

---

# **The cell-type underpinnings of the human functional cortical connectome**

---

In the format provided by the  
authors and unedited

---

**SUPPLEMENT:** The cell-type underpinnings of the human functional cortical connectome

Xi-Han Zhang (张喜寒)<sup>1,\*</sup>, Kevin M. Anderson<sup>1</sup>, Hao-Ming Dong (董昊铭)<sup>1</sup>, Sidhant Chopra<sup>1</sup>,  
Elvisha Dhamala<sup>2</sup>, Prashant S. Emani<sup>3</sup>, Mark B. Gerstein<sup>3,4,5,6</sup>, Daniel Margulies<sup>7</sup>, Avram J.  
Holmes<sup>1,8,9,\*</sup>

<sup>1</sup> Department of Psychology, Yale University, New Haven, CT, USA

<sup>2</sup> Institute of Behavioral Science, Feinstein Institutes for Medical Research, Manhasset, New York, USA

<sup>3</sup> Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA

<sup>4</sup> Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA;

<sup>5</sup> Department of Computer Science, Yale University, New Haven, CT, USA;

<sup>6</sup> Department of Statistics and Data Science, Yale University, New Haven, CT, USA

<sup>7</sup> Centre National de la Recherche Scientifique, Integrative Neuroscience and Cognition Center (UMR 8002), Université de Paris, Paris, France

<sup>8</sup> Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA

<sup>9</sup> Department of Psychiatry, Brain Health Institute, Rutgers University, Piscataway, NJ, USA

**Correspondence:**

\*Xihan Zhang, Yale University, Department of Psychology, 100 College Street, New Haven, CT 06520, Email: xihan.zhang@yale.edu.

\*Avram J. Holmes, Rutgers University, Department of Psychiatry, Brain Health Institute, 119 Staged Research Building, 661 Hoes Lane West, Piscataway, NJ 08854, Email: avram.holmes@rutgers.edu.

| <b>Multivariate Cell Type Abundances Correlation with Gradient 1</b> |          |                        |
|----------------------------------------------------------------------|----------|------------------------|
| <b>Removed Cell Types</b>                                            | <b>r</b> | <b>p<sub>FDR</sub></b> |
| Sncg, Sst, L5 IT, L4 IT                                              | 0.503    | 0.002                  |
| Sncg, Sst, L5 IT                                                     | 0.591    | 0.002                  |
| L4 IT                                                                | 0.609    | 0.002                  |
| Sncg, L4 IT                                                          | 0.566    | 0.002                  |
| Sst, L4 IT                                                           | 0.601    | 0.002                  |
| L5 IT, L4 IT                                                         | 0.559    | 0.002                  |
| Sncg                                                                 | 0.609    | 0.002                  |
| Sst                                                                  | 0.609    | 0.002                  |
| L5 IT                                                                | 0.609    | 0.002                  |

**Table S1. Multivariate cellular (imputed from Jorstad snRNA-seq) correlation with Gradient 1 after iteratively removing cell types individually associated with Gradient 1.**

Cell types that are univariately correlated with the first principal gradient are removed from CCA in different combinations. The top row is the CCA performed without the combination of all the Gradient 1-related cell types surviving from significance tests. The rows coming after are CCA performed excluding different combinations of Gradient 1-related cell types surviving from univariate significance tests. The correlations between multivariate cell types and Gradient 1 remain above 0.503 for all removal combinations. These data suggest the cellular correlation with Gradient 1 is preserved across a host of different cell types.

| Multivariate Cell Type Abundances Correlation with Gradient 1 |       |                  |                                         |       |                  |
|---------------------------------------------------------------|-------|------------------|-----------------------------------------|-------|------------------|
| Frontal Cortex Cell Signature (Lake DFC)                      |       |                  | Visual Cortex Cell Signature (Lake VIS) |       |                  |
| Removed Cell Types                                            | r     | p <sub>FDR</sub> | Removed Cell Types                      | r     | p <sub>FDR</sub> |
| Ex1, In1, PVALB                                               | 0.419 | 0.029            | Ex1, In1, PVALB                         | 0.538 | 0.002            |
| Ex1                                                           | 0.508 | 0.003            | Ex1                                     | 0.561 | 0.002            |
| In1                                                           | 0.502 | 0.003            | In1                                     | 0.548 | 0.002            |
| End                                                           | 0.513 | 0.003            | Ex5                                     | 0.535 | 0.002            |
| PVALB                                                         | 0.505 | 0.003            | PVALB                                   | 0.564 | 0.002            |
| Ex1, End                                                      | 0.493 | 0.003            | Ex5, PVALB                              | 0.529 | 0.002            |
| In1, PVALB                                                    | 0.477 | 0.008            | Ex1, PVALB                              | 0.560 | 0.002            |
| Ex1, PVALB                                                    | 0.469 | 0.008            | In1, PVALB                              | 0.546 | 0.002            |
| In1, End                                                      | 0.491 | 0.003            | Ex5, Ex1, In1                           | 0.519 | 0.002            |
| Ex1, In1                                                      | 0.479 | 0.010            | Ex5, Ex1, In1, PVALB                    | 0.509 | 0.002            |
| End, PVALB                                                    | 0.484 | 0.003            |                                         |       |                  |
| Ex1, In1, End, PVALB                                          | 0.366 | 0.096            |                                         |       |                  |

**Table S2. Multivariate cellular (imputed from Lake dataset) correlation with Gradient 1 after iteratively removing cell types individually associated with Gradient 1.** Cell types that are univariately correlated with the first principal gradient are removed from CCA in different combinations. The top row is the CCA performed without the combination of all the Gradient 1-related cell types surviving from significance tests in both Lake DFC and Lake VIS imputed cell type abundances. The rows coming after are CCA performed excluding different combinations of Gradient 1-related cell types surviving from univariate significance tests in Lake DFC (Ex1, In1, End, PVALB) or Lake VIS (Ex1, In1, Ex5, PVALB) imputed cell type abundances. The correlations between multivariate cell types and Gradient 1 remain above 0.419 for all removal combinations, with one exception (removing all significant univariate cell types in Lake DFC: Ex1, In1, End, PVALB). These data suggest the cellular correlation with Gradient 1 is preserved across a host of different cell types.

| Multivariate Cell Type Abundances Correlation with Gradient 2 |       |                  |                                         |       |                  |
|---------------------------------------------------------------|-------|------------------|-----------------------------------------|-------|------------------|
| Frontal Cortex Cell Signature (Lake DFC)                      |       |                  | Visual Cortex Cell Signature (Lake VIS) |       |                  |
| Removed Cell Types                                            | r     | p <sub>FDR</sub> | Removed Cell Types                      | r     | p <sub>FDR</sub> |
| Ex4                                                           | 0.497 | 0.060            | Ex4                                     | 0.546 | 0.003            |
| End                                                           | 0.509 | 0.060            |                                         |       |                  |
| Ex4, End                                                      | 0.393 | 0.220            |                                         |       |                  |

**Table S3. Multivariate cellular (imputed from Lake dataset) correlation with Gradient 2 after iteratively removing cell types individually associated with Gradient 2.** Cell types that are univariately correlated with the second principal gradient are removed from CCA. The top row is the CCA performed without the combination of all the Gradient 2-related cell types surviving from significant tests in both Lake DFC and Lake VIS imputed cell type abundances (Ex4 only). The rows coming after are CCA performed without different combinations of Gradient 2-related cell types surviving from significance tests in Lake DFC (Ex4 and End) or Lake VIS (Ex4 only) imputed cell type abundances. The correlations between multivariate cell types and Gradient 2 remain above 0.497 for all combinations, with one exception (removing all significant univariate cell types in Lake DFC: Ex4 and End).

