scDRS

Zhang & Hou et al.

Supplementary Information

Supplementary Note

Estimation of gene-specific technical noise variance of gene expression

We estimate the technical variance similar to previous works[1,2]. Specifically, we first compute the mean expression and expression variance for each gene in the original non-log-transformed space. Next, we fit a non-linear tread to the log10-scale variance/mean relationship using local polynomial regression (span=0.3, degree=2). The estimated trend models the expected technical variance based on the mean expression; observed variance values above this expected trend reflect biological variance. Given this trend, the proportion of technical variance is computed as the ratio of predicted technical variance to observed variance (in the original non-log-transformed space). Note that it is possible for this ratio to be greater than 1 if the observed variance is less than the expected variance. Finally, the technical variance of the log-transformed data $\sigma^2_{\text{tech},g}$ is computed as the product of the variance of the gene (computed using the log-transformed data) and the proportion of technical variance as estimated above.

Distribution of raw disease scores and raw control scores

Here we characterize the distribution of the raw disease scores and raw control scores under a simplified model. Assume the expression levels $\{\mathbf{X}_c \in \mathbb{R}^{n_{\text{gene}}}\}_{c=1}^{n_{\text{cell}}}$ are independently and identically distributed (i.i.d.) across cells with mean $\boldsymbol{\mu} \in \mathbb{R}^{n_{\text{gene}}}$ and covariance $\Sigma \in \mathbb{R}^{n_{\text{gene}} \times n_{\text{gene}}}$,

$$\mathbf{X}_c \overset{\text{i.i.d.}}{\sim} (\boldsymbol{\mu}, \Sigma). \tag{1}$$

Without loss of generality, also assume the gene weights are the same. Then the raw disease score and raw control scores (for a given control gene set G_b^{ctrl}) can be written as

$$s_c = \frac{1}{|G|} \sum_{g \in G} X_{cg}, \quad s_{cb}^{\text{ctrl}} = \frac{1}{|G_b^{\text{ctrl}}|} \sum_{g \in G_b^{\text{ctrl}}} X_{cg}, \tag{2}$$

where $|G| = |G_b^{\text{ctrl}}|$. We next compare the distributions of s_c and s_{cb}^{ctrl}. Since the control genes match the mean expression and expression variance of the disease genes, we have

$$\sum_{g \in G} \mu_g = \sum_{g \in G_b^{\text{ctrl}}} \mu_g, \quad \sum_{g \in G} \Sigma_{gg} = \sum_{g \in G_b^{\text{ctrl}}} \Sigma_{gg}. \tag{3}$$

The first equation gives that $\mathbb{E}[s_c] = \mathbb{E}[s_{cb}^{\text{ctrl}}]$.

For covariance, since the disease genes are more positively correlated due to co-expression in the associated cell population, the covariance matrix of the disease genes has larger off-diagonal elements than that of the control genes. Since the sum of the diagonal elements are equal between the two matrices due to the second part of Eq. (3), we have

$$\sum_{g,g' \in G} \Sigma_{g,g'} > \sum_{g,g' \in G_b^{\text{ctrl}}} \Sigma_{g,g'} \tag{4}$$

Since the variances of the raw disease score and raw control score can be written as

$$\text{Var}[s_c] = \frac{1}{|G|^2} \sum_{g,g' \in G} \Sigma_{g,g'}, \quad \text{Var}[s_{cb}^{\text{ctrl}}] = \frac{1}{|G_b^{\text{ctrl}}|^2} \sum_{g,g' \in G_b^{\text{ctrl}}} \Sigma_{g,g'}, \tag{5}$$

we have $\text{Var}[s_c] > \text{Var}[s_{cb}^{\text{ctrl}}]$. In summary,

$$\mathbb{E}[s_c] = \mathbb{E}[s_{cb}^{\text{ctrl}}], \quad \text{Var}[s_c] > \text{Var}[s_{cb}^{\text{ctrl}}]. \tag{6}$$

Since the raw disease scores have the same mean and a higher variance than the raw control scores, the top values of the raw disease scores are larger than all raw control scores. These top values correspond to the disease-associated cells, which have higher expression of the disease genes. Therefore, the disease-associated cells have larger raw disease scores than all raw control scores.

Normalization of disease scores and control scores

First gene set alignment. This step aims to correct for the mismatch that the control genes do not have exactly the same mean expression and expression variance as the disease genes. The variance level is estimated based on a heuristic that assumes the genes in the given gene set are independent. Specifically, if $Y = \sum_{i=1}^n w_i X_i$ for weights $w_1, \cdots, w_n \in \mathbb{R}$ and independent random variables $X_1, \cdots, X_n$, then $\text{Var}[Y] = \sum_{i=1}^n w_i^2 \text{Var}[X_i]$.

Cell-wise standardization. Consider the conditional distribution of the raw control score s_{cb}^{ctrl} given the expression matrix $\mathbf{X}$, where the randomness only comes from MC sampling of the control gene set G_b^{ctrl} but not from the data generation process of $\mathbf{X}$. Again without loss of generality, assume all genes have the same mean and variance, and the gene weights are the same. Then the raw control score can be written as

$$s_{cb}^{\text{ctrl}} = \frac{1}{|G_b^{\text{ctrl}}|} \sum_{g \in G_b^{\text{ctrl}}} X_{cg}. \tag{7}$$

Since all genes have the same mean and variance, G_b^{ctrl} corresponds to randomly and uniformly sampling $|G|$ genes from the set of all genes $\{1, \cdots, n_{\text{gene}}\}$ without replacement. When $|G| \ll n_{\text{gene}}$, sampling without replacement can be well approximated by sampling with replacement ($|G| < 20\% n_{\text{gene}}$ is usually a good heurestic[3]). In other words,

$$s_{cb}^{\text{ctrl}} \overset{d}{\approx} \frac{1}{|G|} \sum_{g=1}^{|G|} Y_{cg}, \tag{8}$$

where Y_{cg}'s are i.i.d. random variables uniformly sampled from $\{X_{c1}, \cdots, X_{cn_{\text{gene}}}\}$ (meaning sampling with replacement) and $\overset{d}{\approx}$ means approximately equal in distribution. Furthermore, when $|G|$ is large (e.g., >50), by the central limit theorem, being an average of $|G|$ i.i.d. random variables, s_{cb}^{ctrl} is close to a normal distribution, which depends only on its first two moments. Therefore, it is sufficient to only match the mean and variance of the control score distributions of different cells.

As a remark, the raw control score distribution considered here is defined with respect to randomly sampling control gene sets for a given cell, where the expression matrix $\mathbf{X}$ is fixed. It is different from the distribution of scores computed from a given gene set S (disease gene set or control gene set) across cells, where the randomness can be viewed as coming from the data generation process of the expression matrix $\mathbf{X}$. In this latter case, since the gene expression levels are correlated, the score of a given cell $s = \frac{1}{|S|} \sum_{g \in S} X_{cg}$ can not be viewed as an average of i.i.d. random variables (because X_{cg}'s may be dependant) and can be very different from the normal distribution especially when the gene set size $|S|$ is small.

Second gene set alignment. This step aims to correct for the differences of the mean values of scores from different gene sets introduced by cell-wise standardization. The differences are relatively small (Supp. Fig. 1F). Hence, this step is less important. For this reason and also because it is hard to find a good heuristic for estimating the variance levels of scores from different gene sets without down-weighting the disease scores due to the higher correlation between disease genes, we do not correct for the difference of the variance of scores from different gene sets.

Related works

Previous methods for identifying disease-critical tissues and cell types. Many types of data that assay gene regulation have been integrated with GWAS data to identify disease-relevant tissues and cell types, including chromatin and histone modifications[4–13] and gene expression measurements[14–23]. Studies using gene expression data have generally either used tissue-level data derived from DNA microarrays / bulk RNA-Seq[14–17,20], or focused on predefined cell types (usually classical cell types based on known marker genes) in scRNA-seq data by aggregating cells from the same cell type[18,21,22]; these cell type annotations may be hard to obtain especially for less well-studied cell populations or subtle cell states within a cell type. One exception is Jagadeesh and Dey et al.[23], who associated intra- and inter-cell type cellular processes identified in scRNA-seq data to disease.

Previous methods for scoring individual cells. Previous works analyzing scRNA-seq data alone have used individual cell-level scores to characterize cellular heterogeneity and subtle cell states within classically defined cell types[2,24–29], where the cell scores were computed based on the expression of a predefined set of genes such as cell cycle signature genes or biological pathways. These works did not integrate GWAS data. In addition, they generally did not provide individual cell-level p-values for associating individual cells to the gene set (VAM[2] is the only exception, but we show that VAM suffers severely inflated type I error; Fig. 2A). We further note that two studies have associated individual cells in scATAC-seq data to disease[12,13]. However, scATAC-seq and scRNA-seq data have different data structures and require different treatments (e.g., matching nucleotide GC content and fragment accessibility for scATAC-seq data[12] vs. matching gene expression mean and variance in our paper). To our knowledge, no previous study has associated individual cells in scRNA-seq data to disease.

References

1. Tim Stuart, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M Mauck III, Yuhan Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. Comprehensive integration of single-cell data. *Cell*, 177(7):1888–1902, 2019.

2. Hildreth Robert Frost. Variance-adjusted mahalanobis (vam): a fast and accurate method for cell-specific gene set scoring. *Nucleic acids research*, 48(16):e94–e94, 2020.

3. Rémi Bardenet and Odalric-Ambrym Maillard. Concentration inequalities for sampling without replacement. *Bernoulli*, 21(3):1361–1385, 2015.

4. Jason Ernst, Pouya Kheradpour, Tarjei S Mikkelsen, Noam Shoresh, Lucas D Ward, Charles B Epstein, Xiaolan Zhang, Li Wang, Robbyn Issner, Michael Coyne, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345):43–49, 2011.

5. Matthew T Maurano, Richard Humbert, Eric Rynes, Robert E Thurman, Eric Haugen, Hao Wang, Alex P Reynolds, Richard Sandstrom, Hongzhu Qu, Jennifer Brody, et al. Systematic localization of common disease-associated variation in regulatory dna. *Science*, 337(6099):1190–1195, 2012.

6. Gosia Trynka, Cynthia Sandor, Buhm Han, Han Xu, Barbara E Stranger, X Shirley Liu, and Soumya Raychaudhuri. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nature genetics*, 45(2):124–130, 2013.

7. Joseph K Pickrell. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *The American Journal of Human Genetics*, 94(4):559–573, 2014.

8. Gleb Kichaev, Wen-Yun Yang, Sara Lindstrom, Farhad Hormozdiari, Eleazar Eskin, Alkes L Price, Peter Kraft, and Bogdan Pasaniuc. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS genetics*, 10(10):e1004722, 2014.

9. Alexander Gusev, S Hong Lee, Gosia Trynka, Hilary Finucane, Bjarni J Vilhjálmsson, Han Xu, Chongzhi Zang, Stephan Ripke, Brendan Bulik-Sullivan, Eli Stahl, et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *The American Journal of Human Genetics*, 95(5):535–552, 2014.

10. Kyle Kai-How Farh, Alexander Marson, Jiang Zhu, Markus Kleinewietfeld, William J Housley, Samantha Beik, Noam Shoresh, Holly Whitton, Russell JH Ryan, Alexander A Shishkin, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, 518(7539):337–343, 2015.

11. Yue Li and Manolis Kellis. Joint bayesian inference of risk variants and tissue-specific epigenomic enrichments across multiple complex human diseases. *Nucleic acids research*, 44(18):e144–e144, 2016.

12. Jacob C Ulirsch, Caleb A Lareau, Erik L Bao, Leif S Ludwig, Michael H Guo, Christian Benner, Ansuman T Satpathy, Vinay K Kartha, Rany M Salem, Joel N Hirschhorn, et al. Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nature genetics*, 51(4):683–693, 2019.

13. Joshua Chiou, Chun Zeng, Zhang Cheng, Jee Yun Han, Michael Schlichting, Michael Miller, Robert Mendez, Serina Huang, Jinzhao Wang, Yinghui Sui, et al. Single-cell chromatin accessibility identifies pancreatic islet cell type–and state-specific regulatory programs of diabetes risk. *Nature Genetics*, 53(4):455–466, 2021.

14. Xinli Hu, Hyun Kim, Eli Stahl, Robert Plenge, Mark Daly, and Soumya Raychaudhuri. Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *The American Journal of Human Genetics*, 89(4):496–506, 2011.

15. Kamil Slowikowski, Xinli Hu, and Soumya Raychaudhuri. Snpsea: an algorithm to identify cell types, tissues and pathways affected by risk loci. *Bioinformatics*, 30(17):2496–2497, 2014.

16. Tune H Pers, Juha M Karjalainen, Yingleong Chan, Harm-Jan Westra, Andrew R Wood, Jian Yang, Julian C Lui, Sailaja Vedantam, Stefan Gustafsson, Tonu Esko, et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nature communications*, 6(1):1–9, 2015.

17. Padhraig Gormley, Verneri Anttila, Bendik S Winsvold, Priit Palta, Tonu Esko, Tune H Pers, Kai-How Farh, Ester Cuenca-Leon, Mikko Muona, Nicholas A Furlotte, et al. Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nature genetics*, 48(8):856–866, 2016.

18. Diego Calderon, Anand Bhaskar, David A Knowles, David Golan, Towfique Raj, Audrey Q Fu, and Jonathan K Pritchard. Inferring relevant cell types for complex traits by using single-cell gene expression. *The American Journal of Human Genetics*, 101(5):686–699, 2017.

19. Halit Ongen, Andrew A Brown, Olivier Delaneau, Nikolaos I Panousis, Alexandra C Nica, and Emmanouil T Dermitzakis. Estimating the causal tissues for complex traits and diseases. *Nature genetics*, 49(12):1676–1683, 2017.

20. Hilary K Finucane, Yakir A Reshef, Verneri Anttila, Kamil Slowikowski, Alexander Gusev, Andrea Byrnes, Steven Gazal, Po-Ru Loh, Caleb Lareau, Noam Shoresh, et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nature genetics*, 50(4):621–629, 2018.

21. Kyoko Watanabe, Maša Umićević Mirkov, Christiaan A de Leeuw, Martijn P van den Heuvel, and Danielle Posthuma. Genetic mapping of cell type specificity for complex traits. *Nature communications*, 10(1):1–13, 2019.

22. Julien Bryois, Nathan G Skene, Thomas Folkmann Hansen, Lisette JA Kogelman, Hunna J Watson, Zijing Liu, Leo Brueggeman, Gerome Breen, Cynthia M Bulik, Ernest Arenas, et al. Genetic identification of cell types underlying brain complex traits yields insights into the etiology of parkinson's disease. *Nature genetics*, 52(5):482–493, 2020.

23. Karthik A Jagadeesh, Kushal K Dey, Daniel T Montoro, Steven Gazal, Jesse M Engreitz, Ramnik J Xavier, Alkes L Price, and Aviv Regev. Identifying disease-critical cell types and cellular processes across the human body by integration of single-cell profiles and human genetics. *bioRxiv*, 2021.

24. Meromit Singer, Chao Wang, Le Cong, Nemanja D Marjanovic, Monika S Kowalczyk, Huiyuan Zhang, Jackson Nyman, Kaori Sakuishi, Sema Kurtulus, David Gennert, et al. A distinct gene module for dysfunction uncoupled from activation in tumor-infiltrating t cells. *Cell*, 166(6):1500–1511, 2016.

25. Jean Fan, Neeraj Salathia, Rui Liu, Gwendolyn E Kaeser, Yun C Yung, Joseph L Herman, Fiona Kaper, Jian-Bing Fan, Kun Zhang, Jerold Chun, et al. Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nature methods*, 13(3):241–244, 2016.

26. Sara Aibar, Carmen Bravo González-Blas, Thomas Moerman, Hana Imrichova, Gert Hulselmans, Florian Rambow, Jean-Christophe Marine, Pierre Geurts, Jan Aerts, Joost van den Oord, et al. Scenic: single-cell regulatory network inference and clustering. *Nature methods*, 14(11):1083–1086, 2017.

27. Sidharth V Puram, Itay Tirosh, Anuraag S Parikh, Anoop P Patel, Keren Yizhak, Shawn Gillespie, Christopher Rodman, Christina L Luo, Edmund A Mroz, Kevin S Emerick, et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell*, 171(7):1611–1624, 2017.

28. Andrew Butler, Paul Hoffman, Peter Smibert, Efthymia Papalexi, and Rahul Satija. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature biotechnology*, 36(5):411–420, 2018.

29. David DeTomaso, Matthew G Jones, Meena Subramaniam, Tal Ashuach, J Ye Chun, and Nir Yosef. Functional interpretation of single cell similarity maps. *Nature communications*, 10(1):1–11, 2019.

30. Brendan K Bulik-Sullivan, Po-Ru Loh, Hilary K Finucane, Stephan Ripke, Jian Yang, Nick Patterson, Mark J Daly, Alkes L Price, and Benjamin M Neale. Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics*, 47(3):291–295, 2015.

31. Hilary K Finucane, Brendan Bulik-Sullivan, Alexander Gusev, Gosia Trynka, Yakir Reshef, Po-Ru Loh, Verneri Anttila, Han Xu, Chongzhi Zang, Kyle Farh, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics*, 47(11):1228–1235, 2015.

32. Steven Gazal, Hilary K Finucane, Nicholas A Furlotte, Po-Ru Loh, Pier Francesco Palamara, Xuanyao Liu, Armin Schoech, Brendan Bulik-Sullivan, Benjamin M Neale, Alexander Gusev, et al. Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. *Nature genetics*, 49(10):1421–1427, 2017.

