

¹ RESEARCH

² **Modelling the cell-type specific mesoscale murine connectome with
3 anterograde tracing experiments**

⁴ **Samson Koelle^{1,2}, Jennifer Whitesell¹, Karla Hirokawa¹, Hongkui Zeng¹, Marina Meila², Julie Harris¹, Stefan Mihalas¹**

⁵ ¹Allen Institute for Brain Science, Seattle, WA, USA

⁶ ²Department of Statistics, University of Washington, Seattle, WA, USA

⁷ **Keywords:** [a series of capitalized words, separated with commas]

ABSTRACT

⁸ The Allen Brain Atlas contains of thousands of anterograde tracing experiments targeting diverse
⁹ structures and classes of projecting neurons. This paper describes the conversion of these
¹⁰ experiments into class-specific connectivity matrices representing the connection between source
¹¹ and target structures. We introduce and validate a novel statistical model for creation of connectivity
¹² matrices that combines spatial and categorical smoothing to share information between similar
¹³ neuron classes. We then show that our connectivities display expected cell-type and structure specific
¹⁴ connectivities, and factor the wild-type connectivity matrix to uncover the underlying latent structure.

AUTHOR SUMMARY

1 INTRODUCTION

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

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

29 Several statistical models for the conversion of such experiment-specific signals into estimates of
 30 connectivity strength have been proposed (K. D. Harris, Mihalas, & Shea-Brown, 2016; Knox et al.,
 31 2019; Oh et al., 2014). Of these, Oh et al. (2014) and Knox et al. (2019) model **structural connectivities**
 32 between structures. Intuitively, these models provide some improvement over simply averaging the
 33 projection signals of injections in a given region. However, these works model connectivities observed
 34 in wild-type mice transduced with constitutive promoters, and so are poorly suited for extension to
 35 recently developed tracing experiments that induce cell-type specific fluorescence (J. A. Harris et al.,
 36 2019). In particular, GFP promotion is induced by Cre-recombinase expression in cell-types specified
 37 by transgenic strain. Thus, this paper introduces a **cell class**-specific statistical model to deal with the
 38 diverse set of **cre-lines** described in J. A. Harris et al. (2019).

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

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

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

56 Section 2 gives information on the data and statistical methodology, and Section 3 presents our
57 results. These include connectivities, assessments of model fit, and subsequent analyses. Additional
58 information on our dataset, methods, and results are given in Supplemental Sections 5, 6, and 7,
59 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.

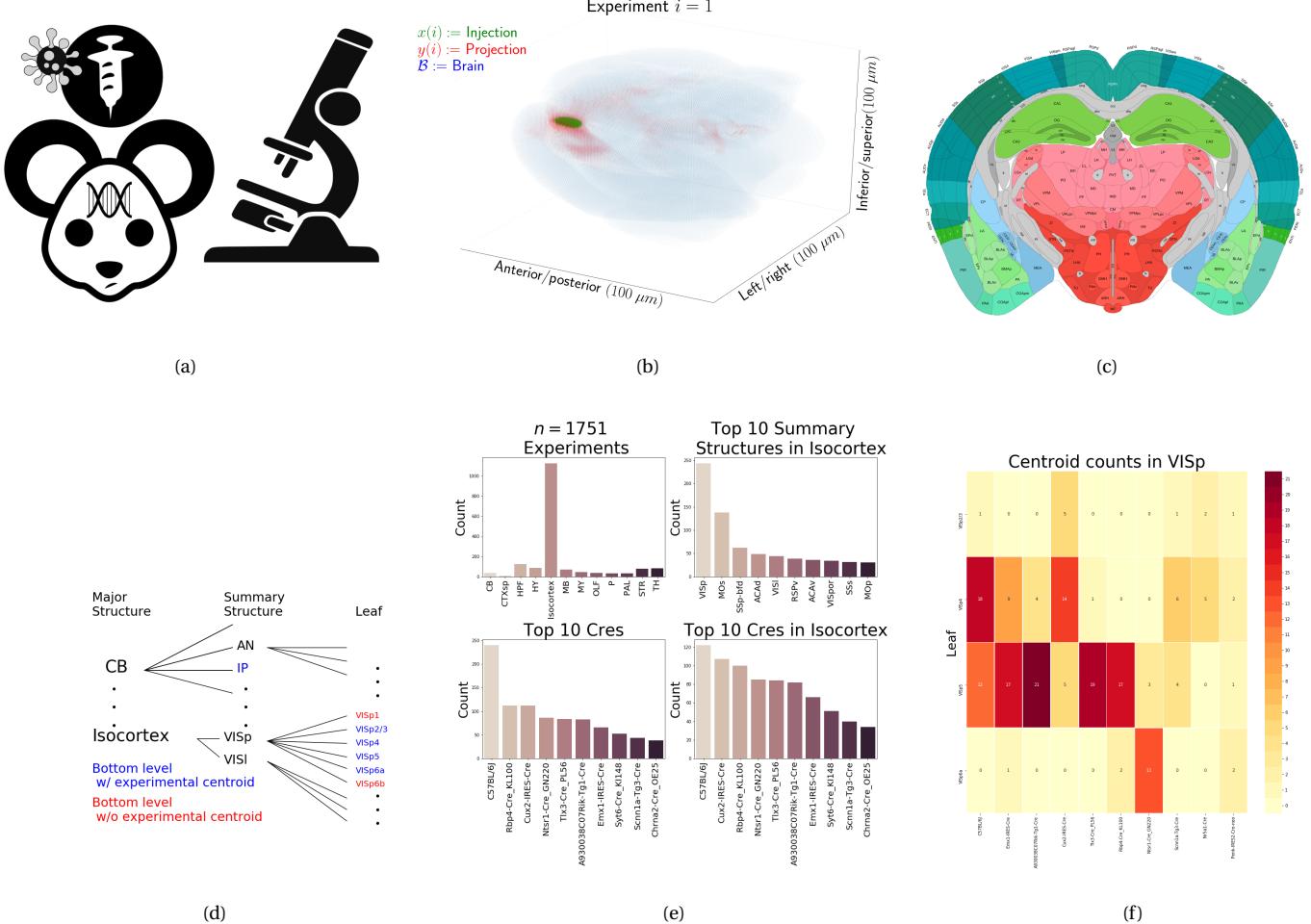


Figure 1: Experimental setting. 1a For each experiment, a potentially Cre-recombinase promoted GFP-expressing transgene cassette is transduced after stereotaxic injection into a Cre-driver mouse, followed by two-photon tomography imaging. 1b An example of the segmentation of projection and injection 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 Coccurrence of layer-specific centroids and Cre-line within VISp

66 Mice

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

69 Data

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

80 After injection, the resultant fluorescent signal is imaged, and aligned into the Allen Common
 81 Coordinate Framework (CCF) v3, a three-dimensional idealized model of the brain that is consistent
 82 between animals. This imaging and alignment procedure (described in detail in (J. A. Harris et al.,
 83 2019)) records fluorescent intensity discretized at the $100 \mu\text{m}$ voxel level. Given an experiment, this
 84 image is histologically segmented into *injection* and *projection* areas corresponding to areas of
 85 transduction and transduction/transfection, respectively (SK's comment:check). An example for a
 86 single experiment is given in Figure 1b.

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

⁹⁴ interested in particular combinations, or interested in the amount of data used to generate a
⁹⁵ particular connectivity estimate, we present this information about all experiments in Supplemental
⁹⁶ Section 5.

⁹⁷ At an essential level, cell-class specific neural connectivity is a function $f : \mathcal{V} \times \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}_{\geq 0}$ giving
⁹⁸ the directed connection of a particular cell class from a one position in the brain to another. However,
⁹⁹ what we will actually estimate are structural connectivities defined with respect to a set of source
¹⁰⁰ regions $\mathcal{S} := \{s\}$, target regions $\mathcal{T} := \{t\}$, and cell classes $\mathcal{V} := \{v\}$. Thus, we preprocess this data in
¹⁰¹ several ways. We discretize florescent signals like injections and projections into $100\mu\text{m}^3$ **voxels**.
¹⁰² Given an experiment i , we represent injections and projections as maps $x(i), y(i) : \mathcal{B} \rightarrow \mathbb{R}_{\geq 0}$, where
¹⁰³ $\subset [1 : 132] \times [1 : 80] \times [1 : 104]$ corresponds to the subset of the $(1.32 \times 0.8 \times 1.04)$ cm rectangular space
¹⁰⁴ occupied by the standard mouse brain. A structure s then contains $|s|$ voxels at locations $\{l(s_j) \in \mathbb{R}^3\}$,
¹⁰⁵ and similarly for targets. We calculate injection centroids $c(i) \in \mathbb{R}^3$ and regionalized projections
¹⁰⁶ $y_{\mathcal{T}}(i) \in \mathbb{R}^T$ giving the sum of $y(i)$ in each region. In contrast to Knox et al. (2019), we also $l1$ normalize
¹⁰⁷ these projection vectors. A detailed mathematical description of these steps, including data quality
¹⁰⁸ control, is given in Supplemental Section 6.

109 ***Modeling Structural Connectivity***

We define

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

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

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

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

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

$$\hat{\mathcal{C}}(v, s, t) = f^*(\hat{f}(f_*(\mathcal{D}(v, s, t))). \quad (1)$$

113 This makes explicit the data featurization f_* , statistical estimator \hat{f} , and any potential subsequent
 114 transformation f^* such as averaging over the source region, as well as the fact that different data may
 115 be used to estimate different connectivities. Table 1 reviews estimators used for this data-type used in
 116 previous work, as well as our two main extensions. Additional information on these estimators is
 117 given in Supplemental Section 6.

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

Table 1: Estimation of \mathcal{C} using connectivity data. The regionalization, estimation, and featurization steps are denoted by f^* , \hat{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_l , respectively. $D(v, s, t) = I_m$ means that the model $f^*(\hat{f}((f_*(I)))$ generates prediction for a structure in m using only data in I_m .

118 Our contributions - the Cre-NW and **Expected Loss** (EL) models - have several differences from the
 119 previous methods. In contrast to the Oh et al. (2014) non-negative least squares and Knox et al. (2019)
 120 Nadaraya-Watson estimators that take into account s and t , but not v , our new estimators specifically
 121 account for cell class. The Cre-NW estimator only uses experiments from a particular class to predict
 122 connectivity for that class, while the EL estimator shares information between classes. This estimator
 123 takes into account two types of covariate information about each experiment: the centroid of the
 124 injection, and the Cre-line. Like the NW and Cre-NW estimator, the EL estimator generates
 125 predictions for each voxel in a structure, and then sums them together to get the overall connectivity.
 126 However, in contrast to these alternative approaches, when predicting the projection pattern of a
 127 certain cell-class at a particular location, the EL estimator weights the average behavior of the class in
 128 the structure containing the location in question against the locations of the various proximal
 129 experiments. Thus, nearby experiments with similar Cre-lines can help generate the prediction, even
 130 when there are few nearby experiments of the cell-class in question. A detailed mathematical
 131 description of our new estimator is given in Appendix 6. Finally, in order to synthesize information
 132 about leafs targeted by different Cre-lines, we also generate an average connectivity matrix over all

¹³³ Cre-lines. This model is not evaluated in our testing, and is only a general heuristic for overall
¹³⁴ behavior, but provides a useful summary of results.

135 ***Model evaluation***

136 We select optimum functions from within and between our estimator classes using empirical risk
 137 minimization. Equation 1 includes a deterministic step f^* included without input by the data. The
 138 performance of $\hat{\mathcal{C}}$ is therefore determined by performance of the model $\hat{f}(f_*(\mathcal{D}(s, t, v)))$. We can then
 139 evaluate $\hat{f}(v, s, t)$ using leave-one-out cross validation, in which the accuracy of the model is assessed
 140 by its ability to predict experiments excluded from the training data.

141 Another main estimation question is what combinations of v , s , and t we actually want to generate
 142 a prediction for. Our EL and Cre-NW models are leaf specific. They only generates predictions for a
 143 cell class in a leaf when at least one experiment with a Cre-line targeting that class has a centroid in
 144 the leaf. To compare our contributions accurately with less-restrictive models such as used in Knox et
 145 al. (2019), we therefore restrict to the smallest set of evaluation experiments suggested by any
 146 of our models. The sizes of these evaluation sets are given in Supplemental Section 5.

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. We use weighted l_2 -loss to evaluate these predictions.

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

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

147 This is a somewhat different loss from Knox et al. (2019), both because of the normalization of
 148 projection, and because of the increased weighting of rarer combinations of s and v implicit in the
 149 loss. As a final modeling step, we establish a lower limit of detection. This is covered in Supplemental
 150 Section 6

151 ***Connectivity analyses***

152 We show neuronal processes underlying our estimated connectome using a variety of types of
 153 unsupervised learning. Our use of hierarchical clustering is standard, and so we do not review it here.
 154 However, our application of non-negative matrix factorization (NMF) to decompose the estimated
 155 long-range connectivity into *connectivity archetypes* that linearly combine to reproduce the observed
 156 connectivity is novel and technically of some independent interest. Non-negative matrix factorization
 157 refers to a collection of **dictionary-learning** algorithms for decomposing a positively-valued matrix
 158 such as \mathcal{C} into positively-valued matrices called, by convention, weights $W \in \mathbb{R}_{\geq 0}^{S \times q}$ and hidden units
 159 $H \in \mathbb{R}_{\geq 0}^{q \times T}$. This H is typically used to identify latent structures with interpretable biological meaning,
 160 and the choice of matrix factorization method reflects particular scientific subquestions and
 161 probabilistic interpretations.

162 Our application of NMF has several interesting characteristics. Our algorithm is

$$\text{NMF}(\mathcal{V}, \lambda, q) = \arg \min_{W, H} \frac{1}{2} \| \mathbf{1}_{d > 1500\mu m} \otimes \hat{\mathcal{C}} - WH \|_2^2 + \lambda (\| H \|_1 + \| W \|_1).$$

163 We ignore connections between source and target regions less than $1500\mu m$ apart, and set
 164 $\alpha = 0.001$ to encourage sparsity. This is because short-range projections resulting from diffusion
 165 dominate the matrices $\hat{\mathcal{C}}$, and represent a less-interesting type of biological structure. We use
 166 unsupervised cross-validation to determine q , and show the top 15 stable components. Stability
 167 analysis accounts for the difficult-to-optimize NMF optimization problem by clustering the resultant
 168 H from multiple replicates. We then determine a superset of clusters appearing frequently across
 169 NMF replicates, and select the median-vectors of the q most common clusters as **connectivity**
 170 **archetypes**. Details of these approaches are given in Appendix 6 and 7.

3 RESULTS

¹⁷¹ Our results include evaluation of model fit, the Cre-specific connectivity matrices themselves, and
¹⁷² retrospective analyses of these matrices for patterns related to cre-type and source and target regions.

