

**1 RESEARCH**

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

**4 Samson Koelle<sup>1,2</sup>, Jennifer Whitesell<sup>1</sup>, Karla Hirokawa<sup>1</sup>, Hongkui Zeng<sup>1</sup>, Marina Meila<sup>2</sup>, Julie Harris<sup>1</sup>, Stefan Mihalas<sup>1</sup>**

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

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

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

**ABSTRACT**

**8** The Allen Brain Connectivity Atlas consists of anterograde tracing experiments targeting diverse  
**9** structures and classes of projecting neurons. Beyond anterograde tracing done in C57BL6 wildtype  
**10** mice, a large fraction of these experiments are performed using transgenic cre lines. This allows access  
**11** to cell-class specific connectivity information, with class defined by the transgenic lines. However,  
**12** even though the number of experiments is large, it does not come close to covering all existing cell  
**13** classes in every area where they exist. We study how much we can fill in these gaps and construct a  
**14** voxel-based connectivity given the observations that nearby voxels have similar connections when  
**15** they are in the same area, that connections can change dramatically at area boundaries, and that  
**16** particular cell classes can have similar connections, but that this similarity is region-dependent.

**17** This paper describes the conversion of these experiments into class-specific connectivity matrices  
**18** representing the connection between source and target structures. We introduce and validate a novel  
**19** statistical model for creation of connectivity matrices. We expand a Nadaraya-Watson kernel learning  
**20** method which we previous used to fill in spatial gaps, to also fill in a gaps in cell class connectivity

21 information. To do this, we construct a "cell-class space" and combine smoothing in 3D spatial as well  
22 as in this abstract space to share information between similar neuron classes. Using this method we  
23 construct a set of connectivity matrices using multiple levels of resolution at which discontinuities in  
24 connectivity are assumed. We show that the connectivities obtained from this model display expected  
25 cell-type and structure specific connectivities. Inspired by how this complexity arises from a much  
26 smaller set of genetic information during development, we also show that the wild-type connectivity  
27 matrix can be factored using a small set of factors, and we uncover the underlying latent structure.

## AUTHOR SUMMARY

28 Large scale studies have described the connections between areas in multiple mammalian models in  
29 ever expanding detail. Standard connectivity studies focus on the connection strength between areas.  
30 However, when describing functions at a local circuit level, there is an increasing focus on cell types.  
31 We have recently described the importance of connection types in the cortico-thalamic system, which  
32 allows an unsupervised discovery of its hierarchical organization. In this study we focus on adding a  
33 dimension of connection type for a brain-wide mesoscopic connectivity model. Even with our  
34 massive dataset, the data in the type direction for the connectivity is quite sparse, and we had to  
35 develop methods to more reliably extrapolate in such directions, and to estimate when such  
36 extrapolations are impossible. This allows us to fill in such a connection type specific inter-areal  
37 connectivity matrix to the extent our data allows us to. While analysing this complex connectivity, we  
38 observed that it can be described via a small set of factors. While not complete, this connectivity  
39 matrix represents a large leap forward in mouse connectivity models.

## 1 INTRODUCTION

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

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

53 Several statistical models for the conversion of such experiment-specific signals into estimates of  
54 connectivity strength have been proposed (K. D. Harris, Mihalas, & Shea-Brown, 2016; Knox et al.,  
55 2019; Oh et al., 2014). Of these, Oh et al. (2014) and Knox et al. (2019) model **structural connectivities**,  
56 which are voxel connectivities integrated by structure. The value of these models is that they provide  
57 some improvement over simply averaging the projection signals of injections in a given region.  
58 However, these previous works only model connectivities observed in wild-type mice transduced with  
59 constitutive promoters, and so are poorly suited for extension to recently developed tracing  
60 experiments that induce cell-type specific fluorescence (J. A. Harris et al., 2019). In particular, GFP  
61 promotion is induced by Cre-recombinase expression in cell-types specified by transgenic strain.  
62 Thus, this paper introduces a **cell class**-specific statistical model to deal with the diverse set of  
63 **cre-lines** described in J. A. Harris et al. (2019).

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

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

74 The resulting cell-type specific connectivity is a directed weighted multigraph which can be  
75 represented as a tensor. We do not attempt an exhaustive analysis of this data, but do manually verify  
76 several cell-type specific connectivity patterns found elsewhere in the literature, and show that these  
77 cell-type specific signals are behaving in expected ways. Finally, we decompose the wild-type  
78 connectivity matrix into factors representing archetypal connective patterns using non-negative  
79 matrix factorization. These components are themselves novel and of some independent interest.

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

## 2 METHODS

<sup>84</sup> We create and analyze cell class-specific connectivity matrices using models trained on murine  
<sup>85</sup> viral-tracing experiments. This section describes the data used to generate the model, the model  
<sup>86</sup> itself, the evaluation of the model, and the use of the model in creation of the connectivity matrices. It  
<sup>87</sup> also includes background on the non-negative matrix factorization method used for decomposing the  
<sup>88</sup> wild-type connectivity matrix into latent structures. Additional information on our data is given in  
<sup>89</sup> Supplemental Section 5 methods is given in Supplemental Section 6.

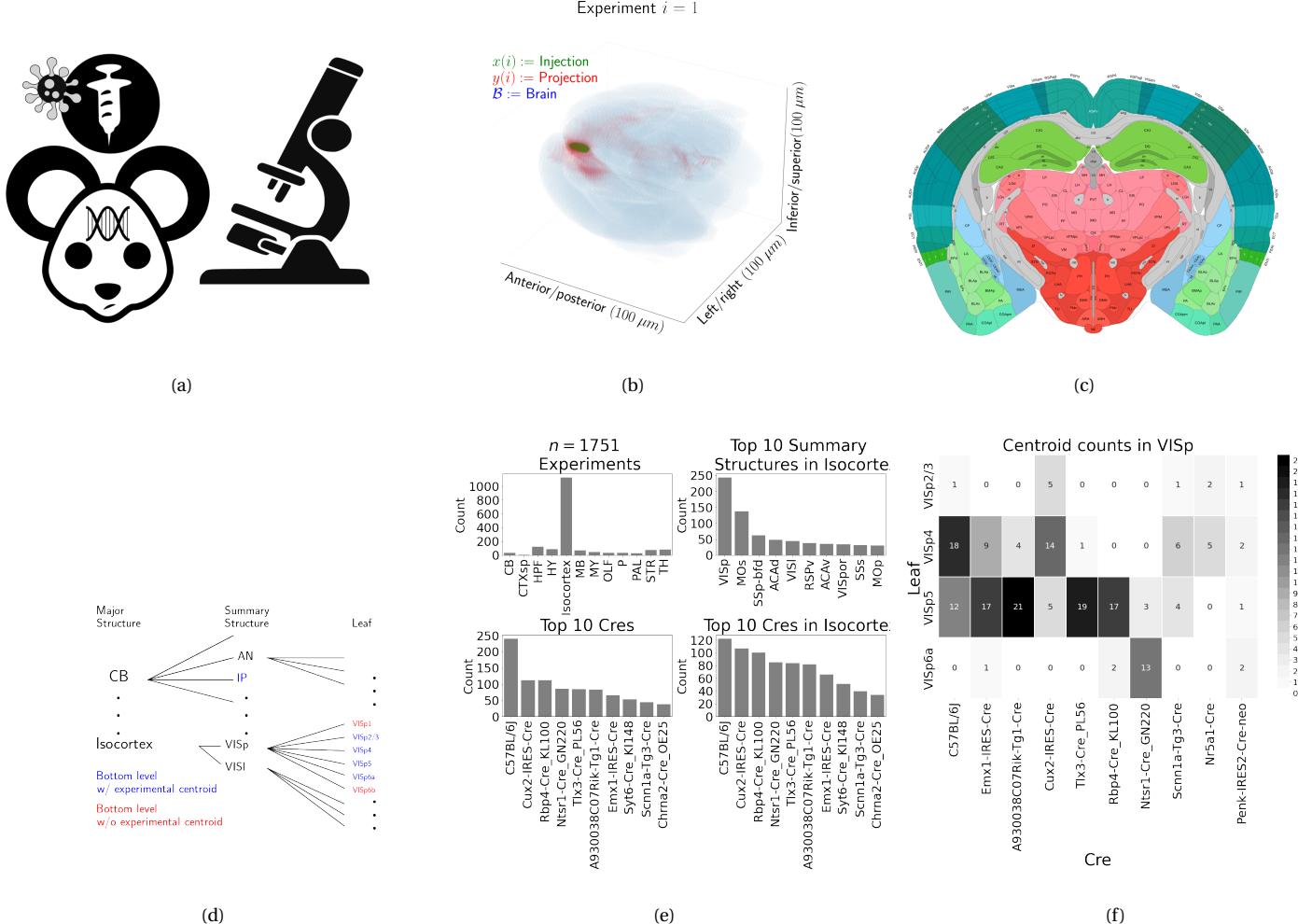


Figure 1: Experimental setting. 1a For each experiment, a potentially Cre-recombinase promoted GFP-expressing transgene cassette is transduced after stereotaxic injection into a Cre-driver mouse, followed by two-photon tomography imaging. 1b An example of the segmentation of projection and injection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

**90 Data**

91 Our dataset  $\mathcal{D}$  consists of  $n = 1751$  publicly available murine viral-tracing experiments from the Allen  
92 Brain Connectivity Atlas. Figure 1a summarizes the multistage experimental process used to generate  
93 this data. In each experiment, a GFP-labelled transgene cassette with a potentially Cre-inducible  
94 promoter is injected into a particular location in a Cre-driver mouse. This causes fluorescence that  
95 depends on the localization of Cre-recombinase expression within the mouse. While frequently this  
96 localization corresponds to a specific cell-type, it can also correspond to a combination of cell-types.  
97 In wild-type mice injected with non-Cre specific promoters, fluorescence is observed in all areas  
98 projected to from the injection site, regardless of cell-type. Thus, we use the term cell class to describe  
99 neurons expressing cre in a specific mouse line. This is the notion of cell-type specificity that we  
100 model.

101 The fluorescent signal imaged after injection is aligned into the Allen Common Coordinate  
102 Framework (CCF) v3, a three-dimensional idealized model of the brain that is consistent between  
103 animals. This imaging and alignment procedure (described in detail in (J. A. Harris et al., 2019))  
104 records fluorescent intensity discretized at the  $100 \mu\text{m}$  voxel level. Given an experiment, this image is  
105 histologically segmented into *injection* and *projection* areas corresponding to areas containing somas,  
106 dendrites and axons or exclusively axons of the transfected neurons. An example for a single  
107 experiment is given in Figure 1b.

108 Our goal is the estimation of **structural connectivity** from one structure to another. A visual  
109 depiction of this structural regionalization for a slice of the brain is given in Figure 1c. For different  
110 areas of the brain, the Allen Brain Atlas contains different depths of regionalization. We denote these  
111 levels as Major Structures, Summary Structures, and Leafs. As indicated in Figure 1d, the dataset used  
112 to generate the connectivity model reported in this paper contains certain combinations of structure  
113 and cell class  $(v, s)$  frequently, and others not at all. A summary of the most frequently assayed cell  
114 classes and structures is given in Figures 1e and 1f. Since users of the connectivity matrices may be  
115 interested in particular combinations, or interested in the amount of data used to generate a  
116 particular connectivity estimate, we present this information about all experiments in Supplemental  
117 Section 5.

118 A cell-class specific neural connectivity is a function  $f: \mathcal{V} \times \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}_{\geq 0}$  giving the directed  
 119 connection of a particular cell class from a one position in the brain to another. However, what we will  
 120 actually estimate are structural connectivities defined with respect to a set of  $S$  source regions  
 121  $\mathcal{S} := \{s\}$ ,  $T$  target regions  $\mathcal{T} := \{t\}$ , and  $V$  cell classes  $\mathcal{V} := \{v\}$ . In contrast to Knox et al. (2019), which  
 122 only uses wild type C57BL/6J mice, these experiments utilize  $V = 114$  different Cre-lines. We generally  
 123 consider  $S = 564$  leaf sources and  $T = 1123$  leaf targets, where 559 are contralateral and 5 are  
 124 mediolateral, but other structuralizations could be used.

