Journal: NETWORK NEUROSCIENCE

RESEARCH

Modelling the cell-type specific mesoscale murine connectome with anterograde tracing experiments

Samson Koelle [1,2], **Jennifer Whitesell** [1], **Karla Hirokawa** [1], **Hongkui Zeng**[1], **Marina Meila**[2], **Julie Harris**[1], **Stefan Mihalas**[1]

[1]Allen Institute for Brain Science, Seattle, WA, USA

[2]Department of Statistics, University of Washington, Seattle, WA, USA

Keywords: [a series of capitalized words, separated with commas]

ABSTRACT

The Allen Brain Atlas contains of thousands of anterograde tracing experiments targeting diverse structures and classes of projecting neurons. This paper describes the conversion of these experiments into class-specific connectivity matrices representing the connection between source and target structures. We introduce and validate a novel statistical model for creation of connectivity matrices that combines spatial and categorical smoothing to share information between similar neuron classes. We then show that our connectivities display expected cell-type and structure specific connectivities, and factor the wild-type connectivity matrix to uncover the underlying latent structure.

AUTHOR SUMMARY

1 INTRODUCTION

The animal nervous system enables an extraordinary range of natural behaviors, and has inspired much of modern artificial intelligence. Neural connectivities - axon-dendrite connections from one region to another - form the architecture underlying this capability. These connectivities vary by neuron type, as well as axonic source and dendritic target structure. Thus, characterization of the relationship between neuron type and source and target structure is an important for understanding the overall nervous system.

Viral tracing experiments - in which a viral vector expressing GFP is transduced into neural cells through stereotaxic injection - are a useful tool for understanding these connections on the mesoscale (Chamberlin, Du, de Lacalle, & Saper, 1998; Daigle et al., 2018; J. A. Harris, Oh, & Zeng, 2012). The GFP protein moves from axon to dendrite through the process of anterograde projection, so neurons 'downstream' of the injection site will also fluoresce. Two-photon tomography imaging can then determine the location and strength of the fluorescent signals in two-dimensional slices. These locations can then be mapped back into three-dimensional space, and the signal is partitioned into the transduced source and merely transfected target regions (SK's comment:**Check**).

Several statistical models for the conversion of such experiment-specific signals into estimates of connectivity strength have been proposed (K. D. Harris, Mihalas, & Shea-Brown, 2016; Knox et al., 2019; Oh et al., 2014). Of these, Oh et al. (2014) and Knox et al. (2019) model **structural connectivities** between structures. Intuitively, these models provide some improvement over simply averaging the projection signals of injections in a given region. However, these works model connectivities observed in wild-type mice transduced with constitutive promoters, and so are poorly suited for extension to recently developed tracing experiments that induce cell-type specific fluorescence (J. A. Harris et al., 2019). In particular, GFP promotion is induced by Cre-recombinase expression in cell-types specified by transgenic strain. Thus, this paper introduces a **cell class**-specific statistical model to deal with the diverse set of **cre-lines** described in J. A. Harris et al. (2019).

Our model is a to-our-knowledge novel estimator that takes into account both the spatial position of the labelled source, as well as the categorical cell class. Like the previously state-of-the-art model in Knox et al. (2019), this model predicts structural connectivity as an average over positions within the

structure, with nearby experiments given more weight. However, our model weighs class-specific behavior in a particular structure against spatial position, so a nearby experiment targeting a similar cell class would be relatively upweighted, while a nearby experiment targeting a dissimilar class would be downweighted. This model outperforms the model of Knox et al. (2019) based off of their ability to predict held-out experiments in leave-one-out cross-validation. We then establish a lower-limit of detection, and use the trained model to estimate overall connectivity matrices for assayed each cell class.

The resulting cell-type specific connectivity matrices form a multi-way **structural connection tensor** of information about neural structure. We do not attempt an exhaustive analysis of this data, but do manually verify several cell-type specific connectivity patterns found elsewhere in the literature, and show that these cell-type specific signals are behaving in expected ways. Finally, we decompose the wild-type connectivity matrix into factors representing archetypal connective patterns using non-negative matrix factorization. These components are themselves novel and of some independent interest.

Section 2 gives information on the data and statistical methodology, and Section 3 presents our results. These include connectivities, assessments of model fit, and subsequent analyses. Additional information on our dataset, methods, and results are given in Supplemental Sections 5, 6, and 7, respectively.

2 METHODS

We create and analyze cell class-specific connectivity matrices using models trained on murine viral-tracing experiments. This section describes the data used to generate the model, the model itself, the evaluation of the model, and the use of the model in creation of the connectivity matrices. It also includes background on the non-negative matrix factorization method used for decomposing the wild-type connectivity matrix into latent structures. Additional information on our data is given in Supplemental Section 5 methods is given in Supplemental Section 6.

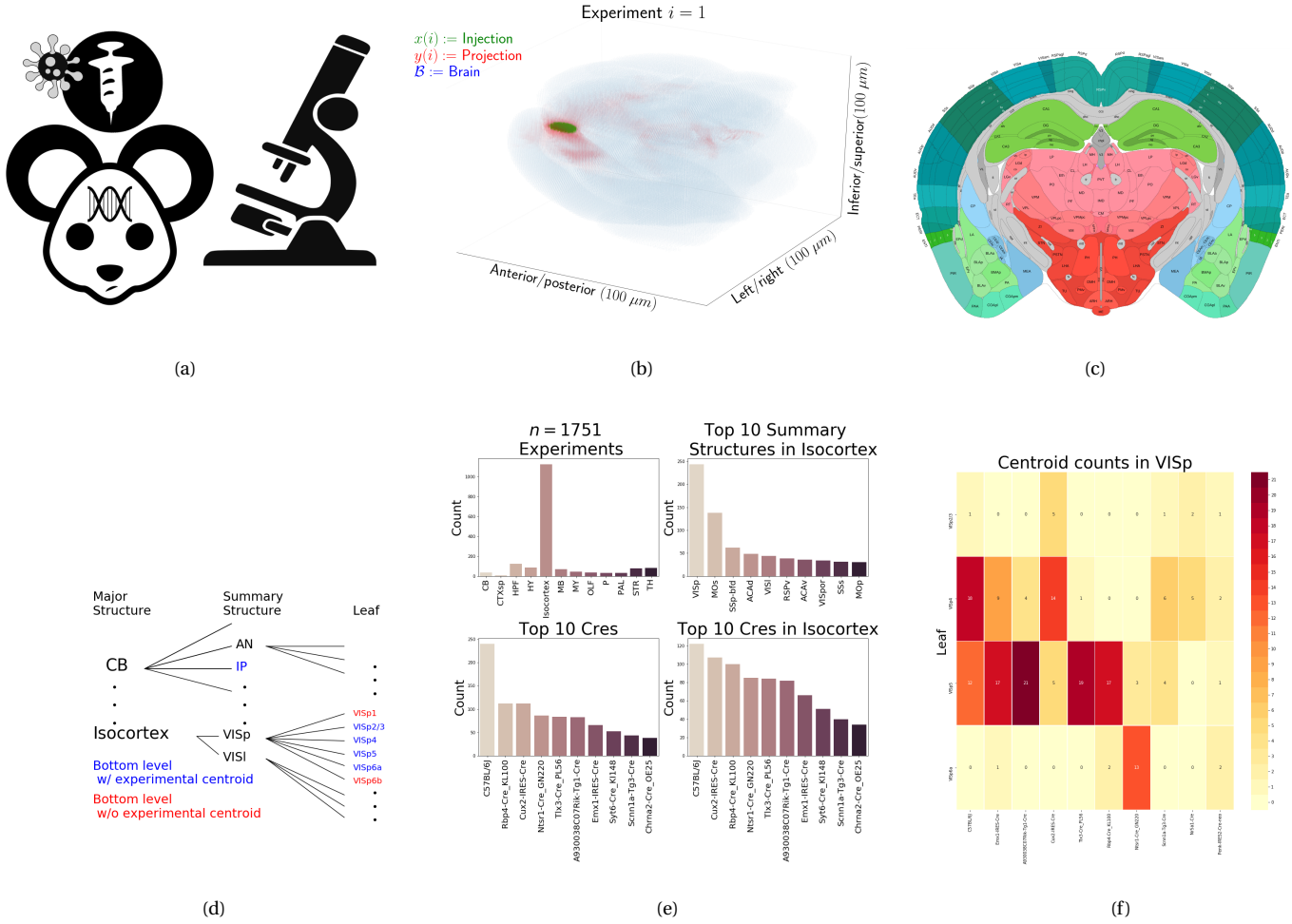


Figure 1: Experimental setting. 1a For each experiment, a potentially Cre-recombinase promoted GFP-expressing transgene casette is transduced after stereotaxic injection into a Cre-driver mouse, followed by two-photon tomography imaging. 1b An example of the segmentation of projection and injection for a single experiment. Within each assayed brain (blue), injection (green) and projection (red) areas are determined via histological analysis and alignment to the Allen Common Coordinate Framework (CCF). 1c Example of structural segmentation within a horizontal plane. 1d Explanation of nested structural ontology highlighting various levels of structural ontology. Lowest-level (leaf) structures are colored in blue, and structures containing an injection centroid are colored in red. 1e Abundances of Cre-lines and structural injections. 1f Coccurence of layer-specific centroids and Cre-line within VISp

Mice

(SK's comment:**Experiments involving mice were approved by the Institutional Animal Care and Use Committees of the Allen Institute for Brain Science in accordance with NIH guidelines.**)

Data

Our dataset $\mathscr{D}$ consists of $n = 1751$ publicly available murine viral-tracing experiments from the Allen Brain Atlas. Figures 1a summarizes the multistage experimental process used to generate this data. In each experiment, a GFP-labelled transgene casette with a potentially Cre-inducible promoter is injected into a particular location in a Cre-driver mouse. This causes fluorescence thats depends on the localization of Cre-recombinase expression within the mouse. While frequently this localization corresponds to a specific cell-type, it can also correspond to a combination of cell-types. For example, in wild-type mice injected with non-Cre specific promoters, fluorescence is observed in all areas projected to from the injection site, regardless of cell-type. Thus, we use the term cell class to describe the neurons targeted by a specific combination (or absence) of transgene and mouse-line. This is the notion of cell-type specificity that we model.

After injection, the resultant fluorescent signal is imaged, and aligned into the Allen Common Coordinate Framework (CCF) v3, a three-dimensional idealized model of the brain that is consistent between animals. This imaging and alignment procedure (described in detail in (J. A. Harris et al., 2019)) records fluorescent intensity discretized at the 100 μm *voxel* level. Given an experiment, this image is histologically segmented into *injection* and *projection* areas corresponding to areas of transduction and transduction/transfection, respectively (SK's comment:**check**). An example for a single experiment is given in Figure 1b.

Our goal is the estimation of *s*tructural connectivity from one structure to another. Thus, a visual depiction of this structural regionalization for a slice of the brain is given in Figure 1c. For different areas of the brain, the Allen Brain Atlas contains different depths of regionalization. We denote these levels as Major Structures, Summary Structures, and Leafs. As indicated in Figure 1d, the dataset used to generate the connectivity model reported in this paper contains certain combinations of structure and cell class (v, s) frequently, and others not at all. A summary of the most frequently assayed cell classes and structures is given in Figures 1e and 1f. Since users of the connectivity matrices may be

interested in particular combinations, or interested in the amount of data used to generate a particular connectivity estimate, we present this information about all experiments in Supplemental Section 5.

At an essential level, cell-class specific neural connectivity is a function $f : \mathcal{V} \times \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}_{\geq 0}$ giving the directed connection of a particular cell class from a one position in the brain to another. However, what we will actually estimate are structural connectivities defined with respect to a set of S source regions $\mathcal{S} := \{s\}$, T target regions $\mathcal{T} := \{t\}$, and V cell classes $\mathcal{V} := \{v\}$. In contrast to Knox et al. (2019), which only uses wild type $C57BL/6J$ mice, these experiments utilize $V = 114$ different Cre-lines. We generally consider $S = 564$ leaf sources and $T = 1123$ leaf targets, where 559 are contralateral or mediolateral.

We preprocess our data in several ways. We discretize flourescent signals like injections and projections into $100\mu m^3$ **voxels**. Given an experiment i, we represent injections and projections as maps $x(i), y(i) : \mathcal{B} \to \mathbb{R}_{\geq 0}$, where $\mathcal{B} \subset [1:132] \times [1:80] \times [1:104]$ corresponds to the subset of the $(1.32 \times 0.8 \times 1.04)$ cm rectangular space occupied by the standard mouse brain. A structure s then contains $|s|$ voxels at locations $\{l_{s_j} \in \mathbb{R}^3\}$, and similarly for targets. We calculate injection centroids $c(i) \in \mathbb{R}^3$ and regionalized projections $y_{\mathcal{T}}(i) \in \mathbb{R}^T$ giving the sum of $y(i)$ in each region. In contrast to Knox et al. (2019), we also $l1$ normalize these projection vectors. This accounts for differences in the cre-driven expression of eGFP via the various transgene promoters. A detailed mathematical description of these steps, including data quality control, is given in Supplemental Section 6.

Modeling Structural Connectivity

We define

$$\textit{structural connectivity strength } \mathscr{C} : \mathcal{V} \times \mathcal{S} \times \mathcal{T} \to \mathbb{R}_{\geq 0} \text{ with } \mathscr{C}(v,s,t) = \sum_{l_{s_j} \in s} \sum_{l_{j'} \in t} f(v, l_j, l_{j'}),$$

$$\textit{normalized structural connectivity strength } \mathscr{C}^S : \mathcal{V} \times \mathcal{S} \times \mathcal{T} \to \mathbb{R}_{\geq 0} \text{ with } \mathscr{C}^S(v,s,t) = \frac{1}{|s|}\mathscr{C}(v, l_j, l_{j'}),$$

$$\textit{normalized structural projection density } \mathscr{C}^D : \mathcal{V} \times \mathcal{S} \times \mathcal{T} \to \mathbb{R}_{\geq 0} \text{ with } \mathscr{C}^D(v,s,t) = \frac{1}{|s||t|}\mathscr{C}(v, l_j, l_{j'}).$$

These represent the strength of the connection from source to target regions for each class. Since the normalized strength and density are computable from the strength via a fixed normalization, our main statistical goal is to estimate $\mathscr{C}(v,s,t)$ for all v, s and t.We call this estimator $\widehat{\mathscr{C}}$.

Construction of such an estimator raises the questions of what data to use for estimating which connectivity, how to featurize the dataset, what statistical estimator to use, and how to reconstruct the connectivity using the chosen estimator. Mathematically, we represent these considerations as

$$\widehat{\mathscr{C}}(v,s,t) = f^*(\widehat{f}(f_*(\mathscr{D}(v,s,t)))). \tag{1}$$

This makes explicit the data featurization f_*, statistical estimator $\widehat{f}$, and any potential subsequent transformation f^* such as summing over the source and target regions. Denoting $\mathscr{D}$ as a function of v, s, and t reflects that different data may be used to estimate different connectivities. Table 1 reviews estimators used for this data-type used in previous work, as well as our two main extensions: the Cre-NW and **Expected Loss** (EL) models. Additional information on these estimators is given in Supplemental Section 6.

Name	f^*	$\widehat{f}$	f_*	$\mathscr{D}(v,s,t)$
NNLS (Oh et al., 2014)	$\widehat{f}(S)$	NNLS(X,Y)	$X = x_{\mathscr{S}}, Y = y_{\mathscr{T}}$	I_m / I_m
NW (Knox et al., 2019)	$\sum_{l_s \in s} \widehat{f}(l_s)$	NW(X,Y)	$X = c, Y = y_{\mathscr{T}}$	I_m / I_m
Cre-NW	$\sum_{l_s \in s} \widehat{f}(l_s)$	NW(X,Y)	$X = c, Y = y_{\mathscr{T}}$	$(I_s \cap I_v) / I_m$
Expected Loss (EL)	$\sum_{l_s \in s} \widehat{f}(s)$	EL(X, Y, v)	$X = c, Y = y_{\mathscr{T}}, v$	I_s / I_m

Table 1: Estimation of $\mathscr{C}$ using connectivity data. The regionalization, estimation, and featurization steps are denoted by $f^*, \widehat{f}$, and f_*, respectively. The training data used to fit the model is given by I. We denote experiments with centroids in particular major brain divisions and leafs as I_m and I_s, respectively. Data I_s / I_m means that, given a location $l_s \in s \in m$, the model $\widehat{f}$ is trained on all of I_m, but only uses I_s for prediction.

Our contributions have several differences from the previous methods. In contrast to the non-negative least squares (Oh et al., 2014) and Nadaraya-Watson (Knox et al., 2019) estimators that take into account s and t, but not v, our new estimators specifically account for cell class. The Cre-NW estimator only uses experiments from a particular class to predict connectivity for that class, while the EL estimator shares information between classes within a structure. A detailed mathematical description of our new estimator is given in Appendix 6. This estimator takes into account two types of covariate information about each experiment: the centroid of the injection, and the Cre-line. Like the NW and Cre-NW estimator, the EL estimator generates predictions for each voxel in a structure, and then sums them together to get the overall connectivity. However, in contrast to these alternative approaches, when predicting the projection pattern of a certain cell-class at a particular location, the EL estimator weights the average behavior of the class in the structure containing the location in question against the locations of the various proximal experiments. Thus, nearby experiments with similar Cre-lines can help generate the prediction, even when there are few nearby experiments of the cell-class in question.

Model evaluation

We select optimum functions from within and between our estimator classes using empirical risk minimization. Equation 1 includes a deterministic step f^* included without input by the data. The performance of $\widehat{\mathscr{C}(v,s,t)}$ is therefore determined by performance of the model $\widehat{f}(f_*(\mathscr{D}(v,s,t)))$. We can thus evaluate $\widehat{f}(v,s,t)$ using *l*eave-one-out cross validation, in which the accuracy of the model is assessed by its ability to predict experiments excluded from the training data. This method is robust to the trivial overfitting in Nadaraya-Watson bandwidth selection.

Another main estimation question is what combinations of v, s, and t to generate a prediction for. Our EL and Cre-NW models are leaf specific; they only generate predictions for cell classes in leafs where at least one experiment with a Cre-line targeting that class has a centroid. To compare our new estimators accurately with less-restrictive models such as used in Knox et al. (2019), we therefore restrict restrict to the smallest set of evaluation experiments suggested by any of our models: virus-leaf combinations that are present at least twice. The sizes of these evaluation sets are given in Supplemental Section 5.