¹⁷³ ***Model evaluation***

¹⁷⁴ Table ?? contains weighted losses from leave-one-out cross-validation of candidate models. Our EL
¹⁷⁵ model generally performs better than the other Nadaraya-Watson estimators that we consider. For
¹⁷⁶ example, the NW Major-WT model is the model from Knox et al. (2019). The EL model combines the
¹⁷⁷ good performance of class-specific models like NW Leaf-Cre in regions like Isocortex with the good
¹⁷⁸ performance of class-agnostic models in regions like Thalamus. Additional information on model
¹⁷⁹ evaluation, including class and structure specific performance, is given in Appendix 5 In particular,
¹⁸⁰ Supplementary Table ?? contains the sizes of these evaluation sets in each major structure, and
¹⁸¹ 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
\hat{f}	Mean	NW					EL
\mathcal{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.

182 Connectivities

183 Our main result is the estimation of matrices $\hat{\mathcal{C}}_v$ representing connections of source structures to
184 target structures for particular cre-lines v . We exhibit several characteristics of interest, and confirm
185 the detection of several well-established connectivities within our tensor. Many additional interesting
186 biological processes are visible within this matrix - more than we can report in this paper - and it is
187 our expectation that these will be identified by users of our results. The connectivity tensor and code
188 to reproduce it are available at

189 https://github.com/AllenInstitute/mouse_connectivity_models/tree/2020.

190 *Overall connectivity* The connectivity matrix for wild-type connectivities from leaf sources to
191 summary structure targets is illustrated in Figure 2. Several expected biological processes are evident.
192 For example, intraareal connectivities are clear, as are ipsilateral connections between cortex and
193 thalamus. The clear intraareal connectivities mirror previous estimates in Oh et al. (2014) and Knox et
194 al. (2019) and descriptive depictions of individual experiments in J. A. Harris et al. (2019). Compared
195 with Knox et al. (2019), our more discretized source smoothing and greater number of experiments
196 leads to a significantly more discretized connectivity matrix.

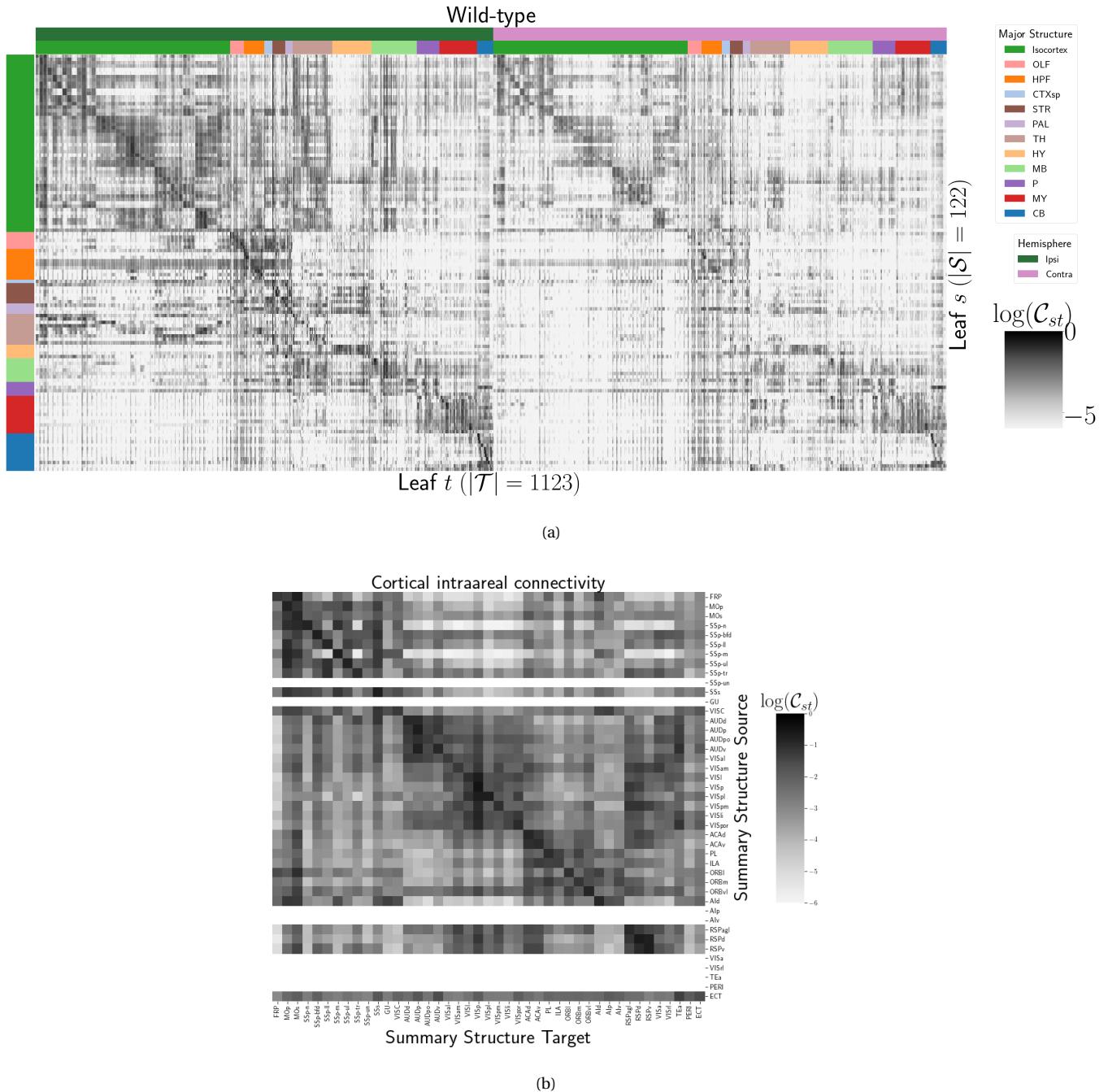


Figure 2: Wild-type connectivities. 2a Log wild-type connectivity matrix $\log \mathcal{C}(s, t, v_{wt})$. 2b Log wild-type intracortical connectivity matrix at the summary structure level.

¹⁹⁷ *Class-specific connectivities* We have generated $V = 114$ cell-class specific structural connectivities
¹⁹⁸ $\mathcal{C}(v) \in \mathbb{R}^{S \times T}$. Exhaustive comparison of this estimated behavior is prohibitive, but we do exhibit
¹⁹⁹ several examples of our class specific connectivities conforming to well-known behaviors. These
²⁰⁰ validation cases are given in Figure 3.

²⁰¹ We begin by plotting subsets of the estimated connectivities in the well-studied VISp and MO
²⁰² regions in Figure 3a. The localization of Rbp4-Cre and Ntsr1-Cre injection centroids to layers 5 and 6
²⁰³ respectively is evident (see also Supplemental Figure ??). These layers project to their expected targets
²⁰⁴ Jeong et al. (2016). In VISp, the Ntsr1-Cre line strongly targets the thalamic LP nuclei, and in MO, layer
²⁰⁵ 5 projects to anterior basolateral amygdala (BLA) and capsular central amygdala (CEA), while layer 6
²⁰⁶ does not.

²⁰⁷ Figure 3b shows a collection of connectivity strengths generated using cre-specific models for
²⁰⁸ wild-type, Cux2, Ntsr1, Rbp4, and Tlx3 cre-lines from visual signal processing leafs in the cortex to
²⁰⁹ cortical and thalamic nucleii. This shows that cell-class has a dominating effect on projection in
²¹⁰ certain regions. We use hierarchical clustering to sort source structure/cell-class combinations by the
²¹¹ similarity of their structural projections, and sort target structures by the structures from which they
²¹² receive projections. Examining the former, we can see that the Ntsr1 Cre-line distinctly projects to
²¹³ thalamic nucleii, regardless of summary structure. This contrasts with the tendency of other cell
²¹⁴ classes to project intracortically in a manner determined by the source structure. Similarly, layer 6
²¹⁵ targets are not strongly projected to by any of the displayed Cre-lines. There are too many targeted
²¹⁶ summary structures to plot here, but we expect that the source profile of each target clusters by
²¹⁷ structure.

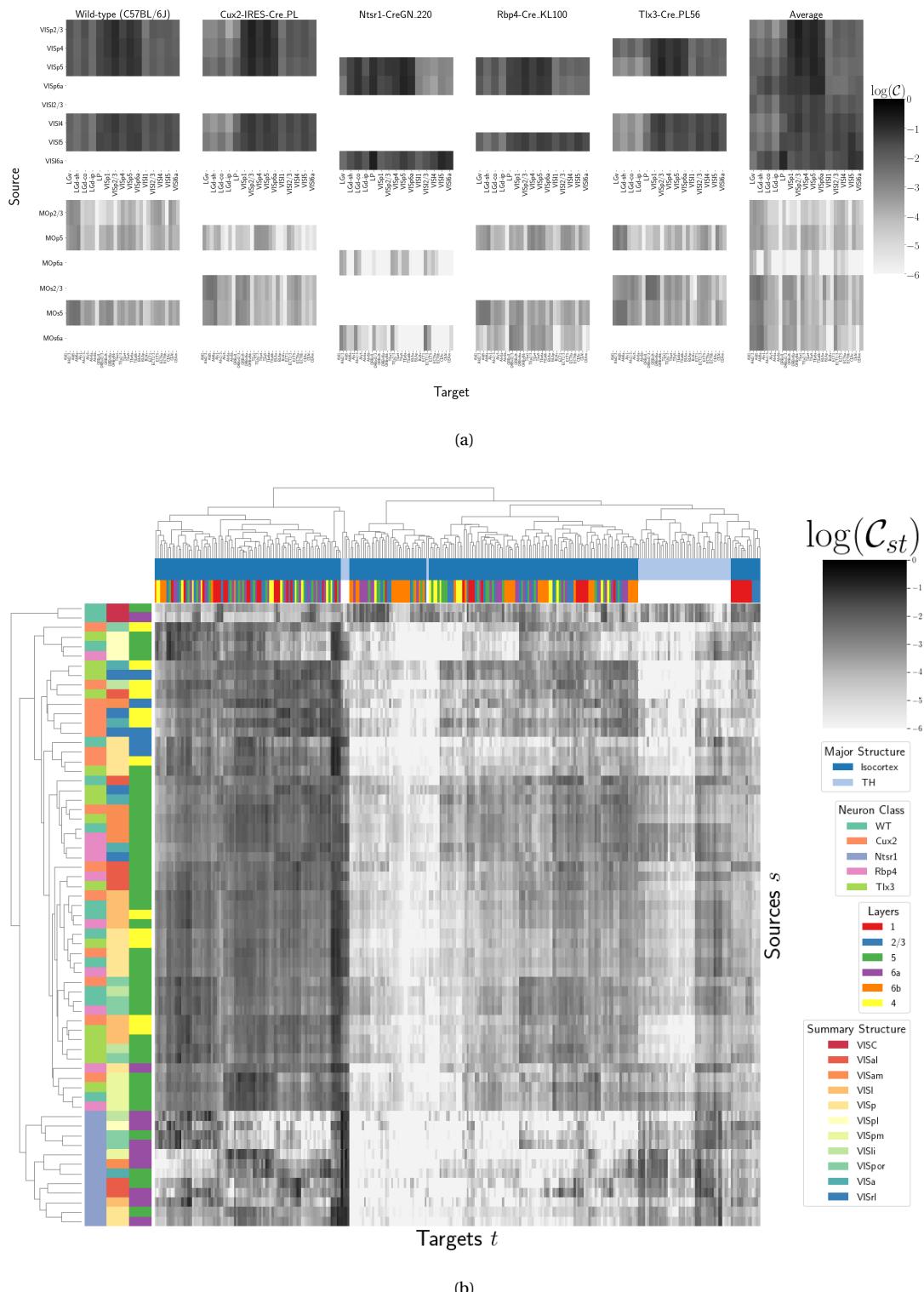


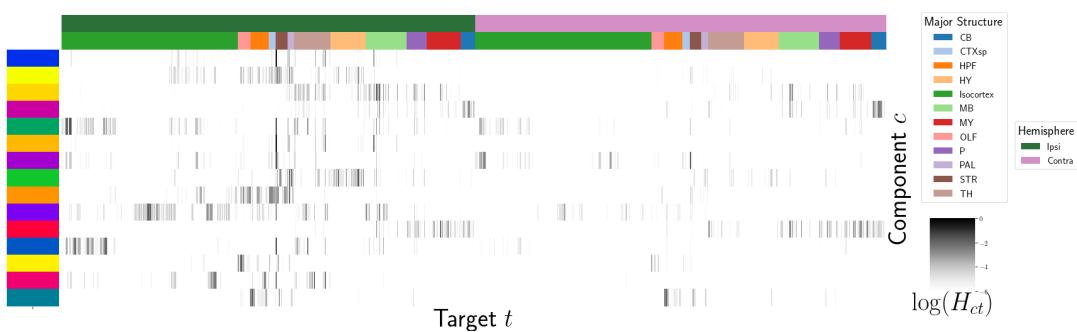
Figure 3: Cell-class and layer specific connectivities from VISp and MO. This figure shows a preselected subset of putatively interesting connectivities from VISp and MO. 3 Heirarchical clustering of connectivity strengths from visual signal processing cell-types to cortical and thalymic targets. Cre-line, 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 experiment of that cre-line in that particular leaf.

218 *Connectivity Analyses*

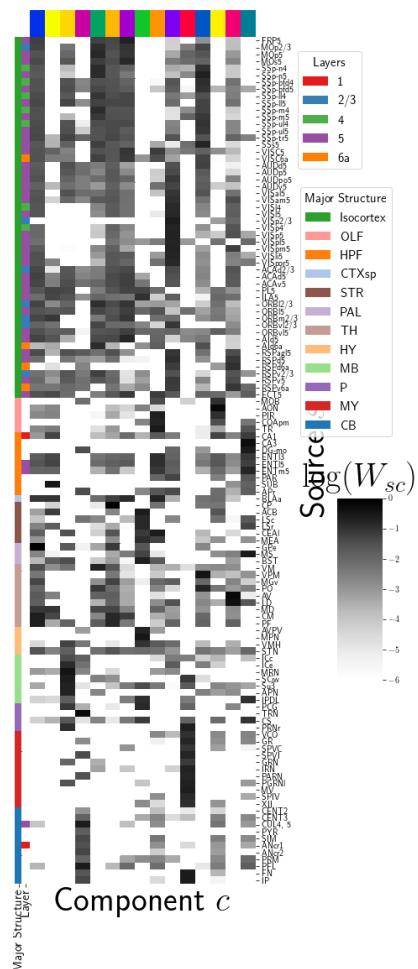
219 Each structural connectivity matrix is a high-dimensional representation of relatively few biological
220 processes. As discussed in Knox et al. (2019), one of the most basic processes underlying the observed
221 connectivity is the tendency of each source region to predominantly project to proximal regions. For
222 example, the heatmap in 7 shows intraregion distances clearly contains an overall pattern reminiscent
223 of the connectivity matrix in 2. These connections are biologically meaningful, but also unsurprising,
224 and their relative strength biases learned latent coordinate representations away from long-range
225 structures. For this reason, we establish a $1500\mu\text{m}$ 'distal' threshold within which to exclude
226 connections for our analysis.

