

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

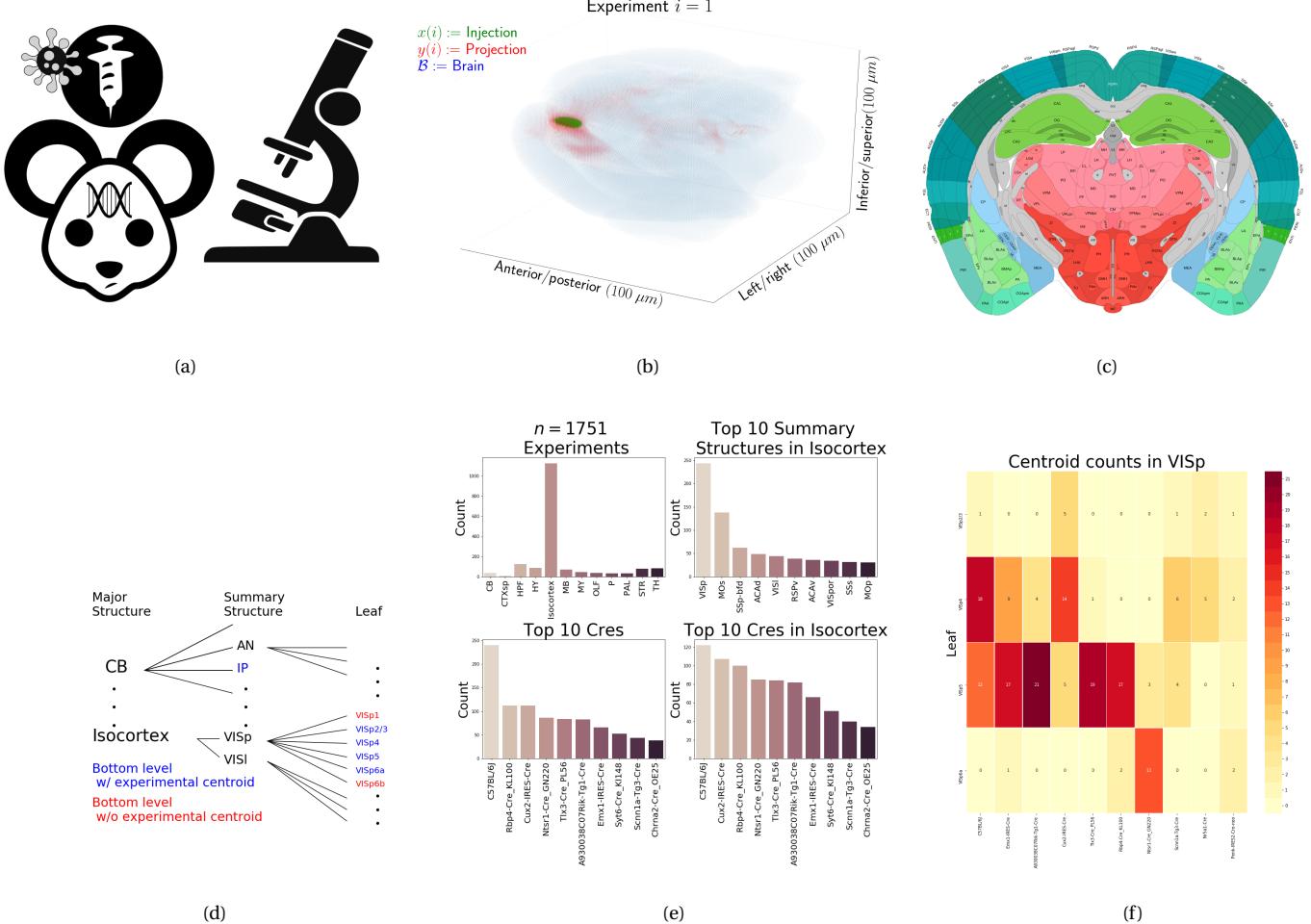
42 structure, with nearby experiments given more weight. However, our model weighs class-specific
43 behavior in a particular structure against spatial position, so a nearby experiment targeting a similar
44 cell class would be relatively upweighted, while a nearby experiment targeting a dissimilar class 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.



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 S source
¹⁰⁰ regions $\mathcal{S} := \{s\}$, T target regions $\mathcal{T} := \{t\}$, and V cell classes $\mathcal{V} := \{v\}$. In contrast to Knox et al. (2019),
¹⁰¹ which only uses wild type *C57BL/6J* mice, these experiments utilize $V = 114$ different Cre-lines. We
¹⁰² generally consider $S = 564$ leaf sources and $T = 1123$ leaf targets, where 559 are contralateral or
¹⁰³ mediolateral.

¹⁰⁴ We preprocess our data in several ways. We discretize florescent signals like injections and
¹⁰⁵ projections into $100\mu m^3$ **voxels**. Given an experiment i , we represent injections and projections as
¹⁰⁶ maps $x(i), y(i) : \mathcal{B} \rightarrow \mathbb{R}_{\geq 0}$, where $\mathcal{B} \subset [1 : 132] \times [1 : 80] \times [1 : 104]$ corresponds to the subset of the
¹⁰⁷ ($1.32 \times 0.8 \times 1.04$) cm rectangular space occupied by the standard mouse brain. A structure s then
¹⁰⁸ contains $|s|$ voxels at locations $\{l_{s_j} \in \mathbb{R}^3\}$, and similarly for targets. We calculate injection centroids
¹⁰⁹ $c(i) \in \mathbb{R}^3$ and regionalized projections $y_{\mathcal{T}}(i) \in \mathbb{R}^T$ giving the sum of $y(i)$ in each region. In contrast to
¹¹⁰ Knox et al. (2019), we also $l1$ normalize these projection vectors. This accounts for differences in the
¹¹¹ cre-driven expression of eGFP via the various transgene promoters. A detailed mathematical
¹¹² description of these steps, including data quality control, is given in Supplemental Section 6.

113 ***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 $\mathcal{C}^D(v, s, t) = \frac{1}{|s||t|} \mathcal{C}(v, l_j, l_{j'})$.

114 These represent the strength of the connection from source to target regions for each class. Since the
 115 normalized strength and density are computable from the strength via a fixed normalization, our
 116 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

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

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

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

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_s , respectively. Data I_s / I_m means that, given a location $l_s \in s \in m$, the model \hat{f} is trained on all of I_m , but only uses I_s for prediction.

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

137 ***Model evaluation***

138 We select optimum functions from within and between our estimator classes using empirical risk
 139 minimization. Equation 1 includes a deterministic step f^* included without input by the data. The
 140 performance of $\widehat{\mathcal{C}}(\nu, s, t)$ is thus determined by performance of $\widehat{f}(f_*(\mathcal{D}(\nu, s)))$. Furthermore, in
 141 practice we can represent f as $f_{\mathcal{T}} : \mathbb{R}^3 \rightarrow \mathbb{R}_{\geq 0}^T$ giving the structural connection strength at a given
 142 location. We can then evaluate the accuracy of $\widehat{f}_{\mathcal{T}}(f_*(\mathcal{D}(\nu, s)))$ using leave-one-out cross validation, in
 143 which the accuracy of the model is assessed by its ability to predict experiments excluded from the
 144 training data.

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

We use weighted l_2 -loss to evaluate these predictions.

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

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

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

156 ***Connectivity analyses***

157 We show neuronal processes underlying our estimated connectome using two types of unsupervised
 158 learning. Our use of hierarchical clustering is standard, and so we do not review it here. However, our
 159 application of non-negative matrix factorization (NMF) to decompose the estimated long-range
 160 connectivity into *connectivity archetypes* that linearly combine to reproduce the observed
 161 connectivity is novel and technically of some independent interest. Non-negative matrix factorization
 162 refers to a collection of **dictionary-learning** algorithms for decomposing a non-negatively-valued
 163 matrix such as \mathcal{C} into positively-valued matrices called, by convention, weights $W \in \mathbb{R}_{\geq 0}^{S \times q}$ and hidden
 164 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
 165 orthant. This H is typically used to identify latent structures with interpretable biological meaning,
 166 and the choice of matrix factorization method reflects particular scientific subquestions and
 167 probabilistic interpretations.

168 Our algorithm is

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

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

3 RESULTS

¹⁷⁷ Our results include a mix of quantitative and qualitative evaluations of model fit, the Cre-specific
¹⁷⁸ connectivity matrices themselves, and retrospective analyses of these matrices for patterns related to
¹⁷⁹ Cre-line and source and target region.

¹⁸⁰ ***Model evaluation***

¹⁸¹ Table ?? contains weighted losses from leave-one-out cross-validation of candidate models. Our EL
¹⁸² model generally performs better than the other Nadaraya-Watson estimators that we consider. For
¹⁸³ example, the NW Major-WT model is the model from Knox et al. (2019). The EL model combines the
¹⁸⁴ good performance of class-specific models like NW Leaf-Cre in regions like Isocortex with the good
¹⁸⁵ performance of class-agnostic models in regions like Thalamus. Additional information on model
¹⁸⁶ evaluation, including class and structure specific performance, is given in Appendix 5 In particular,
¹⁸⁷ Supplementary Table 3 contains the sizes of these evaluation sets in each major structure, and
¹⁸⁸ 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.

189 ***Connectivities***

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

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

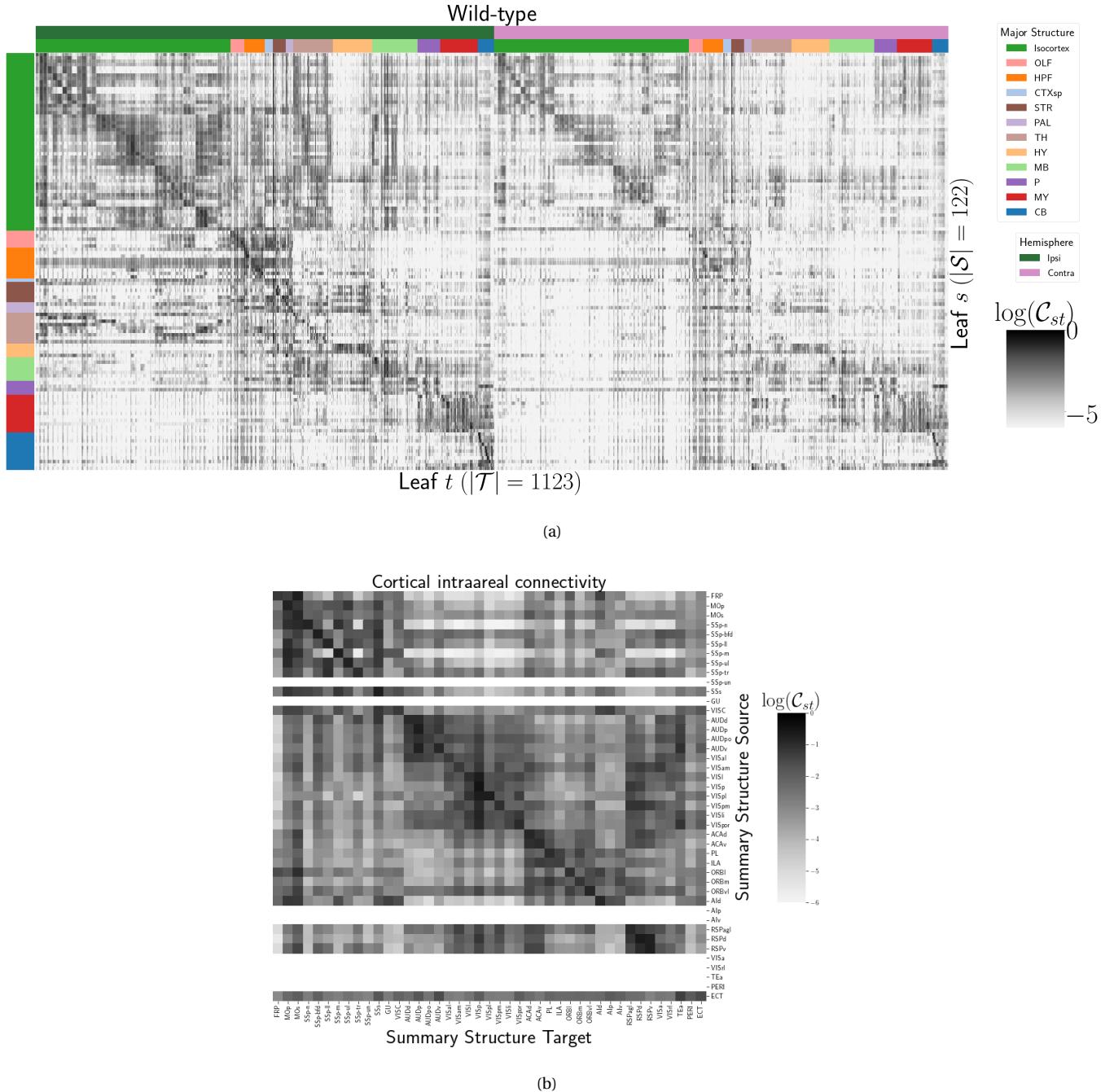


Figure 2: Wild-type connectivities. 2a Log wild-type connectivity matrix $\log \mathcal{C}(s, t, v_{wt})$. 2b Log wild-type intracortical connectivity matrix at the summary structure level.

²⁰⁷ *Class-specific connectivities* We have generated $V = 114$ cell-class specific structural connectivities
²⁰⁸ \mathcal{C}_v . A reasonable question is which source and cell-type combinations behave similarly, and which
²⁰⁹ target projections tend to co-occur. Exhaustive comparison of this estimated behavior is prohibitive,
²¹⁰ but we do exhibit several examples of our class specific connectivities conforming to well-known
²¹¹ behaviors. These validation cases are given in Figure 3.

²¹² We begin by plotting subsets of the estimated connectivities in the well-studied VISp and MO
²¹³ regions in Figure 3a. The localization of Rbp4-Cre and Ntsr1-Cre injection centroids to layers 5 and 6
²¹⁴ respectively is evident (see also Supplemental Figure ??). These layers project to their expected targets
²¹⁵ Jeong et al. (2016). In VISp, the Ntsr1-Cre line strongly targets the thalamic LP nuclei, and in MO, layer
²¹⁶ 5 projects to anterior basolateral amygdala (BLA) and capsular central amygdala (CEA), while layer 6
²¹⁷ does not. As a heuristic alternative model, we also synthesize information about leafs targeted by
²¹⁸ different Cre-lines, we also generate an average connectivity matrix over all Cre-lines. This model is
²¹⁹ not evaluated in our testing, and is only a general stand-in for overall behavior, but provides a useful
²²⁰ summary of results.

