The Geography of Gene Regulation in Alzheimer's Disease

Inferring and Analysing Spatial Gene Regulatory Networks with ScReNI and Tangram

Duco F. Lam [1]

Supervisor(s): Marcel Reinders[1], **Bram Pronk**[1], **Timo Verlaan**[1]

[1]**EEMCS, Delft University of Technology, The Netherlands**

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 21, 2026

Name of the student: Duco F. Lam
Final project course: CSE3000 Research Project
Thesis committee: Marcel Reinders, Bram Pronk, Timo Verlaan, Sicco Verwer

Abstract

Alzheimer's disease progresses in a spatially heterogeneous manner across the brain, yet how the gene regulatory networks (GRNs) of individual cells vary across space in affected tissue remains largely unstudied. This thesis investigates whether changes in the spatial location of cells in human Alzheimer's disease tissue can be related to changes in GRN structure and the activity of key regulators. Cell-specific GRNs were inferred from paired single-nucleus RNA and ATAC data using pyScReNI, a Python port of the ScReNI algorithm, and projected onto MERFISH spatial coordinates using Tangram, for four cell subtypes across ten overlapping donors from the Seattle Alzheimer's Disease Brain Cell Atlas. First, the spatial mapping was validated as a feasible proof of concept, placing 22–29% of held-out cells within 500 μm of their true location. GRN structure was found to be non-randomly organised in tissue space: no signal was present at the level of the whole network, but a subset of leading GRN components was significantly spatially autocorrelated, and the resulting GRN-defined clusters were spatially contiguous beyond chance. This spatial signal was carried by specific regulators concentrated in the neuronal subtypes and absent in the glial subtypes. The most spatially autocorrelated of these, *AC106845.1*, reached roughly 2.4 times the subtype-median spatial autocorrelation, followed by *ADAM28* and *FAM189A2*, and these regulators operated through an on and off change in activity rather than a rewiring of regulatory targets. A directional but non-significant association with disease severity was observed, which the ten-donor cohort left underpowered. These findings show that single-cell GRNs can be inferred and spatially analysed at a resolution finer than previously available methods, and that their structure is spatially organised by specific regulators in Alzheimer's disease tissue.

1 Introduction

Alzheimer's disease is the most common form of dementia, representing 60% to 80% of all cases [1]. Although the disease has been extensively studied, there are still many unanswered questions about its causes and possible treatments. Alzheimer's disease is known to progress in a spatially heterogeneous manner across brain regions, as evidenced by classical neuropathological staging systems such as Braak staging, which classifies disease progression based on the spatial spread of neurofibrillary tangles across brain regions [2]. Recent single-cell and spatial transcriptomic studies have begun to characterise this heterogeneity at the molecular level, revealing disease-associated changes that vary across brain regions and cell types, suggesting that the spatial contexts of cells may shape how Alzheimer's disease affects them [3].

Understanding which genes are actively regulated in Alzheimer's-affected cells, and by which transcription factors, may help to identify the molecular mechanisms that drive Alzheimer's pathology, and thereby reveal potential targets for drug development. To capture these regulatory processes, gene regulatory networks (GRNs) can be used to model interactions between target genes and their regulators, the most prominent of which are transcription factors (TFs)[4]. Until recently, this was mostly done at the cell-type level. However, methods such as SCENIC [5] pioneered the inference of GRNs at the level of individual cells, and more recent work has continued to develop this direction. Inferring GRNs at single-cell resolution is particularly valuable when analysing spatial contexts as it captures regulatory heterogeneity within cell populations, which cell-type-level approaches obscure by averaging over potentially distinct cellular states. A practical challenge in multi-modal single-cell GRN inference is that the two most informative data modalities, single-cell RNA sequencing (scRNA-seq), which measures gene expression, and single-cell ATAC sequencing (scATAC-seq), which measures chromatin accessibility and indicates which parts of the genome are potentially available for regulation [6], are often not measured in the same cell. To address this, Xu et al. introduced ScReNI [7], a single-cell regulatory network inference algorithm that can integrate both paired and unpaired scRNA-seq and scATAC-seq data to infer cell-specific GRNs.

Even though ScReNI enables cell-specific GRN inference from single-cell multi-omics data, the use of spatial transcriptomics data to understand the impact of spatial variation on the GRNs of individual cells has yet to be studied. Unlike conventional single-cell sequencing, spatial transcriptomics uses specialised sequencing techniques such as in situ sequencing [8] or image-based methods like MERFISH [9], to preserve the physical coordinates of cells within tissue, enabling the study of how gene expression varies across spatial locations [10]. Specifically for MERFISH [9], data is collected per *section*, a thin physical slice of brain tissue on which in situ RNA imaging is performed, simultaneously profiling thousands of cells and assigning each a 2D spatial coordinate. Given the novelty of single-cell GRN research and the suggested spatial heterogeneity in Alzheimer's disease pathology, investigating how these regulatory networks vary across spatial locations in Alzheimer's disease tissue is both relevant and promising. This study therefore asks whether changes in the spatial location of cells in human Alzheimer's disease tissue can be related to changes in GRN structure and the activity of key regulators. By addressing this question, the study aims to provide insight into how local tissue context, such as the regulatory state of spatially close cells, may shape disease-related regulatory processes.

To address this, the study proceeds in four stages. First, the reliability of cell-specific spatial GRN inference is established, asking whether GRNs annotated with cell-specific locations can be reliably inferred for cells in the selected Alzheimer's disease samples. Second, the relationship between GRN structure and spatial location is examined, investigating whether differences in GRN structure between cells relate to differences in their spatial location. Third, the observed spatial differences in GRN structure are tested for correlation with Alzheimer's disease severity. Finally, the contribution of specific regulators to spatial GRN variation is investigated, as well as which other factors correlate with spatial differentiation when no significant structural differences are found.

2 Background and Related Works

2.1 Gene regulatory networks and ScReNI

According to Maizels et al. [4], a GRN can be described as a system of molecular regulators that act on each other to achieve particular cellular outcomes. The basic form of a GRN is a set of transcription factors, that act on cis-regulatory elements of other genes. Computationally, GRNs can be modelled as a graph, where nodes consist of genes and transcription factors, and edges represent a regulatory relationship between them. An example of this representation can be found in Figure 1. Inferring these networks from biological data, however, is a central challenge in computational biology. Many algorithms have been developed to achieve this. The algorithm addressed in this paper is *ScReNI*: a novel algorithm created by Xu et al. for inferring cell-specific GRNs from paired or unpaired scRNA-seq and scATAC-seq data [7]. ScReNI's core assumption is that cells with similar molecular profiles also share similar regulatory networks, implying that the GRN of a given single cell can be inferred from the cell and its nearest neighbours in a shared analytical space. Inference of

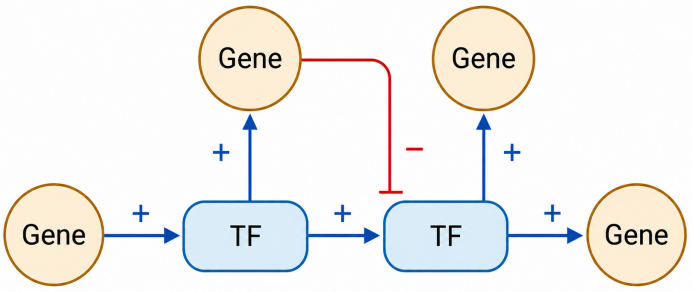

Figure 1: An example of a gene regulatory network (GRN), where nodes represent genes and transcription factors, and edges encode regulatory relationships between them.

GRNs using ScReNI operates in four steps, illustrated in Appendix A (Figure A1). First, cells from both modalities are jointly clustered into a shared analytical space. For paired data, this is done directly using Seurat's weighted nearest neighbour analysis. For unpaired data, the two modalities are first aligned before clustering using Harmony [11]. Second, each cell's neighbours are identified using a weighted k-Nearest-Neighbours algorithm to infer a cell-specific regulatory dataset. Third, gene-peak associations are established. Peaks are first linked to genes located within a 500kb window, centred on their transcription start sites. A Spearman correlation threshold of $|r| > 0.1$ is then applied across all cells, to keep only the peaks with sufficient co-variation with their associated gene. TF binding sites within these peaks are then identified via motif scanning. Finally, a random forest model is trained on the cell and its neighbours, predicting the expression of each target gene from the expression of all other genes and the accessibility of associated peaks. The importance scores returned by this random forest are aggregated into a regulatory weight of each regulator-gene pair, resulting in the construction of a GRN. Although transcription factors are the canonical regulators, ScReNI treats every gene as a candidate regulator, so the inferred networks contain regulator-gene edges beyond canonical TF-gene relationships.

2.2 Single-cell RNA-seq and ATAC-seq

Given that ScReNI's key contribution lies in the integration of multi-modal data, a more nuanced explanation of the two modalities used is warranted. Single-cell RNA sequencing (scRNA-seq) quantifies gene expression at single-cell resolution by capturing and sequencing the messenger RNA molecules present in each cell, producing a per-cell vector of expression levels across all measured genes. Historically, scRNA-seq has been the main modality used for GRN inference, with methods such as SCENIC [5] and GENIE3 [12] relying solely on gene expression data to infer GRNs. In contrast, single-cell chromatin accessibility data (scATAC-seq) is a more recently developed modality [13], with single-cell resolution only becoming available around 2015 [14]. To understand what chromatin accessibility means, it is important to understand that DNA in the cell nucleus is tightly packed around proteins. The degree to which it is packed determines whether regulatory elements are physically reachable for TF binding. scATAC-seq measures which regions of the genome are in an open, accessible state, providing a per-cell vector of accessibility scores across all measured genomic peaks. Together, these two modalities capture both what genes are being expressed and what could potentially be regulated.

2.3 Imputing spatial data using Tangram

One major drawback for investigating spatial transcriptomics data is that most spatial transcriptomics methods either exhibit low mRNA detection efficiency or contain smaller targeted gene panels [15]. *Tangram* is an algorithm developed by Biancalani et al. [16] that addresses this limitation of smaller gene panels by aligning scRNA-seq data to spatial transcriptomics data using differentiable optimisation. Specifically, it optimises an objective function that maximises the similarity between predicted and observed gene expression across shared genes, while matching the expected cell density distribution, to infer a probabilistic mapping matrix M of size $n_{\text{cells}} \times n_{\text{locations}}$, where each entry represents the probability of a given cell being assigned to a given spatial location. Once learned, this matrix can be used to effectively project the full transcriptomic resolution of single-cell data onto a spatial tissue map.

2.4 Related Works

Several methods have been developed to investigate the relation of spatial transcriptomics to GRNs; Table 1 summarises how they compare to the present work. *SpaGRN*[17] infers spatially annotated GRNs by using spatial proximity as a regulatory prior. Thus assuming that cells that are close to each other in space, will be more likely to regulate each other. This assumption is investigated here. Additionally, *SpaGRN* relies solely on gene expression data, therefore leaving out chromatin accessibility data, a limitation addressed in this paper by incorporating the ScReNI algorithm. *SCING* [18] is an algorithm developed by Littman et al., that reconstructs co-expression networks from single-cell RNA data with spatial context. It aims to link spatial organisation to regulatory structures, however, its working resolution is limited to cell-cluster level. The research conducted in this paper

Method	Data	Resolution
SpaGRN [17]	RNA	Single-cell
SCING [18]	RNA	Cell-cluster
SCRIPro [19]	RNA + ATAC	Cell-type

Table 1: Existing methods relating spatial transcriptomics to gene regulatory networks, along with used modalities and output resolution.

will address this limitation by enhancing to the single-cell resolution. Finally, *SCRIPro* [19], is an algorithm that incorporates both RNA and ATAC data to predict transcription factor activity spatially. Aiming to infer GRNs more mechanistically using multi-omics data. This algorithm, however, is also limited to cell-type resolution, thus lacking insight into single-cell resolution results. Despite these advances, SCING and SCRIPro aggregate regulatory states to the cell-type or cell-cluster level, obscuring the heterogeneity within cell populations, while SpaGRN, although single-cell, relies on RNA alone and assumes that spatial proximity drives regulation. The present work infers GRNs at single-cell resolution from both RNA and ATAC using ScReNI and maps them to spatial coordinates via Tangram, without assuming that spatial closeness implies regulatory similarity.

3 Methodology

3.1 Data

The datasets that were used in this study are sourced from the Seattle Alzheimer's Disease Brain Cell Atlas, also known as SEA-AD [20], a large-scale human brain transcriptomics dataset focusing on the Middle Temporal Gyrus (MTG) region. From the SEA-AD datasets two modalities were used, summarised in Table 2. The first one was a paired snMultiome dataset, containing paired scATAC-seq and scRNA-seq data. Each observation contained the full RNA transcriptome with gene expression information, along with chromatin accessibility locations. The dataset also includes donor metadata, including Braak, AD neuropathological change and Continuous Pseudoprogression Scores. What is not included, however, are control samples of donors without AD.