227 Perhaps more interestingly in our setting, certain cell-types and layers have a characteristic
228 connectivity pattern. We therefore perform non-negative matrix factorization on distal wild-type
229 connectivities to estimate these characteristic patterns in a probabilistic way. This decomposes the
230 remaining censored connectivity matrix into a relatively small number of distinct signals. These
231 signals are plotted in Figure 4, and technical details and intermediate results are given in
232 Supplemental Sections 6 and 7, respectively.

233 The plotted decomposition shows that these underlying connectivity archetypes correspond
234 strongly to major brain division. However, certain components that predominantly represent
235 connectivity from a given major brain division may also be accessed from other areas. For example,
236 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 $\mathcal{C}_{wt} = WH$ for $q = 15$ components. 4a Latent space coordinates H of \mathcal{C} . Target major structure and hemisphere are plotted. 4b Loading matrix W . Source major structure and layer are plotted.

4 DISCUSSION

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

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

ACKNOWLEDGMENTS

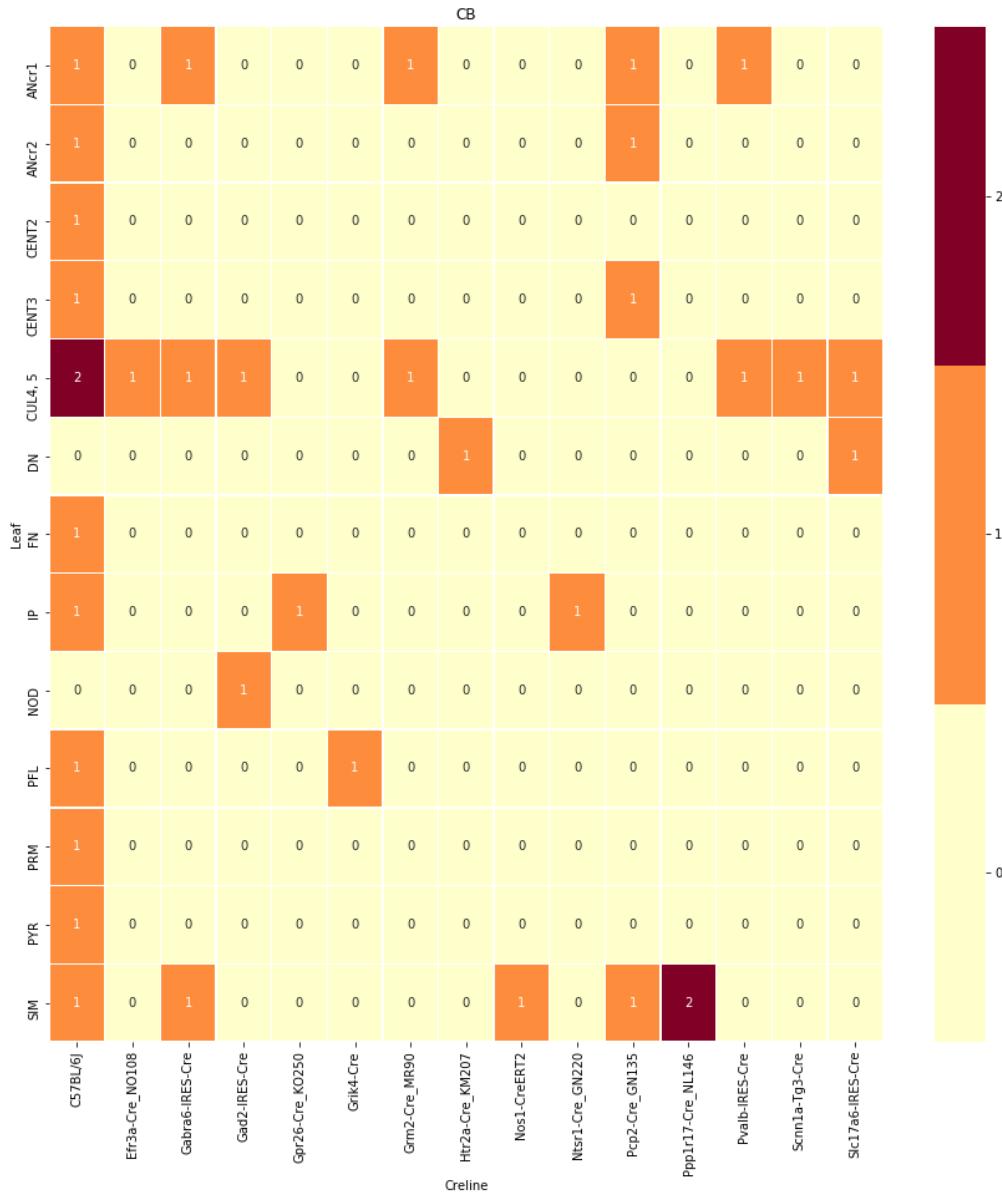
²⁵² The Funder and award ID information you input at submission will be introduced by the publisher
²⁵³ under a Funding Information head during production. Please use this space for any additional
²⁵⁴ acknowledgements and verbiage required by your funders.

5 SUPPLEMENTAL INFORMATION

255 Cre/structure combinations in \mathcal{D}

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

centroid densityoct12.png



centroid densityoct12.png

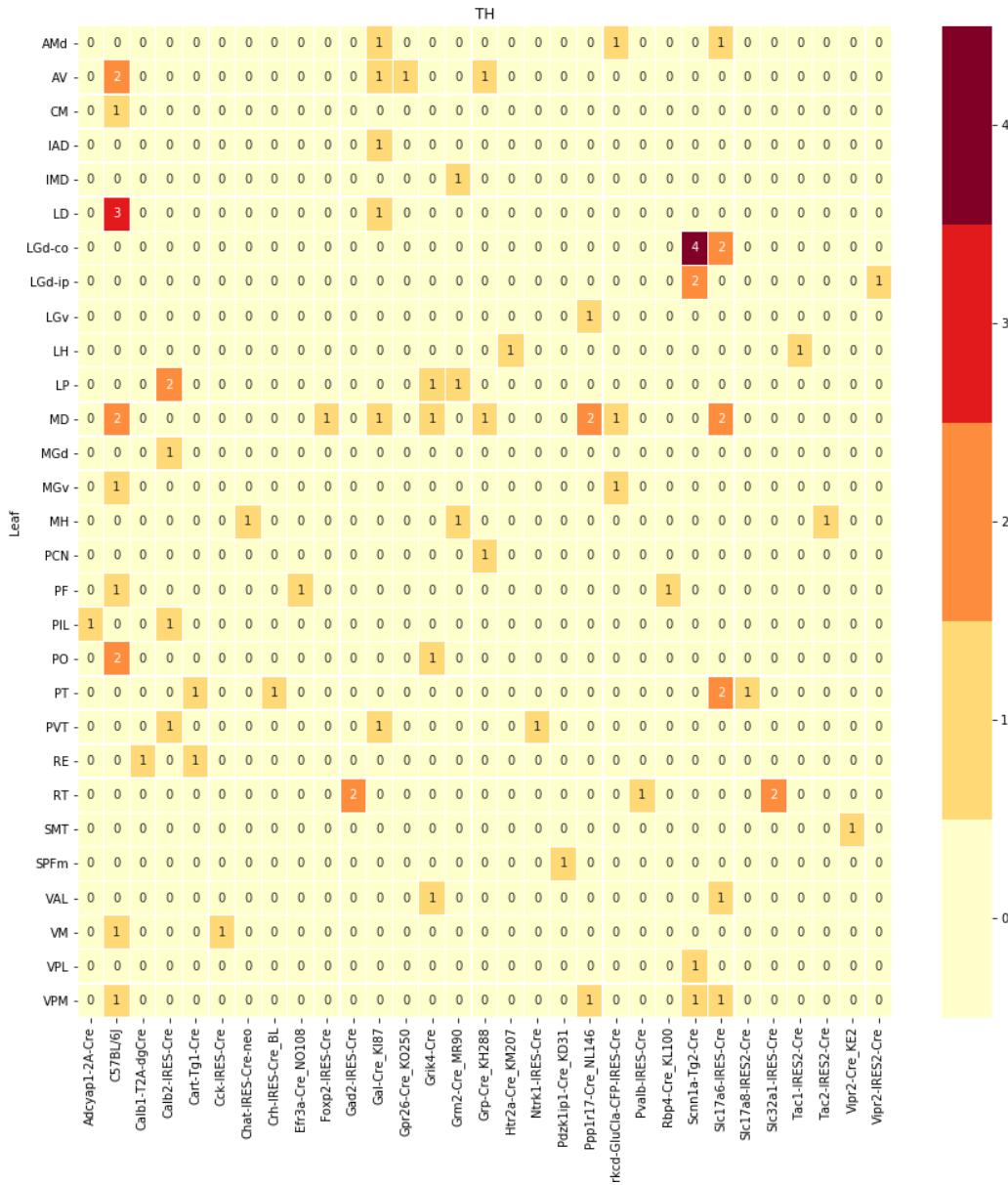


Figure 5: Caption

centroid densityoct12.png

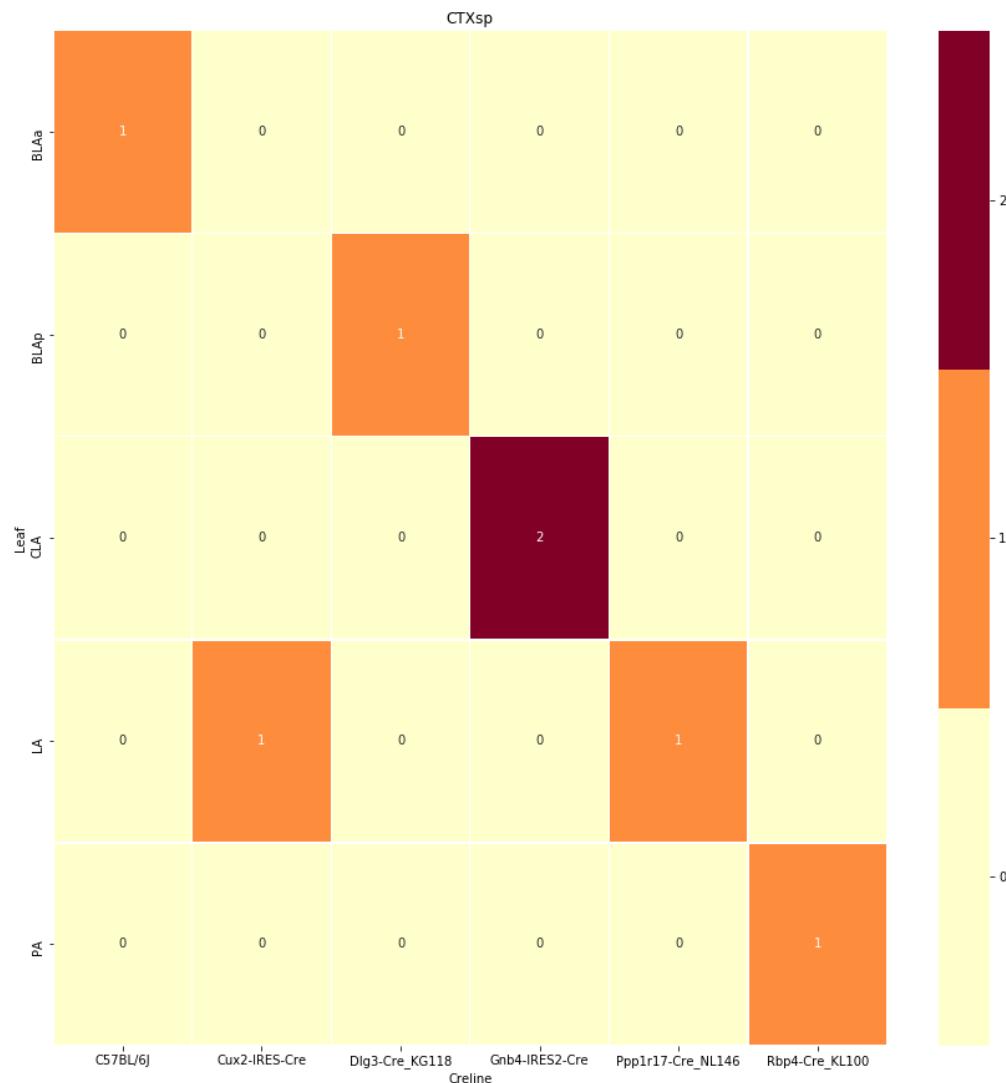
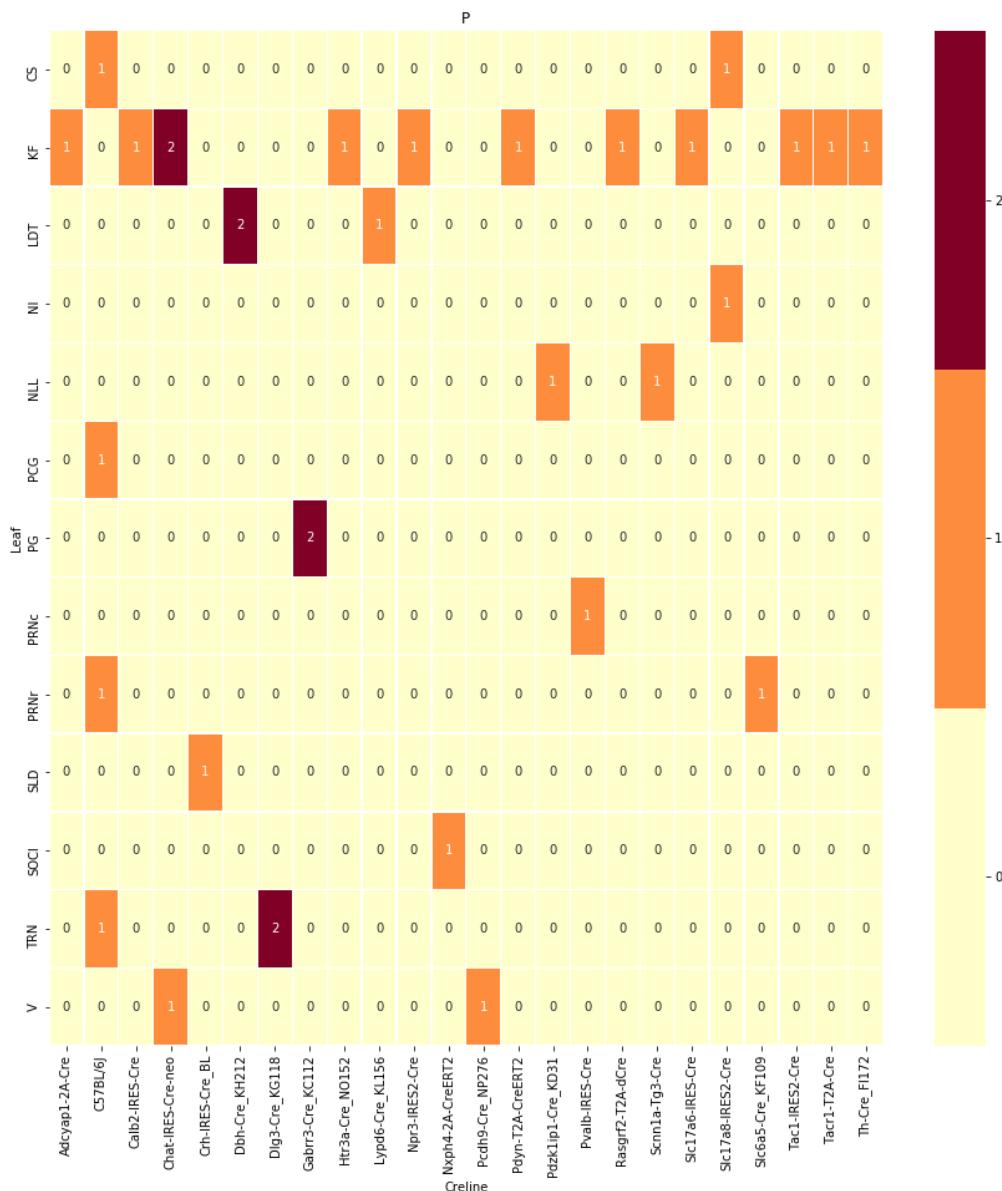
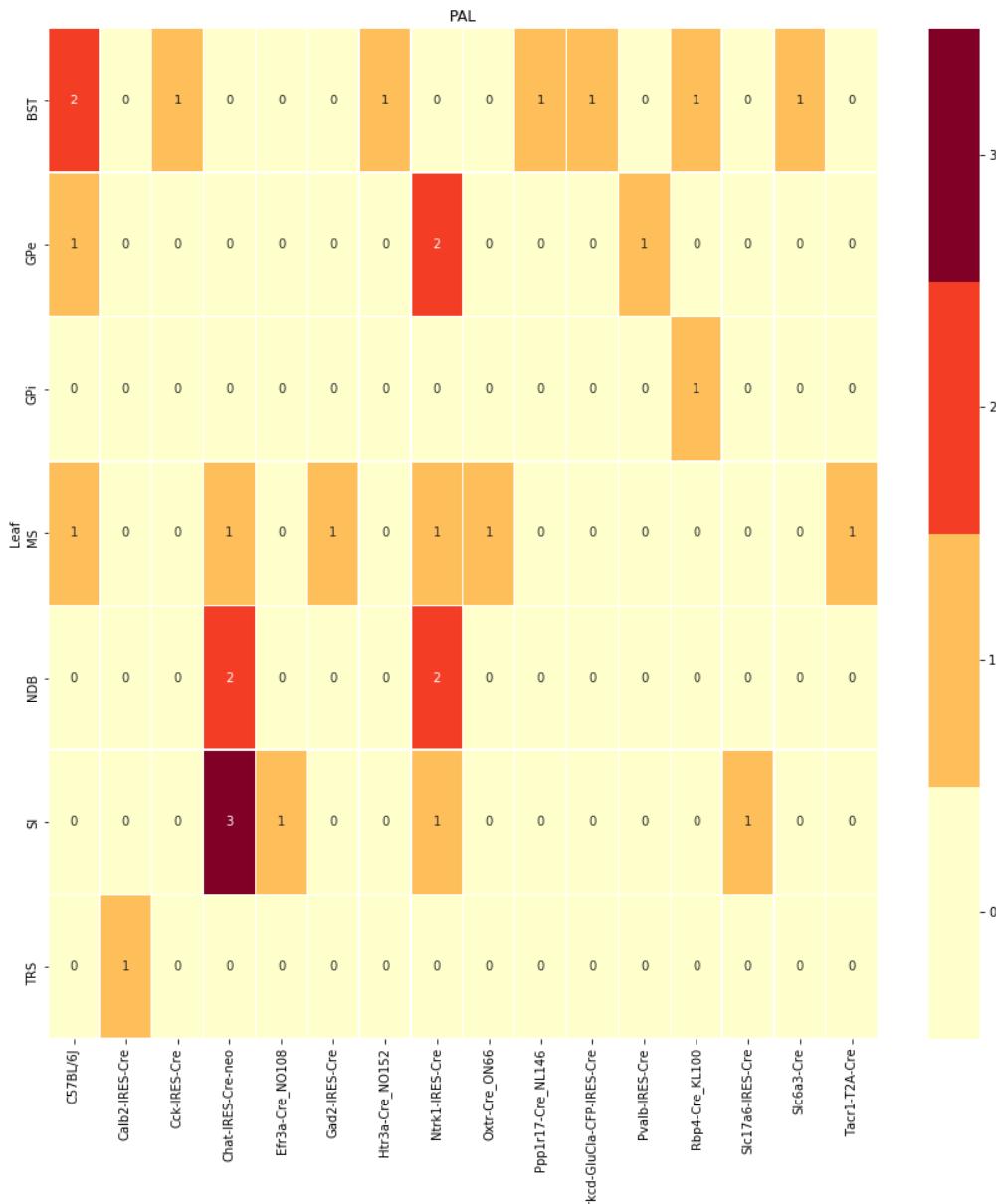