²²¹ Figure 3b shows a collection of connectivity strengths generated using cre-specific models for
²²² wild-type, Cux2, Ntsr1, Rbp4, and Tlx3 cre-lines from visual signal processing leafs in the cortex to
²²³ cortical and thalamic nucleii. This shows that cell-class has a dominating effect on projection in
²²⁴ certain regions. We use hierarchical clustering to sort source structure/cell-class combinations by the
²²⁵ similarity of their structural projections, and sort target structures by the structures from which they
²²⁶ receive projections. Examining the former, we can see that the Ntsr1 Cre-line distinctly projects to
²²⁷ thalamic nucleii, regardless of summary structure. This contrasts with the tendency of other cell
²²⁸ classes to project intracortically in a manner determined by the source structure. Similarly, layer 6
²²⁹ targets are not strongly projected to by any of the displayed Cre-lines. There are too many targeted
²³⁰ summary structures to plot here, but we expect that the source profile of each target clusters by
²³¹ structure.

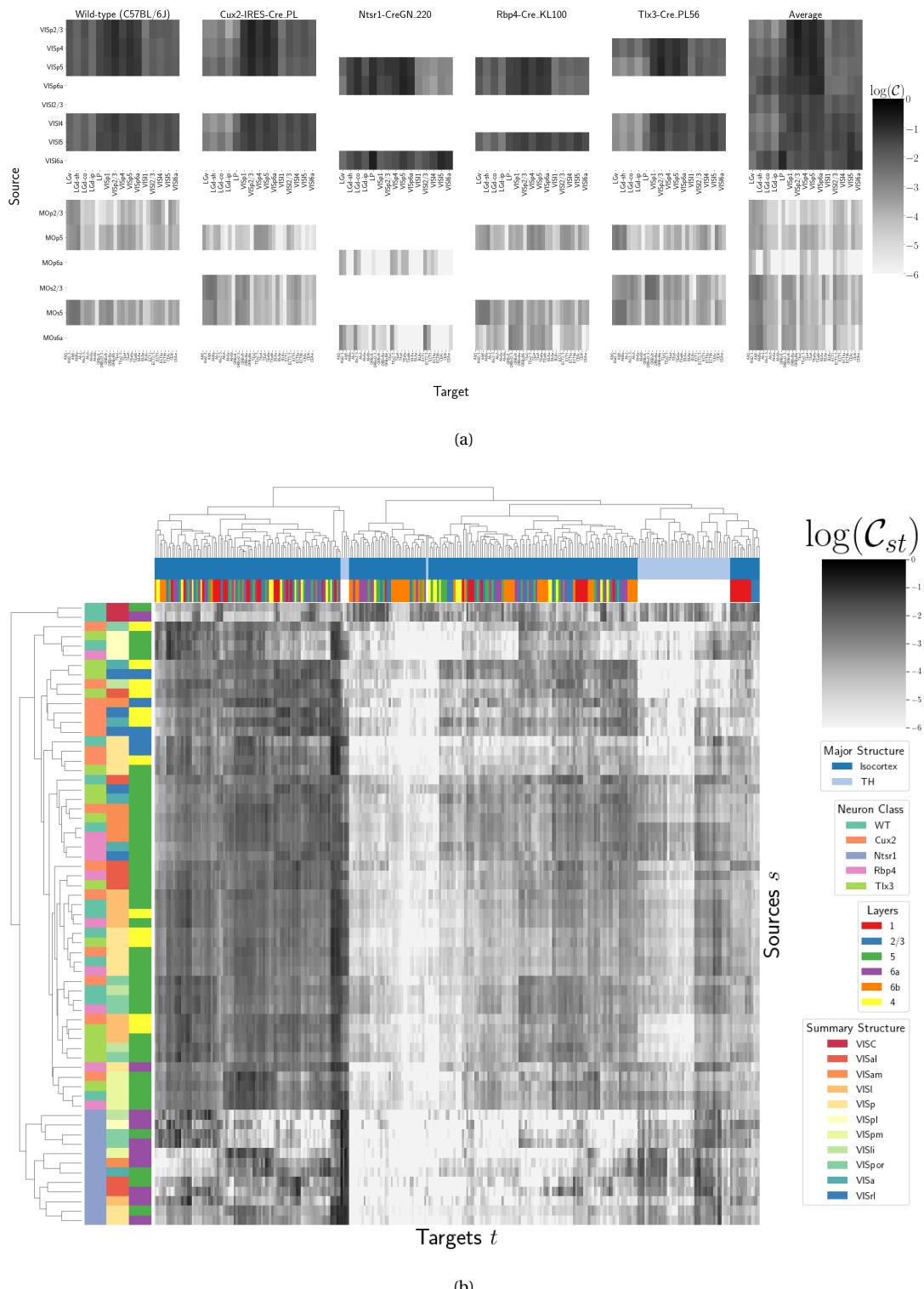
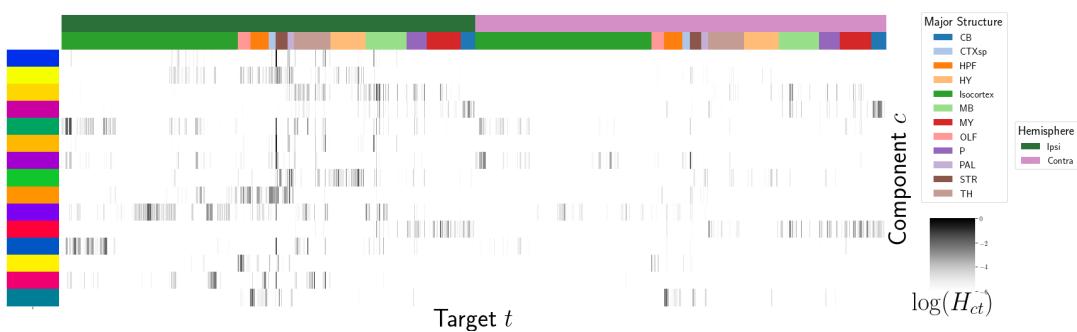


Figure 3: Cell-class and layer specific connectivities from VISp and MO. This figure shows a preselected subset of putatively interesting connectivities from VISp and MO. Sources without a injection of that Cre-type are not estimated due to lack of data for that Cre-line in that structure. 3 Heirarchical clustering of connectivity strengths from visual signal processing cell-types to cortical and thalamic 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

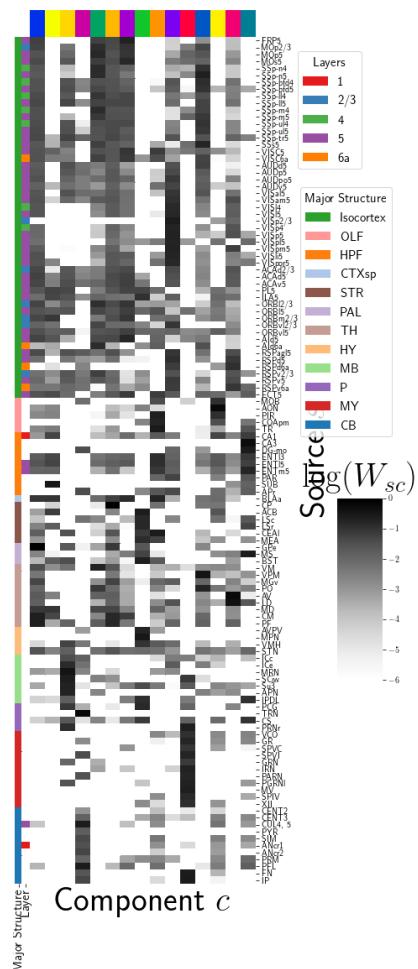
232 Connectivity Analyses

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

242 Since certain cell-types and layers have a characteristic connectivity pattern, we perform
243 non-negative matrix factorization on distal wild-type connectivities to estimate these characteristic
244 patterns in a probabilistic way. This decomposes the remaining censored connectivity matrix into a
245 linear model based off a relatively small number of distinct signals. These signals are plotted in Figure
246 4, and technical details and intermediate results are given in Supplemental Sections 6 and 7,
247 respectively. The plotted decomposition shows that these underlying connectivity archetypes
248 correspond strongly to major brain division. However, certain components that predominantly
249 represent connectivity from a given major brain division may also be accessed from other areas. For
250 example, the IP and FN regions of CB are strongly associated in 4b with the component projecting to
251 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

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

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

ACKNOWLEDGMENTS

²⁶⁷ The Funder and award ID information you input at submission will be introduced by the publisher
²⁶⁸ under a Funding Information head during production. Please use this space for any additional
²⁶⁹ acknowledgements and verbiage required by your funders.

270 This supplement is divided into information about our dataset, supplemental methods, and
271 supplemental results. However, note that there is substantial interconnectedness between these
272 sections. Thus, if a reader is interested in, say, non-negative matrix factorization, they may find
273 relevant information in both methods and results.

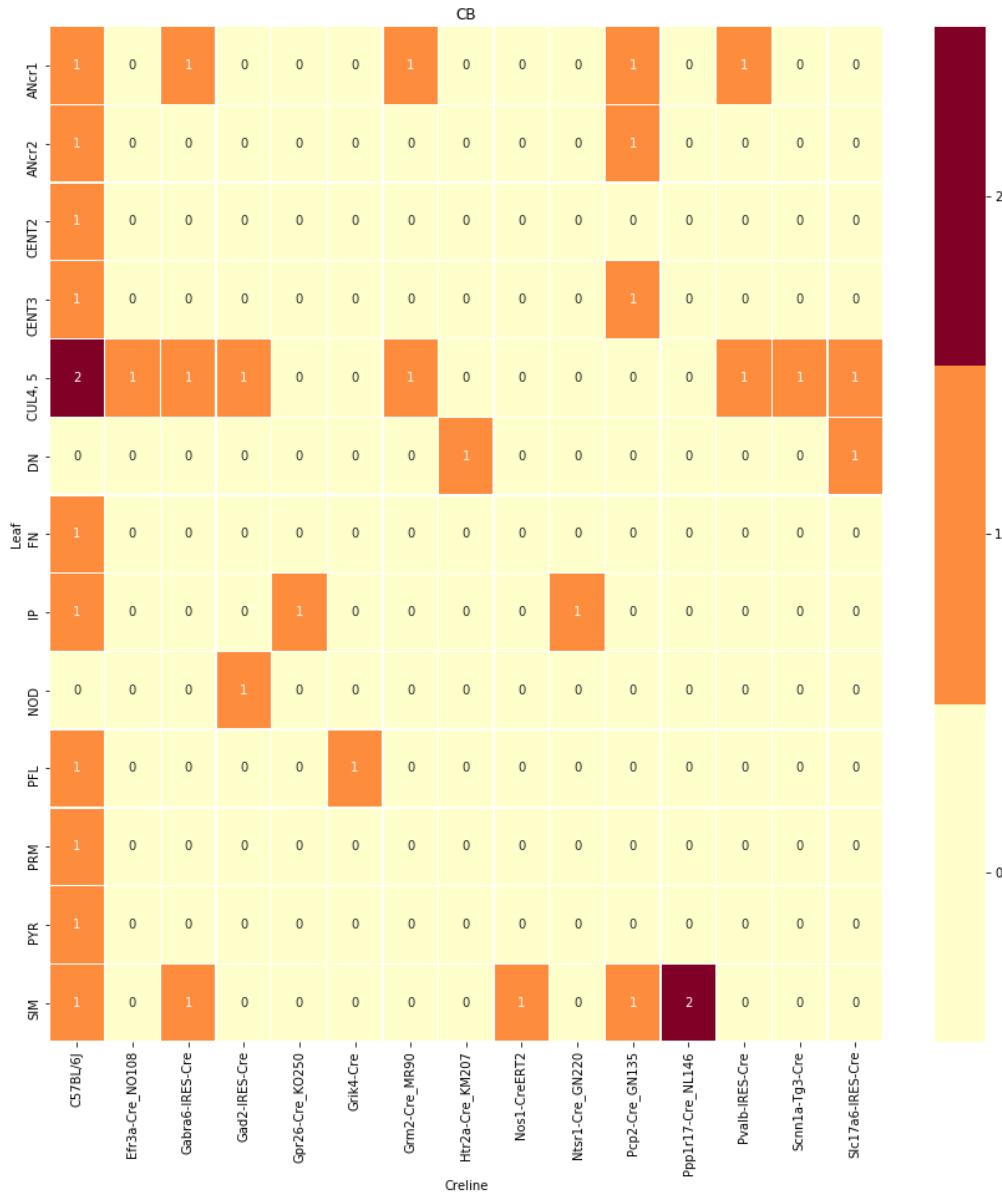
5 SUPPLEMENTAL INFORMATION

274 Our supplementary information consists of abundances of leaf/Cre-line combinations, information
275 about distances between structures, and the size of our restricted evaluation dataset.