125 We preprocess our data in several ways. We discretize fluorescent signals like injections and  
 126 projections into  $100\mu m^3$  **voxels**. Given an experiment  $i$ , we represent injections and projections as  
 127 maps  $x(i), y(i) : \mathcal{B} \rightarrow \mathbb{R}_{\geq 0}$ , where  $\mathcal{B} \subset [1 : 132] \times [1 : 80] \times [1 : 104]$  corresponds to the subset of the  
 128  $(1.32 \times 0.8 \times 1.04)$  cm rectangular space occupied by the standard mouse brain. As an abuse of  
 129 notation, a structure  $s$  then contains  $|s|$  voxels at locations  $\{l_{s_j} \in \mathbb{R}^3\}$ , and similarly for targets. We  
 130 calculate injection centroids  $c(i) \in \mathbb{R}^3$  and regionalized projections  $y_{\mathcal{T}}(i) \in \mathbb{R}^T$  giving the sum of  $y(i)$   
 131 in each region. In contrast to Knox et al. (2019), we generally  $L1$  normalize the projection vectors. This  
 132 accounts for differences in the cre-driven expression of eGFP via the various transgene promoters.  
 133 However, we also for completeness include models of projections normalized by injection signal. A  
 134 detailed mathematical description of these steps, including data quality control, is given in  
 135 Supplemental Section 6.

136 ***Modeling Structural Connectivity***

We define

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

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

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

137 Since the normalized strength and densities are computable from the strength via a fixed  
 138 normalization, our main statistical goal is to estimate  $\mathcal{C}(v, s, t)$  for all  $v, s$  and  $t$ . In other words, we  
 139 want to estimate matrices  $\mathcal{C}_v \in \mathbb{R}_{\geq 0}^{S \times T}$ . We call this estimator  $\widehat{\mathcal{C}}$ .

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

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

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

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

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  $I$ . 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 structural projection signal as a function of structural injection signal. It generates estimated connectivities for individual structures. The Nadaraya-Watson model (NW) is a local smoothing model that generates a prediction for each voxel within a structure. These predictions are averaged to create estimate the structure-specific connectivity.

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

158 experiments. Thus, nearby experiments with similar Cre-lines can help generate the prediction, even  
159 when there are few nearby experiments of the cell-class in question.

160 ***Model evaluation***

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

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

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

$$\begin{aligned} \text{l2-loss } \ell(y_{\mathcal{T}}(i)), \widehat{y_{\mathcal{T}}(i)} ) &:= \|y_{\mathcal{T}}(i)) - \widehat{y_{\mathcal{T}}(i)}\|_2^2. \\ \text{weighted l2-loss } \mathcal{L}(\widehat{f}(f_*)) &:= \frac{1}{|\{s, \nu\}|} \sum_{s, \nu \in \{\mathcal{S}, \mathcal{V}\}} \frac{1}{|I_s \cap I_\nu|} \sum_{i \in (I_s \cap I_\nu)} \ell(y_{\mathcal{T}}(i)), \widehat{f}_{\mathcal{T}}(f_*(\mathcal{D}(\nu, s) \setminus i)). \end{aligned}$$

173 This is a somewhat different loss from Knox et al. (2019), both because of the normalization of  
 174 projection, and because of the increased weighting of rarer combinations of  $s$  and  $\nu$  implicit in the  
 175 loss. Since the number of parameters fit is very low (at least two orders of magnitude) relative to the  
 176 size of the evaluation set, we do not make use of a formal validation-test split. As a final modeling  
 177 step, we establish a lower limit of detection. The EL model also contains a separate cross-validation  
 178 step. These approaches are covered in Supplemental Section 6

179 ***Connectivity analyses***

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

191 Our algorithm is

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

192 For this decomposition we ignore connections between source and target regions less than  
 193  $1500\mu m$  apart. This is because short-range projections resulting from diffusion dominate the  
 194 matrices  $\hat{\mathcal{C}}$ , and represent a less-interesting type of biological structure. We explored different values  
 195 and set  $\lambda = 0.002$  to encourage sparser and therefore more interpretable components. We use  
 196 unsupervised cross-validation to determine an optimum  $q$ , and show the top 15 stable components.  
 197 Stability analysis accounts for the difficult-to-optimize NMF program by clustering the resultant  $H$   
 198 from multiple replicates. The medians of the component clusters appearing frequently across NMF  
 199 replicates are selected as **connectivity archetypes**. Details of these approaches are given in  
 200 Supplementary Sections 6 and 7.

### 3 RESULTS

<sup>201</sup> We provide several types of results. First, we show that the novel expected-loss (EL) estimator  
<sup>202</sup> performs best in our validation assays. Second, qualitative exploratory analysis confirms that the  
<sup>203</sup> Cre-specific connectivity matrices generated using this model are consistent with known biology.  
<sup>204</sup> Third, statistical decomposition of the wild-type connectivity matrix using unsupervised learning  
<sup>205</sup> shows how archetypal components can combine to produce observed signals.

<sup>206</sup> ***Model evaluation***

<sup>207</sup> Our EL model generally performs better than the other estimators that we consider. Table 5 contains  
<sup>208</sup> weighted losses from leave-one-out cross-validation of candidate models, such as the NW Major-WT  
<sup>209</sup> model from Knox et al. (2019). The EL model combines the good performance of class-specific  
<sup>210</sup> models like NW Leaf-Cre in regions like Isocortex with the good performance of class-agnostic models  
<sup>211</sup> in regions like Thalamus. Additional information on model evaluation, including class and structure-  
<sup>212</sup> specific performance, is given in Appendix 5. In particular, Supplementary Table 4 contains the sizes  
<sup>213</sup> of these evaluation sets in each major structure, and Supplementary Section 7 contains the structure-  
<sup>214</sup> and class specific losses.

$\hat{f}$	Mean Leaf-Cre	NW Major-Cre	NW Leaf-Cre	NW Leaf	NW Major-WT	NW Major	EL
$\mathcal{D}$	Mean	NW					EL
	$I_c \cap I_L$	$I_c \cap I_M$	$I_c \cap I_L$	$I_L$	$I_{wt} \cap I_M$	$I_M$	$I_L$
Isocortex	0.239	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

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

215 ***Connectivities***

216 Our main result is the estimation of matrices  $\hat{\mathcal{C}}_v \in \mathbb{R}_{\geq 0}^{S \times T}$  representing connections of source structures  
 217 to target structures for particular cre-lines  $v$ . We confirm the detection of several well-established  
 218 connectivities within our tensor, although it is our expectation that additional interesting biological  
 219 processes are also manifest. The connectivity tensor and code to reproduce it are available at  
 220 [https://github.com/AllenInstitute/mouse\\_connectivity\\_models/tree/2020](https://github.com/AllenInstitute/mouse_connectivity_models/tree/2020).

221 *Overall connectivity* Several expected biological processes are evident in the wild-type connectivity  
 222 matrix  $\mathcal{C}_{wt}$  from leaf sources to leaf targets shown in Figure 2a. Intraareal connectivities are clear, as  
 223 are ipsilateral connections between cortex and thalamus. The clear intrastructural and intraareal  
 224 connectivities mirror previous estimates in Oh et al. (2014) and Knox et al. (2019) and descriptive  
 225 depictions of individual experiments in J. A. Harris et al. (2019). These short-range connectivities  
 226 define a

227 Our estimated wild-type connectivities appear more variable than those in Knox et al. (2019), which  
 228 used the NW Major-WT model whose accuracy is evaluated in Table 5. This is plausibly because of  
 229 both the layer-specific targeting of the different cre-lines, and also the layer-specificity of the selected  
 230 model. Although layer-specificity is a major advantage of including distinct cre-lines, for comparison,  
 231 we also plot coarser projections between summary-structure sources and targets in the cortex in  
 232 Figure 2b. These are averages over component layers weighted by layer size. Grossly congruent with  
 233 the previous work, these results also exhibit a larger range of connectivities than those in Knox et al.  
 234 (2019). Importantly, as shown in Table 5 this finer spatial resolution corresponds to the increased  
 235 accuracy of our EL model over the NW Major-WT model.

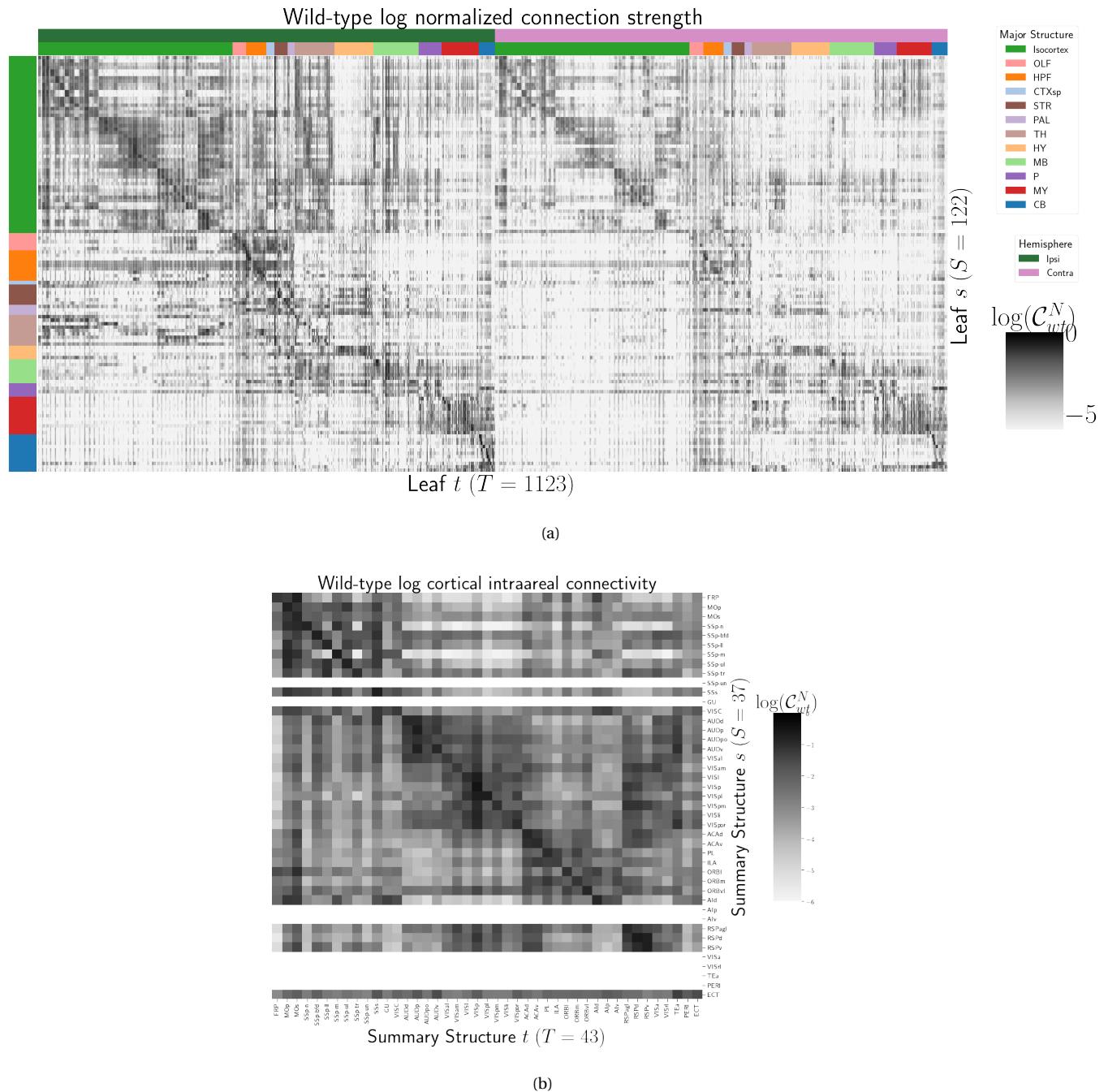


Figure 2: Wild-type connectivities. 2a Log wild-type connectivity matrix  $\log \mathcal{C}(s, t, v_{wt})$ . 2b Log wild-type intracortical connectivity matrix at the summary structure level.