| Networks     | Imputed from Lake |       |       |       | Imputed from Jorstad |       |       |       |
|--------------|-------------------|-------|-------|-------|----------------------|-------|-------|-------|
|              | Mean              | Max   | Min   | Std   | Mean                 | Max   | Min   | Std   |
| Som/Motor    | 0.605             | 1.000 | 0.000 | 0.390 | 0.504                | 1.000 | 0.000 | 0.366 |
| Visual       | 0.674             | 1.000 | 0.000 | 0.390 | 0.566                | 1.000 | 0.000 | 0.417 |
| DorsAttn     | 0.271             | 0.789 | 0.007 | 0.238 | 0.351                | 0.921 | 0.000 | 0.296 |
| Sal/VentAttn | 0.323             | 0.991 | 0.000 | 0.351 | 0.348                | 0.986 | 0.000 | 0.346 |
| Limbic       | 0.654             | 1.000 | 0.000 | 0.371 | 0.559                | 0.993 | 0.007 | 0.362 |
| Control      | 0.225             | 0.819 | 0.000 | 0.236 | 0.254                | 0.865 | 0.000 | 0.255 |
| Default      | 0.264             | 0.864 | 0.000 | 0.246 | 0.223                | 0.965 | 0.000 | 0.230 |

**Table S4. Summary statistics for parcel-level classification accuracy across the seven functional networks.** Parcel-level accuracy was obtained from SVM classifiers trained and tested on the samples from both hemispheres. On the left are the results of SVMs ensembling Lake DFC and Lake VIS imputed cell type abundance. Right are the results of SVMs trained from Jorstad imputed cell type abundance. Statistics were calculated from prediction accuracy of parcels belonging to the same network.

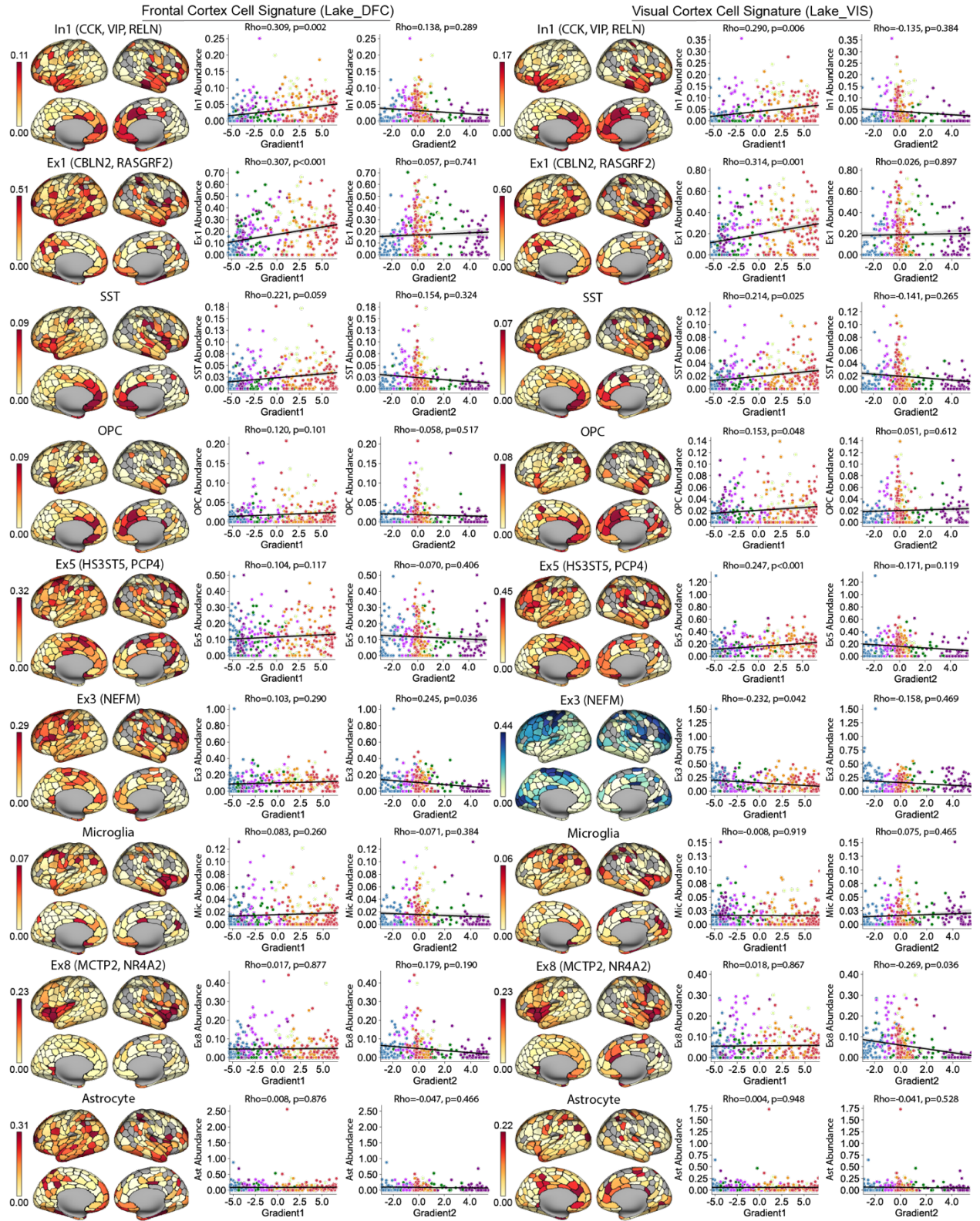


**Fig. S1 Spatial distribution of cell types (imputed from Lake snRNA-seq) preferentially distributed in association cortex.** The imputed cell type abundance across cortex in both hemispheres are aggregated in 400 Schaffer parcels. Cell types imputed from single-nucleus droplet (snDrop) samples in the frontal lobe (Lake DFC) are on the left panel and the right is imputed from visual cortex-sourced snDrop samples (Lake VIS). Within each panel, the scatter plot on the left shows the correlation between cell abundance and the first principle gradient (Gradient 1) across cortical parcels, and the correlation with the second gradient (Gradient 2) is displayed on the right. Correlations were estimated by Spearman's Rho, but for visual reference, the fitted linear regression lines (95% CI) predicting cellular abundance from gradient values are also displayed. P-values were constructed from two-sided spin-tests that control spatial autocorrelation. Dots are color-coded by the functional network each parcel belongs to. A positive correlation with Gradient 1 indicates the cell is preferentially spaced on the transmodal association cortex. Such distribution preference reflected on the scatter plot of Gradient 2 is a peak near 0. Though only the top two cells (Ex1 and In1) have significant positive relationships with Gradient 1, the patterns are visually identifiable in the rest of the top five cells (SST, OPC, and Ex5).

