Title
Development and Application of Single Cell Multi-omics Methods for Complex Disease

Author
Littman, Russell

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Development and Application of Single Cell Multi-omics Methods for Complex Disease

A dissertation submitted in partial satisfaction of

the requirements for the degree Doctor of Philosophy in

Bioinformatics

by

Russell Littman

2022

ABSTRACT OF THE DISSERTATION

Development and Application of

Single Cell Multi-omics

Methods for Complex Disease

by

Russell Littman

Doctor of Philosophy in Bioinformatics

University of California, Los Angeles, 2022

Professor Xia Yang, Chair

Complex diseases such as Alzheimer's disease are driven by molecular changes in many cell types in different tissues. Recent advances in scRNAseq and spatial transcriptomics provide tools to determine cell type specific effects in individual tissues resulting from genetic and environmental perturbations. Properly interpreting these data require computational tools and biologically rooted analyses to identify key mechanisms underlying complex diseases. Here we design, develop, and apply computational methods for integrating scRNAseq and spatial transcriptomics data to identify mechanisms underlying pathogenesis of disease and potential therapeutics. First, we designed a deep learning approach, JSTA, for integrating scRNAseq and spatial transcriptome data from multiplexed FISH for cell segmentation and cell type annotation, revealing spatially distributed cell subtypes and spatially differentially expressed genes in the mouse hippocampus. Next, we developed a gradient-boosting machine based approach, SCING, for identifying cell type specific gene regulatory networks (GRN) using scRNAseq and spatial transcriptomics data. This tool provides GRN subnetworks annotated with biological pathways for

associating subnetwork expression with disease phenotypes and spatial domains. We applied these and other existing tools to scRNAseq and spatial transcriptomics datasets to understand the mechanism underlying diverse types of diseases or physiological traits, including Alzheimer's disease and heart innervating neurons and satellite glial cells in the stellate ganglion in the context of dilated cardiomyopathy. Our studies established new computational tools applicable to diverse types of single cell omics data and revealed biological insights to complex diseases.

DEDICATION

This dissertation is dedicated to my family and friends, who have made the time during my PhD a very fun experience.

The dissertation of Russell Littman is approved.

Weizhe Hong

Jingyi Li

Roy Wollman

Xia Yang, Committee Chair

University of California, Los Angeles

2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENT

My dissertation work would not be possible without the endless support of my advisor, Professor Xia Yang. She has gone above and beyond to continually support my scientific ideas. I want to particularly highlight her ability to keep computational methods development grounded in biological application, as well as her endless care for her students' wellbeing. She is very supportive of any career aspirations and goals and works very hard to support her students.t I would like to thank my committee members, Weizhe Hong, Jessica Li, and Roy Wollman, who I have continuously learned from throughout my PhD work for advice on projects as well as career goals. Throughout my time at UCLA I have worked with many incredible collaborators, graduate, and undergraduate students who have shaped my research and personal development. Finally, this journey would not have been possible without the support of my family and friends who have made the time outside of my PhD work truly amazing.

Chapter 2 is a version of Littman, Russell, Zachary Hemminger, Robert Foreman, Douglas Arneson, Guanglin Zhang, Fernando Gómez-Pinilla, Xia Yang[&], and Roy Wollman[&]. 2021. "Joint Cell Segmentation and Cell Type Annotation for Spatial Transcriptomics." https://doi.org/10.15252/msb.202010108. Molecular Systems Biology 17 (6): e10108. This work was supported by NIH grants R01NS117148, T32CA201160.

Chapter 3 is a version of our preprint Littman, Russell, Ning Wang, Chao Peng, and Xia Yang. "SCING: Single Cell INtegrative Gene Regulatory Network Inference Elucidates Robust, Interpretable Gene Regulatory Networks." https://doi.org/10.1101/2022.09.07.506959. This work was funded by NIH grants R01NS117148, T32CA201160, T32HL13945001, UCLA DYF.

Chapter 4 is a version of Gao, Jie, Russell Littman, Graciel Diamante, Xu Xiao, In Sook Ahn, Xia Yang, Tracy A. Cole, and Peter Tontonoz. 2020. "Therapeutic IDOL Reduction Ameliorates

Amyloidosis and Improves Cognitive Function in APP/PS1 Mice." Molecular and Cellular Biology 40 (8). https://doi.org/10.1128/MCB.00518-19. This research was supported by NIH grants R00AG054736 (to J.G.). R.L. is supported by NIH-NCI National Cancer Institute grant T32CA201160. X.Y. is supported by NIH grants R01 DK104363 and R21 NS103088.

Chapter 5 is a version of unpublished work Sharma, Sachin*, Littman, Russell*, John Tompkins, Douglas Arneson, Jaime Contreras, Al-Hassan Dajani, Kaitlyn Ang, Amit Tsanhani, Patrick Y. Jay, Herbert Herzog, Xia Yang, and Olujimi A. Ajijola. "Defined Subsets of Sympathetic Neurons Innervate the Heart and Differentially Respond to Cardiac Pathology" The authors wish to thank Timothy Davis, Zhiwei Li, Amiksha Gandhi MS, and Pratap Meera PhD for technical assistance; One Legacy for human stellate ganglia; and Felix Schweizer PhD, Emilie Marcus PhD, Xin Sun PhD, and Kalyanam Shivkumar MD PhD for helpful comments on the studies and manuscript. This study is supported by NIH OD DP2HL142045 to OAA, and T32CA201160, T32HL13945001 for RL.

Chapter 6 is a version of Weperen, Valerie Y. H. van, Russell J. Littman, Douglas V. Arneson, Jaime Contreras, Xia Yang, and Olujimi A. Ajijola. 2021. "Single-Cell Transcriptomic Profiling of Satellite Glial Cells in Stellate Ganglia Reveals Developmental and Functional Axial Dynamics." https://doi.org/10.1002/glia.23965. Glia 69 (5): 1281–91. This work was supported by NIH grants DP2HL142045 and T32HL13945001.

VITA

EDUCATION

2022	PhD Candidate Bioinformatics
	University of California, Los Angeles
2018	Bachelor of Science, Biochemistry
	Minor in Bioinformatics
	University of California, Los Angeles

RESEARCH EXPERIENCE

2018-Present	PhD trainee
	Department of Integrative Biology and Physiology
	University of California, Los Angeles
2021	PHC Data Science Imaging Intern
	Genentech Product Development
	San Francisco, California

PEER-REVIEWED PUBLICATIONS (&co-corresponding author)

Littman, Russell, …, Xia Yang[&], and Roy Wollman[&]. 2021. "Joint Cell Segmentation and Cell Type Annotation for Spatial Transcriptomics." Molecular Systems Biology 17 (6): e10108.

Weperen, Valerie Y. H. van, **Russell J. Littman**, … , Olujimi A. Ajijola. 2021. "Single-Cell Transcriptomic Profiling of Satellite Glial Cells in Stellate Ganglia Reveals Developmental and Functional Axial Dynamics." Glia 69 (5): 1281–91.

Gao, Jie, **Russell Littman**, …, Peter Tontonoz. 2020. "Therapeutic IDOL Reduction Ameliorates Amyloidosis and Improves Cognitive Function in APP/PS1 Mice." Molecular and Cellular Biology 40 (8). https://doi.org/10.1128/MCB.00518-19.

SELECTED MANUSCRIPTS UNDER REVIEW OR IN REVISION (*co-first author)

Littman, Russell, … , Xia Yang. "SCING: Single Cell INtegrative Gene Regulatory Network Inference Elucidates Robust, Interpretable Gene Regulatory Networks." https://doi.org/10.1101/2022.09.07.506959.

Sharma, Sachin*, **Littman, Russell*,** …, Olujimi A. Ajijola. "Defined Subsets of Sympathetic Neurons Innervate the Heart and Differentially Respond to Cardiac Pathology"

SELECTED CONFERENCE PRESENTATIONS

"Leveraging Visium and scRNAseq for Integrative Analysis of Alzheimer's Disease" **10X webinar 2022**

"SCING: Single Cell INtegrative Gene regulatory network inference", oral presentation, **UCLA Quantitative and Computational Biology Retreat, Fall 2022**

"SCING: A Single Cell INtegrative Gene regulatory network inference algorithm elucidates robust, interpretable regulatory networks", poster presentation**, RECOMB 2022**

"fastGRN: Gene regulatory network inference from scRNAseq", oral presentation**, iDiscover Seminar Series, Winter 2022**

Joint Cell Segmentation and Cell Type Annotation for Spatial Transcriptomics, oral presentation, **Genentech AI/ML in Biology Seminar 2021**

Joint Cell Segmentation and Cell Type Annotation for Spatial Transcriptomics, oral presentation, **Genentech Imaging and Machine Learning Seminar 2021**

"Inter Cell Type Gene Regulatory Network Construction in Single Cell Data", oral presentation, **iDiscover Seminar Series, Spring 2021**

"Cell Segmentation in Spatial Transcriptomics", oral presentation**, UCLA Quantitative and Computational Biology Seminar, Spring 2020**

Chapter 1. Introduction

It is widely recognized that most common diseases such as Alzheimer's disease (AD) are complex and involve many different factors including both genetic and environmental changes[1,2]. These diseases require systems level multi-omic approaches to understand mechanisms underlying their pathophysiology to further therapeutic discovery[3]. As transcriptomics represents downstream changes of both genotype and environmental changes, studying the transcriptome can help reveal important biological insights. In recent years, single cell RNA sequencing (scRNAseq) has become a widespread and powerful tool for understanding cell type specific gene expression alterations underlying these diseases[4]. Spatial transcriptomics, a newer technology, elucidates spatial distribution of cells and genes throughout a tissue, informing on cell microenvironment, cell-cell interaction[5], cell type specific spatial variation[6], and cell type pathology interaction[7]. These data types require efficient, robust computational tools to integrate and extract underlying pathophysiological mechanisms of disease and to identify key drivers and potential targets.

My dissertation is comprised of two specific aims, one focusing on single cell multi-omics integration tool development (**Chapter 2, Chapter 3**) and the other focusing on application studies to understand cell-type specific mechanisms of IDOL knockdown treatment in APP/PS1 mice, as well as heart innervating neuronal subpopulations and satellite glia in the stellate ganglion in the context of DCM (**Chapter 4, Chapter 5, Chapter 6**).

In **Chapter 2** we develop JSTA, a computational framework for joint cell segmentation and cell type annotation that utilizes prior knowledge of cell type-specific gene expression. RNA hybridization-based spatial transcriptomics provides unparalleled detection sensitivity. However, inaccuracies in segmentation of image volumes into cells cause misassignment of mRNAs which is a major source of errors. Simulation results show that leveraging existing cell type taxonomy

increases RNA assignment accuracy by more than 45%. Using JSTA, we were able to classify cells in the mouse hippocampus into 133 (sub)types revealing the spatial organization of CA1, CA3, and Sst neuron subtypes. Analysis of within cell subtype spatial differential gene expression of 80 candidate genes identified 63 with statistically significant spatial differential gene expression across 61 (sub)types. Overall, our work demonstrates that known cell type expression patterns can be leveraged to improve the accuracy of RNA hybridization-based spatial transcriptomics while providing highly granular cell (sub)type information. The large number of newly discovered spatial gene expression patterns substantiates the need for accurate spatial transcriptomic measurements that can provide information beyond cell (sub)type labels.

In **Chapter 3**, we present Single Cell INtegrative Gene regulatory network inference (SCING), a gradient boosting and mutual information based approach for identifying robust GRNs from scRNAseq, snRNAseq, and spatial transcriptomics data. Gene regulatory network (GRN) inference is an integral part of understanding physiology and disease. Single cell/nuclei RNAseq (scRNAseq/snRNAseq) data has been used to elucidate cell-type GRNs; however, the accuracy and speed of current scRNAseq-based GRN approaches are suboptimal. Performance evaluation using held-out data, Perturb-seq datasets, and the mouse cell atlas combined with the DisGeNET database demonstrates the improved accuracy and biological interpretability of SCING compared to existing methods. We applied SCING to the entire mouse single cell atlas, human AD, and mouse AD spatial transcriptomics. SCING GRNs reveal unique disease subnetwork modeling capabilities, have intrinsic capacity to correct for batch effects, retrieve disease relevant genes and pathways, and are informative on spatial specificity of disease pathogenesis.

In **Chapter 4,** we utilized an antisense oligonucleotide (ASO) to reduce IDOL expression therapeutically in the brains of APP/PS1 male mice. Brain lipoprotein receptors have been shown

to regulate the metabolism of ApoE and β-amyloid (Aβ) and are potential therapeutic targets for AD. Previously, we identified E3 ubiquitin ligase IDOL as a negative regulator of brain lipoprotein receptors. Genetic ablation of *Idol* increases low-density lipoprotein receptor protein levels, which facilitates Aβ uptake and clearance by microglia. ASO treatment led to decreased Aβ pathology and improved spatial learning and memory. Single-cell transcriptomic analysis of hippocampus revealed that IDOL inhibition upregulated lysosomal/phagocytic genes in microglia. Furthermore, clustering of microglia revealed that IDOL-ASO treatment shifted the composition of the microglia population by increasing the prevalence of disease-associated microglia. Our results suggest that reducing IDOL expression in the adult brain promotes the phagocytic clearance of Aβ and ameliorates Aβ-dependent pathology. Pharmacological inhibition of IDOL activity in the brain may represent a therapeutic strategy for the treatment of AD.

In **Chapter 5,** we leveraged retrograde tracing techniques using adeno-associated virus (AAV) expressing fluorescent proteins (GFP or Td-tomato) with single cell RNA sequencing to identify and resolve the transcriptomic profiles of SGNs innervating the heart. The soma of postganglionic sympathetic neurons innervating the heart primarily reside in the stellate ganglion (SG), along with those of neurons innervating other organs and tissue beds. Whether cardiac-innervating stellate ganglionic neurons (SGNs) exhibit diversity and distinction from those innervating other tissues is not known. We investigated electrophysiologic, morphologic, and physiologic roles for subsets of cardiac-specific neurons and found that three of five adrenergic SGN subtypes innervate the heart, differentiated by NPY expression. We also report that these subtypes exhibit distinct morphological, neurochemical, and electrophysiologic characteristics, including differential roles in cardiac physiologic control. In a mouse model of non-ischemic heart failure, we demonstrate that cardiac disease differentially impacts these cardiac neuronal subtypes and alters their distribution to yield a dominant subtype, which represents a viable cell-specific anti-adrenergic target in heart failure. These findings provide novel insights into the unique properties

of neurons responsible for cardiac sympathetic regulation, with implications for novel strategies to target specific neuronal subtypes for sympathetic blockade in cardiac disease.

Stellate ganglion neurons, important mediators of cardiopulmonary neurotransmission, are surrounded by satellite glial cells (SGCs), which are essential for the function, maintenance, and development of neurons. However, it remains unknown whether SGCs in adult sympathetic ganglia exhibit any functional diversity, and what role this plays in modulating neurotransmission.

In **Chapter 6,** we performed single-cell RNA sequencing of mouse stellate ganglia (*n* = 8 animals), focusing on SGCs (*n* = 11,595 cells). SGCs were identified by high expression of glial-specific transcripts, *S100b* and *Fabp7*. Microglia and Schwann cells were identified by expression of *C1qa/C1qb/C1qc* and *Ncmap/Drp2*, respectively, and excluded from further analysis. Dimensionality reduction and clustering of SGCs revealed six distinct transcriptomic subtypes, one of which was characterized the expression of pro-inflammatory markers and excluded from further analyses. The transcriptomic profiles and corresponding biochemical pathways of the remaining subtypes were analyzed and compared with published astrocytic transcriptomes. This revealed gradual shifts of developmental and functional pathways across the subtypes, originating from an immature and pluripotent subpopulation into two mature populations of SGCs, characterized by upregulated functional pathways such as cholesterol metabolism. As SGCs aged, these functional pathways were downregulated while genes and pathways associated with cellular stress responses were upregulated. These findings were confirmed and furthered by an unbiased pseudo-time analysis, which revealed two distinct trajectories involving the five subtypes that were studied. These findings demonstrate that SGCs in mouse stellate ganglia exhibit transcriptomic heterogeneity along maturation or differentiation axes. These subpopulations and their unique biochemical properties suggest dynamic physiological adaptations that modulate neuronal function.

In summary, the bioinformatics methods I developed and the scRNAseq analysis projects I completed will be useful to the broader community for advancing the knowledge of complex disease by comprehensively utilizing scRNAseq and spatial transcriptomics data. JSTA and SCING will enable analysis on high quality spatial transcriptomics data from multiplexed FISH experiments, and the inference of high quality, interpretable GRNs for physiology and disease interpretation respectively. Furthermore, we utilize scRNAseq data to show cell type specific pathogenesis in dilated cardiomyopathy, and a potential therapeutic avenue in AD. These computational methods and analyses help advance the use of scRNAseq and spatial transcriptomics for profiling physiology and complex disease.

Chapter 2. Joint cell segmentation and cell type annotation for spatial transcriptomics

2.1 Introduction

Spatial transcriptomics has been employed to explore the spatial and cell-type specific gene expression to better understand physiology and disease[8-9]. Compared to other spatial transcriptomics methods, RNA hybridization based approaches provided the highest RNA detection accuracies with capture rates > 95%[10]. With the development of combinatorial approaches for RNA hybridization, the ability to measure the expression of hundreds to thousands of genes makes hybridization based methods an attractive platform for spatial transcriptomics [10-13]. Nonetheless, unlike dissociative approaches, such as single-cell RNA sequencing (scRNAseq) where cells are captured individually, RNA hybridization based approaches have no a priori information of which cell a measured RNA molecule belongs to. Segmentation of image volumes into cells is therefore required to convert RNA detection into spatial single-cell data. Assigning mRNA to cells remains a challenging problem that can substantially compromise the overall accuracy of combinatorial FISH approaches.

Generation of spatial single-cell data from imaging based spatial transcriptomics relies on algorithmic segmentation of images into cells. Current combinatorial FISH work uses watershed based algorithms with nuclei as seeds, and the total mRNA density to establish cell borders [11,12,13]. Watershed algorithm was proposed more than 40 years ago[14] and newer segmentation algorithms that utilize state of the art machine learning approaches have been shown to improve upon classical watershed approach [15,16]. However, their performance is inherently bounded by the quality of the "ground truth" dataset used for training. In tissue regions with dense cell distributions, there is simply not enough information in the images to perform accurate manual labeling and create a sufficiently accurate ground truth training datasets. Therefore, there is an urgent need for new approaches that can combine image information with external datasets to improve image segmentation and thereby the overall accuracy of spatial transcriptomics.

Due to the deficiency in existing image segmentation algorithms, a few segmentation free spatial transcriptomics approaches were proposed. pciSeq's primary goal is to assign cell types to nuclei by using proximity to mRNA, and an initialized segmentation map to compute the likelihood of each cell type[17]. Similarly, SSAM creates cell type maps based on RNA distributions, without creating a cell segmentation map because it ignores cellular boundaries[18]. Therefore while both pciSeq and SSAM leverage cell type catalogs to provide insights into the spatial distribution of different cell types they do not produce a high quality cell segmentation map. More recently, an approach for updating cell boundaries in spatial transcriptomics data has been developed[19]. Baysor uses neighborhood composition vectors and markov random fields to segment spatial transcriptomics data and identify cell type clusters.

Here we present JSTA, a computational framework for jointly determining cell (sub)types and assigning mRNAs to cells by leveraging previously defined cell types through scRNAseq. Our approach relies on maximizing the internal consistency of pixel assignment into cells to match known expression patterns. We compared JSTA to watershed in assigning mRNAs to cells through simulation studies to evaluate their accuracy. Application of JSTA to MERFISH measurements of gene expression in the mouse hippocampus together with Neocortical Cell Type Taxonomy[20] (NCTT) provides a highly granular map of cell (sub)type spatial organization and identified many spatially differentially expressed genes (spDEGs) within these (sub)types[21].

2.2 Results

JSTA overview and method

Our computational framework of JSTA is based on improving initial watershed segmentation by incorporating cell (sub)type probabilities for each pixel and iteratively adjusting the assignment of boundary pixels based on those probabilities (Figure 2.1a).

To evaluate JSTA we chose to use the mouse hippocampus for two reasons. 1) the mouse hippocampus has high cell (sub)type diversity as it includes more than 35% of all cell (sub)types defined by the NCTT. 2) the mouse hippocampus has areas of high and low cell density. These two reasons make the mouse hippocampus a good test case for the hypothesis that external cell (sub)type specific expression data could be leveraged to increase the accuracy of spatial transcriptomics, as implemented in our approach. We performed Multiplexed Error Robust Fluorescent In Situ Hybridization (MERFISH) of 163 genes which include 83 selected cell marker genes, which show distinct expression between cell types and are used for cell classification and segmentation and 80 genes previously implicated with biological importance in traumatic brain injury. (Figure 2.1b). Combining this MERFISH dataset, DAPI stained nuclei, and the NCTT reference dataset using JSTA, we created a segmentation map that assigns all mRNAs to cells while simultaneously classifying all cells into granular (sub)types based on NCTT.

In JSTA, we leverage the NCTT information to infer probabilities at the pixel level. However, learning these probabilities from NCTT is challenging for two reasons. 1) NCTT data was acquired with scRNAseq technology that has higher sparsity due to low capture rates and needs to be harmonized. 2) NCTT data provides expression patterns at the cell level and not the pixel level. We expect the mean expression among all pixels in a cell to be the same as that of the whole cell. Yet, variance and potentially higher distribution moments of the pixel level distribution are likely different from those of the cell level distribution due to sampling and biological factors such as variability in subcellular localization of mRNA molecules[12]. To address these issues JSTA learns the pixel level cell (sub)type probabilities using two distinct deep neural networks (DNN) classifiers, a cell level type classifier, and a pixel type classifier. Overall, JSTA learns three distinct layers of information: segmentation map, pixel level classifier, and cell level classifier.

Learning of model parameters is done using a combination of NCTT and the MERFISH data. The cell type classifier is learned directly from NCTT data after harmonization. The other two layers are learned iteratively using expectation maximization (EM) approach[22]. Given the current cell type assignment to cells, we train a pixel level DNN classifier to output the cell (sub)type probability of each pixel. JSTA can be applied on any user selected subset of the genes; the local mRNA density of these selected genes around each pixel are used as the input for the pixel level classifier. The selection of genes drives how well the cell type classifier can distinguish between distinct cell types. The updated pixel classifier is used to assign probabilities to all border pixels. The new probabilities are then used to "flip" border pixels' assignment based on their type probabilities. The updating of the segmentation map requires an update of the cell level type classification which triggers a need for an update of pixel level classifier training. This process is then repeated until convergence. Analysis of the mean pixel level cell (sub)type classification accuracy shows an increase in the algorithm's classification confidence over time demonstrating that the NCTT external information gets iteratively incorporated into the tasks of cell segmentation and type annotation (Fig S1). For computational efficiency, we iterate between training, reassignment, and reclassification in variable rates. As this approach uses cell type information to improve border assignment between neighboring cells, in cases where two neighboring cells are of the same type, the border between them will stay the same as the initial watershed segmentation. The final result is a cell type segmentation map that is initialized based on watershed and adjusted to allow pixels to be assigned to cells to maximize consistency between local RNA density and cell type expression priors.

Performance evaluations

Performance evaluation using simulated hippocampus data

To test the performance of our approach we utilized synthetic data generated based on the NCTT[21] (Figure 2.2ab). Details on the synthetic generation of cell position, morphologies, type,

and expression profiles are available in the methods section. Using this synthetic data we evaluated the performance of JSTA in comparison to watershed at different cell type granularities. For example, two cells next to each other that are of subtypes CA1sp1 and CA1sp4 would add to the error in segmentation, but if the cell type resolution decreases to CA1 cells, these would be considered the same type, and misassignment of mRNA between these cells is no longer penalized. Evaluating the methods in this manner allows us to explore the trade-off between cell type granularity and mRNA assignment accuracy. Our analysis shows that JSTA consistently outperforms watershed at assigning spots to cells (Figure 2.2c). Interestingly, the benefit of JSTA was evident even with a small number of genes (Figure 2.2d). With just 12 genes, the performance jumps to 0.50 at the highest cell type granularity, which is already higher than watershed's accuracy; at a granularity of 16 cell types, the accuracy reached 0.62 (Figure 2.2cd). Overall the synthetic data showed that JSTA outperforms watershed approach, and at physiologically relevant parameters, can increase mRNA assignment accuracy by > 45%. We additionally compared JSTA to pciSeq[17], in the assignment of mRNA molecules to cells. We note that pciSeq is mainly designed to assign cell types to nuclei based on surrounding mRNA, and therefore is not primarily focused on assigning most mRNA molecules to cells as JSTA does. Furthermore, since pciSeq is not designed to operate on 3D data, we simulated 2D data and applied both JSTA and pciSeq. We found that JSTA was more accurate at assigning mRNA molecules to cells than pciSeq (Figure SF2.1a). pciSeq tends to incorrectly assign many spots to background, as segmentation is not its primary goal. However, when ignoring mRNAs assigned to background in a true positive calculation, pciSeq performs well as it primarily assigns mRNAs close to the nuclei, which is an easier task. In this case, JSTA has comparable performance (Figure SF2.1b).

Time requirements of JSTA

We simulated data of different sizes and ran JSTA to determine how the run time scales with larger datasets. We simulated 3 replicates of data with a width and height of 100, 200, 300, 400, 500, and 1000 microns. The run time of JSTA scales linearly with both the area, and number of cells in the section (Figure SF2.7ab).

Performance evaluation using empirical spatial transcriptomics of mouse hippocampus

We next tested the performance of JSTA using empirical data and evaluated its ability to recover the known spatial distribution of coarse neuron types across the hippocampus (Figure 2.3). First, we subset the NCTT scRNAseq data to the shared genes we have in our MERFISH data and harmonized the MERFISH and scRNAseq datasets[23]. Using the cell type annotations from the single-cell data, we trained a DNN to classify cell types. As expected, our classifier derived a cell type mapping agreeing with known spatial patterns in the hippocampus (Figure 2.3a). For example, CA1, CA3, and DG cells were found with high specificity to their known subregions (Figure 2.3b). We found that the gene expression of the segmented cells in MERFISH data highly correlated with their scRNAseq counterparts, and displayed similar correlation patterns between different cell types (Figure 2.3c) as seen in scRNAseq data (Figure 2.3d). These results show that our data and JSTA algorithm can recover existing knowledge on the spatial distribution of cell types and their gene expression patterns in the mouse hippocampus.

JSTA performs high resolution cell type mapping in the mouse hypothalamic preoptic region

We applied JSTA to a MERFISH from a previously published mouse hypothalamic preoptic region with 134 genes provided[23]. Using the provided scRNAseq reference dataset, we accurately mapped 87 high resolution cell types in this region (Figure SF2.2a). The mapped cell types follow spatial distributions of high resolution cell types of this region previously annotated through clustering and marker gene annotation. We find the gene expression profiles of the cell types from the MERFISH data is highly correlated with their scRNAseq counterparts (Figure SF2.2b).

JSTA performs high resolution cell type mapping in the mouse somatosensory cortex

Next, we applied JSTA to an osmFISH dataset from the mouse somatosensory cortex with the 35 genes provided[24]. Using the NCTT reference, we mapped 142 high resolution cell types in this region. We found that the glutamatergic neuronal populations follow known spatial organization (Figure SF2.3a), and that the gene expression patterns of high resolution cell types in the osmFISH data are highly correlated with their NCTT counterparts (Figure SF2.3b).

Applications of JSTA for biological discovery

JSTA identifies spatial distribution of highly granular cell (sub)types in the hippocampus

A key benefit of JSTA is its ability to jointly segment cells in images and classify them into highly granular cell (sub)types. Our analysis of mouse hippocampus MERFISH data found that these subtypes, defined only based on their gene expression patterns, have high spatial localization in the hippocampus. From lateral to medial hippocampus, the subtypes transitioned spatially from CA1sp10 to CA1sp6 (Figure 2.4a). Likewise, JSTA revealed a non-uniform distribution of subtypes in the CA3 region. From lateral to medial hippocampus, the subtypes transitioned from CA3sp4 to CA3sp6 (Figure 2.4b). This gradient of subtypes reveals a high level of spatial organization and points to potentially differential roles for these subtypes.

JSTA shows that spatially proximal cell subtypes are transcriptionally similar

Next, we tested whether across different cell types spatial patterns match their expression patterns by evaluating the colocalization of cell subtypes and their transcriptional similarity. Indeed, spatially proximal CA1 subtypes showed high transcriptional similarity (Figure 2.5a, S5). For example, cells in the subtypes CA1sp3, CA1sp1, and CA1sp6 are proximal to each other and show a high transcriptional correlation. Interestingly, this relationship was not bidirectional, and transcriptional similarity by itself is not necessarily predictive of spatial proximity. For example,

subtypes CA1sp10, CA1sp7, and CA1sp4, show >0.95 correlation but are not proximal to each other. Similar findings were seen in the CA3 region as well (Figure 2.5b, SF2.5).

To test if this principle goes beyond subtypes of the same type we compared CA1 neurons and the Sst interneurons. We found that many Sst subtypes have high specificity in their localization and are transcriptionally related to their non Sst neighbors. Using permutation tests we found that subtypes Sst12, Sst19, Sst20, Sst28 subtypes are significantly colocalized with these same subtypes and are specific to the CA1 region (Figure 2.5cd, methods). Analysis of their transcriptional similarity showed that these subtypes are highly correlated in their gene expression to all CA1 subtypes (Figure 2.5e) but not to CA3 subtypes. These results show that both within a cell type and across cell types spatial proximity indicates similarity in expression patterns.

JSTA identifies spatial differential gene expression (spDEGs)

Given our results on the relationship between spatial localization and gene expression patterns across cell subtypes, we next tested whether spDEGs within the highly granular cell subtypes can be identified. We focused our analysis on the 80 genes in our dataset that were not genes used to classify cells into cell (sub)types. We identified spDEGs by determining if the spatial expression pattern of a given gene was statistically different from a null distribution by permuting the gene expression values. Importantly, the null model was restricted to the permutation of only the cells within that subtype. As a result our spDEGs analysis specifically identifies genes whose expression within a specific subtype has a spatial distribution that is different than random. We found that within hippocampal cell subtypes, many genes were differentially expressed based on their location (Figure 2.6). For example, *Tox* in CA1sp1 shows higher expression on the medial side of the hippocampus and decreases to the lateral side. *Leng8* in subtype CA3sp3 is highly expressed closer to the CA1 region, and lower in the medial CA3. *Hecw1* in the DG2 subtype has varying spatial distribution in the DG region. The lower portion of the DG has clusters of higher

expression, while the upper portion has lower expression. These spatial differences in gene expression are not limited to neuronal subtypes. Astrocytes subtype “Astro1” shows spatial heterogeneity in expression of *Thra*, with large patches of high expression levels and other patches of little to no expression (Figure 2.6a). Overall, we tested for spDEGs in 61 (sub)types with more than 40 cells. We found that all 61 of the tested hippocampal cell subtypes have spDEGs (Figure 2.6b, SF2.6b), with more than 50% (63 of 80) of the tested genes showing non-random spatial pattern (Figure 2.6c, SF2.6c). Certain genes also show spatial patterns in many subtypes (e.g., *Thra* S4ac), while others are more specific to a one or a few subtypes (e.g., *Farp1*, S6ac). Identification of spDEGs highlights an interesting application of highly accurate cell type and mRNA assignment in spatial transcriptomics data.

2.3 Discussion

Spatial transcriptomics provides the coordinates of each transcript without any information on the transcript cell of origin[25]. Here we present JSTA, a new method to convert raw measurements of transcripts and their coordinates into spatial single cell expression maps. The key distinguishing aspect of our approach is its ability to leverage existing scRNAseq based reference cell type taxonomies to simultaneously segment cells, classify cells into (sub)types and assign mRNAs to cells. The unique integration of spatial transcriptomics with existing scRNAseq information to improve the accuracy of image segmentation and enhance the biological applications of spatial transcriptomics, distinguishes our approach from other efforts that regardless of their algorithmic ingenuity are bounded by the available information in the images themselves. As such, JSTA is not a generalist image segmentation algorithm rather a tool specifically designed to convert raw spatial transcriptomics data into single cell level spatial expression maps. We show the benefits of using a dedicated analysis tool through the insights it provides into spatial organization of distinct (sub)types in the mouse hippocampus and the hundreds of newly discovered cell (sub)type specific spDEGs. These insights into the molecular and cellular level structural

architecture of the hippocampus demonstrates the types of biological insights provided by highly accurate spatial transcriptomics.

The promise of single cell and spatial biology lends itself to intense focus on technological and computational development and large scale data collection efforts. We anticipate that JSTA will benefit these efforts while at the same time benefit from them. On the technology side, we have demonstrated the performance of JSTA for one specific variant of spatial transcriptomics, MERFISH. However, the algorithm is extendable and could be applied to other spatial transcriptomics approaches that are based on in situ sequencing [26,27,28], subcellular spatial barcoding [29,30], and potentially any other spatial "omics" platforms [31–36]. Additionally, cell segmentation results from JSTA can be used as input for other tools such as GIOTTO[37] and TANGRAM[38] to facilitate single cell and spatial transcriptomic data analysis. The benefits of JSTA are evident even with a small number of measured genes. This indicates that it is applicable to a broad range of platforms across all multiplexing capabilities. JSTA is limited by its ability to harmonize technical differences between spatial transcriptomics data modalities and the scRNAseq reference. Harmonization between datasets is an active area of research, and JSTA will benefit from these advances[39–43]. JSTA relies on initial seed identification (nuclei or cell centers), and incorrect identification can lead to split or merged cells. JSTA currently does not split or merge cells, but this post processing step could be added to further improve segmentation[44–47]. On the data side, as JSTA leverages external reference data, it will naturally increase in its performance as both the quality and quantity of reference cell type taxonomies improve[48]. We see JSTA as a dynamic analysis tool that could be reapplied multiple times to the same dataset each time external reference data is updated to always provide highest accuracy segmentation, cell (sub)type classification, spDEG identification.

Due to the nascent status of spatial transcriptomics there are many fundamental questions related to the interplay between cell (sub)types and other information gleaned from dissociative technologies and tissue and organ architecture[49,50]. Our results show that strong co-dependency between spatial position and transcriptional state of a cell in the hippocampus, these results mirror findings from other organs[51–53]. This codependency supports the usefulness of the reference taxonomies that were developed without the use of spatial information. Agreements between cell type taxonomies developed solely based on scRNAseq and other measurement modalities, i.e. spatial position, corroborates the relevance of the taxonomical definitions created for mouse brain[20]. At the same time, the spatial measurements demonstrate the limitation of scRNAseq. We discovered many spatial expression patterns within most cell (sub)types that prior to these spatial measurements would have been considered biological heterogeneity or even noise but in fact they represent key structural features of brain organization. High accuracy mapping at the molecular and cellular level will allow us to bridge cell biology with organ anatomy and physiology pointing towards a highly promising future for spatial biology.

2.4 Materials and Methods

Tissue Preparation. B6 mouse was euthanized using carbon dioxide with cervical dislocation. Its brain was harvested and flash frozen in Optimal Cutting Temperature Compound (OCT) using liquid nitrogen. 15 um sections were prepared and placed on pretreated coverslips.

Coverslip Functionalization. Coverslips were functionalized to improve tissue adhesion and promote gel attachment [54]. Briefly, 40 mm No.1 coverslips were cleaned with a 50:50 mixture of concentrated 37% hydrochloric acid and methanol under sonication for 30 minutes. Coverslips were silanized to improve gel adhesion with 0.1% triethylamine and 0.2% allyltrichlorosiloxane in chloroform under sonication for 30 minutes then rinsed once with chloroform then twice with ethanol. Silanization was cured at 70C for 1 hour. An additional coating of 2%

aminopropyltriethoxysilane to improve tissue adhesion was applied in acetone under sonication for 2 minutes then washed twice with water and once with ethanol. Coverslips were dried at 70C for 1 hour then stored in a desiccator for less than 1 month.

Probe Design and Synthesis. A total of 18 readout probes were used to encode the identity of each gene. Each gene was assigned 4 of the possible 18 probes such that each combination was a minimum hamming distance of 4 away from any other gene. This provides classification that is robust up to 2 bit errors. 80 to 120 encoder probes were designed for each target gene. Encoder probes contained a 30 bp region complementary to the transcript of interest with a melting point of 65C and less than 17 bp homology to off target transcripts including highly expressed ncRNA and rRNA. Probes also contained 3 of 4 readout sequences assigned to each gene. Sequences are available in supplementary material. Probes were designed using modified MATLAB code developed by the Zhuang Lab [54]. Probes were ordered from custom arrays as a single strand pool. A T7 promoter was primed into each sequence with a limited cycle qPCR to allow amplification through in vitro transcription and reverse transcription[54].

Hybridization. Hybridization was performed using a modified MERFISH protocol [54]. Briefly, Tissue sections were fixed in 4% PFA in 1xPBS for 15 minutes and washed 3 times with 1xPBS for 5 minutes each. Tissue was permeabilized with 1% Triton X-100 in 1xPBS for 30 minutes and washed 3 times with 1x PBS. Tissue was incubated in 30% formamide in 2xTBS at 37C for 10 minutes. Encoding probes were hybridized at 5 nM per probe in 30% Formamide 10% dextran sulfate 1 mg/mL tRNA 1 uM poly T acridite anchor probed and 1% Murine RNAse inhibitor in 2xTBS. A 30 uL drop of this encoding hybridization solution was placed directly on the coverslip and a piece of parafilm was placed on the coverslip to prevent evaporation. Probes were hybridized for 30-40 hours at 37C in a humidity chamber. Tissue was washed twice with 30% formamide in 2xTBS for 30 minutes each at 45C. Tissue was washed 3 times with 2xTBS. Tissue

was embedded in a 4% polyacrylamide hydrogel with 0.5uL/mL TEMED 5uL 10% APS and 200nm blue beads for 2 hours (Can expand if wanted). Tissue was cleared with 1% SDS, 0.5% Triton x-100, 1 mM EDTA, 0.8 M guanidine HCl 1% proteinase K in 2xTBS for 48 hours at 37C replacing clearing solution every 24 hours. Sample was washed with 2xTBS and mounted for imaging. Readout hybridization was automated using a custom fluidics system. Sample was rinsed with 2xTBS and buffer exchanged into 10% dextran sulfate in 2xTBS for hybridization. Hybridization was performed in 10% dextran sulfate in 2xTBS with a probe concentration of 3nM per probe. Sample was washed with 10% dextran sulfate then 2xTBS. Sample chamber was filled with a 2mM pca 0.1& rPCO 2mM VRC 2mM Trolox in 2xTBS Imaging Buffer. Sample was imaged at 63X using a custom epifluorescent microscope. After imaging fluorophores were stripped using 50mM TCEP in 2xTBS and the next round of readout probes were hybridized.

Image Analysis. Image analysis was performed using custom python code[55]. To register multiple rounds of imaging together with sub pixel resolution, fiduciary markers were found and a rigid body transformation was performed. Images were preprocessed using hot pixel correction, background subtraction, chromatic aberration correction, and deconvolution. An 18 bit vector was generated for each pixel where each bit represented a different round and fluorophore. Each bit was normalized so that background approached 0 and spots approached 1. An L2 normalization was applied to the vector and the euclidean distance was calculated to the 18 bit gene barcode vectors. Pixels were classified if their euclidean distance was less than a 2 bit error away from the nearest gene barcode. Individual pixels that were physically connected were merged into a spot. Dim spots and spots that contained 1 pixel were removed.

Nuclei Segmentation. Nuclei were stained using dapi and imaged after MERFISH acquisition. Each 2D image was segmented using cellpose with a flow threshold of 1 and a cell probability threshold of 0[56] CITE. 2D masks of at least 10um^2 area were merged if there was at least 30

percent overlap between frames. 3D masks that were present in less than 5 z frames (2um) were removed.

Simulation. *scRNAseq reference preparation.* The NCTT was subset to the cells found in the hippocampus, and to the genes from our MERFISH data. Expression levels of simulated genes were taken from scRNAseq reference and were harmonized to qualitatively match the variance observed in measured in MERFISH data. These were then rounded to create a scaled counts matrix. For each of the 133 hippocampal cell types from the NCTT, we computed a mean vector and covariance matrix of gene expression. We additionally computed the cell type proportions in the single cell data for later use in cell type assignment.

Creating the cell map. Initially, the cell centers were placed in a 200 X 200 X 30 micron grid, equidistant from one another, with an average distance between cell centers of 4 microns. The cell centers were then moved around in each direction (x, y, z) based on a gaussian function with mean 0 and standard deviation 0.6. Pixels were then assigned to their closest center with a minimum distance of 5 microns and maximum distance of 7 microns. Cell's with less than 30 pixels were removed due to small unrealistic sizes. To create more realistic and non-round cells, we merged neighboring, touching cells twice. Each cell was assigned a (sub)type uniformly across all 133 types in our dataset. Nuclei were randomly placed within each cell with 20 pixels. Nuclei pixels placed on the border were removed. We simulated 10 independent replicates in each simulation study.

Generating cell transcriptional profiles, and placing spots. Each cell's gene expression profile was drawn from a multivariate gaussian using the mean vector, and covariance matrix computed from the scRNAseq reference. This vector and matrix are cell type specific, and each cell's gene expression profile is sampled from these cell type specific distributions. The mRNA spots were

then placed inside of each cell, slightly centered around the nucleus, but mostly uniform throughout.

Simulated data on limited genes. To perform feature selection and extract a limited number of important genes (4, 12, 20, 28, 36, 44), we used a random forest classifier with 100 trees to predict cell types in the reference dataset. The top n important features for classifying cell types were used. Other simulation parameters were the same as above.

K-Nearest Neighbors (KNN) based Density Estimation Method. We used a KNN approach to estimate density for many genes at each point[57]. The volume required to reach the 5th spot was computed and used to compute the density estimation (Equation 1). Where r is the radius to the 5th closest spot of that gene. We repeated this process for all genes.

$$density = \frac{5}{\frac{4}{3}\pi r^3}$$

JSTA Overview. Expectation Maximization (EM) can be used to jointly classify the identity of an observation of interest, while learning the parameters that describe the class distributions. In EM, the object classes are initialized with a best guess. The parameters of the classifying function are learned from this distribution of initialized classes (M-step). The objects are reclassified according to the updated function parameters (E-step). These steps are repeated until the function parameters converge. JSTA is designed with an EM approach for reclassifying border pixels in the 3 dimensional grid of pixels based on their estimated transcriptional densities. First, we initialize the spatial map with watershed, in euclidean space with a maximum radius. Next we classify cell types of the segmented cells based on the computed count matrix. We then randomly sample a fraction of the pixels' gene expression vectors, and train a pixel classifier (M-step). The

pixel classifier is used to reclassify the cell identity of pixels that are at the border between different cell types, or between a cell and empty space (E-step).

Cell Type Classification. *Data preparation.* To match the distributions of both scRNAseq and MERFISH, we centered and scaled each cell across all genes. We then subsequently centered and scaled each gene across all cells. We note that other harmonization approaches could be applied here.

Cell type classifier. We parameterized the cell type classifier as a neural network, with 3 intermediate layers with 3 times the number of input genes as nodes. We used a tanh activation function with L1 regularization (1e-4) allowing for the influence of negative numbers in the scaled values and parameter space sparsity[58]. Batch normalization was used on each layer[59], and a softmax activation was used for the output layer[60] (Table S1).

Training the classifier. The network parameters were initialized with Xavier initialization[61]. The neural network was trained with two steps with learning rates of 5e-3 and 5e-4 for 20 epochs each, with batch size of 64, and the Adam optimizer was used [62]. A 75/25 train validation split was used to tune the L1 regularization parameter and reduce overfitting. We used 75/25 to increase the representation of lower frequency cell classes. Cross entropy loss was used to penalize the model and update parameters accordingly (Figure SF2.4ab).

Pixel Classification. *Pixel classifier.* We parameterized the pixel classifier as a neural network with 3 intermediate layers. Each layer was twice the size of the last to increase the modeling power of this network, and indirectly model the other genes not in the MERFISH dataset. Each layer used the tanh activation function and used an l2 regularizer (1e-3). Each layer was centered

and scaled with batch normalization, and the output activation was an l2 regularized softmax function (Table S2).

Training the classifier. Each time cell types are reclassified, a new network was reinitialized with Xavier initialization. The network was initially trained with learning rates or 1e-3 and 1e-4 for 25 epochs. After the first round of classifying and flipping the assignment of pixels, the network was retrained on a new sample of pixels starting from the previous parameter values. This was then trained with a learning rate of 1e-4 for 15 epochs. All training was performed with the Adam optimizer and a batch size of 64. We used an 80/20 train validation split to help monitor any overfitting that might be occurring, and adjust the hyperparameter selection accordingly. We used cross entropy loss (Figure SF2.4cd).

Identifying border pixels. Border pixels are defined as pixels that are between two cells of different types, or between a cell and empty space. To enhance the smoothness of cells' borders, we require a border pixel to have 5 of its surroundings be from a different cell, and 2 of its surroundings be from the same cell.

Classifying pixels. The trained classifier was then used to estimate the cell type class of border pixels. The pixel classifier outputs a probability vector for each cell type, and the probabilities are scaled by a distance metric based on the distance to the cells' nuclei that it could flip to. Probabilities less than 0.05 are set to 0. The classification is sampled from that probability vector subset to cell types of its neighbors, and renormalized to 1. If the subset probability vector only contains 0, the pixel identity is set to background. To balance the exploration and exploitation of pixel classification map, we anneal the probability of selecting a non-maximum probability cell type by multiplying the maximum probability by (1 + number of iterations run * *0.05)*. If this is

selected as 0, complete stochasticity presides, and if it is large, the maximum probability will be selected.

JSTA Formalization. *Definitions and background.* The gene expression level of n_c cells and n_p pixels are described by the matrices E_c (cells) and E_p (pixels) which are $n_c \times m$ and $n_p \times m$ matrices respectively, where $m$is the number of genes. Likewise, cell type probability distributions of all cells or pixels can be described by matrices. These distributions for cells and pixels are P_c and P_p respectively, represented as $n_c \times k$ and $n_p \times k$ matrices, where k is the number of cell types. We aim to learn θ and ϕ, such that f_θ and g_ϕ, accurately map from E_c to P_c and E_p to P_p. We used the cross-entropy loss function for penalizing our models.

Cell type classification. First, we learn the parameters of f_θ by:

$$\theta = argmin_\theta[Loss(f_\theta(E_{ref}), T_{ref})]$$

Where E_{ref} is an $n_{ref} \times m$ gene expression matrix representing the harmonized NCTT data and T_{ref} is an n_{ref} vector of cell types labels provided by NCTT. We then use the newly learned mapping to infer the cell type probability distributions in the initialized dataset E_c with:

$$P_c = f_\theta(E_c)$$

We classify each cell as the highest classification probability for that cell:

$$T_c = argmax_k(P_c)$$

Where T_c are the predicted cell types for each of the cells in the matrix E_c.

Joint pixel and parameter updates. We initialize the labels T_p for all pixels based on the current segmentation map that assigns pixels to cells. We then learn the parameters of the mapping function g_ϕ (maximization). Learning is performed by updating the parameters of the mapping function g_ϕ with:

$$\phi = argmax_\phi[Loss(g_\phi(E_p), T_p)]$$

The updated mapping function is then used to infer the probability of observing a type T_p given expression E_p in all pixels:

$$P_p = P(T_p|E_p) = g(E_p)$$

The next step is to update P_p based on spatial proximity to cells of each type. Using the notation q for the vector of probabilities of a single pixel ($q = P_{p_j} = [q_0 \ldots, q_i, \ldots q_k]$) we next update the elements in the vector q based on neighborhood information. We scaled the values of q_i based on its distance from the nuclei and its neighbors. $q^{'}$ is intermediate in the calculation that does not represent true probabilities.

$$q_i^{'} = \begin{cases} 10 & r < d \\ \frac{q_i * d * 5}{2(r-d)} & r \geq d \end{cases}$$

Where r is the distance from the nucleus of the closest cell of cell type i, d is the distance threshold for which a pixel should automatically be assigned to that nucleus. The values 10 and 5 were determined empirically to modify the sharpness of probability decline based on distance. 10 was chosen to be much bigger than probabilities produced by g_ϕ and 5 was chosen to allow the probability to decay to half over $5d$.

We then only kept probabilities for cell types of neighboring cells:

$$q_i^{'} = \begin{cases} 0 & \text{if } i \text{ is neighbor} \\ q_i^{'} & \text{otherwise} \end{cases}$$

We then used the intermediate q' to recalculate the pixel type probabilities:

$$q_i = \frac{q_i^{'}}{\sum_{i=0}^{i=k} q_i^{'}}$$

The updated values per cell (q_j) are then used to update the probability matrix P_p. The type per pixel (T_p). The assignment of pixel to cells is then stochastically assigned according to:

the inferred probability P_p per pixel basis.

$$T_{p_j} \sim multinomial(P_{p_j})$$

We then repeat updating g_ϕ and T_p until convergence.

Segmentation. *Density estimation.* The 3 dimensional space was broken into a grid of pixels with the edge of each pixel 2 microns in length (1 micron in simulation). The density was estimated at the center of each pixel, for each gene. The volume required to reach 5 mRNA molecules was used as the denominator of the density estimation.

Segmentation with JSTA. The cell assignment map was initialized with watershed on the distance transform with a maximum distance from the nucleus of 2 microns. The cells were only classified once. The pixel classifier was trained 6 times (5 in simulation) on 10% of the pixels excluding pixels without assignment. After each training step, we reassigned pixels for 10 iterations (5 in simulation). The lowest probability kept in the predicted pixel assignment vector was 0.05 (0.01 in simulation.

Segmentation with watershed. The overall gene density was the sum of each gene in a given pixel. To smooth the range of the density, we $\log_2$ transformed the density values. Log transformed density values less than 1 were masked. The segmentation used the nuclei locations as seeds and watershed from the skimage python package, with *compactness* of 10. Using compactness of 10 was the highest performing value for watershed. A watershed line was used to separate cells from one another.

Evaluation of Segmentation in Simulated data. mRNA spot call accuracy was evaluated at different taxonomic levels. For a given cell the accuracy was defined as the number of mRNA spots correctly assigned to that cell divided by the total number of mRNA spots assigned to that

cell. To match the algorithm's ability to segment based on cell type information, RNAs that were assigned to a neighboring cell of the same (sub)type were also considered correct assignment. The overall segmentation accuracy was the mean accuracy across all cells in a given sample. To evaluate accuracy at different levels, we utilized the NCTT dendrogram. We used dendrogram heights at 0 through 0.8 with a step size of 0.05 (133, 71, 32, 16, 11, 8, 5, 4, 3, 2 cell types).

Correlation of Segmented MERFISH with scRNAseq. The NCTT scRNAseq data was subset to the genes from our MERFISH data. Cells in the segmented MERFISH dataset were assigned to canonical hippocampus cell types (Astrocyte, CA1 pyramidal neuron, CA2 Pyramidal neuron, CA3 Pyramidal, Dentate Gyrus, Inferior temporal cortex, Macrophage, Oligodendrocyte, Subiculum, Interneuron) based on their high resolution cell type classification. In each cell type the average expression in each gene was calculated. Only genes were kept that had an average expression of at least 5 counts in one of the cell types. Values were centered and scaled across all cell types. The Pearson correlation was computed for each gene for the matching cell types between scRNAseq and MERFISH.

Distribution of High Resolution Celltypes in the Hippocampus. CA1 and CA3 subtypes were projected onto the lateral medial axis. The smoothed density across this dimension was plotted for each of the subtypes.

Colocalization of High Resolution Celltypes. Significant colocalization of subtypes was determined through a permutation test. First, the 20 nearest cell types around each cell were determined. We counted the number of cells from each type that surround each cell type and computed the fraction of neighbors coming from each subtype. This created a matrix with the fraction of colocalizations per cell between each cell type combination. We then permuted the labels of the cell types 1000 times, and recomputed this interaction matrix to create a null

distribution. For each cell type colocalization, we determined the percentage of colocalizations in the null distribution that are higher than the true colocalization number to create a p-value for each colocalization. We corrected for multiple testing with the benjamini hochberg procedure and determined significance using FDR < 0.05.

Identification of Spatial Differential Gene Expression (spDEGs). spDEGs were calculated in cell types with more than 40 cells. Within each cell type, we computed a local expression of each gene for each cell. The local expression was the mean expression of a gene in the cell and its 9 nearest neighbors. We then built a null distribution by permuting gene expression values within the cell type, and repeating the local expression process for 100 permutations. Determining if a gene was spatially differentially expressed, we compared the variance of the null distribution within a cell type with the variance of the true distribution of local expression to get a p-value. We corrected for multiple testing with benjamini hochberg procedure and determined significance using FDR < 0.05.

Python packages used.

python (3.8.3), numpy (1.18.5), pandas (1.0.5), matplotlib (3.2.2), scipy (1.5.0), scikit-learn (0.23.1), scikit-image (0.16.2), tensorflow (2.2.0). seaborn (0.10.1)

Data Availability

Source code: GitHub (https://github.com/wollmanlab/JSTA)

2.5 Figures

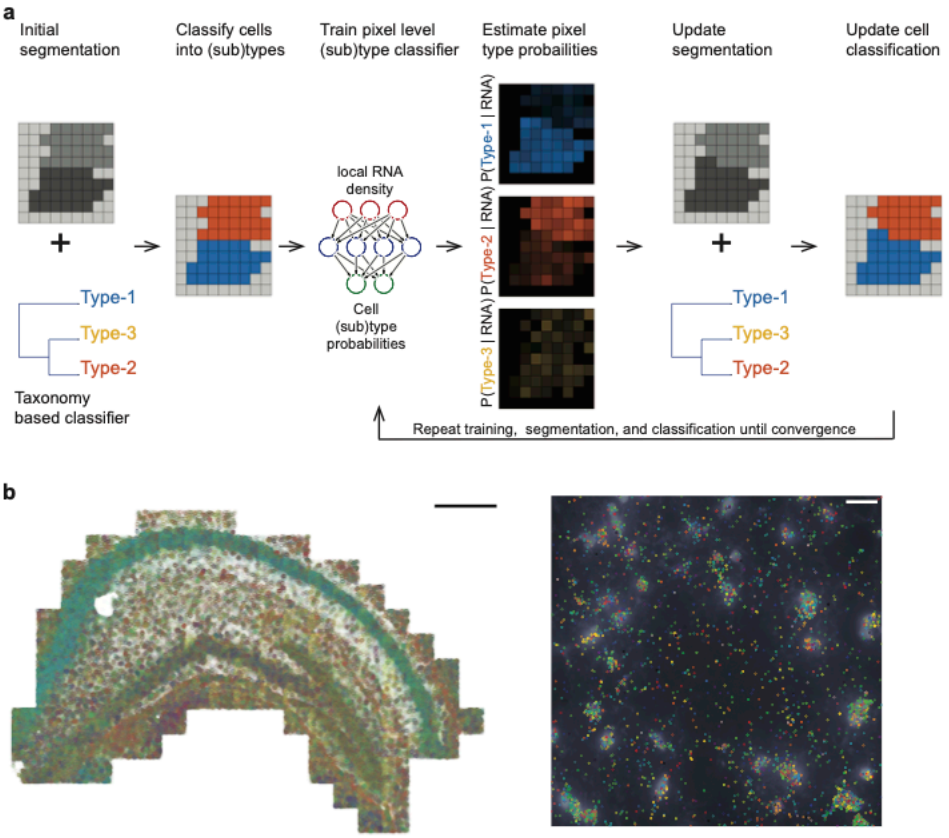


Figure 2.1. Overview of JSTA and the spatial transcriptomics data used for performance evaluation.

a. Joint cell segmentation and cell type annotation (JSTA) overview. Initially, watershed based segmentation is performed and a cell level type classifier is trained based on the Neocortical Cell Type Taxonomy (NCTT) data. The deep neural network (DNN) parameterized cell level classifier then assigns cell (sub)types (red and blue in this cartoon example). Based on the current assignment of pixels to cell (sub)types, a new DNN is trained to estimate the probabilities that each pixel comes from each of the possible (sub)types given the local RNA density at each pixel. In this example, two pixels that were initially assigned to the “red” cells got higher probability to be of a blue type. Since the neighbor cell is of type “blue” they were reassigned to that cell during

segmentation update. Using the updated segmentation and the cell type classifier cell types are reassigned. The tasks of training, segmentation, and classification are repeated over many iterations until convergence.

b. Multiplexed Error Robust Fluorescent in situ hybridization (MERFISH) and DAPI stained nuclei in the mouse hippocampus. Each gene is represented by a different color. For the entire hippocampus (left), only the mRNA spots are shown with a scale bar of 500 microns. On the zoomed-in section (right), each gene is represented by a different color dot, and the DAPI intensity is displayed in white. The scale bar is 20 microns.

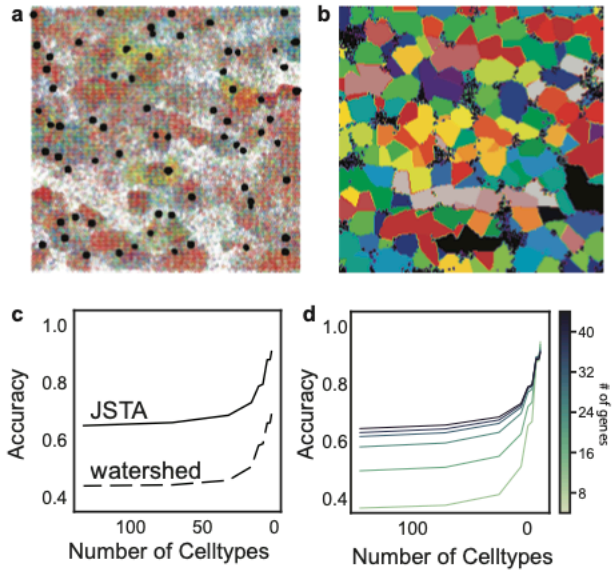


Figure 2.2 Performance evaluation of JSTA using simulated data.

a. Representative synthetic dataset of nuclei (black) and mRNAs, where each color represents a different gene.

b. Ground truth segmentation map of the cells in the representative synthetic dataset. Each color represents a different cell.

c. Average Accuracy of calling mRNA spots to cells at different cell type resolutions using 83 genes across 10 replicates. Accuracy was determined by the assignment of each mRNA molecule to the correct cell type. JSTA (solid line) is more accurate than watershed (dashed line) at assigning mRNA molecules to the correct cells (FDR < 0.05). Statistical significance was determined with a Mann-Whitney test and false discovery rate correction.

d. Accuracy (as described in c) of calling mRNA spots to cells when using JSTA to segment cells with a lower selection of cell type marker genes (8-44 genes tested). The color of the line gets progressively darker as the number of genes used increases.

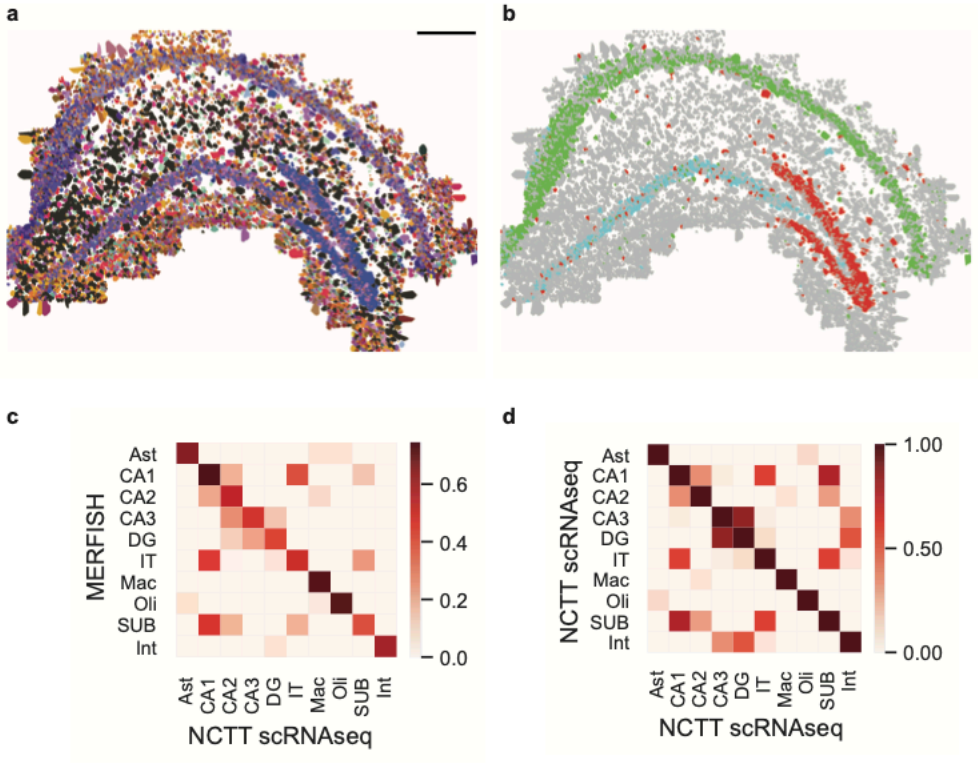


Figure 2.3 Segmentation of MERFISH data from the hippocampus using JSTA.

a. High resolution cell type map of 133 cell (sub)types segmented and annotated by JSTA. Colors match those defined by Neocortical Cell Type Taxonomy (NCTT). Scale bar is 500 microns.

b. JSTA based classification of CA1 (green), CA3 (cyan), and DG (red) neurons matches their known domains.

c. Correlation of the average expression of 163 genes across major cell types between MERFISH measurements to scRNAseq data from NCTT.

d. Correlation of the average expression of the same genes as in c between expression of types in scRNAseq data from NCTT. The correlation structure in c panel C closely mirrors the structure in dpanel D.

