Unsupervised pattern discovery in spatial gene expression atlas reveals mouse brain regions beyond established ontology

Robert Cahill[1,2,#], Yu Wang[3,#], Alex Lee[1,2], Hongkui Zeng[4], Bin Yu[3], Bosiljka Tasic[4], Reza Abbasi-Asl[1,2,*]

[1]University of California, San Francisco
[2]Weill Institute for Neurosciences
[3]University of California, Berkeley
[4]Allen Institute for Brain Science

[#]indicates equal contribution
*Corresponding author: Reza Abbasi-Asl (Reza.AbbasiAsl@ucsf.edu)

Abstract

The growth of large-scale spatial gene expression data requires new computational tools to extract major trends in gene expression in their native spatial context. Here, we describe an unsupervised and interpretable computational framework to (1) pre-process 3D spatial gene expression datasets by imputation of missing voxels, (2) identify principal patterns (PPs) of 3D spatial gene expression profiles using the stability-driven non-negative matrix factorization (staNMF) technique, and (3) systematically compare these PPs to known anatomical regions and ontology. This framework, referred to as osNMF (ontology discovery via staNMF), identifies PPs that are derived purely from thousands of 3D spatial gene expression profiles in the Allen Mouse Brain Atlas. These 3D PPs present stable and spatially coherent regions of the mouse brain, potentially without human labor and bias. We demonstrate that osNMF PPs offer new brain patterns that are highly correlated with combinations of expert-annotated brain regions, while also identifying a unique ontology based purely on spatial gene expression data. Compared to principal component analysis (PCA) and other clustering algorithms, our PPs exhibit better spatial coherence, more accurately match expert labeling and are more stable across multiple bootstrapped simulations. We also used osNMF to define marker genes and build putative spatial gene interaction networks. Our findings highlight the capability of osNMF to rapidly generate new atlases from a large set of spatial gene expression data without supervision and uncover novel relationships between brain regions that were difficult to discern using conventional manual approaches.

Introduction

In the past decade, unsupervised explorations of large-scale single-cell transcriptomics datasets enabled by machine learning tools have offered an unbiased definition of cell types—groups of cells with similar gene expression patterns[1–4]. To obtain these data, however, cells are usually dissociated and information on their precise location in the tissue and organ is lost. The precise locations of cell types and their combinations usually define distinct functional

regions of different organs. Therefore, new machine learning tools are needed to investigate spatial gene expression data to extract major trends in gene expression in their native spatial context. These newly defined patterns will likely correlate with specific cell types or cell type combinations and are likely to reflect organ regions with distinct functions. Perturbations in gene expression in specific organs or organ regions are associated with a variety of diseases[5–9]; the tools defined here apply to both healthy and diseased tissues and may point to gene and organ function and provide hypotheses for disease mechanisms.

Different regions within an organ are often defined by manual/expert-guided judgment based on various but frequently limited data. For the adult mouse brain, the Allen Common Coordinate Framework (CCF 3.0) has been a widely useful atlas and ontology built on the Allen Mouse Brain Atlas (ABA)[10]. However, these expert-labeled atlases and ontologies are time-intensive, hard-to-scale, and potentially biased by human judgment. Identifying completely data-driven organ patterns based on high-dimensional data would avoid human error, speed up the process, and reveal patterns and regions that may not be obvious to the human eye[11]. With the availability of many different modalities for whole-organ imaging[12–14], it is essential to combine these data and derive organ patterns that are consistent across modalities. The discovery of consistent patterns within and across these modalities would not only avoid human labor and bias but are also more likely to be informative for investigating the functions of these regions. The specific sets of genes and their spatial co-expression that contribute to principal patterns are also likely to contribute to the unique functions of brain regions they delineate[15]. Moreover, the specific genes or their combinations identified through these analyses will be highly informative for making genetic tools for experimental access to specific cell types and regions within an organ of interest[16–18].

Techniques based on machine learning (ML) principles have been widely used to segment or cluster spatial gene expression datasets[11,15,19–31]. These segmentation and clustering techniques include but are not limited to k-nearest neighbors, hierarchical clustering, spectral clustering, and deep learning based methods. While a segmentation technique provides a set of non-overlapping or in some cases overlapping clusters in the spatial gene expression dataset, it fails to provide a systematic model-based representation of the entire gene atlas. By contrast, matrix decomposition techniques such as non-negative matrix factorization (NMF) provide a model-based representation of an entire dataset as a combination of a set of dictionary elements or principal patterns (PPs). These models could optimally capture the complex patterns as PPs and represent each data point as a combination of these PPs. This form of decomposition provides a more interpretable representation of each data point and its relationship to PPs compared to segmentation or clustering-based approaches. Note that Principal Component Analysis (PCA) as a matrix decomposition technique is limited in the analysis of spatial gene expression datasets because these datasets require biologically-realistic assumptions that are not addressed by PCA (e.g., non-negative PPs and non-negative combination coefficients).

Moreover, while all the existing ML toolboxes are promising for the segmentation and clustering of gene expression datasets, the stability of the patterns hasn't been the central criterion in designing them. Stability is a measure of scientific reproducibility and statistical robustness; it asks whether each step of the ML pipeline produces consistent results with slight perturbations in the model or data[32,33]. Stability is a minimum requirement for trustworthy interpretability[33,34]. Without demonstrated stability and reproducibility of segments or PPs, it is not clear if they can be used for further biological interpretations. In this paper, we describe a

3D PP and ontology identification framework, referred to as osNMF (ontology discovery via stability-driven non-negative matrix factorization). osNMF builds on the staNMF technique[15] to identify meaningful, stable, and spatially-distinct patterns in spatial gene expression datasets in the adult mouse brain. Previous work has demonstrated the promise of staNMF in finding biologically meaningful pre-organ regions using 2D spatial gene expression images from *Drosophila* embryos[15]. Here we leverage these findings in osNMF and carry out similar analysis for 3D spatial gene expression data in the mouse brain. Moreover, osNMF is able to systematically compare PPs with known anatomical regions and develop new ontology maps from spatial gene expression and spatial transcriptomics (ST) datasets for any biological system or organ.

Results

Unsupervised identification of 3D principal patterns in the Allen Mouse Brain Atlas

We designed and implemented a computational framework called osNMF to (1) pre-process 3D spatial gene expression datasets by imputation of missing voxels, (2) identify principal patterns (PPs) of 3D gene expression profiles, and (3) systematically compare these PPs to known anatomical regions (**Fig. 1A**). osNMF consists of three processing modules based on interpretable machine learning techniques. First, we used a machine learning module with a default as the k-nearest neighbors (kNN) algorithm[35,36] to impute missing voxels in the 3D gene expression atlas. Second, we used the stability-driven non-negative matrix factorization (staNMF)[15] technique to find spatially coherent principal patterns (PPs) in the imputed gene expression data. For each 3D gene expression profile, this model is an additive non-negative linear combination of the PPs that reconstructs each gene profile. In other words, staNMF is trained to learn a latent space for the dataset where the latent space features can accurately reconstruct the whole dataset. Non-negativity ensures a more biologically plausible pattern recognition with sensible non-negative spatial patterns. We then trained the staNMF using Coordinate Descent, a well-known optimization technique in the field of machine learning that is especially efficient for non-negative matrix factorization[37]. After training the staNMF, a PP is calculated as one of K dictionary elements learned via staNMF. Note that the dimension of the coefficient matrix is K by the number of voxels in each input image, hence yielding K distinct PPs. A gene profile is then reconstructed by the weighted linear combination of these K PPs. The weights of each PP or dictionary element for each gene are determined by the coefficients of staNMF. Therefore, staNMF constructs two outputs: (1) K PPs for the whole imaging data and (2) the coefficients or PP weights for each gene image. Finally, the third processing module in osNMF is a search algorithm that systematically identifies the combinations of known anatomical regions that are most similar to a PP. osNMF is an automatic, unsupervised, data-driven process to find the PPs, and therefore, no manual annotations or labels are needed for the voxels in the brain.

We used osNMF to determine the PPs for 4,345 3D spatial gene expression profiles in the adult mouse brain (56 days old) from the ABA dataset[14]. In this dataset, each gene was examined by whole-brain serial sectioning and RNA *in situ* hybridization (ISH). We resized this dataset to 200 μm isotropic resolution for all analysis in this study. To make a machine-learning-ready dataset, we first implemented a kNN-based voxel imputation module in osNMF to impute approximately

10% missing voxel data in ABA. This imputation algorithm accurately estimates the missing voxel values in the gene expression atlas enabling osNMF to handle datasets with missing values. On a hold-out test set of 1,000 random voxels for each of the 4,345 genes from the ABA dataset (for 4,435,000 total hold-out data points), the mean error was smaller than 0.01. The Pearson correlation coefficient between the measured and imputed gene expression data was 0.52, with a p-value < 0.01.

We then used the second module in the osNMF framework (the staNMF core) to automatically identify stable PPs in the ABA dataset. Similar to staNMF, a PP instability score was used to select K (the number of PPs) with the most stable PPs. The instability score was defined as the average dissimilarity of all learned dictionary pairs using their cross-correlation matrix. The dissimilarity was quantified by: (1) Hungarian matching method[38] (**Fig. 1B**) and (2) an Amari-type error function[15] (**Fig. S1**). This instability score measures the instability of the learned PPs over 100 runs of the same algorithm across a range of 8 to 30 possible numbers of PPs. For both dissimilarity methods, the lowest instability (and thus highest stability) was found for 11 PPs (K=11) for the ABA dataset. At 11 PPs, the instability score is 0.020 ± 0.002 (1 is the maximum instability). K=13 and K=12 are the next two lowest instability scores (0.03 and 0.04, respectively, with standard deviations < 0.01). **Fig. 1B** also compares the instability of PPs between osNMF and PCA. osNMF PPs have higher stability and lower standard deviation vs. PCA PPs at every value of K tested. At 11 PPs, PCA PPs have an instability score of 0.25 ± 0.01 (mean ± SD), which is considerably higher (less stable) than the instability score for osNMF (0.020 ± 0.002). osNMF's lower instability score and standard deviation suggest greater stability, repeatability, and interpretability compared to PCA PPs. In terms of computational cost, osNMF takes longer to run than PCA, though both are fast-running models. On a 2021 MacBook Pro M1 laptop CPU, it takes 26 seconds to run osNMF to create one set of PPs on the ABA dataset vs. 4 seconds for PCA.

osNMF's 11 automatically-extracted PPs appear spatially distinct and cohesive (**Fig. 1C**). The spatial coherence in these PPs is important and could support their biological interpretability. Therefore, we explore quantitative measures of spatial coherence in this section. To quantify spatial coherence of PPs, we used the Moran's Index, frequently referred to as Moran's I[39]. Moran's I was originally used in geostatistics and has more recently been used in spatial gene expression literature[40–42]. Moran's I ranges in value from –1 to 1. A value close to -1 indicates little spatial organization, similar to a chess board with black and white spots distributed across the board. A value close to 1 indicates a clear spatially distinct pattern, such as if all the black blocks in a chess board were on one side and all white blocks on the other. osNMF PPs have an average Moran's I of 0.58 ± 0.12 which is considerably higher than that of PCA at 0.47 ± 0.15 (p-value<0.001) across 20 bootstrapped simulations for each of the 11 PPs (**Fig. 1D**, see **Fig. S2** for 11 boxplots of Moran's I, one each of the 11 PPs). This suggests a strong spatial separation and coherence of PPs for osNMF compared to PCA PPs which appear less visually distinct and cohesive (See **Fig. S3A** for visualization of PCA PPs). Although osNMF PPs are spatially coherent, a large number of PPs tend to be present in most gene expression profiles (58% of all genes are represented in 9 or more PPs), suggesting the heterogeneous spatial expression of the genes in the adult mouse brain (**Fig. 1E**). Only two genes are represented in a single PP (<0.1% of all 4,345 genes), while 438 genes are represented in all 11 PPs (10.1% of all genes).

