

**1 RESEARCH**

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

**4 Samson Koelle** <sup>1,2</sup>, **Dana Mastrovito** <sup>1</sup>, **Jennifer D Whitesell** <sup>1</sup>, **Karla E Hirokawa** <sup>1</sup>, **Hongkui Zeng** <sup>1</sup>, **Marina Meila** <sup>2</sup>, **Julie A**  
**5 Harris** <sup>1</sup>, **Stefan Mihalas** <sup>1</sup>

**6** <sup>1</sup> Allen Institute for Brain Science, Seattle, WA, USA

**7** <sup>2</sup>Department of Statistics, University of Washington, Seattle, WA, USA

**8** **Keywords:** [Connectivity, Cell-type, Mouse]

**ABSTRACT**

**9** The Allen Mouse Brain Connectivity Atlas (MCA) consists of anterograde tracing experiments  
**10** targeting diverse structures and classes of projecting neurons. Beyond regional anterograde tracing  
**11** done in C57BL/6 wild-type mice, a large fraction of experiments are performed using transgenic  
**12** Cre-lines. This allows access to cell-class specific whole brain connectivity information, with class  
**13** defined by the transgenic lines. However, even though the number of experiments is large, it does not  
**14** come close to covering all existing cell classes in every area where they exist. Here, we study how  
**15** much we can fill in these gaps and estimate the cell-class specific connectivity function given the  
**16** simplifying assumptions that nearby voxels have smoothly varying projections, but that these  
**17** projection tensors can change sharply depending on the region and class of the projecting cells.

**18** This paper describes the conversion of Cre-line tracer experiments into class-specific connectivity  
**19** matrices representing the connection strengths between source and target structures. We introduce  
**20** and validate a novel statistical model for creation of connectivity matrices. We extend the

21 Nadaraya-Watson kernel learning method which we previously used to fill in spatial gaps to also fill in  
22 gaps in cell-class connectivity information. To do this, we construct a "cell-class space" based on  
23 class-specific averaged regionalized projections and combine smoothing in 3D space as well as in this  
24 abstract space to share information between similar neuron classes. Using this method we construct a  
25 set of connectivity matrices using multiple levels of resolution at which discontinuities in connectivity  
26 are assumed. We show that the connectivities obtained from this model display expected cell-type  
27 and structure specific connectivities. We also show that the wild-type connectivity matrix can be  
28 factored using a sparse set of factors, and analyze the informativeness of this latent variable model.

## AUTHOR SUMMARY

29 While standard connectivity studies in mammalian models focus on the overall connection strength  
30 between areas, there is an increasing focus on specific connection types between cell classes when  
31 describing functions at a local circuit level. Having recently described the importance of such classes  
32 in the cortico-thalamic system, we now investigate their importance for estimating brain-wide  
33 mesoscopic connectivity. Even within our relatively large dataset, the connectivity data across cell  
34 classes is sparse, and so we introduce a method to more reliably extrapolate across classes and  
35 estimate connection-type specific inter-areal connectivity. We observe that this complex connectivity  
36 may be described via a relatively small set of factors. While not complete, this connectivity matrix  
37 represents a categorical and quantitative improvement in mouse mesoscale connectivity models.

## 1 INTRODUCTION

38 The mammalian nervous system enables an extraordinary range of natural behaviors, and has  
39 inspired much of modern artificial intelligence. Neural connections including those from one region  
40 to another form the architecture underlying this capability. These connectivities vary by neuron type,  
41 as well as source (cell body) location and target (axonal projection) structures. Thus, characterization  
42 of the relationship between neuron type and source and target structure is important for  
43 understanding the overall nervous system.

44 Viral tracing experiments - in which a viral vector expressing GFP is transduced into neural cells  
45 through stereotaxic injection - are a useful tool for mapping these connections on the mesoscale  
46 (Chamberlin, Du, de Lacalle, & Saper, 1998; Daigle et al., 2018; J. A. Harris, Oh, & Zeng, 2012). The long  
47 range connections between different areas are generally formed by axons which travel from one  
48 region to another, and the GFP protein moves into the axon of the projecting neurons. Two-photon  
49 tomography imaging can be used to determine the location and strength of the fluorescent signals in  
50 two-dimensional slices. These locations can then be mapped back into three-dimensional space, and  
51 the signal may then be integrated over area into cubic voxels to give a finely-quantized  
52 three-dimensional fluorescence.

53 Several statistical models for the conversion of such experiment-specific signals into generalized  
54 estimates of connectivity strength have been proposed (Gămănuț et al., 2018; K. D. Harris, Mihalas, &  
55 Shea-Brown, 2016; Knox et al., 2019; Oh et al., 2014). Of these, Oh et al. (2014) and Knox et al. (2019)  
56 provide a model for **regionalized connectivities**, which are voxel connectivities integrated by region.  
57 The value of these models is that they provide some improvement over simply averaging the  
58 projection signals of injections in a given region. However, these previous works only model  
59 connectivities observed in wild-type mice which are suboptimally suited to assessment of cell-type  
60 specific connectivity compared with fluorescence from Cre-recombinase induced eGFP expression in  
61 cell-types specified by the combination of transgenic mouse strain and transgene promoter  
62 (J. A. Harris et al., 2019). We generally refer to sets of so-targeted eGFP-expressing cells in tracing  
63 experiments as a **cell class** since they may contain multiple types. For example, use of both wild-type  
64 and transgenic mice would give rise to cell-class specific experiments, albeit with different yet  
65 perhaps overlapping classes of cells.

66 Thus, this paper introduces a class-specific statistical model for anterograde tracing experiments  
67 that synthesizes the diverse set of **Cre-lines** described in J. A. Harris et al. (2019), and expands this  
68 model to the entire mouse brain. Our model is a to-our-knowledge novel statistical estimator that  
69 takes into account both the spatial position of the labelled source, as well as the categorical cell class.  
70 Like the previously state-of-the-art model in Knox et al. (2019), this model predicts regionalized  
71 connectivity as an average over positions within the structure, with nearby experiments given more  
72 weight. However, our model weighs class-specific behavior in a particular structure against spatial  
73 position, so a nearby experiment specific to a similar cell-class is relatively up-weighted, while a  
74 nearby experiment specific to a dissimilar class is down-weighted. This model outperforms the model  
75 of Knox et al. (2019) based on its ability to predict held-out experiments in leave-one-out  
76 cross-validation. We use the trained model to estimate overall connectivity matrices for each assayed  
77 cell class.

78 The resulting cell-class specific connectivity is a directed weighted multigraph which can be  
79 represented as a tensor with missing values. We do not give an exhaustive analysis of this data, but do  
80 establish a lower-limit of detection, verify several cell-type specific connectivity patterns found  
81 elsewhere in the literature, and show that these cell-type specific signals are behaving in expected  
82 ways. We also decompose the wild-type connectivity matrix into factors representing latent  
83 connectivity patterns, which we call archetypes. These components allow approximation of the  
84 regionalized connectivity using linear combinations of a small set of components.

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

## 2 METHODS

88 We estimate and analyze cell class-specific connectivity functions using models trained on murine  
89 brain viral tracing experiments. This section describes the data used to generate the model, the model  
90 itself, the evaluation of the model against its alternatives, and the use of the model in creation of the  
91 connectivity estimate matrices. It also includes background on the non-negative matrix factorization  
92 method used for decomposing the wild type connectivity matrix into latent factors. Additional  
93 information about our data and methods are given in Supplemental Sections 5 and 6, respectively.  
94

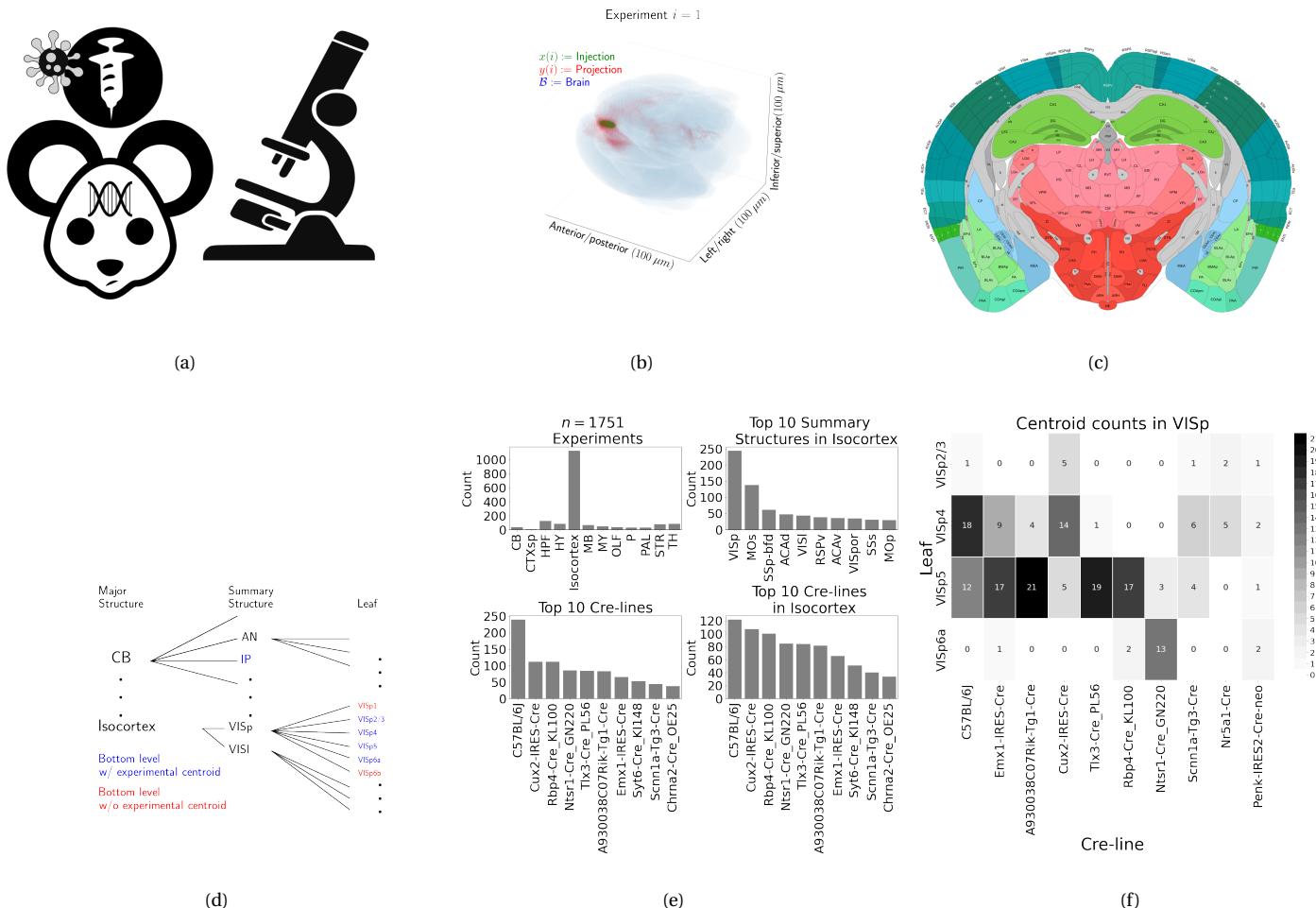


Figure 1: Experimental setting. 1a For each experiment, a Cre-dependent eGFP-expressing transgene cassette is transduced by stereotaxic injection into a Cre-driver mouse, followed by serial two-photon tomography imaging. 1b An example of the segmentation of projection (targets) and injection (source) for a single experiment. Within each brain (blue), injection (green) and projection (red) areas are determined via histological analysis and alignment to the Allen Common Coordinate Framework (CCF). 1c Brain region parcellations within a coronal plane of CCFv3. 1d Explanation of nested structural ontology highlighting various levels of CCFv3 structure ontology. Lowest-level (leaf) structures are colored in blue, and structures without an injection centroid are colored in red. 1e Abundances of tracer experiments by Cre-line and region of injection. 1f Co-occurrence of layer-specific centroids and Cre-lines within VISp.

95 **Data**

96 Our dataset  $\mathcal{D}$  consists of  $n = 1751$  publicly available murine brain viral tracing experiments from the  
 97 Allen Mouse Brain Connectivity Atlas. Figure 1a summarizes the experimental process used to  
 98 generate this data. In each experiment, a mouse is injected with an adeno-associated virus (AAV)  
 99 encoding green fluorescent protein (GFP) into a single location in the brain. Location of fluorescence  
 100 is mediated by the location of the injection, the characteristics of the transgene, and the genotype of  
 101 the mouse. In particular, Cre-driver or, equivalently, Cre-line mice are engineered to express Cre  
 102 under the control of a specific and single gene promoter. This localizes expression of Cre to regions  
 103 with certain transcriptomic cell-types signatures. In such Cre-driver mice, we used a double-inverted  
 104 floxed AAV to produce eGFP fluorescence that depends on Cre expression in infected cells. To account  
 105 for the complex cell-type targeting induced by a particular combination of Cre-driver genotype and  
 106 GFP promoter, we refer to the combinations of cell-types targeted by a particular combination of AAV  
 107 and Cre-driver mice as cell-classes. For example, we include experiments from Cre-driver lines that  
 108 selectively label cell classes located in distinct cortical layers or other nuclei across the whole brain.  
 109 For injections in the wild type mice, we used the Synapsin I promoter (Jackson, Dayton, Deverman, &  
 110 Klein, 2016; Kügler, Kilic, & Bähr, 2003). For injections into Cre mice, we used the CAG promoter with  
 111 a Flex cassette for Cre-mediated recombination control (Saunders, Johnson, & Sabatini, 2012).  
 112 Additional details on are given in J. A. Harris et al. (2019).

113 For each experiment, the fluorescent signal imaged after injection is aligned into the Allen  
 114 Common Coordinate Framework (CCF) v3, a three-dimensional average template brain that is fully  
 115 annotated with regional parcellations (Wang et al., 2020). The whole brain imaging and registration  
 116 procedures described in detail in Kuan et al. (2015); Oh et al. (2014) produce quantitative metrics of  
 117 fluorescence discretized at the  $100 \mu\text{m}$  voxel level. Given an experiment, this image was histologically  
 118 segmented by an analyst into *injection* and *projection* areas corresponding to areas containing somas,  
 119 dendrites and axons or exclusively axons of the transfected neurons. An example of a single  
 120 experiment rendered in 3D is given in Figure 1b. Given an experiment  $i$ , we represent injections and  
 121 projections as functions  $x(i), y(i) : \mathcal{B} \rightarrow \mathbb{R}_{\geq 0}$ , where  $\mathcal{B} \subset [1 : 132] \times [1 : 80] \times [1 : 104]$  corresponds to the  
 122 subset of the  $(1.32 \times 0.8 \times 1.04)$  cm rectangular space occupied by the standard voxelized mouse brain.

123 We also calculate injection centroids  $c(i) \in \mathbb{R}^3$  and regionalized projections  $y_{\mathcal{T}}(i) \in \mathbb{R}^T$  given by the  
124 sum of  $y(i)$  in each region. A description of these steps is in Supplemental Section 6.

125 Our goal is the estimation of **regionalized connectivity** from one region to another. A visual  
126 depiction of this region parcellation for a two-dimensional slice of the brain is given in Figure 1c. All  
127 structures annotated in the CCF belong to a hierarchically ordered ontology, with different areas of the  
128 brain are parcellated to differing finer depths within a hierarchical tree. We denote the main levels of  
129 interest as major structures, summary structures, and layers. Not every summary structure has a layer  
130 decomposition within this ontology, so we typically consider the finest possible regionalization - for  
131 example, layer within the cortex, and summary structure within the thalamus, and denote these  
132 structures as leafs. As indicated in Figure 1d, the dataset used to generate the connectivity model  
133 reported in this paper contains certain combinations of region and cell class frequently, and others  
134 not at all. A summary of the most frequently assayed cell classes and structures is given in Figures 1e  
135 and 1f. Since users of the connectivity matrices may be interested in particular combinations, or  
136 interested in the amount of data used to generate a particular connectivity estimate, we present this  
137 information about all experiments in Supplemental Section 5.

138 ***Modeling Regionalized Connectivity***

We define voxelized cell-class specific connectivity  $f : \mathcal{V} \times \mathcal{B} \times \mathbb{R}_{\geq 0}$  as giving the voxelized connectivity strength of a particular cell class from a source voxel to a target voxel. In contrast to Knox et al. (2019), which only uses wild type C57BL/6J mice, our dataset has experiments targeting  $|\mathcal{V}| = 114$  different combinations of Cre-driver mice and Cre-regulated AAV transgenes jointly denoted as  $\mathcal{V} := \{v\}$ . As in Knox et al. (2019), we ultimately estimate an integrated regionalized connectivity defined with respect to a set of  $S = 564$  source leafs  $\mathcal{S} := \{s\}$  and  $T = 1123$  target leafs  $\mathcal{T} := \{t\}$ , of which  $1123 - 564 = 559$  are contralateral. That is, we define

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

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

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

139 where  $l_j$  and  $l_{j'}$  are the locations of source and target voxels, and  $|s|$  and  $|v|$  are defined to be the  
 140 number of voxels in the source and target structure, respectively. Since the normalized strength and  
 141 densities are computable from the strength via a fixed normalization, our main statistical goal is to  
 142 estimate  $\mathcal{C}(v, s, t)$  for all  $v, s$  and  $t$ . In other words, we want to estimate matrices  $\mathcal{C}_v \in \mathbb{R}_{\geq 0}^{S \times T}$ . We call this  
 143 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. We represent these considerations as

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

144 This makes explicit the data featurization  $f_*$ , statistical estimator  $\hat{f}$ , and any potential subsequent  
 145 transformation  $f^*$  such as summing over the source and target regions. Denoting  $\mathcal{D}$  as a function of  $v$   
 146 and  $s$  reflects that we consider using different data to estimate connectivities for different cell-classes  
 147 and source regions. Table 1 reviews estimators used for this data-type used in previous work, as well  
 148 as our two main extensions: the Cre-NW and **Expected Loss** (EL) models. The main differences in our  
 149 data featurization from (Knox et al., 2019) are that we regionalize our data at the leaf level where

150 available so that its layer-specific behavior is visible, and normalize our data by projection signal in  
 151 order to account for differences between cell class. Additional model selection results are given in  
 152 Supplemental Section 5 for alternative normalization strategies, and more detail on estimation is  
 153 given in Supplemental Section 6.

Name	$f^*$	$\hat{f}$	$f_*$	$\mathcal{D}(v, s)$
NNLS (Oh et al., 2014)	$\hat{f}(1_s)$	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)	$X = l_s, Y = y_{\mathcal{T}}$	$I_m/I_m$
Cre-NW	$\sum_{l_s \in s} \hat{f}(l_s)$	NW(X,Y)	$X = l_s, Y = y_{\mathcal{T}}$	$(I_l \cap I_v)/I_m$
Expected Loss (EL)	$\sum_{l_s \in s} \hat{f}(s)$	EL( $X, Y, v$ )	$X = l_s, Y = y_{\mathcal{T}}, v$	$I_l/I_m$

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 index set  $I$ . We denote experiments with centroids in particular major brain divisions and leafs as  $I_m$  and  $I_l$ , respectively. Data  $I_l/I_m$  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  $I_l$  for prediction. The non-negative least squares estimator (NNLS) fits a linear model that predicts regionalized projection signal  $y_{\mathcal{T}}$  as a function of regionalized injection signal  $x_{\mathcal{S}}$ . Thus, the regionalization step for a region  $s$  is given by applying the learned matrix  $\hat{f}$  to the  $s$ -th indicator vector. In contrast, the Nadaraya-Watson model (NW) is a local smoothing model that generates a prediction for each voxel within the source structure that are then averaged to create estimate the structure-specific connectivity.

154 Our contributions - the Cre-NW and Expected Loss (EL) models - have several differences from the  
 155 previous methods. In contrast to the non-negative least squares (Oh et al., 2014) and  
 156 Nadaraya-Watson (Knox et al., 2019) estimators that account only for source region  $s$ , our new  
 157 estimators account cell class  $v$ , The Cre-NW estimator only uses experiments from a particular class  
 158 to predict connectivity for that class, while the EL estimator shares information between classes  
 159 within a structure. Both of these estimator take into account both the cell-class and the centroid  
 160 position of the experimental injection. Like the NW and Cre-NW estimator, the EL estimator generates

<sup>161</sup> predictions for each voxel in a structure, and then sums them together to get the overall connectivity.  
<sup>162</sup> However, in contrast to the NW approaches, the EL estimate of the projection vector for a cell-class at  
<sup>163</sup> a location weights the average projection of that cell-class in the region containing the location  
<sup>164</sup> against the relative locations of all experimental centroids in the region regardless of class. That is,  
<sup>165</sup> cell-class and source region combinations with similar average projection vectors will be upweighted  
<sup>166</sup> when estimating  $\hat{f}$ . Thus, all experiments that are nearby in three-dimensional space can help  
<sup>167</sup> generate the prediction, even when there are few nearby experiments for the cell-class in question. A  
<sup>168</sup> detailed mathematical description of our new estimator is given in Supplemental Section 6.

169 ***Model evaluation***

170 We select optimum functions from within and between our estimator classes using **leave-one-out**  
 171 **cross validation**, in which the accuracy of the model is assessed by its ability to predict projection  
 172 vectors experiments excluded from the training data on the basis of their cell class and experimental  
 173 centroid. Equation 1 includes a deterministic step  $f^*$  included without input by the data. The  
 174 performance of  $\hat{\mathcal{C}}(v, s, t)$  is thus determined by performance of  $\hat{f}(f_*(\mathcal{D}(v, s)))$ . Thus, we evaluate  
 175 prediction of  $f_{\mathcal{T}} : \mathbb{R}^3 \rightarrow \mathbb{R}_{\geq 0}^T$  - the regionalized connection strength at a given location.

176 Another question is what combinations of  $v$ ,  $s$ , and  $t$  to generate a prediction for. Our EL and  
 177 Cre-NW models are leaf specific. They only generate predictions for cell-classes in leafs where at least  
 178 one experiment with a Cre-line targeting that class has a centroid. To accurately compare our new  
 179 estimators with less-restrictive models such as used in Knox et al. (2019), we restrict our evaluation set  
 180 to Cre driver/leaf combinations that are present at least twice. The sizes of these evaluation sets are  
 181 given in Supplemental Section 5.

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

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

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

182  $I_s$  refers to the set of experiments with centroid in structure  $s$ , and  $I_v$  refers to the set of experiments  
 183 with Cre-line  $v$ , so  $|I_s \cap I_v|$  is the number of experiments of Cre-line  $v$  with injection centroid in  
 184 structure  $s$ . This is a somewhat different loss from Knox et al. (2019) because of the increased  
 185 weighting of rarer combinations of  $s$  and  $v$  implicit in the  $\frac{1}{|I_s \cap I_v|}$  term in the loss. The establishment of  
 186 a lower limit of detection and the extra estimation step used in the EL model to establish the relative  
 187 importance of regionally averaged cell-class projection and injection centroid position are covered in  
 188 Supplemental Section 6.