Figure 6: Caption

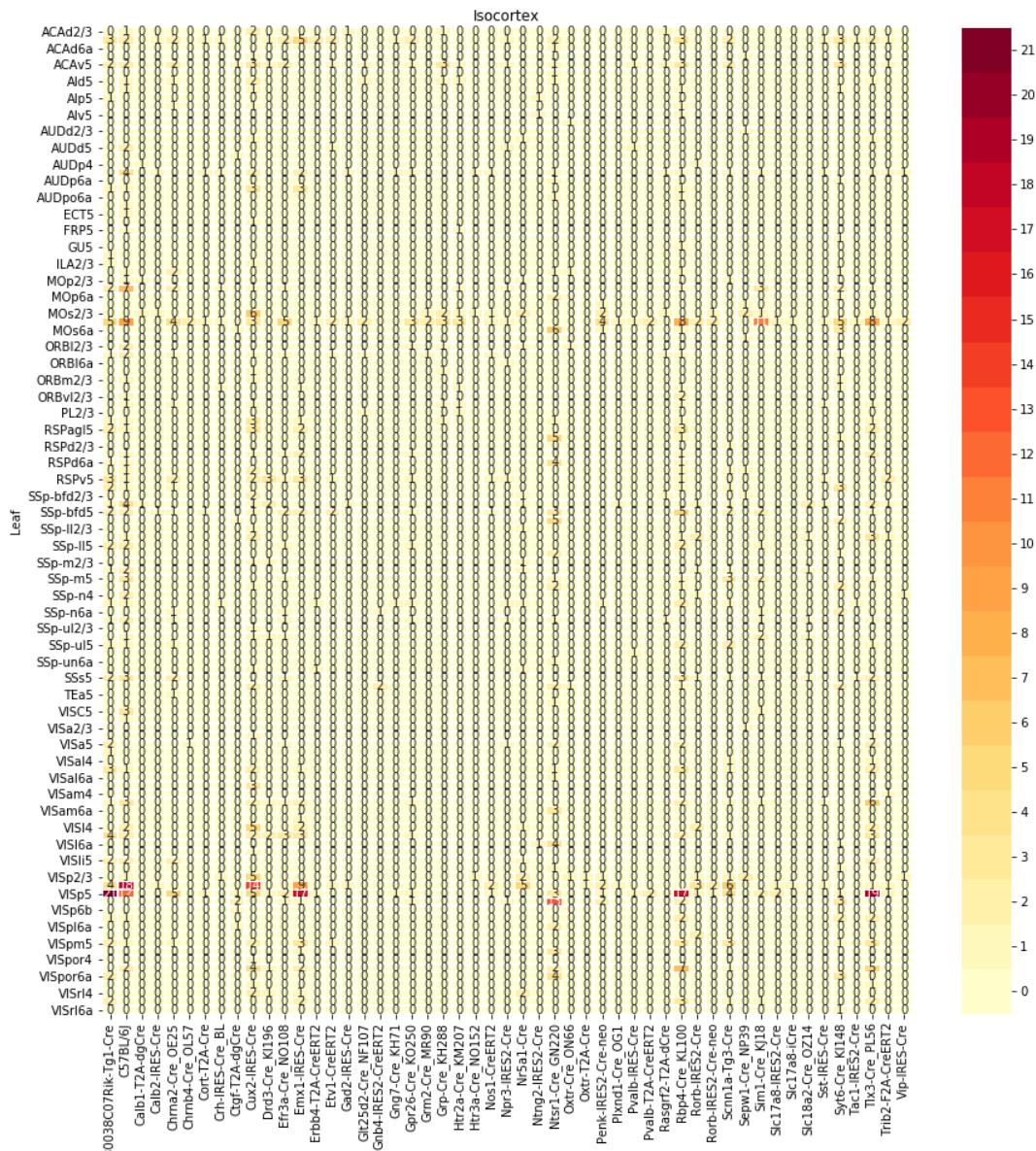
centroid densityoct12.png



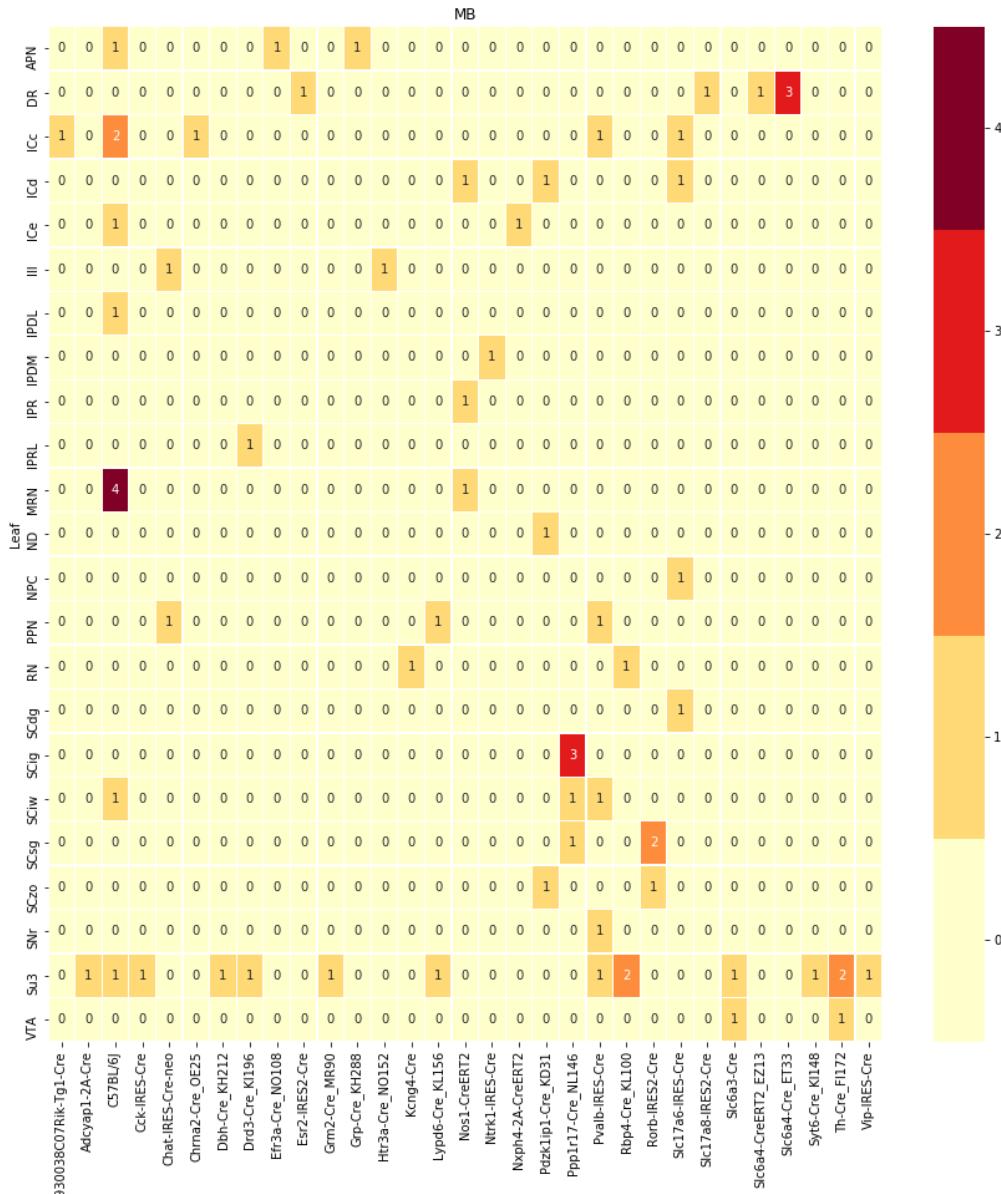
centroid densityoct12.png



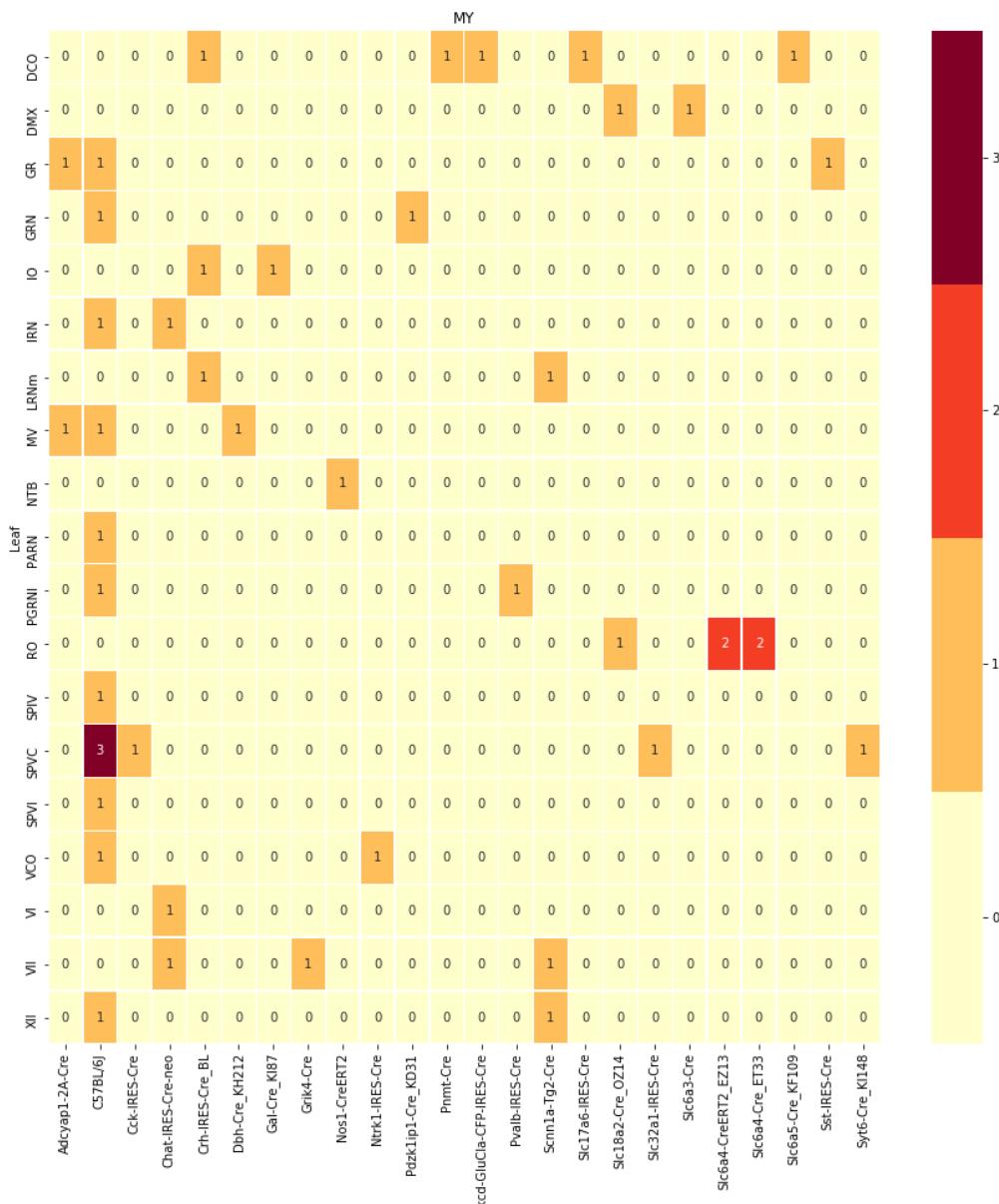
centroid densityoct12.png



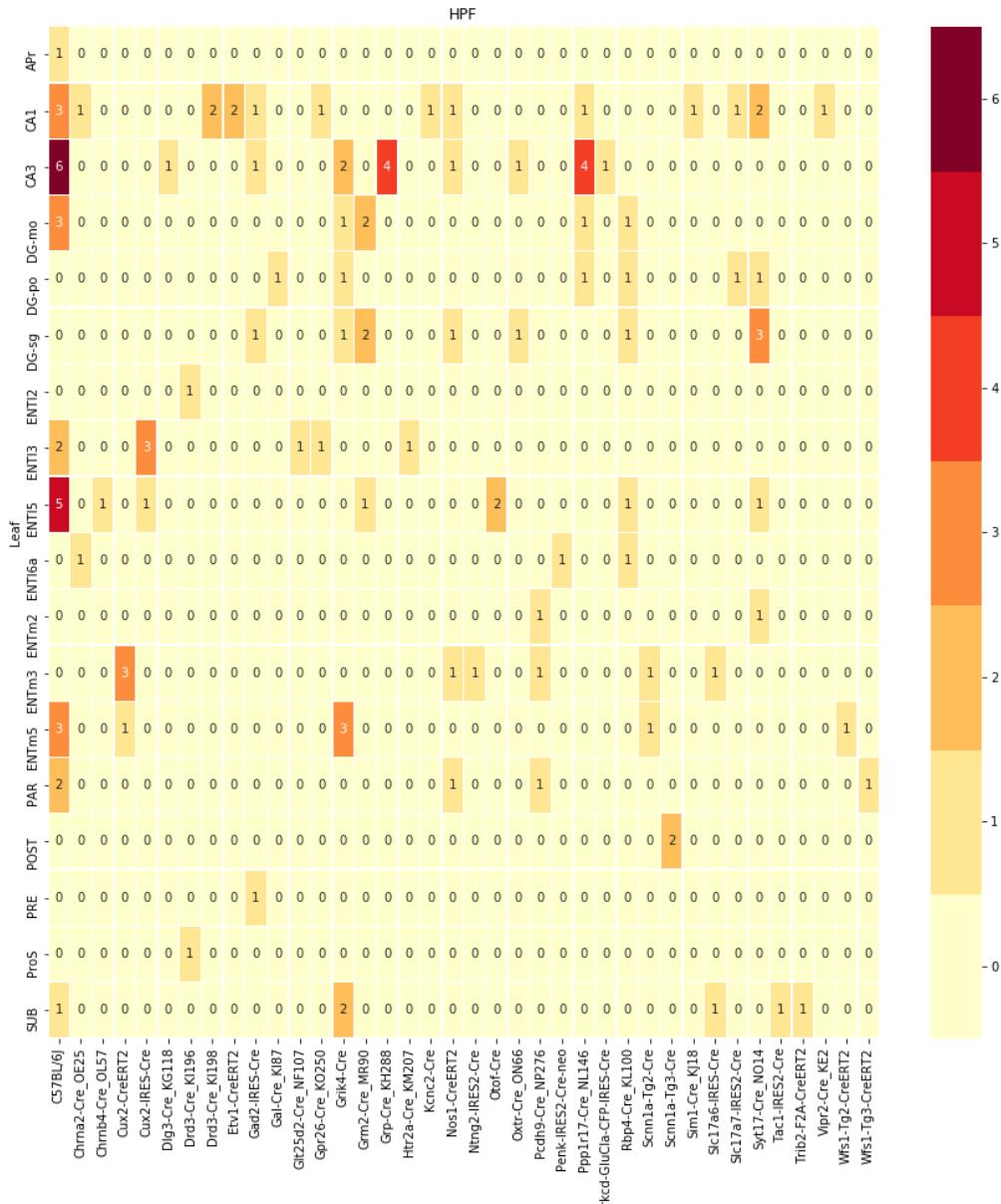
centroid densityoct12.png



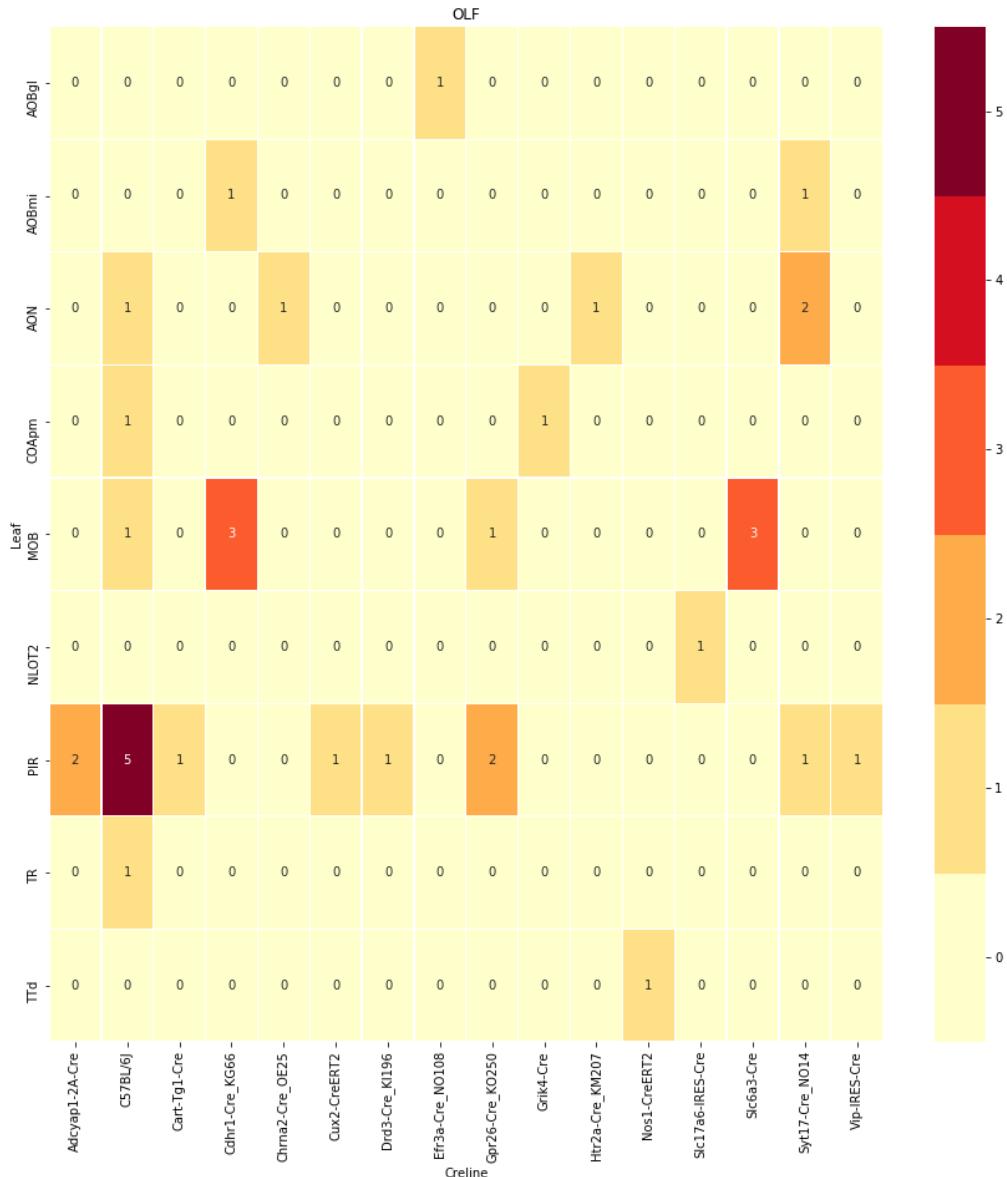
centroid densityoct12.png



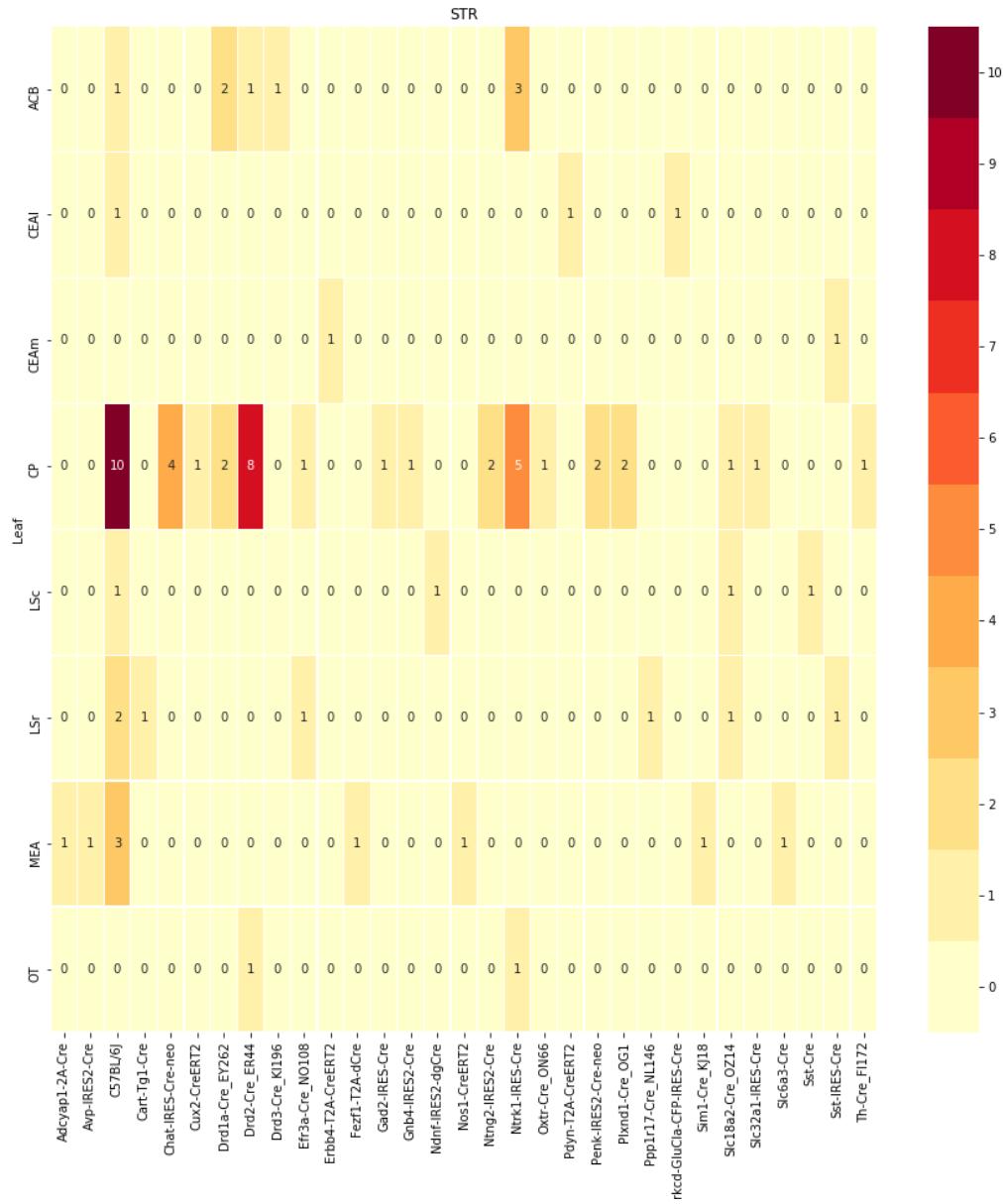
centroid densityoct12.png



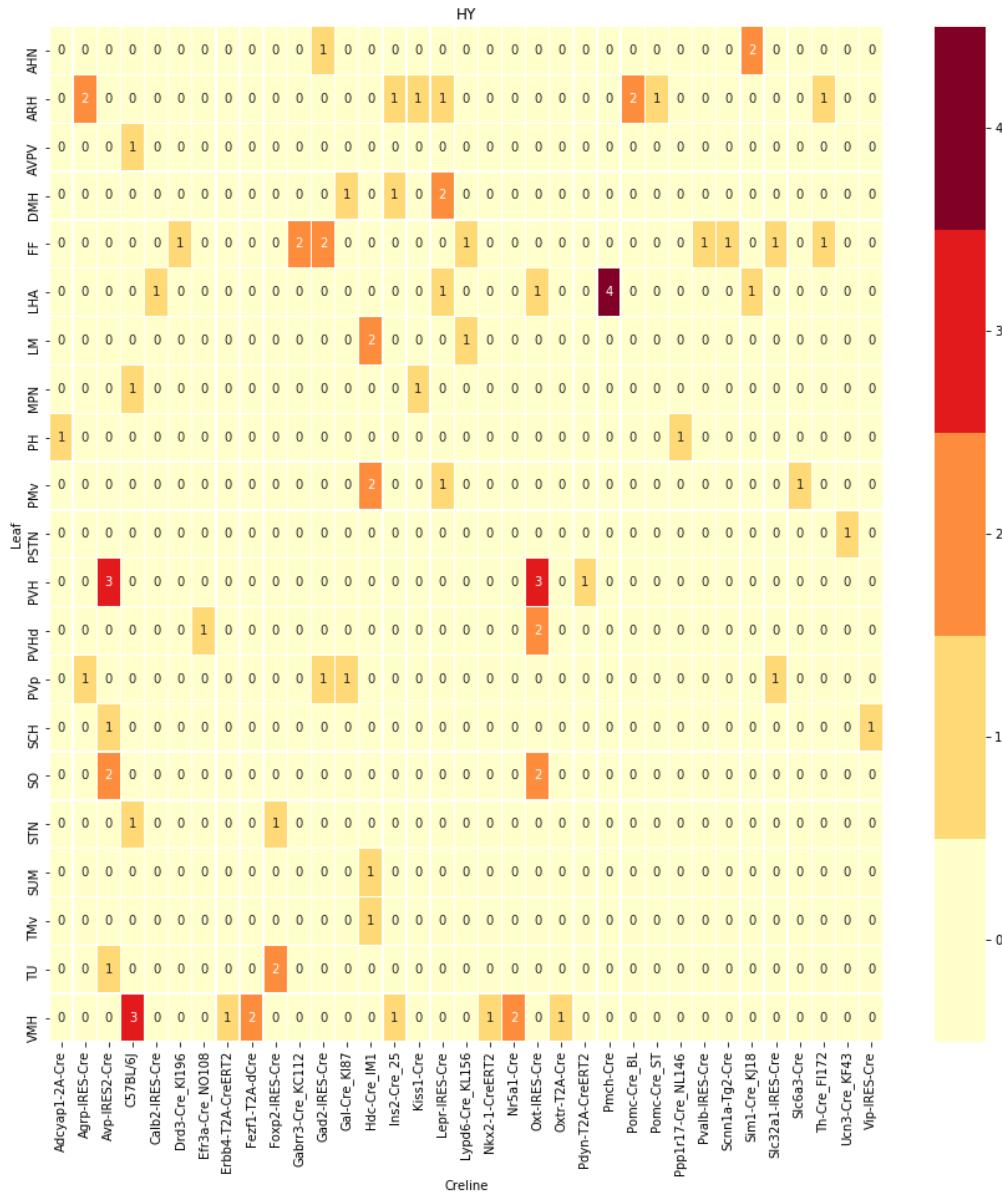
centroid densityoct12.png



centroid densityoct12.png



centroid densityoct12.png



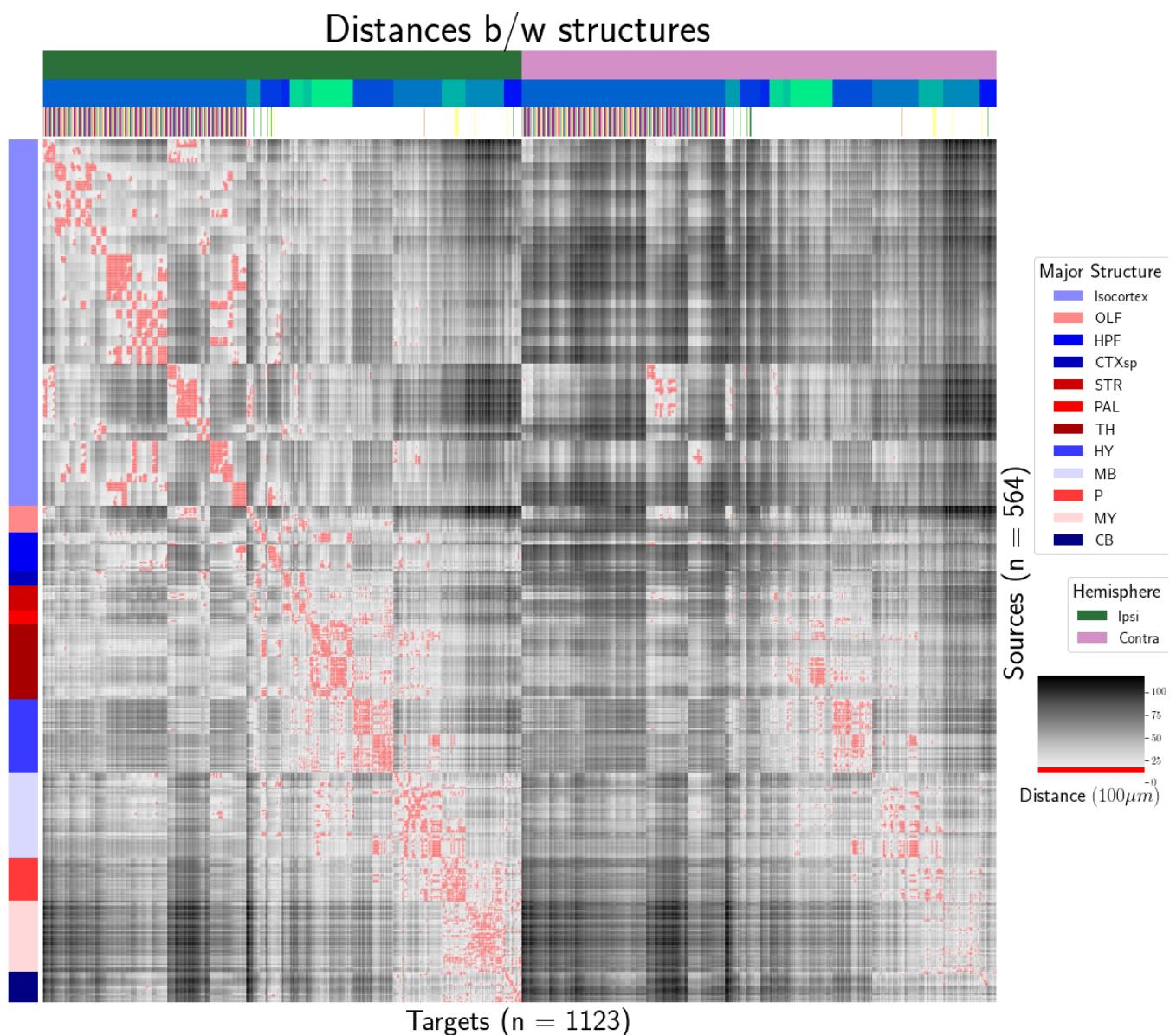
259 ***Distances between structures***

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

260 ***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

²⁶¹ This section consists of additional information on preprocessing of the neural connectivity data,
²⁶² estimation of connectivity, and matrix factorization.

²⁶³ ***Data preprocessing***

²⁶⁴ Several data preprocessing steps take place prior to evaluations of the connectivity matrices. These
²⁶⁵ steps are described in Algorithm ?? . The arguments of this normalization process - injection signals
²⁶⁶ $x(i)$, projection signals $y(i)$, injection fraction $F(i)$, and data quality mask $q(i)$ - were downloaded
²⁶⁷ using the Allen SDK. The injections and projection signals $\in \mathcal{B} \times [0, 1]$ were segmented manually in
²⁶⁸ histological analysis. The projection signal gives the proportion of pixels within the voxel displaying
²⁶⁹ fluorescence, and the injection signal gives the proportion of pixels within the histologically-selected
²⁷⁰ injection subset displaying fluorescence. The injection fraction $\in \mathcal{B} \times [0, 1]$ gives the proportion of
²⁷¹ pixels within each voxel in the injection subset. Finally, the data quality mask $\in \mathcal{B} \times \{0, 1\}$ gives the
²⁷² voxels that have valid data.

²⁷³ Our preprocessing makes use of the above ingredients, as well as several other essential steps. First,
²⁷⁴ we compute the weighted injection centroid

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

²⁷⁵ Given a regionalization \mathcal{R} , we also have access to a regionalization map as $R : \mathcal{B} \rightarrow \mathcal{R}$ which induces a
²⁷⁶ map of connectivities

$$\begin{aligned} R_* : \mathcal{F} &\rightarrow \mathcal{R} \times \mathbb{R}^+ \\ (\nu, y) &\mapsto \sum_{\nu' \in R} y' \text{ for } (\nu, y') \text{ s.t. } R \ni \nu. \end{aligned}$$

²⁷⁷ This map depends on the choice of regionalization; we regionalize at the leaf level. We also can
²⁷⁸ restrict a signal to a individual structure

$$\begin{aligned} S_* : \mathcal{F} &\rightarrow \mathcal{F} \\ (\nu, y) &= \begin{cases} (\nu, y) & \text{if } \nu \in S \\ (\nu, 0) & \text{otherwise} \end{cases} \end{aligned}$$

²⁷⁹ 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 $\tilde{y}(i), c(i)$

280 ***Estimators***281 Our estimators model a connectivity vector $f(v, s) \in \mathbb{R}_{\geq}^T$, and so we may write

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

282 Thus, for the remainder of this section, we will discuss only $f(v, s)$.283 *Centroid-based Nadaraya-Watson* In the Nadaraya-Watson approach of Knox et al. (2019), the injection
284 is considered only through its centroid $c(i) := c(x(i))$, and the projection is considered regionalized.

285 That is,

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

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

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

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

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

290 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
291 position l .292 Several facets of the estimator are visible here. A smaller γ corresponds to a greater amount of
293 smoothing, and index set $I \subseteq \{1 : n\}$ indicates which experiments to use to generate the prediction.
294 Fitting γ via empirical risk minimization therefore bridges between 1-nearest neighbor prediction and
295 averaging of all experiments in I . In Knox et al. (2019), I consisted of experiments sharing the same
296 brain division. Restricting of index set to only include experiments with the same neuron class gives
297 the class-specific Cre-NW model.

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

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

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

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

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

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

315 where

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

316 Note that g must be a concave, non-decreasing function of its arguments with $g(0, 0) = 0$, then g
 317 defines a metric on the product of the metric spaces defined by experiment centroid and
 318 target-encoded cre-line, and \hat{f}_{EL} is a Nadaraya-Watson estimator. A derivation of this fact is given in
 319 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 $\nu(I_s)$, g , location l , cell-class

ν

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

Target encode $\nu(1:n)$ and ν with $n(r(y(1:n)))$

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

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

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