We use weighted $l2$-loss to evaluate these predictions.

$$\text{l2-loss } \ell(y_{\mathscr{T}}(i)), \widehat{y_{\mathscr{T}}(i)}) = \|y_{\mathscr{T}}(i)) - \widehat{y_{\mathscr{T}}(i)}\|_2^2.$$

$$\text{weighted l2-loss } \mathscr{L}(\widehat{f(f_*)}) = \frac{1}{|\{s,v\}|} \sum_{s,v \in \{\mathscr{S},\mathscr{V}\}} \frac{1}{|I_s \cap I_v|} \sum_{i \in (I_s \cap I_v)} \ell(y_{\mathscr{T}}(i)), \hat{f}(f_*(\mathscr{D}(s,t,v) \setminus i)).$$

This is a somewhat different loss from Knox et al. (2019), both because of the normalization of projection, and because of the increased weighting of rarer combinations of s and v implicit in the loss. Since the number of parameters fit is quite low relative to the size of the evaluation set, we do not make use of a formal validation-test split. As a final modeling step, we establish a lower limit of detection. This is covered in Supplemental Section 6

Connectivity analyses

We show neuronal processes underlying our estimated connectome using two types of unsupervised learning. Our use of heirarchical clustering is standard, and so we do not review it here. However, our application of non-negative matrix factorization (NMF) to decompose the estimated long-range connectivity into *connectivity archetypes* that linearly combine to reproduce the observed connectivity is novel and technically of some independent interest. Non-negative matrix factorization refers to a collection of **dictionary-learning** algorithms for decomposing a non-negatively-valued matrix such as $\mathscr{C}$ into positively-valued matrices called, by convention, weights $W \in \mathbb{R}_{\geq 0}^{S \times q}$ and hidden units $H \in \mathbb{R}_{\geq 0}^{q \times T}$. Unlike PCA, NMF specifically accounts for the fact that data are all in the positive orthant. This H is typically used to identify latent structures with interpretable biological meaning, and the choice of matrix factorization method reflects particular scientific subquestions and probabilistic interpretations.

Our algorithm is

$$\text{NMF}(\mathscr{C}, \lambda, q) := \arg\min_{W,H} \frac{1}{2} \| 1_{d(s,t)>1500\mu m} \odot \mathscr{C} - WH \|_2^2 + \lambda(\|H\|_1 + \|W\|_1).$$

We ignore connections between source and and target regions less than $1500\mu m$ apart. This is because short-range projections resulting from diffusion dominate the matrices $\hat{\mathscr{C}}$, and represent a less-interesting type of biological structure. We also set $\lambda = 0.002$ to encourage sparser and therefore more interpretable components. We use unsupervised cross-validation to determine an optimum q, and show the top 15 stable components. Stability analysis accounts for the difficult-to-optimize NMF optimization problem by clustering the resultant H from multiple replicates. The medians of the component clusters appearing frequently across NMF replicates are selected as **connectivity archetypes**. Details of these approaches are given in Supplementary Sections 6 and 7.

3 RESULTS

Our results include a mix of quantitative and qualitative evaluations of model fit, the Cre-specific connectivity matrices themselves, and retrospective analyses of these matrices for patterns related to Cre-line and source and target region.

Model evaluation

Table **??** contains weighted losses from leave-one-out cross-validation of candidate models. Our EL model generally performs better than the other Nadaraya-Watson estimators that we consider. For example, the NW Major-WT model is the model from Knox et al. (2019). The EL model combines the good performance of class-specific models like NW Leaf-Cre in regions like Isocortex with the good performance of class-agnostic models in regions like Thalymus. Additional information on model evaluation, including class and structure specific performance, is given in Appendix 5 In particular, Supplementary Table 3 contains the sizes of these evaluation sets in each major structure, and Supplementary Section 7 contains the structure- and class specific losses.

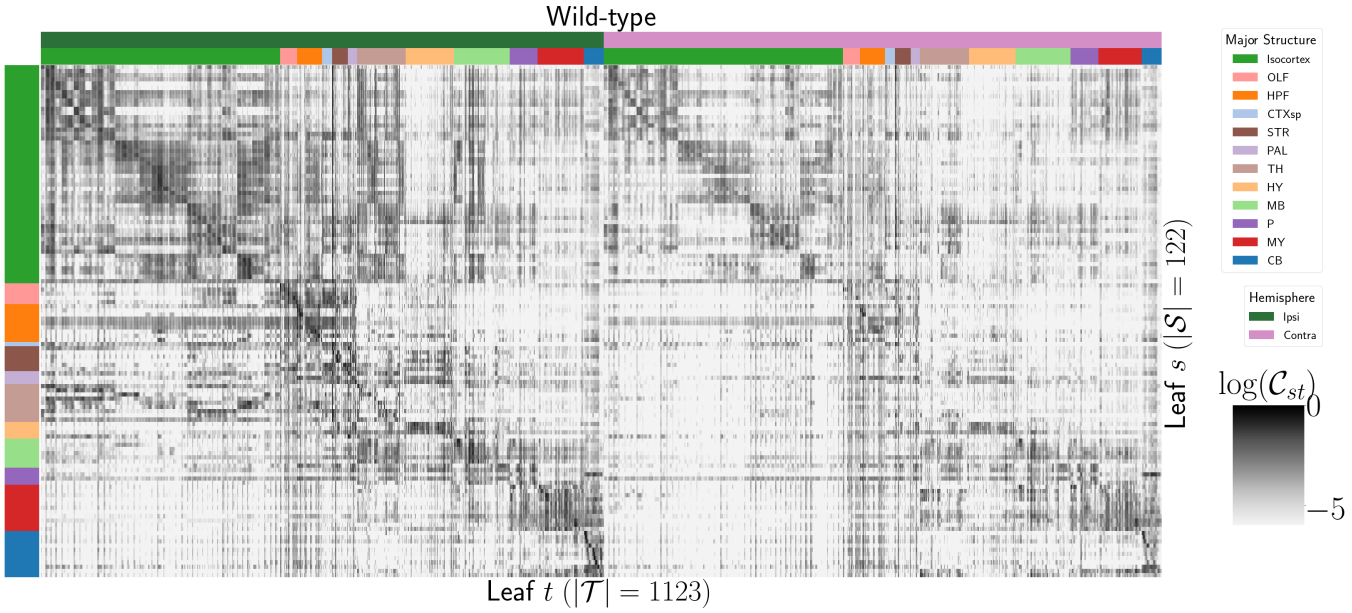
	Mean Leaf-Cre	NW Major-Cre	NW Leaf-Cre	NW Leaf	NW Major-WT	NW Major	EL
$\widehat{f}$	Mean	NW					EL
$\mathscr{D}$	$I_c \cap I_L$	$I_c \cap I_M$	$I_c \cap I_L$	I_L	$I_{wt} \cap I_M$	I_M	I_L
Isocortex	0.264	0.256	0.257	0.358	0.370	0.370	**0.246**
OLF	0.185	0.215	0.184	**0.131**	0.175	0.175	0.136
HPF	0.176	0.335	0.170	0.201	0.235	0.235	**0.148**
CTXsp	**0.758**	**0.758**	**0.758**	**0.758**	**0.758**	**0.758**	**0.758**
STR	0.131	**0.121**	0.129	0.173	0.236	0.236	0.125
PAL	0.220	0.223	0.220	0.339	0.324	0.324	**0.197**
TH	0.634	0.626	0.634	0.362	**0.360**	**0.360**	0.366
HY	0.388	0.392	0.381	0.359	0.338	0.338	**0.331**
MB	0.213	0.232	0.201	0.276	0.285	0.285	**0.195**
P	0.309	0.309	0.309	0.404	0.402	0.402	**0.306**
MY	0.261	0.340	0.261	0.188	**0.187**	**0.187**	0.198
CB	0.062	**0.061**	0.062	0.067	0.111	0.111	0.068

Table 2: Losses from leave-one-out cross-validation of candidate models. **Bold** numbers are best for their major structure.

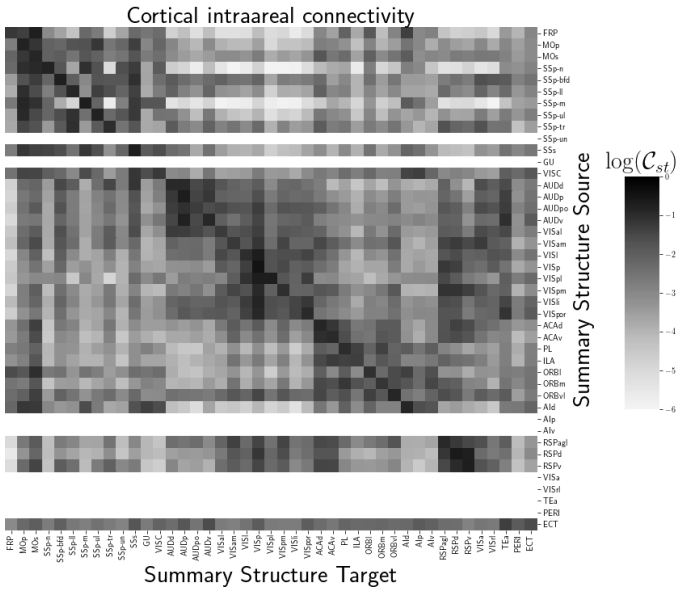
Connectivities

Our main result is the estimation of matrices $\hat{\mathscr{C}}_v \in \mathbb{R}_{\geq 0}^{S \times T}$ representing connections of source structures to target structures for particular cre-lines v. We exhibit several characteristics of interest, and confirm the detection of several well-established connectivities within our tensor. Many additional interesting biological processes are visible within this matrix - more than we can report in this paper - and it is our expectation that these will be identified by users of our results. The connectivity tensor and code to reproduce it are available at https://github.com/AllenInstitute/mouse_connectivity_models/tree/2020. Note that many entries of these matrices are missing due to lack of experiments.

Overall connectivity The connectivity matrix $\mathscr{C}_{wt}$ for wild-type connectivities from leaf sources to leaf targets is illustrated in Figure 2a. Several expected biological processes are evident. For example, intraareal connectivities are clear, as are ipsilateral connections between cortex and thalymus. The clear intraareal connectivities mirror previous estimates in Oh et al. (2014) and Knox et al. (2019) and descriptive depictions of individual experiments in J. A. Harris et al. (2019). Although a major advantage of including distinct Cre-lines is layer-specific targeting, for comparison with Figure 3 Knox et al. (2019), we also plot connectivity between summary-structure sources and targets in the cortex in Figure 2b. These coarser projections are simply averages over component layers weighted by layer size. Our results exhibit a much larger range of connectivities, and therefore are more dense.



(a)



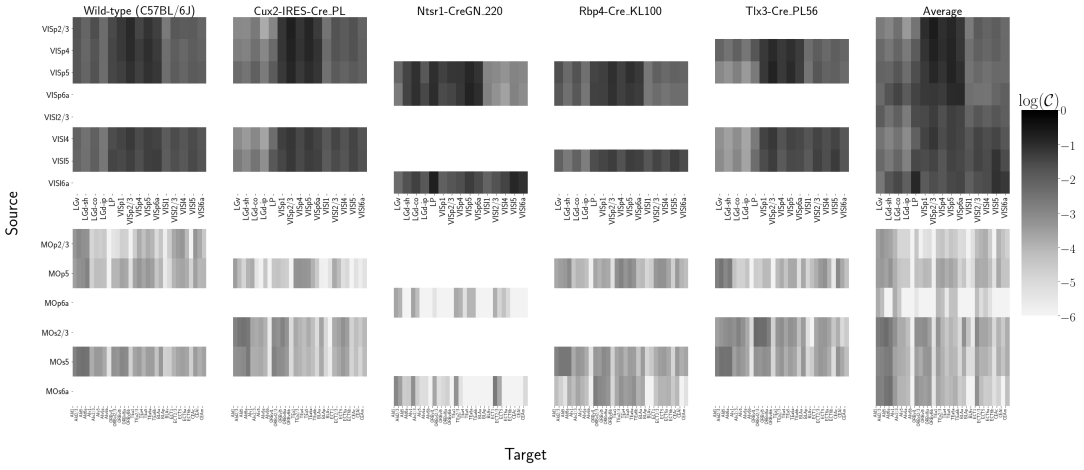
(b)

Figure 2: Wild-type connectivities. 2a Log wild-type connectivity matrix $\log \mathscr{C}(s,t,v_{wt})$. 2b Log wild-type intracortical connectivity matrix at the summary structure level.

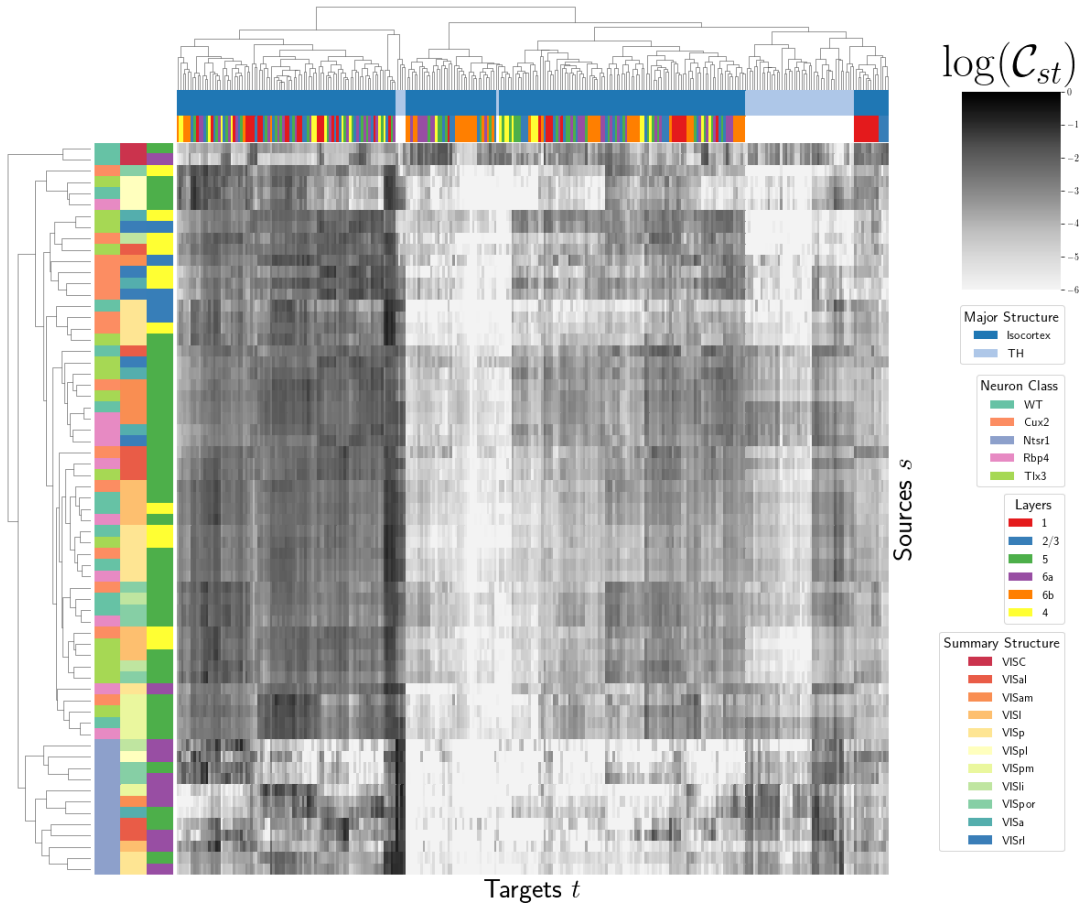
Class-specific connectivities We have generated $V = 114$ cell-class specific structural connectivities $\mathscr{C}_v$. A reasonable question is which source and cell-type combinations behave similarly, and which target projections tend to co-occur. Exhaustive comparison of this estimated behavior is prohibitive, but we do exhibit several examples of our class specific connectivities conforming to well-known behaviors. These validation cases are given in Figure 3.

We begin by plotting subsets of the estimated connectivities in the well-studied VISp and MO regions in Figure 3a. The localization of Rbp4-Cre and Ntsr1-Cre injection centroids to layers 5 and 6 respectively is evident (see also Supplemental Figure **??**). These layers project to their expected targets Jeong et al. (2016). In VISp, the Ntsr1-Cre line strongly targets the thalymic LP nuclei, and in MO, layer 5 projects to anterior basolateral amygdala (BLA) and capsular central amygdala (CEA), while layer 6 does not. As a heuristic alternative model, we also synthesize information about leafs targeted by different Cre-lines, we also generate an average connectivity matrix over all Cre-lines. This model is not evaluated in our testing, and is only a general stand-in for overall behavior, but provides a useful summary of results.

Figure 3b shows a collection of connectivity strengths generated using cre-specific models for wild-type, Cux2, Ntsr1, Rbp4, and Tlx3 cre-lines from visual signal processing leafs in the cortex to cortical and thalymic nucleii. This shows that cell-class has a dominating effect on projection in certain regions. We use heirarchical clustering to sort source structure/cell-class combinations by the similarity of their structural projections, and sort target structures by the structures from which they receive projections. Examining the former, we can see that the Ntsr1 Cre-line distinctly projects to thalymic nucleii, regardless of summary structure. This contrasts with the tendency of other cell classes to project intracortically in a manner determined by the source structure. Similarly, layer 6 targets are not strongly projected to by any of the displayed Cre-lines. There are too many targeted summary structures to plot here, but we expect that the source profile of each target clusters by structure.