189 ***Connectivity analyses***

190 We examine our connectome estimates with both comparisons to known biology and statistical  
 191 decompositions. As an exploratory analysis, we use hierarchical clustering to compare outputs from  
 192 connectivities from different Cre-lines. Details of and results from this approach are given in Section  
 193 7. We then use non-negative matrix factorization (NMF) to factor the wild-type connectivity matrix  
 194 into a small set of underlying components that can be linearly combined to reproduce the observed  
 195 long-range wild-type connectivity. Inspired by Mohammadi, Ravindra, Gleich, and Grama (2018), we  
 196 refer to these latent coordinates as **connectivity archetypes** since they represent underlying patterns  
 197 from which we can reconstruct a broad range of observed connectivities, although we note that the  
 198 genomic archetypal analysis in that paper is slightly methodologically distinct.

199 NMF refers to a collection of **dictionary-learning** algorithms for decomposing a  
 200 non-negatively-valued matrix such as  $\mathcal{C}$  into positively-valued matrices called, by convention,  
 201 weights  $W \in \mathbb{R}_{\geq 0}^{S \times q}$  and hidden units  $H \in \mathbb{R}_{\geq 0}^{q \times T}$ . NMF assumes a simple linear statistical model: that the  
 202 observed matrix is composed of linear combinations of latent coordinates (Devarajan, 2008). Unlike  
 203 PCA, NMF specifically accounts for the fact that data are all in the positive orthant, and it is more  
 204 stable and interpretable in assays of complex biological systems than hierarchical clustering (Brunet,  
 205 Tamayo, Golub, & Mesirov, 2004). The choice of matrix factorization method reflects particular  
 206 scientific subquestions and probabilistic interpretations, and the matrix  $H$  may be used to identify latent  
 207 structures with interpretable biological meaning.

208 Our application of NMF to decompose the estimated long-range connectivity is some independent  
 209 interest, since we ignore connections between source and target regions less than  $1500 \mu m$  apart. This  
 210 is because short-range projections resulting from diffusion and traveling fibers dominate the matrices  
 211  $\hat{\mathcal{C}}$ . Our NMF algorithm solves the following optimization problem

$$\text{NMF}(\mathcal{C}, \lambda, q) := \arg \min_{W \in \mathbb{R}_{\geq 0}^{S \times q}, H \in \mathbb{R}_{\geq 0}^{q \times T}} \frac{1}{2} \| \mathbf{1}_{d(s,t) > 1500 \mu m} \odot \mathcal{C} - WH \|_2^2 + \lambda (\|H\|_1 + \|W\|_1).$$

212 We set  $\lambda = 0.002$  to encourage sparser and therefore more interpretable components. We use  
 213 unsupervised cross-validation to determine an optimum  $q$ , and show the top 15 stable components  
 214 (Perry, 2009). Since the NMF objective is difficult to optimize and sensitive to initialization, we follow

<sup>215</sup> up with a stability analysis via clustering the resultant  $H$  from multiple replicates. The medians of the  
<sup>216</sup> component clusters appearing frequently across NMF replicates are selected as **connectivity**  
<sup>217</sup> **archetypes**. Details of these approaches are given in Supplementary Sections 6 and 7.

### 3 RESULTS

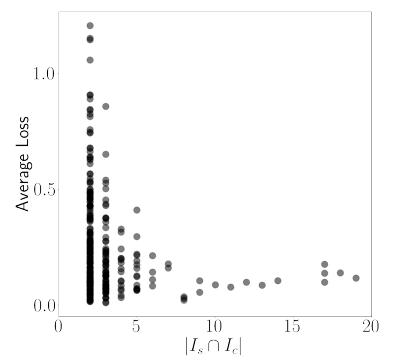
<sup>218</sup> The main result of this paper is the creation of cell-type specific connectivity estimates from the Allen  
<sup>219</sup> Mouse Brain Connectivity Atlas (MCA) experiments. We first establish that our new expected-loss (EL)  
<sup>220</sup> estimator performs best in validation assays for estimating wild-type and cell-type specific  
<sup>221</sup> connectivities. We then show that Cre-specific connectivity matrices generated using this model are  
<sup>222</sup> consistent with known biology. Finally, we factor some of these connectivity matrices to show how  
<sup>223</sup> connectivity arises from latent components, and that these latent components may be associated  
<sup>224</sup> with cell-types.

<sup>225</sup> ***Model evaluation***

<sup>226</sup> Our EL model generally performs better than the other estimators that we consider. Table 2a contains  
<sup>227</sup> weighted losses from leave-one-out cross-validation of candidate models, such as the NW Major-WT  
<sup>228</sup> model from Knox et al. (2019). The EL model combines the good performance of class-specific  
<sup>229</sup> models like NW Leaf-Cre in regions like Isocortex with the good performance of class-agnostic models  
<sup>230</sup> in regions like Thalamus. Additional information on model evaluation, including class and structure-  
<sup>231</sup> specific performance, is given in Appendix 5. In particular, Supplementary Table 4 contains the sizes  
<sup>232</sup> of these evaluation sets in each major structure, and Supplementary Section 7 contains the structure-  
<sup>233</sup> and class specific losses.

$\hat{f}$	Mean Leaf-Cre Mean $I_c \cap I_L$	NW Major-Cre NW $I_c \cap I_M$	NW Leaf-Cre NW $I_c \cap I_L$	NW Leaf NW $I_L$	NW Major-WT NW $I_{wt} \cap I_M$	NW Major NW $I_M$	EL EL $I_L$
Isocortex	0.239	0.252	0.234	0.279	0.274	0.274	<b>0.228</b>
OLF	0.193	0.233	0.191	<b>0.135</b>	0.179	0.179	0.138
HPF	0.175	0.332	0.170	0.205	0.228	0.228	<b>0.153</b>
CTXsp	<b>0.621</b>	<b>0.621</b>	<b>0.621</b>	<b>0.621</b>	<b>0.621</b>	<b>0.621</b>	<b>0.621</b>
STR	0.131	<b>0.121</b>	0.128	0.169	0.232	0.232	0.124
PAL	0.203	0.205	0.203	0.295	0.291	0.291	<b>0.188</b>
TH	0.673	0.664	0.673	<b>0.358</b>	0.379	0.379	0.369
HY	0.360	0.382	0.353	0.337	0.317	0.317	<b>0.311</b>
MB	0.168	0.191	0.160	0.199	0.202	0.202	<b>0.159</b>
P	0.292	0.292	0.292	0.299	0.299	0.299	<b>0.287</b>
MY	0.268	0.347	0.268	0.190	<b>0.189</b>	<b>0.189</b>	0.204
CB	<b>0.062</b>	<b>0.062</b>	<b>0.062</b>	0.068	0.112	0.112	0.068

(a)



(b)

Table 2: 2a Losses from leave-one-out cross-validation of candidate models. **Bold** numbers are best for their major structure. 2b Empirical performance of selected EL model by data abundance. The model is more accurate in Cre-leaf combinations where it draws on more data. The dataset variable  $\mathcal{D}$  indicates the set of experiments used to model a given connectivity. For example,  $I_c \cap I_L$  means only experiments with a given Cre-line in a given leaf are used to model connectivity for the corresponding cell-class in that leaf, while  $I_L$  means that all experiments in that leaf are used.  $I_{wt} \cap I_M$  means all wild-type experiments in the major structure are used - this was the model in Knox et al. (2019).

### 234 **Connectivities**

235 We estimate matrices  $\hat{\mathcal{C}}_v \in \mathbb{R}_{\geq 0}^{S \times T}$  representing connections of source structures to target structures for  
 236 particular Cre-lines  $v$ . We confirm the detection of several well-established connectivities within our  
 237 tensor, although we expect additional interesting biological processes to be identifiable. The  
 238 connectivity tensor and code to reproduce it are available at <https://github.com/>  
 239 AllenInstitute/mouse\_connectivity\_models/tree/class-specific.

240 **Overall connectivity** Several known biological projection patterns are evident in the wild-type  
 241 connectivity matrix  $\mathcal{C}_{wt}$  shown in Figure 2a. This matrix shows connectivity from leaf sources to leaf  
 242 targets. Large-scale patterns like intraareal connectivities and ipsilateral connections between cortex  
 243 and thalamus are clear, as in previous estimates in J. A. Harris et al. (2019); Knox et al. (2019); Oh et al.

<sup>244</sup> (2014). However, the layer-specific targeting of the different Cre-lines enables our estimated wild-type  
<sup>245</sup> connectivities to display heterogeneity at the layer level. This contrasts the model in Knox et al.  
<sup>246</sup> (2019), which is denominated as the NW Major-WT model whose accuracy is evaluated in Table 2a.  
<sup>247</sup> For comparison, we also plot averages over component layers weighted by layer size projections  
<sup>248</sup> between summary-structure sources and targets in the cortex in Figure 2b. Importantly, as shown in  
<sup>249</sup> Table 2a this finer spatial resolution corresponds to the increased accuracy of our EL model over the  
<sup>250</sup> NW Major-WT model.

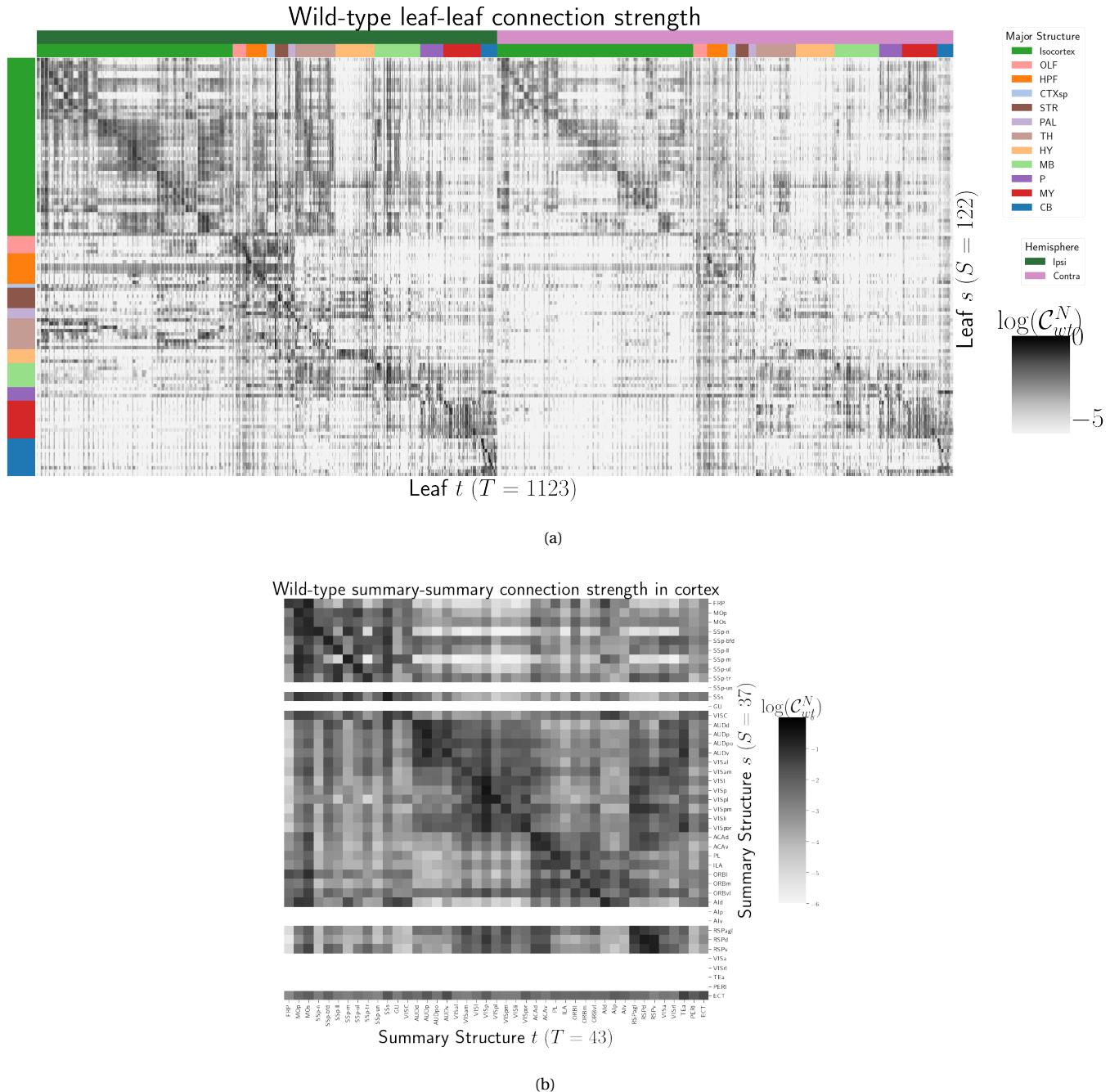


Figure 2: Wild-type connectivities. 2a Log wild-type leaf-to-leaf connectivity matrix  $\log \mathcal{C}(s, t, v_{wt})$ . Sources and target major brain division structure are shown. 2b Log wild-type intracortical connectivity matrix at the summary structure level. Summary structures without an injection centroid are left blank.

251 *Class-specific connectivities* We investigate the presence of known biological processes within our  
 252 connectivity estimates and confirm that these class-specific connectivities exhibit certain known  
 253 behaviors. Although there is a rich anatomical literature using anterograde tracing data to describe  
 254 projection patterns from subcortical sources to a small set of targets of interest, much of the  
 255 accessible whole brain projection data is from the MCA project used here to generate the connectome  
 256 models. Thus, we compare to external studies to validate our results while avoiding a circular  
 257 validation of the data used to generate the model weights. The cell types and source areas with  
 258 extensive previous anatomical descriptions of projections using both bulk tracer methods with cell  
 259 type specificity and single cell reconstructions that we investigate are 1) thalamic-projecting neurons  
 260 in the visual and motor cortical regions, 2) cholinergic neurons in the medial septum and nucleus of  
 261 the diagonal band (MS/NDB); and 3) serotonergic neurons of the dorsal raphe nucleus (DR). Our  
 262 estimated connections are in agreement with literature on these cell types.

263 DEPENDENCE OF THALAMIC CONNECTION ON CORTICAL LAYER. Visual cortical areas VISp and VISl and  
 264 cortical motor areas MOp and MOs have established layer-specific projection patterns that can be  
 265 labeled with the layer-specific Cre-lines from the Allen datasets and others (J. A. Harris et al., 2019;  
 266 Jeong et al., 2016). Figure 3a shows that in VISp, the Ntsr1-Cre line strongly targets the core part of the  
 267 thalamic LGd nucleus while in VISl, it a strong projector to the LP nucleus. In VISp, the Rbp4-Cre line  
 268 strongly targets LP as well. Rbp4-Cre and Ntsr1-Cre injections target layers 5 and 6 respectively. Since  
 269 we only generate connectivity estimates for structures with at least one injection centroid, this is  
 270 shown by the position of non-zero rows in Figure 3a. To fill these gaps, as a heuristic alternative  
 271 model, we also display an average connectivity matrix over all Cre-lines.

272 MS AND NDB PROJECTIONS IN THE CHAT-IRES-CRE-NEO MODEL. Cholinergic neurons in the MS and  
 273 NDB are well-known to strongly innervate the hippocampus, olfactory bulb, piriform cortex,  
 274 entorhinal cortex, and lateral hypothalamus (Watson, Paxinos, & Puelles, 2012; Zaborszky et al., 2015).  
 275 In the Allen MCA, cholinergic neurons were labeled by injections into Chat-IRES-Cre-neo mice.  
 276 Figure 3b checks the estimated connectome weights to targets in these major brain divisions from MS  
 277 and NDB. We observed that all these expected divisions were represented above the 90th percentile of  
 278 weights from these source structures.

279 We also compared our Chat-IRES-Cre connectome model data for MS and NDB with the targets  
280 identified by Li et al. (2018). This single cell whole brain mapping project using Chat-Cre mice fully  
281 reconstructed n=50 cells to reveal these same major targets and also naming additional targets from  
282 MS/NDB (Li et al., 2018). We identified 150 targets at the fine leaf structure level among the top decile  
283 of estimated weights. To directly compare our data across studies, we merged structures as needed to  
284 get to the same ontology level, and remove ipsilateral and contralateral information. After formatting  
285 our data, we found 51 targets in the top 10%; Li et al. (2018) reported 47 targets across the 50 cells.  
286 There was good consistency overall between the target sets; 35 targets were shared, 12 were unique to  
287 the single cell dataset, and 16 unique to our model data. We checked whether targets missing from  
288 our dataset were because of the threshold level. Indeed, lowering the threshold to the 75th percentile  
289 confirmed 6 more targets-in-common, and all but 2 targets from Li et al. (2018) were above the 50th  
290 percentile weights in our model. Of note, the absence of a target in the single cell dataset that was  
291 identified in our model data is most likely due to the sparse sampling of all possible projections from  
292 only n=50 MS/NDB cells.

293 DR PROJECTIONS IN THE SLC6A4-CRE'ET33 MODEL. Serotonergic projections from cells in the  
294 dorsal raphe (DR) are widely distributed and innervate many forebrain structures including isocortex  
295 and amygdala. In the Allen MCA, serotonergic neurons were labeled using Slc6a4-Cre\_ET33 and  
296 Slc6a4-CreERT2\_EZ13 mice. This small nucleus appears to contain a complex mix of molecularly  
297 distinct serotonergic neuron subtypes with some hints of subtype-specific projection patterns (Huang  
298 et al., 2019; Ren et al., 2018, 2019). We expect that the Cre-lines used here in the Allen MCA, which use  
299 the serotonin transporter promoter (Slc6a4-Cre and -CreERT2), will lead to expression of tracer in all  
300 the serotonergic subtypes recently described in an unbiased way, but this assumption has not been  
301 tested directly. We compared our model data to a single cell reconstruction dataset consisting of n=50  
302 serotonergic cells with somas in the DR that also had bulk tracer validation 3c. After processing our  
303 data to match the target structure ontology level across studies, we identified 37 targets from the DR  
304 with weights above the 90th percentile, whereas Ren et al. (2019) listed 55 targets across the single cell  
305 reconstructions. Twenty seven of these targets where shared.

306 Overall there was good consistency between targets in olfactory areas, cortical subplate, CP, ACB  
307 and amygdala areas, as well in palidum and midbrain, while the two major brain divisions with the

308 least number of matches are the isocortex and thalamus. There are a few likely reasons for these  
309 observations. First, in the isocortex, there is known to be significant variation in the density of  
310 projections across different locations, with the strongest innervation in lateral and frontal orbital  
311 cortices Ren et al. (2019). Indeed, when we lower the threshold and check for weights of the targets  
312 outside of the 90%, we see all but one of these regions (PTLp, parietal cortex which is not frontal or  
313 lateral) has a weight assigned in the top half of all targets. In the thalamus, our model predicted strong  
314 connections to several medial thalamic nuclei (i.e., MD, SMT) that were not targeted by the single  
315 cells. This discrepancy may be at least partially explained by the complex topographical organization  
316 of the DR that, like the molecular subtypes, is not yet completely understood. A previous bulk tracer  
317 study that specifically targeted injections to the central, lateral wings, and dorsal subregions of the DR  
318 reported semi-quantitative differences in projection patterns (Muzerelle, Scotto-Lomassese, Bernard,  
319 Soiza-Reilly, & Gaspar, 2016). Notably, Muzerelle et al. (2016) report that cells in the ventral region of  
320 DR project more strongly to medial thalamic nuclei, whereas the lateral and dorsal DR cells innervate  
321 more lateral regions (e.g., LGd). Thus, it is possible that the single cell somas did not adequately  
322 sample the entire DR.

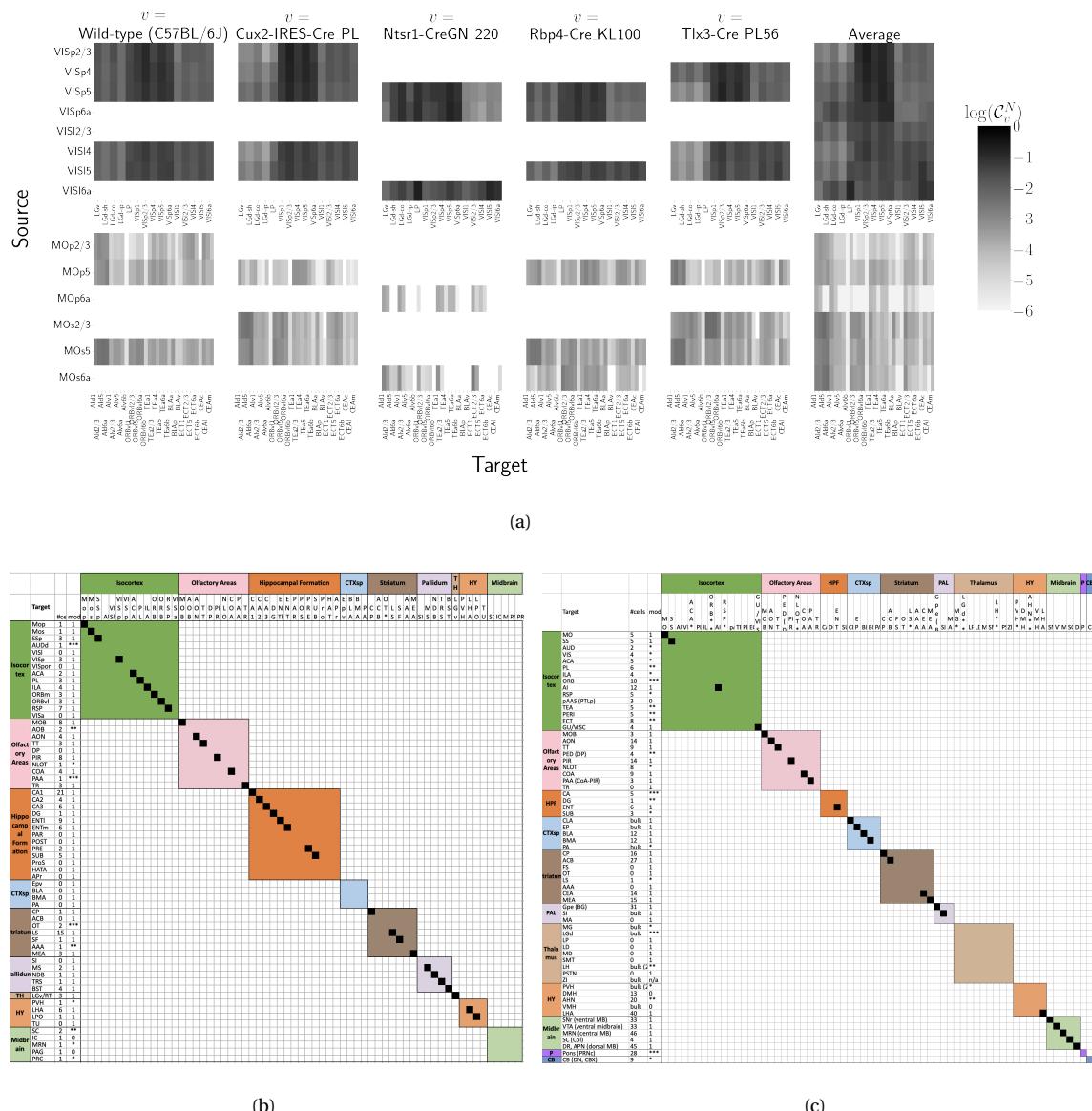


Figure 3: Cell-class specificity. 3a Selected cell-class and layer specific connectivities from two visual and two motor areas. Sources without an injection of a Cre-line are not estimated due to lack of data. 3b Targets reported for cholinergic cells in MS/NDB in Li et al. (2018) and above the 90th percentile in our Chat-IRES-Cre-neo model. A black box along the diagonal indicates the target was identified in both studies ( $n = 35$ ). The number of single cells (out of 50) with a projection to each target is shown in the first column after the target acronym and the next column shows whether a target appeared in the 90% thresholded model weights (1= present, 0=absent). Some of these targets only appeared at lower thresholds as indicated by asterisks, \* \* \* > 85th%, \*\* > 75th%, \* > 50th%. 3c Targets reported for serotonergic cells in DR in Ren et al. (2018, 2019) and above the 90th percentile in our Slc6a4-Cre\_ET33 model.

