. S. E., Nathan, A., Raychaudhuri, S., MacArthur, D. G. & Powell, J. E. Single-cell genomics meets human genetics. *Nature Reviews Genetics 2023 24:8* **24**, 535–549 (2023).
6. Heath, J. R., Ribas, A. & Mischel, P. S. Single-cell analysis tools for drug discovery and development. *Nature Reviews Drug Discovery 2015 15:3* **15**, 204–216 (2015).
7. Mosquera, J. V. *et al.* Integrative single-cell meta-analysis reveals disease-relevant vascular cell states and markers in human atherosclerosis. *Cell Rep* **42**, (2023).
8. Masuda, T., Sankowski, R., Staszewski, O. & Prinz, M. Microglia Heterogeneity in the Single-Cell Era. *Cell Rep* **30**, 1271–1281 (2020).
9. Zhang, M. J. *et al.* Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *Nature Genetics 2022 54:10* **54**, 1572–1580 (2022).
10. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics 2015 47:11* **47**, 1228–1235 (2015).
11. Gazal, S. *et al.* Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. *Nature Genetics 2017 49:10* **49**, 1421–1427 (2017).
12. Skene, N. G. *et al.* Genetic identification of brain cell types underlying schizophrenia. *Nature Genetics 2018 50:6* **50**, 825–833 (2018).
13. Timshel, P. N., Thompson, J. J. & Pers, T. H. Genetic mapping of etiologic brain cell types for obesity. *Elife* **9**, 1–45 (2020).
14. Jagadeesh, K. A. *et al.* Identifying disease-critical cell types and cellular processes by integrating single-cell RNA-sequencing and human genetics. *Nature Genetics 2022*

54:10 **54**, 1479–1492 (2022).

15. Ma, Y. *et al.* Polygenic regression uncovers trait-relevant cellular contexts through pathway activation transformation of single-cell RNA sequencing data. *Cell Genomics* **3**, 100383 (2023).
16. Shang, L., Smith, J. A. & Zhou, X. Leveraging gene co-expression patterns to infer trait-relevant tissues in genome-wide association studies. *PLoS Genet* **16**, e1008734 (2020).
17. Wang, R., Lin, D. Y. & Jiang, Y. EPIC: Inferring relevant cell types for complex traits by integrating genome-wide association studies and single-cell RNA sequencing. *PLoS Genet* **18**, e1010251 (2022).
18. Calderon, D. *et al.* Inferring Relevant Cell Types for Complex Traits by Using Single-Cell Gene Expression. *Am J Hum Genet* **101**, 686–699 (2017).
19. Watanabe, K., Taskesen, E., Van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications 2017 8:1* **8**, 1–11 (2017).
20. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* **101**, 5 (2017).
21. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nature Genetics 2019 51:9* **51**, 1339–1348 (2019).
22. Luecken, M. D. *et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nature Methods 2021 19:1* **19**, 41–50 (2021).
23. Kim, M. C. *et al.* Method of moments framework for differential expression analysis of single-cell RNA sequencing data. *Cell* **187**, 6393-6410.e16 (2024).
24. Pullin, J. M. & McCarthy, D. J. A comparison of marker gene selection methods for single-cell RNA sequencing data. *Genome Biol* **25**, 1–37 (2024).
25. Jardine, L. *et al.* Blood and immune development in human fetal bone marrow and Down syndrome. *Nature 2021 598:7880* **598**, 327–331 (2021).
26. Hocker, J. D. *et al.* Cardiac cell type-specific gene regulatory programs and disease risk association. *Sci Adv* **7**, (2021).
27. Sikkema, L. *et al.* An integrated cell atlas of the lung in health and disease. *Nature Medicine 2023 29:6* **29**, 1563–1577 (2023).
28. Kamath, T. *et al.* Single-cell genomic profiling of human dopamine neurons identifies a population that selectively degenerates in Parkinson's disease. *Nature Neuroscience 2022 25:5* **25**, 588–595 (2022).

29. Cheng, J. B. *et al.* Transcriptional Programming of Normal and Inflamed Human Epidermis at Single-Cell Resolution. (2018) doi:10.1016/j.celrep.2018.09.006.
30. Smillie, C. S. *et al.* Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell* **178**, 714-730.e22 (2019).
31. Fasolino, M. *et al.* Single-cell multi-omics analysis of human pancreatic islets reveals novel cellular states in type 1 diabetes. *Nature Metabolism 2022 4:2* **4**, 284–299 (2022).
32. Guilliams, M. *et al.* Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell* **185**, 379-396.e38 (2022).
33. Jones, R. C. *et al.* The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, (2022).
34. Schaum, N. *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature 2018 562:7727* **562**, 367–372 (2018).
35. Skene, N. G. & Grant, S. G. N. Identification of vulnerable cell types in major brain disorders using single cell transcriptomes and expression weighted cell type enrichment. *Front Neurosci* **10**, 179460 (2016).
36. Kim, H. J. *et al.* Uncovering cell identity through differential stability with Cepo. *Nature Computational Science 2021 1:12* **1**, 784–790 (2021).
37. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput Biol* **11**, e1004219 (2015).
38. Liu, Y. *et al.* ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in Sequencing Studies. *Am J Hum Genet* **104**, (2019).
39. Liu, Y. & Xie, J. Cauchy Combination Test: A Powerful Test With Analytic p-Value Calculation Under Arbitrary Dependency Structures. *J Am Stat Assoc* **115**, 393–402 (2020).
40. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet* **50**, 621 (2018).
41. Kim, A. *et al.* Inferring causal cell types of human diseases and risk variants from candidate regulatory elements. *medRxiv* 2024.05.17.24307556 (2024) doi:10.1101/2024.05.17.24307556.
42. Li, A. *et al.* mBAT-combo: A more powerful test to detect gene-trait associations from GWAS data. *The American Journal of Human Genetics* **110**, 30–43 (2023).
43. Ward, L. D. & Kellis, M. Interpreting noncoding genetic variation in complex traits and human disease. *Nature Biotechnology 2012 30:11* **30**, 1095–1106 (2012).

44. Mattick, J. S. *et al*. Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nature Reviews Molecular Cell Biology 2023 24:6* **24**, 430–447 (2023).