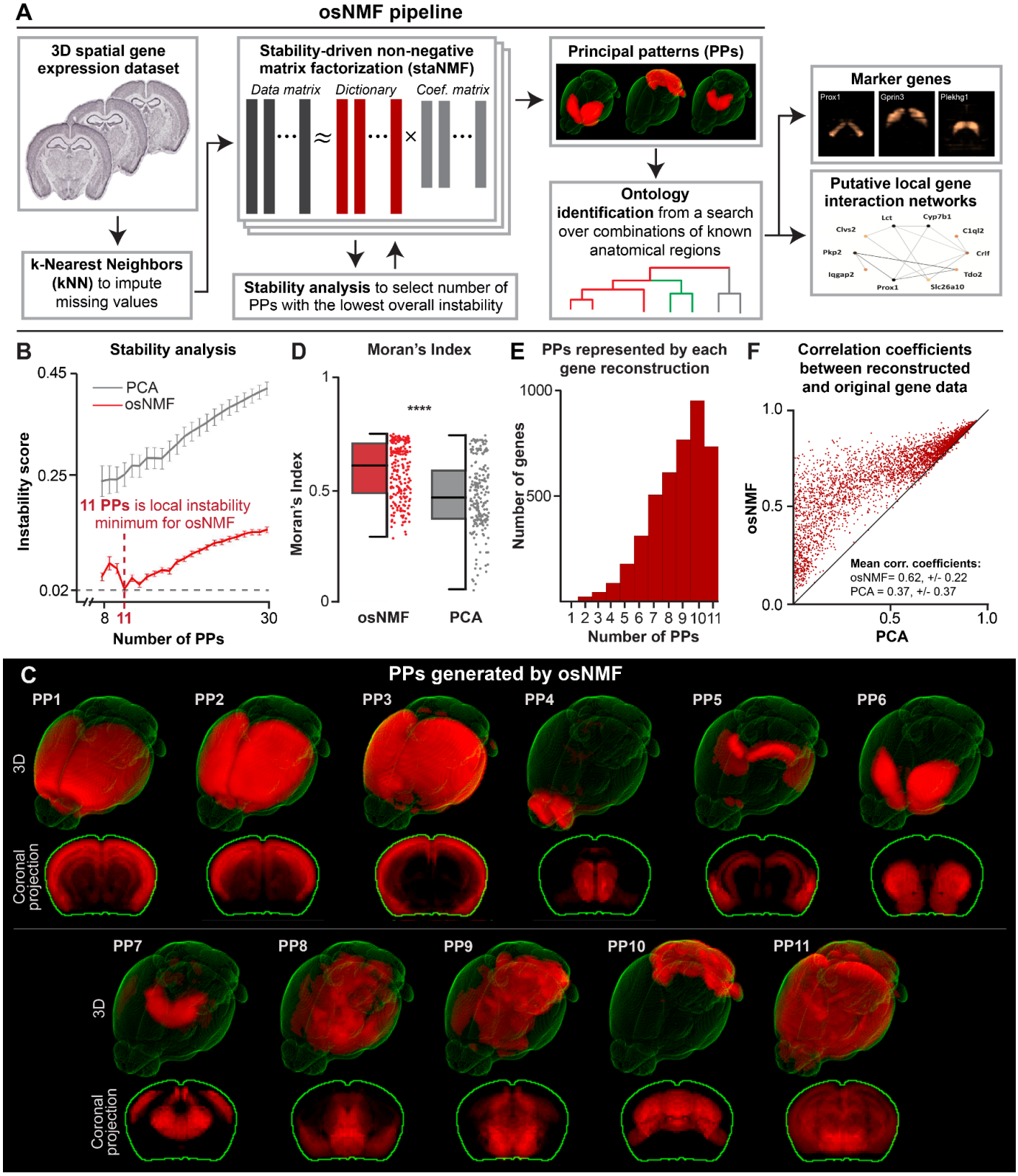

Figure 1: The architecture of the osNMF framework and its application in revealing PPs in spatial gene expression datasets. A. The computational framework for osNMF. **B**. Stability analysis for osNMF PPs and PCA PPs across 100 runs for each K value, from 8 to 30 for ABA dataset, using the Hungarian matching method. The ABA dataset contains 4,345 3D in situ hybridization (ISH) images at 200 μm^3 resolution of the adult mouse brain. Error bars show the standard deviation. **C**. 11 PPs generated by osNMF from the ABA dataset in 3D and projected on the coronal plane. **D.** Boxplots of Moran's Index for osNMF vs. PCA PPs across 220 bootstrapped simulations

(p-value<0.001). Data from each individual point is shown in a vertical column to the right of the boxplot. **E.** Pearson correlation coefficients of osNMF and PCA gene reconstructions vs. the original ABA data. Each dot represents the correlation coefficient osNMF and PCA reconstructions of one of the 4,345 genes versus the original ABA gene-by-gene data. **F.** Number of PPs represented by each osNMF gene reconstruction of the 4,345 ABA genes.

To quantify the accuracy of the osNMF reconstruction, we defined the gene reconstruction accuracy as the Pearson correlation coefficient between the PP-reconstructed 3D gene image and the original gene image. We compared the osNMF and PCA reconstruction accuracy in the scatterplot, where each point represents one of the 4,345 genes in the dataset (**Fig. 1F**). The analysis of the scatterplot shows that osNMF considerably outperforms PCA in the reconstruction performance (0.62 ± 0.22 for osNMF compared to 0.37 ± 0.37 for PCA; 24% higher accuracy for osNMF). We also found that our kNN imputation of missing values improves osNMF's reconstruction accuracy of the original data set from 0.59 to 0.62. It is worthwhile to note that the reconstruction accuracy will slightly increase with a higher value of K (e.g. reconstruction accuracy is 0.69 for K=30). However, the instability score tends to significantly decrease for the higher values of K (e.g. instability score for K=30 is 0.14 vs. 0.02 at K=11, which is roughly 7x higher instability). Here, we use the stability of PPs as a minimum requirement for the biological interpretation of the PPs and assess the reconstruction accuracy to ensure an acceptable reconstruction performance. Computational tools are often optimized purely for reconstruction accuracy, which may lead to less reproducible and biologically relevant results. Overall, our findings suggest that osNMF outperforms PCA in automatically generating biologically-relevant patterns from spatial gene expression profiles.

osNMF PPs map to combinations of known anatomical regions in the mouse brain beyond the established ontology

Next, we investigated how our PPs, which were automatically-discovered from the gene expression profiles, overlap with known anatomical brain regions. To characterize the mapping between PPs and known brain regions, we first calculated the Pearson correlation coefficient between all 868 expert-annotated brain regions from the Allen Common Coordinate Framework (CCF v3.0)[10] to each of the osNMF PPs. To match the registered ABA gene expression dataset, we downsampled the CCF to 200 μm isotropic resolution. To facilitate visualization for this comparison, we show 66 of the 868 regions in **Fig. 2**. These 66 regions provide a complete medium-level representation of the mouse brain CCF. We selected these 66 regions by including all "child" regions for the 12 coarse CCF regions (isocortex, olfactory areas, hippocampal formation, cortical subplate, striatum, pallidum, thalamus, hypothalamus, midbrain, pons, medulla, and cerebellar cortex/nuclei). In this paper, we define "coarse-level" regions as these 12 CCF regions, "medium-level" regions as their 66 children, and "fine-level" regions as all regions that are finer than medium-level.

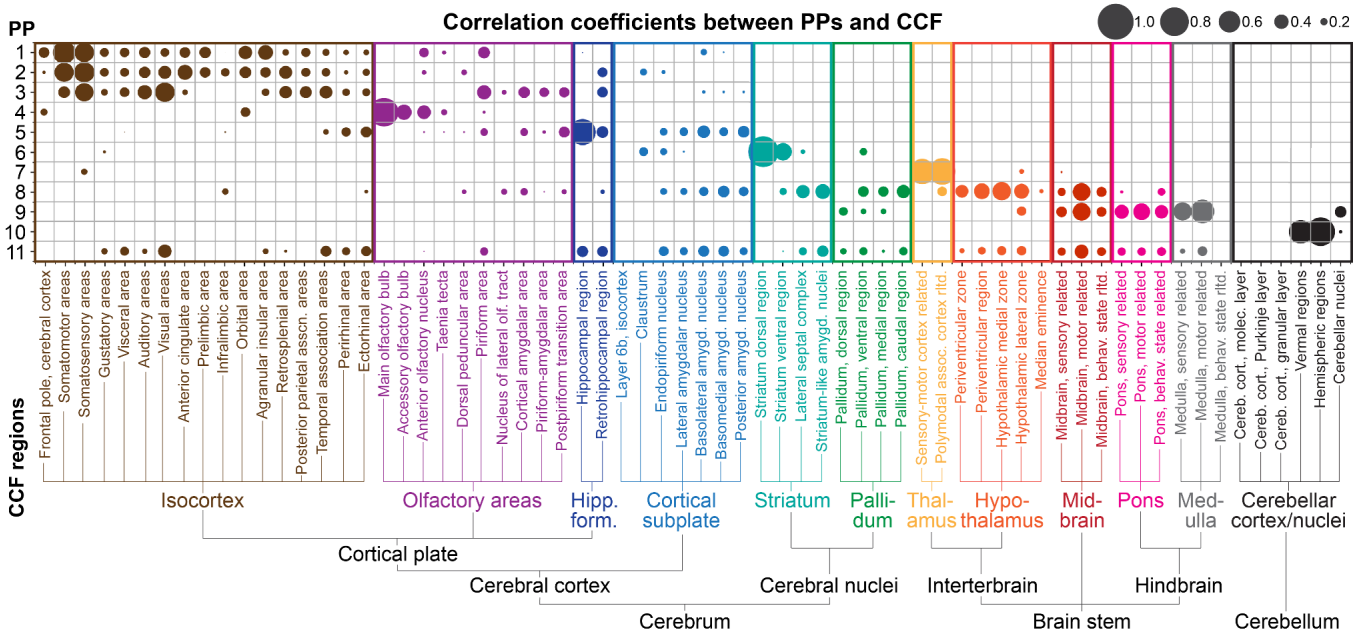

Figure 2: Correlation map between osNMF PPs and expert-annotated anatomical regions from the Allen Common Coordinate Framework (CCF 3.0). Map of the Pearson correlation coefficients between PPs (y-axis) and expert-annotated regions from the CCF in the adult mouse brain (x-axis). Each circle represents the value of the correlation coefficient between each PP and CCF region. The CCF regions labeled in the vertical text are the complete set of 66 children of the 12 coarse CCF regions and are organized left-to-right based on the CCF ontology map. PPs are organized top-to-bottom based on their correlation coefficient to the CCF coarse regions.