276 ***Cre/structure combinations in \mathcal{D}***

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

centroid densityoct12.png



centroid densityoct12.png

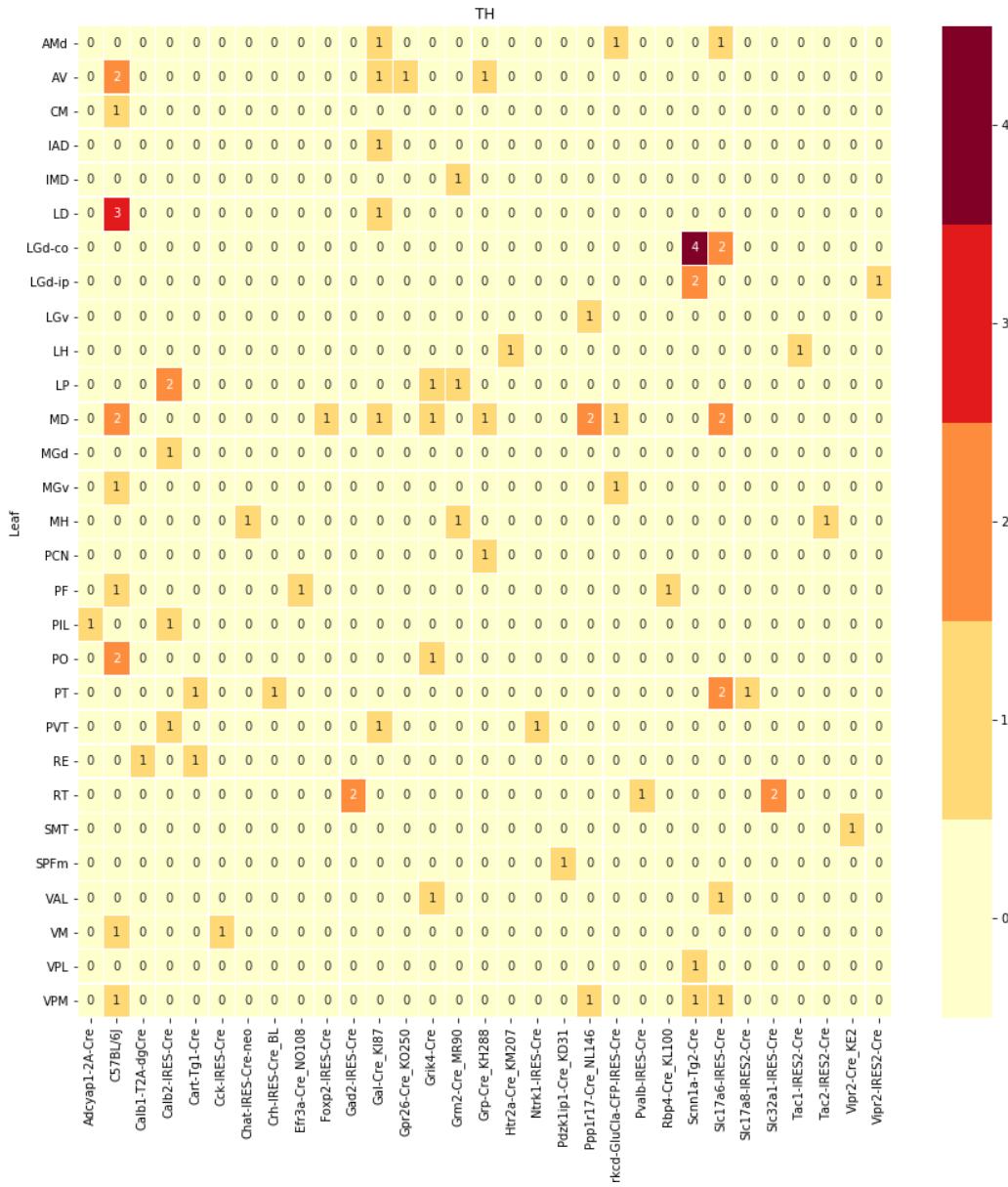


Figure 5: Caption

centroid densityoct12.png

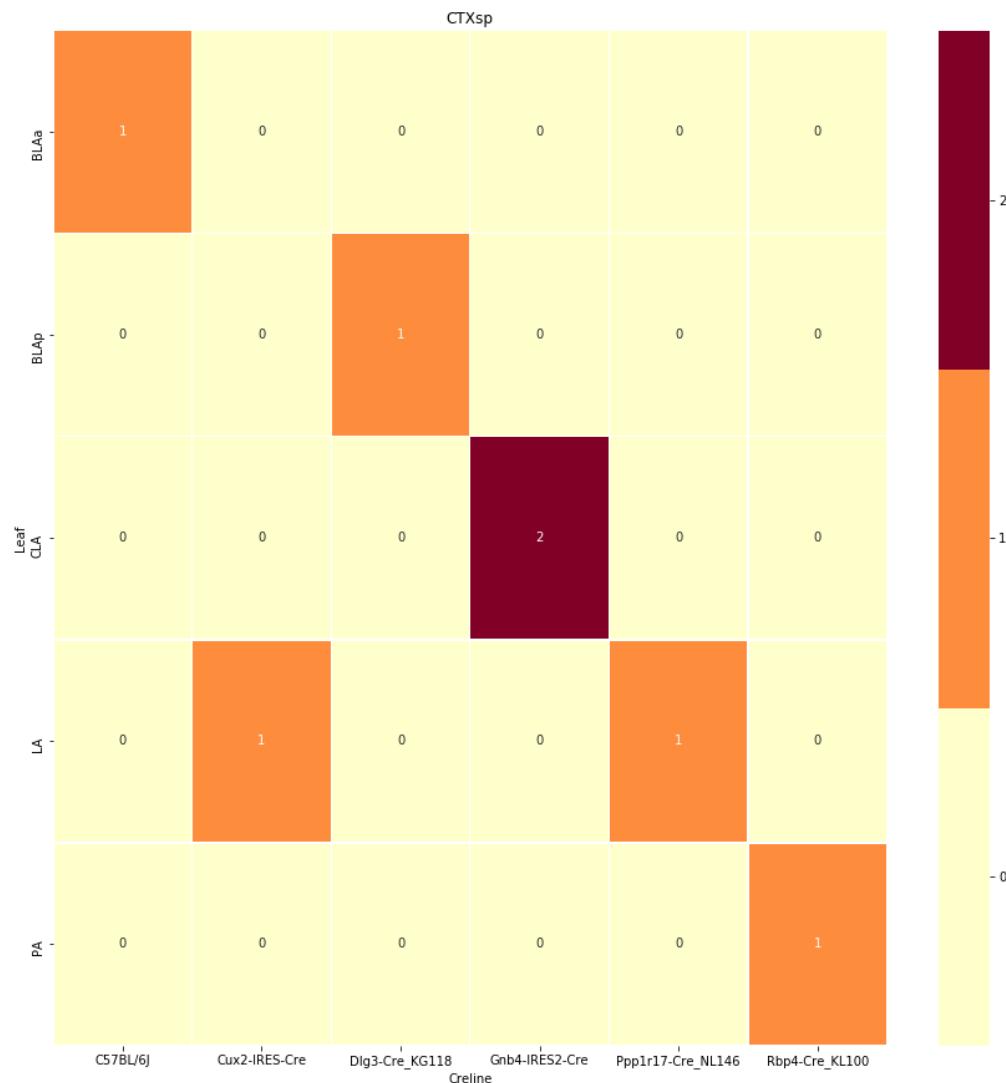
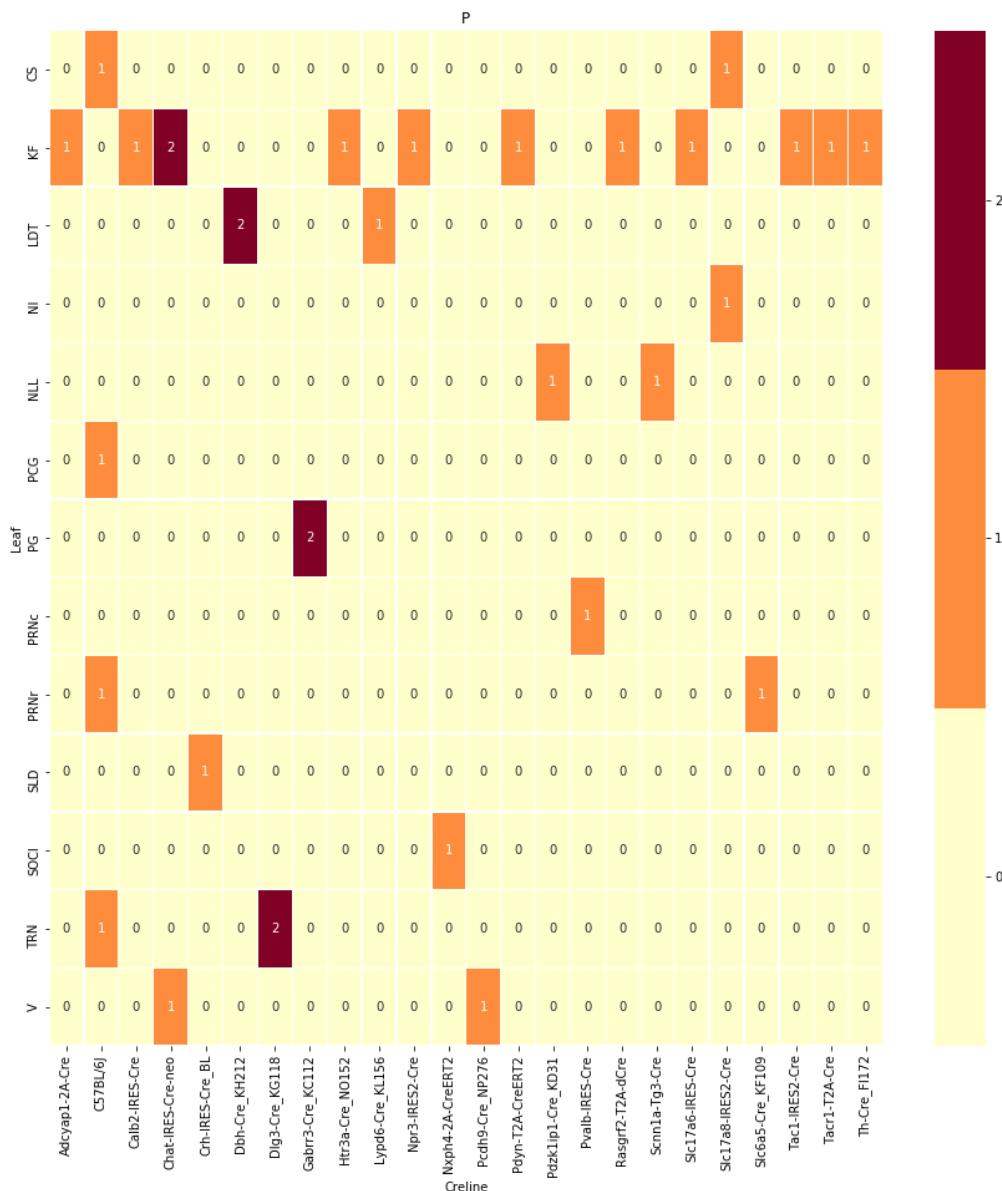
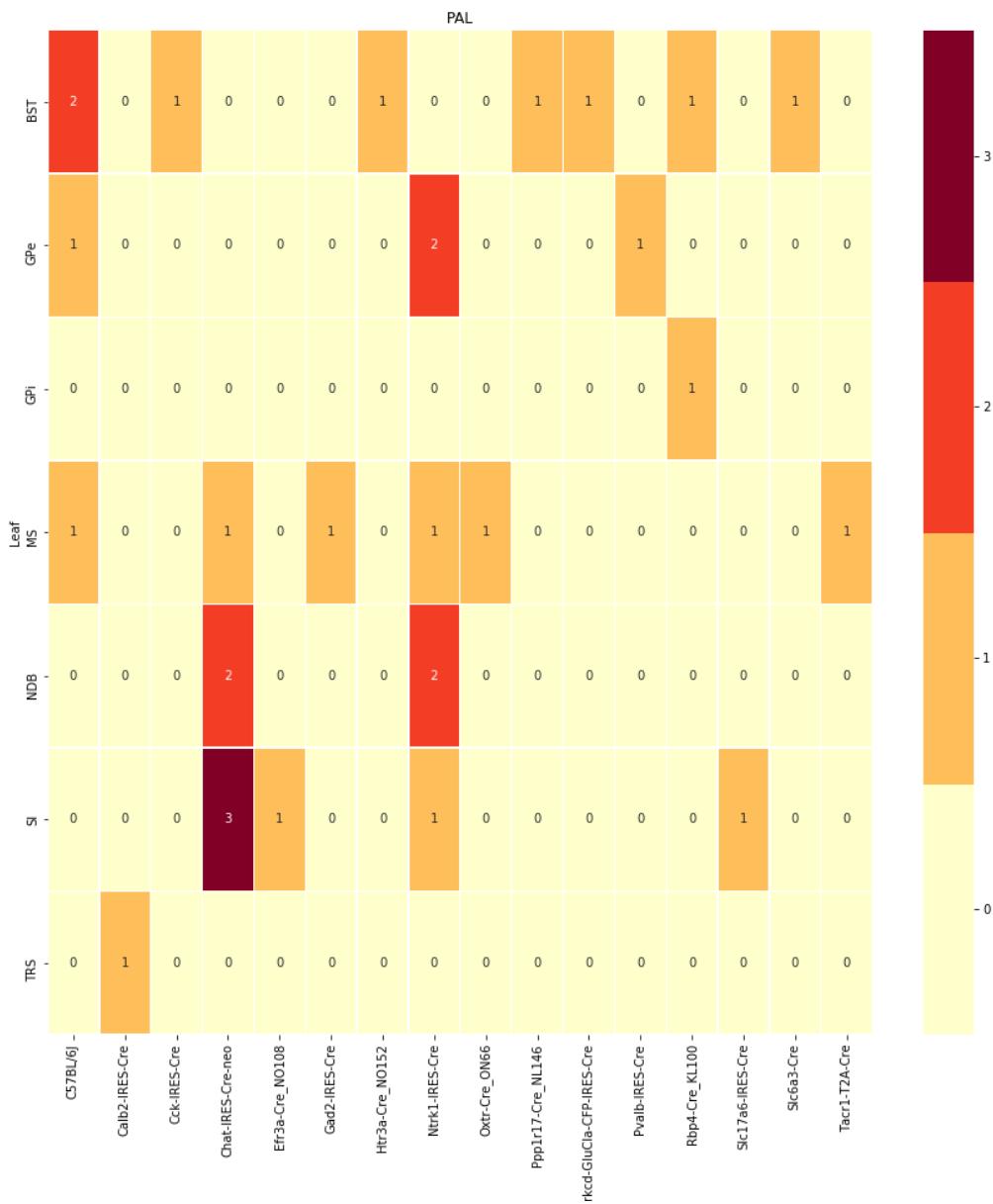