33. Luke J O'Connor, Armin P Schoech, Farhad Hormozdiari, Steven Gazal, Nick Patterson, and Alkes L Price. Extreme polygenicity of complex traits is explained by negative selection. *The American Journal of Human Genetics*, 105(3):456–476, 2019.

34. Farhad Hormozdiari, Steven Gazal, Bryce Van De Geijn, Hilary K Finucane, Chelsea J-T Ju, Po-Ru Loh, Armin Schoech, Yakir Reshef, Xuanyao Liu, Luke O'connor, et al. Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. *Nature genetics*, 50(7):1041–1047, 2018.

35. Steven Gazal, Po-Ru Loh, Hilary K Finucane, Andrea Ganna, Armin Schoech, Shamil Sunyaev, and Alkes L Price. Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. *Nature genetics*, 50(11):1600–1607, 2018.

36. The Tabula Muris Consortium. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature*, 583(7817):590–595, 2020.

37. Stephen R Quake, Tabula Sapiens Consortium, et al. The tabula sapiens: a single cell transcriptomic atlas of multiple organs from individual human donors. *bioRxiv*, 2021.

38. Eddie Cano-Gamez, Blagoje Soskic, Theodoros I Roumeliotis, Ernest So, Deborah J Smyth, Marta Baldrighi, David Willé, Nikolina Nakic, Jorge Esparza-Gordillo, Christopher GC Larminie, et al. Single-cell transcriptomics identifies an effectorness gradient shaping the response of cd4+ t cells to cytokines. *Nature communications*, 11(1):1–15, 2020.

39. Aparna Nathan, Jessica I Beynor, Yuriy Baglaenko, Sara Suliman, Kazuyoshi Ishigaki, Samira Asgari, Chuan-Chin Huang, Yang Luo, Zibiao Zhang, Kattya Lopez, et al. Multimodally profiling memory t cells from a tuberculosis cohort identifies cell state associations with demographics, environment and disease. *Nature Immunology*, 22(6):781–793, 2021.

40. Amit Zeisel, Ana B Muñoz-Manchado, Simone Codeluppi, Peter Lönnerberg, Gioele La Manno, Anna Juréus, Sueli Marques, Hermany Munguba, Liqun He, Christer Betsholtz, et al. Cell types in the mouse cortex and hippocampus revealed by single-cell rna-seq. *Science*, 347(6226):1138–1142, 2015.

41. Amit Zeisel, Hannah Hochgerner, Peter Lönnerberg, Anna Johnsson, Fatima Memic, Job Van Der Zwan, Martin Häring, Emelie Braun, Lars E Borm, Gioele La Manno, et al. Molecular architecture of the mouse nervous system. *Cell*, 174(4):999–1014, 2018.

42. Naomi Habib, Yinqing Li, Matthias Heidenreich, Lukasz Swiech, Inbal Avraham-Davidi, John J Trombetta, Cynthia Hession, Feng Zhang, and Aviv Regev. Div-seq: Single-nucleus rna-seq reveals dynamics of rare adult newborn neurons. *Science*, 353(6302):925–928, 2016.

43. Naomi Habib, Inbal Avraham-Davidi, Anindita Basu, Tyler Burks, Karthik Shekhar, Matan Hofree, Sourav R Choudhury, François Aguet, Ellen Gelfand, Kristin Ardlie, et al. Massively parallel single-nucleus rna-seq with dronc-seq. *Nature methods*, 14(10):955–958, 2017.

44. Fatma Ayhan, Ashwinikumar Kulkarni, Stefano Berto, Karthigayini Sivaprakasam, Connor Douglas, Bradley C Lega, and Genevieve Konopka. Resolving cellular and molecular diversity along the hippocampal anterior-to-posterior axis in humans. *Neuron*, 2021.

45. Zizhen Yao, Cindy TJ van Velthoven, Thuc Nghi Nguyen, Jeff Goldy, Adriana E Sedeno-Cortes, Fahimeh Baftizadeh, Darren Bertagnolli, Tamara Casper, Megan Chiang, Kirsten Crichton, et al. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell*, 184(12):3222–3241, 2021.

46. Suijuan Zhong, Wenyu Ding, Le Sun, Yufeng Lu, Hao Dong, Xiaoying Fan, Zeyuan Liu, Ruiguo Chen, Shu Zhang, Qiang Ma, et al. Decoding the development of the human hippocampus. *Nature*, 577(7791):531–536, 2020.

47. Nadim Aizarani, Antonio Saviano, Laurent Mailly, Sarah Durand, Josip S Herman, Patrick Pessaux, Thomas F Baumert, Dominic Grün, et al. A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature*, 572(7768):199–204, 2019.

48. Keren Bahar Halpern, Rom Shenhav, Orit Matcovitch-Natan, Beáta Tóth, Doron Lemze, Matan Golan, Efi E Massasa, Shaked Baydatch, Shanie Landen, Andreas E Moor, et al. Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature*, 542(7641):352–356, 2017.

49. ML Richter, IK Deligiannis, K Yin, A Danese, E Lleshi, P Coupland, Catalina A Vallejos, KP Matchett, NC Henderson, M Colome-Tatche, et al. Single-nucleus rna-seq2 reveals functional crosstalk between liver zonation and ploidy. *Nature communications*, 12(1):1–16, 2021.

50. Aris Taychameekiatchai and Bruce Wang. Tentative title. *Manuscript in preparation*, 2021.

51. Malika Kumar Freund, Kathryn S Burch, Huwenbo Shi, Nicholas Mancuso, Gleb Kichaev, Kristina M Garske, David Z Pan, Zong Miao, Karen L Mohlke, Markku Laakso, et al. Phenotype-specific enrichment of mendelian disorder genes near gwas regions across 62 complex traits. *The American Journal of Human Genetics*, 103(4):535–552, 2018.

52. Peng Huang, Yongzhong Zhao, Jianmei Zhong, Xinhua Zhang, Qifa Liu, Xiaoxia Qiu, Shaoke Chen, Hongxia Yan, Christopher Hillyer, Narla Mohandas, et al. Putative regulators for the continuum of erythroid differentiation revealed by single-cell transcriptome of human bm and ucb cells. *Proceedings of the National Academy of Sciences*, 117(23):12868–12876, 2020.

53. Sonia Olivia Spitzer, Sergey Sitnikov, Yasmine Kamen, Kimberley Anne Evans, Deborah Kronenberg-Versteeg, Sabine Dietmann, Omar de Faria Jr, Sylvia Agathou, and Ragnhildur Thóra Káradóttir. Oligodendrocyte progenitor cells become regionally diverse and heterogeneous with age. *Neuron*, 101(3):459–471, 2019.

54. Giselle Dominguez Gutierrez, Jesper Gromada, and Lori Sussel. Heterogeneity of the pancreatic beta cell. *Frontiers in genetics*, 8:22, 2017.

55. F Alexander Wolf, Philipp Angerer, and Fabian J Theis. Scanpy: large-scale single-cell gene expression data analysis. *Genome biology*, 19(1):1–5, 2018.

56. Amy Li, Rebecca H Herbst, David Canner, Jason M Schenkel, Olivia C Smith, Jonathan Y Kim, Michelle Hillman, Arjun Bhutkar, Michael S Cuoco, C Garrett Rappazzo, et al. Il-33 signaling alters regulatory t cell diversity in support of tumor development. *Cell reports*, 29(10):2998–3008, 2019.

57. Maxim V Kuleshov, Matthew R Jones, Andrew D Rouillard, Nicolas F Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, Sherry L Jenkins, Kathleen M Jagodnik, Alexander Lachmann, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, 44(W1):W90–W97, 2016.

58. Zhuorui Xie, Allison Bailey, Maxim V Kuleshov, Daniel JB Clarke, John E Evangelista, Sherry L Jenkins, Alexander Lachmann, Megan L Wojciechowicz, Eryk Kropiwnicki, Kathleen M Jagodnik, et al. Gene set knowledge discovery with enrichr. *Current protocols*, 1(3):e90, 2021.

Supplementary Tables

See Supplementary Excel file

Supplementary Table 1. GWAS diseases and complex trait data sets. We report the name, identifier, code, category, reference, sample size, number of variants, estimated heritability using LD Score regression (LDSC)[30–32] for the 74 diseases/traits analyzed in the paper. We also report polygenicity for a subset of 21 diseases/traits (M_e for common SNPs; Table 1 in O'Connor et al.[33]). The disease called "Auto Immune Traits" (in UK Biobank) is based on the following codes and disease names that characterize autoimmune physiopathogenic etiology: 1222 (t1d; type 1 diabetes); 1256 (guillainBarre); 1260 (myasthenia); 1261 (ms); 1372 (vasculitis); 1378 (Wegener's); 1381 (sle); 1382 (Sjogren); 1384 (sysSclerosis); 1437 (myasthenia); 1456 (celiac); 1464 (ra); 1522 (grave); 1661 (vitiligo)[34,35].

Data set	Species	N_{cell}	N_{tissue}	$N_{\text{cell type}}$	Description
TMS FACS[36]	Mus musculus	110,096	23	120	Mouse cell atlas (FACS + Smart-seq2)
TMS droplet[36]	Mus musculus	245,389	16	123	Mouse cell atlas (10x microfluidic droplets)
TS FACS[37]	Homo sapiens	26,813	24	134	Human cell atlas (FACS + Smart-seq2)
Cano-Gamez & Soskic et al.[38]	Homo sapiens	43,112	1	22	Subtypes of naive, memory, and activated $CD4^+$ T cells from the blood
Nathan et al.[39]	Homo sapiens	500,089	1	30	T cells from the blood
Zeisel & Muñoz-Manchado et al.[40]	Mus musculus	3,005	1	9	Cortex and hippocampus; 827 CA1 pyramidal cells
Zeisel et al.[41]	Mus musculus	160,797	1	265	Whole nervous system; 304 CA1 pyramidal cells
Habib & Li et al.[42]	Mus musculus	1,367	1	7	Hippocampal regions from adult mice; 155 CA1 pyramidal cells; snRNA-seq
Habib, Avraham-Davidi, & Basu et al.[43]	Homo sapiens	14,963	1	21	Archived brain sample; 421 CA1 pyramidal cells; snRNA-seq
Ayhan et al.[44]	Homo sapiens	129,908	1	24	Surgically resected anterior and posterior hippocampus from epilepsy patients; 5,454 CA1 pyramidal cells; snRNA-seq
Yao et al.[45]	Mus musculus	74,974	1	388	Cortex and hippocampus; 1,701 CA1 pyramidal cells using SMART-Seq v4 technology
Zhong et al.[46]	Homo sapiens	30,416	1	11	Hippocampus at gestational weeks 16–27; 5,972 CA1 pyramidal cells
Aizarani et al.[47]	Homo sapiens	10,372	1	11	Hepatocytes, endothelial cells, and other common cell types from the liver
Halpern & Shenhav et al.[48]	Mus musculus	1,415	1	1	Hepatocytes
Richter & Deligiannis et al.[49]	Mus musculus	1,649	1	1	Sorted 2n and 4n hepatocytes (Hoechst dye + FACS); snRNA-seq
Taychameekiatchai et al.[50]	Mus musculus	19,254	1	15	Hepatocytes, endothelial cells, and other common cell types from the liver

Supplementary Table 2. ScRNA-seq and snRNA-seq data sets. We report the reference, species, number of cells, number of tissues, number of cell types, and a short description for each scRNA-seq/snRNA-seq data set analyzed in the paper. Data sets without "snRNA-seq" in the description are scRNA-seq data sets. The 16 data sets contain more than 1.3 million cells from 31 tissues and organs, including aorta, brown adipose tissue (BAT), bladder, blood, bone marrow, brain myeloid, brain non-myeloid, diaphragm, eye, gonadal adipose tissue (GAT), heart, kidney, large intestine, limb muscle, liver, lung, lymph node, mesenteric adipose tissue (MAT), mammary gland, pancreas, prostate, subcutaneous adipose tissue (SCAT), salivary gland, skin, small intestine, spleen, thymus, tongue, trachea, uterus, vasculature. For clarification, Zeisel & Muñoz-Manchado et al. refers to the data from Zeisel & Muñoz-Manchado et al. 2015 *Science*[40] and Zeisel et al. refers to the data from Zeisel et al. 2018 *Cell*[41].

See Supplementary Excel file

Supplementary Table 3. Cell types in the TMS FACS data. We report 120 cell types in the TMS FACS data, the corresponding number of cells, and the corresponding tissue composition (for tissues consisting >1% of cells from the cell type).

See Supplementary Excel file

Supplementary Table 4. Cell types in the TMS droplet data. We report 123 cell types in the TMS droplet data, the corresponding number of cells, and the corresponding tissue composition (for tissues consisting >1% of cells from the cell type).

See Supplementary Excel file

Supplementary Table 5. Cell types in the TS FACS data. We report 132 cell types in the TS FACS data, the corresponding number of cells, and the corresponding tissue composition (for tissues consisting >1% of cells from the cell type).

See Supplementary Excel file

Supplementary Table 6. MAGMA gene sets. We report MAGMA gene sets for the 74 diseases and traits.

See Supplementary Excel file

Supplementary Table 7. MSigDB terms for curating signature gene sets. We report the MSigDB terms used to curate the signature gene sets in the paper.

See Supplementary Excel file

Supplementary Table 8. Signature gene sets. We report the signature gene sets used in the paper.

See Supplementary Excel file

Supplementary Table 9. Numerical results for null simulations in Fig. 2A. We report the mean and SE of p-value quantiles for different cell-scoring methods over 100 repetitions.

Nominal FDR level	0.05	0.1	0.2
Actual FDR level	0.00±0.00	0.02±0.03	0.18±0.08

Supplementary Table 10. Results for null simulations for testing cell type-disease association. We assessed calibration of the MC test for cell type-disease association based on the output of `scDRS`. We used the same subsampled data (10,000 cells from TMS FACS) and 1,000 randomly-selected disease genes. We report the actual FDR for multiple testing across all 118 cell types in the subsampled data at various nominal FDR levels. 95% confidence intervals are provided based on the 100 repetitions.

See Supplementary Excel file

Supplementary Table 11. Numerical results for causal simulations in Fig. 2B. We report the mean and SE of power at various effect sizes for different cell-scoring methods over 100 repetitions.

See Supplementary Excel file

Supplementary Table 12. Numerical results for cell type-level analyses for the TMS FACS data in Fig. 3. For each pair of cell type and disease/trait, we report the proportion of significantly associated cells (FDR<0.1), FDR for cell type-disease association, and FDR for within-cell type disease association heterogeneity.

See Supplementary Excel file

Supplementary Table 13. Numerical results for cell type-level analyses for the TMS droplet data. For each pair of cell type and disease/trait, we report the proportion of significantly associated cells (FDR<0.1), FDR for cell type-disease association, and FDR for within-cell type disease association heterogeneity.

See Supplementary Excel file

Supplementary Table 14. Numerical results for cell type-level analyses for the TS FACS data. For each pair of cell type and disease/trait, we report the proportion of significantly associated cells (FDR<0.1), FDR for cell type-disease association, and FDR for within-cell type disease association heterogeneity.

Trait	Positive control	Negative control
Schizophrenia	Neuron	T cell
Major depressive disorder	Neuron	T cell
Atrial fibrillation	Atrial & ventricular myocyte	Neuron
Rheumatoid arthritis	T cell	Hepatocyte
LDLdirect	Hepatocyte	Neuron

Supplementary Table 15. Traits and control cell types for evaluating parameters when constructing putative disease gene sets. We report the 5 traits and the corresponding positive and negative control cell types in the TMS FACS data for evaluating the performance of `scDRS` under different parameter settings. "Neuron" refers to all sub-types of neurons, including MSNs and interneurons; "T cell" refers to all sub-types of T cells.

Window size	**10kb**				**0kb**	**50kb**
Number of genes	**100**	**500**	**1000**	**2000**	**1000**	
Schizophrenia	30.0	47.8	54.3	**60.4**	**56.7**	52.8
Major depressive disorder	40.5	47.4	**52.4**	**53.1**	52.0	51.9
Atrial fibrillation	**33.5**	24.7	**25.3**	21.7	24.2	23.4
Rheumatoid arthritis	124.4	120.8	**133.1**	117.2	108.3	**137.8**
LDLdirect	61.4	**91.7**	71.0	76.8	71.7	**78.2**

Supplementary Table 16. Evaluating parameters for constructing putative disease gene sets. We report t-statistics for comparing the normalized disease scores from the positive control and negative control cell types for each trait and each configuration. For each trait, the top two configurations yielding the highest t-statistics are highlighted in bold font.

Trait	CD4.P	CD4.Var	CD4.P.joint	CD8.P	CD8.Var	CD8.P.joint
IBD	**0.001**	0.310	**0.002**	0.009	0.113	0.010
CD	**0.001**	0.222	**0.002**	0.019	0.096	**0.003**
UC	**0.001**	0.192	0.032	0.378	0.000	0.489
RA	0.343	0.002	0.468	0.370	0.001	0.716
MS	0.339	0.002	0.390	0.202	0.009	0.492
AIT	**0.001**	0.177	**0.004**	0.060	0.049	0.208
HT	0.011	0.100	0.053	0.165	0.014	0.384
Eczema	0.020	0.066	0.298	0.804	0.012	0.870
ASM	0.023	0.062	0.294	0.263	0.004	0.532
RR-ENT	0.021	0.081	0.100	0.180	0.014	0.196
Height	0.291	0.004	0.747	0.828	0.010	0.635

Supplementary Table 17. Numerical results for correlations between `scDRS` disease scores and T cell effectorness gradients in Fig. 4C. We first regressed the `scDRS` disease score against the CD4 (resp., CD8) effectorness gradient for each of the 10 autoimmune diseases and the negative control trait height. We report p-values for significant positive correlation between the `scDRS` disease score and the effectorness gradients ("CD4.P"/"CD8.P"; MC test) and variance explained ("CD4.Var"/"CD8.Var"). We then jointly regressed the `scDRS` disease score against the CD4 (resp., CD8) effectorness gradient and the cluster labels (encoded as dummy variables). We report p-values for significant positive correlation between the `scDRS` disease score and the effectorness gradients ("CD4.P.joint"/"CD8.P.joint"; MC test). P-values smaller than 0.005 were highlighted in bold font.