236 *Class-specific connectivities* Source and cell-type combinations which project similarly indicate the  
 237 network structure underpinning cognition. Our estimates of these class-specific connectivities exhibit  
 238 certain known behaviors. In Figure 3, we display results for the VISp and MO cortical areas. These are  
 239 ideal testbeds for our connectivities because they have well-established layer-specific projection  
 240 patterns that can be detected with our layer-specific cre-line based targeting Jeong et al. (2016), and  
 241 are also well-represented in our dataset.

242 Our results are consistent with anterograde tracing experiments outside our dataset Jeong et al.  
 243 (2016). Figure 3a shows that in VISp, the Ntsr1-Cre line strongly targets the thalamic LP nuclei, and in  
 244 MO, layer 5 projects to anterior basolateral amygdala (BLA) and capsular central amygdala (CEA),  
 245 while layer 6 does not. Recall that we display connectivity estimates for structures with at least one  
 246 injection centroid in the structure. Thus, the position of non-zero rows in Figure 3a shows the  
 247 localization of Rbp4-Cre and Ntsr1-Cre injection centroids to layers 5 and 6 respectively (this is  
 248 further examined in Supplemental Section ??). Thus, as a heuristic alternative model, to also  
 249 synthesize information about leafs targeted by different cre-lines, we also generate an average  
 250 connectivity matrix over all cre-lines. This model is not evaluated in our testing, and is only a general  
 251 stand-in for overall behavior, but provides a useful summary of results.

252 Cell-class, while often correlated with cortical layer, is often a stronger driver of connectivity than  
 253 summary structure. Figure 3b shows a collection of connectivity strengths generated using  
 254 cre-specific models for wild-type, Cux2, Ntsr1, Rbp4, and Tlx3 cre-lines from visual signal processing  
 255 leafs in the cortex to cortical and thalamic nucleii. We use hierarchical clustering to sort source  
 256 structure/cell-class combinations by the similarity of their structural projections, and sort target  
 257 structures by the structures from which they receive projections. Examining the former, we can see  
 258 that the Ntsr1 Cre-line distinctly projects to thalamic nucleii, regardless of summary structure. This  
 259 contrasts with the tendency of other cell-classes to project intracortically in a manner determined by  
 260 the source structure. Similarly, layer 6 targets are not strongly projected to by any of the displayed  
 261 Cre-lines. There are too many targeted summary structures to plot here, but we expect that the source  
 262 profile of each target clusters by structure.

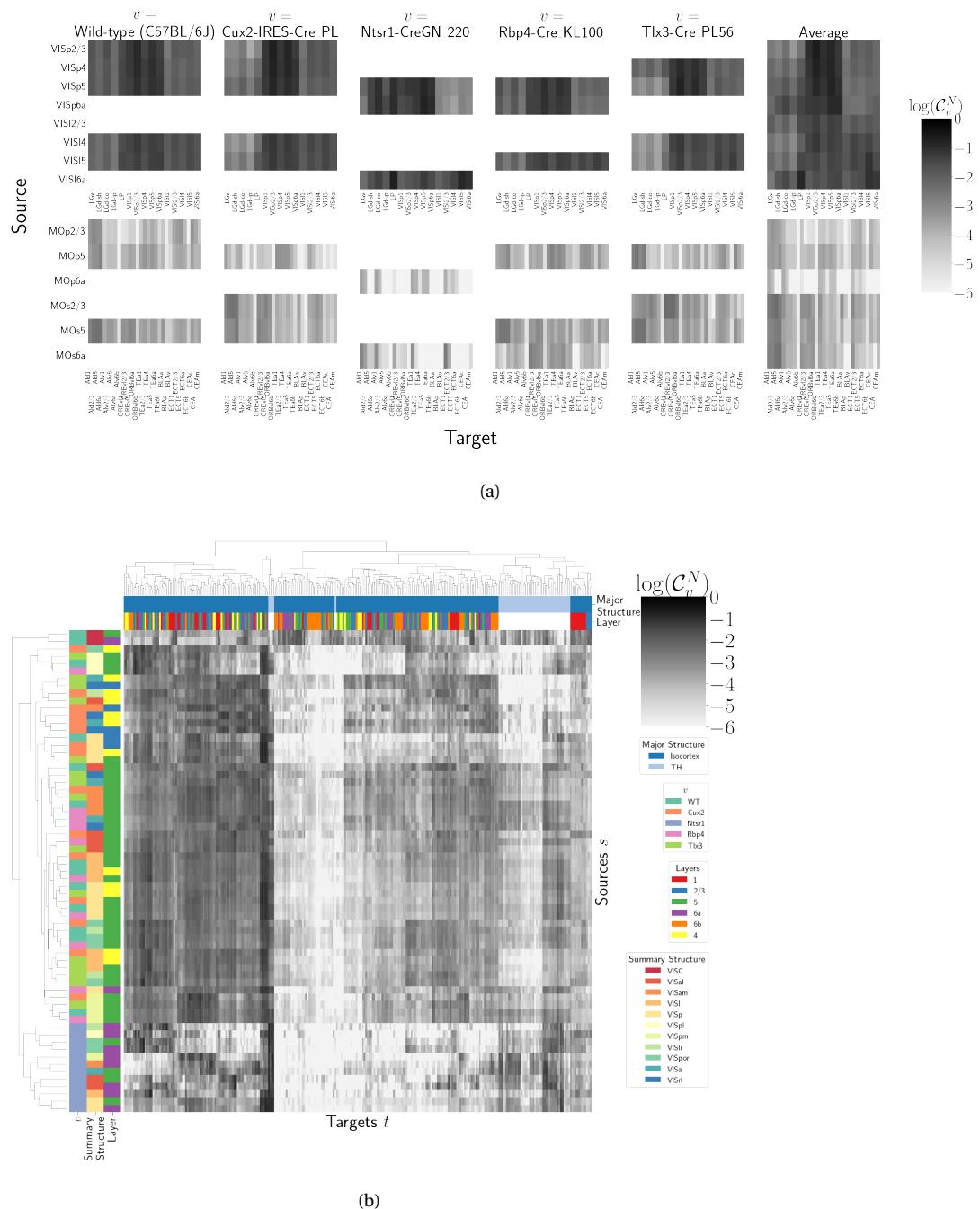


Figure 3: Cell-class specificity. 3a Selected cell-class and layer specific connectivities from VISp and MO. Sources without a injection of that Cre-type are not estimated due to lack of data for that Cre-line in that structure. 3b Heirarchical clustering of connectivity strengths from visual signal processing cell-types to cortical and thalamic targets. Cre-line, summary structure, and layer are labelled on the sources. Major brain division and layer are labelled on the targets.

**263 *Connectivity Analyses***

264 Each structural connectivity matrix is a high-dimensional realization of relatively few biological  
265 processes, and decomposition of neural signals to recover these processes is a fundamental goal in  
266 neuroscience. In this section, we apply non-negative matrix factorization to decompose the  
267 long-range wild-type connectivities into linear combinations of archetypal connectivities. This  
268 decomposes the remaining censored connectivity matrix into a linear model based off a relatively  
269 small number of distinct signals. This model is able to capture a large amount of the observed  
270 variability, and recovers structure-specific archetypal signals.

271 These signals are plotted in Figure 4, and technical details and intermediate results are given in  
272 Supplemental Sections 6 and 7, respectively. These details include a cross-validation based method  
273 for selecting the number of components, a masking method for focusing only on long range  
274 connections, and a stability method for ensuring that the decomposition is reliable across  
275 computational replicates. The plotted decomposition shows that these underlying connectivity  
276 archetypes correspond strongly to major brain division. However, certain components that  
277 predominantly represent connectivity from a given major brain division may also be accessed from  
278 other areas. For example, the IP and FN regions of CB are strongly associated in 4b with the  
279 component projecting to MY in 4a.

280 Inspection of the reconstructed distal normalized connection strength using the top 15  
281 components shows qualitatively shows that this relatively sparse decomposition is able to capture  
282 much of the observed variability. Layer-specific targeting is evident, indicating that the factorization  
283 method is detecting cell-type specific signals, even though it is trained only on the wild-type  
284 connectivity. Other connectivity patterns like cortical-cortical and cortical-thalamic are also detected.

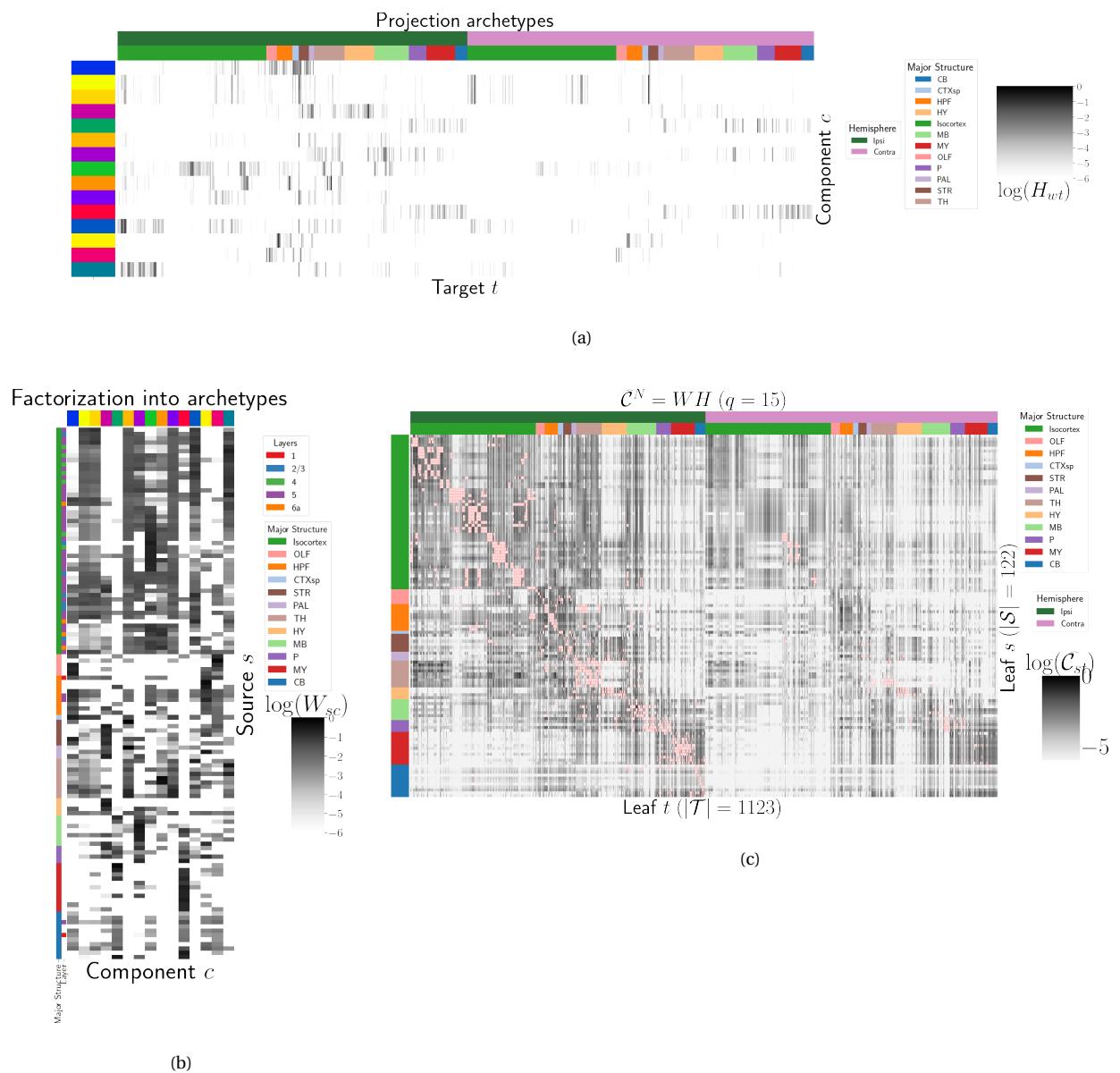


Figure 4: Non-negative matrix factorization results  $\mathcal{C}_w^N = W H$  for  $q = 15$  components. 4a Latent space coordinates  $H$  of  $\mathcal{C}$ . Target major structure and hemisphere are plotted. 4b Loading matrix  $W$ . Source major structure and layer are plotted. 4c Reconstruction of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

## 4 DISCUSSION

<sup>285</sup> The model presented here is a milestone in characterization of connectomes. It is the first cell-type  
<sup>286</sup> specific whole brain projectomme for a mammalian species, and it opens the door for a large number  
<sup>287</sup> of models linking brain structure to computational architectures.