Figure 6: Caption

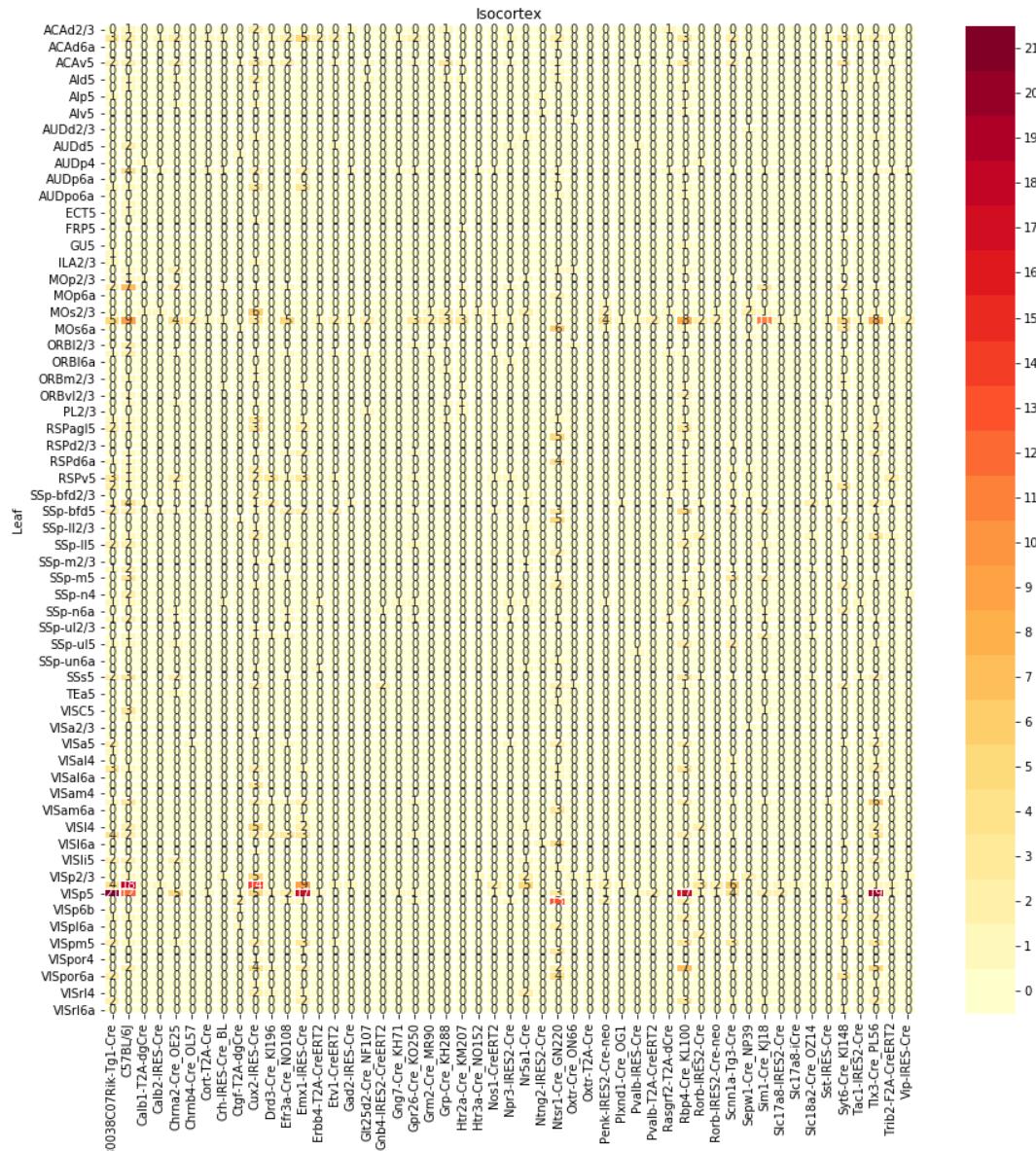
centroid densityoct12.png



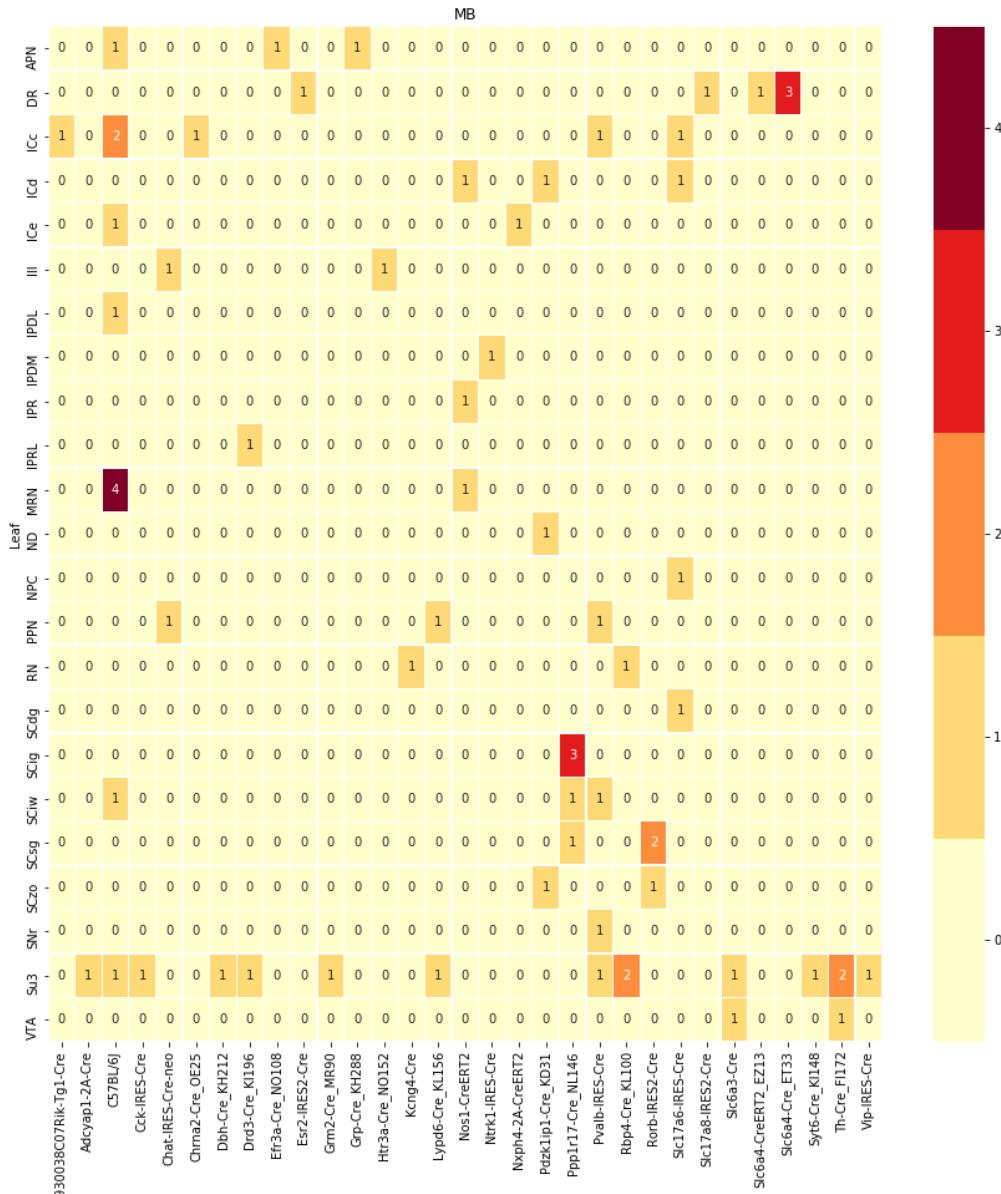
centroid densityoct12.png



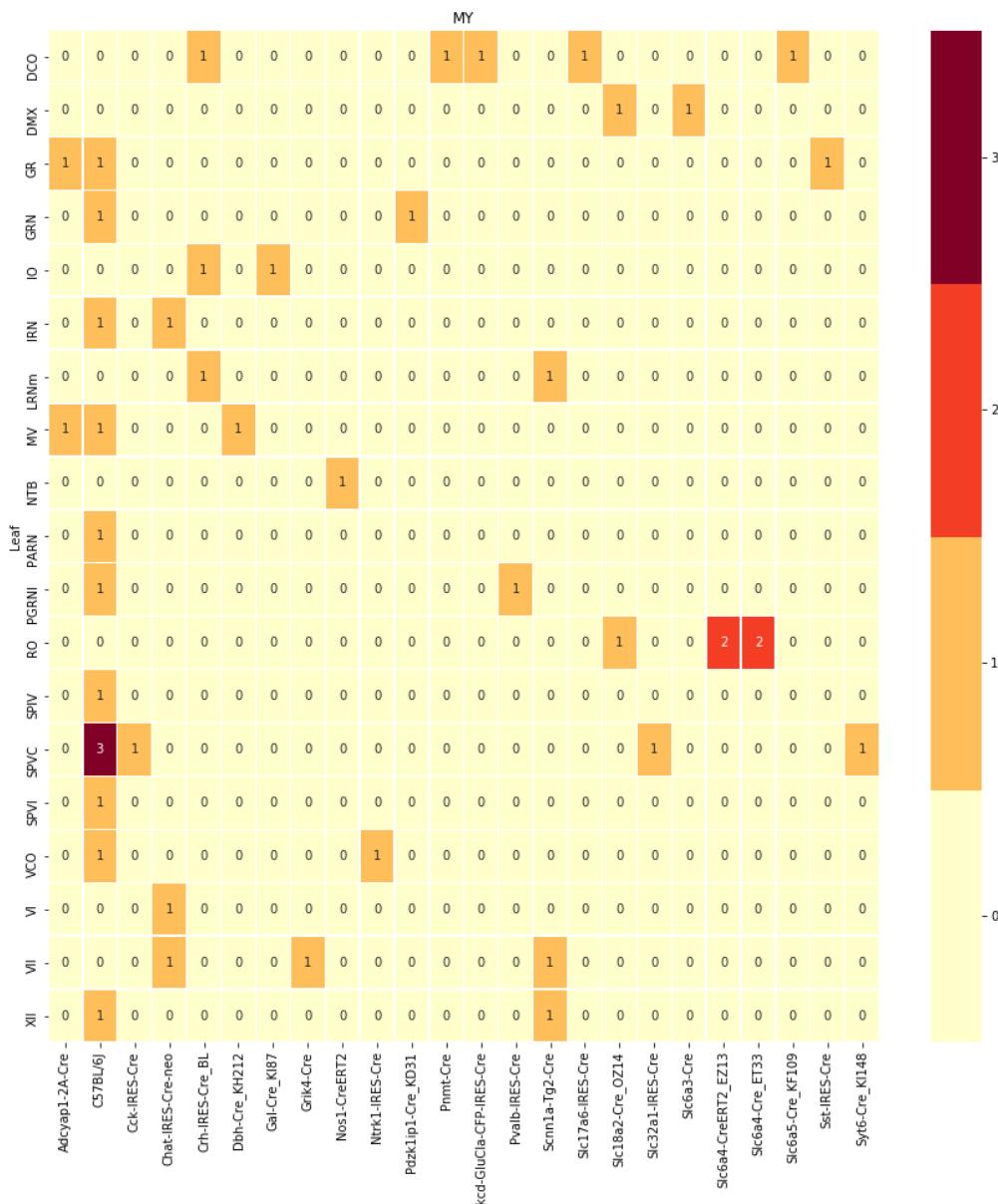
centroid densityoct12.png



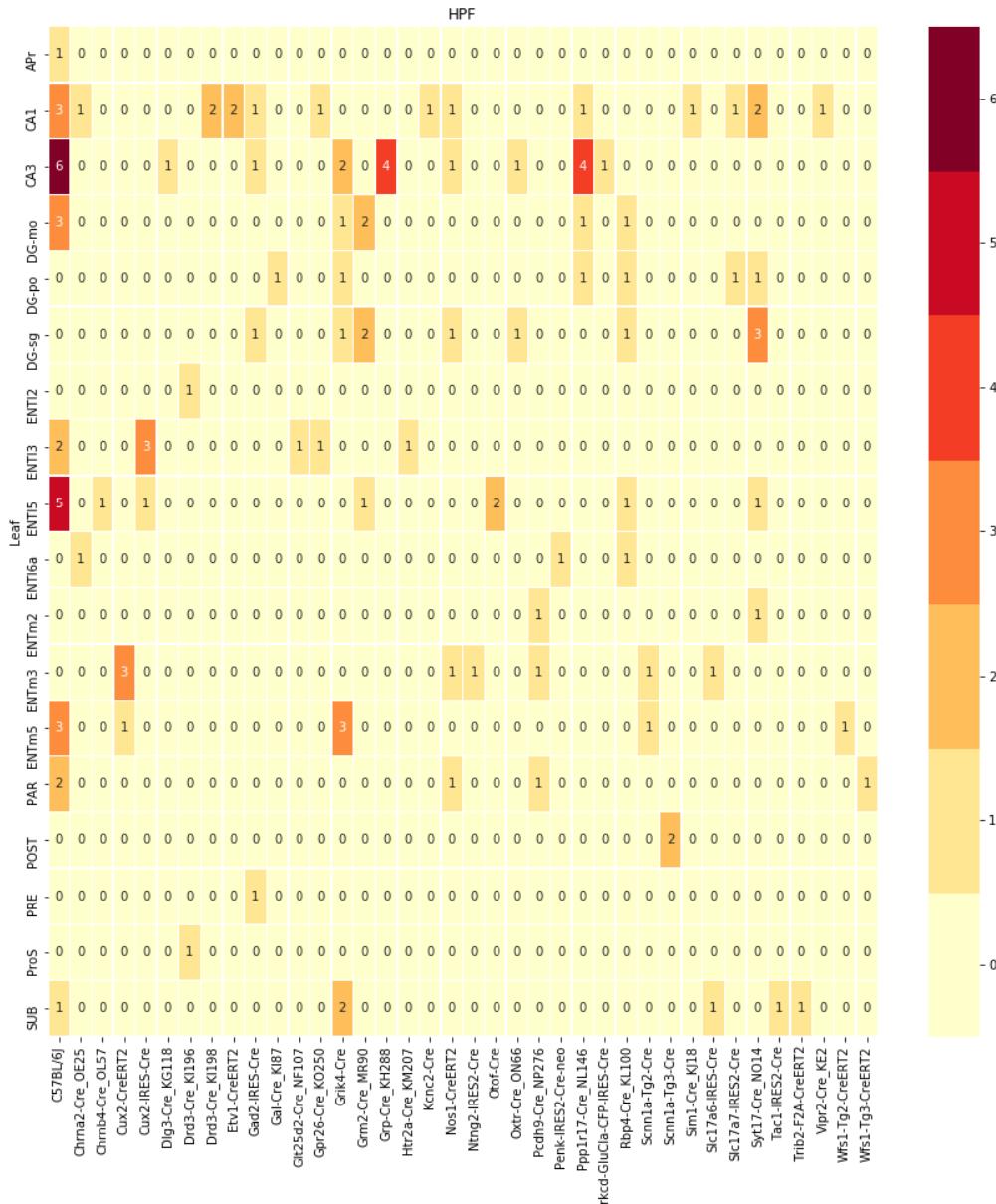
centroid densityoct12.png



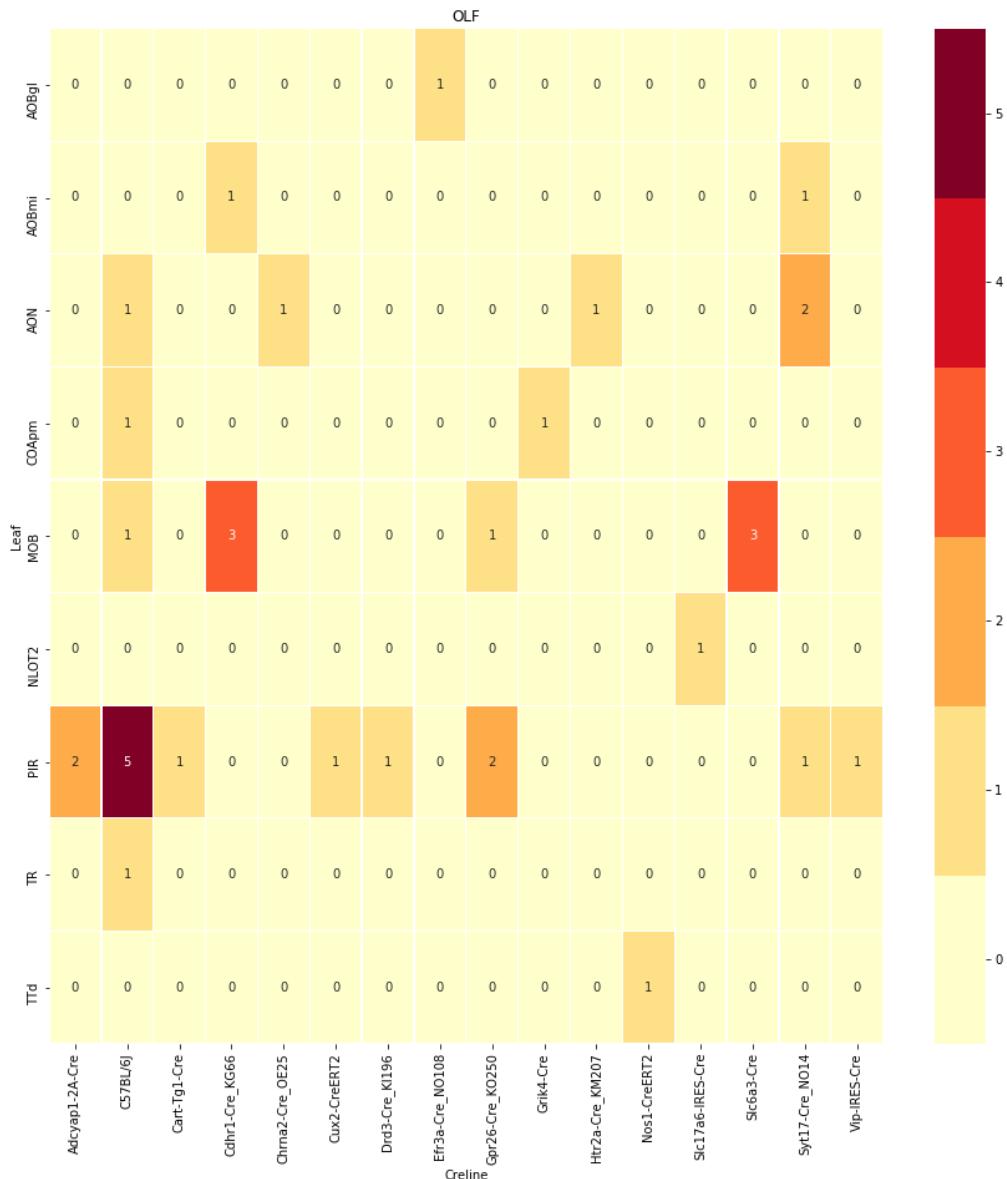
centroid densityoct12.png



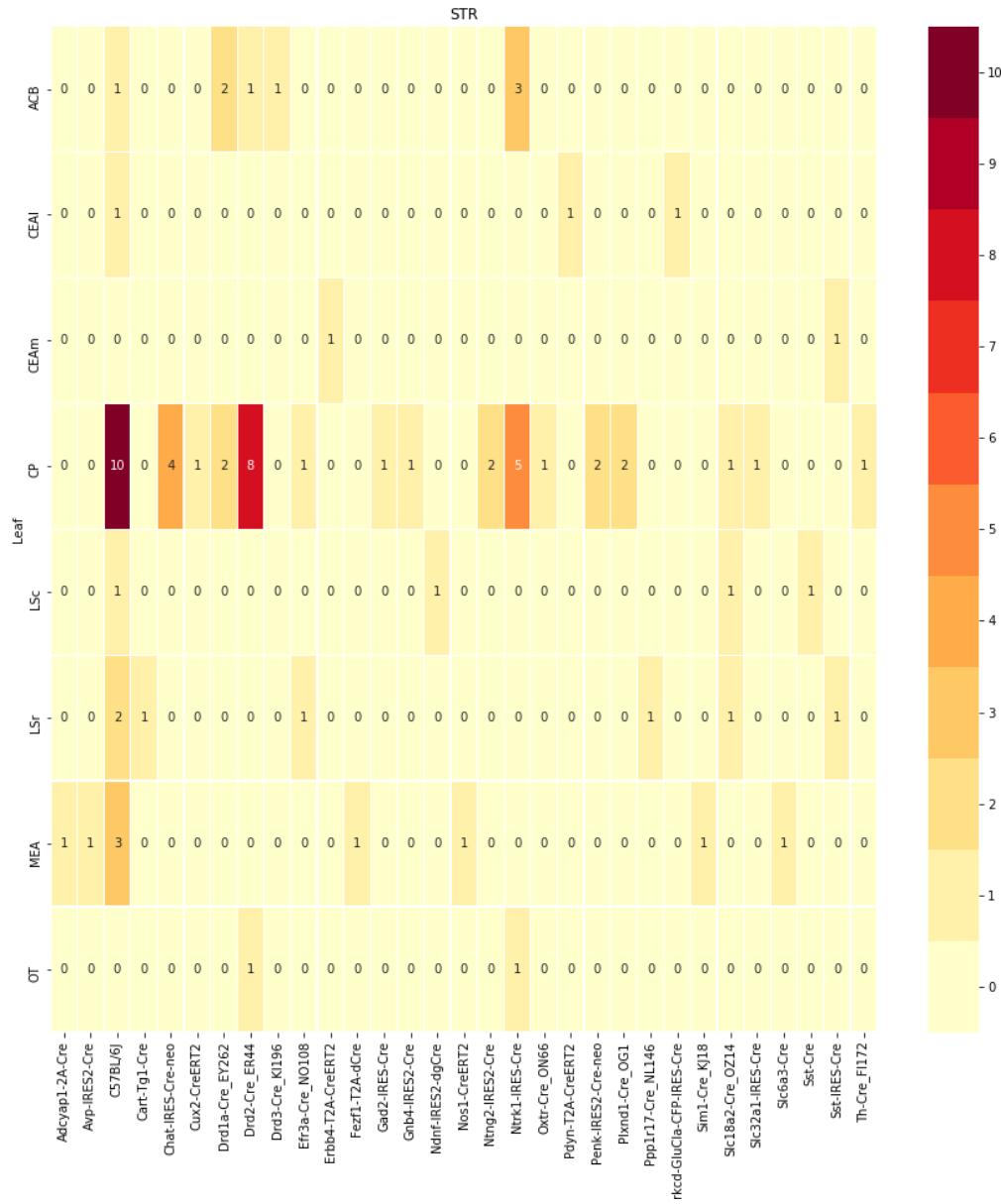
centroid densityoct12.png



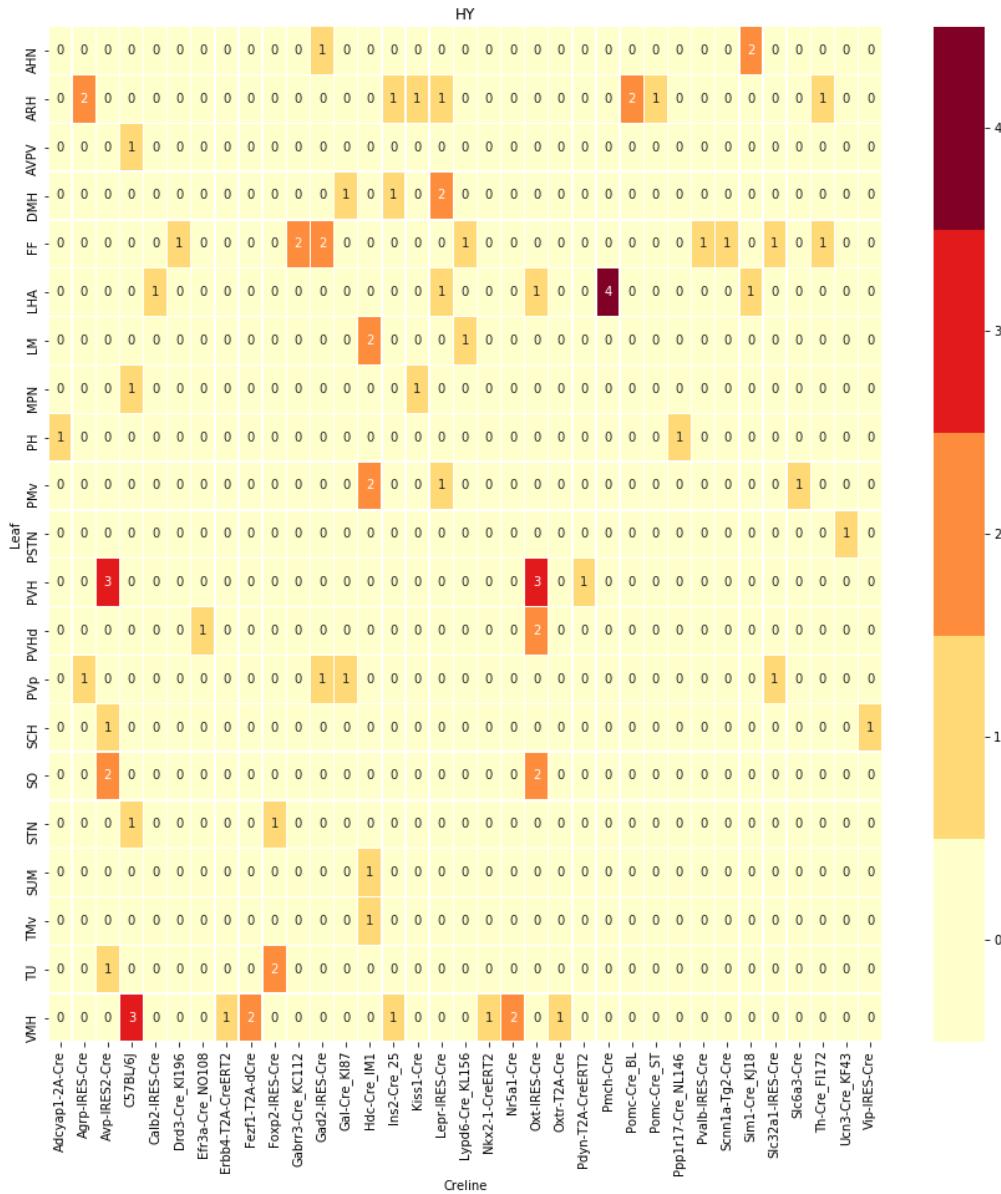
centroid densityoct12.png



centroid densityoct12.png



centroid densityoct12.png



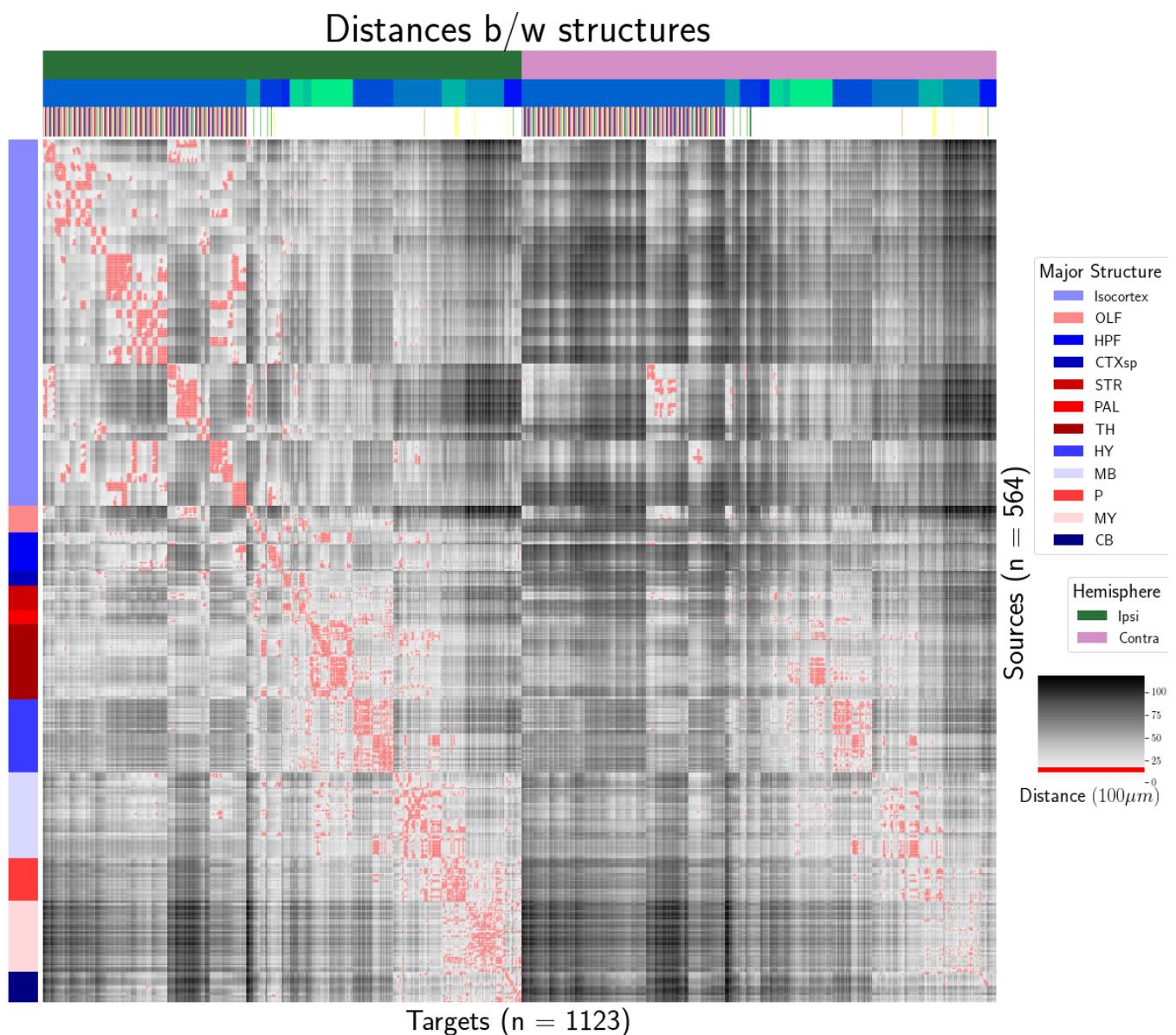
280 ***Distances between structures***

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

281 ***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

²⁸² This section consists of additional information on preprocessing of the neural connectivity data,
²⁸³ estimation of connectivity, and matrix factorization.

²⁸⁴ ***Data preprocessing***

²⁸⁵ Several data preprocessing steps take place prior to evaluations of the connectivity matrices. These
²⁸⁶ steps are described in Algorithm PREPROCESS. The arguments of this normalization process - injection
²⁸⁷ signals $x(i)$, projection signals $y(i)$, injection fraction $F(i)$, and data quality mask $q(i)$ - were
²⁸⁸ downloaded using the Allen SDK. The injections and projection signals $\mathcal{B} \rightarrow [0, 1]$ were segmented
²⁸⁹ manually in histological analysis. The projection signal gives the proportion of pixels within the voxel
²⁹⁰ displaying fluorescence, and the injection signal gives the proportion of pixels within the
²⁹¹ histologically-selected injection subset displaying fluorescence. The injection fraction $\mathcal{B} \rightarrow [0, 1]$ gives
²⁹² the proportion of pixels within each voxel in the injection subset. Finally, the data quality mask
²⁹³ $\mathcal{B} \rightarrow \{0, 1\}$ gives the voxels that have valid data.