See Supplementary Excel file

Supplementary Table 18. Gold standard gene sets. We report matched Experimental Factor Ontology (EFO), corresponding EFO label, corresponding putative drug target gene set, number of putative drug target genes, matched Mendelian disorder, corresponding Mendelian disease gene set, and number of Mendelian disease genes for 27 diseases with putative drug target gene sets (from Open Targets) and 45 diseases/traits with Mendelian gene sets (from Freund et al.[51]); a disease/trait may have both drug target and Mendelian disease gene set. Specifically, for the 10 autoimmune diseases in Fig. 4D, the immune dysregulation Medelian disease gene set was used for HT and RR-ENT while the corresponding drug target gene sets were used for the other 8 autoimmune diseases.

See Supplementary Excel file

Supplementary Table 19. Numerical results for comparison to gold standard gene sets in Fig. 4D. We report excess overlap and $-\log_{10}$ p-value for comparison to drug target and Mendelian disease gene sets respectively for each disease gene prioritization method (scDRS/MAGMA) and each of the 27 GWAS diseases with drug target gene sets and 45 diseases/traits with Mendelian disease gene sets.

Cano-Gamez et al.	IBD	CD	UC	RA	MS	AIT	HT	Eczema	ASM	RR-ENT	Height
TCM1 (Th17/iTreg)	.05	.05	ns	ns	ns	.04	ns	.04	ns	ns	ns
TCM2 (Th0)	ns	.05	ns	ns	ns	ns	ns	ns	ns	ns	ns
TEM (Th17/iTreg)	.01	.05	ns	ns	ns	.04	.02	ns	ns	ns	ns
TN (Th17/iTreg)	.01	.03	.01	ns	ns	ns	.04	.01	ns	.02	ns
nTreg (Th0)	.01	.02	.01	.02	.02	.02	.04	.01	ns	.03	ns
Nathan et al.											
$CD4^+$ $CCR4^+$	ns	ns	ns	ns	ns	ns	ns	<.01	ns	ns	ns
$CD4^+$ $CCR4^+ICOS^+$ central	ns	ns	.05	ns	ns	ns	ns	.02	ns	ns	ns
$CD4^+$ $CD38^+ICOS^+$ central	ns	ns	ns	<.01	ns	ns	ns	ns	ns	ns	ns
$CD4^+$ $CD161^+$ Th2	<.01	ns	<.01	ns	ns	ns	ns	<.01	.03	<.01	ns
$CD4^+$ HLA-DR^+	ns	ns	<.01	ns	ns	ns	ns	ns	ns	ns	ns
$CD4^+$ $RORC^+$ Treg	<.01	<.01	<.01	<.01	.01	<.01	<.01	<.01	<.01	<.01	ns
$CD4^+$ Th2	<.01	ns	<.01	ns	ns	ns	ns	<.01	.04	.03	ns
$CD4^+$ Th17	<.01	ns	.01	ns	ns	ns	<.01	<.01	<.01	<.01	ns
$CD4^+$ Th17/1	<.01	<.01	.03	ns	ns	ns	ns	<.01	<.01	<.01	ns
$CD4^+$ Treg	ns	ns	<.01	<.01	ns	<.01	<.01	ns	ns	ns	ns
$CD4^+$ activated	<.01	ns	<.01	<.01	ns	ns	<.01	.02	<.01	<.01	ns
$CD4^+$ central	ns	ns	ns	ns	ns	<.01	<.01	ns	ns	ns	ns
$CD4^+$ lncRNA	ns	ns	ns	<.01	ns	ns	<.01	ns	ns	<.01	ns
$CD8^+$ central	ns	ns	ns	ns	ns	ns	ns	.02	ns	ns	ns
Vd1	ns	ns	ns	<.01	ns	ns	ns	ns	ns	ns	ns

Supplementary Table 20. Cell type-disease associations for T cell subtypes in Cano-Gamez & Soskic et al. and Nathan et al. data sets. We report FDR for significant cell type-disease associations (FDR<0.05) in the Cano-Gamez & Soskic et al. and Nathan et al. data sets (Supp. Table 2). We performed cell type-disease associations between each cell type in the two data sets and each of the 10 autoimmune diseases (and height, a negative control trait) using scDRS-based MC test. The Nathan et al. data has 500,089 cells, which we randomly divided into 4 batches and processed them using scDRS separately, followed by combining the MC p-values across batches for each cell type-disease pair using Fisher's combined probability test. We applied FDR control for each data set and each disease separately across all cell types in the data set (22 cell types in Cano-Gamez & Soskic et al. and 29 cell types in Nathan et al.).

See Supplementary Excel file

Supplementary Table 21. Results for cell type-level analyses for the Zeisel & Muñoz-Manchado et al. data. For each pair of cell type and disease/trait, we report the proportion of significantly associated cells (FDR<0.1), FDR for cell type-disease association, and FDR for within-cell type disease association heterogeneity.

Trait	Dorsal.P	Dorsal.Var	Ventral.P	Ventral.Var	Proximal.P	Proximal.Var	Distal.P	Distal.Var	Deep.P	Deep.Var	Superficial.P	Superficial.Var
MDD	**0.001**	0.245	0.75	0.006	**0.002**	0.151	0.25	0.003	**0.001**	0.189	0.82	0.004
SCZ	**0.001**	0.178	0.67	0.002	**0.001**	0.194	0.81	0.004	**0.001**	0.114	0.83	0.004
NRT	**0.002**	0.130	0.28	0.004	**0.001**	0.205	0.65	0.001	**0.001**	0.146	0.55	0.000
Smoking	0.009	0.099	0.13	0.017	**0.001**	0.176	0.61	0.001	**0.001**	0.139	0.27	0.002
ECOL	**0.001**	0.165	0.42	0.000	**0.001**	0.216	0.41	0.000	**0.001**	0.168	0.69	0.001
BMI	0.021	0.060	0.071	0.026	**0.001**	0.287	0.61	0.000	**0.001**	0.142	0.44	0.000
Height	0.37	0.002	0.55	0.000	0.008	0.063	0.68	0.001	0.21	0.007	0.98	0.018

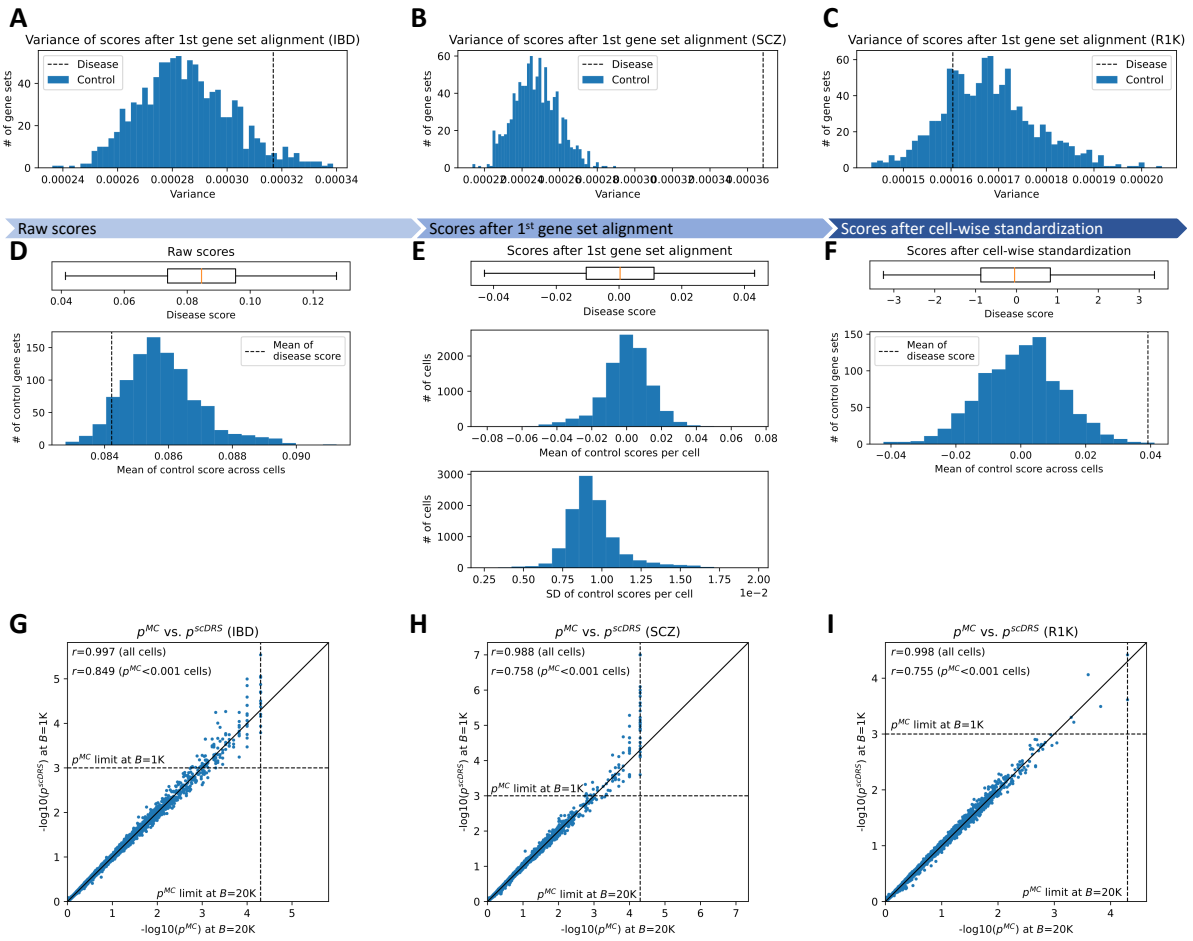
Supplementary Table 22. Numerical results for correlations between scDRS disease scores and inferred spatial coordinates in Fig. 5B. We separately regressed the scDRS scores of each of the 6 brain traits (and height, a negative control trait) against each of the 6 inferred spatial coordinates across the 827 CA1 pyramidal neurons. We report p-values for significant positive correlation between the scDRS disease score and the inferred spatial coordinates (MC test) and variance explained. P-values smaller than 0.005 were highlighted in bold font.

	Polyploidy.P	Pericentral.P	Periportal.P	Variance explained
TG	**0.001**	0.722	0.012	0.556
HDL	**0.003**	0.612	0.036	0.443
LDL	0.010	0.476	0.010	0.415
TC	**0.003**	0.432	0.008	0.457
TST	**0.001**	0.755	0.007	0.584
ALT	**0.001**	0.540	**0.004**	0.632
ALP	**0.001**	0.539	**0.001**	0.573
SHBG	**0.001**	0.643	0.018	0.597
TBIL	**0.001**	0.804	0.186	0.482
Height	0.622	0.940	0.239	0.114

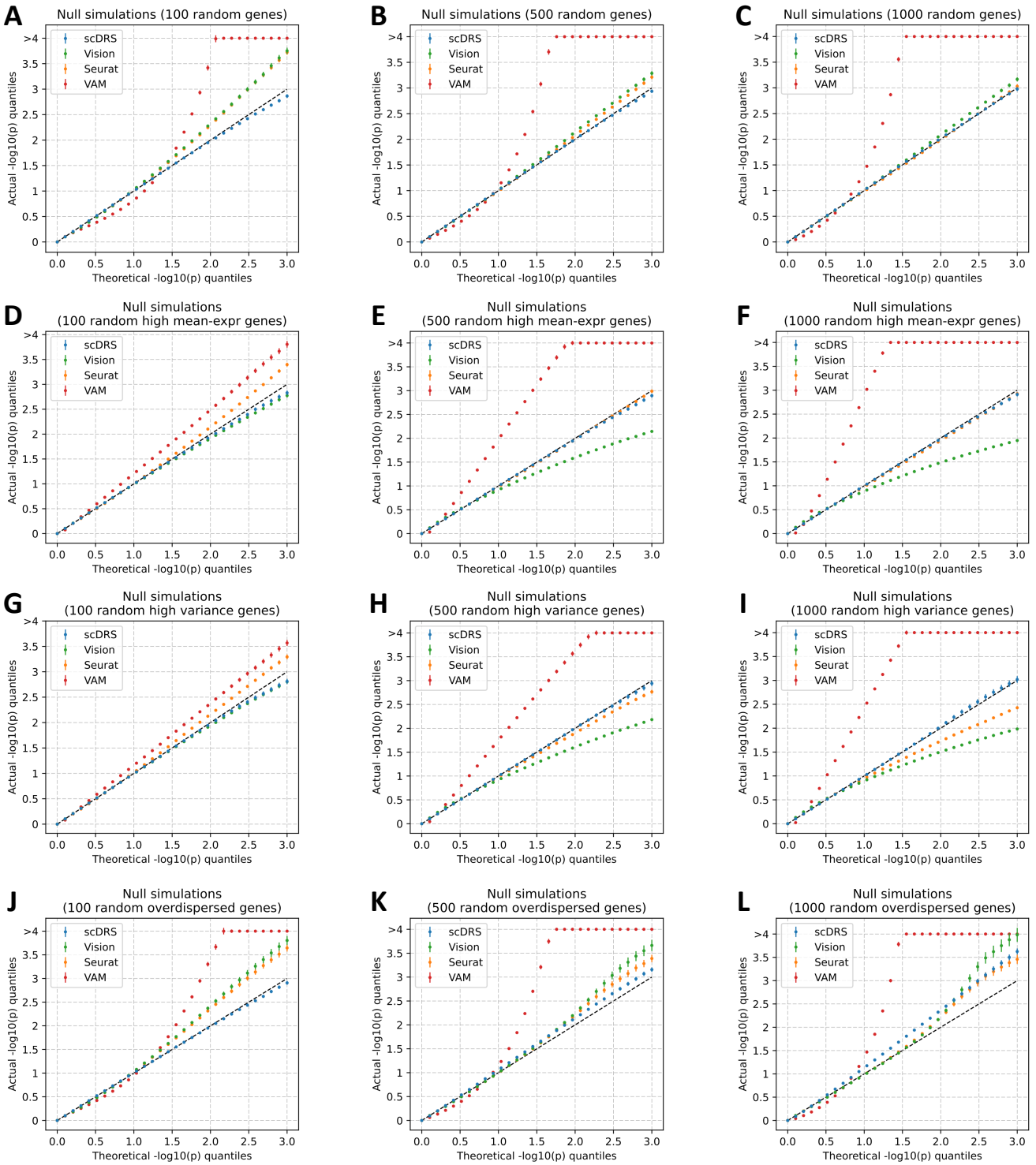
Supplementary Table 23. Numerical results for correlations between scDRS disease scores and inferred ploidy and zonation scores in Fig. 5D. We jointly regressed the scDRS scores of each of the 9 metabolic traits (and height, a negative control trait) on the polyploidy score, pericentral score, and periportal score. We report p-values for significant positive correlation between the scDRS disease score and the polyploidy score, pericentral score, and periportal score (MC test) and variance explained. P-values smaller than 0.005 were highlighted in bold font.