Interestingly, osNMF PPs, driven purely by the gene expression, have similarities to the CCF ontology, but also major differences (**Fig. 2**). Three PPs (PPs 1, 2, and 3) are well, yet in many cases, differentially correlated with select parts of the isocortex. For example, PPs 1, 2, and 3 all have correlation with the somatosensory areas of the isocortex, in addition to differential correlation with other cortical areas (e.g., somatomotor, visual, and orbital areas of the isocortex). In addition, these three PPs have varying representations outside of the isocortex, including in the olfactory areas, hippocampal formation, and cortical subplate. This is an interesting observation because the hippocampal formation, olfactory areas, and the cortical subplate are each viewed as part of the cerebral cortex[10]. PP4 is mostly represented within the olfactory areas, with an especially high correlation to the main olfactory bulb and orbitofrontal areas of the isocortex. PP5 has a strong correlation to hippocampal formation but has some correlations to sub-regions within the isocortex, olfactory areas, and cortical subplate. Thus, we see that PPs 1, 2, 3, 4, and 5 correlate with the cerebral cortex, one of the three highest-level CCF regions (in addition to the brainstem and cerebellum), but do not fit neatly within the coarse- or medium-level CCF regions.

Moving next to PP6, we found a considerably high correlation between this PP and the striatum with minor expression in the cortical subplate. PP7 exhibits a high correlation only to the thalamus, showing good agreement between this PP and CCF's thalamus in the overall ontology. Unlike PP7, PP8 shows correlations spread across multiple regions, especially the hypothalamus, midbrain, striatum, pallidum, and cortical subplate (in descending order of correlation), which suggests that these CCF regions share gene expression patterns. Similarly, PP9 is highly correlated with multiple regions in the brain stem areas including the medulla,

midbrain, and pons, as well as a minor expression in cerebellar nuclei. Finally, PP10 is highly correlated with the cerebellum, with major expression in cerebellar vermal and hemispheric regions but not in the cerebellar nuclei. A comparison between PP9 and PP10 suggests that there are significant gene expression differences between the cerebellar nuclei and the vernal/hemispheric regions of the cerebellum. Genes that are expressed in cerebellar nuclei tend to also be expressed in the brain stem areas while genes that are expressed in cerebellar vernal/hemispheric regions tend to be exclusively present in the cerebellum. Finally, note that PP11 is correlated to most CCF regions. Visual inspection of PP11 (**Fig. 1C**) suggests that this PP corresponds to the noisy gene expression profiles throughout the brain.

To further delineate the similarities and differences between CCF ontology and PPs, we asked which combination of CCF regions is best aligned with each PP. To answer this question, we ran a search of all possible neighboring combinations of 1, 2, or 3 CCF regions (the third processing module in osNMF). Out of 868 total CCF regions, we found 22,711 neighboring combinations of 2 regions and 1,834,540 neighboring combinations of 3 regions. We did not consider neighboring combinations of higher than 3 regions due to exponentially higher computational demand. We then identified the maximum Pearson correlation coefficient between each PP and the superset of all single CCF regions, all combinations of 2 CCF regions, and all combinations of 3 CCF regions. We found that our PPs tend to be aligned with combinations of the coarse, medium, and/or fine CCF regions, but these combinations may exhibit different ontology than CCF (**Fig. 3A**). PP7 and PP10 are the only PPs that are each maximally correlated with only one single CCF region: PP7 has a correlation coefficient of 0.88 to the thalamus, while PP10 has a correlation coefficient of 0.92 to the cerebellar cortex. PPs 1, 2, and 3 have their highest correlation to combinations of three CCF regions (correlation coefficients of 0.73, 0.71, and 0.63, respectively), which includes the isocortex for each PP. In addition to isocortex, PP1 adds the anterior olfactory nucleus and the olfactory piriform area, PP2 adds two finer-level retrohippocampal regions including the subiculum and the fine-level layer 6a of the lateral entorhinal area (ENTl6a), and PP3 adds the olfactory piriform area and the entorhinal area (ENT) of the retrohippocampal region. PP4 has its highest correlation (correlation coefficient of 0.78) with the combination of olfactory bulb and accessory olfactory bulb with a fine-level cortical region (Orbital area, ventrolateral part, layer 1, referred to as ORBvl1). PP5 is maximally correlated with a combination of two medium-level regions from hippocampal formation (hippocampal region and ENTl), and the high-level cortical subplate region. PP6 has its highest correlation (correlation coefficient of 0.89) to a combination of three CCF regions: 1) striatum: dorsal region; 2) striatum: nucleus accumbens; 3) striatum: olfactory tubercle. This PP does not include the striatum: amygdalar nuclei. Instead, the combination of hypothalamus, amygdalar nuclei, and midbrain is maximally correlated to PP8. Single-cell gene expression research has suggested that the amygdalar nucleus, midbrain, and hypothalamus contain cell types that are in fact highly related[43]. Finally, PP9 is maximally correlated with the combination of hindbrain, midbrain, and cerebellar nuclei (correlation coefficient of 0.84). This organizes the midbrain and hindbrain together, and suggests a relatively high similarity of gene expression between the midbrain, medulla, and pons, as observed with single-cell transcriptomics and clustering[43].

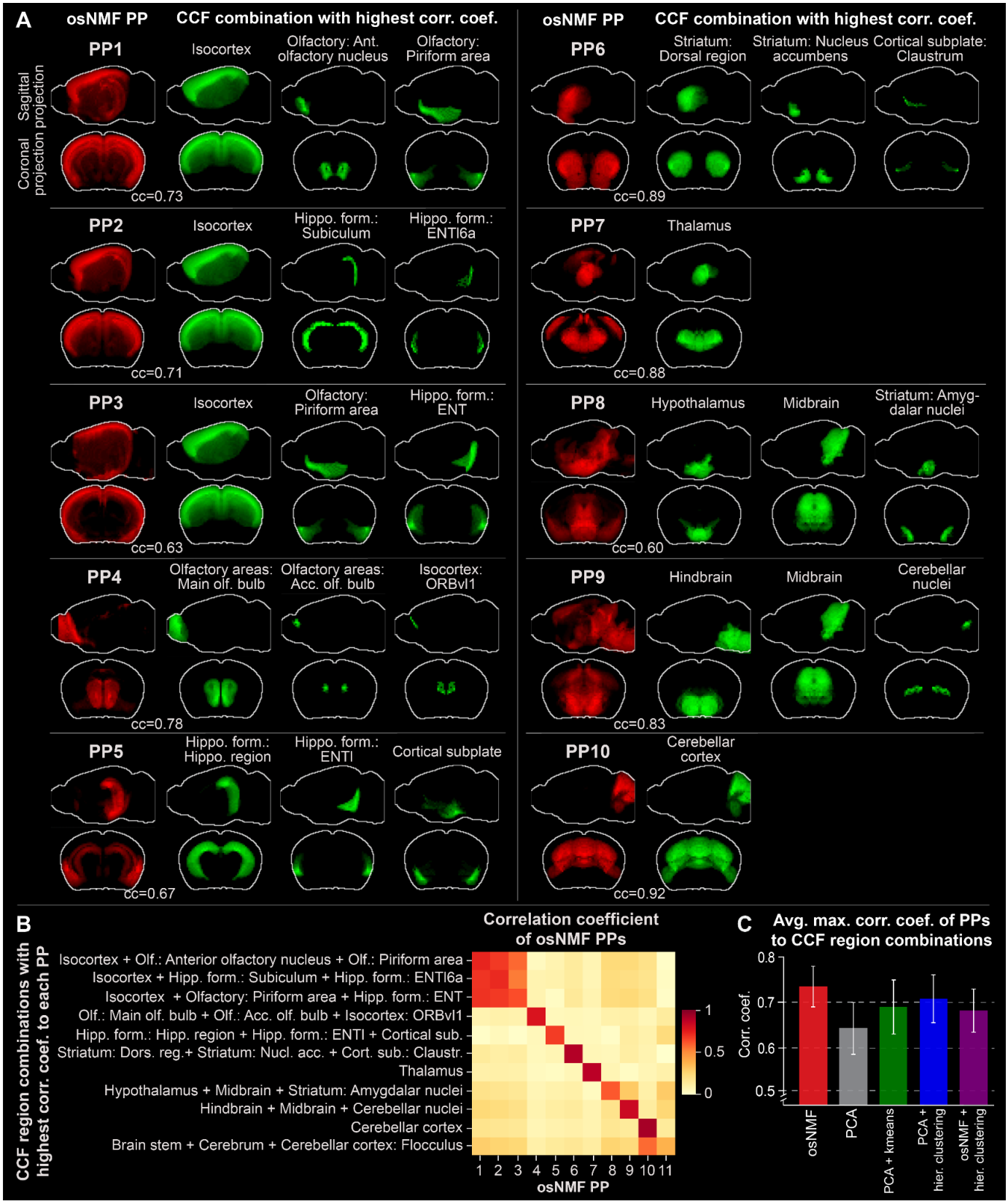

Figure 3: Alignment between osNMF PPs and combinations of CCF regions. A. PPs (in red) and the most similar combination of expert-annotated regions (in green) from CCF in the adult mouse brain projected on the sagittal and coronal planes. To identify the most similar combination of regions, we looked at all possible neighboring combinations of 1, 2, or 3 regions from all 868 CCF regions, and then identified the combination with the highest Pearson correlation coefficient to each PP. The top 10 PPs with the highest correlation coefficients are

shown. PPs are organized by correlation coefficient in descending order. **B**. Heatmap of the correlation coefficient between osNMF PPs and each PP's combination of CCF regions with the highest correlation coefficient. **C**. Comparison of the average maximum correlation coefficient of CCF region combinations to each PP for five matrix decomposition and segmentation methods: osNMF, PCA, PCA followed by k-means, PCA followed by hierarchical clustering, and osNMF followed by hierarchical clustering.

To explore how uniquely each PP represents the combination of CCF brain regions, we constructed the correlation matrix between all PPs and all associated CCF combinations (**Fig. 3B**). Most PPs (except PPs 1, 2, 3, and 11) exclusively map to their associated CCF region combinations, suggesting low overlap between these PPs. The average maximum correlation coefficient between PPs and their respective CCF region combination is 0.74±0.04. By contrast, the average correlation coefficient between each PP and the CCF region combinations except for its highest correlation region is 0.10±0.17. Note that PPs 1, 2, and 3 form a cluster that is highly correlated with the isocortex and is further explored in the next section. Finally, PP11 has the lowest maximum correlation coefficient (0.37, vs. 0.60 as the next lowest) with comparable correlation coefficients to other CCF regions, further suggesting its role in explaining the noisy gene expression profiles.