<sup>288</sup> The Nadaraya-Watson estimator presented here is novel. Beyond using a Nadaraya-Watson kernel  
<sup>289</sup> regression defined in physical space, we define a cell-type space based on similarities of projections,  
<sup>290</sup> and theoretically justify the use of an intermediate shape-constrained estimator. While methods like  
<sup>291</sup> non-negative least squares can also account for covariates, the centroid method from Knox et al.  
<sup>292</sup> (2019) was shown that the more precise notion of injection location than the non-negative least  
<sup>293</sup> squares in Oh et al. (2014). Furthermore, our sample size seems too low to utilize a fixed or mixed  
<sup>294</sup> effect, particularly since the impact of the virus depend on the particular injection region. In a sense  
<sup>295</sup> both the NNLS and NW models can be thought of as improvements over the structure-specific  
<sup>296</sup> average, and so is also possible that a yet undeveloped residual-based data-driven blend of these  
<sup>297</sup> models could provide improved performance.

<sup>298</sup> We see several other opportunities for improving on our model. Ours is certainly not the first  
<sup>299</sup> cross-validation based model averaging method Gao, Zhang, Wang, and Zou (2016). However, our use  
<sup>300</sup> of shape-constrained estimator in target-encoded feature space is novel and fundamentally different  
<sup>301</sup> from Nadaraya-Watson estimators that use an optimization method for selecting the weights (Saul &  
<sup>302</sup> Roweis, 2003). The properties of this estimator, as well as its relation to estimators fit using an  
<sup>303</sup> optimization algorithm, are a possible future avenue of research. A deep model such as Lotfollahi,  
<sup>304</sup> Naghipourfar, Theis, and Alexander Wolf (2019) could be appropriate, provided enough data was  
<sup>305</sup> available. Finally, a Wasserstein-based measure of injection similarity per structure would combine  
<sup>306</sup> both the physical simplicity of the centroid model while also incorporating the full distribution of the  
<sup>307</sup> injection signal.

<sup>308</sup> The factorization of the connectivity matrix could also be improved and better utilized. From a  
<sup>309</sup> statistical perspective, stability-based method for establishing archetypal connectivities in NMF is  
<sup>310</sup> similar to those applied to genomic data Kotliar et al. (2019); Wu et al. (2016). However, non-linear  
<sup>311</sup> data transformations or matrix decompositions, or tensor factorizations that account for correlations

<sup>312</sup> between cell-types could better capture the true nature of archetypal neural connections. It would  
<sup>313</sup> also be of great interest to associate the archetypal signals detected from connectivity analysis with  
<sup>314</sup> undergirding gene expression patterns or functional information.

## ACKNOWLEDGMENTS

<sup>315</sup> The Funder and award ID information you input at submission will be introduced by the publisher  
<sup>316</sup> under a Funding Information head during production. Please use this space for any additional  
<sup>317</sup> acknowledgements and verbiage required by your funders.

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

## 5 SUPPLEMENTAL INFORMATION

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

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

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

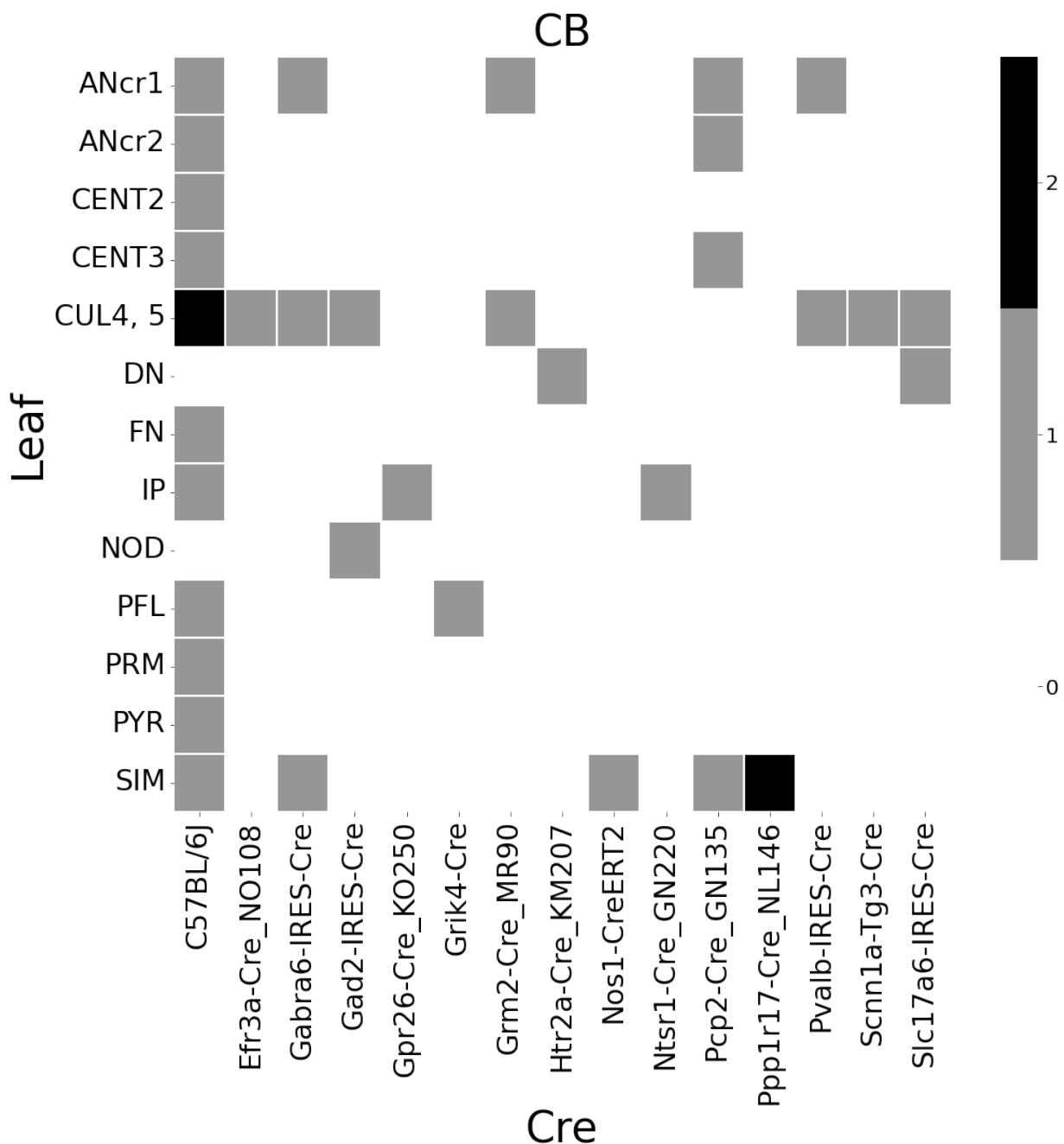


Figure 5: Abundances of cre-line and leaf-centroid combinations.

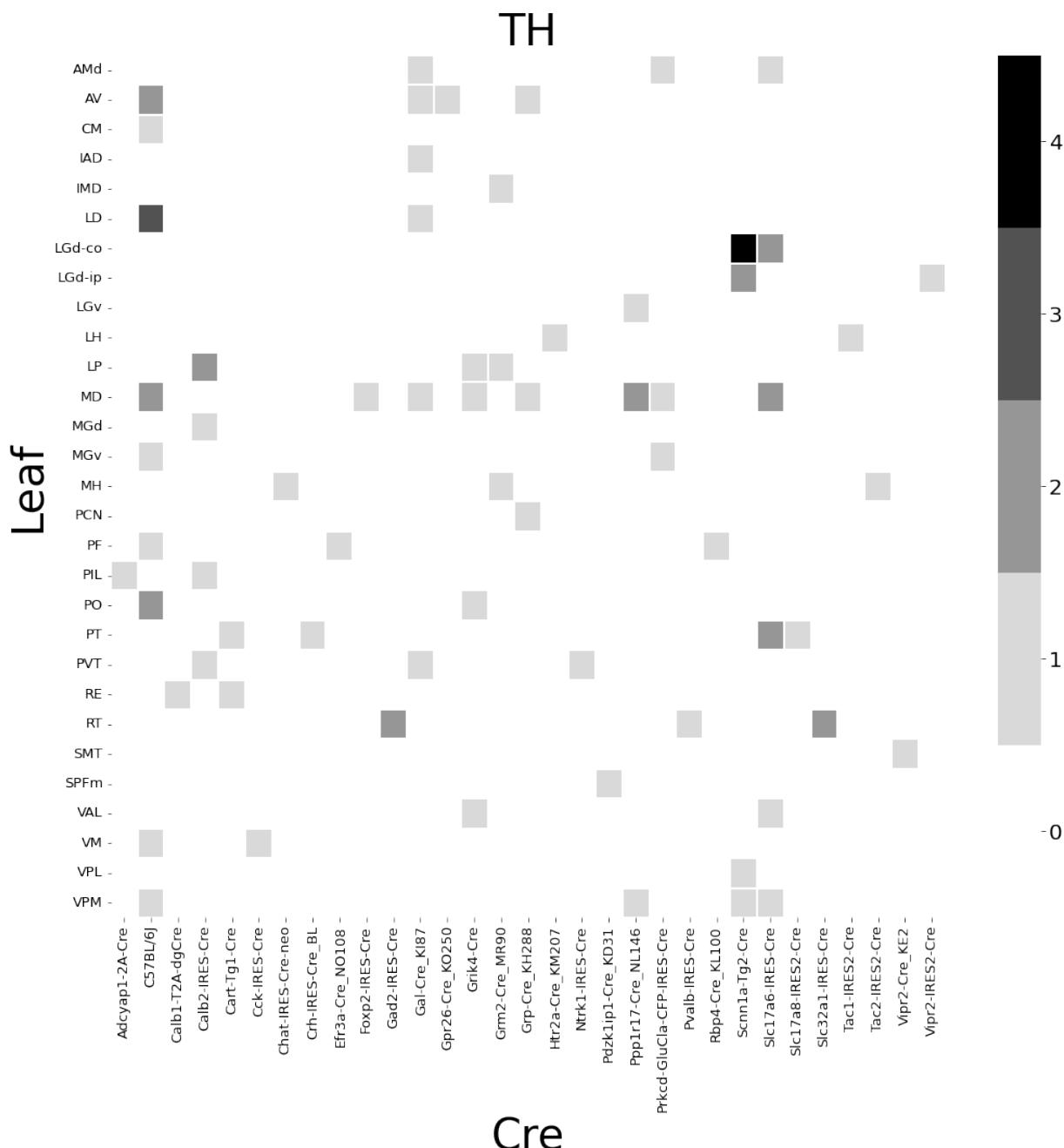


Figure 6: Abundances of cre-line and leaf-centroid combinations.