(a)



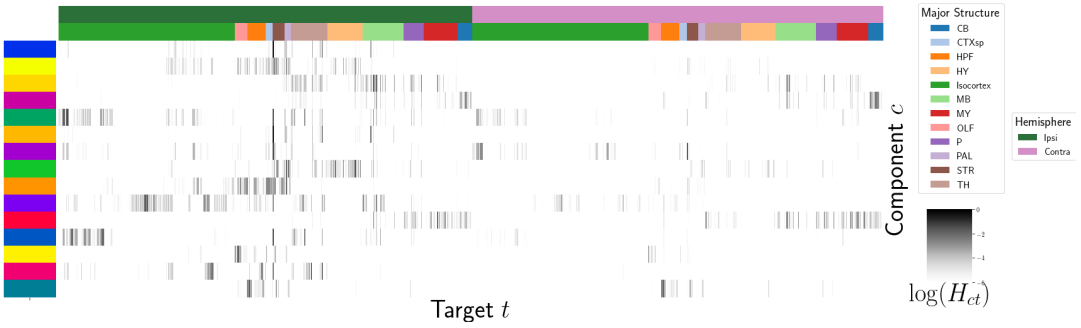
(b)

Figure 3: Cell-class and layer specific connectivities from VISp and MO. This figure shows a preselected subset of putatively interesting connectivities from VISp and MO. Sources without a injection of that Cre-type are not estimated due to lack of data for that Cre-line in that structure. 3 Heirarchical clustering of connectivity strengths from visual signal processing cell-types to cortical and thalymic targets. Cre-line, summary structure, and layer are labelled on the sources. Major brain division and layer are labelled on the targets. Note that sources/cre combinations are only included if there is at least one

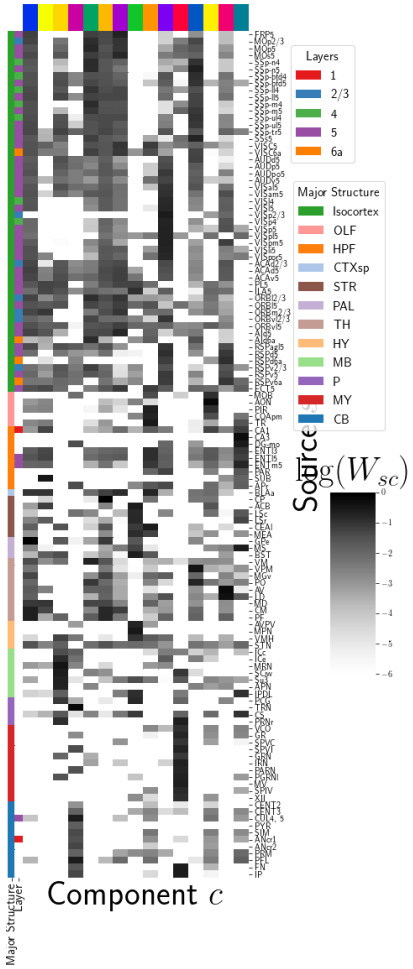
Connectivity Analyses

Each structural connectivity matrix is a high-dimensional representation of relatively few biological processes, and decomposition of neural signals to recover these processes is a fundamental goal in neuroscience. As discussed in Knox et al. (2019), one of the most basic processes underlying the observed connectivity is the tendency of each source region to predominantly project to proximal regions. For example, the heatmap in 7 shows intraregion distances clearly contains an overall pattern reminscent of the connectivity matrix in 2. These connections are biologically meaningful, but also unsurprising, and their relative strength biases learned latent coordinate representations away from long-range structures. For this reason, we establish a 1500μm 'distal' threshold within which to exclude connections for our analysis.

Since certain cell-types and layers have a characteristic connectivity pattern, we perform non-negative matrix factorization on distal wild-type connectivities to estimate these characteristic patterns in a probabilistic way. This decomposes the remaining censored connectivity matrix into a linear model based off a relatively small number of distinct signals. These signals are plotted in Figure 4, and technical details and intermediate results are given in Supplemental Sections 6 and 7, respectively. The plotted decomposition shows that these underlying connectivity archetypes correspond strongly to major brain division. However, certain components that predominantly represent connectivity from a given major brain division may also be accessed from other areas. For example, the IP and FN regions of CB are strongly associated in 4b with the component projecting to MY in 4a.



(a)



(b)

Figure 4: Non-negative matrix factorization results $\mathscr{C}_{wt} = WH$ for $q = 15$ components. 4a Latent space coordinates H of $\mathscr{C}$. Target major structure and hemisphere are plotted. 4b Loading matrix W. Source major structure and layer are plotted.

4 DISCUSSION

We see several opportunities for improving on our model. Our particular task of transforming the injection and projection signal depending on cell-type is a non-linear transformation problem with categorical covariate. Model averaging based off of cross-validation has been implemented in Gao, Zhang, Wang, and Zou (2016), but we note that our approach makes use of a non-parametric estimator, rather than an optimization method for selecting the weights (Saul & Roweis, 2003), and is applied specifically to a target-encoded feature space. The properties of this estimator, as well as its relation to estimators fit using an optimization algorithm, are a possible future avenue of research. Therefore, a deep model such as Lotfollahi, Naghipourfar, Theis, and Alexander Wolf (2019) could be appropriate, provided enough data was available. With respect to the model, a Wasserstein-based measure of injection similarity per structure would combine both the physical simplicity of the centroid model while also incorporating structural knowledge. Residual models of the above could also be considered.

The factorization of the connectivity matrix could be similarly improved. Flattening $\mathscr{C}$ prior to unsupervised analysis is not necessarily recommended, but provides an easy solution for this problem.

ACKNOWLEDGMENTS

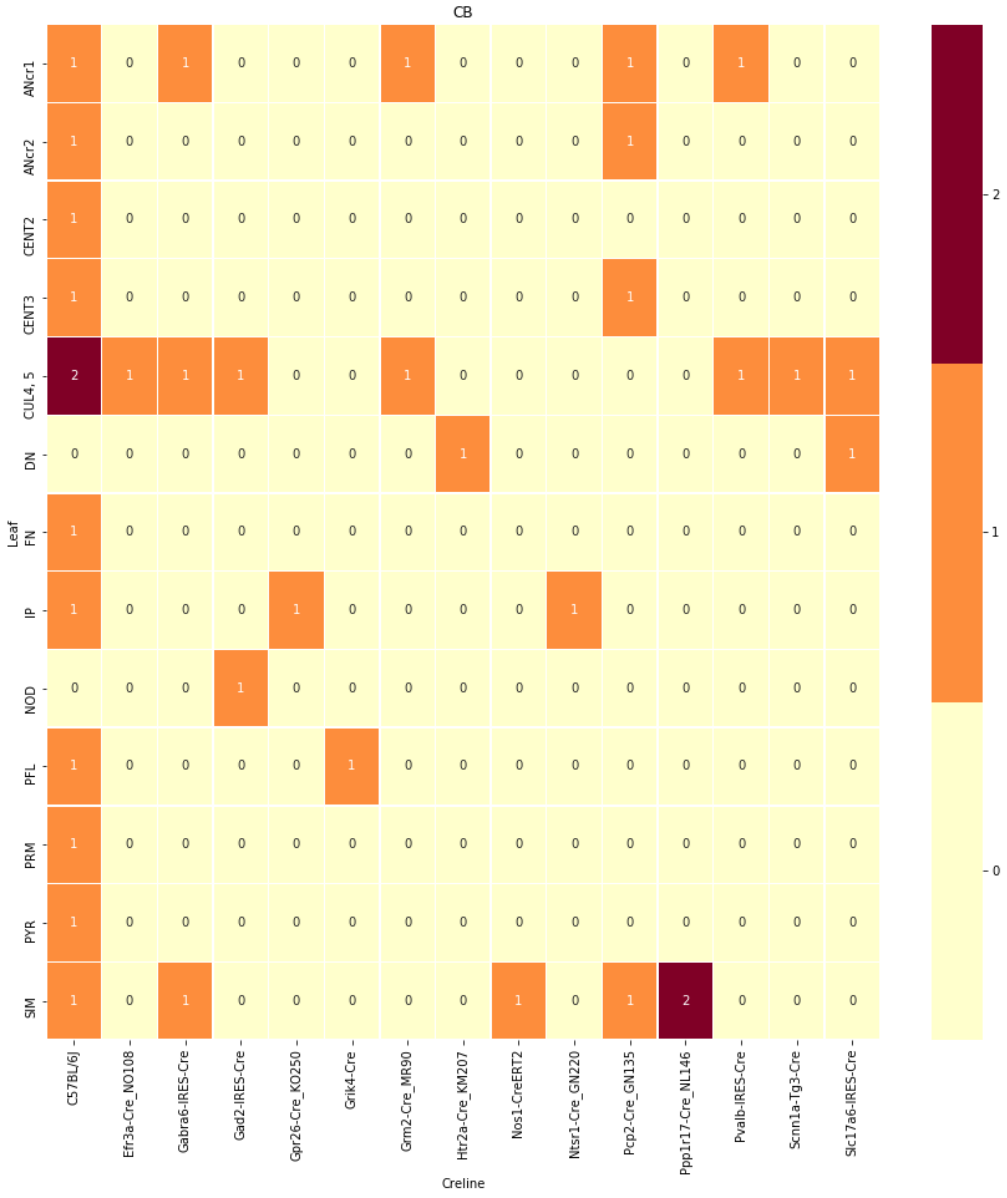
The Funder and award ID information you input at submission will be introduced by the publisher under a Funding Information head during production. Please use this space for any additional acknowledgements and verbiage required by your funders.

5 SUPPLEMENTAL INFORMATION

Cre/structure combinations in $\mathscr{D}$

This section describes the abundances of leaf and cre line combinations in our dataset. Users of the connectivity matrices who are interested in a particular cre line or structure can see the quantity and type of data used to compute and evaluate that connectivity.