Figure 8: The Expected-Loss estimator

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

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

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

- 329 ▪ The first property of a metric is that $f(x, x') = 0 \iff x = x'$. The left implication:
 330 $x = x' \implies f(x_1, x'_1, \dots, x_n, x'_n) = g(0, \dots, 0)$, since d are metrics. Then, since $g(0) = 0$, we have that
 331 $f(x, x') = 0$. The right implication: $f(x, x') = 0 \implies d = 0 \implies x = x'$ since d are metrics.
- 332 ▪ The second property of a metric is that $f(x, x') = f(x', x)$. This follows immediately from the
 333 symmetry of the d_i , i.e. $f(x, x') = f(x_1, x'_1, \dots, x_n, x'_n) = g(d_1(x_1, x'_1), \dots, d_n(x_n, x'_n)) =$
 334 $g(d_1(x'_1, x_1), \dots, d_n(x'_n, x_n)) = f(x'_1, x_1, \dots, x'_n, x_n) = f(x', x)$.
- 335 ▪ The third property of a metric is the triangle inequality: $f(x, x') \leq f(x, x^*) + f(x^*, x')$. To show this
 336 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
 337 non-decreasing and by the triangle inequality of d . Then, since g is concave,
 338 $g(d(x, x^*) + d(x^*, x')) \leq g(d(x, x^*)) + g(d(x^*, x')) = f(x, x^*) + f(x^*, x')$.

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

- 340 ▪ The first property is that $g(d) = 0 \iff d = 0$. We first show the right implication: $g(d) = 0$, and
 341 $g(d) = f(x, x')$, so $x = x'$ (since f is a metric), so $d = 0$. We then show the left implication:
 342 $d = 0 \implies x = x'$, since d is a metric, so $f(x, x') = 0$, since f is a metric, and thus $g(d) = 0$.
- 343 ▪ The second property is that g is non-decreasing. We proceed by contradiction. Suppose g is
 344 decreasing in argument d_1 in some region $[l, u]$ with $0 < l < u$. Then
 345 $g(d_1(0, l), 0) \geq g(d_1(0, 0), 0) + g(d_1(0, u), 0) = g(d_1(0, u), 0)$, which violates the triangle inequality on
 346 f . Thus, decreasing g means that f is not a metric, so f a metric implies non-decreasing g .

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

351 ***Establishing a lower detection limit***

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

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

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

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

363 ***Decomposing the connectivity matrix***

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

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

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

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

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

378 *Cross-validating NMF* Cross-validation for NMF is somewhat standard but not entirely well-known,
 379 and so we review it here. In summary, a NMF model is first fit on a reduced data set, and an evaluation
 380 set is held out. After random masking of the evaluation set, the loss of the learned model is then
 381 evaluated on the basis of successful reconstruction of the held-out values. This procedure is
 382 performed repeatedly, with replicates of random masks at each tested dimensionality q . This
 383 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 \mathcal{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

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

³⁸⁴ 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

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

where

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

³⁸⁵ Here, $1_{M(p)_r}^C$ is the binary complement of $1_{M(p)_r}$.

Theoretically, the optimum number of components is then

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

³⁸⁶ However, the low decrease in error at higher values of q will motivate us to empirically select a slightly
³⁸⁷ smaller number of components.

³⁸⁸ *Stabilizing NMF* The NMF program is non-convex, and, empirically, individual replicates will not
³⁸⁹ converge to the same optima. One solution therefore is to run multiple replicates of the NMF
³⁹⁰ algorithm and cluster the resulting vectors. This approach raises the questions of how many clusters
³⁹¹ to use, and how to deal with stochasticity in the clustering algorithm itself. We address this issue
³⁹² through the notion of clustering stability (von Luxburg, 2010a).

The clustering stability approach is to generate L replicas of k-cluster partitions $\{C_{kl} : l \in 1 \dots L\}$ 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

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

393 A review of this approach is found in von Luxburg (2010b). Intuitively, archetype vectors that cluster
394 together frequently over clustering replicates indicate the presence of a stable clustering. For d , we
395 utilize the adjusted Rand Index - a simple dissimilarity measure between clusterings. Note that we
396 expect to select slightly more than the q components suggested by cross-validation, since archetype
397 vectors which appear in one NMF replicate generally should appear in others. We then select the q
398 clusters with the most archetype vectors - the most stable NMF results - and take the median of each
399 cluster to create a sparse representative archetype Kotliar et al. (2019); Wu et al. (2016). Experimental
400 results for these cross-validation and stability selection approaches are given in Supplemental Section
401 7.