The second dataset was a spatial transcriptomics dataset, captured using MERFISH imaging technologies [9]. Due to the limitations of spatial transcriptomics mentioned in subsection 2.3, the MERFISH dataset crucially provided spatial coordinates (x, y) of each cell observation. Between the two datasets, 10 overlapping donors

Name	Cells	Modalities	Genes	Donors	Sections
snMultiome	138,118	RNA+ATAC	36,601	28	–
MERFISH	1,887,729	RNA+spatial	140	27	69
Pooled overlap	24,461	RNA+ATAC+ spatial	36,601	10	28

Table 2: SEA-AD MTG datasets used in this study and their 10-donor pooled overlap, which forms the joint analysis set for GRN inference and spatial mapping. snMultiome has no sections as it carries no spatial annotation.

exist. The donors comprise 9 males and 1 female, aged 75-90+ years with a median age of 87, spanning Braak stages 0-V and a continuous pseudoprogression score (CPS) range of 0.15 - 0.88. Per-donor specifics regarding meta-data may be inspected in Appendix B.

3.2 Selected Cell-types

To limit the scope of this study, only 4 out of the 27 existing cell subtypes were used, which are shown in Table 3: two excitatory neuronal subtypes (*L2/3 IT* and *L4 IT*) and two glial subtypes (*Astrocyte* and *Oligodendrocyte*). The snMultiome dataset was pooled to these four subtypes, reducing the set from 138,118 to 24,461 observations, with per-subtype counts in Table 3.

Subtype	Cells	Share
L2/3 IT	14,224	58.1%
L4 IT	6,012	24.6%
Oligodendrocyte	2,369	9.7%
Astrocyte	1,856	7.6%
Total	**24,461**	**100%**

Table 3: Cell counts for the four AD-relevant subtypes in the analysis set (10-donor pooled overlap, Table 2), after ScReNI GRN inference and Tangram spatial mapping. Only cells with both a cell-specific GRN and an inferred spatial location are included. Share = fraction of the 24,461-cell analysis set.

3.3 Spatial Inference

As the pooled snMultiome dataset lacks spatial coordinates, Tangram (Subsection 2.3) was used to infer a spatial location for each cell by mapping it onto the MERFISH dataset, which provides spatial coordinates but lacks the full multi-omics data required for GRN inference using ScReNI. Mapping was performed matched by donor and cell subtype, so that each subtype from each donor was mapped using only the matching subtype and donor as reference. Additionally, only the 140 native MERFISH genes have been used as the overlapping set of genes between the snMultiome and the MERFISH dataset to infer spatial locations.

Although there is an overlap in donors between the two datasets, there is no overlap in tissue sections. This means that given a single cell from a single donor, there are multiple possible tissue sections that it may be mapped to. To account for this, the tissue with the location of the single highest probability has been assigned as not only the designated tissue, but also the designated location. Mapping was performed with Tangram's $map\_cells\_to\_space(...)$ method, in cell mode with a uniform density prior. The method was run for 1000 epochs with a fixed random seed (seed $= 42$). The final result per cell is a coordinate (x, y), a mapped section, and a Tangram confidence, defined as the probability of the assigned location in the mapping matrix.

3.4 ScReNI Implementation Details

As ScReNI was originally developed in R, it was first converted to Python, resulting in the creation of pyScReNI [1], a functionally equivalent Python port of ScReNI. This was primarily done to enable a unified pipeline with Tangram, which is Python-native, and to leverage the broader single-cell Python ecosystem, including AnnData and Scanpy, for seamless data handling at scale. Code was run on the TU Delft DAIC cluster [21]. The pyScReNI source code is available at https://github.com/DelftBioinformaticsLab/bsc-screni [22]

3.5 pyScReNI parameters for GRN inference

The weighted k-Nearest-Neighbours graph was inferred using *k = 20* across the pooled dataset. This value was chosen to balance local neighbourhood resolution for the ScReNI random forest inference with computation efficiency. The gene-peak relationships were then inferred using a Spearman correlation threshold of $|r| > 0.1$, consistent with ScReNI, after which a MOODS motif scan was performed to identify TF binding sites within accessible chromatin peaks [23]. This pipeline produced 5,122 TF→peak→gene triplets, where each triplet encodes a transcription factor, the chromatin accessibility peak it binds to, and the target gene regulated via that peak. These collapse to 3,692 unique transcription factor-gene regulatory relationships, involving 26 transcription factors and 150 target genes via 206 peaks. This motif scaffold does not, however, limit which genes can regulate; it only boosts the edges it supports. The networks themselves were inferred with the weighted variant of ScReNI (wScReNI), which predicts each of the 500 genes from all other genes and the accessibility peaks per cell, yielding 24,461 cell-specific GRNs, each a weighted 500×500 matrix.

3.6 Spatial Link

With cell-specific GRNs inferred from the snMultiome dataset and spatial coordinates inferred via Tangram, the two outputs were joined using an inner join on barcodes, which are shared cell IDs across the Tangram outputs, the GRN files and the h5ad source inputs. The final output after performing the inner-join was a *cells* $\times$ *active edge* matrix of size $(24,461 \times 246,890)$. The $3,110$ missing edges were dropped to save space, as these values were zero in every single GRN that was inferred. In addition to the inferred GRN matrix, was also a baseline expression matrix of all 500 HVGs, and a metadata table containing: subtype, (x, y) coordinates, section, Tangram confidence and AD severity measurements *(BRAAK, AD Progression Score, etc.)*.

3.7 Spatial-GRN Analysis

Variogram

First, a variogram was constructed per section and cell type. Every cell's GRN was expressed as a single vector of ordered gene-pair relationships. Cell pairs were grouped into spatial distance bands, and for each distance band, the average dissimilarity score was computed using Cosine similarity as the metric, defined as $Dissimilarity = 1 - cos\_sim(i, j)$.

Moran's I

Moran's I measures whether cells that are spatially close to each other tend to have more similar values of a given feature than would be expected by chance. Specifically, it compares the deviation of each cell's value from the overall mean with the deviations observed in its spatial neighbours [24]. For a given cell subtype, it evaluates whether feature activity is clustered ($I > 0$), dispersed

[1] Developed in collaboration with fellow students: Eduard Cîmpean, Mihnea Guşu, Ivo Haršáni, and Leo Lin.

($I < 0$), or randomly distributed ($I \approx 0$) across the tissue, relative to the cell's MERFISH coordinates. It is calculated using Formula 1.

$$I = \frac{n}{W} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (1)$$

Where n is the total number of cells, x_i and x_j are the activity values for cells i and j, $\bar{x}$ is the mean activity across all cells, w_{ij} is the spatial weight between cells i and j, non-zero only when j is among i's nearest spatial neighbours, and W is the sum of all spatial weights. In this study, the activity values corresponded to the principal component (PC) scores derived from PCA of the $cell \times edge$ matrix for each subtype. Then, Moran's I was evaluated for each component using a row-standardised spatial weights matrix W, in which each cell is connected to its six nearest spatial neighbours within the same tissue section (a within-section 6-nearest-neighbour graph). Significance was assessed by 999 permutations of the component values across all cells, which asserts whether the observed autocorrelation exceeds random chance.

Spatial GRN Clustering

Additionally GRN clustering analysis was performed to assess whether GRN-defined cell groups are spatially organised. Cells were clustered by their principal components using a k-means algorithm, independently of spatial location, where the $k \in \{2, ..., 8\}$ was selected by silhouette score. After this GRN-based clustering process, the spatial neighbour-purity score was calculated as the fraction of within-section nearest-neighbour pairs whose cells were assigned to the same GRN cluster. Again, significance was assessed by 999 permutations of cluster labels. To ensure the result was done with a robust choice of k, neighbour purity was additionally computed across all $k \in \{2, ..., 8\}$.

3.8 Linking Spatial GRNs with AD

Continuous Pseudoprogression Score

Building on the spatial GRN structure identified in Subsection 3.7, the spatially structured GRN components were tested for correlation with Alzheimer's disease severity. To limit the risk of false discoveries from multiple testing [25], correlation testing was restricted to the Continuous Pseudoprogression Score (CPS) [26] as the sole AD severity metric. CPS is defined as a variable $s \in [0, 1]$ inferred from a Bayesian model over multiple neuropathological metrics used to quantify AD pathology. These include Braak stage (spatial spread of neurofibrillary tangles), Thal phase (spatial spread of amyloid-β plaques), CERAD score (neuritic plaque density in neocortex) and overall neuritic plaque density. Together, these metrics are used to order donors along a continuous disease progression trajectory [26].

Moran's I vs CPS

Per cell subtype, the global GRN principal components were derived from the $cell \times edge$ matrix, and Moran's I was computed for each principal component following the same procedure as described in Subsection 3.7. One change, however, is that instead of computing the Moran's I per PC and subtype, the Moran's I was computed per donor. This was necessary as the CPS is annotated per donor, not per section, implying that one donor is one observation instead of one section or cell. After obtaining the donor-wide spatial autocorrelation score per PC, the top 5 PCs with the highest Moran's I score, so the ones that were most spatially correlated, were selected. Per PC, the Moran's I score was then correlated with the donor's CPS using a Spearman correlation. All resulting p-values were adjusted using the Benjamini-Hochberg procedure [27] to control the False Discovery Rate (FDR), guarding against false discoveries arising from multiple simultaneous correlation tests.

3.9 Linking Spatial GRNs with Regulators

Multivariate Moran's I

To identify which regulators contribute the most to the spatial variation of GRN components, each regulator was individually analysed by computing a multivariate Moran's I for every regulator's set of outgoing regulatory edges, per cell type. Since these sub-regulation networks may change in either shape, magnitude, or both, a signed generalised Jaccard (min-max) kernel[2] was used as a similarity metric between two cells $\mathbf{a}$ and $\mathbf{b}$ using the Equation 2,

$$K(\mathbf{a}, \mathbf{b}) = \frac{\sum_i \text{sign}(a_i)\ \text{sign}(b_i)\ \min(|a_i|,\ |b_i|)}{\sum_i \max(|a_i|,\ |b_i|)}, \quad (2)$$

where a_i and b_i are the two cells' weights for target edge i. For each edge, the kernel compares the magnitudes, summing the shared minimum and dividing by the total maximum, while the sign product $\text{sign}(a_i) \times \text{sign}(b_i)$ makes matching activation/repression count as agreement and opposite signs as disagreement. As a result, two cells that regulate the same targets in the same direction with the same strength score near 1, cells with disjoint edges score near 0, and cells with systematically opposite regulation score near -1.

This kernel supplied the value-similarity for a multivariate Moran's I, computed with the same row-standardised, within-section 6-nearest-neighbour weight matrix as in the Moran's I conducted in Subsection 3.7, and pooled across sections. Regulators were ranked by spatial autocorrelation, with significance from a within-section label permutation null (999 permutations) and a Benjamini-Hochberg FDR per cell type. I_R was pooled both variance-weighted and section-balanced, and a regulator was considered *robust* only when significant under both poolings.

Component Shape vs Magnitude

The top 20 most auto-correlated regulators were then compared by their outgoing regulatory edges averaged over all active cells, between the GRN clusters found in Subsection 3.7. Following this, each difference was split into a *magnitude* component (total edge weight) and a *shape* component (the L1-normalised weights compared by cosine similarity).

Global Gradient Validation

Finally, a global activity-gradient control was computed across all 500 regulators by taking the average difference in edge weights between the two GRN clusters of Subsection 3.7, across all edges and sections. This serves as a background gradient against which the cluster differences of the top spatially autocorrelated regulators were compared. An edge-level scan across all individual regulator→gene links was additionally performed to detect any single edge whose direction reverses between clusters.

4 Results

4.1 Reliability of spatial GRN inference

Before any spatial analysis could be conducted, the underlying data and analysis set had to be chosen. Paired data was used, as the RNA and ATAC modalities are measured in the same nucleus and their alignment is therefore directly observed rather than computationally inferred through ScReNI's *Harmony* step, which improved alignment accuracy. As for the analysis set, the four cell sub-types were all chosen based on established relevance to AD pathology in the MTG. According to Gabitto *et al.* [26], L2/3 IT

[2]Here the wScReNI weights are always non-negative, so the kernel reduces to the ordinary weighted Jaccard in $[0, 1]$; the signed form is kept for generality with future signed networks.

neurons and Oligodendrocytes were identified as the most vulnerable cell types along the AD progression score trajectory, while Astrocytes were found to increase in abundance, reflecting a reactive response to disease progression. L4 IT neurons were selected as they represent one of the most abundant neuron cell types in the datasets, along with L2/3 IT neurons. With the data and subtypes fixed, each design choice in the spatial mapping pipeline could then be validated. Validation was done by holding out 5% of cells per section and subtype, re-inferring their spatial locations with Tangram, and measuring the Euclidean distance in μm between each inferred location and the cell's true MERFISH coordinate.