²⁹⁴ Our preprocessing makes use of the above ingredients, as well as several other essential steps. First,
²⁹⁵ we compute the weighted injection centroid

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

²⁹⁶ where $x(i)(l)$ is the injection density at location $l \in \mathbb{R}^3$. Given a regionalization \mathcal{R} from the Allen SDK,
²⁹⁷ we can also access regionalization map $R: \mathcal{B} \rightarrow \mathcal{R}$. This induces a functional of connectivities from
²⁹⁸ the space of maps $\{\mathcal{X} = x: \mathcal{B} \rightarrow [0, 1]$

$$\begin{aligned} 1_{\mathcal{R}}: \mathcal{X} &\rightarrow \mathcal{R} \times \mathbb{R}_{\geq 0} \\ x &\mapsto \sum_{l \in r} x(l) \text{ for } r \in \mathcal{R}. \end{aligned}$$

²⁹⁹ We also can restrict a signal to a individual structure as

$$\begin{aligned} 1|_s: \mathcal{X} &\rightarrow \mathcal{X} \\ x(l) &= \begin{cases} x(l) & \text{if } l \in S \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

³⁰⁰ Finally, given a vector or array $a \in \mathbb{R}^T$, we have the $L1$ normalization map

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

³⁰¹ We define these objects as functions and functionals, but this is for notational convenience and
³⁰² non-essential. A function $x(i) : \mathcal{B} \rightarrow [0, 1]$ is mathematically equivalent to the graph
³⁰³ $\mathcal{G}(x(i)) \in \mathcal{B} \times [0, 1]$. As an abuse of notation, we define $x \odot x' := z$ such that $z(l) = x(l)x'(l)$ for all $l \in \mathcal{B}$.
³⁰⁴ Also, denote $m(i)$ as the major structure containing experiment i . We then can write the
³⁰⁵ preprocessing algorithm.

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 $y_q(i) \leftarrow \odot y(i) \odot q(i), x_q(i) \leftarrow x_F(i) \odot q(i)$

Restrict injection $x_m(i) = 1|_{m(i)} x_q(i)$.

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

Regionalize $\tilde{y}_{\mathcal{T}}(i) \leftarrow 1_{\mathcal{T}}(y_q(i))$

Normalize $y_{\mathcal{T}}(i) \leftarrow n(\tilde{y}_{\mathcal{T}}(i))$

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

306 **Estimators**

307 As mentioned previously, we can consider our estimators as modelling a connectivity vector
 308 $f_{\mathcal{T}}(\nu, s) \in \mathbb{R}_{\geq 0}^T$. Thus, for the remainder of this section, we will discuss only $f(\nu, s)$. We review the
 309 Nadaraya-Watson estimator from Knox et al. (2019), and describe its conversion into our cell-class
 310 specific Expected Loss estimator.

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

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

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

$$f^*(\hat{f}(f_*(\mathcal{D})))(\nu, s) = \sum_{l \in s} \hat{f}(f_*(\mathcal{D}(I)))(\nu, l).$$

316 Here, I is the training data, and \hat{f} is the Nadaraya-Watson estimator

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

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

319 Several facets of the estimator are visible here. A smaller γ corresponds to a greater amount of
 320 smoothing, and the index set $I \subseteq \{1 : n\}$ generally depends on s and ν . Fitting γ via empirical risk
 321 minimization therefore bridges between 1-nearest neighbor prediction and averaging of all
 322 experiments in I . In Knox et al. (2019), I consisted of experiments sharing the same brain division, i.e.
 323 $I = I_m$, while restricting of index set to only include experiments with the same cell class gives the
 324 class-specific Cre-NW model. Despite this restriction, we fit γ for each m rather than a smaller subset
 325 like s or ν . That is,

$$\hat{\gamma}_m = \arg \min_{\gamma \in \mathbb{R}_{\geq 0}} \frac{1}{|\{s, \nu\}|} \sum_{s, \nu \in \{m, \mathcal{V}\}} \frac{1}{|I_s \cap I_\nu|} \sum_{i \in (I_s \cap I_\nu)} \ell(y_{\mathcal{T}}(i)), \hat{f}_{\mathcal{T}}(f_*(\mathcal{D}(\nu, s) \setminus i)). \quad (2)$$

³²⁶ *The Expected-Loss estimator* Besides the injection location, the targeted cell class also influences
³²⁷ projection. Since Cre-lines that target similar classes are induce similar projections, and including
³²⁸ similar Cre-lines in the Nadaraya-Watson estimator increases effective sample size, we introduce an
³²⁹ estimator that assigns a predictive weight to each training point that depends both on its
³³⁰ centroid-distance and Cre-line. This weight is determined by the expected prediction error of each of
³³¹ the two feature types, as determined by cross-validation. For this reason, we call this the Expected
³³² Loss Estimator. The resulting weights are then utilized in a Nadaraya-Watson estimator in a final
³³³ prediction step.

³³⁴ We formalize Cre-line behavior as the average regionalized projection of a Cre-line in a given
³³⁵ structure (i.e. leaf). This vectorization of categorical information is known as **target encoding**, and we
³³⁶ define this as $\bar{y}_{\mathcal{T},s,v} := \frac{1}{|I_s \cap I_v|} \sum_{i \in (I_s \cap I_v)} y_{\mathcal{T}}(i)$. We define a **Cre-distance** in a leaf to be the distance
³³⁷ between the target-encoded projections of two Cre-lines. The relative predictive accuracy of
³³⁸ Cre-distance and centroid distance is determined by fitting a surface of projection distance as a
³³⁹ function of Cre-distance and centroid distance.

³⁴⁰ In mathematical terms, our full feature set consists of the centroid coordinates and the
³⁴¹ target-encoded means of the combinations of virus type and injection-centroid structure. That is,

$$f_*(\mathcal{D}_i) = \{c(i), \{\bar{y}_{\mathcal{T},s,v} \forall v\}, y_{\mathcal{T}}(i)\}.$$

³⁴² f^* is defined as in (2). The expected loss estimator is then

$$\hat{f}_{EL}(c(I), y_{\mathcal{T}}(I))(l, v) := \sum_{i \in I} \frac{v_{ilv}}{\sum_{i \in I} v_{ilv}} y_{\mathcal{T}}(i)$$

³⁴³ where

$$v_{ilv} := \exp(-\gamma g(d(l, c(i))^2, d(\bar{y}_{\mathcal{T},s,v}, \bar{y}_{\mathcal{T},s,v(i)})^2))$$

³⁴⁴ and s is the structure containing l .

³⁴⁵ The key step therefore is finding a suitable g with which to weight the positional and Cre
³⁴⁶ information. Note that g must be a concave, non-decreasing function of its arguments with with
³⁴⁷ $g(0, 0) = 0$, then g defines a metric on the product of the metric spaces defined by experiment centroid
³⁴⁸ and target-encoded cre-line, and \hat{f}_{EL} is a Nadaraya-Watson estimator. A derivation of this fact is given

³⁴⁹ later in this section, and we therefore use shape-constrained B-splines to estimate g . Similarly to the
³⁵⁰ Nadaraya-Watson model, we make the decision to fit a g separately for each major brain division. We
³⁵¹ can then select $\hat{\gamma}$ as in 2.

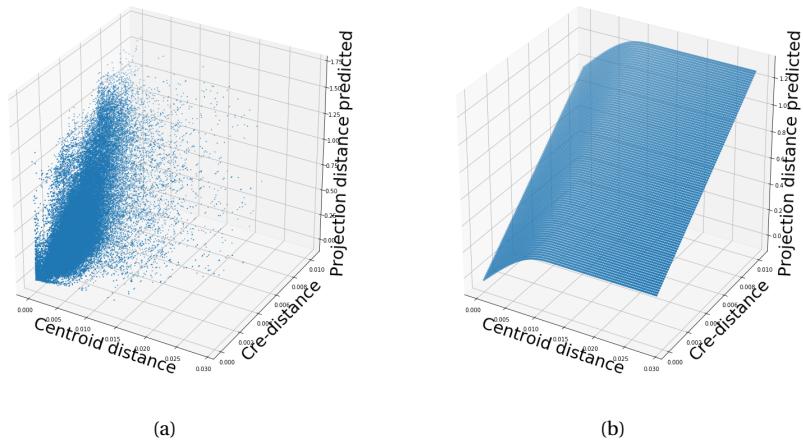


Figure 8: Fitting g . 8a Distribution of projection errors against centroid distance and cre-distance in Isocortex. 8b \hat{g} found using B-splines.

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

357 **Proposition 1.** *Given a collection of metric spaces X_1, \dots, X_n with metrics d_1, \dots, d_n (e.g. $d_{centroid}, d_{cre}$),
 358 and a 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 then f is a metric iff g is
 359 concave, non-decreasing and $g(d) = 0 \iff d = 0$.*

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

- 361 ▪ The first property of a metric is that $f(x, x') = 0 \iff x = x'$. The left implication:
 362 $x = x' \implies f(x_1, x'_1, \dots, x_n, x'_n) = g(0, \dots, 0)$, since d are metrics. Then, since $g(0) = 0$, we have that
 363 $f(x, x') = 0$. The right implication: $f(x, x') = 0 \implies d = 0 \implies x = x'$ since d are metrics.
- 364 ▪ The second property of a metric is that $f(x, x') = f(x', x)$. This follows immediately from the
 365 symmetry of the d_i , i.e. $f(x, x') = f(x_1, x'_1, \dots, x_n, x'_n) = g(d_1(x_1, x'_1), \dots, d_n(x_n, x'_n)) =$
 366 $g(d_1(x'_1, x_1), \dots, d_n(x'_n, x_n)) = f(x'_1, x_1, \dots, x'_n, x_n) = f(x', x)$.
- 367 ▪ The third property of a metric is the triangle inequality: $f(x, x') \leq f(x, x^*) + f(x^*, x')$. To show this
 368 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
 369 non-decreasing and by the triangle inequality of d . Then, since g is concave,
 370 $g(d(x, x^*) + d(x^*, x')) \leq g(d(x, x^*)) + g(d(x^*, x')) = f(x, x^*) + f(x^*, x')$.

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

- 372 ▪ The first property is that $g(d) = 0 \iff d = 0$. We first show the right implication: $g(d) = 0$, and
 373 $g(d) = f(x, x')$, so $x = x'$ (since f is a metric), so $d = 0$. We then show the left implication:
 374 $d = 0 \implies x = x'$, since d is a metric, so $f(x, x') = 0$, since f is a metric, and thus $g(d) = 0$.
- 375 ▪ The second property is that g is non-decreasing. We proceed by contradiction. Suppose g is
 376 decreasing in argument d_1 in some region $[l, u]$ with $0 < l < u$. Then
 377 $g(d_1(0, l), 0) \geq g(d_1(0, 0), 0) + g(d_1(0, u), 0) = g(d_1(0, u), 0)$, which violates the triangle inequality on
 378 f . Thus, decreasing g means that f is not a metric, so f a metric implies non-decreasing g .

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

383

□

384 Establishing a lower detection limit

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

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

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

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

396 ***Decomposing the connectivity matrix***

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

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

$$\text{NMF}(\mathcal{C}, \lambda, q, 1_M) = \arg \min_{W \in \mathbb{R}_{\geq 0}^{S \times q}, H \in \mathbb{R}_{\geq 0}^{q \times T}} \frac{1}{2} \|1_M \odot \mathcal{C} - WH\|_2^2 + \lambda(\|H\|_1 + \|W\|_1).$$

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

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

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

417 The differentiating feature of cross-validation for NMF compared with supervised learning is the
 418 randomness of the masking matrix 1_M . Cross-validation for supervised learning generally leaves out
 419 entire observations, but this is insufficient for our situation. This is because, given W , our H is the

420 solution of a regularized non-negative least squares optimization problem

$$H := \widehat{e}_W(1_M \odot \mathcal{C}) = \arg \min_{\beta \in \mathbb{R}_{\geq 0}^{q \times T}} \|1_M \odot \mathcal{C} - W\beta\|_2^2 + \|\beta\|_1. \quad (3)$$

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

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

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

NMF is then performed using the mask $1_{M(p)}$ to get W . The cross-validation error is then

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

where $1_{M(p)_r}^C$ is the binary complement of $1_{M(p)_r}$ and R is a number of replicates. Theoretically, the optimum number of components is then

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

422 *Stabilizing NMF* The NMF program is non-convex, and, empirically, individual replicates will not
 423 converge to the same optima. One solution therefore is to run multiple replicates of the NMF
 424 algorithm and cluster the resulting vectors. This approach raises the questions of how many clusters
 425 to use, and how to deal with stochasticity in the clustering algorithm itself. We address this issue
 426 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

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

427 A review of this approach is found in von Luxburg (2010b). Intuitively, archetype vectors that cluster
 428 together frequently over clustering replicates indicate the presence of a stable clustering. For d , we
 429 utilize the adjusted Rand Index - a simple dissimilarity measure between clusterings. Note that we

430 expect to select slightly more than the q components suggested by cross-validation, since archetype
431 vectors which appear in one NMF replicate generally should appear in others. We then select the q
432 clusters with the most archetype vectors - the most stable NMF results - and take the median of each
433 cluster to create a sparse representative archetype Kotliar et al. (2019); Wu et al. (2016). We then find
434 the according H using Program 3. Experimental results for these cross-validation and stability
435 selection approaches are given in Supplemental Section 7.

7 SUPPLEMENTAL EXPERIMENTS

436 *Establishing a lower limit of detection*

437 We give results on the false detection rate at different limits of detection. These conclusively show that

438 10^{-6} is the good threshold for our normalized data.