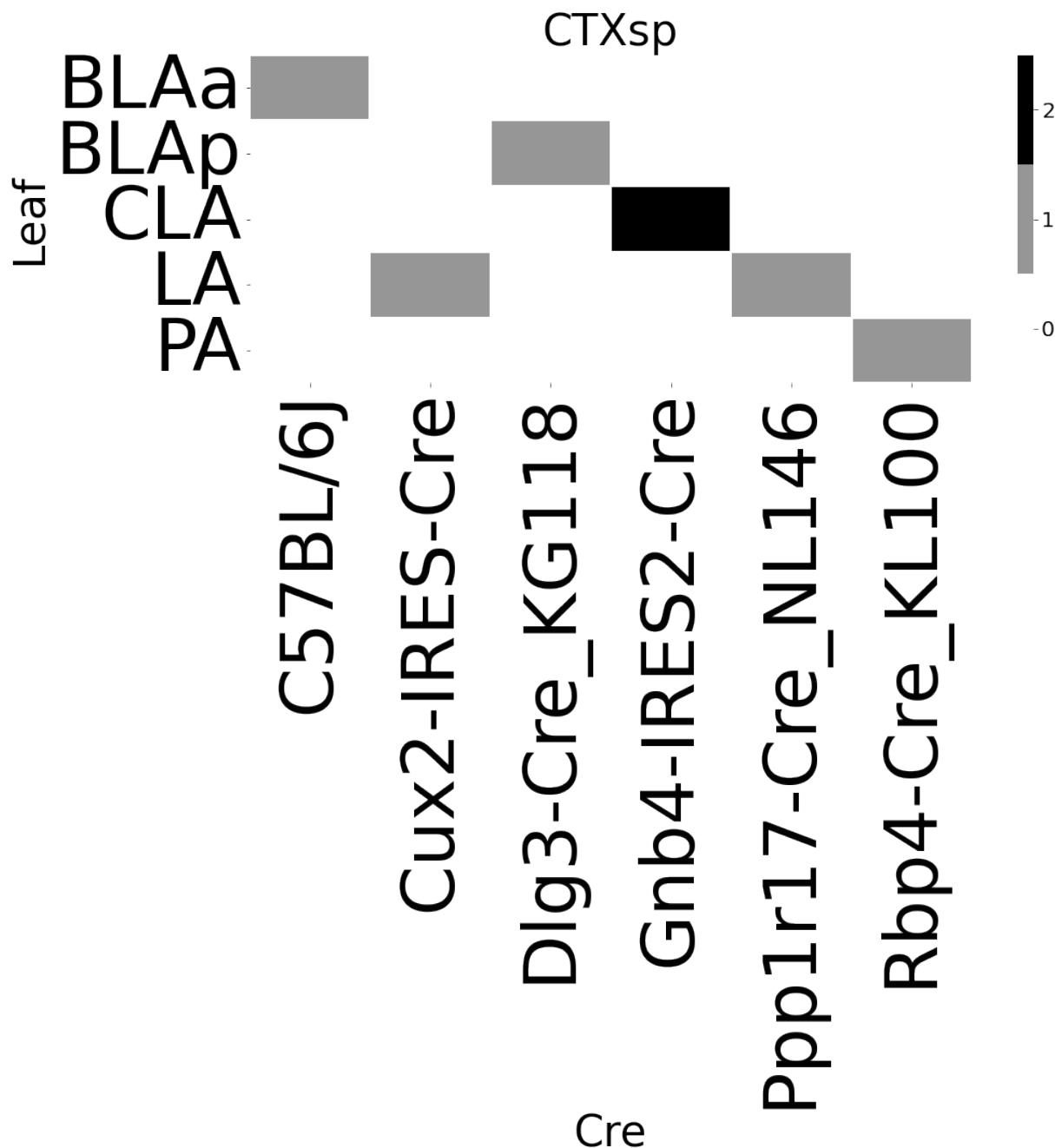


Figure 7: Abundances of cre-line and leaf-centroid combinations.

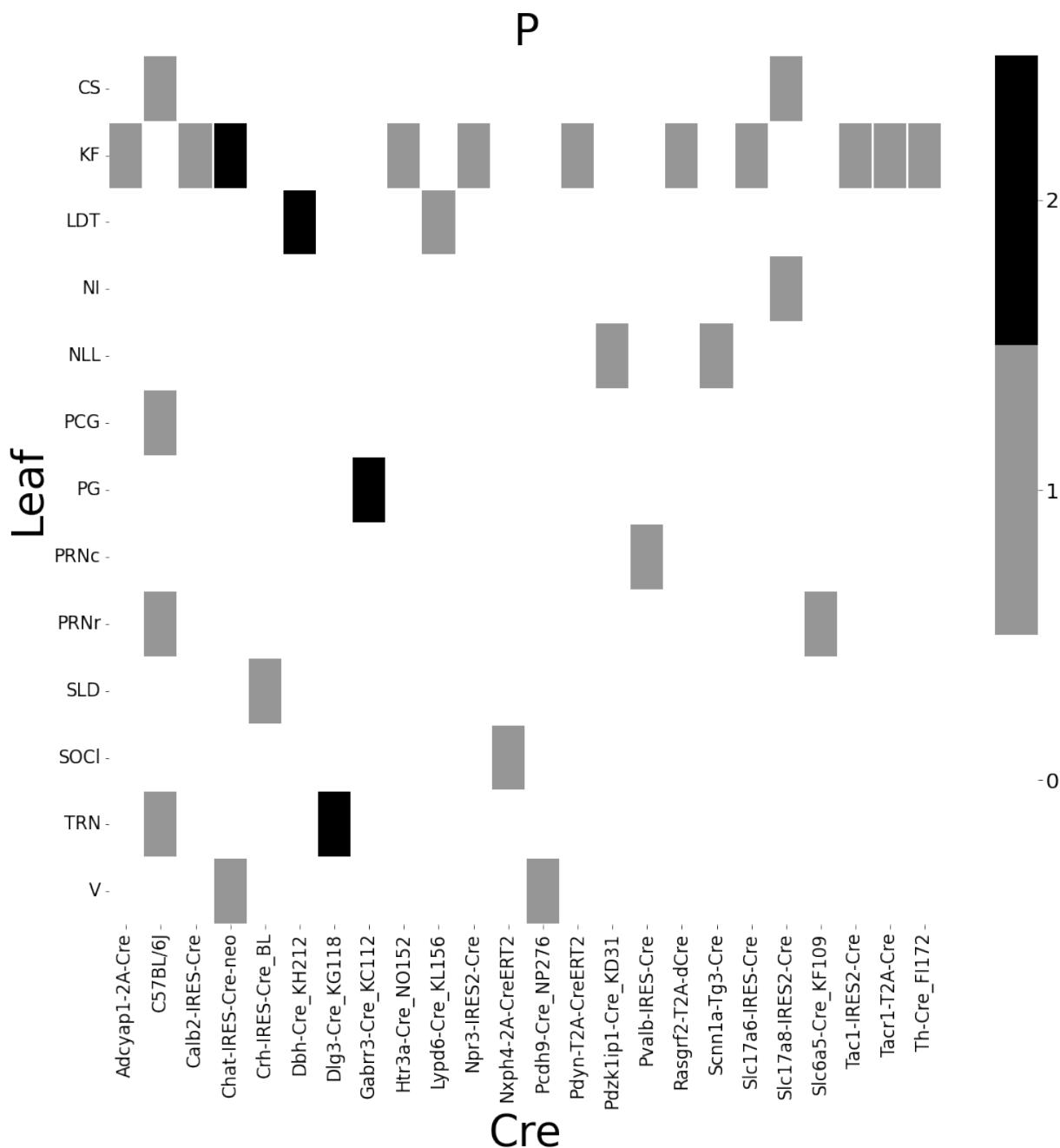


Figure 8: Abundances of cre-line and leaf-centroid combinations.

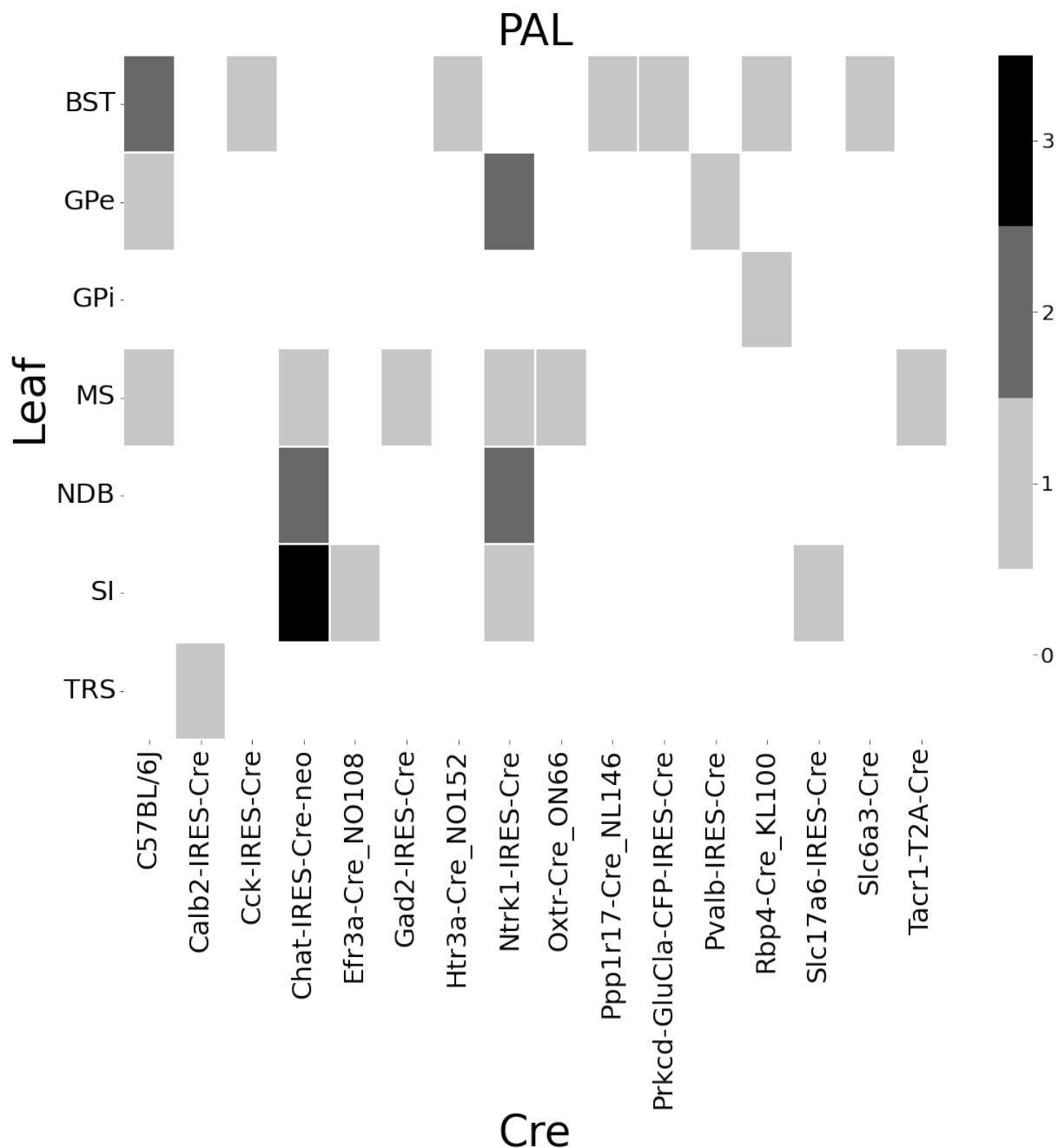


Figure 9: Abundances of cre-line and leaf-centroid combinations.