centroid densityoct12.png



centroid densityoct12.png

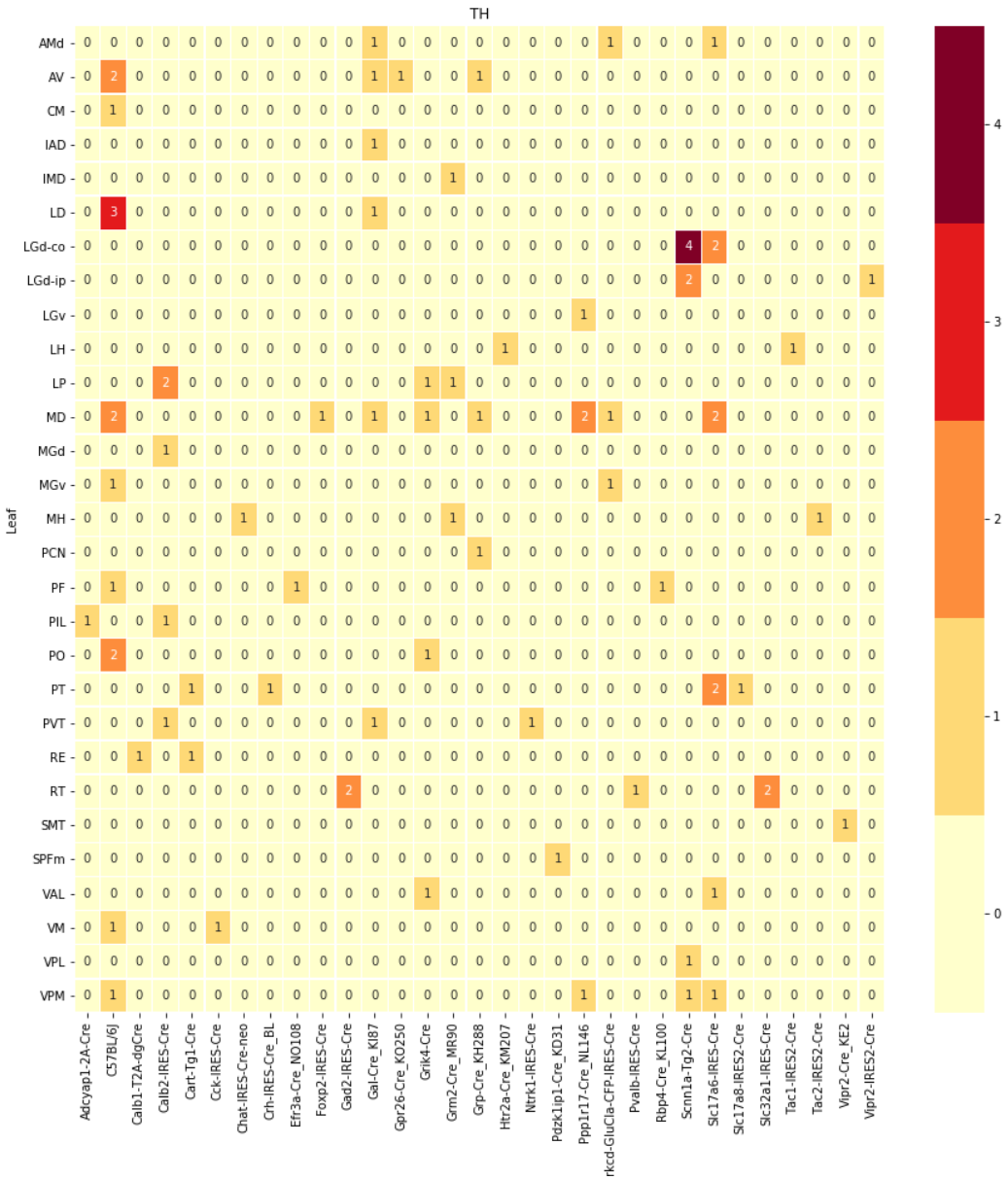


Figure 5: Caption

centroid densityoct12.png

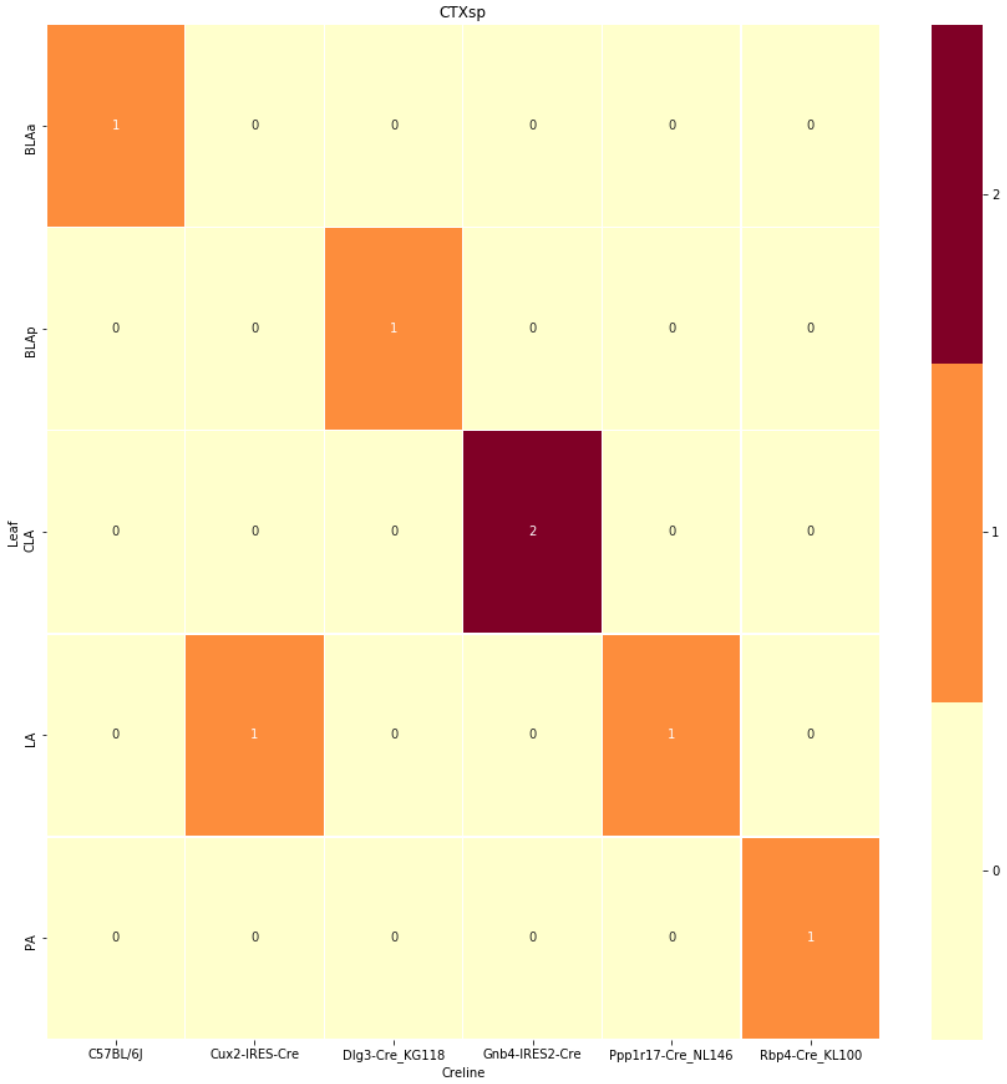
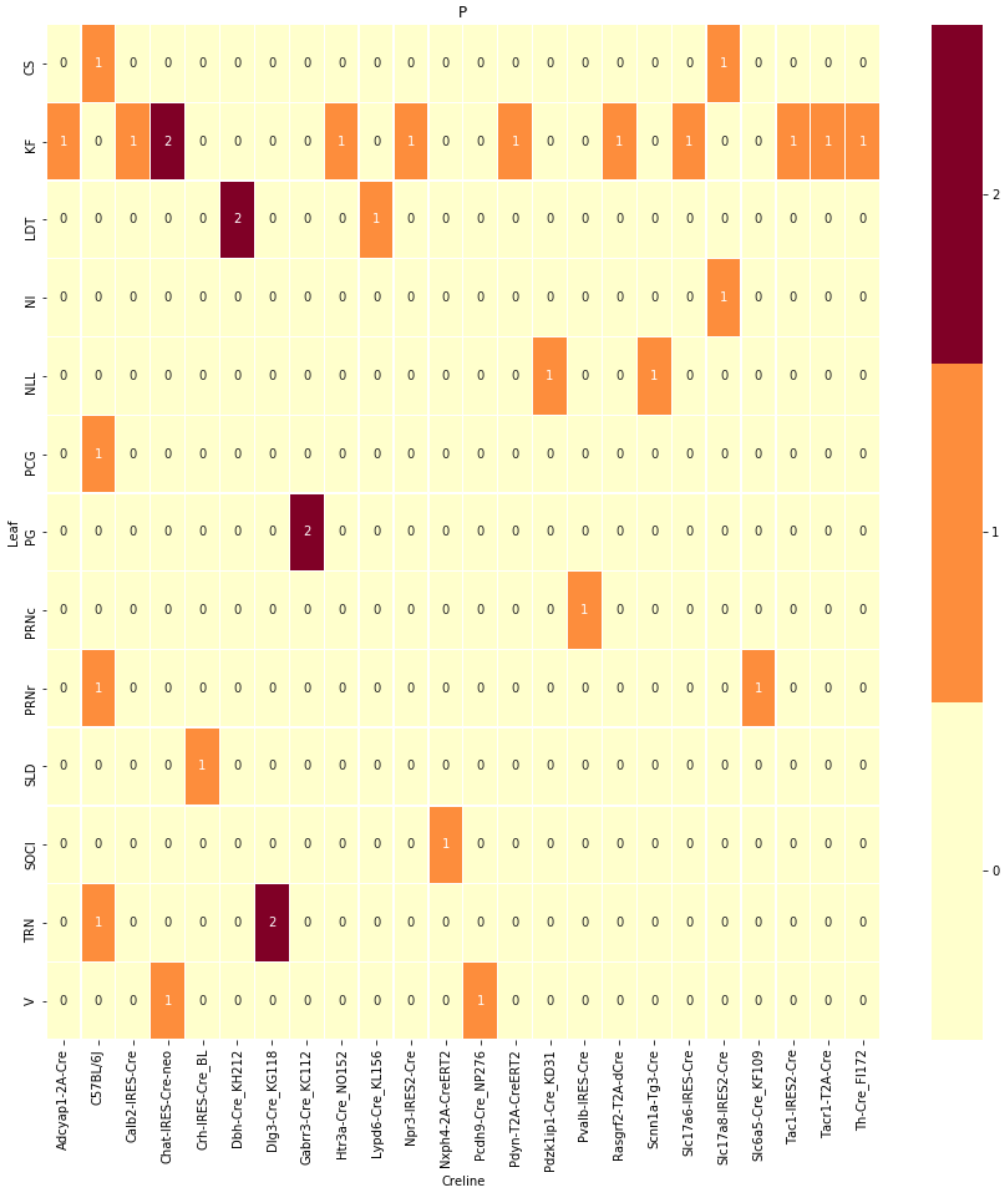
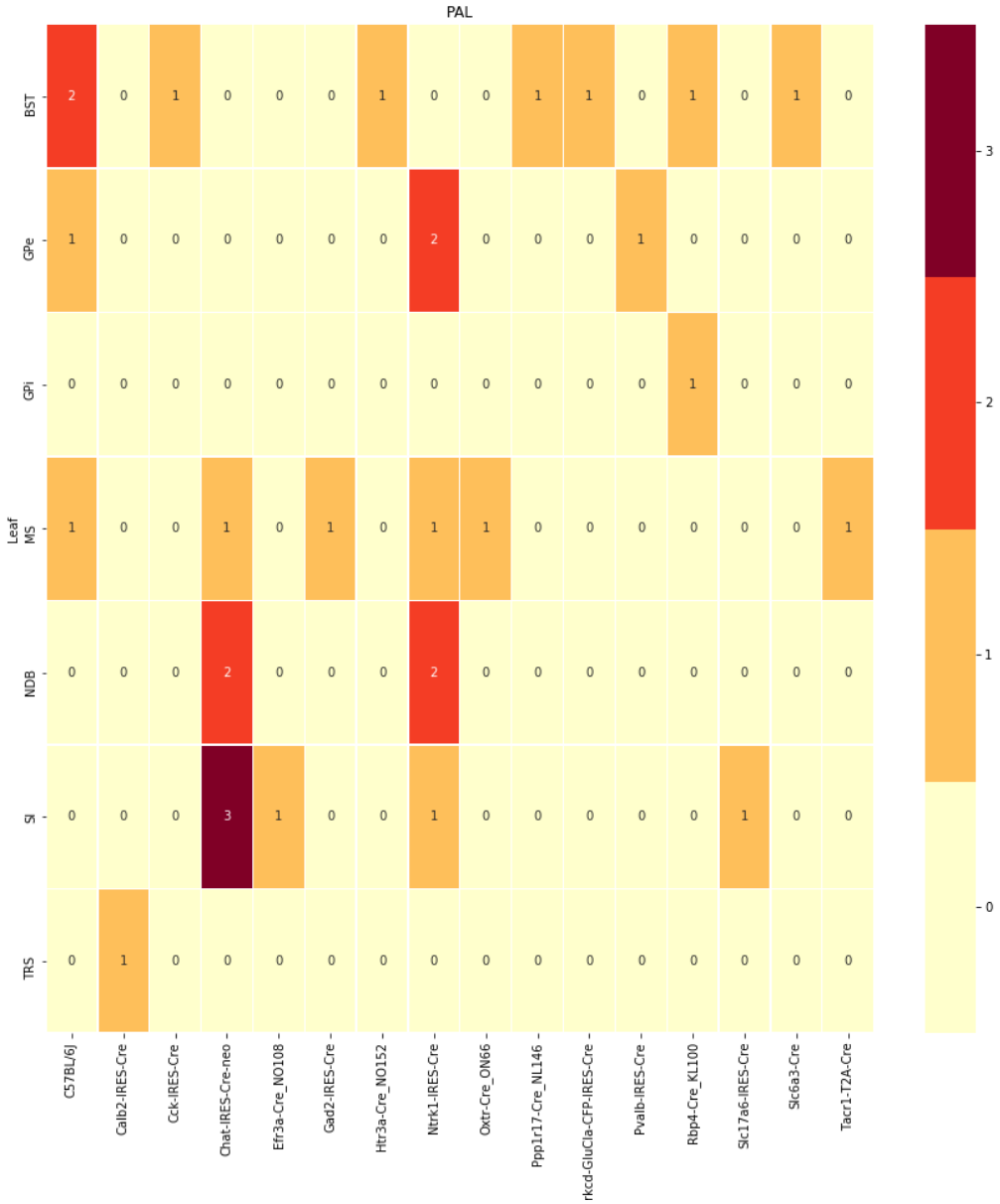