**323 Connectivity Analyses**

324 While the manual analysis in Section 3 is valuable for validation, scaling our interpretation of our  
 325 connectivity estimates motivated us to apply dimension reduction methods to our connectivity  
 326 estimates and understand whether the learned structure agrees with the expected biology. For  
 327 example, Supplemental Section 7 shows a collection of connectivity strengths generated using  
 328 Cre-specific models for wild-type, Cux2, Ntsr1, Rbp4, and Tlx3 cre-lines from VIS areas at leaf level in  
 329 the cortex to cortical and thalamic nuclei. Sorting source and target structure/cell-class combinations  
 330 hierarchical clustering shows, for example, that layer 6-specific Ntsr1 Cre-lines source regions cluster  
 331 together. This makes sense, since layer 6-specific Ntsr1 Cre-line distinctly projects to thalamic nuclei,  
 332 regardless of source summary structure, and in contrast with the tendency of other cell-classes to  
 333 project to nearby regions within the cortex.

334 In contrast with hierarchical clustering, non-negative matrix factorization provides a simpler  
 335 linear-model based factorization (Hastie, Tibshirani, & Friedman, 2009). The low-dimensional  
 336 coordinates returned by NMF nevertheless highlight important features within the connectivity  
 337 matrix in a data-driven way. The learned projection archetypes  $H$  and model weights  $W$  are plotted in  
 338 Figure 4. Infrastructure connections such as MB-MB MY-MY are visible in the 7th and 11th  
 339 archetypes, but there is no obvious layer-6 specific signal. These factors may be used to generate a  
 340 reconstructed connectivity matrix using the implied statistical model. Despite the relatively small  
 341 number of learned additive factors, comparing with Figure 2a shows that this reconstructed matrix  
 342 has a relatively globally plausible structure. Supplemental Sections 6 and 7 contain quantitative  
 343 performance of this model across choices of  $q$ , and assessment of factorization stability to ensure the  
 344 decomposition is reliable across computational replicates. The supplement also quantitatively shows  
 345 differential association of projection archetypes in this model with projection vectors of sources from  
 346 the Cux2, Ntsr1, Rbp4, and Tlx3 Cre-lines.

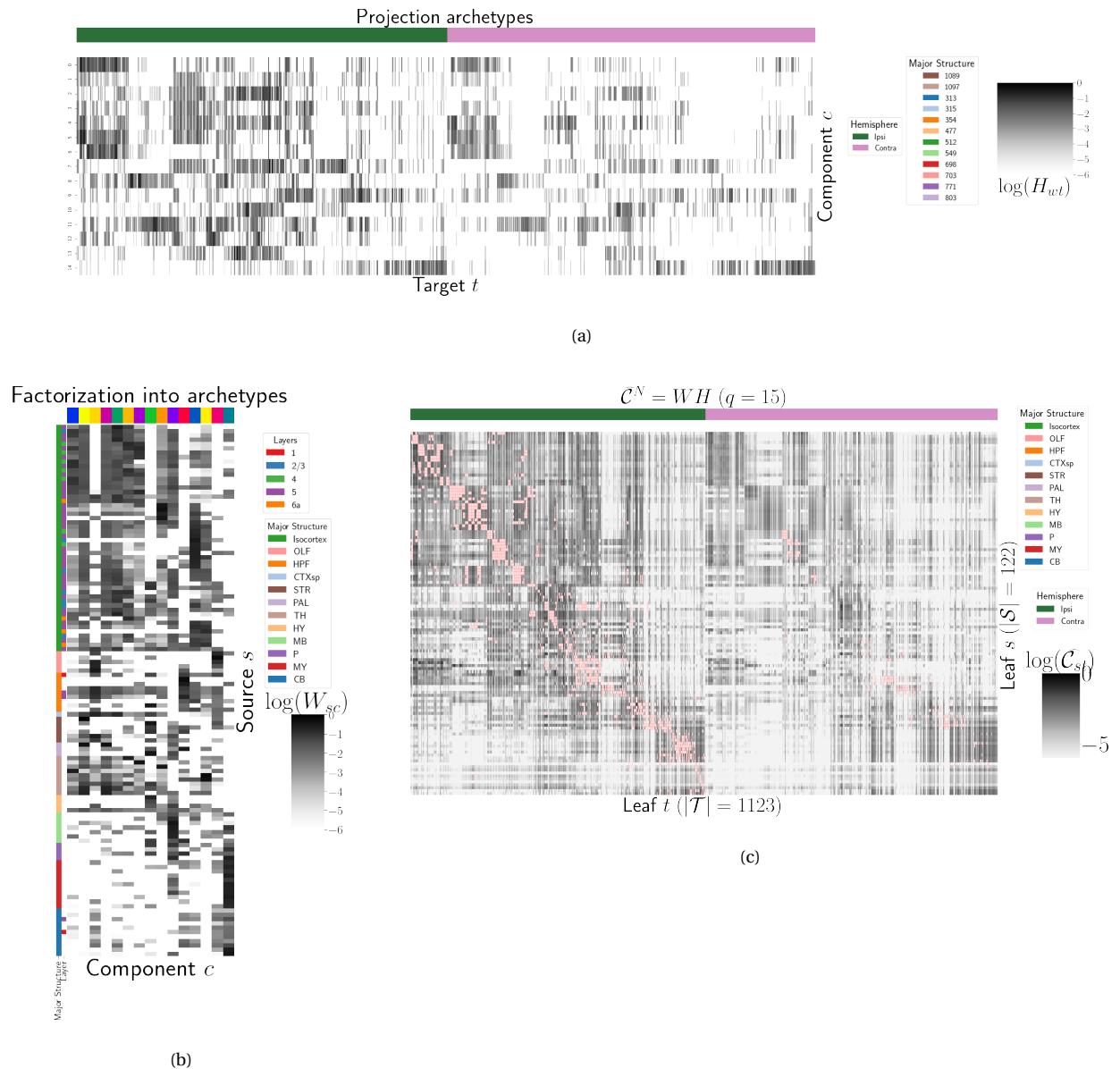


Figure 4: Non-negative matrix factorization results  $\mathcal{C}_{wt}^N = WH$  for  $q = 15$  components. 4a Latent space coordinates  $H$  of  $\mathcal{C}$ . Target major structure and hemisphere are plotted. 4b Loading matrix  $W$ . Source major structure and layer are plotted. 4c Reconstruction of the normalized distal connectivity strength using the top 15 archetypes. Areas less than  $1500 \mu m$  apart are not modeled, and therefore shown in pink.

## 4 DISCUSSION

347 The model presented here is among the first cell-type specific whole brain projectome models for a  
 348 mammalian species, and it opens the door for a large number of models linking brain structure to  
 349 computational architectures. Overall, we find expected targets, based on our anatomical expertise  
 350 and published reports, but underscore that the core utility of this bulk connectivity analysis is not  
 351 only in validation of existing connection patterns, but also in identification of new ones. We note that  
 352 although the concordance appeared stronger for the cholinergic cells than the serotonergic cells, any  
 353 differences might still be explained by the lack of high quality “ground-truth” datasets to validate  
 354 these Cre-line connectome models. It is important to note several limitations of the current analyses.  
 355 Short-range connections can be affected by saturating signals near injection sites, as well as the  
 356 segmentation algorithm capturing dendrites as well as axons. Furthermore, larger numbers of single  
 357 cell reconstructions that saturate all possible projection types would be a better gold standard than  
 358 the small number of cells reported here. Future iterations of connectome models may also take into  
 359 account single cell axon projection data, or synthesize with retrograde tracing experiments.

360 The Nadaraya-Watson estimator using the cell-type space based on similarities of projections, and  
 361 theoretical justification of the use of an intermediate shape-constrained estimator, provides an  
 362 empirically useful new tool for categorical modeling. Ours is not the first cross-validation based  
 363 model averaging method (Gao, Zhang, Wang, & Zou, 2016), but our use of shape-constrained  
 364 estimator in target-encoded feature space is novel and fundamentally different from  
 365 Nadaraya-Watson estimators that use an optimization method for selecting the weights (Saul &  
 366 Roweis, 2003). The properties of this estimator and its relation to estimators fit using an optimization  
 367 algorithm are therefore a possible future avenue of research (Groeneboom & Jongbloed, 2018; Salha &  
 368 El Shekh Ahmed, 2015). Since in truth the impact of the viral tracer likely depends on the injection  
 369 location in a non-linear way, a deep-learning model could be appropriate, provided enough data was  
 370 available, but our sample size seems too low to utilize a fixed-effect generative model (Lotfollahi,  
 371 Naghipourfar, Theis, & Alexander Wolf, 2019). In a sense both the non-negative least squares Oh et al.  
 372 (2014) and NW models can be thought of as improvements over the structure-specific average, and so  
 373 is also possible that a yet undeveloped residual-based data-driven blend of these models could  
 374 provide improved performance. Finally, we note that a Wasserstein-based measure of injection

375 similarity per structure could naturally combine both the physical simplicity of the centroid model  
376 while also incorporating the full distribution of the injection signal.

377 The factorization of the connectivity matrix could also be improved. Non-linear data  
378 transformations or matrix decompositions, or tensor factorizations that account for correlations  
379 between cell-types could better capture the true nature of latent neural connections (K. D. Harris et  
380 al., 2016). Cre- or layer-specific signal recovery as performed here could be used to evaluate a range of  
381 matrix decompositions. This could help for example to understand the influence of traveling fibers on  
382 the observed connectivity (Llano & Sherman, 2008). From a statistical perspective, stability-based  
383 method for establishing archetypal connectivities in NMF is similar to those applied to genomic data  
384 Kotliar et al. (2019); Wu et al. (2016). Regardless of statistical approach, as in genomics, latent  
385 low-dimensional organization in connectivity should inspire search for similarly parsimonious  
386 biological correlates.

## ACKNOWLEDGMENTS

<sup>387</sup> We thank the Allen Institute for Brain Science founder, Paul G. Allen, for his vision, encouragement,  
<sup>388</sup> and support. Samson Koelle and Marina Meila have been in part supported by NSF grants DMS  
<sup>389</sup> 2015272 and DMS 1810975. Stefan Mihalas has in part been supported by NIH grants  
<sup>390</sup> 1RF1DA055669-01 and 1R01EB029813-01.

391 This supplement is divided into information about our dataset, supplemental methods, and  
392 supplemental results. However, certain topics are revisited between sections. Thus, if a reader is  
393 interested in, say, non-negative matrix factorization, they may find relevant information in both  
394 methods and results.

## 5 SUPPLEMENTAL INFORMATION

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

### 397 *Cre/structure combinations in $\mathcal{D}$*

398 This section describes the abundances of structure and Cre-line combinations in our dataset. That is,  
399 it indicates how many experiments in our dataset with a particular Cre-line have an injection centroid  
400 in a particular structure. Users of the connectivity matrices who are interested in a particular Cre-line  
401 or structure can see the quantity and type of data used to compute and evaluate that connectivity.