45. French, J. D. & Edwards, S. L. The Role of Noncoding Variants in Heritable Disease. *Trends in Genetics* **36**, (2020).

46. Liu, Y. & Xie, J. Cauchy Combination Test: A Powerful Test With Analytic p-Value Calculation Under Arbitrary Dependency Structures. *J Am Stat Assoc* **115**, 393–402 (2020).

47. Hu, S. *et al.* Fine-scale population structure and widespread conservation of genetic effect sizes between human groups across traits. *Nature Genetics 2025 57:2* **57**, 379–389 (2025).

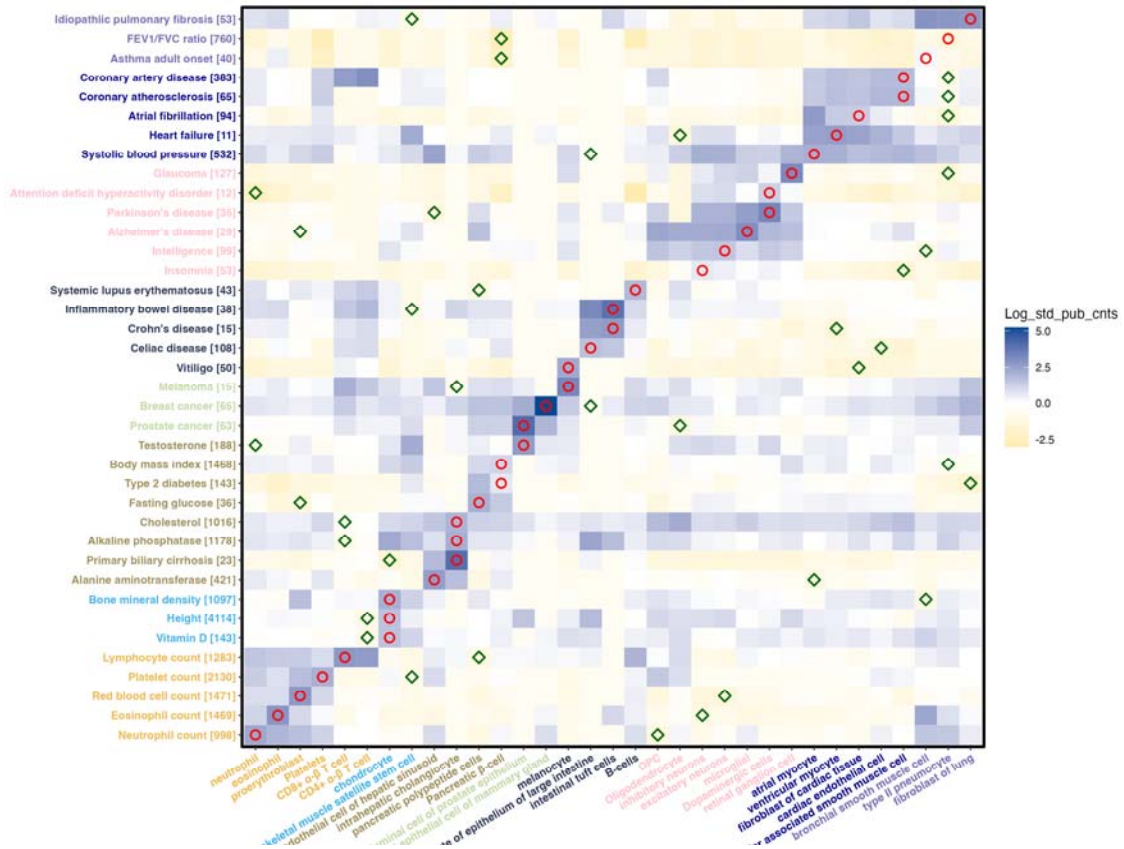
48. Thrupp, N. *et al.* Single-Nucleus RNA-Seq Is Not Suitable for Detection of Microglial Activation Genes in Humans. *Cell Rep* **32**, 108189 (2020).

49. Rouhana, J. M. *et al.* ECLIPSER: identifying causal cell types and genes for complex traits through single cell enrichment of e/sQTL-mapped genes in GWAS loci. *bioRxiv* 2021.11.24.469720 (2021) doi:10.1101/2021.11.24.469720.

50. Li, Y. E. *et al.* A comparative atlas of single-cell chromatin accessibility in the human brain. *Science* **382**, eadf7044 (2023).

51. Farbehi, N. *et al*. Integrating population genetics, stem cell biology and cellular genomics to study complex human diseases. *Nature Genetics 2024 56:5* **56**, 758–766 (2024).

a.


Idiopathic pulmonary fibrosis [53]
FEV1/FVC ratio [760]
Asthma adult onset [40]
Coronary artery disease [383]
Coronary atherosclerosis [65]
Atrial fibrillation [94]
Heart failure [11]
Systolic blood pressure [532]
Glaucoma [127]
Attention deficit hyperactivity disorder [12]
Parkinson's disease [35]
Alzheimer's disease [29]
Intelligence [99]
Insomnia [53]
Systemic lupus erythematosus [43]
Inflammatory bowel disease [38]
Crohn's disease [15]
Celiac disease [108]
Vitiligo [50]
Melanoma [15]
Breast cancer [66]
Prostate cancer [63]
Testosterone [188]
Body mass index [1468]
Type 2 diabetes [143]
Fasting glucose [36]
Cholesterol [1016]
Alkaline phosphatase [1178]
Primary biliary cirrhosis [23]
Alanine aminotransferase [421]
Bone mineral density [1097]
Height [4114]
Vitamin D [143]
Lymphocyte count [1283]
Platelet count [2130]
Red blood cell count [1471]
Eosinophil count [1469]
Neutrophil count [998]
neutrophil
eosinophil
proerythroblast
Platelets
CD8+ α-β T cell
CD4+ α-β T cell
chondrocyte
satellite stem cell
pancreatic polypeptide cells
pancreatic β-cell
melanocyte
intestinal tuft cells
B-cells
OPC
Oligodendrocyte
atrial myocyte
ventricular myocyte
fibroblast of cardiac tissue
cardiac endothelial cell
associated smooth muscle cell
bronchial smooth muscle cell
type II pneumocyte
fibroblast of lung
Log_std_pub_cnts
5.0
2.5
0.0
-2.5

b.