Figure 6: Caption

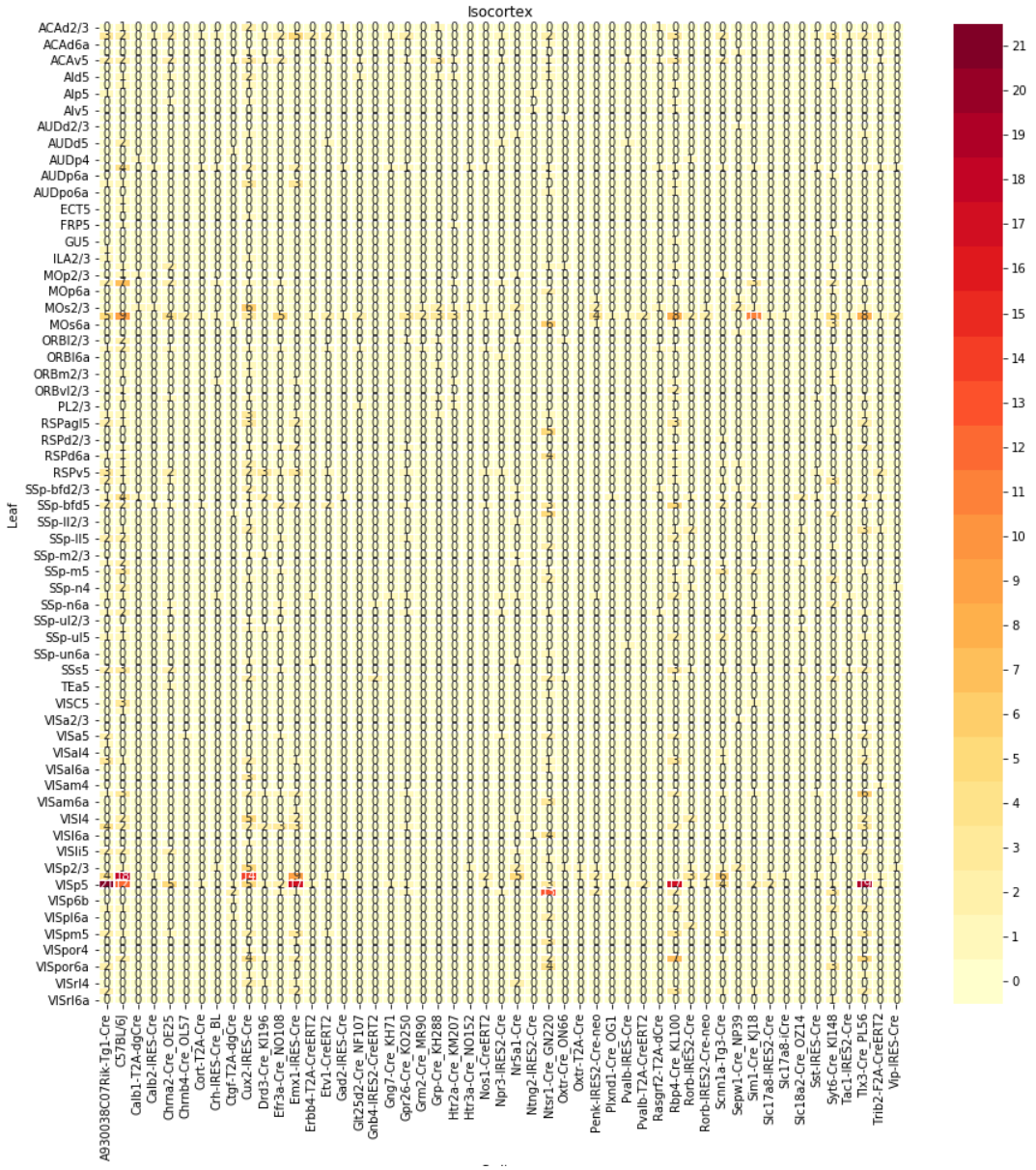
centroid densityoct12.png



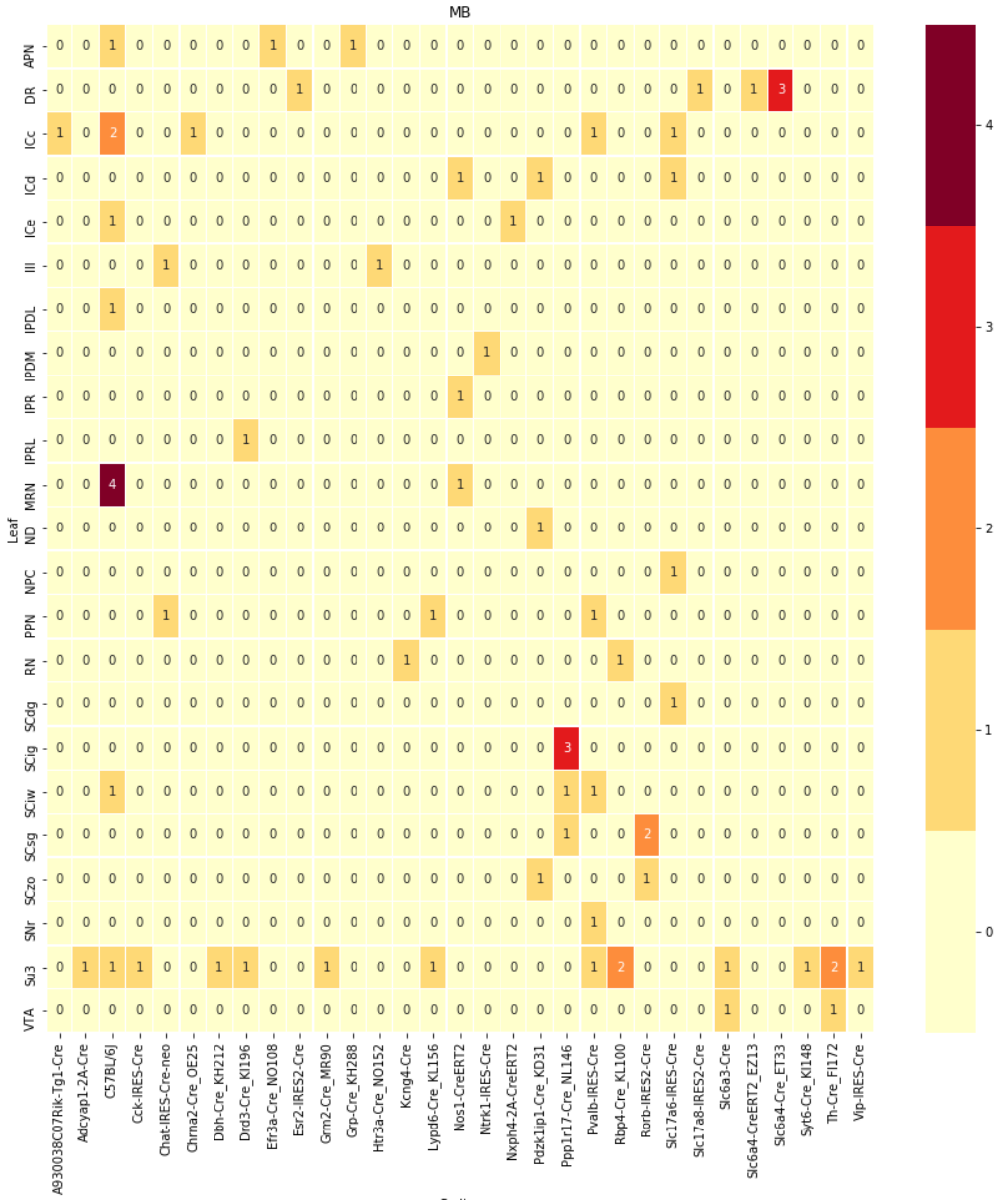
centroid densityoct12.png



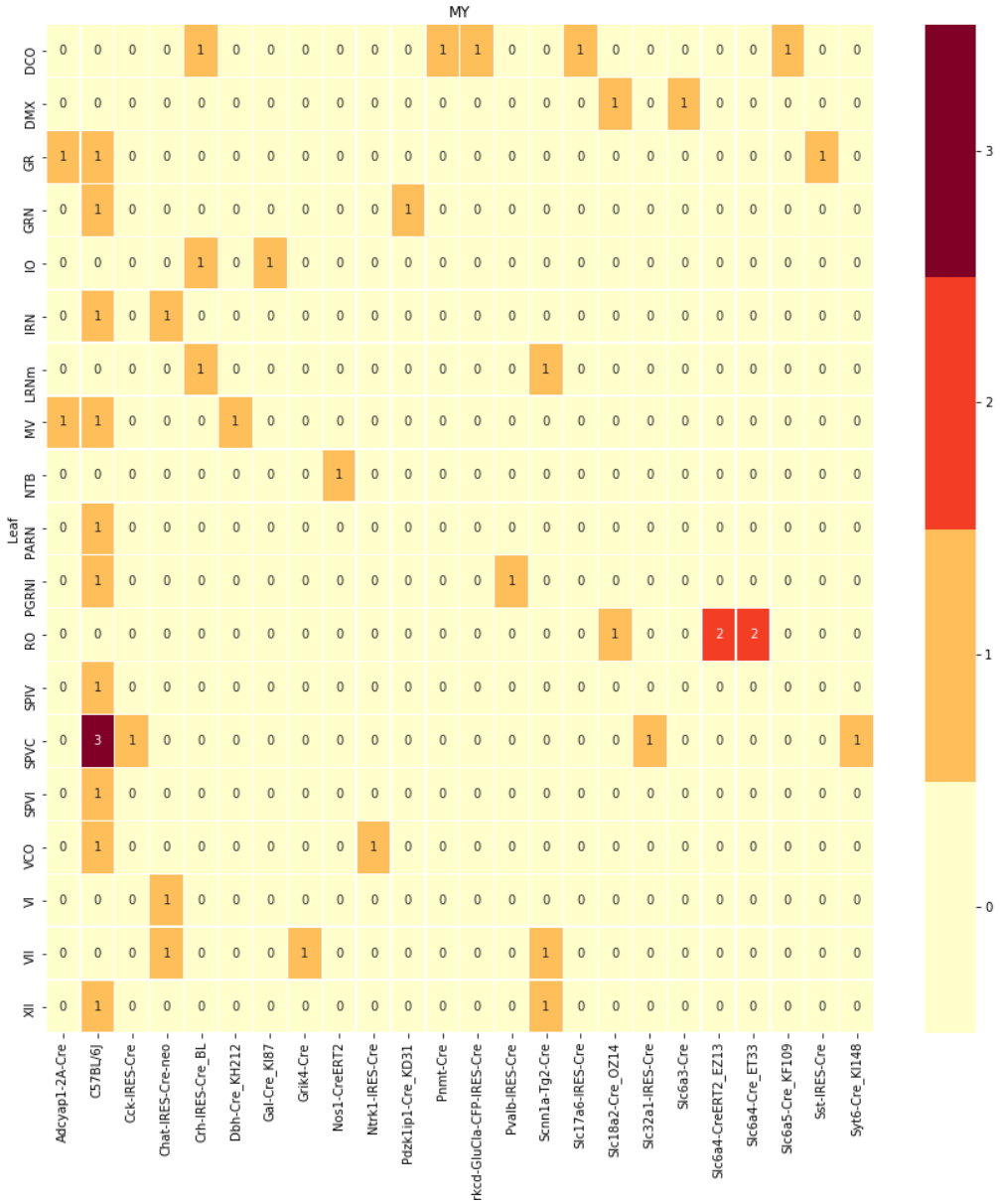
centroid densityoct12.png



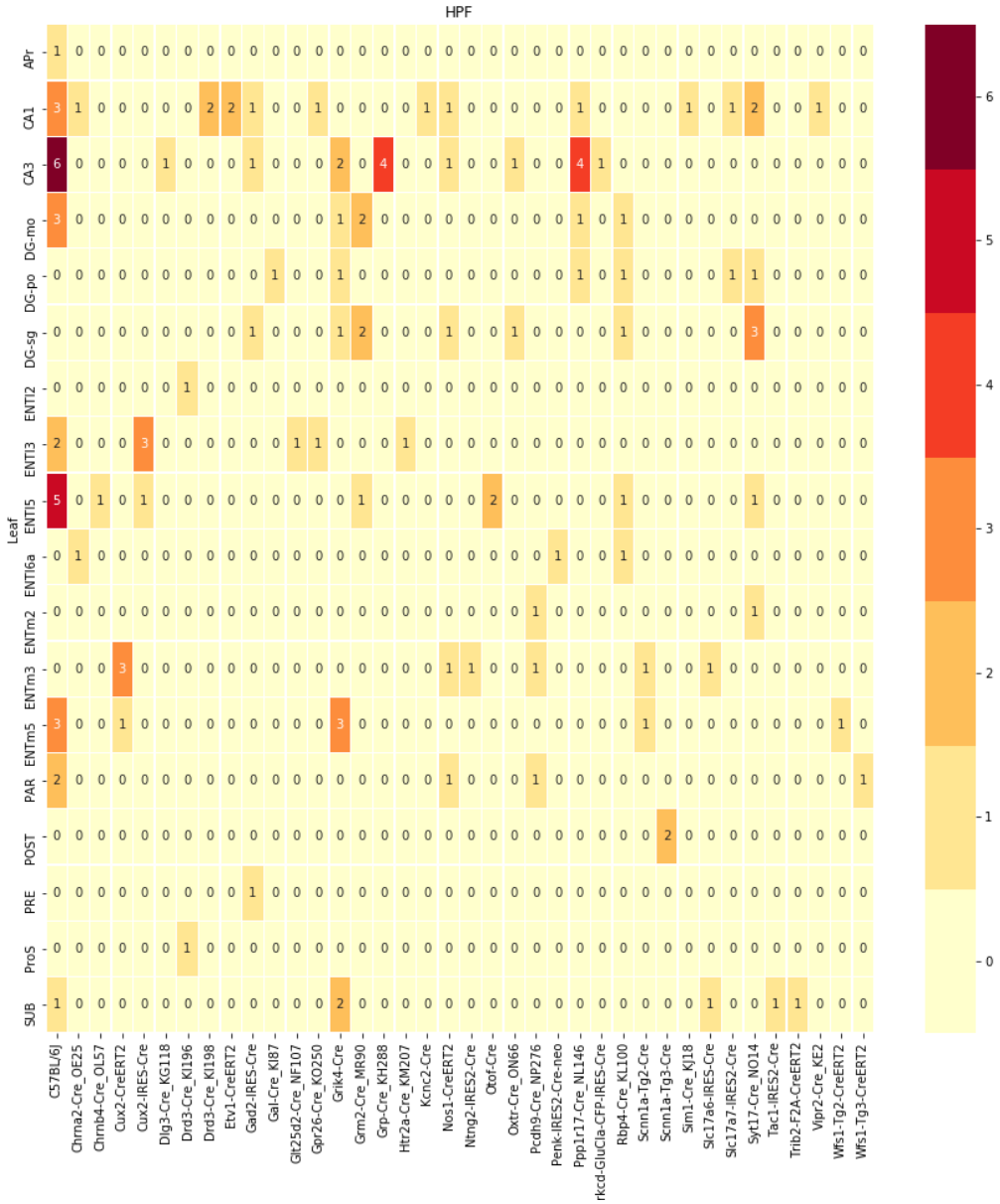
centroid densityoct12.png



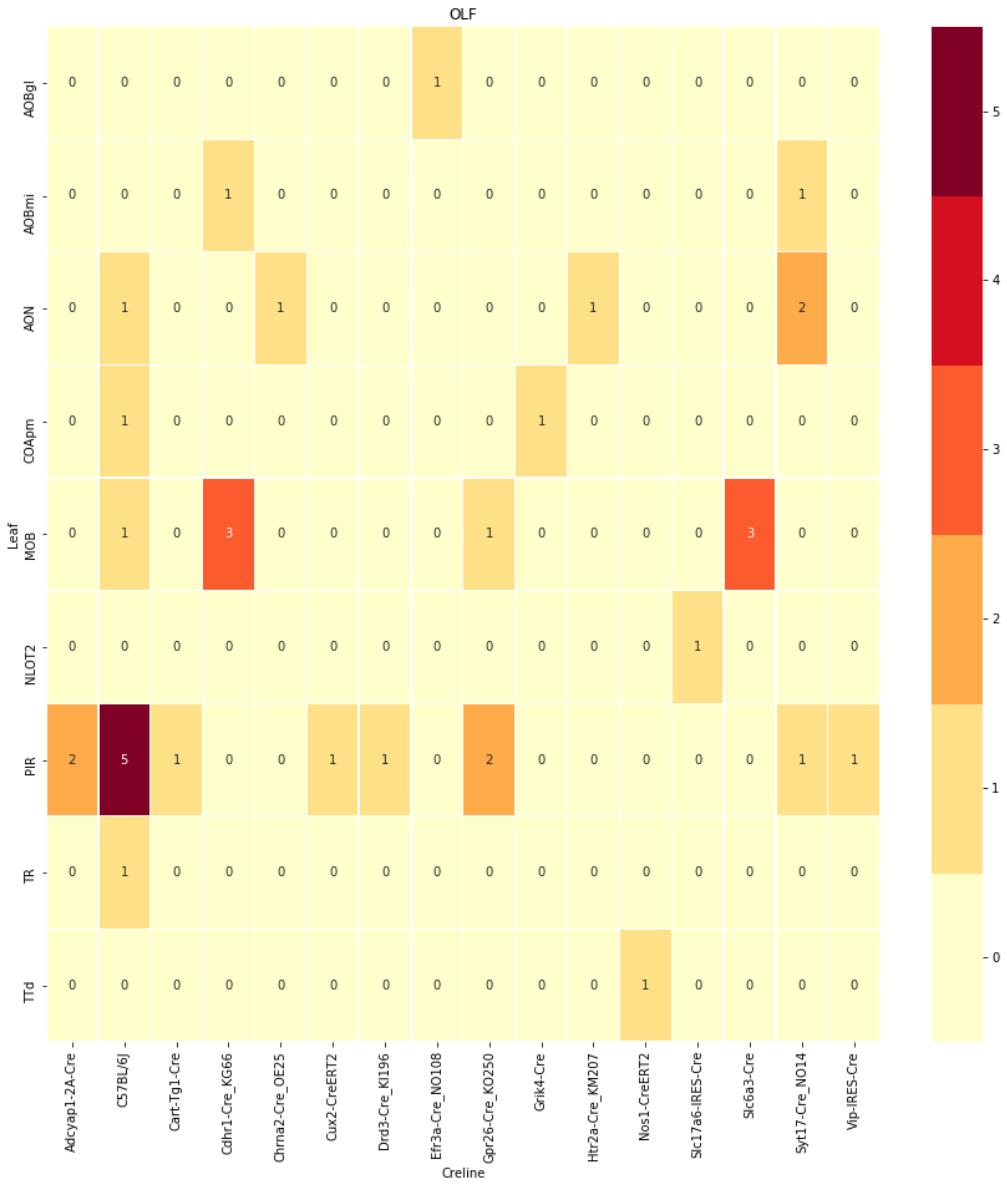
centroid densityoct12.png



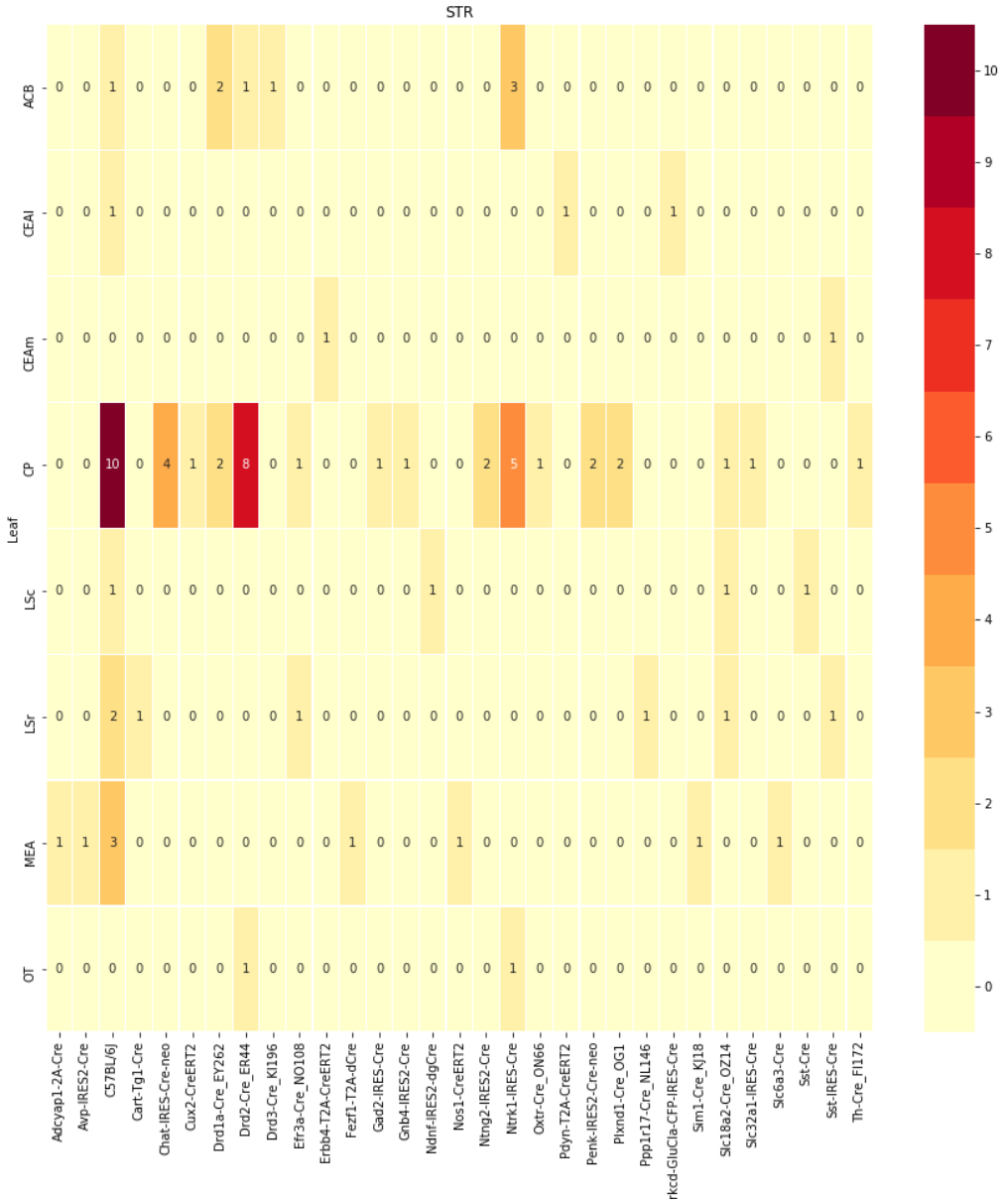
centroid densityoct12.png



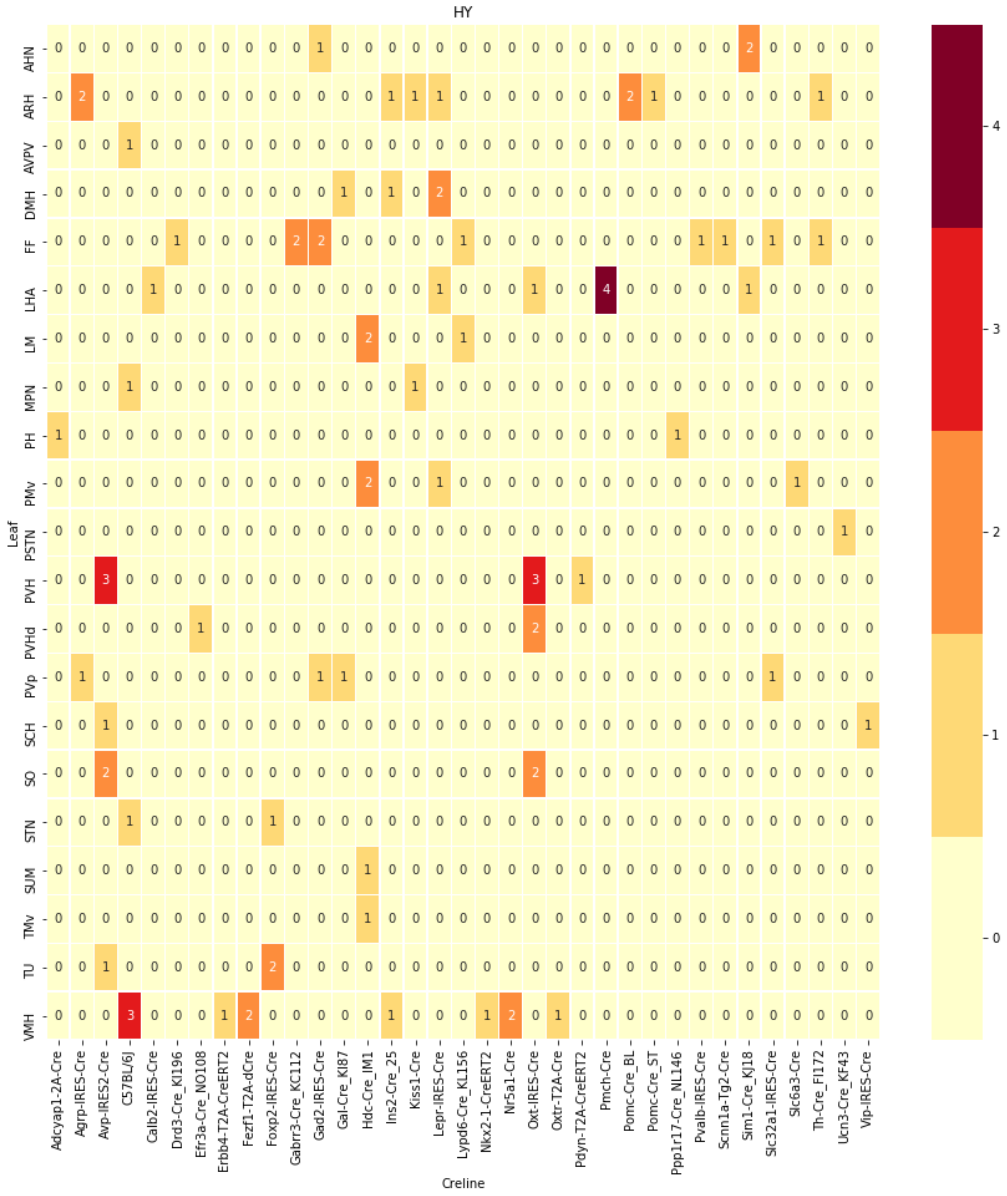
centroid densityoct12.png



centroid densityoct12.png



centroid densityoct12.png



Distances between structures

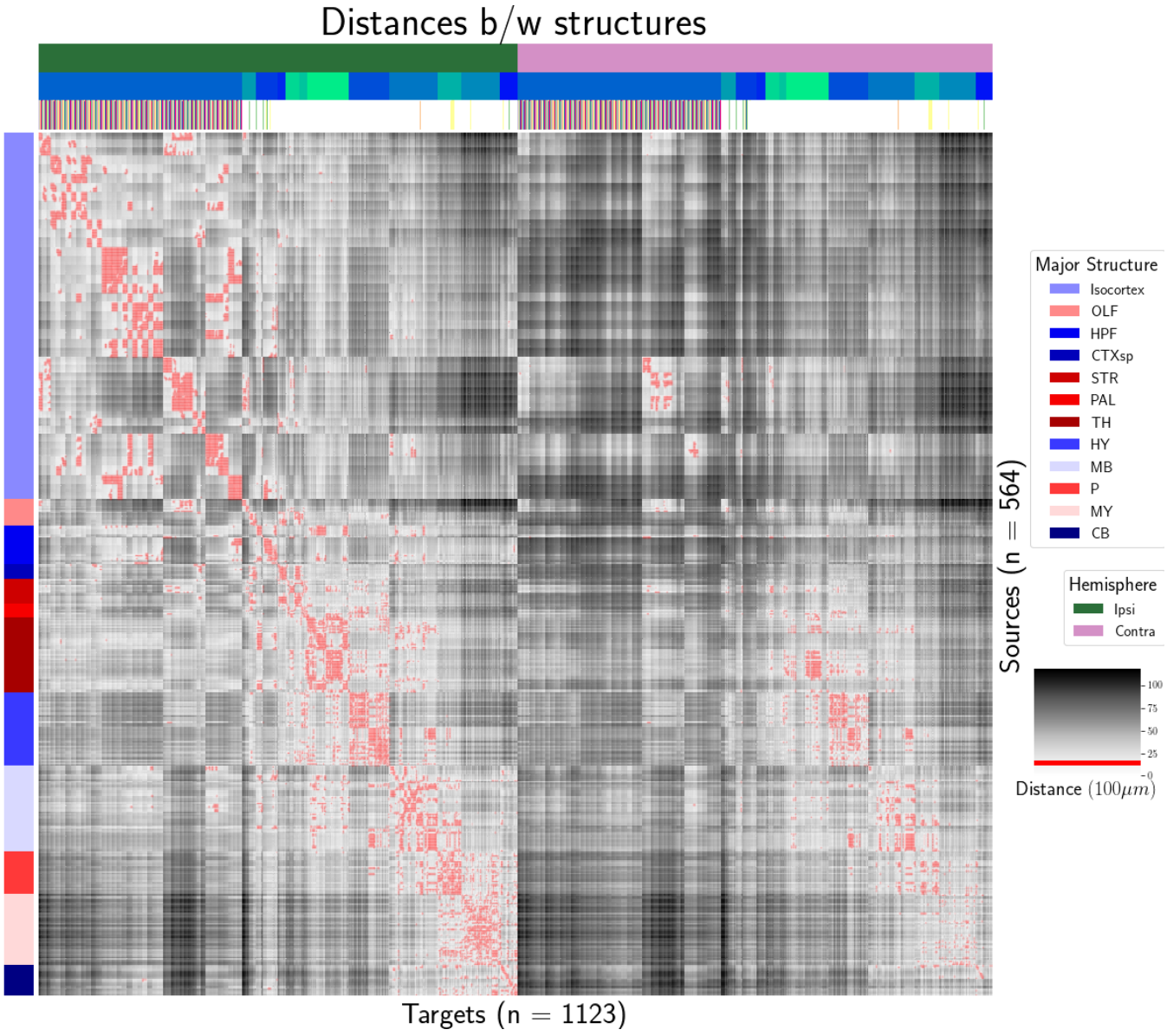


Figure 7: Distance between structures. Short-range connections are masked in red

Model evaluation

	Total	Cre-Leaf
Isocortex	36	4
OLF	7	2
HPF	122	62
CTXsp	85	41
STR	1128	732
PAL	68	18
TH	46	7
HY	35	17
MB	33	8
P	30	11
MY	78	45
CB	83	29

Table 3: Number of experiments available to evaluate models in leave-one-out cross validation. Models that rely on a finer granularity of modeling have less data available to validate with.