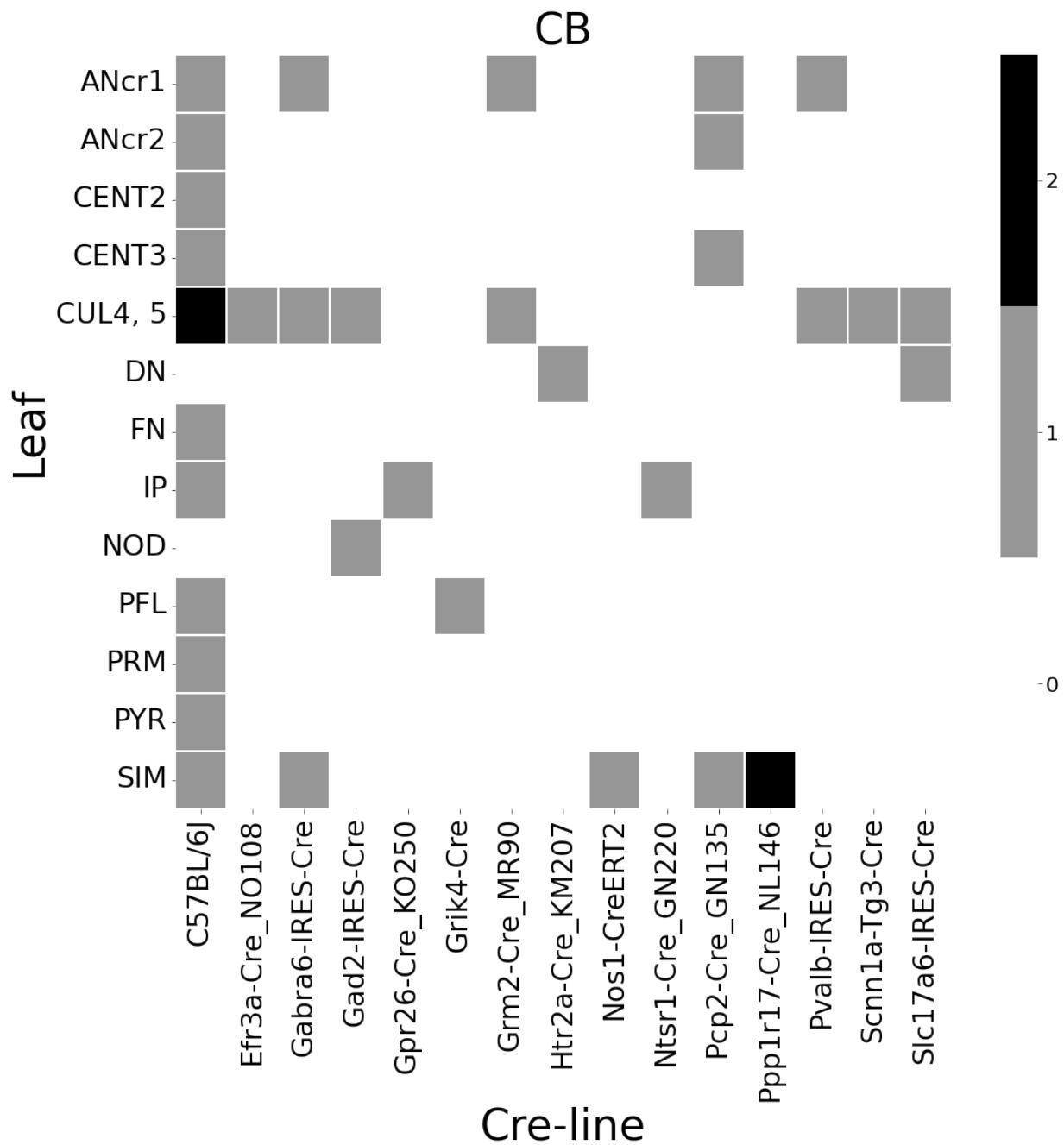


Figure 5: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

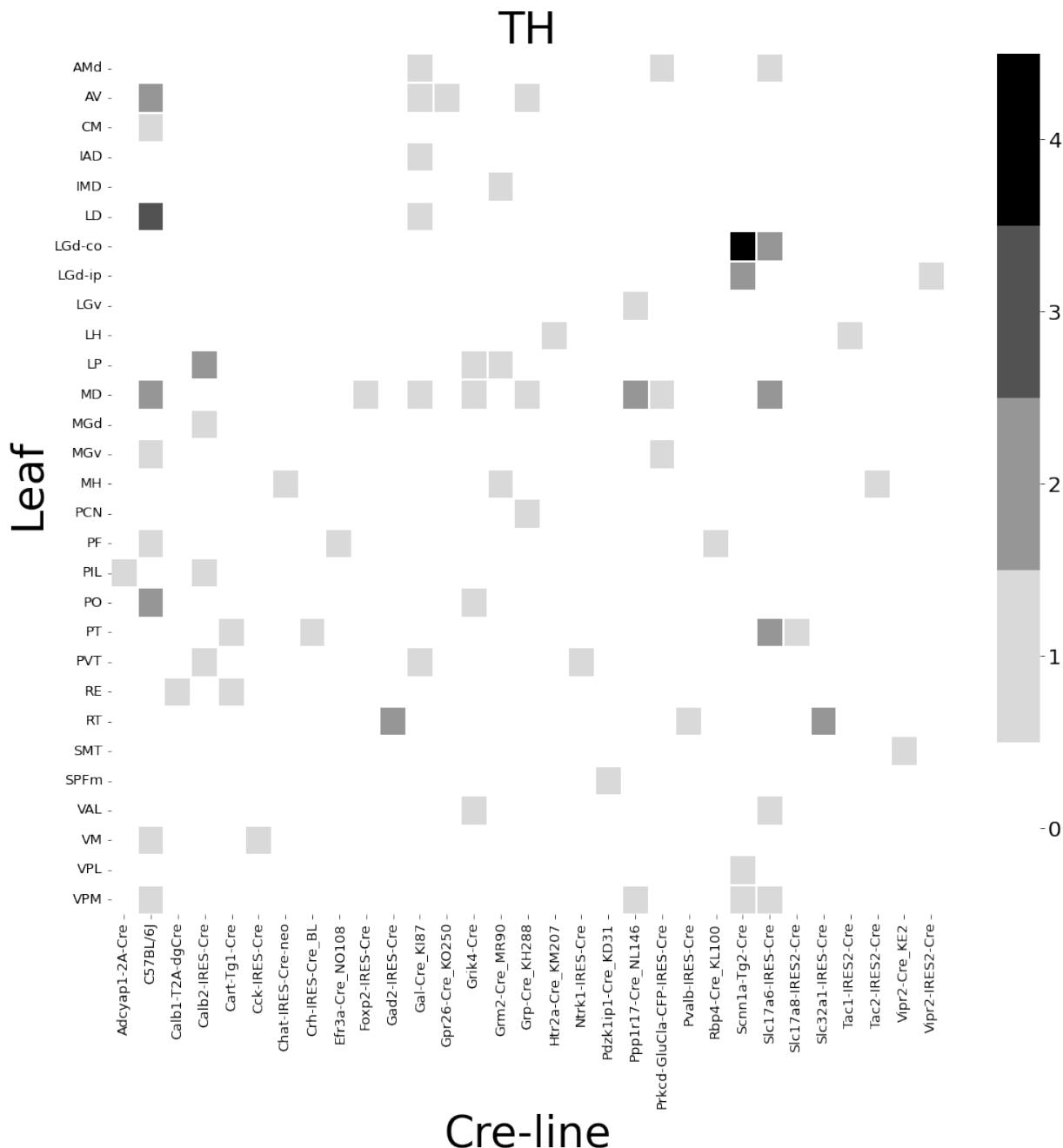


Figure 6: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

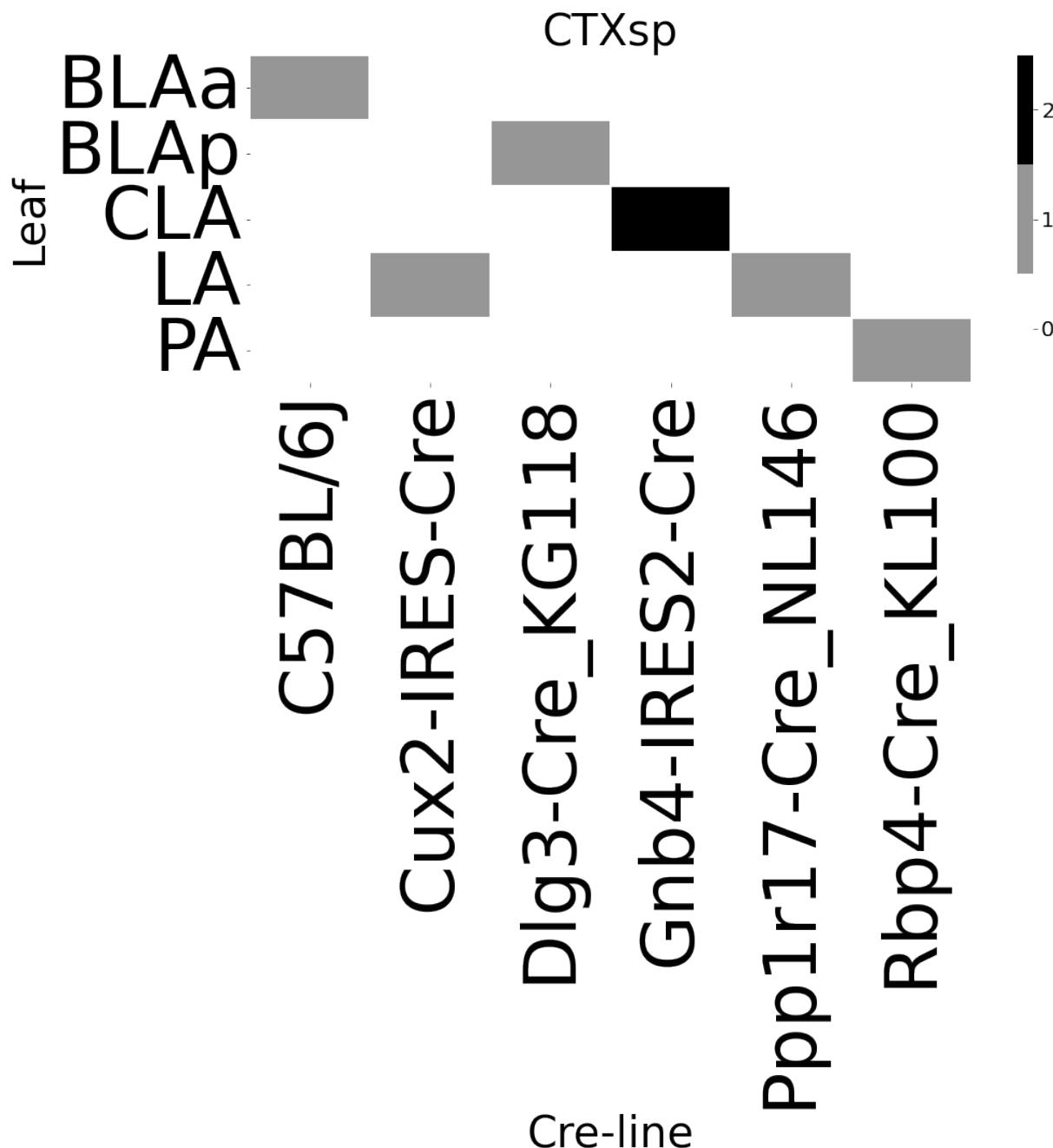


Figure 7: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

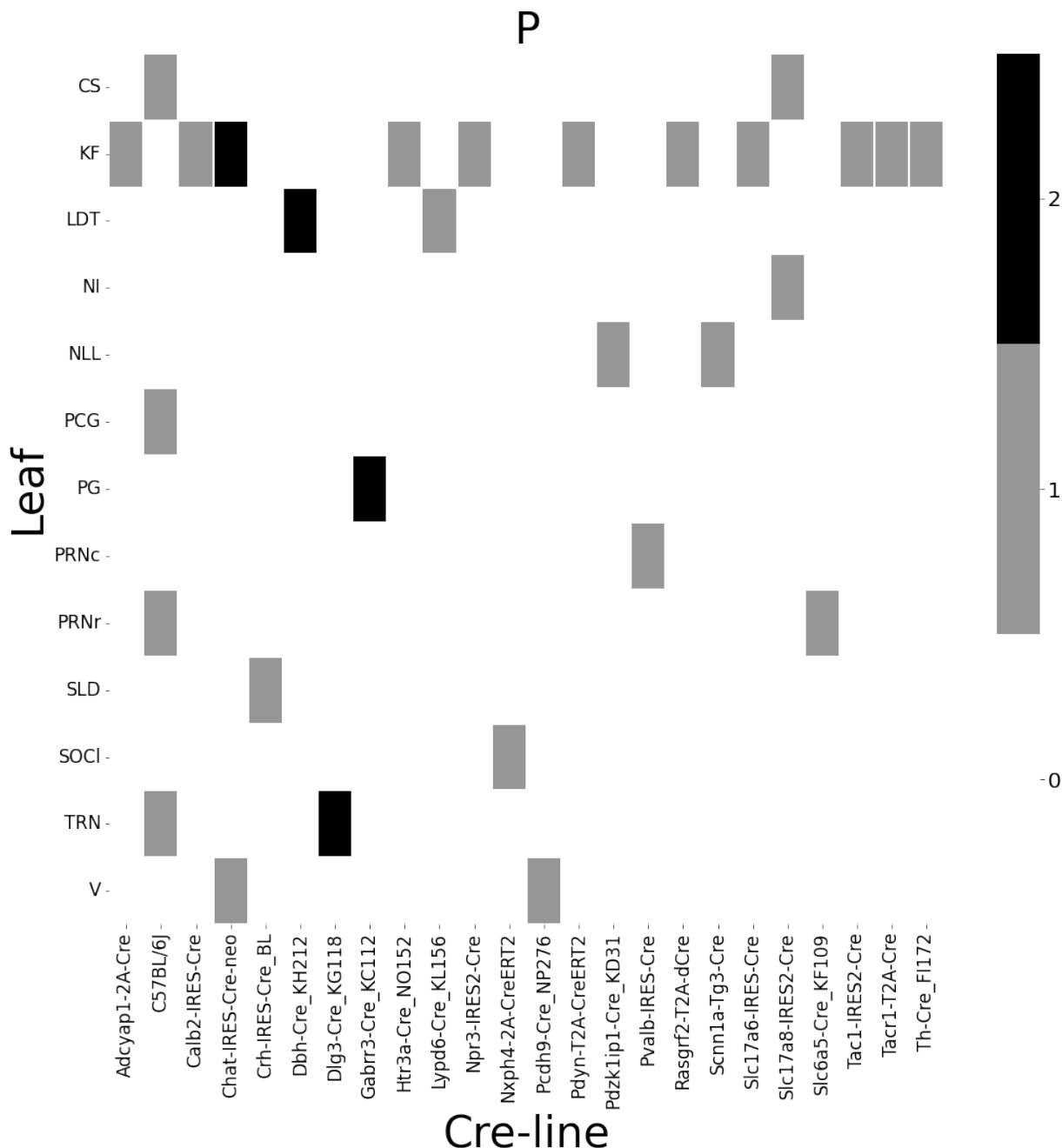


Figure 8: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

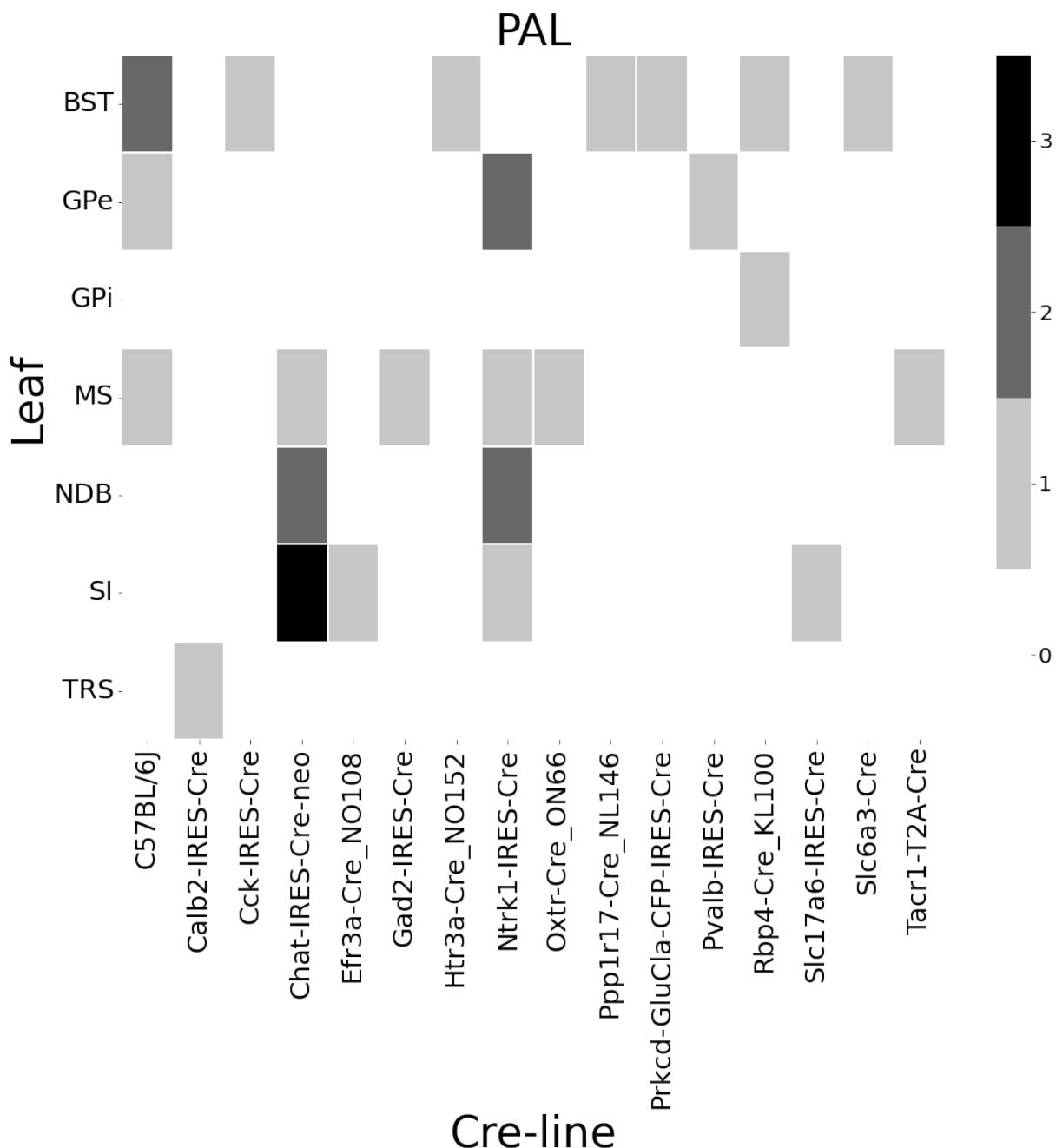


Figure 9: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

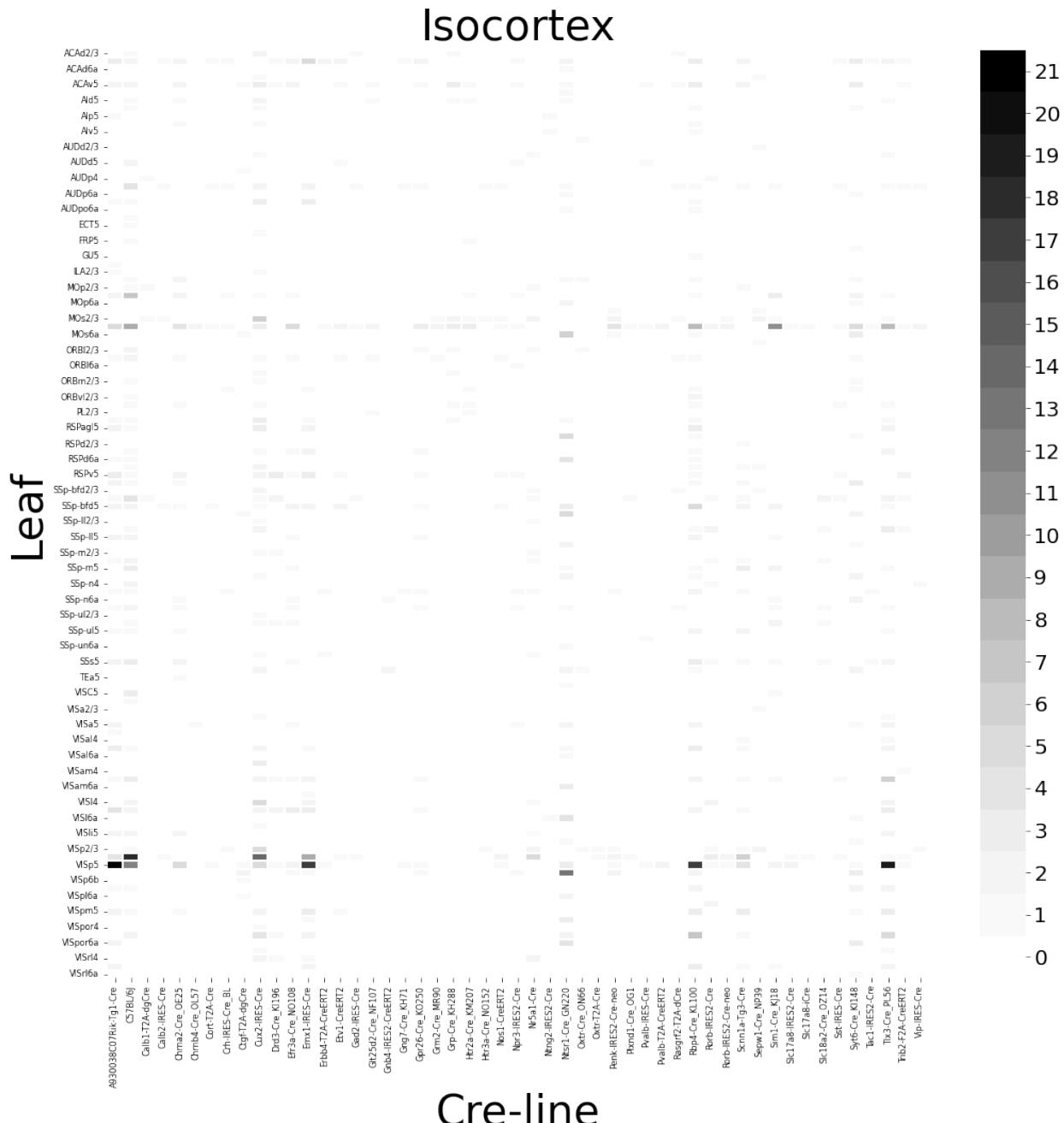


Figure 10: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

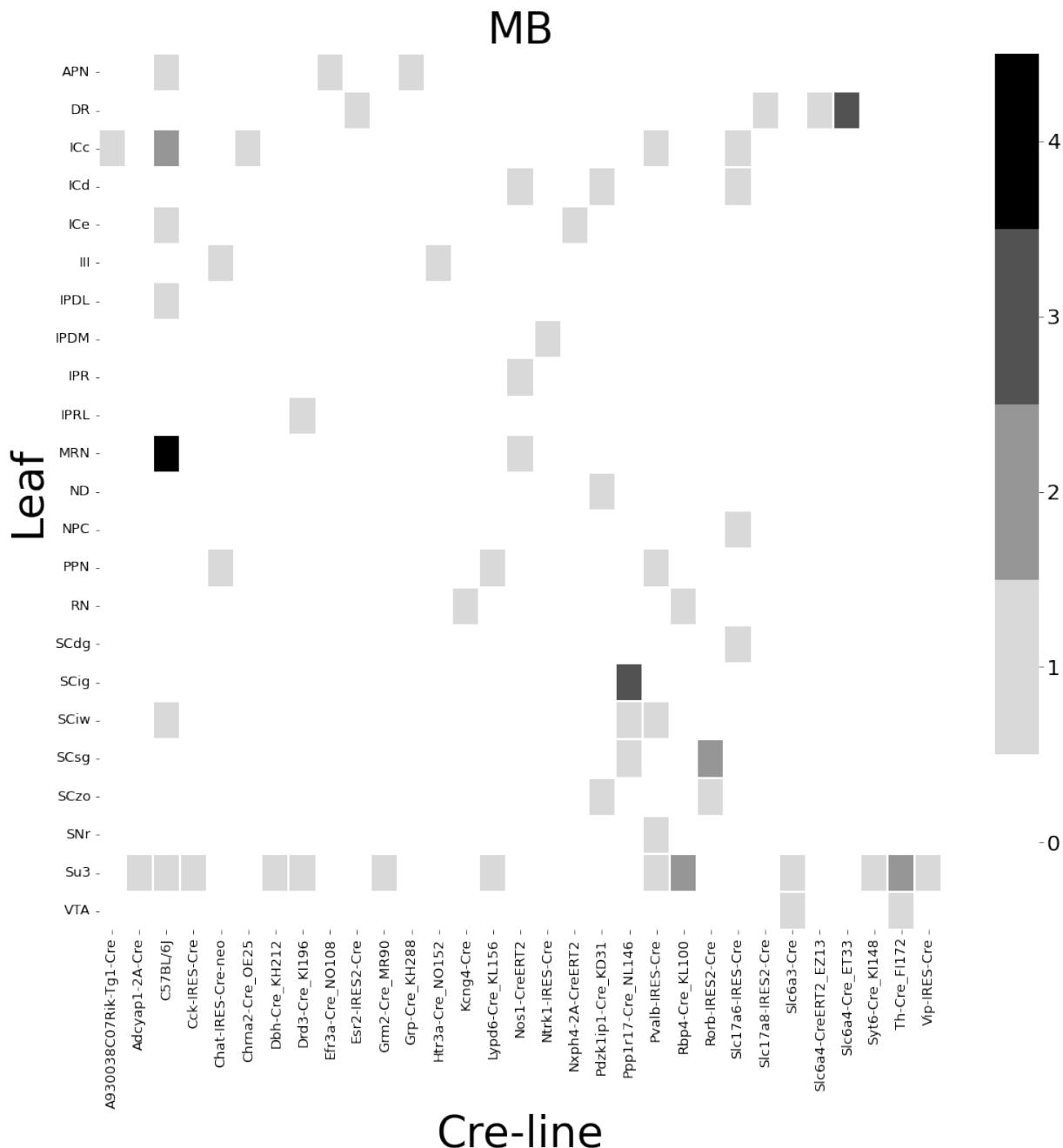


Figure 11: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

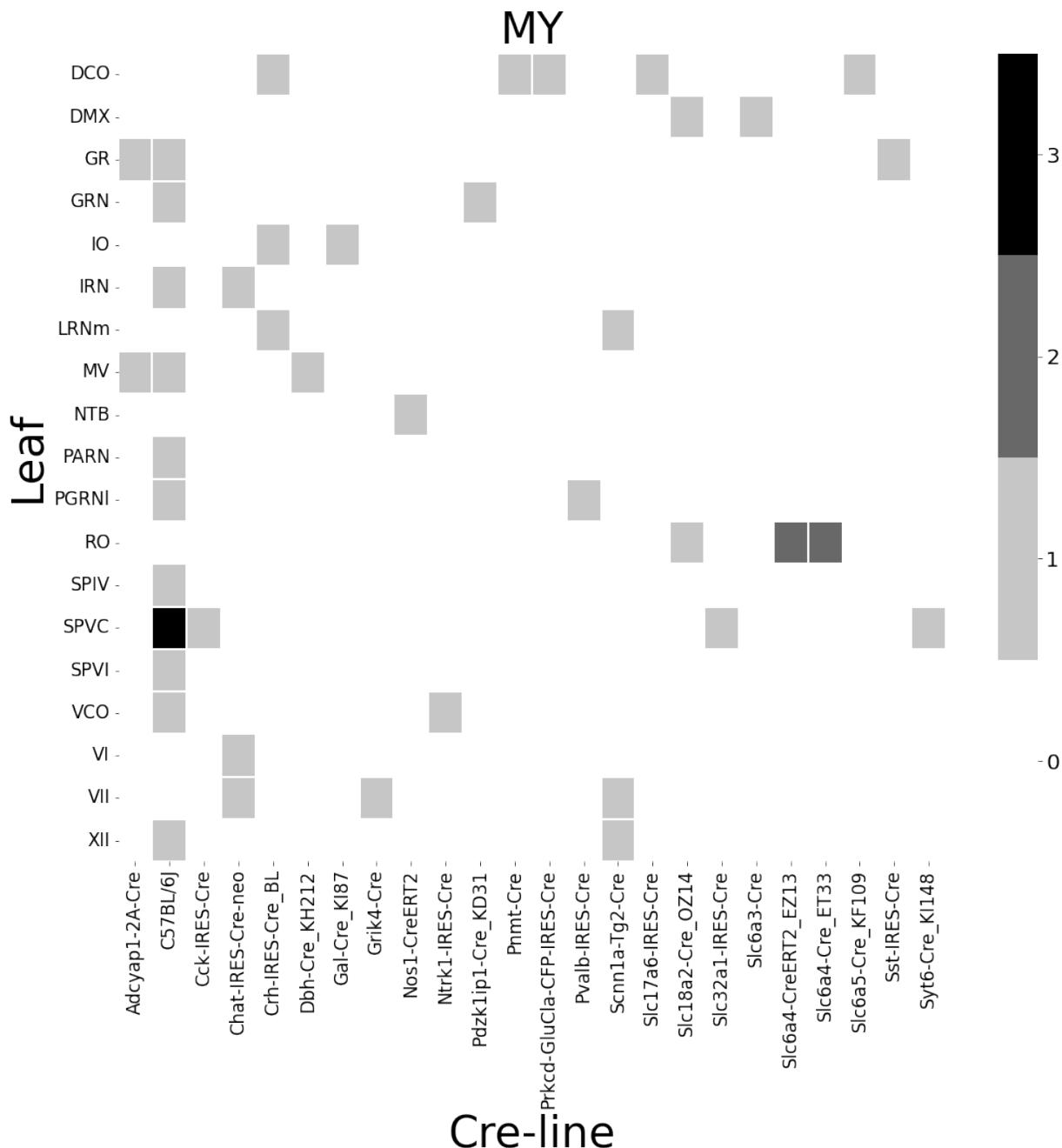


Figure 12: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

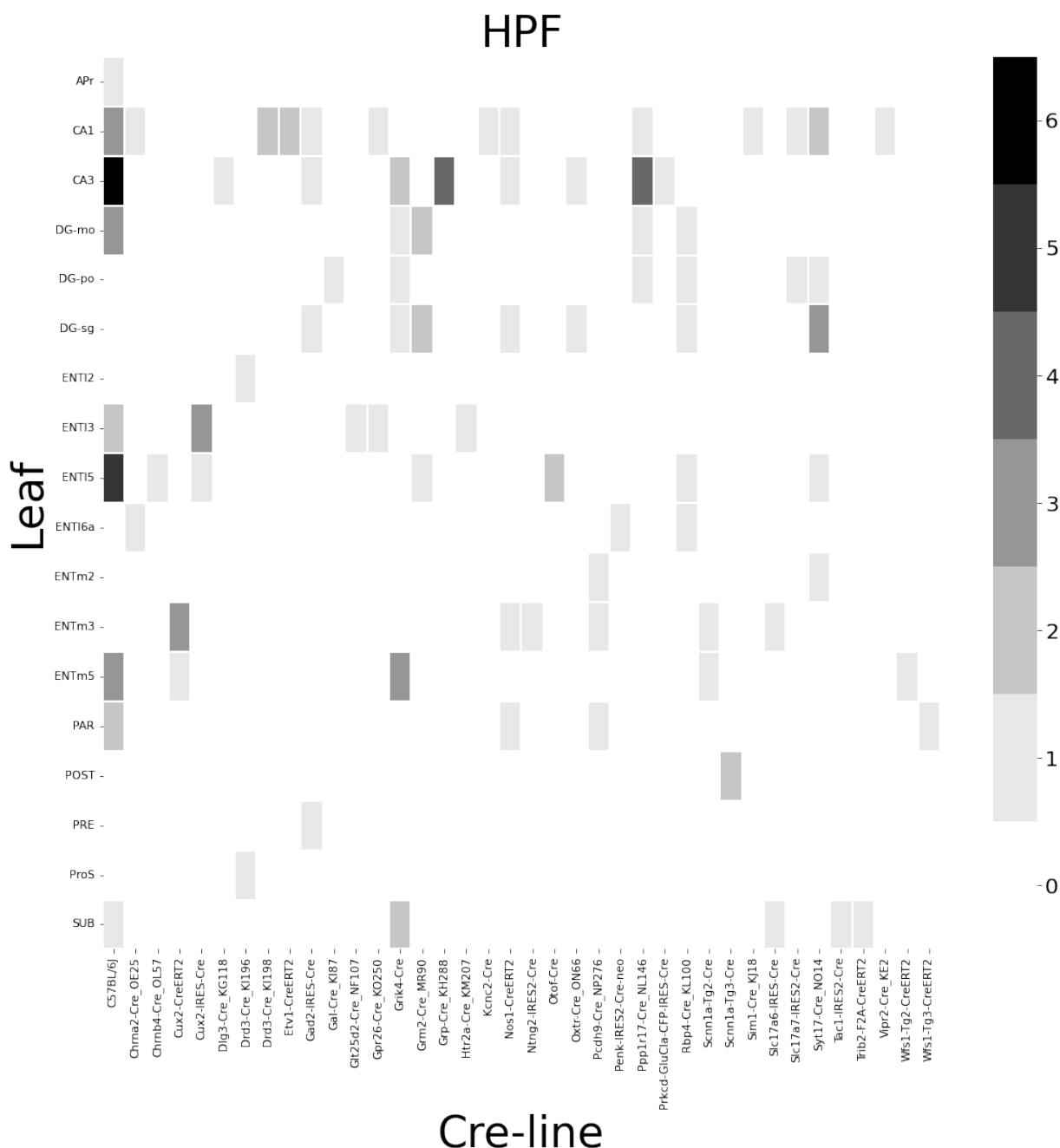


Figure 13: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

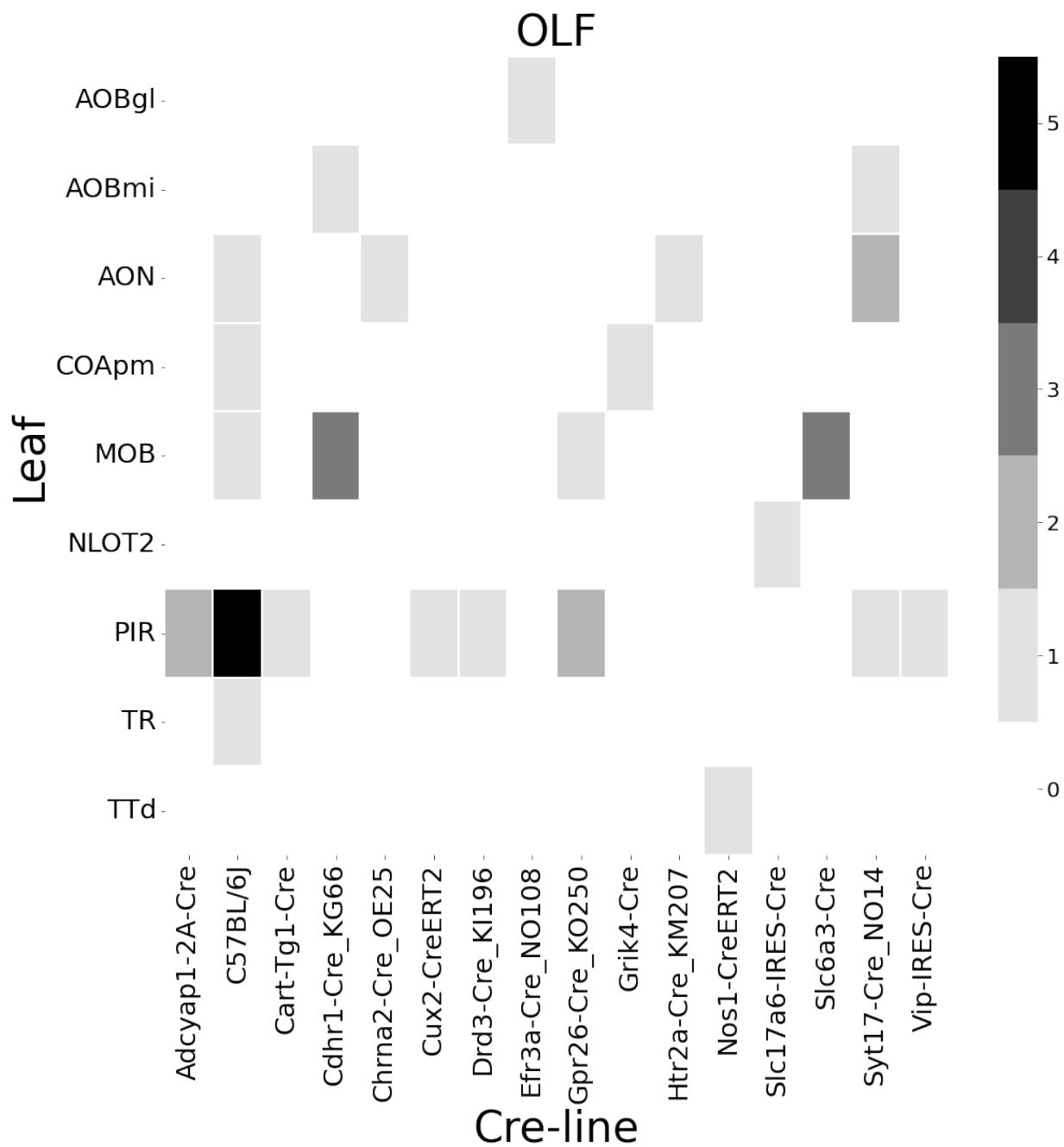


Figure 14: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

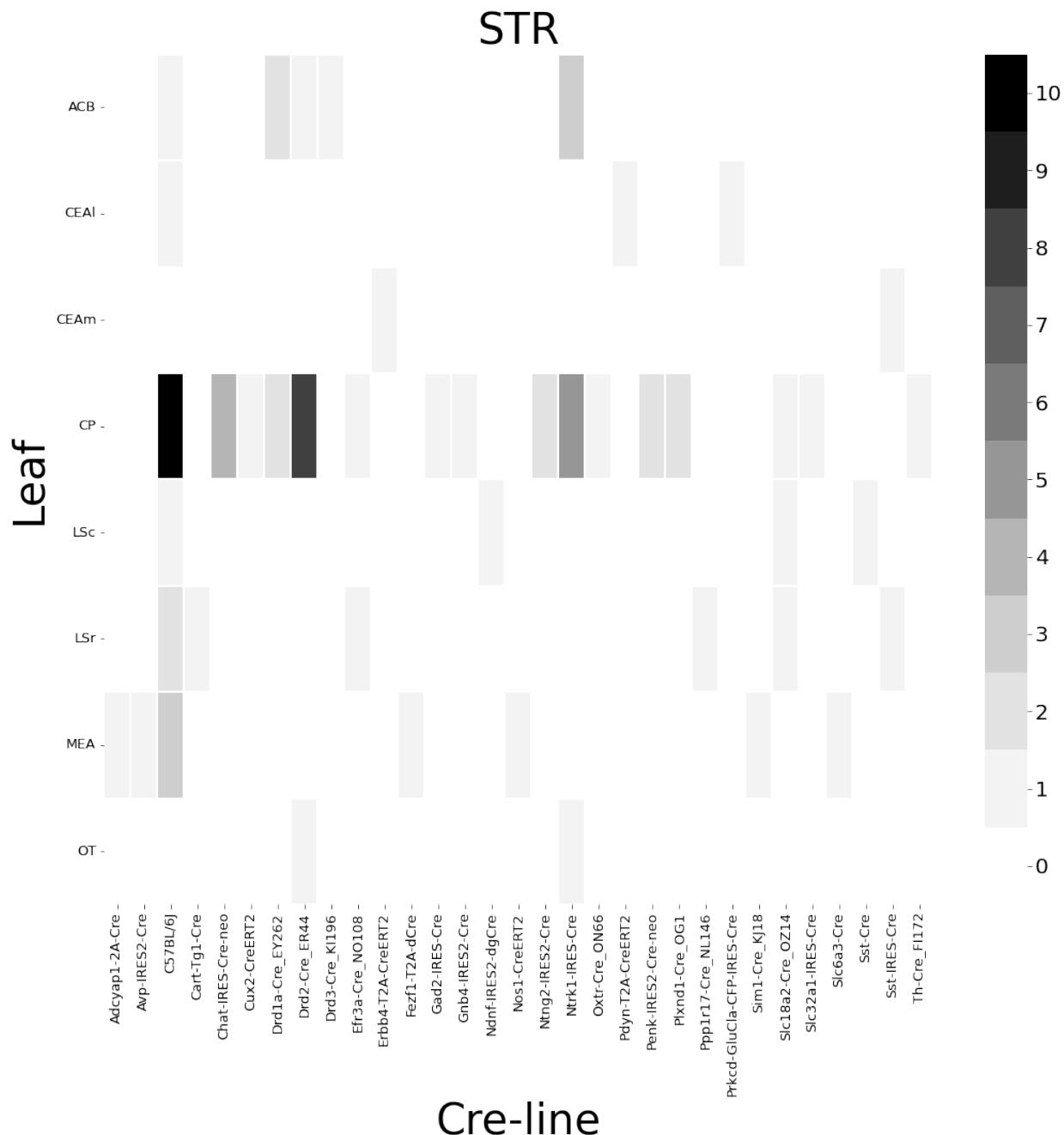


Figure 15: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

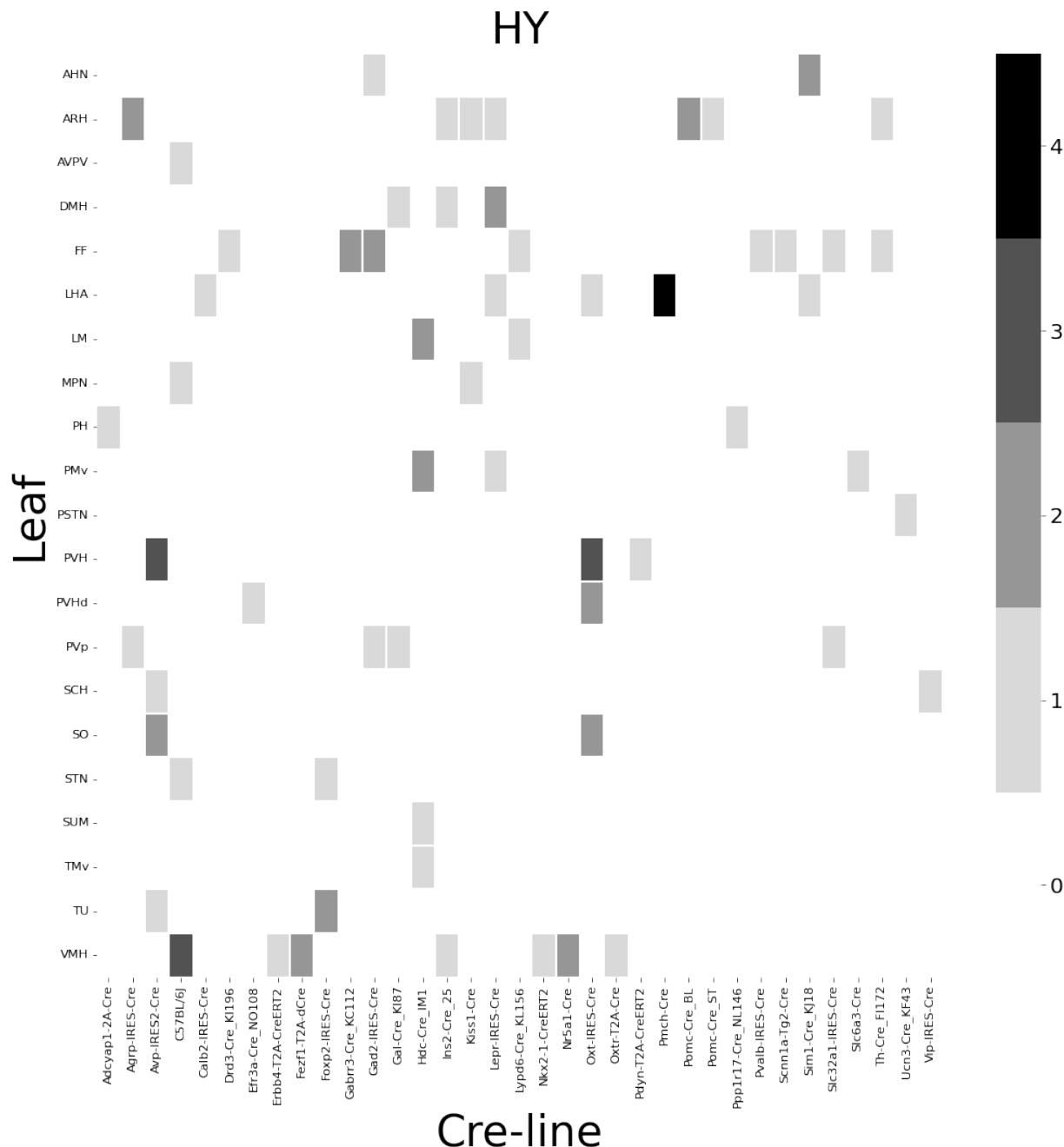


Figure 16: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

402 ***Distances between structures***

403 The distance between structures has a strong effect on the connectivity (Knox et al., 2019). For  
 404 reference, we show these distances here. Short range distances are not used in our matrix  
 405 factorization approach. This masking is methodologically novel.

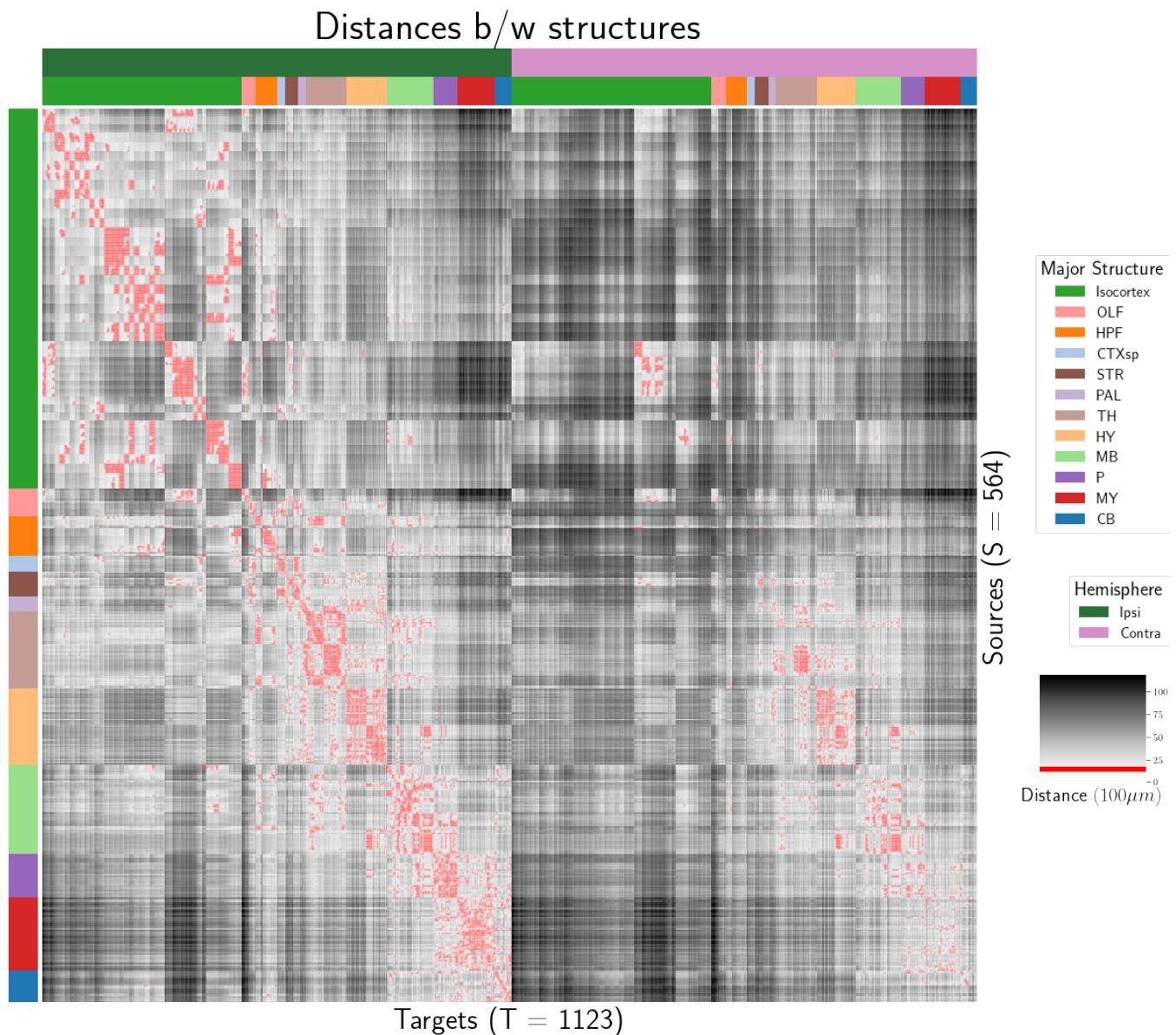


Figure 17: Distance between structures. Short-range connections are masked in red.

406 ***Model evaluation***

407 We give the sizes of our evaluation sets in leave-one-out cross-validation and additional losses using  
408 the injection-based normalization scheme from Knox et al. (2019).

409 NUMBER OF EXPERIMENTS IN EVALUATION SETS In order to compare between methods, we restrict to  
410 the smallest set of evaluation indices. That is, the set of experiments used to validate our models  
411 combinations that are those whose combination of Cre-line and injection centroid leaf are present at  
412 least twice, since one experiment at least must be held out. This means that our evaluation set is  
413 smaller in size than our overall list of experiments.

	Total	Evaluation set	Injection-thresholded evaluation
Isocortex	1128	732	709
OLF	35	17	17
HPF	122	62	59
CTXsp	7	2	2
STR	78	45	44
PAL	30	11	11
TH	83	29	29
HY	85	41	38
MB	68	18	18
P	33	8	8
MY	46	7	7
CB	36	4	4

Table 3: Number of experiments available to evaluate models in leave-one-out cross validation. The size of the evaluation set is lower than the total number of experiments since models that rely on a finer granularity of modeling have less data available to validate with, and we restrict all models to the smallest evaluation set necessitated by any of the modes. In this case, the Expected Loss and Cre-NW models require at least two experiments to be present with a combination of injection centroid structure and Cre-line for leave-one-out cross-validation. We also include results with a slightly smaller evaluation set that removes experiments without a sufficiently strong injection signal, as in Knox et al. (2019)

414 INJECTION-NORMALIZED LOSSES To compare with the injection-normalization procedure from Knox  
 415 et al. (2019), we also remove experiments with small injection, and here give results for this slightly  
 416 reduced set using injection-normalization. That is, instead of dividing the projection signal of each  
 417 experiment by its  $l_1$  norm (as we have used throughout the study), we divide by the  $l_1$  norm of the  
 418 corresponding injection signal. We find that setting a summed injection-signal threshold of 1 is  
 419 sufficient for evading pathological edge cases in this normalization, while still retaining a large  
 420 evaluation set.

$\hat{f}$	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	0.413	0.453	0.408	0.538	0.528	0.528	<b>0.396</b>
OLF	0.499	0.504	0.494	0.441	0.543	0.543	<b>0.437</b>
HPF	0.336	0.483	0.332	0.444	0.501	0.501	<b>0.321</b>
CTXsp	0.497	0.497	0.497	0.497	0.497	0.497	0.497
STR	0.359	0.386	0.359	0.364	0.433	0.433	<b>0.322</b>
PAL	0.519	0.497	0.519	0.436	0.459	0.459	<b>0.434</b>