Next, we compared osNMF PPs to other common methodologies for matrix decomposition and segmentation/clustering. First, we explored PCA, one of the most frequently used tools for matrix decomposition. We used the same computational pipeline as osNMF but replaced non-negative matrix factorization with PCA to generate 11 PCA-based PPs (**Fig. S3A**). We find that PCA PPs are less spatially coherent and distinct than osNMF PPs (**Fig. S3A** for PCA vs. **Fig. 1C** for osNMF, also quantified by Moran's Index in **Figs. 1D** and **S2**). For example, PCA PPs 3, 8, and 11 do not appear to identify unique CCF regions (similar to osNMF PP11). osNMF PPs have a higher average correlation coefficient to their respective CCF regions (0.73±0.05) compared to PCA PPs (0.63±0.06). Further, the stronger diagonal pattern in the correlation matrix for osNMF (**Fig. 3B**) compared to PCA (**Fig. S3B**) suggests that osNMF PPs have a better alignment with the annotated brain regions. Overall, osNMF is able to better recognize previously known annotated brain regions with higher accuracy than PCA.

In addition to PCA, we conducted the same CCF similarity analysis on PPs from well-established clustering techniques. We clustered the ABA dataset using (1) PCA followed by k-means clustering (similar to stLearn framework[44]), (2) PCA followed by agglomerative hierarchical clustering (similar to the AGEA framework[11]), and (3) osNMF followed by hierarchical clustering as a point of comparison (**Fig. 3C**). osNMF has the most similar PPs to their optimal CCF regions (correlation coefficient of 0.73±0.05), whereas PCA, PCA followed by k-means, and PCA followed by hierarchical clustering have lower correlation coefficients (0.63±0.06, 0.68±0.06, and 0.70±0.06, respectively). Additionally, osNMF PPs have a higher similarity to CCF region combinations compared to osNMF followed by hierarchical clustering (0.67±0.05). This highlights the advantage of osNMF PP identification over the clustering of voxels in the spatial gene expression datasets. Overall, these findings suggest that osNMF PPs are more similar to the combinations of known brain regions compared to PCA or standard clustering techniques.

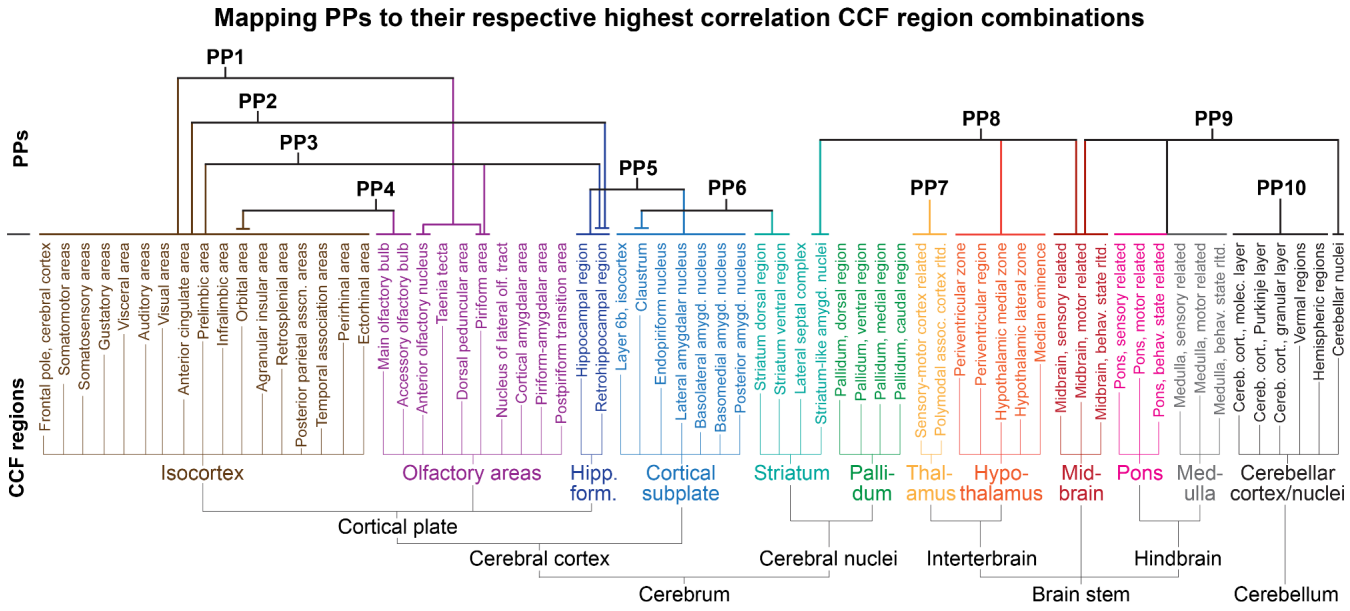

Figure 4: Summary of osNMF PPs mapping to the best-fit combination of Allen CCF regions. The 10 PPs from Fig. 3 mapped to their best-fit combinations of CCF regions.

Overall, osNMF PPs show a high correlation to known adult mouse brain regions, while also identifying a unique ontology based purely on spatial gene expression data. **Fig. 4** illustrates a summary of this proposed ontology overlayed on the CCF tree structure. Our findings highlight the capability of osNMF for rapid generation of new atlases from a large set of spatial gene expression data without supervision, and uncover novel relationships between brain regions that were difficult to discern using conventional manual approaches.

osNMF identifies subregions of isocortex in the adult mouse brain

The correlation matrix between PPs and the combination of CCF regions (**Fig. 3B**) suggests that PPs 1, 2, and 3 are highly correlated with the isocortex CCF region (**Fig. 3A**). For each of these PPs, the correlation coefficient to the isocortex dominates the overall correlation vs. the other two regions that make up the highest correlated combination (**Fig. 5A**). For example, PP2 has a correlation coefficient of 0.70 to the isocortex, while it only has a 0.13 and 0.08 correlation coefficient to the other two regions that make up its highest correlated combination. Similarly, PPs 2 and 3 have correlation coefficients of 0.70 and 0.54 to the isocortex, respectively, while their correlation coefficients to other regions are considerably lower (**Fig. 5A**). Visualization of these three PPs suggests that they represent different spatial regions of the isocortex, in addition to minor components of the hippocampus and olfactory areas (**Fig. 5B**). Overall, PP1 represents the superficial layers in frontal areas of the cortex, in addition to a partial representation of the anterior olfactory nucleus and the piriform area of the olfactory areas. PP2 represents the deeper layers of the isocortex in dorsolateral regions, and has a minor correlation to the subiculum and entorhinal area (lateral part, layer 6a) within the retrohippocampal region. Finally, PP3 represents the superficial layers of isocortex in dorsal regions as well as the piriform area of the olfactory areas and the entorhinal area of the

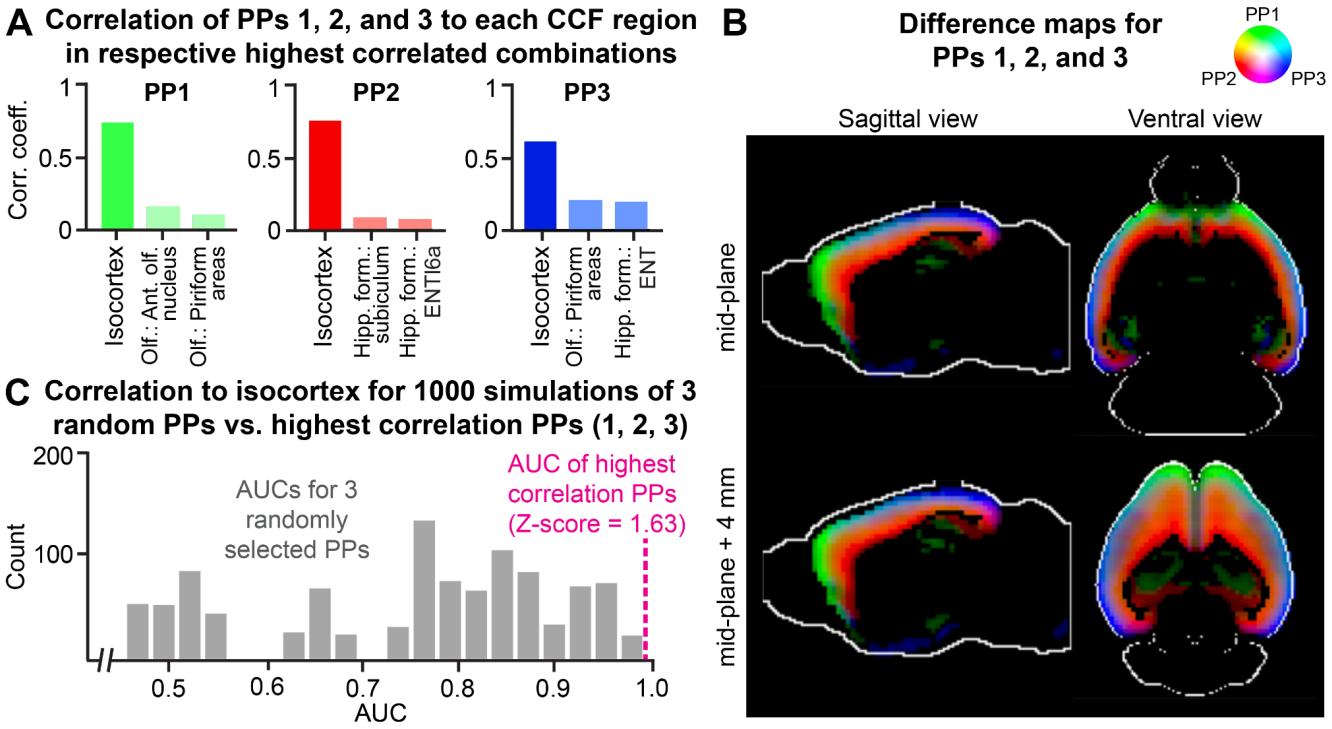

Figure 5. A. Correlation coefficient of PPs 1, 2, and 3 to each of their respective highest correlated CCF regions. Isocortex is the most correlated region to each PP. **B.** Difference map between the three PPs that are most correlated to the isocortex (PP1, PP2, PP3). Each image is a 2D cross-sectional representation. The columns represent different views (sagittal, horizontal, and coronal), as indicated above the image. The rows represent two different 2D planes for each view: The first row is a mid-plane representation and the second row is a mid-plane representation + 4 mm. **C**. Histograms of AUC values for isocortex for 1,000 runs of a logistic regression randomly fitting three PPs to isocortex CCF regions, with the magenta vertical dashed line showing the AUC for PP1, PP2, and PP3. These three PPs are the best predictors of the isocortex compared to any other combination of three PPs.

retrohippocampal region. PP1 and PP3 have slight gradual overlap in superficial layers, as seen by the cyan color in **Fig. 5B**.