7 SUPPLEMENTAL EXPERIMENTS

402 Establishing a lower limit of detection

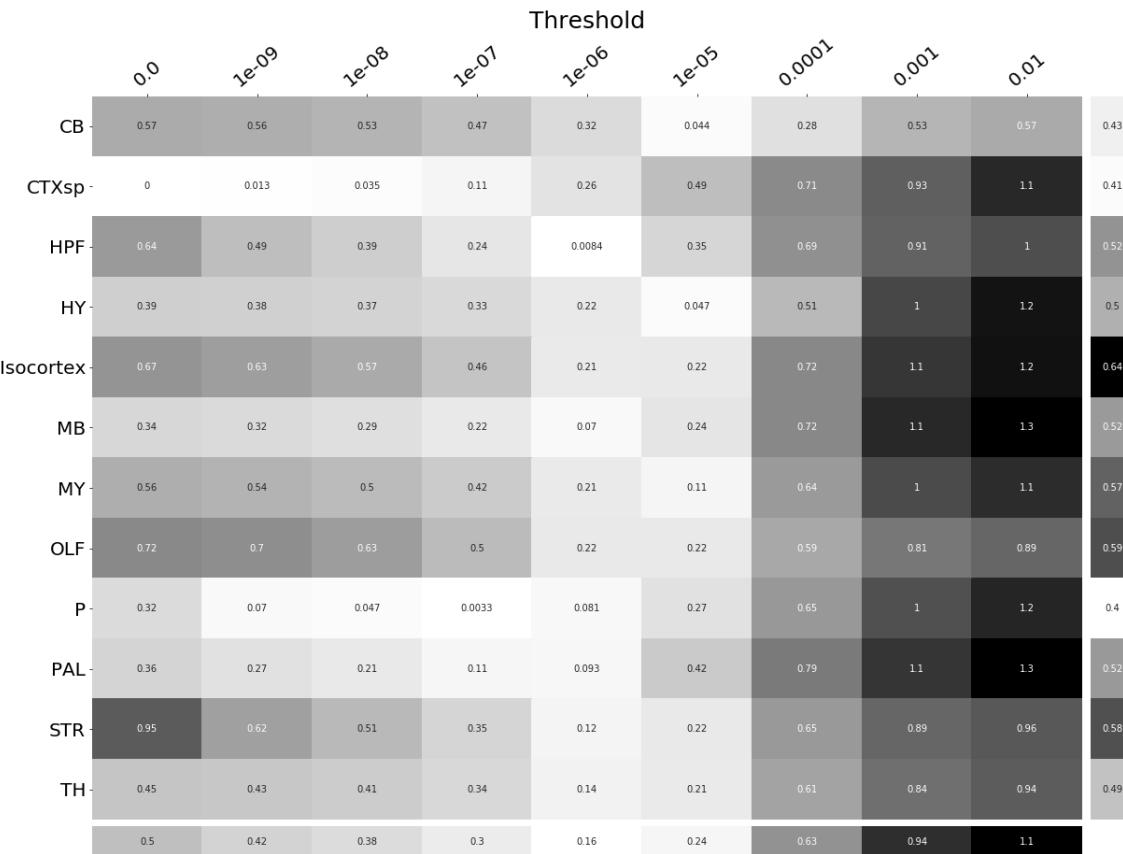


Figure 9: τ at different limits of detection.