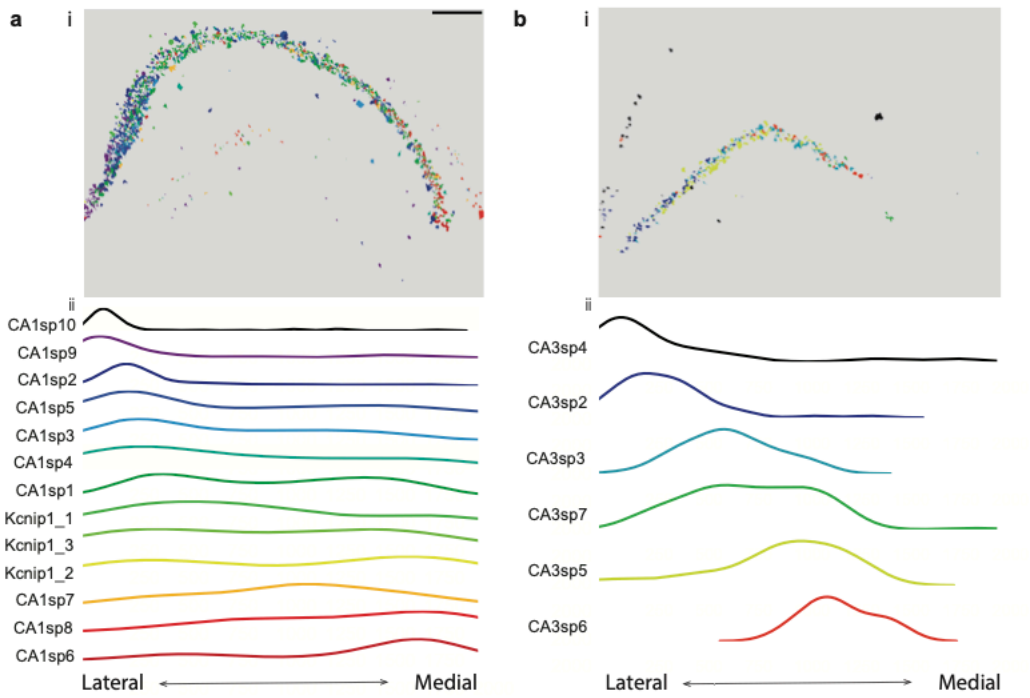


Figure 2.4 Spatial distribution of neuronal subtypes in the hippocampus.

a(i). (i) Cell subtype map of CA1 neurons in the hippocampus as annotated by JSTA. Scale bar is 500 micron. Distribution of CA1 subtypes in the hippocampus, computed by projecting cell centers to the lateral to medial axis. CA1 neuronal subtypes show a non-uniform distribution across the whole CA1 region. **a(ii)** Smoothed histogram highlighting the density of CA1 subtypes across the CA1 region**.**

b(i). (i) Cell subtype map of CA3 neurons in the hippocampus as annotated by JSTA. Distribution of CA3 subtypes in the hippocampus, computed by projecting the cell centers to the lateral to medial axis. CA3 neuronal subtypes show a non-uniform distribution across the whole CA3 region. **b(ii)** Smoothed histogram highlighting the density of CA3 subtypes across the CA3 region**.**

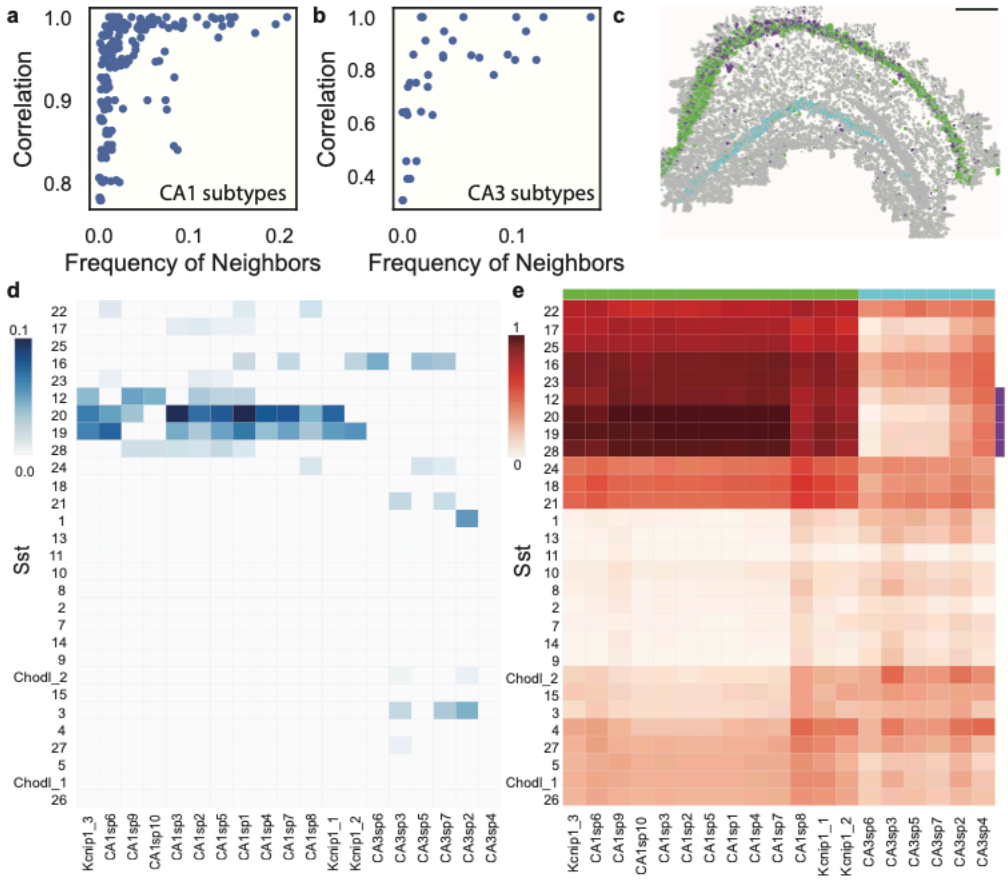


Figure 2.5 Agreement between spatial proximity and gene coexpression in highly granular cell subtypes in the hippocampus.

a-b. Relationship between the frequency of a (sub)type's neighbors and it's transcriptional pearson correlation between CA1 subtypes (**a**) and between CA3 subtypes (**b**).

c. Cell type map in the hippocampus shows specific colocalization patterns between a subset of Sst subtypes (purple) and CA1 neurons (green); these Sst subtypes do not colocalize with CA3 neurons (cyan).

d. Colocalization patterns of Sst subtypes with CA1 and CA3 subtypes. Sst subtypes that colocalize with the CA1 subtypes have high transcriptional similarity. Colocalization was defined as the percent of neighbors that are of that subtype (Materials and Methods)

e Transcriptional correlation patterns between Sst subtypes and CA1 and CA3 neurons. Green, purple and cyan sidebars highlight the subset of Sst co-localized with CA1 (purple), CA1 (green) and CA3 (cyan).

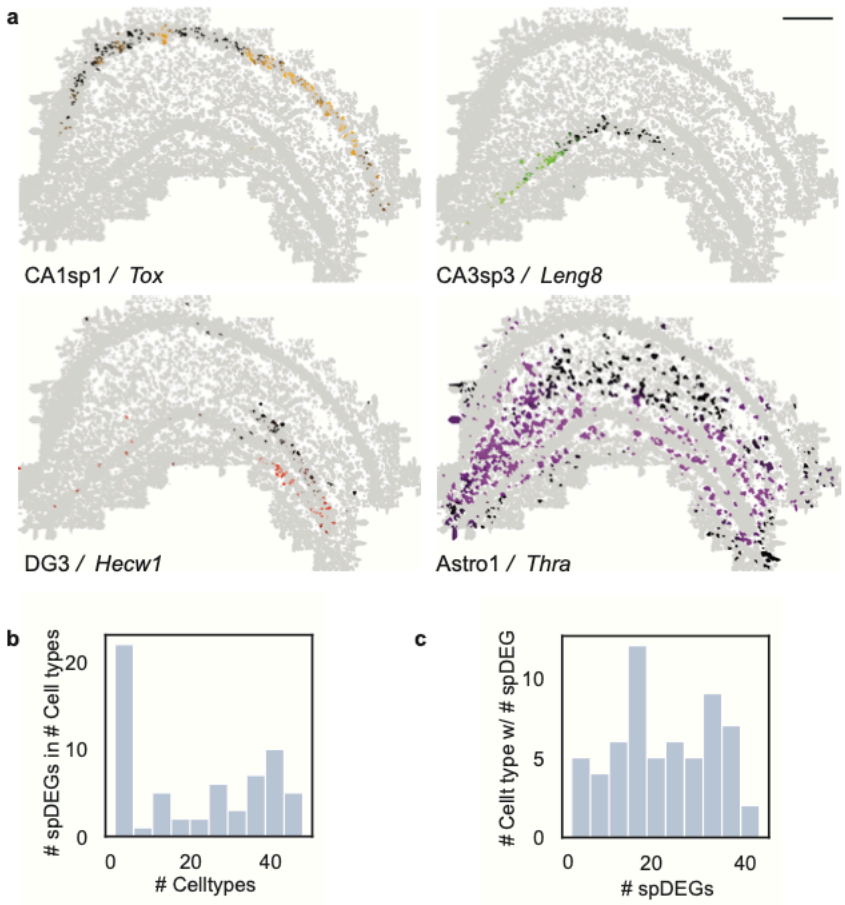


Figure 2.6 Identification of spatial differential gene expression (spDEGs).

a. Normalized expression of Tox in CA1sp1, Leng8 in CA3sp3, Hecw1 in DG3, and Thra in Astro1 show variable expression throughout the hippocampus. Scale bar is 500 microns. spDEGs were computed by comparing the true variance in gene expression between cell subtype neighborhoods to that of randomly permuted cell (sub)type neighborhoods.

b. Histogram of the number of statistically significant spDEGs (Benjamini-Hochberg corrected FDR < 0.05) in each subtype.

c. Histogram of the number of subtypes that have an spDEG for each gene.

Expanded View Figures

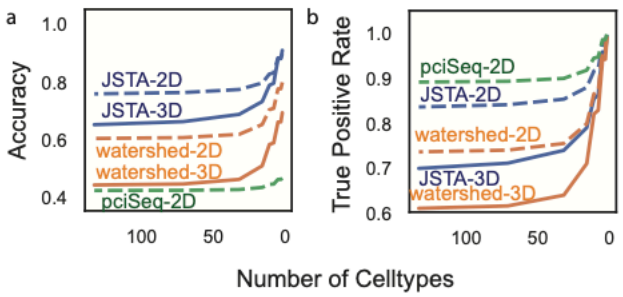


Figure SF2.1. Performance evaluation of JSTA, pciSeq, and watershed.

A, B pciSeq is unable to run on 3D data (solid line), so we simulated additional 2D data (dotted line). We evaluated these methods on the performance of accuracy of assigning mRNA to the correct cell (a). JSTA is more accurate than pciSeq on the accuracy metric. pciSeq is not very accurate here, because many mRNA are incorrectly assigned to background. We additionally tested these methods on their performance of assigning mRNA to the correct cell while ignoring mRNA assigned to background (b). pciSeq is highlighted here, because it mainly assigns spots close to the nucleus; JSTA is comparable.

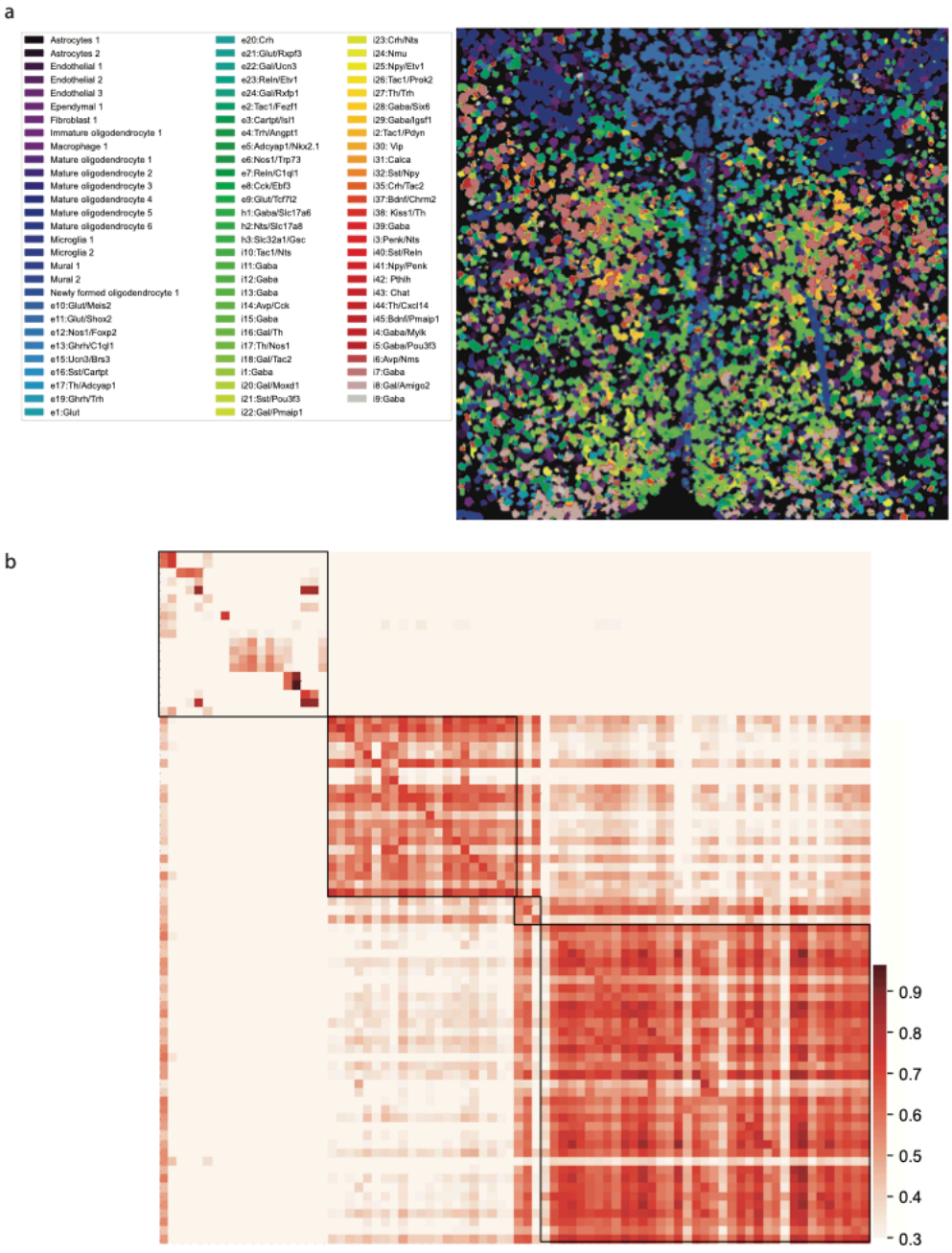


Figure SF2.2. Application of JSTA to MERFISH data from the mouse hypothalamic preoptic region.

A. High resolution cell types identified by JSTA (a). The spatial mappings of these high resolution cell types are consistent with the manually annotated data from Moffit et al.

B. JSTA mapped high resolution (sub)types are highly correlated to their scRNAseq reference counterparts in terms of gene expression patterns (b; Supplementary table 3). Cell types with at least 5 cells were kept.

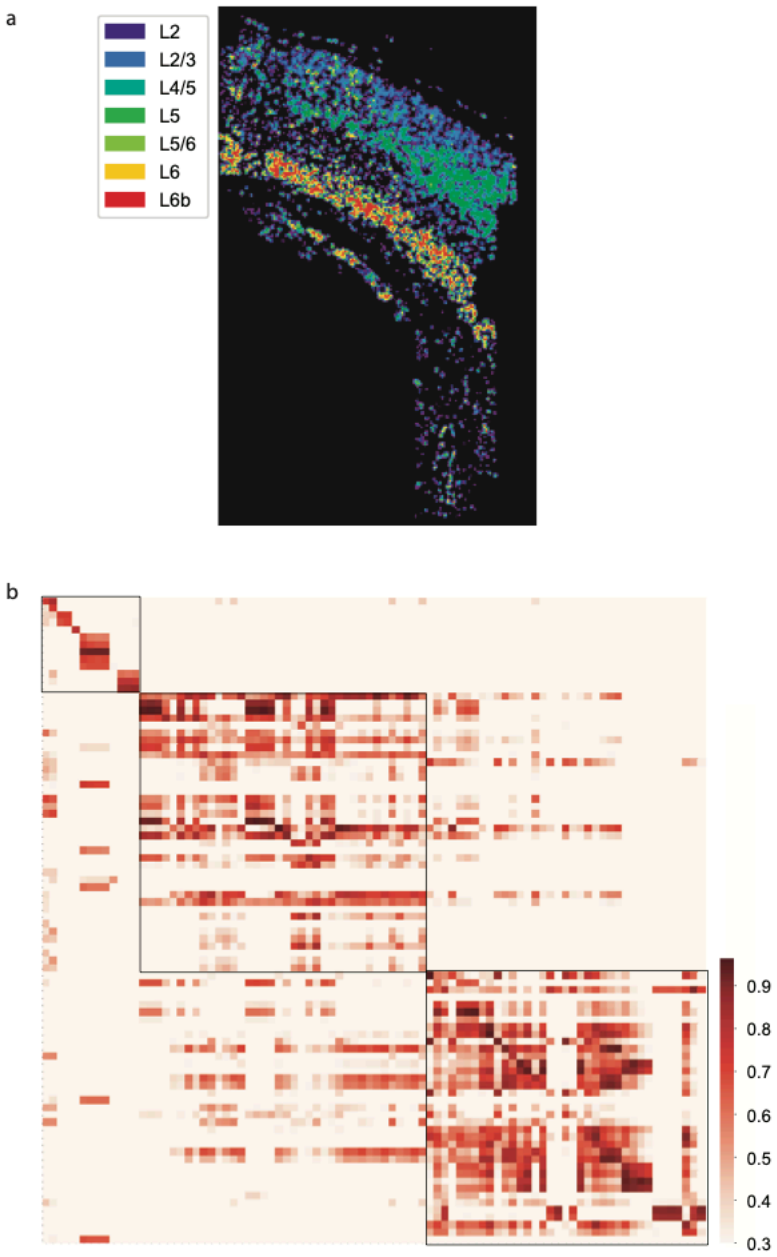


Figure SF2.3. Application of JSTA to osmFISH data from the mouse somatosensory cortex.

A. Glutamatergic neurons are consistent with previously identified spatial patterns of the somatosensory cortex (a).

B. JSTA mapped high resolution (sub)types are correlated with their NCTT counterparts in terms of gene expression patterns (b; Supplementary table 4). Cell types with at least 5 cells were kept.

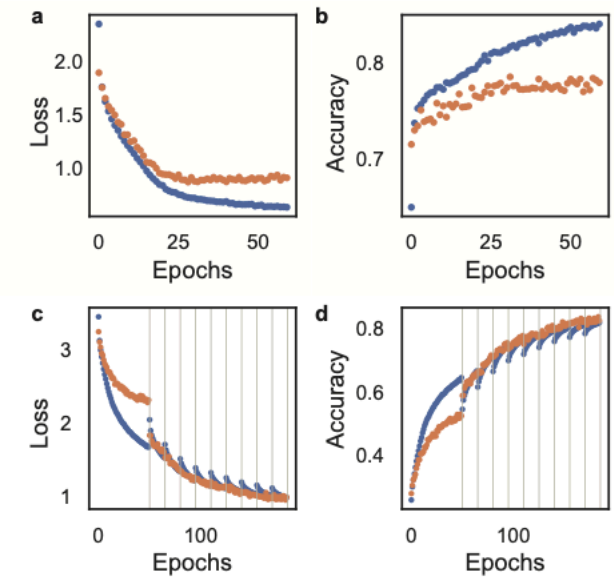


Figure SF2.4. Cross entropy loss and accuracy of cell type (a-, b) and pixel (c-, d) classifier during training for the train (blue) and validation (orange) data sets.

a-, b. Cross entropy (**a**) loss and accuracy (**b**) during training cell type classifier. The cell type classifier overfits the training data, and is mitigated by stopping training after 40 epochs.

c-, d. Cross entropy loss (**c**) and Accuracy (**d**) during training of the pixel classifier. Black lines indicate new training iteration after pixel reassignment**.**

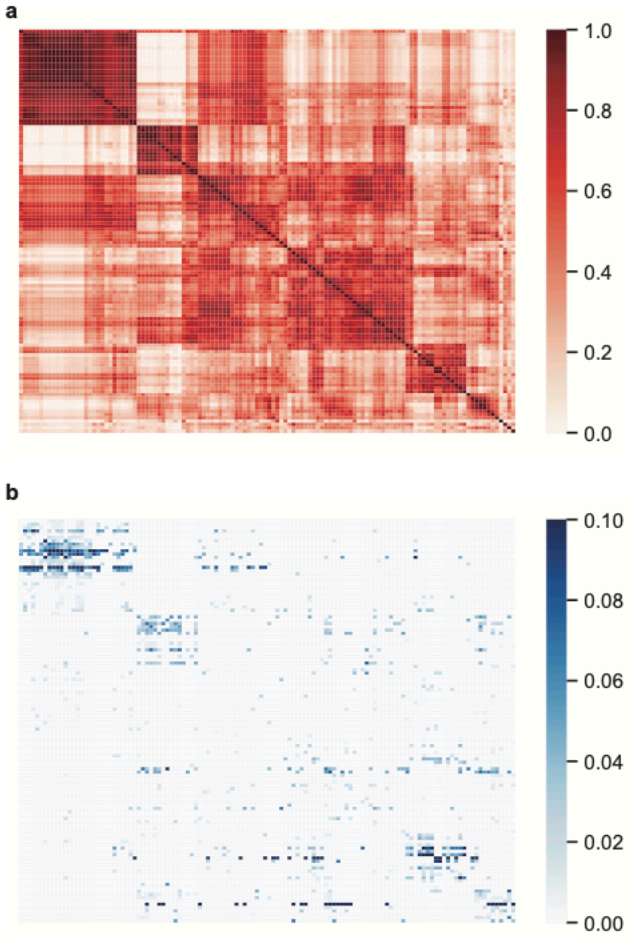


Figure SF2.5. Correlation structure of cell types compared to their colocalization.

Neuronal subtypes that are highly collocalized are often correlated in their gene expression. Cell types with more than 10 cells were included.

a. pearson Pearson correlation of 122 (sub)types across 83 selected genes.

b. Frequency of neighbors between each of 122 (sub)types. Only significant (FDR < 0.05) colocalizations are shown. Labels and values are detailed in supplementary table 5 and 6.

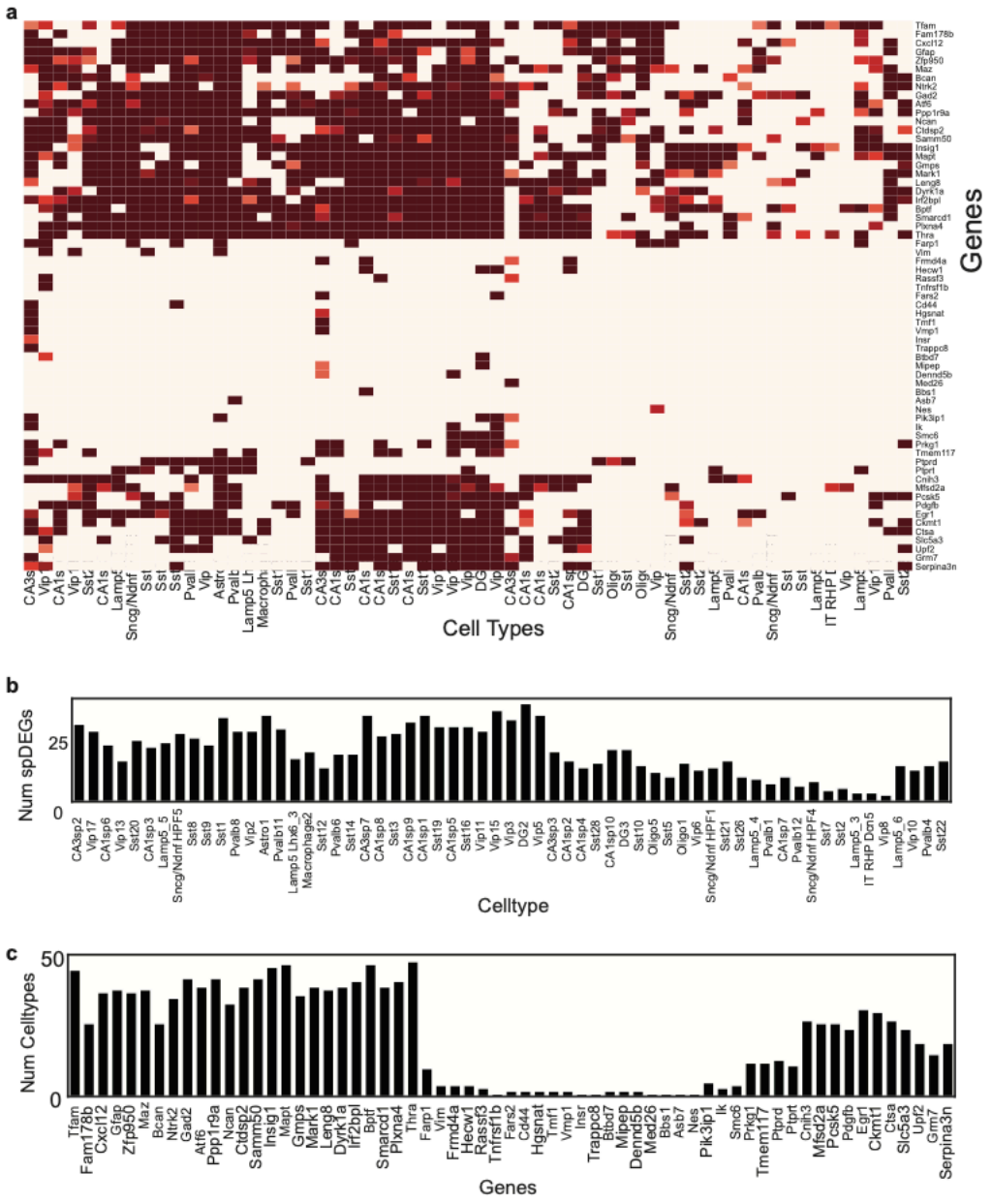


Figure SF2.6. Identification of spatial differentially expressed genes (spDEGs).

spDEGs were computed by comparing the true variance in gene expression between cell subtype neighborhoods to that of randomly permuted cell (sub)type neighborhoods.

a. 63 genes across 61 cell types show significant spDEGs. Heatmap values correspond to -$\log_2$(p-value).

b. Number of spDEGs in each of the 61 cell types.

c. Number of cell types with each of the 63 spDEGs.

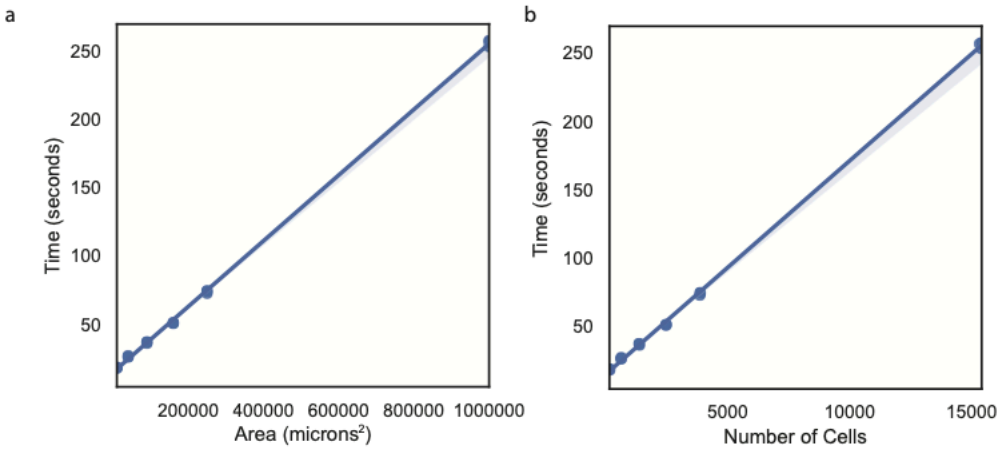


Figure SF2.7. Run time evaluation of JSTA on simulated data.

A, B We ran JSTA on data simulated with a width and height of 100, 200, 300, 400, 500, and 1000 microns, with 3 replicates each. We evaluated the time taken to run JSTA by the area of the section (a), and the number of cells in each section (b).

2.6 Tables

Table S2.1. Cell type classifier architecture.

The network was initialized with Xavier initialization. Learning rates of $5e^{-3}$ and $5e^{-4}$ for 20 epochs each. A batch size of 64 was used. The Adam optimizer was used to update parameters. Cross entropy loss was used.

Layer Type	Num Nodes	Activation	Regularization
Input	83	-	-
Dense	249	tanh	L1 ($5e^{-3}$)
Batch Normalization	-	-	-
Dense	249	tanh	L1 ($5e^{-3}$)
Batch Normalization	-	-	-
Output	133	softmax	L1 ($5e^{-3}$)

Table S2.1. Pixel classifier architecture.

The network was initialized with Xavier initialization. Initially the model was trained for 25 epochs with $1e^{-4}$ and $1e^{-3}$ learning rate. Subsequent updates were done on 15 epochs with a learning rate of $1e^{-4}$. We used the Adam optimizer to update parameters. Cross entropy loss was used.

Layer Type	Num Nodes	Activation	Regularization
Input	83	-	-
Dense	166	tanh	L1 ($1e^{-3}$)

Batch Normalization	-	-	-
Dense	166	tanh	L1 ($1e^{-3}$)
Batch Normalization	-	-	-
Dense	332	tanh	L1 ($1e^{-3}$)
Batch Normalization	-	-	-
Output	133	softmax	L1 ($1e^{-3}$)

Chapter 3. SCING: Single Cell INtegrative Gene regulatory network inference elucidates robust, interpretable gene regulatory networks

3.1 Introduction

Understanding pathophysiology is necessary for the diagnosis and treatment of complex diseases, which involve the perturbation of hundreds or thousands of genes[3,63–65]. Identifying perturbed gene pathways and key drivers of complex diseases requires the elucidation of gene regulatory networks (GRN) from high dimensional omics data[66,67]. Previous approaches have been developed and applied to identify these GRNs through bulk transcriptomic data and to determine causal mechanisms of disease[68–70]. More recently, with the advent of single cell RNA sequencing (scRNAseq) and spatial transcriptomics, the contributions of numerous genes in individual cell types have been implicated in diseases across many disciplines of biology and medicine[4,71,72].

GRN construction from scRNAseq data has been tackled with limited success[73–75]. Existing GRN tools utilize scRNAseq data with thousands of pre-select genes and cells, because GRNs from full transcriptomes in large scRNAseq datasets are often computationally intensive and intractable[73,75]. Additionally, benchmarking studies have shown limited accuracy of existing methods on both synthetic and real data[73]. While tools such as ppcor[76] and PIDC[77] use partial correlation and partial information decomposition, respectively, to identify gene coexpression modules, few methods are able to identify directed networks[74]. One tool that can identify directed networks is GRNBOOST2[78], a gradient boosting based method that has been successful in benchmarking studies[73,75]. However, most of the regulatory edges from GRNBOOST2 point in both directions, and the resulting GRNs contain too many edges in the range of 34,000 to 47,000 edges for networks with only 3,000 input genes, making this approach impractical for datasets with thousands of cells and full transcriptomes[78]. SCENIC[79], an extension of GRNBOOST2, prunes edges based on known transcription factor binding sites (TFBS). However, it only focuses

on regulatory behavior between transcription factors (TF) and its downstream target genes[79], thereby missing other non-TF gene regulatory mechanisms[80,81], as well as regulatory patterns due to low TF expression[82]. While there exist other GRN inference approaches based on pseudotime analysis[83–85], their performance is generally inferior to those that rely solely on the scRNAseq data[73].

Here, we present Single Cell INtegrative gene regulatory network inference (SCING), a gradient boosting based approach to efficiently identify GRNs in full single cell transcriptomes. The robustness of SCING is achieved via i) merging and taking consensus of GRNs through bagging, and ii) further directing and pruning edges through the use of edge importance and conditional mutual information. SCING GRNs are then partitioned into modules, to compute module specific expression for each cell. These modules can be used for clustering, phenotypic association, and biological annotation through pathway enrichment. We show our approach is both efficient and robust on large scRNAseq datasets, able to predict perturbed downstream genes of high throughput perturbation experiments, and produces gene subnetworks with biologically meaningful pathway annotations. We evaluate our approach against GRNBOOST2, ppcor, and PIDC through perturbation target prediction in Perturb-seq data, goodness of fit, network characteristic metrics, and disease modeling accuracy. Furthermore, we apply SCING to the mouse single cell atlas[86], snRNAseq[87], and spatial transcriptomics[7] datasets to demonstrate its versatility in datatype accommodation and its biological interpretability of high-throughput transcriptomics datasets. Our code and tutorials for running SCING are publicly available at https://github.com/XiaYangLabOrg/SCING.

3.2 Results

SCING method and evaluation overview

SCING leverages the power of abundant cell-level transcriptome data from scRNAseq/snRNAseq to identify potential directional regulatory patterns between genes. However, single cell transcriptomics data has technical issues, such as high sparsity and low sequencing depth[88,89], which make the use of traditional linear or correlative approaches challenging. Additionally, these datasets often contain tens of thousands of cells each with hundreds to thousands of genes, making identifying GRNs using complex non-linear approaches on full transcriptomes difficult[73]. To address these limitations, we employ a combination of supercell, gene neighborhood based connection pruning, bagging, gradient boosting regression, and conditional mutual information approaches to identify robust regulatory relationships between genes based on single cell transcriptomics data (Figure 3.1a; Methods).

We evaluated our approach against PIDC, a partial information decomposition approach; ppcor, a partial correlation approach; and GRNBOOST2, a gradient boosting approach. We selected these particular methods for comparison due to their overall better performance in recent benchmarking studies[73,75] and diverse approaches. We compared these methods on the ability to predict downstream gene targets of large scale perturb-seq studies (Figure 3.1b), robustness of the network on training and test data (Figure 3.1c), metrics including the consistency of edge overlap on GRNs built on independent cells, and the ability to model disease subnetworks (Figure 3.1d). Furthermore, we demonstrated the utility of using SCING on the full mouse cell atlas and a human prefrontal-cortex snRNAseq dataset[87] with AD and control patients to perform snRNAseq batch harmonization, gene module identification with biological annotation (Figure 1e), and module-trait association analysis (Figure 3.1f). Data from Morabito et al. has snRNAseq from 11 AD and 7 control human prefrontal cortex samples, with 61,472 nuclei across 7 cell types, which provides a high quality dataset for benchmarking. Furthermore, we applied SCING to a visium mouse dataset with AD vs control samples[7]. We show that the SCING subnetworks are versatile in data type accomodation (scRNAseq, snRNAseq, spatial transcriptomics), can resolve

spatial biology, and are powerful in retrieving biologically meaningful pathways, gene connections, and disease associations.

SCING extends network node inclusion capacity and improves computing speed

SCING builds many GRNs for each dataset and the speed of such computation is paramount to reasonable computation for a whole dataset. The use of supercells and gene covariance based potential edge pruning enables faster performance of SCING. We show that SCING improves computational speed over GRNBOOST2 and PIDC, when increasing the number of genes (SF3.1a, Table 3.1), and number of cells (SF3.1b, Table 3.2). GRNBOOST2 scales exponentially on the number of genes, while PIDC scales exponentially on the number of cells, making whole transcriptome and large dataset GRN inference difficult. Supercells in SCING ensure the network building run time does not increase as a function of cells and potential edge pruning enables linear increase in computation with respect to genes. We note that ppcor's fast general matrix formulation improves GRN inference time compared to all other approaches, including SCING. While SCING is slower than ppcor, it performs inference on 4,000 genes in ~21 seconds for all cell types, which is reasonable to compute hundreds of GRNs for any given sample.

SCING GRNs better predict downstream genes of perturbed genes in Perturb-seq

We tested whether GRNs from each approach can predict gene expression changes in downstream genes from gene knockdown treatments. Here, we used perturb-seq datasets, which enabled us to identify the effects of many perturbed genes in parallel. We utilize previously published datasets with THP-1, dendritic (DC), and K562 cells with 25, 24, and 21 genes perturbed, respectively[90,91]. The DC cells were split into lipopolysaccharide (LPS) stimulated and non-stimulated cells with perturbations targeting transcription factors (TFs) as two datasets, and the K562 cells were split into two datasets based on the genes initially perturbed (TFs or cell

cycle related genes) in the Perturb-seq experiments. THP-1 cells contain perturbations targeting PD-L1 regulators.

We identified genes downstream of each perturbation through an elastic net regression framework to determine the effect of RNA guides on each gene while regressing out cell state[91] (Methods). We compared GRNs generated from SCING to GRNs generated from GRNBOOST2, PIDC, and ppcor in predicting genes downstream of each target gene in each perturb-seq experiment. For any given network, we iterated through the downstream genes of an initially perturbed gene by the RNA guide and determined if the predicted downstream genes were significantly altered. We determined the true positive rate (TPR) and false positive rate (FPR) at each network depth to compute the area under the receiver operating characteristic (AUROC) curve, as well as the TPR at an FPR of 0.05. We examine TPR at FPR 0.05 to show perturbation prediction accuracy in a setting more relevant to biological analysis (controlling for FPR 0.05). We first examine the prediction performance of each GRN approach when building GRNs on datasets with cells removed that have zero expression of the target gene. Removing these cells mitigates performance effects from sparsity (SF3.2a). Since ppcor and PIDC produce undirected graphs, and GRNBOOST2 generally has bidirectional edges, we first evaluated SCING against the other methods without considering edge direction, showing a higher AUROC for SCING (Figure 3.2a). However, edge direction only affects prediction accuracy in TPR at FPR 0.05 for dc 3hr cells (Figure 3.2b). We additionally show SCING improves TPR at FPR of 0.05 (Figure 3.2c), and similarly, edge direction typically does not affect this metric (Figure 3.2d).

Across all perturb-seq datasets tested and when no cells were removed, SCING either outperformed, or met the performance of the other tools in both AUROC (SF3.2b) and TPR at FPR 0.05 (SF3.2d) and performance was minimally affected by edge direction (SF3.2ce). For AUROC, SCING outperforms ppcor across all datasets, PIDC across 4 datasets, and

GRNBOOST2 across 2 datasets. For TPR at FPR 0.05, SCING outperforms ppcor and PIDC across 4 datasets and GRNBOOST2 across all datasets (SF3.2d). SCING performed best in the Papalexi et al. THP-1 dataset, although there is a large variance in both AUROC (SF3.2b) and TPR at FPR 0.05 (SF3.2d) for both ppcor and SCING.

Overall, SCING outperforms all other approaches at predicting perturbation effects in perturb-seq data when sparsity is adjusted (Figure 3.2) and outperforms select methods when all cells are used (SF3.2). Thus, we recommend removing cells with sparse gene expression when using SCING.

SCING mitigates overfitting and builds more robust GRNs

Models often overfit to their data and fail to properly perform on new datasets. In the GRN context, we aimed to identify connections and networks that are able to capture biological variation rather than sample or batch specific effects. To test the performance of GRN inference approaches and their ability to capture robust biological signals, we tested the ability of a model trained on parents of each gene in training data to predict the gene expression of the downstream target gene in testing data. We split the scRNAseq data from control human prefrontal cortex[87] into training and testing sets (Methods). First, we built GRNs on the training data from oligodendrocytes, astrocytes, and microglia using SCING, ppcor, PIDC, and GRNBOOST2. Subsequently, we trained gradient boosting regressors for each gene based on the parents in a given network using the training data. The trained regressors were then used to predict the gene expression of cells in testing data based on the expression of the parent genes in those cells. We evaluated the performance of each GRN approach by averaging the cosine similarity score over all downstream genes that have parents in the network. This process was repeated for 10 replicates on random subsamples of 3,000 genes to reduce runtime.

To measure overfitting, we used the cosine similarity score (ratio between the test and training sets) with a higher ratio indicating lower overfitting. We found that SCING GRNs had less overfitting than the other approaches (Figure 3.2e). In terms of performance in the test sets, SCING performed similarly to ppcor and PIDC, outperforming GRNBOOST2 (Figure 3.2f). On training data, GRNs from ppcor, PIDC, and GRNBOOST2 had higher cosine similarity scores compared to SCING, reflecting overfitting on the training data by the other methods (Figure 3.2g). We noted that the number of genes in the resulting network to be very low in the ppcor oligodendrocyte network (Figure 3.2h), which likely affected the results of ppcor as evaluated by the cosine similarity measure here. These results support the robustness of SCING GRNs, highlighting its ability to identify a robust GRN without overfitting or sacrificing performance.

SCING fits scale free model, and shows edge consistency

As another measure of GRN quality and performance, we compared GRNs generated by each method by various standard network metrics (scale-free network fit, number of edges, number of genes, and betweenness centrality), as well as robustness of network edges between networks on 50/50 split datasets. We tested this on 10 replicates for each of the 3 cell types (oligodendrocyte, astrocyte, and microglia) in the scRNAseq data from control human prefrontal context, with 3,000 different genes randomly selected for each subsample of cells.

GRNs are thought to follow a scale-free network structure, in which there are few nodes with many connections, and many nodes with few connections[92]. We computed scale-free network structure through the R-squared coefficient of a linear regression model regressing on the log of each node's degree and the log of the proportion of nodes with that given degree. We show that the R-squared value for SCING is significantly higher than that of the other methods, indicating SCING networks more closely follow a scale free network structure (Figure 3.3a). In a typical scale-free network plot of log10 node count vs log10 degree, we expect a power law distribution.

However, we found that networks built on scRNAseq data have a parabolic distribution, with only the right half following a power law distribution whereas the left portion of the plot is driven by genes that were very sparse and likely do not fit typical distributions (Figure 3.3b). Therefore, we excluded the sparse genes from the regression calculation.

Among all 4 methods tested, PIDC produced the largest networks (Figure 3.3cd), followed by ppcor, GRNBOOST2, and SCING. Since SCING aims to find robust edges, as expected, SCING networks have much fewer edges than the other approaches while keeping similar numbers of genes (Figure 3.2h). One exception is that the ppcor oligodendrocytes network contains only 214.8 edges on average compared to 14,545.8, 46,891.9, and 449,850 in SCING, GRNBOOST2, and PIDC, respectively. Many genes without regulatory edges were not included in the final ppcor network. The smaller ppcor oligodendrocyte network has implications for the betweenness centrality and edge overlap metrics, as follows.

Betweenness centrality is often used as another metric to determine the overall connectedness of a graph[93]. For a given node, betweenness centrality is the number of shortest paths that pass through that node, indicating how much information that node presents to the graph. High betweenness centrality indicates that a node conveys a lot of information to a given graph. We found that SCING networks generally have higher variance of betweenness for the nodes in the networks (Figure 3.3e). This indicates that some nodes are more centralized than others when compared to other approaches, again consistent with the scale-free network model.

To determine network consistency, we split each cell type into two groups of non-overlapping cells. We built networks for each dataset using all methods and calculated the fraction of total edges that overlap between the two networks. While larger networks tend to have more edge overlap, this also holds for larger random networks. We designed a normalized overlap score: the

fraction of edges overlapped divided by the expected number of overlapping edges of a random network of the same size (Methods). When controlling for network size, SCING has significantly more overlap, or higher reproducibility, between networks of 50/50 split data than the other approaches (Figure 3.3f).

SCING more accurately models disease subnetworks

To evaluate the performance of GRNs on disease modeling, we applied an approach developed by Huang et al[94]. Briefly, given a known disease gene set and a GRN, we evaluate the ability of the GRN to reach held out disease genes by starting from select disease genes in the network through random walks. We then compared the performance for each network to that of a random network in which the nodes follow similar degree characteristics to derive a performance gain measurement. Here, we selected known gene sets for 3 classes of diseases from DisGeNET (Immune, Metabolic, and Neuronal) (Supplementary Table 3.1) and obtained scRNAseq data for cell types from 3 tissues relevant to each disease class from the mouse cell atlas[86] (bone marrow for immune diseases, brain for neuronal diseases, and liver for metabolic diseases) (Supplementary Table 3.2). First, to reduce the number of genes, we filtered the scRNAseq data by removing genes expressed in fewer than 5% of cells and added expressed disease genes from all DisGeNET disease gene sets. We built GRNs using each method and evaluated the performance gain over random networks on the disease gene sets. We found that across all tissues and all disease types, SCING outperformed all other approaches (Figure 3.4a).

Application case 1: constructing SCING GRNs using Mouse Cell Atlas (MCA) scRNAseq datasets to interpret diseases

After establishing the performance of SCING GRNs using the various approaches described above, we established a SCING GRN resource for diverse cell types and tested the broader utility of SCING to produce biologically meaningful GRNs. To this end, we applied SCING to generate

GRNs for all cell types with at least 100 cells in all tissues of the MCA. We constructed a total of 273 cell-type specific networks, across 33 tissues and 106 cell types. To identify which GRN informs on which disease, we applied the above random walk approach from Huang et al. and summarized the results in (SF3.3abc, Supplementary Table 3.3). We found clusters of cell type GRNs defined by DisGeNET diseases that had similar patterns (SF3.3a, Figure 3.4b). Some disease genes can be modeled well using GRNs from numerous cell types (SF3.3a) whereas others are more cell type or tissue specific (SF3.3b). Additionally, some cell type GRNs are able to model a broad range of diseases (SF3.3ac). We found that immune cell type (light blue squares in Figure 3.4b) GRNs can model a wide range of diseases, whereas non-immune cell type GRNs (light purple squares in Figure 3.4b) are more specific to vasculature related diseases.

We further explored the dynamics of GRNs of immune cell types across all diseases in DisGeNET. We clustered the cell types in the performance gain matrix with only immune cell types included, and sorted the GRNs by the number of diseases they can accurately model (Figure 3.4cd, SF3.4ab). We noticed that cell types of the innate immune system can model a broader range of diseases than those of the adaptive immune system[95] (Figure 3.4e).

Our SCING cell type GRNs resource and the above patterns of relationships between cell type GRNs and diseases support the utility of the SCING cell type GRN in disease interpretation. The networks can be accessed at https://github.com/XiaYangLabOrg/SCING to facilitate further biological mining of complex diseases.

Application case 2: Using SCING GRNs to interpret Alzheimer's disease (AD)

We next applied SCING to a single nuclei RNAseq (snRNAseq) dataset from Morabito et al. that examined human prefrontal cortex samples from AD and control patients to evaluate the applicability of SCING GRNs in understanding AD pathogenesis[87]. We focused on microglia, due

to their strong implication in AD[96,97] and a better current understanding of the genes and biological pathways in microglia in AD, to demonstrate that SCING GRNs can retrieve known biology.

The SCING microglia GRN contained 10,159 genes and 63,056 edges. Using the Leiden clustering algorithm[98], we partitioned the SCING microglia GRN into 21 network modules. Next, we summarized module-level expression for each cell using the AUCell method from SCENIC on the partitioned GRN modules[79]. When cells were clustered based on the raw gene expression values, as is typical with human samples, cells from individual samples clustered together (Figure 3.5a), making it difficult to isolate sample heterogeneity from biological variability. However, when using SCING module expression to cluster cells, the sample, batch, and RNA quality effects were mitigated (Figure 3.5ab). In contrast, biologically relevant variation, such as sex, AD diagnosis, and mitochondrial fraction were better retained (Figure 3.5cd). In the UMAP control cells tend to localize to the right side, while cells from females tend to localize to the top part. These results suggest that SCING GRNs have intrinsic ability to correct for non-biological variations.

To quantitatively evaluate how well SCING GRNs can be used for batch effect correction and biological preservation, we compared SCING GRNs with commonly used batch correction methods, such as FastMNN[99], Harmony[100], and Seurat[40], chosen based on their better performance in previous benchmarking studies[39]. We first performed dimension reduction and clustering of cells based on corrected data from each batch correction method. For SCING, the values used were SCING GRN module AUCell scores. We took each cell, determined how many neighbors in the PC space had the same annotation of interest (sample, batch, diagnosis, etc.), and then scored each batch correction approach by the fraction of cells that had the same annotation. We removed batch and sample specific effects using the F1 score (Methods)[39]. We found that the SCING GRN module based dimensionality reduction carried the ability to correct for batch effect and retain biological information (Figure 3.5ef) in a similar manner to dedicated

batch correction methods such as FastMNN[99], Harmony[100], and Seurat[40]. Although SCING GRN based batch correction was not as optimized as the dedicated batch correction methods, SCING is unique in that each GRN module has direct biological interpretability, since each SCING module can be associated with phenotypic traits and annotated with pathways, as described below.

We identified SCING GRN modules associated with AD diagnosis, plaque stage, and tangle stage through linear regression of the phenotypic traits and each module's expression across cells, while regressing out sex specific differences. We found that ~43% of modules were significantly (FDR < 0.05) associated with at least one trait and ~24% of modules were significantly associated with all three traits. To examine the biological interpretation of these modules, we performed pathway enrichment on the genes in each module utilizing the GO biological process, DisGeNET, Reactome, BioCarta, and KEGG knowledge bases. We found that 78% of the significantly trait-associated modules were significantly enriched for biological pathways (Figure 3.5g, Supplementary Table 3.4). These modules recapitulated pathways related to Alzheimer's disease (module 9), immune processes (modules 0, 2, 9, 13, and 19), cytokine triggered gene expression (modules 12, 18, 19), and endocytosis (modules 2, 9, and 13). These are expected perturbed pathways for microglia in AD[96].

We dug into the vesicle mediated transport pathway from module 2 and visualized the network with the differential gene expression of each gene (SF3.5). This microglia pathway is important in AD[101]. Additionally, we found the APOE and APP subnetwork within the vesicle mediated transport pathway which are among the most significant genetic risk factors for AD (Figure 3.5h)[102–104].

Therefore, our results demonstrate that SCING GRNs can correct for batch effects intrinsic to scRNAseq studies and can recapitulate known cell type specific genes, pathways, and network connections.

Application case 3: Using SCING to model GRNs based on 10x Genomics Visium spatial transcriptomics data to interpret AD

To evaluate the applicability of SCING beyond scRNAseq and snRNAseq, we next applied SCING to spatial transcriptomics data from Chen et al.[7] as a new approach for spatial transcriptomics analysis. We built a SCING GRN on all spots in the visium data to obtain a global GRN with 128,720 edges across 15,432 genes. Additionally, Chen et al. profiled beta amyloid plaque with immunohistochemistry, and we included this protein expression value in the SCING GRN, highlighting the possibility of constructing multi-omics networks using SCING.

We partitioned the genes in the resulting SCING GRN with the leiden graph partition algorithm into 33 modules and performed the AUCell score from SCENIC[79] to obtain module specific expression for each spot and annotated the enriched pathways for each module (Supplementary Table 3.5). We found module specific expression in the mouse brain subregions (Figure 3.6ab, SF3.6), based on clustering of the average module expression across spots in each brain subregion (SF3.7). We found cortical subregions to cluster together, as well as the thalamus and hypothalamus (SF3.7), based on GRN module expression patterns. We also identified modules more specifically expressed in the cortex and hippocampus (CS, HP) (module 12), or the fiber tract, thalamus, and hypothalamus (BS) (module 14) (Figure 3.6ab). Module 12 was highly enriched for genes involved in neuronal system, axonogenesis, and chemical synapse, which might reflect the dynamic status of hippocampus and cortex neurons for memory formation and cognitive function. In contrast, module 14 was enriched in genes involved in myelination[105] and blood-brain barrier[106], consistent with the high enrichment of oligodendrocyte populations in the

fiber tracts, and indicated blood-brain barrier changes in the thalamus. We also found modules that separate more similar subregions from one another such as module 27 (enriched for axonogenesis) and 21 (enriched for calcium ion transport) expressed much higher in the thalamus than in the hypothalamus (SF3.7). By contrast, modules, such as module 5 (enriched for chromatin organization and peroxisomal lipid metabolism) and module 19 (enriched for ribosomal biogenesis and protein processing), are much less specific to subregions (SF3.6, SF3.7). These are general cellular functions that are expected to have broad expression across the brain.

We found many modules to be AD associated (Figure 3.6c), in particular, module 9 (enriched for neurodegeneration) (Figure 3.6d) and module 25 (enriched for microglial activation, lysosome, and cell migration) (Figure 3.6e). We further explored these subnetworks and found most of the genes in module 9 to be neuronal marker genes, while most of the genes in module 25 to be microglial marker genes (SF3.9). The module 25 subnetwork to contain the *Trem2* and *C1q* subnetworks, highly profiled in AD microglial cells (Figure 3.6f). Interestingly, we found a cross-module edge and several cross-cell-type edges, likely revealing intercellular communications.

We also identified all but two modules that were significantly correlated with amyloid beta plaque (Figure 3.6g). Modules most highly correlated with plaque were module 25 (enriched for immune function), module 19 (enriched for nonsense mediated decay, and infectious disease), and module 9 (enriched for neurodegeneration), which were all also associated with AD (Figure 3.6c).

We found that amyloid plaque staining could be partitioned based on the expression of module 10 in the SCING GRN (SF3.8), which had significantly higher expression in AD mice than in WT (Figure 3.6c), and is also quantitatively correlated with plaque (Figure 3.6g). Module 10 contains genes highly related to metabolism, neurodegenerative diseases, and immune function (Supplementary Table 3.5). We show the subnetwork for the amyloid beta plaque in module 10

(SF3.9). Additionally, module 25 (enriched for microglial activation, lysosome, and cell migration) was also highly associated with plaque (Figure 3.6g), as expected in AD pathogenesis.

In addition to plaque association, we explored the age related modules through pearson correlation of age (6, 12, and 18 months) and module expression in WT samples. We found most modules to be age related (Figure 3.6h); however, some modules such as module 21 (enriched for muscle contraction and calcium ion transport) have a lower correlation coefficient. The top 2 positively correlated modules with age, module 10 (subnetwork with amyloid protein) and 13, have age related pathways. Module 10 is involved in protein degradation, neurodegeneration, and cell cycle, and module 13 is involved in metabolism, cellular response to stress, and immune system (Figure 3.6h, Supplementary Table 3.5).

We found certain modules such as module 30 (enriched for neuropeptide signaling) to be spatially variable (SF3.7) and associated with AD (Figure 3.6b). Based on the locations from the Allen Brain Atlas[107] (SF3.10a), we find module 30 to be more specific to the hypothalamus than other regions of the BS (SF3.10b). We also find that within the hypothalamus, module 30 expression is higher in 18 month old AD mice compared to the WT mice (SF3.10c). Hypothalamic alterations have been observed in the hypothalamus in AD development[108]. The module 30 subnetwork (SF3.10d) shows key hypothalamic neuropeptides, such as Pomc[109] and Pnoc[110] which is consistent with their enrichment in the hypothalamus.

Our applications of SCING to spatial transcriptomics data demonstrate its broader utility beyond scRNAseq/snRNAseq and revealed spatial network patterns of AD.

3.3 Discussion

Single cell multiomics has become a powerful tool for identifying regulatory interactions between genes, but the performance of existing tools is limited in both accuracy and scalability[73]. Here, we present SCING, a gradient boosting, bagging, and conditional mutual information based approach for efficiently extracting robust GRNs on full transcriptomes for individual cell types. We validate our networks using a novel perturb-seq based approach, held-out data prediction, and established network characteristic metrics (network size, network overlap, scale-free network, and betweenness centrality) to determine performance and network features against other existing tools. SCING not only offers robust and accurate GRN inference and improved gene coverage and speed compared to previous approaches[70,78], but also versatile GRN inference with scRNAseq, snRNAseq, and spatial transcriptomics data. Using various application examples, we show that SCING infers robust GRNs that inform on cell type specific genes and pathways underlying pathophysiology while simultaneously removing non-biological signals from data quality, sample, and batch effects through gene regulatory module detection and functional annotation. We also provide a comprehensive SCING GRN resource for 106 cell types across 33 tissues using data from MCA to facilitate future applications of single cell GRNs in our understanding of pathophysiology.

SCING efficiently identifies robust networks using supercells, a bagging approach, and mutual information based edge pruning, to remove redundant edges in the network. The supercell and reduction in potential edges make the bagging approach possible by removing computational time for each GRN. GRNBOOST2 uses a similar framework as SCING (gradient boosting regression) but overfits the data by generating too many edges to be interpretable, which are also undirected, making them less useful for biological interpretation. Meanwhile, ppcor and PIDC use partial correlation and partial information decomposition approaches, which are more accurately described as measures of coexpression, rather than gene regulation. In contrast, the directed graphs built with SCING show better perturbation prediction and consistency across replicates.

Validation of GRN inference tools has remained challenging[73]. Our novel perturb-seq based approach provides a unique way to determine GRN accuracy. Prediction of perturbed genes is a very powerful aspect of GRN construction, and SCING stands out above all other methods in this regard[111].

We demonstrate that SCING networks are applicable to scRNAseq, snRNAseq, and spatial transcriptomics data. Our network-based module expression provides batch corrected, biologically annotated expression values for each cell that can be directly used for disease modeling (Figure 3.4), phenotypic correlation (Figure 3.5), and spatially resolved analysis (Figure 3.6) to boost our ability to interpret single cell data.

At the intersection of single cell omics and complex diseases, SCING provides sparse but robust, directional, and interpretable GRN models for understanding biological systems and how they change through pathogenesis. GRNs can be analyzed to identify and predict perturbed subnetworks, and as a result, be used to investigate key drivers of disease[112]. Identifying key drivers of disease by teasing apart biology and technical variation from high throughput, high dimensional datasets will lead to more successful drug and perturbation target identification, as well as robust drug development[113–115].

We note that SCING is currently tested to infer GRNs based on individual scRNAseq, snRNAseq, and spatial transcriptomics data. Other types of omic information such as scATACseq, scHi-C, or cell type specific trans-eQTL information can be included in SCING to further inform on regulatory structure to refine and improve on GRNs[70,116–119]. Information from multiple data types will become an integral part of the systems biology, and future efforts to properly model multiomics data simultaneously to inform on complex disease are warranted.

3.4 Materials and Methods

SCING method overview

To reduce the challenges from data sparsity from single cell omics, as well as reduce computational time, we first used supercells which combine gene expression data from subsets of cells sharing similar transcriptome patterns. To improve the robustness of GRNs, we built GRNs on subsamples before merging the networks, keeping edges that appear in at least 20% of networks. Lastly, we removed cycles and bidirectional edges where one direction was >25% stronger than the other direction and pruned the network using conditional mutual information to reduce redundancy. These steps are described in more detail below. To benchmark the performance of SCING, we selected three existing methods, namely GRNBOOST2, ppcor, and PIDCm with default parameters. These methods were selected based on previous benchmarking studies where superior performance of these methods were supported[73].

Supercell construction

For each dataset, we first normalized the data for a total count number of 10,000 per cell[120]. We then took the log of the gene expression values, identifying and subsetting to the top 2000 highly variable genes. Data were then centered and scaled and used to compute the nearest neighbor embedding with 10 neighbors. We then used the Leiden graph partitioning algorithm to separate cells into groups. The leiden resolution is determined by the user input specifying the final number of supercells. Here, we used 500 supercells which balances runtime, with dataset summarization. We then merged each group of cells into a supercell by averaging the gene expression within each group.

GRN inference in SCING

We normalized the total number of counts in each cell to 10,000 and took the natural log of the gene expression. We removed genes not expressed in any cells and any duplicate genes. We transposed, centered, and scaled the data before running principal component analysis (PCA). This provides us with low dimensional embeddings for each gene. A nearest neighbor algorithm from scikit-learn[121] was used to find the nearest neighbors of each gene. The potential regulatory relationship between genes was limited to the 100 nearest neighbors. For each gene, we trained a gradient boosting regressor[78] to predict the expression based on its nearest neighbors. For each gradient boosting regressor, we used 500 estimators, max depth of 3, learning rate of 0.01, 90 percent subsample, the square root of the total number of features as max features, and an early stop window of 25 trees, if the regressor was no longer showing improved performance[78].

SCING parameter selection

Parameter selection was used to retain biological accuracy while limiting computational cost. We determined that using 500 supercells and 100 GRNs per merged network balanced computational resourcefulness with robustness. If computational cost is not an issue, supercells are not necessary and more GRNs can be built as intermediates. Since GRNs are typically built within a single cell type, we use 10 principal components (PCs) for determining gene covariance. Typically in scRNAseq analysis more PCs are used, but there is less variation overall within one cell type. We determined 100 neighbors to be chosen for each gene, again balancing computational cost.

Merging GRNs from data subsamples

For each network from subsampled data, we kept the top 10 percent of edges based on the edge weight, as well as the top 3 edges for each downstream gene and edges that appeared in greater than 20% of all networks. We also removed reversed edges if the edge with a higher weight was at least 25% stronger than that of the reverse direction. Otherwise, we kept the edge bidirectional. We removed cycles in the graph by removing the edge with the lowest edge weight. We removed

triads in the network based on the significance of the conditional mutual information. The p-value of the conditional mutual information is based on the chi-squared distribution[122]. If an edge between two genes is not statistically significant given a parent of both of the genes, then the edge is removed.

Other GRN methods

We benchmarked SCING against GRNBOOST2, ppcor, and PIDC. Default parameters were used for all existing approaches unless otherwise specified.

For GRNBOOST2[78], we ran this approach by predicting the expression of all genes from all other genes. We then took the top 10% of edges to reduce the number of edges with extremely low importance (e.g. 10^{-17}).

For ppcor[76], due to the sparse non-linear nature of scRNAseq connections, we ran the approach using spearman correlation. We only kept edges with a Benjamini-Hochberg FDR < 0.05.

For PIDC[77], according to their tutorial, we used a threshold of 0.1, to keep the top 10% of highest scoring edges.

Datasets

For the perturb-seq validation, we used datasets from Dixit et al.[91] and Papalexi et al.[90]. Dixit et al. has 24 transcription factors perturbed in dendritic cells and 25 cell cycle genes targeted in K562 cells, while Papalexi et al. has 25 PD-L1 regulators perturbed in THP-1 cells.

For the train-test split and network consistency assessment, we used the human AD snRNAseq data from Morabito et al.[87]. This adds another slightly different data type from the scRNAseq in the perturb-seq and MCA and is later used for biological application with microglial cells in AD.

We used the mouse cell atlas[86] scRNAseq database, since it has a large number of cell types (106 cell types) across numerous tissues (33 tissues), to test 446 disease associations with the random walk approach from Huang et al.[94] This additionally provides a resource of GRNs throughout cell types of the entire mouse.

Finally, to test the applicability of SCING on spatial transcriptomics data, we used the mouse AD dataset from Chen et al.[7] This dataset contains AD and WT mice from various age groups (3, 6, 12, 18 months), in addition to amyloid beta plaque staining.

Computation of time requirements

We determined the run time to build a GRN from each approach on subsets of cells and genes with varying cell numbers and gene numbers. All tests were performed on a ryzen 9 3900X 12-Core processor with 64Gb RAM. We determined the speed on 10 iterations of randomly selected genes (1000, 2000, 4000) with 1000 cells each, and on randomly selected cells (250, 500, 1000) with 1000 genes each.

Overview of network robustness evaluation based on Perturb-seq datasets

Briefly, to determine the accuracy of the GRNs from each approach with perturb-seq data, we first identified significantly altered genes downstream of each guide RNA perturbation through an elastic net regression approach[91], as detailed below. We then determine the accuracy of a given network by identifying the true positive rate (TPR) and false positive rate (FPR) at each depth in

the network. The AUROC and TPR at FPR 0.05 were determined for each network on a given perturbation. More details on each step are below.

Computation of guide RNA perturbation coefficients

First, we downloaded the perturb-seq data from Papalexi et al.[90] and Dixit et al[91]. We then followed the steps as described by Dixit et al[91] to compute guide RNA perturbation coefficients, which indicate the effects of a specific guide RNA (single perturbation) on other genes. As cell state can affect gene perturbation efficiency, to determine cell states and remove state specific perturbations, we clustered the non-perturbed cells in each dataset through Leiden clustering. The leiden resolution was determined by identifying unique subclusters in the data (Table 3.3). We subset the data to highly variable genes (min_mean=0.0125, max_mean=3, min_disp=0.5). We centered and scaled the data and performed PCA to get the first 50 PCs. We then trained a linear support vector machine (C=1) on the 50 PCs of the data and determined the probability of each cell in the dataset being from each state. We used these continuous state probabilities in our regression equation to regress out state specific effects on gene expression. To identify the perturbation effect of a given guide RNA on genes other than the target gene, we utilized elastic net regression (l1_ratio=0.5, alpha=0.0005)[91]. We fit the elastic net model to predict the gene expression of all genes from the binary matrix determined by the guide sequenced in each cell, combined with the continuous state values determined for each cell. To remove the effect of synergistic perturbations, we removed cells with multiple perturbations. We determined each guide's perturbation effect on a given gene by the regression coefficient.

Determination of significant perturbation effects

To determine the significance of a given perturbation coefficient, we employed a permutation test as in Dixit et al[91]. For each guide, we permuted the vector of perturbations to randomize which

cells received the given perturbation of interest. The elastic net regression model was trained with the same hyperparameters to determine the coefficients of perturbation. This approach was repeated 100 times to generate a null distribution of the perturbation effect of a given guide on each gene. The p-value was calculated as the fraction of null coefficients that were greater than or less than the true coefficient, determined by the sign of the coefficient. Significant perturbations were determined at a false discovery rate of 0.05 using the Benjamini-Hochberg procedure. This permutation approach was repeated for each guide RNA. A gene was determined as a downstream perturbation if at least one guide had a significant perturbation for the given gene.

Selection of genes and cells for perturb-seq networks

To reduce computational cost and enable network building for all approaches, we first took the top 3,000 highly variable genes using the variance stabilizing transform method[123], including the differentially perturbed genes. We built two networks for each dataset, one using all cells and the other using only cells with non-zero expression of the gene of interest.

Evaluation of perturbation predictions

We built networks for each perturbed gene separately using SCING, GRNBOOST2, ppcor, and PIDC. Starting from the perturbed gene of interest, at each depth in the network, we determined the TPR and FPR based on the perturbed genes computed above. This gives a TPR vs FPR graph, from which an AUROC was computed. For each perturbed gene, we calculated the AUROC and the TPR at a FPR of 0.05.

Network robustness evaluation based on training and held out testing data

We performed a train test split (75/25) on the dataset from Morabito et al on each random subsample of 3000 genes. We built GRNs on the training data and trained a gradient boosting regressor for each gene based on the predicted regulatory parents. We used the trained gradient

boosting regressor to predict the expression of all genes in the test dataset and evaluated the performance based on the cosine similarity metric. We performed this on the training and testing data separately and computed the test to train ratio of the cosine similarity, with a smaller test to train ratio indicating potential overfitting of the training data.

Computation of network characteristics

We built GRNs on oligodendrocytes, astrocytes, and microglia from snRNAseq data from Morabito et al[87]. For each cell type, we randomly selected 3000 genes (reduce computational time of methods) for each sample and generated 10 GRNs, in which 3000 genes were randomly subsampled from the full transcriptome. To compute scale-free network characteristics for each network, we fit a linear regression model on the log of each node degree with the log of the proportion of nodes at each degree. We removed low degree data points that are an artifact of scRNAseq sparsity. We also characterized each network by the number of edges in the network, number of genes remaining in the resulting network, and the mean betweenness centrality of nodes across the network.

Computation of network overlap

We used the Morabito et al datasets described above and split each dataset in half and generated GRNs on each subset of cells. We checked the overlapping edges between the two networks and normalized for the expected number of overlapping edges based on the number of total edges in each network and the hypergeometric distribution. The overlap score measured the fraction of overlap between the two networks, divided by the expected number of overlapping edges.