TH	0.769	0.767	0.769	<b>0.514</b>	0.539	0.539	0.556
HY	0.414	0.439	0.414	0.441	0.452	0.452	<b>0.399</b>
MB	0.459	0.396	0.397	0.358	<b>0.324</b>	<b>0.324</b>	0.403
P	<b>0.562</b>	<b>0.562</b>	<b>0.562</b>	0.758	0.764	0.764	<b>0.562</b>
MY	0.699	0.552	0.621	<b>0.439</b>	0.578	0.578	<b>0.439</b>
CB	0.849	0.689	0.849	0.500	0.615	0.615	<b>0.495</b>

Table 4: Losses from leave-one-out cross-validation of candidate for injection-normalized regionalized connectivity on injection-thresholded evaluation set. **Bold** numbers are best for their major structure.

421 PROJECTION-NORMALIZED LOSSES ON THRESHOLDED SET We also give results for the  
 422 projection-normalization procedure from the main text on this reduced subset.

$\hat{f}$	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	0.229	0.248	0.224	0.274	0.269	0.269	<b>0.217</b>
OLF	0.193	0.233	0.191	<b>0.135</b>	0.179	0.179	0.138
HPF	0.178	0.342	<b>0.172</b>	0.212	0.235	0.235	<b>0.172</b>
CTXsp	<b>0.621</b>	<b>0.621</b>	<b>0.621</b>	<b>0.621</b>	<b>0.621</b>	<b>0.621</b>	<b>0.621</b>
STR	0.128	<b>0.117</b>	0.124	0.171	0.234	0.234	0.125
PAL	0.203	0.205	0.203	0.295	0.291	0.291	<b>0.188</b>
TH	0.673	0.664	0.673	<b>0.358</b>	0.379	0.379	0.417
HY	0.358	0.378	0.351	0.331	0.312	<b>0.312</b>	0.314
MB	0.168	0.191	<b>0.160</b>	0.199	0.202	0.202	<b>0.160</b>
P	0.292	0.292	0.292	0.299	0.299	0.299	<b>0.287</b>
MY	0.268	0.347	0.268	<b>0.167</b>	0.189	0.189	0.196
CB	0.062	0.062	0.062	0.068	0.108	0.108	<b>0.061</b>

Table 5: Losses from leave-one-out cross-validation of candidate for normalized regionalized connectivity on injection-thresholded evaluation set. **Bold** numbers are best for their major structure.

## 6 SUPPLEMENTAL METHODS

<sup>423</sup> This section consists of additional information on preprocessing of the neural connectivity data,  
<sup>424</sup> estimation of connectivity, and matrix factorization.

<sup>425</sup> ***Data preprocessing***

<sup>426</sup> Several data preprocessing steps take place prior to evaluations of the connectivity matrices. These  
<sup>427</sup> steps are described in Algorithm PREPROCESS. The arguments of this normalization process - injection  
<sup>428</sup> signals  $x(i)$ , projection signals  $y(i)$ , injection fraction  $F(i)$ , and data quality mask  $q(i)$  - were  
<sup>429</sup> downloaded using the Allen SDK, a programmatic interface to the brain connectivity data. The  
<sup>430</sup> injections and projection signals  $\mathcal{B} \rightarrow [0, 1]$  were segmented manually in histological analysis. The  
<sup>431</sup> projection signal gives the proportion of pixels within the voxel displaying fluorescence, and the  
<sup>432</sup> injection signal gives the proportion of pixels within the histologically-selected injection subset  
<sup>433</sup> displaying fluorescence. The injection fraction  $F(i) : \mathcal{B} \rightarrow [0, 1]$  gives the proportion of pixels within  
<sup>434</sup> each voxel in the injection subset. Finally, the data quality mask  $q(i) : \mathcal{B} \rightarrow \{0, 1\}$  gives the voxels that  
<sup>435</sup> have valid data.

<sup>436</sup> Our preprocessing makes use of the above ingredients, as well as several other essential steps. First,  
<sup>437</sup> we compute the weighted injection centroid

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

<sup>438</sup> where  $x(i)|_l$  is the injection density at location  $l \in \mathbb{R}^3$ . Given a regionalization  $\mathcal{R}$  from the Allen SDK,  
<sup>439</sup> we can also access regionalization map  $R : \mathcal{B} \rightarrow \mathcal{R}$ . This induces a functional of connectivities from  
<sup>440</sup> the space of maps  $\{\mathcal{X} = x : \mathcal{B} \rightarrow [0, 1]\}$

$$1_{\mathcal{R}} : \mathcal{X} \rightarrow \mathcal{R} \times \mathbb{R}_{\geq 0}$$

$$x \mapsto \sum_{l \in r} x(l) \text{ for } r \in \mathcal{R}.$$

<sup>441</sup> We also can restrict a signal to a individual structure as

$$1|_S : \mathcal{X} \rightarrow \mathcal{X}$$

$$x(l) = \begin{cases} x(l) & \text{if } l \in S \\ 0 & \text{otherwise.} \end{cases}$$

<sup>442</sup> Finally, given a vector or array  $a \in \mathbb{R}^T$ , we have the  $l1$  normalization map

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

<sup>443</sup> Denote  $m$  as the major structure containing an experiment, and define  $\odot$  for maps  $\mathcal{B} \rightarrow [0, 1]$  by e.g.