6 SUPPLEMENTAL METHODS

This section consists of additional information on preprocessing of the neural connectivity data, estimation of connectivity, and matrix factorization.

Data preprocessing

Several data prepreprocessing steps take place prior to evaluations of the connectivity matrices. These steps are described in Algorithm **??**. The arguments of this normalization process - injection signals $x(i)$, projection signals $y(i)$, injection fraction $F(i)$, and data quality mask $q(i)$ - were downloaded using the Allen SDK. The injections and projection signals $\in \mathscr{B} \times [0,1]$ were segmented manually in histological analysis. The projection signal gives the proportion of pixels within the voxel displaying fluorescence, and the injection signal gives the proportion of pixels within the histologically-selected injection subset displaying fluorescence. The injection fraction $\in \mathscr{B} \times [0,1]$ gives the proportion of pixels within each voxel in the injection subset. Finally, the data quality mask $\in \mathscr{B} \times \{0,1\}$ gives the voxels that have valid data.

Our preprocessing makes use of the above ingredients, as well as several other essential steps. First, we compute the weighted injection centroid

$$c(i) \quad = \sum_{l \in \mathscr{B}} x(i) l(v)$$

Given a regionalization $\mathscr{R}$, we also have access to a regionalization map as $R : \mathscr{B} \to \mathscr{R}$ which induces a map of connectivities

$$\begin{aligned} R_* : \mathscr{F} \qquad & \to \mathscr{R} \times \mathbb{R}^+ \\ (v, y) \quad & \mapsto \textstyle\sum_{v' \in R} y' \text{ for } (v, y') \text{ s.t. } R \ni v. \end{aligned}$$

This map depends on the choice of regionalization; we regionalize at the leaf level. We also can restrict a signal to a individual structure

$$\begin{aligned} S_* : \mathscr{F} \qquad & \to \mathscr{F} \\ (v, y) \quad & = \begin{cases} (v, y) \text{ if } v \in S \\ (v, 0) \text{ otherwise} \end{cases} \end{aligned}$$

294 Finally, given a vector or array a, we have the $L1$ normalization map

$$n : a \mapsto \frac{a}{\sum_{j=1}^{p} a_j}$$

PREPROCESS 1 Input Injection $x(i)$, Projection $y(i)$, Injection centroid $c(i) \in \mathbb{R}^3$, injection fraction $F(i)$, data quality mask $q(i)$

Injection fraction $x_F(i) \leftarrow x(i) \odot F(i)$

Data-quality censor $y_M(i) \leftarrow \odot y(i) \odot q(i), x_M(i) \leftarrow x_F(i) \odot F(i)$

Restrict injection $x_M(i) \odot S(i)$.

Compute centroid $c(i)$ from $x_M(i)$

Regionalize $y_S(i) \leftarrow R_*(y_M(i))$

Normalize $\tilde{y}(i) \leftarrow n(Y_S(i))$

Output $\tilde{y}(i)$, $c(i)$)

Estimators

Our estimators model a connectivity vector $f(v,s) \in \mathbb{R}^T_{\geq}$, and so we may write

$$f(v,s,t) = f(v,t)[t].$$

Thus, for the remainder of this section, we will discuss only $f(v,s)$.

Centroid-based Nadaraya-Watson In the Nadaraya-Watson approach of Knox et al. (2019), the injection is considered only through its centroid $c(i) := c(x(i))$, and the projection is considered regionalized. That is,

$$f_*(\mathscr{D}_i) = \{c(i), y_{\mathscr{T}}(i)\}.$$

Since the injection is considered only by its centroid, this model only generates predictions for particular locations c, and the prediction for a structure s is given by integrating over locations within the structure

$$f^*(\hat{f}(f_*(\mathscr{D})))(v,s) = \sum_{l_{s_j} \in s} \hat{f}(f_*(\mathscr{D}))(v, l_{s_j}),$$

This $\hat{f}$ is the Nadaraya-Watson estimator

$$\hat{f}_{NW}(c(I), y_{\mathscr{T}}(I))(l) := \sum_{i \in I} \frac{\omega_{c(i)l}}{\sum_{i \in I} \omega_{c(i),l}} y_{\mathscr{T}}(i)$$

where $\omega_{c(i)l} = \exp(-\gamma d(l, c(i))^2)$ and d is the Euclidean distance between centroid $c(i)$ and voxel with position l.

Several facets of the estimator are visible here. A smaller γ corresponds to a greater amount of smoothing, and index set $I \subseteq \{1:n\}$ indicates which experiments to use to generate the prediction. Fitting γ via empirical risk minimization therefore bridges between 1-nearest neighbor prediction and averaging of all experiments in I. In Knox et al. (2019), I consisted of experiments sharing the same brain division. Restricting of index set to only include experiments with the same neuron class gives the class-specific Cre-NW model.

The expected-loss estimator The response induced by each of the Cre-lines is effected by both the injection location and the targeted cell types. Since Cre-lines that target similar cell classes are therefore expected to induce similar projections, and including similar Cre-lines in the Nadaraya-Watson estimator increases the effective sample size, we introduce an estimator that assigns a predictive weight to each training point that depends both on its centroid-distance and Cre-line. This weight is determined by the expected prediction error of each of the two feature types, as determined by cross-validation. These weights are then utilized in a Nadaraya-Watson estimator in a final prediction step. Estimating $\hat{f}(v, c)$ shares the advantage of fine-scale spatial resolution with Knox et al. (2019), but in addition enables us to model a particular cell-class v.

We formalize Cre-line behavior as the average regionalized projection of a Cre-line in a given structure (i.e. leaf). This vectorization of categorical information is known as **target encoding**. We define a **Cre-distance** in a leaf to be the distance between the target-encoded projections of two Cre-lines. The relative predictive accuracy of Cre-distance and centroid distance is determined by fitting a surface of projection distance as a function of Cre-distance and centroid distance.

In mathematical terms, our full feature set consists of the centroid coordinates and the target-encoded means of the combinations of virus type and injection-centroid structure. That is,

$$f_*(\mathscr{D}_i) = \{c(i), \bar{y}_{\mathscr{T}}(I_v \cap I_s), y_{\mathscr{T}}(i)\}.$$

f^* is defined as in (2). The expected loss estimator is then

$$\hat{f}_{EL}(c, c(i), v, y_{\mathscr{T}}(I_v \cap I_s)) = \sum_{i \in I} \frac{\nu(c(i), c, v(i), v)}{\sum_{i \in I} \nu(c(x_i), c, v_i, v)} r(y_i)$$

where

$$\nu_i = \exp(-\gamma g(d(c, c(x_i))^2, d(\bar{r}(v), \bar{r}(v_i))^2))$$

Note that g must be a concave, non-decreasing function of its arguments with with $g(0,0) = 0$, then g defines a metric on the product of the metric spaces defined by experiment centroid and target-encoded cre-line, and $\hat{f}_{EL}$ is a Nadaraya-Watson estimator. A derivation of this fact is given in Appendix 6, and we therefore use shape-constrained B-splines to estimate g.

EL 2 Input Projection $y_{\mathcal{T}}(I_s)$, Injection centroids $c(I_s) \in \mathbb{R}^3$, Cell-classes $v(I_s)$, g, location l, cell-class v

Get structures $s(1:n) = r(c(1:n))$, $s = r(c)$

Target encode $v(1:n)$ and v with $n(r(y(1:n)))$

Estimate expected losses $X = [g(\|l - c(i')\|_2^2, \|\bar{y}_{\mathcal{T}}(v,s) - \bar{y}_{\mathcal{T}}(v(i'),s)\|_2^2) : i' \in (I_s)]$

Predict $\widehat{y}_{\mathcal{T}} = NW(X, y_{\mathcal{T}}(I_s)$

Output $\tilde{y}(i)$, $c(i)$)

Figure 8: The Expected-Loss estimator

JUSTIFICATION OF SHAPE CONSTRAINT The shape-constrained expected-loss estimator introduced in this paper is, to our knowledge, novel. It should be considered an alternative method to the classic weighted kernel method. While we do not attempt a detailed theoretical study of this estimator, we do establish the need for the shape constraint in our spline estimator. Though this fact is probably well known, we prove a (slightly stronger) version here for completeness.

Given a collection of metric spaces $X_1, ...X_n$ with metrics $d_1...d_n$ (e.g. $d_{centroid}, d_c re$), and a function $f : (X_1 \times X_1)... \times (X_n \times X_n) = g(d_1(X_1 \times X_1), ...d_n(X_n \times X_n))$, then then f is a metric iff g is concave, non-decreasing and $g(d) = 0 \iff d = 0$.

We first show g satisfying the above properties implies that f is a metric.

- The first property of a metric is that $f(x, x') = 0 \iff x = x'$. The left implication: $x = x' \implies f(x_1, x_1', ...x_n, x_n') = g(0, ..., 0)$, since d are metrics. Then, since $g(0) = 0$, we have that $f(x, x') = 0$. The right implication: $f(x, x') = 0 \implies d = 0 \implies x = x'$ since d are metrics.
- The second property of a metric is that $f(x, x') = f(x', x)$. This follows immediately from the symmetry of the d_i, i.e. $f(x, x') = f(x_1, x_1', ...x_n, x_n') = g(d_1(x_1, x_1'), ...d_n(x_n, x_n')) = g(d_1(x_1', x_1), ...d_n(x_n', x_n)) = f(x_1', x_1, ...x_n', x_n) = f(x', x)$.
- The third property of a metric is the triangle inequality: $f(x, x') \leq f(x, x^*) + f(x^*, x')$. To show this is satisfied for such a g, we first note that $f(x, x') = g(d(x, x')) \leq g(d(x, x^*) + d(x^*, x'))$ since g is non-decreasing and by the triangle inequality of d. Then, since g is concave, $g(d(x, x^*) + d(x^*, x')) \leq g(d(x, x^*)) + g(d(x^*, x')) = f(x, x^*) + f(x^*, x')$.

We then show that f being a metric implies that g satisfies the above properties.

- The first property is that $g(d) = 0 \iff d = 0$. We first show the right implication: $g(d) = 0$, and $g(d) = f(x, x')$, so $x = x'$ (since f is a metric), so $d = 0$. We then show the left implication: $d = 0 \implies x = x'$, since d is a metric, so $f(x, x') = 0$, since f is a metric, and thus $g(d) = 0$.
- The second property is that g is non-decreasing. We proceed by contradiction. Suppose g is decreasing in argument d_1 in some region $[l, u]$ with $0 < l < u$. Then $g(d_1(0, l), 0) \geq g(d_1(0, 0), 0) + g(d_1(0, u), 0) = g(d_1(0, u), 0)$, which violates the triangle inequality on f. Thus, decreasing g means that f is not a metric, so f a metric implies non-decreasing g.

- The final property is that g is concave. We proceed by contradiction. Suppose g is strictly convex. Then there exist vectors d, d' such that $g(d+d') < g(d) + g(d')$. Assume that d and d' only are non-zero in the first position, and $d = d(0,x), d' = d(0,x')$. Then, $f(0,x) + f(0,x') < f(0,x+x')$, which violates the triangle inequality on f. Therefore, g must be concave.

Establishing a lower detection limit

The lower detection limit of our approach is a complicated consequence of our experimental and analytical protocols. For example, the Nadaraya-Watson estimator is likely to generate many small false positive connections, since the projection of even a single experiment within the source region to a target will cause a non-zero connectivity in the Nadaraya-Watson weighted average. On the other hand, the complexities of the experimental protocol itself and the image analysis and alignment can also cause spurious signals. Therefore, it is of interest to establish a lower-detection threshold below which we have very little power-to-predict, and set estimated connectivities below this threshold to zero. This should make our estimated connectivities more accurate, especially in the biologically-important sense of sparsity.

We establish this limit with respect to the sum of Type 1 and Type 2 errors

$$\iota = \sum_{i \in \mathscr{E}} 1^T_{y_{\mathscr{T}}(i)=0} 1_{\hat{f}(v(i),c(i))>\tau} + 1^T_{y_{\mathscr{T}}(i)>0} 1_{\hat{f}(v(i),c(i))<\tau}.$$

We then select the τ that minimizes ι. Results for this approach are given in Supplemental Section 7.

Decomposing the connectivity matrix

We utilize non-negative matrix factorization (NMF) to analyze the principal signals in our connectivity matrix. Here, we review this approach as applied to decomposition of the distal elements of the estimated connectivity matrix $\widehat{\mathscr{C}}$ to identify q connectivity archetypes. Aside from the NMF program itself, the key elements are selection of the number of archetypes q and stabilization of the tendency of NMF to give random results over different initialization.

Non-negative matrix factorization Given a matrix $X \in \mathbb{R}^{a\times b}_{\geq 0}$ and a desired latent space dimension q, the non-negative matrix factorization is

$$\mathrm{NMF}(\mathscr{V}, \lambda, q) = \arg\min_{W,H} \frac{1}{2}\|1_M \odot \mathscr{C} - WH\|_2^2 + \lambda(\|H\|_1 + \|W\|_1).$$

We note the existence of NMF with alternative norms for certain marginal distributions, but leave utilization of this approach for future work (Brunet, Tamayo, Golub, & Mesirov, 2004).

The mask $1_M \in \{0,1\}^{S\times T}$ serves two purposes. First, it enables computation of the NMF objective while excluding self and nearby connections. These connections are both strong and linearly independent, and so would unduly influence the NMF reconstruction error over more biologically interesting or cell-type dependent long-range connections. Second, it enables cross-validation based selection of the number of retained components.

Cross-validating NMF Cross-validation for NMF is somewhat standard but not entirely well-known, and so we review it here. In summary, a NMF model is first fit on a reduced data set, and an evaluation set is held out. After random masking of the evaluation set, the loss of the learned model is then evaluated on the basis of successful reconstruction of the held-out values. This procedure is performed repeatedly, with replicates of random masks at each tested dimensionality q. This determines the point past which additional hidden units provide no reconstructive value.

The differentiating feature of cross-validation for NMF compared with supervised learning is the random masking of the matrix $\mathscr{C}$. Cross-validation for supervised learning generally leaves out entire observations, but this is insufficient for our situation. This is because, given W, our H is the solution

of a regularized non-negative least squares optimization problem

$$H := e_W(X) = \arg\min_{\beta \in \mathbb{R}^{q\times T}_{\geq 0}} \|X - W\beta\|_2^2 + \|\beta\|_1.$$

The negative effects of an overfit model can therefore be optimized away from on the evaluation set.

The standard solution is to generate uniformly random masks $1_{M(p)} \in \mathbb{R}^{S\times T}$ where

$$1_{M(p)}(s,t) \sim \text{Bernoulli(p)}.$$

Our cross-validation error is then

$$\epsilon_q = \frac{1}{R}\sum_{r=1}^{R} (\|1_{M(p)_r^C} \odot X - \widehat{d}_q(\widehat{e}_q(1_{M(p)_r^C} \odot X))\|_2^2$$

where

$$\widehat{d}_q, \widehat{e}_q = \widehat{\text{NMF}}(1_{M(p)_r} \odot X, q)).$$

Here, $1^C_{M(p)_r}$ is the binary complement of $1_{M(p)_r}$.

Theoretically, the optimum number of components is then

$$\widehat{q} = \arg\min_q \epsilon_q.$$

However, the low decrease in error at higher values of q will motivate us to empirically select a slightly smaller number of components.

Stabilizing NMF The NMF program is non-convex, and, empirically, individual replicates will not converge to the same optima. One solution therefore is to run multiple replicates of the NMF algorithm and cluster the resulting vectors. This approach raises the questions of how many clusters to use, and how to deal with stochasticity in the clustering algorithm itself. We address this issue through the notion of clustering stability (von Luxburg, 2010a).

The clustering stability approach is to generate L replicas of k-cluster partitions $\{C_{kl} : l \in 1\ldots L\}$ and then compute the average dissimilarity between clusterings

$$\xi_k = \frac{2}{L(L-1)}\sum_{l=1}^{L}\sum_{l'=1}^{l} d(C_{kl}, C_{kl'}).$$

Then, the optimum number of clusters is

$$\hat{k} = \arg\min_k \xi_k.$$

A review of this approach is found in von Luxburg (2010b). Intuitively, archetype vectors that cluster together frequently over clustering replicates indicate the presence of a stable clustering. For d, we utilize the adjusted Rand Index - a simple dissimilarity measure between clusterings. Note that we expect to select slightly more than the q components suggested by cross-validation, since archetype vectors which appear in one NMF replicate generally should appear in others. We then select the q clusters with the most archetype vectors - the most stable NMF results - and take the median of each cluster to create a sparse representative archetype Kotliar et al. (2019); Wu et al. (2016). Experimental results for these cross-validation and stability selection approaches are given in Supplemental Section 7.