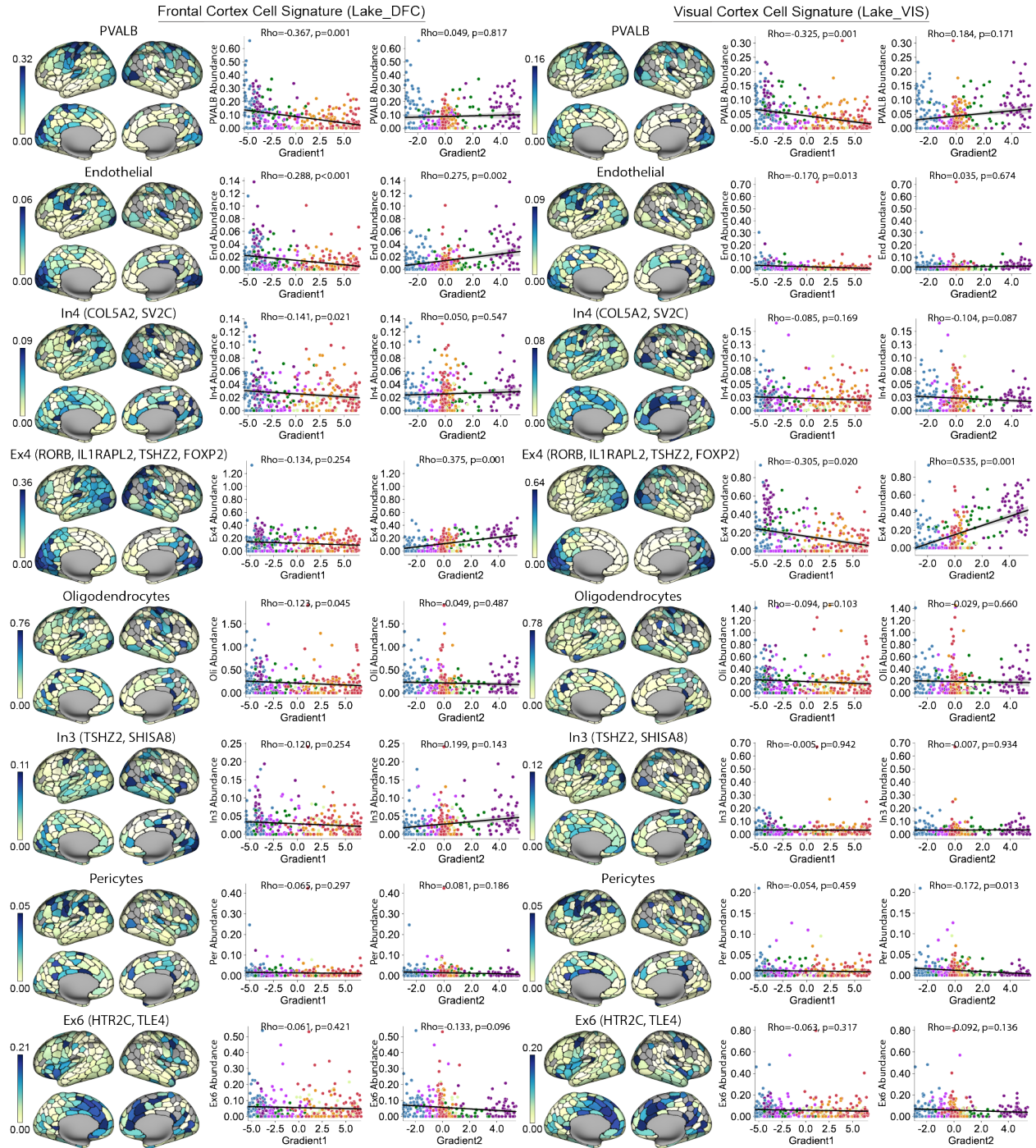


**Fig. S2 Spatial distribution of cell types (imputed from Lake snRNA-seq) preferentially distributed in the unimodal area.** The imputed cell type abundance across cortex in both hemispheres are aggregated in 400 Schaffer parcels. Cell types imputed from single-nucleus droplet (snDrop) samples in the frontal lobe (Lake DFC) are on the left panel and the right is imputed from visual cortex-sourced snDrop samples (Lake VIS). Within each panel, the scatter

plot on the left shows the correlation between cell abundance and the first principle gradient (Gradient 1) across cortical parcels, and the correlation with the second gradient (Gradient 2) is displayed on the right. Correlations were estimated by Spearman's Rho, but for visual reference, the fitted linear regression lines (95% CI) predicting cellular abundance from gradient values are also displayed. P-values were constructed from two-sided spin-tests that control spatial autocorrelation. Dots are color-coded by the functional network each parcel belongs to. A negative correlation with Gradient 1 indicates the cell is preferentially spaced on the unimodal area. On the scatter plots, this is the peak near the negative end of Gradient 1. On Gradient 2, this is the peak on the visual or somato/motor or both ends. Though only the top two cells (PVALB and End) have significant negative relationships with Gradient 1, the patterns are visually evident in the rest of the top four cells (In4 and Ex4).

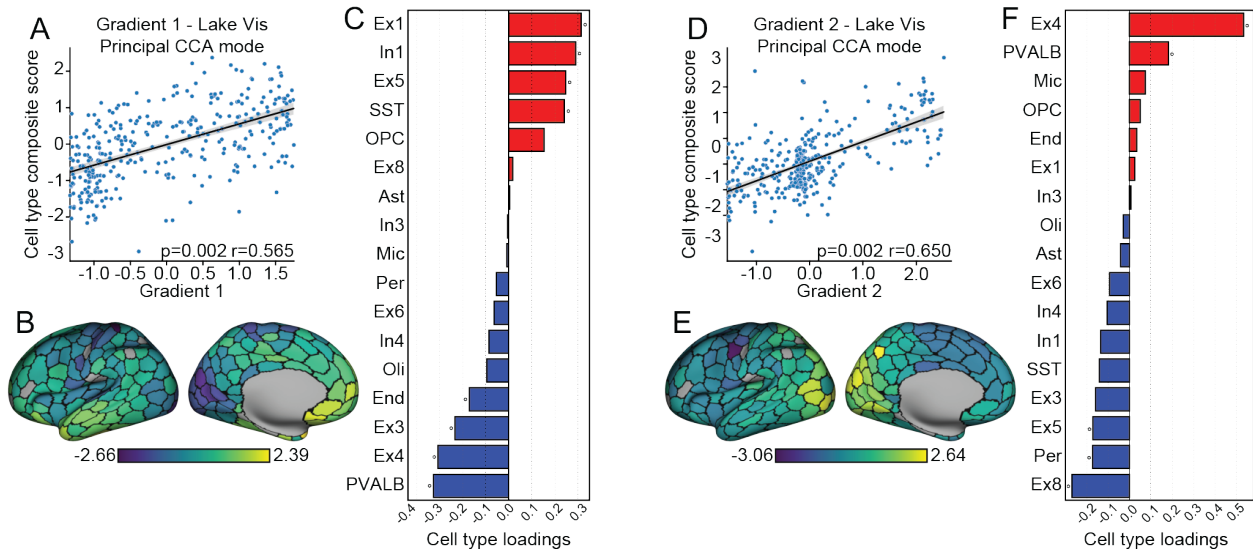


**Fig. S3 Multivariate cellular (imputed from Lake VIS) profiles follow the macroscale organization of cortex.** The analysis performed based on the Lake VIS imputed cell type abundance showed similar correlation strength and pattern with functional gradients, consistent with the results of Lake DFC (in **Extended Data Fig. 7**). **A.** The scatter plot displays permutational canonical correlation analysis modeling results. Across cortical parcels, the first functional gradient was positively associated with the cell type composite score (95% CI). **B.** Cell type composite score associated with the first functional gradient projected to the cortical surface. **C.** Cell type loadings from the first gradient indicate the correlation is strongest within previously identified Ex1, In1, and PVALB (red indicates positive associations; blue, negative associations). **D.** The second functional gradient was positively associated with the cell type composite score (95% CI). **E.** Cell type composite score associated with the second functional gradient mapped to the cortical surface. **F.** Ex4 and Ex8 (MCTP2, NR4A2) excitatory neurons remain strong contributors to the correlation with Gradient 2. P-values in **A** and **D** are calculated from two-sided spin-tests after correcting the two comparisons.