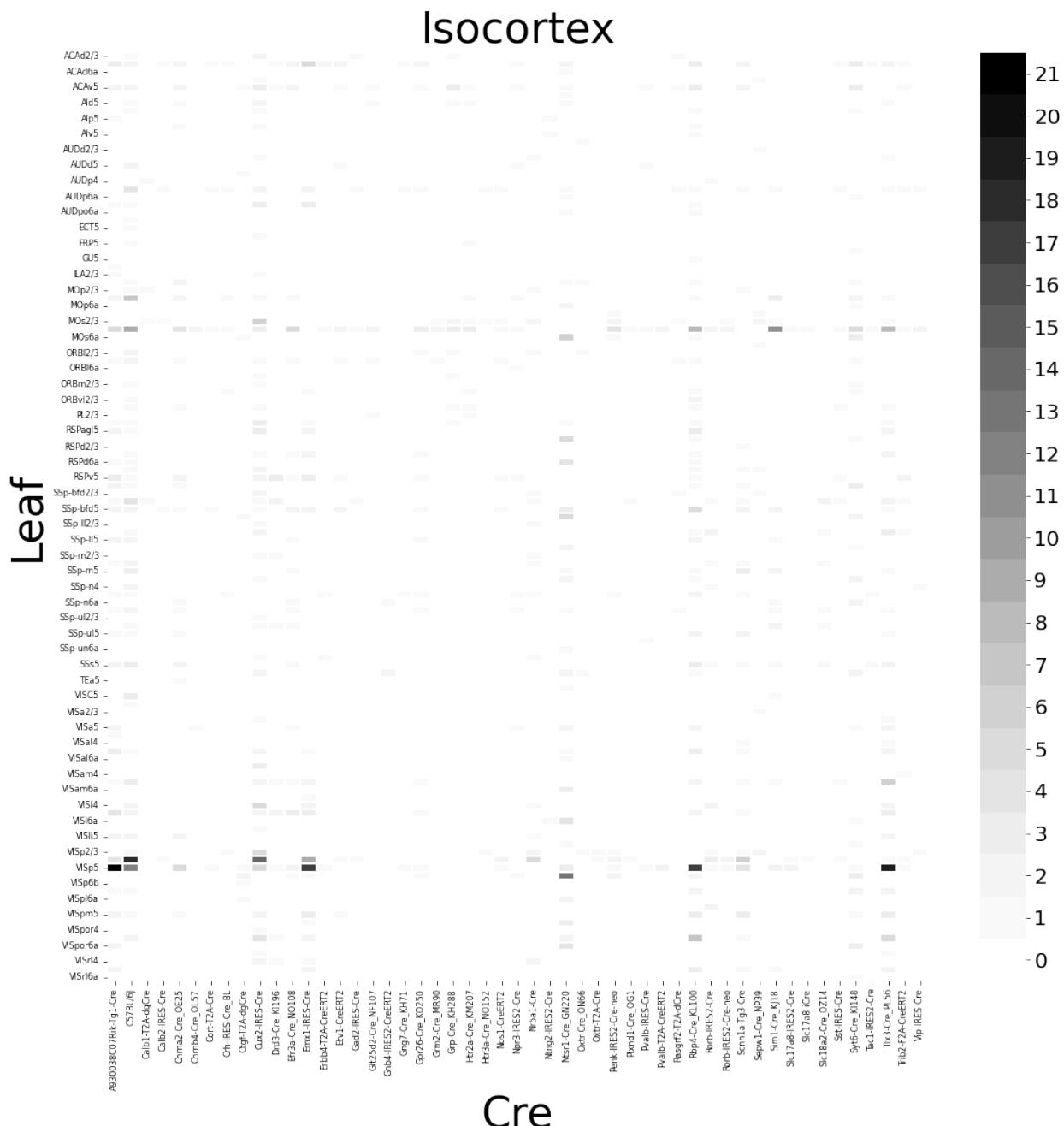


Figure 10: Abundances of cre-line and leaf-centroid combinations.

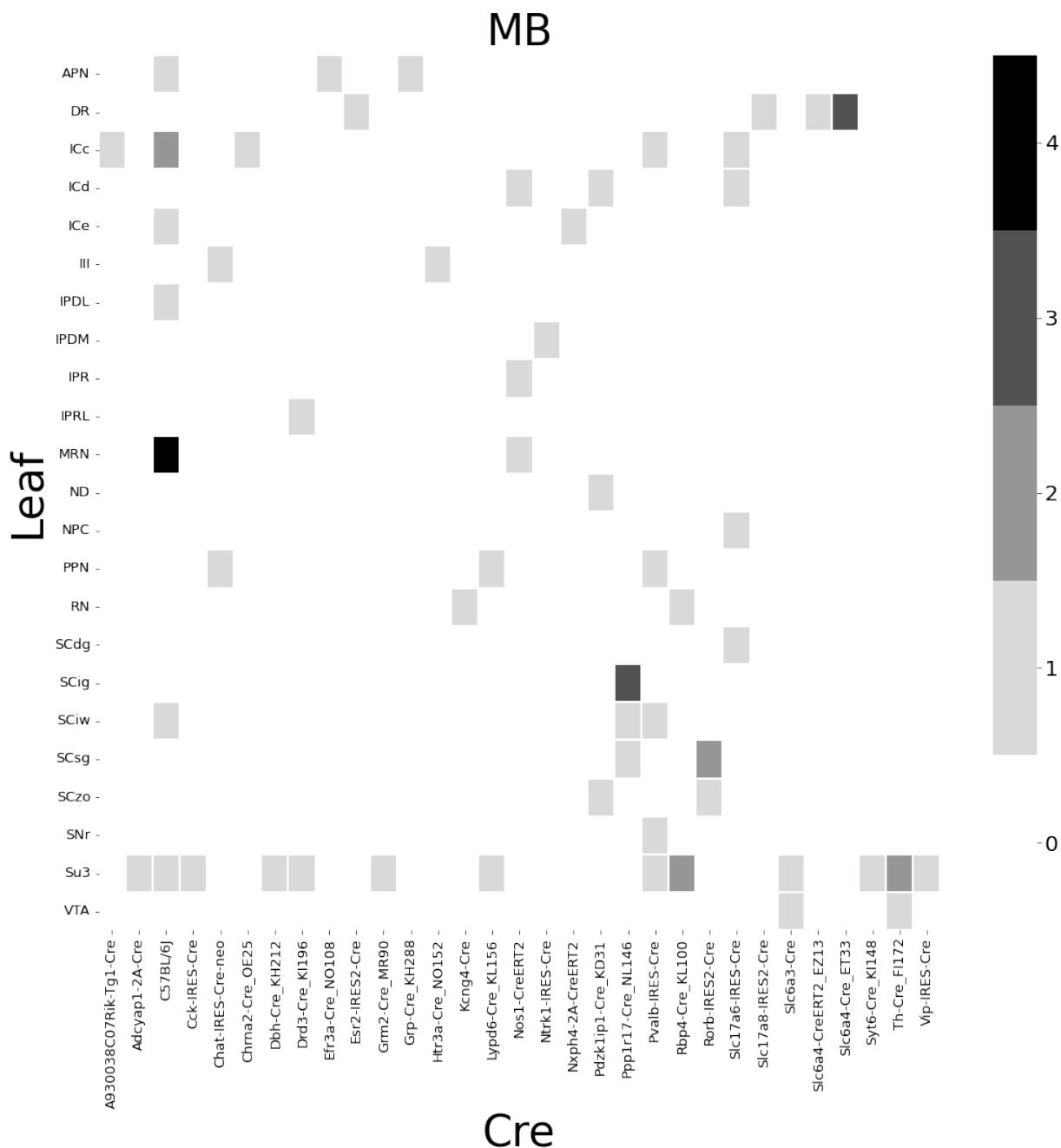


Figure 11: Abundances of cre-line and leaf-centroid combinations.

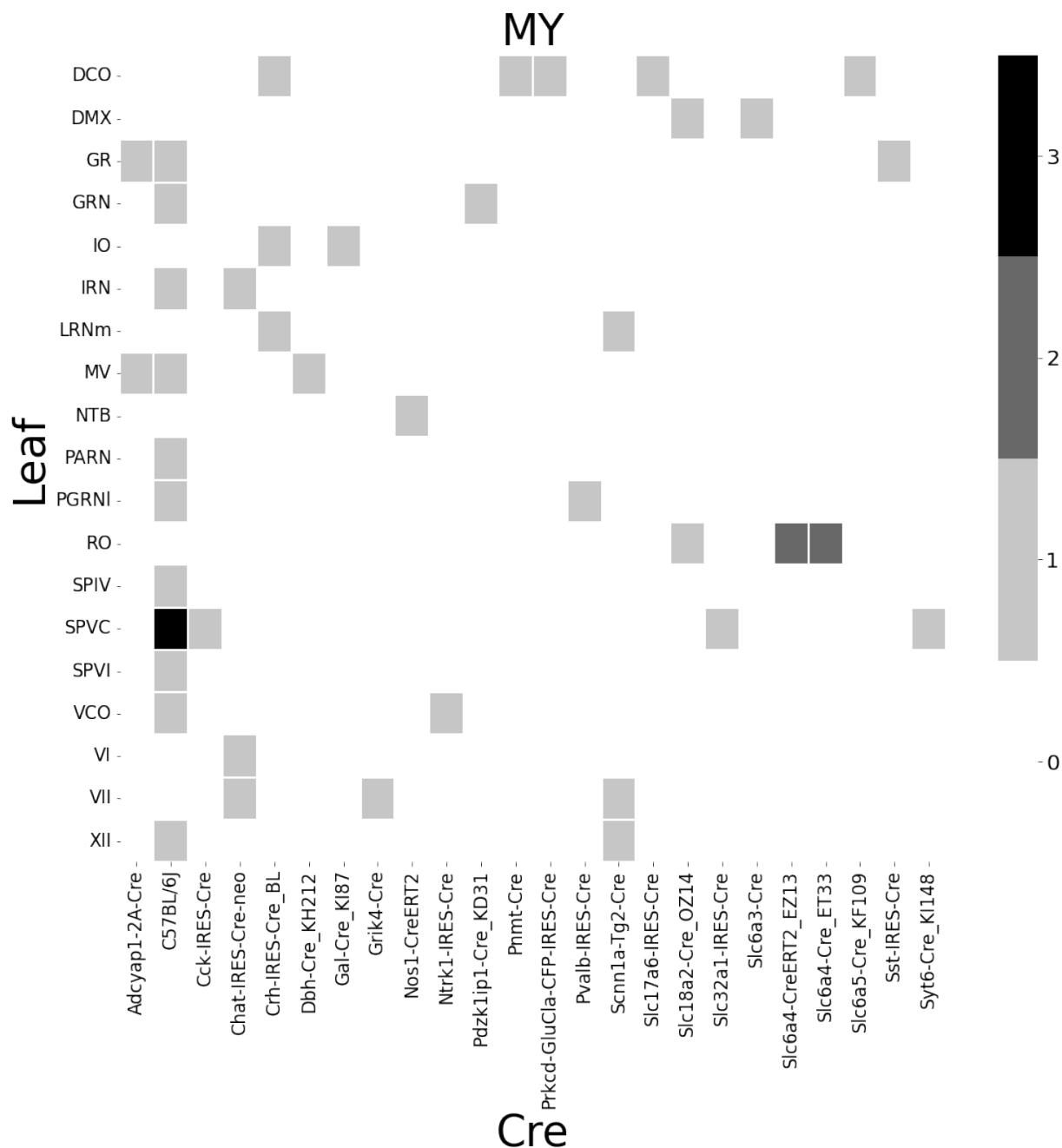


Figure 12: Abundances of cre-line and leaf-centroid combinations.