Gene panel selection

The first decision concerned the use of the 140 native MERFISH genes as training data for spatial inference. Only 34 of the 140 native genes overlapped with the 500 HVGs, suggesting that the remaining 106 carried predominantly sparse, low-signal expression insufficient for reliable spatial inference. This was further supported by Tangram documentation [28], which suggests that typical gene panels for training data should contain ∼100–1,000 differentially expressed genes, far more than the 34 HVGs present in the dataset. Artificial imputation of additional HVGs was therefore explored as a remedy using SpaGE [29], which imputes genes into a spatial dataset from paired single-cell data through a shared gene pool. In testing, all 466 HVGs not yet present were inferred for the MERFISH dataset, using the snMultiome dataset as reference and the 140 genes present in both datasets as the shared gene pool. The native 140-gene panel outperformed all imputed alternatives across every subtype, with SpaGE-imputed panels yielding consistently higher median errors (Table C1). The native panel was therefore retained.

Assignment method and spatial frame

With the gene panel fixed, the next decision concerned how a single spatial location should be selected from Tangram's mapping matrix. The comparison between using the cell location with the maximal probability ($argmax(...)$), and using the centroid of the top-k% cell location was performed, with results showing consistently that $argmax(...)$ results in the lowest localisation error (Table C2). Additionally, validation between section-aware and pooled matching has been performed, showing far more favourable results toward section-aware matching (Table C2).

Donor matching

A final decision concerned the mapping reference itself. This is crucial as gene expression information can vary from donor to donor, especially for the use case of this study, since different AD Severity scores may also be present. For this reason, validation between donor matching and cross-donor matching was performed, with the matched setup yielding the highest accuracy compared to non-matched alternatives, in particular cross-donor mapping and an unmatched, all-subtype reference (Table C3).

Operating configuration and accuracy

Combining these decisions, the accuracy of the final operating configuration per cell subtype is reported in Table 4. The fraction of cells placed within 500 μm is reported as the primary accuracy metric, in preference to the median localisation error (Table 4). Under this metric, 22–29% of cells fall within 500 μm of their true location, at a median error of 1.4–1.9 mm. These results define the most accurate spatial mapping achievable from the available data, on which the following spatial analyses are based.

Subtype	Med. (μm)	≤500 μm	≤1000 μm
L2/3 IT	1435	28.9%	40.3%
L4 IT	1585	27.3%	38.7%
Oligodendrocyte	1787	24.1%	33.9%
Astrocyte	1947	22.5%	31.0%

Table 4: Spatial mapping accuracy for the operating configuration (donor-matched, section-aware, argmax, native 140-gene panel). Median localisation error and the fraction of held-out cells placed within 500 μm and 1000 μm of their true position, per subtype.

4.2 GRN structure vs spatial location

Variogram

To first assess whether GRN structure varies smoothly with space, a variogram of GRN dissimilarity against spatial distance was examined per subtype. The curves are flat across all subtypes, indicating no smooth spatial gradient in whole-GRN structure, as shown in Figure 2. A variogram compares whole-GRN vectors globally, so a signal confined to a few edges would be obscured.

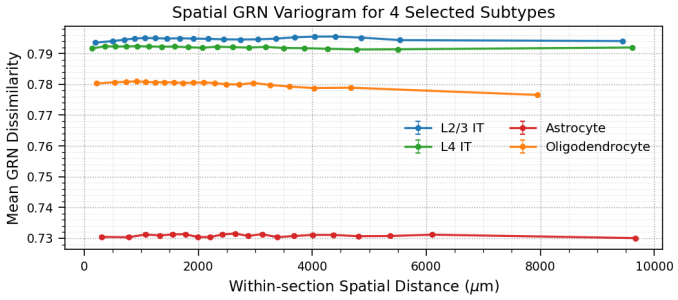

Figure 2: Empirical variograms of GRN dissimilarity ($1-\cos\_sim$) versus spatial distance band, per cell subtype. The flat curves indicate no smooth spatial gradient in whole-GRN structure.

Moran's I on GRN components

While the variogram revealed no global gradient, decomposing each subtype's $cell \times edge$ matrix into principal components and testing each for spatial autocorrelation reveals structure that the global comparison obscured. A subset of the leading GRN principal components is significantly spatially autocorrelated (Appendix D, Figure D1). This pattern holds across subtypes: most

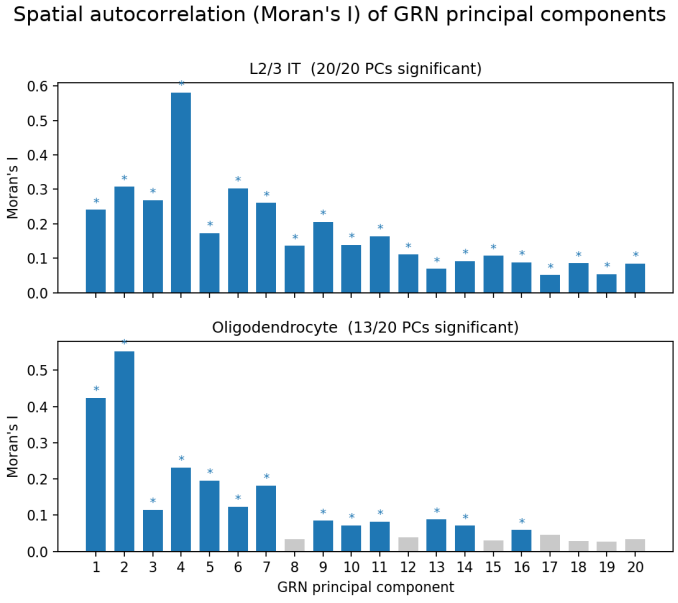

Figure 3: Spatial autocorrelation (Moran's I) of each GRN principal component for the L2/3 IT and Oligodendrocyte subtypes. Blue bars are significantly spatially autocorrelated ($p < 0.05$, 999 permutations) and marked with an asterisk, while grey bars are not, and each panel heading reports the number of significant components. Positive values indicate spatial clustering of GRN states.

leading principal components are significantly and positively autocorrelated for both the L2/3 IT and Oligodendrocyte subtypes (Figure 3), indicating that spatially proximate cells share more similar GRN states than expected under spatial randomness.

Spatial GRN clustering

To test whether this spatial organisation extends to discrete groups of cells, cells were clustered by their GRN principal components, independently of spatial location, and the spatial contiguity of the resulting clusters was assessed by neighbour-purity. The GRN-defined clusters are spatially contiguous beyond chance, with within-section neighbour-purity significant for the L2/3 IT subtype ($p = 0.001$, 999 permutations), and this result is robust across all tested values of $k \in \{2, \ldots, 8\}$. This local segregation is visible at the level of a single section (Figure 4), where GRN-defined clusters occupy partially distinct regions rather than mixing uniformly. Whether this spatial organisation relates to disease, however, cannot be determined from structure alone, motivating the comparison with Alzheimer's disease severity that follows.

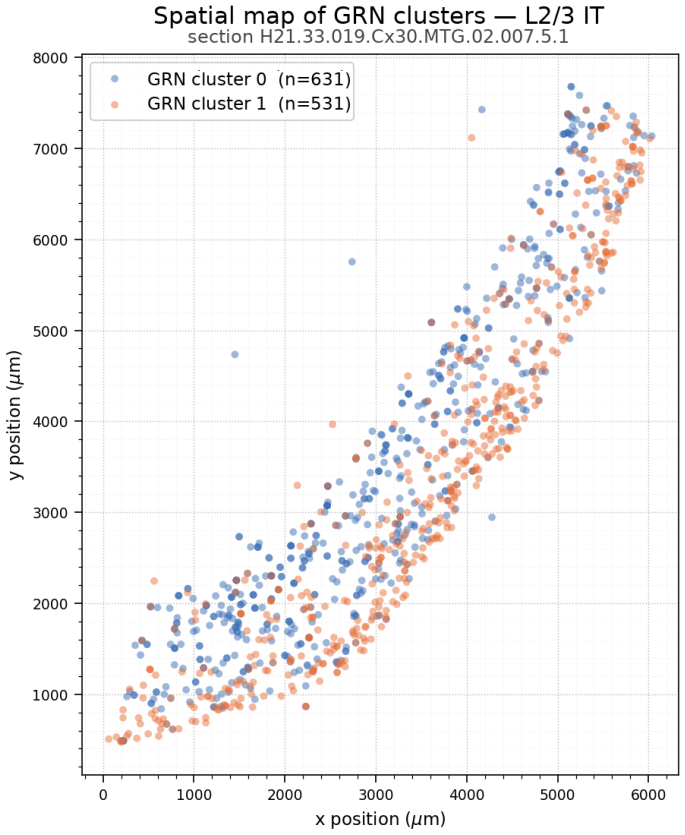

Figure 4: Spatial map of GRN-defined clusters for L2/3 IT cells in section H21.33.019.Cx30.MTG.02.007.5.1. Each point is a cell at its inferred MERFISH coordinate, coloured by its GRN-based k-means cluster ($k = 2$), where clusters are assigned from the GRN principal components independently of spatial location.

4.3 Spatial GRN differences vs AD Severity

Building on the spatial GRN structure identified above, the GRN components were tested for correlation with Alzheimer's disease severity, measured by the Continuous Pseudoprogression Score (CPS). For each subtype, the five most spatially autocorrelated components were selected independently of CPS, and their per-donor Moran's I was correlated with the donor's CPS using a Spearman correlation. No correlation survives Benjamini–Hochberg correction in any subtype. Among the five tested components, the strongest association in L2/3 IT is PC3 (Spearman $\rho = 0.57$; Figure 5), which does not survive correction and is therefore reported as a directional trend rather than a result. With the donor as the unit of replication, ten donors leave this question underpowered, and no reliable relationship between spatial GRN organisation and AD severity can be claimed. Per-PC correlations for all four subtypes are reported in Appendix E.

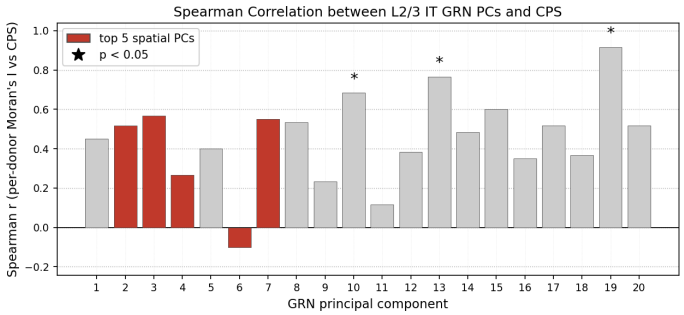

Figure 5: Per-donor Spearman correlation between each L2/3 IT GRN principal component's Moran's I and donor CPS, across the ten donors. Red bars are the five most spatially autocorrelated components (PCs 2, 3, 4, 6 and 7), selected independently of CPS; the strongest, PC3 ($\rho = 0.57$), does not survive Benjamini–Hochberg correction. Asterisks mark components reaching raw $p < 0.05$, none of which belong to the pre-selected spatial set.

Per-section validation

Due to the limited number of donors, the same analysis was repeated at per-section resolution as a corroboration check rather than an independent result: because CPS is annotated per donor, all sections of a donor share one CPS value, so treating them as independent would constitute pseudoreplication [30] and inflate the apparent significance. Significance was therefore assessed against a donor-level permutation null, shuffling CPS across donors, and the same five spatially selected components were carried forward rather than the most CPS-correlated ones, to avoid selecting on the outcome being tested. Under this framing the section-level analysis does not strengthen the signal: the L2/3 IT PC3 association weakens from $\rho = 0.57$ at donor level to $\rho = 0.34$ at section level (Appendix E.1), corroborating that no reliable relationship with severity is detectable at the current scale.

4.4 Spatial Variability in Regulators

Robust spatial regulators

To identify which regulators drive the spatial variation in GRN structure, a multivariate Moran's I was computed over the set of regulatory edges originating from each regulator, that is, its target sub-network, per subtype. Since the per-cell sample size makes nearly all p-values significant, regulators were ranked by spatial autocorrelation rather than by significance alone, and a regulator was considered robust only when significant under both variance-weighted and section-balanced pooling, guarding against a result driven by a single large section. Under this criterion, hundreds of regulators are robustly spatial in the L2/3 IT and L4 IT subtypes (366/500 and 168/500 regulators respectively), whereas none are found in the Astrocyte (0/498) or Oligodendrocyte (0/500) subtypes (Figure 6).