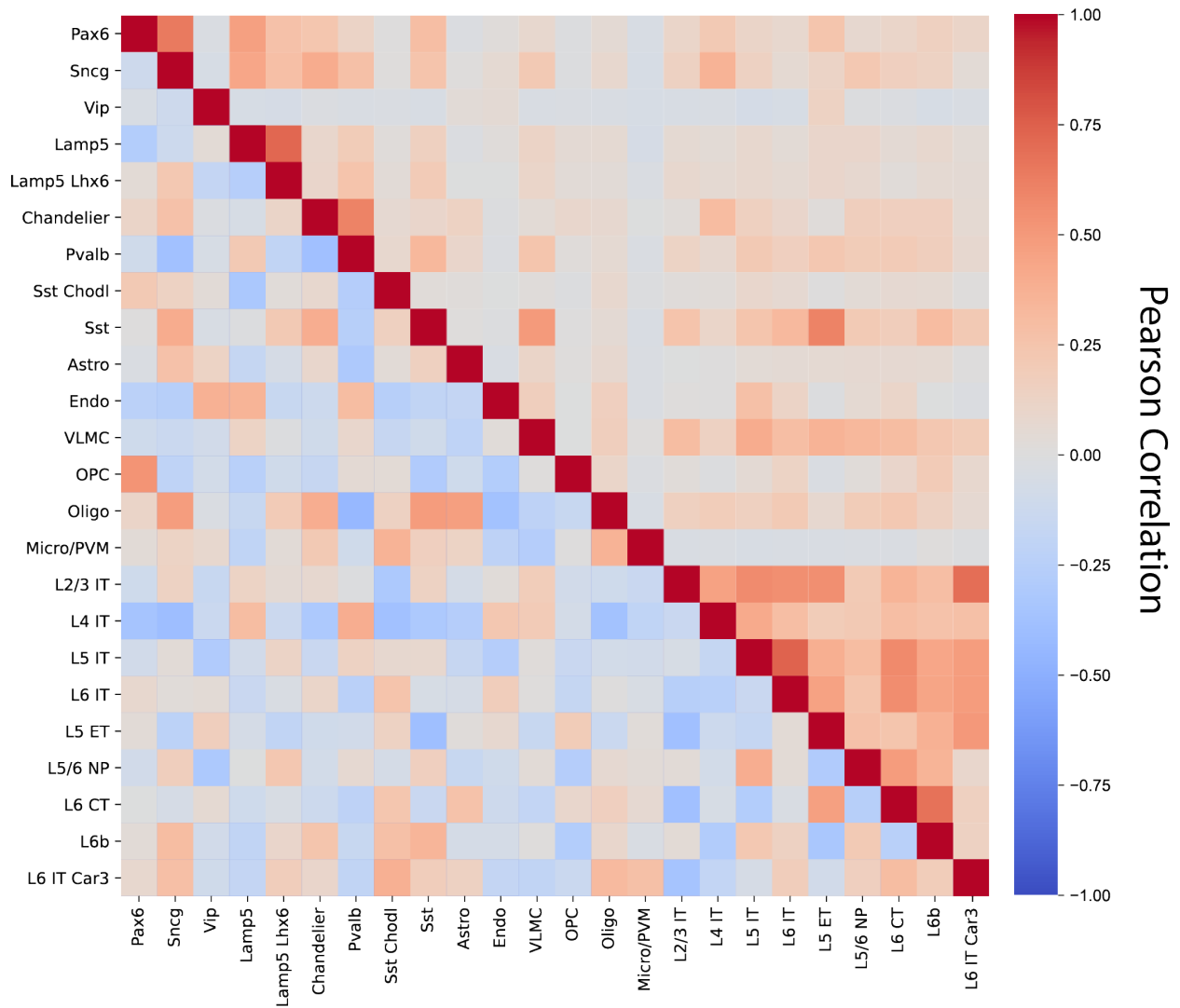


**Fig. S4 Spatial correlation of cortical cell type abundances and gene signature correlation (imputed from Jorstad snRNA-seq).** Elements in the lower triangle represent the spatial correlation of cell abundance across cortical parcels. Upper triangle displays the expression correlation across the marker genes. The abundance of each cell type across cortical parcels is correlated with other cell types, whereas the gene signatures have subtle collinearity. A higher correlation (in red) indicates more similar spatial distribution or gene expression similarity between cell types.

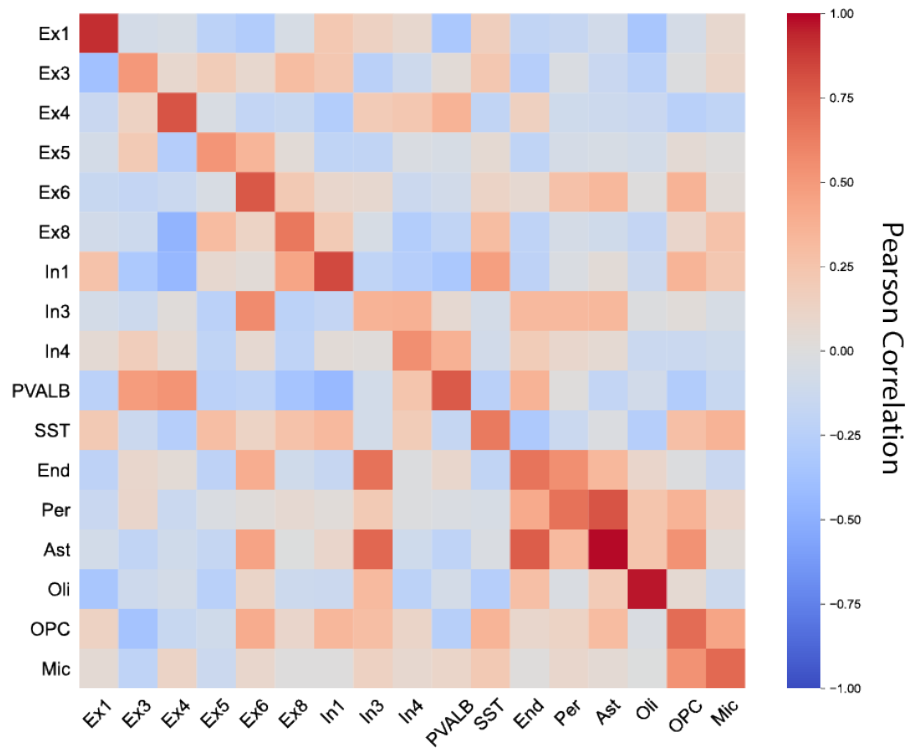


**Fig. S5 Spatial correlation of cortical cell type abundances (imputed from Lake snRNA-seq).** The abundance of each cell type across cortical parcels is correlated with other cell types. A higher correlation (in red) indicates more similar spatial distribution between cell types. Elements in the diagonal represent the correlation between the same cell type imputed from Lake DFC and Lake VIS. Elements in the upper triangle represent the correlation between two different cell types imputed from Lake DFC and the lower triangle are the correlations imputed from Lake VIS.