Assessment of disease subnetwork retrieval of GRNs

We utilized a random walk approach from Huang et al[94] to determine the ability for GRNs from different methods to accurately model disease gene subnetworks. This approach provides a

biologically relevant benchmarking approach to determine a GRNs ability to model disease subnetworks. Briefly, the approach splits a known disease gene set into two groups, to attempt to reach the held out gene set starting from the selected disease genes through random walks. An improvement score is computed by calculating the z-score for a given network relative to 50 degree-preserved randomized networks.

We built networks from the MCA on immune cells from bone marrow, neurons from the brain, and hepatocytes from the liver. To accommodate less efficient tools, we subsetted the transcriptome to genes that are expressed in more than 5% of the cells in the dataset. We used the method from Huang et al. using relevant immune, neuronal, and metabolic disease gene sets from DisGeNET. We kept these genes with >5% percent expression and included the genes from the disease gene sets. We determined performance of each subnetwork based on the improved performance compared to the random network distribution.

Application of SCING to construct GRNs for all MCA cell types and assessment of network relevance to all DisGeNET disease gene sets

We applied SCING to all cell types for all tissues in the MCA and utilized the approach from Huang et al. to determine the ability of each network to accurately model each disease gene set. We clustered the disease gene sets and cell types using hierarchical clustering with complete linkage. We determined the number of disease sets accurately modeled by each cell type based on a performance gain of at least 0.1. We subsequently computed the number of cell types that can accurately model each disease set. To compare the number of diseases modeled by cell types from the adaptive and innate immune system on tissue relevant subsets of the DisGeNET diseases, we performed a t-test between the distributions of the number of disease gene sets each cell type can accurately model.

Biological application to microglia in Alzheimer's disease patients

We built a SCING network for the microglia on the genes expressed in at least 2.5 percent of cells in the Morabito et al. dataset[87]. For the SCING pipeline, we used 500 supercells, 70 percent of cells in each subsample, 100 neighbors, 10 PCs, and 100 subsamples. We utilized the Leiden graph partitioning algorithm to divide genes in the resulting GRNs into modules. We performed Leiden clustering at different resolutions and performed pathway enrichment analysis on the modules using the enrichr[124] R package, using the GO biological process, DisGeNET, Reactome, BioCarta, and KEGG knowledge bases. We selected the resolution (0.0011) that had the highest fraction of modules annotated for between 20 and 50 modules per network. This avoids clustering too many modules with few genes while maintaining enough separate modules to have biological interpretation. We used the AUCell method from the SCENIC workflow[79], to retrieve module specific expressions (AUCell scores) for each cell. We found trait (diagnosis, plaque stage, tangle stage) associated modules by fitting a linear regression model to predict the trait based on the module score, while regressing out the effects of sex. For each trait, multiple testing was controlled at FDR < 0.05 with the Benjamini-Hochberg procedure. The subnetwork for vesicle-mediated transport in module 2 was visualized using Cytoscape[125]. We determined marker genes using the Allen Brain Atlas whole brain Smartseq2 data[20].

Batch correction comparison

We compared top batch correction methods from Seurat, Harmony, and fastMNN with SCING module embeddings. To evaluate each method, we determined the average proportion of cells with the same group assignment (sample, batch, diagnosis, tangle stage, plaque stage, and sex), using 20 PCs and a variable number of neighbors (0.25, 0.5, 1, 2, 4, 8, and 16 percent of the dataset) (Equation 1). We determined the ability of each approach to remove batch and sample specific differences while retaining biologically relevant differences (diagnosis, tangle stage,

plaque stage, and sex) by removing the batch and sample differences with an F1-score[39] (Equation 2).

$$\text{neigh\_score} = \frac{1}{n_{cell}} \sum_{cells} \frac{\text{similar\_neighbors}}{\text{n\_neighbors})}$$

Equation 1.

neigh_score: neighborhood score used to find the average fraction of neighbors of the same type (i.e. batch).

similar_neighbors: number of neighbors of a given cell that have the same identity (i.e. batch)

n_neighbors: number of total neighbors checked

n_{cell}*:* total number of single cells

$$F1_{phenotype} = \frac{2 * (1 - \text{neigh\_score}_{sample}) * (1 - \text{neigh\_score}_{batch}) * (\text{neigh\_score}_{phenotype})}{(1 - \text{neigh\_score}_{sample}) + (1 - \text{neigh\_score}_{batch}) + (\text{neigh\_score}_{phenotype})}$$

Equation 2.

neigh_score: neighborhood score computed in Equation 1 for a given identity

Application of SCING to Visium spatial transcriptomics data for mouse AD and WT brain

To determine the applicability and interpretability of SCING to spatial transcriptomics data, we applied SCING to mouse whole brain AD and WT data[7]. Since the network was built on the whole brain rather than a single cell type, we expect more variance amongst networks from subsamples, therefore we built 1,000 GRNs to be merged into the final network. We partitioned the genes with the Leiden graph partitioning algorithm into 33 modules. Using AUCell from SCENIC[79], we obtained module specific expression for each spot. We determined regional specificity between pairs of larger regions (cortex, hippocampus, brainstem) through t-tests and overall variance for

the smaller subregions through ANOVA. We determine differential module expression between AD and WT through t-tests, and correlation with age or plaque with Pearson correlation. Finally, the module 9, 10, 25, and 30 subnetworks were visualized using Cytoscape[125]. We determined marker genes using the Allen Brain Atlas whole brain Smartseq2 data[20].

3.5 Figures

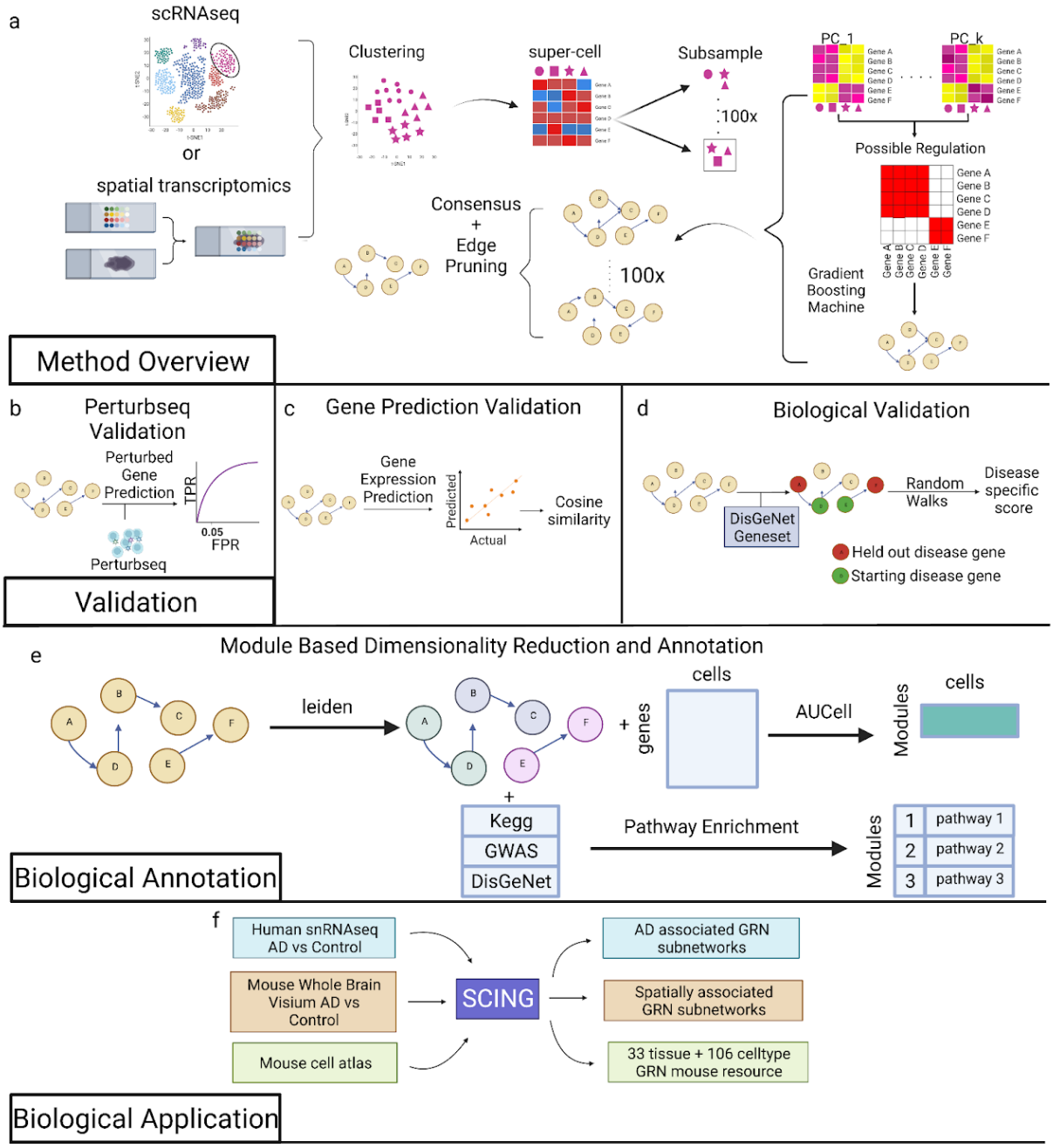

a
scRNAseq
or
spatial transcriptomics
Clustering
super-cell
Subsample
100x
PC_1
PC_k
Possible Regulation
Gradient Boosting Machine
Consensus + Edge Pruning
100x
Method Overview
b
Perturbseq Validation
Perturbed Gene Prediction
TPR
0.05
FPR
Perturbseq
c
Gene Prediction Validation
Gene Expression Prediction
Predicted
Actual
Cosine similarity
d
Biological Validation
DisGeNet Geneset
Random Walks
Disease specific score
Held out disease gene
Starting disease gene
Validation
e
Module Based Dimensionality Reduction and Annotation
leiden
cells
genes
AUCell
Modules
cells
Kegg
GWAS
DisGeNet
Pathway Enrichment
Modules
1 pathway 1
2 pathway 2
3 pathway 3
Biological Annotation
f
Human snRNAseq AD vs Control
Mouse Whole Brain Visium AD vs Control
Mouse cell atlas
SCING
AD associated GRN subnetworks
Spatially associated GRN subnetworks
33 tissue + 106 celltype GRN mouse resource
Biological Application

Figure 3.1. SCING overview, benchmarking, and application.

SCING overview (a). First, we select a specific cell type, or use whole visium data. We then cluster the cells using the leiden graph partitioning algorithm and merge subclusters into supercells. We utilize bagging through subsamples of supercells to keep robust edges in the final GRN. For each subsample, the genes are clustered to only operate on likely regulatory edges. We then identify edges through gradient boosting regressors (GBR). We find the consensus as edges that show up in 20% of the networks. We then prune edges and cycles using conditional mutual information petrics. perturb-seq validation (b). We identified downstream perturbed genes of guides for specific genes. We then predict perturbed genes at each depth in the network from the perturbed gene. True positive rate, and false positive rate are determined at each depth in the network. We utilize AUROC and TPR at FPR 0.05 as metrics for evaluation. Gene prediction validation, to determine network overfitting (b). We split data into train and test sets and build a network on the train set. A GBR is trained for each gene based on its parents in the train data. We then predict the expression of each gene in the test set and determine the distance from the true expression through cosine similarity. Biological validation through disease subnetwork modeling (c). We utilize a random walk framework from Huang et al. to determine the increase in performance of a GRN to model disease subnetworks versus random GRNs with similar node attributes (d). We utilize the leiden graph partitioning algorithm to identify GRN subnetworks. We combine these subnetworks with the AUCell method to get module specific expression for each cell and further combine the gene modules with pathway knowledge bases to annotate modules with biological pathways (e). We apply SCING to human prefrontal cortex snRNAseq data with AD and Control patients, whole brain visium data, for AD vs WT mice at different ages, and to the mouse cell atlas in 33 tissues and 106 cell types (f).

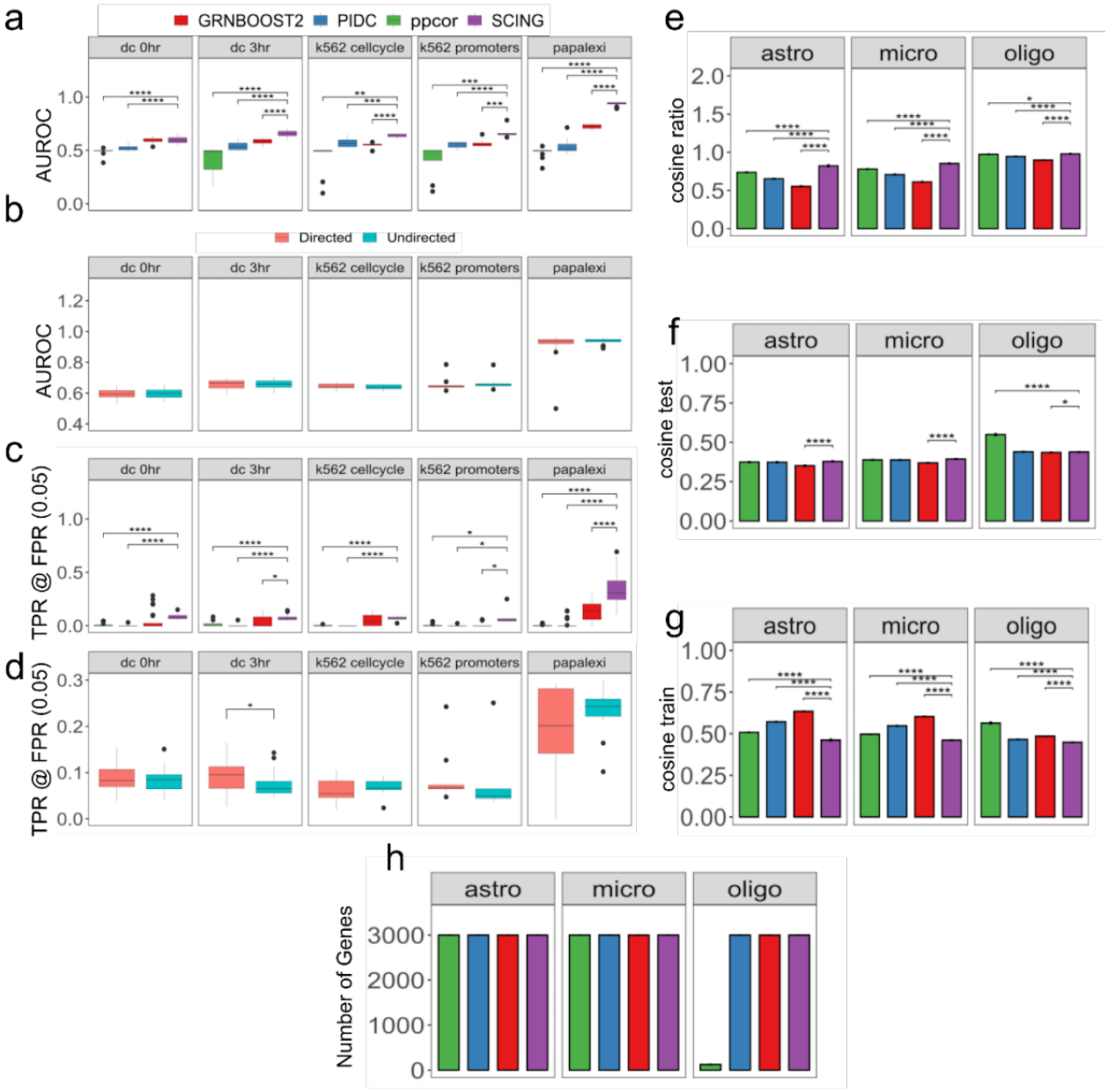


Figure 3.2. Performance evaluation.

Predicted downstream affected genes of perturb-seq based perturbation in 5 datasets with GRNs built on cells with non-zero expression of the perturbation of interest. Area under receiver operator characteristic (AUROC) curve for prediction of downstream perturbations using undirected GRNs (a). AUROC for prediction of downstream perturbation on directed GRNs for SCING (b). True

positive rate (TPR) at a false positive rate (FPR) of 0.05 for the prediction of downstream perturbations on undirected GRNs (c). TPR at FPR of 0.05 for the prediction of downstream perturbations on directed GRNs for SCING. (d) Measure of network overfitting by similarity of predicted gene expression and actual in held out data for astrocytes, microglia, and oligodendrocytes. Cosine similarity of predicted gene expression and actual in testing data show few differences between SCING and others (e). Cosine similarity of predicted gene expression and actual in training data shows other methods overfit to the training data (f). Ratio of test to train shows SCING does not overfit to the data as much as other approaches (g).

(*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$)

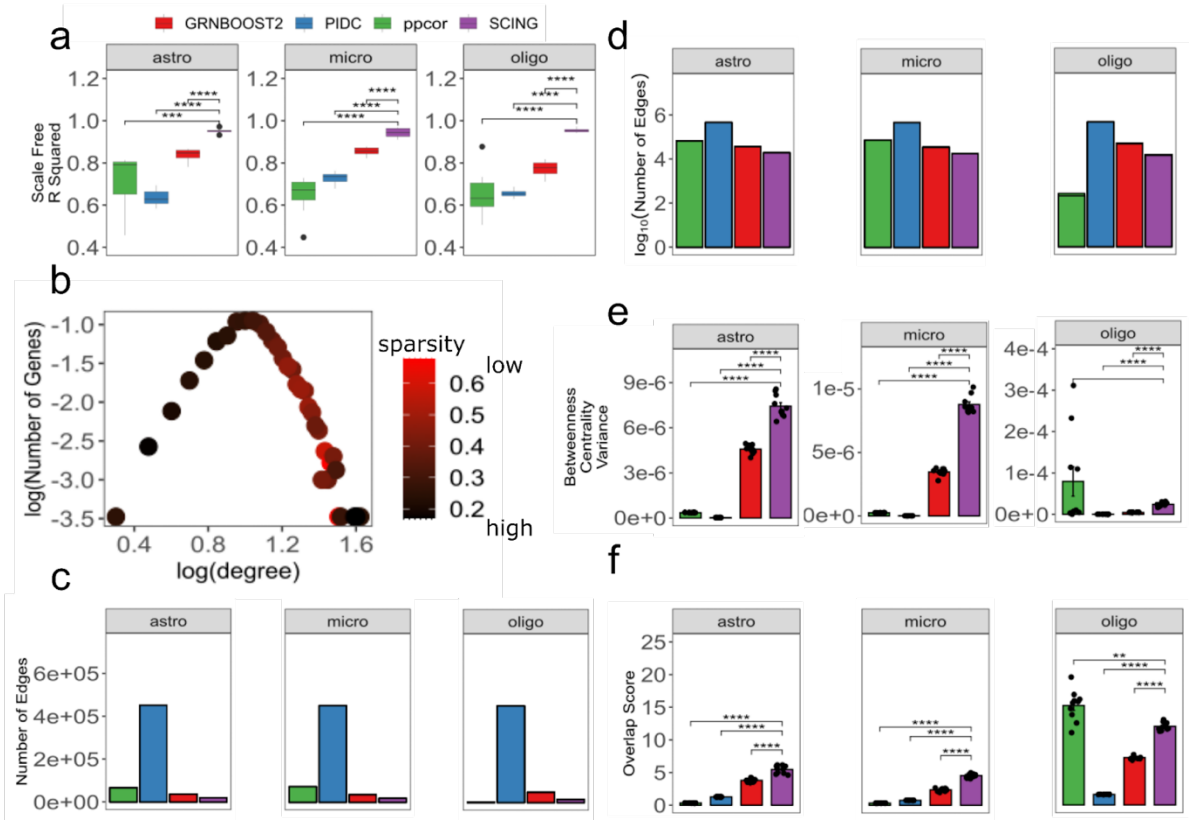


Figure 3.3. Network features and consistency.

Descriptive features of networks across SCING and other approaches for 10 networks on astrocytes, microglia, and oligodendrocytes. Linear regression R-squared for log degree vs log count for goodness of fit metric of scale-free network (a). Example scatter plot of log degree vs log count with the average sparsity of genes in each dot. Brighter red indicates less sparse. This shows highly sparse genes tend to have lower degrees (b). Average number of edges (c) and log of the number of edges (d) for each method across all cell types. The variance of the betweenness centrality across nodes in each graph (e). The overlap score (number of overlapping edges/expected number of edges) in independent sets of cells (f).

(*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$)

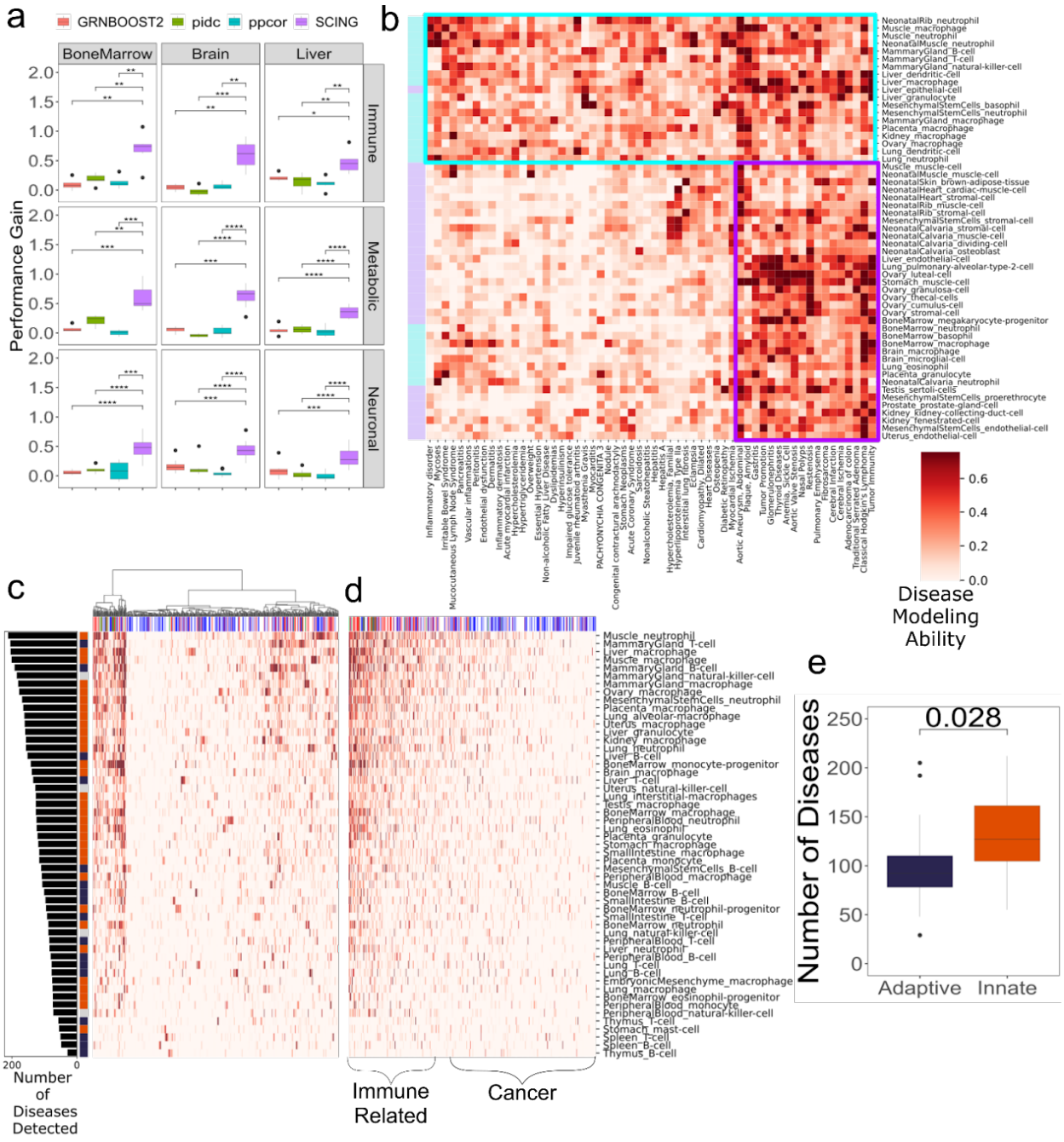


Figure 3.4. Application case 1: constructing and using SCING GRNs based on Mouse Cell Atlas scRNAseq datasets to interpret diseases.

Performance of modeling disease subnetworks for DisGeNET gene sets related to the immune, metabolic, and neuronal diseases with GRNs built on bone marrow, brain, and liver cells, reveals SCING models disease subnetworks more accurately than other methods (a). Clustermap

depicting GRNs built with SCING from immune cell types (light blue), model disease subnetworks from many different disease gene sets, while vascular cell types (purple) are more specific to vascular diseases (b). Cell types (rows) from the adaptive (blue) and innate (orange) immune systems, show variability in the number of diseases (columns) they model (>0.1). Clustermap shows diseases clustered with hierarchical clustering (c) and sorted by the number of cell types that can accurately model that disease subnetwork (d). Diseases are colored by disease category (immune related: red; cardiothoracic: green; cancer: blue; immune related cancer: purple), and cell types are colored by innate (orange), and adaptive immune system (dark blue). Innate immune system cell types better model disease subnetworks from more diseases (e).

(*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$)

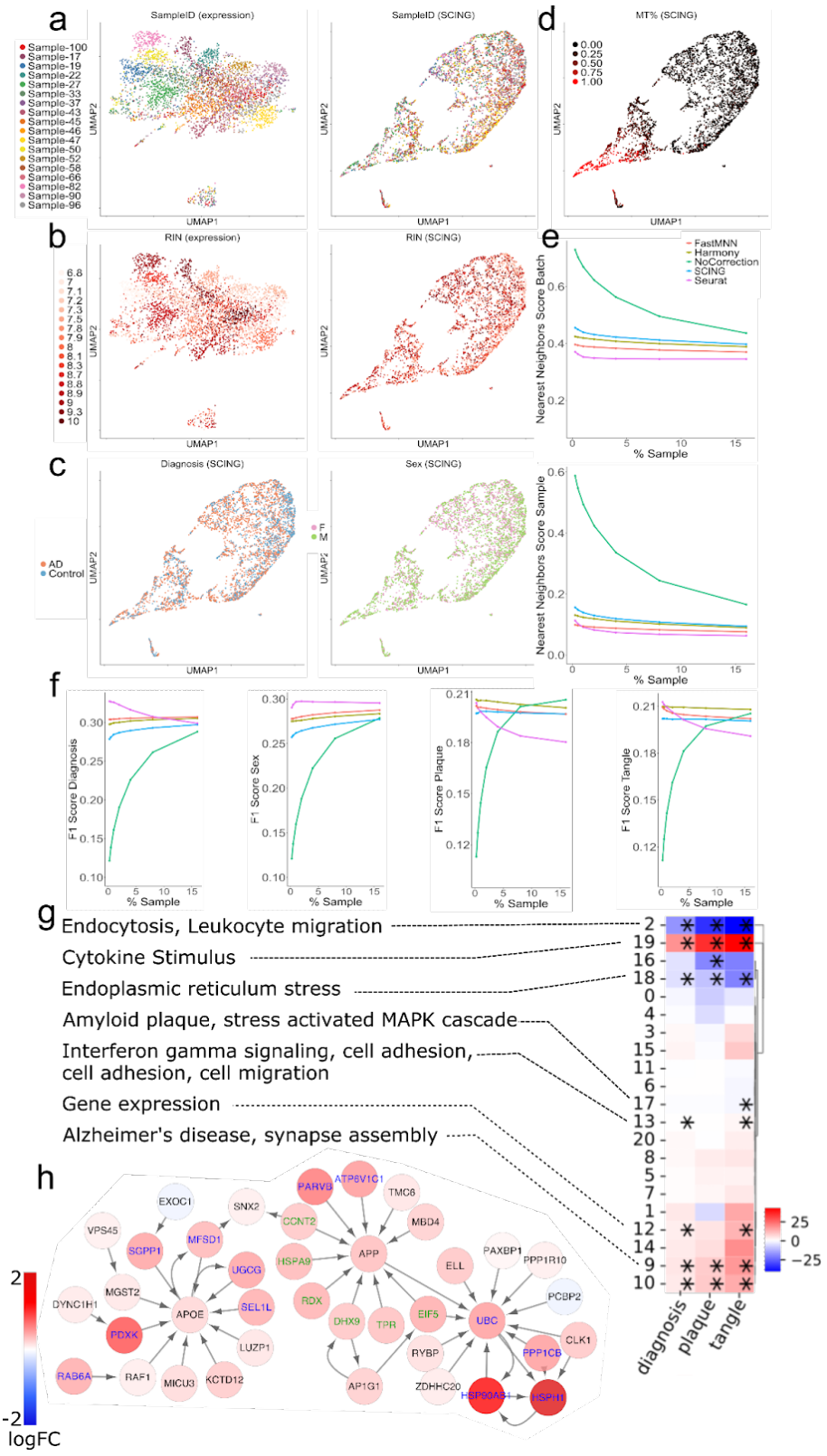

a
SampleID (expression)
SampleID (SCING)
d
MT% (SCING)
b
RIN (expression)
RIN (SCING)
e
FastMNN
Harmony
NoCorrection
SCING
Seurat
Nearest Neighbors Score Batch
% Sample
c
Diagnosis (SCING)
Sex (SCING)
AD
Control
Nearest Neighbors Score Sample
f
F1 Score Diagnosis
F1 Score Sex
F1 Score Plaque
F1 Score Tangle
g
Endocytosis, Leukocyte migration
Cytokine Stimulus
Endoplasmic reticulum stress
Amyloid plaque, stress activated MAPK cascade
Interferon gamma signaling, cell adhesion, cell adhesion, cell migration
Gene expression
Alzheimer's disease, synapse assembly
diagnosis
plaque
tangle
h
logFC

Figure 3.5. Application case 2: Using SCING GRNs to interpret Alzheimer's disease (AD).

UMAP representation of scRNAseq data shows sample specific differences when operating on gene expression space (left panel). Dimensionality reduction on SCING module embeddings removes sample specific effects (right panel). SCING removes RNA quality effects on gene expression clustering (b). Clustering on SCING modules, keeps biologically relevant features such as AD status (left panel) and sex (right panel) (c), as well as mitochondrial fraction (d). SCING corrects batch (top) and sample (bottom) specific effects based on the fraction of neighbors with the same batch or sample status (lower is better) for a given cell (e). This was performed at different fractions of the cells used as neighbors for a given cell. Dedicated batch correction techniques have better batch correction. F1 score of nearest neighbor score with batch and sample effects corrected (higher is better), shows SCING corrects sample, and batch effects while retaining biologically relevant features (f). Heatmap showing coefficients of linear regression of diagnosis, plaque, and tangle status, on module expression, while regressing out sex (g). Pathway annotations for significant (*: FDR < 0.05) modules are provided. Subnetwork reveals collocalization of canonical Alzheimer's genes and subnetworks such as APOE and APP (h). Colors indicate the linear regression coefficients between a phenotype and module expression.

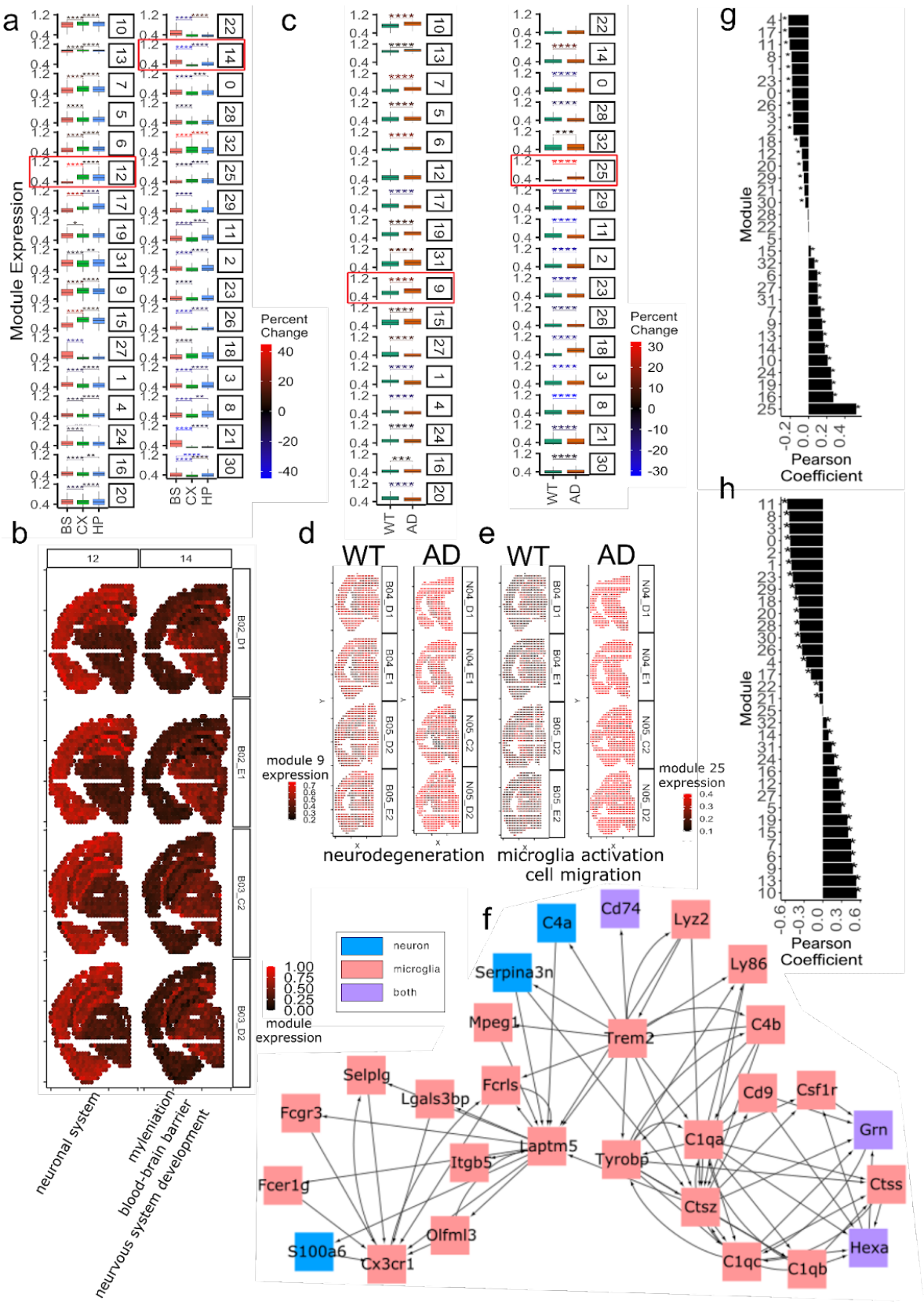
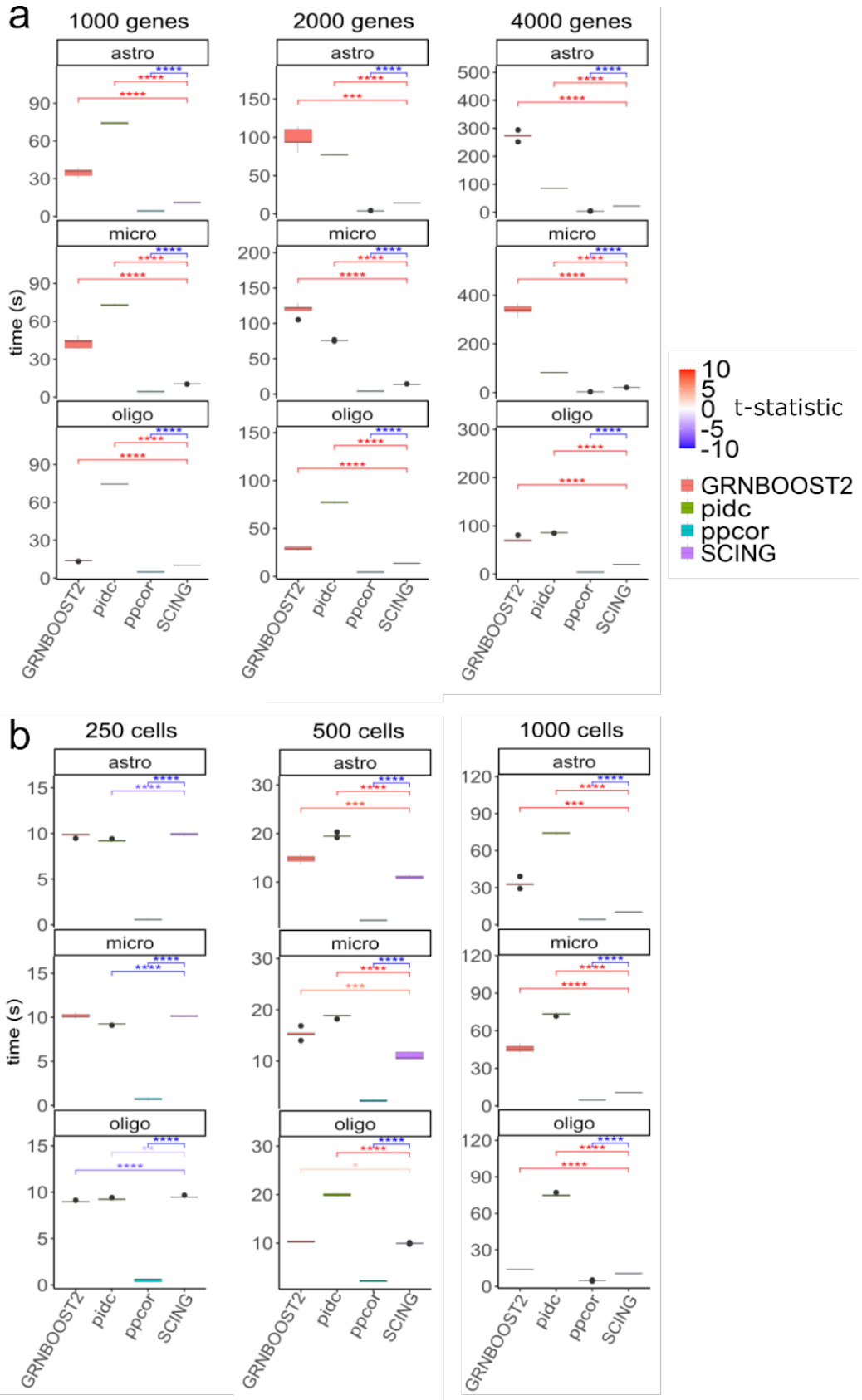

a
Module Expression
BS
CX
HP
Percent Change
40
20
0
-20
-40
c
WT
AD
Percent Change
30
20
10
0
-10
-20
-30
g
Module
Pearson Coefficient
h
Module
Pearson Coefficient
b
12
14
B02_D1
B02_E1
B03_C2
B03_D2
neuronal system
myleniation
blood-brain barrier
neurvous system development
module expression
d
WT
AD
B04_D1
B04_E1
B05_D2
B05_E2
N04_D1
N04_E1
N05_C2
N05_D2
module 9 expression
neurodegeneration
e
WT
AD
module 25 expression
microglia activation
cell migration
neuron
microglia
both
f
C4a
Cd74
Lyz2
Ly86
Serpina3n
Mpeg1
Trem2
C4b
Selplg
Fcrls
Lgals3bp
Fcgr3
Cd9
Csf1r
Grn
Laptm5
Itgb5
Tyrobp
C1qa
Fcer1g
Ctss
Ctsz
Olfml3
S100a6
Cx3cr1
C1qc
C1qb
Hexa

Figure 3.6. Application case 3: Using SCING to model GRNs based on 10x Genomics Visium spatial transcriptomics data to interpret AD.

Boxplots show regional specificity of specific modules, namely module 12 and 14 (red boxes) (a). Scale bar shows the t-test statistic of the test between regions. Regional specificity of module 12 (hippocampus and cortex) and module 14 (hypothalamus, thalamus, and fiber tract), visualized on brain samples of 3 month old WT mice (b). Module association in AD vs WT mice (c). Scale bar shows the t-test statistic of the test between mouse groups. Visualization of module 9 in 18 month old WT and AD mice (d). Visualization of module 25 in 18 month old WT and AD mice (e). Module 25 subnetwork of Trem2 and complement proteins shows microglial association of module 25 (f). Full subnetwork in supplement 9. Nodes are colored by marker gene status of neuronal (blue), microglial (red), or both (purple), as determined from the Allen Brain Atlas. Cross cell type communication edges seen between red and blue. Pearson correlation between module expression and plaque (g) and age (h), shows with significance (*: $p<0.05$)

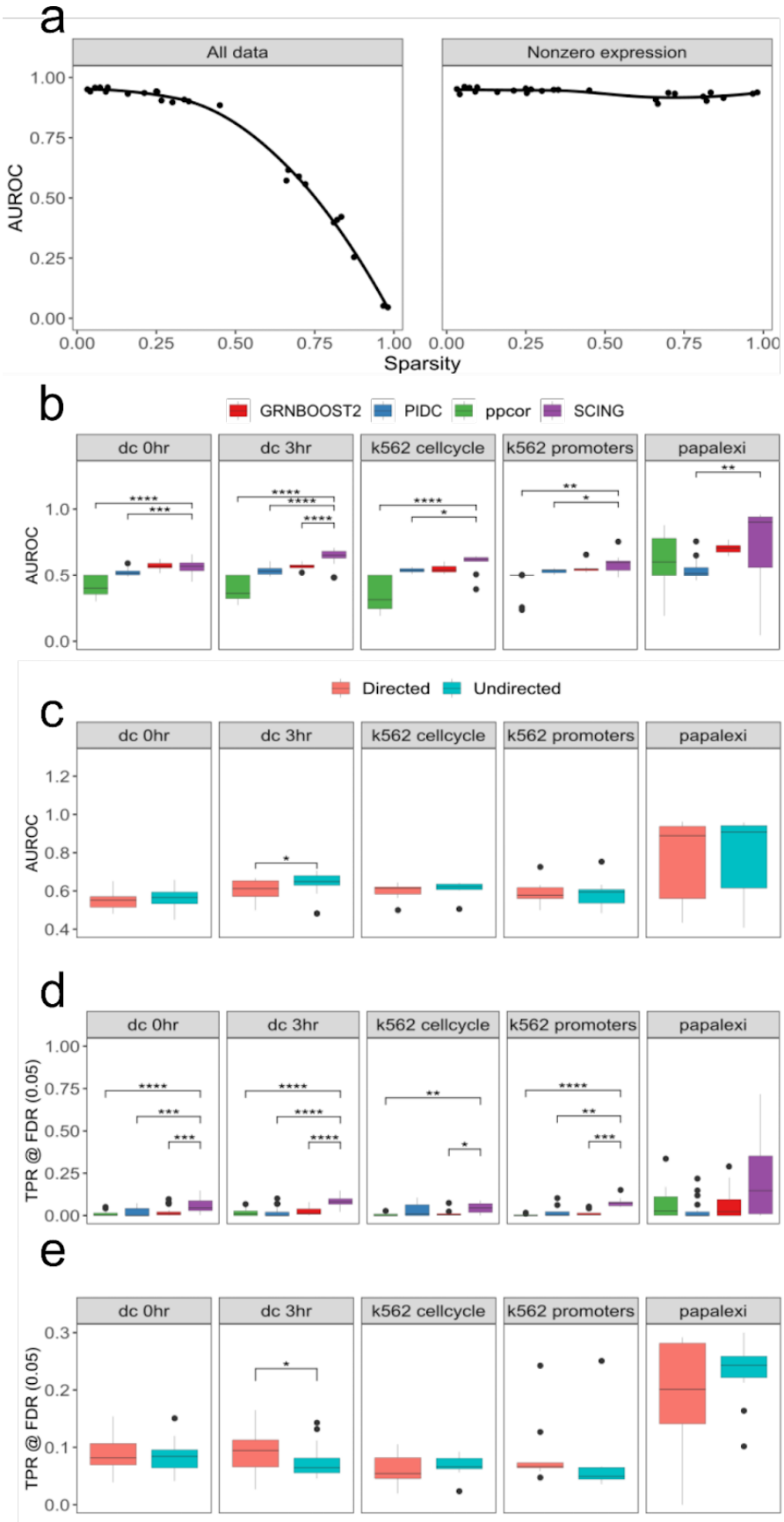
Subpanel a,c: (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$)


a
1000 genes
2000 genes
4000 genes
astro
micro
oligo
time (s)
GRNBOOST2
pidc
ppcor
SCING
10
5
0
-5
-10
t-statistic
b
250 cells
500 cells
1000 cells

SF 3.1. Run time comparison of methods for building a single GRN.

SCING is faster than GRNBOOST2 and PIDC as the number of genes increases (a). The test increasing the number of genes was performed on 1000 cells. SCING is faster than GRNBOOST2 and PIDC as the number of cells increases (b). The test increasing the number of cells was performed on 1000 genes. ppcor is faster than all other methods. Red significance indicates that SCING is faster, while blue indicates SCING is slower.

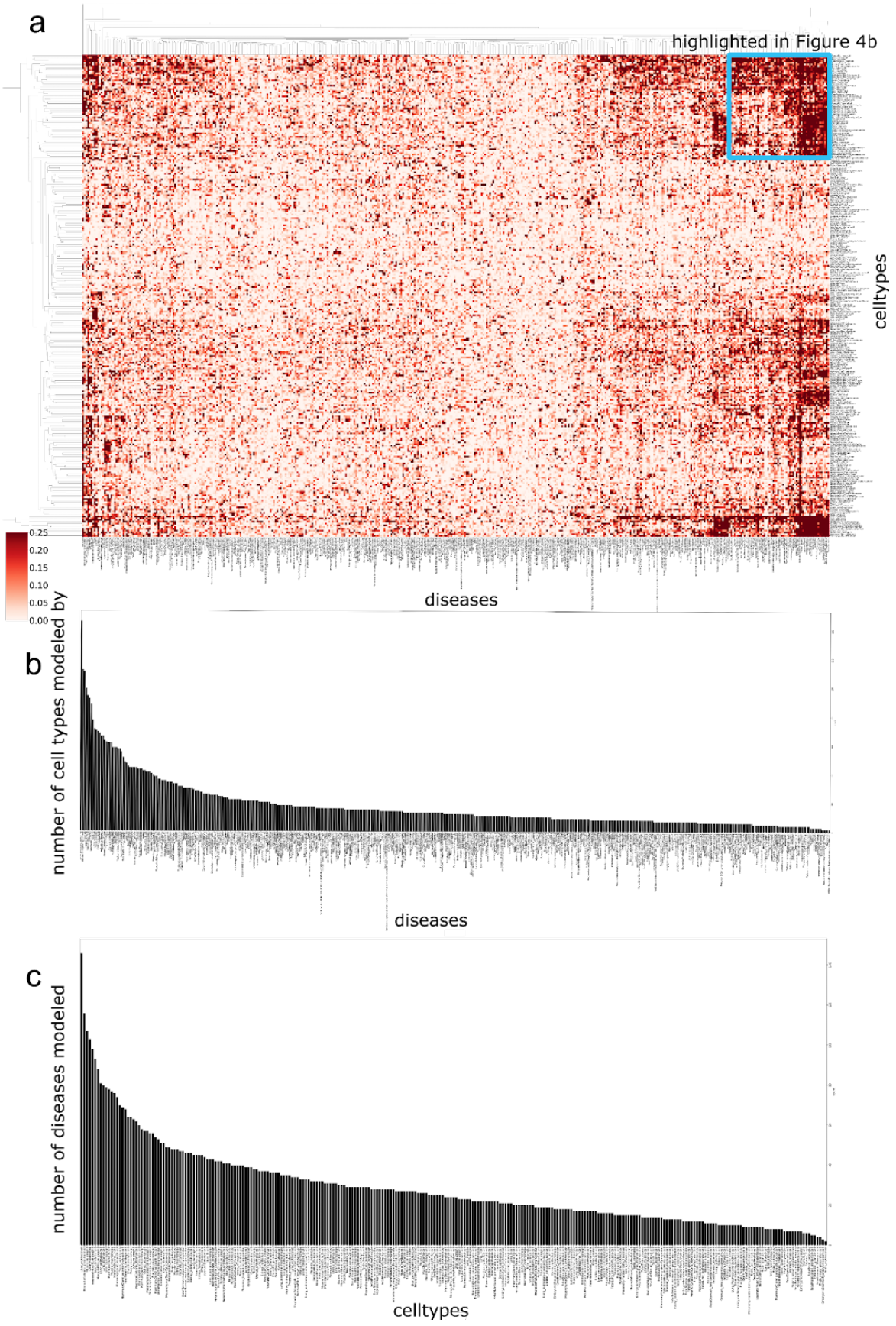
(*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$)


a
All data
Nonzero expression
AUROC
Sparsity
b
GRNBOOST2
PIDC
ppcor
SCING
dc 0hr
dc 3hr
k562 cellcycle
k562 promoters
papalexi
AUROC
c
Directed
Undirected
AUROC
d
TPR @ FDR (0.05)
e
TPR @ FDR (0.05)

SF 3.2. Predicted downstream affected genes of perturb-seq based perturbation in 5 datasets with GRNs built on all cells in each dataset.

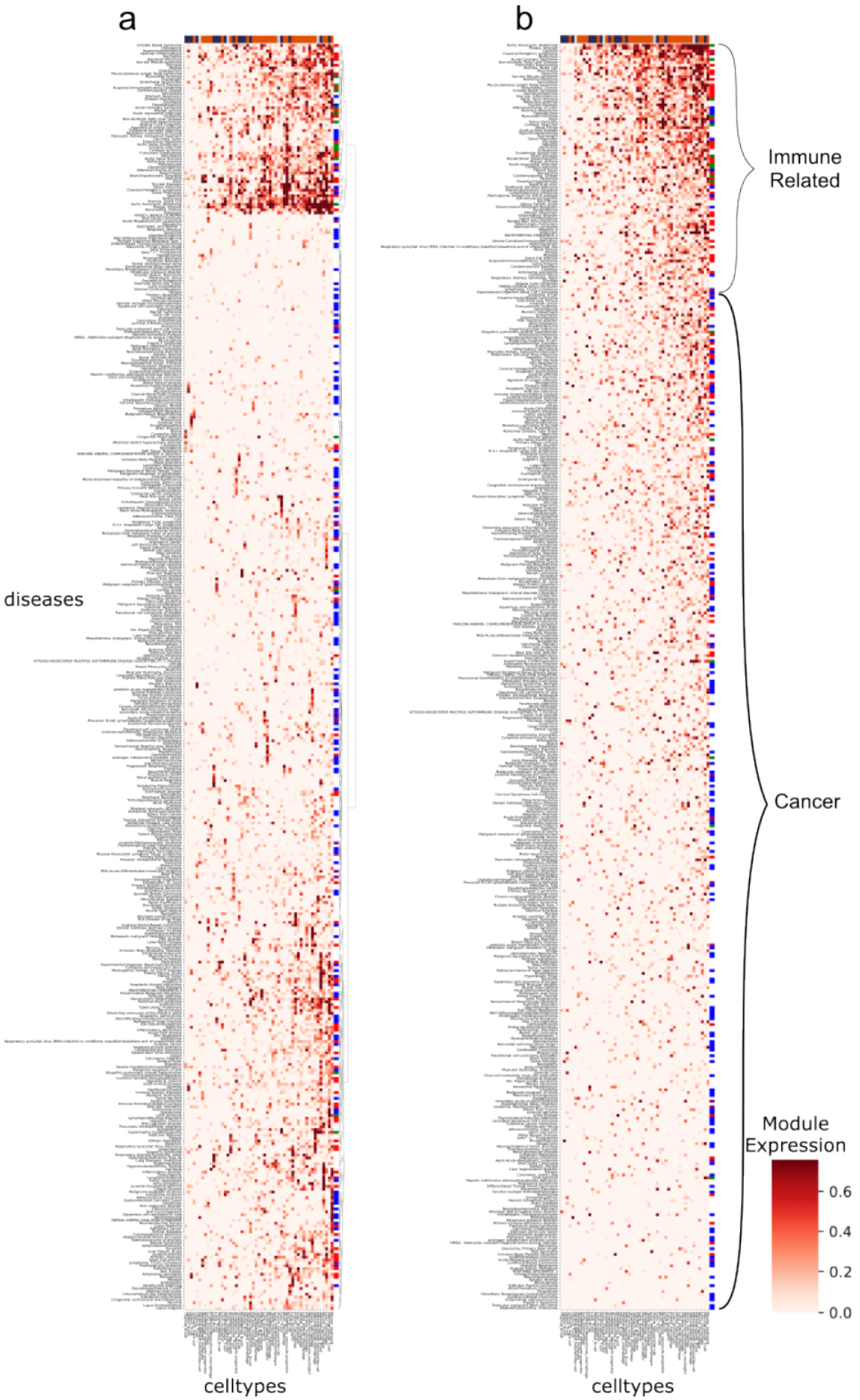
Performance of SCING in predicting downstream gene perturbation as a function of the fraction of zeros in each perturbation, before and after removing cells with zero expression of the gene of interest (a). Area under receiver operator characteristic (AUROC) curve for prediction of downstream perturbations using undirected GRNs (b). AUROC for prediction of downstream perturbation on directed GRNs for SCING (c). True positive rate (TPR) at a false discovery rate (FDR) of 0.05 for the prediction of downstream perturbations on undirected GRNs (d). TPR at FDR of 0.05 for the prediction of downstream perturbations on directed GRNs for SCING (e).

(*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$)

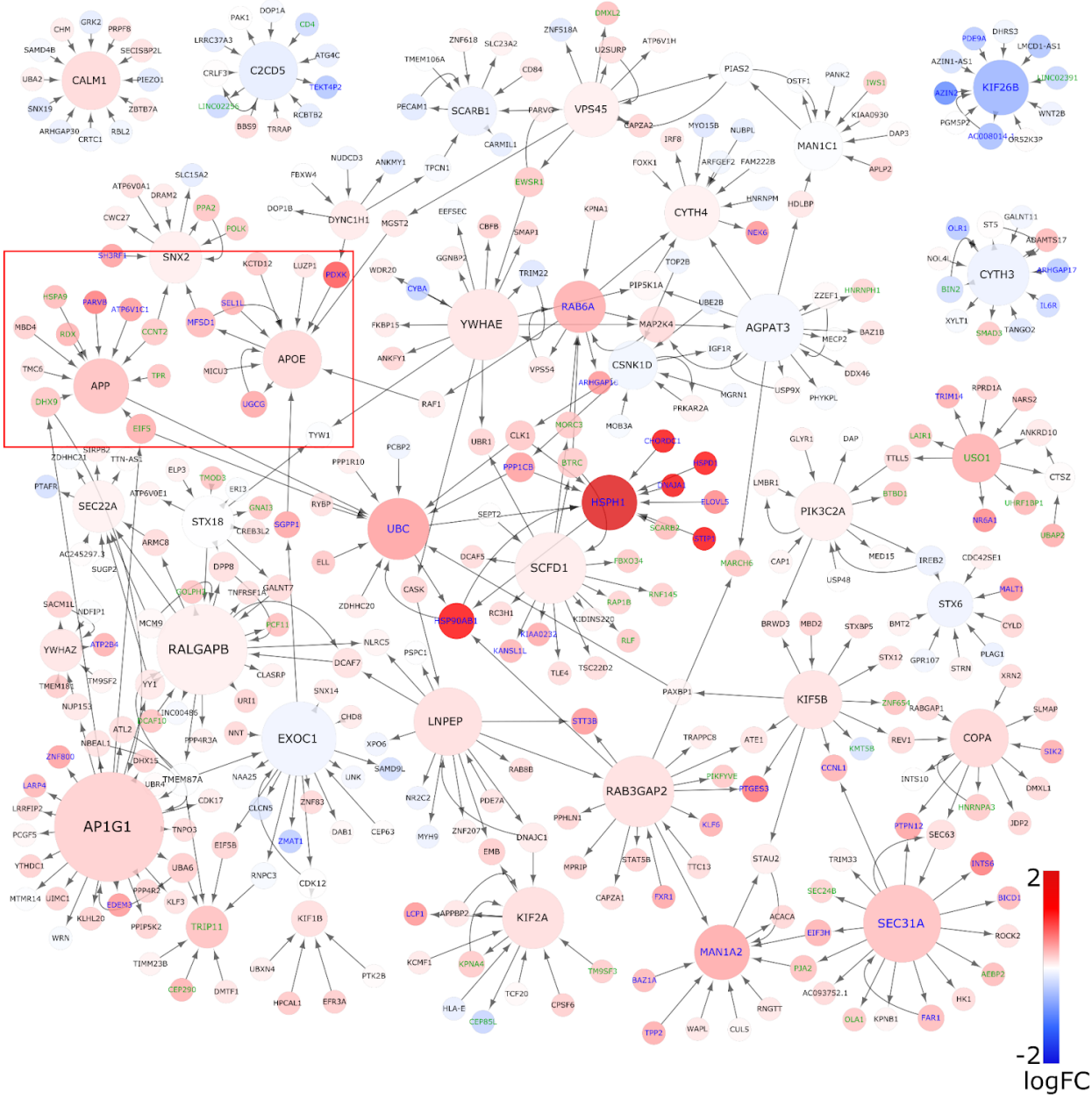

a
highlighted in Figure 4b
celltypes
0.25
0.20
0.15
0.10
0.05
0.00
diseases
b
number of cell types modeled by
diseases
c
number of diseases modeled
celltypes

SF 3.3. SCING GRNs' modeling capabilities of DisGeNET diseases on entire MCA.

Clustermap of all cell types in mouse cell atlas and SCING GRNs modeling of disease subnetworks across all diseases in DisGeNET (a). Number of cell types each disease subnetwork can be accurately (>0.1) modeled by (b). Number of disease subnetworks accurately modeled (> 0.1) by each cell type (c).

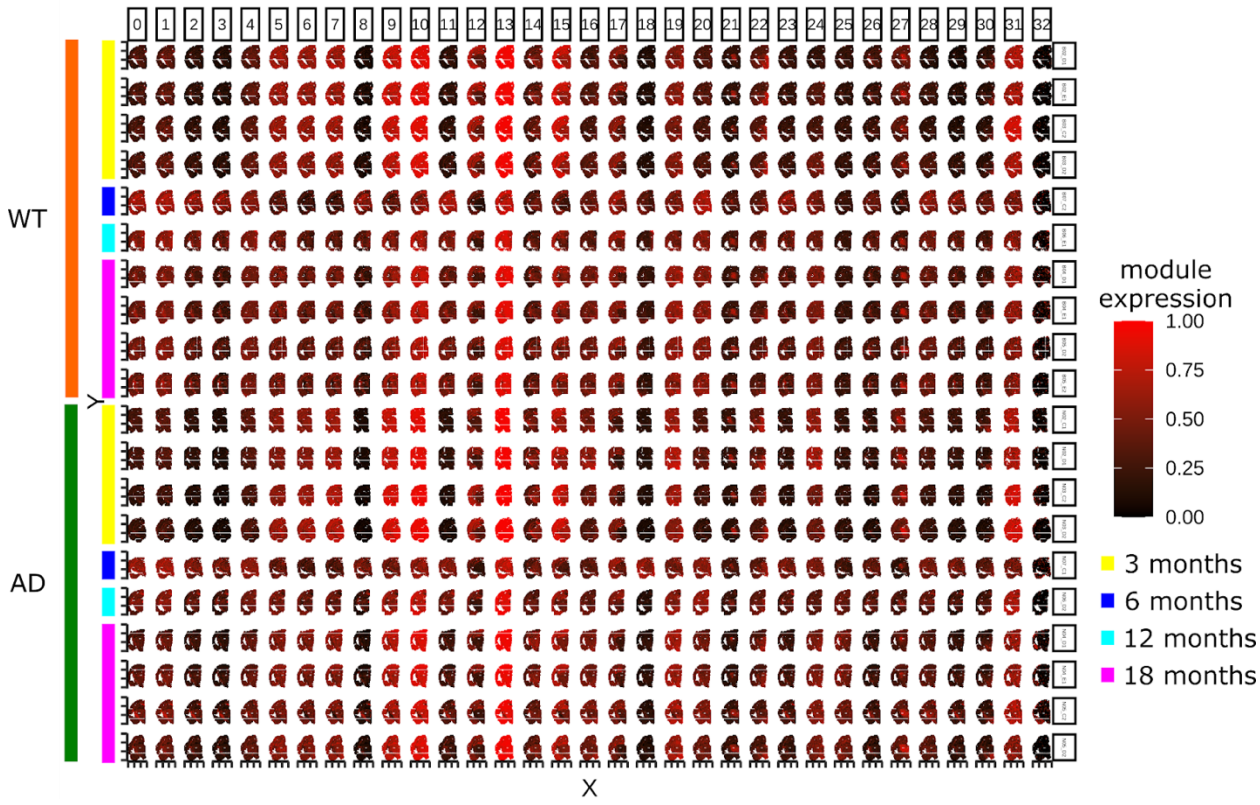

a
b
diseases
celltypes
celltypes
Immune
Related
Cancer
Module
Expression
0.6
0.4
0.2
0.0

SF 3.4. Clustermap of immune cell types and all diseases. Cell types are ordered by the number of diseases they accurately model, and diseases are either clustered with hierarchical clustering (a), or ordered by the number of cell types they are modeled by (b). Diseases are colored by disease category (immune related: red; cardiothoracic: green; cancer: blue; immune cell cancer: purple), and cell types are colored by innate (orange), and adaptive immune system (dark blue).



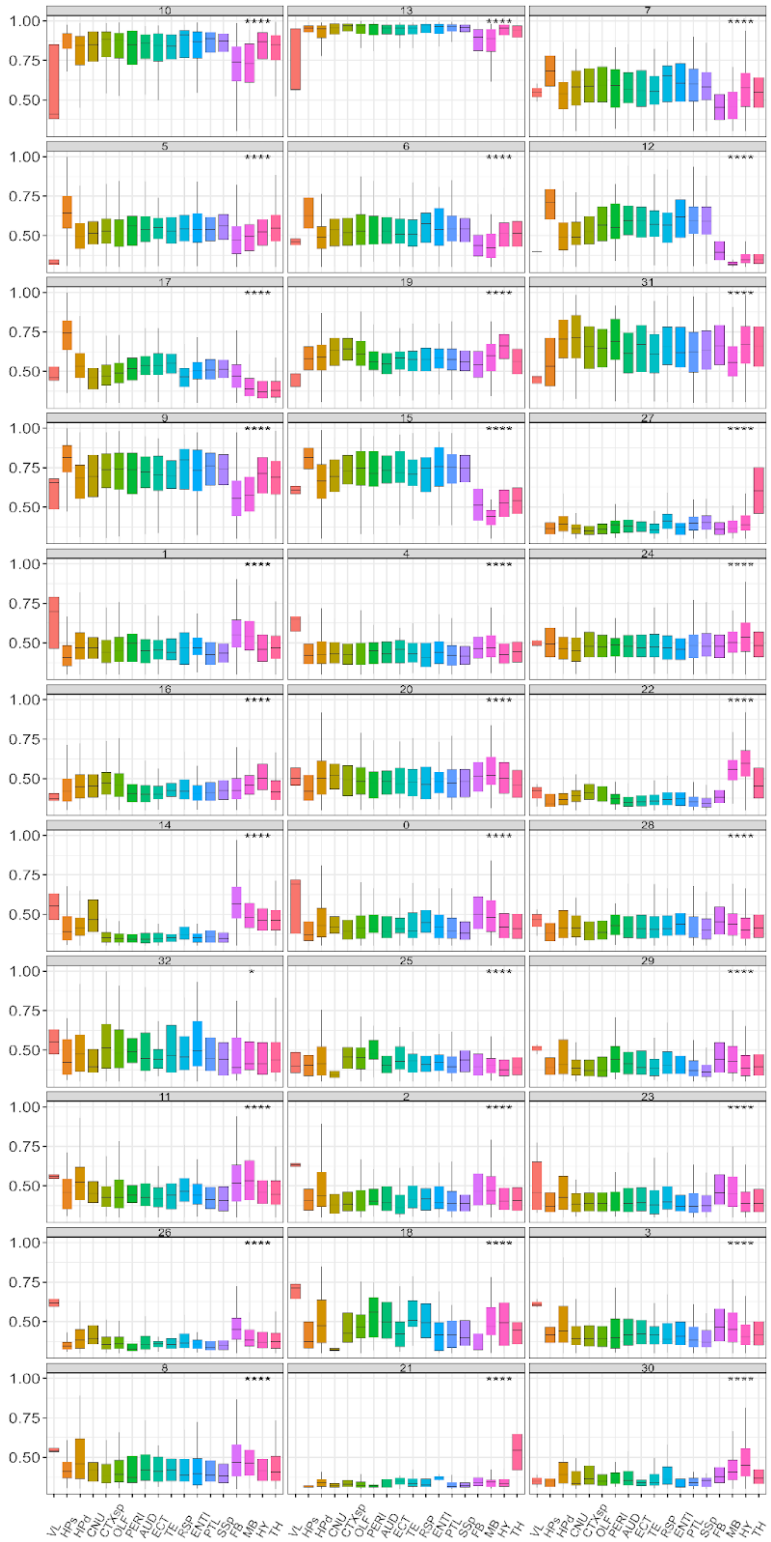
SF 3.5. Subnetwork of the vesicle-mediated transport pathway from module 2 in the snRNAseq microglia GRN.