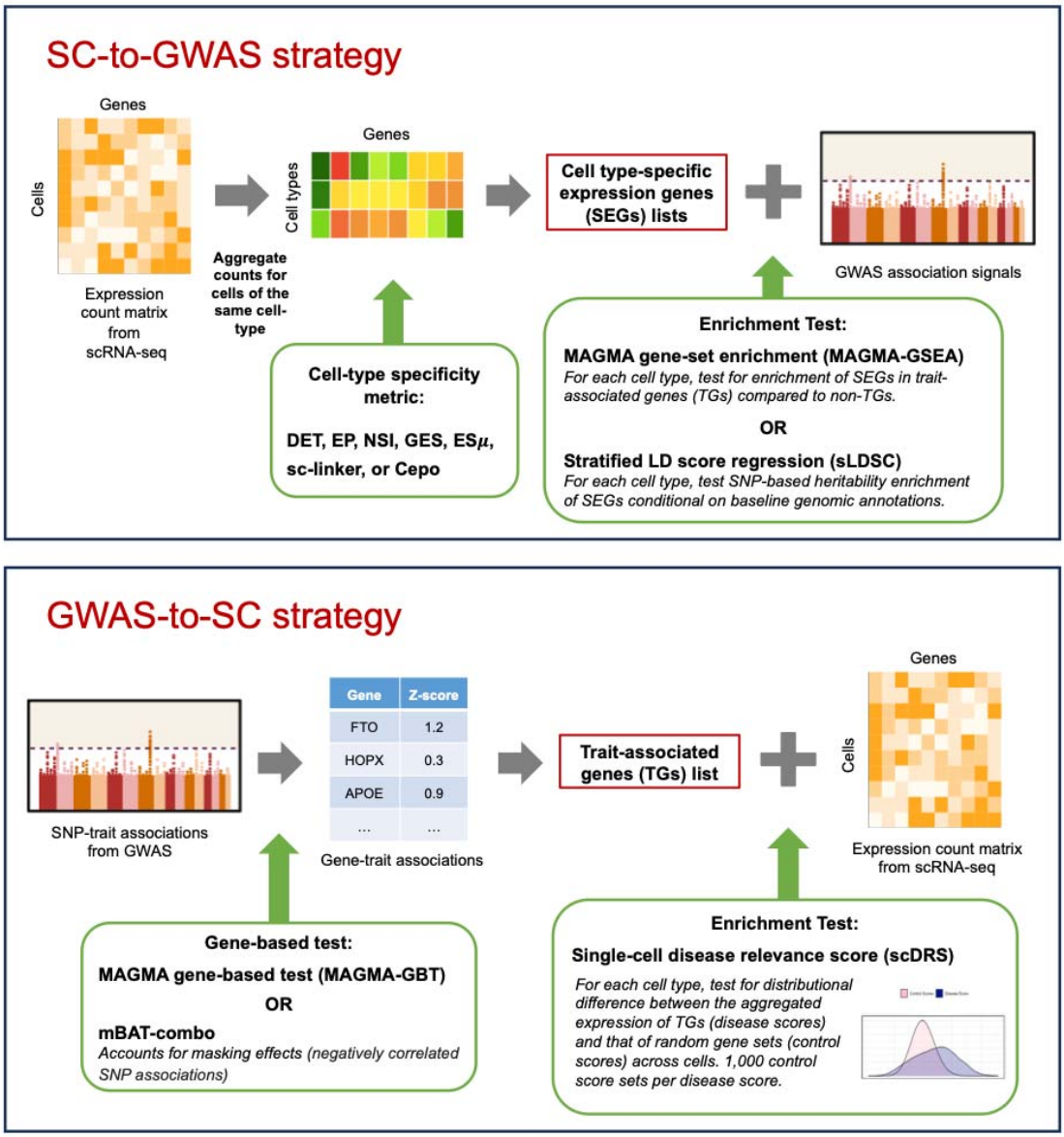


Figure.1. Study design overview. **a. Cell type – trait pairs as ground truth combinations.** One putatively critical cell type (true-positive, red circle) and one control cell type (true-negative, green diamond) were selected for each trait based on expert knowledge and publication records. These selected trait-cell type pairs were used as the "ground truth" for method evaluations and comparisons. X-axis denotes cell types ordered by tissue indicated by colour. Y-axis denotes traits or diseases, with the number of independent GWAS loci in parentheses; traits are ordered and coloured corresponding to their relevant tissues. Colour within matrix entries shows the value of scaled publication count for each trait-cell type pair. **b. Schematic of the approach to compare methods**. Summary of input data types and data integration.