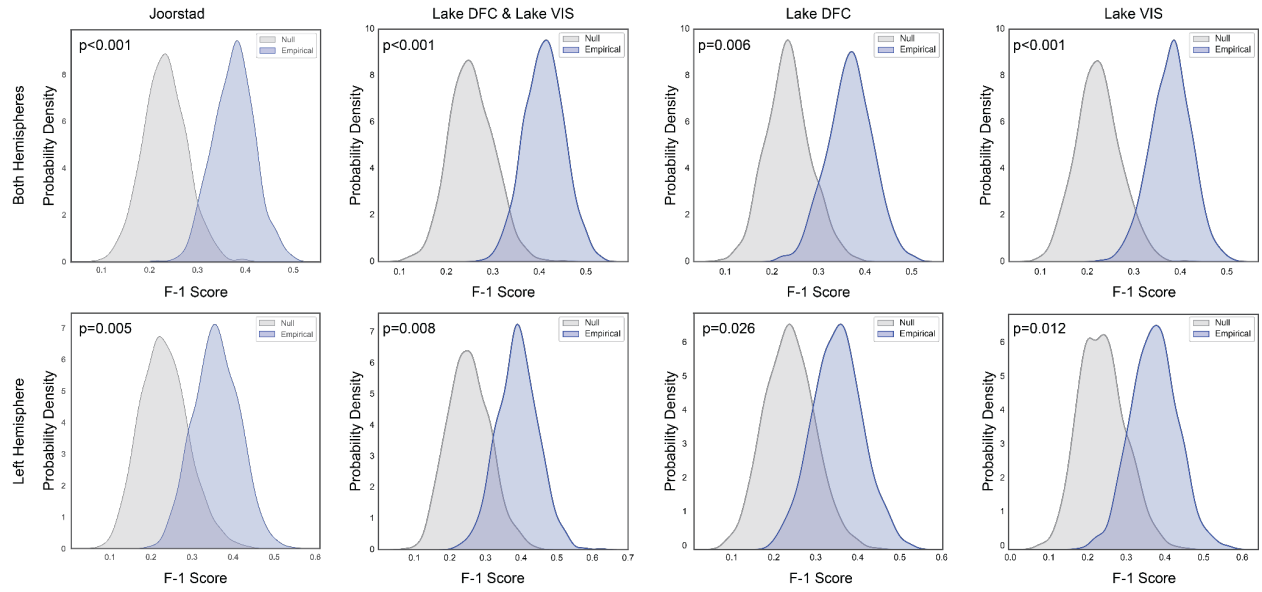


**Fig. S6 F-1 scores for models trained and tested in the six different groups (imputed from Joorstad and Lake snDrop-seq).** Distributions in blue were constructed from 1000 classifiers trained on real network labels, and the gray colored distribution represent classifiers trained on network labels shuffled by a Hungarian spinning method that controls for spatial autocorrelation. P-values were calculated as the percentage of null F-1 scores that are greater than the median of the empirical F-1 scores. The first row are the models trained and tested from parcels of both hemispheres, and the second row are the models trained and tested from parcels of the left hemisphere only. The first column are the models trained from the Joorstad dataset. The second column are the models that assemble information from both Lake DFC and Lake VIS. The third and the fourth column are the models trained from Lake DFC and Lake VIS only. Each one of the 8 ways of training and testing models yield a significant p-value for F-1 score.

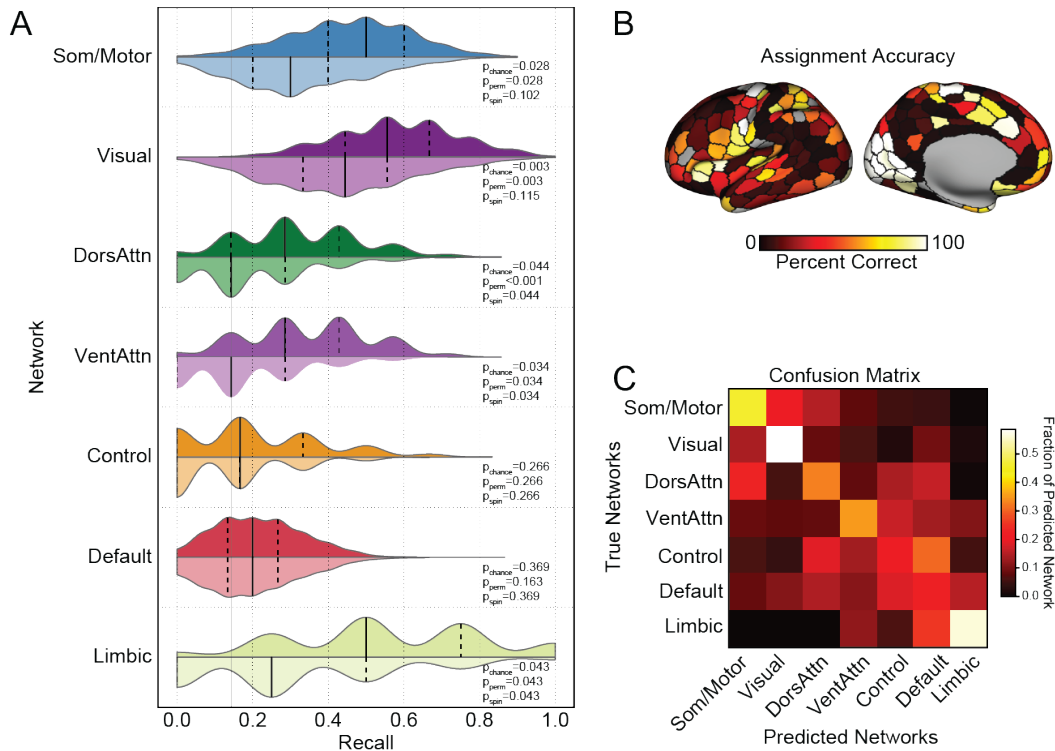


**Fig. S7 Trained and predicted on the left hemisphere only (imputed from Jorstad snRNA-seq): Large-scale functional network assignment can be predicted by cell-type fractions in post-mortem tissue. A.** Histograms display the SVM recall, or the probability of correctly classifying a parcel to the associated network. These data suggest the classifiers were able to predict dorsal attention, ventral attention, and limbic networks significantly above chance. Distributions in darker color were constructed from 1000 classifiers trained on real network labels, and the lighter colored distribution represent classifiers trained on network labels shuffled by spin-test that controls for spatial autocorrelation. The solid lines indicate median and the dashed lines represent quartiles of the distribution. P-value of empirical recall were constructed from two-sided tests of increasing stringency, (1) theoretical chance ( $1/7$ ,  $p_{\text{chance}}$ ); (2) models trained from randomly permuted network labels ( $p_{\text{perm}}$ ) (3) or labels shuffled by spin-test ( $p_{\text{spin}}$ ). **B.** Accuracy for network assignment across cortical parcels, calculated from all testing sets. **C.** Each row of the confusion matrix represents the fraction of parcels within the specific network that were predicted as belonging to each of the 7 networks. The diagonal represents the percentage of correctly classified parcels within each network. Here, the confusion matrix suggests a preferentially distinct cellular profile for somato/motor, visual, and limbic networks. While classification accuracies were low for the remaining association cortex networks, dorsal attention, default, and control networks display a higher rate of misclassifications among each other, relative to unimodal networks.