Nodes are colored by their log-fold change (logFC) between AD and Control patients. APOE and APP subnetwork is highlighted (red box).

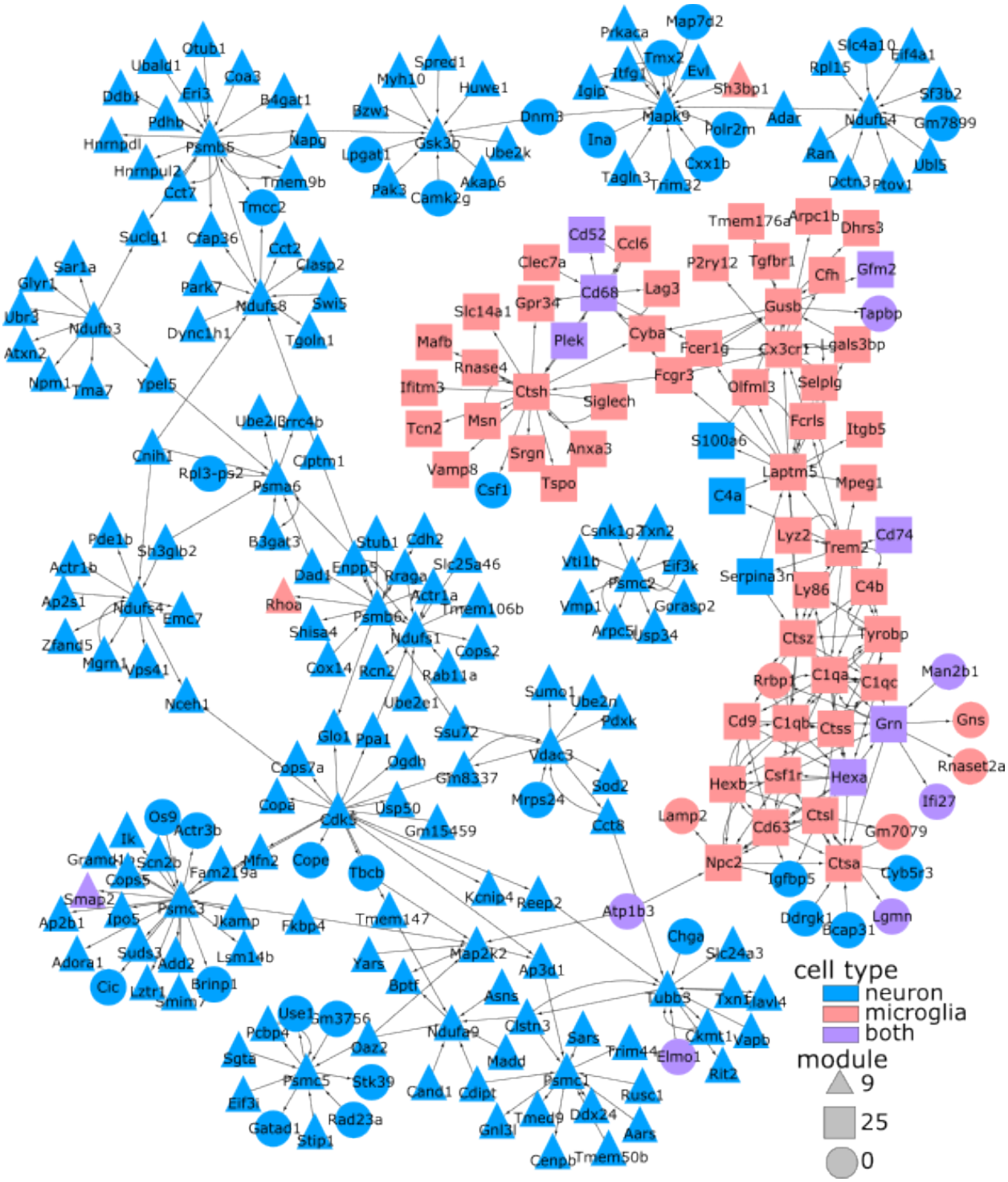


SF 3.6. Expression of each module visualized in each brain.

Mice are ordered by age (3, 6, 12, 18 months) and genotype (WT vs AD) mice.

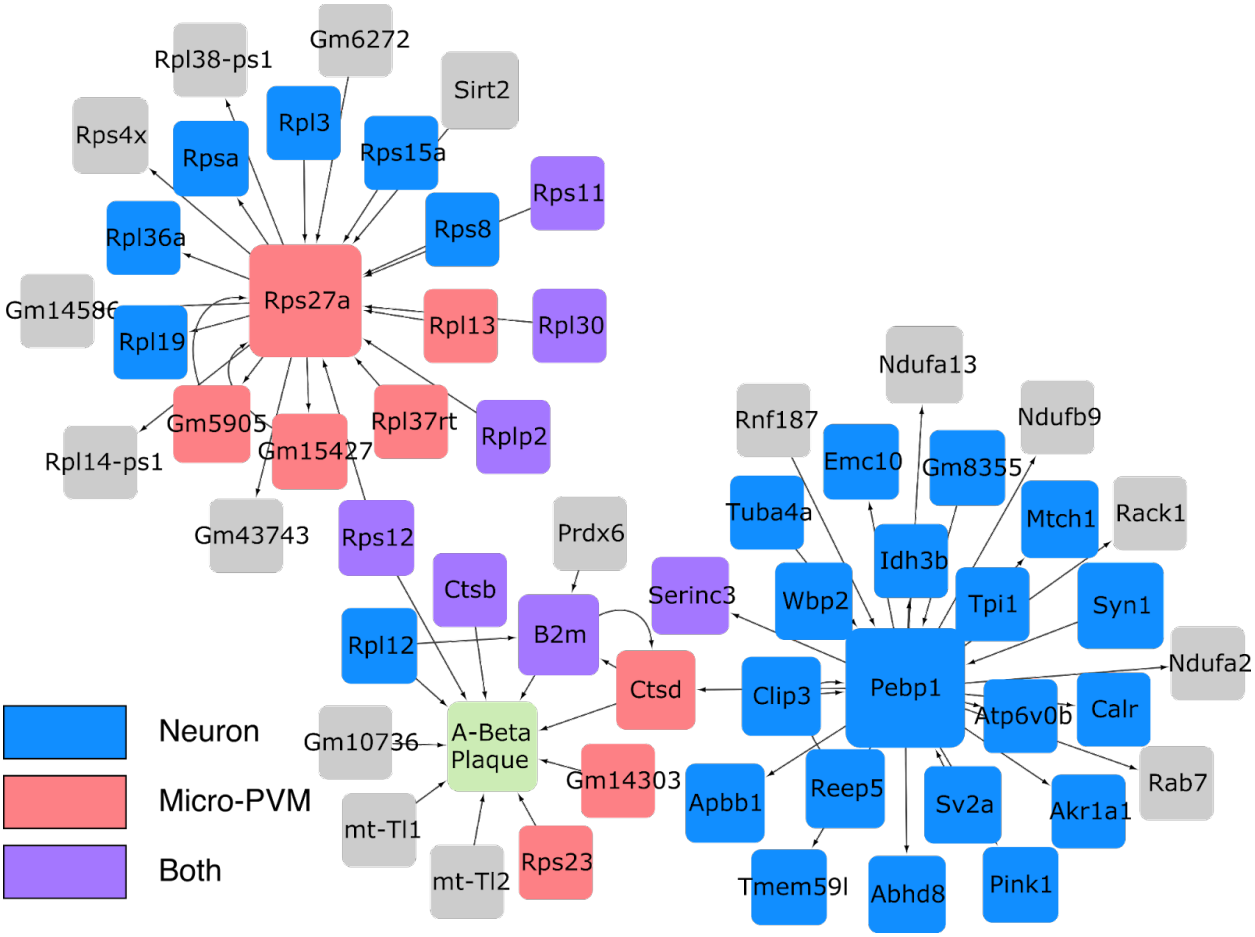


SF 3.7. Distribution of module expression in each region of the brain. Significantly variable expression is determined by ANOVA ($p<0.05$).

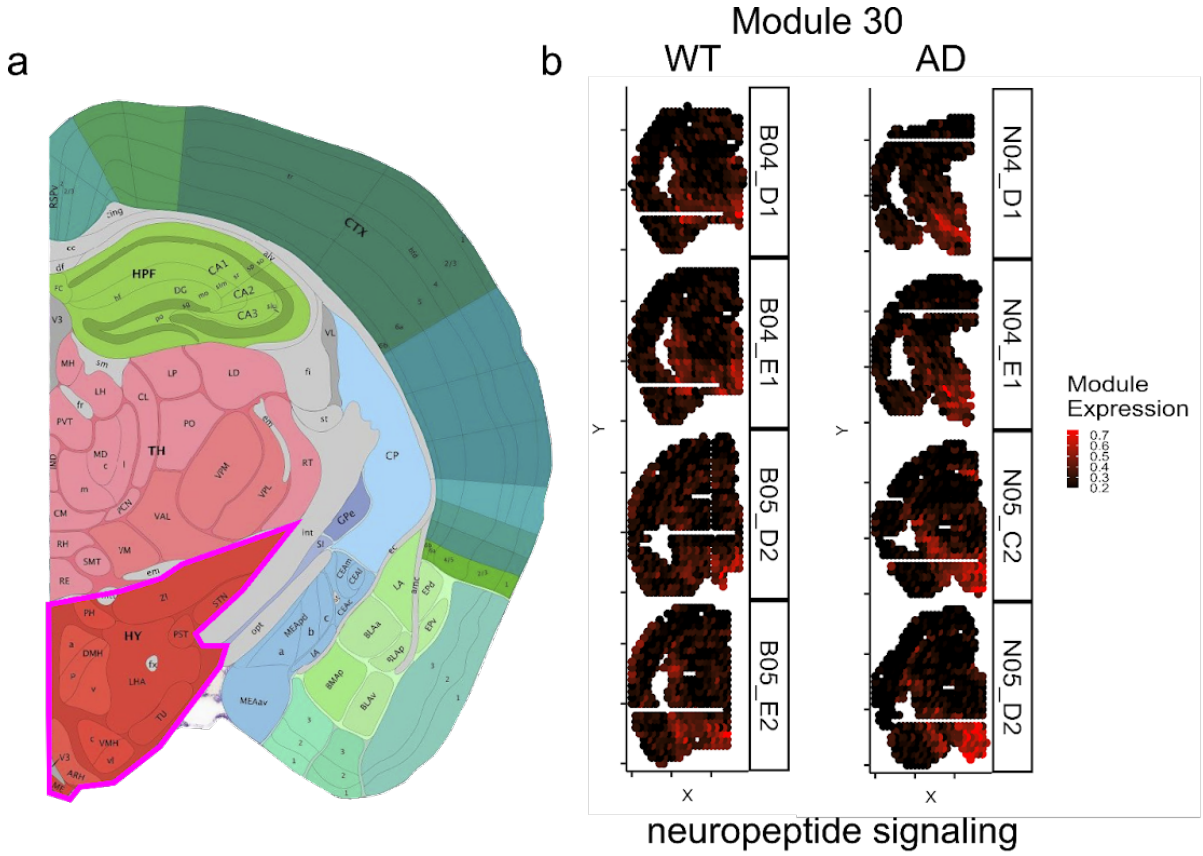


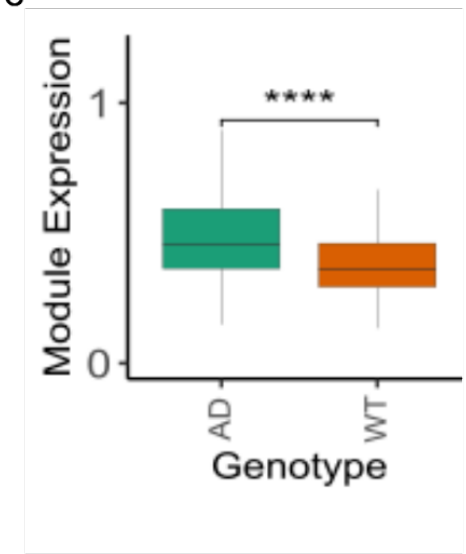
SF 3.8. Subnetwork of module 9 (triangles) and 25 (squares), from the visium AD dataset.

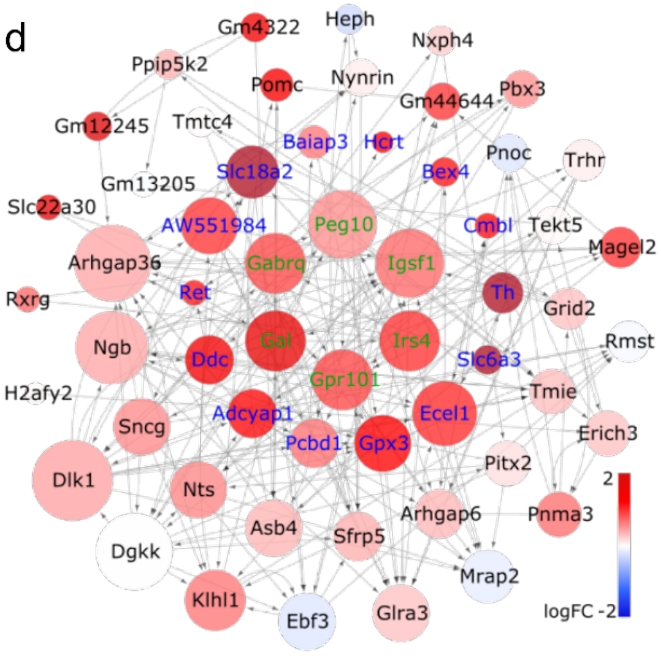
Marker genes were determined from the Allen Brain Atlas whole brain smart-seq dataset. Nodes are colored by marker gene status (microglia:red; neuronal:blue; both:purple). Cross cell type communication edges are between microglia and neuronal nodes. Module 9 is composed mostly of neuronal genes, while module 25 is composed mostly of mostly microglia genes.



SF 3.9. Subnetwork containing amyloid beta plaque in module 10 from the visium data. Nodes are colored by marker gene status (microglia:red; neuronal:blue; both:purple). Hypothesized cross cell type communication edges are between microglia and neuronal nodes. This contains the amyloid-beta plaque stain (colored green).


a
b
Module 30
WT
AD
B04_D1
B04_E1
B05_D2
B05_E2
N04_D1
N04_E1
N05_C2
N05_D2
Module Expression
0.7
0.6
0.5
0.4
0.3
0.2
X
Y
CTX
HPF
CA1
CA2
CA3
DG
TH
HY
CP
GPe
STN
PST
LHA
DMH
VMH
ARH
MEAav
BLAv
BMAp
LA
BLAa
BLAp
CEAm
CEAl
CEAc
MEApd
opt
int
st
fi
VL
VPM
VPL
RT
PO
LP
LD
LH
CL
PVT
MD
VAL
PCN
CM
RH
SMT
IM
RE
PH
ZI
fx
TU
V3
MH
sm
cc
df
fc
alv
bsc
ccg
BSTv
neuropeptide signaling


c
Module Expression
1
0
****
AD
WT
Genotype


d
Gm4322
Heph
Nxph4
Ppip5k2
Pomc
Nynrin
Gm44644
Pbx3
Gm12245
Tmtc4
Baiap3
Hcrt
Pnoc
Trhr
Slc18a2
Bex4
Gm13205
Slc22a30
AW551984
Peg10
Cmbl
Tekt5
Magel2
Arhgap36
Gabrq
Igsf1
Rxrg
Ret
Th
Grid2
Ngb
Ddc
Gal
Irs4
Slc6a3
Rmst
H2afy2
Gpr101
Tmie
Sncg
Adcyap1
Ecel1
Erich3
Pcbd1
Gpx3
Pitx2
Dlk1
Nts
Arhgap6
Pnma3
Asb4
Sfrp5
Dgkk
Mrap2
Klhl1
Ebf3
Glra3
2
logFC -2

SF 3.10. Localization and association of neuropeptide signaling in the hypothalamus.

Diagram from the Allen Brain Atlas shows the location of the hypothalamus (outlined in pink) (a). Module 30 has localizes to the hypothalamus (b) and has higher expression in AD (c). The module 30 subnetwork (d) is enriched for neuropeptide signaling. Color of each node represents the logFC in AD vs WT mice. The color of the text represents significance (green: adjusted p-value < 0.05, blue: p-value < 0.01).

3.6 Tables

Table 3.1. Average run time (seconds) of GRN building methods on variable number of genes for 1,000 cells across 10 iterations.

		1000 genes	2000 genes	4000 genes
astrocytes	SCING	10.8	14.2	21.4
	ppcor	4.3	4.1	4.1
	PIDC	74.3	77.2	85.1
	GRNBOOST2	35.0	98.2	272.9
microglia	SCING	10.5	13.8	20.8
	ppcor	4.4	3.9	4.2
	PIDC	73.0	75.8	83.0
	GRNBOOST2	43.0	340.5	340.5
oligodendrocytes	SCING	10.3	13.5	20.7
	ppcor	4.9	4.6	4.1
	PIDC	74.5	77.3	86.0
	GRNBOOST2	13.8	29.0	71.3

Table 3.2. Average run time (seconds) of GRN building methods on variable number of cells for 1,000 genes across 10 iterations.

		250 cells	500 cells	1,000 cells
astrocytes	SCING	9.9	11.0	10.6
	ppcor	0.6	2.1	4.3
	PIDC	0.2	19.6	74.1
	GRNBOOST2	9.8	14.8	33.3
microglia	SCING	10.1	11.0	10.5
	ppcor	0.7	2.3	4.6
	PIDC	9.2	18.7	73.1
	GRNBOOST2	10.2	15.3	45.8
oligodendrocytes	SCING	9.5	9.9	10.5
	ppcor	0.52	2.2	4.7
	PIDC	9.2	19.9	75.0
	GRNBOOST2	9.0	10.3	13.9

Table 3.3. Hyperparameters for perturb-seq clustering

Dataset	Resolution
DC 0hr	1.2
DC 3hr	0.7
K562 cell cycle	1.0
K562 promoters	1.0
Papalexi	1.2

Chapter 4. Therapeutic IDOL reduction ameliorates amyloidosis and improves cognitive function in APP/PS1 mice

4.1 Introduction

The *APOE* genotype is the strongest genetic risk factor for Alzheimer's disease (AD). ApoE has been shown to independently influence several key factors that drive pathogenesis of AD, including β-amyloidosis, tauopathy, and synaptic dysfunction[126–128]. The impact of ApoE on amyloidosis has been the subject of intensive research, since β-amyloid (Aβ) accumulation and aggregation are key initiators of complex pathological changes in the brain that culminate in neurodegeneration years later. Mounting evidence suggests that ApoE primarily influences AD pathology via its effects on Aβ metabolism. ApoE exerts the greatest impact on amyloidosis during the initial seeding stage; accordingly, lowering ApoE levels prior to the formation of Aβ plaque in APP/PS1 mice reduces Aβ plaque pathology[129]. ApoE has also been reported to promote Aβ aggregation[130] and to impair its clearance from the brain interstitial fluid[131].

In the brain, ApoE functions as a ligand for members of the lipoprotein receptor family, including low-density lipoprotein receptor (LDLR), LDL receptor-related protein 1 (LRP1), very low-density lipoprotein receptor (VLDLR), and ApoE receptor 2 (ApoER2). Among ApoE receptors, LDLR and neuronal LRP1 are the principal regulators of ApoE metabolism, acting to mediate the uptake and degradation of ApoE-containing lipoprotein particles by brain cells[132]. Overexpression of the LDLR in glia cells reduces brain ApoE and Aβ deposition level by enhancing Aβ clearance[133], suggesting that increasing glial LDLR levels may represent a therapeutic strategy to treat AD.

We previously identified E3 ubiquitin ligase IDOL as a negative regulator of LDLR in microglia. Loss of IDOL in microglia increases LDLR protein levels, which in turn facilitates ApoE and Aβ uptake and clearance by microglia. Ablation of IDOL in both male and female APP/PS1 mice—a transgenic mouse model of Aβ amyloidosis—led to decreased soluble and insoluble Aβ, reduced

amyloid plaque burden, and ameliorated neuroinflammation [134]. Whether pharmacological inhibition of IDOL in the adult brain can serve as a safe and effective therapeutic strategy to ameliorate Aβ-related pathology remains to be determined. In this study, we utilized an antisense oligonucleotide (ASO) to therapeutically inhibit IDOL activity in the adult brain of APP/PS1 mouse model of AD amyloidosis. IDOL ASO treatment reduced soluble and insoluble Aβ and amyloid plaque load in the brain and also decreased neuritic dystrophy around plaques. Importantly, IDOL ASO treatment also improved the cognitive performance of APP/PS1 mice in the Morris water maze. Our results provide validation of the potential utility of IDOL as a therapeutic target for AD pathogenesis.

4.2 Results

ASO treatment reduces IDOL expression *in vivo*.

We optimized a central nervous system (CNS)-acting ASO against IDOL to enable the inhibition of IDOL activity in the adult mouse brain. To evaluate the efficacy of the anti-IDOL ASO (here referred to as IDOL ASO) and determine the optimal dose, five groups of C57BL/6J male mice (n = 5 for each group) received intracerebroventricular (i.c.v.) injection of various doses of IDOL ASO or PBS (vehicle control) into the lateral ventricle. After a 2-week incubation, we measured IDOL mRNA level in total brain lysates. IDOL ASO showed high potency with half-maximal inhibitory concentration (IC_{50}) of 12.5 μg/mice and prolonged stability with an estimated half-life ($t_{1/2}$) of 9 weeks (Figure 4.1A). A single bolus dose of 40 μg of IDOL ASO reduced IDOL mRNA level by at least 80% relative to controls by qPCR, and knockdown of IDOL was seen in all the brain regions examined, including the hippocampus, cortex, thalamus, and cerebellum (Figure 4.1B). We chose a 40 μg of IDOL ASO i.c.v. injection every 2 to 3 months as the optimal treatment strategy for the following studies. This approach achieves close to maximum IDOL knockdown efficiency with minimal amount of ASO, thereby minimizing nonspecific effects. Since IDOL is mainly expressed in the microglia and neurons, we examined the cellular uptake of ASO by

immunohistochemical staining of brain sections with a pan-ASO antibody (Figure 4.1C). This analysis showed that the ASO was taken up by both microglia and neurons. Consistent with reduced IDOL expression, the protein level of IDOL substrates—LDLR, VLDLR, and ApoER2—were all increased in the brains of IDOL ASO-treated mice (Figure 4.1D), despite the fact that their mRNA levels remain unchanged (data not shown). We did not observe changes in body weight or gross behavior associated with IDOL ASO treatment in our study. When evaluated in contextual fear conditioning and the Morris water maze, IDOL ASO treatment did not affect the cognitive performance of wild-type mice (Figure 4.1E and F).

IDOL-ASO treatment ameliorates Aβ plaque pathology.

To assess the effects of brain IDOL inhibition on amyloidosis in an AD mouse model, we administered IDOL ASO (40 μg/mouse) or vehicle (phosphate-buffered saline [PBS]) via i.c.v. injection to a cohort of APP/PS1 male mice (n = 8 to 10 per group) at 3 months of age (before the onset of plaque formation), followed by a booster dose at 6 months of age. Mice were sacrificed at 8 to 9 months of age and brains were collected for pathological analysis. We first assessed Aβ deposition by immunostaining with an anti-Aβ antibody (82E1). Aβ plaque area in IDOL ASO-treated APP/PS1 mice was reduced >50% compared to controls (Figure 4.2A). Quantification of fibrillar plaque load using X-34 dye also showed a marked reduction (>50%) in the ASO-treated group (Figure 4.2B). Next, we biochemically assessed Aβ40 and Aβ42 levels in radioimmunoprecipitation assay (RIPA; soluble) and Guan (insoluble) fractions of brain lysates. IDOL ASO treatment reduced Aβ40 and Aβ42 load in both soluble and insoluble fractions in APP/PS1 mice (Figure 4.2C and D). Western blot analysis of RIPA fractions of brain lysates further confirmed that IDOL ASO treatment led to reduction of total amyloid species and ApoE levels (Figure 4.2E).

To examine the effects of IDOL knockdown on plaque size distribution, we analyzed the X34-stained data set by grouping individual plaques based on size. We found reduced plaque density and average size in the IDOL ASO group compared to controls (Figure 4.2F). We also observed a shift in the size distribution of the plaques between groups. The total area covered by larger plaques (>1,000 μm^2) was dramatically reduced, and plaques larger than 2,000 μm^2 were only rarely observed in the ASO group (Figure 4.2G). Together, these results suggested that pharmacological inhibition of brain IDOL activity is sufficient to reduce Aβ levels and plaque burdens in APP/PS1 mice. These consequences of acute IDOL knockdown are consistent with our previous findings of reduced AD-like pathology in IDOL-deficient APP/PS1 mice[134].

Plaque-associated neuritic dystrophy—a well-described consequence of Aβ-induced neurotoxicity—is a key feature of pathological processes downstream Aβ in AD[135]. To investigate the effects of brain IDOL inhibition on plaque-associated neurotoxicity *in vivo*, brain sections of control or IDOL ASO-treated APP/PS1 mice were costained with X-34 and an antibody against lysosomal-associated membrane protein 1 (LAMP1), a marker of dystrophic neurites. IDOL ASO-treated mice showed a reduction in the volume of dystrophic neurites (Figure 4.3A). We also costained brain sections with X-34 and amyloid precursor protein (APP; another marker of dystrophic neurites)[136] and again found a reduction in the volume of plaque-associated dystrophic neurites in IDOL ASO-treated mice (Figure 4.3B). These results suggest that IDOL inhibition in brain ameliorates plaque-associated neuritic dystrophy.

Glial cells, including microglia and astrocytes, play essential roles in neuroinflammation and are believed to be responsible for neuronal toxicity and synaptic dysfunction downstream of Aβ. To determine the extent of Aβ-mediated gliosis in control and IDOL ASO-treated mice, we examined anti-glial fibrillary acidic protein (anti-GFAP)-positive reactive astrocytes in mice by immunostaining. The IDOL ASO group had fewer GFAP-positive astrocytes than controls (Figure

4.4A), consistent with the effects of IDOL ASO on amyloid plaque load. To evaluate the changes in the microglial response to amyloid plaque, we performed histological staining for ionized calcium-binding receptor 1 (Iba1), a marker of activated microglia. Plaque-associated microglial activation was evident around Aβ deposits in both groups. However, when corrected with plaque area, analysis of the Iba1-positive area did not reveal differences between treatment and control groups (Figure 4.4B). Microglia localized in close proximity to Aβ plaques were distributed in a circumferential pattern around the plaque periphery and showed an enlarged cell body size (Figure 4.4C), a feature typical of amoeboid microglia with enhanced phagocytic capacity. Interestingly, microglia in the IDOL ASO-treated group were less hypertrophic and displayed a reduced cell body size compared to those in the vehicle-treated group (Figure 4.4C).

IDOL ASO treatment improves cognitive function in APP/PS1 mice.

To determine whether reduced plaque load in ASO-treated APP/PS1 mice affected cognitive functions, including the ability to learn, we treated another cohort of APP/PS1 male mice (n = 12 per group) with IDOL ASO. Two groups of mice were administered IDOL ASO (40 μg/mice) or vehicle (PBS) at 3 months of age (before the onset of plaque formation), followed by two booster doses at 6 and 9 months of age. Mice were then subjected to behavioral testing at 10 months of age. We first examined hippocampus-dependent spatial memory acquisition with the Morris water maze[137]. APP/PS1 mice were trained to memorize the location of a hidden, submerged platform in a water-filled pool over a 5-day period. The ASO-treated group required less time to reach the platform compared to the control group on days 4 and 5 (Figure 4.5A). To determine the degree of reliance of the mice on spatial versus nonspatial cues to find the platform, we performed probe trials on day 6 in which the platform was removed. The ASO-treated group spent more time and traveled longer distances (Figure 4.5B) in the quadrant of the submerged platform. These results suggested that IDOL-ASO treatment improves spatial memory acquisition in APP/PS1 mice. This improvement in performance could not be attributed to vision or locomotor differences, since both

groups performed similarly in the visible platform test (Figure 4.5C, left) and exhibited comparable swimming speed (Figure 4.5C, right).

We next tested cued and contextual-fear conditioning, a paradigm that assesses hippocampus-dependent (context) and amygdala-dependent (context and cued) fear learning[138] in vehicle- and IDOL ASO-treated APP/PS1 mice. On day 1, mice were trained to associate environment and sound with a foot shock. On day 2, mice were placed back into the same environment and assessed for freezing for 5 min to evaluate the contextual association with the foot shock. On day 3, mice were placed in a novel environment for 6 min and assessed for freezing during the last 3 min, when the same tone was played to measure the association of tone with foot shock. Both groups of mice displayed similar freezing behavior during training (Figure 4.5D). On days 2 and 3, both groups showed a high level of freezing when returned to the same context or presented with the same tone. ASO-treated mice were not different from control mice in their ability to memorize the association between shock and the context (Figure 4.5E, left) or auditory cue (Figure 4.5E, right). The failure to detect a difference in this test may be due to the fact that performance in both groups was comparable to wild-type mice (Figure 4.1E).

Single cell RNA sequencing reveals enhanced phagocytic function of microglia in IDOL knockdown mice.

Local environmental cues are critical for shaping and maintaining brain microglial phenotypes in both mice and humans[139,140]. Furthermore, recent studies have highlighted the heterogeneity of microglial populations during brain development[141], in different brain regions[142], and in various disease states[143]. We previously showed that loss of IDOL facilitates Aβ uptake and clearance by microglia *in vitro*[134]. However, the role of IDOL in the function of microglia within physiological environments has not been investigated. To better capture heterogeneity of microglia and other brain cells, we used single-cell RNA sequencing (scRNA-seq) to profile the transcriptome of individual cells in mouse hippocampus. Hippocampal cells from PBS- and ASO-treated APP/PS1

mice (n = 3/group) were isolated 1 week after behavior testing, and subjected to scRNA-seq with the Drop-seq platform[144]. We obtained data for a total of 14,682 cells (6,888 from the PBS group and 7,794 from the ASO-treated group) after quality control filtering.

We clustered the hippocampal cells based on their similarities in gene expression pattern and projected the cells onto two dimensions using uniform manifold approximation and projection (UMAP)[145], and determined cell-type identities of the main clusters using a mouse hippocampal reference panel[146] (Figure 4.6A). Each cell cluster contained cells from all six samples without batch effects, and there were no clear clustering differences between the PBS and ASO groups (Figure 4.6B). To further confirm cell cluster specificity, we highlighted known marker genes for each cell type and found these key genes to be uniquely expressed in their respective clusters, such as *C1qc* for microglia and *Snap25* for neurons (Figure 4.6C to F). By assessing IDOL expression in each cell type, we confirmed that IDOL was highly expressed in microglia and that its expression level was effectively reduced by ASO treatment (data not shown). To determine specific genes and pathways that were altered by IDOL knockdown, we identified differentially expressed genes (DEGs) between PBS- and ASO-treated groups within each cell cluster (Table 4.1) at $P < 0.05$ and at a false discovery rate (FDR) of <0.05. Microglia and oligodendrocytes had the largest numbers of DEGs between the PBS and ASO groups. Annotation of the DEGs with curated biological pathways revealed the lysosomal pathway as the top enriched biological pathway in microglia. IDOL-ASO treatment upregulated lysosomal/phagocytic genes, such as those for cathepsins (*Ctsb* and *Ctsd*) and CD63, strongly suggesting that knockdown of IDOL enhanced phagocytic function of microglia *in vivo* (Table 4.1).

To further examine the impact of IDOL ASO treatment on microglia heterogeneity in APP/PS1 mice, we reclustered and projected the microglia cells in two dimensions using t-distributed stochastic neighbor embedding (t-SNE)[147] and identified four distinct subgroups (Figure 4.7A) that did not show apparent clustering distinction between PBS and ASO mice (Figure 4.7B). The gene

expression patterns of subgroups 0 and 1 resembled the signatures of homeostatic microglia, with high expression of marker genes such as *P2ry12* and *Cx3cr1* (Figure 4.7C, F, and G), whereas subgroups 2 and 3 resembled disease-associated microglia (DAM), with high expression of marker genes such as *Cst7*, *Trem2*, and *ApoE*[96] (Figure 4.7D, F, and G). The separation of subgroups 0 and 3 from subgroups 1 and 2 was mainly due to differential expression of inflammatory genes, such as the cytokine genes *Il1a*, *Il1b*, and *Tnf* (Figure 4.7E to G). Interestingly, IDOL-ASO treatment shifted the microglia population to a higher proportion of DAM-like subgroups 2 and 3 (22.45% in PBS group versus 35.36% in ASO group; $P = 0.04154$ [Fisher exact test]) (Figure 4.7A). Consistently, microglia in the ASO group displayed higher expression levels of DAM markers, including *B2m*, *Ctsb*, *Ctsd*, *Lpl*, and *Cd63* (Table 4.1). It has been shown that DAM cells are abundant near amyloid plaques and actively participate in phagocytic removal of amyloid plaques[96]. Thus, this shift in microglial populations in IDOL-ASO-treated mice could plausibly be responsible for the reduced amyloid plaque load.

We also reclustered and projected the general neuronal population onto two dimensions using t-SNE to identify neuronal subtypes (Figure 4.8A and B). The CA neurons and the subiculum neurons did not clearly separate, since their gene profiles and physical locations are similar to one another. However, neurons of the dentate gyrus, interneurons, and Cajal-Retzius neurons were clearly separated from the others and demonstrated unique gene expression profiles (Figure 4.8C and D). Pathway analysis of the DEGs ($P < 0.05$) revealed that ASO treatment changed the expression of genes enriched in the Wnt signaling pathway in Cajal-Retzius neurons, the oxidative phosphorylation pathway in subiculum neurons (mostly pyramidal cells), and the homeostatic responses in interneurons (Table 4.1).

4.3 Discussion

ApoE4 has a powerful impact on the development of late-onset AD. Aβ turnover and clearance in the mouse brain *in vivo* is strongly dependent upon ApoE isoform[131], suggesting

that *APOE* alleles contribute to human AD risk by differentially regulating clearance of Aβ from the brain and that enhancing Aβ clearance may be a therapeutic strategy for AD prevention and treatment. It is reasonable to speculate that the biological effects of ApoE are at least partially mediated by the receptors to which it binds. LDLR is an ApoE receptor known to regulate brain Aβ clearance. LDLR deficiency is associated with increased Aβ deposition in Tg2576 APP-transgenic mice[148,149], while overexpression of the LDLR in the brain enhances Aβ clearance and decreases Aβ deposition[133]. These findings suggest that increasing LDLR levels in brain cells could be beneficial in AD, but it is pivotal to validate tractable approaches to regulate brain LDLR levels in order to translate basic science knowledge to clinical application. We previously identified E3 ubiquitin ligase IDOL as a negative regulator of LDLR proteins levels in microglia and neurons[134]. In this study, we examined the effects of IDOL reduction in adult APP/PS1 mice using an anti-IDOL ASO. IDOL inhibition led to a reduction in Aβ pathology, a decrease in neuritic dystrophy around the plaques, and an improvement in spatial memory. This study suggests that pharmacologically inhibiting IDOL could be a feasible approach to ameliorate Aβ-related pathology.

Microglia are professional phagocytes capable of clearing targeted pathogens, cellular debris, and pathogenic Aβ in the brain. Previous *in vitro* studies suggested that loss of IDOL in microglia enhances the uptake and clearance of Aβ. In this study, our scRNA-seq data showed that ASO treatment increased the expression of lysosomal/phagocytic cell-related genes in microglia *in vivo*, corresponding to the enhanced clearance of soluble Aβ and reduced plaque deposition in the brains of APP/PS1 mice. High-level expression of genes in phagocytic and lipid metabolism pathways is the molecular signature of disease-associated microglia (DAM), a recently identified subset of CNS-resident microglia in the 5XFAD mouse model[96]. The existence of DAM in mice and humans has recently been confirmed in tauopathy AD models[150,151], aging[151,152], and other neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS)[153] and multiple

sclerosis[154]. DAM has been shown to associate with amyloid plaques and actively participate in the dismantling and digestion of the amyloid plaques[96]. In our exploration of the impact of IDOL on microglia heterogeneity, we found that IDOL-ASO treatment shifted the microglial populations to a higher proportion of DAM-like subgroups (Figure 4.6), despite the fact that ASO-treated APP/PS1 mice had lower levels of Aβ and plaque deposition (Figure 4.1 and 4.2). These findings suggest that the IDOL-LDLR pathway regulates the phagocytotic function in microglia in response to Aβ challenge. The underlying mechanisms through which increased LDLR levels enhanced phagocytic function in microglia remain to be clarified.

Since LDLR is a major metabolic receptor for ApoE, the IDOL-LDLR pathway is expected to regulate ApoE uptake and recycling in microglia. It has been reported that ApoE facilitates the microglial response to amyloid plaque pathology[155]. We reason that LDLR may enhance phagocytic activity of microglia through the ApoE pathway. Given that ApoE also serves as ligand for TREM2[156], we further speculate that ApoE may increase phagocytosis of Aβ by enhancing TREM2 signaling. Future experiments will address whether the effect of the IDOL-LDLR pathway on microglial phagocytic function is dependent on ApoE and different ApoE alleles.

When reclustering microglia, we detected two previously identified subtypes: the homeostatic microglia (subgroups 0 and 1) and DAM (subgroups 2 and 3). Within each of these subtypes, we also identified subgroups with different expressions of inflammatory genes, such as *Il1b* and *Tnfa*. Compared to subgroups 1 and 2, microglial subgroups 0 and 3 were enriched for pathways involved in tumor necrosis factor alpha (TNF-α) signaling, interferon gamma (IFN-γ) response, and the inflammatory response (FDR <0.05). The exact role of these microglia subgroups in amyloid pathology remains unclear and will be a topic for future investigation. Interestingly, inflammatory cytokine genes (*Il1a*, *Il1b*, and *Tnfa*) were positively associated with the DAM-associated gene *Csf1* in our study (Figure 4.6B). CSF-1 is considered a key regulator of

inflammatory responses in the brain. Several lines of evidence suggest that microglial activation is mediated by a CSF-1/CSF-1R autocrine loop, which results in reactive microgliosis and the release of inflammatory cytokines[157,158]. Our results are consistent with these previous findings. It has been shown that CSF-1 increases microglial phagocytic activity and stimulates clearance of Aβ[159]; however, CSF-1 also mediates microglial-induced neurotoxicity by promoting the release of inflammatory cytokines. Our result showed both homeostatic microglia and DAM can be further divided into subgroups with different *Csf1* expression levels (Figure 4.6B). It is plausible that these four subgroups may contribute differently to the pathogenesis of AD. Discovering ways of precisely modulating microglial inflammation and phagocytosis in AD is an exciting area for future investigation.

IDOL is also expressed in neurons, where it posttranslationally regulates the level of neuronal lipoprotein receptors, including ApoER2/LRP8 and VLDLR. IDOL-dependent regulation of synaptic ApoER2 has been implicated in the modulation of dendritic filopodium initiation and synapse maturation during the early postnatal stage[160]. ApoER2 has also been shown to functionally couple to NMDA receptors[160] and to protect against the loss of cortical neurons during normal aging[161]. IDOL ASO treatment increased brain ApoER2 levels in our study (Figure 4.1C). Our scRNAseq results yielded only small number of DEGs in neurons between PBS and ASO groups when comparing to microglia, suggesting that knockdown of IDOL has more a prominent impact on microglial functions. However, we cannot rule out the possibility that enhanced neuronal ApoER2 levels also contribute to the neuroprotective effects of IDOL knockdown, such as the decreased neuritic dystrophy around the plaques. We are in the process of evaluating the impact of neuron- and microglia-specific IDOL deletion on neuropathological progression and cognitive function in AD mouse models. The results from these studies will offer a better understanding of the mechanisms through which IDOL impacts AD-related pathology.

4.4 Materials and Methods

Animals.

Male C57BL/6J (000664) and B6.Cg-Tg(APPswe,PSEN1dE9)85Dbo/Mmjax (005864) mice were purchased from The Jackson Laboratory. All mouse experiments were approved and performed under the guidelines of the Animal Care and Research Advisory Committees at the University of California, Los Angeles (UCLA).

Antibodies and reagents.

Primary antibodies used in this study are: anti-human amyloid β (N) (82E1) from IBL Co., Ltd.; anti-amyloid precursor protein antibody [Y188] (ab32136), anti-apolipoprotein E antibody (ab183596), anti-ApoER2 antibody (ab108208), and anti-LAMP1 antibody (ab25245) from Abcam; anti-LDL receptor polyclonal antibody from Cayman; anti-Iba1 antibody from Wako; and anti-glial fibrillary acidic protein (anti-GFAP) from Dako. Anti-VLDLR antibody was a gift from Joachim Herz, University of Texas Southwestern Medical Center. All secondary antibodies were purchased from Thermo Fisher or Jackson Immunoresearch. Antisense oligonucleotides (ASOs) targeting IDOL (5′-TTCCTTTTTTCCACACGCCA-3′) were provided by Ionis Pharmaceuticals, Carlsbad CA. Complete ASO chemistry information was as follows. IDOL ASO (Tes Tes mCes mCeo Tes Tds Tds Tds Tds Tds mCds mCds Ads mCds Ads mCeo Ges mCes mCes Ae), where capital letters indicate base abbreviation, m is 5-methylcytosine, e is 2′-*O*-methoxyethylribose (MOE), k is (*S*)-2′,4′onstrained 2′-*O*-ethyl (cEt), d is deoxyribose, s is phosphorothioate, and o is phosphodiester. When administrating IDOL ASO to a cohort of wild-type mice (two groups, $n = 5$ for each group), IDOL mRNA levels were downregulated 92 and 62%, respectively, 1 week or 2 months after ASO administration. The half-life of ASO was estimated around 9 weeks.

mRNA analysis.

mRNA level was determined by real-time RT-qPCR (Diagenode, Denville, NJ) from RNA isolated from frozen tissues with TRIzol (Life Technologies, Carlsbad, CA) according the manufacturer's

instructions. Statistical analysis was conducted using a two-tailed unpaired *t* test or one-way ANOVA.

Protein analysis.

Proteins were sequentially extracted from brain tissues with RIPA, and 5 M guanidine buffer in the presence of protease inhibitors as described previously[161]. For Western blots, equal amounts of proteins (10 to 40 mg) were separated on NuPAGE bis-tris gels (Invitrogen), and membranes were probed with primary and secondary antibodies. Signals were visualized by chemiluminescence (ECL Plus; GE Healthcare). Blots were quantified by densitometry with ImageJ software (National Institutes of Health [NIH]). To quantify Aβ40 (KHB3481) and Aβ42 (KHB3441), human ELISA kits from Life Technologies were used. Statistical analysis was conducted using a two-tailed unpaired *t* test or one-way analysis of variance (ANOVA).

Histological analysis.

Brains were sectioned on a cryostat at a 40-mm thickness. For immunofluorescence staining, free-floating sections were blocked with PBS containing 10% normal goat serum (NGS) at room temperature for 30 min, incubated with primary antibody in blocking solution at 4°Covernight, and then incubated with secondary antibody at room temperature for 1 h. Sections were mounted on slides with ProLong Diamond (Life Technologies). For immunohistochemistry, free-floating sections were treated with 0.3% H_2O_2 and blocked with 3% NGS. Sections were then incubated with biotinylated 82E1 (1:500) or anti-Iba1 (1:1,000) in blocking solution at 4°C for 24 h. Sections were incubated with biotinylated anti-rat or anti-rabbit IgG antibody (Vector Laboratories) in blocking solution at room temperature for 1 h. Antibody binding was detected with Vectastain ABC Elite (Vector Laboratories) and DAB peroxidase (horseradish peroxidase) substrate kits (Vector Laboratories) supplemented with nickel solution. Sections were dehydrated and mounted on slides with Permount (Fisher Scientific). Images were captured on a confocal microscopy (Zeiss) and quantified using ImageJ software. Statistical analysis was conducted using a two-tailed unpaired *t* test or one-way ANOVA.

Behavior tests.

(i) Morris water maze test. The Morris water maze is a stainless-steel circular tank with a 200-cm diameter. The tank is filled water dyed with white, using liquid tempera paint to make the water opaque. A platform is placed in one of the quadrants of the pool and submerged 1 cm below the surface. During the trial, mice were placed in the desired start position facing the tank wall and released into the water at water level. If a mouse failed to find the platform within 60 s, it was then placed on the platform for 15 s before being removed. Mice were then placed at a new start location, and the trial repeated four times/day. On day 6, the probe test (wherein the mouse was allowed to free swim for 60 s with the submerged platform removed) was conducted. The percent time spent in the goal quadrant was calculated and analyzed for each mouse. The experimenter was blind to the genotype of the animals. Behavioral data from the training period were analyzed using repeated-measures ANOVA. Data from the probe test were analyzed using one-way ANOVA. Data from the visual cue test were analyzed using a two-tailed unpaired *t* test or repeated-measures ANOVA. All behavioral data were analyzed in Prism 7 (GraphPad Software, Inc.).

(ii) Fear conditioning. To assess fear conditioning, mice were placed in a shock chamber (Med Associates, Inc.) on day 1 for a 7-min training period; during the last 5 min, the mice were exposed to three pairings (1-min intervals) of a 20-s tone immediately followed by a 2-s, 0.38-mA foot shock. On day 2, the mice were placed in the training context for 5 min, and the level of freezing was recorded. On day 3, the mice were placed in a different context for 6 min, and during the last 3 min the tone was played. Mouse freezing was recorded with the FreezeFrame program and analyzed using the FreezeView program.

Normal nociception of all animal strains was ascertained by registering. One-way ANOVA was used to analyze the percent freezing scores of the contextual and cue-dependent freezing. All behavioral data were analyzed in Prism 7.

(iii) Single cell preparation. Single-cell suspensions from bulk tissue were generated as previously described[162] at a final concentration of 100 cells/μl in 0.01% bovine serum albumin (BSA)-PBS by digesting freshly dissected hippocampus tissue with papain (Worthington, Lakewood, NJ). Briefly, the hippocampus was rapidly dissected and transferred into 4 ml of ice-cold Hibernate A (HA; BrainBits LLC, Springfield, IL)-B27 (Invitrogen, Carlsbad, CA)-GlutaMAX (Fisher Scientific, Hampton, NH) (HABG) and then incubated in a water bath at 30°C for 8 min [163]. The supernatant was discarded, and the remaining tissue was incubated with papain (12 mg in 6 ml of HA-Ca) at 30°C for 30 min. After incubation, the papain solution was removed from the tissue and washed with HABG three times. Using a siliconized 9-in Pasteur pipette with a fire-polished tip, the suspension was triturated approximately ten times in 45 s. Next, the cell suspension was carefully applied to the top of the prepared Opti-Prep density gradient (Sigma-Aldrich, St. Louis, MO). The gradient was then centrifuged at 800 × *g* for 15 min at 22°C. We aspirated the top 6 ml containing cellular debris. To dilute the gradient material, we mixed the desired cell fractions with 5 ml of HABG. The cell suspension containing the desired cell fractions was centrifuged for 3 min at 22°C at 200 × *g*, and the supernatant containing the debris was discarded. Finally, the cell pellet was loosened by flicking the tube, and the cells were resuspended in 1 ml of 0.01% BSA (in PBS). This final cell suspension solution was passed through a 40-μm strainer (Fisher Scientific) to discard debris, followed by cell counting.

(iv) Drop-seq single-cell barcoding, library preparation, and sequencing. Drop-seq was performed as previously described[144], with the following modifications: (i) the number of beads in a single PCR tube was increased to 4,000/tube, (ii) the number of PCR cycles was increased to 4 + 11 cycles, and (iii) multiple PCR tubes were pooled prior to size selection and purification with AMPure XP (Beckman Coulter, Brea, CA). The amplified cDNAs were then checked using the Agilent TapeStation system (Agilent, Santa Clara, CA) for library quality, average size, and

concentration estimation. The samples were then tagmented using a Nextera DNA library preparation kit (Illumina, San Diego, CA), and multiplex indices were added. The Drop-seq library molar concentration was determined by Qubit fluorometric quantitation (Thermo Fisher, Canoga Park, CA), and library fragment length was estimated using a TapeStation. Sequencing was performed on an Illumina HiSeq 4000 (Illumina, San Diego, CA) instrument using the Drop-seq custom read 1B primer (IDT, Coralville, IA). 100-bp paired-end reads were generated with an 8-bp index read for multiplexing. Read 1 consists of the 12-bp cell barcode, followed by the 8-bp unique molecular identifier (UMI). Read 2 contains the single cell transcripts.

(v) Drop-seq data preprocessing and quality control. The demultiplexed fastq files from the Drop-seq sequencing data were processed into a digital gene expression matrix using the dropSeqPipe (https://github.com/Hoohm/dropSeqPipe) SnakeMake wrapper for Drop-seq tools version 1.13. Briefly, read pairs with quality score of <10 were discarded, transcript reads were trimmed to remove TSO contamination, and poly(A) tails and then aligned to mm10 using STAR v2.5.0c with default settings. A digital gene expression matrix for each sample was generated, wherein each row is the read count of a gene, and each column represents a unique cell. The transcript counts of each cell were normalized by the total number of UMIs for that cell. These values were then multiplied by 10,000 and Ln transformed. Digital gene expression matrices from the six samples (three PBS- and three ASO-treated samples) were combined to create a pooled digital gene expression matrix. Single cells were identified from background noise by using a threshold of at least 250 genes and 500 transcripts.

(vi) Identification of cell clusters. The Seurat R package (version 2.3.1; https://github.com/satijalab/seurat) was used to project all sequenced cells onto two dimensions using uniform manifold approximation and projection (UMAP), and Louvain modularity clustering[164] was used to define clusters. To further refine the microglia and neuronal

cell clusters, clusters expressing either microglia- or neuron-specific markers were pooled, projected onto two dimensions using t-distributed stochastic neighbor embedding (t-SNE), and reanalyzed separately in a similar fashion, only considering this microglia or neuronal subset. Briefly, the most highly variable genes were identified using the mean and dispersion (variance/mean), which were used to scale and center the data. Principal component analysis was performed on this normalized data, and significant principal components were identified using the JackStraw permutation-based approach[165]. These significant PCs were used in t-SNE and UMAP to project the data onto two dimensions, and graph-based clustering was used to identify cell clusters.

(vii) Resolving cell identities of the cell clusters. To resolve the identities of the cell clusters, the single-cell data were mapped with a hippocampal reference data set[146] using FindTransferAnchors and TransferData in the Seurat Package[40]. Cell types were confirmed by comparing marker genes to known markers for hippocampal cell types and neuronal subtypes[166–168]. These markers were sufficient to confirm all major cell types, as well as neuronal subpopulations.

(viii) Identification of differentially expressed genes (DEGs). Differentially expressed genes were determined using FindAllMarkers in the Seurat Package. The log fold change was 0.25, and the minimum fraction of cells expressing each gene was set at 0.10. The Wilcoxon test was used to determine differential expression. The Benjamini-Hochberg procedure was used to correct multiple testing, and an FDR threshold of 0.05 was used to determine differentially expressed genes.

(ix) Pathway analysis of DEGs. *P* values for enriched pathways were determined using a hypergeometric distribution, taking into account the number of overlapping differentially

expressed genes with a pathway, the total number of genes in the pathway, and the total number of genes, and the total number of differentially expressed genes. All pathways were determined significant or suggestive with a Bonferroni corrected *P* value of 0.05. Significant pathways were determined with Benjamini-Hochberg corrected differentially expressed genes at a threshold of 0.05. Suggestive pathways were determined with differentially expressed genes at a *P* value threshold of 0.01.

Data availability.

The sequencing data have been deposited to GEO under accession number GSE142535.

4.5 Figures

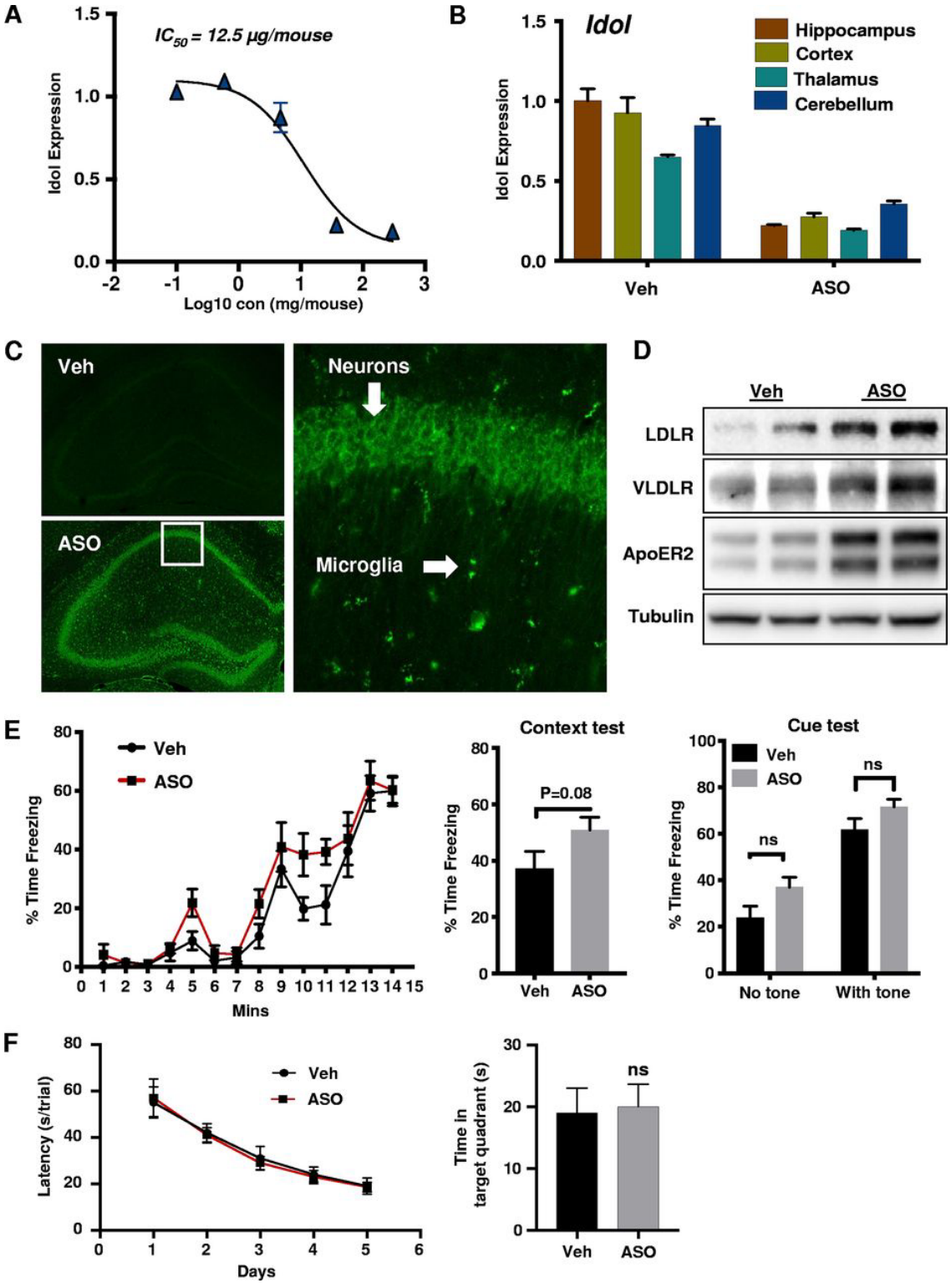

A
IC50 = 12.5 µg/mouse
Idol Expression
Log10 con (mg/mouse)
B
Idol
Hippocampus
Cortex
Thalamus
Cerebellum
Idol Expression
Veh
ASO
C
Veh
ASO
Neurons
Microglia
D
Veh
ASO
LDLR
VLDLR
ApoER2
Tubulin
E
Veh
ASO
% Time Freezing
Mins
Context test
P=0.08
Cue test
ns
ns
No tone
With tone
F
Latency (s/trial)
Days
Time in target quadrant (s)
ns

Figure 4.1 IDOL expression reduced following ASO injection

(A) IDOL ASO administration via i.c.v. bolus injection led to dose-dependent reduction of IDOL mRNA levels in the brains of C56BL/6j mice. (B) IDOL ASO (40 μg) administration suppresses the IDOL mRNA level across different brain regions detected by RT-PCR. (C) Immunofluorescent staining of brain sections from PBS- or ASO-treated APP/PS1 mice. Primary antibody against the general backbone of ASOs (a gift from Ionis Pharmaceuticals) was used to visualize the distribution of ASOs in brain cells. ASOs (green) are taken up by both neurons and microglia (arrows). (D) Representative Western blot analysis of total protein lysate from brains of WT mice. (E) Average freezing behavior during the fear conditioning training (left), average freezing behavior over 5 min, 24 h after training, in the same context in which training was carried out on day 1 (middle), and average freezing behavior in the different context 48 h after training (right), The tone was played for the last 2 min. All values are expressed in means ± the standard errors of the mean (SEM), using two-way repeated-measures ANOVA for statistical analysis (*, $P < 0.05$; **, $P < 0.01$). (F) Escape latency to find the hidden platform during training trials of wild-type mice in the Morris water maze ($n = 5$/group) (left) and time spent in the target quadrant searching for the hidden platform within a 1-min test duration ($n = 5$/group) (right).

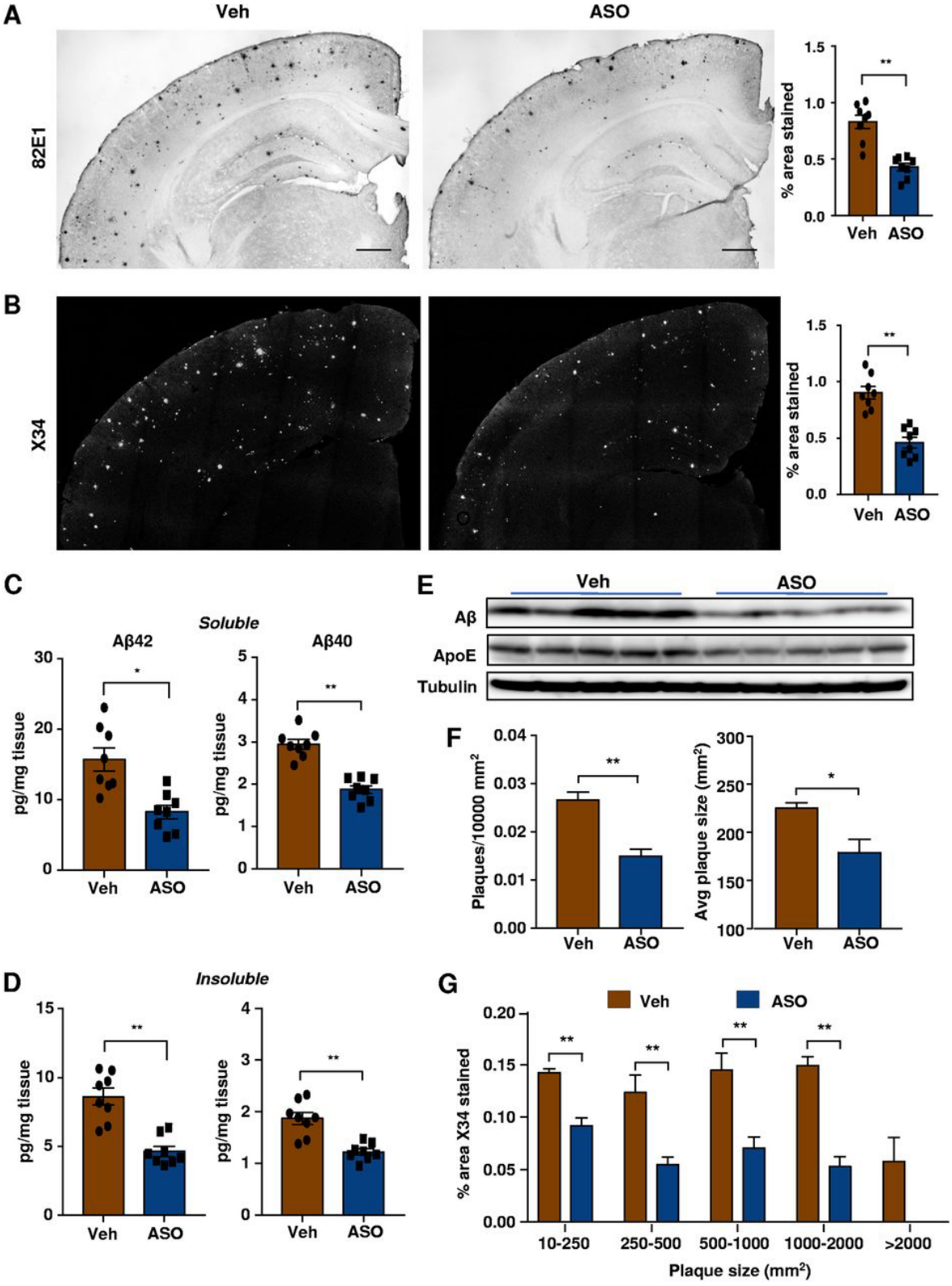

A
Veh
ASO
82E1
% area stained
Veh
ASO
B
X34
% area stained
Veh
ASO
C
Soluble
Aβ42
Aβ40
pg/mg tissue
Veh
ASO
D
Insoluble
pg/mg tissue
Veh
ASO
E
Veh
ASO
Aβ
ApoE
Tubulin
F
Plaques/10000 mm²
Avg plaque size (mm²)
Veh
ASO
G
Veh
ASO
% area X34 stained
10-250
250-500
500-1000
1000-2000
>2000
Plaque size (mm²)

Figure 4.2 ASO injection reduces plaque

(A) Brain sections from APP/PS1 mice were immunostained with anti-Aβ antibody 82E1, and the extent of Aβ deposition was quantified from the cortex (right). (B) Brain sections from the same cohort were stained with X-34 dye, and the fibrillar plaque load was quantified (right). *, $P < 0.05$; **, $P < 0.01$. (C) Soluble (RIPA fraction) Aβ40 and Aβ42 levels were measured from the cortex. (D) Insoluble (guanidine fraction) Aβ40 and Aβ42 levels were measured from the same cohort ($n = 8$ per group). *, $P < 0.05$; **, $P < 0.01$. (E) Western blot analysis of Aβ and ApoE from RIPA fractions of cortical lysates. (F) The densities of Aβ antibody-stained plaques and average plaque sizes were analyzed in the same cohort of mice. (G) Analysis of plaque distribution based on size and the total area covered by plaques in each group. *, $P < 0.05$; **, $P < 0.01$.

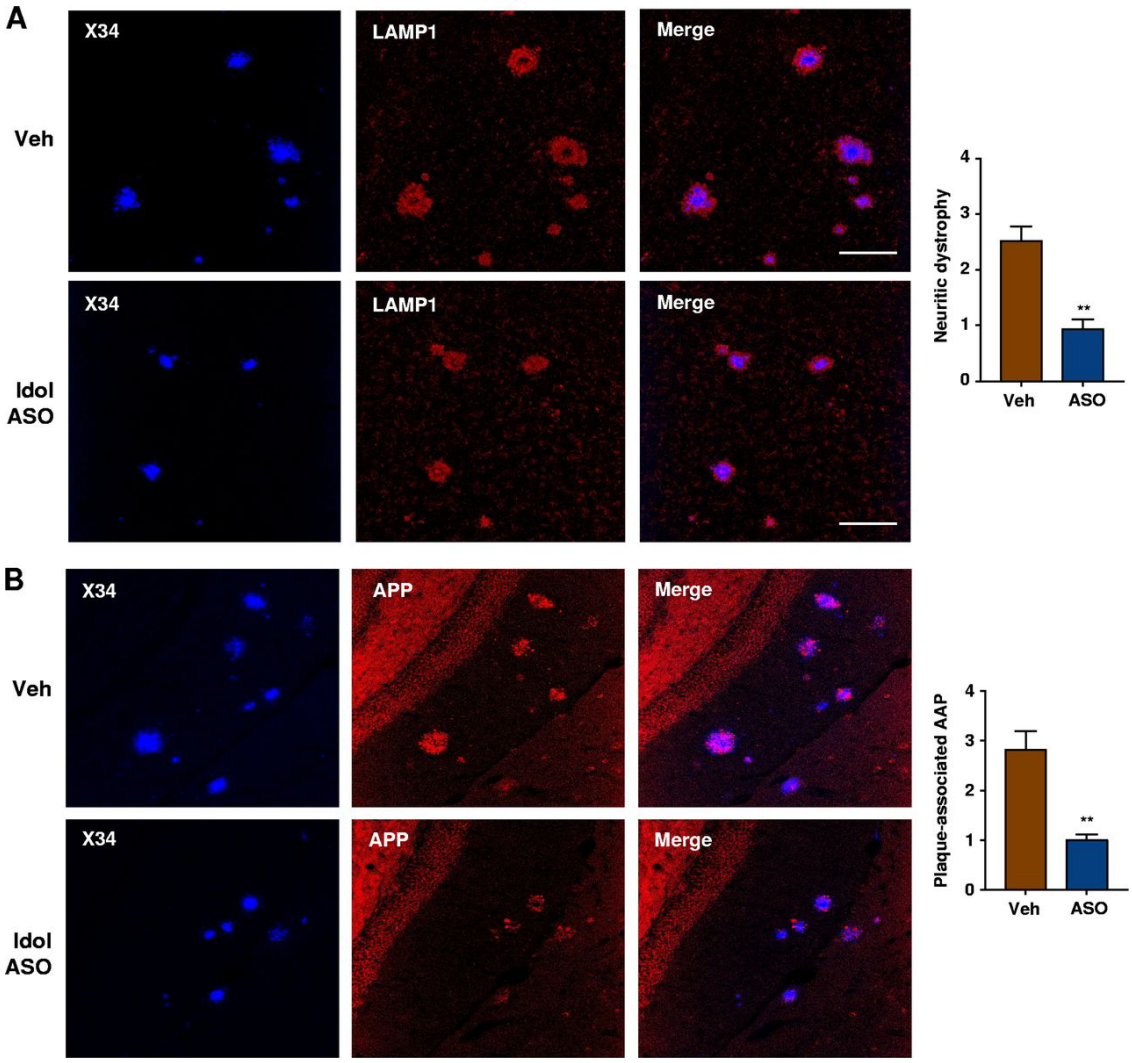


Figure 4.3 ASO injection reduces neuritic dystrophy and plaque-associated AAP

(A) Representative images of brain sections from PBS- or IDOL ASO-treated APP/PS1 mice stained with LAMP1 (scale bars, 100 μm), and the volume of LAMP1 staining in cortex was quantified (n = 8 per group). (B) Representative images of brain sections from PBS- or IDOL ASO-treated APP/PS1 mice costained with X-34 and APP, and the volume of APP staining associated with amyloid plaques was quantified (n = 8 per group).