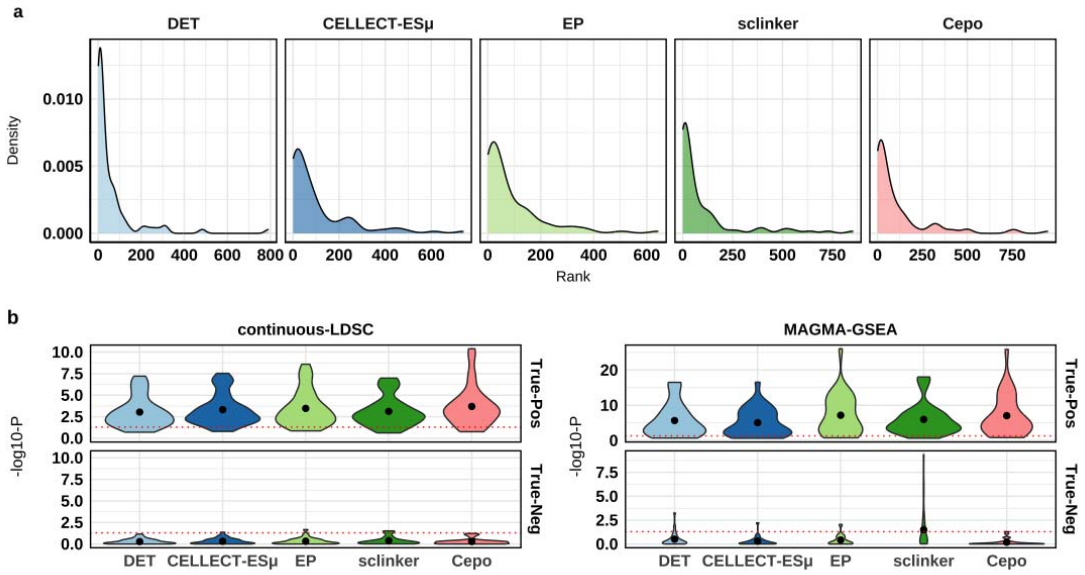


Figure 2. Comparisons of different SC-to-GWAS cell type specificity metrics. a. Rank of gold-standard cell type marker genes identified by different SEG metrics. X-axis represents the rank (amongst all genes) of cell type marker genes ranked by the specificity metrics (panels). The ranks are for a total of 125 unique marker genes across 28 cell-types. Y-axis represents the density of rank values based on metrics across all cell types. **b. Violin plots representing dentification of true positive and true-negative cell type -trait pairs when integrated with GWAS results through sLDSC SNP-based heritability enrichment (left) and MAGMA-gene-set enrichment (right).** The sLDSC baseline-LD model is used for annotation with different cell type specificity metrics**.** X-axis represents different metrics. Y-axis shows -$\log_{10}$(*P*-value) for trait-cell type association using different metrics. Red dash line denotes the significance threshold value 0.05. Upper panels represent results of true-positive pairs. Bottom panels represent results of true-negative pairs. Colours represent different metrics (see Supplementary Table 1). True-Pos and True-Neg represent results from the putatively critical and control trait-cell type pairs (Figure 1b) respectively and equate to power and false discovery of the tests (-log10 p-values above the red-dashed line).

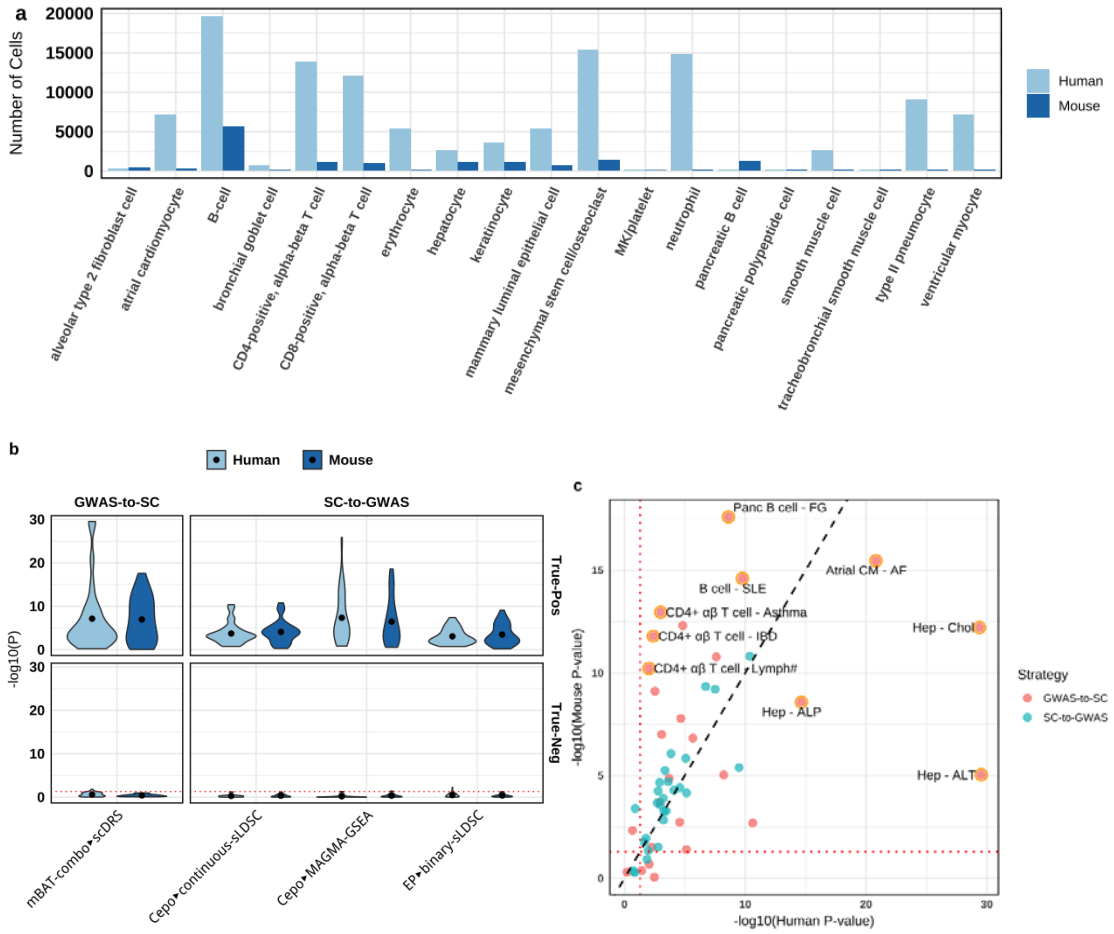
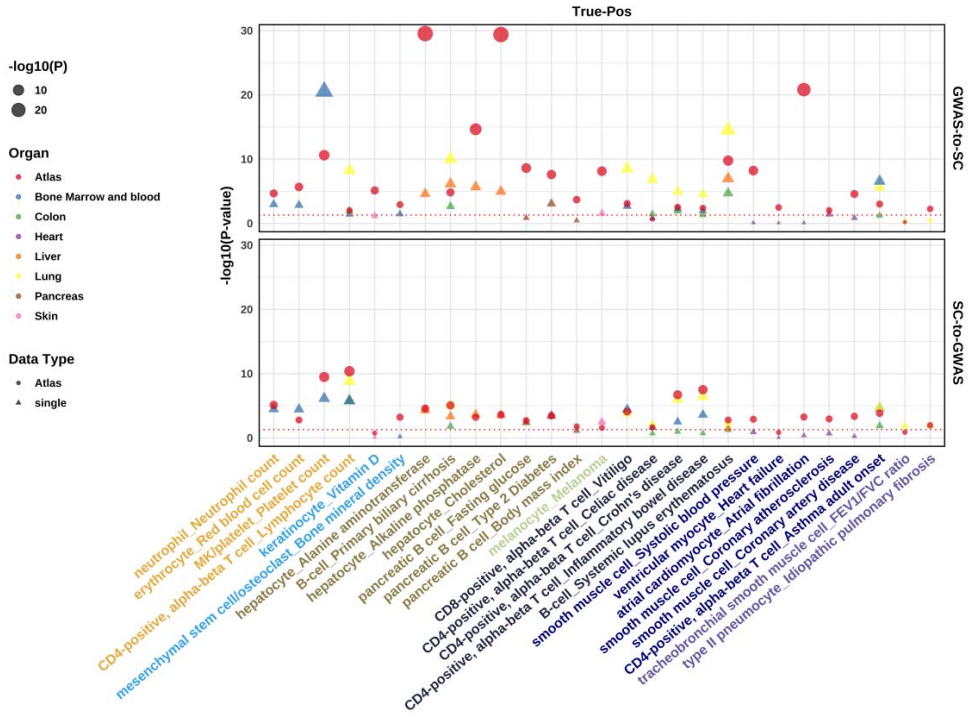