	TMS FACS	TMS Droplet	Aizarani	Halpern	Richter	Taychameekiatchai
4n hepatocyte (vs. 2n)	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
polyploid (Cdk1 ko)	**0.001**	**0.001**	**0.001**	0.038	**0.001**	**0.001**
large hepatocyte (vs. small)	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
2n hepatocyte (vs. 4n)	0.470	0.009	0.542	**0.003**	**0.001**	0.014
diploid (Cdk1 ko)	**0.001**	**0.001**	0.321	0.028	**0.001**	0.148
diploid (PH)	**0.001**	**0.001**	0.156	**0.001**	**0.002**	**0.001**

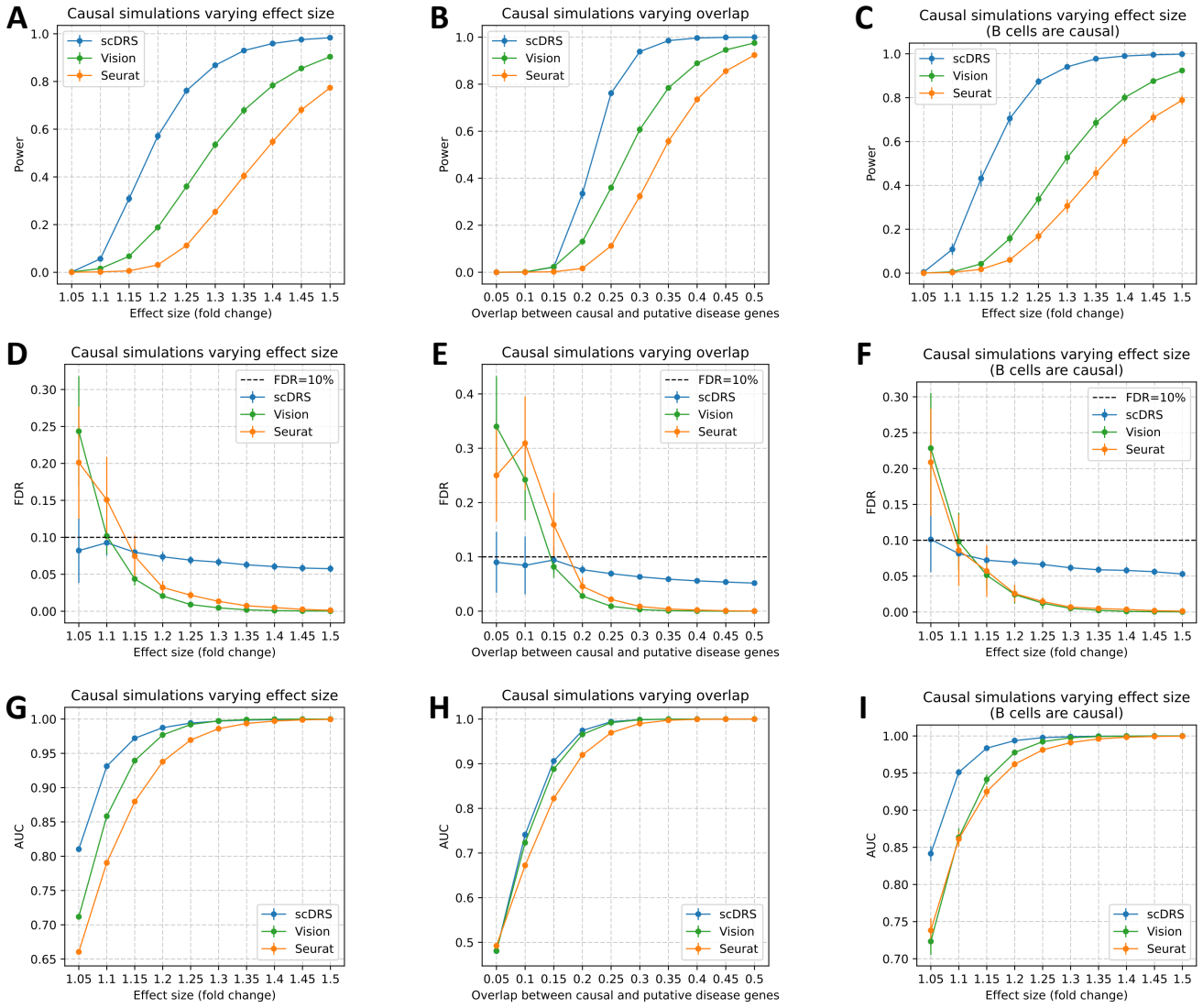
Supplementary Table 24. Correlation between scDRS scores and different ploidy signatures. We correlated our polyploidy score (based on DEGs for PH vs. pre-PH) with other ploypoidy and diploidy signatures for the 6 data sets. We report p-values for significant positive correlation for the polyploidy signatures (first three rows) and significant negative correlation for the diploidy signatures (last three rows). P-values smaller than 0.005 were highlighted in bold font.



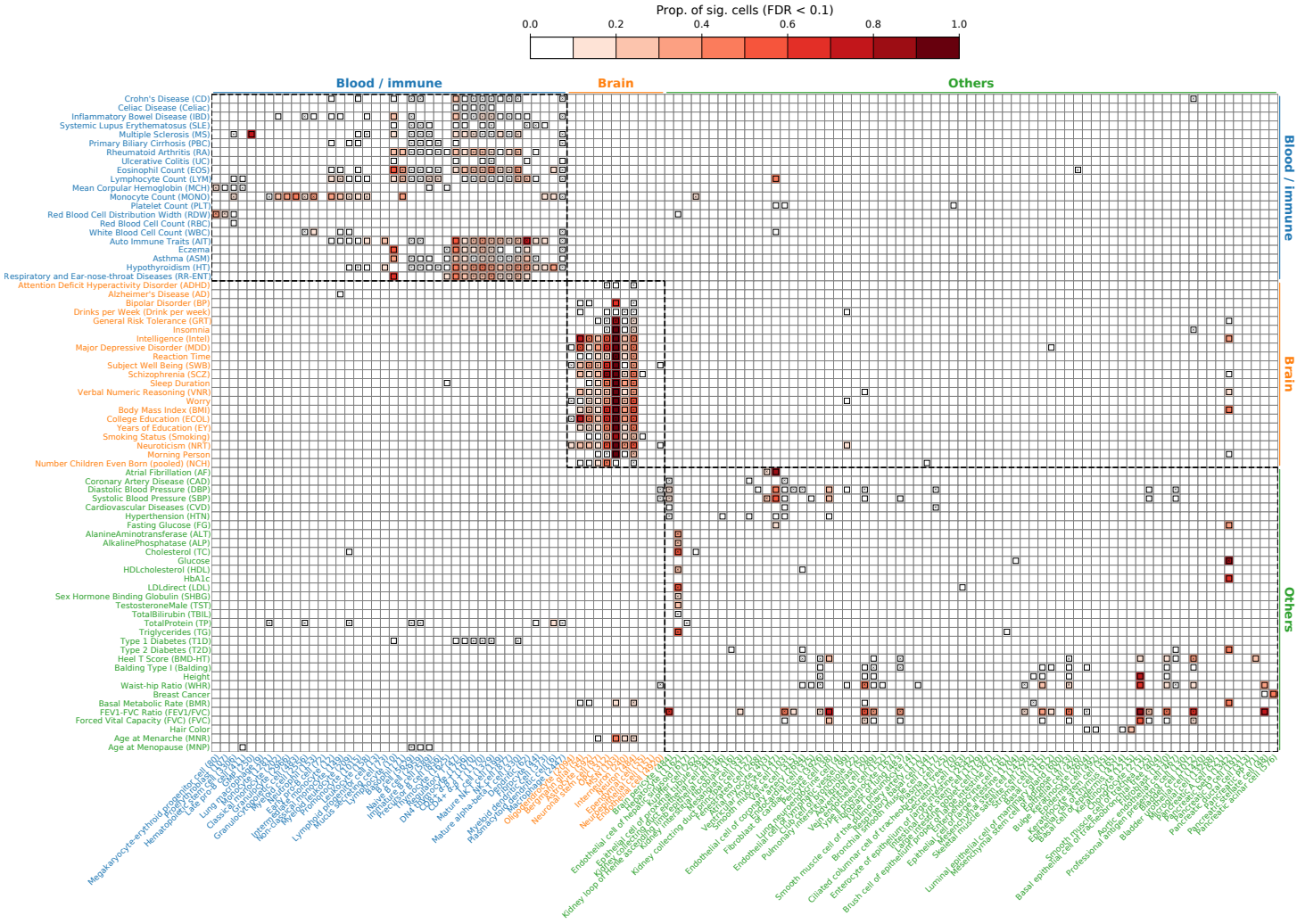
Supplementary Figure 1. Details of `scDRS` method. Results are based 10,000 cells subsampled from TMS FACS and 3 sets of putative disease genes: IBD, SCZ, and R1K (1,000 random genes). **(A-C)** Variance of raw disease scores (dashed line) and raw control scores (histogram) computed across cells for different gene sets. Informative disease gene sets (IBD and SCZ) have higher disease score variances while the uninformative gene set (R1K) has comparable values for the disease score variance and control score variances. We considered raw scores after the first gene set alignment to remove potential mismatch of expression mean and variance across gene sets. **(D)** Comparison between raw disease scores and raw control scores (before Box 1, step 3a). The upper panel shows the distribution of disease scores and the lower panel shows the histogram of the mean of control scores (computed for each control gene set across cells). There is a moderate level of mismatch of mean expression across control gene sets (ratio between the SD of mean control scores across control gene sets and SD of disease scores across cells is 6.8%); this mismatch is corrected by the first gene set alignment (Box 1, step 3a). **(E)** Comparison between disease scores and control scores after the first gene set alignment (before Box 1, step 3b). The upper panel shows the distribution of disease scores and the middle (resp. lower) panel shows the histogram of the per-cell mean (resp. SD) of control scores (computed for each cell across control gene sets). There is a high level of mismatch of the control score distribution (mean and SD) across cells; this mismatch is corrected by the cell-wise standardization (Box 1, step 3b). **(F)** Comparison between disease scores and control scores after cell-wise standardization (before Box 1, step 3c). The upper panel shows the distribution of disease scores and the lower panel shows the histogram of the mean of control scores (computed for each control gene set across cells). There is a mild level of mismatch of mean expression across control gene sets (ratio between the SD of mean control scores across control gene sets and SD of disease scores across cells is 1%); this mismatch is corrected by the second gene set alignment (Box 1, step 3c). Panels D-F are based on the IBD results. **(G-I)** Comparison between MC p-values with B =20,000 and `scDRS` p-values with B =1,000 for IBD, SCZ, and R1K. Each dot denotes a cell and the p^{MC} limit 1/(1+B) is the smallest MC p-value that an MC test with B MC samples can achieve. The `scDRS` p-values are highly correlated with the MC p-values obtained with B =20,000, both across all cells and across cells with p^{MC} <0.001 (p^{MC} limit at $B = 1{,}000$; corresponding to cells whose ideal MC p-values (with $B = \infty$) are small and require `scDRS` to extrapolate beyond the p^{MC} limit at the given number of MC samples), suggesting that `scDRS` can reliably approximate the ideal p-values obtained with an infinite number of MC samples. The dots on the vertical dashed line correspond to those whose ideal MC p-values are smaller than the p^{MC} limit at B =20,000; it is not surprising that the `scDRS` p-values, approximating the ideal MC p-values, are on average smaller than the MC p-values for these cells.



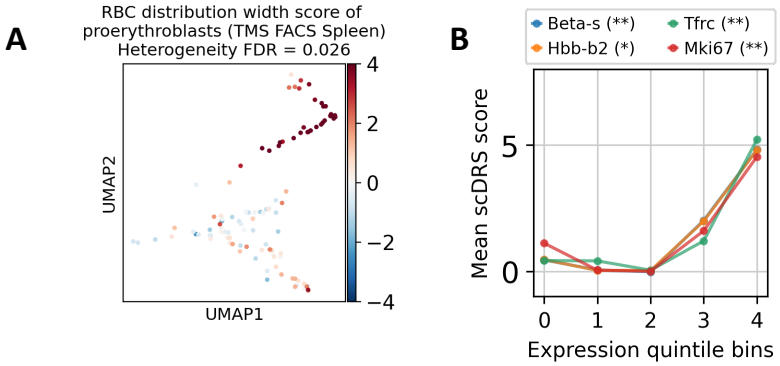
Supplementary Figure 2. Additional null simulations We performed null simulations for various numbers of putative disease genes (100, 500, 1,000 for the three columns respectively) and various types of genes to randomly sample from: all genes (first row), and top 25% genes with high expression (second row), top 25% genes with high expression variance (third row), top 25% overdispersed genes (fourth row). In each panel, the x-axis denotes theoretical $-\log_{10}$ p-value quantiles and the y-axis denotes actual $-\log_{10}$ p-value quantiles for different methods. `scDRS` produced well-calibrated p-values in most settings and suffered slightly inflated type I error in panel L, possibly because it is hard to match a large number of overdispersed putative disease genes using the remaining set of genes. In comparison, all other methods are less well-calibrated and are particularly problematic when the numbers of putative disease genes are small. All experiments were repeated 100 times and 95% confidence intervals were provided.



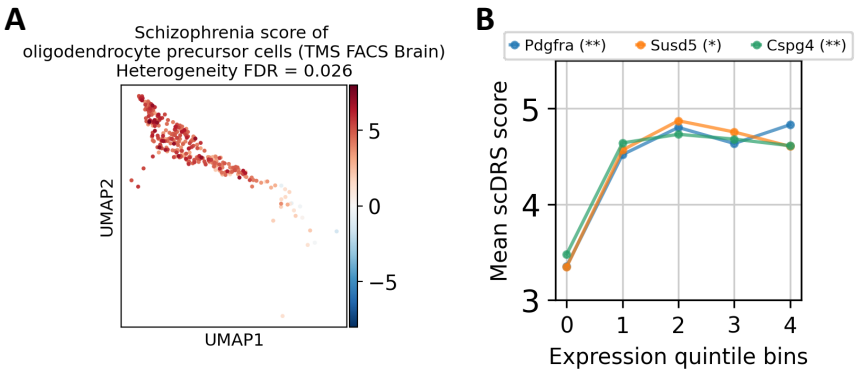
Supplementary Figure 3. Additional causal simulations. We performed three sets of causal simulations: (1) varying effect size from 5% to 50% while fixing 25% overlap (first column), (2) varying level of overlap from 5% to 50% while fixing 25% effect size (second column), (3) assigning the 528 B cells in the subsampled data to be causal (instead of the 500 randomly selected cells; varying effect size while fixing 25% overlap; third column). We report the power (first row), FDR (second row), and AUC for classifying causal from non-causal cells based on the p-values (third row). `scDRS` outperformed other methods under all metrics. All experiments were repeated 100 times and 95% confidence intervals were provided.



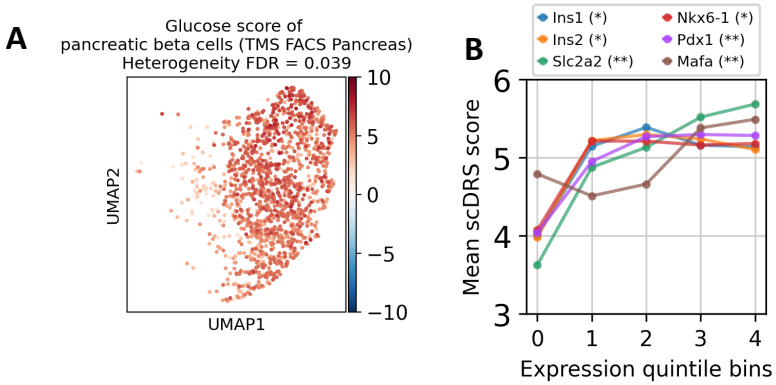
Supplementary Figure 4. Complete results for disease associations at the cell type level for 74 diseases/traits and 120 cell types in the TMS FACS data in Fig. 3. Each row represents a disease/trait and each column represents a cell type (with number of cells indicated in parentheses). Heatmap colors for each cell type-disease pair denote the proportion of significantly associated cells (FDR<0.1 across all cells for a given disease). Squares denote significant cell type-disease associations (FDR<0.05 across all pairs of the 120 cell types and 74 diseases/traits; 577 significant pairs; p-values via MC test; Methods). Cross symbols denote significant heterogeneity in association with disease across individual cells within a given cell type (FDR<0.05 across all pairs; 247 significant pairs; p-values via MC test; Methods). Heatmap colors and cross symbols are omitted for cell type-disease pairs with non-significant cell type-disease associations. Within the blood/immune block (40 cell types and 21 diseases/traits), 120 of 249 cell type-disease pairs with significant association also had significant heterogeneity. Within the brain block (11 cell types and 21 diseases/traits), 56 of 130 cell type-disease pairs with significant association also had significant heterogeneity. Within the other block (69 cell types and 32 diseases/traits), 51 of 143 cell type-disease pairs with significant association also had significant heterogeneity. Numerical results are reported in Supp. Table 12.



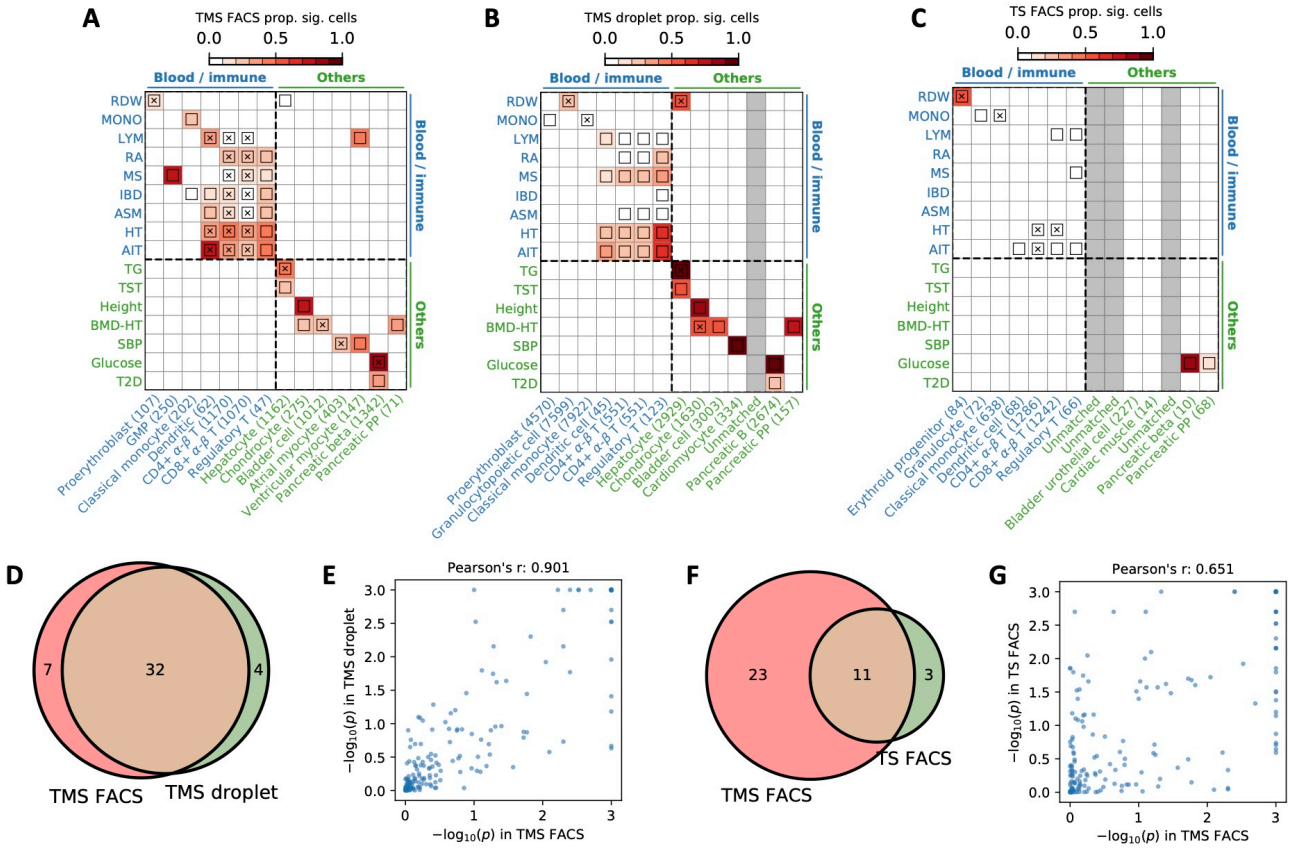
Supplementary Figure 5. Heterogeneous subpopulations of proerythroblasts associated with red blood cell distribution width (RDW). (A) Significant heterogeneity (FDR=0.026) of proerythroblasts (in the spleen) in association with RDW. **(B)** Expression levels of proerythroblast marker genes are significantly positively correlated with the `scDRS` disease score. The x-axis denotes marker gene expression quintile bins and the y-axis denotes average `scDRS` disease score for each bin. * denotes P <0.05 and ** denotes P <0.005. The heterogeneous association levels of proerythroblasts with RDW may correspond to the different differentiation stages of proerythroblasts[52]. Of note, the `scDRS` disease score was not correlated with age (P =0.13) or sex (P =0.39).