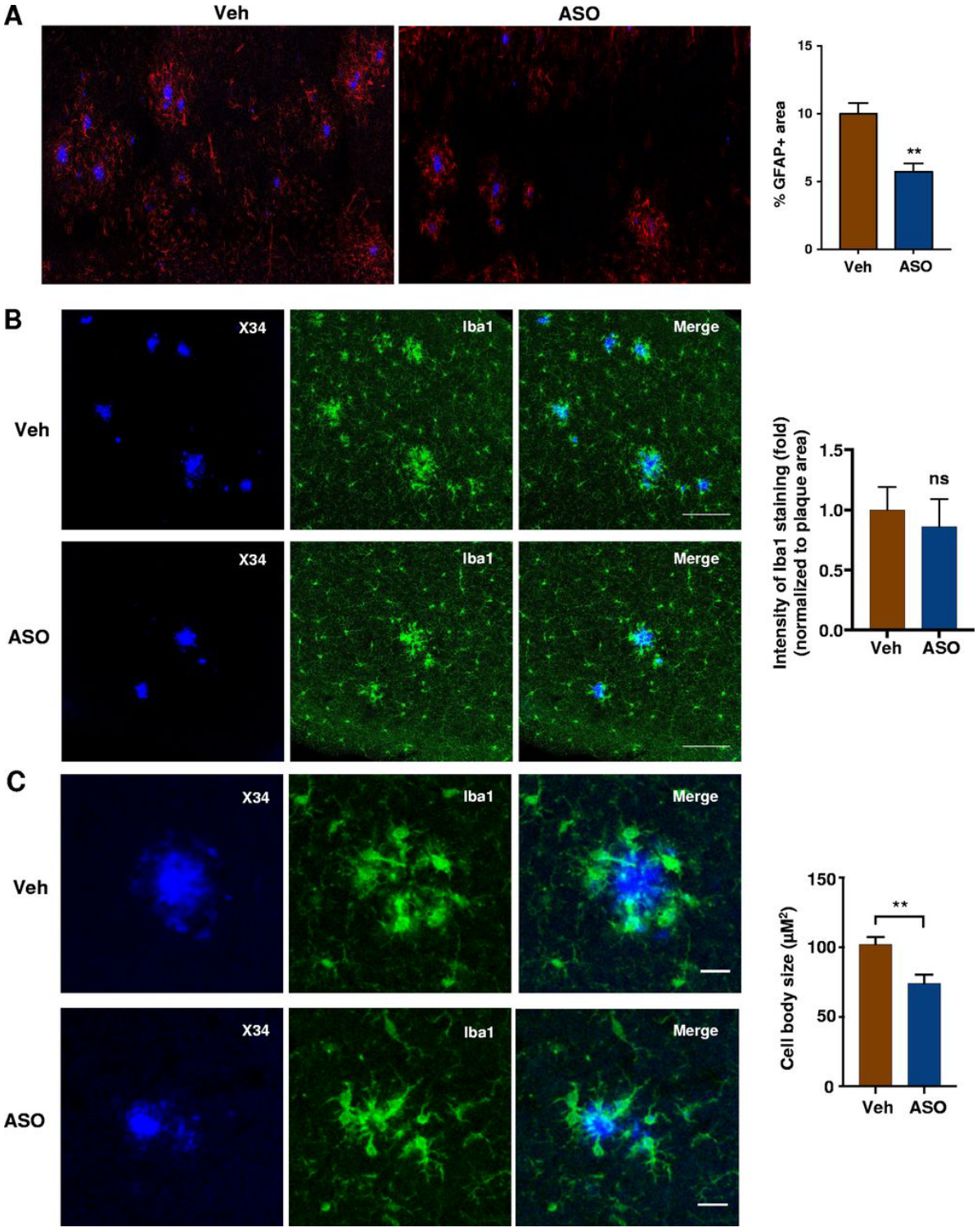


Figure 4.4 ASO injection reduces GFAP-positive astrocytes

(A) Representative images of brain sections from PBS- or IDOL ASO-treated APP/PS1 mice costained with X-34 and GFAP, and the volume of GFAP staining was quantified ($n = 8$ per group). *, $P < 0.05$; **, $P < 0.01$. (B) Representative images of brain sections from PBS- or IDOL ASO-treated APP/PS1 mice stained with Iba1 (scale bars, 100 μm), and the volume of Iba1 staining normalized to the plaque size was quantified ($n = 8$ per group). (C) Representative images of brain sections from PBS- or IDOL ASO-treated APP/PS1 mice stained with Iba1 (scale bars, 10 μm), and the diameters of the microglia cell bodies around the plaques were quantified ($n > 50$ microglia from each mouse, and eight mice for each group were measured). *, $P < 0.05$; **, $P < 0.01$.

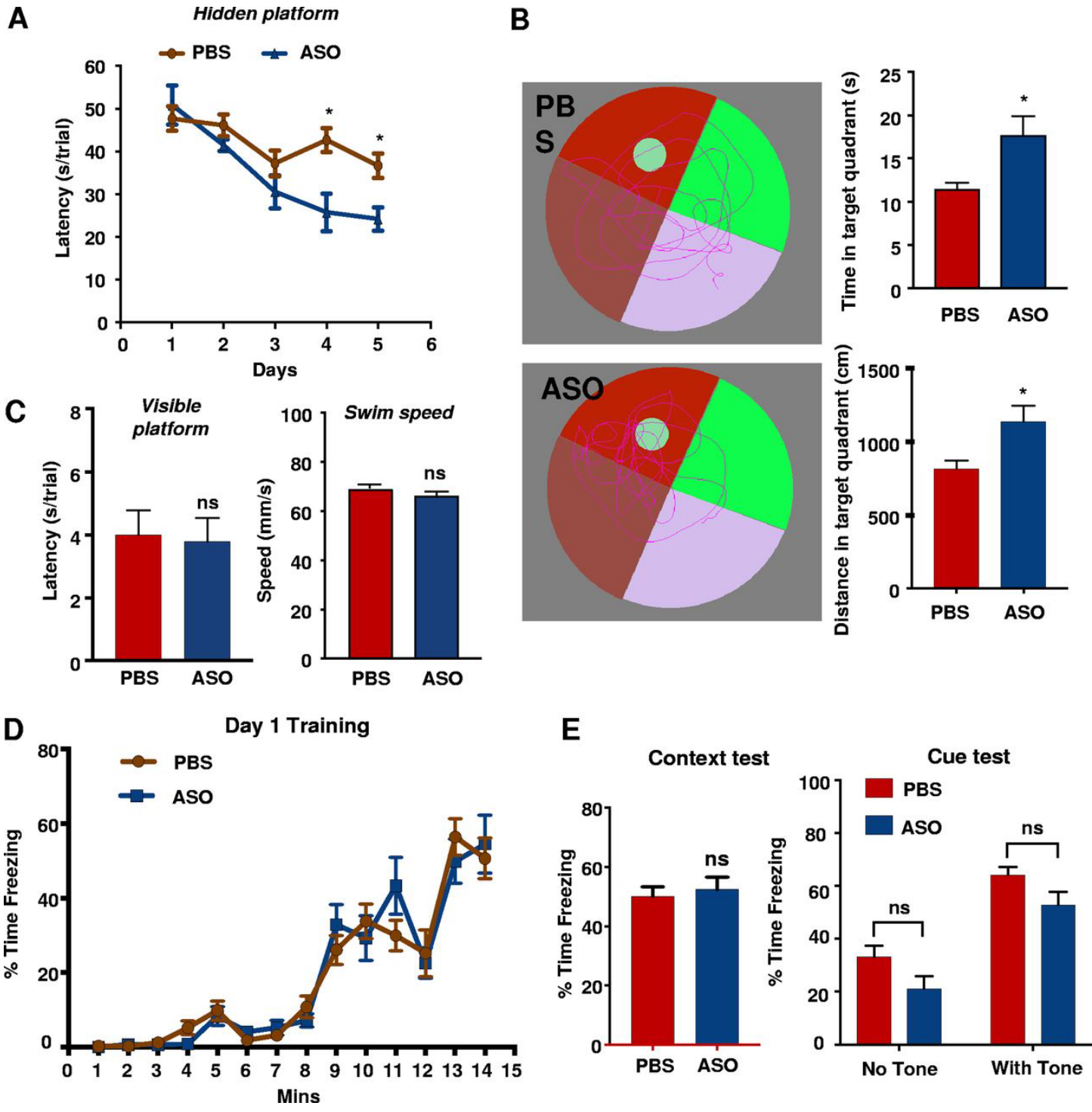


Figure 4.5 ASO injection recovers learning and memory deficits

(A) Escape latency to find the hidden platform during training trials of APP/PS1 mice in the Morris water maze (n = 12/group). (B) Representative swimming path of APP/PS1 mice treated with PBS or IDOL ASO (left) and time or distance spent in the target quadrant searching for the hidden platform within 1 min test duration (n = 12/group) (right). *, $P < 0.05$ (one-way ANOVA with repeated measures). (C) Escape latency to find the visible platform and swimming speeds of APP/PS1 mice in each group. (D) Average freezing behavior during the fear conditioning training. (E) Average freezing behavior over 5 min, 24 h after training, in the same context in which training was carried out on day 1 (left), and average freezing behavior in the different context 48 h after training. A tone was played for the last 2 min (right). All values are expressed as means ± the SEM, using two-way repeated-measures ANOVA for statistical analysis. *, $P < 0.05$; **, $P < 0.01$.

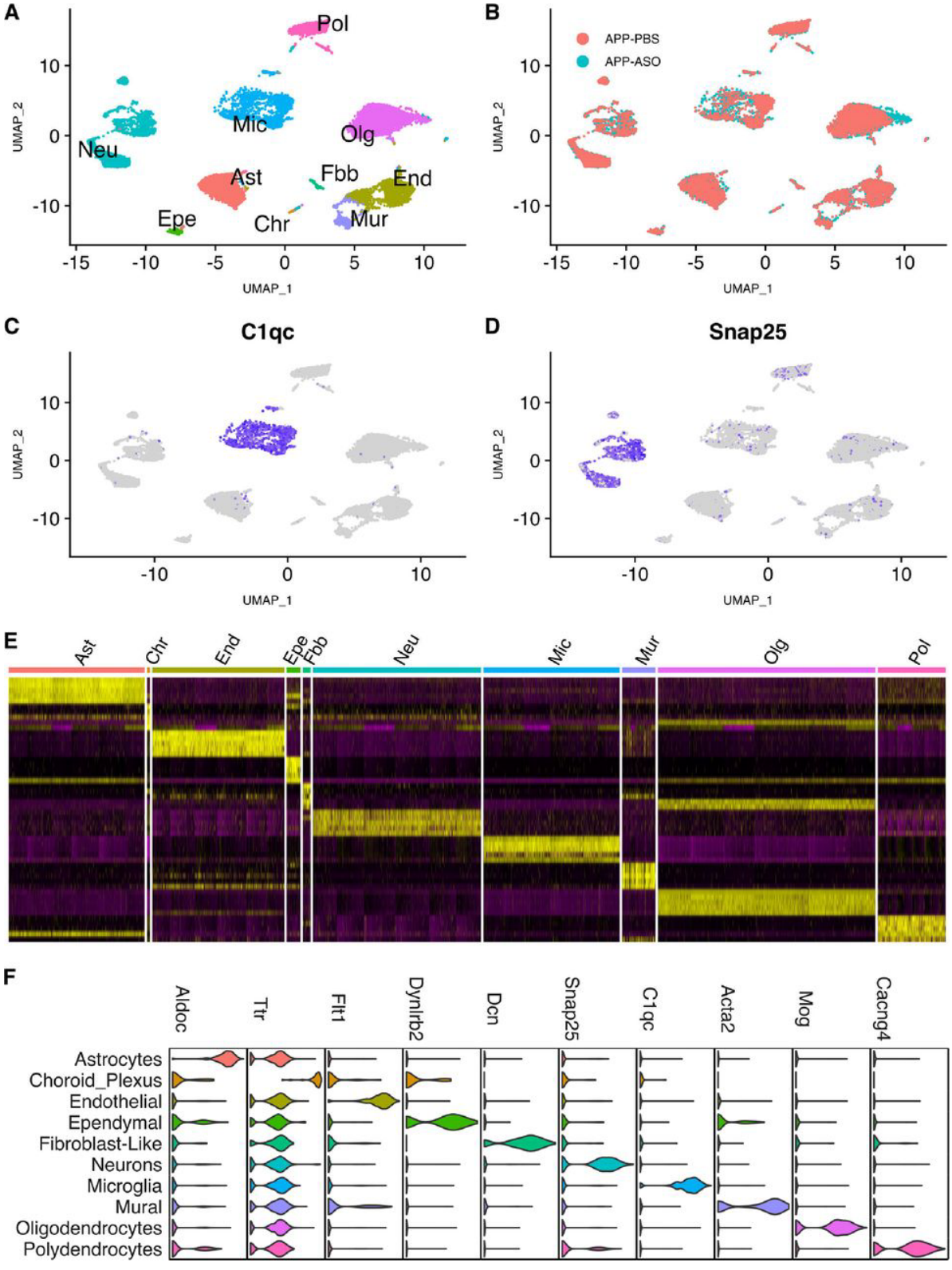

A
Pol
Mic
Olg
Neu
Ast
Fbb
End
Epe
Chr
Mur
UMAP_2
UMAP_1
B
APP-PBS
APP-ASO
C
C1qc
D
Snap25
E
Ast
Chr
End
Epe
Fbb
Neu
Mic
Mur
Olg
Pol
F
Aldoc
Ttr
Flt1
Dynlrb2
Dcn
Snap25
C1qc
Acta2
Mog
Cacng4
Astrocytes
Choroid_Plexus
Endothelial
Ependymal
Fibroblast-Like
Neurons
Microglia
Mural
Oligodendrocytes
Polydendrocytes

Figure 4.6 scRNAseq reveals distinct hippocampal cell types

(A) Uniform manifold approximation and projection (UMAP) dimensionality reduction plot showing hippocampal cell type cluster separation. Each colored dot is a cell and different cell types are labeled with different colors. (B) UMAP plot showing no clear cell type separation by treatment. Red cells originated from PBS-treated mice, and blue cells originated from IDOL ASO-treated mice. (C and D) Feature plot highlighting the expression of known cell markers: *C1qc* for Microglia (C) and *Snap25* for neuron (D). (E) Heat map showing distinct cluster-specific gene expression patterns by plotting the top ten marker genes from each cluster (*y* axis) against the cell types (*x* axis). (F) Normalized expression values of top cell-type-specific marker genes are plotted as violin plots, with cell types as rows and genes as columns. Cells were from three PBS- and three ASO-treated mice.

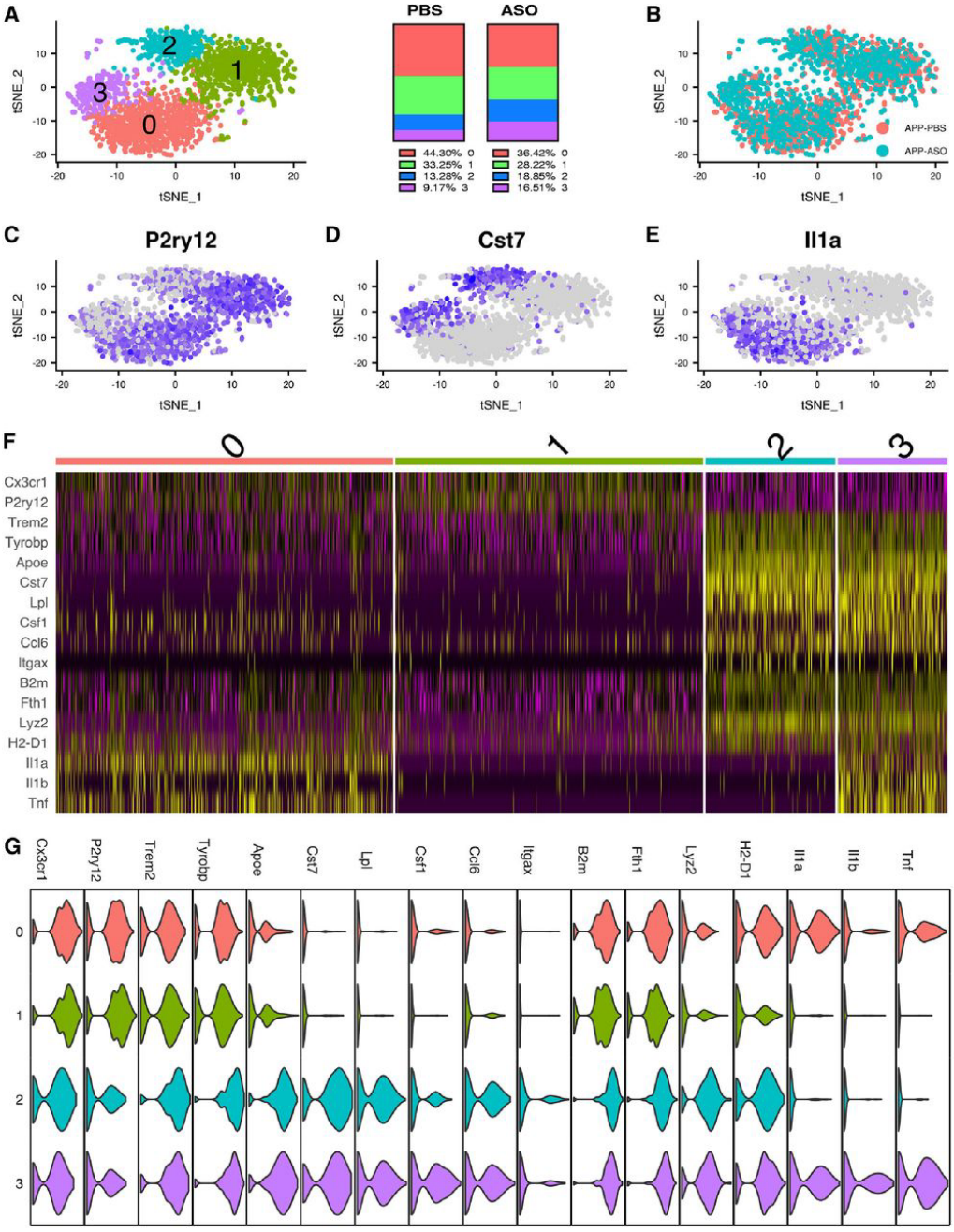


Figure 4.7 ASO treatment shifts the microglia population

(A) t-SNE plot showing microglia subtypes. Each colored dot is a cell, microglia subtypes are labeled with different colors (left), and the percentages of each subtypes in each treatment group are indicated (right). (B) t-SNE plot shows no clear separation of microglia subtypes by treatment

group. Microglia originating from PBS-treated mice are labeled in red, and microglia originating from IDOL ASO-treated mice are labeled in blue. (C to E) Feature plot highlighting key marker genes for microglia subtypes. *P2ry12*, a homeostatic microglia marker gene; *Cst7*, a disease associated microglia marker gene; *Il1a*, an inflammatory microglia marker gene. (F) Heat map showing the distinct transcriptional patterns of the four microglia subgroups. (G) Normalized expression values of top marker genes for microglia subtypes are plotted as violin plots, with subtypes as rows and genes as columns.

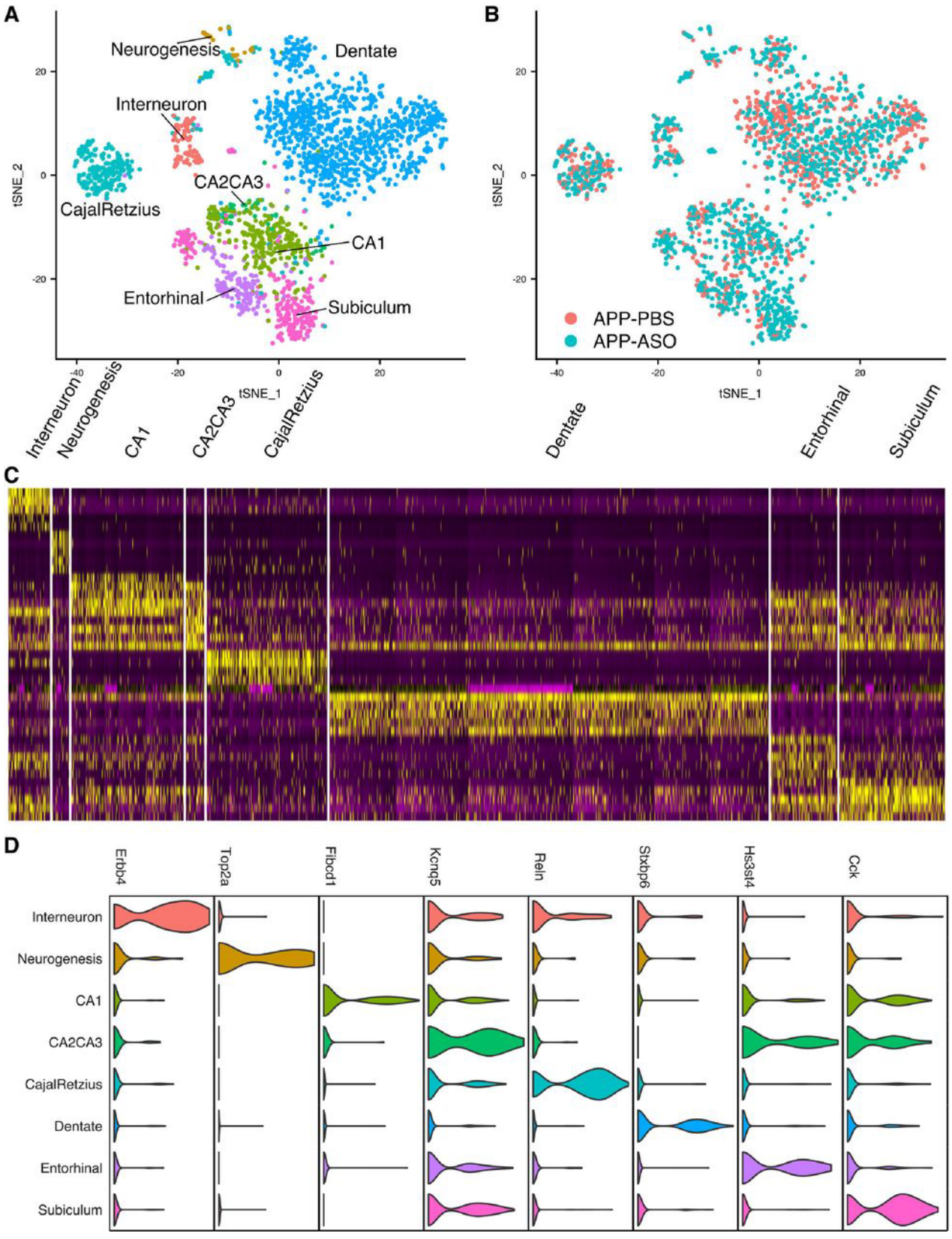

A
Neurogenesis
Dentate
Interneuron
CA2CA3
CajalRetzius
CA1
Entorhinal
Subiculum
tSNE_2
tSNE_1
Interneuron
Neurogenesis
CA1
CA2CA3
CajalRetzius
B
APP-PBS
APP-ASO
Dentate
Entorhinal
Subiculum
C
D
Erbb4
Top2a
Fibcd1
Kcnq5
Reln
Stxbp6
Hs3st4
Cck
Interneuron
Neurogenesis
CA1
CA2CA3
CajalRetzius
Dentate
Entorhinal
Subiculum

Figure 4.8 scRNAseq hippocampal neuronal subtypes

(A) t-SNE plot showing neuronal subtypes. Each color indicates a different cell type cluster identified. (B) t-SNE plot showing no clear separation of neuronal subtypes by treatment group. Red cells originated from PBS-treated mice, and blue cells originated from IDOL ASO-treated mice. (C) A heat map of Drop-seq defined marker genes of major neuronal subtypes shows distinct gene expression patterns between each neuronal subtype. (D) Normalized expression values of top marker genes for neuronal subtypes are plotted as violin plots, with subtypes as rows and genes as columns.

4.6 Tables

Table 4.1. DEGs and Pathways

Cell type	No. of DEGs	Top DEG pathway	Top four or five representative DEGs
Astrocyte	108 (6)	Cholesterol homeostasis	Up: Clu. Down: Hmgcs1, Cd9, Alcam
		Mtorc1 signaling	Up: Bhlhe40, Hspa9. Down: Gapdh, Hmgcs1, Cd9
Endothelial	129 (6)	TNFA signaling via NF-κB	Up: Fosb, Edn1, Egr1. Down: Klf9, Tsc22d1
		Angiogenesis	Up: Thbd. Down: Vtn, App, Kcnj8
		P53 pathway	Up: Fos. Down: Fas, Ccnd3, Tsc22d1, App
		Androgen response	Up: B2m, Elk4. Down: Sepp1, Ccnd3, Tsc22d1
		UV response up	Up: Fosb, Tfrc, Fos, Ptprd. Down: Ccnd3
		Amyloids	Up: Ttr, Cst3, B2m. Down: App, H2afz
		Heme metabolism	Up: Tfrc, Pqlc1. Down: Sec14l1, Slc2a1, Tcea1
		Focal adhesion	Up: Cav2, Actb. Down: Itga6, Vtn, Ccnd3
Fibroblast-like	61	Signaling by insulin receptor	Down: Sos1, Atp6v1e1, Atp6v0b, Eif4g1
		mRNA splicing	Down: Hnrnpf, Srsf1, Srsf7, Prpf6
		G_2/M checkpoint	Down: Srsf1, Ythdc1, Notch2, Amd1
		Processing of capped intron containing pre-mRNA	Down: Hnrnpf, Srsf1, Srsf7, Prpf6
Interneuron	39	Hemostasis	Up: Ppp2r5c. Down: App, Zfpm2, Atp1b1, L1cam
Microglial macrophage	170 (16)	Lysosome	Up: Npc2, Ctsd, Ctsb, Cd63, Hexa
		UV response up	Up: Cxcl2. Down: Ccnd3, Rhob, Lyn, Junb
		GPVI-mediated activation cascade	Down: Rhob, Lyn, Plcg2, Vav1, Pik3r1
		B cell receptor signaling pathway	Down: Lyn, Inpp5d, Plcg2, Jun, Vav1
		Phosphatidylinositol signaling system	Up: Dgkd. Down: Inpp4b, Inpp5d, Plcg2, Pip4k2a
		Androgen response	Up: B2m. Down: Ccnd3, Inpp4b, Fkbp5, Mak
		Biocarta Ppara pathway	Up: Hsp90aa1, Lpl. Down: Nrip1, Jun, Pik3r1
		Allograft rejection	Up: B2m, Tlr2. Down: Ccnd3, Ccr5, Lyn
		Antigen activates B cell receptor leading to generation of second messengers	Down: Lyn, Plcg2, Vav1, Pik3r1
		Hemostasis	Up: Cd63. Down: Rhob, Slc8a1, Lyn, Dock4
		Biocarta Fcer1 pathway	Down: Lyn, Jun, Vav1, Pik3r1
		Fc epsilon Ri signaling pathway	Down: Lyn, Inpp5d, Plcg2, Vav1, Pik3r1
		Leukocyte transendothelial migration	Up: Cyba. Down: Rhoh, Arhgap5, Plcg2, Vav1
		TNFA signaling via NF-κB	Up: Tlr2, Cxcl2, Cdkn1a. Down: Rhob, Junb
		Complement	Up: Ctsd, Ctsb. Down: Lyn, Dock4, Lgmn
		Inflammatory response	Up: Tlr2, Rgs1, Cdkn1a. Down: Lyn, Stab1
		Interleukin-2 Stat5 signaling	Up: Cst7, Ctsz. Down: Ccnd3, Rhob, Rhoh
		Pathways in cancer	Up: Cdkn1a. Down: Tpm3, Tgfbr1, Runx1, Plcg2
		Erbb signaling pathway	Up: Cdkn1a. Down: Plcg2, Jun, Nrg3, Pik3r1
		Antigen processing and presentation	Up: B2m, Ctsb, Hsp90aa1, Tapbp. Down: Lgmn
		Apoptosis	Up: Cdkn1a. Down: Rhob, Txnip, Jun, Btg2
		Signaling by Rho GTPases	Down: Srgap2, Rhob, Rhoh, Ophn1, Arhgap25
		FcγR-mediated phagocytosis	Down: Lyn, Inpp5d, Plcg2, Vav1, Pik3r1
		Chemokine signaling pathway	Up: Cxcl2, Ccl3. Down: Elmo1, Ccr5, Lyn
		Platelet activation signaling and aggregation	Up: Cd63, Dgkd. Down: Rhob, Lyn, Plcg2
		Toll receptor cascades	Up: Ctsb, Tlr2. Down: Plcg2, Lgmn, Jun
		Glioma	Up: Cdkn1a, Igf1. Down: Plcg2, Pik3r1
		Leishmania infection	Up: Tlr2, Cyba. Down: Jun, Ifngr1
		P53 pathway	Up: Ctsd, Cdkn1a. Down: Ccnd3, Txnip, Jun
		Chronic myeloid leukemia	Up: Cdkn1a. Down: Tgfbr1, Runx1, Pik3r1
Mural	88 (1)	Smooth muscle contraction	Up: Lmod1, Calm2, Acta2, Myh11. Down: Tpm4
		Muscle contraction	Up: Lmod1, Calm2, Acta2, Myh11. Down: Tpm4
		Vascular smooth muscle contraction	Up: Rock1, Calm2, Acta2, Myh11. Down: Ramp2
		Myogenesis	Up: Myh11, Smtn. Down: App, Igfbp7, Ablim1
		Integrin cell surface interactions	Down: Pecam1, Lamc1, Bsg, Vtn
		Focal adhesion	Up: Rock1, Mylk. Down: Flt1, Lamc1, Vtn
Cajal-Retzius neuron	46 (1)	Wnt signaling pathway	Down: Csnk1a1, Tbl1x, Ccnd2, Btrc
Neuron dentate	90 (2)	Alzheimer's disease	Up: Itpr1. Down: Gapdh, Apbb1, Ndufb4, Ndufc2
Neuron subiculum	93	Oxidative phosphorylation	Up: Vdac1, Atp5g3, Afg3l2, Cox6b1. Down: Pdhb
Oligodendrocyte	261 (54)	Cardiac muscle contraction	Up: Cox6a1, Cox7a2, Cacnb4, Cox4i1. Down: Atp1b3
		Pathogenic *E. coli* infection	Up: Arpc1a, Actb, Arhgef2, Actg1, Ywhaz
		Biocarta Chrebp2 pathway	Up: Ywhaz, Ywhab. Down: Ppp2r2a, Ywhae
Polydendrocyte	95 (4)	Cardiac muscle contraction	Up: Cox4i1, Cox6a1. Down: Cacna2d3, Cacna2d1
		Parkinson's disease	Up: Cox4i1, Cox6a1, Ubb, Atp5e, Pink1
		Oxidative phosphorylation	Up: Cox4i1, Ldhb, Cox6a1, Atp5e. Down: Timm13
		Alzheimer's disease	Up: Apoe, Cox4i1, Cox6a1, Atp5e. Down: Gsk3b
		Amyloids	Up: Ttr, Cst3, Itm2b, B2m

Chapter 5. Defined subsets of sympathetic neurons innervate the heart and differentially respond to cardiac pathology

5.1 Introduction

Cardiac autonomic regulation is exerted by reflex loops consisting of afferent vagal and spinal afferent inputs to the nervous system, and descending control by efferent parasympathetic and sympathetic control[169,170]. Postganglionic sympathetic neurons (PGNs) responsible for direct cardiac sympathetic neurotransmission are located within the cervico-thoracic sympathetic chain, particularly, the stellate ganglia and middle cervical ganglia[171,172]. These stellate ganglion neurons (SGNs) release neurotransmitters that regulate all aspects of cardiac function including inotropy and chronotropy. However, following cardiac injury such as myocardial infarction (MI) and heart failure (HF), SGNs undergo profound structural, neurochemical, and electrophysiologic remodeling, contributing directly to progressive cardiac dysfunction and lethal ventricular arrhythmias such as ventricular tachycardia/fibrillation[173–177]. As a result, anti-adrenergic therapies such as beta-adrenergic receptor and neurohormonal blockers remain cornerstone treatments of MI and VT/VF[178–180]. Further, interventional and surgical therapies targeting the stellate ganglion (e.g., percutaneous anesthetic blockade and surgical sympathetic denervation, respectively) are increasingly applied clinically[181–183].

Despite strong clinical data supporting therapies targeting SGNs to treat cardiac disease, the efficacy of available therapies remains limited, particularly by systemic side effects such as hypotension or renal dysfunction[184–188]. Little is known about the subtypes and properties of neurons that innervate the heart, compared to neurons innervating other tissue beds e.g., skin and paw, that might facilitate the development of more specific targets for therapeutic purposes. Prior work investigating SGN subpopulations focusing on nipple and piloerector neurons utilized single cell RNA sequencing (scRNAseq) to identify eight subpopulations of neurons within the stellate ganglion[189].Specifically, five noradrenergic, two cholinergic sympathetic neurons, and one

glutaminergic neuronal subtypes were identified, each with unique combinations of transcription factors, neuropeptides, and ion channels. Whether a specific subtype of neurons innervates the heart, or whether various subtypes contribute to cardiac adrenergic control remains unknown.

In this study, we utilized retrograde viral tracing using adeno-associated viruses (AAVs) injected into the subepicardium in mice, to identify and label cardiac-specific SGNs with the reporter green fluorescent protein (GFP), and in contrast, the front paws were injected with AAVs encoding the Td-tomato protein. Ganglia were collected after 4 weeks and subjected to dissociation and scRNAseq, with cells containing GFP or tdTomato identifying cardiac and paw-innervating neurons respectively. We found that the heart and paws are innervated by three and four subtypes of noradrenergic neurons respectively, which largely segregate by neuropeptide Y (NPY expression). These subtypes were also identified in porcine and human stellate ganglia. We find that NPY+ vs. NPY- neurons exhibit unique transcriptional profiles, along with unique morphological, neurochemical, and electrophysiologic properties. Functional studies in NPY ablated mice indicate that NPY is required for maximal cardiac sympatho-excitation. Further, heart failure reduces the transcriptomic diversity of cardiac sympathetic neuronal subtypes, yielding a dominant neuronal subtype with an altered transcriptome. Collectively, these findings shed light on cell-specific cardiac adrenergic regulation, with implications for novel therapeutic targeting of adrenergic signaling following cardiac injury.

5.2 Materials and Methods

Mouse strains

Animal experiments complied with institutional guidelines and ethical regulations. The study protocol was approved by the UCLA institutional Animal Care and Use Committee. The ethical approvals for the use of AAVs were taken from Institutional Biosafety Committee (IBC), UCLA. Adult male mice, aged between 8 to 10 weeks, were housed according to the standard laboratory

conditions (12 h light/dark) with ad libitum access to food and water. Mice strains C57BL/6J (000664), NPY-hrGFP (006417), and Npy-IRES-Cre (027851) mice were purchased from the Jackson Laboratory. Dilated cardiomyopathy mice (DCM-Tg9) were a gift from Dr. Patrick Jay.

Retrograde neuronal tracing experiments

We used C57BL/6J, NPY-hrGFP, and DCM-Tg9 mice strains for tracing the target organs-innervated neurons in the stellate ganglion. We injected 10 µl of each retrograde tracers Adeno-associated virus (AAV) or cholera toxin-B (CTB) either from heart or paw. AAVs serotype AAV2/retro-CAG-flex-tdTomato (1e+12 gc/ml) and AAV2/retro-CAG-flex-GFP (1e+12 gc/ml) were procured from Viral vector core, Boston children's Hospital. The cholera toxin subunit-B conjugated; Alexa Fluor-488 (C-34775), Alexa Fluor-555; (C-34776), Alexa Fluor-647 (C-34778) were procured from Molecular Probes.

Single cell clustering of neuronal subpopulation

To identify the neuronal subtypes from neuronal populations, we processed scRNAseq data as mentioned in the supplementary method section 3. First, we integrated the data as described, reduced the dimensionality with PCA on 30 PCs, found the nearest neighbor embedding, and determined the 8 subclusters with the Louvain community detection algorithm. We identified marker genes for each cluster by the Wilcoxon test (FDR < 0.05). These neuronal subclusters were annotated according to previously defined neuronal cell subtypes[189]. We identified 5 neuro adrenergic clusters (NA1a, NA1b, NA1c, NA2, NA3), 2 acetylcholine (Ach1, Ach2), and 1 glutamatergic subtype (Supplementary Figure 5.2). We identified cardiac and paw innervating neurons based on the expression of GFP and tdTomato transcripts respectively. We determined *Npy* (low, medium, and high) subpopulations of neurons through the bottom 30%, middle 40%, and top 30% quantile. We identified Npy associated genes in the cardiac (NA1a, NA1b, NA3) and paw (NA1a, NA1b, NA1c, NA3) innervating subclusters. For each gene, to eliminate the effects

of high zero content, we identified cells that were expressing at least one transcript of both Npy and the other gene of interest. We then computed the Pearson correlation coefficient, and corrected p-values based on the Benjamini-Hochberg procedure. Significant associations (FDR<0.05) were kept for further analysis. To visualize the association of receptors, transcription factors, neuropeptides, and ion channels, we split the neurons into 3 categories based on their Npy expression. The bottom 30% based on Npy were categorized as low/negative, the middle 40% were categorized as medium, and the top 30% were categorized as high Npy expressions. We then visualized the relative expression of each gene with heatmaps using the seaborn python package.

Statistical Analysis. Microsoft Excel and GraphPad Prism 9.0.1 were used for data handling, analysis, and graphs. Sample sizes and statistical tests performed are indicated in the legend for each figure. Normality of data was assessed using the Shapiro-Wilk or Kolmogorov–Smirnov test. Normally distributed data was compared through a Welch's t-test or ANOVA, while data not normally distributed were analyzed through a Mann-Whitney test or Kruskal-Wallis test. Legend for statistical significance: *$*p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001, ****p ≤0.0001.$* All data are shown as individual data points.

Software packages. scRNAseq processing: R, cell ranger, Seurat v3. Visualization: python, scaNpy, pandas, numpy, matplotlib, seaborn. Npy association: python scipy.

5.3 Results

Three unique stellate ganglion neuronal subtypes innervate the heart

We performed scRNAseq on stellate ganglion neurons from eight retrogradely labeled C57BL/6 mice. We clustered scRNAseq data, identified cell types based on their canonical markers, and

separated out the neuronal population based on Dbh, Th, and Snap25 (Supplementary Figure 5.1, Supplementary Figure 5.2, Supplementary Figure 5.3, Supplementary Table 1, Supplementary Table 2). To identify cardiac innervating neuronal subpopulation in stellate ganglia, we labeled mice retrogradely from the heart using AAV-GFP (Figure 5.1a, left panel) and from front paws using AAV-tdTomato as control (Supplementary Figure 5.4a, left panel). We observed regional specificity in stellate ganglia for cardiac vs. paw innervating neurons. Most cardiac-labeled neurons were localized in the cranio-medial region (cardiac pole) of the stellate ganglion (Figure 5.1a, right panel) while paw labeled neurons were distributed across the stellate ganglion (Supplementary Figure 5.4a, right panel). We identified five dopaminergic, two cholinergic, and one glutamatergic neuronal subtype in the stellate ganglia (Figure 5.1b, c, Supplementary Figure 5.2).

In our scRNAseq studies, we identified labeled cells in the stellate ganglion either as cardiac- or paw- innervating neuronal subtypes based on the presence of GFP or tdTomato transcript, respectively. We found a total of 3 GFP-expressing (Cardiac) neuronal subpopulations NA1a, NA1b, and NA3 (Figure 5.1d) and 4 tdTomato-expressing (Paw) neuronal subtypes NA1a, NA1b, NA1c, and NA3 (Supplementary Figure 5.4b).

Interestingly, we found that similar neuronal subpopulations (NA1a, NA1b, and NA3) innervate both heart and front paws, however the front paws are innervated by an additional neuronal subpopulation NA1c (Supplementary Figure 5.4b). This suggests that similar subsets of neurons in the stellate ganglion are responsible for neural control of various tissues. We observed significant differences in the transcriptome of each neuronal subtype (Figure 5.1e, Supplementary Figure 5.4c), with distinct relative expression levels of secreted neuropeptides, transcription factors, and receptors (Figure 5.1f-h and Supplementary Figure 5.4e-g).

Next, we validated the retrograde labeling of SGNs from heart by staining AAV-GFP injected heart-sections with antibodies against the pan-neuronal marker PGP9.5. In multiple cardiac cross-sections, we found nerve fibers co-labeled with GFP and PGP9.5 antibodies (Supplementary Figure 5.5). Since this expression largely requires retrograde transport of AAV to the soma from nerve endings in the heart, GFP expression, and antegrade transport of GFP to cardiac nerve fibers, it confirms that we successfully labeled postganglionic neurons in the stellate ganglion by cardiac injections.

To validate the neuronal subtypes identified by scRNAseq and their localization in heart and paw neurons (Figure 5.1i and Supplementary Figure 5.4h), we performed retrograde labeling from the heart and paw using CTB conjugates to either 555 or 647 in wildtype or NPY-hrGFP transgenic mice. We then performed immunohistochemistry (IHC) using antibodies against markers specific each transcriptomic subtype of neurons. We performed combinatorial IHC as follows; *Npy*^{high}, *Cntn5*^{high} for NA1a, *Npy*^{low}, *Kcnt1*^{high} for NA1b, and *Npy*^{low}, *Vim*^{high} for NA3, and *Npy*^{low}, *Sctr*^{high} for NA1c (Figure 5.1i and Supplementary Figure 5.2g). As illustrated in the figure, we identified all expected markers in cardiac and paw neuronal subtypes. We also carried out combinatorial immunohistochemistry for the non-cardiac innervating neuronal subtypes using specific markers for each subtype. We found that less than ≤6% of cells co-labeled with the cardiac retrograde label (CTB-647) when stained for these combinations of markers (Supplementary Figure 5.6), confirming the accuracy of our AAV-labeling and scRNAseq approach.

Finally, we validated the presence of these cardiac-innervating neuronal subpopulations in porcine (Figure 5.2a) and human (Figure 5.2b) stellate ganglia to support the translational potential of the neuronal subtypes we identified. We used the following combination of subtype-specific markers; *NPY*^{high}, *CNTN5*^{high} (NA1a), *NPY*^{low}, *CASZ1*^{high} (NA1b), and *NPY*^{low}, *VIM*^{high} (NA3) in pig stellate ganglia (Figure 5.2a) and *NPY*^{high}, *CNTN5*^{high} (NA1a); *NPY*^{low}, *CASZ1*^{high} (NA1b); and *NPY*^{low}, *FABP7*^{high} (NA3) in human stellate ganglia (Figure 5.2b).

Cardiac subtypes differentially express NPY

NPY is a co-transmitter released by the sympathetic nerve terminals along with norepinephrine to modulate cardiac function dynamically[190]. Elevated circulating NPY levels are associated with adverse outcomes chronic heart-failure patients[191]. Our scRNAseq findings demonstrate high *Npy* expression in NA1a neurons, but lower expression in NA1b and NA3, independent of cardiac (or paw) innervation (Figure 5.3a and Supplementary Figure 5.7a). Furthermore, we identified 114 genes positively and 527 genes negatively associated with *Npy* (FDR<0.05; methods) for cardiac subtypes. We show the top 10 positively and negatively correlated genes with *Npy* in cardiac subtypes (Figure 5.3b) and in paw subtypes (Supplementary Figure 5.7b).

We confirmed previously known associations between *Npy* and *Gal* (Pearson correlation: 0.64; Figure 5.3c) and uncovered new correlation such as that between *Col5a3* and *Npy* (Pearson correlation: -0.52; Figure 5.3d) in cardiac neuronal subtypes. We stratified cardiac and paw subtypes by *Npy* expression level (low, medium, and high) and found unique expression patterns of neuropeptides, transcription factors, receptors, and ion channels by NPY expression level (Figure 5.3e-g, Figure 5.4a-d, and Supplementary Figure 5.7c-f). We corroborated this by performing gene set enrichment analysis on the combined, positive, and negative set of *Npy* associated genes (methods). This analysis revealed that biochemical pathways associated with high *Npy* expression included dopamine neurotransmitter release cycle (catecholamine biosynthesis and release), inotropic activity of kainate receptors (excitatory glutaminergic receptors), and respiratory electron transport (mitochondrial metabolism and function) suggesting that NPY expression is associated with neuronal metabolism and excitability (Figure 5.3h-j and Supplementary Figure 5.7g-i). These findings suggest that the high *Npy* expressing neuronal subtype NA1a and lower expressing subtypes, NA1b and NA3, are indeed distinct cardiac-innervating neuronal subtypes in stellate ganglia.

High *Npy*-expressing cardiac neurons exhibit greater excitability

To determine whether NPY-rich cardiac neuronal subtypes exhibit unique electrophysiologic properties, we examined the transcriptome of these subtypes for ion channel expression. We identified differential expression of potassium, sodium, calcium, and proton channels in cardiac NA1a compared to the other subtypes (Figure 5.4a-d). We hypothesized that the unique ion channel expression patterns in NA1a conferred differential cellular electrophysiologic properties. To test this, we studied the electrophysiological properties of *Npy* expressing neurons in the cardiac pole of the stellate ganglion. We used a reporter mouse line NPY-hrGFP that expresses humanized Renilla Green Fluorescent protein under control of the *Npy* promoter. Cells with (GFP+) and without GFP expression (GFP-), were selectively targeted in ganglion whole-mounts using epifluorescence (Figure 5.4e, f). Resting membrane potential values were -51.91 ± 3.14 mV for GFP+ and -52.45 ± 2.45 mV for GFP- cells ($p = 0.899$). Membrane input resistance values were 106.9 ±12.63 M in GFP+ and 103.5 ± 10.40 M in GFP- neurons ($p = 0.8391$) (Figure 5.4g). Current-voltage relationship (Figure 5.4g) were also similar for both cell types over a range of tested currents (-500pA to +500pA, Δ100pA) and showed a mostly linear I-V relationship with little rectification ($F_{(9, 128)} = 0.562$, $P = 0.825$). In response to depolarizing current steps, both cell types elicited an increasing number of action potentials in response to increasing magnitude of depolarization, while the excitability curves were similar ($p = 0.5392$) up to 300 pA current, NPY+ cells showed significantly a greater number of action potentials ($p < 0.0001$) than NPY- cells at 500 pA current (NPY+ = 4.250 ±0.901, n = 8; NPY- = 1.818 ± 0.1820, n = 11) (Figure 5.4i). These findings further support the notion that high NPY expressing cardiac sympathetic neurons (NA1a) possess distinct electrophysiological properties from NA1b and NA neurons which express low levels or are negative for NPY.

NPY expressing subpopulation in stellate ganglia are necessary to achieve maximal cardiac sympathoexcitation

Next, we sought to examine how high *Npy*-expressing subtype, NA1a, might impact cardiac function in physiologic studies in mice lacking NPY. We injected AAVs carrying DIO-EGFP or DIO-EGFP+taCasp 3 in right stellate ganglia of Npy-IRES-Cre mice. We electrically stimulated RSG from NPYCre mice injected either DIO-EGFP (n=3) or DIO-EGFP+taCasp 3 (n=5) at 1, 5, 10, 15, and 20 Hz and compared change in heart rate from baseline in response to escalating RSG stimulation intensities (Figure 5.5a,b). Stimulation frequency and heart rate change showed a showed a linear relationship for both control and NPY ablated mice. However, overall response in heart-rate change was significantly lower for NPY ablated mice compared to controls ($p = 0.0240$) (Figure 5.5c). We observed maximal significant heart rate change difference for NPY ablated mice (91.061 +2.628 bpm; n = 5) vs. controls (27.017 ±11.888 bpm; n = 3) at 10 Hz ($p = 0.0240$) stimulation frequency. However, the baseline heart rate remains the same for both the groups at 1, 5, 10, 15, and 20 Hz stimulation frequency (Figure 5.5d). Further to confirm the transfection efficiency of AAVs virus in stellate ganglia, we visualized the whole stellate ganglia from controls and taCasp 3 injected mice (Figure 5.5e,f). We observed EGFP labeling in the stellate ganglia from both the group but EGFP+ cells were very few in DIO-EGFP+taCasp 3 compared to those injected with only DIO-EGFP virus (Figure 5.5e,f). Moreover, in a magnified view, blebs or tiny holes appeared in taCasp 3 injected stellate ganglia indicating neuronal degradation initiated by caspase3 (Figure 5.5f). These findings suggest that NA1a neuron which express high NPY, potentially required to achieve maximal cardiac sympathoexcitation mediated by right stellate ganglia.

High *Npy* expressing neurons uniformly innervate disparate regions of the heart

Next, we sought to determine whether cardiac neuronal subtypes differentially innervate disparate regions of the heart, such as the apex vs. base of the left ventricle. We performed retrograde labeling from the apex (using CTB-555) and base (using CTB-647) in NPY-hrGFP mice (Figure 5.6a). First, we found that neurons traced from the apex or base colocalized to the craniomedial

pole of the stellate ganglion irrespective of their site of injection in heart (Figure 5.6b). Next, we examined the subtype distribution (i.e. NPY+ vs. NPY-) of neurons innervating the apex (CTB-555+), base (CTB-647+), or both (CTB-555 & CTB-647 double positive). We observed that more NPY+ neurons (vs. NPY-) were labeled from apex+base (Figure 5.6f, $p < 0.001$), apex alone (Figure 5.6g, $p < 0.001$) or base alone (Figure 5.6h, $p = 0.0026$). Across animals, ~70% of cardiac innervating neurons express high *Npy* (Figure 5.6i), compared to 30% innervating the paw. These results suggest that cardiac neuronal subtypes NA1a (high NPY expressing) and NA1b/NA3 (low to negative NPY expressing) uniformly innervate the cardiac apex and base, however, most neurons innervating the heart are NPY+.

Interestingly, we assessed whether NPY expression was associated with morphologic differences in cardiac-specific neurons identified by retrograde tracing. We found that the soma of NPY+ neurons were physically larger than those of NPY- neurons (Figure 5.6j). Mean soma size for NPY+ vs. NPY- neurons were 546±252μm2 vs.342±170 μm2 (mean±SD, $p<0.0001$).

Heart failure differentially impacts cardiac sympathetic neurons to produce a dominant subtype

To understand how cardiac SGN subtypes are altered by heart failure, we performed scRNAseq on stellate ganglia from an established transgenic mouse model of chronic nonischemic heart failure (DCM TG9)[192–194] and control littermates (WT). In this model cardiac dysfunction begins at 7 weeks of age, and over heart failure at 10 weeks, confirmed by transthoracic echocardiography (Supplementary Figure 5.8a-d).

We performed scRNAseq of stellate ganglia from 10-week old DCM animals and WT littermates. We identified the same neuronal subtypes from our prior analysis with a support vector machine (SVM) classifier. However, we observed that unlike WT animals where cardiac subtypes are equally distributed among NA1a, NA1b, and NA3 subtypes, the NA1b subtype was predominant

in DCM animals (Figure 5.7a,b). We found that NA1b neurons represented a greater proportion of DCM cardiac neurons, while NA1a and NA3 subpopulations were reduced (Figure 5.7b). Since NPY has an important role in heart failure, we characterized the distribution of NPY expression in tertiles (low, medium, and high) across the cardiac subtypes in DCM mice (Figure 5.7c) compared to WT animals. While most NA1a and NA1b neurons remained high and low NPY expressing, respectively, most NA3 neurons were high NPY expressing, suggesting remodeling of this subtype of neurons towards greater NPY expression.

We analyzed a subset of differentially expressed genes (DEGs) in cardiac SGNs from DCM and WT animals. We found significant DEGs only in the NA1b subtype with the top 10 upregulated and downregulated genes in NA1b subtype in DCM animals shown in Figure 5. 7d,e. Our data suggest that lack of DEGs in other cardiac subtypes (NA1a, NA3) is likely due to relatively low cell numbers of these subtypes in WT and DCM animals (Supplementary Figure SF 5.9c). Interestingly, we observed a specific and significant increase in the expression of M-current conducting ion channel Kcnq2 in DCM mice (Figure 5.7f). Along with Kcnq2, a slowly activating and deactivating potassium channel that plays a critical role in the regulation of neuronal excitability and synaptic transmission across the chemical synapses, we found observed trends indicating increased expression of voltage gated sodium channels (Nav1.1, Nav1.2), potassium channel (Kv11.3), calcium channel (Cav2.2), and vesicle transport proteins (Ap2a2, Snap25) (Figure 5.7f). These alterations in ion channel expression in this subtype implicate the NA1b subtype in the alterations of sympathetic neuronal excitability demonstrated in the setting of heart failure[195–200].

We performed pathway enrichment analysis on the suggestive DEGs ($p < 0.05$) stellate ganglion neurons from DCM mice and found enriched pathways (FDR < 0.05) in only NA1a and NA1b subtypes. We identified relevant significantly enriched pathways from NA1b subtypes where three or more than three genes were affected. We selected over-representative pathways: oxidative

phosphorylation (FDR=6.49e-12), neurodegenerative disease (FDR=1.90e-11), and neurotrophin signaling (FDR=2.04e-2) pathways from NA1b subtype (Figure 5.7g). In these pathway enrichment studies, we found significant downregulation of mitochondrial genes in cardiac neuronal subtypes (NA1a, NA1b, NA3), suggesting an impairment in the metabolic function during the heart failure (Figure 5.7g). Based on the DEGs and pathways enrichment analysis, our studies suggest that perturbed excitability of NA1b sympathetic neurons may enhance release of neurotransmitters that amplify chronic sympathetic signaling in heart failure (Figure 5.7h).

5.4 Discussion

The major findings of the present study are; 1) three unique subtypes of stellate ganglion neurons innervate the heart, although similar subtypes of neurons may innervate other tissue beds; 2) these three subtypes are subdivided by high NPY expression, such that neuronal subtypes with high NPY expression (i.e. NA1a subtype) are morphologically, transcriptionally, and electrophysiologically distinct cells; 3) NPY release, presumably from NA1a neurons is required for maximal cardiac sympatho-excitation; and 4) Heart failure shifts cardiac neuronal subtypes such that one subtype (NA1b) becomes dominant, and exhibits robust changes in transcriptomic architecture relative to controls. These findings reveal the subtypes of postganglionic sympathetic neurons controlling the heart and the impact of chronic cardiac disease. These findings lay the groundwork for new therapeutic approaches targeting a specific neuronal subtype in heart failure.

Stellate ganglion neurons innervate a variety of tissue beds and organs with principal neurotransmitters being cholinergic (Ach) or adrenergic (NE)[189]. However, whether SGN subtypes are restricted to particular tissue beds is not understood. In our study, we identified significant overlap between subtypes that innervate the heart (NA1a, NA1b, & NA3) and the paw (NA1a, NA1b, NA1c, and NA3), although paw neurons consisted of more subtypes than the heart. These findings suggest that multiple SGN subtypes contribute to peripheral nervous system control of

organs and tissues. These findings are consistent with those of Furlan et al[189], who used scRNAseq to explore SGNs innervating nipple- and pilo-erector muscles and found that these neurons acquired cell type specification during organogenesis via the acquisition of unique combinations of growth factor receptors, and expression of ligands by nascent target tissues.

In the present study, we identified specific subsets of cardiac-innervating neurons in stellate ganglia distinguished by NPY expression. NPY, a key co-transmitter released from sympathetic nerve endings are implicated in a diverse array of metabolic functions including temperature regulation, hunger and satiety, and cardiac adrenergic regulation[191,201–204]. NPY receptors are expressed by various tissues[205–207], and in the heart, couple with downstream signaling mechanisms that amplify cardiomyocyte calcium transients and increase diastolic calcium release from the sarcoplasmic reticulum[208–210]. Furthermore, the importance of NPY as a cardiac neurotransmitter is underscored by the recent finding that cardiac sympathetic blockade using beta-adrenergic receptor antagonists without NPY antagonists did not significantly raise ventricular fibrillation threshold, however, concomitant antagonism of both NPY and beta-receptor signaling did[202]. Additionally, it is recognized that low frequency sympathetic ganglion stimulation induced cardiac norepinephrine release, however, high frequency stimulation of such ganglia induced norepinephrine and NPY release into the myocardium from distal axonal projections from sympathetic neurons[211]. These data, interpreted in light of the new findings in this study, suggest that cardiac sympathetic regulation exists as a two-tiered system, where low-level cardiac sympathoexcitation is mediated by NPY-low or NPY-negative neurons, while high-level sympathoexcitation requires the recruitment of NPY expressing neurons.

A few findings in the present study support this notion. First, distinct cells are high or low in NPY, suggesting that when robust cardiac sympathetic activation is required for the fight or flight responses, subsets of neurons rich in NPY are recruited. Second, mice with NPY ablated neurons

in stellate ganglia are unable to reach maximal levels of cardiac chronotropic response during escalating sympathetic stimulation frequencies, unlike wildtype littermates.

In the setting of cardiac injury such as myocardial infarction or chronic heart failure, elevated circulating or coronary sinus levels of NPY are associated with adverse outcomes in patients including death, ventricular arrhythmias, and need for orthotopic heart transplantation[191]. Findings from human stellate ganglia suggest that in chronic heart failure, expression of NPY remains at levels comparable to those in humans without heart failure, however, there is increased axonal transport for distal release to maintain chronic symapthoexcitation characteristic of heart failure[191]. Strikingly, another study showed that NPY is protective and preventive during heart failure development[212].

As such, we investigated how heart failure influences the distribution of cardiac SGN subtypes, and how NPY expression is altered in NPY high and NPY low/negative subsets. Two main findings from our study provide insights into potential mechanisms driving elevated NPY levels in the setting of chronic cardiac injury. First, we find that the 3 cardiac SGN subtypes (NA1a, NA1b, and NA3) remain preserved, however, heart failure shifts the relative distribution across the 3 subtypes towards one dominant one (NA1b). Importantly, NA1b, which had low levels of NPY expression in control littermates, exhibited increased NPY expression albeit to levels lower than that seen in NA1a. Taken together, our findings suggest that the transcriptomes of cardiac SGNs generally shift to amplify the subtypes and numbers of cells that express NPY. Given the profound actions of NPY as a potentiator of adrenergic signaling, our findings provide a basis for the prognostic role of elevated circulating or coronary sinus levels of NPY in identifying patients with severe sympathoexcitation for whom outcomes are dismal. Our findings are in line with those of Davis et al[213], who performed scRNAseq in rats with spontaneous hypertension (SHR) and Wistar controls. In this study, cardiovascular SGNs were not specifically identified for characterization, however,

SGNs in the SHR model exhibited functional (increased excitability and tonic firing) and transcriptomic alterations (reduced transcript levels of genes encoding the M-current, subunits KCNQ2, KCNQ3, and KCNQ5). In our study, we found increased expression of KCNQ2 in heart failure animals, contrary to decreased expression observed in the SHR model. This suggests that although sympathetic activation is characteristics of both hypertension and chronic heart failure, the mechanisms, at the neuronal level, that drive sympathetic activation may differ fundamentally at the neuronal level.

Our findings of a dominant neuronal subtype with greater NPY expression in the setting of heart failure may have significant clinical implications. In patients with chronic heart failure or arrhythmias, systemic pharmacologic blockade of neurohormonal activation and sympathetic signaling, is a mainstay of treatment[179,195,197]. While these drugs have proven lifesaving benefits, they result in substantial side effects such as hypotension, mental incapacitation, fatigue, weakness, and sexual side effects to name a few[214]. Of note, therapies that target the stellate ganglion, rather than systemic blockade, have shown significant benefit and are increasingly used clinically. Yet, these approaches also target the entire stellate ganglion with potential off target issues to the other tissue beds and organs innervated by neurons in the stellate ganglion. Hence, more specific targeting of neurons that innervate the heart, and drive the excessive chronic sympathetic signaling characteristic of chronic cardiac injury are needed. Our study presents one such candidate, the cardiac SGN subtype NA1b, which becomes the dominant cardiac neuronal subtype in heart failure. Analysis of the transcriptome of this cell identifies potential targets that can be modulated directly, or used to identify this cell type for other forms of intervention, for example, biologic agents. Coupling such cell specific targeting with localized delivery or such agents to the stellate ganglion via minimally invasive approach (for example as used with stellate ganglion block) provides novel avenues to target sympathetic excess in chronic cardiac injury.

The present study has limitations. First, while we carefully optimized tissue-specific labeling, we cannot completely exclude non-specific labeling of a few cells in stellate ganglia. However, our IHC studies validated the subpopulations we identified by scRNAseq. Second, changes in transcript levels in DCM vs control stellate ganglia identified by scRNAseq do not connote functional changes, hence our findings imply but do not prove functional differences in the setting of heart failure. However, the extensive prior published literature showing functional changes in SGNs following cardiac injury support functional consequences of the differences in gene expression we identified in this study. Second, although we find that the NA1b subtype becomes dominant in the setting of heart failure, we postulate but do not have direct evidence that they are responsible for chronic hyperactive cardiac sympathetic tone.

5.5 Conclusions

In summary, we have identified the distinct subtypes of SGNs that innervate the heart and have shown that the subtypes fall into two groups based on expression of NPY. We present evidence that the absence of NPY expressing subpopulations in the stellate ganglia results in the inability to achieve maximal sympathoexcitation, and finally, that cardiac neuronal subtypes in the SGN are differentially impacted by heart failure, yielding a dominant cardiac SGN subtype that may represent a novel target for adrenergic blockade in heart failure. These findings expand our understanding of cardiac sympathetic control mechanisms and suggests that cell-specific targeting of sympathetic neurons may offer new therapies for heart failure.