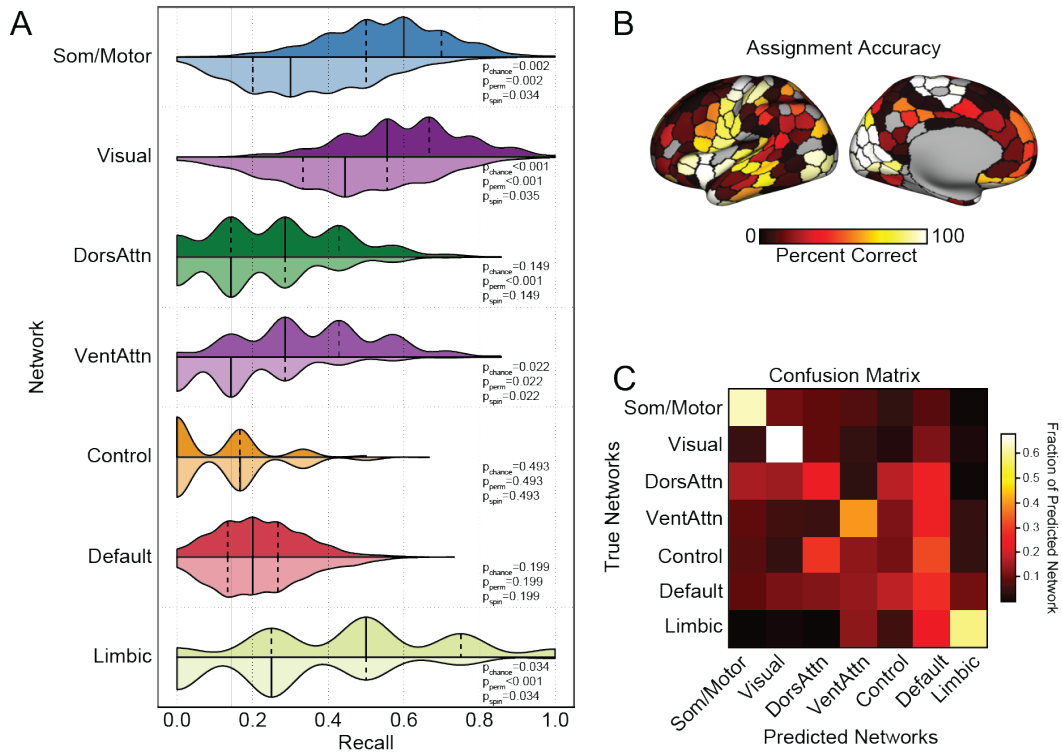


**Fig. S8 Trained and predicted on the left hemisphere only, ensembling Lake DFC and Lake VIS: Large-scale functional network assignment can be predicted by cell-type fractions in post-mortem tissue. A.** Histograms display the SVM recall, or the probability of correctly classifying a parcel to the associated network. Consistent with **Fig. 5**, these data suggest the classifiers were able to predict somato/motor, visual, ventral attention, and limbic networks significantly above chance. Distributions in darker color were constructed from 1000 classifiers trained on real network labels, and the lighter colored distribution represent classifiers trained on network labels shuffled by spin-test that controls for spatial autocorrelation. The solid lines indicate median and the dashed lines represent quartiles of the distribution. P-value of empirical recall were constructed from two-sided tests of increasing stringency, (1) theoretical chance ( $1/7$ ,  $p_{\text{chance}}$ ); (2) models trained from randomly permuted network labels ( $p_{\text{perm}}$ ) (3) or labels shuffled by spin-test ( $p_{\text{spin}}$ ). **B.** Accuracy for network assignment across cortical parcels, calculated from all testing sets. **C.** Each row of the confusion matrix represents the fraction of parcels within the specific network that were predicted as belonging to each of the 7 networks. The diagonal represents the percentage of correctly classified parcels within each network. Here, the confusion matrix suggests a preferentially distinct cellular profile for somato/motor, visual, and limbic networks. While classification accuracies were low for the remaining

association cortex networks, dorsal attention, default, and control networks display a higher rate of misclassifications among each other, relative to unimodal networks.

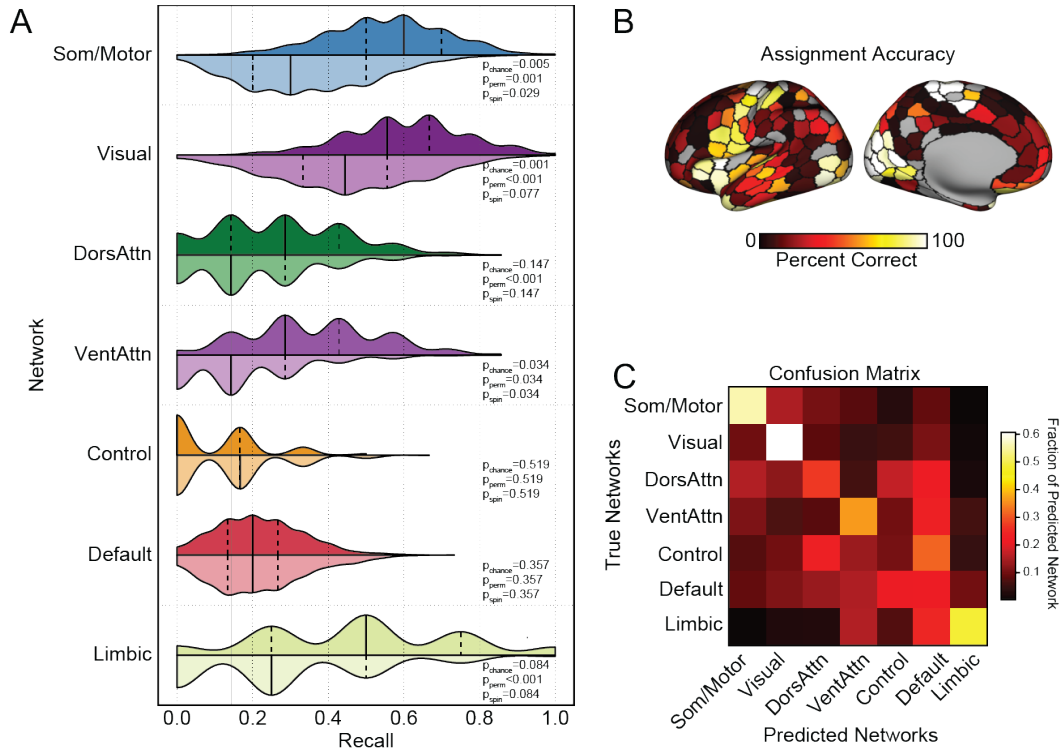


**Fig. S9 Trained and predicted on the left hemisphere only, using Lake DFC: Large-scale functional network assignment can be predicted by cell-type fractions in post-mortem tissue.** **A.** Histograms display the SVM recall, or the probability of correctly classifying a parcel to the associated network. These data suggest the classifiers were able to predict somato/motor and ventral attention networks significantly above chance. Distributions in darker color were constructed from 1000 classifiers trained on real network labels, and the lighter colored distribution represent classifiers trained on network labels shuffled by a Hungarian spinning method that controls for spatial autocorrelation. The solid lines indicate median and the dashed lines represent quartiles of the distribution. P-value of empirical recall were constructed from two-sided tests of increasing stringency, (1) theoretical chance ( $1/7$ ,  $p_{\text{chance}}$ ); (2) models trained from randomly permuted network labels ( $p_{\text{perm}}$ ) (3) or labels shuffled by spin-test ( $p_{\text{spin}}$ ). **B.** Accuracy for network assignment across cortical parcels, calculated from all testing sets. **C.** Each row of the confusion matrix represents the percentage of parcels within the specific network that were predicted as belonging to each of the 7 networks. The diagonal represents the fraction of correctly classified parcels within each network. Consistent with **Fig. 5**, the confusion matrix suggests a preferentially distinct cellular profile for somato/motor, visual, and limbic networks. While classification accuracies were low for the remaining association cortex networks, dorsal attention, default, and control networks display a higher rate of misclassifications among each other, relative to unimodal networks.