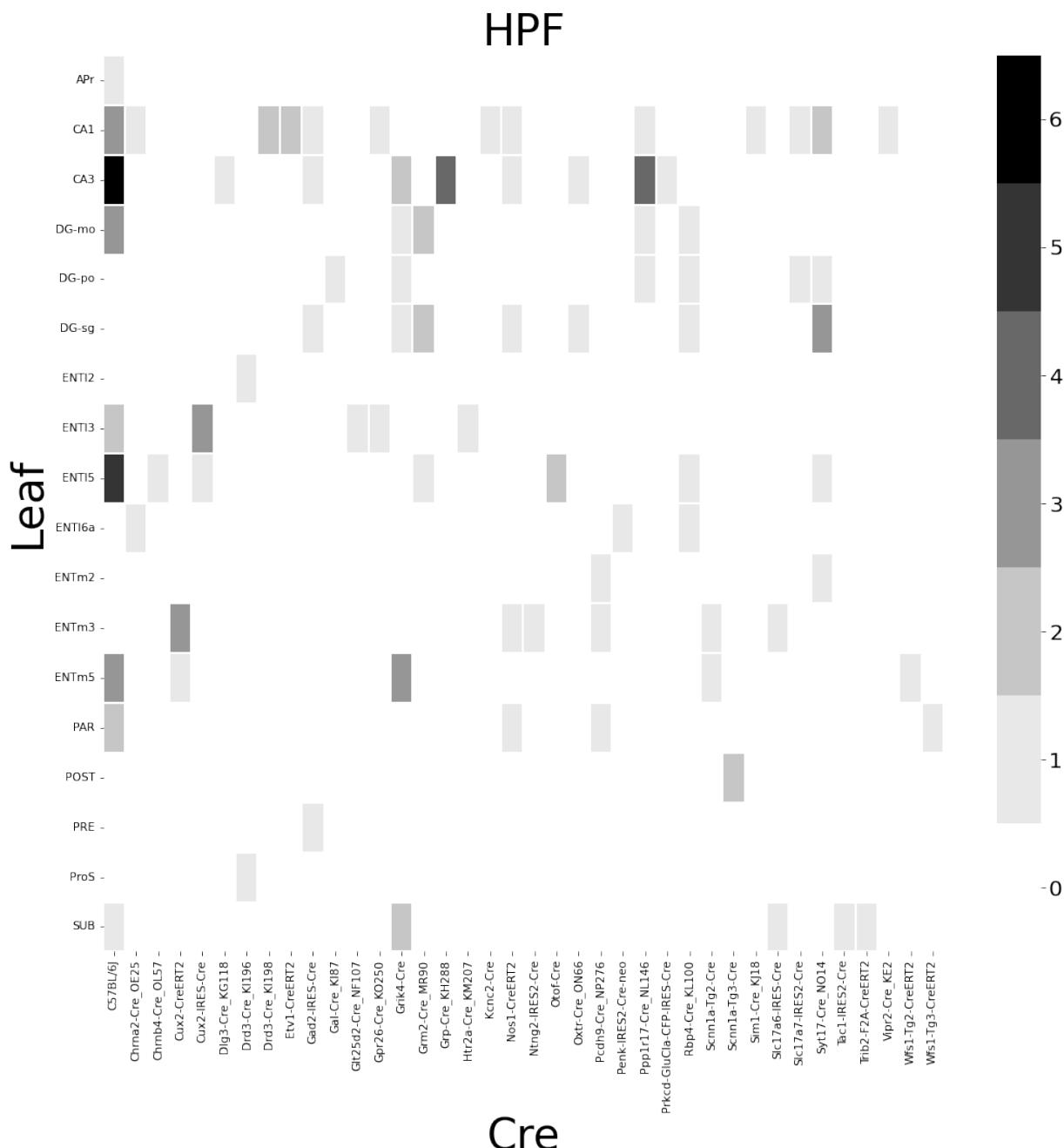


Figure 13: Abundances of cre-line and leaf-centroid combinations.

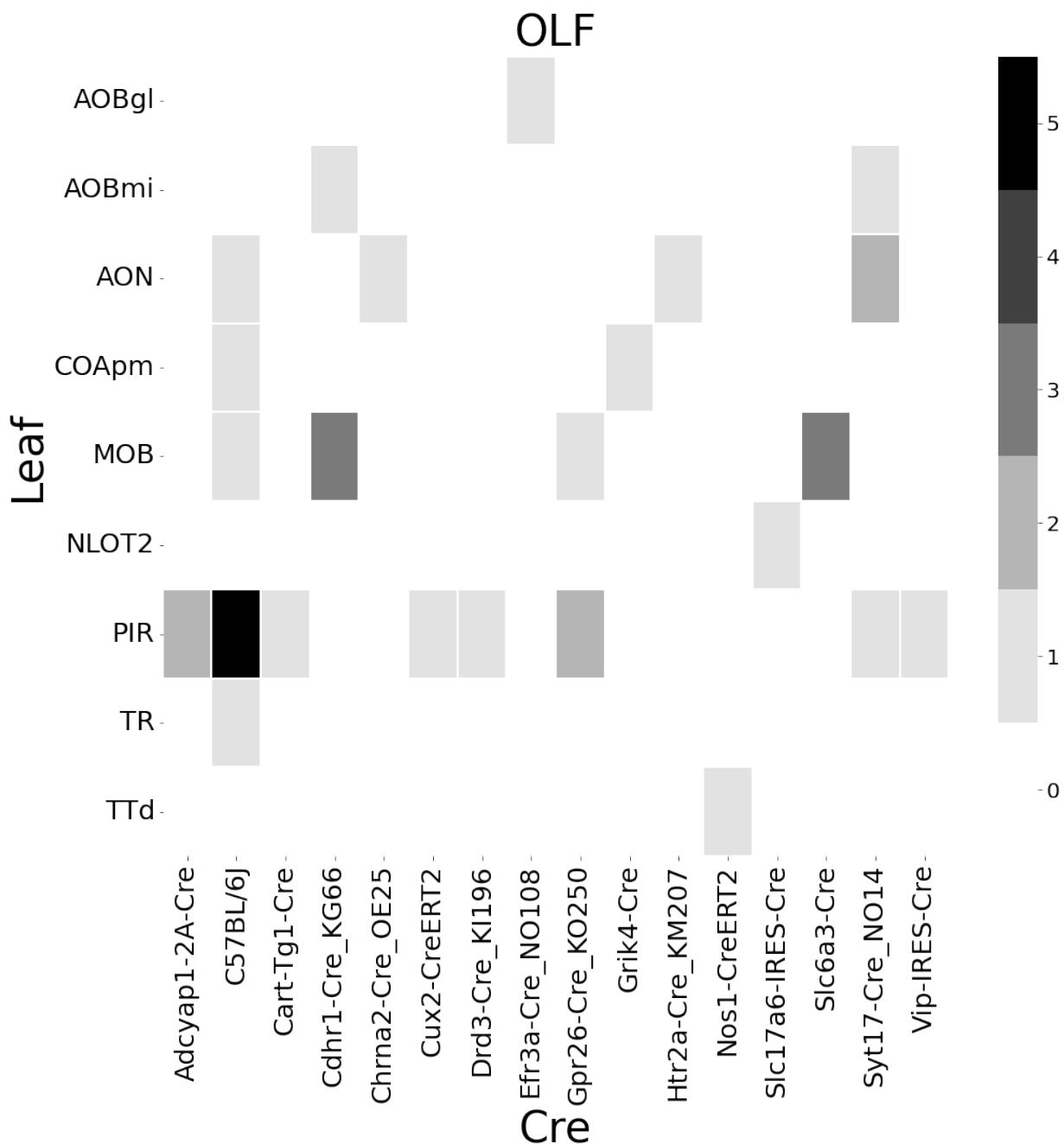


Figure 14: Abundances of cre-line and leaf-centroid combinations.

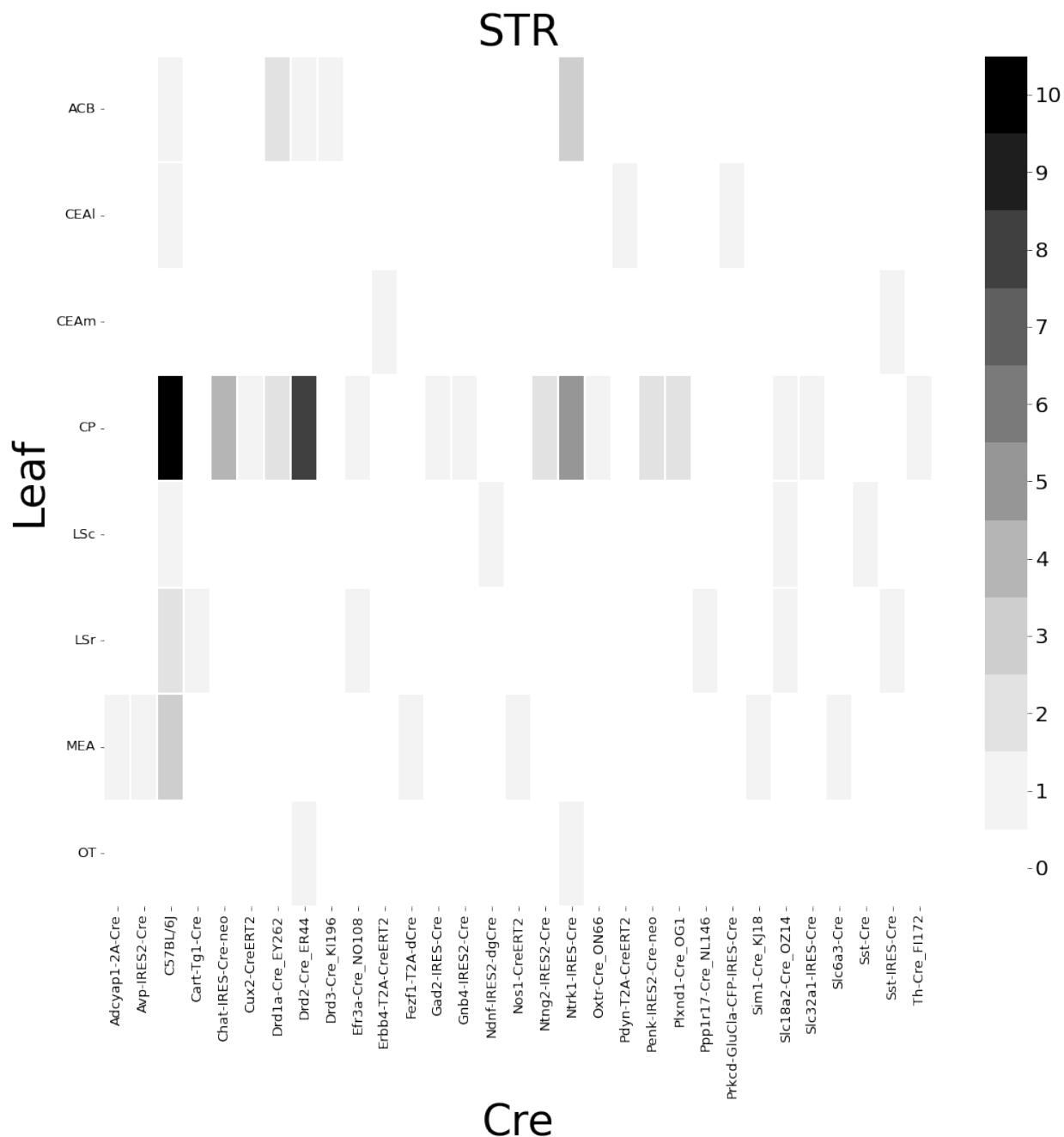


Figure 15: Abundances of cre-line and leaf-centroid combinations.