5.6 Figures

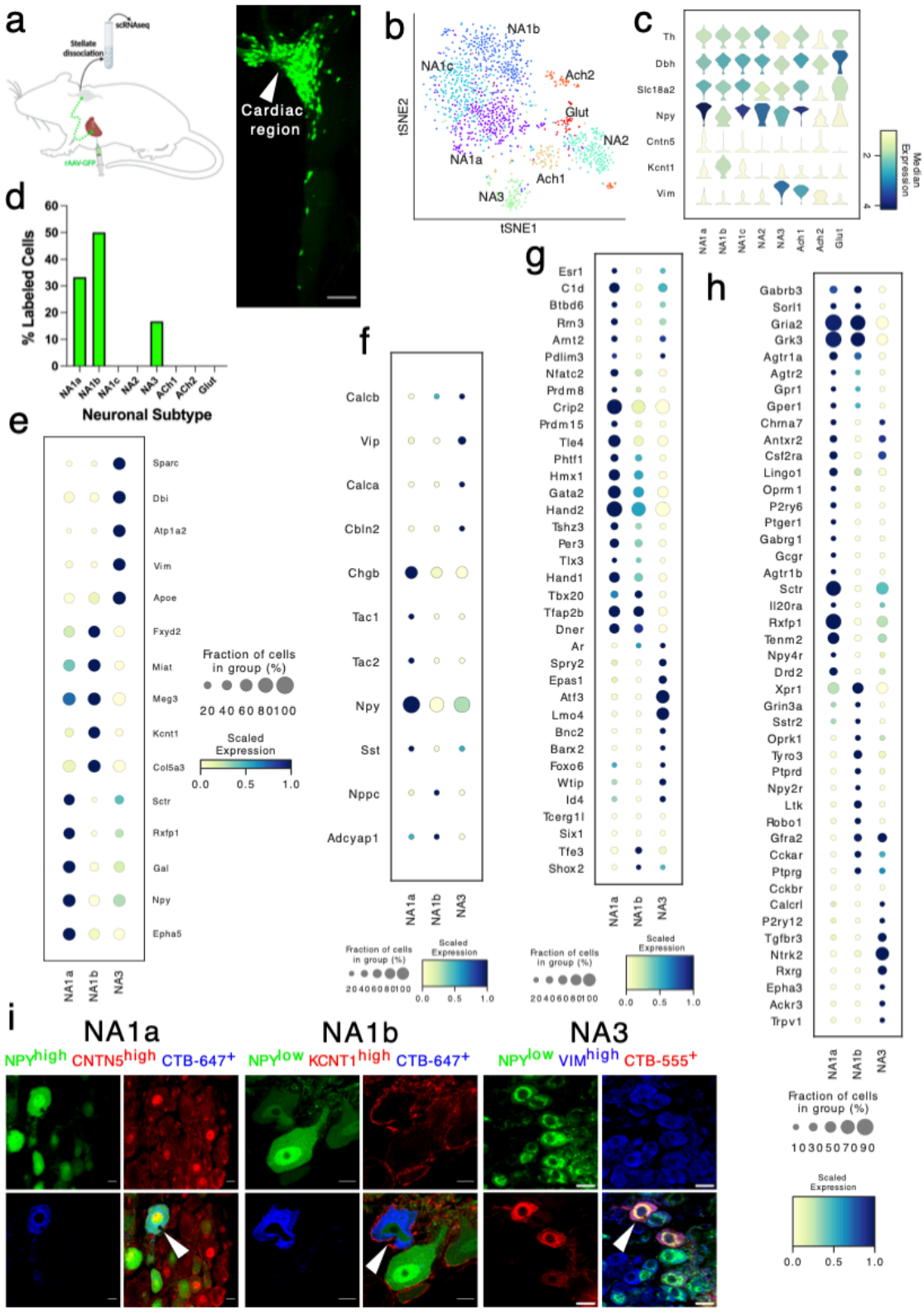

a
scRNAseq
Stellate dissociation
rAAV-GFP
Cardiac region
b
NA1b
NA1c
Ach2
Glut
NA2
NA1a
Ach1
NA3
tSNE2
tSNE1
c
Th
Dbh
Slc18a2
Npy
Cntn5
Kcnt1
Vim
Median Expression
NA1a NA1b NA1c NA2 NA3 Ach1 Ach2 Glut
d
% Labeled Cells
Neuronal Subtype
e
Sparc
Dbi
Atp1a2
Vim
Apoe
Fxyd2
Miat
Meg3
Kcnt1
Col5a3
Sctr
Rxfp1
Gal
Npy
Epha5
Fraction of cells in group (%)
20 40 60 80100
Scaled Expression
f
Calcb
Vip
Calca
Cbln2
Chgb
Tac1
Tac2
Npy
Sst
Nppc
Adcyap1
g
Esr1
C1d
Btbd6
Rrn3
Arnt2
Pdlim3
Nfatc2
Prdm8
Crip2
Prdm15
Tle4
Phtf1
Hmx1
Gata2
Hand2
Tshz3
Per3
Tlx3
Hand1
Tbx20
Tfap2b
Dner
Ar
Spry2
Epas1
Atf3
Lmo4
Bnc2
Barx2
Foxo6
Wtip
Id4
Tcerg1l
Six1
Tfe3
Shox2
h
Gabrb3
Sorl1
Gria2
Grk3
Agtr1a
Agtr2
Gpr1
Gper1
Chrna7
Antxr2
Csf2ra
Lingo1
Oprm1
P2ry6
Ptger1
Gabrg1
Gcgr
Agtr1b
Sctr
Il20ra
Rxfp1
Tenm2
Npy4r
Drd2
Xpr1
Grin3a
Sstr2
Oprk1
Tyro3
Ptprd
Npy2r
Ltk
Robo1
Gfra2
Cckar
Ptprg
Cckbr
Calcrl
P2ry12
Tgfbr3
Ntrk2
Rxrg
Epha3
Ackr3
Trpv1
i
NA1a
NA1b
NA3
NPYhigh CNTN5high CTB-647+
NPYlow KCNT1high CTB-647+
NPYlow VIMhigh CTB-555+

Figure 5.1 Distribution of cardiac specific neuronal sub-population in the mouse stellate ganglia.

a, Experimental overview, diagrammatic representation of tracing from cardiac (AAV-GFP) neurons to the stellate ganglion, followed by single cell RNA sequencing (scRNAseq) analysis (left). The CTB labeled stellate ganglion shows the regional specificity of cardiac neurons (right) in the stellate ganglion. **b**, tSNE plot visualize eight Louvain identified neuronal subpopulations with prior annotations. **c**, Violin plot of key marker genes that were used for the validation of the sympathetic cardiac neuronal subtypes. **d** Distribution of cardiac neuronal subpopulations in stellate ganglia. n = 8 mice. **e,** Dot-plot shows the transcriptomic profile in each cardiac subtype. Dot-plot shows relative expression of neuropeptides (**f**); transcription factors (**g**); and receptors (**h**) associated with subtypes. I, Immunohistochemistry validation of cardiac innervating neuronal subtypes (NA1a, NA1b, NA3). *Scale bar: 10 μm***.**

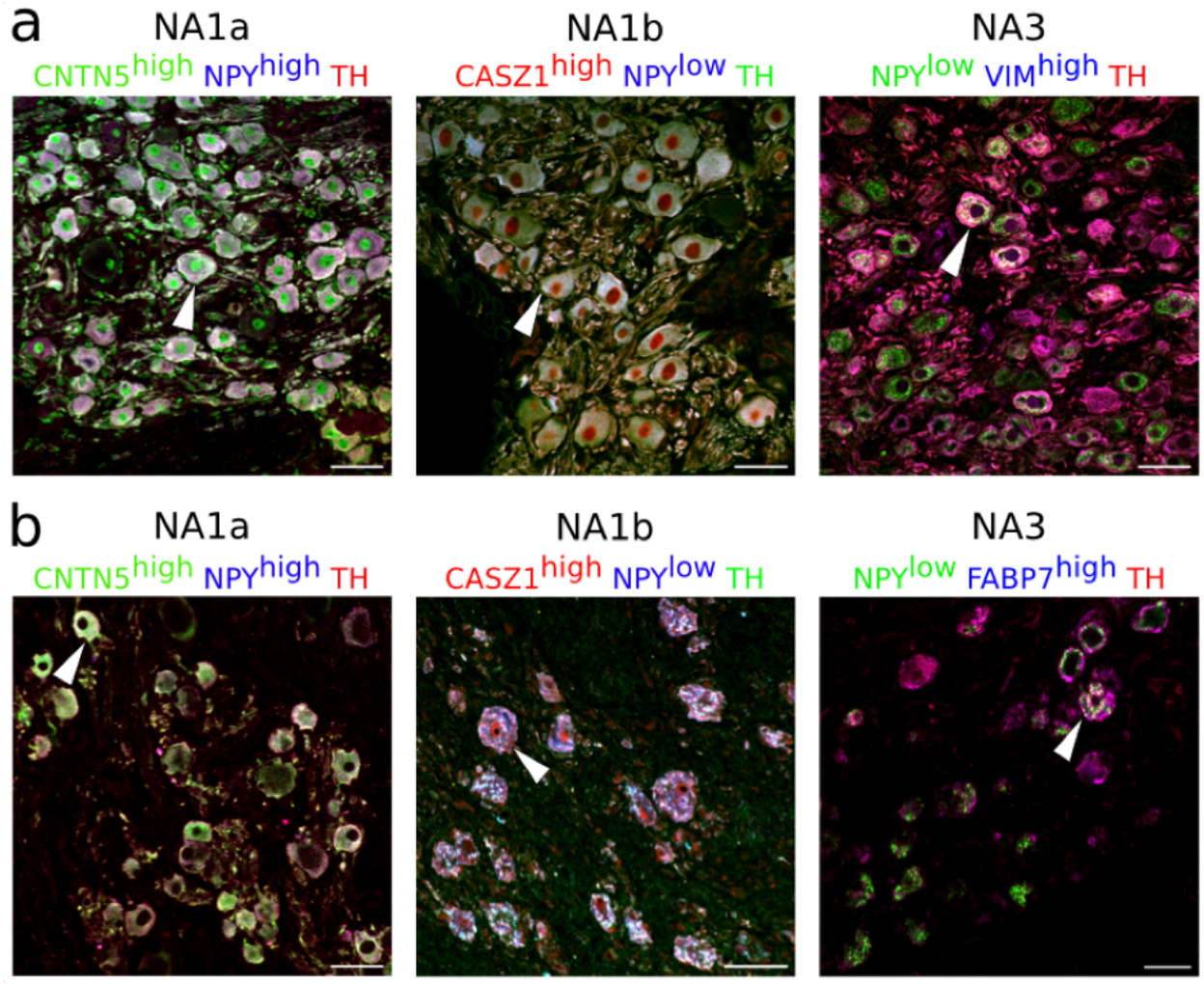


Figure 5.2 Cardiac neuronal subtypes identified in stellate ganglia

from porcine (**a**) and humans (**b**). Cardiac-specific neuronal subtypes were identified by immunostaining with antibodies against specific key marker genes. n =3. *Scale bars: 50* µ*m*.

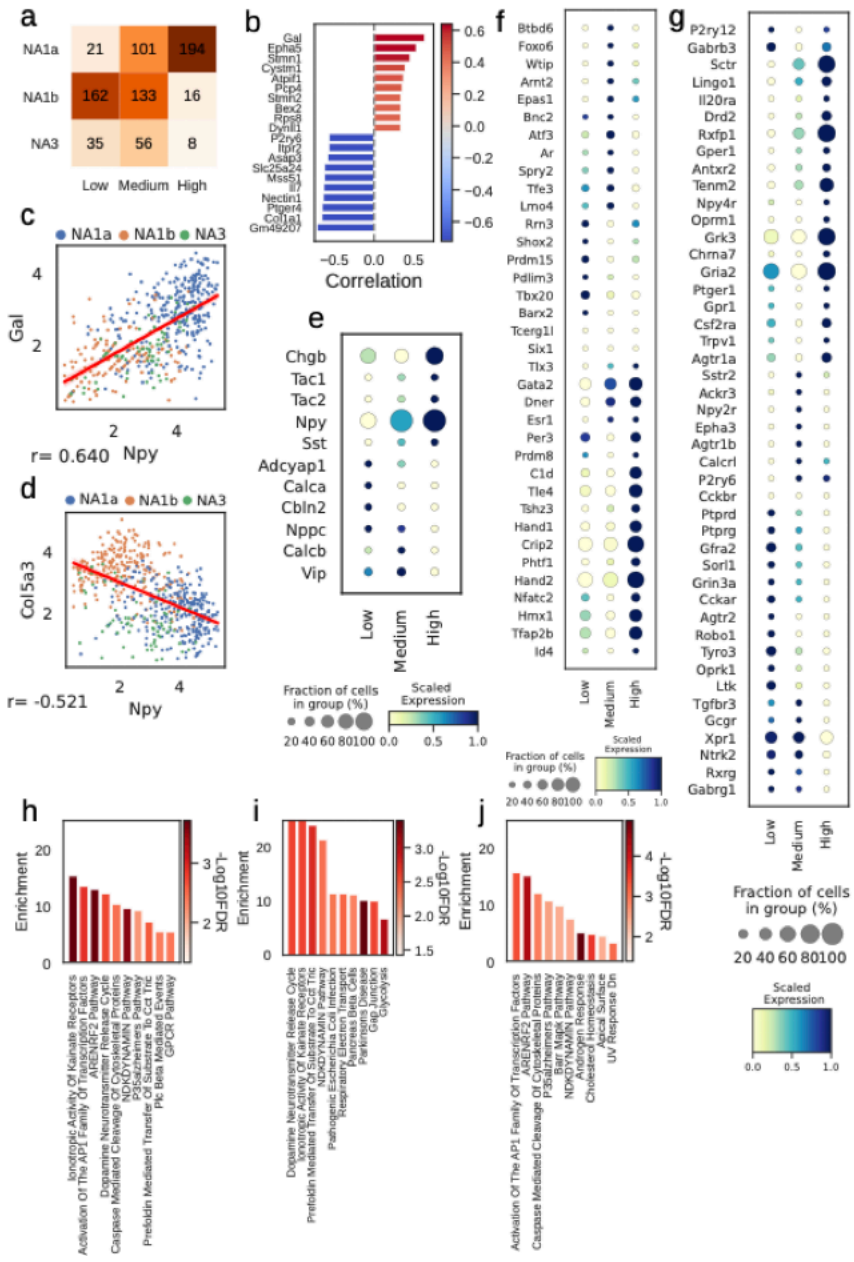


Figure 5.3 Npy Association in cardiac neuronal subtypes

Npy gene distribution and the associated genes in cardiac neuronal subtypes. **a**, Distribution of subtypes across quantiles of *Npy* expression. **B**, Top 10 positively (red) and negatively (blue) associated genes with *Npy*. Intensity of the color indicates the strength of the Pearson correlation coefficient. **c**,**d** Scatter plots showing the positive correlation of *Gal* l and, negative correlation of *Col5a3* (**d**) with *Npy* expression on the log scale. The dots in the scatter plots are colored to represent each cardiac neuronal subtypes. Cells containing zeros for either gene were removed. **e**,**f**,**g,** Dot-plots show relative expression of neuropeptide (**e)**, receptors (**f)**, and transcription factors (**g**) in cells with low, medium, and high *Npy* expression. **h**, Top 10 combined; **i**, Positively and; **j**, Negatively enriched pathways associated with *Npy*. n = 8 mice.

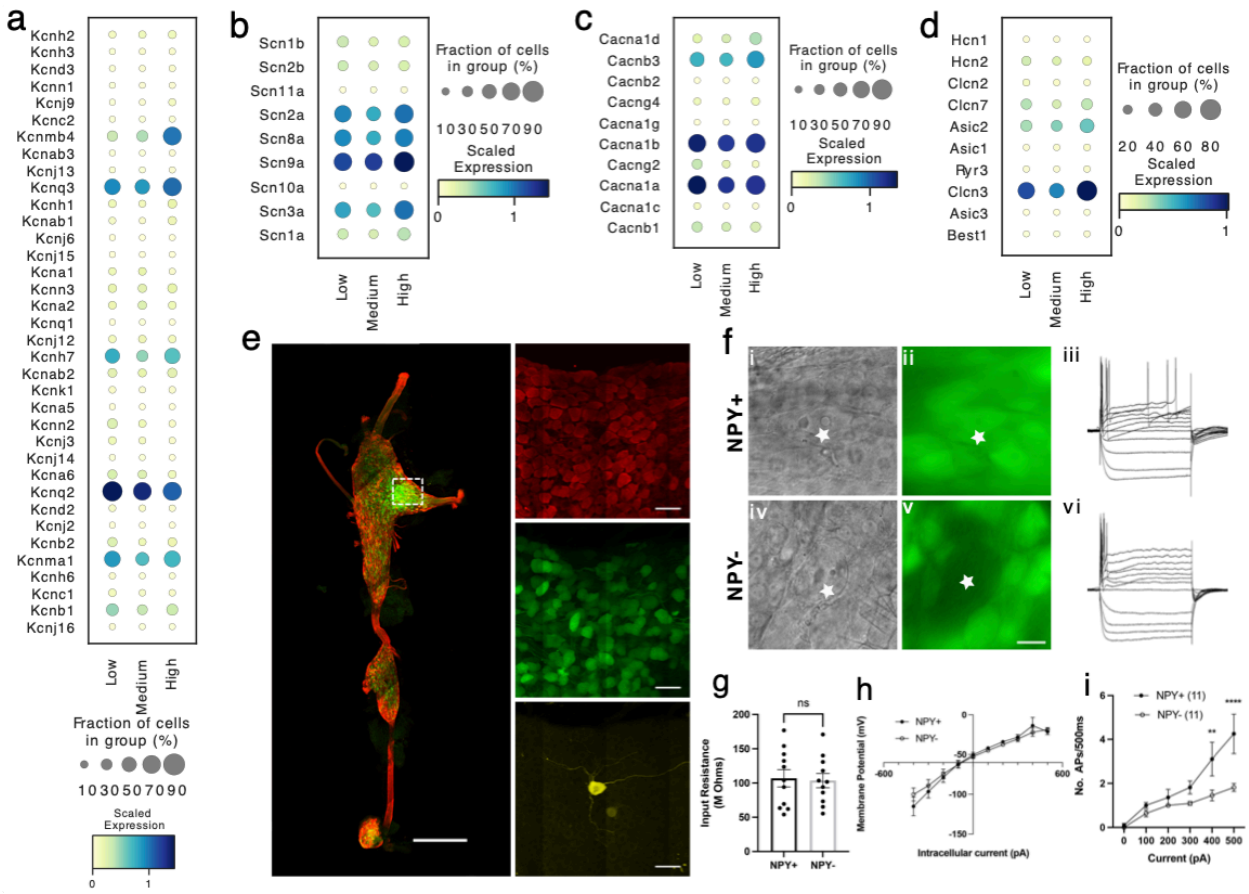


Figure 5.4 Transcriptomic and electrophysiological properties of *Npy*^{high} vs. *Npy*$^{low/neg}$ neurons.

a,**b**,**c**,**d** Distribution of different ion channels associated with *Npy*^{high} and *Npy*$^{low/neg}$ expressing cells: Potassium channels (**a**); Sodium channels (**b**); Calcium channels (**c**); and several other chloride, acid sensing ion channels (**d**) in high, medium, and low *Npy* expressing Cells. **e**, Confocal image of a fixed stellate ganglion whole-mount. PGP9.5 (red), NPY-GFP+ (green), and Neurobiotin (yellow). Magnified images from boxed region shown in right hand panels. **f**, NPY+ cells identified by epifluorescence during targeting with intracellular microelectrodes. Left panels

DIC image (**i**,**iv**), right panel GFP (green) (**ii**,**v**). **iii**,**vi;** Membrane potential tracings from a GFP+ (**iii**) and a GFP- (**vi**) cells in response to 500ms hyperpolarizing and depolarizing pulses. Resting membrane potential for both cells was more or less similar to ~ -52.45mV. Dash is 0mV. **g**,**h**,**i** Summary electrophysiology data for the NPY+ (n = 11 mice) and NPY- cells (n = 11 mice). GFP+ cells had similar input resistance (**g**), current-voltage relationship (**h**), and increased excitability (**i**). Data are shown as means ± SEM; **p<0.01, **p<0.001* compared with NPY- cells. Panel (g) was analyzed using an unpaired t test. The two-way ANOVA followed by Sidak's multiple comparison test was used for statistical analyses of panel h and i.

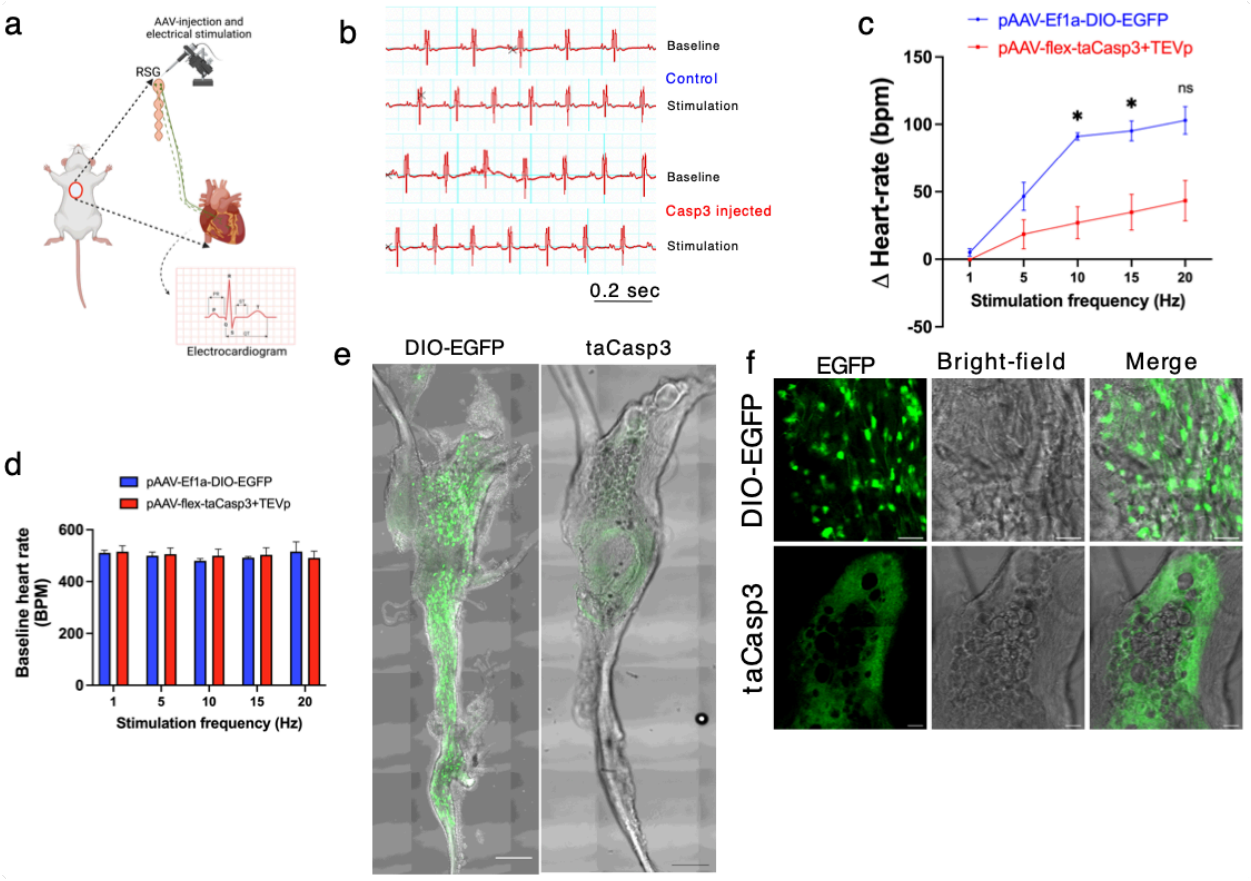


Figure 5.5 Cardiac sympathetic activation by stellate ganglion neurons requires Npy expressing subpopulations.

a, A schematic view of the chemogenetic based approach used for ablating Npy expressing neurons. **b**, The representative electrocardiogram snaps during right stellate ganglion stimulation (RSGS) from control and Npy ablated mice. **c**, Heart rate change at 1, 5, 10, 15, and 20 Hz stimulation frequencies in control and Npy ablated mice. **d**, Baseline heart rate for control and Npy ablated mice at 1, 5, 10, 15, and 20 Hz stimulation frequencies. **e**, Stellate ganglia isolated from EGFP and EGFP+taCasp3 virus injected mice. **f**, High magnification images obtained from control and taCasp3 injected stellate ganglia. Data are shown as means ± SEM. n = 3 for control mice and n = 5 for caspase-3 injected mice. **p<0.05*; *Scale bar: 200 and 50* µ*m*. The two-way

repeated measures ANOVA followed by Sidak's multiple comparison test was used for statistical analyses.

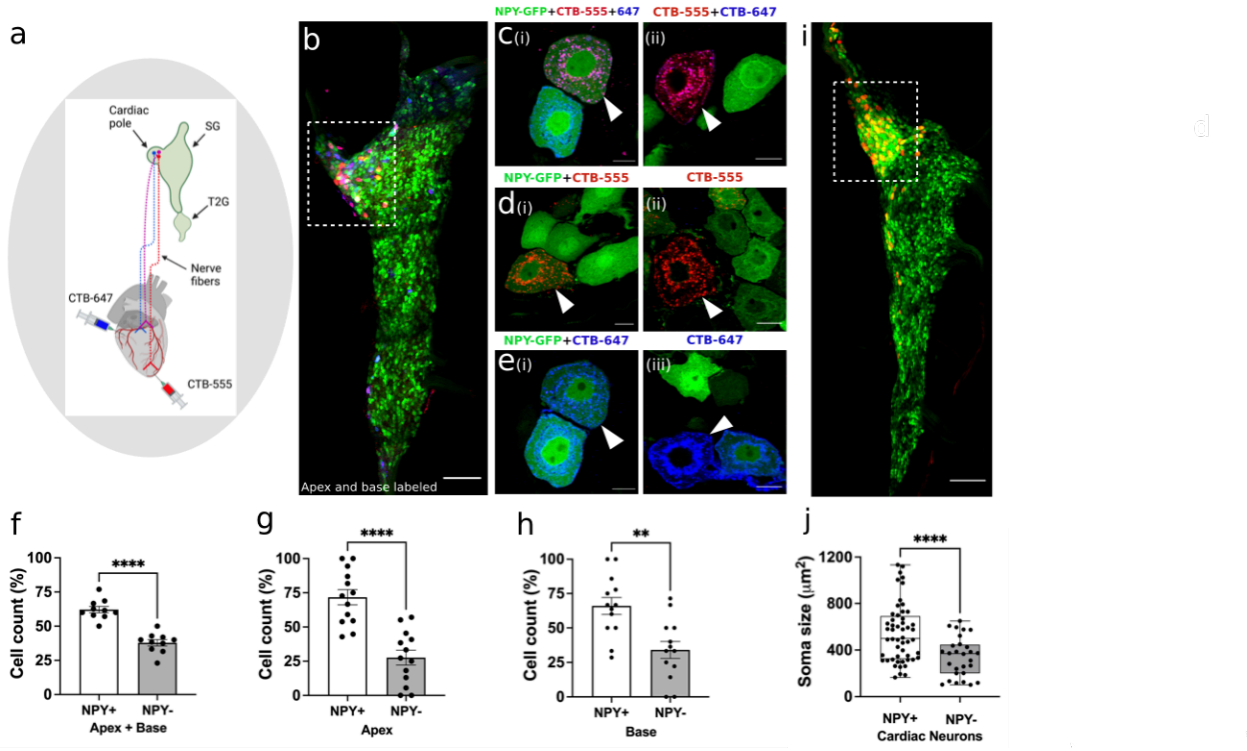


Figure 5.6 Distribution of NPY innervating neurons in disparate regions of the heart.

Cholera toxin subunit B (CTB) conjugated Alexa Fluor 555 and 647 (CTB-555 or -647) were injected into the apex and base of the heart, respectively. **a**, A schematic depiction of the approach used to retrogradely label the stellate ganglion neurons (SGNs). **b**, An image of the right stellate ganglia after CTB-555 and CTB-647 retrograde labeling from the heart. *Scale bars: 200 μm*. **c**,**d**,**e**, High-resolution images of several types of co-labeled cells appeared in the stellate ganglia. *Scale bars: 10 μm*. **f**, Cell count for NPY+ and NPY- neurons that innervate apex and base regions of the heart simultaneously. NPY+ and NPY- neurons that innervate either only the apex (**g**) or base (**h**) regions of the heart. **i**, Stellate ganglia from NPY-GFP animals labeled retrograde with CTB-555 from heart showing that majority of cardiac neurons are NPY expressing neurons. High-GFP expression is seen in the cardiac pole (box). **j**, Soma size measurements for cardiac innervating NPY+ and NPY- neurons. n = 7 mice/group. Data are shown as means±SEM and individual data points in **f**, **g**, and **h** represent each stellate ganglion. Mann-Whitney non-parametric test was used for statistical analyses for **f**, **g**, and **h**. A box and

whisker plot representing data points as individual cells was used for visualization of the trend in the soma size for cardiac NPY+ and NPY- neurons and Welch's t-test was used for finding the statistical significance in panel **j**. ** = $p < 0.01$, **** = $p < 0.001$.

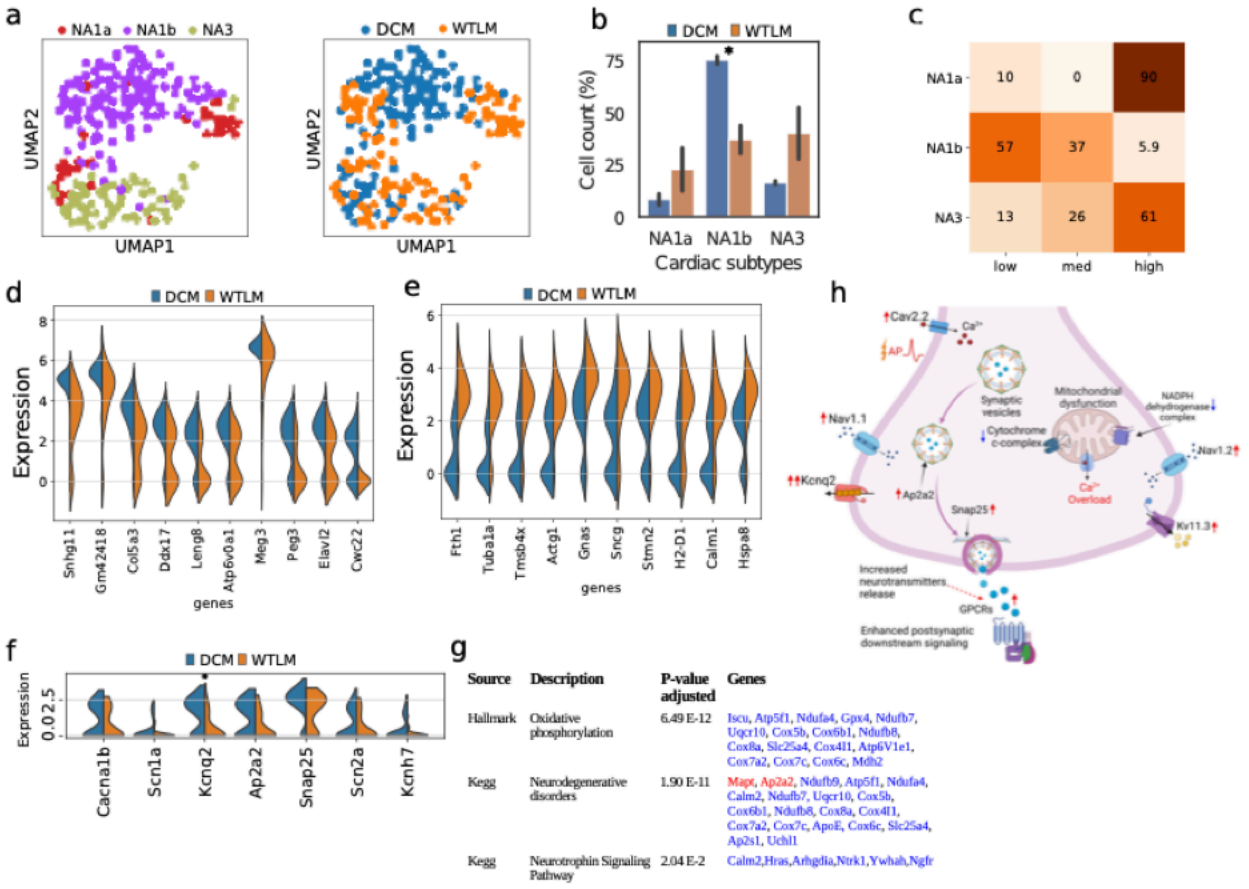


Figure 5.7 Cardiac neuronal subtypes and their transcriptomic profile in the stellate ganglia from dilated cardiomyopathy (DCM) mouse

a, UMAP plot shows the cardiac subpopulations in DCM and WT mice. In DCM mice, NA1b becomes the predominant neuronal subtypes. **b**, Proportion of neuronal subtypes cell count in DCM and WT littermates. **c**, Distribution of cardiac subtypes across the quantiles of NPY expression in DCM mice. **d**,**e** Violin plots show top ten significantly upregulated genes (**d**) and downregulated genes (**e**) in NA1b subtypes in heart failure. **f**, Violin plot indicating differential expression of ion channels and vesicular transport genes in NA1b subtype in heart failure. **g**, Relevant significantly enriched pathways in NA1b subtype. **h**, A schematic diagram of the potential mechanism underlying the increased neuronal activity in NA1b subtype, elucidated

based on the DEGs and pathway analysis in heart failure mice. Data are shown as means ±SD. * $p < 0.05$ compared to control. For each group (WT and DCM), n = 4 mice were used.

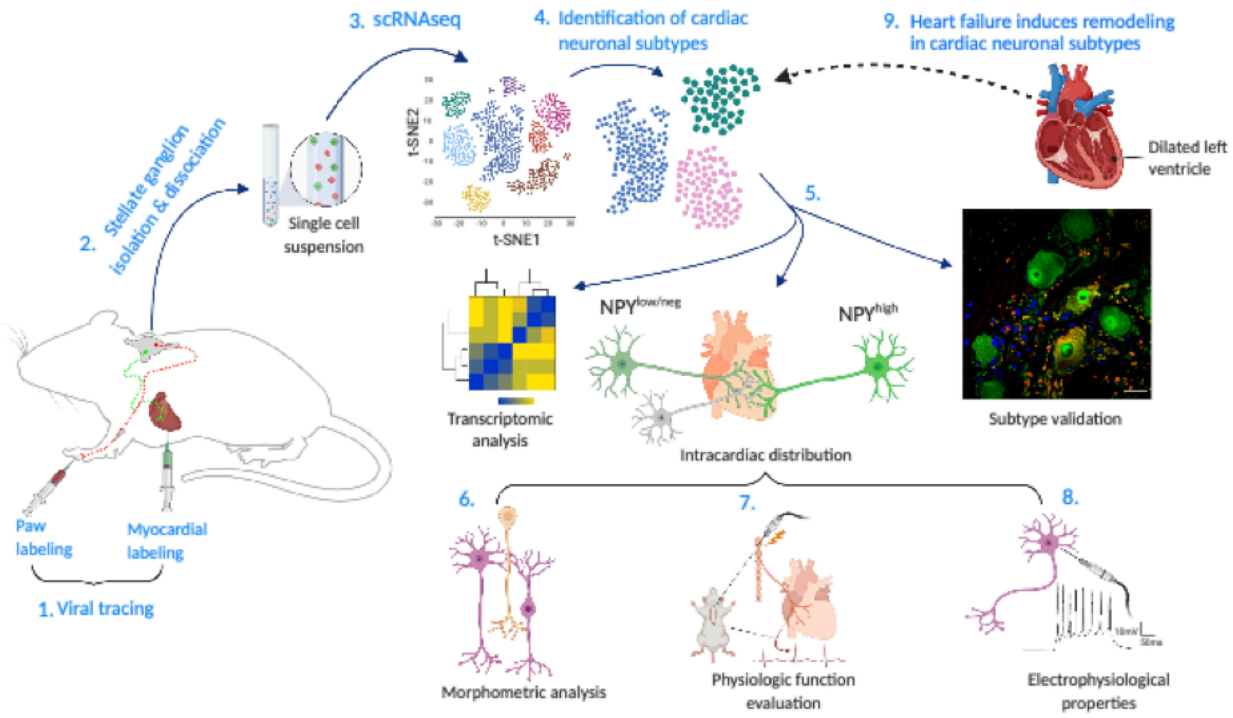


Figure 5.8 A schematic diagram of the approach taken in the current study

scRNAseq analysis was carried out for stellate ganglia tissue from normal and DCM mice. We identified and characterized the cardiac innervating neuronal subtypes either high or low *Npy* expressing cells. We identified intracardiac distribution of *Npy* expressing cells and characterized their structural and electrophysiological properties and determined the physiological function of *Npy* in relation to cardiac contractility. We also determined that heart failure causes remodeling in the cardiac-innervating neuronal subtypes.

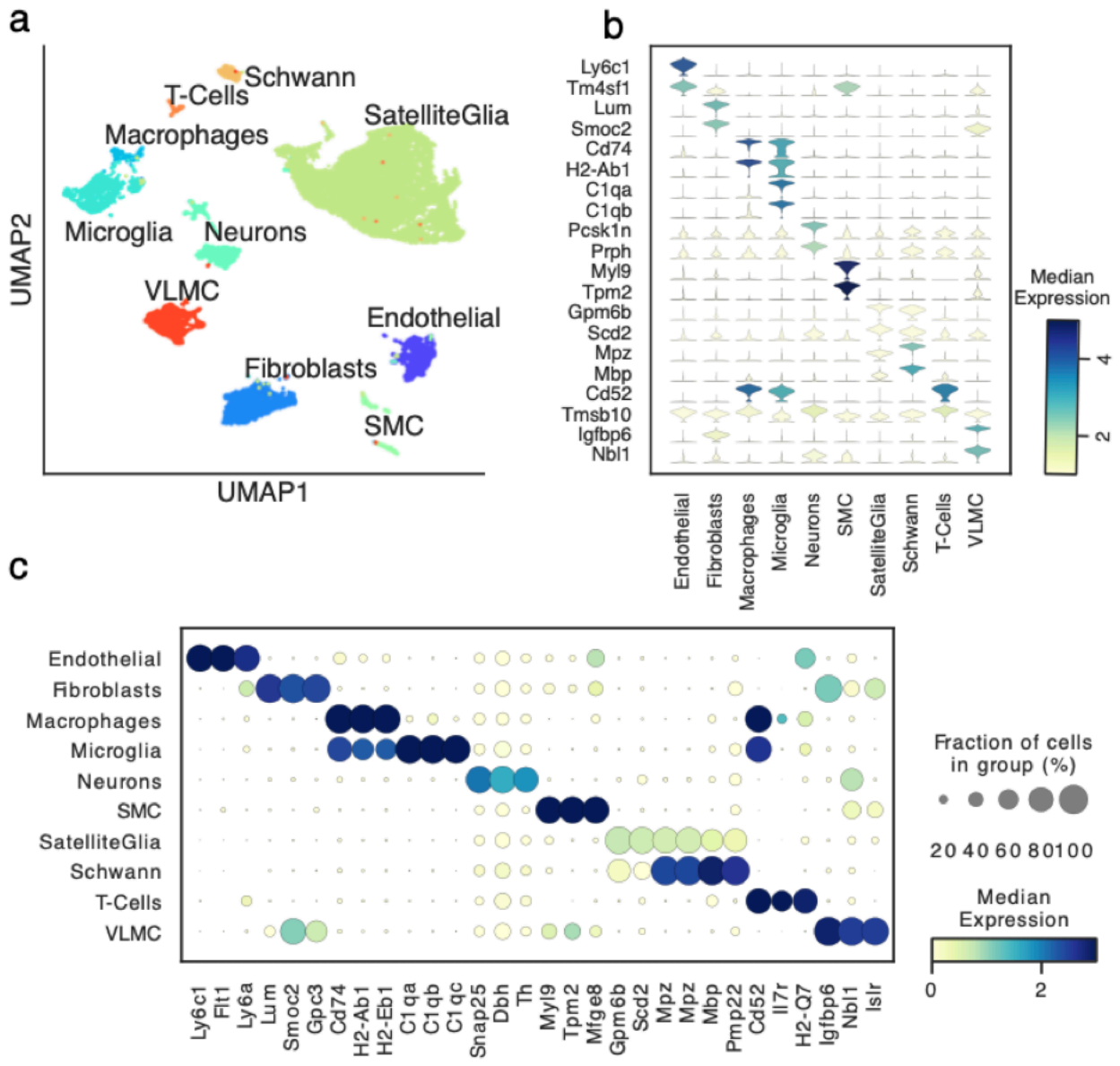


Figure SF 5.1 scRNAseq clustering of all cell types in the stellate ganglion, showing the heterogeneous populations of cells.

a, UMAP clearly shows separation between major cell types. **b**, Violin plot; **c**, Dot plot shows that different cell types are transcriptionally separable by curated and data driven cell type marker genes. n = 8 mice.

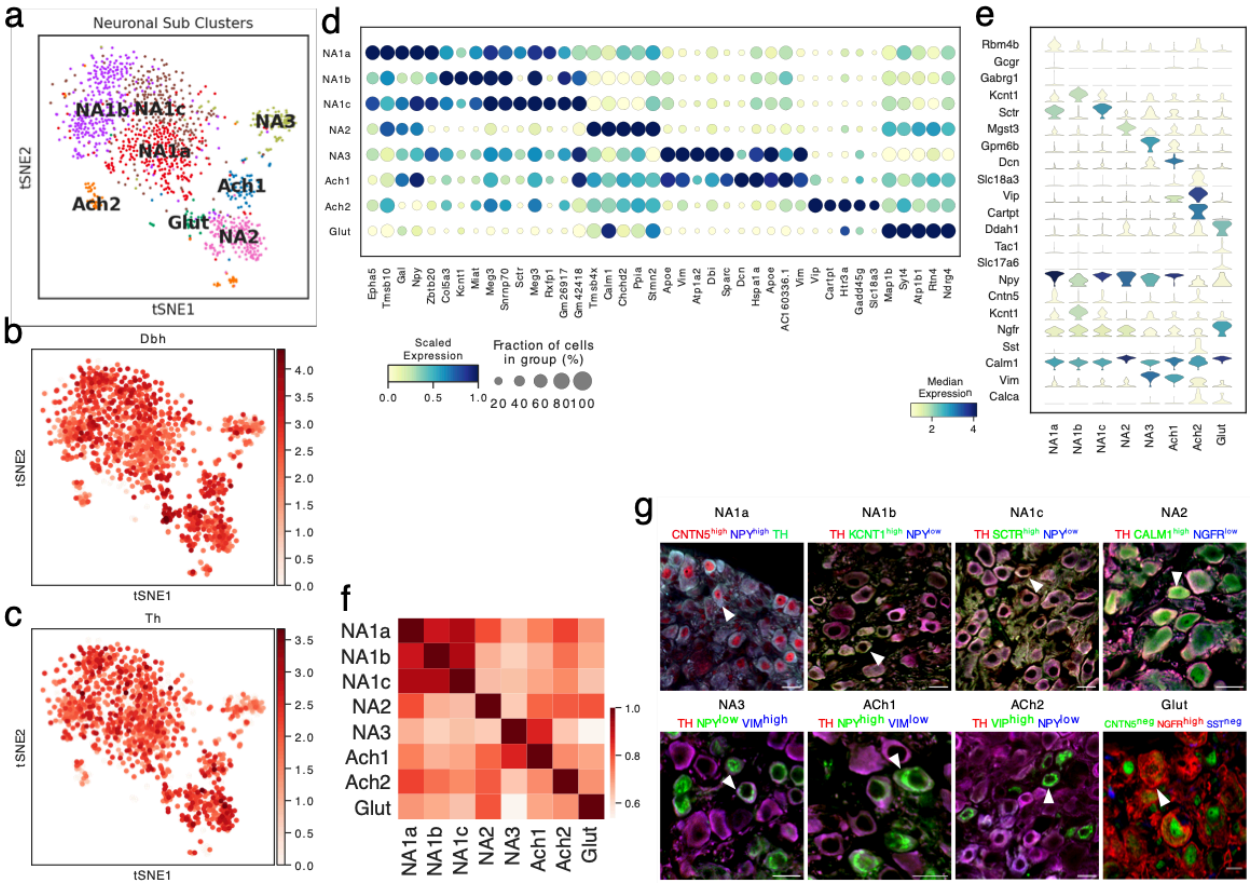
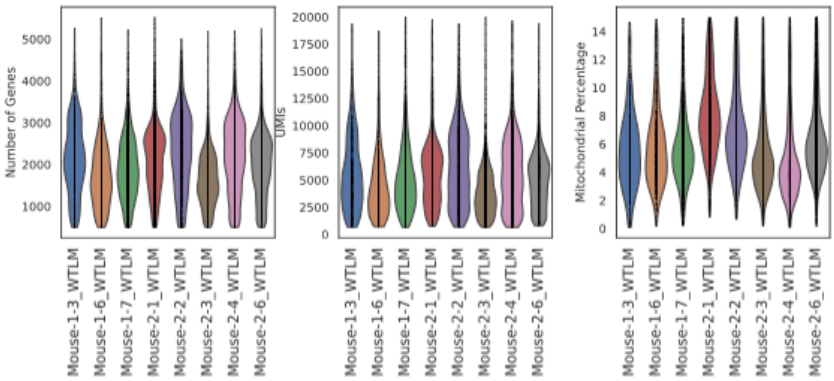


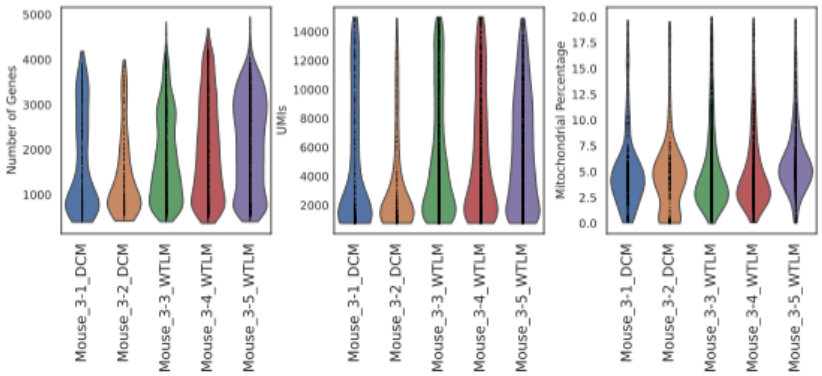
Figure SF 5.2 Neuronal subtypes in the stellate ganglion.

a, tSNE visualizes eight louvain identified neuronal subpopulations with prior annotation (n = 8 mice). **b**,**c** Feature plots containing Dbh and Th (sympathetic neuron markers) for the neuronal population on the log scale. **d**, Dot-plot shows different transcriptomic profiles of each subtype. **e**, Violin plot for the expression levels of previously known, and novel marker genes. Genes that were used to validate the clusters by immunohistochemistry were included in the plot. **f**, Pearson correlation between neuronal subtypes based on mean gene expression in each cluster. **g**, Immunohistochemistry validation of scRNAseq identified neuronal subtypes in the stellate ganglion. n = 6 mice. *Scale bar: 20 μm*.

a



b



c

Samples	Endothelial	Fibroblasts	Macrophages	Microglia	Neurons	SMC	SatelliteGlia	Schwann	T-Cells	VLMC
Mouse-1-3_WTLM	131	28	3	14	60	14	554	14	3	15
Mouse-1-6_WTLM	187	63	4	32	70	12	666	29	4	29
Mouse-1-7_WTLM	199	256	38	132	171	316	1447	8	8	116
Mouse-2-1_WTLM	136	286	31	226	170	22	1711	67	14	107
Mouse-2-2_WTLM	63	268	34	211	235	20	1581	45	7	197
Mouse-2-3_WTLM	199	552	123	419	165	50	2446	62	47	552
Mouse-2-4_WTLM	148	512	64	291	235	30	2266	65	41	312
Mouse-2-6_WTLM	233	722	49	259	197	68	2699	80	44	363

Celltypes

Samples	Endothelial	Fibroblasts	Macrophages	Microglia	Neurons	SMC	SatelliteGlia	Schwann	T-Cells	VLMC
Mouse_3-1_DCM	11	51	14	28	211	6	198	7	19	23
Mouse_3-2_DCM	3	9	4	3	235	2	57	3	0	12
Mouse_3-3_WTLM	32	40	540	743	235	9	393	44	563	4
Mouse_3-4_WTLM	25	58	224	148	147	2	508	29	439	54
Mouse_3-5_WTLM	31	78	127	95	160	16	951	60	203	20

Celltypes

Figure SF 5.3 scRNAseq quality control.

Samples are indicated with their batch, sample number, and disease status. Violin plots showing the distribution of the number of genes, UMIs, and mitochondrial percentage per cell in the first experiment with 8 WT mice **(a)** and in the second experiment with WT and DCM mice **(b)**. The number of cells in each cell type for each sample in the first experiment **(c)** and second experiment **(d)**.

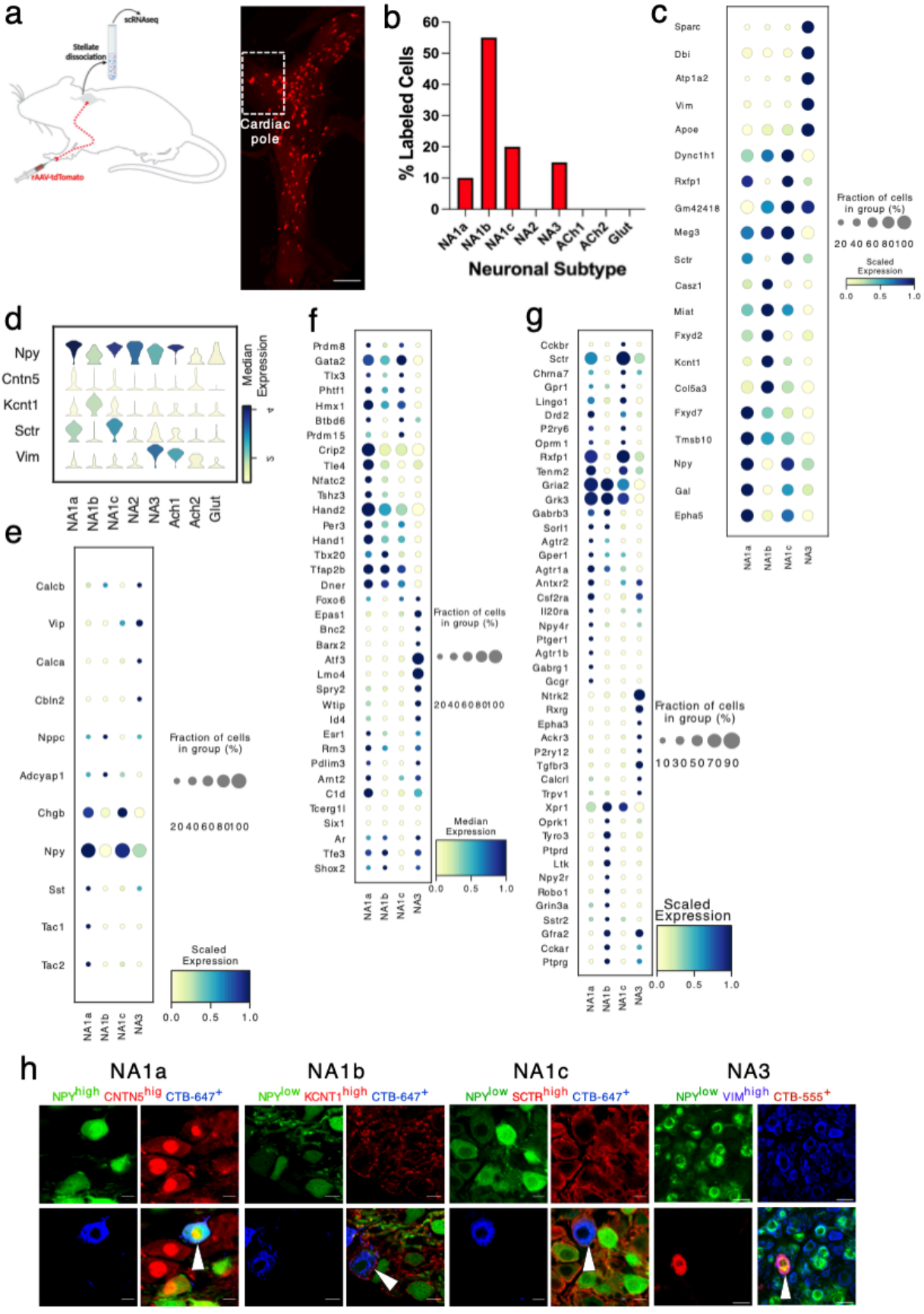

a
scRNAseq
Stellate dissociation
rAAV-tdTomato
Cardiac pole
b
% Labeled Cells
Neuronal Subtype
NA1a
NA1b
NA1c
NA2
NA3
ACh1
ACh2
Glut
c
Sparc
Dbi
Atp1a2
Vim
Apoe
Dync1h1
Rxfp1
Gm42418
Meg3
Sctr
Casz1
Miat
Fxyd2
Kcnt1
Col5a3
Fxyd7
Tmsb10
Npy
Gal
Epha5
Fraction of cells in group (%)
Scaled Expression
d
Npy
Cntn5
Kcnt1
Sctr
Vim
Median Expression
e
Calcb
Vip
Calca
Cbln2
Nppc
Adcyap1
Chgb
Npy
Sst
Tac1
Tac2
Fraction of cells in group (%)
Scaled Expression
f
Prdm8
Gata2
Tlx3
Phtf1
Hmx1
Btbd6
Prdm15
Crip2
Tle4
Nfatc2
Tshz3
Hand2
Per3
Hand1
Tbx20
Tfap2b
Dner
Foxo6
Epas1
Bnc2
Barx2
Atf3
Lmo4
Spry2
Wtip
Id4
Esr1
Rrm3
Pdlim3
Amt2
C1d
Tcerg1l
Six1
Ar
Tfe3
Shox2
Fraction of cells in group (%)
Median Expression
g
Cckbr
Sctr
Chrna7
Gpr1
Lingo1
Drd2
P2ry6
Oprm1
Rxfp1
Tenm2
Gria2
Grk3
Gabrb3
Sorl1
Agtr2
Gper1
Agtr1a
Antxr2
Csf2ra
Il20ra
Npy4r
Ptger1
Agtr1b
Gabrg1
Gcgr
Ntrk2
Rxrg
Epha3
Ackr3
P2ry12
Tgfbr3
Calcrl
Trpv1
Xpr1
Oprk1
Tyro3
Ptprd
Ltk
Npy2r
Robo1
Grin3a
Sstr2
Gfra2
Cckar
Ptprg
Fraction of cells in group (%)
Scaled Expression
h
NA1a
NA1b
NA1c
NA3
NPYhigh CNTN5hig CTB-647+
NPYlow KCNT1high CTB-647+
NPYlow SCTRhigh CTB-647+
NPYlow VIMhigh CTB-555+

Figure SF 5.4 Paw innervating neuronal subpopulation in the mouse stellate ganglion.

a, Experimental overview, diagrammatic representation of tracing the paw-specific neurons to the stellate ganglion, followed by scRNAseq analysis (left). Unlike cardiac-specific neurons, CTB labeled paw neurons did not show regional specificity in the stellate ganglia (right). **b**, Distribution of paw innervating neuronal subtypes identified and characterized using scRNAseq data analysis. (n = 8 mice). **c**, Dot-plot shows the transcriptomic profile for each paw innervating neuronal subtypes. **d**, Violin plot of the key marker gene that were used for the validation of the paw neuronal subtypes. **e**,**f**,**g** Dot-plot show relative expression of neuropeptides (**e**), receptors (**f**), and transcription factors (**g**) associated with paw neuronal subtypes. **h**, Immunohistochemistry validation of paw innervating neuronal subtypes (NA1a, NA1b, NA1c, NA3). n = 6. *Scale bar: 10 μm*.

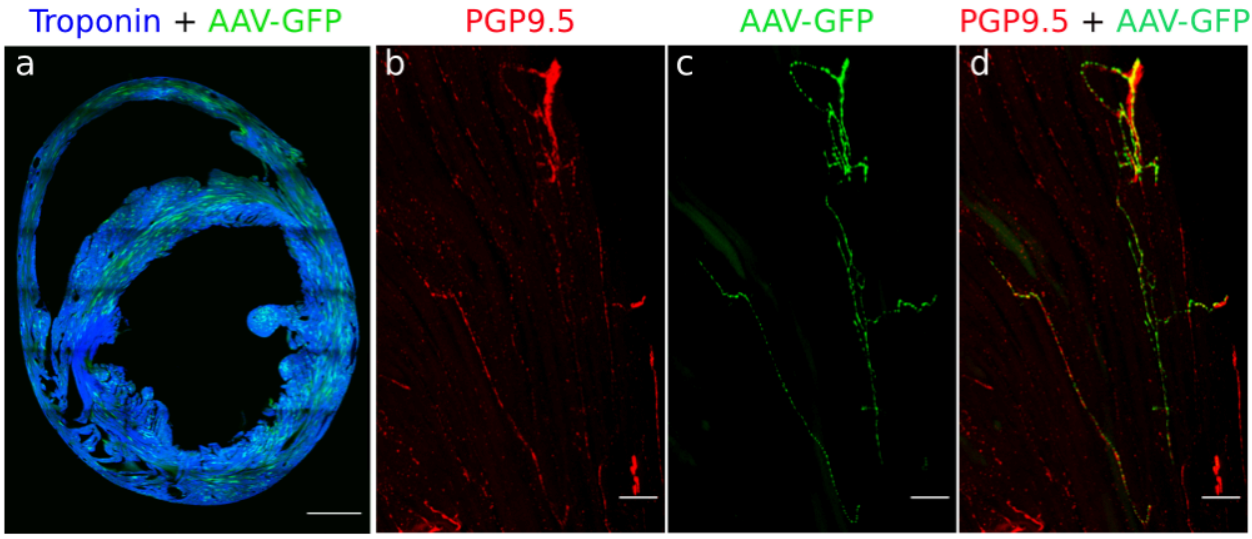


Figure SF 5.5 Validation of retrograde tracing of AAVs from myocardium (Heart injections). **a**, Cross-section of the heart is labeled with AAV-GFP (green) and cardiac troponin I (blue). *Scale bar: 1000 µm*. **b**, PGP9.5 staining in the nerve fibers of the heart section. **c**, AAV-GFP labeling in the nerve fibers of the heart section. **d**, Merge image shows the colabeling of PGP9.5 and AAV-GFP in nerve fibers of the heart-section that travels to the stellate ganglia. n = 3 mice. *Scale bar: 50 µm*

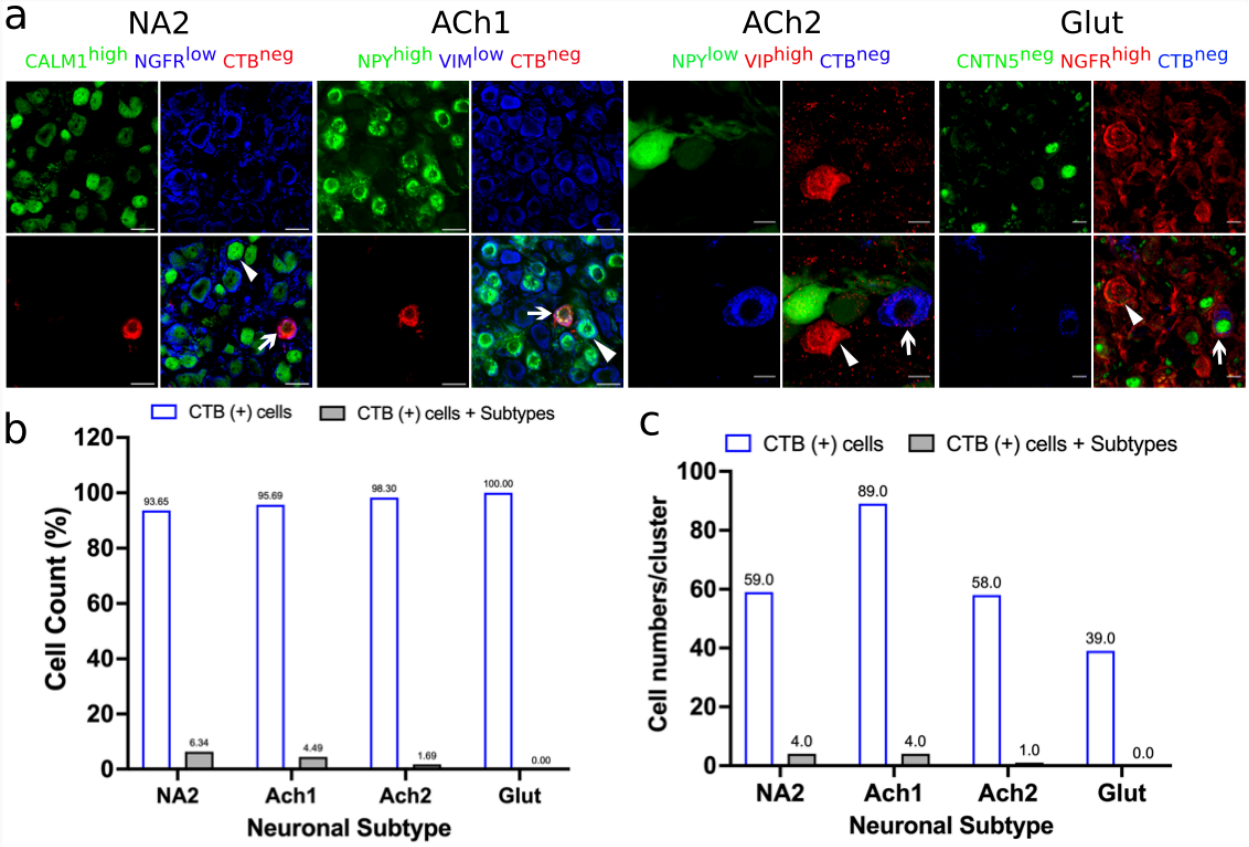


Figure SF 5.6 Confirming the accuracy of cardiac and paw innervating neuronal subtypes in stellate ganglia identified using scRNAseq analysis.

Stellate ganglia were labeled retrogradely using either CTB-647 or CTB-555 from paw, and stained with specific key markers for paw non-innervating neuronal subtypes. **a**, Immunohistochemistry validation of cardiac and paw non-innervating neuronal subtypes. **b**, Percent cell count for cardiac and paw non-innervating neuronal subtypes. **c**, Total cell count for cardiac and paw non-innervating subtypes. Note that only CTB labeled cells from paw were significantly higher compared to CTB colabeled cells with cardiac and paw non-innervating neuronal subtypes. n = 3 mice. *Scale bar: 10 and 20 μm*.