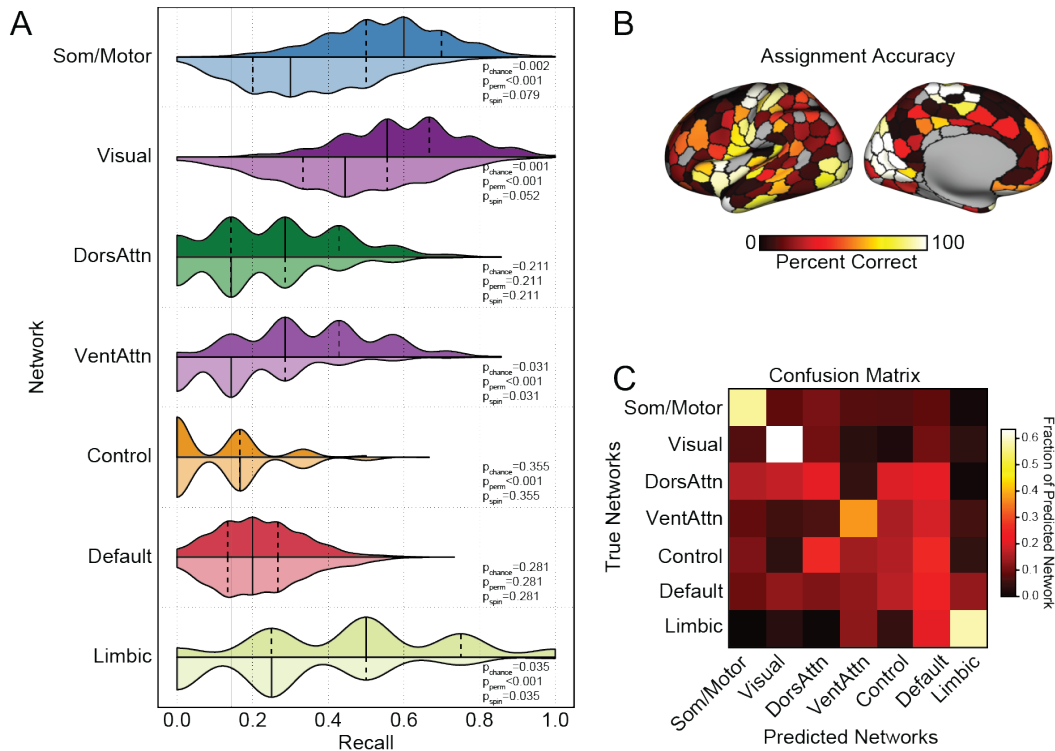


**Fig. S10 Trained and predicted on the left hemisphere only, using Lake VIS: Large-scale functional network assignment can be predicted by cell-type fractions in post-mortem tissue.** **A.** Histograms display the SVM recall, or the probability of correctly classifying a parcel to the associated network. These data suggest the classifiers were able to predict ventral attention and limbic networks significantly above chance. Distributions in darker color were constructed from 1000 classifiers trained on real network labels, and the lighter colored distribution represent classifiers trained on network labels shuffled by a Hungarian spinning method that controls for spatial autocorrelation. The solid lines indicate median and the dashed lines represent quartiles of the distribution. P-value of empirical recall were constructed from two-sided tests of increasing stringency, (1) theoretical chance ( $1/7$ ,  $p_{\text{chance}}$ ); (2) models trained from randomly permuted network labels ( $p_{\text{perm}}$ ) (3) or labels shuffled by spin-test ( $p_{\text{spin}}$ ). **B.** Accuracy for network assignment across cortical parcels, calculated from all testing sets. **C.** Each row of the confusion matrix represents the fraction of parcels within the specific network that were predicted as belonging to each of the 7 networks. The diagonal represents the percentage of correctly classified parcels within each network. Consistent with **Fig. 5**, the confusion matrix suggests a preferentially distinct cellular profile for somato/motor, visual, and limbic networks. While classification accuracies were low for the remaining association cortex networks, dorsal attention, default, and control networks display a higher rate of misclassifications among each other, relative to unimodal networks.

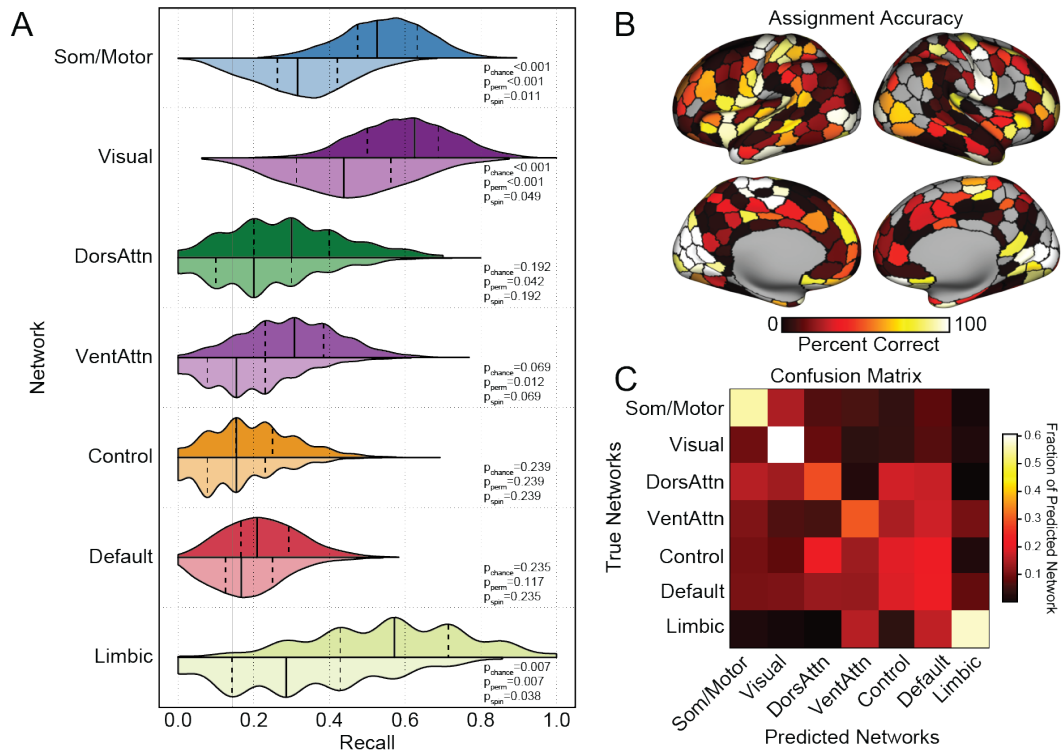


**Fig. S11 Trained and predicted on both hemispheres, using Lake DFC: Large-scale functional network assignment can be predicted by cell-type fractions in post-mortem tissue.** **A.** Histograms display the SVM recall, or the probability of correctly classifying a parcel to the associated network. These data suggest the classifiers were able to predict somato/motor, visual, and limbic networks significantly above chance. Distributions in darker color were constructed from 1000 classifiers trained on real network labels, and the lighter colored distribution represent classifiers trained on network labels shuffled by a Hungarian spinning method that controls for spatial autocorrelation. The solid lines indicate median and the dashed lines represent quartiles of the distribution. P-value of empirical recall were constructed from two-sided tests of increasing stringency, (1) theoretical chance ( $1/7$ ,  $p_{\text{chance}}$ ); (2) models trained from randomly permuted network labels ( $p_{\text{perm}}$ ) (3) or labels shuffled by spin-test ( $p_{\text{spin}}$ ). **B.** Accuracy for network assignment across cortical parcels, calculated from all testing sets. **C.** Each row of the confusion matrix represents the fraction of parcels within the specific network that were predicted as belonging to each of the 7 networks. The diagonal represents the percentage of correctly classified parcels within each network. Consistent with **Fig. 5**, the confusion matrix suggests a preferentially distinct cellular profile for somato/motor, visual, and limbic networks. While classification accuracies were low for the remaining association cortex networks, dorsal attention, default, and control networks display a higher rate of misclassifications among each other, relative to unimodal networks.