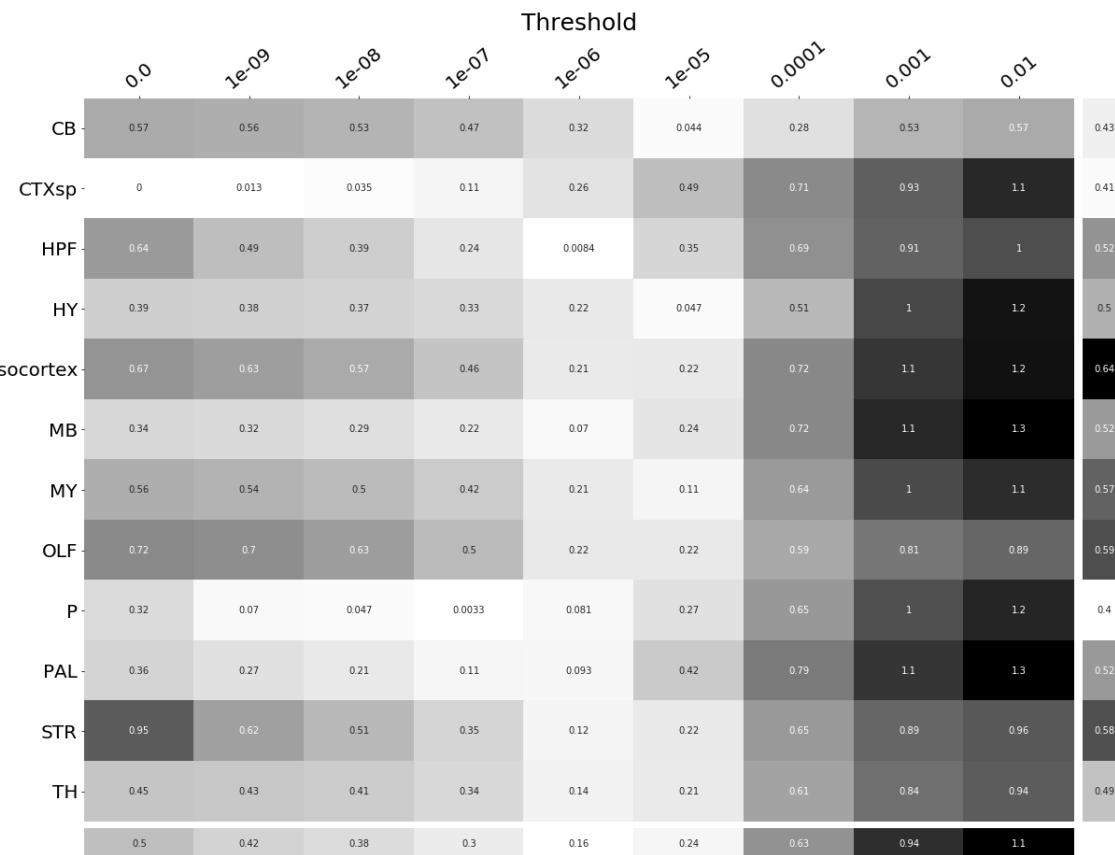


Figure 9: τ at different limits of detection in different major structures. 10^{-6} is clearly the optimal detection threshold.