Figure 3. Comparison between human and mouse scRNA-seq datasets. a. Cell counts per cell type across cell-types in our selected positive and negative cell types for complex traits. X-axis denotes different cell types. Y-axis is the number of cells within each cell type. Bars are color-coded to distinguish species. **b. Comparison of identifying true positive and true negative cell type and trait pairs.** X-axis represents metrics selected from prior comparisons. Y-axis represents different metric-based -log10(P-value); red dash line denotes the significance threshold value -log10(0.05). Upper panel represent results of causal pairs group. Bottom panel represent results of control pairs group. True-Neg and True -Pos represent results from the control and putatively causal trait-cell type pairs (Figure 1b) respectively and equate to false discovery and power of the tests (-log10 p-values above the red-dashed line). **c. Comparison on detection of putatively causal cell type and trait pairs between species.** X-axis represents -log10(P-values) of results with human datasets. Y-axis represents -log10(P-values) of results with mouse datasets. Red dotted line denotes the significance threshold value -log10(0.05). Black dash line shows y=x. Outliners are highlighted with orange circle. Abbreviations for the cell types: "CD4+ αβ T cell", "CD4-positive, alpha-beta T cell"; "Panc B cell", "pancreatic B cell"; "Hep", "hepatocyte"; "Atrial CM", "atrial cardiomyocyte". Abbreviations for phenotype (trait): "SLE", "Systemic lupus erythematosus"; "FG", "Fasting glucose"; "Lymph#",

"Lymphocyte count"; "IBD", "Inflammatory bowel disease"; "ALT", "Alanine aminotransferase"; "ALP", "Alkaline phosphatase"; "Chol", "Cholesterol"; "AF", "Atrial fibrillation".

a



b

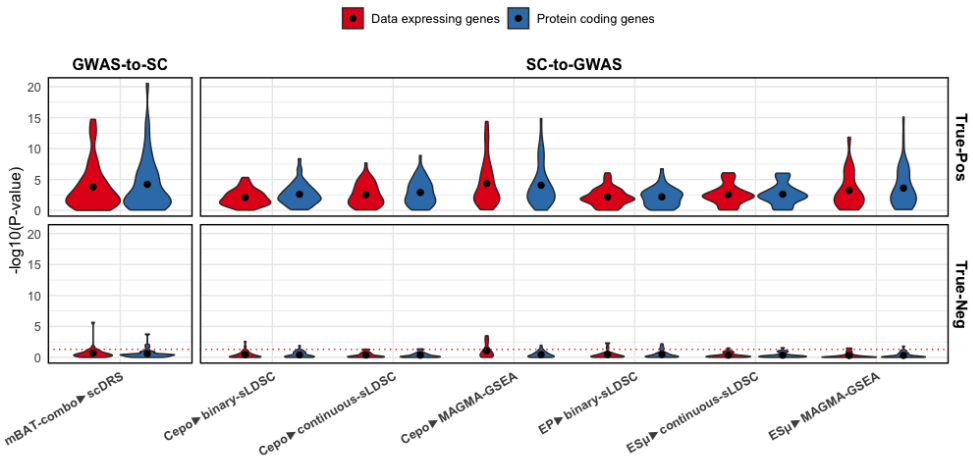


Figure 4. Impact of using single- and multi-organ (i.e., atlas) scRNA-seq datasets. a. Significance of tests identifying true-positive cell type–trait pairs comparing SC-to-GWAS and GWAS-to-SC frameworks with single-organ and atlas scRNA-seq datasets. X-axis notes the cell-type and trait pairs defined. Y-axis represents P-value from each method; red dash line denotes the significance threshold value 0.05. Colour represents the organ composition of the scRNA-seq datasets type being single organ (circle) or atlas (triangle). Point size corresponds to the -log10(P-value) of the methods' results. The figure including control cell type pairs is presented in Supplementary Figure 7. Based on previous results presented the SC-to-GWAS method is Cepo based binary-MAGMA-GSEA, and the GWAS-to-SC method is mBAT-combo based scDRS with by statistical P. **b. Impact of using data**

comprising all genes vs. protein coding genes in the single-organ based scRNA-seq datasets. X-axis notes different methods. mBAT-combo based scDRS with by MC-P by default. Y-axis represents the number of trait-cell type pairs being detected significantly in each method(top). Y-axis represents the P-values of trait-cell type pairs being detected in each method (bottom).