<sup>444</sup>  $(y(i) \odot q(i))|_l := (y(i)|_l)(q(i)|_l)$ . We then can write the preprocessing algorithm.

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

Injection fraction  $x_F \leftarrow x \odot F$

Data-quality censor  $y_q \leftarrow y \odot q, x_q \leftarrow x_F \odot q$

Restrict injection  $x_m = 1|_m x_q$ .

Compute centroid  $c$  from  $x_m$

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

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

**Output**  $\tilde{y}_{\mathcal{T}}, c$

**445 Estimators**

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

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

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

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

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

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

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

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

458 Several facets of the estimator are visible here. A smaller  $\gamma$  corresponds to a greater amount of  
 459 smoothing, and the index set  $I \subseteq \{1 : n\}$  generally depends on  $s$  and  $\nu$ . Varying  $\gamma$  bridges between  
 460 1-nearest neighbor prediction and averaging of all experiments in  $I$ . In Knox et al. (2019),  $I$  consisted  
 461 of experiments sharing the same brain division, i.e.  $I = I_m$ , while restricting the index set to only  
 462 include experiments with the same cell class gives the class-specific Cre-NW model. Despite this  
 463 restriction, we fit  $\gamma$  by leave-one-out cross-validation for each  $m$  rather than a smaller subset like  $s$  or  
 464  $\nu$ . That is,

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

<sup>465</sup> *The Expected-Loss estimator* Besides location of the injection centroid, cell class also influences  
<sup>466</sup> projection. Thus, we introduce method for estimating the effect of Cre-distance, which we define as  
<sup>467</sup> the distance between the projections of the mean experiment of one (Cre,leaf) pair with another.  
<sup>468</sup> Equivalently, relatively small Cre-distance defines what we call similar cell classes. This method  
<sup>469</sup> assigns a predictive weight to each pair of training points that depends both on their  
<sup>470</sup> centroid-distance and Cre-distance. This weight is determined by the expected prediction error of  
<sup>471</sup> each of the two feature types

<sup>472</sup> We define Cre-line behavior as the average regionalized projection of a Cre-line in a given structure  
<sup>473</sup> (i.e. leaf). The vectorization of categorical information is known as **target encoding**

$$\bar{y}_{\mathcal{T},s,v} := \frac{1}{|I_s \cap I_v|} \sum_{i \in (I_s \cap I_v)} y_{\mathcal{T}}(i)$$

<sup>474</sup> We then define a **Cre-distance** in a leaf to be the distance between the target-encoded projections of  
<sup>475</sup> two Cre-lines. The relative predictive accuracy of Cre-distance and centroid distance is determined by  
<sup>476</sup> fitting a surface of projection distance as a function of Cre-distance and centroid distance. For this  
<sup>477</sup> reason, we call this the Expected Loss Estimator. When we use shape-constrained B-splines to  
<sup>478</sup> estimate this weight, the weights then may be said to be used in a Nadaraya-Watson estimator. The  
<sup>479</sup> resulting weights are then utilized in a Nadaraya-Watson estimator in a final prediction step.

<sup>480</sup> In mathematical terms, our full feature set consists of the centroid coordinates and the  
<sup>481</sup> target-encoded means of the combinations of virus type and injection-centroid structure. That is,

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

<sup>482</sup>  $f^*$  is defined as in (2). The expected loss estimator is then

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

<sup>483</sup> where

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

<sup>484</sup> and  $s$  is the structure containing  $l$ .

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

490 We therefore use a linear generalized additive model of shape-constrained B-splines to estimate  $g$   
 491 (Eilers & Marx, 1996). This is a method for generating a predictive model  $g$  that minimizes the loss of

$$\sum_{i, i' \in S} \| \|y_{\mathcal{T}}(i) - y_{\mathcal{T}}(i')\|_2 - \sum_{q=1}^Q \rho_q B_q(\|c(i') - c(i)\|_2, \|\bar{y}_{\mathcal{T}, s, v} - \bar{y}_{\mathcal{T}, s, v}(i)\|) \|_2$$

492 given the constraints on  $g$ . That is, given all pairs of experiments with injection centroid in the same  
 493 structure,  $g$  gives a prediction of the distance between their projections made using the distance  
 494 between the average behavior of their Cre-lines given their injection centroid, and the distance  
 495 between their injection centroids. In particular,  $g$  is the empirically best such function within the  
 496 class of  $B$ -splines, which Similarly to the Nadaraya-Watson model, we make the decision to fit a  $g$   
 497 separately for each major brain division, and select  $\gamma$  as in 2. We set  $Q = 10$  and leave validation of this  
 498 parameter, as well as the precise nature of the polynomial  $B$ -spline terms  $B_q$  out of the scope of this  
 499 paper. Empirically this leads to a smooth surface using the pyGAM Python package (Servén &  
 500 Brummitt, 2018).

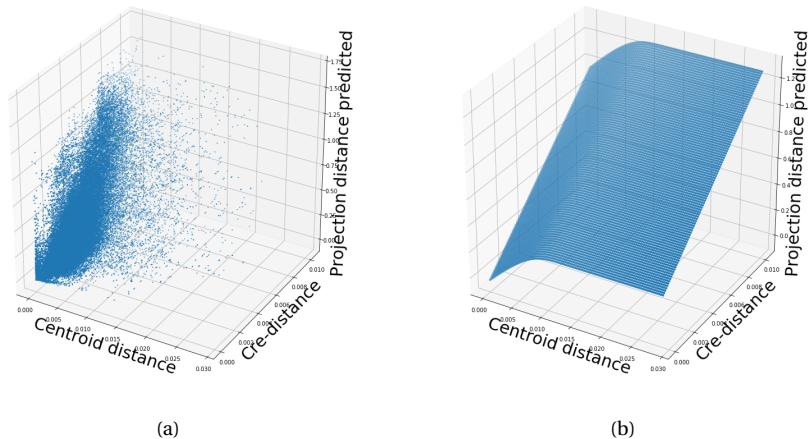


Figure 18: Fitting  $g$ . 18a Distribution of projection errors against centroid distance and cre-distance in Isocortex. 18b estimated  $\hat{g}$  using B-splines. Projection distance is  $\|y_{\mathcal{T}}(i) - y_{\mathcal{T}}(i')\|_2$ , Cre-distance is  $\|\bar{y}_{\mathcal{T},s,v} - \bar{y}_{\mathcal{T},s,v}(i)\|_2$ , and centroid-distance is  $\|c(i') - c(i)\|_2$ .

501 JUSTIFICATION OF SHAPE CONSTRAINT     The shape-constrained expected-loss estimator introduced  
 502 in this paper is, to our knowledge, novel. It should be considered an alternative method to the classic  
 503 weighted kernel method (Cai, 2001; Salha & El Shekh Ahmed, 2015). While we do not attempt a  
 504 detailed theoretical study of this estimator, we do establish the need for the shape constraint in our  
 505 spline estimator. Though this fact is probably well known, we prove a (slightly stronger) version here  
 506 for completeness.

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

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

- 511    ▪ The first property of a metric is that  $f(x, x') = 0 \iff x = x'$ . The left implication:  
 512        $x = x' \implies f(x_1, x'_1, \dots, x_n, x'_n) = g(0, \dots, 0)$ , since  $d$  are metrics. Then, since  $g(0) = 0$ , we have that  
 513        $f(x, x') = 0$ . The right implication:  $f(x, x') = 0 \implies d = 0 \implies x = x'$  since  $d$  are metrics.
- 514    ▪ The second property of a metric is that  $f(x, x') = f(x', x)$ . This follows immediately from the  
 515       symmetry of the  $d_i$ , i.e.  $f(x, x') = f(x_1, x'_1, \dots, x_n, x'_n) = g(d_1(x_1, x'_1), \dots, d_n(x_n, x'_n)) =$   
 516        $g(d_1(x'_1, x_1), \dots, d_n(x'_n, x_n)) = f(x'_1, x_1, \dots, x'_n, x_n) = f(x', x)$ .
- 517    ▪ The third property of a metric is the triangle inequality:  $f(x, x') \leq f(x, x^*) + f(x^*, x')$ . To show this  
 518       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  
 519       non-decreasing and by the triangle inequality of  $d$ . Then, since  $g$  is concave,  
 520        $g(d(x, x^*) + d(x^*, x')) \leq g(d(x, x^*)) + g(d(x^*, x')) = f(x, x^*) + f(x^*, x')$ .

522 ***Setting a lower detection threshold***

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

531 We set this threshold with respect to the sum of Type 1 and Type 2 errors

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

532 We then select the  $\tau$  that minimizes  $\iota$ . Results for this approach are given in Supplemental Section 7.

533 ***Decomposing the connectivity matrix***

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

539 *Non-negative matrix factorization* As discussed in Knox et al. (2019), one of the most basic processes  
 540 underlying the observed connectivity is the tendency of each source region to predominantly project  
 541 to proximal regions. For example, the heatmap in Supplemental Figure 17 shows that the pattern of  
 542 infrastructure distances resembles the connectivity matrix in 2. These connections are biologically  
 543 meaningful, but also unsurprising, and their relative strength biases learned latent coordinate  
 544 representations away from long-range structures. For this reason, we establish a  $1500 \mu\text{m}$  'distal'  
 545 threshold within which to exclude connections for our analysis.

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

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

548 The mask  $\mathbf{1}_M$  specifies this objective for detecting patterns in long-range connections. We note the  
 549 existence of NMF with alternative norms for certain marginal distributions, but leave utilization of  
 550 this approach for future work (Brunet et al., 2004).

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

<sup>556</sup> *Cross-validating NMF* We review cross-validation for NMF following (Perry, 2009). In summary, a  
<sup>557</sup> NMF model is first fit on a reduced data set, and an evaluation set is held out. After random masking  
<sup>558</sup> of the evaluation set, the loss of the learned model is then evaluated on the basis of successful  
<sup>559</sup> reconstruction of the held-out values. This procedure is performed repeatedly, with replicates of  
<sup>560</sup> random masks at each tested dimensionality  $q$ . This determines the point past which additional  
<sup>561</sup> hidden units provide no additional value for reconstructing the original signal.

<sup>562</sup> The differentiating feature of cross-validation for NMF compared with supervised learning is the  
<sup>563</sup> randomness of the masking matrix  $1_M$ . Cross-validation for supervised learning generally leaves out  
<sup>564</sup> entire observations, but this is insufficient for our situation. This is because, given  $W$ , our  $H$  is the  
<sup>565</sup> solution of a regularized non-negative least squares optimization problem

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

<sup>566</sup> The negative effects of an overfit model can therefore be optimized away from on the evaluation set.

We therefore 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

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

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

$$\hat{q} = \operatorname{arg\,min}_q \epsilon_q.$$

<sup>567</sup> *Stabilizing NMF* The NMF program is non-convex, and, empirically, individual replicates will not  
<sup>568</sup> converge to the same optima. One solution therefore is to run multiple replicates of the NMF  
<sup>569</sup> algorithm and cluster the resulting vectors. This approach raises the questions of how many clusters  
<sup>570</sup> to use, and how to deal with stochasticity in the clustering algorithm itself. We address this issue  
<sup>571</sup> 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

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

572 A review of this approach is found in von Luxburg (2010b). Intuitively, archetype vectors that cluster  
 573 together frequently over clustering replicates indicate the presence of a stable clustering. For  $d$ , we  
 574 utilize the adjusted Rand Index - a simple dissimilarity measure between clusterings. Note that we  
 575 expect to select slightly more than the  $q$  components suggested by cross-validation, since archetype  
 576 vectors which appear in one NMF replicate generally should appear in others. We then select the  $q$   
 577 clusters with the most archetype vectors - the most stable NMF results - and take the median of each  
 578 cluster to create a sparse representative archetype (Kotliar et al., 2019; Wu et al., 2016). We then find  
 579 the according  $H$  using Program 3. Experimental results for these cross-validation and stability  
 580 selection approaches are given in Supplemental Section 7.

## 7 SUPPLEMENTAL EXPERIMENTS

581 The supplemental experiments show results on lower limit of detection, performance of our estimator  
582 for different regions and cell-classes, hierarchical clustering of connectivities, and stability and  
583 component analysis of our NMF results.

584 *Setting detection threshold  $\tau$*

585 We give results on the false detection rate at different limits of detection. These conclusively show that  
586  $10^{-6}$  is the good threshold for our normalized data.

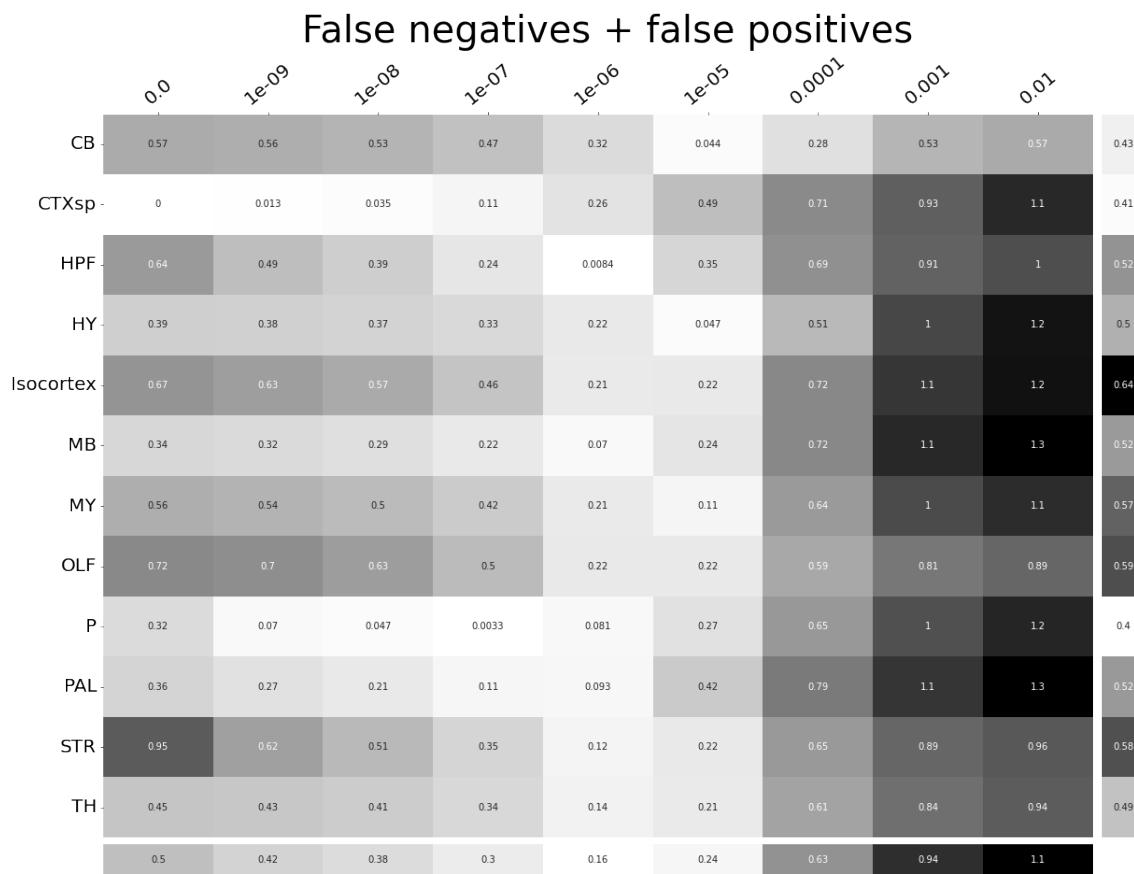


Figure 19:  $\tau$  at different limits of detection in different major structures.  $10^{-6}$  is the optimal detection threshold.

587 ***Loss subsets***

588 We report model accuracies for our *EL* model by neuron class and structure. These expand upon the  
 589 results in Table 5 and give more specific information about the quality of our estimates. CTXsp is  
 590 omitted due to the small evaluation set.



Figure 20: Weighted loss for Cre-leaf combinations in CB. Missing values are omitted. For example, this figure has one present and three missing values. Row and column averages are also plotted.

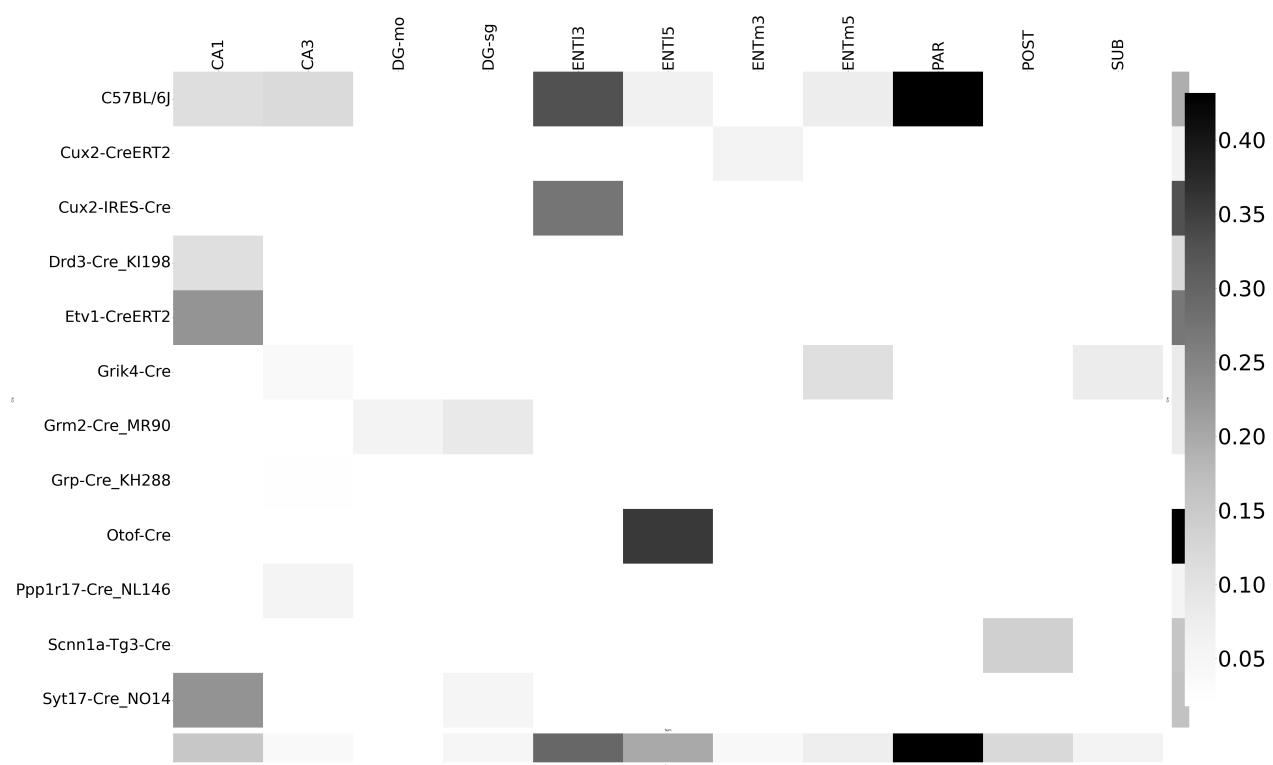


Figure 21: Weighted loss for Cre-leaf combinations in HPF. Missing values are omitted. Row and column averages are also plotted.

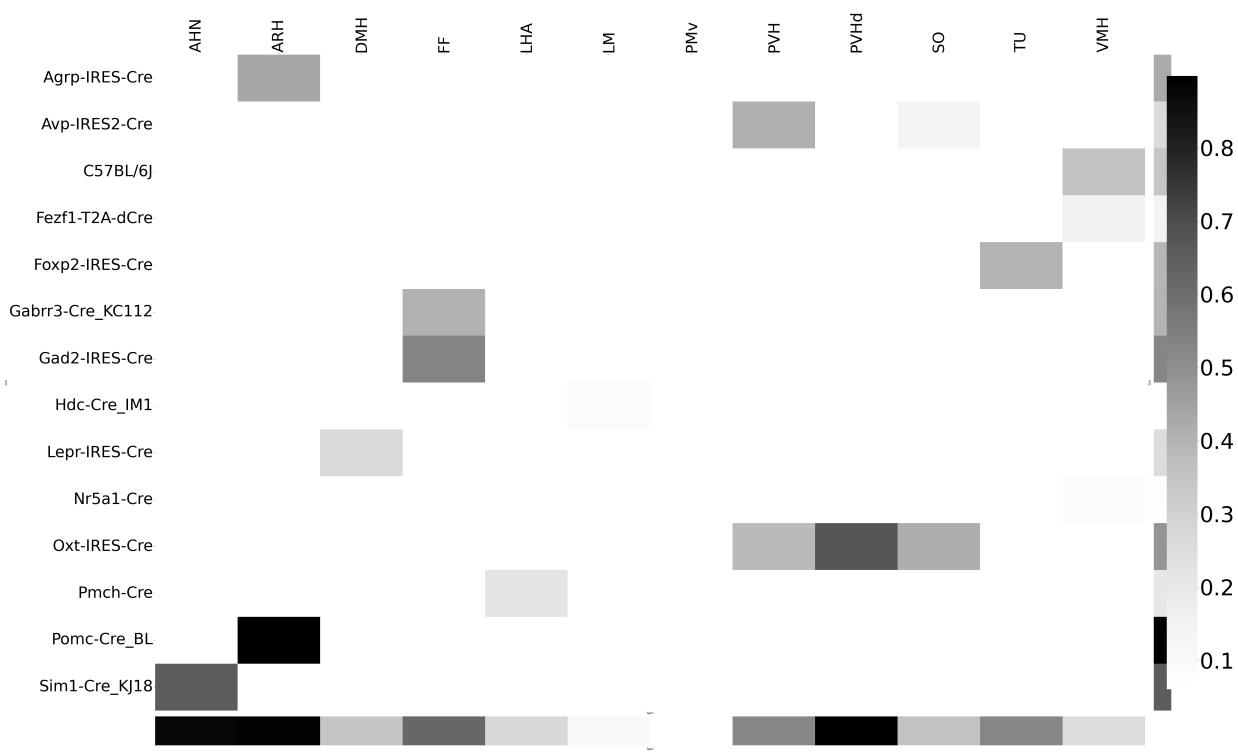


Figure 22: Weighted loss for Cre-leaf combinations in HY. Missing values are omitted. Row and column averages are also plotted.