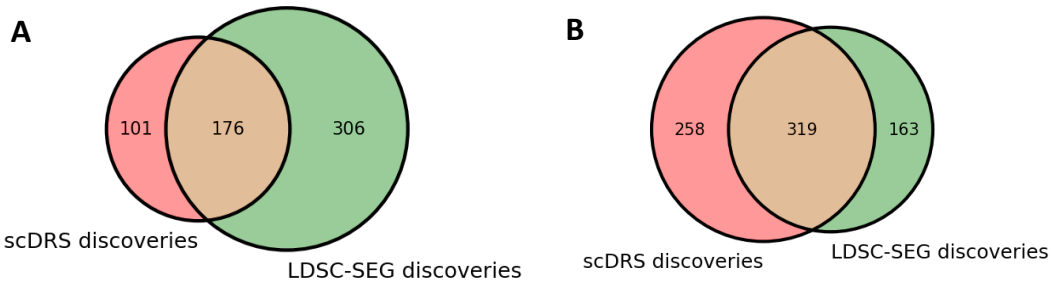
Supplementary Figure 6. Heterogeneous subpopulations of oligodendrocyte precursor cells associated with schizophrenia (SCZ). (A) Significant heterogeneity (FDR=0.039) of oligodendrocyte precursor cells (in the brain non-myeloid) in association with SCZ. **(B)** Expression levels of oligodendrocyte precursor cell marker genes are significantly positively correlated with the `scDRS` disease score. The x-axis denotes marker gene expression quintile bins and the y-axis denotes average `scDRS` disease score for each bin. * denotes P <0.05 and ** denotes P <0.005. The heterogeneous association levels of oligodendrocyte precursor cells with SCZ may correspond to the different developmental stages of oligodendrocyte precursor cells[53]. Of note, the `scDRS` disease score was slightly higher in male than female (P =0.012), and was not correlated with age (P =0.083).



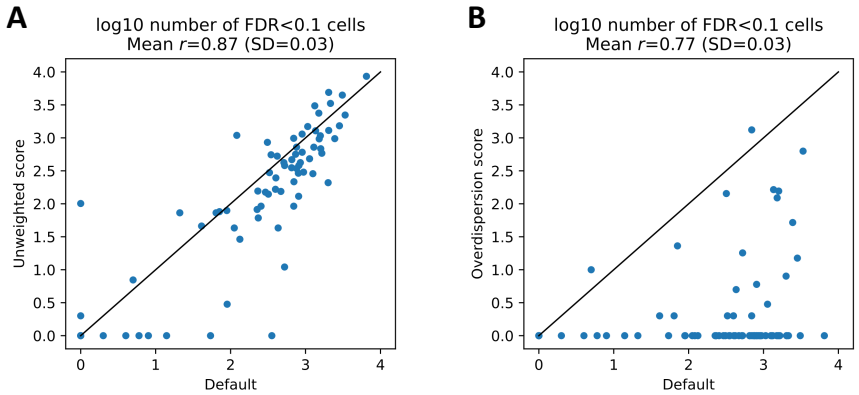
Supplementary Figure 7. Heterogeneous subpopulations of pancreatic beta cells associated with glucose. (A) Significant heterogeneity (FDR=0.039) of pancreatic beta cells (in the pancreas) in association with glucose. **(B)** Expression levels of pancreatic beta cell marker genes are significantly positively correlated with the `scDRS` disease score. The x-axis denotes marker gene expression quintile bins and the y-axis denotes average `scDRS` disease score for each bin. * denotes $P < 0.05$ and ** denotes $P < 0.005$. The heterogeneous association levels of pancreatic beta cells with glucose level may correspond to different states of pancreatic beta cells with varying insulin production levels[54]. Of note, the `scDRS` disease score was higher in female than male ($P = 0.003$) and was negatively correlated with age ($P = 0.002$).



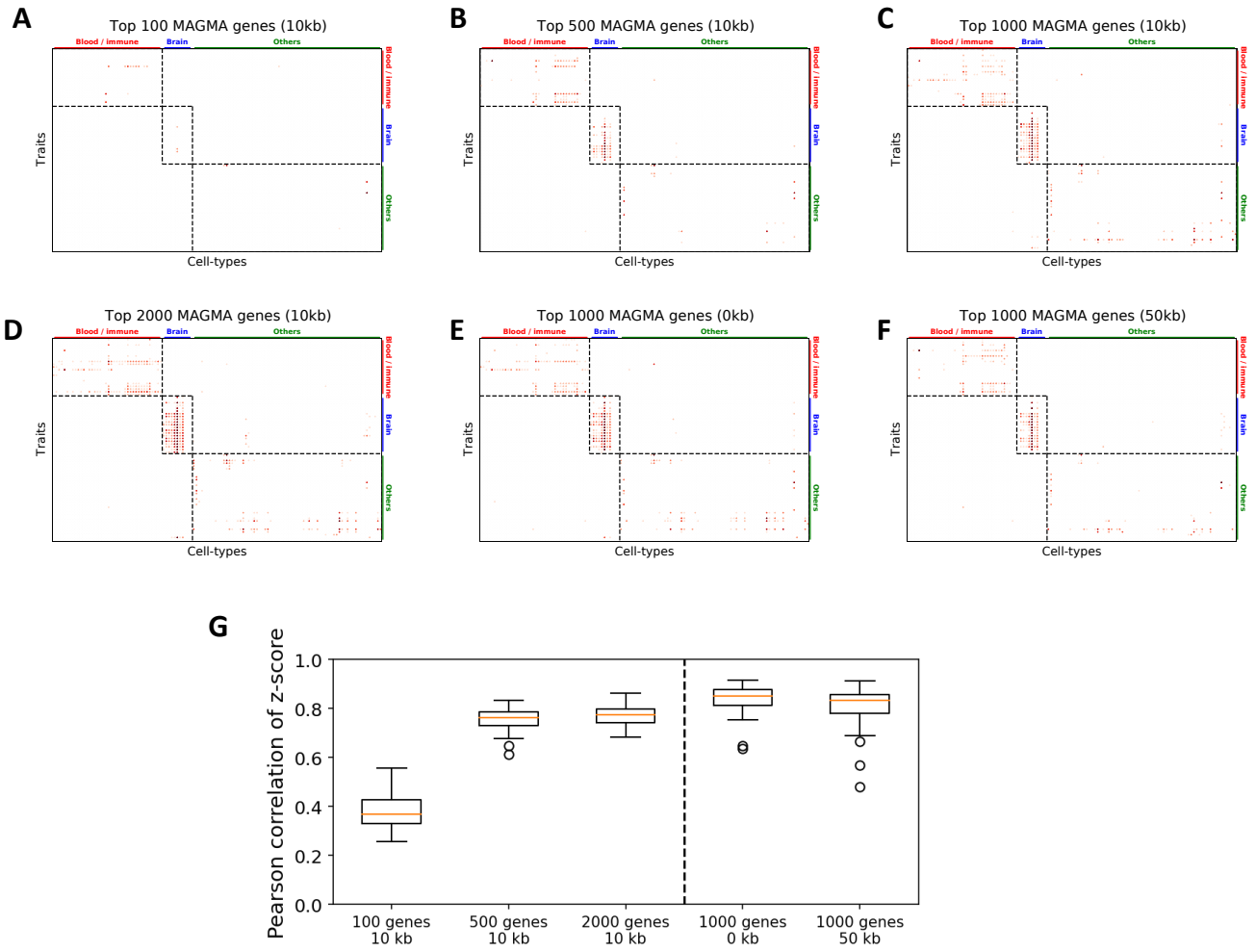
Supplementary Figure 8. Comparison of cell type-level disease association results between TMS FACS and TMS droplet (different technology), TS FACS (different species). (A-C) Results for disease association at the cell type-level for TMS FACS, TMS droplet, and TS FACS for diseases and cell types in the blood/immune block (upper left) and the other cell types/diseases block (lower right) in Fig. 3 (TMS droplet and TS FACS do not contain brain data). The plotting style is same as Fig. 3. Heatmap colors for each cell type-disease pair denote the proportion of significantly associated cells (FDR<0.1); squares denote significant cell type-disease associations (FDR<0.05); and cross symbols denote significant heterogeneity in association with disease across individual cells within a given cell type (FDR<0.05). Heatmap colors (>10% of cells associated) and cross symbols are omitted for cell type-disease pairs with non-significant cell type-disease associations via MC test. We matched each TMS FACS cell type using the closest cell type in the TMS droplet and TS FACS data; unmatched cell types were colored in grey. **(D)** Overlap of significant cell type-disease associations between TMS FACS and TMS droplet ($P = 1.7 \times 10^{-26}$, Fisher's exact test). **(E)** Correlation of $-\log_{10}$ p-values for cell type-disease associations between TMS FACS and TMS droplet. **(F)** Overlap of significant cell type-disease associations between TMS FACS and TS FACS ($P = 7.5 \times 10^{-7}$, Fisher's exact test). **(G)** Correlation of $-\log_{10}$ p-values for cell type-disease associations between TMS FACS and TS FACS. We determined that the results are highly consistent between TMS FACS and TMS droplet, and are reasonably consistent between TMS FACS and TS FACS. Our method is underpowered in the TS FACS data, possibly due to the smaller sample size (27K cells in TS FACS vs. 110K cells in TMS FACS). The current TS FACS data corresponds to the initial data release and there will likely be more cells in future releases[37].



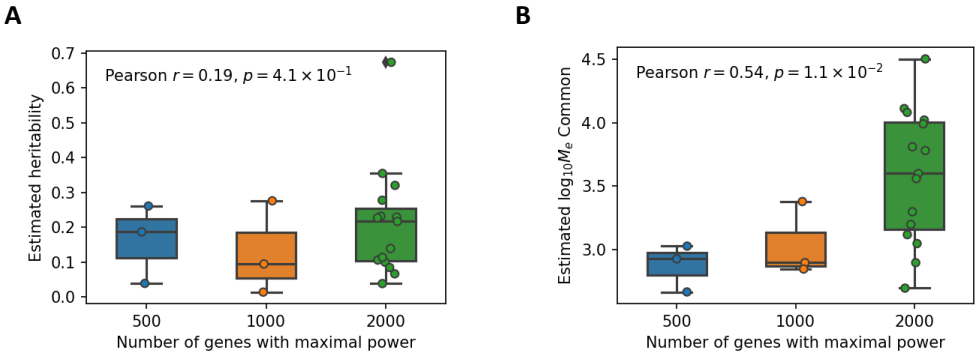
Supplementary Figure 9. Comparison of cell type-disease associations to LDSC-SEG. We used LDSC-SEG to analyze the same 120 TMS FACS cell types and 74 diseases and compared the cell type-disease associations identified by the two methods. We considered two versions of `scDRS` cell type-disease association discoveries: (1) cell type-disease pairs with more than 10% of individual cells associated to disease (non-white color in Supp. Fig. 4), (2) cell type-disease pairs with FDR<0.05 for cell type-disease association via MC test (squares in Supp. Fig. 4). For LDSC-SEG, we performed one-versus-rest differential expression analysis ("rank_genes_groups" with option "t-test_overestim_var" in scanpy[55]) to identify specifically-expressed genes for each of the 120 cell types (top 1,000 genes to be consistent with `scDRS`). We used 100-kb windows around the gene body to map genes to variants (default setting in LDSC-SEG) and applied LDSC-SEG[20] conditional on the 52 baseline annotations (baseline v1.2) to identify disease-relevant cell types. We computed FDR across all pairs of cell types and diseases/traits and used a significance threshold of 0.05. **(A)** Venn diagram between the `scDRS` discoveries (based on `scDRS` individual cell-level associations) and LDSC-SEG discoveries ($P = 1.4 \times 10^{-162}$, Fisher's exact test). **(B)** Venn diagram between the `scDRS` discoveries (based on `scDRS` cell type-level associations via MC test) and LDSC-SEG discoveries ($P = 8.7 \times 10^{-295}$, Fisher's exact test).



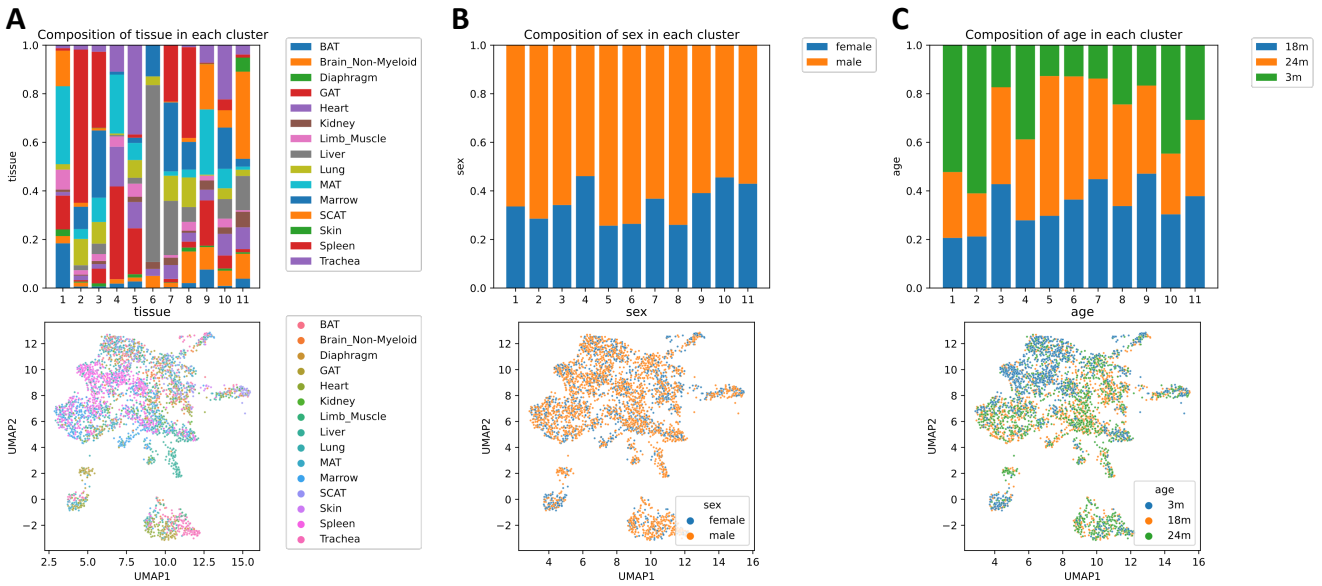
Supplementary Figure 10. Comparison of `scDRS` to alternative cell-level scores. We considered other cell-level scores, including the unweighted average and the overdispersion score, which tests for both overexpression and underexpression of the disease genes in the relevant cell population (Methods). Each dot corresponds to one of the 74 diseases, the x-axis denotes the number of associated cells (FDR<0.1) using the default weighted average score (used in `scDRS`) and the y-axis denotes results using the alternative cell-level scores. For each of these comparisons, we also computed the Pearson's correlation of the $-\log_{10}$ p-values across cells for each trait, and report the average correlation across traits (along with SD). We determined that the default weighted average score attained higher power while being consistent with the alternative cell-level scores.