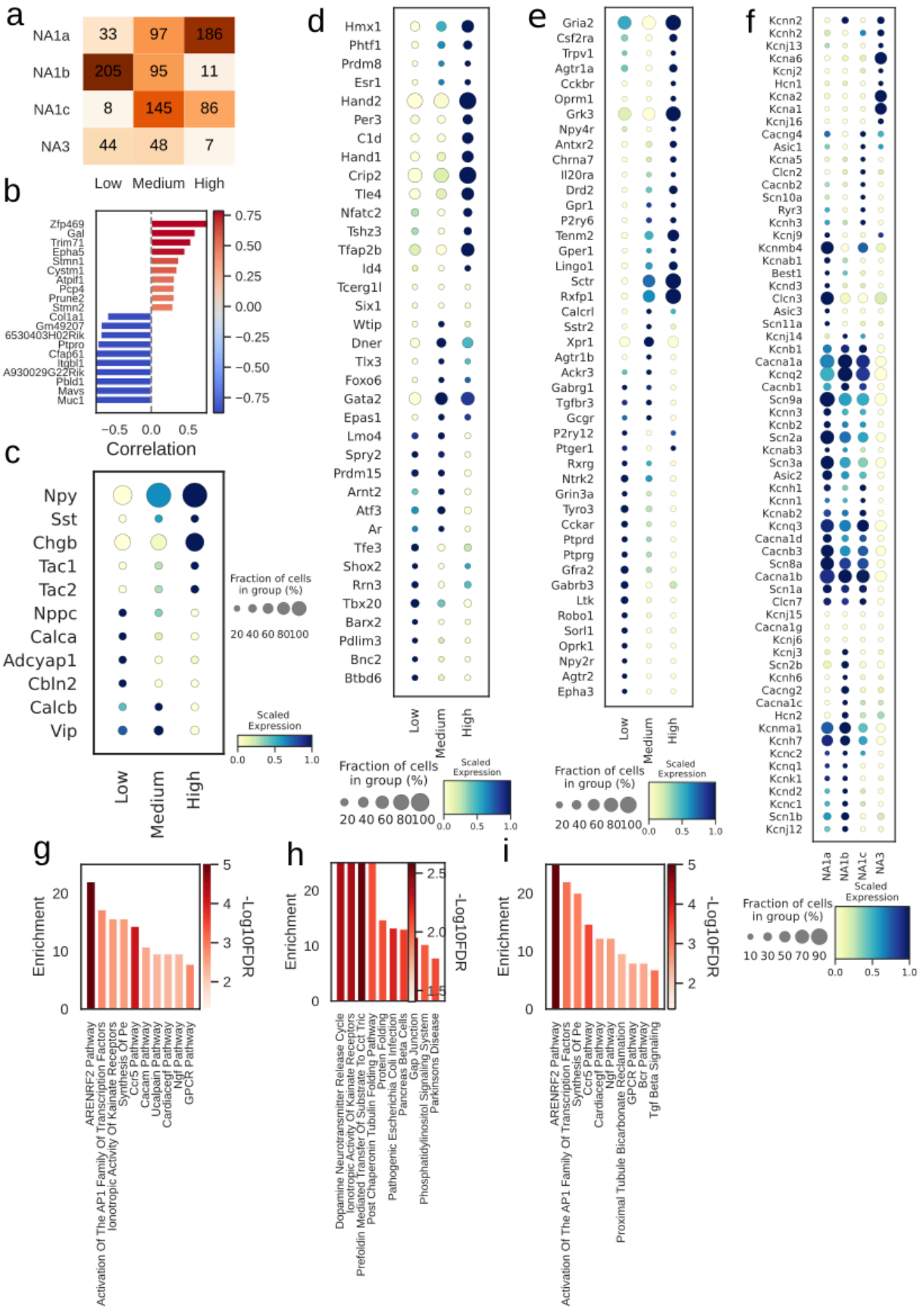

a
NA1a
NA1b
NA1c
NA3
33
97
186
205
95
11
8
145
86
44
48
7
Low
Medium
High
b
Zfp469
Gal
Trim71
Epha5
Stmn1
Cystm1
Atpif1
Pcp4
Prune2
Stmn2
Col1a1
Gm49207
6530403H02Rik
Ptprb
Cfap61
Itgbl1
A930029G22Rik
Pbld1
Mavs
Muc1
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−0.5
0.0
0.5
Correlation
c
Npy
Sst
Chgb
Tac1
Tac2
Nppc
Calca
Adcyap1
Cbln2
Calcb
Vip
Fraction of cells in group (%)
20 40 60 80100
Scaled Expression
0.0
0.5
1.0
Low
Medium
High
d
Hmx1
Phtf1
Prdm8
Esr1
Hand2
Per3
C1d
Hand1
Crip2
Tle4
Nfatc2
Tshz3
Tfap2b
Id4
Tcerg1l
Six1
Wtip
Dner
Tlx3
Foxo6
Gata2
Epas1
Lmo4
Spry2
Prdm15
Arnt2
Atf3
Ar
Tfe3
Shox2
Rrn3
Tbx20
Barx2
Pdlim3
Bnc2
Btbd6
Low
Medium
High
Fraction of cells in group (%)
20 40 60 80100
Scaled Expression
0.0
0.5
1.0
e
Gria2
Csf2ra
Trpv1
Agtr1a
Cckbr
Oprm1
Grk3
Npy4r
Antxr2
Chrna7
Il20ra
Drd2
Gpr1
P2ry6
Tenm2
Gper1
Lingo1
Sctr
Rxfp1
Calcrl
Sstr2
Xpr1
Agtr1b
Ackr3
Gabrg1
Tgfbr3
Gcgr
P2ry12
Ptger1
Rxrg
Ntrk2
Grin3a
Tyro3
Cckar
Ptprd
Ptprg
Gfra2
Gabrb3
Ltk
Robo1
Sorl1
Oprk1
Npy2r
Agtr2
Epha3
Low
Medium
High
Fraction of cells in group (%)
20 40 60 80100
Scaled Expression
0.0
0.5
1.0
f
Kcnn2
Kcnh2
Kcnj13
Kcna6
Kcnj2
Hcn1
Kcna2
Kcna1
Kcnj16
Cacng4
Asic1
Kcna5
Clcn2
Cacnb2
Scn10a
Ryr3
Kcnh3
Kcnj9
Kcnmb4
Kcnab1
Best1
Kcnd3
Clcn3
Asic3
Scn11a
Kcnj14
Kcnb1
Cacna1a
Kcnq2
Cacnb1
Scn9a
Kcnn3
Kcnb2
Scn2a
Kcnab3
Scn3a
Asic2
Kcnh1
Kcnn1
Kcnab2
Kcnq3
Cacna1d
Cacnb3
Scn8a
Cacna1b
Scn1a
Clcn7
Kcnj15
Cacna1g
Kcnj6
Kcnj3
Scn2b
Kcnh6
Cacng2
Cacna1c
Hcn2
Kcnma1
Kcnh7
Kcnc2
Kcnq1
Kcnk1
Kcnd2
Kcnc1
Scn1b
Kcnj12
NA1a
NA1b
NA1c
NA3
Fraction of cells in group (%)
10 30 50 70 90
Scaled Expression
0.0
0.5
1.0
g
Enrichment
0
10
20
-Log10FDR
2
3
4
5
ARENRF2 Pathway
Activation Of The AP1 Family Of Transcription Factors
Ionotropic Activity Of Kainate Receptors
Synthesis Of Pe
Ccr5 Pathway
Cacam Pathway
Ucalpain Pathway
Cardiacegf Pathway
Ngf Pathway
GPCR Pathway
h
Enrichment
0
10
20
-Log10FDR
1.5
2.0
2.5
Dopamine Neurotransmitter Release Cycle
Ionotropic Activity Of Kainate Receptors
Prefoldin Mediated Transfer Of Substrate To Cct Tric
Post Chaperonin Tubulin Folding Pathway
Protein Folding
Pathogenic Escherichia Coli Infection
Pancreas Beta Cells
Gap Junction
Phosphatidylinositol Signaling System
Parkinsons Disease
i
Enrichment
0
10
20
-Log10FDR
2
3
4
5
ARENRF2 Pathway
Activation Of The AP1 Family Of Transcription Factors
Synthesis Of Pe
Ccr5 Pathway
Cardiacegf Pathway
Ngf Pathway
Proximal Tubule Bicarbonate Reclamation
GPCR Pathway
Bcr Pathway
Tgf Beta Signaling

Figure SF 5.7 Npy distribution and associated genes in paw innervating neuronal subtypes.

a, Distribution of subtypes across quantiles of Npy expression. b, Top 10 positively (red) and negatively (blue) associated genes with Npy. Intensity of the color indicates the strength of the Pearson correlation coefficient. c, Relative expression of neuropeptide; d, transcription factors; e, receptors and; f, ion channels in cells with low, medium, and high Npy expression. Top 10 enriched pathways for the: g, combined; h, Positive; and i, negative associated genes with Npy.

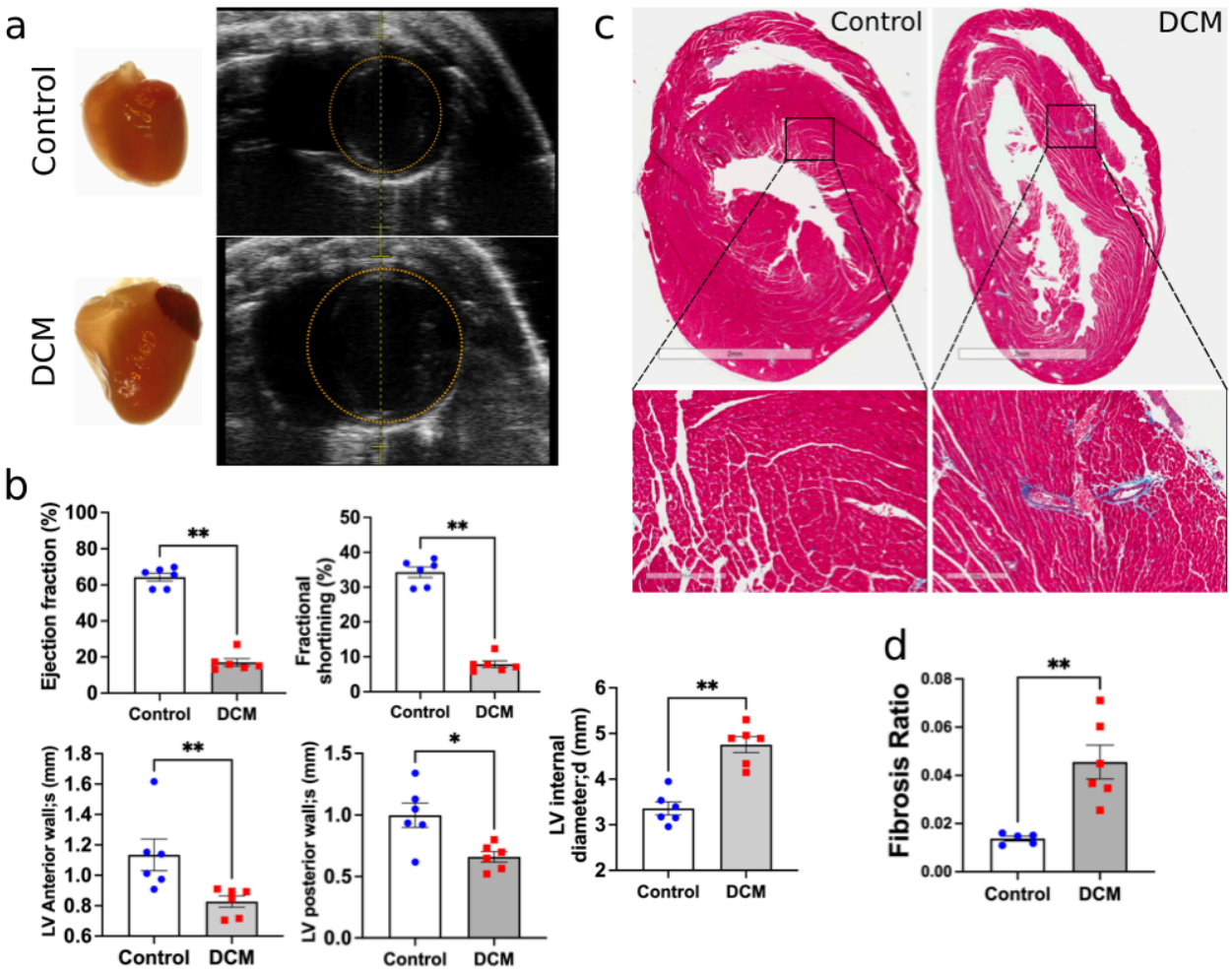


Figure SF 5.8 Echocardiography and fibrosis analysis in dilated cardiomyopathy mice (DCM).

a, DCM hearts grow bigger in size due to increased left ventricular (LV) chamber size. **b**, Ejection fraction, fractional shortening, LV anterior wall, LV posterior wall, and LV internal diameter in wildtype (control) and DCM mice. **c**, Masson's trichrome staining in heart sections from control and DCM mice. **d**, Fibrosis analysis in control and DCM mice. Data are shown as means ± SEM. For each group n = 6 mice were used. * $p < 0.05$, ** $p < 0.01$. Mann-Whitney test was used for the statistical analyses.

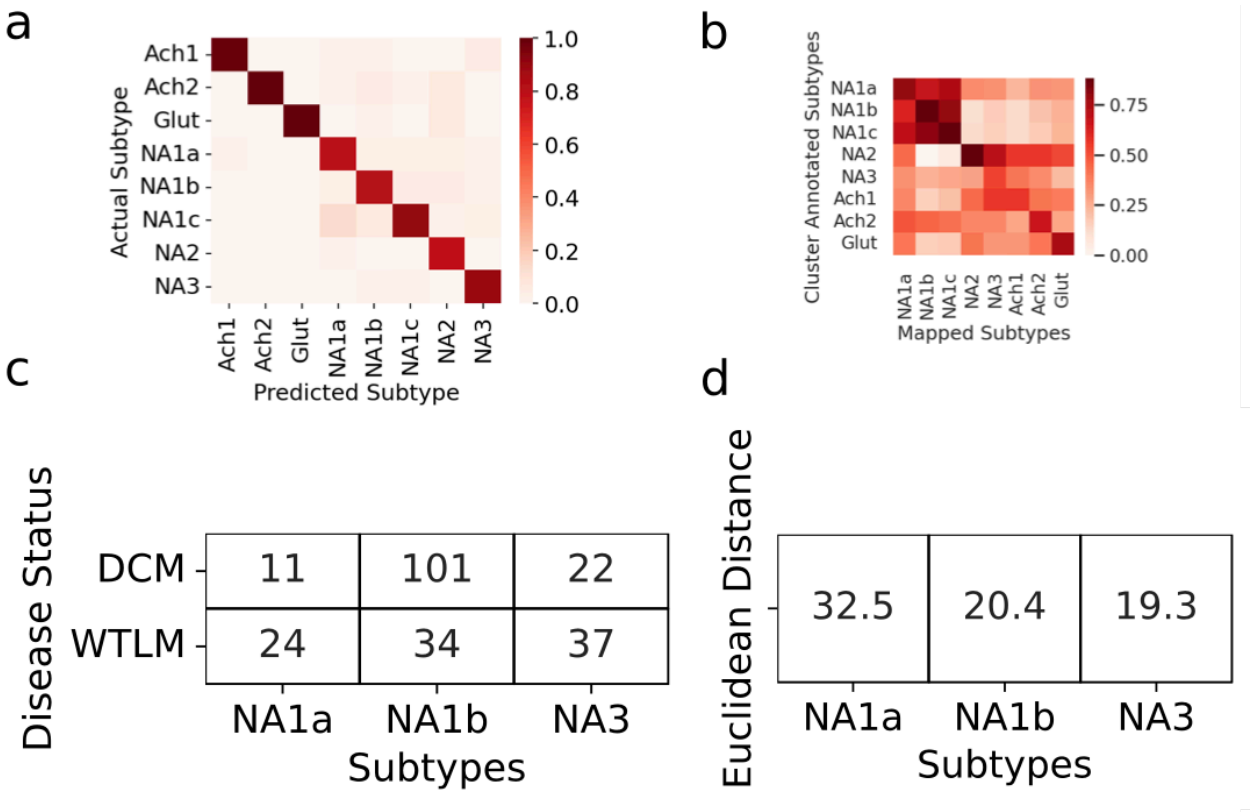


Figure SF 5.9 Neuronal subtype prediction accuracy in scRNAseq

a, Confusion matrix showing the percentage of SVM predicted subtypes that are correctly classified (diagonal) and incorrectly classified (off-diagonal). The percentage is separate for each predicted subtype. **b**, Pearson correlation between the average gene expression in the subtypes of the DCM vs WT data set, and the subtypes in the WT dataset used to define the neuronal subtypes. **c**, Tables showing the number of cardiac subtypes in DCM and WT mice, and; **d**, Euclidean distance based on the average gene expression of DCM vs WT mice in each cardiac subtypes.

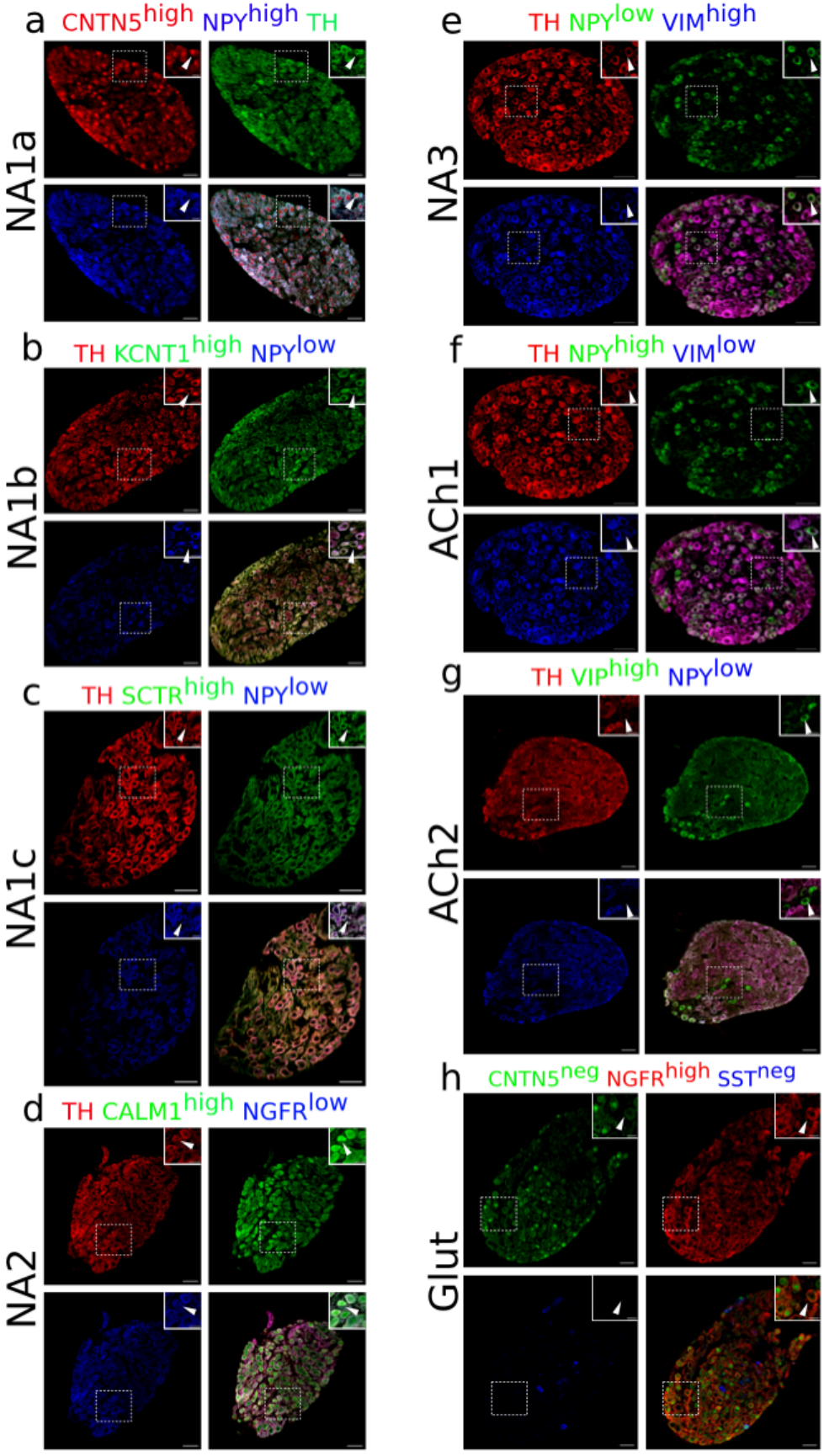

a CNTN5high NPYhigh TH
NA1a
b TH KCNT1high NPYlow
NA1b
c TH SCTRhigh NPYlow
NA1c
d TH CALM1high NGFRlow
NA2
e TH NPYlow VIMhigh
NA3
f TH NPYhigh VIMlow
ACh1
g TH VIPhigh NPYlow
ACh2
h CNTN5neg NGFRhigh SSTneg
Glut

Figure SF 5.10 Immunohistochemistry validation of the neuronal subtypes in the stellate ganglion identified from scRNAseq data analysis.

Except NA2 and Glut, all other neuronal subtypes express either high *Npy* or low *Npy*. NA1a, NA1b, NA1c neurons appeared more in numbers compared to the NA2, NA3, ACh1, ACh2, and Glut. **a**) NA1a neuronal subtypes in the stellate ganglia. NA1a neurons express high *Cntn5* and *Npy* (Inset), and can be easily distinguished from the low Npy expressing neurons colabeled with $CNTN5^{high}$. **b**) Stellate section contain neurons that stained for *Kcnt1* and express either low or high *Npy* but NA1b subtype can be distinguished by low *Npy* and high *Kcnt1* expression (inset). **c**) NA1c subtype similar to NA1b contain low *Npy* but higher SCTR labeled on the cell borders (inset). **d**). NA2 neurons are differentiated from the Glut with expression of high *Calm1* and low *Ngfr* while Glut neuronal subtype express low *Calm1* and high *Ngfr*. **e**) NA3 subtype can be separated from ACh1 based on low *Npy* and high *Vimentin* gene expression whereas ACh1 **f;** express high *Npy* and low *Vimentin*. **g**) ACh2 subtype was separated from others with low *Npy* and high *VIP* expression. **h**) Gluts expressed high *Ngfr* expression and are negative for *Cntn5* and *Sst*. In insets, all subtypes are marked for each group and their magnified view is shown in the upper right corner of the figure. n = 6. Scale bars *50 and 20 µm*.

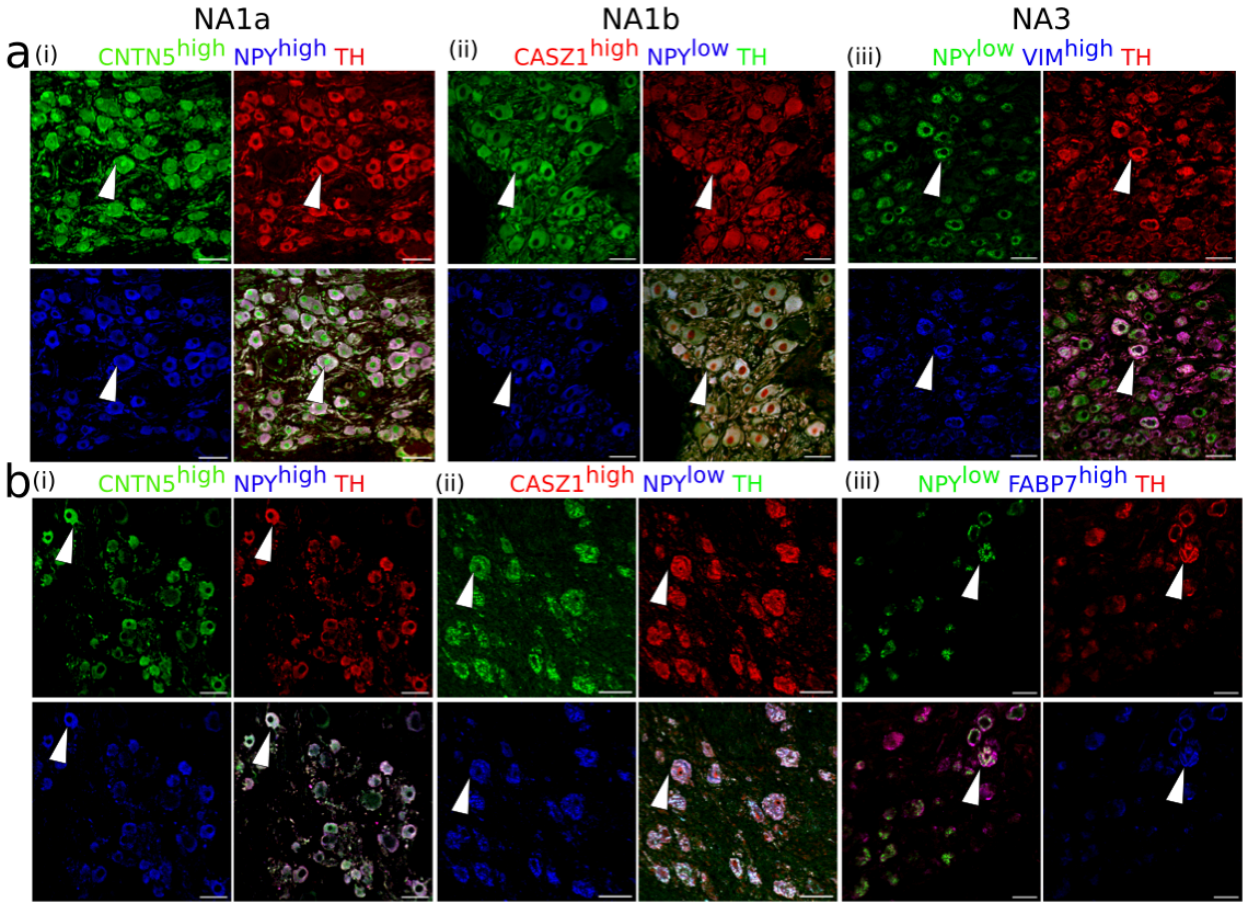


Figure SF 5.11 Cardiac cluster validation in porcine and human stellate ganglia

Cardiac clusters were also validated in porcine (a) and human (b) stellate ganglia using immunohistochemistry by combining the specific marker genes for each neuronal subtype. **a,b**(**i**) NA1a neurons express high *NPY* and *CNTN5*. **a,b**(**ii**) NA1b subtype have low *NPY* and high *CASZ1* expression level. **a**(**iii**) NA3 similar to NA1b has low *NPY* but unique higher *Vimentin* expression in porcine tissues. **b**(**iii**) NA3 like NA1b express low NPY and high FABP7 expression in humans. *Scale bar: 50 µm*. n = 3 samples.

Chapter 6. Single-cell RNA sequencing reveals molecular heterogeneity of glia within mouse sympathetic ganglia

6.1 Introduction

Glia are non-neuronal cells that are present throughout the entire nervous system. In the central nervous system (CNS), glial cells have been demonstrated to play important roles in both health and disease[215]. Astrocytes, the most abundant glial type in the CNS, are of particular importance in the development and functioning of the CNS, but can become pathologically activated in response to injury[215,216]. Even though astrocytes are widely studied, much less is known about their peripheral counterpart, satellite glial cells (SGCs). These cells are found in sensory, parasympathetic, and sympathetic ganglia of the peripheral nervous system (PNS), where they envelope the neurons[217]. SGCs and neurons have been demonstrated to reciprocally modulate each other's activity *via* gap junctions as well as through the release of signaling molecules, such as adenosine triphosphate (ATP) and gamma aminobutyric acid (GABA)[217–220]. Furthermore, SGCs have been described to become reactive in response to neuronal injury, which is reflected by an increase in size[221] and number[222] and is accompanied by changes in modulatory activity[223,224]. This reactivity is often objectified by the upregulation of glial fibrillary acidic protein (GFAP)[225,226]. Despite this clear evidence that SGCs play a dynamic and important role in the PNS, studies into SGC biology are limited. Furthermore, most studies of SGCs have focused on their role in the sensory nervous system.

Nevertheless, it has been shown that SGCs in the sympathetic ganglia are able to modulate efferent sympathetic cardiac outflow[227,228]. Moreover, Ajijola *et al.* (2017) demonstrated that GFAP was upregulated in the stellate ganglia SGCs of humans with heart failure (HF) and arrhythmias. As sympathetic outflow from the stellate ganglia is known to be increased in HF, these observations indicated a potential role for SGC activation in the development and/or progression of HF.

As much remains unknown about the molecular background of SGCs and their role in health and disease, we used single-cell RNA sequencing (scRNAseq) to study SGCs from healthy adult murine stellate ganglia to characterize their transcriptomic diversity. We demonstrate that SGCs in murine stellate ganglia can be divided into subpopulations that reflect different stages of maturation and activation, similar to astrocytes in the CNS. Moreover, we highlight the various developmental trajectories within the subtypes.

6.2 Materials and Methods

Animals

Eight C57BL/6J male mice (Jackson Laboratory, Bar Harbor, ME, USA), 10 to 12 weeks of age, were used for scRNAseq. Animals were sedated in an induction chamber (3-5% isoflurane) and sacrificed by decapitation. All experiments were performed in accordance with guidelines set forth by the University of California Institutional Animal Care and Use Committee (IACUC), and The National Institutes of Health Guide for the Care and Use of Laboratory Animals. Data from study animals were also used for neuronal scRNAseq, and cardiac/paw neurons were identified by retrograde labeling using adeno-associated viruses (AAV, subtype 2). Those data are being reported elsewhere. AAVs are known to trigger very low immune responses and are thus unlikely to have caused much reactive response in the stellate ganglia[229,230].

Cellular Dissociation

Bilateral stellate ganglia were identified, isolated and collected in artificial cerebrospinal fluid (ACSF; Figure 1A). Next, ganglia were incubated for an hour at 37°C in a digestion solution prepared with 500 µl TrypLE Express (ThermoFisher Scientific. Waltham, Massachusetts), 2000 µl Papain solution (Worthington Biochemical Corporation. Lakewood, New Jersey; 25 units/mL in ACSF), 100 µl Collagenase-Dispase (Millipore Sigma. Burlington, Massachusetts; 20 mg/mL in

ACSF) and 270 µl DNAse I (Worthington Biochemical Corporation. Lakewood, New Jersey; 200 units/mL in ACSF). Following this first hour, the cells were carefully triturated with fire-blown Pasteur pipettes every thirty minutes. After the second trituration, 100 µl fresh Collagenase-Dispase (20 mg/mL in ACSF) was added to the solution and cells were incubated for another hour at 37°C, triturating every thirty minutes. Next, the suspension was filtered through a 40mm filter (ThermoFisher Scientific. Waltham, Massachusetts) and ACSF was added to stop the enzymatic digestions. The suspension was spun down at 100*g* for four minutes at room temperature, and the pellet was collected and resuspended in 500 µL ACSF and 500 µL supplemented Neurobasal-A medium (ThermoFisher Scientific. Waltham, Massachusetts; 250 µL B27, 250 µL Penicillin-streptomycin, 31.3 µL L-Glutamine). The cell suspension was transferred onto a density gradient (Millipore Sigma. Burlington, Massachusetts) and centrifuged at 100*g* for 10 minutes at room temperature. The supernatant was carefully removed until the solution was concentrated to 500 µL.

Single Cell RNA Sequencing

Single cell RNAseq was performed by microfluidic capture-based encapsulation, barcoding, and library preparation (10x Genomics Chromium scRNAseq system (10x Genomics, Pleasanton, California). Cells were loaded into a Chromium Chip B along with partitioning oil, reverse transcription reagents, and a mix of hydrogel beads containing 3,500,000 unique 10X Barcodes. Paired end sequencing was performed on a Novaseq S4 system, using the v3 Illumina platform. Coverage depth was 20K per cell, and the read length was 2x50. Analysis including demultiplexing, reference based mapping (GRCm38.98), and UMI identification, was performed according to the 10X Cell Ranger pipeline.

scRNAseq data quality control, normalization, and integration

To eliminate lowly expressed genes, a gene was required to be expressed in at least 5 cells per sample, or was removed. Cells with more than 20,000 UMIs, 6,000 expressed genes, or less than 500 expressed genes were considered outliers, and removed. Cells were removed with more than 15 percent mitochondrial gene content. All of the following processing steps were performed in the R package Seurat (v3.0.0)[40]. Each sample was normalized separately, scaling the total UMI count to 10,000 per cell and log transformed. 2000 variable genes were identified among each sample with *FindVariableFeatures* function. The samples were integrated with Canonical correlation analysis[231] and mutual nearest neighbors with *FindIntegrationAnchors* and *IntegrateData* functions.

scRNA processing and clustering

The integrated scRNAseq data was scaled, regressing out the total number of genes, and percent mitochondrial content in each cell. Principal component analysis was performed reducing the dimensionality to 30 dimensions. A graph of cell neighbors was created with the first 20 principal components (PCs), and Louvain clustering[232] was used to find cell clusters. We performed Uniform Manifold Approximation and Projection (UMAP)[145] on the first 20 PCs to visualize the clustering in 2 dimensions.

Identification of cell clusters and marker genes

SGCs were identified by the expression of SGC markers *S100B* and *Fabp7* (Figure 1A; supplementary figure 1)[233,234]. The feature counts of individual cells were divided by the total number of counts for that cell, multiplied by 10,000 and natural-log transformed. 50 principal components were used for *Louvain* clustering, which was visualized using T-distributed stochastic neighbor embedding (t-SNE) dimensionality reduction analysis from the Seurat R Package (version 3.6.2) (Figure 1A). Differentially expressed genes (DEGs) were defined as genes that (1)

were expressed in ≥ 30% of the cells from the cluster of interest and (2) if there was a ≥ 0.25 log fold change in gene expression between this cluster and the other groups and (3) adjusted *p*-value < 0.05 using the Benjamini–Hochberg method to correct for multiple testing.

Resolving cell identities of the different clusters

We determined cell identities by (1) manual assessment of the cellular function of the proteins encoded by the DEGs, (2) comparing DEGs of each cluster with expression patterns of astrocytes in different stages of maturation and (3) in-depth analysis of canonical pathways associated with DEGs of the clusters. Astrocyte-specific markers were retrieved from multiple studies that have previously published the transcriptome of astrocytes in different stages of development [235–242].

Cell lineage trajectory analysis

Trajectory analysis was performed using the Slingshot R package[243]. This method first identifies lineages based on the minimum spanning tree between nodes, which are cell clusters here. Using a user defined root node, Slingshot subsequently computes individual cell pseudotimes for each inferred trajectory. We used cluster 1 as the root node, because it expressed genes suggestive of astrocyte progenitor cells such as *Slc12a2*, *Ptprz1* and *Itgb8* (supplementary figure 2)[235,237]. To identify trajectory associated genes for each predicted trajectory, we trained a random forest model (parsnip R package) on the top 1000 highly variable genes to predict cell pseudo times. We used 1,400 trees, 200 predictors sampled per split, and 15 as the minimum number of data points per node to be split again. This trained random forest model was used to identify the pseudotime associated genes based on their regression coefficients. A different random forest model was trained for each trajectory.

Pathway annotation of cluster marker genes

Ingenuity Pathway Analysis (IPA; QIAGEN's Ingenuity Pathway Analysis, Qiagen Redwood City, Build and content version: 49932394) was used to annotate the enriched Ingenuity pathways

among cluster marker genes. Lists containing the normalized average expression of the 500 highest expressed genes or cluster-specific DEGs were used to establish associated canonical pathways. All analyses were performed with the following settings: “Reference set: Ingenuity Knowledge Base (Genes only), Relationship to include: Direct and Indirect, Networks: Interaction and causal, Data sources: All, Species: All, Tissues and Cell lines: Nervous system, Consider only relationships where confidence = Experimentally Observed”. Multiple testing was corrected for using the Benjamini–Hochberg method. Pathways were considered to be significantly associated with a cluster when -log(BH-adjusted *p*-value) ≥ 1.3.

6.3 Results

Identification of SGC clusters in the mouse SG

Using acutely dissociated stellate ganglion preparations from 8 mice, we performed single cell RNA sequencing using the 10x Genomics scRNAseq platform. In total, 11,595 single SGCs were captured, expressing 5538 ± 2486 genes on average. Using the bi-dimensional t-distributed stochastic neighbor embedding (t-SNE) algorithm, we transformed the multidimensional data and visualized the relationships between the cells in a two-dimensional t-SNE plot. Louvain clustering produced 6 subclusters of SGCs (Figure 6.1B; supplementary file 1). All clusters were similarly represented in the stellate ganglia of the 8 mice (supplementary file 2; supplementary figure 3). The top 10 differentially expressed genes (DEGs) per cluster with the highest discriminatory power in terms of p-value are depicted in the heatmap in Figure 6.1C. The sixth cluster was excluded from further analyses due to relatively high expression of reactive markers, including interferon-related genes which may be related to an endogenous low-level reactive subtype, or transient exposure to AAVs.

Similarities between SGC clusters in the mouse SG

We began by examining similarities across the subclusters in their gene expression. A correlation map showing the similarities in average gene expression, demonstrated a high level of correlation between all clusters (Figure 6.2A; Supplementary file 3). Biochemical pathways associated with the 500 highest expressed genes for each cluster are listed in Supplementary file 4. As shown in Figure 2B, 36 ± 14% of the pathways were similar across clusters, representing cellular signaling, growth and development, metabolism, immune response, cell cycle, and cellular stress and injury. These pathways, corresponding to glia-glia and glia-neuron communication, indicate that these functions were important for cell types across all clusters.

Determination of SGC identities in the mouse SG

Next, we investigated the cellular heterogeneity among the subclusters by examining differentially expressed genes (DEGs). In-depth analysis of the DEGs in each cluster suggested that the cell populations were distinct from each other based on state of maturation or functionality (Figure 6.3A; supplementary figure 6.4).

Cluster 1, representing 14.57% of all SGCs, was identified as immature SGCs. This population was enriched in genetic cell cycle and pluripotency markers, corresponding to developing astrocytes in the CNS (Figure 6.3B)[235,238–242]. This corresponded with the, though insignificant, upregulation of the pluripotency pathway in this cluster (Figure 6.3C). Cluster 2 and 3 were the two largest populations representing 25.27 and 32.00% of all SGCs, respectively. Both clusters had high expression of genes associated with mature astrocytes (Figure 6.3A and B)[235–238]. Consistent with this, functional pathways, such as cholesterol synthesis were enriched in clusters 2 and 3 (Figure 6.3C). These clusters were therefore identified as mature SGCs.

In contrast, clusters 4 and 5, containing 13.00 and 8.20% of the SGCs population, respectively, were characterized by downregulation of the functional pathways (Figure 6.3C) and showed activated stress response pathways. As such, both clusters were classified as aged SGCs since

cellular stress responses and downregulation of metabolism pathways are indicative of aging[244]. However, this upregulation of cellular stress response pathways was less clear in cluster 5 (Figure 6.3C). Moreover, whereas the NRF2 -mediated stress response seemed to be slightly upregulated, the production of nitric oxide and reactive oxygen species was downregulated (Figure 6.3C). Therefore, we identified cluster 5 as a quiescent, aged group of SGCs.

Supplementary file 5 lists the pathways associated with the DEGs of each cluster. DEGs of the different clusters were also compared to external transcriptomic datasets of astrocytes in the CNS (Figure 6.3B)[235,236]. Combined, these results led to the grouping of the clusters in different states of maturation and activation (Figure 6.3A).

Maturation and activation trajectories revealed by pseudo-time analysis

To test our hypothesis that the different clusters compose developmental and functional trajectories, we performed pseudo-time analysis on the SGC clusters. Using cluster 1 as the starting cluster, three different trajectories were obtained (Figure 6.4A-C; supplementary figure 6.5). Similar to the aforementioned developmental progression, cluster 1 seemed to evolve to cluster 2 (and 3), which were identified as the mature SGCs. From there, SGC were observed to progress to cluster 4 or 5 which were identified as the (quiescent) aged SGCs (Figure 6.4A, B). Genes that trace these trajectories include Adamts5 which was progressively downregulated as SGCs matured to subsequent clusters. Adamts5 is a metalloproteinase involved in cellular development and cell migration[245], causing it to be upregulated in developing astrocytes[235], which corresponds to the developmental function associated with cluster 1. Moreover, Txnip and Hspb1, which are both involved in cellular stress responses[246 247] and associated with aging astrocytes [235], were traced in the trajectories leading up to the aged clusters 4 and 5.

Signaling

Biochemical pathways associated with cell junction signaling, such as adherens and gap junctions signaling, were identified in all clusters, indicating the importance of direct cell-cell adhesion and communication pathways in fundamental functions of SGCs. In addition, the mature SGC in cluster 3 were enriched for genes involved in aldosterone, endothelial nitric oxide synthase (eNOS), purinergic and gap junction signaling (Figure 6.5A, B; bold).

However, as the SGCs mature and age, distinct signaling pathways are employed. In general, signaling pathways associated with proliferation and metabolism, such as the aldosterone[248] (Figure 6.5A; blue) and mTOR signaling pathways[249] (Figure 6.5A, B; blue), and pathways associated with cell proliferation and neuronal guidance, such as 14-3-3 mediated signaling[250] and Reelin signaling[251], were decreased (Figure 6.5B; blue). Also, adherens junction signaling seemed most important in the immature SGCs (Figure 6.5B; bold).

Moreover, aging (Figure 6.5A; arrows) was associated with the gradual upregulation of signaling pathways involved in cell death and cellular stress, *e.g.* androgen[252] and BAG2 signaling pathways[253].

Furthermore, as the SGCs transitioned to a more quiescent state (Figure 6.5B; arrow), the Sirtuin signaling pathway was upregulated, which is involved in aging and stress resistance[254].

6.4 Discussion

Major findings

In the present study, we assessed the transcriptomic profiles of SGCs in the stellate ganglia at single cell resolution. We show that the SGCs in the murine stellate ganglia are 1) a fairly heterogenous population of cells, that can be separated into five subpopulations of SGCs based on their state of maturation or activation, and 2) that these subpopulations result from two

nonlinear trajectories. In addition, we show that 3) signaling pathways change as SGCs progress through the different trajectories, indicating different functions of the subclusters.

SGC subpopulations in the murine stellate ganglia

To our knowledge, this is first study to elucidate the complexity of SGCs in the stellate ganglia based on transcriptomic profiles. We observed that the population of SGCs comprised five subpopulations that shared many characteristics, including signaling and metabolism pathways (Figure 6.2A). However, in-depth evaluation of cluster-specific DEGs and pathways revealed that the clusters represented SGCs in different states of maturation and/or states of functionality. This corresponds with the CNS, where astrocytes are known to co-exist in different states of maturation[255]. Moreover, our results also corroborate studies by Pannese *et al.* (1960) and Zhang *et al.* (2019) which suggested subclustering within the SGCs based on microscopic imaging and immunohistochemical experiments, respectively. Using scRNAseq, we were able to validate the existence of SGC subpopulations and indicate that these cells can assume different states of activation that presumably change over time. We defined five subclusters and provided transcriptomic markers that allude to a functional identity.

We found that SGCs go through developmental progression within the SG and identified cluster 1 as the most immature cluster. As pluripotency markers such as *Ptprz1* and *Itgb8* were enriched in these cells (Figure 6.3A), we suspect that this subpopulation remains undifferentiated and serves to locally replenish the SGC population in the SG. Moreover, *NG2* was upregulated in cluster 1. In the CNS, *NG2* positive glia have been established to generate astrocytes as well as oligodendrocytes, strengthening the identification of cluster 1 as a pluripotent population of SGCs (Zhu et al. 2008).

As the SGCs progress to more mature states, represented by the progression from cluster 2 to 3 in our data, metabolic functions such as cholesterol syntheses become increased (Figure 6.3C). Starting from cluster 3 and up, cells become increasingly enriched in cellular stress pathways, such as the unfolded protein response. Therefore, we established an aging trajectory as the cluster numbers go up. Moreover, though insignificant, the senescence pathway became upregulated as SGCs progress from cluster 2 to cluster 4 (Figure 3C). In addition, cholesterol-associated pathways that were upregulated in the mature SGCs, became increasingly downregulated in the aged SGCs.

With regard to the highest expressed genes, cluster 5 was highly similar to cluster 4 (Figure 6.2A) and was also classified as aged SGCs. However, the enrichment in cellular stress pathways in cluster 4 was one of the main characteristics that distinguished the two clusters. The lesser inflammatory response combined with the low activity of functional metabolic and cell division pathways, led to the identification of cluster 5 as quiescent, aged SGCs. However, it could also indicate that this cluster comprises recovering reactive SGCs and/or a mid-state in between the aging SGCs of cluster 3 and 4.

Signaling

As glia-glia and glia-neuron communication is of fundamental importance to the functioning of SGCs, we performed a more detailed evaluation of the signaling pathways that were active and evolving in the SGCs.

We demonstrated that signaling pathways associated with the cell cycle and neuronal guidance seemed to be enriched in the immature SGCs in cluster 1, and progressively downregulated as SGCs progressed to a more mature or activated state (Figure 6.5; left columns). Moreover, compared to all other clusters, cell junction signaling seemed to be increased in these immature

SGCs (Figure 6.5B; Bold). As cell-cell adhesion and communication has been shown to be important in stem cell behavior, the increased correlation of these pathways and cluster 1 further corroborate its pluripotent identity[256,257].

With maturation, SGCs increased their aldosterone, eNOS, purinergic and gap junction signaling (Figure 6.5A; Bold). These pathways correspond with experimental data on SGC communication in the trigeminal and dorsal root ganglia[217]. Aldosterone signaling has been demonstrated to result in astrocyte proliferation[248], but can also alter astrocyte function and activity[258], which might cause neuronal death[259]. Comparison of these studies indicate a dose-dependent effect of aldosterone, the higher aldosterone levels causing neurotoxic effects. As aldosterone levels are also increased in HF, this could possibly indicate a (direct) connection between (neurohumoral) changes during the development and progression of HF and SGC function and activity[173].

As astrocytes aged, (oxidative) stress signaling pathways, such as through the BAG2 protein[253] pathway, became progressively enriched (Figure 6.5A). Androgen signaling, which has been demonstrated to promote death in primary cortical astrocytes[252], and α-adrenergic signaling were also increased (Figure 6.5A). Although the exact meaning of the latter pathway remains to be elucidated, Paukert *et al.* (2014) showed that norepinephrine enhanced the reactivity of astrocytes to neuronal activity in the CNS. Surprisingly, melatonin signaling was also increased as SGCs aged (Figure 6.5A). Signaling through melatonin has been shown to protect astrocytes from oxidative-stress induced stress responses[260], and might therefore reflect a compensatory reaction of aging SGCs.

Clinical relevance

SGCs have been demonstrated to modulate efferent sympathetic outflow to the heart and to be more activated in the stellate ganglia of cardiomyopathy patients[173]. Using scRNAseq, we offer

unique insights into SGC biology and diversity. Especially the knowledge on glial communication, its dynamic response to injury and how this injury-response translates to modulation of sympathetic outflow, might be an interesting target for future HF therapies.

Future directions

The present study is the first to identify the co-existence of SGC subtypes in murine stellate ganglia and to describe their (functional) characteristics. Hence, the field of SGC behavior in stellate ganglia is wide open. The change in SGC characteristics and populations over time remains to be fully disseminated. Understanding how SGC adapt to aging and changing conditions, would give more insight on the biological functions of SGCs and their modulatory and supportive roles in the stellate ganglia. Moreover, the current study only included male mice to exclude the possible effects of female sex hormones. Nevertheless, the degree to which sex differences affect SGC behavior would be an interesting direction for future studies.

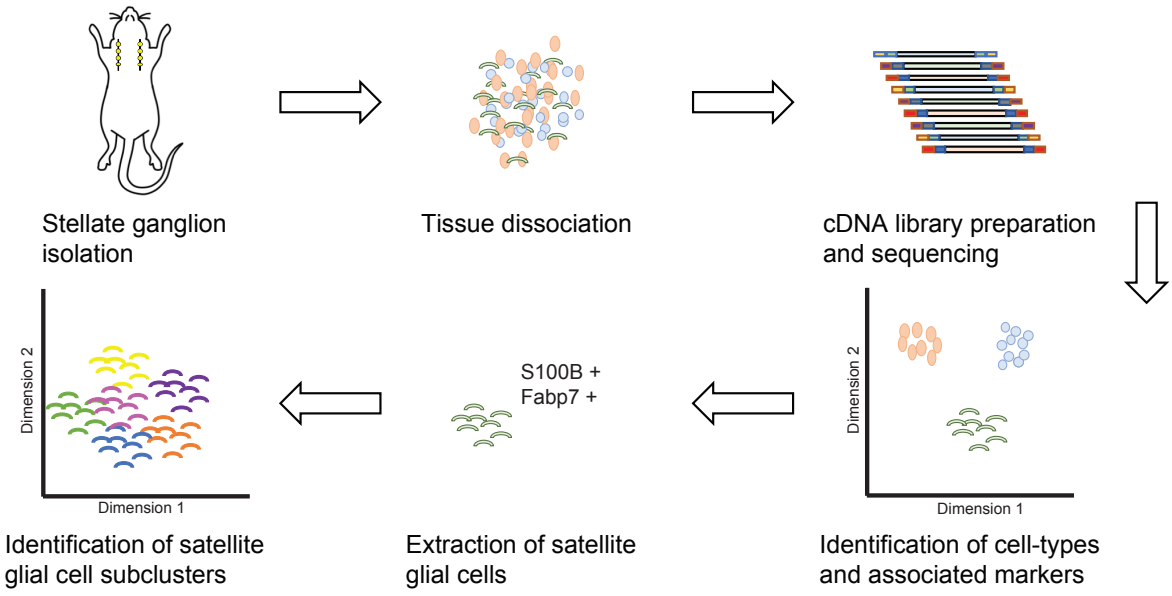
Limitations

While we have presented comprehensive transcriptomic analyses, we have not demonstrated physiological distinctions across the transcriptomes as this is beyond the scope of this report. The mice included in the study had received AAVs through paw and intracardiac injections, which were retrogradely transported to the stellate ganglia. Nevertheless, it is unlikely that these AAVs have disrupted normal SGC dynamics in the stellate ganglia as AAVs are known to trigger very low immune responses and to be rapidly cleared[229,230]. Moreover, the AAVs in this study were injected locally with minimal systemic exposure and studies were performed three weeks after injection, which is well beyond the half-life and functional clearance of the viruses[261,262]. Even though it is unlikely to have disturbed SGC dynamics by AAV exposure, the sixth cluster, which was characterized by increased expression of pro-inflammatory markers, was excluded from further analyses.

6.5 Conclusion

SGCs are a heterogeneous population of neuronal modulatory cells in the SG. Their diversity is based on state of maturation or differentiation, which might be dynamically responding to environmental changes in health and disease. Functional studies in control and pathologic conditions are warranted to validate these findings.

6.6 Figures



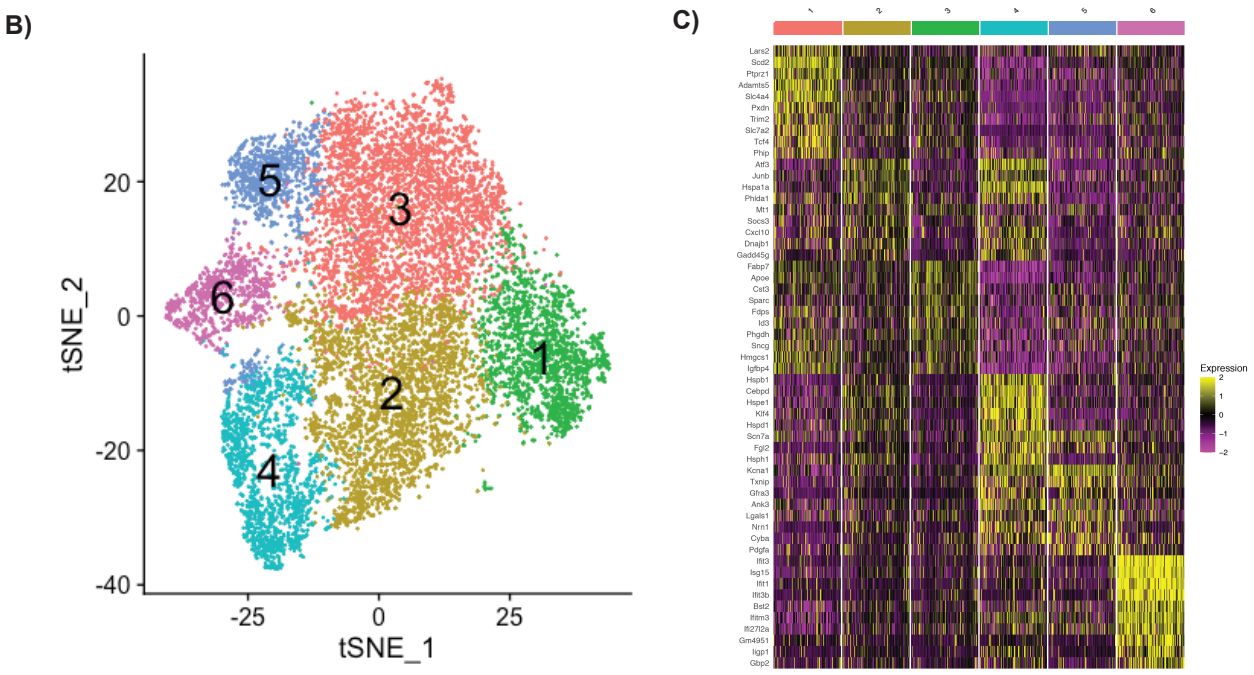
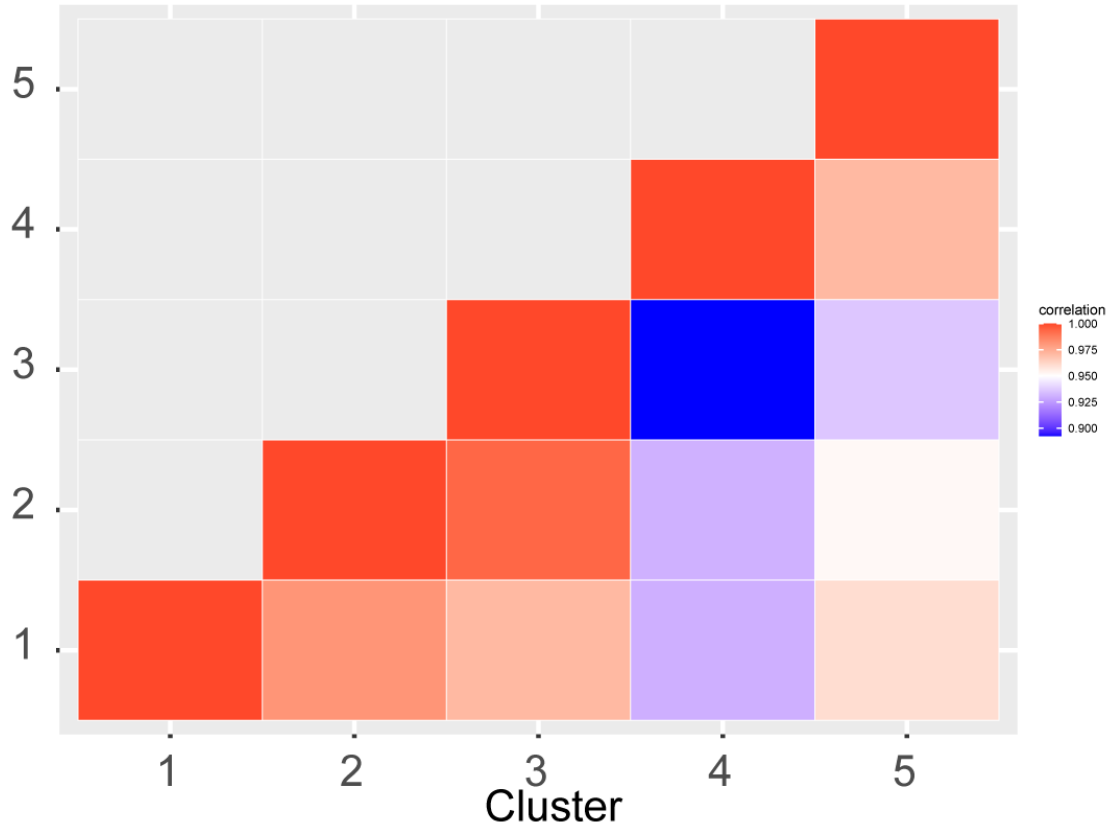


Figure 6.1. Single-cell RNA sequencing of murine stellate ganglia shows six subpopulations.

(A) Pipeline depicting the derivation of stellate ganglion cells which were used for single cell RNA sequencing. Cell types within the stellate ganglia were identified and satellite glial cells (SGCs) were identified by the expression of genetic markers *S100B* and *Fabp7.* SGCs were distinguished from other cell types and used for subclustering. **(B)** Unsupervised clustering of SGC resulted in the identification of six SGC subclusters **(C)** Heatmap of the most differentially expressed genes, displayed as scaled expression levels, of the different SGC clusters.

A)


5
4
3
2
1
1
2
3
4
5
Cluster
correlation
1.000
0.975
0.950
0.925
0.900

B)

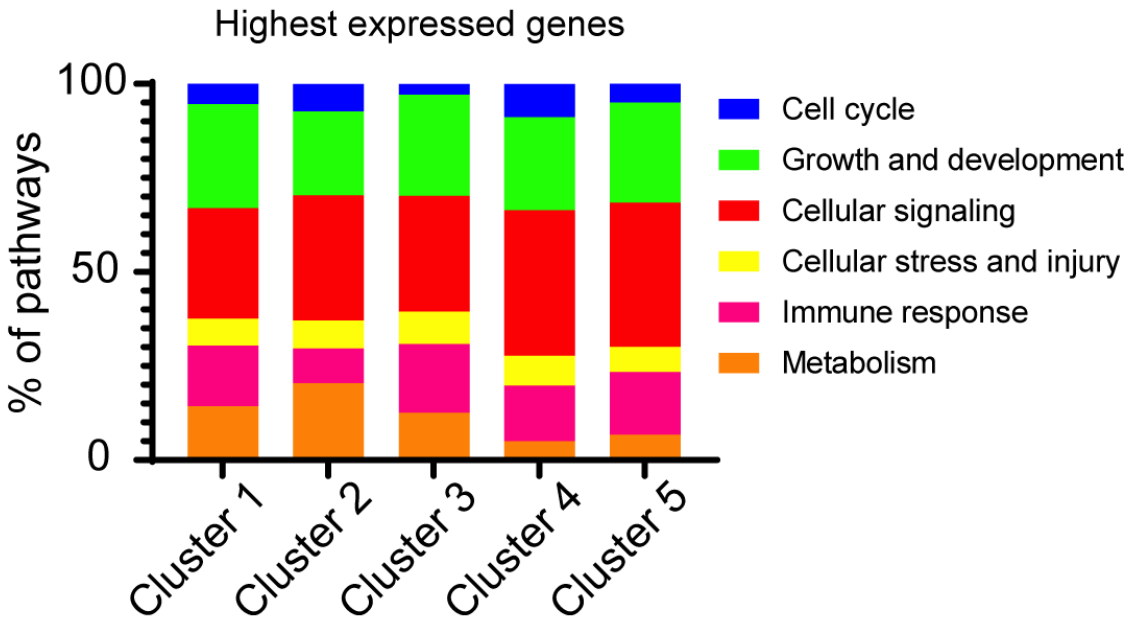

Highest expressed genes
% of pathways
100
50
0
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cell cycle
Growth and development
Cellular signaling
Cellular stress and injury
Immune response
Metabolism

Figure 6.2. Average gene expression and highest expressed pathways show many similarities between the different subclusters of satellite glial cells.

(A) Correlation map of the average gene expression of the five clusters. Extent of similarity is depicted by color; the blue and red colors corresponding with a continuous progression from lower to higher levels of similarities, respectively **(B)** Classification of the pathways corresponding to the 500 highest expressed genes of the different clusters to different biological functions.

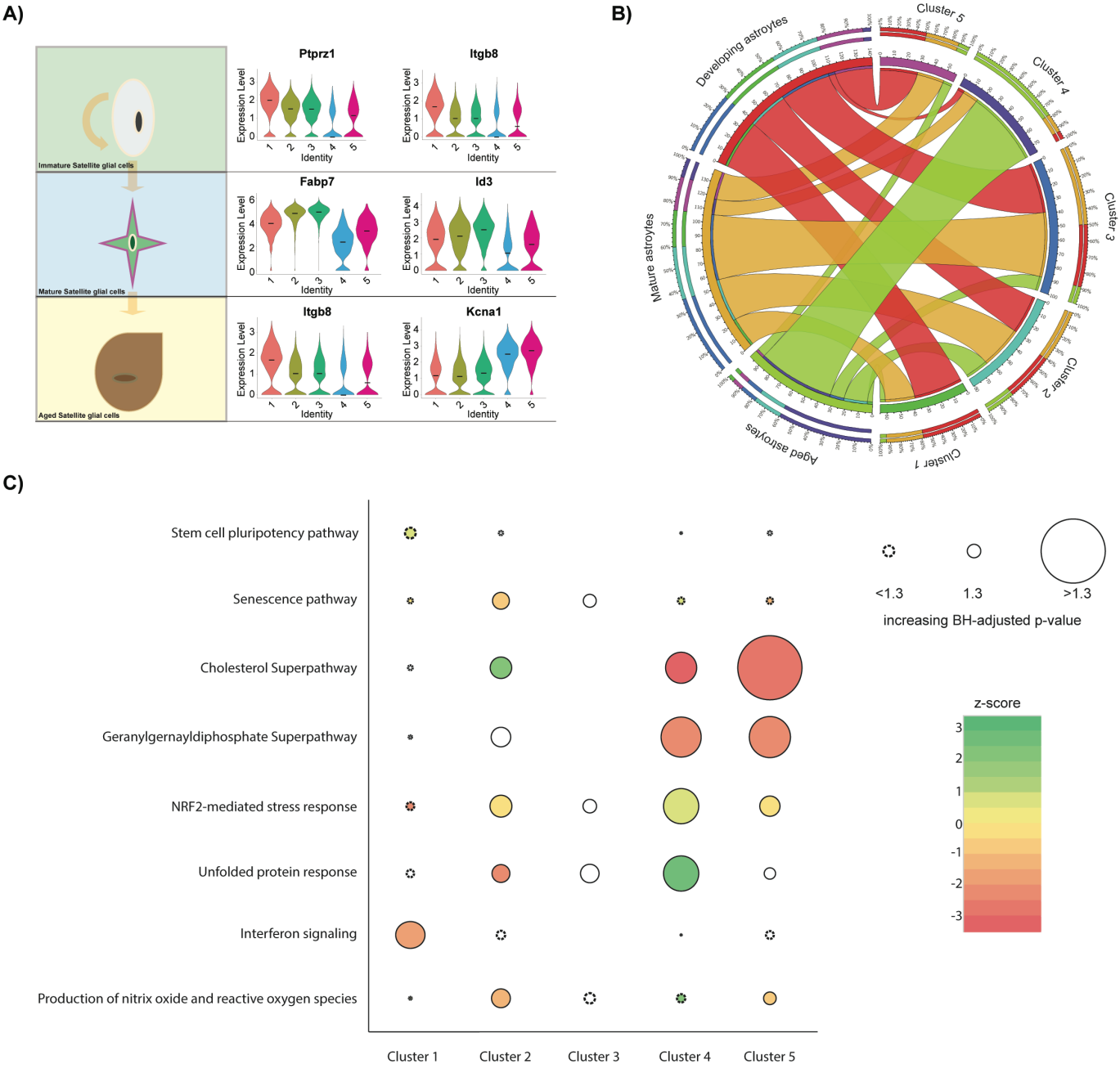


Figure 6.3. Satellite glial cell (SGC) subcluster identification.

(A) The clusters were identified as SGCs in different stages of maturation or activation. As such, immature SGC are believed to either proliferate or progress to mature and aged SGC. The violin plots depict marker genes associated with the different states of SGC. **B)** Comparison of the differentially expressed genes (DEGs) of the different clusters with marker genes of astrocytes in

different stages of maturation. Extend of overlap between the different clusters of SGCs and developing astrocytes (red), mature astrocytes (orange) and aged astrocytes (yellow) corresponds with the thickness of the bands. **C)** Pathways associated with the DEGs of the different clusters. P-value correlates with dot size, a dashed line indicating insignificance. Colours of the dots represent the z-score of the pathway in the respective cluster, when white, z-score could not be established.

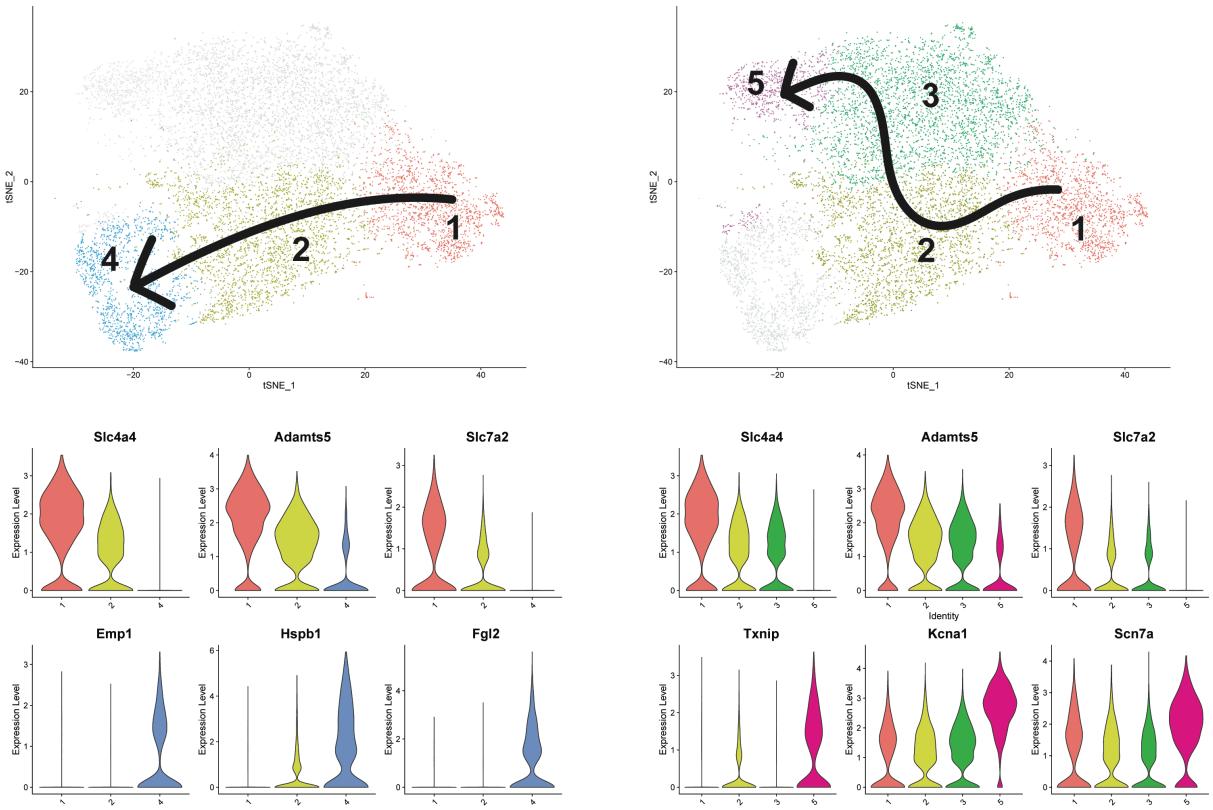
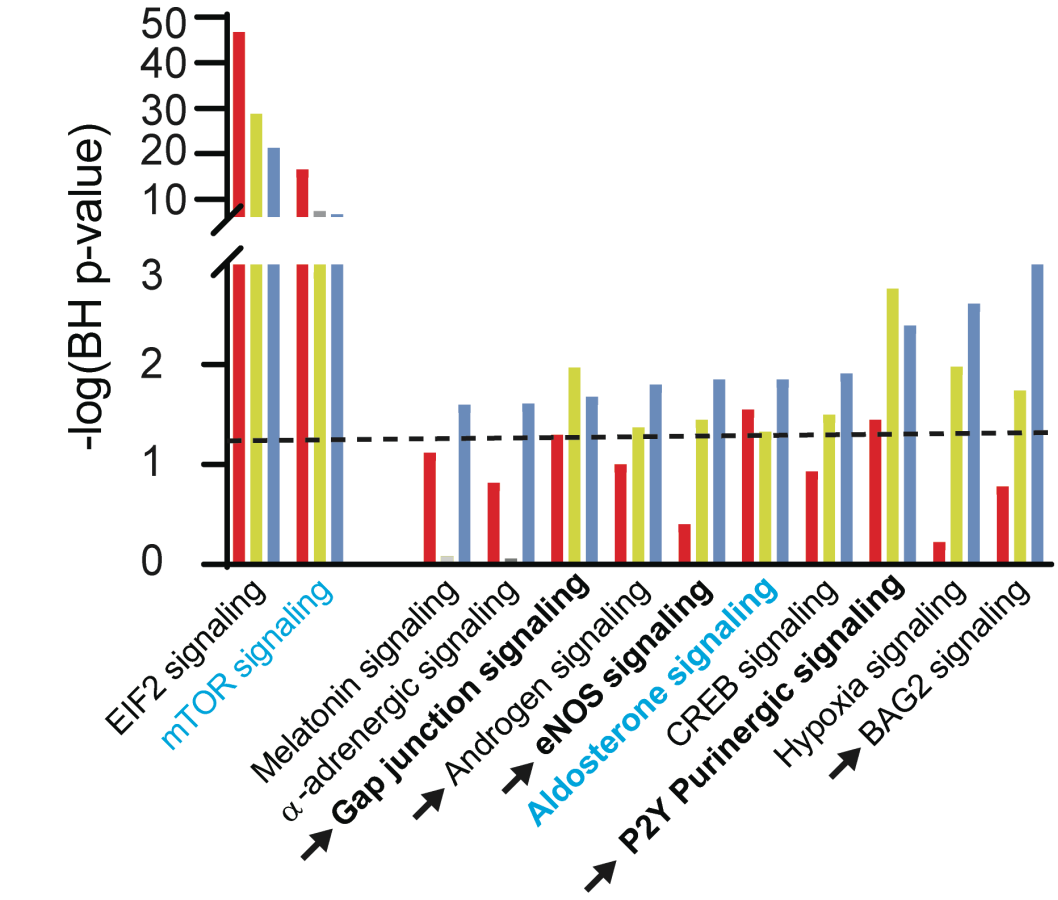


Figure 6.4. Pseudotime analysis of satellite glial cells (SGCs) results in three separate trajectories.