7 SUPPLEMENTAL EXPERIMENTS

Establishing a lower limit of detection

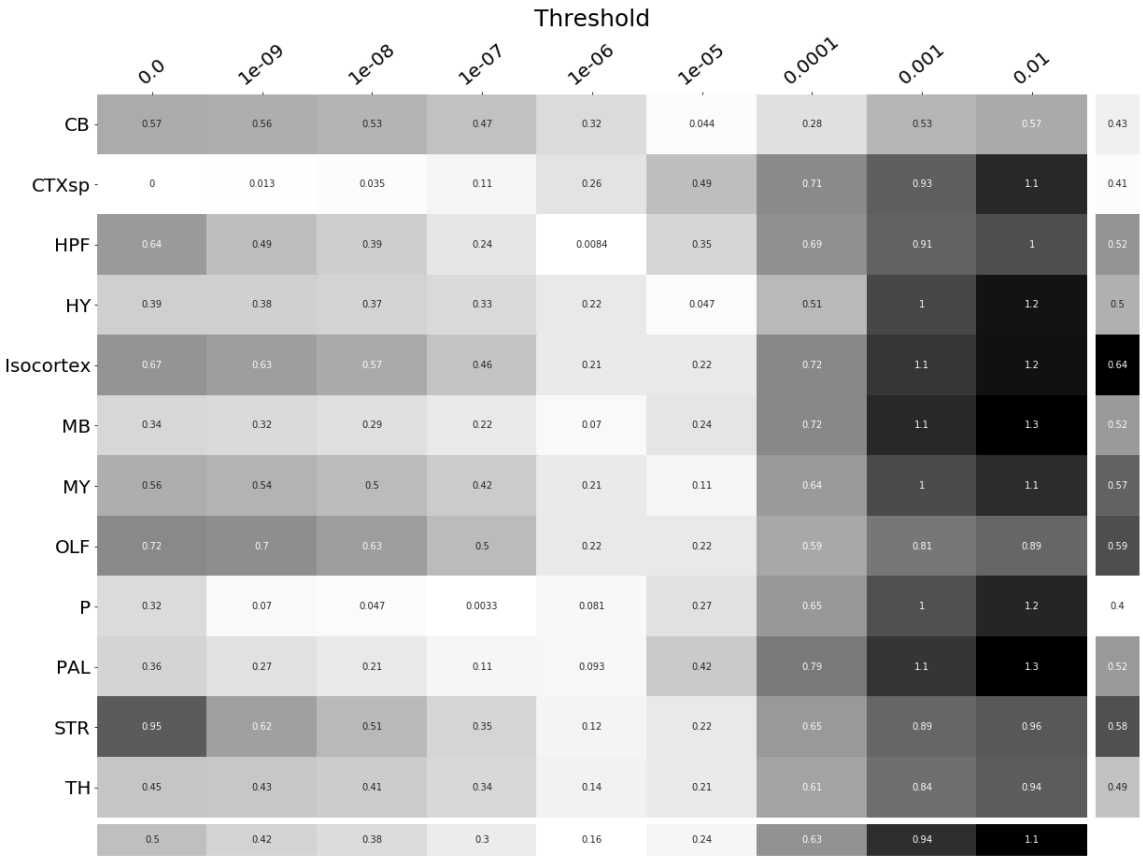


Figure 9: τ at different limits of detection.

Loss subsets

We report model accuracies for our *EL* model by neuron class and structure. These expand upon the results in Table ?? and give more specific information about the quality of our estimates.

Figure 10

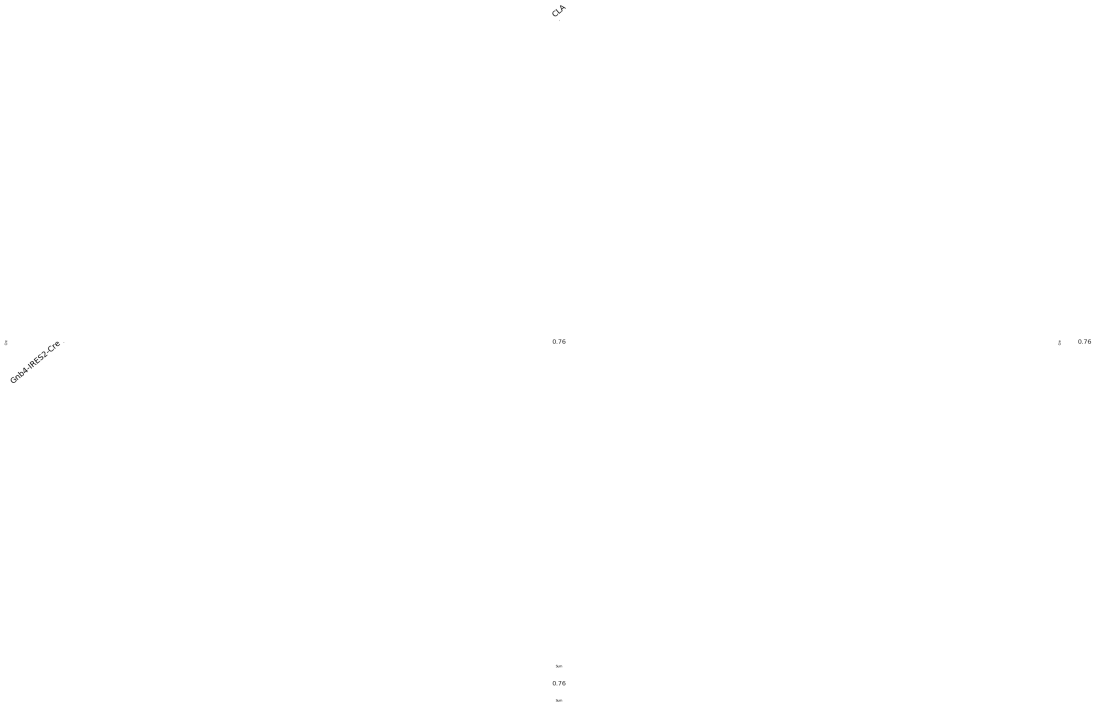


Figure 11

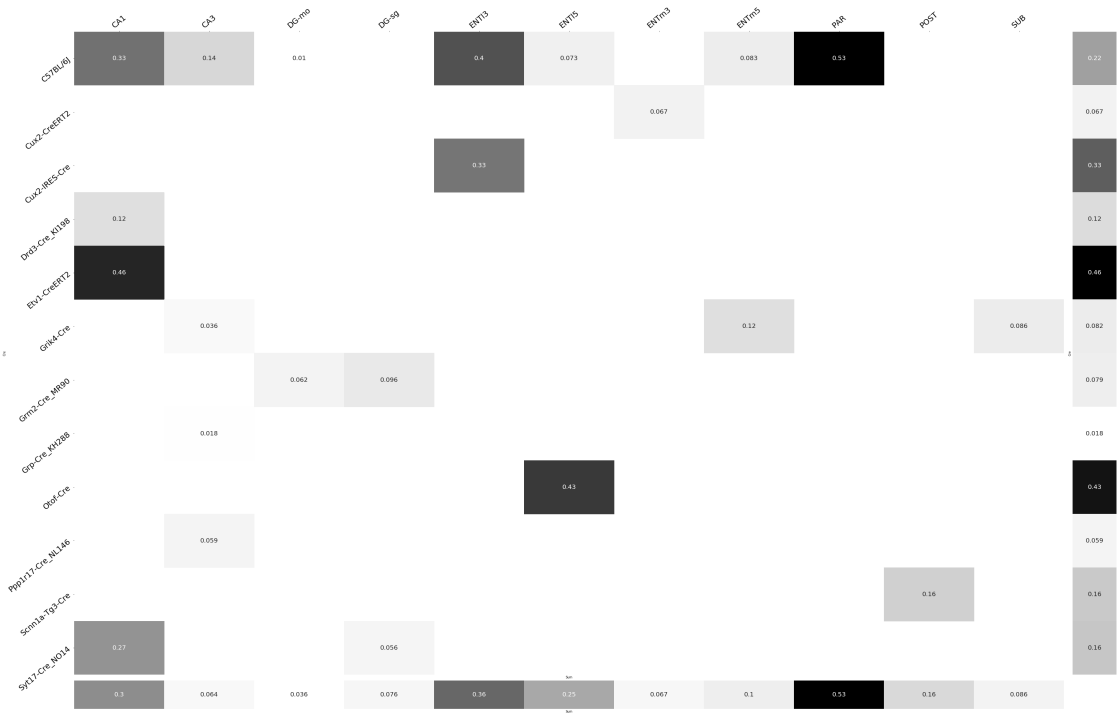

CA1
CA3
DG-mo
DG-sg
ENTl3
ENTl5
ENTm3
ENTm5
PAR
POST
SUB
C57BL/6J
Cux2-CreERT2
Cux2-IRES-Cre
Drd3-Cre_KI198
Etv1-CreERT2
Grik4-Cre
Grm3-Cre_MR90
Grp-Cre_KH288
Otof-Cre
Pbp1r17-Cre_NL146
Scnn1a-Tg3-Cre
Syt17-Cre_NO14
0.33 0.14 0.01 0.4 0.073 0.083 0.53 0.22
0.067 0.067
0.33 0.33
0.12 0.12
0.46 0.46
0.036 0.12 0.086 0.082
0.062 0.096 0.079
0.018 0.018
0.43 0.43
0.059 0.059
0.16 0.16
0.27 0.056 0.16
0.3 0.064 0.036 0.076 0.36 0.25 0.067 0.1 0.53 0.16 0.086

Figure 12

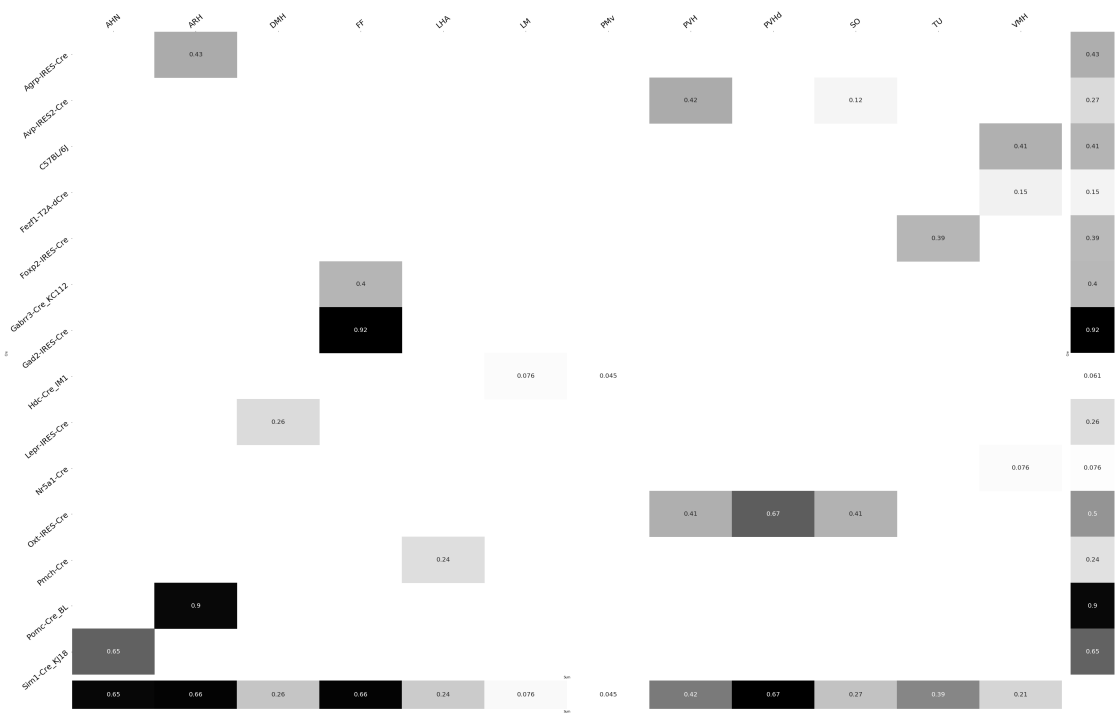

AHN
ARH
DMH
FF
LHA
LM
PMv
PVH
PVHd
SO
TU
VMH
Agrp-IRES-Cre
Avp-IRES2-Cre
C57BL/6J
Fezf1-T2A-dCre
Foxp2-IRES-Cre
Gabrr3-Cre_KC112
Gad2-IRES-Cre
Hdc-Cre_JM1
Lepr-IRES-Cre
Nr5a1-Cre
Oxt-IRES-Cre
Pmch-Cre
Pomc-Cre_BL
Slm1-Cre_KJ18

Figure 13

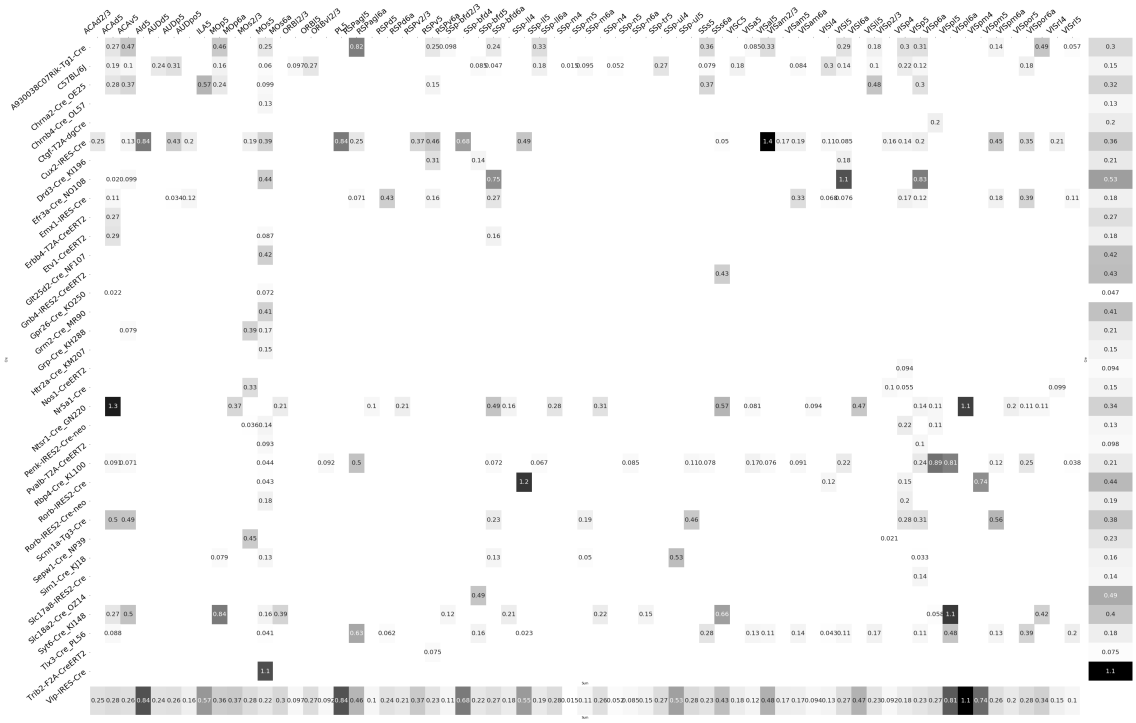


Figure 14


DR
ICc
MRN
SCig
SCsg
Su3
C57BL/6J
Ppp1r17-Cre_NL146
Rbp4-Cre_KL100
Rorb-IRES2-Cre
Slc6a4-Cre_ET33
Th-Cre_FI172
0.059
0.044
0.052
0.18
0.18
0.28
0.28
0.026
0.026
0.12
0.12
0.71
0.71
0.12
0.059
0.044
0.18
0.026
0.5

Figure 15


RO
SPVC
C57BL/6J
Slc6a4-CreERT2_EZ13
Slc6a4-Cre_ET33
0.27
0.35
0.31
0.19
0.19
0.27
0.35
0.19

Figure 16


AON
MOB
PIR
Adcyap1-2A-Cre
C57BL/6J
Cdhr1-Cre_KG66
Gpr26-Cre_KO250
Slc6a3-Cre
Syt17-Cre_NO14
0.21
0.21
0.037
0.037
0.025
0.025
0.04
0.04
0.63
0.63
0.46
0.46
0.46
0.33
0.097

Figure 17

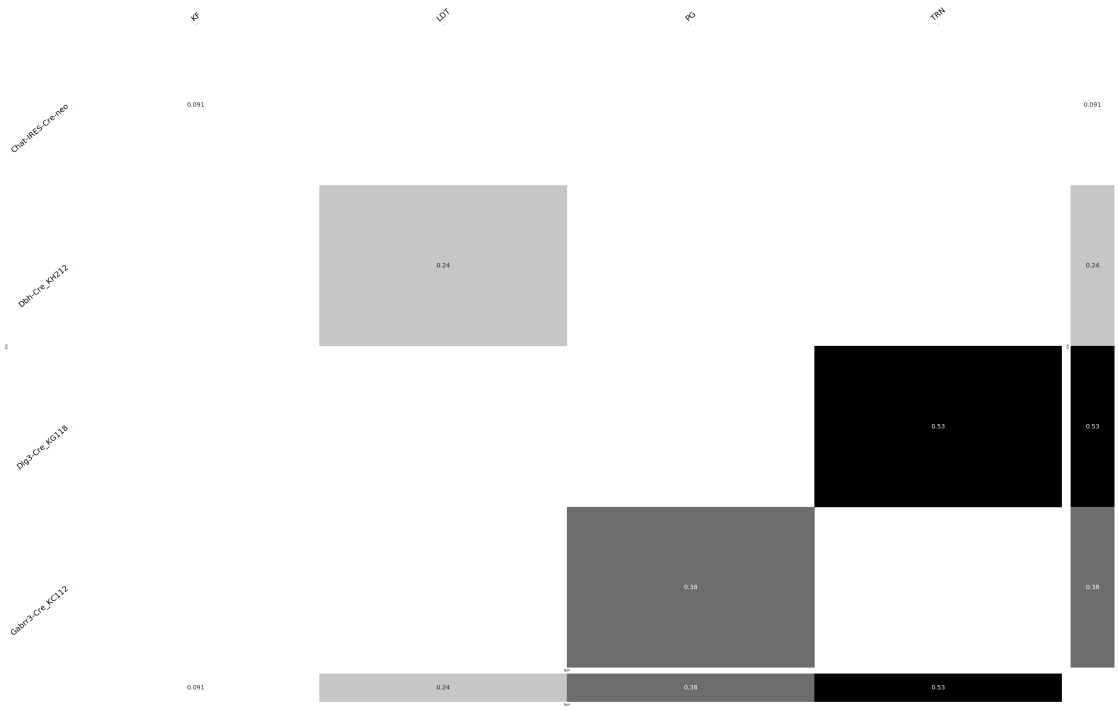

KF
LDT
PG
TRN
Chat-IRES-Cre-neo
Dbh-Cre_KH212
Dlg3-Cre_KG118
Gabrr3-Cre_KC112
0.091
0.24
0.53
0.38
0.091
0.24
0.53
0.38
0.091
0.24
0.38
0.53