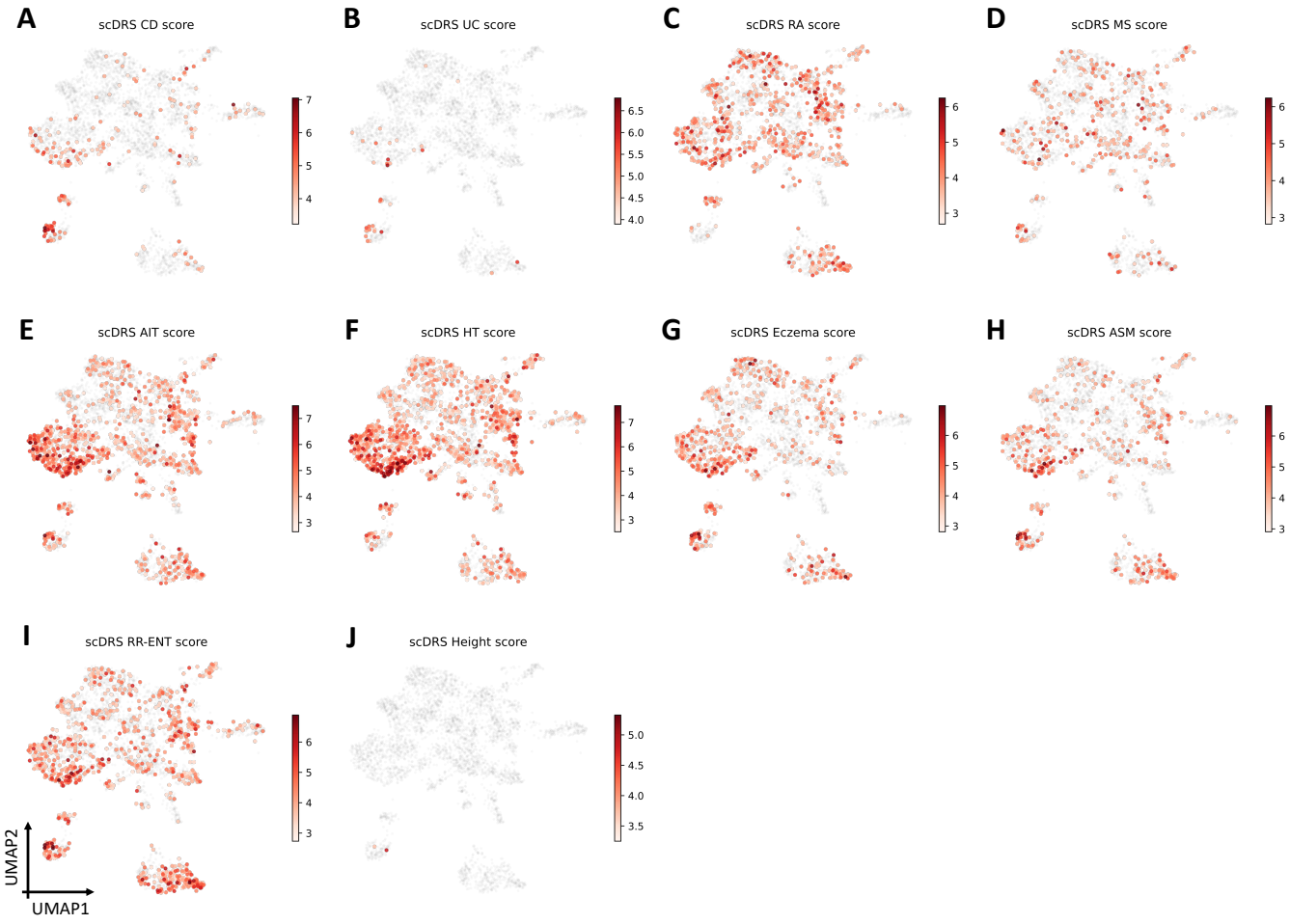
Supplementary Figure 11. Comparison of scDRS results for different parameter settings. (A-F) scDRS disease associations at the cell type level for the 120 cell types in TMS FACS data and 74 diseases/traits using different numbers of putative disease genes (100, 500, 1,000, 2,000 while fixing the 10-kb MAGMA window) and using different MAGMA window sizes (0 kb, 10 kb, 50 kb while fixing the 1,000 putative disease genes). For each heatmap, each row represents a disease/trait and each column represents a cell type. Heatmap colors for each cell type-disease pair denote the proportion of significantly associated cells (FDR<0.1 across all cells for a given disease). Diseases/traits and cell types are in the same order as in Supp. Fig. 4. **(G)** Pearson's correlation of the scDRS normalized disease scores across cells between the default parameter setting (1,000 genes, 10-kb window, panel C) and other parameter settings. Each boxplot represents the distribution of the Pearson's correlations across 74 diseases/traits.



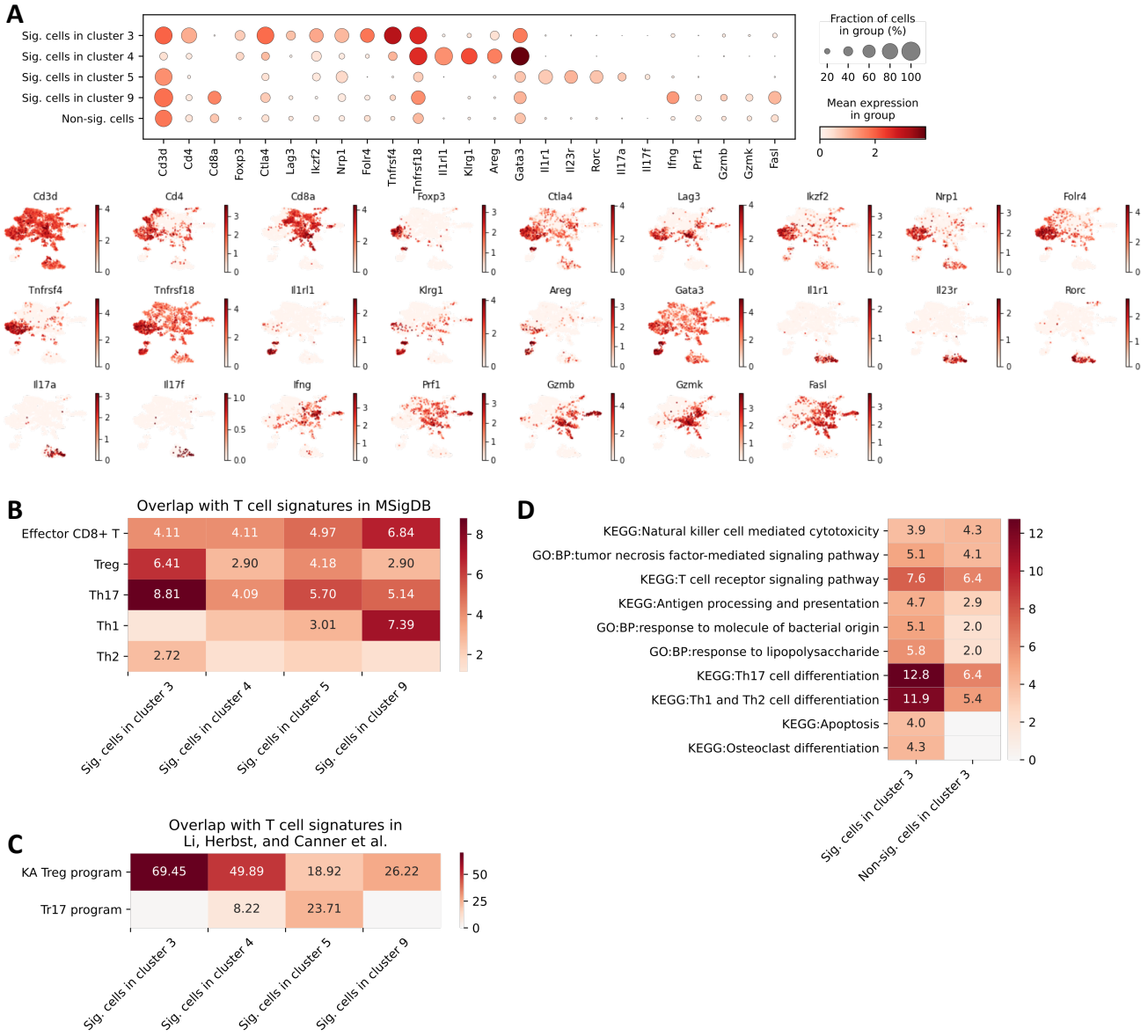
Supplementary Figure 12. Relationship between optimal size of disease gene set with heritability and polygenicity across diseases/traits. **(A)** Heritability. **(B)** Polygenicity as measured by $\log_{10} M_e$ in O'Connor et al.[33] for common SNPs. Each dot corresponds to a trait, the x-axis denotes the optimal size of putative disease genes for the trait and the y-axis denotes heritability/polygenicity. Pearson's correlation and the corresponding p-value are based on correlating the optimal size of disease gene set with heritability/polygenicity across diseases/traits.



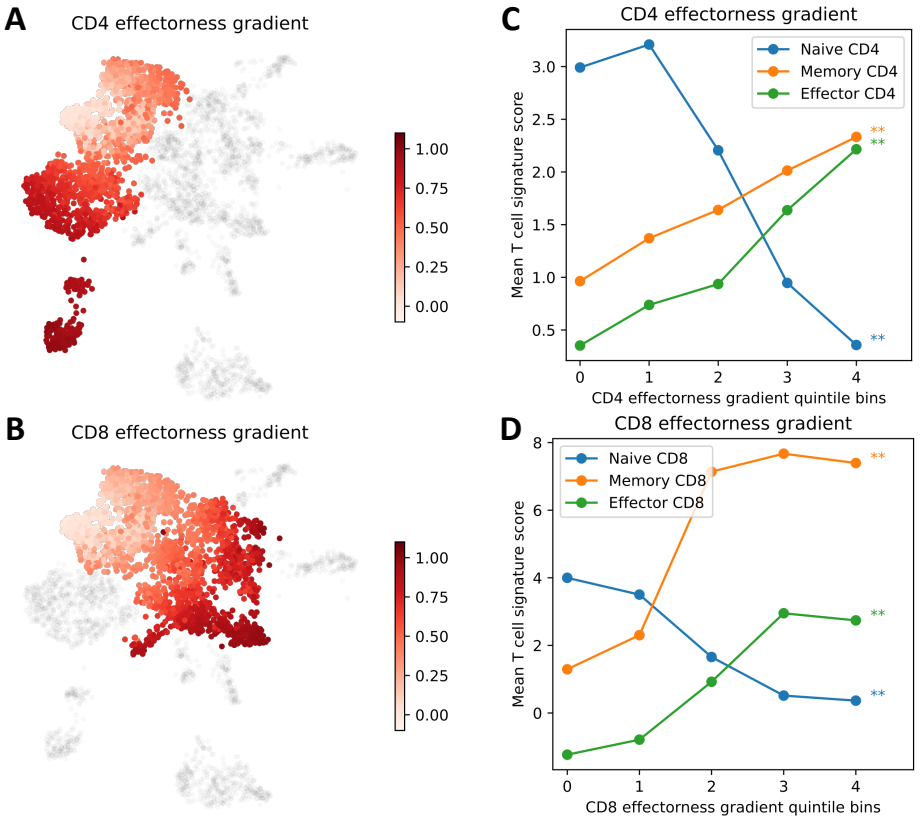
Supplementary Figure 13. Covariate composition for clusters in Fig. 4A and the corresponding UMAP visualization. **(A)** Tissue. **(B)** Sex. **(C)** Age.



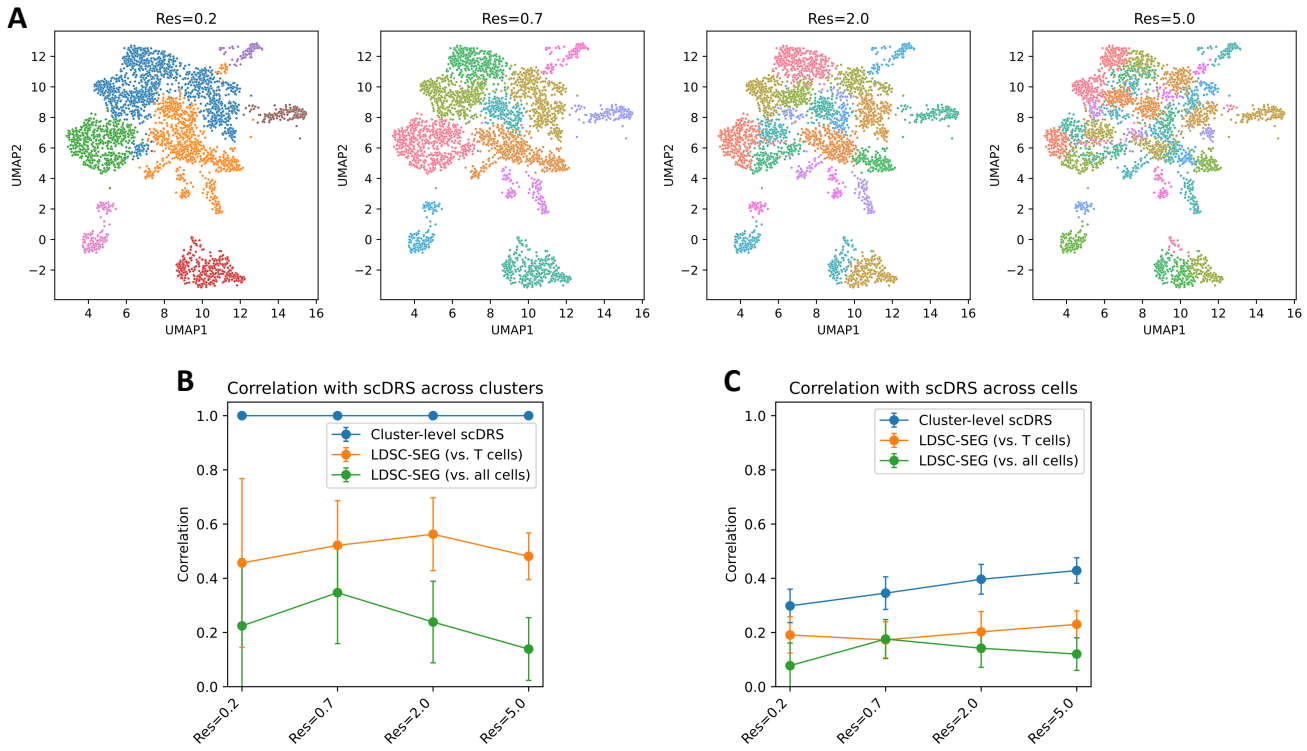
Supplementary Figure 14. Subpopulations of TMS FACS T cells associated with the other 9 autoimmune diseases (besides IBD reported in Fig. 4B) and height, a negative control trait. Significantly associated cells (FDR<0.1) were colored in red with shades of red representing the `scDRS` disease score; non-significant cells were colored in grey.



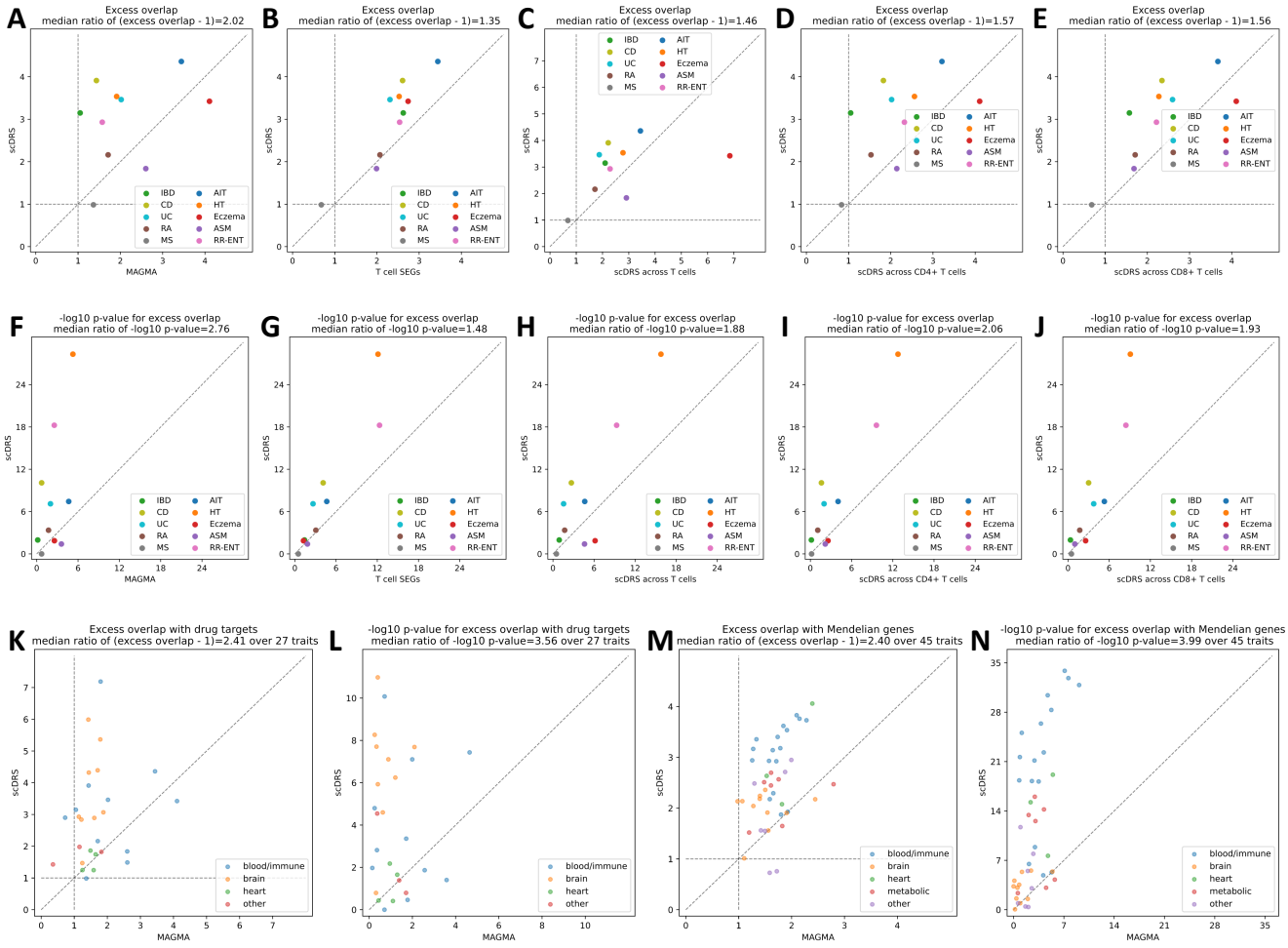
Supplementary Figure 15. Additional results on subpopulations of T cells associated with IBD. **(A)** Expression of marker genes in subpopulations of T cells associated with IBD, including general T cell markers (*CD3D*, *CD4*, *CD8A*), Treg markers (*FOXP3*, *CTLA4*, *LAG3*, *IKZF2* (*Helios*), *NRP1* (*Neuropilin-1*), *FLOR4*, *TNFRSF4* (*OX40*), *TNFRSF18* (*GITR*)), signatures of a *KLRG1*$^{+}$ *AREG*$^{+}$ effector-like Treg program[56] (*IL1RL1* (*ST2*), *KLRG1*, *AREG1*, *GATA3*), Th17 markers (*IL1R1*, *IL23R*, *RORC*, *IL17A*, *IL17F*), and effector CD8^{+} T cell signatures (*IFNG*, *PRF1*, *GZMB*, *GZMK*, *FASL*). **(B)** Overlap between T cell signatures and the top 300 specifically expressed genes for each IBD-associated T cell subpopulation (Supp. Table 8; Methods). The color and number in each cell represent the $-\log_{10}$ enrichment p-values (Fisher's exact test). **(C)** Overlap between Treg programs[56] (KA Treg: *KLRG1*$^{+}$ *AREG*$^{+}$ effector-like Treg program; Tr17: Th17-like Treg program; Supp. Table 8) and the top 300 specifically expressed genes of each IBD-associated T cell subpopulation. The color and number in each cell represent the $-\log_{10}$ enrichment p-values (Fisher's exact test). **(D)** Pathways enriched in the top 300 specifically expressed genes of the IBD-associated cells and non-associated cells in cluster 3 respectively. The color and number in each cell represent the $-\log_{10}$ enrichment p-values (based on Enrichr implemented in GSEAPY[57,58]).