403 **Loss subsets**

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



Figure 10

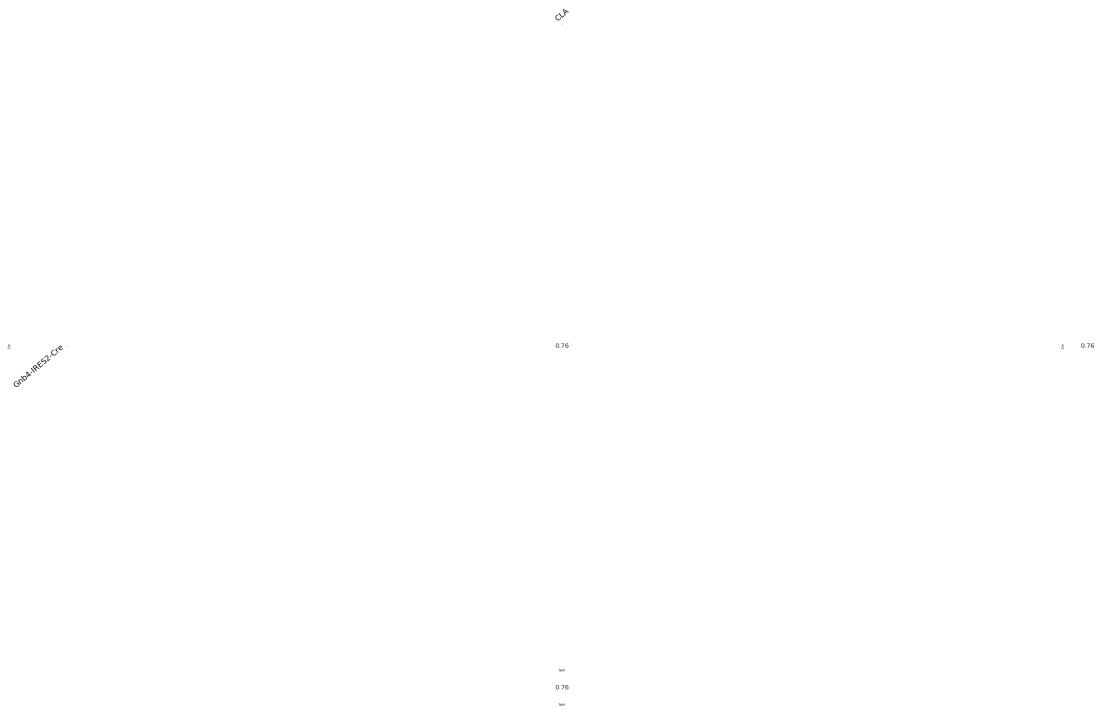


Figure 11

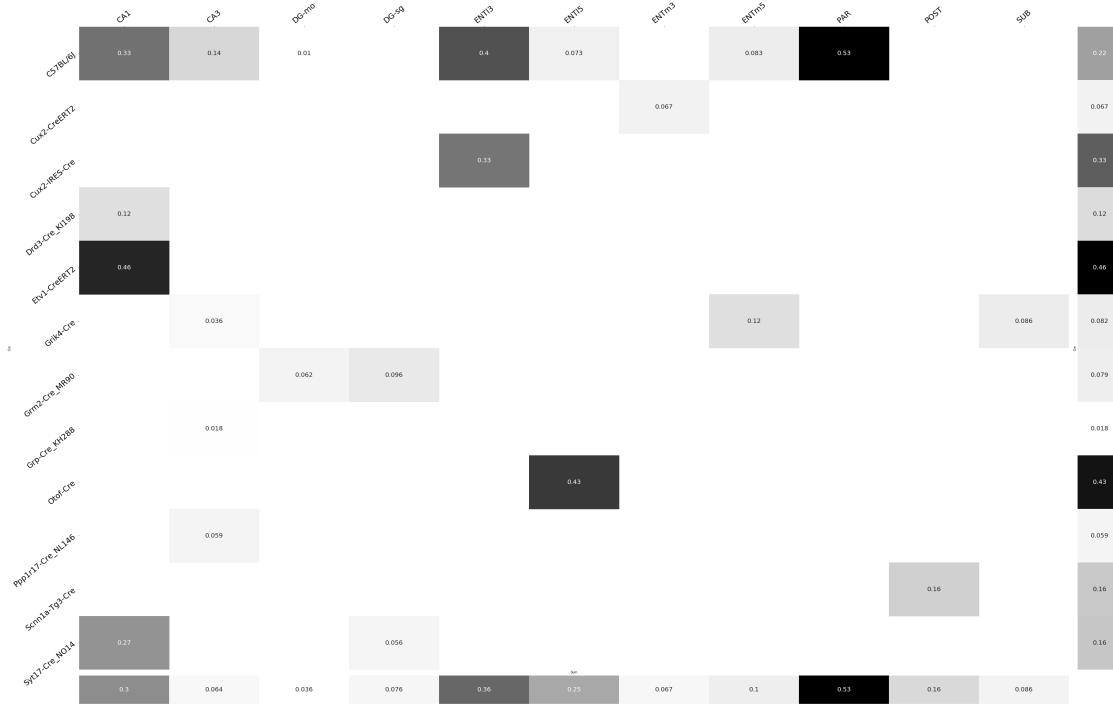


Figure 12

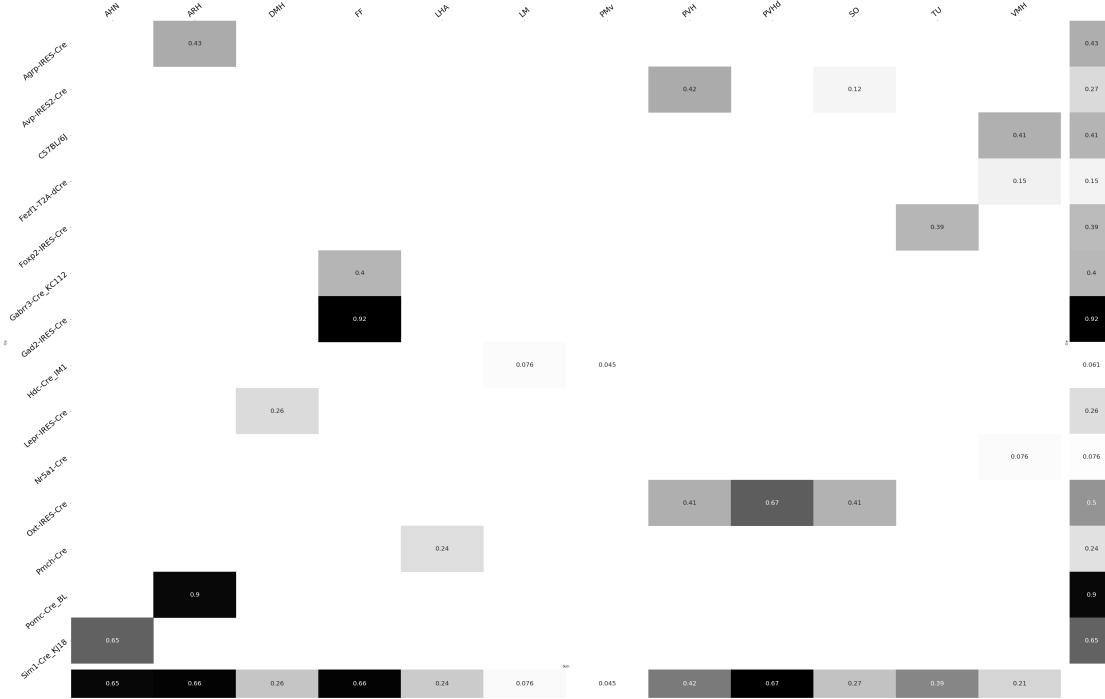


Figure 13

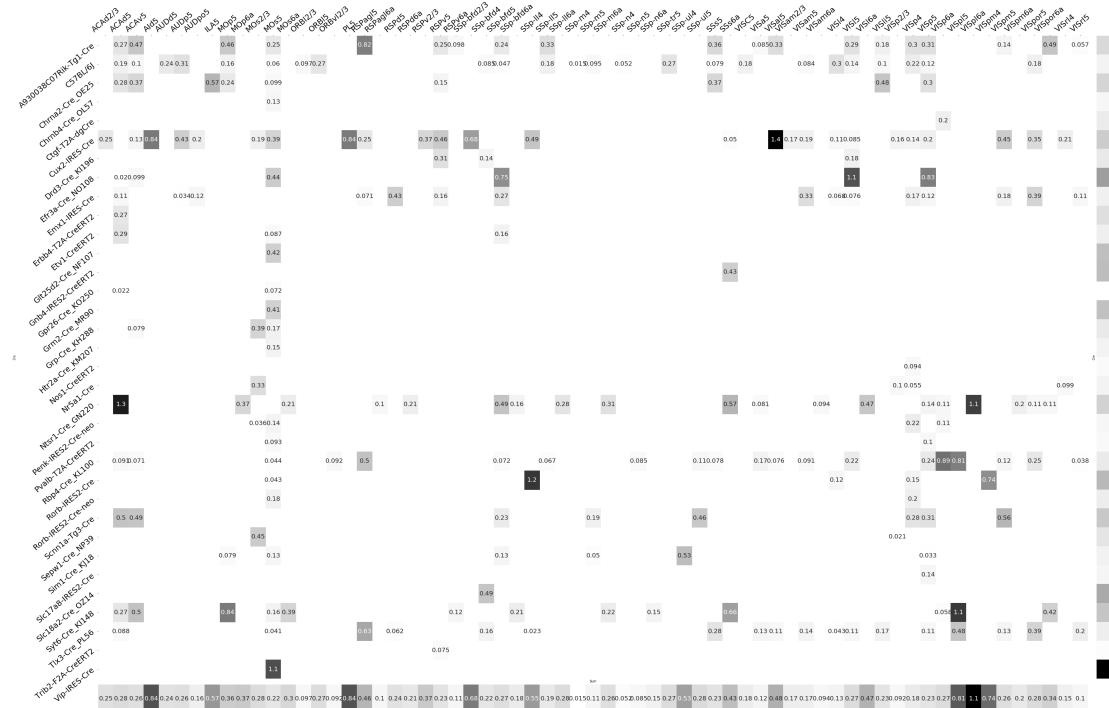


Figure 14



Figure 15



Figure 16



Figure 17

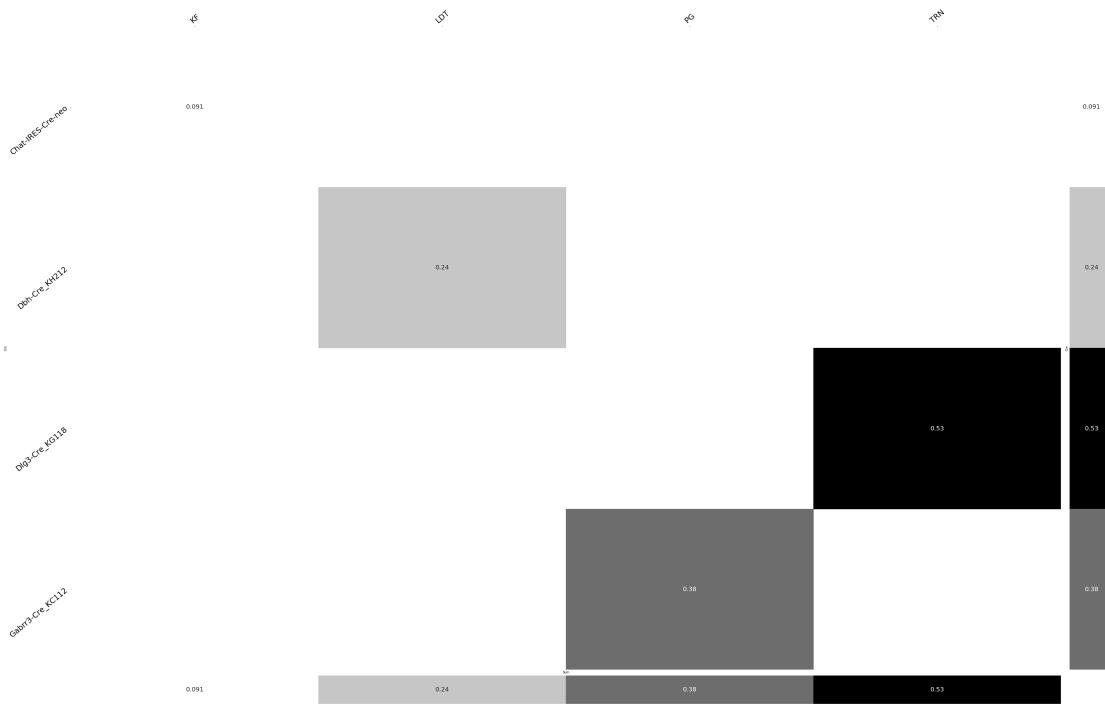


Figure 18

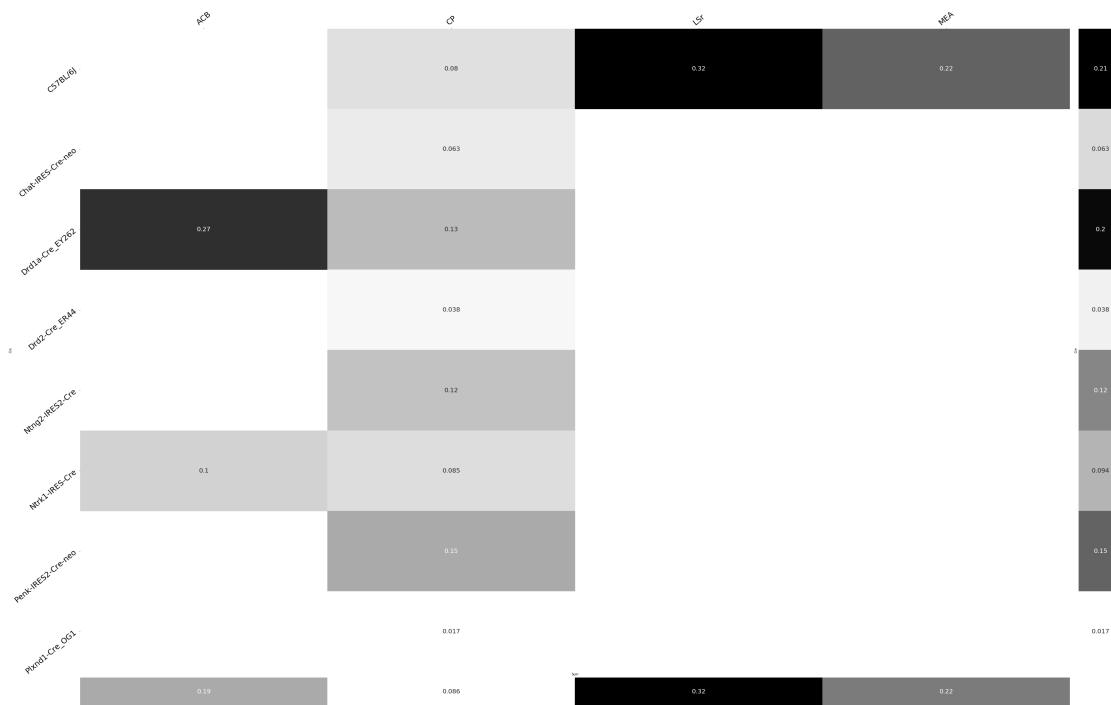


Figure 19

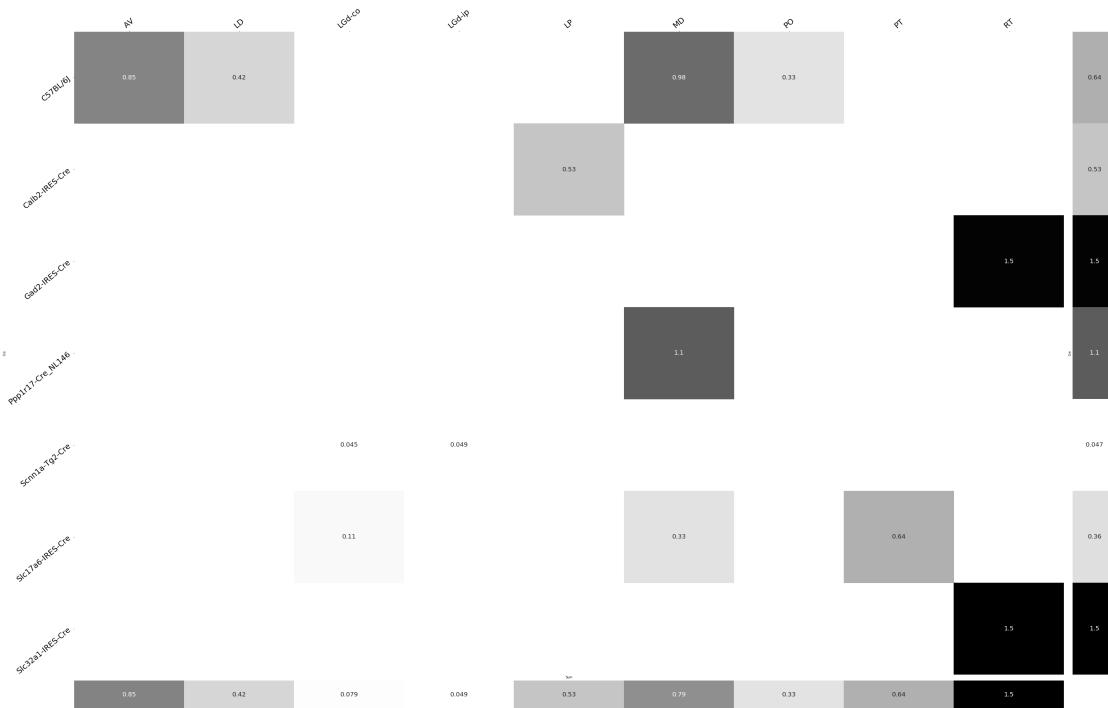


Figure 20

406 **Matrix Factorization**

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

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

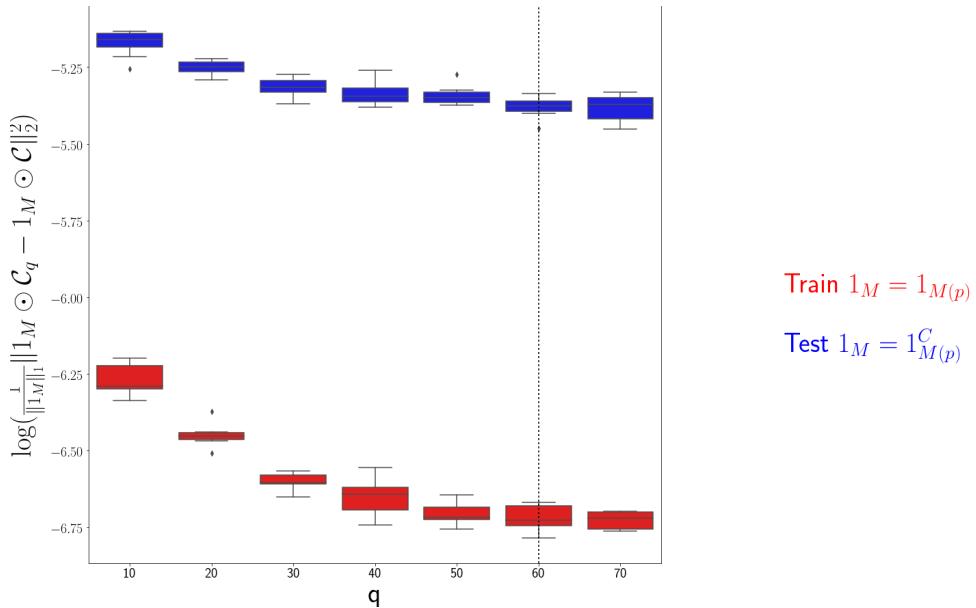


Figure 21: Train and test error using NMF decomposition.

415 *Stability* For the purposes of visualization and interpretability, we restrict to a $q = 15$ component
 416 model. To address the instability of the NMF algorithm in identifying components, we $k - \text{means}$
 417 cluster components over $R = 10$ replicates with $k \in \{10, 15, 20, 25, 30\}$. Since the clustering is itself
 418 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

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

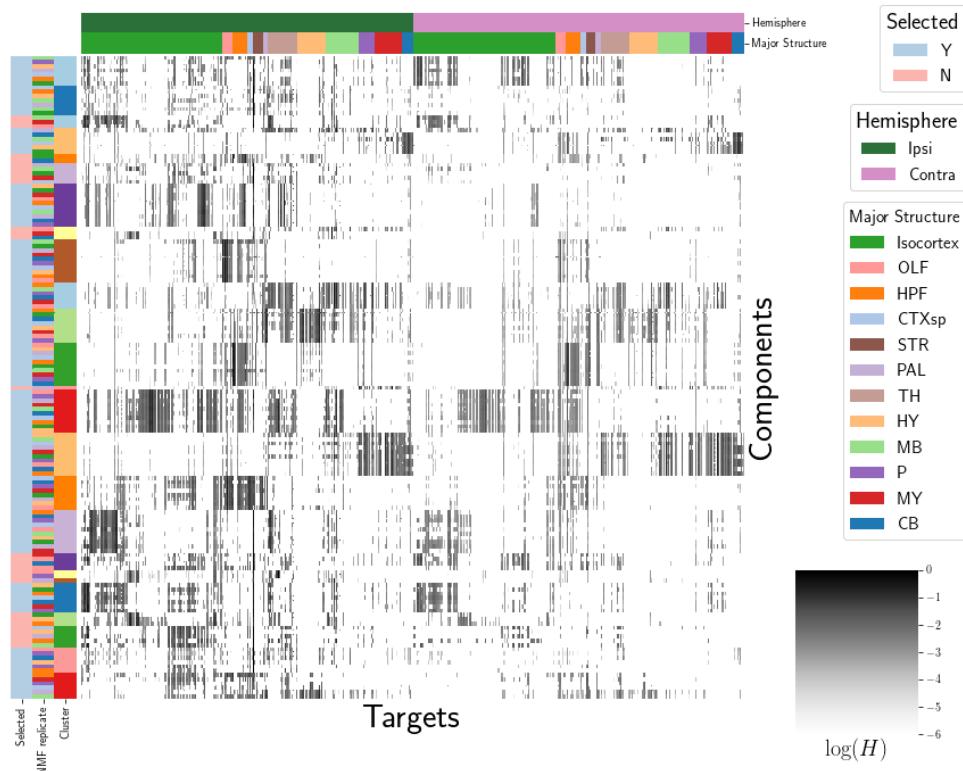


Figure 22: Stability of NMF results across replicates. Replicate and NMF component are shown on rows. Components that are in the top 15 are also indicated.

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

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

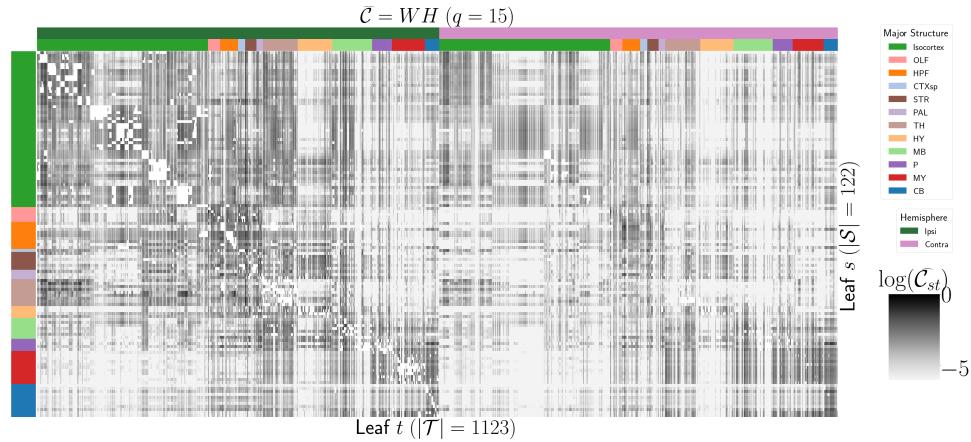


Figure 23: Reconstruction of \mathcal{C} from H and W with $q = 15$ in Figure ??.

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

8 GLOSSARY OF SYMBOLS

9 COMPETING INTERESTS

⁴²⁴ This is an optional section. If you declared a conflict of interest when you submitted your manuscript,
⁴²⁵ please use this space to provide details about this conflict.

426

427

REFERENCES

428

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

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

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

435 Gao, Y., Zhang, X., Wang, S., & Zou, G. (2016). Model averaging based on leave-subject-out cross-validation. *J. Econom.*,
 436 192(1), 139–151.