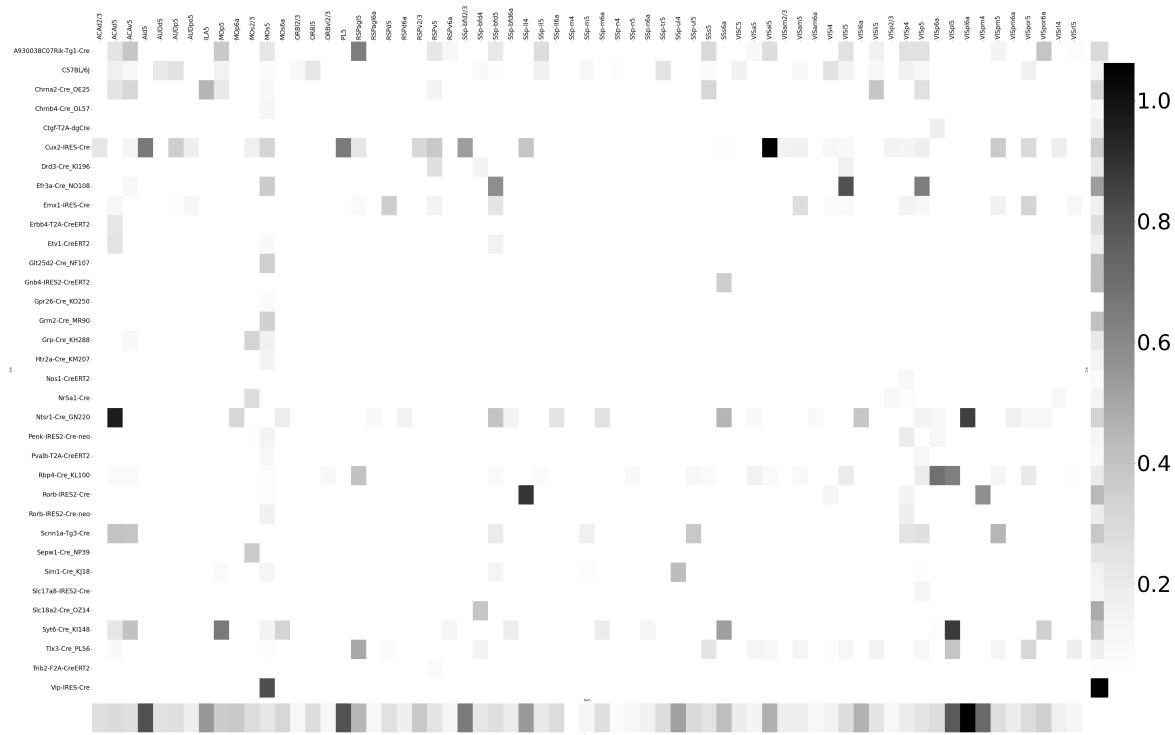


Figure 23: Weighted loss for Cre-leaf combinations in Isocortex. Missing values are omitted. Row and column averages are also plotted.

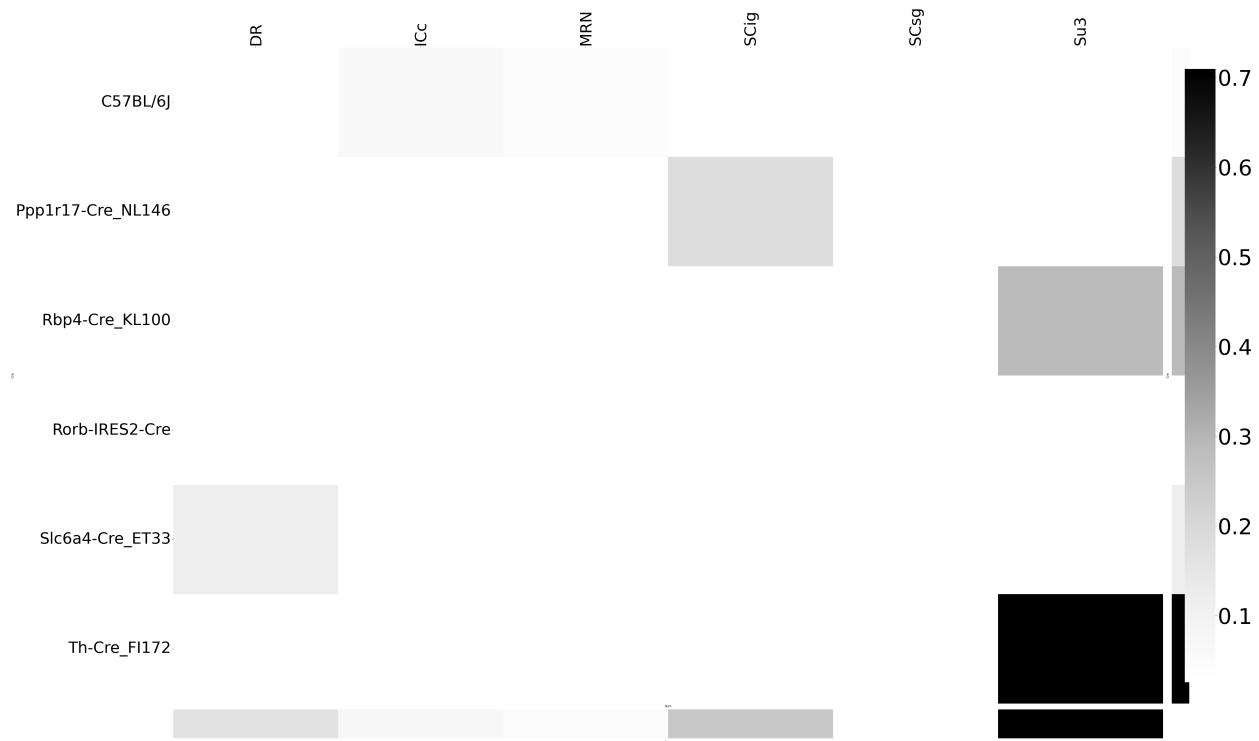


Figure 24: Weighted loss for Cre-leaf combinations in MB. Missing values are omitted. Row and column averages are also plotted.

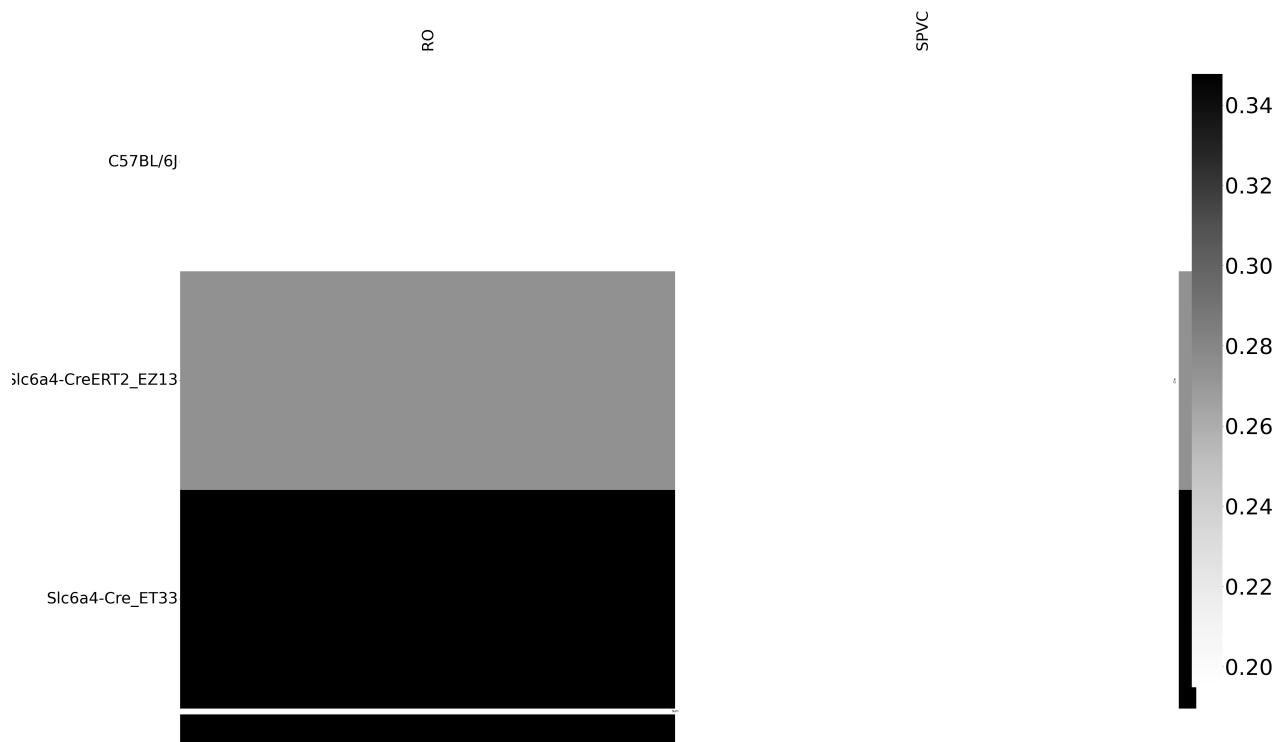


Figure 25: Weighted loss for Cre-leaf combinations in MY. Missing values are omitted. Row and column averages are also plotted.

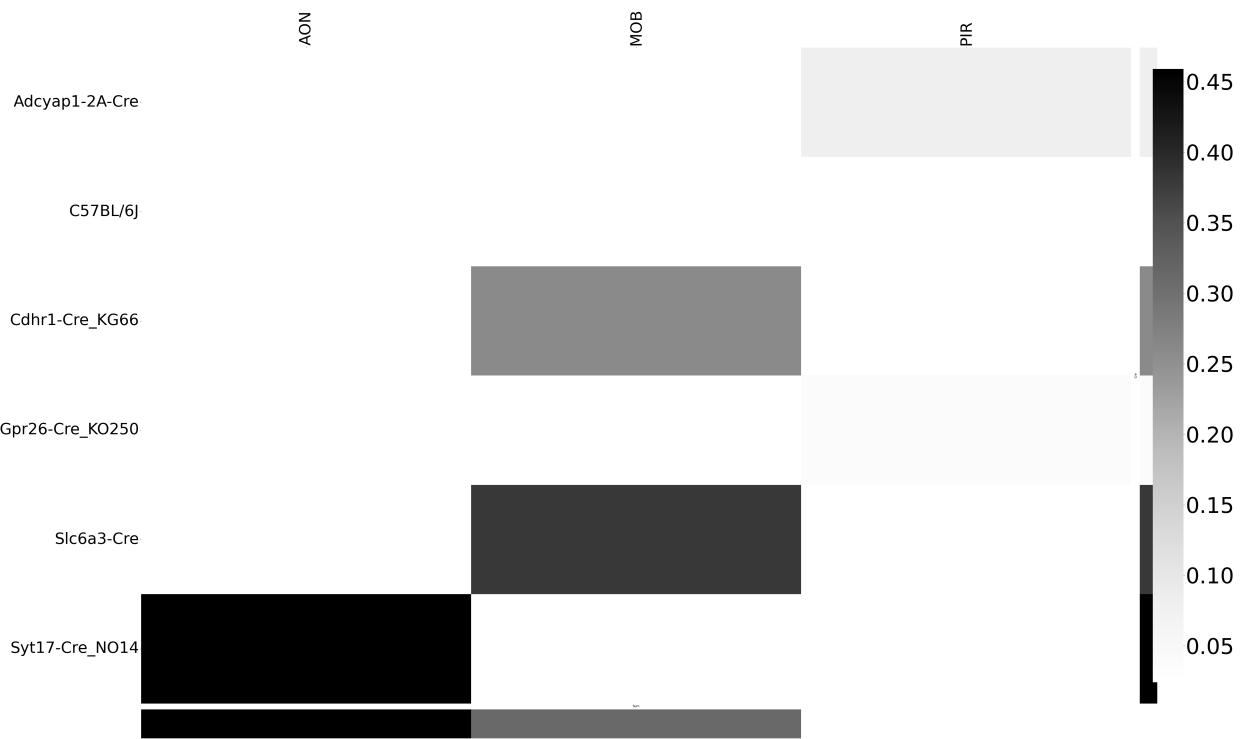


Figure 26: Weighted loss for Cre-leaf combinations in OLF. Missing values are omitted. Row and column averages are also plotted.

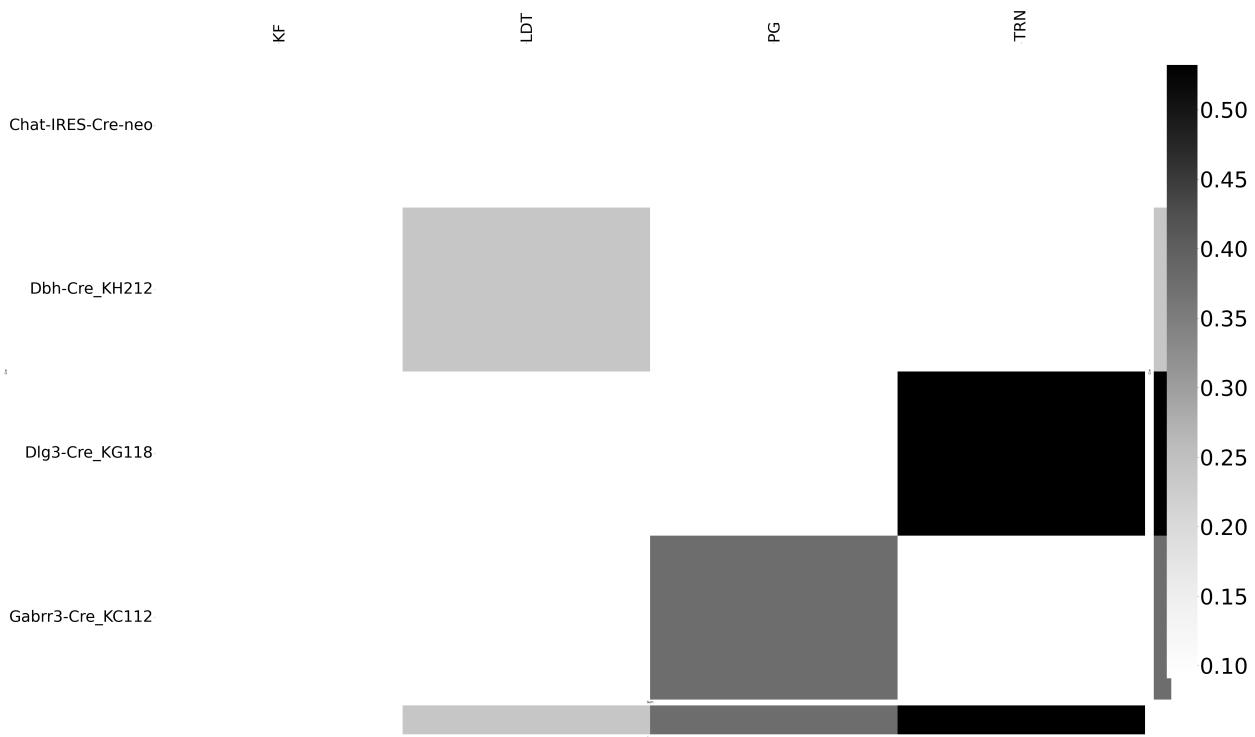


Figure 27: Weighted loss for Cre-leaf combinations in P. Missing values are omitted. Row and column averages are also plotted.

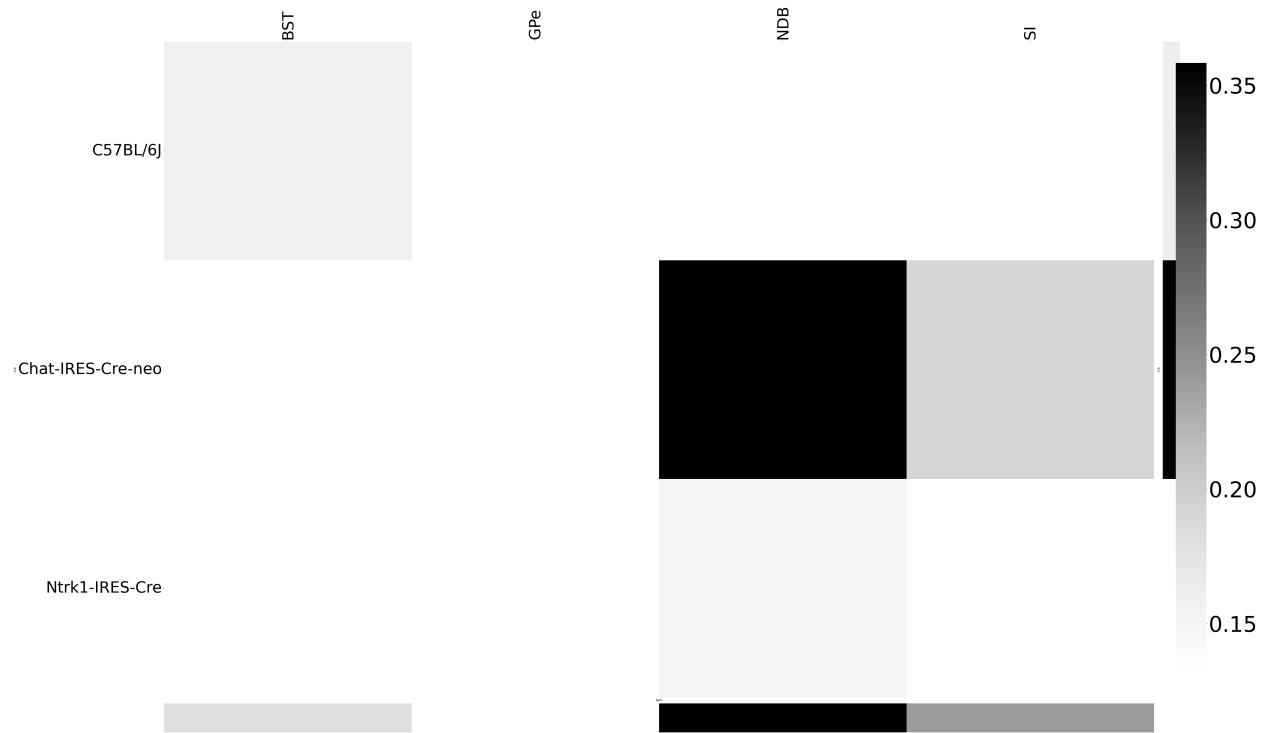


Figure 28: Weighted loss for Cre-leaf combinations in PAL. Missing values are omitted. Row and column averages are also plotted.

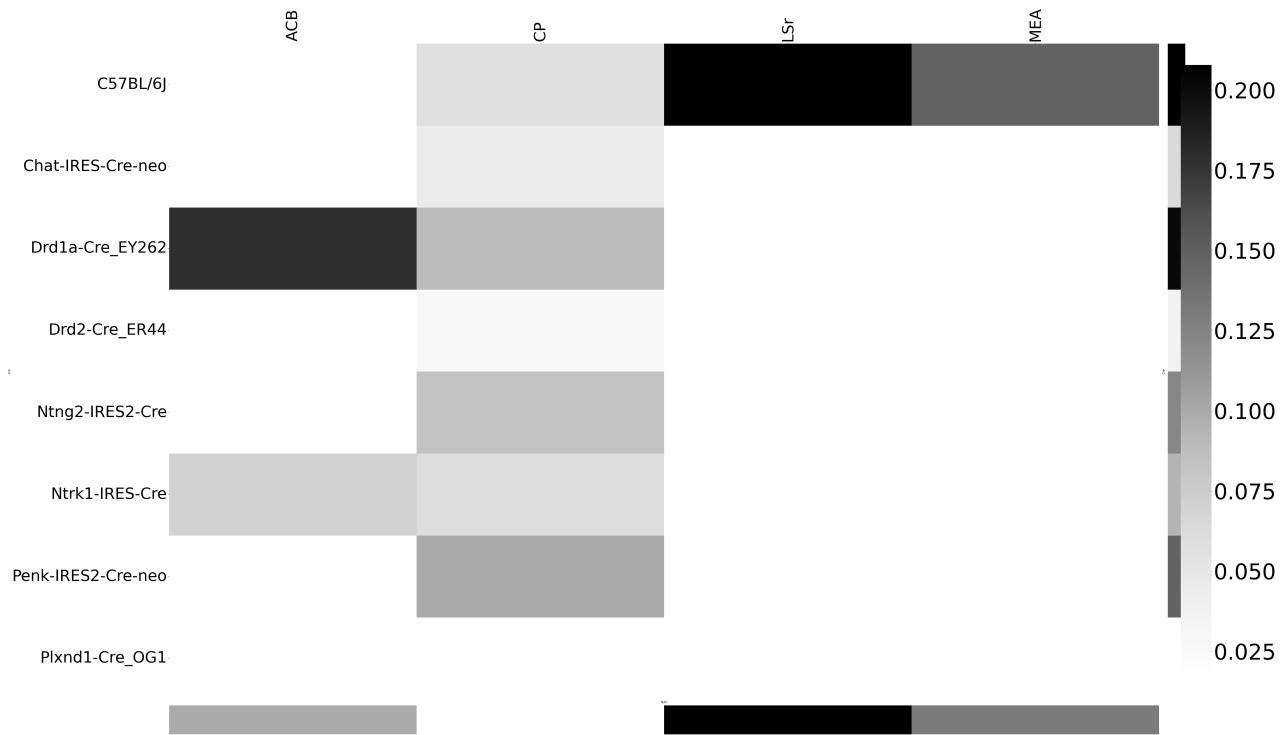


Figure 29: Weighted loss for Cre-leaf combinations in STR. Missing values are omitted. Row and column averages are also plotted.