Regulator-specific vs global gradient

To confirm that this spatial signal reflects specific regulators rather than a global activity gradient, in which one spatial domain simply carries heavier GRNs overall, the per-regulator effect sizes were compared against a background across all 500 regulators. The background domain gradient is only $\sim$1.3$\times$, whereas the top spatially autocorrelated regulators reach 3–7$\times$ (90th–99th percentile of the background). Moreover, *LINC02248* leans in the opposite

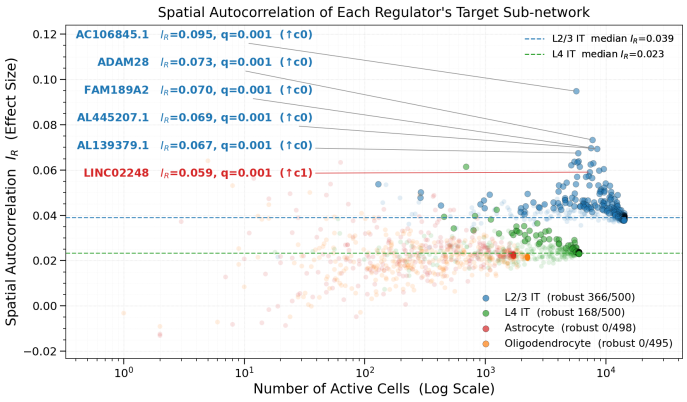

Figure 6: Spatial autocorrelation (I_R) of each regulator's target sub-network versus the number of cells in which the regulator is active (log scale), coloured by subtype. Each point is one regulator; the five most autocorrelated regulators and *LINC02248* are labelled with their Benjamini–Hochberg q-values, and the arrows (↑ 0/↑ 1) indicate the GRN cluster (Subsection 4.2) in which the regulator is stronger. Dashed lines mark the per-subtype median I_R.

spatial direction to the bulk trend, confirming that the effect is carried by specific regulators rather than a uniform background gradient. Consistent with this, the top spatial regulators concentrate their loading on a small number of GRN components rather than spreading uniformly across them (Figure 7).

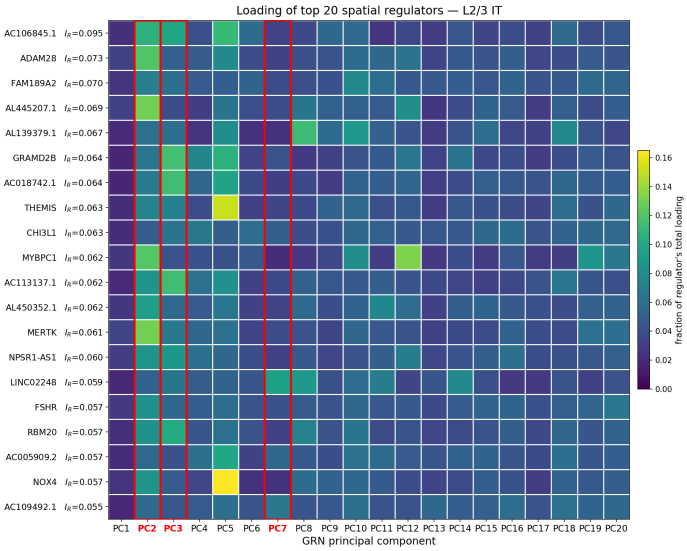

Figure 7: Distribution of each top-20 spatial regulator's total loading across the 20 GRN principal components (row-normalised), for the L2/3 IT subtype. Red-boxed columns mark the components with the strongest (directional, non-significant) CPS association (Subsection 4.3).

On/off vs rewiring

Finally, the spatial differences between the top regulators were decomposed into a magnitude component (total edge weight) and a shape component (the relative distribution of weights, compared by cosine similarity), between the GRN clusters identified in Subsection 3.7. Each spatial difference is dominated by its magnitude: when a regulator is active, the shape of its target sub-network is preserved between the two clusters (cosine similarity ∼0.8–0.9), while its total edge weight changes. An edge-level scan further finds no single regulator→gene link that reliably reverses direction across space. Spatially patterned regulation therefore operates by regulators switching on and off across space, rather than by re-targeting which genes they regulate (Figure 8).

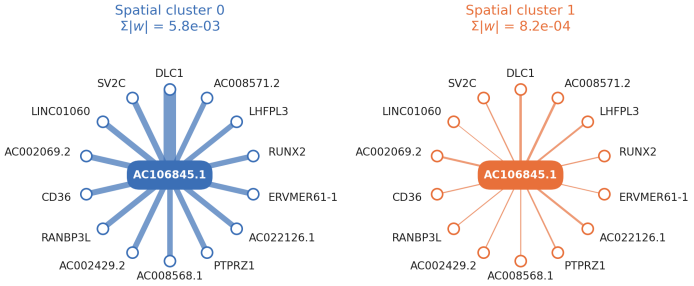

Figure 8: Mean target sub-network of regulator *AC106845.1* across the two GRN clusters (Subsection 4.2) in L2/3 IT, drawn on a shared edge-width scale. The same target genes are present in both clusters and the network shape is preserved, while the overall edge magnitude, proportional to edge width, differs markedly between clusters (mean $|w| = 0.0058$ versus 0.0008).

5 Discussion

5.1 Spatial Organisation of Gene Regulation

Following the results of Subsection 4.2, it can be concluded that GRN structure is non-randomly organised in tissue space, and is thus spatially correlated, as spatially proximate cells share more similar GRN states than chance. This organisation is not a smooth, tissue-wide gradient, as the flat variograms show, but is instead concentrated in a subset of leading GRN components and expressed as discrete, spatially contiguous groups of cells. That the GRN-defined clusters segregate in space beyond chance, robustly across the number of clusters, indicates that this spatial signal is a stable property of the regulatory state rather than an artefact of a particular partition.

This spatial organisation is, however, confined to the neuronal subtypes. Hundreds of regulators are robustly spatial in the L2/3 IT and L4 IT subtypes, whereas none are found in the Astrocyte or Oligodendrocyte subtypes, locating spatially structured regulation within the excitatory neuronal layers rather than across all cell types. This asymmetry is notable, as Oligodendrocytes were identified as a vulnerable cell type along the AD progression trajectory [26] yet show no spatial regulatory structure here, indicating that spatial patterning of regulation does not simply track cell-type vulnerability. Taken together, these findings provide molecular-level support for the premise that local tissue context shapes regulatory state, the regulatory counterpart of the spatial heterogeneity through which Alzheimer's disease is known to progress [2][3].

5.2 Mechanism of Spatial Regulation

Having established that spatial regulation is carried by specific regulators, the question remains how those regulators differ across space. The spatial differences between the top regulators are dominated by their magnitude rather than their shape: across these regulators the relative distribution of weights across the target sub-network is largely preserved between spatial clusters (cosine similarity ∼0.8–0.9), while the total edge weight changes. This is exemplified by *AC106845.1*, the most spatially autocorrelated regulator, whose mean edge weight differs by almost an order of magnitude between the two clusters (∼0.0058 versus ∼0.0008; Figure 8), leaving its target sub-network effectively switched on in one cluster and off in the other. Since a change in magnitude rather than in which targets are engaged is what dominates, it can be concluded that spatially patterned regulation operates by regulators switching on and off across space, rather than by re-targeting which genes they regulate.

That this on/off pattern reflects specific regulators, rather than a single global activity gradient in which one spatial domain simply carries heavier GRNs overall, is confirmed by the effect sizes: the background domain gradient is only $\sim$1.3$\times$, whereas the top spatially autocorrelated regulators reach 3–7$\times$. Consistent with a regulator-specific effect, *LINC02248* is switched on in the opposite spatial cluster to the bulk of regulators, against the direction the global gradient would impose.

5.3 Relationship to Disease Severity

The spatial organisation of GRN structure can be related to Alzheimer's disease severity descriptively, but no statistical significance can be attached to that relationship. Correlating the per-donor spatial autocorrelation of the GRN components with the Continuous Pseudoprogression Score yields a directional association in L2/3 IT, strongest at PC3 ($\rho = 0.57$), but no correlation survives Benjamini–Hochberg correction in any subtype, and this same PC3 association weakens to $\rho = 0.34$ when the analysis is repeated at per-section resolution. It can therefore only be reported as a directional trend rather than a result. This absence of significance is best interpreted as a lack of statistical power rather than as evidence for the absence of a relationship: with the donor as the unit of replication, ten donors leave the question underpowered, and treating the more numerous sections as independent would constitute pseudoreplication [30]. A link between spatial GRN organisation and AD severity can thus be observed as a trend, but neither confirmed nor excluded at the current scale.

5.4 Limitations

Inferred Positions and Networks

The spatial positions and the gene regulatory networks analysed in this study are both inferred rather than measured: cell locations were mapped with Tangram at a median error of 1.4–1.9 mm, with only 22–29% of the held-out cells in the validation test placed within 500 μm of their true location, and the cell-specific GRNs were inferred with pyScReNI, leaving the Continuous Pseudoprogression Score as the only directly measured variable. The results are therefore best interpreted as a proof-of-concept for integrating cell-specific GRN inference with spatial mapping, rather than as a precise spatial reconstruction. The localisation margin, on the order of hundreds of micrometres, is moreover coarse relative to cortical microstructure and bounds how fine a spatial signal can be resolved. This is mitigated in part because Tangram localises cells by expression profile, so the lower-error cells are also the most spatially informative ones on which the analysis rests.

Limitations in Data & Cohort Composition

As SEA-AD includes no neurotypical controls, all donors carry some AD pathology, so the analysis characterises spatial GRN variation across a severity gradient rather than disease-vs-control. The 9 to 1 male skew follows from modality overlap, not selection, and thus limits sex-specific interpretation. Additionally, the relatively low donor observation count (10 donors) further limits the statistical power of the donor-level analyses. Therefore no concrete claims can be made about any donor-level statistics, such as the correlations between spatial GRN variation and AD severity. Finally, given that the dataset is limited to the middle temporal gyrus, conclusions resulting from the analyses done in this thesis cannot be claimed to generalise to other regions, making claims about brain-wide processes such as the spatial spread captured by Braak staging beyond the scope of this study.

6 Conclusion and Future Work

6.1 Conclusion

This thesis set out to determine whether changes in the spatial location of cells in human Alzheimer's disease tissue can be related to changes in GRN structure and the activity of key regulators, and showed that GRN structure is non-randomly organised in tissue space and that this organisation is driven by specific regulators in the neuronal subtypes, although it could not be reliably linked to disease severity at the current scale. Spatial GRN inference using Tangram and pyScReNI was shown to be a feasible proof of concept, placing 22–29% of held-out cells within 500 μm of their true location at a median error of 1.4–1.9 mm. The spatial organisation of GRN structure was observed to be composed of GRN-defined clusters rather than smooth spatial gradients: no spatial signal was detectable at the level of the whole GRN, but a subset of its leading principal components was shown to be significantly spatially autocorrelated, and the resulting clusters were spatially contiguous beyond chance (within-section neighbour-purity $p = 0.001$, robust across $k \in \{2, \ldots, 8\}$). This spatial signal was shown to be carried by specific regulators concentrated in the neuronal subtypes (366/500 in L2/3 IT and 168/500 in L4 IT, versus none in the glial subtypes), operating through differences in the magnitude of regulation, with network shape preserved at a cosine similarity of $\sim$0.8–0.9, rather than a rewiring of regulatory targets. Finally, comparing the spatial autocorrelation of single-cell GRNs with AD severity (CPS) revealed indications of a possible relationship (strongest at L2/3 IT PC3, $\rho = 0.57$); however, with only ten donors, none survived multiple-testing correction and no conclusions can be drawn about its significance.

Altogether, these findings establish the main contributions of this thesis: showing that single-cell gene regulatory networks can be inferred and spatially analysed through an integration of ScReNI and Tangram not previously applied to Alzheimer's disease tissue, and doing so at single-cell resolution, finer than the cell-type- and cell-cluster-level methods previously available. A further contribution lies in the results of these analyses: showing that GRN structure is spatially correlated, identifying the specific regulators that drive this spatial variation, and indicating a possible correlation between GRN spatial variability and AD severity.