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

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

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

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

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

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

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

- 451 Oh, S. W., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., ... Zeng, H. (2014). A mesoscale connectome of the
452 mouse brain. *Nature*, 508(7495), 207–214.
- 453 Saul, L. K., & Roweis, S. T. (2003). Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *J.
454 Mach. Learn. Res.*, 4(Jun), 119–155.
- 455 von Luxburg, U. (2010a). Clustering stability: An overview.
- 456 von Luxburg, U. (2010b). Clustering stability: An overview.
- 457 Wu, S., Joseph, A., Hammonds, A. S., Celtniker, S. E., Yu, B., & Frise, E. (2016). Stability-driven nonnegative matrix
458 factorization to interpret spatial gene expression and build local gene networks. *Proc. Natl. Acad. Sci. U. S. A.*, 113(16),
459 4290–4295.

10 TECHNICAL TERMS

⁴⁶⁰ **Technical Term** a key term that is mentioned in an NETN article and whose usage and definition may
⁴⁶¹ not be familiar across the broad readership of the journal.

⁴⁶² **Cre-line** Refers to the combination of cre-recombinase expression in transgenic mouse and
⁴⁶³ cre-induced promotion in the vector that induces labelling of cell-class specific projection.

⁴⁶⁴ **Cell class** The projecting neurons targeted by a particular cre-line

⁴⁶⁵ **structural connectivities** connectivity between structures

⁴⁶⁶ **Voxel** A $100\mu m$ cube of brain.

⁴⁶⁷ **structural connection tensor** Connectivities between structures given a neuron class

⁴⁶⁸ **dictionary-learning** A family of algorithms for finding low-dimensional data representations.

⁴⁶⁹ **shape constrained estimator** A statistical estimator that fits a function of a particular shape (e.g.
⁴⁷⁰ monotonic increasing, convex).

⁴⁷¹ **Nadaraya-Watson** A simple smoothing estimator.

⁴⁷² **connectivity archetypes** Typical connectivity patterns

⁴⁷³ **Expected Loss** Our new estimator that weights different features by their estimated predictive
⁴⁷⁴ power.