439 **Loss subsets**

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



Figure 10

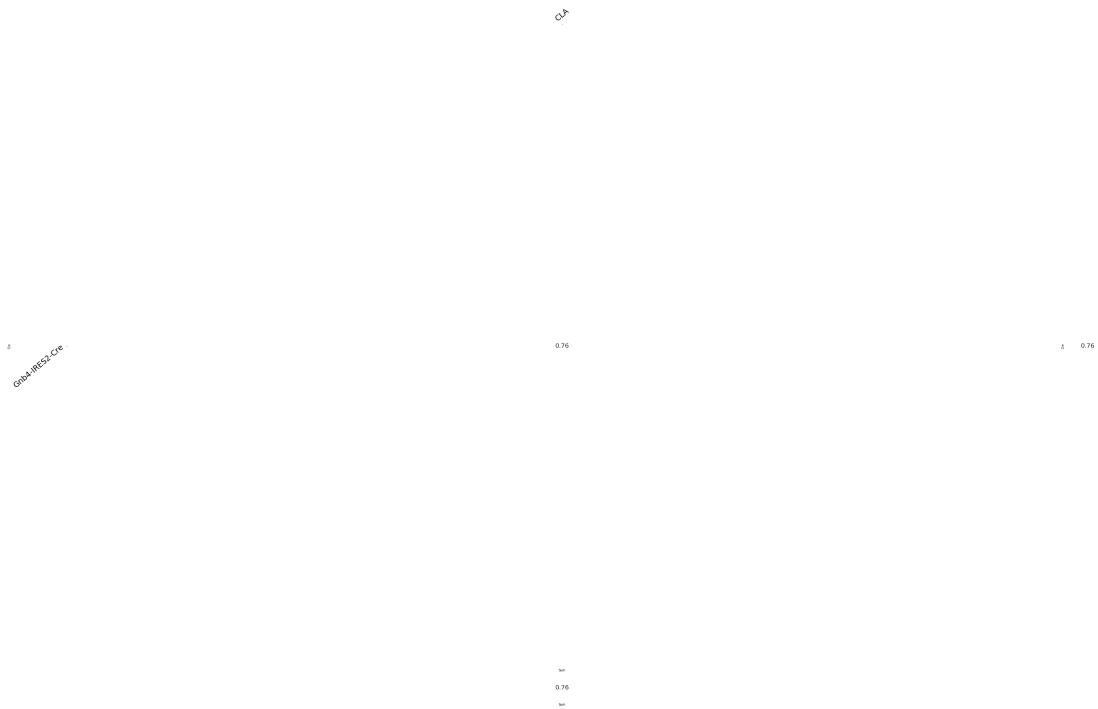


Figure 11

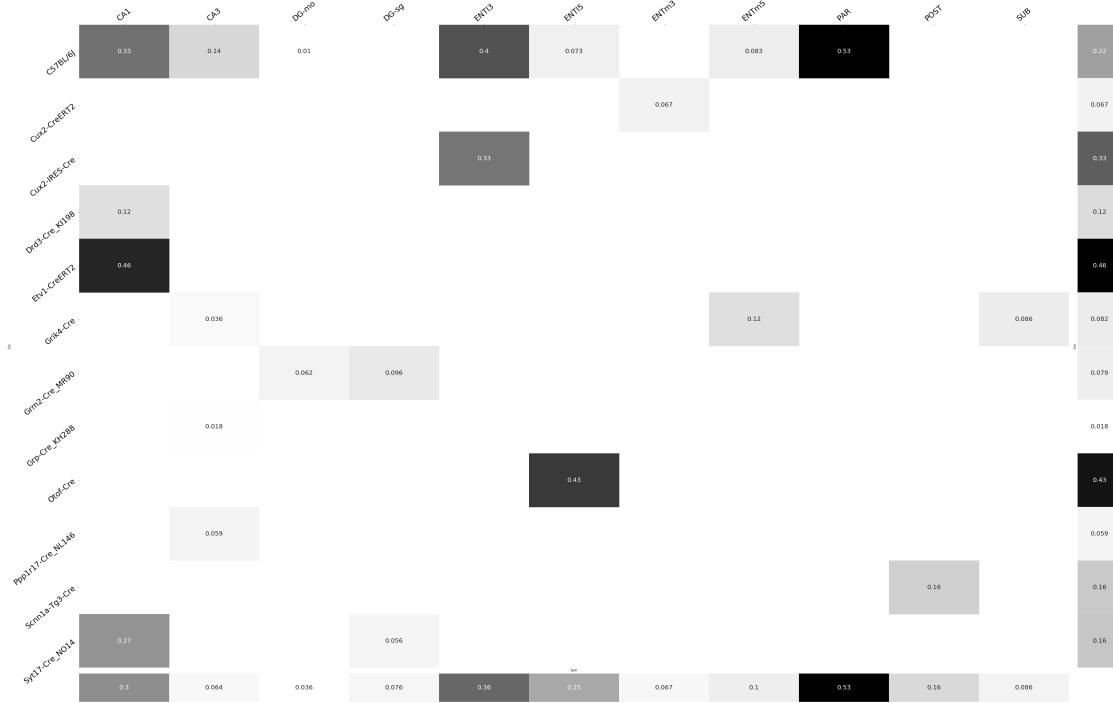


Figure 12

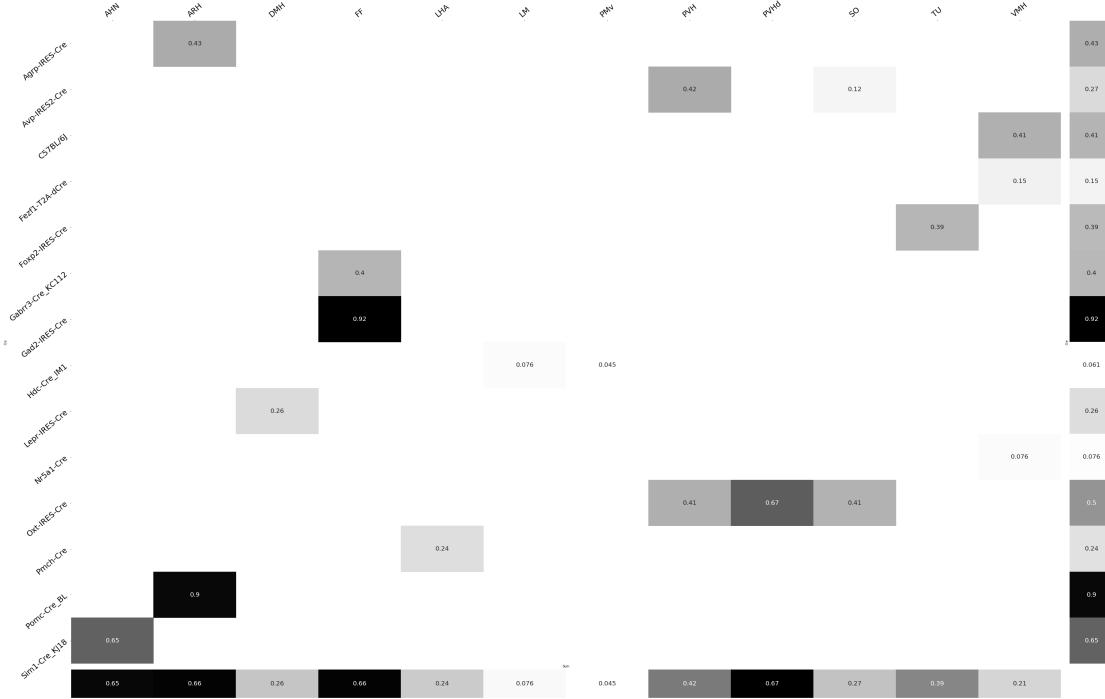


Figure 13

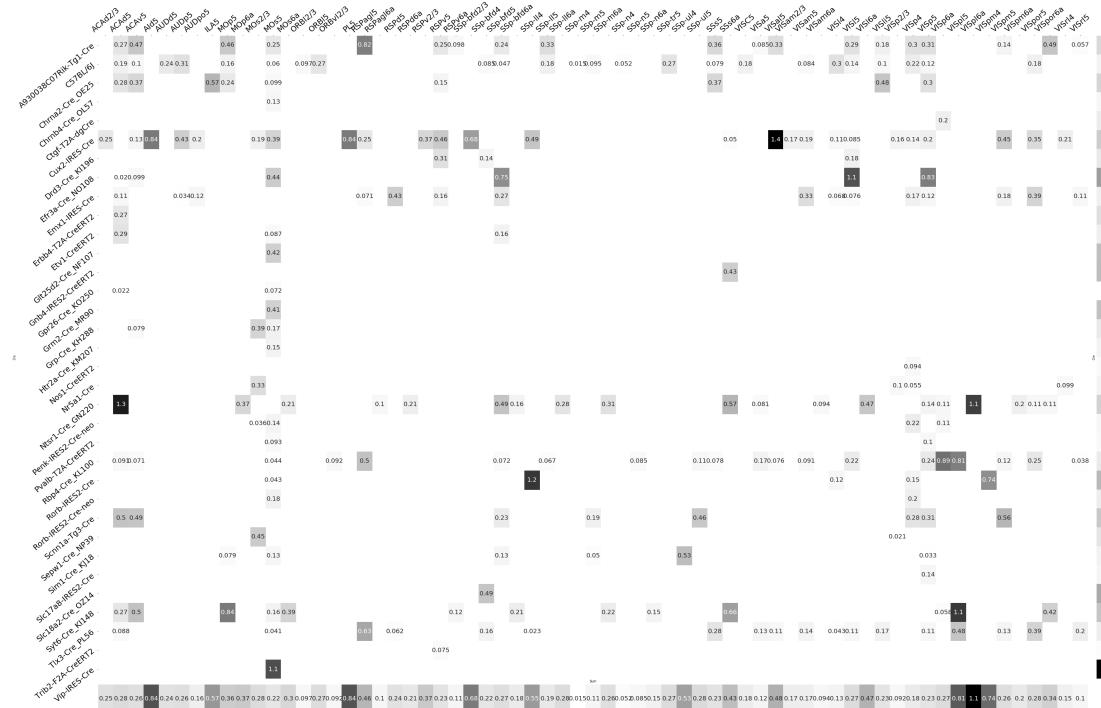


Figure 14



Figure 15



Figure 16



Figure 17

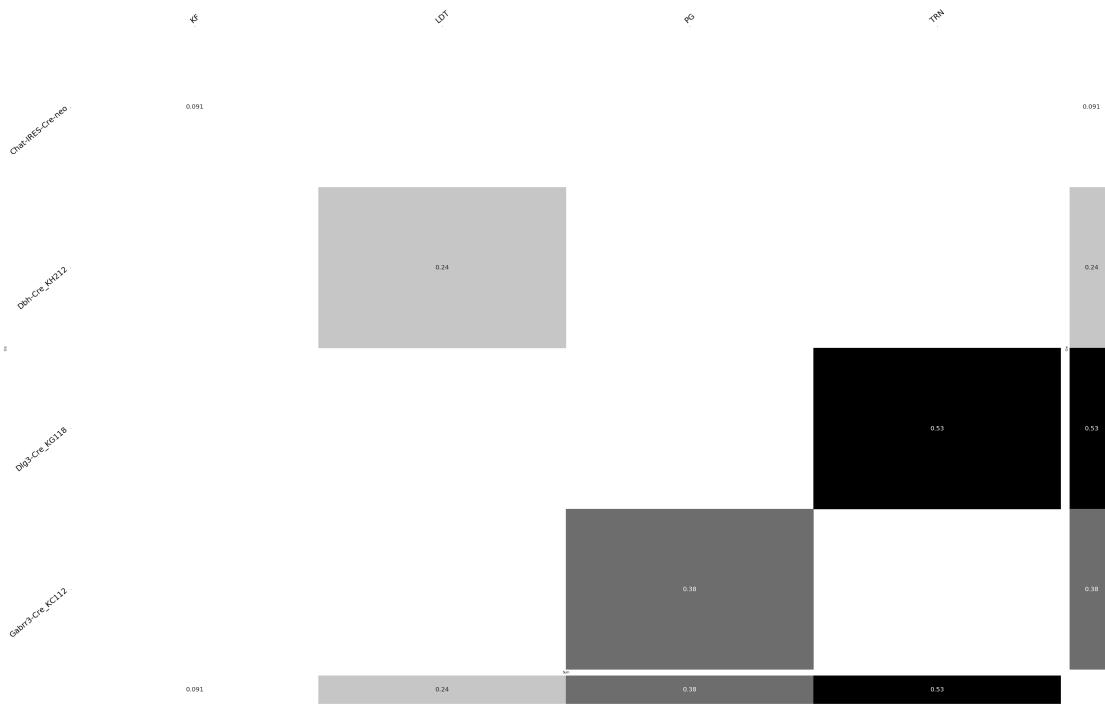


Figure 18

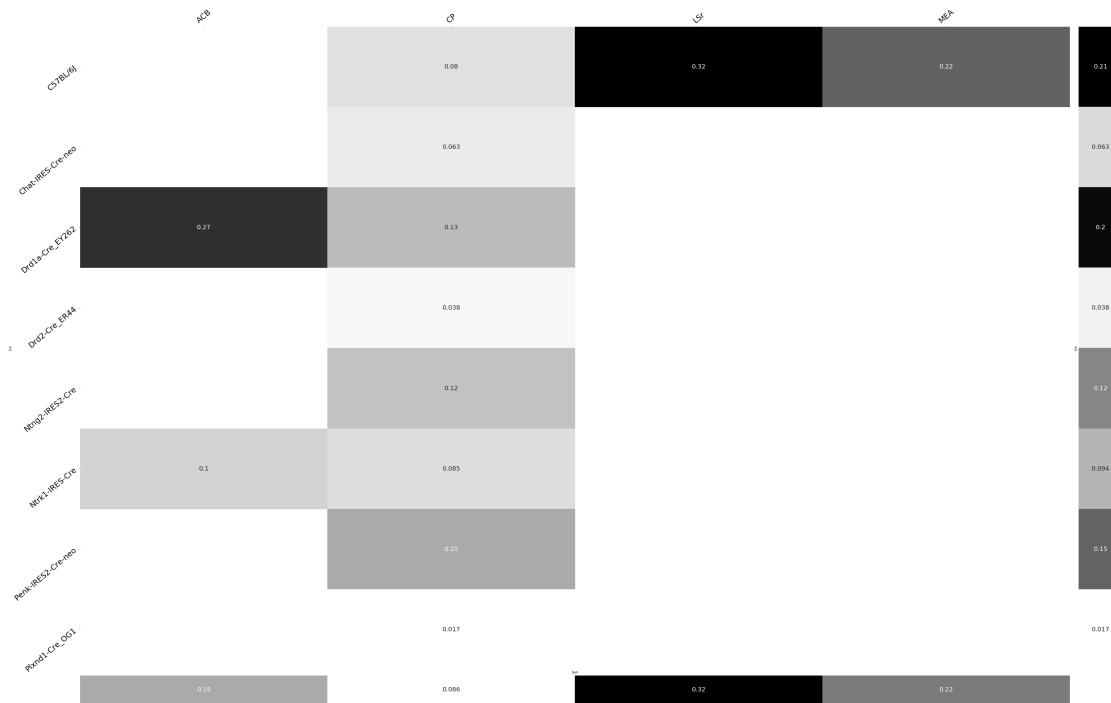


Figure 19

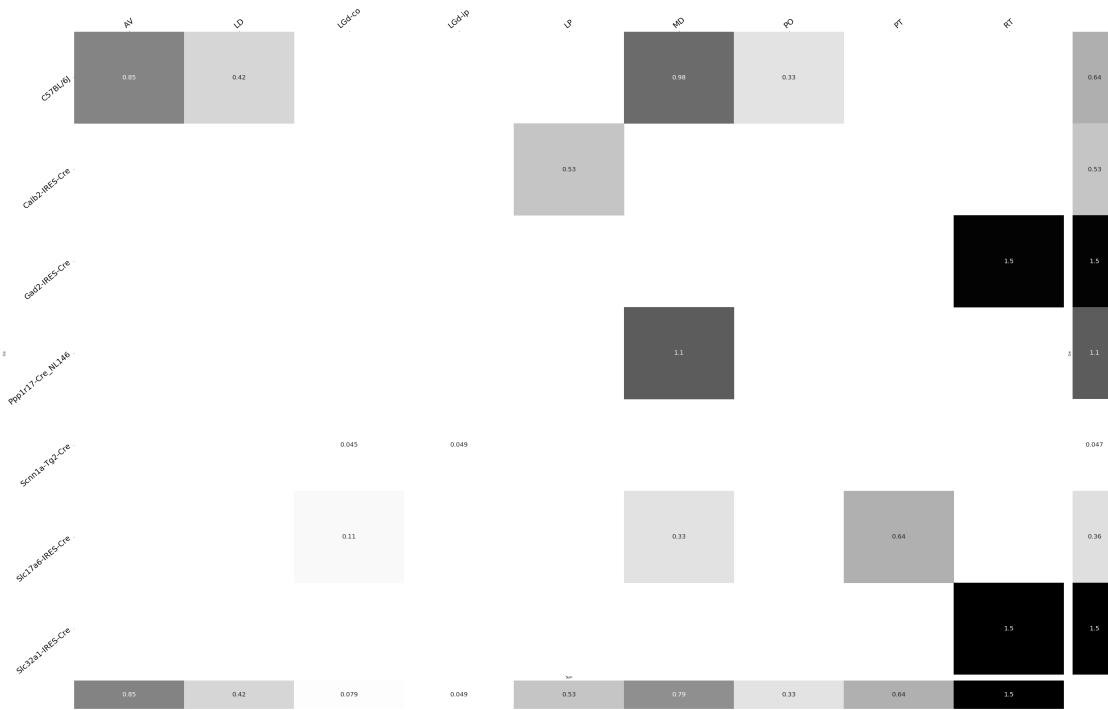


Figure 20

442 **Matrix Factorization**

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

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

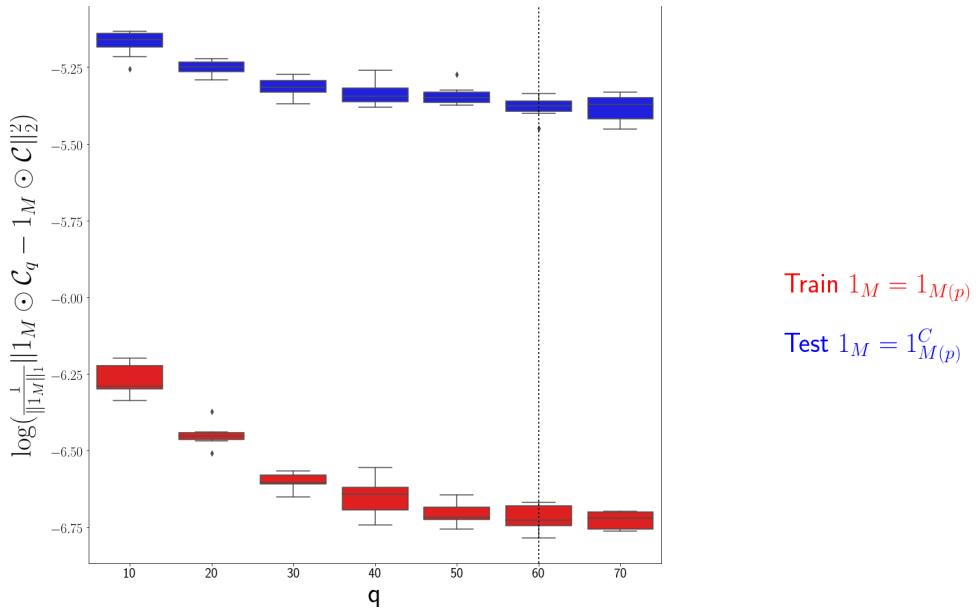


Figure 21: Train and test error using NMF decomposition.

451 *Stability* For the purposes of visualization and interpretability, we restrict to a $q = 15$ component
 452 model. To address the instability of the NMF algorithm in identifying components, we $k - \text{means}$
 453 cluster components over $R = 10$ replicates with $k \in \{10, 15, 20, 25, 30\}$. Since the clustering is itself
 454 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

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

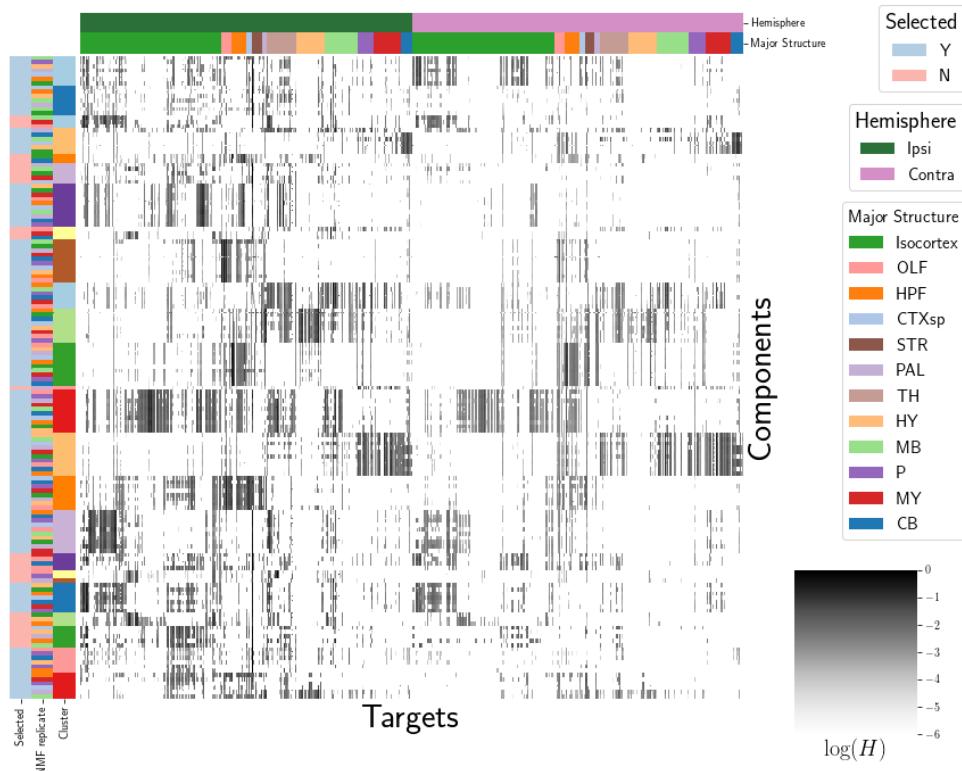


Figure 22: Stability of NMF results across replicates. Replicate and NMF component are shown on rows. Components that are in the top 15 are also indicated.

457

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

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

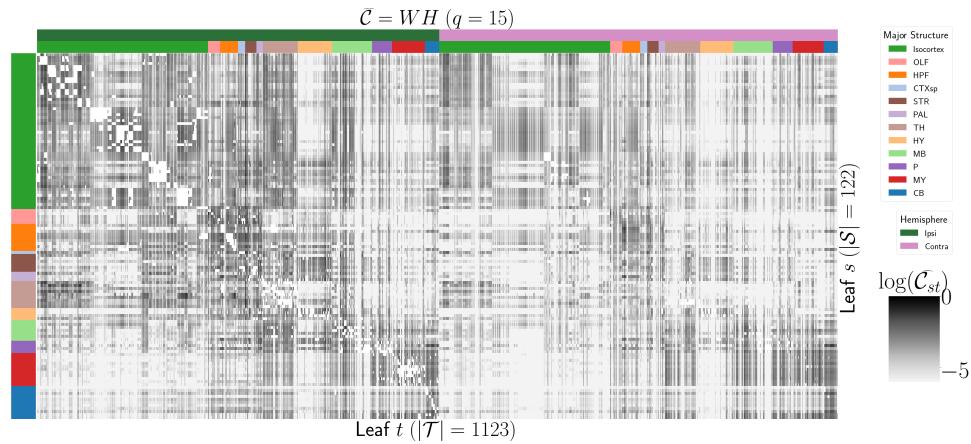


Figure 23: Reconstruction of \mathcal{C} from H and W with $q = 15$ in Figure 4.

8 GLOSSARY OF SYMBOLS

Symbol	Mathematical setting	Meaning
\mathcal{D}	$[1:n]$	Dataset
n		Number of data points
f	$\mathcal{V} \times \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}_{\geq 0}$	
$x(i)$	$\mathcal{B} \rightarrow [0, 1]$	Injection
$y(i)$	$\mathcal{B} \rightarrow [0, 1]$	Projection
\mathcal{C}	$\mathcal{V} \times \mathcal{S} \times \mathcal{T} \rightarrow \mathbb{R}_{\geq 0}$	structural connection strength
\mathcal{C}^D	$\mathcal{V} \times \mathcal{S} \times \mathcal{T} \rightarrow \mathbb{R}_{\geq 0}$	normalized structural connectivity strength
\mathcal{C}^S	$\mathcal{V} \times \mathcal{S} \times \mathcal{T} \rightarrow \mathbb{R}_{\geq 0}$	normalized structural connectivity density
\mathcal{C}_v	$\in \mathbb{R}^{S \times T}$	Structural connection matrix for Cre-line v .
$y_{\mathcal{T}}(i)$	$\in \mathbb{R}_{\geq 0}^T$	Regionalized projection
I_m	$\subset [1 : n]$	Experiments in major subdivision
I_s	$\subset [1 : n]$	Experiments in structure
I_v	$\subset [1 : n]$	Experiments with Cre-line v
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	

9 COMPETING INTERESTS

- ⁴⁶⁰ This is an optional section. If you declared a conflict of interest when you submitted your manuscript,
⁴⁶¹ please use this space to provide details about this conflict.

462

463

REFERENCES

464

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

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

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

471 Gao, Y., Zhang, X., Wang, S., & Zou, G. (2016). Model averaging based on leave-subject-out cross-validation. *J. Econom.*,
 472 192(1), 139–151.

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

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

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

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

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

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

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

- 487 Oh, S. W., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., ... Zeng, H. (2014). A mesoscale connectome of the
488 mouse brain. *Nature*, 508(7495), 207–214.
- 489 Saul, L. K., & Roweis, S. T. (2003). Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *J.
490 Mach. Learn. Res.*, 4(Jun), 119–155.
- 491 von Luxburg, U. (2010a). Clustering stability: An overview.
- 492 von Luxburg, U. (2010b). Clustering stability: An overview.
- 493 Wu, S., Joseph, A., Hammonds, A. S., Celtniker, S. E., Yu, B., & Frise, E. (2016). Stability-driven nonnegative matrix
494 factorization to interpret spatial gene expression and build local gene networks. *Proc. Natl. Acad. Sci. U. S. A.*, 113(16),
495 4290–4295.

10 TECHNICAL TERMS

⁴⁹⁶ **Technical Term** a key term that is mentioned in an NETN article and whose usage and definition may
⁴⁹⁷ not be familiar across the broad readership of the journal.

⁴⁹⁸ **Cre-line** Refers to the combination of cre-recombinase expression in transgenic mouse and
⁴⁹⁹ cre-induced promotion in the vector that induces labelling of cell-class specific projection.

⁵⁰⁰ **Cell class** The projecting neurons targeted by a particular cre-line

⁵⁰¹ **structural connectivities** connectivity between structures

⁵⁰² **Voxel** A $100\mu m$ cube of brain.

⁵⁰³ **structural connection tensor** Connectivities between structures given a neuron class

⁵⁰⁴ **dictionary-learning** A family of algorithms for finding low-dimensional data representations.

⁵⁰⁵ **shape constrained estimator** A statistical estimator that fits a function of a particular shape (e.g.
⁵⁰⁶ monotonic increasing, convex).

⁵⁰⁷ **Nadaraya-Watson** A simple smoothing estimator.

⁵⁰⁸ **connectivity archetypes** Typical connectivity patterns

⁵⁰⁹ **Expected Loss** Our new estimator that weights different features by their estimated predictive
⁵¹⁰ power.