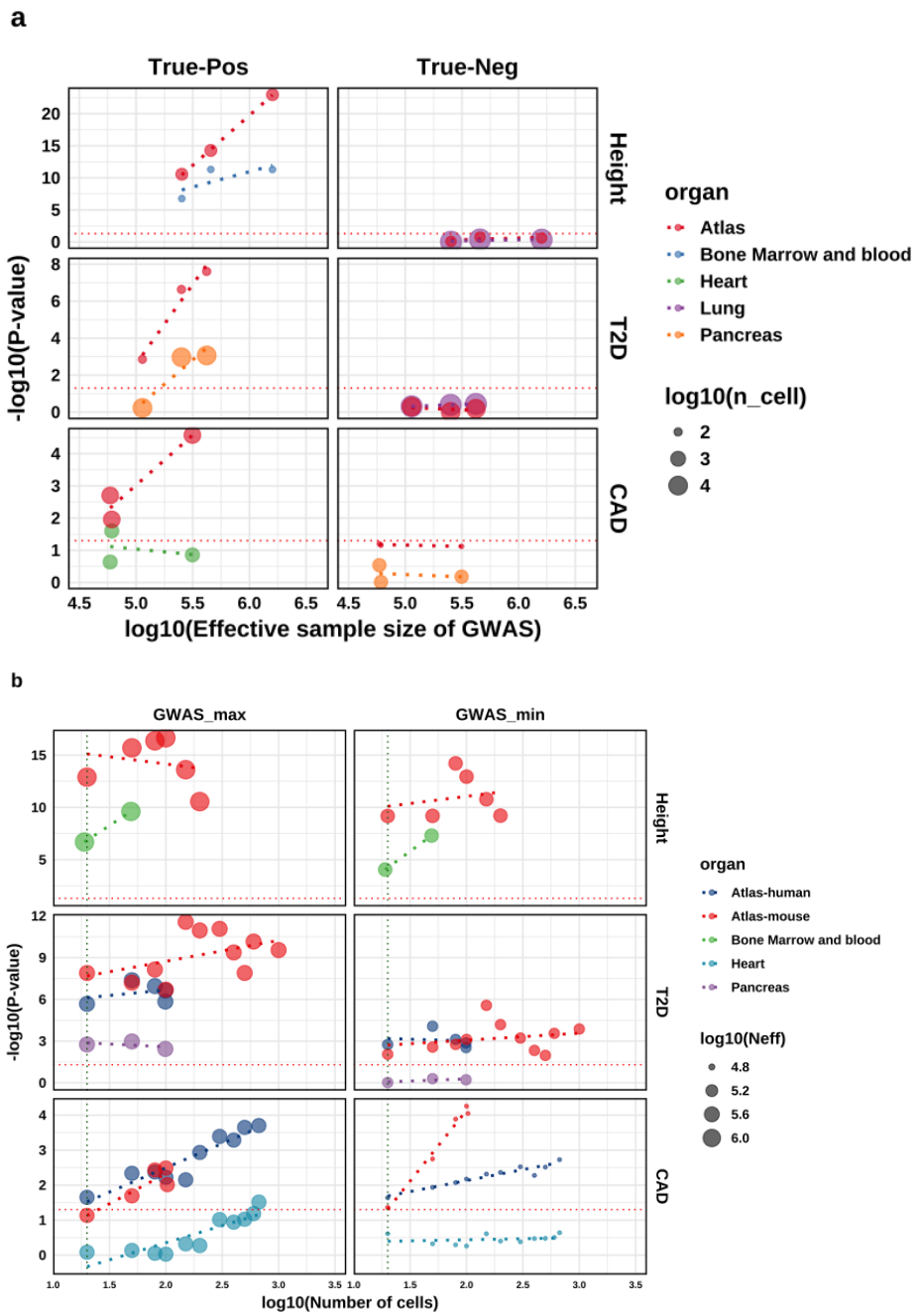


Figure 5. Impact of sample size of GWAS and scRNA-seq data. a. Performance comparison of scDRS (GWAS-to-SC) in identifying true negative and true positive cell type-trait pairs across different GWAS sample sizes for the same traits (Height, T2D, and CAD), measured as -log10(P-values). The X-axis represents the effective GWAS sample size, while the Y-axis shows the -log10(P-value) of the trait-cell type associations. Colour indicates organ composition, and point size reflects the number of cells within the detected cell type. **b. Performance comparison of scDRS in identifying true positive cell type-trait pairs using the largest and smallest GWAS sample size for the same traits while down sampling the number of cells within the same cell types.** The X-

axis represents the number of cells per cell type, and the Y-axis shows the -log10(P-value) of the trait-cell type associations. Colour denotes organ composition, and point size represents the effective GWAS sample size. Red dash line denotes the significance threshold value -log10(0.05). Green vertical dash line denotes the number of cells = 20.