We then investigated how effectively the combination of PP1, PP2, and PP3 can recreate the isocortex alone. To estimate this, we trained a logistic regression model to predict the CCF reference map for the isocortex from these three PPs. The area under the receiver operating characteristic (ROC) curve or AUC measure for this prediction is 0.99. This regression model is the most accurate model amongst 1,000 other models that uses a random selection of three PPs to predict isocortex (**Fig. 5C**). The median AUC for these 1,000 models is 0.78 (compared to 0.99 for the model that uses PP1, PP2, and PP3, as shown in the magenta vertical dashed line), demonstrating that PP1, PP2, and PP3 represent the isocortex as a whole.

osNMF identifies marker genes of PPs

Marker genes for an organ or tissue region are a set of genes with high expression within that region and relatively low expression in other regions. Marker genes are frequently used as

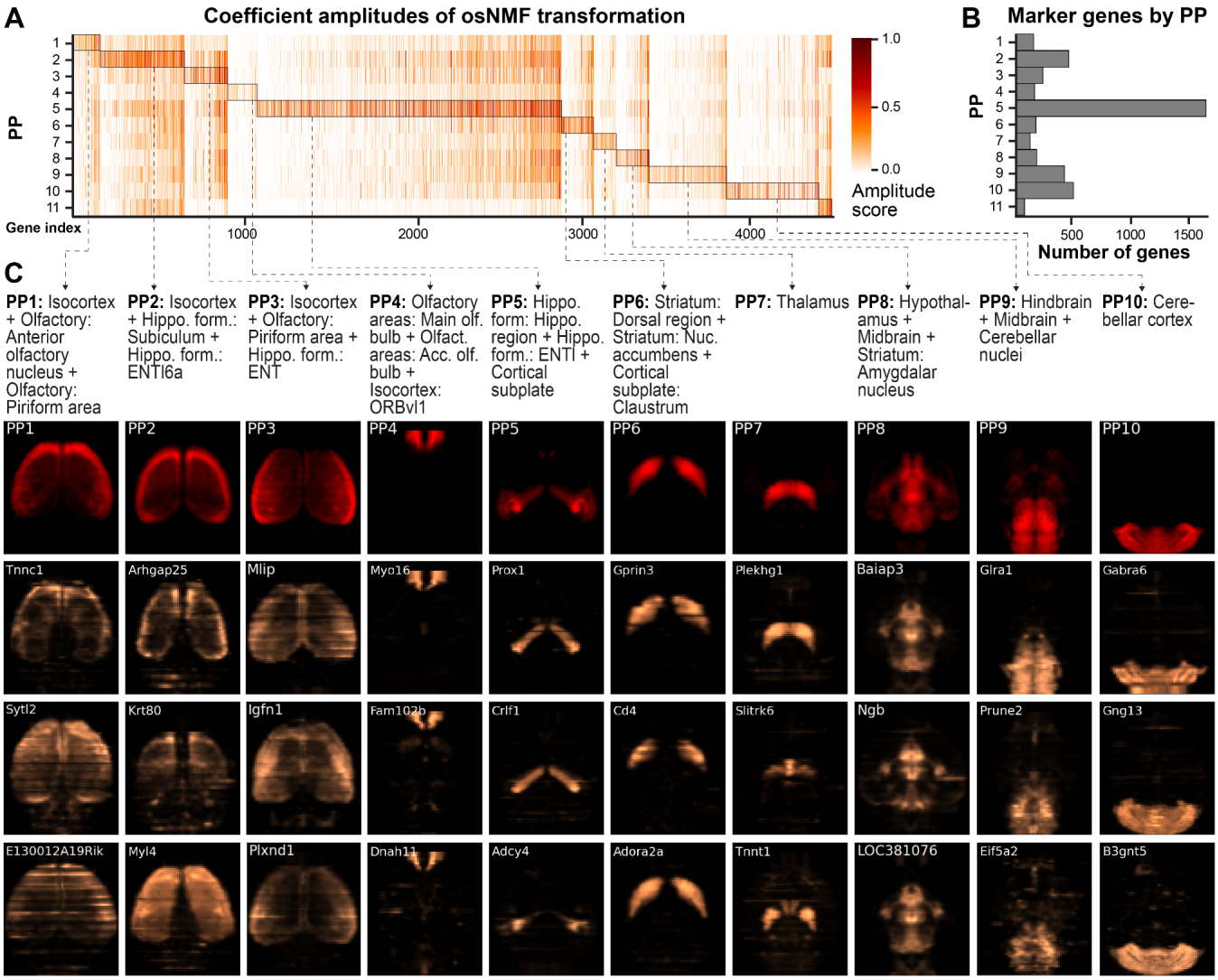

Figure 6: Identifying robust marker genes for PPs in the adult mouse brain. A. Heatmap of coefficients for each of the 4,345 genes in the dataset calculated for each of the 11 PPs. PPs are ordered in descending order by the maximum correlation coefficient to a combination of CCF regions, as described in Fig. 3. Genes are assigned to the PP for which they have the largest coefficients. **B**. Number of genes for each PP that have the highest amplitude score associated with that PP. **C**. The horizontal projection of gene expression of the top three marker genes (in the copper color) from ABA for 10 PPs and their respective CCF region combinations. The marker gene name is displayed in the top left corner of each horizontal gene expression image. The horizontal view of the respective PP (in red) is displayed below the marker genes.

starting points for understanding functions of cells and regions they are expressed in and are widely used in designing genetic tools for experimental access to those cell types and regions for further studies[45,46]. Given the relationship between PPs and regions established in the previous section, one can use PPs to identify the most robust marker genes for a region within any organ or tissue. Inspired by[15], we used the following process to identify the marker genes for each PP: For each gene, we first extracted the staNMF coefficients. Each coefficient quantifies the contribution of each PP in explaining the expression of the gene. We then assigned each gene to a specific PP based on the PP for which the gene has the highest coefficient. Next, we defined an importance score for each gene by taking the coefficient of

that gene to its assigned PP divided by the sum of the coefficient for that gene. To select unique marker genes for each PP, we picked the genes with the top 3 highest importance scores for each PP.

We used this pipeline to identify marker genes of each osNMF PP in the adult mouse brain. The coefficients of each PP for each gene are visualized in **Fig. 6A**. Intuitively, this coefficient map represents the importance of each of the 11 PPs for each of the 4,345 genes. For this plot, the genes are grouped together based on the PP for which they have the largest coefficient and then sorted based on their importance scores for that PP. The number of genes selective to each PP is not uniform across PPs (**Fig. 6B**). PP5 (correlated with the hippocampal region), has by far the most unique genes, with over 1,500 of the 4,345 total genes. PP2 (correlated with the isocortex region), PP5 (correlated with the hypothalamus region), and PP10 (correlated with cerebellar cortex) also have an especially large number of associated genes (represented by darker orange and red in the heatmap). The genes with the highest coefficients for each PP are candidate marker genes for that PP and consequently its associated brain regions (**Fig. 6C**). The comparison of gene expression with the corresponding PP shows convincing visual alignment suggesting a successful identification of marker genes for each PP. For example, *Prox1*, the top identified marker gene for PP5 (associated with hippocampal formation and cortical subplate), is known to be widely expressed across the brain during development, but primarily in the hippocampus and cerebellum in adulthood[47]. As another example, *Gabra6*, the top identified marker gene for PP10 (associated with the cerebellar cortex) is known to be preferentially expressed in the cerebellum as part of a program related to differentiation[48].

osNMF identifies putative spatial gene interaction networks in the adult mouse brain

It is known that the spatial co-expression of genes yields meaningful biological relationships. For example, a spatial co-expression network has successfully reconstructed the gap gene regulatory network in Drosophila[15]. However, few existing computational tools incorporate spatial information in identifying gene co-expression networks, and the ones that do, leverage existing expert-led ontologies vs. data-driven ontologies like osNMF PPs[49–52]. Data-driven ontologies from tools like osNMF will allow better identification and exploration of 3D spatial gene networks. Building on a similar analysis for *Drosophila* embryos[15], we used the following process to construct putative local gene interaction networks for the PPs in the adult mouse brain: We first identified the top marker genes for each PP by selecting the genes with the top 0.025% highest normalized coefficients for that PP. For these top marker genes of each PP, we computed the Pearson correlation coefficient between the staNMF coefficients of the two genes. We then drew an edge between two genes if their similarity score is among the top 5% of all similarity scores for that gene subset.

We identified the putative local gene interaction networks for each PP. This analysis resulted in the selection of 10 or 11 top marker genes for each PP. We created the putative spatial gene interaction networks based on these top marker genes identified for each PP (**Fig. 7** for PPs 1-7, and **Fig. S4** for the remaining PPs). Interestingly, these data-driven putative gene networks identify some of the regulatory relationships that are recently found via experimental research.

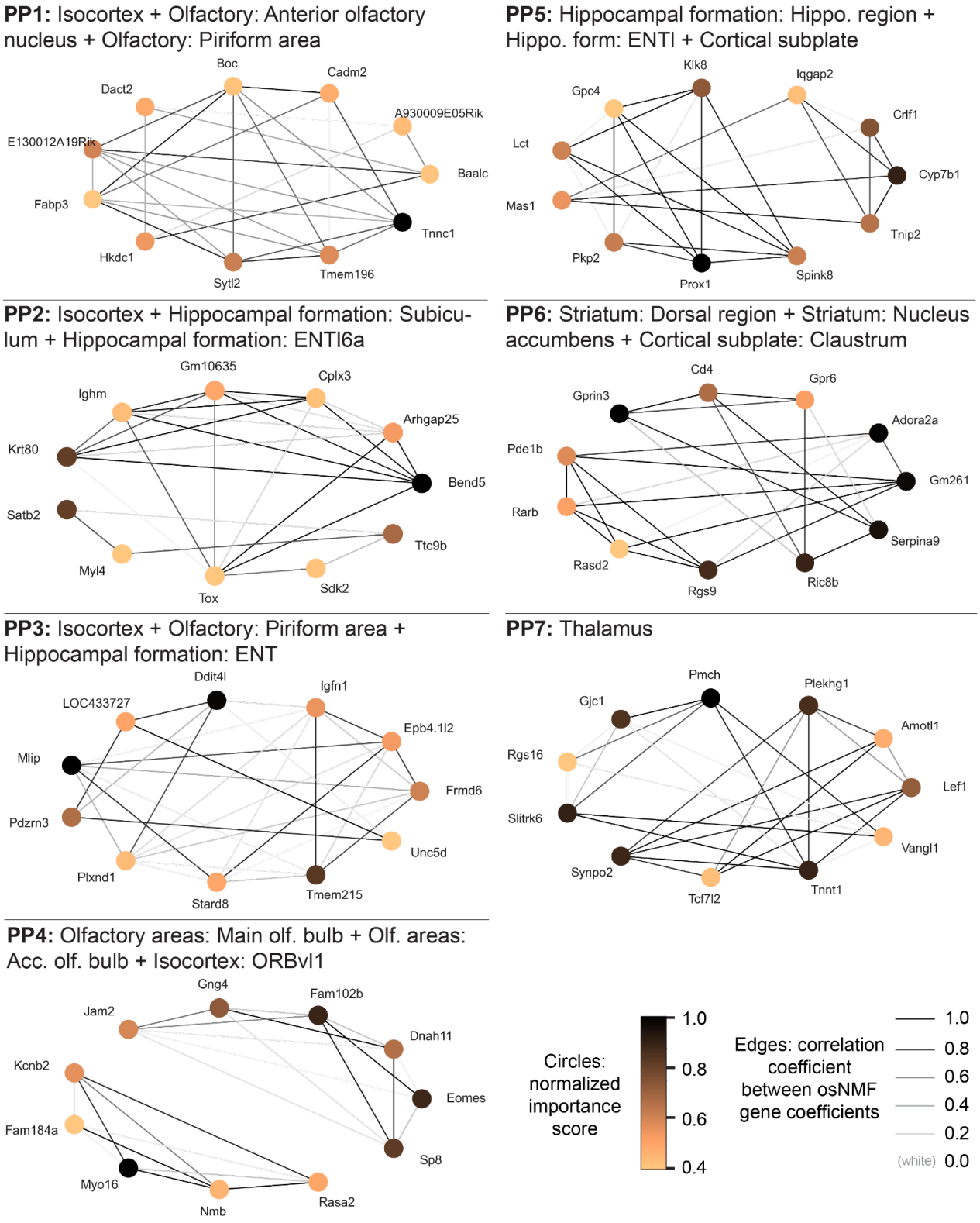

Figure 7: Putative local gene interaction networks constructed for seven PPs with high similarity to the coarse CCF regions in the adult mouse brain. The node color presents the selectivity of the gene to the PP associated with the brain region. An edge is drawn between genes if the similarity score is among the top 5% of all similarity scores for that gene subset. The edge thickness is proportional to the similarity scores between the osNMF representation of the two genes.