6.2 Future work

Given these main contributions, the suggested future work would first aim to strengthen the current validation and results. Expanding the analysis to more cell subtypes and applying the Tangram and pyScReNI pipeline to additional data sources to test how well the present findings generalise, and to validate the approach beyond the setting examined here. Incorporating AD control data and a more sex-balanced cohort would additionally enable a disease-versus-control comparison and allow sex to be eliminated as a confounding variable, neither of which was possible with the present data. A closer study of the error statistics of Tangram inference could refine this further, particularly given the observation that lower-error inferred locations tend to carry more spatial information. The results themselves also invite deeper analysis. Investigating the spatially correlated regulators discovered in this study, such as the most spatially autocorrelated regulators *AC106845.1*, *ADAM28*, and *FAM189A2*, as well as *LINC02248* (Figure 6), whose spatial activity ran counter to the dominant trend, may reveal new insights into the disease. In addition, further investigation into the relationship between AD severity and GRN spatial autocorrelation may allow firmer conclusions to be drawn regarding its significance, although this would require a larger number of donor-level observations.

7 Responsible Research

Reproducibility

To aid reproducibility, the spatial-GRN pipeline and all analysis code developed for this thesis are available at https://github.com/DucoLam/spatial-grn-alzheimers, building on the pyScReNI implementation at https://github.com/DelftBioinformaticsLab/bsc-screni. Moreover, the source data used in this experiment has been publicly available under the Seattle Alzheimer's Disease Brain Cell Atlas [20]. The entire pipeline was run in Python using open-source libraries inside a versioned Apptainer container managed by Pixi. Package versions are noted in the project's pixi.lock file. Additionally, all randomised computations were run using a set $seed = 42$ to ensure consistently reproducible results.

Limitations

While every attempt was made to results, the tight duration of the project (10 weeks) meant the scope of this thesis remains limited. This limited both the depth of analysis and the depth of validation possible in this research. Mainly, the size of error noted in Subsection 4.1 limits the results of this research. Additionally, the limited sample size of donors also limit the significance of the work. Given the fact that there are 6 viable donors for relating GRN variability with AD for Astrocytes and Oligodendrocytes, the lowest possible significance is $\frac{2}{6!} = 0.00278$ with a two-sided Spearman correlation. Including the assumption that single-cell biological data tends to be noisy, the set up does not allow for drawing conclusions relating to significance.

Use of Generative AI

Generative AI was used throughout research and development to increase productivity and improve output quality. The used models include: Claude Sonnet 4.6 and Claude Opus 4.7 and 4.8. These tools aided in writing code, handling a significant portion of the syntax and documentation workload. Beyond programming, these models aided with early-stage research and brainstorming, along with LaTeX syntax editing, grammatical checks and sentence-flow.

Ethical Considerations

All data used in this thesis are publicly available and were collected under informed consent obtained by the Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD) consortium prior to public release. No additional ethical approval is required for the use of anonymised, publicly available datasets of this kind, and no attempt was made to re-identify individual donors.

Given the sensitive nature of Alzheimer's disease research, findings are reported carefully and qualified where uncertainty exists. It should be noted that the results of this thesis are exploratory and are not intended to be interpreted as clinical or diagnostic conclusions, as further independent validation would be required before any such claims could be made. Additionally, the demographic composition of the SEA-AD donor cohort may limit how broadly the findings generalise to other populations. Methodological validation was performed to the best of our ability within the constraints of the available data, computational resources, and the ten-week duration of the project.

References

[1] Alzheimer's Association. Alzheimer's disease facts and figures, 2025.

[2] Heiko Braak and Eva Braak. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica*, 82(4):239–259, 1991.

[3] Hansruedi Mathys, Carles A. Boix, Leyla Anne Akay, Ziting Xia, Jose Davila-Velderrain, Ayesha P. Ng, Xueqiao Jiang, Ghada Abdelhady, Kyriaki Galani, Julio Mantero, Neil Band, Benjamin T. James, Sudhagar Babu, Fabiola Galiana-Melendez, Kate Louderback, Dmitry Prokopenko, Rudolph E. Tanzi, David A. Bennett, Li-Huei Tsai, and Manolis Kellis. Single-cell multiregion dissection of alzheimer's disease. *Nature*, 632(8026):858–868, 2024.

[4] Rory J. Maizels and James Briscoe. Gene regulatory networks: from correlative models to causal explanations. *Nature Reviews Genetics*, mar 2026.

[5] Sara Aibar, Carmen Bravo González-Blas, Thomas Moerman, Vân Anh Huynh-Thu, Hana Imrichova, Gert Hulselmans, Florian Rambow, Jean-Christophe Marine, Pierre Geurts, Jan Aerts, et al. Scenic: single-cell regulatory network inference and clustering. *Nature Methods*, 14(11):1083–1086, 2017.

[6] Yang Chen, Rui Liang, Yong Li, Lingli Jiang, Di Ma, Qing Luo, and Guanbin Song. Chromatin accessibility: biological functions, molecular mechanisms and therapeutic application. *Signal Transduction and Targeted Therapy*, 9(1):340, 2024.

[7] Xueli Xu, Yanran Liang, Miaoxiu Tang (), Jiongliang Wang (), Xi Wang (), Yixue Li (), and Jie Wang (). Screni: Single-cell regulatory network inference through integrating scrna-seq and scatac-seq data. *Genomics, Proteomics Bioinformatics*, 23(4):qzaf060, 08 2025.

[8] Rongqin Ke, Marco Mignardi, Alexandra Pacureanu, Jessica Svedlund, Johan Botling, Carolina Wählby, and Mats Nilsson. In situ sequencing for RNA analysis in preserved tissue and cells. *Nature Methods*, 10(9):857–860, Sep 2013.

[9] Kok Hao Chen, Eric Lubeck, Shahar Shah, and Long Cai. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science*, 348(6233):aaa6090, 2015.

[10] Boxiang Liu, Yanjun Li, and Liang Zhang. Analysis and visualization of spatial transcriptomic data. *Frontiers in Genetics*, 12:785290, January 2022.

[11] Ilya Korsunsky, Nathaniel Millard, Jean Fan, Kamil Slowikowski, Fan Zhang, Kevin Wei, Yuriy Baglaenko, Michael Brenner, Po-ru Loh, and Soumit Raychaudhuri. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods*, 16(12):1289–1296, 2019.

[12] Vân Anh Huynh-Thu, Alexandre Irrthum, Louis Wehenkel, and Pierre Geurts. Inferring regulatory networks from expression data using tree-based methods. *PLoS One*, 5(9):e12776, 2010.

[13] Jason D Buenrostro, Paul G Giresi, Lisa C Zaba, Howard Y Chang, and William J Greenleaf. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, 10(12):1213–1218, 2013.

[14] Darren A Cusanovich, Riza Daza, Andrew Adey, Hannah A Pliner, Lena Christiansen, Kevin L Gunderson, Frank J Steemers, Cole Trapnell, and Jay Shendure. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science*, 348(6237):910–914, 2015.

[15] Cameron G. Williams, Hyun Jae Lee, Takahiro Asatsuma, Roser Vento-Tormo, and Ashraful Haque. An introduction to spatial transcriptomics for biomedical research. *Genome Medicine*, 14(1):68, 2022.

[16] Tommaso Biancalani, Gabriele Scalia, Lorenzo Buffoni, Raghav Avasthi, Ziqing Lu, Aman Sanger, Neriman Tokcan, {Charles R.} Vanderburg, Åsa Segerstolpe, Meng Zhang, Inbal Avraham-Davidi, Sanja Vickovic, Mor Nitzan, Sai Ma, Ayshwarya Subramanian, Michal Lipinski, Jason Buenrostro, {Nik Bear} Brown, Duccio Fanelli, Xiaowei Zhuang, {Evan Z.} Macosko, and Aviv Regev. Deep learning and alignment of spatially resolved single-cell transcriptomes with tangram. *Nature Methods*, 18(11):1352–1362, November 2021. Publisher Copyright: © 2021, The Author(s).

[17] Y. Li, X. Liu, L. Guo, K. Han, S. Fang, X. Wan, D. Wang, X. Xu, L. Jiang, G. Fan, and M. Xu. Spagrn: Investigating spatially informed regulatory paths for spatially resolved transcriptomics data. *Cell Systems*, 16:101243, 2025.

[18] Russell Littman, Michael Cheng, Ning Wang, Chao Peng, and Xia Yang. Scing: Inference of robust, interpretable gene regulatory networks from single cell and spatial transcriptomics. *iScience*, 26(7):107124, June 2023.

[19] Pengcheng Yuan and Zhana Duren. Single-cell and spatial multiomic inference of gene regulatory networks using SCRIPro. *Bioinformatics*, 40(7):btae466, July 2024.

[20] Michael Hawrylycz, Eitan S. Kaplan, Kyle J. Travaglini, Mariano I. Gabitto, Jeremy A. Miller, Lydia Ng, Jennie L. Close, Rebecca D. Hodge, Brian Long, Tyler Mollenkopf, Shoaib Mufti, Nicole M. Gatto, Eric B. Larson, Paul K. Crane, Thomas J. Grabowski, C. Dirk Keene, and Ed S. Lein. SEA-AD: A multimodal cellular atlas and resource for Alzheimer's disease. *Nature Aging*, 4(10):1331–1334, October 2024.

[21] Delft AI Cluster (DAIC). The delft ai cluster (daic), 2024.

[22] DelftBioinformaticsLab. bsc-screni: RP applying ScReNI to SEA-AD. https://github.com/DelftBioinformaticsLab/bsc-screni, 2025. Accessed: 2026-06-01.

[23] Janne Korhonen, Petri Martinmäki, Cinzia Pizzi, Pasi Rastas, and Esko Ukkonen. Moods: fast search for position weight matrix matches in dna sequences. *Bioinformatics*, 25(23):3181–3182, 2009.

[24] Patrick Alfred Pierce Moran. Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2):17–23, 1950.

[25] Joseph P. Simmons, Leif D. Nelson, and Uri Simonsohn. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11):1359–1366, 2011.

[26] Mariano I. Gabitto, Kyle J. Travaglini, Victoria M. Rachleff, Eitan S. Kaplan, Brian Long, Jeanelle Ariza, Yi Ding, Joseph T. Mahoney, et al. Integrated multimodal cell atlas of Alzheimer's disease. *Nature Neuroscience*, 27(12):2366–2383, October 2024.

[27] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, 1995.

[28] Ziqing Lu and Tommaso Biancalani. Tutorial for spatial mapping using Tangram. https://tangram-sc.readthedocs.io/

en/latest/tutorial_sq_link.html, 2021. Tangram v0.4.0 documentation. Accessed: 2026-06-08.

[29] Tamim Abdelaal, Soufiane Mourragui, Ahmed Mahfouz, and Marcel J T Reinders. Spage: Spatial gene enhancement using scrna-seq. *Nucleic Acids Research*, 48(18):e107–e107, 10 2020.

[30] Stuart H. Hurlbert. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, 54(2):187–211, 1984.

A ScReNI workflow

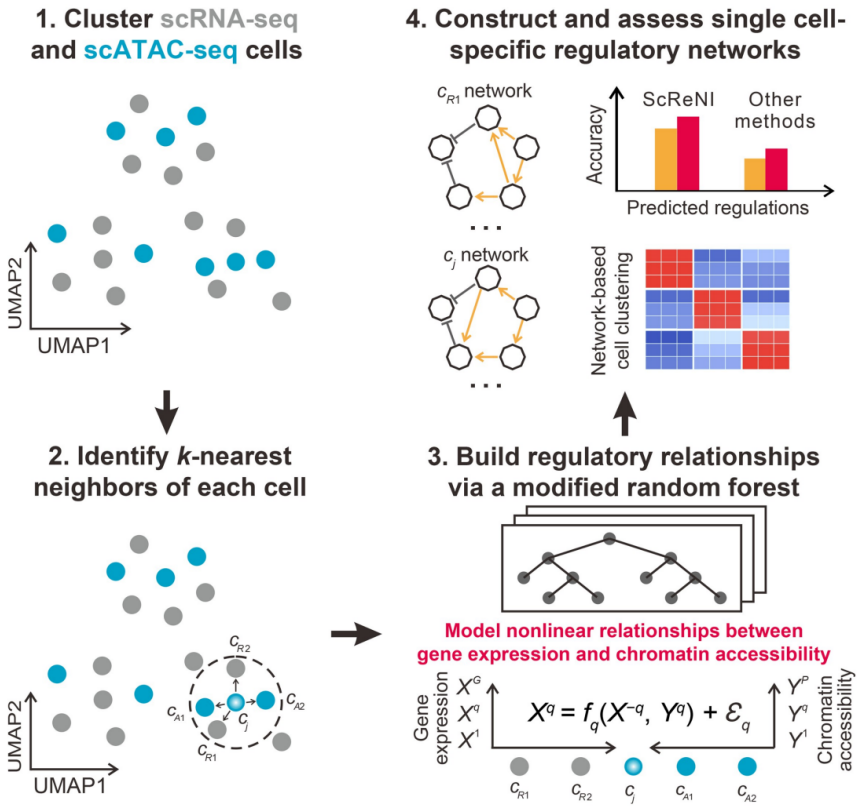

Figure A1: Workflow of ScReNI consisting of co-clustering single cells from scRNA-seq and scATAC-seq data, identifying cell neighbours via the k-nearest neighbours algorithm, calculating regulatory weights of regulators on target genes using a random forest model, and validating the resulting cell-specific regulatory networks [7].