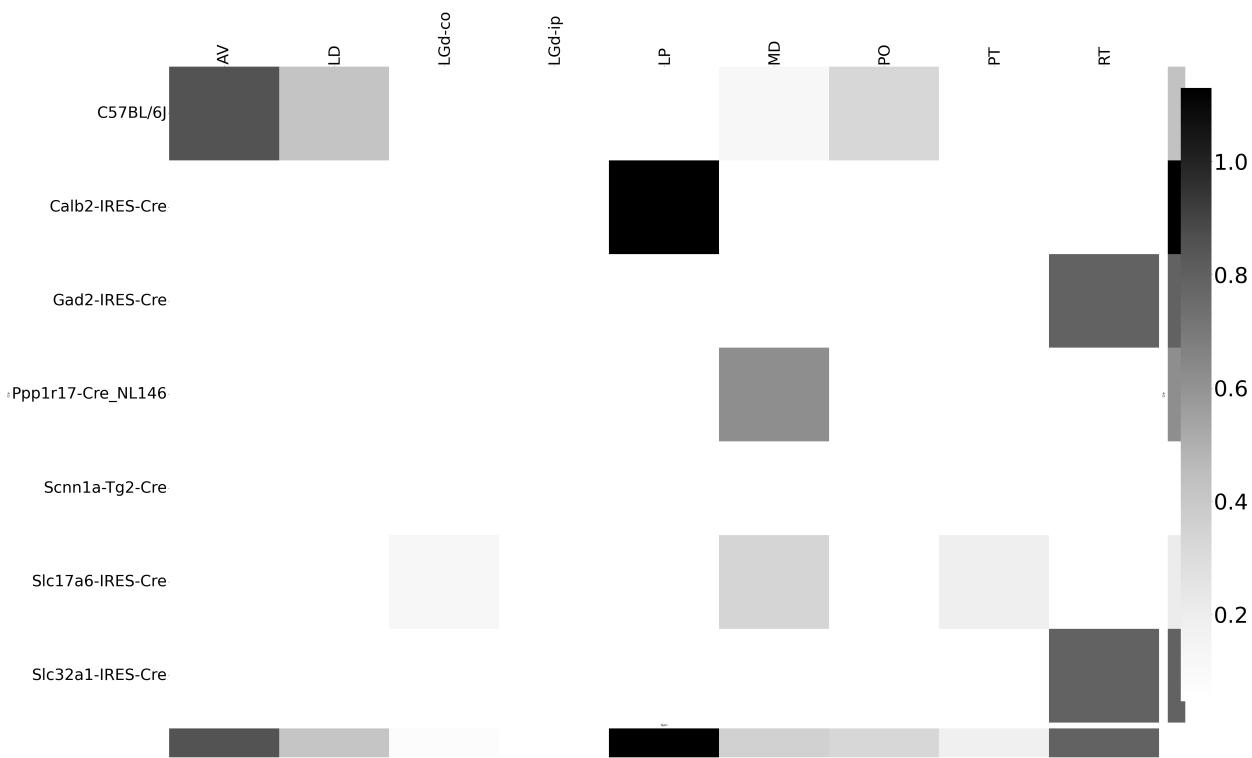


Figure 30: Weighted loss for Cre-leaf combinations in TH. Missing values are omitted. Row and column averages are also plotted.

591 ***Cell-type specificity***

592 We performed hierarchical clustering using the default method in Seaborn (Waskom, 2021) to  
 593 investigate shared projection patterns across Cre-lines. That is, we used agglomerative clustering with  
 594 Ward's criterion Hastie et al. (2009); Lalloué et al. (2013). This showed clustering of Ntsr1 projections  
 595 to Thalamic nuceli.

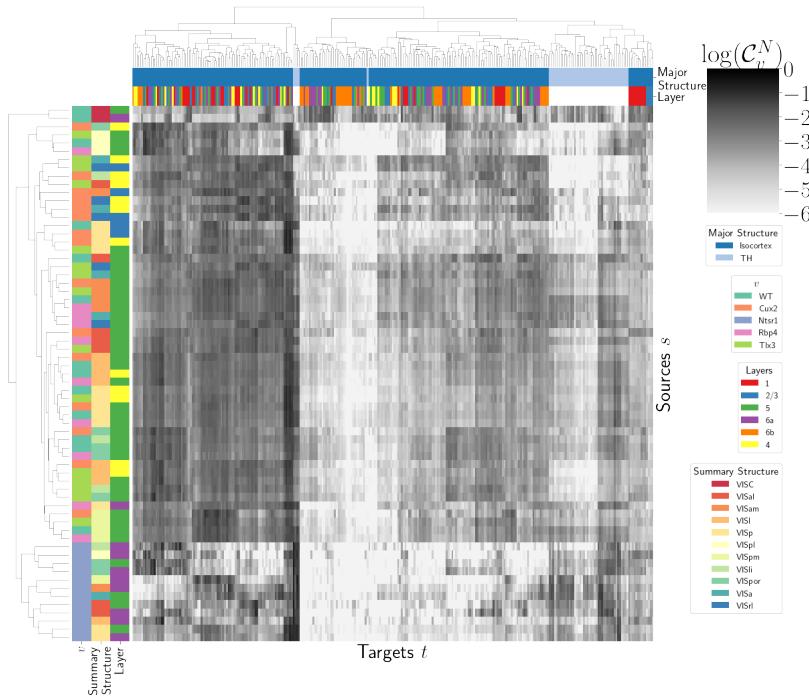


Figure 31: Hierarchical clustering of connectivity strengths from visual cortex cell-types to cortical and thalamic targets. Cre-line, summary structure, and layer are labelled on the sources. Major brain division and layer are labelled on the targets.

596 **Matrix Factorization**

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

600 *Cross-validation* We set  $\alpha = 0.002$  and run Program 2 on  $\mathcal{C}_{wt}$ . We use a random mask with  $p = .3$  to  
 601 evaluate prediction accuracy of models trained on the unmasked data on the masked data. To  
 602 account for stochasticity in the NMF algorithm, we run  $R = 8$  replicates at each potential dimension  $q$ .  
 603 The lowest mean test error was observed at  $\hat{q} = 70$ , indicating that even more components could be  
 604 estimated. However, the low decrease in reconstruction error at higher values of  $q$  and need for  
 605 brevity in our figures motivated us to choose  $q = 15$  for the purposes of display.

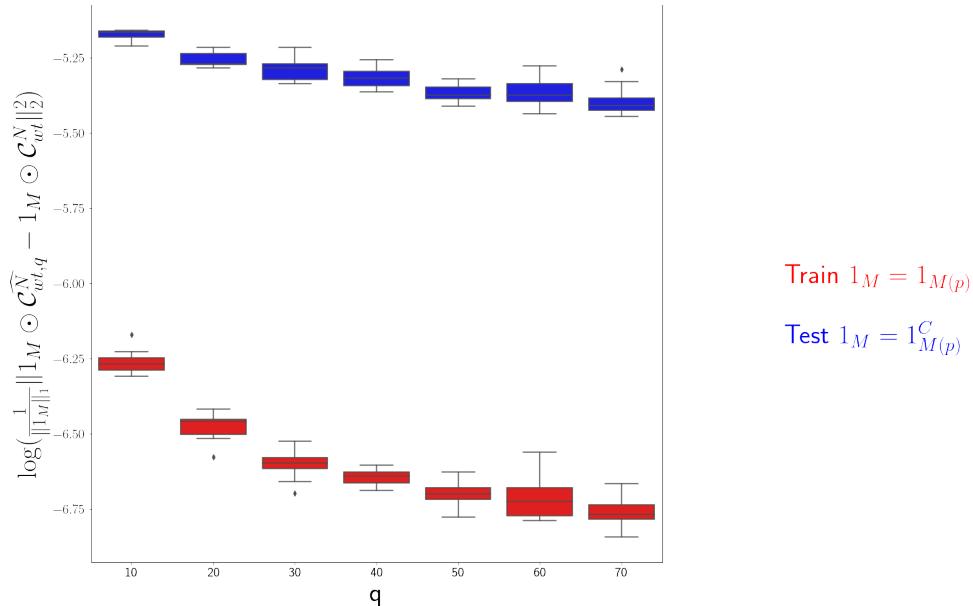


Figure 32: Train and test error using NMF decomposition.

606 *Stability* To address the instability of the NMF algorithm in identifying components, we k-means  
 607 cluster components over  $R = 10$  replicates with  $k \in \{10, 15, 20, 25, 30\}$ . Since the clustering is itself

608 unstable, we repeat the clustering 25 times and select the  $k$  with the largest Rand index, a standard  
 609 method of clustering stability (Meila, 2007; Rand, 1971).

610	q	10.000000	20.000000	30.000000	40.000000	50.000000
	Rand index	0.772544	0.844981	<b>0.932957</b>	0.929827	0.885862

611 Since  $k$ -means is most stable at  $k = 30$ , we cluster the  $qR = 150$  components into 30 clusters and  
 612 select the 15 clusters appearing in the most replicates.

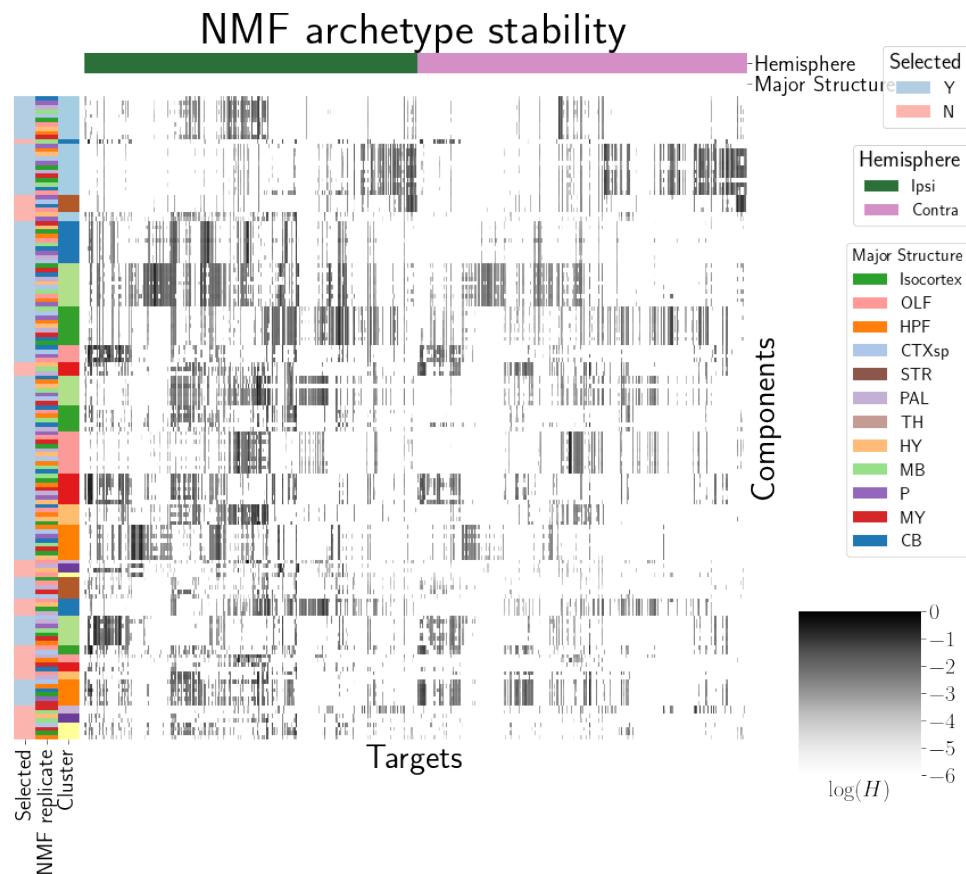


Figure 33: Stability of NMF results across replicates. Replicate and NMF component are shown on rows. Components that are in the top 15 are also indicated.

613 We plot the medians of these components in Figure 4a and in the main text. These are the  
 614 connectivity archetypes. We then fit a non-negative least squares (the second step in the standard  
 615 NMF optimization algorithm) to determine  $W$  (Lee & Seung, 2000)

616 *Association with Cre-line* Finally, we show the association of our learned archetypes with projections  
 617 from sources with injection centroids from the Ntsr1, Cux2, Rbp4, and Tlx3 Cre-lines. While we make  
 618 no statistical claims on these associations, the distribution of cosine similarities of sources from each  
 619 of the Cre-line lines shows an association of learned archetypes with Cre-line.

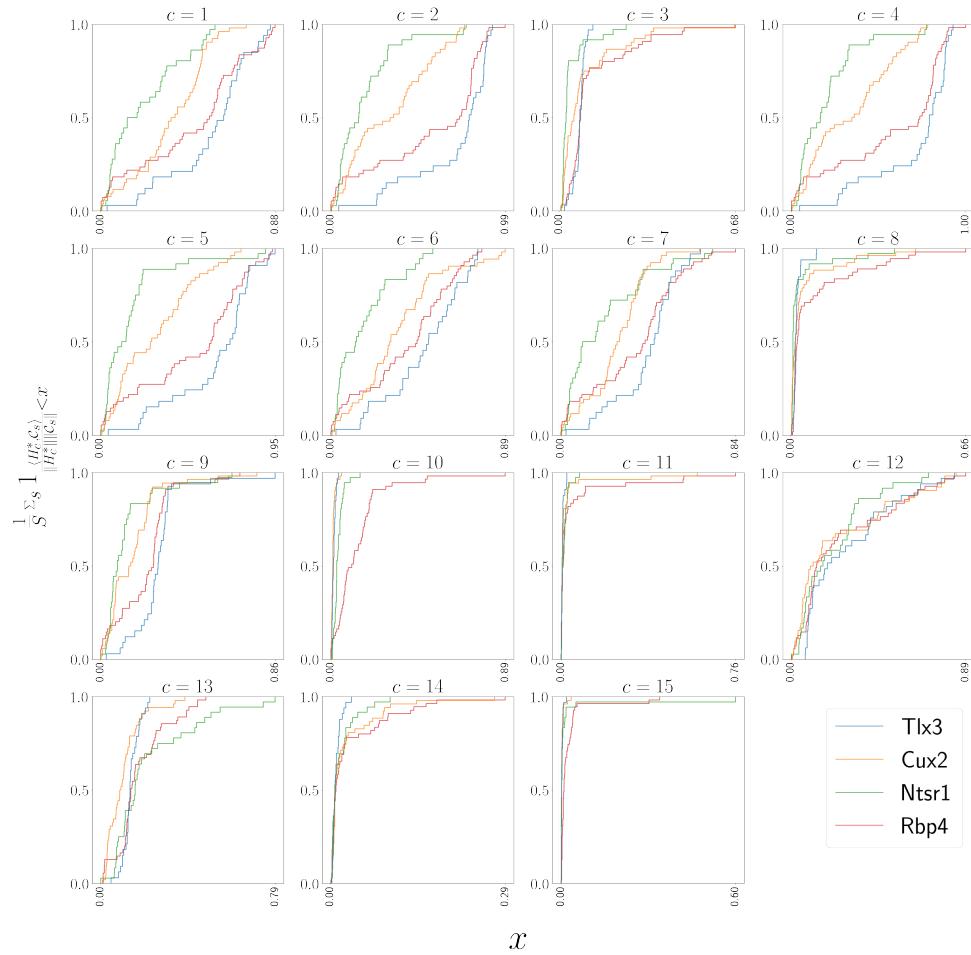


Figure 34: Empirical cumulative distributions of cosine similarities between source structures and connectivity components for four different Cre-lines.

620

**621 REFERENCES**

622

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

625 Cai, Z. (2001). Weighted Nadaraya-Watson regression estimation. *Stat. Probab. Lett.*, 51(3), 307–318.

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

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

630 Devarajan, K. (2008). Nonnegative matrix factorization: an analytical and interpretive tool in computational biology.  
631 *PLoS Comput. Biol.*, 4(7), e1000029.

632 Eilers, P. H. C., & Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *SSO Schweiz. Monatsschr.*  
633 *Zahnheilkd.*, 11(2), 89–121.

634 Gämänuť, R., Kennedy, H., Toroczkai, Z., Ercsey-Ravasz, M., Van Essen, D. C., Knoblauch, K., & Burkhalter, A. (2018). The  
635 mouse cortical connectome, characterized by an Ultra-Dense cortical graph, maintains specificity by distinct  
636 connectivity profiles. *Neuron*, 97(3), 698–715.e10.

637 Gao, Y., Zhang, X., Wang, S., & Zou, G. (2016). Model averaging based on leave-subject-out cross-validation. *J. Econom.*,  
638 192(1), 139–151.

639 Groeneboom, P., & Jongbloed, G. (2018). Some developments in the theory of shape constrained inference. *Stat. Sci.*,  
640 33(4), 473–492.

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

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

- 645 Harris, K. D., Mihalas, S., & Shea-Brown, E. (2016). Nonnegative spline regression of incomplete tracing data reveals high  
646 resolution neural connectivity.
- 647 Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer New York.
- 648 Huang, K. W., Ochandarena, N. E., Philson, A. C., Hyun, M., Birnbaum, J. E., Cicconet, M., & Sabatini, B. L. (2019).  
649 Molecular and anatomical organization of the dorsal raphe nucleus. *Elife*, 8.
- 650 Jackson, K. L., Dayton, R. D., Deverman, B. E., & Klein, R. L. (2016). Better targeting, better efficiency for Wide-Scale  
651 neuronal transduction with the synapsin promoter and AAV-PHP.B. *Front. Mol. Neurosci.*, 9, 116.
- 652 Jeong, M., Kim, Y., Kim, J., Ferrante, D. D., Mitra, P. P., Osten, P., & Kim, D. (2016). Comparative three-dimensional  
653 connectome map of motor cortical projections in the mouse brain. *Sci. Rep.*, 6, 20072.
- 654 Knox, J. E., Harris, K. D., Graddis, N., Whitesell, J. D., Zeng, H., Harris, J. A., ... Mihalas, S. (2019). High-resolution  
655 data-driven model of the mouse connectome. *Netw Neurosci*, 3(1), 217–236.
- 656 Kotliar, D., Veres, A., Nagy, M. A., Tabrizi, S., Hodis, E., Melton, D. A., & Sabeti, P. C. (2019). Identifying gene expression  
657 programs of cell-type identity and cellular activity with single-cell RNA-Seq. *Elife*, 8.
- 658 Kuan, L., Li, Y., Lau, C., Feng, D., Bernard, A., Sunkin, S. M., ... Ng, L. (2015). Neuroinformatics of the allen mouse brain  
659 connectivity atlas. *Methods*, 73, 4–17.
- 660 Kügler, S., Kilic, E., & Bähr, M. (2003). Human synapsin 1 gene promoter confers highly neuron-specific long-term  
661 transgene expression from an adenoviral vector in the adult rat brain depending on the transduced area. *Gene Ther.*,  
662 10(4), 337–347.
- 663 Lalloué, B., Monnez, J.-M., Padilla, C., Kihal, W., Le Meur, N., Zmirou-Navier, D., & Deguen, S. (2013). A statistical  
664 procedure to create a neighborhood socioeconomic index for health inequalities analysis. *Int. J. Equity Health*, 12, 21.
- 665 Lee, D., & Seung, H. S. (2000). Algorithms for non-negative matrix factorization. *Adv. Neural Inf. Process. Syst.*, 13.
- 666 Li, X., Yu, B., Sun, Q., Zhang, Y., Ren, M., Zhang, X., ... Qiu, Z. (2018). Generation of a whole-brain atlas for the  
667 cholinergic system and mesoscopic projectome analysis of basal forebrain cholinergic neurons. *Proc. Natl. Acad. Sci.  
668 U. S. A.*, 115(2), 415–420.

- 669 Llano, D. A., & Sherman, S. M. (2008). Evidence for nonreciprocal organization of the mouse auditory  
670 thalamocortical-corticothalamic projection systems. *J. Comp. Neurol.*, 507(2), 1209–1227.
- 671 Lotfollahi, M., Naghipourfar, M., Theis, F. J., & Alexander Wolf, F. (2019). Conditional out-of-sample generation for  
672 unpaired data using trVAE.
- 673 Meila, M. (2007). Comparing clusterings—an information based distance. *J. Multivar. Anal.*, 98(5), 873–895.
- 674 Mohammadi, S., Ravindra, V., Gleich, D. F., & Grama, A. (2018). A geometric approach to characterize the functional  
675 identity of single cells. *Nat. Commun.*, 9(1), 1516.
- 676 Muzerelle, A., Scotto-Lomassese, S., Bernard, J. F., Soiza-Reilly, M., & Gaspar, P. (2016). Conditional anterograde tracing  
677 reveals distinct targeting of individual serotonin cell groups (B5-B9) to the forebrain and brainstem. *Brain Struct.  
678 Funct.*, 221(1), 535–561.
- 679 Oh, S. W., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., ... Zeng, H. (2014). A mesoscale connectome of the  
680 mouse brain. *Nature*, 508(7495), 207–214.
- 681 Perry, P. O. (2009). Cross-Validation for unsupervised learning.
- 682 Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *J. Am. Stat. Assoc.*, 66(336), 846–850.
- 683 Ren, J., Friedmann, D., Xiong, J., Liu, C. D., Ferguson, B. R., Weerakkody, T., ... Luo, L. (2018). Anatomically defined and  
684 functionally distinct dorsal raphe serotonin sub-systems. *Cell*, 175(2), 472–487.e20.
- 685 Ren, J., Isakova, A., Friedmann, D., Zeng, J., Grutzner, S. M., Pun, A., ... Luo, L. (2019). Single-cell transcriptomes and  
686 whole-brain projections of serotonin neurons in the mouse dorsal and median raphe nuclei. *Elife*, 8.
- 687 Salha, R. B., & El Shekh Ahmed, H. I. (2015). Reweighted Nadaraya-Watson estimator of the regression mean.  
688 *International Journal of Statistics and Probability*, 4(1).
- 689 Saul, L. K., & Roweis, S. T. (2003). Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *J.  
690 Mach. Learn. Res.*, 4(Jun), 119–155.
- 691 Saunders, A., Johnson, C. A., & Sabatini, B. L. (2012). Novel recombinant adeno-associated viruses for cre activated and  
692 inactivated transgene expression in neurons. *Front. Neural Circuits*, 6, 47.

- 693 Servén, D., & Brummitt, C. (2018). *pygam: Generalized additive models in python*. doi: 10.5281/zenodo.1208723
- 694 von Luxburg, U. (2010a). Clustering stability: An overview.
- 695 von Luxburg, U. (2010b). Clustering stability: An overview.
- 696 Wang, Q., Ding, S.-L., Li, Y., Royall, J., Feng, D., Lesnar, P., ... Ng, L. (2020). The allen mouse brain common coordinate  
697 framework: A 3D reference atlas. *Cell*, 181(4), 936–953.e20.
- 698 Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60), 3021. Retrieved  
699 from <https://doi.org/10.21105/joss.03021> doi: 10.21105/joss.03021
- 700 Watson, C., Paxinos, G., & Puelles, L. (2012). The mouse nervous system..
- 701 Wu, S., Joseph, A., Hammonds, A. S., Celiker, S. E., Yu, B., & Frise, E. (2016). Stability-driven nonnegative matrix  
702 factorization to interpret spatial gene expression and build local gene networks. *Proc. Natl. Acad. Sci. U. S. A.*, 113(16),  
703 4290–4295.
- 704 Zaborszky, L., Csordas, A., Mosca, K., Kim, J., Gielow, M. R., Vadász, C., & Nádasdy, Z. (2015). Neurons in the basal  
705 forebrain project to the cortex in a complex topographic organization that reflects corticocortical connectivity  
706 patterns: an experimental study based on retrograde tracing and 3D reconstruction. *Cereb. Cortex*, 25(1), 118–137.

## 8 TECHNICAL TERMS

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

709 **Cre-line** The combination of Cre-recombinase expression in transgenic mouse and Cre-induced  
710 expression in the vector that induces labelling of projection

711 **Cell class** The projecting neurons targeted by a particular Cre-line

712 **Structural connectivities** The matrix of connectivities between structures

713 **Voxel** A  $100 \mu\text{m}$  cube of brain

714 **Structural connection tensor** Connectivities between structures across cell classes

715 **Dictionary-learning** A family of algorithms for finding low-dimensional data representations.

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

718 **Nadaraya-Watson** A simple local smoothing estimator with one parameter

719 **Connectivity archetypes** Projection patterns from which we can linearly reconstruct the  
720 connectome

721 **Expected loss** Our new estimator that weights location against cell-class by their estimated  
722 predictive power