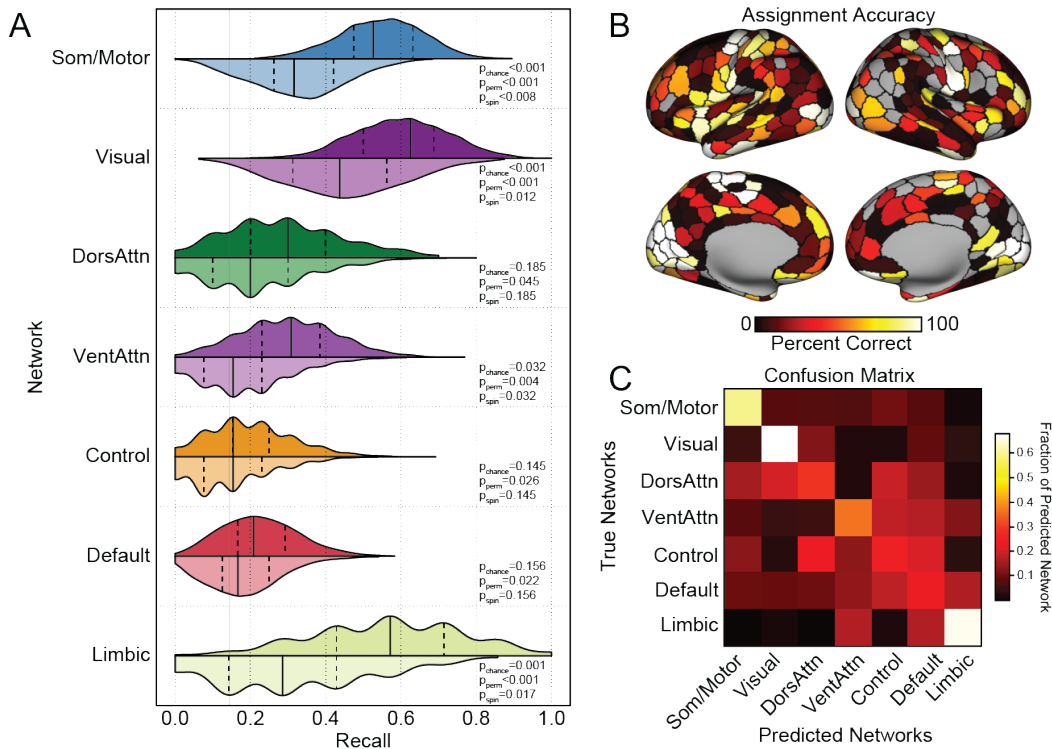


**Fig. S12 Trained and predicted on both hemispheres, using Lake VIS: Large-scale functional network assignment can be predicted by cell-type fractions in post-mortem tissue.** **A.** Histograms display the SVM recall, or the probability of correctly classifying a parcel to the associated network. Consistent with **Fig. 5**, these data suggest the classifiers were able to predict somato/motor, visual, ventral attention, and limbic networks significantly above chance. Distributions in darker color were constructed from 1000 classifiers trained on real network labels, and the lighter colored distribution represent classifiers trained on network labels shuffled by a Hungarian spinning method that controls for spatial autocorrelation. The solid lines indicate median and the dashed lines represent quartiles of the distribution. P-value of empirical recall were constructed from two-sided tests of increasing stringency, (1) theoretical chance ( $1/7$ ,  $p_{\text{chance}}$ ); (2) models trained from randomly permuted network labels ( $p_{\text{perm}}$ ) (3) or labels shuffled by spin-test ( $p_{\text{spin}}$ ). **B.** Accuracy for network assignment across cortical parcels, calculated from all testing sets. **C.** Each row of the confusion matrix represents the fraction of parcels within the specific network that were predicted as belonging to each of the 7 networks. The diagonal represents the percentage of correctly classified parcels within each network. The confusion matrix suggests a preferentially distinct cellular profile for somato/motor, visual, and limbic networks. While classification accuracies were low for the remaining association cortex networks, dorsal attention, default, and control networks display a higher rate of misclassifications among each other, relative to unimodal networks.

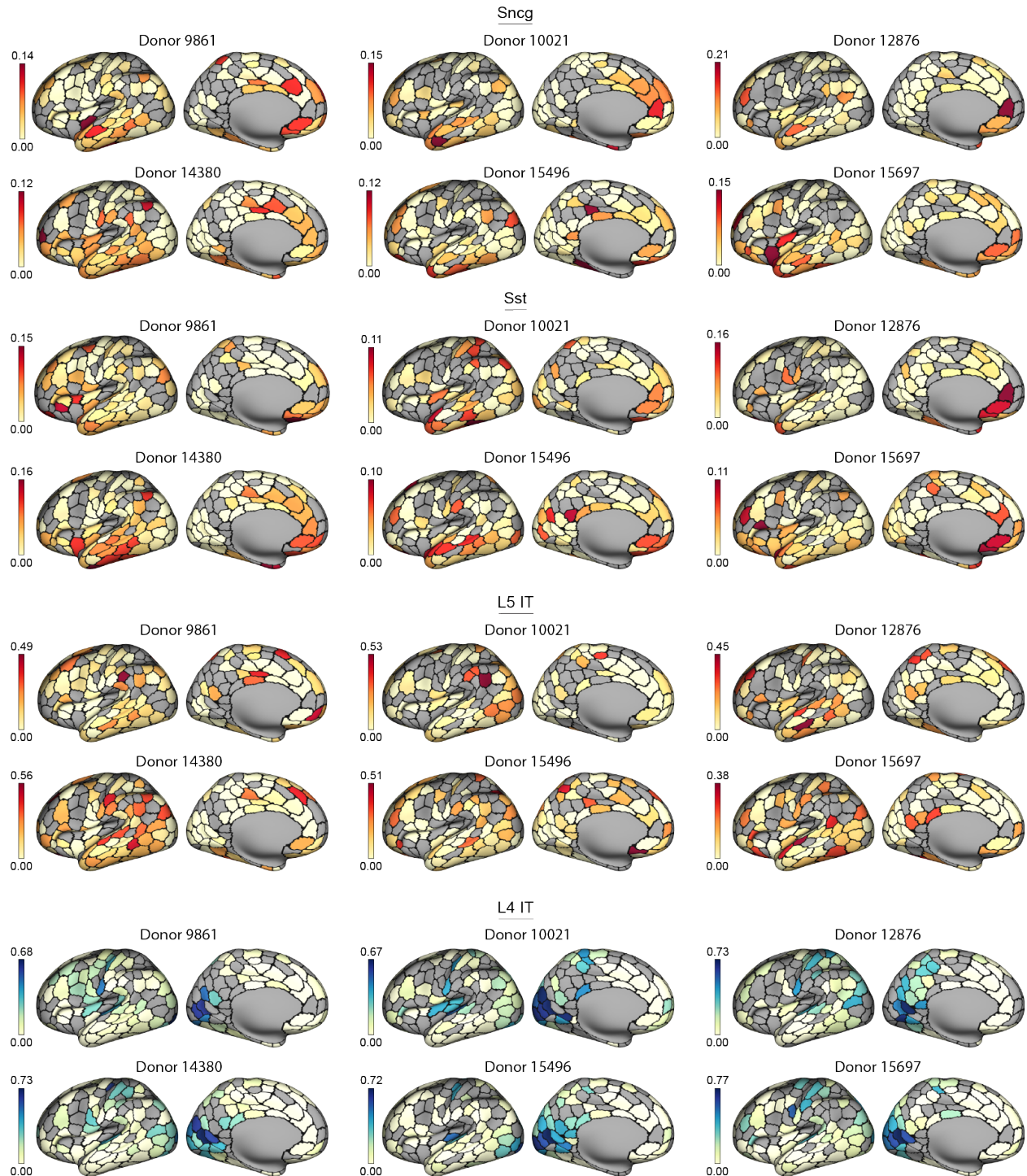


**Fig. S13 Gradient 1 associated cell type abundances at donor level (imputed from Jorstad snRNA-seq).** Surface maps display the parcel-level cellular abundances of cell types exhibiting significant relationships with Gradient 1 for each AHBA donor.