B Donor-specific metadata

Donor ID	Sex	Age	Braak	CPS
H20.33.001	M	82	IV	0.52
H20.33.004	M	86	V	0.73
H20.33.012	F	90+	III	0.24
H20.33.015	M	88	V	0.88
H20.33.025	M	90+	V	0.59
H20.33.040	M	90+	IV	0.44
H20.33.044	M	81	V	0.36
H21.33.001	M	80	III	0.15
H21.33.006	M	90+	IV	0.84
H21.33.019	M	75	0	0.52

Table B1: Demographic and neuropathological characteristics of the 10 donors overlapping between the snMultiome and MERFISH datasets (Table 2). Sex (M/F), age at death (donors $\geq$90 years de-identified as "90+"), Braak stage (tau-based neurofibrillary staging), and Continuous Pseudo-progression Score (CPS). The cohort comprises 9 male and 1 female donor, median age 87, spanning Braak 0–V and CPS 0.15–0.88; as SEA-AD contains no neurotypical controls, all donors carry some degree of AD pathology.

C Spatial Mapping Validation

The spatial mapping (Tangram) was validated by per-section 5% holdout across four design axes: gene panel, assignment method, spatial frame, and donor matching. Tables C1–C3 report median localisation error and the fraction of held-out cells placed within 500 μm and 1000 μm of their true position.

Subtype	**Panel**	**Median (μm)**	**$\leq$500 μm**	**$\leq$1000 μm**
Astrocyte	Native 140	1947	22.5%	31.0%
Astrocyte	Imputed 100	2522	8.5%	18.2%
Astrocyte	Imputed 500	2424	8.3%	18.1%
L2/3 IT	Native 140	1435	28.9%	40.3%
L2/3 IT	Imputed 100	1848	18.6%	31.8%
L2/3 IT	Imputed 500	1939	16.4%	29.5%
L4 IT	Native 140	1585	27.3%	38.7%
L4 IT	Imputed 100	1795	19.5%	32.4%
L4 IT	Imputed 500	1907	17.3%	30.1%
Oligodendrocyte	Native 140	1787	24.1%	33.9%
Oligodendrocyte	Imputed 100	–	–	–
Oligodendrocyte	Imputed 500	–	–	–

Table C1: Effect of gene panel on spatial mapping accuracy (section-aware, argmax, donor-matched). The native 140-gene MERFISH panel outperforms the SpaGE-imputed 100- and 500-gene panels for every subtype. Oligodendrocyte imputed panels were not computed due to reference size and computation time.

		Section-aware			Pooled		
Subtype	**Method**	**Med.**	**≤500**	**≤1000**	**Med.**	**≤500**	**≤1000**
Astrocyte	argmax	1947	22.5%	31.0%	2378	21.2%	27.7%
Astrocyte	top 0.1%	1975	6.2%	19.6%	2314	3.3%	13.2%
Astrocyte	top 0.5%	1937	5.3%	18.5%	2297	3.4%	12.4%
Astrocyte	top 1%	1931	4.9%	17.8%	2305	3.3%	12.5%
Astrocyte	top 5%	1981	4.3%	16.3%	2325	3.1%	12.0%
Astrocyte	top 10%	2004	4.1%	15.6%	2327	3.0%	11.9%
L2/3 IT	argmax	1435	28.9%	40.3%	2223	24.4%	32.3%
L2/3 IT	top 0.1%	1742	10.2%	27.2%	2443	3.5%	12.0%
L2/3 IT	top 0.5%	1723	8.9%	26.2%	2444	3.1%	11.0%
L2/3 IT	top 1%	1737	8.7%	25.4%	2458	3.1%	10.7%
L2/3 IT	top 5%	1765	8.5%	24.7%	2475	2.9%	10.2%
L2/3 IT	top 10%	1775	8.4%	24.6%	2479	2.9%	10.0%
L4 IT	argmax	1585	27.3%	38.7%	2272	21.2%	28.7%
L4 IT	top 0.1%	1720	13.7%	29.8%	2121	5.0%	16.6%
L4 IT	top 0.5%	1653	11.7%	28.0%	2088	4.6%	15.0%
L4 IT	top 1%	1659	11.1%	27.8%	2087	4.4%	14.8%
L4 IT	top 5%	1659	10.8%	27.1%	2120	4.5%	14.6%
L4 IT	top 10%	1663	10.2%	26.6%	2125	4.4%	14.5%
Oligodendrocyte	argmax	1787	24.1%	33.9%	2508	17.6%	24.6%
Oligodendrocyte	top 0.1%	1828	6.8%	22.3%	2282	3.9%	13.9%
Oligodendrocyte	top 0.5%	1745	6.6%	23.2%	2276	4.0%	13.9%
Oligodendrocyte	top 1%	1736	6.9%	23.6%	2277	3.9%	13.8%
Oligodendrocyte	top 5%	1733	6.6%	23.8%	2281	3.9%	13.7%
Oligodendrocyte	top 10%	1733	6.6%	23.9%	2281	3.9%	13.8%

Table C2: Assignment method (argmax vs. top-k centroid) and spatial frame (section-aware vs. pooled across sections), native 140-gene panel, donor-matched. Values are median error in μm and the fraction of held-out cells within 500 μm and 1000 μm. Argmax with a section-aware frame gives the best within-distance precision; top-k averaging and section pooling both degrade accuracy.

Subtype	**Configuration**	**Median (μm)**	**≤500 μm**	**≤1000 μm**
Astrocyte	Donor-matched	1947	22.5%	31.0%
Astrocyte	Cross-donor	3641	1.3%	5.4%
L2/3 IT	Donor-matched	1435	28.9%	40.3%
L2/3 IT	Cross-donor	4139	1.2%	4.4%
L4 IT	Donor-matched	1585	27.3%	38.7%
L4 IT	Cross-donor	3400	1.6%	6.1%
Oligodendrocyte	Donor-matched	1787	24.1%	33.9%
Oligodendrocyte	Cross-donor	3677	1.3%	5.2%

Table C3: Donor-matched vs. cross-donor mapping (section-aware, argmax, native 140-gene panel). Cross-donor mapping uses a reference from a different donor as a negative control: median error roughly doubles and within-500 μm accuracy drops 15–25-fold, confirming the mapping is donor-specific.

D Spatial GRN correlation results

This appendix reports the figures supporting the analysis of GRN structure versus spatial location (Subsection 4.2).

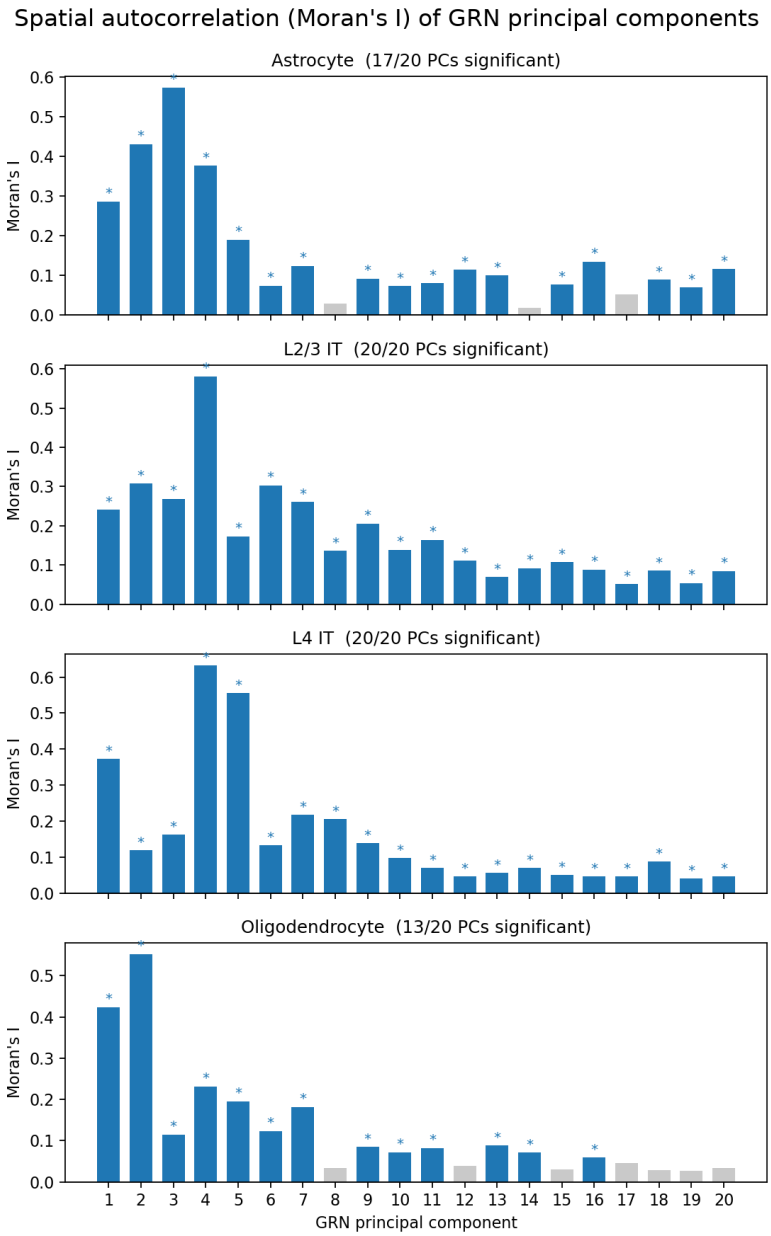

Figure D1: Moran's I per GRN principal component, by subtype. Blue bars are significant ($p < 0.05$, 999 permutations); grey bars are not. Spatial autocorrelation concentrates in the leading components.

E Spatial GRN and AD severity correlation results

Per-PC Spearman correlation, across donors, between each GRN component's per-donor Moran's I and CPS. Red bars are the five most spatially autocorrelated PCs (selected independently of CPS) and the only ones tested; grey bars are non-spatial PCs shown for context. Asterisks mark $p < 0.05$.

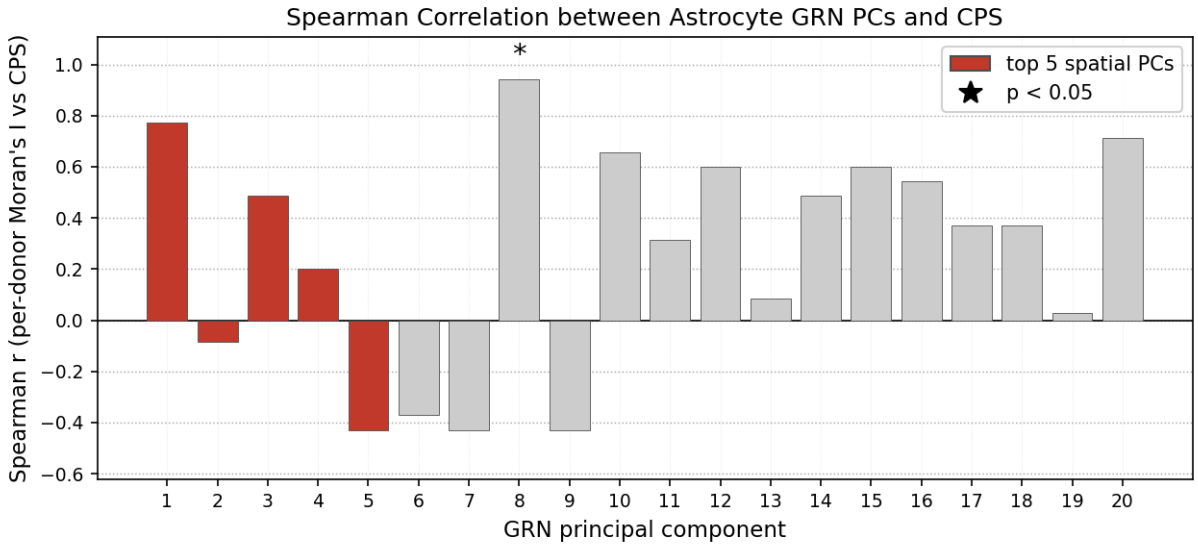

Figure E1: Astrocyte: per-PC Moran's I versus CPS.