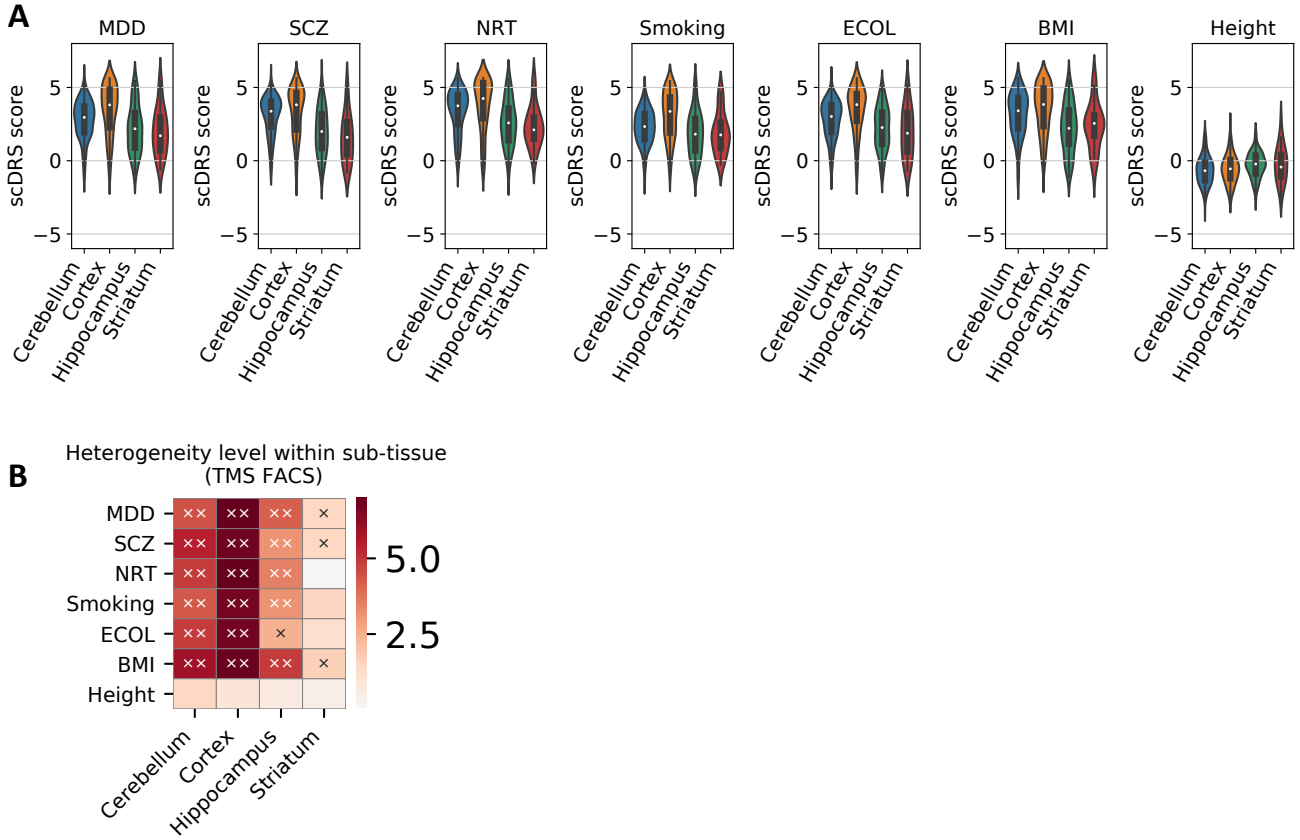
Supplementary Figure 16. Results for CD4 and CD8 effectorness gradients. **(A)** UMAP visualization of CD4 effectorness gradient across CD4$^+$ T cells. The effectorness gradient was represented in the red color and non-CD4$^+$ T cells were colored in grey. **(B)** UMAP visualization of CD8 effectorness gradient across CD8$^+$ T cells. The effectorness gradient was represented in the red color and non-CD8$^+$ T cells were colored in grey. **(C)** Correlation across CD4$^+$ T cells between CD4 effectorness gradient and signatures for naive, memory, and effector CD4$^+$ T cells (`scDRS` disease scores for applying these signature gene sets to the TMS FACS data). The x-axis denotes CD4 effectorness gradient quintile bins and the y-axis denotes average `scDRS` disease score for each bin and each signature gene set. * denotes P <0.05 and ** denotes P <0.005. **(D)** Correlation across CD8$^+$ T cells between CD8 effectorness gradient and signatures for naive, memory, and effector CD8$^+$ T cells. The x-axis denotes CD8 effectorness gradient quintile bins and the y-axis denotes average `scDRS` disease score for each bin and each signature gene set. * denotes P <0.05 and ** denotes P <0.005.



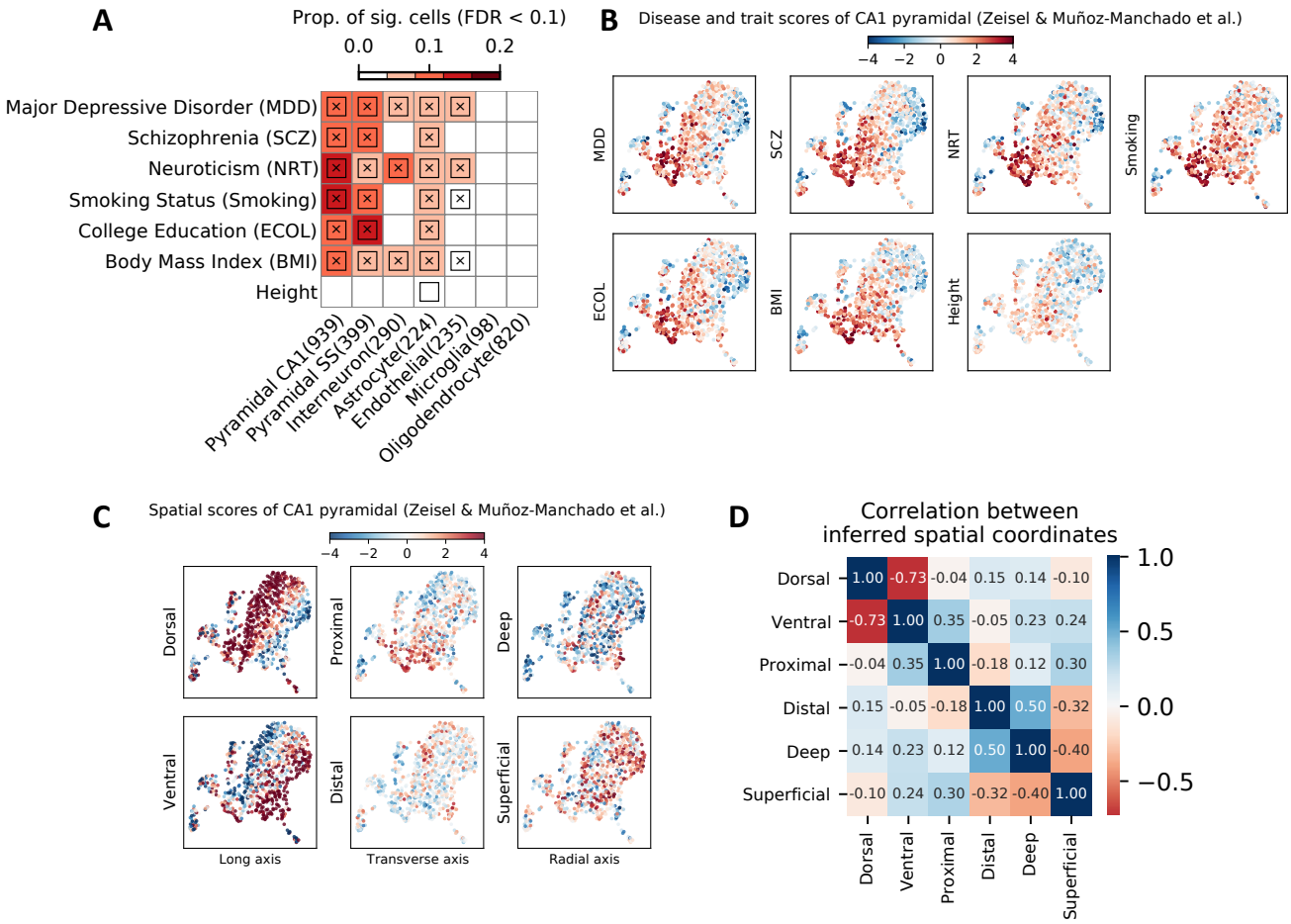
Supplementary Figure 17. Comparison of individual cell level T cell results to cluster-level analyses. We reclustered the set of T cells with different cluster resolutions (0.2, 0.7, 2, 5; 0.7 corresponds to Fig. 4A), followed by performing LDSC-SEG analysis on specifically expressed genes (SEGs) for each cluster and each of the 10 autoimmune diseases; we considered two ways for computing SEGs for a given cluster: by comparing to other T cells (vs. T cells) or to all other cells in the TMS FACS data (vs. all cells). We considered a third baseline that obtains $-\log_{10}$ p-values for a given cluster by averaging the `scDRS` log10 p-values of cells within the cluster (cluster-level `scDRS`). We note that the gap between `scDRS` and cluster-level `scDRS` is due to finite clustering resolution, while the gap between cluster-level `scDRS` and LDSC-SEG is due to the difference between `scDRS` and LDSC-SEG for capturing cluster-level disease heritability enrichment. **(A)** T cells clustered at different resolutions. **(B)** Correlation of $-\log_{10}$ p-value *across clusters* between `scDRS` and the 3 comparison methods. For `scDRS`, the cluster-level $-\log_{10}$ p-values were obtained by averaging the $-\log_{10}$ p-value of cells within the same cluster (identical to cluster-level `scDRS`). The three methods were highly correlated with `scDRS` at cluster level, suggesting LDSC-SEG and `scDRS` produced similar results. **(C)** Correlation of $-\log_{10}$ p-value *across cells* between `scDRS` and the 3 comparison methods. For the 3 comparison methods, cell-level $-\log_{10}$ p-values were obtained by assigning the same cluster-level $-\log_{10}$ p-value to all cells within the cluster. The 3 methods were less correlated with `scDRS` at individual cell level, suggesting cluster-level analyses were not able to capture the individual cell-disease associations detected in the `scDRS` individual cell-level analysis.



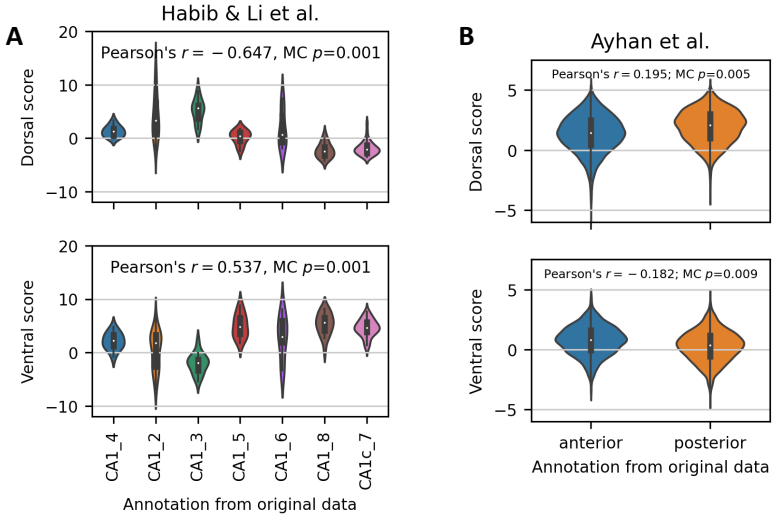
Supplementary Figure 18. Additional results on disease gene prioritization. (A-J) Comparison to alternative disease gene prioritization methods for the 10 autoimmune diseases. The first row shows levels of excess overlap between the prioritized disease genes and the gold standard gene sets while the second row shows the corresponding $-\log_{10}$ p-values for excess overlap. Each dot corresponds to a disease, the y-axis shows results for the proposed prioritization method (correlating gene expression levels with the `scDRS` disease score across all TMS FACS cells), and the x-axis shows results from comparison methods, including (from left to right) top 1,000 MAGMA genes, top 1,000 genes specifically expressed in T cells (vs. the rest of cells in TMS FACS), prioritization based on correlation across T cells (instead of all TMS FACS cells), prioritization based on correlation across $CD4^+$ T cells (instead of all TMS FACS cells), and prioritization based on correlation across $CD8^+$ T cells (instead of all TMS FACS cells). **(K-L)** Overlap with drug target genes for 27 diseases. **(M-N)** Overlap with Mendelian disease genes for 45 diseases. The median ratio of $-\log_{10}$ p-values and (excess overlap $-$ 1) between the y- and x-values (median of ratios) was provided in the figure title.



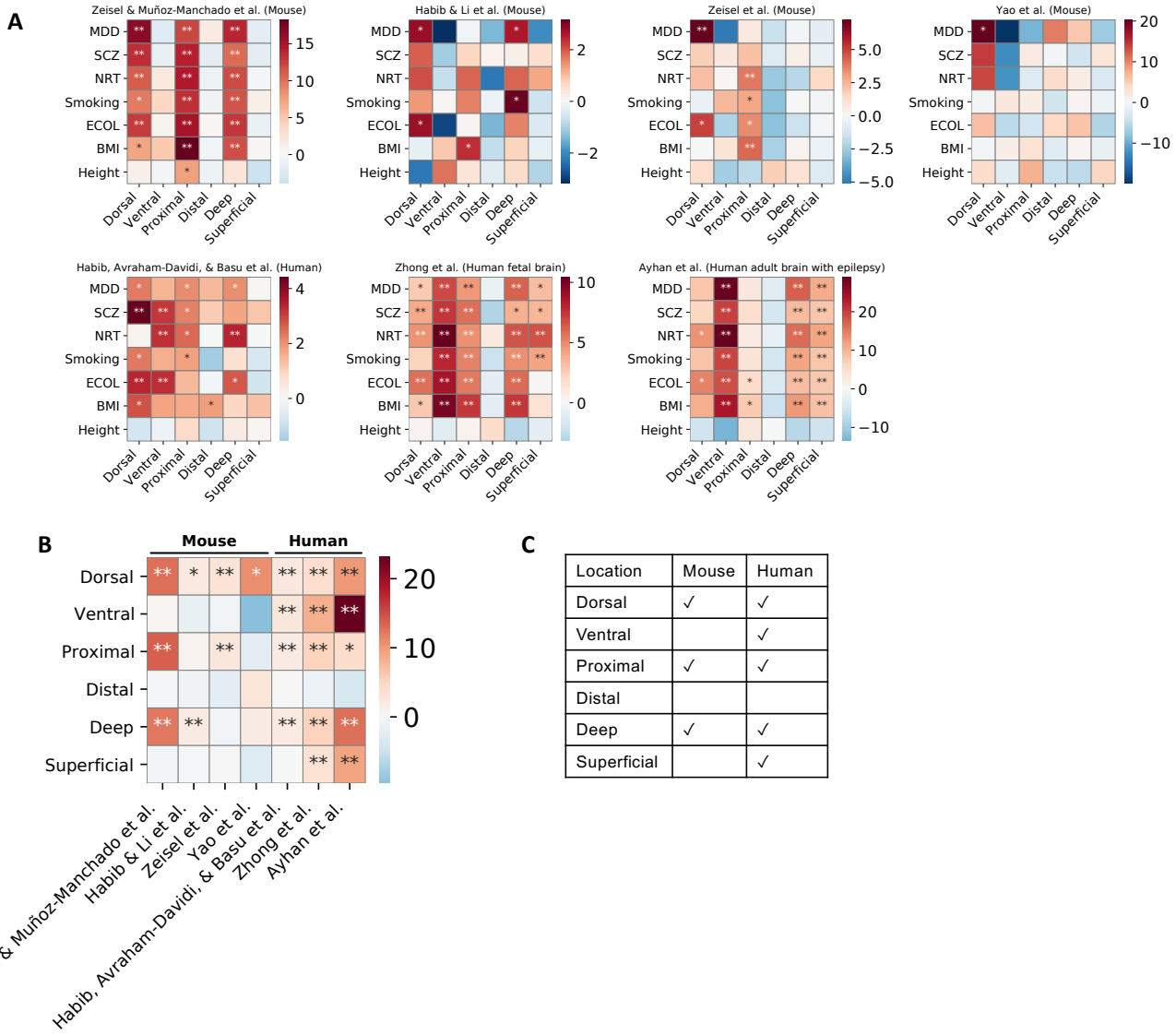
Supplementary Figure 19. Associations of TMS FACS neurons with brain-related diseases/traits. **(A)** Violin plots of the `scDRS` disease score for TMS FACS brain neurons (undetermined neurons, excluding interneurons and MSNs) in different brain subtissues and for different diseases/traits. Height was included as a negative control. **(B)** Within-subtissue heterogeneity of neurons in association with different diseases/traits. Heatmap colors represent the heterogeneity z-score (MC z-score) and cross symbols represent significant within-subtissue disease association heterogeneity ($\times$ denotes $P < 0.05$ and $\times\times$ denotes $P < 0.005$, MC test).