For example, in PP6, which is correlated to striatum, seven marker genes show especially strong edges (*Gprin3*, *CD4*, *Gpr6*, *Ric8b*, *Rgs9*, *Serpina9,* and *Gm261*) and seem to form a hub of connections. Interestingly, a 2019 experimental study in mice found that *Gprin3* controls striatal neuronal phenotypes including excitability and morphology, as well as behaviors dependent on the striatal indirect pathway and mediates G-protein-coupled receptor (GPCR) signaling[53]. *Gpr6* is a GPCR gene, and *Rgs9* and *Ric8b* are regulators of GPCR genes. In addition, *Gm261* and *Serpina9* are known to impact synapse development. In addition, *Prox1* and *PKP2* appear as interactions in PP5, which is related to hippocampal formation. Interestingly, a recently published experimental study has identified *Prox1* as a transcription factor associated with *PKP2* expression[54]. These relationships could be used as leads for experimental validation when studying specific genes in their tissue context.

Discussion

Here, we describe osNMF, an unsupervised data-driven framework for (1) pre-processing of 3D spatial gene expression datasets by imputation of missing values, (2) identifying principal patterns (PPs) from spatial gene expression datasets, and (3) systematically comparing these PPs to known anatomical regions. Unsupervised techniques to analyze gene expression patterns are increasingly important given the rapid growth in spatial transcriptomics (ST) datasets. For a data-driven technique to be useful in the biological interpretation of gene expression patterns, the technique should be stable, accurate, and interpretable. In this paper, we build on the stability-driven nonnegative matrix factorization (staNMF) technique to reveal stable, spatially coherent, and accurate gene patterns and interactions in the adult mouse brain. Further, these 3D PPs reveal an ontology driven solely by spatial gene expression data that differs from the established expert-annotated region agglomeration/ontology included in the most up-to-date CCF[10]. Compared to other methods, the PPs identified by osNMF have a higher correlation to combinations of CCF regions compared to PCA and other popular clustering methods, suggesting osNMF PPs are more representative of biological reality for spatial gene expression data. Our framework is built on recent developments in the field of data science that establish three principles for data science: predictability, computability, and stability (PCS)[33]. In particular, the stability of patterns generated is critical for reproducibility and interpretability. If patterns are not stable, then each new realization of a model could lead to inconsistent patterns that are divorced from biological reality. In summary, we believe that osNMF can be useful in automatically generating accurate, stable, and spatially-coherent patterns, atlases, and ontologies for any spatial gene expression and ST dataset, potentially avoiding human labor and bias.

Our analysis pipeline and the biological interpretations of the PPs have some limitations. For faster computation, the ABA dataset used in our analysis was reduced to 200 μm isotropic resolution, which is a relatively low resolution compared to emerging ST datasets. There may be errors and biases in the gene expression data both in the values and in the alignment of images, which could affect our analysis. Further, osNMF PPs have been derived only from the ABA dataset, which is a single modality dataset (ISH gene expression). While this allows the investigation of PPs that are automatically derived from gene expression only, it does not leverage the richness of other modalities to fully parcellate the brain. Our future work will

include integration with other modalities such as MRI and axonal projections[55,56] to precisely characterize finer brain regions.

Representation of spatial gene expression profiles by osNMF shows promise in determining putative local gene interaction networks. Some of the marker genes identified match with known biological reality for the brain regions. For example, *Prox1*, the top identified marker gene for PP8 (associated with hippocampal formation), is known to be widely expressed across the brain during development, but primarily in the hippocampus and cerebellum in adulthood. Similarly, our putative gene interaction networks highlight genes that may functionally cooperate to produce specific striatal neuron function (e.g. *Gprin3*, *CD4*, *Gpr6*, *Ric8b*, *Rgs9*, *Serpina9,* and *Gm261*). These marker genes and gene networks may also identify new functions and relationships in the mouse brain. However, the predicted marker genes and gene interaction networks still need to be validated. Our local gene interaction identification pipeline currently leverages the linear relationship between PPs to identify gene networks. This pipeline could be improved by incorporating nonlinear interactions using supervised methods such as iterative random forest[57].

Given the rapid growth of ST datasets and the importance of interpretability, we expect stability-driven methods like osNMF to increasingly be utilized. Future explorations could include applying osNMF to new ST datasets and modalities, as well as introducing stability-driven methods to DNNs and other modern machine learning algorithms. The code for osNMF is freely available to the scientific community at https://github.com/abbasilab/osNMF.

Methods

Data

The primary dataset used in this study is the in situ hybridization (ISH) measurements from 4,345 genes at 200 μm isotropic resolution from the adult mouse brain at 56 days postnatal[14]. The data was collected at the Allen Institute for Brain Science and is publicly available under the Allen Brain Atlas (ABA) [https://mouse.brain-map.org/]. The methods for data collection are described in detail in[14]. An API enables the download of the data at [http://help.brain-map.org/display/mousebrain/API]. The Allen Mouse Brain Common Coordinate Framework (CCF) was used as the 3D reference atlas[10]. We used CCFv3 which is publicly available at [http://help.brain-map.org/display/mousebrain/api]. CCF consists of parcellations of the entire mouse brain in 3D and at 10 μm voxel resolution. CCF provides labeling for every voxel with a brain structure spanning 43 isocortical areas and their layers, 329 subcortical gray matter structures, 81 fiber tracts, and 8 ventricular structures. The methods for constructing the CCF dataset are described in detail in[10].

Data preprocessing

The ISH data was resized at 200 μm isotropic resulting in a matrix of size 67x41x58 for each gene. We used data from 4,345 genes in the ABA dataset. For computational efficiency, we created a support mask representing all the voxels that belong to the mouse brain using the CCF reference map. The 3D gene data was then cropped using this support mask. This allowed our osNMF framework to only analyze the areas representing the brain. This 3D brain

area was 55,954 voxels, vs. the total cube array of 159,326 voxels, reducing the number of voxels fed into osNMF by roughly two-thirds. Once the analysis was run, we unmasked the analysis outcomes and transformed the data back to the full 3D space (67x41x58).

The osNMF framework

As shown in **Figure 1A,** osNMF consists of three processing modules: an imputation module to impute missing voxels in 3D gene expression dataset, the stability-driven NMF (staNMF) core, and a search algorithm that systematically identifies the combinations of known anatomical regions that are most similar to a PP. For the first module, a k-nearest neighbors (kNN) algorithm[35] with 6 neighbors was used to estimate and impute the missing voxels in the gene expression atlas. To test the efficacy of the kNN algorithm, we calculated accuracy on a hold-out test set of 1,000 random voxels for each of the 4,345 genes from the ABA dataset (for a total of 4,435,000 data points). For the second module, we used the stability-driven non-negative matrix factorization (staNMF)[15] technique to decompose data into dictionary elements and associated coefficients for each gene. We first transformed the data into a matrix of voxels by genes (of size 55,954 by 4,345). We then used NMF, a widely-used unsupervised learning algorithm for dimensionality reduction[58], to decompose the gene data matrix into spatially segmented principal patterns (PPs) in the brain. The non-negativity constraint enables the learning of parts-based representations[59]. Formally, let $X = [x_1, x_2, \ldots, x_v]$, be a $v \times n$ matrix representing the input data, where v is the number of unique voxels and n is the number of genes represented. Let $D = [d_1, d_2, \ldots, d_K]$, be a $v \times K$ matrix, representing a dictionary with K atoms, and $A = [a_1, a, \ldots, a_n]$, be a $K \times n$ matrix, representing the coefficient matrix. NMF aims to minimize the following objective function:

$$Loss\ (D, A) = \sum_{ij} \bar{X}$$

where $\bar{X} = (X - DA)^2$. We used the sklearn implementation of NMF, integrated with the staNMF implementation[15] and adapted this to fit within the osNMF pipeline. The Frobenius norm was used for the loss function with 0.0001 as the tolerance of the stopping condition and 200 as the maximum number of iterations before timing out. For the stability analysis to identify the number of dictionary elements (PPs), K, we used a stability-driven criterion[15] for systematic and automated selection of the number of PPs (**Figure 1B**). This instability score measures the instability of the learned PPs over 100 runs of the same algorithm across a range of 8 to 30 possible numbers of PPs, K. We estimated the instability of the PP by computing the average dissimilarity of all learned dictionary pairs ($\mathbf{D}$ and $\mathbf{D}'$) using their cross-correlation matrix ($\mathbf{C}$) and two dissimilarity functions: (1) Hungarian matching method[38] (**Figure 1B**) and (2) an Amari-type error function[15] (**Figure S1**).

Moran's Index

To quantify the spatial coherence of PPs, we used the Moran's Index (Moran's I)[39]. Moran's I was originally used in geostatistics and has more recently been used in spatial gene expression literature[40–42]. Moran's I ranges in value from –1 to 1. A value close to -1 indicates little spatial organization, similar to a chess board with black and white squares distributed across the board. A value close to 1 indicates a clear spatially distinct pattern, such as if all the black squares in a chess board were on one side and all white squares on the other.

We calculated Moran's I using the following formula[40]:

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

x_i and x_j represent the PP coefficient at voxel locations i and j. $\bar{x}$ is the mean expression level of each PP. N is the total number of voxel locations, w_{ij} is the spatial adjacency relationship (based on the adjacency matrix, w) between voxels i and j. W is the sum of w, which represents the cumulative total adjacencies. We mask the dataset to only include the brain region. Then, for each voxel, we select up to 6 voxels for determining adjacency (up, down, left, right, forward, background, where available). This is the "rook" method for identifying adjacency (as compared to the "queen" method which includes diagonal adjacencies). We assign $w_{ij} = 1$ if voxel j is adjacent i, and $w_{ij} = 0$ otherwise.