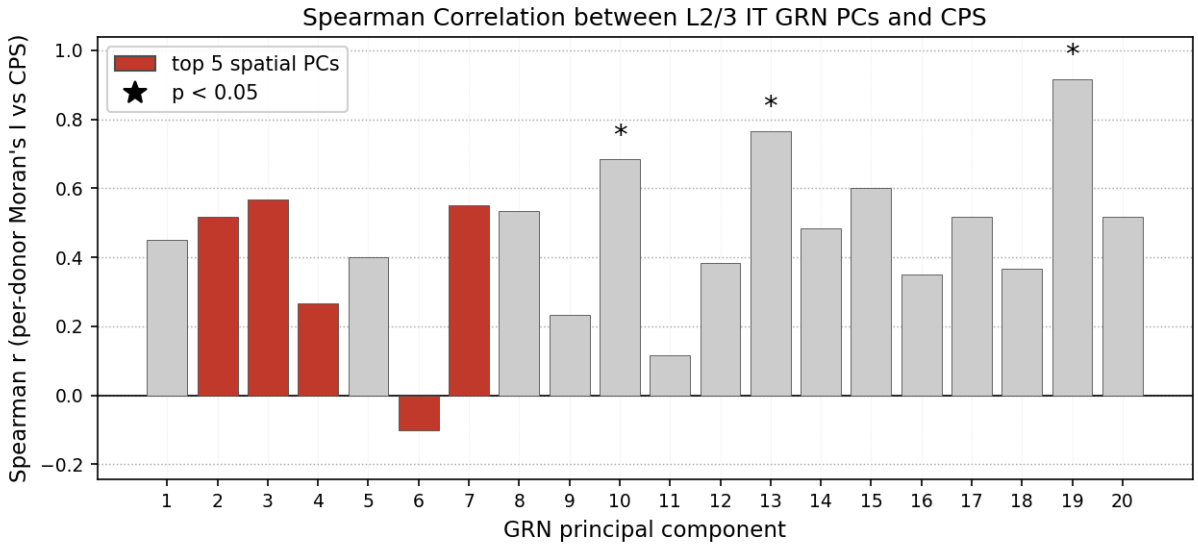

Figure E2: L2/3 IT: per-PC Moran's I versus CPS.

E.1 Per-section validation

Donor-level (solid points) versus per-section (faded points) Moran's I against CPS, for the five most spatially structured PCs per subtype. One colour per donor; sections of a donor share a CPS and so stack vertically. Solid lines show the donor-level trend, dashed lines the section-level trend.

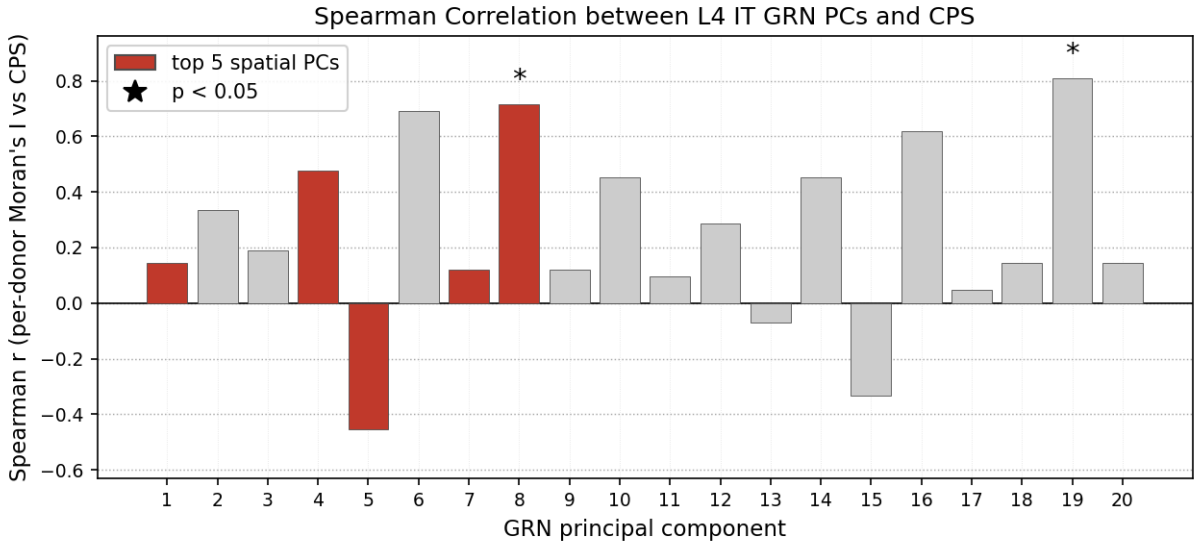

Figure E3: L4 IT: per-PC Moran's I versus CPS.

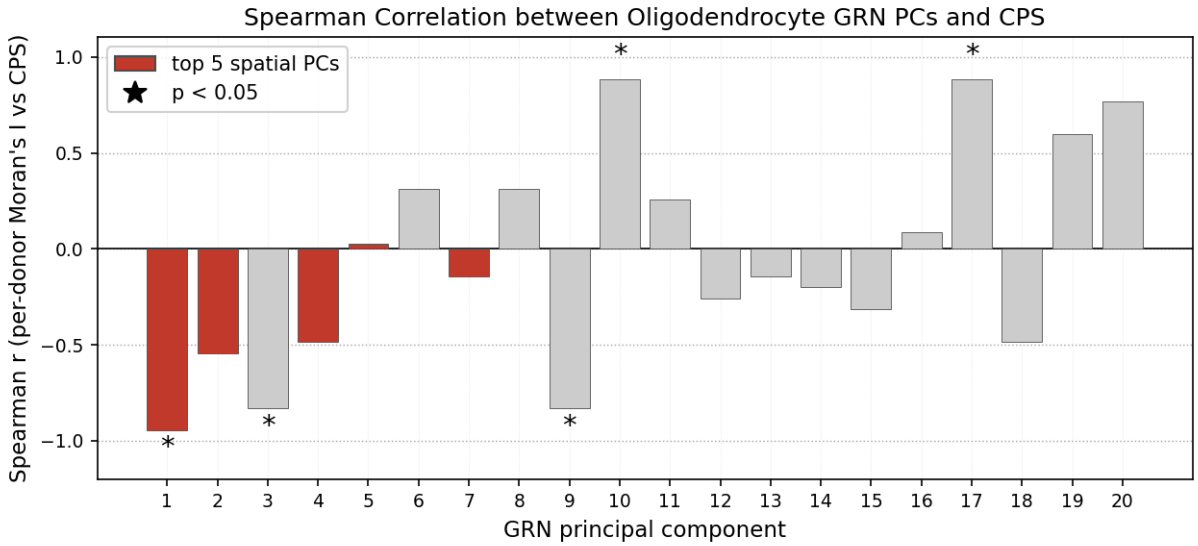

Figure E4: Oligodendrocyte: per-PC Moran's I versus CPS. PC1 shows the strongest association ($\rho = -0.94$).

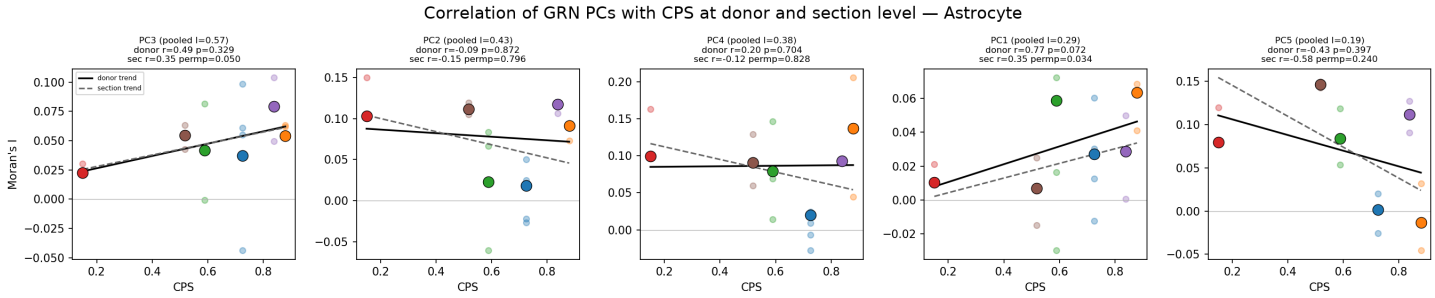

Figure E5: Astrocyte: donor- versus section-level spatial organisation against CPS.

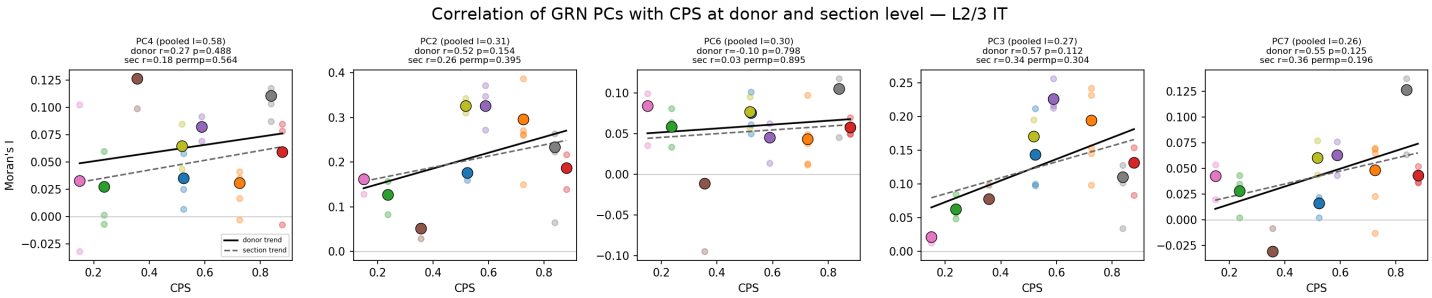

Figure E6: L2/3 IT: donor- versus section-level spatial organisation against CPS.

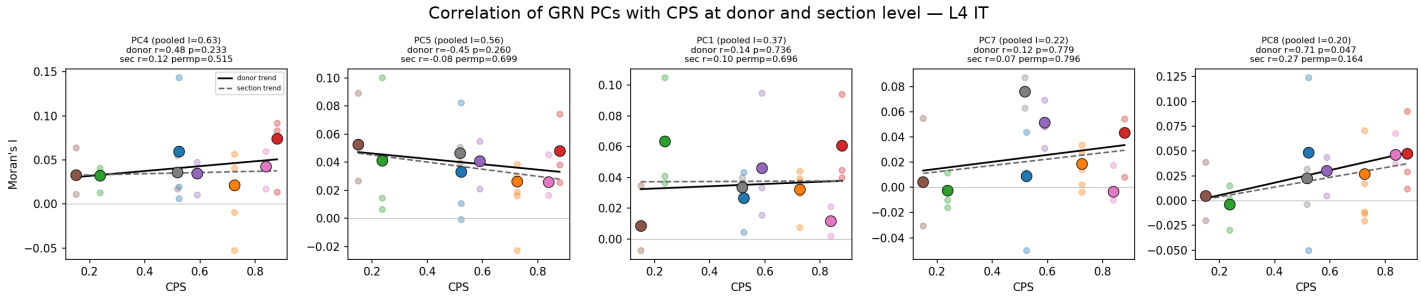

Figure E7: L4 IT: donor- versus section-level spatial organisation against CPS.

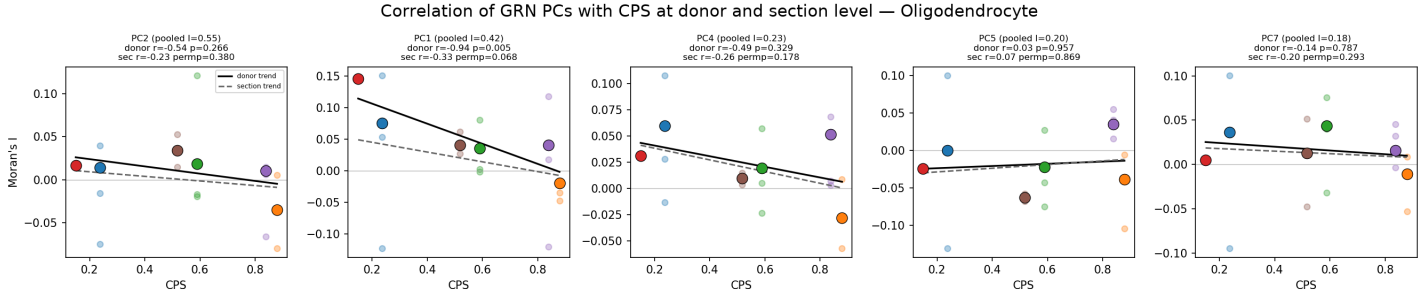

Figure E8: Oligodendrocyte: donor- versus section-level spatial organisation against CPS. PC1's donor-level association ($\rho = -0.94$) weakens at section level ($\rho = -0.33$).