The first trajectory comprised the progression of SGCs from cluster 1 to 2, to 3 and finally to 5 **(A)**. The second trajectory described the progression from cluster 1, to 3 to 4 **(B).** Violin plots of the expression patterns of the genes associated with these trajectories across the clusters are depicted underneath their respective tSNE plots.


-log(BH p-value)
50
40
30
20
10
3
2
1
0
EIF2 signaling
mTOR signaling
Melatonin signaling
α-adrenergic signaling
Gap junction signaling
Androgen signaling
eNOS signaling
Aldosterone signaling
CREB signaling
P2Y Purinergic signaling
Hypoxia signaling
BAG2 signaling

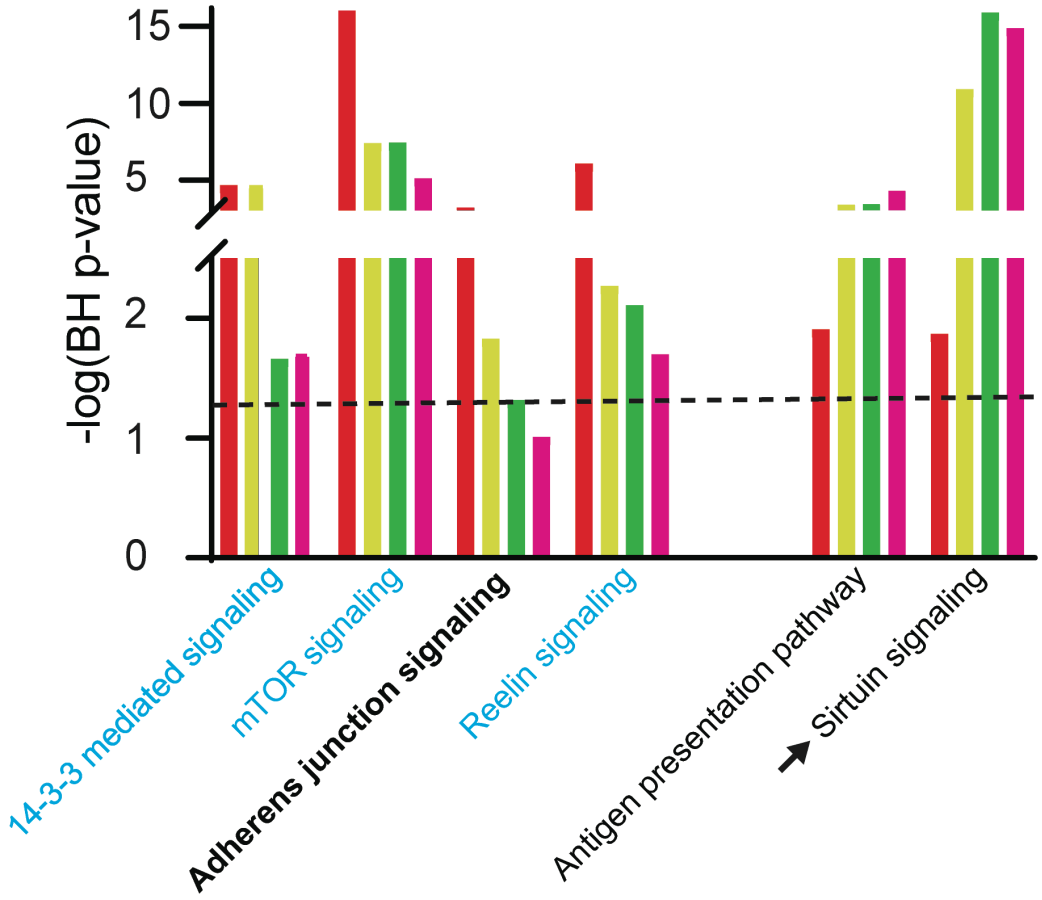
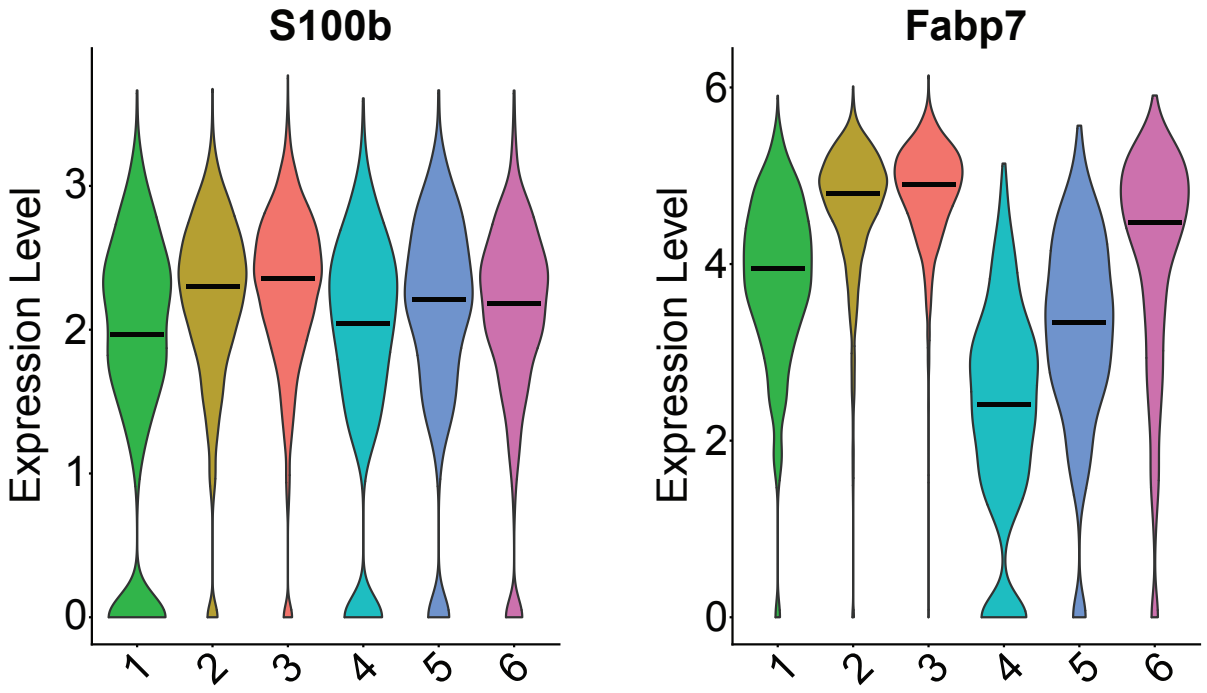
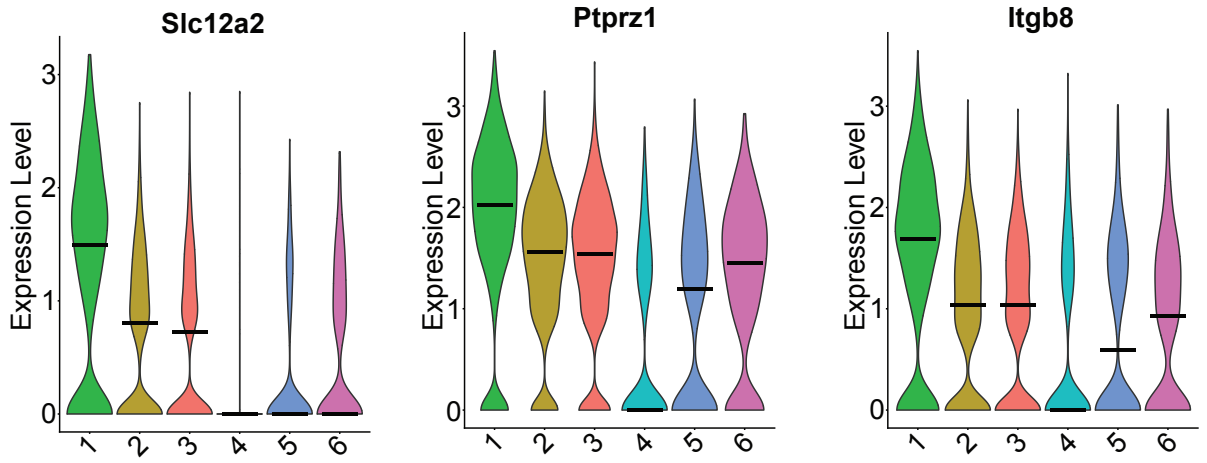

-log(BH p-value)
15
10
5
2
1
0
14-3-3 mediated signaling
mTOR signaling
Adherens junction signaling
Reelin signaling
Antigen presentation pathway
Sirtuin signaling

Figure 6.5. Active signalling pathways of the different trajectories in murine satellite glial cells (SGCs) in the stellate ganglia.

A) Signalling pathways that are down- (left) or upregulated (right) as SGCs progress from immature SGCs (cluster 1; red) to mature (cluster 2; gold) to aged SGCs (Cluster 4; light blue). **B)** Signalling pathways that are down- (left) or upregulated (right) as SGCs progress from immature SGCs (cluster 1; red) to mature (cluster 2 (gold) and 3; green) to aged, quiescent SGCs (Cluster 5; pink).

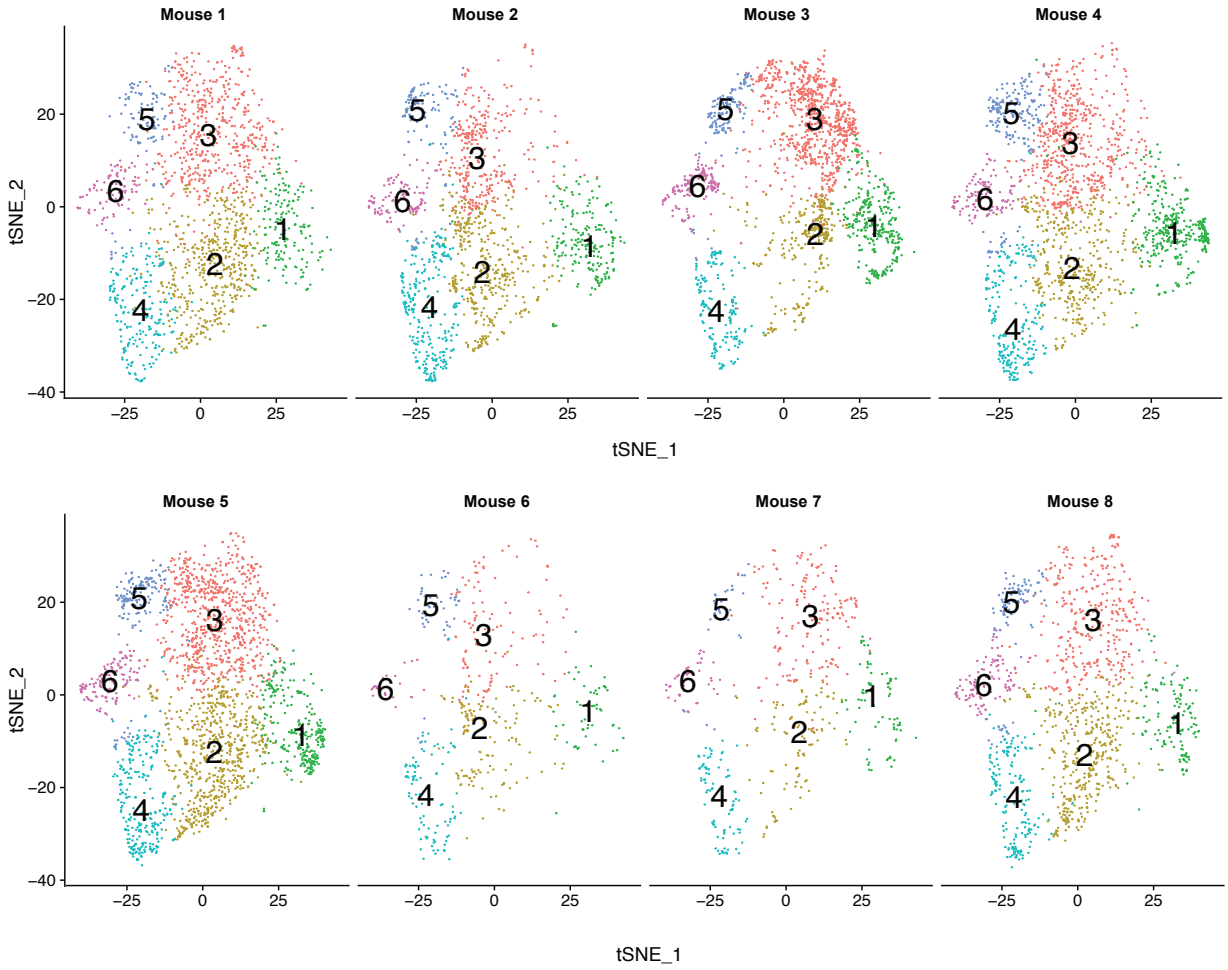


<u>Supplementary figure 6.1:</u> Violin plots showing the expression of the known SGC markers *S100B* and *Fabp7* in the different clusters.



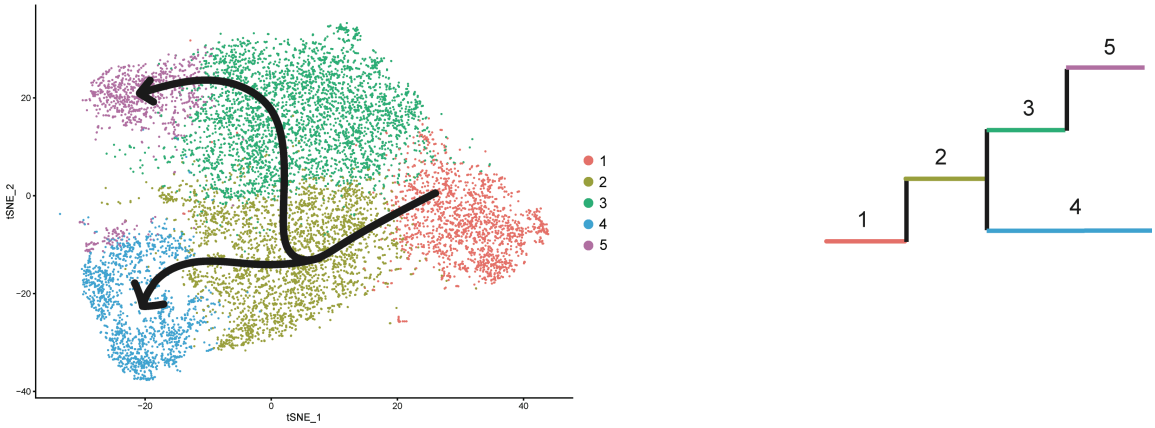
<u>Supplementary figure 6.2:</u> Expression levels of astrocyte immaturity markers in the satellite glial cell clusters.

Violin plots depicting the relative expression levels of genetic markers *Slc12a2*, *Ptprz1* and *Itgb8,* which are highly expressed in astrocyte progenitor cells. As all three markers were highest expressed in cluster 1, this cluster was chosen as the root node for subsequent trajectory analysis.



<u>Supplementary figure 6.3:</u> t-SNE plots showing that all clusters were present and similarly represented in all eight mice.

<u>Supplementary figure 6.4:</u> t-SNE plot of the different clusters with their corresponding state of maturation or reactivity.



<u>Supplementary figure 6.5:</u> Pseudo-time analysis on the satellite glial cell (SGC) clusters resulted in three different trajectories.

Cluster 1, identified as the immature SGC, initially matured to cluster 2. From there, SGC development branched, progressing to either cluster 3 or 4, which were identified as mature and aged SGCs, respectively. SGCs in cluster 3 subsequently progressed to aged SGC in cluster 5.

Chapter 7. Conclusions and Future Directions

Single cell multi-omics technologies are rapidly evolving and data volume is growing exponentially, calling for new methodologies for efficient data analysis, integration, and interpretation. To this end, my dissertation focused on two major aims. (1) To develop robust computational approaches for single cell multi-omics integration and analysis. (2) To apply single cell multi-omics to understand the mechanisms of pathophysiology of complex disease as well as potential therapeutics. In aim 1 we developed JSTA to combine scRNAseq references with single molecular resolution spatial transcriptomics to segment cell boundaries, assign mRNA molecules to cells, and classify cell types in combinatorial FISH data, and SCING to construct cell-type specific GRNs from scRNAseq and spatial transcriptomics data. In aim 2 we applied scRNAseq to identify the mechanism underlying the ameliorative effects of IDOL knockdown in the hippocampus of APP/PS1 mice, to identify cardiac innervating neuronal subtypes from the stellate ganglion, and the satellite glial subpopulations of the stellate ganglion. These methodological approaches and biological findings are important steps in understanding and treating complex diseases such as Alzheimer's disease and cardiovascular disease.

The JSTA method was developed to address the lack of existing tools in the spatial transcriptomics space to simultaneously segment cell boundaries, generate gene expression counts in individual cells, and label each cell with cell type identity. Without accurately segmenting each mRNA molecule into each cell, single cell analysis of spatially resolved transcriptomics is much noisier and less powerful. JSTA is built on an expectation maximization and deep learning framework to segment and classify each cell based on a high quality scRNAseq reference[20]. We apply JSTA to the mouse hippocampus and show spatial distribution of 133 high resolution cell subtypes as well as spatial DEGs within each cell type. JSTA provides a tool for future studies identifying spatially resolved transcriptomic changes in complex disease. For instance, we are currently applying JSTA to compare MERFISH data between sham controls and mice with

traumatic brain injury to understand cell types and genes with spatial alterations in disease. Currently JSTA is designed for multiplexed FISH data. In the future we plan on applying JSTA to MERFISH datasets for understanding spatial consequences of traumatic brain injury, and to other publicly available FISH datasets to provide clean datasets for inferring cell-cell interaction GRNs.

In parallel, we developed SCING for elucidating cell type specific GRNs. GRNs produced by SCING are more robust than existing methods and biologically interpretable through pathway enrichment analysis and regulator identification. Cell specific module expression provides a tool for identifying phenotype associated gene programs and key drivers of disease. Next, we will integrate the current SCING framework with spatial transcriptomics and RL interaction databases for infering cell-cell communication GRNs, and with ATACseq and transcription factor binding site databases for refining GRNs through epigenomic data.

Building on the tools developed, we have carried out a series of collaborative application studies to identify key cell types and regulators in the neuronal system for physiology and disease. We first explored the cell types affected in a mouse AD model and how the inhibition of the IDOL gene in the brain rescues learning and memory by shifting the microglial subpopulation to DAM or phagocytic microglia that clear plaque. Next, we applied SCING to snRNAseq and 10x Visium spatial transcriptomics to identify gene network changes in various cell types including microglia and to resolve the network changes in spatial domains. Further, we utilized JSTA to identify and spatially localize high-resolution cell subtypes in the hippocampus, a key brain region involved in AD. Future investigation of phenotypes (morphological, electrophysiological, etc.) associated with these high-resolution cell subtypes will reveal how changes in their distribution and gene programs influences physiology and disease. Finally, we applied scRNAseq with AAV retrograde tracing to identify heart innervating neurons and the satellite glial subtypes surrounding them in the stellate ganglion. We show the proportions of these neuronal subpopulations are changed in

mice with dilated cardiomytopathy, providing a potential therapeutic avenue for heart failure. These application studies demonstrate the power of single cell multiomics methods in accelerating biological discoveries in diverse disease areas. We are systematically constructing human and mouse cell type specific GRNs using existing multiomics data from cell atlases to enable network studies across diseases at cell type resolution.

In summary, we developed and applied methods for scRNAseq and spatial transcriptomics to understanding mechanisms underlying complex diseases and providing resources for the scientific community to further their research. These tools and the biological insights obtained will be impactful for many studies to follow.

REFERENCES

1. Schork, N. J. Genetics of complex disease: approaches, problems, and solutions. *Am. J. Respir. Crit. Care Med.* **156**, S103-9 (1997).
2. Buchanan, A. V., Weiss, K. M. & Fullerton, S. M. Dissecting complex disease: the quest for the Philosopher's Stone? *Int. J. Epidemiol.* **35**, 562–571 (2006).
3. Yang, X. Multitissue Multiomics Systems Biology to Dissect Complex Diseases. *Trends Mol. Med.* **26**, 718–728 (2020).
4. Nomura, S. Single-cell genomics to understand disease pathogenesis. *J. Hum. Genet.* **66**, 75–84 (2021).
5. Shao, X. *et al.* Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk. *Nat. Commun.* **13**, 4429 (2022).
6. Littman, R. *et al.* Joint cell segmentation and cell type annotation for spatial transcriptomics. *Mol. Syst. Biol.* **17**, e10108 (2021).
7. Chen, W.-T. *et al.* Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer's Disease. *Cell* vol. 182 976-991.e19 Preprint at https://doi.org/10.1016/j.cell.2020.06.038 (2020).
8. Burgess, D. J. Spatial transcriptomics coming of age. *Nature reviews. Genetics* vol. 20 317 (2019).
9. Asp, M. *et al.* A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. *Cell* **179**, 1647-1660.e19 (2019).
10. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nature methods* vol. 11 360–361 (2014).

11. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
12. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235–239 (2019).
13. Najman, L. & Schmitt, M. Watershed of a continuous function. *Signal Processing* vol. 38 99–112 Preprint at https://doi.org/10.1016/0165-1684(94)90059-0 (1994).
14. Beucher S, L. C. Use of Watersheds in Contour Detection. *International Workshop on image processing: Real-time Edge and Motion detection/estimation* (1979).
15. Vu, Q. D. *et al.* Methods for Segmentation and Classification of Digital Microscopy Tissue Images. *Front Bioeng Biotechnol* **7**, 53 (2019).
16. Al-Kofahi, Y., Lassoued, W., Lee, W. & Roysam, B. Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Trans. Biomed. Eng.* **57**, 841–852 (2010).
17. Qian, X. *et al.* Probabilistic cell typing enables fine mapping of closely related cell types in situ. *Nat. Methods* **17**, 101–106 (2020).
18. Park, J. *et al.* Segmentation-free inference of cell types from in situ transcriptomics data. Preprint at https://doi.org/10.1101/800748.
19. Petukhov, V., Soldatov, R. A., Khodosevich, K. & Kharchenko, P. V. Bayesian segmentation of spatially resolved transcriptomics data. *Cold Spring Harbor Laboratory* 2020.10.05.326777 (2020) doi:10.1101/2020.10.05.326777.
20. Yuste, R. *et al.* A community-based transcriptomics classification and nomenclature of neocortical cell types. *Nat. Neurosci.* **23**, 1456–1468 (2020).
21. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
22. Chen, K. H. *Spatially Resolved, Highly Multiplexed RNA Profiling in Single Cells*. (2015).

23. Moffitt, J. R. *et al.* Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, (2018).

24. Codeluppi, S. *et al.* Spatial organization of the somatosensory cortex revealed by osmFISH. *Nature Methods* vol. 15 932–935 Preprint at https://doi.org/10.1038/s41592-018-0175-z (2018).

25. Lee, J. H. Quantitative approaches for investigating the spatial context of gene expression. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **9**, (2017).

26. Turczyk, B. M. *et al.* Spatial Sequencing: A Perspective. *J. Biomol. Tech.* **31**, 44 (2020).

27. Lee, J. H. *et al.* Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat. Protoc.* **10**, 442–458 (2015).

28. Lee, J. H. *et al.* Highly Multiplexed Subcellular RNA Sequencing in Situ. *Science* **343**, 1360–1363 (2014).

29. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).

30. Salmén, F. *et al.* Barcoded solid-phase RNA capture for Spatial Transcriptomics profiling in mammalian tissue sections. *Nat. Protoc.* **13**, 2501–2534 (2018).

31. Lundberg, E. & Borner, G. H. H. Spatial proteomics: a powerful discovery tool for cell biology. *Nat. Rev. Mol. Cell Biol.* **20**, 285–302 (2019).

32. Gerdes, M. J. *et al.* Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 11982–11987 (2013).

33. Lin, J.-R., Fallahi-Sichani, M. & Sorger, P. K. Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. *Nat. Commun.* **6**, 8390 (2015).

34. Goltsev, Y. *et al.* Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **174**, 968-981.e15 (2018).

35. Lin, J.-R. *et al.* Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *Elife* **7**, (2018).

36. Keren, L. *et al.* A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. *Cell* **174**, 1373-1387.e19 (2018).
37. Dries, R. *et al.* Giotto, a pipeline for integrative analysis and visualization of single-cell spatial transcriptomic data. *Cold Spring Harbor Laboratory* 701680 (2019) doi:10.1101/701680.
38. Biancalani, T., Scalia, G., Buffoni, L., Avasthi, R. & Lu, Z. Deep learning and alignment of spatially-resolved whole transcriptomes of single cells in the mouse brain with Tangram. *bioRxiv* (2020).
39. Tran, H. T. N. *et al.* A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol.* **21**, 12 (2020).
40. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
41. Welch, J. D. *et al.* Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. *Cell* **177**, 1873-1887.e17 (2019).
42. Lopez, R. *et al.* A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements. *arXiv [cs.LG]* (2019).
43. Abdelaal, T., Mourragui, S., Mahfouz, A. & Reinders, M. J. T. SpaGE: Spatial Gene Enhancement using scRNA-seq. *Nucleic Acids Res.* **48**, e107 (2020).
44. Gamarra, M., Zurek, E., Escalante, H. J., Hurtado, L. & San-Juan-Vergara, H. Split and merge watershed: A two-step method for cell segmentation in fluorescence microscopy images. *Biomed. Signal Process. Control* **53**, 101575 (2019).
45. Surut, Y. & Phukpattaranont, P. Overlapping cell image segmentation using surface splitting and surface merging algorithms. in *Second APSIPA Annual Summit and Conference* 662–666 (2010).
46. Correa-Tome, F. E. & Sanchez-Yanez, R. E. Integral split-and-merge methodology for real-time image segmentation. *JEI* **24**, 013007 (2015).

47. Chaudhuri, D. & Agrawal, A. Split-and-merge procedure for image segmentation using bimodality detection approach. *Def. Sci. J.* **60**, 290–301 (2010).

48. HuBMAP Consortium. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* **574**, 187–192 (2019).

49. Mukamel, E. A. & Ngai, J. Perspectives on defining cell types in the brain. *Curr. Opin. Neurobiol.* **56**, 61–68 (2019).

50. Trapnell, C. Defining cell types and states with single-cell genomics. *Genome Res.* **25**, 1491–1498 (2015).

51. Egozi, A., Bahar Halpern, K., Farack, L., Rotem, H. & Itzkovitz, S. Zonation of Pancreatic Acinar Cells in Diabetic Mice. *Cell Rep.* **32**, 108043 (2020).

52. Moor, A. E. *et al.* Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis. *Cell* **175**, 1156-1167.e15 (2018).

53. Halpern, K. B. *et al.* Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* **542**, 352–356 (2017).

54. Moffitt, J. R. & Zhuang, X. Chapter One - RNA Imaging with Multiplexed Error-Robust Fluorescence In Situ Hybridization (MERFISH). in *Methods in Enzymology* (eds. Filonov, G. S. & Jaffrey, S. R.) vol. 572 1–49 (Academic Press, 2016).

55. wollmanlab. wollmanlab/PySpots. https://github.com/wollmanlab/PySpots.

56. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. 2020.02.02.931238 (2020) doi:10.1101/2020.02.02.931238.

57. Wasserman, L. *All of Nonparametric Statistics*. (Springer Science & Business Media, 2006).

58. Bach, F., Jenatton, R., Mairal, J. & Obozinski, G. Optimization with Sparsity-Inducing Penalties. *arXiv [cs.LG]* (2011).

59. Ioffe, S. & Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv [cs.LG]* (2015).

60. Goodfellow, I., Bengio, Y., Courville, A. & Bengio, Y. *Deep learning*. vol. 1 (MIT press Cambridge, 2016).
61. Glorot, X. & Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. in *Proceedings of the thirteenth international conference on artificial intelligence and statistics* 249–256 (2010).
62. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *arXiv [cs.LG]* (2014).
63. Somvanshi, P. R. & Venkatesh, K. V. A conceptual review on systems biology in health and diseases: from biological networks to modern therapeutics. *Syst. Synth. Biol.* **8**, 99–116 (2014).
64. Yan, J., Risacher, S. L., Shen, L. & Saykin, A. J. Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data. *Brief. Bioinform.* **19**, 1370–1381 (2018).
65. Hood, L. & Tian, Q. Systems approaches to biology and disease enable translational systems medicine. *Genomics Proteomics Bioinformatics* **10**, 181–185 (2012).
66. Chen, J. C. *et al.* Identification of causal genetic drivers of human disease through systems-level analysis of regulatory networks. *Cell* **159**, 402–414 (2014).
67. Chatterjee, S. & Ahituv, N. Gene Regulatory Elements, Major Drivers of Human Disease. *Annu. Rev. Genomics Hum. Genet.* **18**, 45–63 (2017).
68. Zhu, J. *et al.* Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nature Genetics* vol. 40 854–861 Preprint at https://doi.org/10.1038/ng.167 (2008).
69. Schadt, E. E. *et al.* An integrative genomics approach to infer causal associations between gene expression and disease. *Nat. Genet.* **37**, 710–717 (2005).
70. Zhu, J. *et al.* An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenet. Genome Res.* **105**, 363–374 (2004).

71. Haque, A., Engel, J., Teichmann, S. A. & Lönnberg, T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Medicine* vol. 9 Preprint at https://doi.org/10.1186/s13073-017-0467-4 (2017).

72. Rao, A., Barkley, D., França, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021).

73. Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A. & Murali, T. M. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods* **17**, 147–154 (2020).

74. Blencowe, M. *et al.* Network modeling of single-cell omics data: challenges, opportunities, and progresses. *Emerg Top Life Sci* **3**, 379–398 (2019).

75. Kang, Y., Thieffry, D. & Cantini, L. Evaluating the Reproducibility of Single-Cell Gene Regulatory Network Inference Algorithms. *Front. Genet.* **12**, 617282 (2021).

76. Kim, S. ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. *Commun Stat Appl Methods* **22**, 665–674 (2015).

77. Chan, T. E., Stumpf, M. P. H. & Babtie, A. C. Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures. *Cell Systems* **5**, 251 (2017).

78. Moerman, T. *et al.* GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics* **35**, 2159–2161 (2019).

79. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).

80. Schvartzman, J. M., Thompson, C. B. & Finley, L. W. S. Metabolic regulation of chromatin modifications and gene expression. *J. Cell Biol.* **217**, 2247–2259 (2018).

81. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).

82. Martin, E. & Sung, M.-H. Challenges of Decoding Transcription Factor Dynamics in Terms of Gene Regulation. *Cells* vol. 7 132 Preprint at https://doi.org/10.3390/cells7090132 (2018).

83. Matsumoto, H. *et al.* SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. *Bioinformatics* **33**, 2314–2321 (2017).

84. Qiu, X. *et al.* Inferring Causal Gene Regulatory Networks from Coupled Single-Cell Expression Dynamics Using Scribe. *Cell Syst* **10**, 265-274.e11 (2020).

85. Deshpande, A., Chu, L.-F., Stewart, R. & Gitter, A. Network inference with Granger causality ensembles on single-cell transcriptomics. *Cell Rep.* **38**, 110333 (2022).

86. Han, X. *et al.* Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **173**, 1307 (2018).

87. Morabito, S. *et al.* Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nature Genetics* vol. 53 1143–1155 Preprint at https://doi.org/10.1038/s41588-021-00894-z (2021).

88. Chen, G., Ning, B. & Shi, T. Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. *Front. Genet.* **10**, 317 (2019).

89. Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* **50**, 96 (2018).

90. Papalexi, E. *et al.* Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens. *Nat. Genet.* **53**, 322–331 (2021).

91. Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853-1866.e17 (2016).

92. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 1–13 (2008).

93. Golbeck, J. *Introduction to Social Media Investigation: A Hands-on Approach*. (Syngress, 2015).

94. Huang, J. K. *et al.* Systematic Evaluation of Molecular Networks for Discovery of Disease Genes. *Cell Syst* **6**, 484-495.e5 (2018).

95. Janeway, C. A., Jr, Travers, P., Walport, M. & Shlomchik, M. J. *Principles of innate and adaptive immunity*. (Garland Science, 2001).

96. Keren-Shaul, H. *et al.* A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* vol. 169 1276-1290.e17 Preprint at https://doi.org/10.1016/j.cell.2017.05.018 (2017).

97. Hemonnot, A.-L., Hua, J., Ulmann, L. & Hirbec, H. Microglia in Alzheimer Disease: Well-Known Targets and New Opportunities. *Front. Aging Neurosci.* **11**, 233 (2019).

98. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).

99. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).

100. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

101. Shin, J.-W. & Lee, J. C. Roles of microglial membranes in Alzheimer's disease. *Curr. Top. Membr.* **86**, 301–314 (2020).

102. Bertram, L. & Tanzi, R. E. Genome-wide association studies in Alzheimer's disease. *Hum. Mol. Genet.* **18**, R137-45 (2009).

103. Hooli, B. V. *et al.* Role of common and rare APP DNA sequence variants in Alzheimer disease. *Neurology* **78**, 1250–1257 (2012).

104. Sherva, R. *et al.* Genome-wide association study of the rate of cognitive decline in Alzheimer's disease. *Alzheimers. Dement.* **10**, 45–52 (2014).

105. Bradl, M. & Lassmann, H. Oligodendrocytes: biology and pathology. *Acta Neuropathol.* **119**, 37–53 (2010).

106. Haddad-Tóvolli, R., Dragano, N. R. V., Ramalho, A. F. S. & Velloso, L. A. Development and Function of the Blood-Brain Barrier in the Context of Metabolic Control. *Front. Neurosci.* **11**, 224 (2017).

107. Daigle, T. L. *et al.* A Suite of Transgenic Driver and Reporter Mouse Lines with Enhanced Brain-Cell-Type Targeting and Functionality. *Cell* **174**, 465-480.e22 (2018).

108. Vercruysse, P., Vieau, D., Blum, D., Petersén, Å. & Dupuis, L. Hypothalamic Alterations in Neurodegenerative Diseases and Their Relation to Abnormal Energy Metabolism. *Front. Mol. Neurosci.* **11**, 2 (2018).

109. Shen, Y. *et al.* Stimulation of the Hippocampal POMC/MC4R Circuit Alleviates Synaptic Plasticity Impairment in an Alzheimer's Disease Model. *Cell Rep.* **17**, 1819–1831 (2016).

110. Jais, A. *et al.* PNOCARC Neurons Promote Hyperphagia and Obesity upon High-Fat-Diet Feeding. *Neuron* **106**, 1009-1025.e10 (2020).

111. Jackson, C. A., Castro, D. M., Saldi, G.-A., Bonneau, R. & Gresham, D. Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments. *Elife* **9**, (2020).

112. Zhang & Zhu. Identification of key causal regulators in gene networks. *Proc. World Cong. Neurol. Surg. World Fed. Neurosurg. Soc.*

113. Nussinov, R., Jang, H., Tsai, C.-J. & Cheng, F. Review: Precision medicine and driver mutations: Computational methods, functional assays and conformational principles for interpreting cancer drivers. *PLOS Computational Biology* vol. 15 e1006658 Preprint at https://doi.org/10.1371/journal.pcbi.1006658 (2019).

114. Dugger, S. A., Platt, A. & Goldstein, D. B. Drug development in the era of precision medicine. *Nat. Rev. Drug Discov.* **17**, 183–196 (2018).

115. Emmert-Streib, F., Dehmer, M. & Haibe-Kains, B. Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Front Cell Dev Biol* **2**, 38 (2014).

116. Jansen, C. *et al.* Building gene regulatory networks from scATAC-seq and scRNA-seq using Linked Self Organizing Maps. *PLoS Comput. Biol.* **15**, e1006555 (2019).

117. Wang, C. *et al.* Integrative analyses of single-cell transcriptome and regulome using MAESTRO. *Genome Biol.* **21**, 198 (2020).

118. Zhang, R., Zhou, T. & Ma, J. Multiscale and integrative single-cell Hi-C analysis with Higashi. *Nat. Biotechnol.* **40**, 254–261 (2022).

119. Albert, F. W., Bloom, J. S., Siegel, J., Day, L. & Kruglyak, L. Genetics of trans-regulatory variation in gene expression. *Elife* **7**, (2018).

120. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

121. Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

122. Kubkowski, Mielniczuk & Teisseyre. How to Gain on Power: Novel Conditional Independence Tests Based on Short Expansion of Conditional Mutual Information. *J. Mach. Learn. Res.*

123. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).

124. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90-7 (2016).

125. Gustavsen, J. A., Pai, S., Isserlin, R., Demchak, B. & Pico, A. R. RCy3: Network biology using Cytoscape from within R. *F1000Res.* **8**, 1774 (2019).

126. Corder, E. H. *et al.* Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921–923 (1993).

127. Liu, C.-C., Liu, C.-C., Kanekiyo, T., Xu, H. & Bu, G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat. Rev. Neurol.* **9**, 106–118 (2013).

128. Shi, Y. *et al.* ApoE4 markedly exacerbates tau-mediated neurodegeneration in a mouse model of tauopathy. *Nature* **549**, 523–527 (2017).

129. Huynh, T.-P. V. *et al.* Age-dependent effects of apoE reduction using antisense oligonucleotides in a model of β-amyloidosis. *Neuron* **96**, 1013-1023.e4 (2017).

130. Ma, J., Yee, A., Brewer, H. B., Jr, Das, S. & Potter, H. Amyloid-associated proteins α1-antichymotrypsin and apolipoprotein E promote assembly of Alzheimer β-protein into filaments. *Nature* **372**, 92–94 (1994).

131. Castellano, J. M. *et al.* Human apoE isoforms differentially regulate brain amyloid-β peptide clearance. *Sci. Transl. Med.* **3**, (2011).

132. Holtzman, D. M., Herz, J. & Bu, G. Apolipoprotein E and apolipoprotein E receptors: Normal biology and roles in Alzheimer disease. *Cold Spring Harb. Perspect. Med.* **2**, a006312–a006312 (2012).

133. Kim, J. *et al.* Overexpression of low-density lipoprotein receptor in the brain markedly inhibits amyloid deposition and increases extracellular Aβ clearance. *Neuron* **64**, 632–644 (2009).

134. Choi, J. *et al.* The E3 ubiquitin ligase Idol controls brain LDL receptor expression, ApoE clearance, and Aβ amyloidosis. *Sci. Transl. Med.* **7**, 314ra184 (2015).

135. Gowrishankar, S. *et al.* Massive accumulation of luminal protease-deficient axonal lysosomes at Alzheimer's disease amyloid plaques. *Proc. Natl. Acad. Sci. U. S. A.* **112**, (2015).

136. Cras, P. *et al.* Senile plaque neurites in Alzheimer disease accumulate amyloid precursor protein. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 7552–7556 (1991).

137. Vorhees, C. V. & Williams, M. T. Morris water maze: procedures for assessing spatial and related forms of learning and memory. *Nat. Protoc.* **1**, 848–858 (2006).

138. Rustay, N., Browman, K. & Curzon, P. Cued and contextual fear conditioning for rodents. in *Methods of Behavior Analysis in Neuroscience, Second Edition* 19–37 (CRC Press, 2008).

139. Gosselin, D. *et al.* Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* **159**, 1327–1340 (2014).

140. Gosselin, D. *et al.* An environment-dependent transcriptional network specifies human microglia identity. *Science* **356**, (2017).

141. Li, Q. *et al.* Developmental heterogeneity of microglia and brain myeloid cells revealed by deep single-cell RNA sequencing. *Neuron* **101**, 207-223.e10 (2019).

142. Böttcher, C. *et al.* Human microglia regional heterogeneity and phenotypes determined by multiplexed single-cell mass cytometry. *Nat. Neurosci.* **22**, 78–90 (2019).

143. Hammond, T. R. *et al.* Single-cell RNA sequencing of microglia throughout the mouse lifespan and in the injured brain reveals complex cell-state changes. *Immunity* **50**, 253-271.e6 (2019).

144. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).

145. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [stat.ML]* (2018).

146. Saunders, A. *et al.* Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015-1030.e16 (2018).

147. van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).

148. Cao, D., Fukuchi, K.-I., Wan, H., Kim, H. & Li, L. Lack of LDL receptor aggravates learning deficits and amyloid deposits in Alzheimer transgenic mice. *Neurobiol. Aging* **27**, 1632–1643 (2006).

149. Katsouri, L. & Georgopoulos, S. Lack of LDL receptor enhances amyloid deposition and decreases glial response in an Alzheimer's disease mouse model. *PLoS One* **6**, e21880 (2011).

150. Friedman, B. A. *et al.* Diverse brain myeloid expression profiles reveal distinct microglial activation states and aspects of Alzheimer's disease not evident in mouse models. *Cell Rep.* **22**, 832–847 (2018).

151. Mrdjen, D. *et al.* High-dimensional single-cell mapping of central nervous system immune cells reveals distinct myeloid subsets in health, aging, and disease. *Immunity* **48**, 380-395.e6 (2018).

152. Olah, M. *et al.* A transcriptomic atlas of aged human microglia. *Nat. Commun.* **9**, (2018).

153. Spiller, K. J. *et al.* Microglia-mediated recovery from ALS-relevant motor neuron degeneration in a mouse model of TDP-43 proteinopathy. *Nat. Neurosci.* **21**, 329–340 (2018).

154. Krasemann, S. *et al.* The TREM2-APOE pathway drives the transcriptional phenotype of dysfunctional microglia in neurodegenerative diseases. *Immunity* **47**, 566-581.e9 (2017).

155. Colonna, M. & Butovsky, O. Microglia function in the central nervous system during health and neurodegeneration. *Annu. Rev. Immunol.* **35**, 441–468 (2017).

156. Yeh, F. L., Wang, Y., Tom, I., Gonzalez, L. C. & Sheng, M. TREM2 binds to apolipoproteins, including APOE and CLU/APOJ, and thereby facilitates uptake of amyloid-beta by microglia. *Neuron* **91**, 328–340 (2016).

157. Chitu, V. & Stanley, E. R. Colony-stimulating factor-1 in immunity and inflammation. *Curr. Opin. Immunol.* **18**, 39–48 (2006).

158. Hao, A.-J., Dheen, S. T. & Ling, E.-A. Expression of macrophage colony-stimulating factor and its receptor in microglia activation is linked to teratogen-induced neuronal damage. *Neuroscience* **112**, 889–900 (2002).

159. Mitrasinovic, O. M., Vincent, V. A. M., Simsek, D. & Murphy, G. M., Jr. Macrophage colony stimulating factor promotes phagocytosis by murine microglia. *Neurosci. Lett.* **344**, 185–188 (2003).

160. Gao, J. *et al.* The E3 ubiquitin ligase IDOL regulates synaptic ApoER2 levels and is important for plasticity and learning. *Elife* **6**, (2017).

161. Beffert, U. *et al.* ApoE receptor 2 controls neuronal survival in the adult brain. *Curr. Biol.* **16**, 2446–2452 (2006).

162. Arneson, D. *et al.* Single cell molecular alterations reveal target cells and pathways of concussive brain injury. *Nat. Commun.* **9**, 3894 (2018).

163. Poon, K., Barson, J. R., Fagan, S. E. & Leibowitz, S. F. Developmental changes in embryonic hypothalamic neurons during prenatal fat exposure. *Am. J. Physiol. Endocrinol. Metab.* **303**, E432–E441 (2012).

164. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).

165. Chung, N. C. & Storey, J. D. Statistical significance of variables driving systematic variation in high-dimensional data. *Bioinformatics* **31**, 545–554 (2015).

166. Romanov, R. A. *et al.* Molecular interrogation of hypothalamic organization reveals distinct dopamine neuronal subtypes. *Nat. Neurosci.* **20**, 176–188 (2017).

167. Chen, R., Wu, X., Jiang, L. & Zhang, Y. Single-cell RNA-seq reveals hypothalamic cell diversity. *Cell Rep.* **18**, 3227–3241 (2017).

168. Campbell, J. N. *et al.* A molecular census of arcuate hypothalamus and median eminence cell types. *Nat. Neurosci.* **20**, 484–496 (2017).

169. Shivkumar, K. *et al.* Clinical neurocardiology defining the value of neuroscience-based cardiovascular therapeutics. *J. Physiol.* **594**, 3911–3954 (2016).

170. Ardell, J. L. *et al.* Translational neurocardiology: preclinical models and cardioneural integrative aspects. *J. Physiol.* **594**, 3877–3909 (2016).

171. Hanna, P. *et al.* Cardiac neuroanatomy - Imaging nerves to define functional control. *Auton. Neurosci.* **207**, 48–58 (2017).

172. Rajendran, P. S. *et al.* Identification of peripheral neural circuits that regulate heart rate using optogenetic and viral vector strategies. *Nat. Commun.* **10**, 1944 (2019).

173. Ajijola, O. A. *et al.* Inflammation, oxidative stress, and glial cell activation characterize stellate ganglia from humans with electrical storm. *JCI Insight* **2**, (2017).

174. Ajijola, O. A. *et al.* Remodeling of stellate ganglion neurons after spatially targeted myocardial infarction: Neuropeptide and morphologic changes. *Heart Rhythm* **12**, 1027–1035 (2015).

175. Han, S. *et al.* Electroanatomic remodeling of the left stellate ganglion after myocardial infarction. *J. Am. Coll. Cardiol.* **59**, 954–961 (2012).

176. Alston, E. N. *et al.* Cardiac ischemia-reperfusion regulates sympathetic neuropeptide expression through gp130-dependent and independent mechanisms. *Neuropeptides* **45**, 33–42 (2011).

177. Gardner, R. T. *et al.* Targeting protein tyrosine phosphatase σ after myocardial infarction restores cardiac sympathetic innervation and prevents arrhythmias. *Nat. Commun.* **6**, 6235 (2015).

178. Yancy, C. W. *et al.* 2016 ACC/AHA/HFSA focused update on new pharmacological therapy for heart failure: An update of the 2013 ACCF/AHA guideline for the management of heart failure. *J. Am. Coll. Cardiol.* **68**, 1476–1488 (2016).

179. Al-Khatib, S. M. *et al.* 2017 AHA/ACC/HRS guideline for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: Executive summary. *Heart Rhythm* **15**, e190–e252 (2018).

180. Gardner, R. T., Ripplinger, C. M., Myles, R. C. & Habecker, B. A. Molecular mechanisms of sympathetic remodeling and arrhythmias. *Circ. Arrhythm. Electrophysiol.* **9**, e001359 (2016).

181. Ajijola, O. A. *et al.* Bilateral cardiac sympathetic denervation for the management of electrical storm. *J. Am. Coll. Cardiol.* **59**, 91–92 (2012).

182. Meng, L., Tseng, C.-H., Shivkumar, K. & Ajijola, O. Efficacy of stellate ganglion blockade in managing electrical storm: A systematic review. *JACC Clin. Electrophysiol.* **3**, 942–949 (2017).

183. Bourke, T. *et al.* Neuraxial modulation for refractory ventricular arrhythmias: value of thoracic epidural anesthesia and surgical left cardiac sympathetic denervation. *Circulation* **121**, 2255–2262 (2010).

184. Mets, T., De Bock, V. & Praet, J.-P. First-dose hypotension, ACE inhibitors, and heart failure in the elderly. *Lancet* **339**, 1487 (1992).

185. Rich, M. W. Pharmacotherapy of heart failure in the elderly: adverse events. *Heart Fail. Rev.* **17**, 589–595 (2012).

186. Tang, W. H. W. & Maroo, A. PPARgamma agonists: safety issues in heart failure. *Diabetes Obes. Metab.* **9**, 447–454 (2007).

187. Lisi, F. *et al.* Mineralcorticoid receptor antagonist withdrawal for hyperkalemia and mortality in patients with heart failure. *Cardiorenal Med.* **10**, 145–153 (2020).

188. Sica, D. A. & Black, H. R. Angioedema in heart failure: occurrence with ACE inhibitors and safety of angiotensin receptor blocker therapy. *Congest. Heart Fail.* **8**, 334–41, 345 (2002).

189. Furlan, A. *et al.* Visceral motor neuron diversity delineates a cellular basis for nipple- and pilo-erection muscle control. *Nature Neuroscience* vol. 19 1331–1340 Preprint at https://doi.org/10.1038/nn.4376 (2016).

190. Tan, C. M. J. *et al.* The role of neuropeptide Y in cardiovascular health and disease. *Front. Physiol.* **9**, 1281 (2018).

191. Ajijola, O. A. *et al.* Coronary sinus neuropeptide Y levels and adverse outcomes in patients with stable chronic heart failure. *JAMA Cardiol.* **5**, 318–325 (2020).

192. McMullen, J. R. *et al.* Protective effects of exercise and phosphoinositide 3-kinase(p110alpha) signaling in dilated and hypertrophic cardiomyopathy. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 612–617 (2007).

193. Buerger, A. *et al.* Dilated cardiomyopathy resulting from high-level myocardial expression of Cre-recombinase. *J. Card. Fail.* **12**, 392–398 (2006).

194. Yang, K.-C., Jay, P. Y., McMullen, J. R. & Nerbonne, J. M. Enhanced cardiac PI3Kα signalling mitigates arrhythmogenic electrical remodelling in pathological hypertrophy and heart failure. *Cardiovasc. Res.* **93**, 252–262 (2012).

195. Colucci, W. S. *et al.* Carvedilol inhibits clinical progression in patients with mild symptoms of heart failure. US Carvedilol Heart Failure Study Group. *Circulation* **94**, 2800–2806 (1996).

196. Cuevas, J. Molecular mechanisms of dysautonomia during heart failure. Focus on "Heart failure-induced changes of voltage-gated Ca2+ channels and cell excitability in rat cardiac postganglionic neurons." *American journal of physiology. Cell physiology* vol. 306 C121-2 (2014).

197. Floras, J. S. Sympathetic nervous system activation in human heart failure: clinical implications of an updated model. *J. Am. Coll. Cardiol.* **54**, 375–385 (2009).

198. Packer, M. The neurohormonal hypothesis: a theory to explain the mechanism of disease progression in heart failure. *J. Am. Coll. Cardiol.* **20**, 248–254 (1992).

199. Schwartz, P. J. & De Ferrari, G. M. Sympathetic-parasympathetic interaction in health and disease: abnormalities and relevance in heart failure. *Heart Fail. Rev.* **16**, 101–107 (2011).

200. Yoshie, K. *et al.* Cardiac TRPV1 afferent signaling promotes arrhythmogenic ventricular remodeling after myocardial infarction. *JCI Insight* **5**, (2020).

201. Tatemoto, K., Carlquist, M. & Mutt, V. Neuropeptide Y--a novel brain peptide with structural similarities to peptide YY and pancreatic polypeptide. *Nature* **296**, 659–660 (1982).

202. Kalla, M. *et al.* The cardiac sympathetic co-transmitter neuropeptide Y is pro-arrhythmic following ST-elevation myocardial infarction despite beta-blockade. *Eur. Heart J.* **41**, 2168–2179 (2020).

203. Yan, C. *et al.* Peripheral-specific Y1 receptor antagonism increases thermogenesis and protects against diet-induced obesity. *Nat. Commun.* **12**, 2622 (2021).

204. Engström Ruud, L., Pereira, M. M. A., de Solis, A. J., Fenselau, H. & Brüning, J. C. NPY mediates the rapid feeding and glucose metabolism regulatory functions of AgRP neurons. *Nat. Commun.* **11**, 442 (2020).

205. Pedrazzini, T. *et al.* Cardiovascular response, feeding behavior and locomotor activity in mice lacking the NPY Y1 receptor. *Nat. Med.* **4**, 722–726 (1998).

206. Kuo, L. E. *et al.* Neuropeptide Y acts directly in the periphery on fat tissue and mediates stress-induced obesity and metabolic syndrome. *Nat. Med.* **13**, 803–811 (2007).

207. Marsh, D. J., Hollopeter, G., Kafer, K. E. & Palmiter, R. D. Role of the Y5 neuropeptide Y receptor in feeding and obesity. *Nat. Med.* **4**, 718–721 (1998).

208. Bryant, S. M. & Hart, G. Effects of neuropeptide Y on L-type calcium current in guinea-pig ventricular myocytes. *Br. J. Pharmacol.* **118**, 1455–1460 (1996).

209. Protas, L. *et al.* Neuropeptide Y is an essential in vivo developmental regulator of cardiac ICa,L. *Circ. Res.* **93**, 972–979 (2003).

210. Pellieux, C. *et al.* Neuropeptide Y (NPY) potentiates phenylephrine-induced mitogen-activated protein kinase activation in primary cardiomyocytes via NPY Y5 receptors. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 1595–1600 (2000).

211. Kluge, N. *et al.* Rapid measurement of cardiac neuropeptide dynamics by capacitive immunoprobe in the porcine heart. *Am. J. Physiol. Heart Circ. Physiol.* **320**, H66–H76 (2021).

212. Qin, Y.-Y. *et al.* Neuropeptide Y attenuates cardiac remodeling and deterioration of function following myocardial infarction. *Mol. Ther.* **30**, 881–897 (2022).

213. Davis, H., Herring, N. & Paterson, D. J. Downregulation of M current is coupled to membrane excitability in sympathetic neurons before the onset of hypertension. *Hypertension* **76**, 1915–1923 (2020).

214. Marti, C. N. *et al.* Medication dosing for heart failure with reduced ejection fraction - opportunities and challenges. *Eur. J. Heart Fail.* **21**, 286–296 (2019).

215. Allen, N. J. & Lyons, D. A. Glia as architects of central nervous system formation and function. *Science* **362**, 181–185 (2018).

216. Allen, N. J. & Eroglu, C. Cell Biology of Astrocyte-Synapse Interactions. *Neuron* **96**, 697–708 (2017).

217. Hanani, M. Satellite glial cells: More than just rings around the neuron. *Neuron Glia Biology* **6**, 1–2 (2010).

218. Hanani, M. Satellite glial cells in sympathetic and parasympathetic ganglia: In search of function. *Brain Research Reviews* **64**, 304–327 (2010).

219. Huang, L. Y. M., Gu, Y. & Chen, Y. Communication between neuronal somata and satellite glial cells in sensory ganglia. *Glia* **61**, 1571–1581 (2013).

220. Pannese, E. *The Satellite Cells of the Sensory Ganglia*. vol. 65 (Springer Berlin Heidelberg, 1981).

221. Donegan, M., Kernisant, M., Cua, C., Jasmin, L. & Ohara, P. T. Satellite glial cell proliferation in the trigeminal ganglia after chronic constriction injury of the infraorbital nerve. *Glia* **61**, 2000–2008 (2013).

222. Shinder, V. *et al.* Structural basis of sympathetic-sensory coupling in rat and human dorsal root ganglia following peripheral nerve injury. *Journal of Neurocytology* **28**, 743–761 (1999).

223. Pannese, E., Ledda, M., Cherkas, P. S., Huang, T. Y. & Hanani, M. Satellite cell reactions to axon injury of sensory ganglion neurons: Increase in number of gap junctions and formation of bridges connecting previously separate perineuronal sheaths. *Anatomy and Embryology* **206**, 337–347 (2003).

224. Pannese, E. The structure of the perineuronal sheath of satellite glial cells (SGCs) in sensory ganglia. Preprint at (2010).

225. Woodham, P., Anderson, P. N., Nadim, W. & Turmaine, M. Satellite cells surrounding axotomised rat dorsal root ganglion cells increase expression of a GFAP-like protein. *Neuroscience Letters* **98**, 8–12 (1989).

226. Blum, E., Procacci, P., Conte, V., Sartori, P. & Hanani, M. Long term effects of lipopolysaccharide on satellite glial cells in mouse dorsal root ganglia. *Experimental Cell Research* **350**, 236–241 (2017).

227. Agulhon, C. *et al.* Modulation of the autonomic nervous system and behaviour by acute glial cell G_q protein-coupled receptor activation *in vivo*. *The Journal of Physiology* **591**, 5599–5609 (2013).

228. Xie, A. X., Lee, J. J. & McCarthy, K. D. Ganglionic GFAP+ glial Gq-GPCR signaling enhances heart functions in vivo. *JCI Insight* **2**, e90565 (2017).

229. Basner-Tschakarjan, E. & Mingozzi, F. Cell-mediated immunity to AAV vectors, evolving concepts and potential solutions. *Frontiers in Immunology* **5**, (2014).

230. Mays, L. E. *et al.* AAV8 induces tolerance in murine muscle as a result of poor APC transduction, T cell exhaustion, and minimal MHCI upregulation on target cells. *Molecular Therapy* **22**, 28–41 (2014).

231. Canonical Correlation Analysis. in *Applied Multivariate Statistical Analysis* 321–330 (Springer Berlin Heidelberg, 2007).

232. Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P10008 (2008).

233. Bin Wang, X. *et al.* A novel primary culture method for high-purity satellite glial cells derived from rat dorsal root ganglion. *Neural Regeneration Research* **14**, 339–345 (2019).

234. Weider, M. *et al.* Elevated In Vivo Levels of a Single Transcription Factor Directly Convert Satellite Glia into Oligodendrocyte-like Cells. *PLoS Genetics* **11**, 1–21 (2015).

235. Cahoy, J. D. *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: A new resource for understanding brain development and function. *Journal of Neuroscience* **28**, 264–278 (2008).

236. Boisvert, M. M., Erikson, G. A., Shokhirev, M. N. & Allen, N. J. The Aging Astrocyte Transcriptome from Multiple Regions of the Mouse Brain. *Cell Reports* **22**, 269–285 (2018).
237. Magistri, M. *et al.* A comparative transcriptomic analysis of astrocytes differentiation from human neural progenitor cells. *European Journal of Neuroscience* **44**, 2858–2870 (2016).
238. Zhang, Y. *et al.* An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *Journal of Neuroscience* **34**, 11929–11947 (2014).
239. Jacob, C. Transcriptional control of neural crest specification into peripheral glia. *Glia* **63**, 1883–1896 (2015).
240. Johnson, M. B. *et al.* Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. *Nature Neuroscience* **18**, 637–646 (2015).
241. Pollen, A. A. *et al.* Molecular Identity of Human Outer Radial Glia during Cortical Development. *Cell* **163**, 55–67 (2015).
242. Penisson, M., Ladewig, J., Belvindrah, R. & Francis, F. Genes and Mechanisms Involved in the Generation and Amplification of Basal Radial Glial Cells. *Frontiers in Cellular Neuroscience* **13**, 381 (2019).
243. Street, K. *et al.* Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, (2018).
244. Palmer, A. L. & Ousman, S. S. Astrocytes and aging. *Frontiers in Aging Neuroscience* **10**, 337 (2018).
245. Gottschall, P. E. & Howell, M. D. ADAMTS expression and function in central nervous system injury and disorders. *Matrix Biology* **44–46**, 70–76 (2015).
246. Oberacker, T. *et al.* Enhanced expression of thioredoxin-interacting-protein regulates oxidative DNA damage and aging. *FEBS Letters* **592**, 2297–2307 (2018).
247. Miller, D. J. & Fort, P. E. Heat shock proteins regulatory role in neurodevelopment. *Frontiers in Neuroscience* **12**, 821 (2018).

248. Clark, M. A., Guillaume, G. & Pierre-Louis, H. C. Angiotensin II induces proliferation of cultured rat astrocytes through c-Jun N-terminal kinase. *Brain Research Bulletin* **75**, 101–106 (2008).

249. Latacz, A., Russell, J. A., Ocloń, E., Zubel-Lojek, J. & Pierzchala-Koziec, K. mTOR pathway - novel modulator of astrocyte activity. *Folia Biologica* **63**, 95–105 (2015).

250. Cornell, B. & Toyo-oka, K. 14-3-3 proteins in brain development: Neurogenesis, neuronal migration and neuromorphogenesis. *Frontiers in Molecular Neuroscience* **10**, 318 (2017).

251. Brunkhorst, R., Bock, H. & Derouiche, A. Reelin induces process growth in cultured astrocytes: Implication for Glia-synaptic plasticity. *Archives Italiennes de Biologie* **153**, 249–254 (2015).

252. Gatson, J. W. & Singh, M. Activation of a membrane-associated androgen receptor promotes cell death in primary cortical astrocytes. *Endocrinology* **148**, 2458–2464 (2007).

253. Qin, L., Guo, J., Zheng, Q. & Zhang, H. BAG2 structure, function and involvement in disease. *Cellular and Molecular Biology Letters* **21**, 18 (2016).

254. Cheng, Y. *et al.* Sirtuin 1 attenuates oxidative stress via upregulation of superoxide dismutase 2 and catalase in astrocytes. *Journal of Neuroimmunology* **269**, 38–43 (2014).

255. Brunne, B. *et al.* Origin, maturation, and astroglial transformation of secondary radial glial cells in the developing dentate gyrus. *Glia* **58**, 1553–1569 (2010).

256. Lechler, T. Adherens junctions and stem cells. *Sub-Cellular Biochemistry* **60**, 359–377 (2012).

257. Pieters, T. & van Roy, F. Role of cell-cell adhesion complexes in embryonic stem cell biology. *Journal of Cell Science* **127**, 2603–2613 (2014).

258. Zschocke, J. *et al.* Differential promotion of glutamate transporter expression and function by glucocorticoids in astrocytes from various brain regions. *Journal of Biological Chemistry* **280**, 34924–34932 (2005).

259. Min, L. J. *et al.* Angiotensin II and aldosterone-induced neuronal damage in neurons through an astrocyte-dependent mechanism. *Hypertension Research* **34**, 773–778 (2011).

260. Chen, X. *et al.* Melatonin Prevents Mice Cortical Astrocytes From Hemin-Induced Toxicity Through Activating PKCα/Nrf2/HO-1 Signaling in vitro. *Frontiers in Neuroscience* **13**, 760 (2019).

261. van Gestel, M. A. *et al.* Recombinant Adeno-Associated Virus: Efficient Transduction of the Rat VMH and Clearance from Blood. *PLoS ONE* **9**, e97639 (2014).

262. Mason, M. R. J. *et al.* Comparison of AAV serotypes for gene delivery to dorsal root ganglion neurons. *Molecular Therapy* **18**, 715–724 (2010).