Figure 18

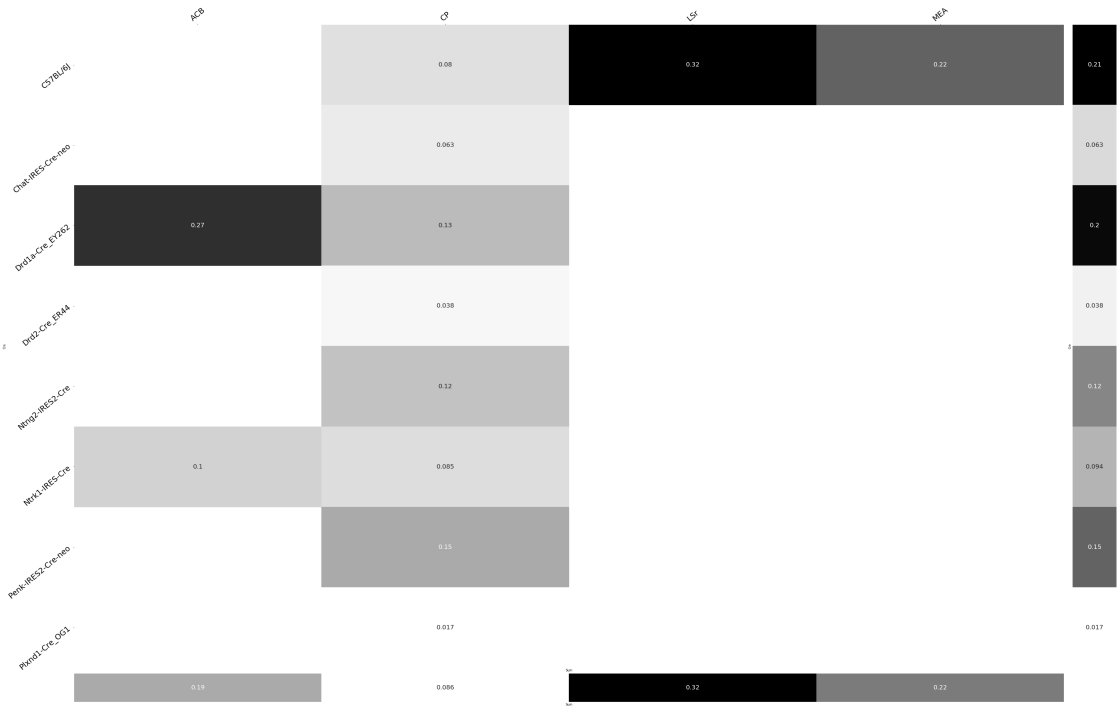

ACB
CP
LSr
MEA
C57BL/6J
Chat-IRES-Cre-neo
Drd1a-Cre_EY262
Drd2-Cre_ER44
Ntng2-IRES2-Cre
Ntrk1-IRES-Cre
Penk-IRES2-Cre-neo
Pdyn[1-Cre_OG1
0.08
0.32
0.22
0.21
0.063
0.063
0.27
0.13
0.2
0.038
0.038
0.12
0.12
0.1
0.085
0.094
0.15
0.15
0.017
0.017
0.19
0.086
0.32
0.22

Figure 19

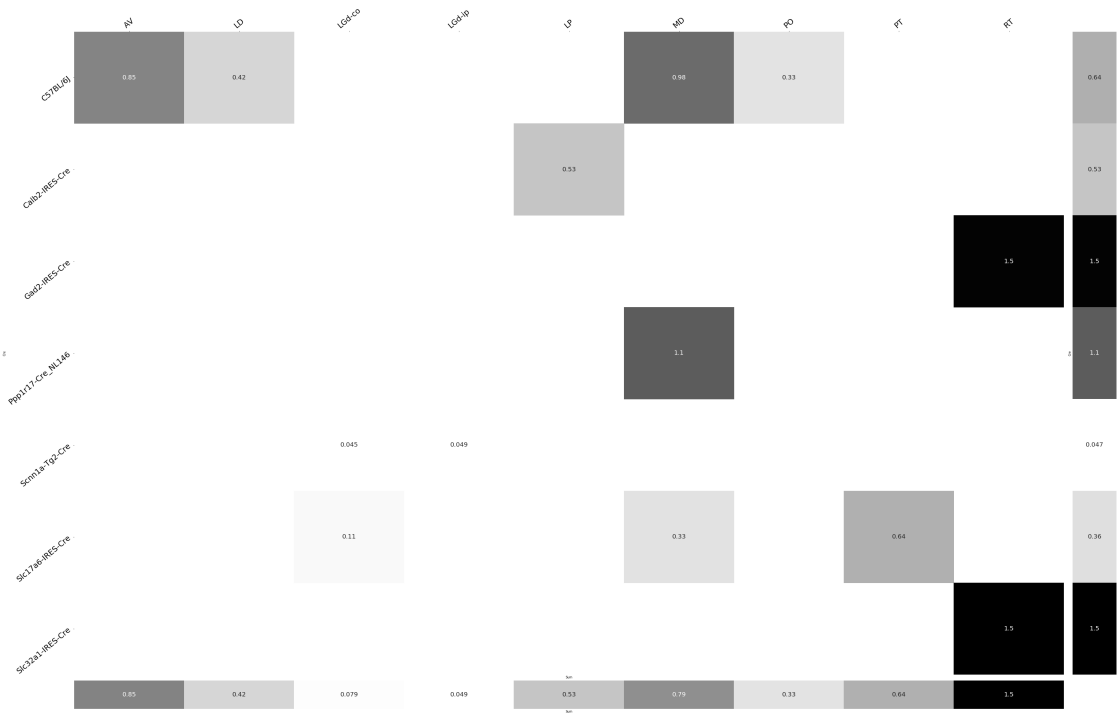

AV
LD
LGd-co
LGd-ip
LP
MD
PO
PT
RT
C57BL6J
Calb2-IRES-Cre
Gad2-IRES-Cre
Ppp1r17-Cre_NL146
Scnn1a-Tg2-Cre
Slc17a6-IRES-Cre
Slc32a1-IRES-Cre
0.85
0.42
0.98
0.33
0.64
0.53
0.53
1.5
1.5
1.1
1.1
0.045
0.049
0.047
0.11
0.33
0.64
0.36
1.5
1.5
0.85
0.42
0.079
0.049
0.53
0.79
0.33
0.64
1.5

Figure 20

Matrix Factorization

We give additional results on the generation of the archetypal connectome patterns. These consist of cross-validation selection of q, the number of latent components, stability analysis, and visualization of the reconstructed wild-type connectivity.

Cross-validation We set $\alpha = 0.002$ and run Program 2 on $\mathscr{C}_{wt}$. We use a random mask with $p = .3$ to evaluate prediction accuracy of models trained on the unmasked data on the masked data. To account for stochasticity in the NMF algorithm, we run $R = 8$ replicates at each potential dimension q. This selects $\hat{q} = 60$. (SK's comment:**Can run longer experient to show larger elbow. Note that training error also increases at high q due to difficulty training model**).

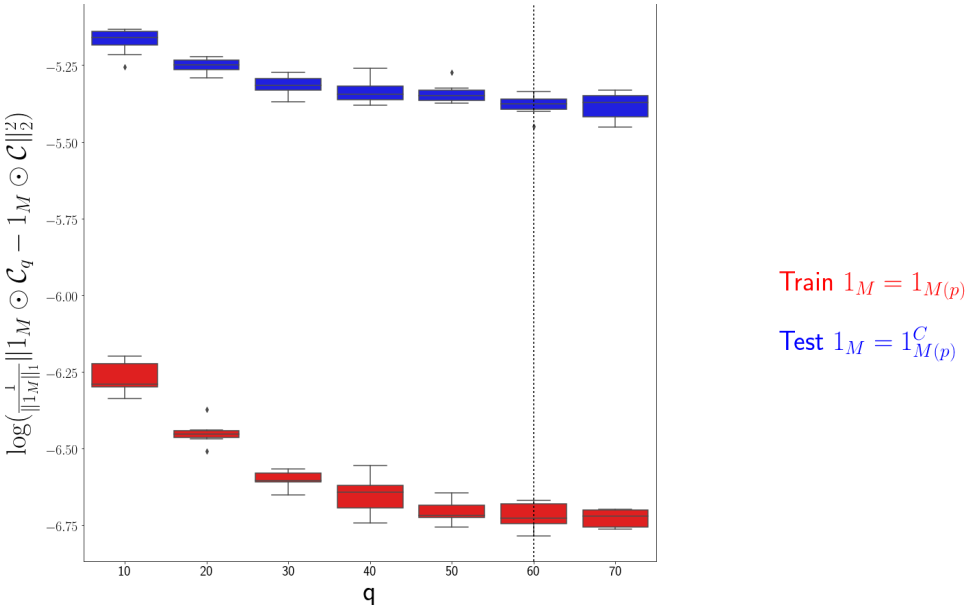


Figure 21: Train and test error using NMF decomposition.

Stability For the purposes of visualization and interpretability, we restrict to a $q = 15$ component model. To address the instability of the NMF algorithm in identifying components, we $k-means$ cluster components over $R = 10$ replicates with $k \in \{10, 15, 20, 25, 30\}$. Since the clustering is itself unstable, we repeat the clustering 25 times and select the k with the largest Rand index.

	0	1	2	3	4
q	10	15	20	25	30
Rand index	0.685081	0.789262	0.921578	**0.94548**	0.914799

Since k-means is most stable at $k = 25$, we cluster the $qR = 150$ components into 25 clusters and select the 15 clusters appearing in the most replicates.

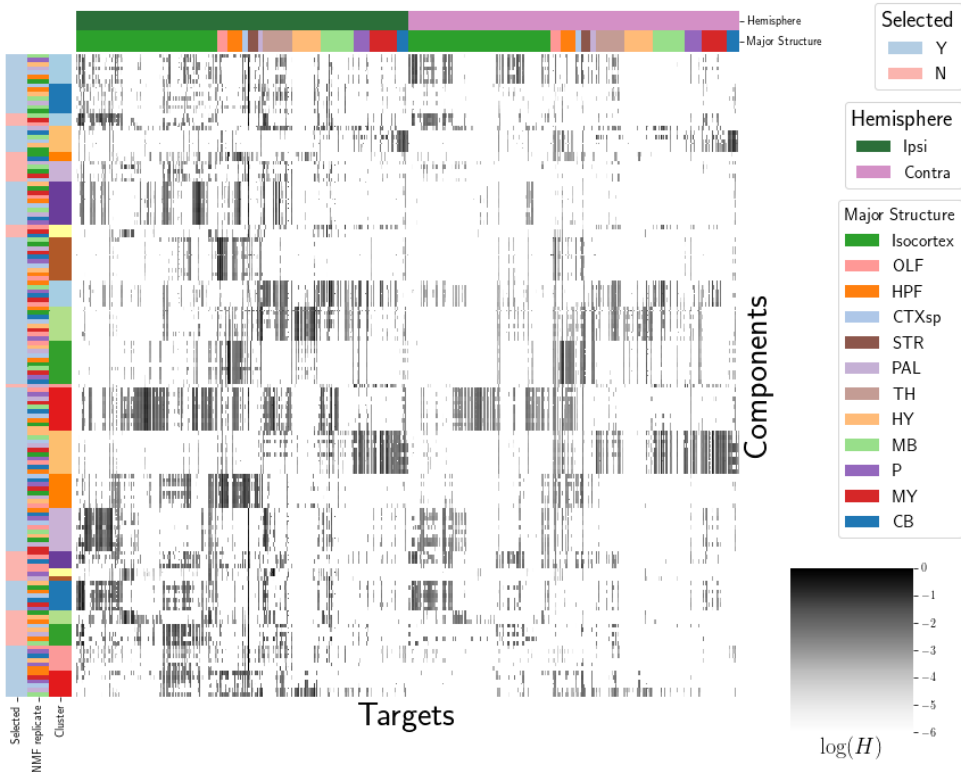


Figure 22: Stability of NMF results across replicates. Replicate and NMF component are shown on rows. Components that are in the top 15 are also indicated.

436 These are the components whose medians are plotted in Figure 4a.

Reconstructed connectivity from archetypes As a simple heuristic validation of our archetypes, we plot the reconstructed wild-type connectivity.

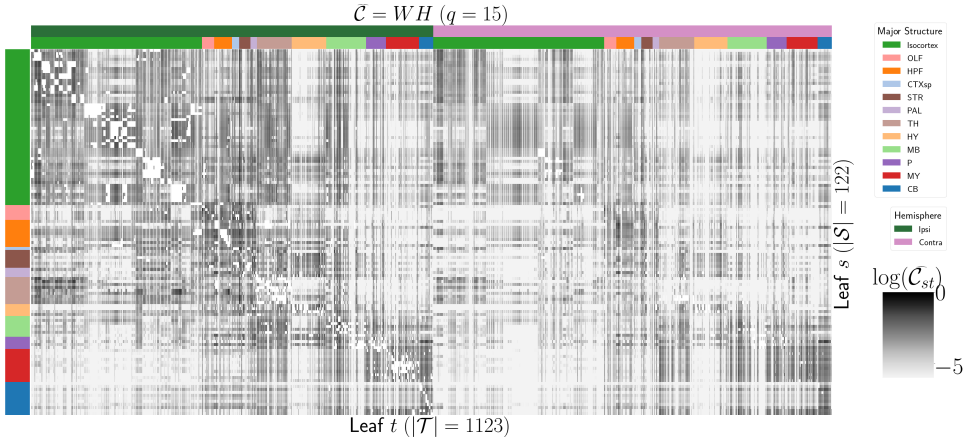


Figure 23: Reconstruction of $\mathscr{C}$ from H and W with $q = 15$ in Figure **??**.

Symbol	Meaning
q	Number of components of latent space
$\mathscr{S}$	Set of source structures
$\mathscr{T}$	Set of target structures
S	$\lvert\mathscr{S}\rvert$
T	$\lvert\mathscr{T}\rvert$
$\mathscr{C}$	Connectivity
R	Number of replicates
r	A replicate index
$\mathscr{R}$	Set of regions

8 GLOSSARY OF SYMBOLS

9 COMPETING INTERESTS

This is an optional section. If you declared a conflict of interest when you submitted your manuscript, please use this space to provide details about this conflict.

441

442 # REFERENCES

443

444 Brunet, J.-P., Tamayo, P., Golub, T. R., & Mesirov, J. P. (2004). Metagenes and molecular pattern discovery using matrix
445 factorization. *Proc. Natl. Acad. Sci. U. S. A.*, *101*(12), 4164–4169.

446 Chamberlin, N. L., Du, B., de Lacalle, S., & Saper, C. B. (1998). Recombinant adeno-associated virus vector: use for
447 transgene expression and anterograde tract tracing in the CNS. *Brain Res.*, *793*(1-2), 169–175.

448 Daigle, T. L., Madisen, L., Hage, T. A., Valley, M. T., Knoblich, U., Larsen, R. S., … Zeng, H. (2018). A suite of transgenic
449 driver and reporter mouse lines with enhanced Brain-Cell-Type targeting and functionality. *Cell*, *174*(2), 465–480.e22.

450 Gao, Y., Zhang, X., Wang, S., & Zou, G. (2016). Model averaging based on leave-subject-out cross-validation. *J. Econom.*,
451 *192*(1), 139–151.

452 Harris, J. A., Mihalas, S., Hirokawa, K. E., Whitesell, J. D., Choi, H., Bernard, A., … Zeng, H. (2019). Hierarchical
453 organization of cortical and thalamic connectivity. *Nature*, *575*(7781), 195–202.

454 Harris, J. A., Oh, S. W., & Zeng, H. (2012). Adeno-associated viral vectors for anterograde axonal tracing with fluorescent
455 proteins in nontransgenic and cre driver mice. *Curr. Protoc. Neurosci.*, *Chapter 1*, Unit 1.20.1–18.

456 Harris, K. D., Mihalas, S., & Shea-Brown, E. (2016). Nonnegative spline regression of incomplete tracing data reveals high
457 resolution neural connectivity.

458 Jeong, M., Kim, Y., Kim, J., Ferrante, D. D., Mitra, P. P., Osten, P., & Kim, D. (2016). Comparative three-dimensional
459 connectome map of motor cortical projections in the mouse brain. *Sci. Rep.*, *6*, 20072.

460 Knox, J. E., Harris, K. D., Graddis, N., Whitesell, J. D., Zeng, H., Harris, J. A., … Mihalas, S. (2019). High-resolution
461 data-driven model of the mouse connectome. *Netw Neurosci*, *3*(1), 217–236.

462 Kotliar, D., Veres, A., Nagy, M. A., Tabrizi, S., Hodis, E., Melton, D. A., & Sabeti, P. C. (2019). Identifying gene expression
463 programs of cell-type identity and cellular activity with single-cell RNA-Seq. *Elife*, *8*.

464 Lotfollahi, M., Naghipourfar, M., Theis, F. J., & Alexander Wolf, F. (2019). Conditional out-of-sample generation for
465 unpaired data using trVAE.

Oh, S. W., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., … Zeng, H. (2014). A mesoscale connectome of the mouse brain. *Nature*, *508*(7495), 207–214.

Saul, L. K., & Roweis, S. T. (2003). Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *J. Mach. Learn. Res.*, *4*(Jun), 119–155.

von Luxburg, U. (2010a). Clustering stability: An overview.

von Luxburg, U. (2010b). Clustering stability: An overview.

Wu, S., Joseph, A., Hammonds, A. S., Celniker, S. E., Yu, B., & Frise, E. (2016). Stability-driven nonnegative matrix factorization to interpret spatial gene expression and build local gene networks. *Proc. Natl. Acad. Sci. U. S. A.*, *113*(16), 4290–4295.

10 TECHNICAL TERMS

Technical Term a key term that is mentioned in an NETN article and whose usage and definition may not be familiar across the broad readership of the journal.

Cre-line Refers to the combination of cre-recombinase expression in transgenic mouse and cre-induced promotion in the vector that induces labelling of cell-class specific projection.

Cell class The projecting neurons targeted by a particular cre-line

structural connectivities connectivity between structures

Voxel A $100\mu m$ cube of brain.

structural connection tensor Connectivities between structures given a neuron class

dictionary-learning A family of algorithms for finding low-dimensional data representations.

shape constrained estimator A statistical estimator that fits a function of a particular shape (e.g. monotonic increasing, convex).

Nadaraya-Watson A simple smoothing estimator.

connectivity archetypes Typical connectivity patterns

Expected Loss Our new estimator that weights different features by their estimated predictive power.