F Regulator loading distribution across GRN components

This appendix reports the loading distribution of the top-20 spatial regulators across the GRN principal components for the three subtypes not shown in the main text. The corresponding figure for the L2/3 IT subtype is given in the main text (Figure 7). In each heatmap a row is one regulator and a column one GRN principal component; the colour is the fraction of that regulator's total loading carried by the component (row-normalised). Red-boxed columns mark the components with the strongest (directional, non-significant) CPS association (Subsection 4.3).

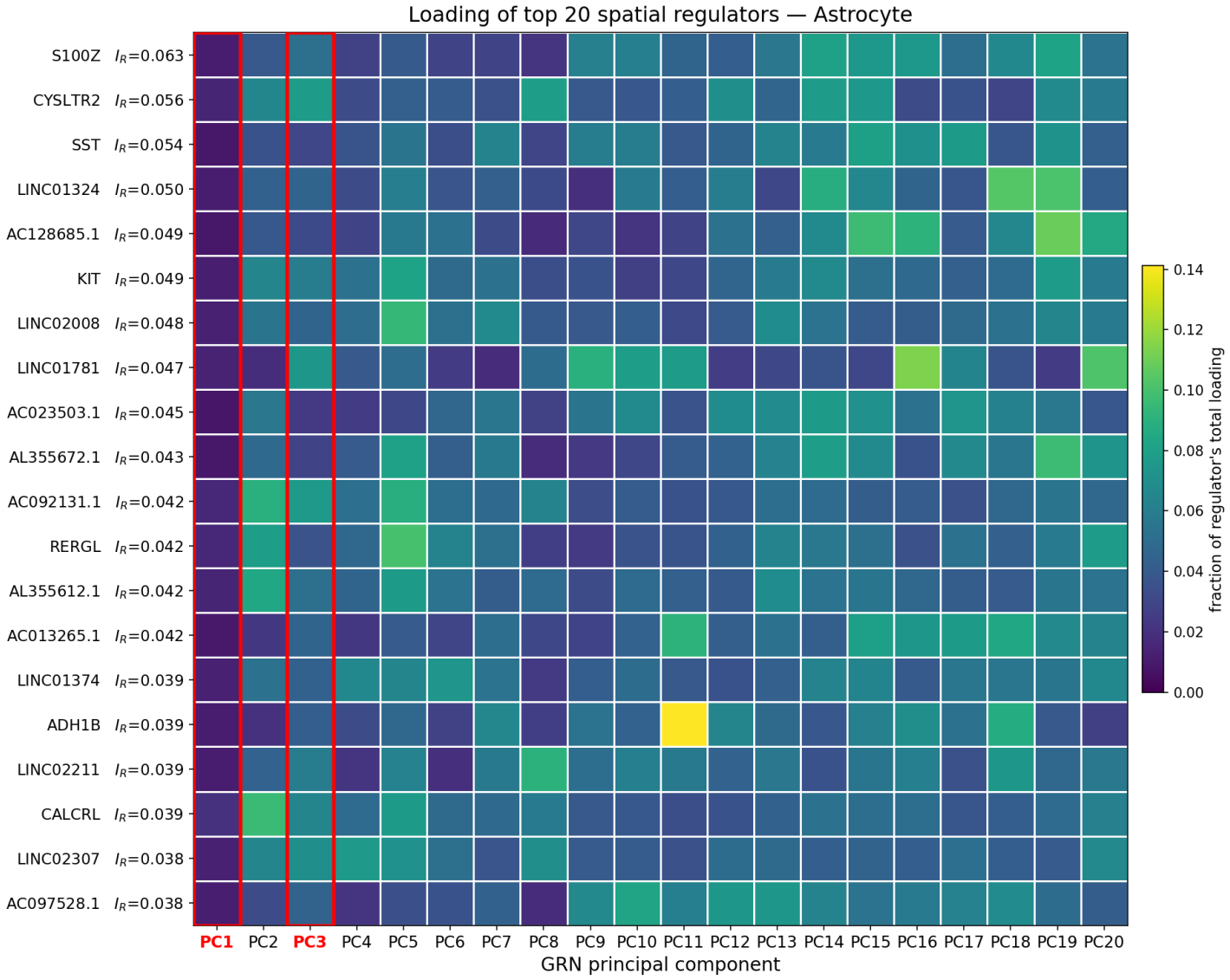

Figure F1: Astrocyte: loading distribution of the top-20 spatial regulators across the GRN principal components.

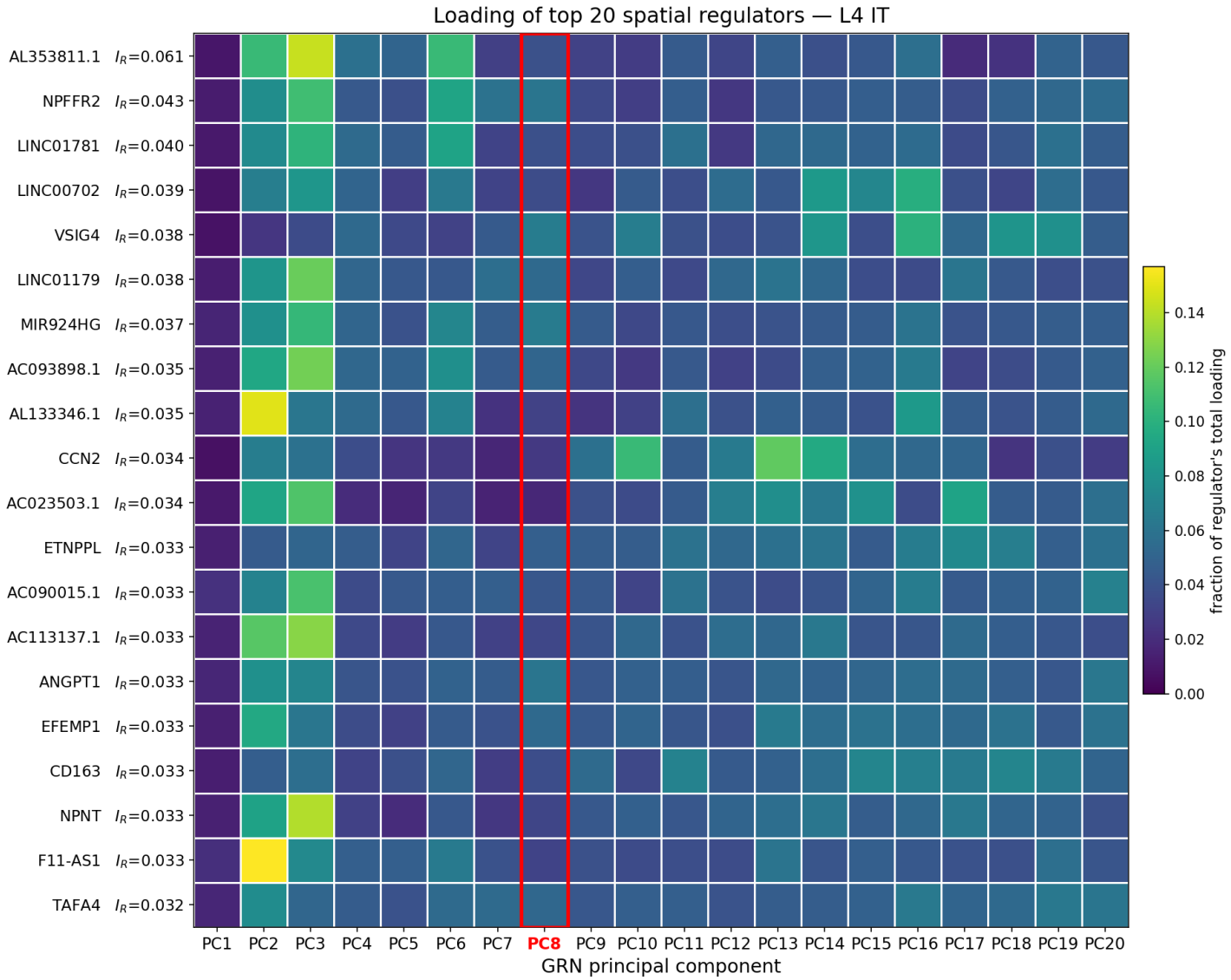

Figure F2: L4 IT: loading distribution of the top-20 spatial regulators across the GRN principal components.

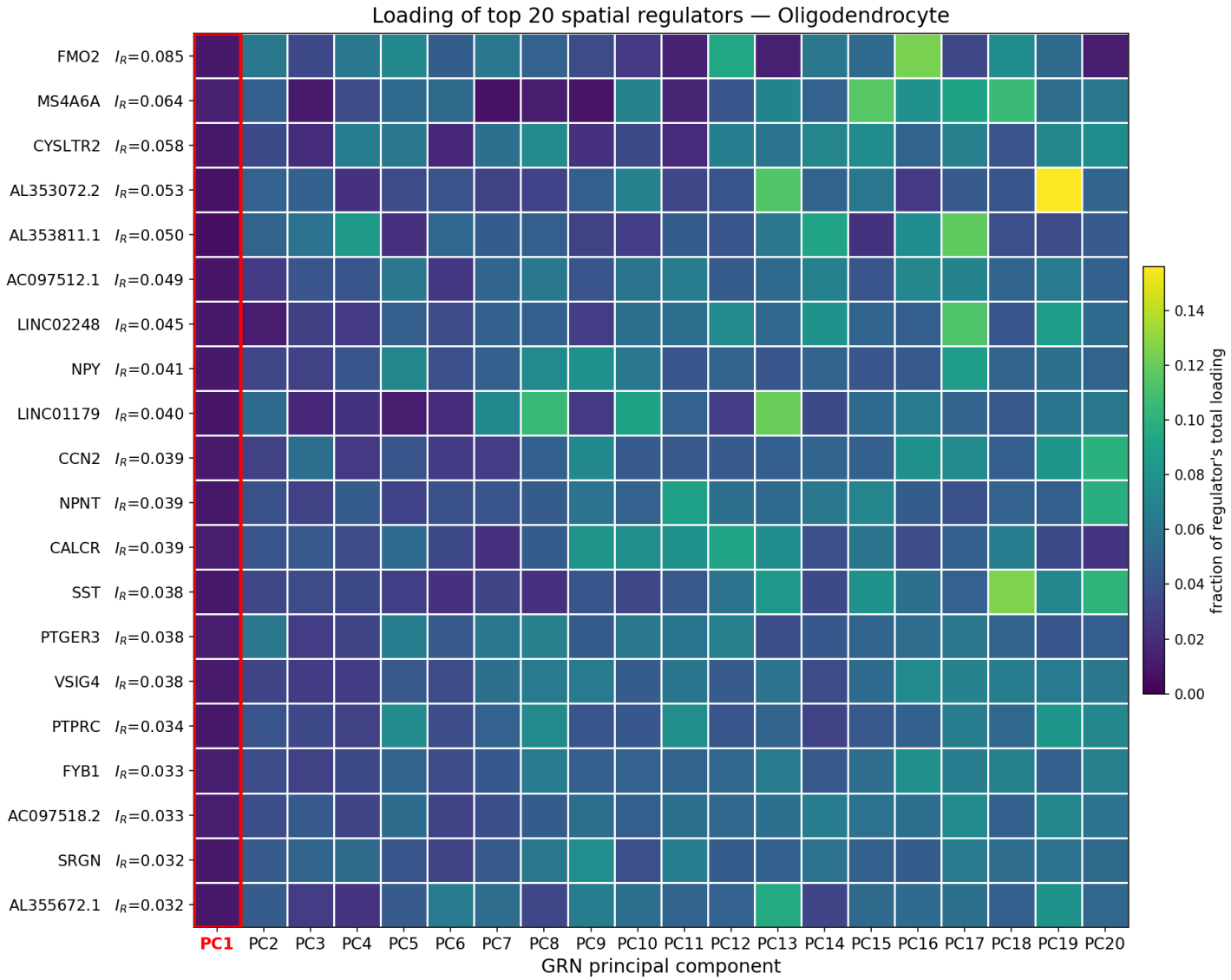

Figure F3: Oligodendrocyte: loading distribution of the top-20 spatial regulators across the GRN principal components.