Given the large size of the adjacency matrix (159,326 x 159,326), we downsampled the PPs by removing every other row in each of the three dimensions to improve computational efficiency. Given certain voxels had multiple PPs with small but non-zero coefficients, we assigned each voxel in the brain map to the PP with the highest coefficient for that voxel. This ensures that unique voxels are not represented by multiple PPs. Thus, the adjacency matrix (which is binary, with 1 for adjacency and 0 for non-adjacency) cleanly reflects the correct adjacencies. A possible extension of this work would be to develop a continuous adjacency matrix.

3D visualizations of PPs

The 3D gene visualizations were performed using Napari viewer, a multi-dimensional image viewer for Python[60]. Key settings in Napari for PPs included: opacity=1, gamma=1, blending='additive', depiction='volume', and rendering='attenuated MIP'. MIP stands for maximum intensity projection, which enhances the 3D representation of objects. We moved the slide bar to 20% from the left side for 'attenuated MIP.'

Data availability

The data used in this study is publicly available under the Allen Brain Atlas (ABA) [https://mouse.brain-map.org/]. The intermediate files are freely available at https://github.com/abbasilab/osNMF.

Code availability

The software package is freely available at https://github.com/abbasilab/osNMF.

Contributions

RA, YW, BY, BT, and HZ conceived the research design. RA and YW fitted the model and implemented the pattern discovery pipeline, marker genes analysis, and gene network reconstruction. RC and RA designed and implemented the ontology discovery pipeline and the spatial coherence analysis. AL implemented the visualization platforms. BY, BT, HZ, and AL

provided input on the analysis. RA and RC wrote the manuscript with inputs from BT, HZ, and BY.

Acknowledgments

The authors would like to thank Lydia Ng and Zizhen Yao for their constructive feedback. RA, BT, BY, and HZ would like to acknowledge support from the Weill Neurohub through the Weill Neurohub's Next Great Ideas Award. RA would like to acknowledge support from Sandler Program for Breakthrough Biomedical Research, which is partially funded by the Sandler Foundation.

Competing Interests

The authors declare no competing interests.

References

1. Tasic, B. Single cell transcriptomics in neuroscience: cell classification and beyond. *Current opinion in neurobiology* **50**, 242–249 (2018).
2. Zeng, H. & Sanes, J. R. Neuronal cell-type classification: challenges, opportunities and the path forward. *Nature Reviews Neuroscience* **18**, 530–546 (2017).
3. Tasic, B. *et al.* Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).
4. Yao, Z. *et al.* A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222–3241 (2021).
5. Cooper-Knock, J. *et al.* Gene expression profiling in human neurodegenerative disease. *Nature Reviews Neurology* **8**, 518–530 (2012).
6. Baechler, E. C. *et al.* Gene expression profiling in human autoimmunity. *Immunological reviews* **210**, 120–137 (2006).
7. Reis-Filho, J. S. & Pusztai, L. Gene expression profiling in breast cancer: classification, prognostication, and prediction. *The Lancet* **378**, 1812–1823 (2011).
8. Cookson, W., Liang, L., Abecasis, G., Moffatt, M. & Lathrop, M. Mapping complex disease traits with global gene expression. *Nature Reviews Genetics* **10**, 184–194 (2009).
9. Mirnics, K., Middleton, F. A., Lewis, D. A. & Levitt, P. Analysis of complex brain disorders with gene expression microarrays: schizophrenia as a disease of the synapse. *Trends in neurosciences* **24**, 479–486 (2001).
10. Wang, Q. *et al.* The Allen mouse brain common coordinate framework: a 3D reference atlas. *Cell* **181**, 936–953 (2020).
11. Ng, L. *et al.* An anatomic gene expression atlas of the adult mouse brain. *Nature*

neuroscience **12**, 356–362 (2009).
12. Close, J. L., Long, B. R. & Zeng, H. Spatially resolved transcriptomics in neuroscience. *Nature Methods* **18**, 23–25 (2021).
13. Lein, E., Borm, L. E. & Linnarsson, S. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science* **358**, 64–69 (2017).
14. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
15. Wu, S. *et al.* Stability-driven nonnegative matrix factorization to interpret spatial gene expression and build local gene networks. *Proceedings of the National Academy of Sciences* **113**, 4290–4295 (2016).
16. Wang, F. *et al.* Developmental topography of cortical thickness during infancy. *Proceedings of the National Academy of Sciences* **116**, 15855–15860 (2019).
17. Fang, X. *et al.* Global transcriptional regulatory network for Escherichia coli robustly connects gene expression to transcription factor activities. *Proceedings of the National Academy of Sciences* **114**, 10286–10291 (2017).
18. Ortiz, C. *et al.* Molecular atlas of the adult mouse brain. *Science Advances* **6**, eabb3446.
19. Yang, Y., Fang, Q. & Shen, H.-B. Predicting gene regulatory interactions based on spatial gene expression data and deep learning. *PLoS computational biology* **15**, e1007324 (2019).
20. He, B. *et al.* Integrating spatial gene expression and breast tumour morphology via deep learning. *Nature biomedical engineering* **4**, 827–834 (2020).
21. Zeng, T., Li, R., Mukkamala, R., Ye, J. & Ji, S. Deep convolutional neural networks for annotating gene expression patterns in the mouse brain. *BMC bioinformatics* **16**, 1–10 (2015).
22. Ng, L. *et al.* Neuroinformatics for genome-wide 3-D gene expression mapping in the mouse brain. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **4**, 382–393 (2007).
23. Morris, J. A. *et al.* Divergent and nonuniform gene expression patterns in mouse brain. *Proceedings of the National Academy of Sciences* **107**, 19049–19054 (2010).
24. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
25. Maynard, K. R. *et al.* Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature Neuroscience* **24**, 425–436 (2021).
26. Dries, R. *et al.* Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biology* **22**, 78 (2021).
27. Lopez, R. *et al.* DestVI identifies continuums of cell types in spatial transcriptomics data. *Nature Biotechnology* **40**, 1360–1369 (2022).
28. Hu, J. *et al.* Statistical and machine learning methods for spatially resolved transcriptomics with histology. *Comput Struct Biotechnol J* **19**, 3829–3841 (2021).
29. Dries, R. *et al.* Advances in spatial transcriptomic data analysis. *Genome Res* **31**, 1706–1718 (2021).
30. Zeng, Z., Li, Y., Li, Y. & Luo, Y. Statistical and machine learning methods for spatially

resolved transcriptomics data analysis. *Genome Biology* **23**, 83 (2022).

31. Li, Y., Stanojevic, S. & Garmire, L. X. Emerging artificial intelligence applications in Spatial Transcriptomics analysis. *Computational and Structural Biotechnology Journal* **20**, 2895–2908 (2022).
32. Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R. & Yu, B. Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences* **116**, 22071–22080 (2019).
33. Yu, B. & Kumbier, K. Veridical data science. *Proceedings of the National Academy of Sciences* **117**, 3920–3929 (2020).
34. Bin Yu. Stability. *Bernoulli* **19**, 1484–1500 (2013).
35. Fix, E. & Hodges, J. L. Discriminatory analysis. Nonparametric discrimination: Consistency properties. *International Statistical Review/Revue Internationale de Statistique* **57**, 238–247 (1989).
36. Altman, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* **46**, 175–185 (1992).
37. Hsieh, C.-J. & Dhillon, I. S. Fast Coordinate Descent Methods with Variable Selection for Non-Negative Matrix Factorization. in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1064–1072 (Association for Computing Machinery, 2011). doi:10.1145/2020408.2020577.
38. Kuhn, H. The Hungarian Method for the assignment problem. *Naval Res Logist* **Q2**, 83–97 (1955).
39. Li, H., Calder, C. & Cressie, N. Beyond Moran's I: Testing for Spatial Dependence Based on the Spatial Autoregressive Model. *Geographical Anal.* 357–375 (2007) doi:https://doi.org/10.1111/j.1538-4632.2007.00708.x.
40. Hu, J. *et al.* SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nature Methods* **18**, 1342–1351 (2021).
41. Cable, D. M. *et al.* Robust decomposition of cell type mixtures in spatial transcriptomics. *Nature Biotechnology* **40**, 517–526 (2022).
42. Abdelaal, T., Mourragui, S., Mahfouz, A. & Reinders, M. J. T. SpaGE: Spatial Gene Enhancement using scRNA-seq. *Nucleic Acids Research* **48**, e107–e107 (2020).
43. Yao, Z. *et al.* A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *bioRxiv* 2023.03.06.531121 (2023) doi:10.1101/2023.03.06.531121.
44. Pham, D. *et al.* stLearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. *bioRxiv* 2020.05.31.125658 (2020) doi:10.1101/2020.05.31.125658.
45. 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 (2018).
46. He, M. *et al.* Strategies and Tools for Combinatorial Targeting of GABAergic Neurons in Mouse Cerebral Cortex. *Neuron* **91**, 1228–1243 (2016).
47. Lavado, A. & Oliver, G. Prox1 expression patterns in the developing and adult murine brain.

Developmental Dynamics **236**, 518–524 (2007).

48. Wang, W. *et al.* A Role for Nuclear Factor I in the Intrinsic Control of Cerebellar Granule Neuron Gene Expression *. *Journal of Biological Chemistry* **279**, 53491–53497 (2004).
49. Acharyya, S., Zhou, X. & Baladandayuthapani, V. SpaceX: gene co-expression network estimation for spatial transcriptomics. *Bioinformatics* **38**, 5033–5041 (2022).
50. Salamon, J., Qian, X., Nilsson, M. & Lynn, D. J. Network Visualization and Analysis of Spatially Aware Gene Expression Data with InsituNet. *Cell Systems* **6**, 626-630.e3 (2018).
51. Wu, Z. *et al.* SPACE-GM: geometric deep learning of disease-associated microenvironments from multiplex spatial protein profiles. *bioRxiv* 2022.05.12.491707 (2022) doi:10.1101/2022.05.12.491707.
52. Yuan, Y. & Bar-Joseph, Z. GCNG: graph convolutional networks for inferring gene interaction from spatial transcriptomics data. *Genome Biology* **21**, 300 (2020).
53. Karadurmus, D. *et al.* GPRIN3 Controls Neuronal Excitability, Morphology, and Striatal-Dependent Behaviors in the Indirect Pathway of the Striatum. *J. Neurosci.* **39**, 7513 (2019).
54. Montnach, J., Agullo-Pascual, E., Tadros, R., Bezzina, C. R. & Delmar, M. Bioinformatic analysis of a plakophilin-2-dependent transcription network: implications for the mechanisms of arrhythmogenic right ventricular cardiomyopathy in humans and in boxer dogs. *EP Europace* **20**, iii125–iii132 (2018).
55. Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).
56. Harris, J. A. *et al.* Hierarchical organization of cortical and thalamic connectivity. *Nature* **575**, 195–202 (2019).
57. Basu, S., Kumbier, K., Brown, J. B. & Yu, B. Iterative random forests to discover predictive and stable high-order interactions. *Proceedings of the National Academy of Sciences* **115**, 1943–1948 (2018).
58. Y. -X. Wang & Y. -J. Zhang. Nonnegative Matrix Factorization: A Comprehensive Review. *IEEE Transactions on Knowledge and Data Engineering* **25**, 1336–1353 (2013).
59. Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **401**, 788–791 (1999).
60. Napari: Multi-dimensional image viewer for python. Information available at https://napari.org/stable/.