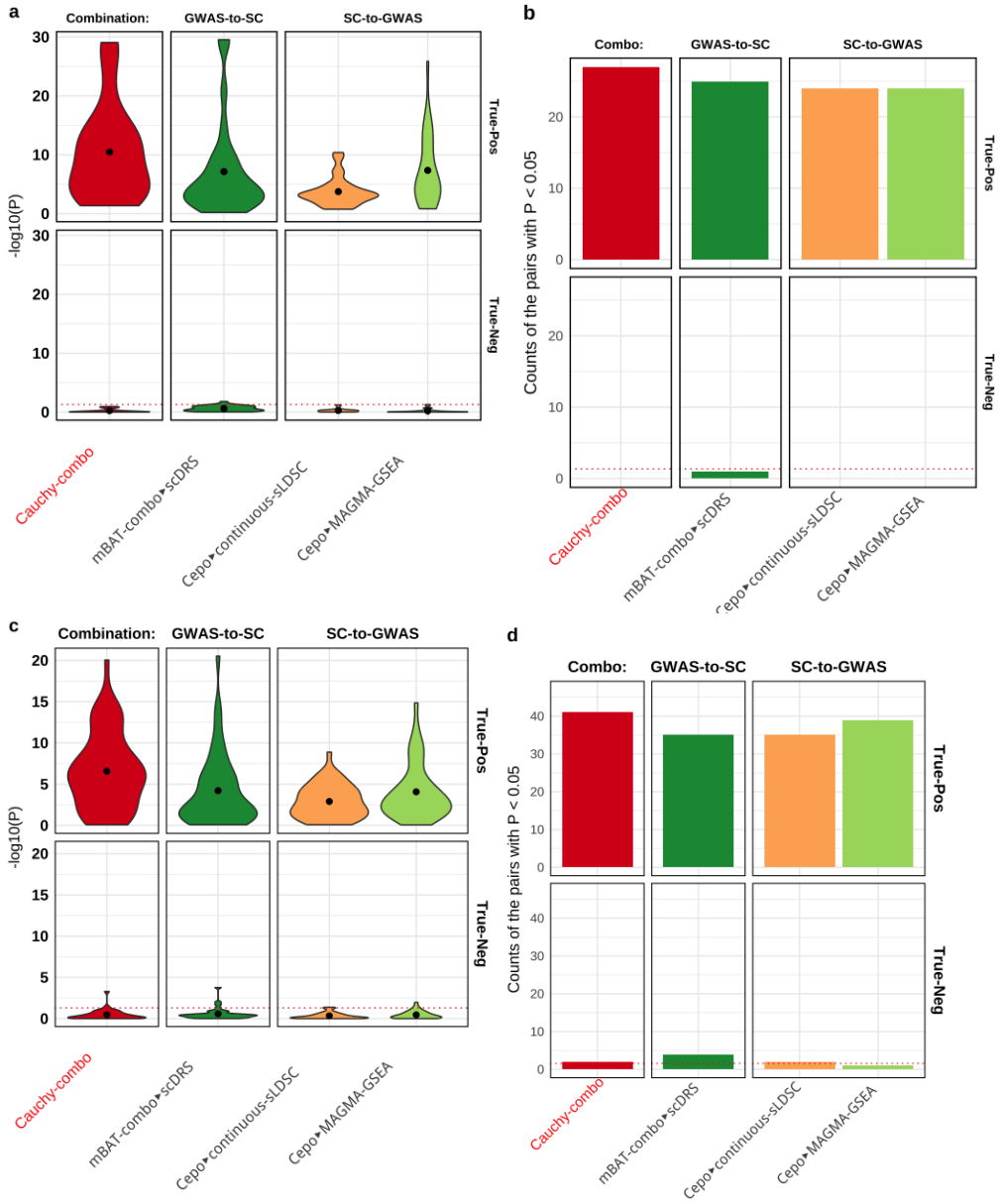


Figure 6. Combination of SC-to-GWAS and GWAS-to-SC under the best setting of each strategy. a. Performance comparison of different strategies in identifying true positive and true negative cell type-trait pairs across our defined benchmark pairs in atlas datasets-based results, measured as -log10(P-values). The X-axis represents different methods while the Y-axis shows the -log10(P-value) of the trait-cell type associations. Red dash line denotes the significance threshold value 0.05. **b. Performance comparison of different strategies in rue positive and true negative cell type-trait pairs across our defined benchmark pairs in atlas datasets-based results, measured as number of pairs being significantly detected.** Red dash line denotes the significance threshold value 28*0.05**. c. Performance comparison of different strategies in identifying rue positive and true negative cell type-trait pairs across our defined benchmark pairs in single**

organ datasets-based results, measured as -log10(P-values). The X-axis represents different methods while the Y-axis shows the -log10(P-value) of the trait-cell type associations. Red dash line denotes the significance threshold value 0.05. **d. Performance comparison of different strategies in identifying rue positive and true negative cell type-trait pairs across our defined benchmark pairs in single organ datasets-based results, measured as number of pairs being significantly detected.** Red dash line denotes the significance threshold value 33*0.05.

Online Methods

scRNA-seq data pre-processing

We used 10 databases of scRNA-seq data (**Supplementary Table 3**). Considering the high level of conservation with one-to-one orthologous relationships for protein-coding genes between mice and humans, we focused on 19,430 such genes with 19,116 expressed in the human (TS) atlas data and 14,879 expressed in the mouse (TMS) atlas data for the analysis. We standardized the cell-by-gene matrix by reversing prior log normalization to ensure dataset consistency, followed by standard UMI-based normalization to adjust for sequencing depth variations. The final expression values are represented as log2(TPM + 1), where TPM is transcripts per million, TPM. Cell annotations were initially based on existing labels in the dataset provided by the group generating and releasing the data. We then manually standardized these labels to ensure identical naming for the same cell types, correcting any inconsistencies in the original annotations (**Supplementary Table 4**). For cross-species analysis, mouse genes were mapped to their orthologous human counterparts using Ensembl (v. 91), with only one-to-one mapping orthologs retained.

GWAS data QC

We performed PLINK clumping to calculate the number of independent loci (--clump-p1 5e-8 --clump-r2 0.1 --clump-kb 500) and SBayesRC (GCTB_v2.5.2) to estimate the SNP-based heritability (on liability scale) for a range of complex traits and diseases. We selected 35 complex traits and diseases for which 38 GWAS summary data with a large sample size (n >13,000) are publicly available and those with a polygenic architecture (number of independent GWAS loci > 10 and SNP-based heritability > 0.05). The trait information can be found in **Supplementary Table 5**.

Establishing the ground truth

To achieve a fair comparison, we set out to find a set of putatively critical cell type and trait/disease pairs as the ground truth based on expert knowledge and empirical evidence from prior studies (**Supplementary Table 6**). Our selection procedure for the putatively critical cell type-trait pairs was as following.

1. For each trait, select up to three most relevant cell types based on biological knowledge and which can be found in the Chan Zuckerberg CELL by GENE Discover (CZ CELLxGENE Discover, https://cellxgene.cziscience.com/), such as candidate critical trait-cell type pairs.
2. Search for the number of publications in PubMed for each pair of trait and candidate cell type.
3. Establish a matrix (traits by cell types) of the PubMed counts for all trait-candidate cell type combinations.
4. Scale the PubMed counts per cell type by first applying a log transformation and then standardization of each matrix column. The scaling step is expected to produce more robust

evidence for trait-cell type pair by accounting for the differences between cell types in the availability of data and the degree of research activity.

5. Select one critical cell type for each trait as the true positive (True-Pos) that has the top scaled PubMed count across cell types or the 2nd/3rd top for a consideration of cell type diversity in the putatively positive set.
6. Select one irrelevant cell type for each trait as the true negative (True-Neg) that has the minimum scaled PubMed count across cell types or the second minimum for a consideration of cell type diversity in the control group.

Generally, the putatively True-Pos pairs had high scaled PubMed publication counts, whereas the controls pairs had low counts. For some traits, like blood cell count and FEV1/FVC ratio, the relevant cell types were readily apparent, but for metabolite traits and physical measures, there are likely to be several cell types relevant to the trait. Nevertheless, we selected one putatively critical cell type and one control cell type per trait as the benchmark for the subsequent analysis of method comparison. We caution that the publication counts are not unbiased, especially when the causal relationship in certain pairs has been well established. In such cases, there may be limited number of investigations available on PubMed. For example, bronchial smooth muscle cells play a causal role in determining the FEV1/FVC ratio by directly influencing airway calibre and reactivity, thereby modulating airflow obstruction in pulmonary diseases. Similarly, the red blood cell count is directly influenced by the production and maturation of proerythroblasts, which are the earliest precursors in the erythroid lineage committed to developing into mature red blood cells.