Supplementary Figure 20. Additional results for the Zeisel & Muñoz-Manchado et al. data on associations to brain-related diseases/traits. (A) Disease associations at the cell type level. Each row represents a disease/trait and each column represents a cell type. Heatmap colors for each cell type-disease pair denote the proportion of significantly associated cells (FDR<0.1 across all cells for a given disease). Squares denote significant cell type-disease associations (FDR<0.05 across all pairs of 7 cell types and 7 diseases/traits). Cross symbols denote significant heterogeneity in association with disease across individual cells within a given cell type (FDR<0.05 across all pairs). Heatmap colors and cross symbols are omitted for cell type-disease pairs with non-significant cell type-disease associations; the plotting style is the same as in Fig. 3. **(B)** UMAP visualization of the CA1 pyramidal neurons in the Zeisel & Muñoz-Manchado et al. data with color representing the `scDRS` disease score (extending results in Fig. 5A). Height was included as a negative control trait. **(C)** UMAP visualization of the CA1 pyramidal neurons with color representing the inferred spatial coordinates. **(D)** Pairwise correlations across cells for the 6 inferred spatial coordinates in the Zeisel & Muñoz-Manchado et al. data set.

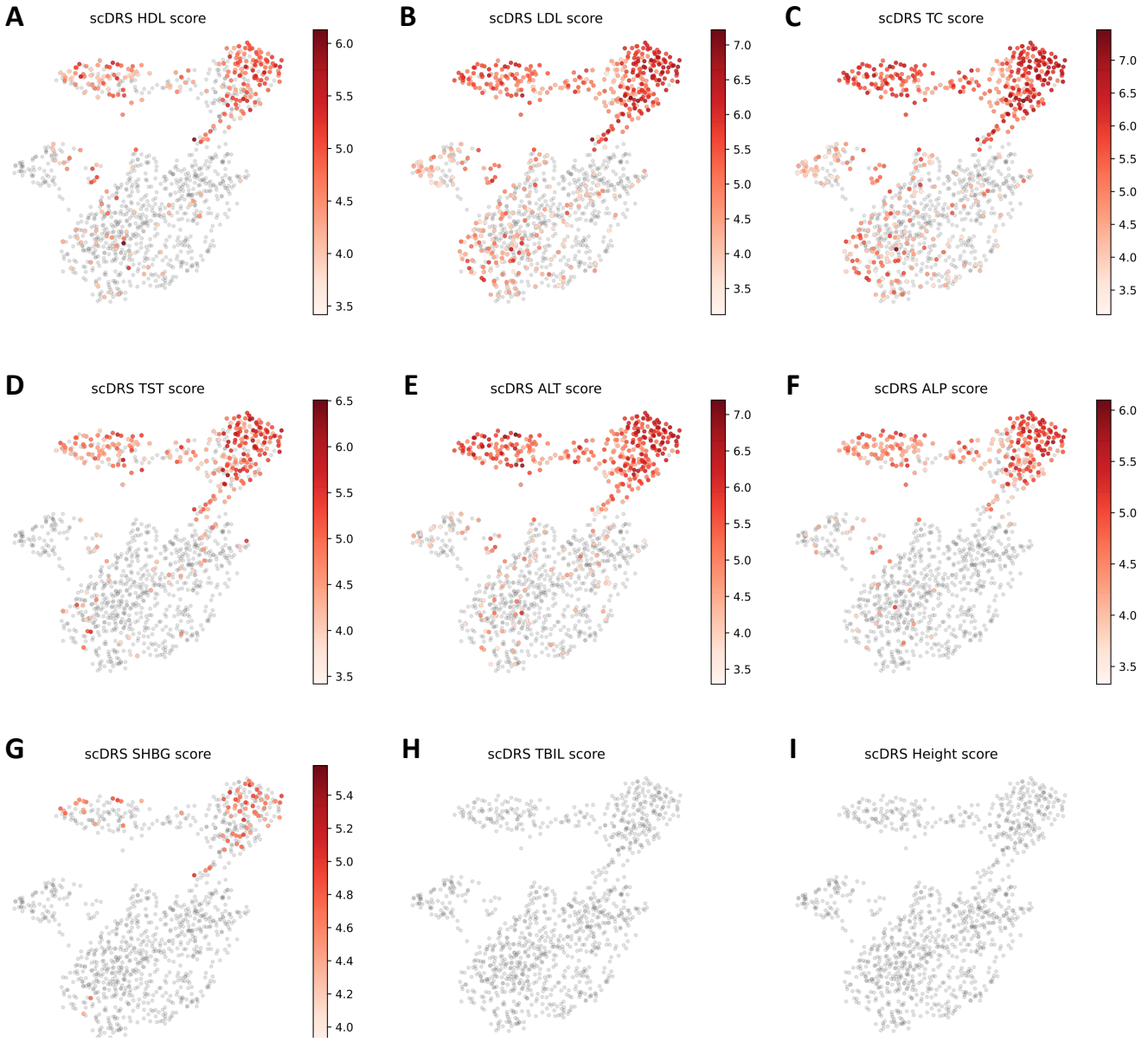


Supplementary Figure 21. Validation of inferred spatial coordinates of CA1 pyramidal neurons using data sets with annotated spatial annotations. Violin plots of the inferred cell-level spatial coordinates (y-axis) against the provided spatial annotations (x-axis). **(A)** Habib & Li et al. data. The x-axis labels were ordered from dorsal to ventral neurons according to the provided spatial annotation (Fig. 2B in Habib & Li et al.[42]). Both the dorsal and ventral scores are significantly associated with the provided spatial annotation (P<0.001, MC test based on Pearson's correlation between the provided spatial annotation (ordinal ranks of the 7 categories) and the inferred spatial score). **(B)** Ayhan et al. data. Anterior corresponds to the ventral region while posterior corresponds to the dorsal region. Both the dorsal and ventral scores are significantly associated with the spatial annotation (P<0.01, MC test based on Pearson's correlation between provided spatial annotation (ordinal ranks of the 2 categories) and the inferred spatial score).

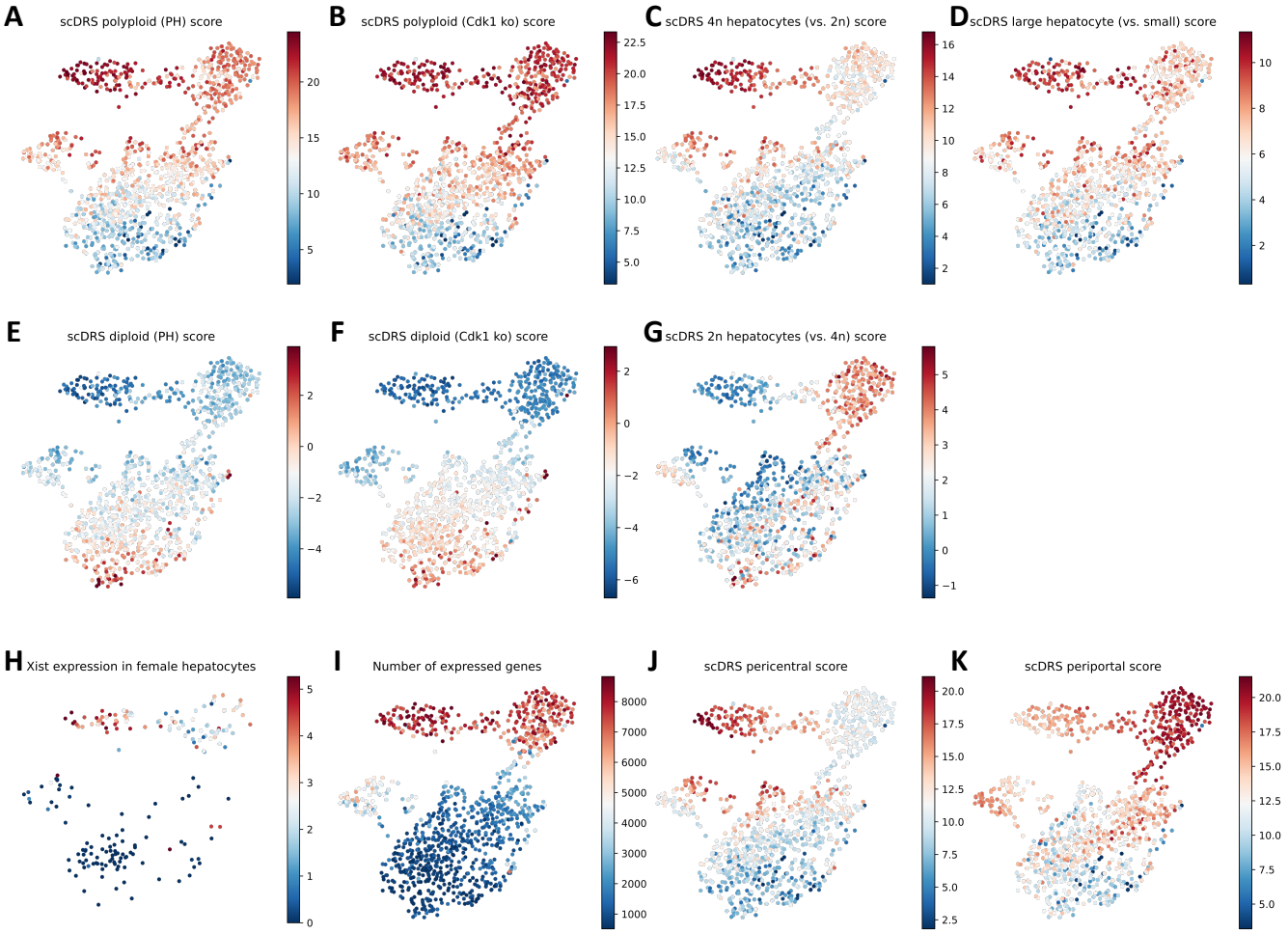


Location	Mouse	Human
Dorsal	✓	✓
Ventral		✓
Proximal	✓	✓
Distal		
Deep	✓	✓
Superficial		✓

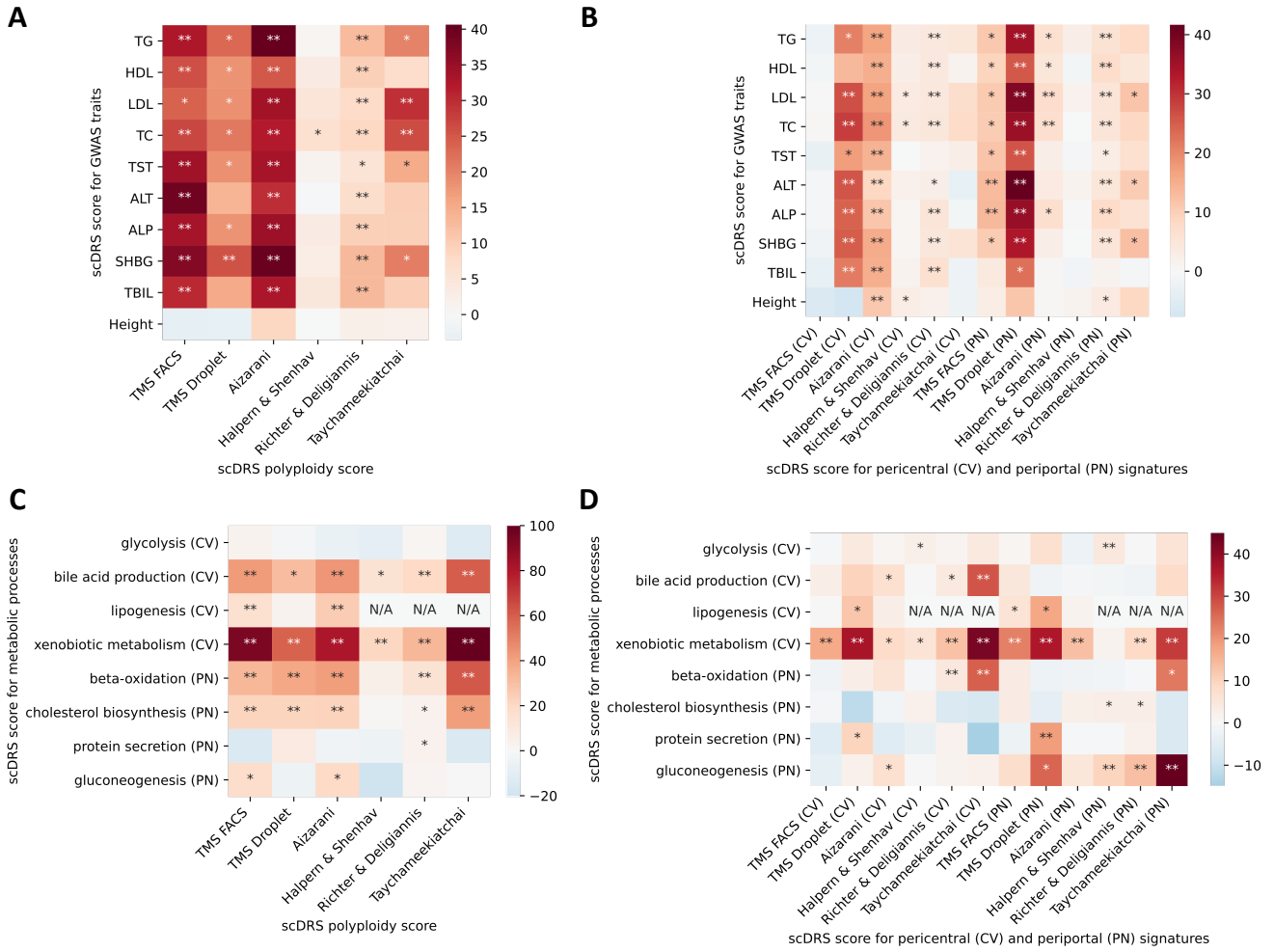
Supplementary Figure 22. Complete results of correlations between scDRS disease scores and inferred spatial coordinates across CA1 pyramidal neurons in 7 single-cell data sets (extending results in Fig. 5B). (A) Results for regressing the scDRS disease scores against the inferred spatial coordinates for each disease/trait and each inferred spatial coordinate. Color represents the t-statistics and stars represent significant associations (* denotes $P < 0.05$ and ** indicates $P < 0.005$, MC test; Methods). For clarification, Zeisel & Muñoz-Manchado et al. refers to the data from Zeisel & Muñoz-Manchado et al. 2015 *Science*[40] and Zeisel et al. refers to the data from Zeisel et al. 2018 *Cell*[41]. **(B)** Summary of results in panel A. Heatmap color represent the average t-statistics across the 6 brain-related diseases/traits (excluding height) for each data set and stars represent significant associations by combining p-values across datasets using Fisher's combined probability test. **(C)** Summary of the association between brain-related diseases and the inferred spatial coordinates for the mouse and human data sets in panel B.



Supplementary Figure 23. Subpopulations of TMS FACS hepatocytes associated with the other 8 metabolic diseases (besides TG reported in Fig. 5C) and height. Significantly associated cells (FDR<0.1) were colored in red with shades of red representing the `scDRS` disease score; non-significant cells were colored in grey. The color bar was removed for traits without a significant cell.



Supplementary Figure 24. Signature scores for hepatocyte ploidy level and zonation. (A-D) `scDRS` score for polyploid hepatocyte signatures: partial hepatectomy (PH) vs. pre-PH, Cdk1 knockout vs. control, 4n vs. 2n hepatocytes, large vs. small hepatocytes. **(E-G)** `scDRS` score for diploid hepatocyte signatures: pre-PH vs. PH, control vs. Cdk1 knockout, 2n vs. 4n hepatocytes. **(H)** Expression of *Xist* in female hepatocytes (expected to have high expression in high-ploidy female hepatocytes). **(I)** Number of genes (expected to be high in high-ploidy hepatocytes). **(J)** `scDRS` score for pericentral hepatocyte signatures. **(K)** `scDRS` score for periportal hepatocyte signatures.



Supplementary Figure 25. Complete results of joint regression analysis for GWAS metabolic traits and putative zonated metablic processes across the 6 data sets (extending results in Fig. 5D). (A-B) Results for the 9 metabolic traits and height, a negative control triat. The polyploidy score (panel A) and both the pericentral and periportal score (panel B) were consistently associated with the 9 metabolic traits across the data sets. The strong association (P <0.005) between the pericentral score and height in the Aizarani et al. data may be because that we inferred the pericentral score using mouse gene signatures, which are less conserved in human (as also mentioned in the original paper[47]). **(C-D)** Results for the 8 metabolic pathways. Overall, as shown in panel D, the pericentral score was associated with pericentral-specific pathways (first 4 rows) while the periportal score was associated with periportal-specific pathways (last 4 rows). * denotes P <0.05 and ** denotes P <0.005.