Supplementary Figures

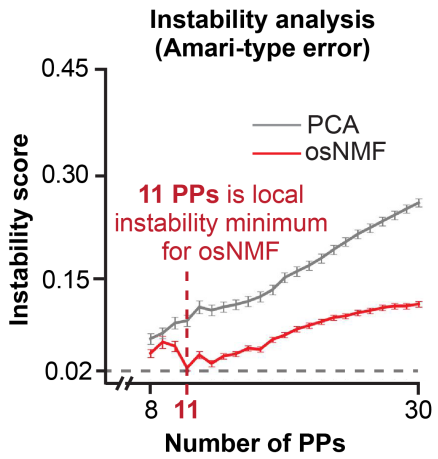

Figure S1: Instability analysis with Amari-type error function. Instability score of osNMF PPs and PCA PPs across 100 runs for each K value, from 8 to 30 for ABA dataset. The ABA dataset contains 4,345 3D in situ hybridization (ISH) images at 200 μm^3 resolution of the adult mouse brain. Error bars show the standard deviation. This figure uses Amari-type error, while Figure 1B uses the Hungarian matching method. Both approaches identify K = 11 for the minimum instability score (and thus most stability) for osNMF PPs.

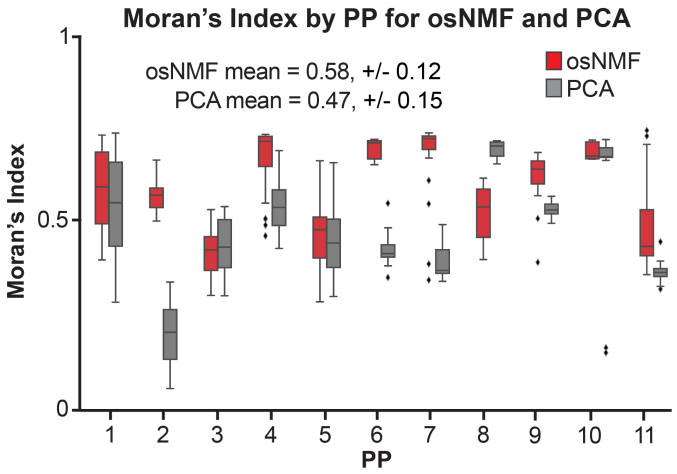

Figure S2: Moran's Index for osNMF and PCA at each PP. Data are from 20 bootstrapped simulations for each PP, for a total of 220 simulations for osNMF PPs and 220 simulations for PCA PPs. The mean Moran's I was 0.58 ± 0.12 for osNMF and 0.47 ± 0.15 for PCA. The p-value between the two samples was <0.001. osNMF PPs show greater spatial coherence than PCA PPs, as measured by a Moran's I value closer to 1.

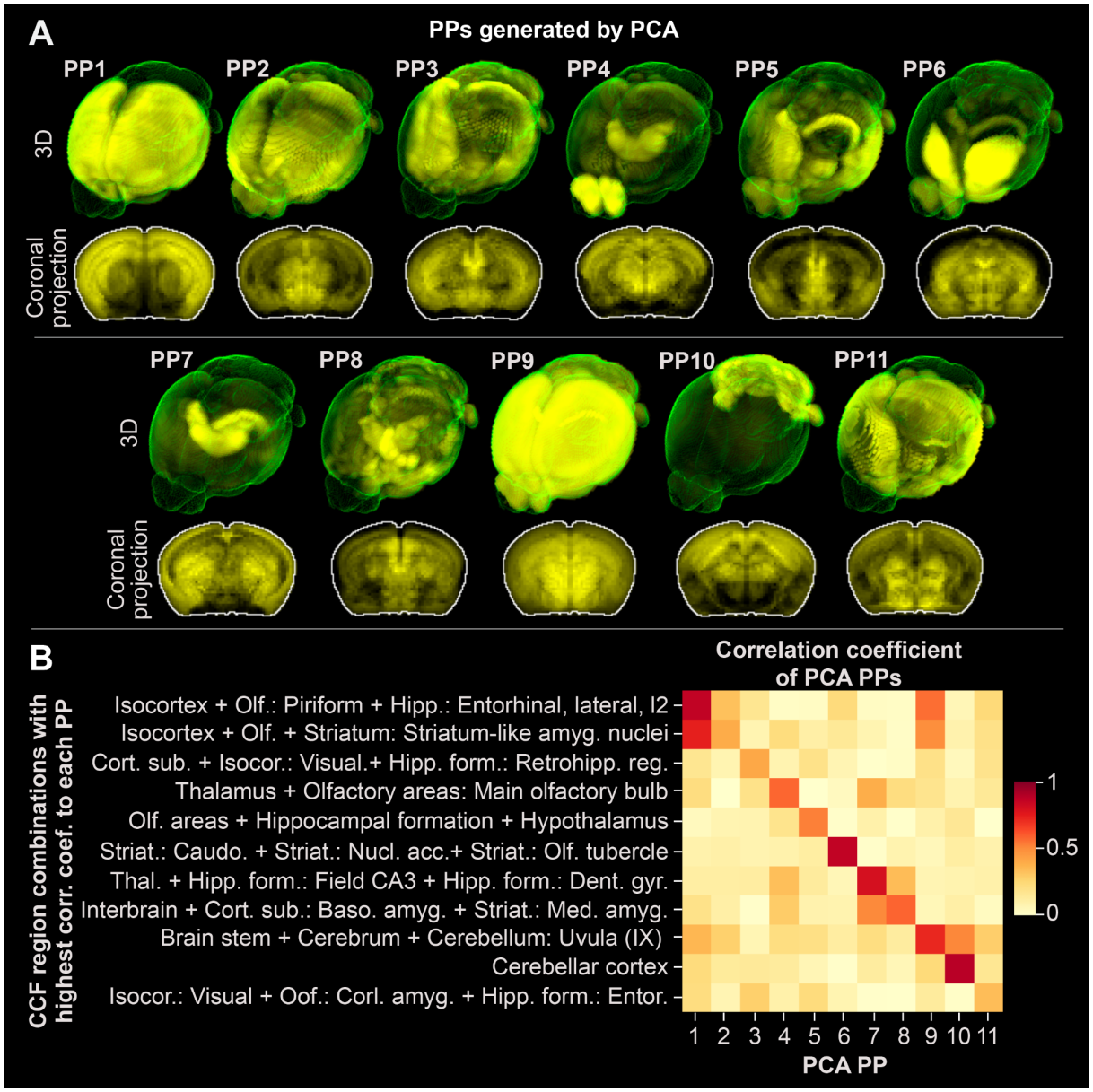

Figure S3: Similarity of PCA PPs to the expert-annotated brain regions. A. 11 PPs generated by PCA, ordered based on highest coarse region correlation to CCF ontology in 3D and projected on the coronal plane. **B.** Heat map of the correlation coefficient between PCA PPs and the most similar combination of CCF regions (with the highest correlation coefficient).

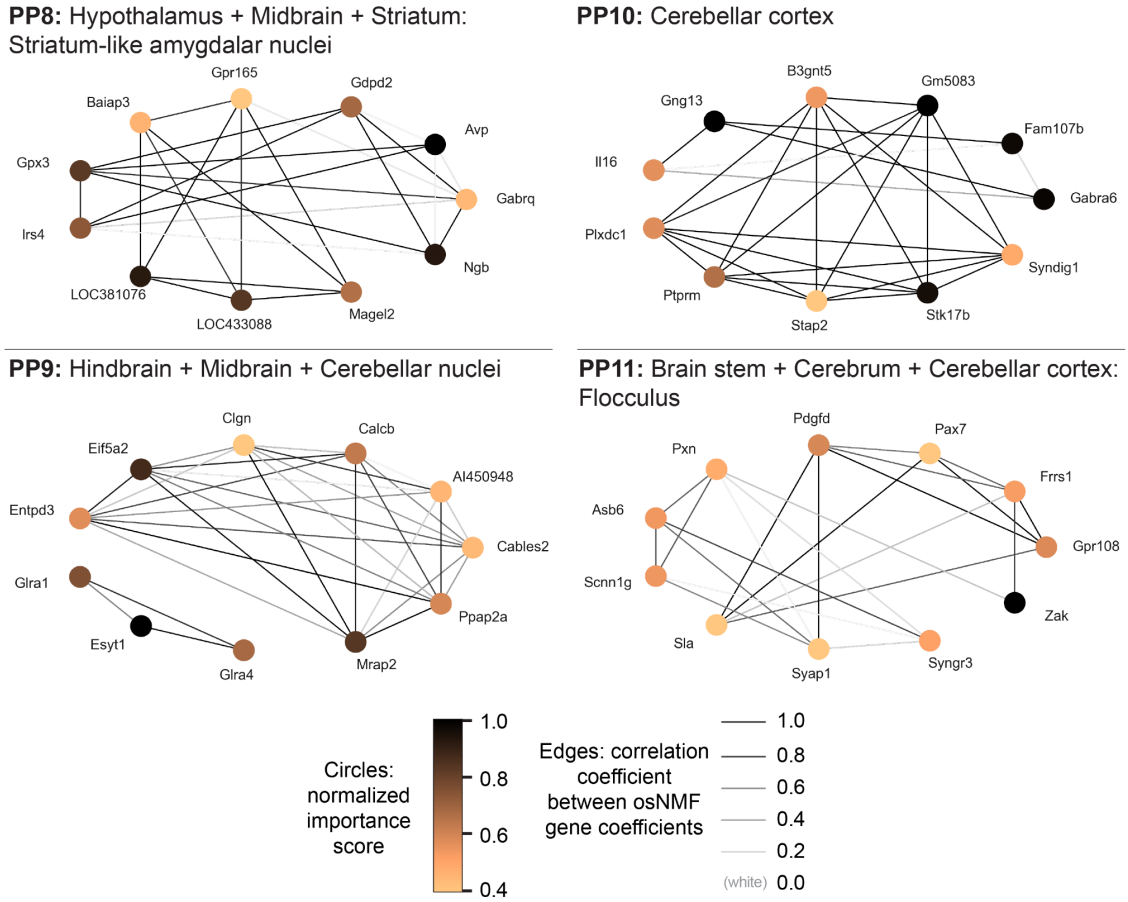

Figure S4: Local gene interaction networks constructed for the 4 remaining brain regions not shown in Figure 7. The node color presents the selectivity of the gene to the PP associated with the brain region. An edge is drawn between genes if the similarity score is among the top 5% of all similarity scores for that gene subset. The edge thickness is proportional to the similarity scores between the osNMF representation of the two genes.