To evaluate the ability of different metrics to rank golden cell type marker genes, we examined at least five gold-standard marker genes for each of the 28 cell types included in our benchmark (**Supplementary Table** 7). We ensured that all selected cell types were present in both our atlas dataset and single-organ datasets. The gold-standard marker genes were chosen based on the data resource papers, manually curated genes, and ChatGPT latest survey.

Identifying associations between traits/diseases and cell types

Calculation of cell-type specificity metrics

The SC-to-GWAS strategy usually includes two steps. The first step is to identify the cell type specifically/differentially expressed genes or cell identity genes. Here, we used SEG (specifically expressed genes) to represent these genes for brevity in the following descriptions. In this strategy, predefined cell type labels are required for the statistical analysis. A cell type label is usually based on very high gene expression of a small number of genes. Various statistical methods are available to calculate the cell type specificity score **Supplementary Table 4**. In each method genes in the top decile (i.e., 1,400 genes) are considered as SEGs (e.g., TDEP, top decile expression proportion) and are used to create either a binary genomic annotation (e.g. LDSC-SEG and MAGMA covariate analysis, also refers to MAGMA-gene-set-enrichment analysis or MAGMA-GSEA) or a continuous genomic annotation (e.g. sc-linker, CELLECT and MAGMA gene property analysis) for each cell type. We ran the following cell-type specificity metric methods using their default parameter settings: EWCE for EP, Cepo, CELLEX, and sc-linker. All methods followed the same pre-processing steps for scRNA-seq data, as described above. Furthermore, CELLEX filtered out sporadically expressed genes using a one-way ANOVA with cell type annotations as the grouping factor. EP and CELLEX do not have the requirement for the minimal number of cells within each cell type. In contrast, scDRS includes cell types with more than 3 cells, sc-linker includes those with more than 10 cells and Cepo includes those with more than 20 cells. Although different metrics have their own criteria for the minimal number of cells within the cell type, we excluded cell types with fewer than 20 cells in our analysis.

GWAS and scRNA-seq integrative analysis

The GWAS-to-SC strategy requires the results of gene-based test using GWAS data as input for the integrative analysis, such as scDRS. We compared the performance of scDRS using MAGMA (as used in the scDRS publication) and mBAT-combo (a new method which is better powered for gene associations when haplotypes have risk alleles with opposite effect) as the gene-based test methods, respectively. For a fair comparison, we used the transcription start and end site +/- 10kb window (suggested by scDRS default settings) to define a gene region when conducting gene-based tests using MAGMA (v1.09) and mBAT-combo. The 1000 Genomes Project phase 3 European genotypes were used as the LD reference data.

For the SC-to-GWAS strategy, we used sLDSC and MAGMA gene-set enrichment methods. sLDSC tests whether the 10% most specifically expressed genes in a given cell type (define from specificity metrics) are enriched in SNP-based heritability for the trait. Only genes with at least 1 TPM or 1 UMI per million in the tested cell type were used for this analysis. To capture potential regulatory elements contributing to the effect of the region on the trait, we extended the gene coordinates by 100 kb upstream and downstream, respectively, for each gene (suggested by all methods based on the sLDSC framework). SNPs located in 100 kb regions surrounding the 10% most specific genes in each cell type were annotated with either binary or continuous values. This cell type-specific genomic annotation was analysed in sLDSC for one cell type at a time, jointly with 53 functional genomic annotations included in the Baseline model. The coefficient z-score P value was used as a measure of the association of the cell type with the trait. The significance threshold was set to be 5% false discovery rate (FDR) across all tissues/cell types and traits within each dataset. We also evaluated the cell-type specificity gene sets in the MAGMA-GSEA approach. In this approach, we examined the association between MAGMA gene-level trait associate statistics and different SEG metrics with either binary or continuous values using MAGMA-GSEA test. We adjusted the gene-level z-score obtained from MAGMA for default covariates, including gene size, gene density (a measure of within-gene LD), inverse mean minor allele count, as well as the logarithm of these variables. Subsequently, we performed a linear regression model, with the MAGMA gene-level z-score as the dependent variable and SEG metrics for each cell type as the independent variable. The outcome of interest was the significance of the contribution of cell type SEG to the regression coefficient of gene-level z-scores of trait association, assessed through a one-sided test (whether the gene-level z-score is positively associated with the SEG metric). In summary, this analysis allowed us to calculate cell type prioritization P-value, indicating the positive impact of SEG on the trait's gene-level z-statistics, as determined by the linear regression model.

Simulations for evaluating the impact of cell type diversities in the scRNA-seq datasets

To evaluate cell type specificity across different datasets (atlas and single organ), we focused on six genes and six cell types that were consistently present in all datasets (Supplementary Table 8). To ensure sufficient statistical power, we selected genes that were expressed in at least 1,000 cells in each dataset. Also, we ensured that the gene was substantially expressed (checked with minimum total expression threshold as 10 counts per cell type) and broadly expressed (checked with minimum number of cells threshold as 20). From this set, we identified the gene with the highest specificity in the dataset for further analysis. For specificity estimation, we performed 100 independent replications of random subsampling, selecting 1,000 cells per replication from each dataset. Gene expression values were extracted and normalized to 1 million molecules per cell type, enabling direct comparisons across cell types. Specificity was then calculated as the proportion of total expression assigned to each cell type. The aggregated results across replications allowed us to robustly characterize gene expression patterns and compare specificity across datasets.

Simulations for evaluating the impact of cell counts per cell type in scRNA-seq

To assess the impact of sample size of the two data types, we subsampled five predefined cell types—chondrocytes, vascular-associated smooth muscle cells, pancreatic beta cells, pancreatic polypeptide

cells, and type II pneumocytes—to either match the corresponding cell counts in another dataset, or to a fixed target count of 50 or 20, while retaining all other cell types. Cells were randomly selected with an equal probability. Three down-sampled datasets were generated, corresponding to these target cell counts, and saved for further analysis using scDRS, Cepo-sLDSC, and MAGMA-GSEA methods. Additionally, we performed the same analysis using different GWAS datasets for the same trait but with varying effective sample sizes to examine the impact of GWAS sample size on the results.