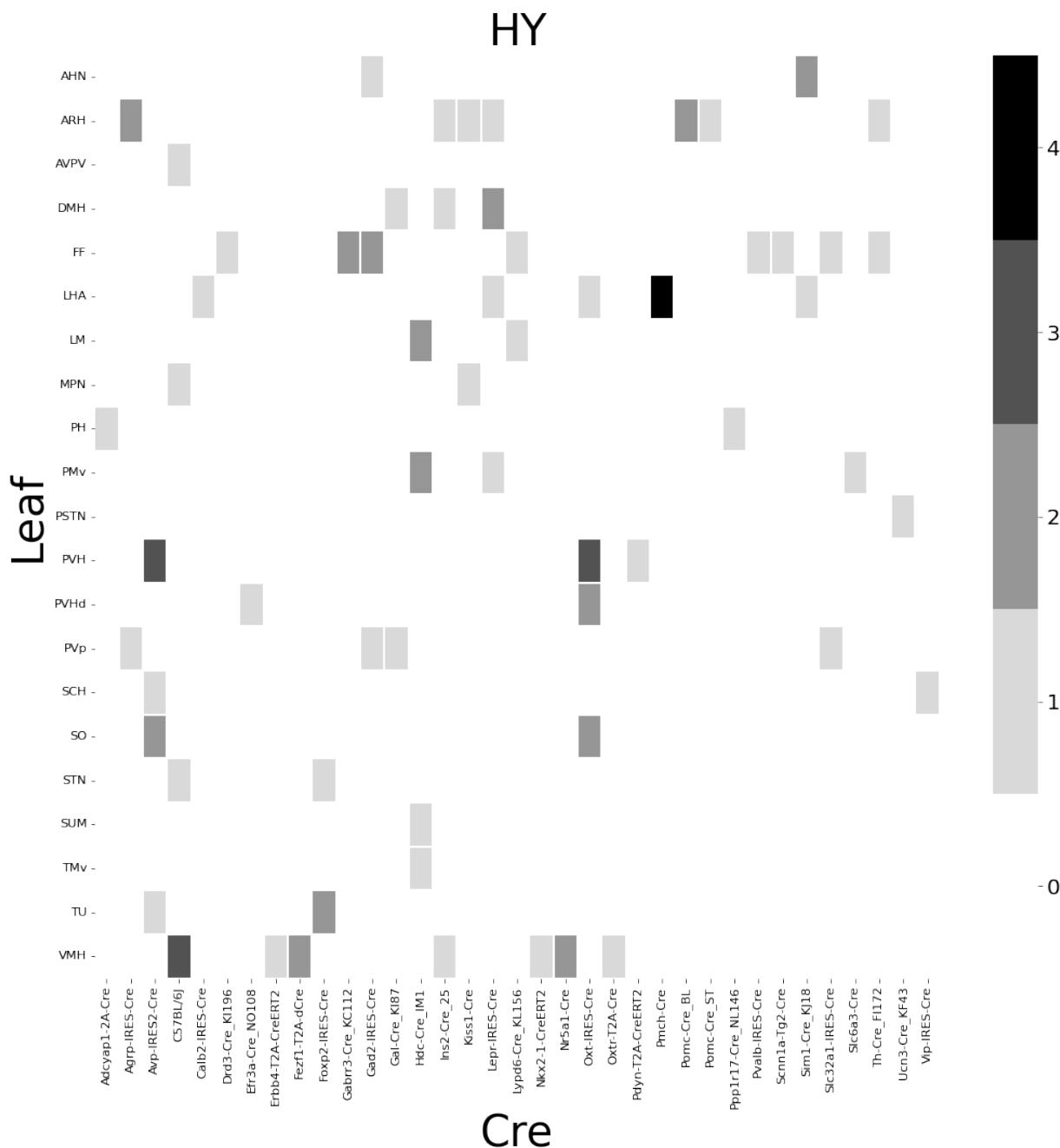


Figure 16: Abundances of cre-line and leaf-centroid combinations.

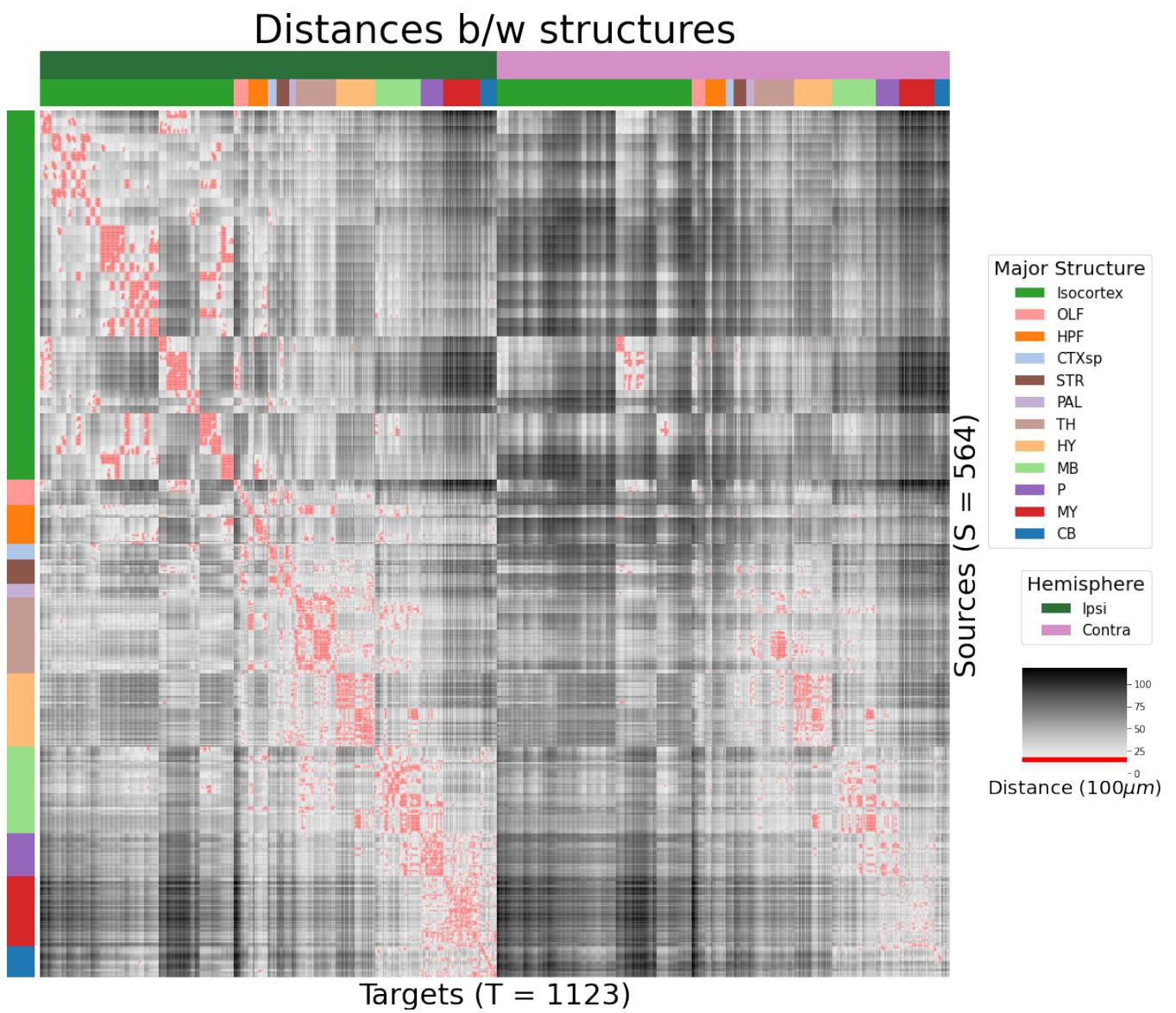
328 ***Distances between structures***

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

**329 Model evaluation**

330 We give information on the quality of our models. This includes the sizes of our evaluation sets in  
 331 leave-one-out cross-validation and additional losses in the injection-normalized case.

332 NUMBER OF EXPERIMENTS IN EVALUATION SETS In order to compare between methods, we therefore  
 333 restrict to the smallest set of evaluation indices, which is to say, virus-leaf combinations that are  
 334 present at least twice. This means that our evaluation set in size is smaller than our evaluation set is  
 335 smaller in size than our overall list of experiments.

	Total	Cre-Leaf	Cre-leaf over threshold
Isocortex	36	4	4
OLF	7	2	2
HPF	122	62	59
CTXsp	85	41	38
STR	1128	732	7
PAL	68	18	17
TH	46	7	7
HY	35	17	17
MB	33	8	8
P	30	11	11
MY	78	45	44
CB	83	29	29

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

336 INJECTION-NORMALIZED LOSSES To compare with the injection-normalization procedure from Knox  
 337 et al. (2019), we also remove experiments with small injection, and here give results for this slightly  
 338 reduced set using injection-normalization. That is, instead of dividing the projection signal of each  
 339 experiment by its  $l_1$  norm, we divide by the  $l_1$  norm of the corresponding injection signal. We find  
 340 that setting a summed injection-signal of threshold of 1 is sufficient for evading pathological edge  
 341 cases in this normalization, while still retaining a large 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 structural connectivity on injection-thresholded evaluation set. **Bold** numbers are best for their major structure.

<sup>342</sup> PROJECTION-NORMALIZED LOSSES ON THRESHOLDED SET We also give results for the  
<sup>343</sup> 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 structural connectivity on injection-thresholded evaluation set. **Bold** numbers are best for their major structure.

## 6 SUPPLEMENTAL METHODS

<sup>344</sup> This section consists of additional information on preprocessing of the neural connectivity data,  
<sup>345</sup> estimation of connectivity, and matrix factorization.

<sup>346</sup> ***Data preprocessing***

<sup>347</sup> Several data preprocessing steps take place prior to evaluations of the connectivity matrices. These  
<sup>348</sup> steps are described in Algorithm PREPROCESS. The arguments of this normalization process - injection  
<sup>349</sup> signals  $x(i)$ , projection signals  $y(i)$ , injection fraction  $F(i)$ , and data quality mask  $q(i)$  - were  
<sup>350</sup> downloaded using the Allen SDK. The injections and projection signals  $\mathcal{B} \rightarrow [0, 1]$  were segmented  
<sup>351</sup> manually in histological analysis. The projection signal gives the proportion of pixels within the voxel  
<sup>352</sup> displaying fluorescence, and the injection signal gives the proportion of pixels within the  
<sup>353</sup> histologically-selected injection subset displaying fluorescence. The injection fraction  $\mathcal{B} \rightarrow [0, 1]$  gives  
<sup>354</sup> the proportion of pixels within each voxel in the injection subset. Finally, the data quality mask  
<sup>355</sup>  $\mathcal{B} \rightarrow \{0, 1\}$  gives the voxels that have valid data.

<sup>356</sup> Our preprocessing makes use of the above ingredients, as well as several other essential steps. First,  
<sup>357</sup> we compute the weighted injection centroid

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

<sup>358</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>359</sup> we can also access regionalization map  $R: \mathcal{B} \rightarrow \mathcal{R}$ . This induces a functional of connectivities from  
<sup>360</sup> the space of maps  $\{\mathcal{X} = x: \mathcal{B} \rightarrow [0, 1]$

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

<sup>361</sup> We also can restrict a signal to a individual structure as

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

<sup>362</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>363</sup> We define these objects as functions and functionals, but this is for notational convenience and  
<sup>364</sup> non-essential. A function  $x(i) : \mathcal{B} \rightarrow [0, 1]$  is mathematically equivalent to the graph  
<sup>365</sup>  $\mathcal{G}(x(i)) \in \mathcal{B} \times [0, 1]$ . As an abuse of notation, we define  $x \odot x' := z$  such that  $z(l) = x(l)x'(l)$  for all  $l \in \mathcal{B}$ .  
<sup>366</sup> Also, denote  $m(i)$  as the major structure containing experiment  $i$ . We then can write the  
<sup>367</sup> preprocessing algorithm.

---

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

---

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

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

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

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

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

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

---

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

---

368 **Estimators**

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

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

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

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

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

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

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

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

381 Several facets of the estimator are visible here. A smaller  $\gamma$  corresponds to a greater amount of  
 382 smoothing, and the index set  $I \subseteq \{1 : n\}$  generally depends on  $s$  and  $\nu$ . Fitting  $\gamma$  via empirical risk  
 383 minimization therefore bridges between 1-nearest neighbor prediction and averaging of all  
 384 experiments in  $I$ . In Knox et al. (2019),  $I$  consisted of experiments sharing the same brain division, i.e.  
 385  $I = I_m$ , while restricting of index set to only include experiments with the same cell class gives the  
 386 class-specific Cre-NW model. Despite this restriction, we fit  $\gamma$  for each  $m$  rather than a smaller subset  
 387 like  $s$  or  $\nu$ . 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>388</sup> *The Expected-Loss estimator* Besides the injection location, the targeted cell class also influences  
<sup>389</sup> projection. Since Cre-lines that target similar classes are induce similar projections, and including  
<sup>390</sup> similar Cre-lines in the Nadaraya-Watson estimator increases effective sample size, we introduce an  
<sup>391</sup> estimator that assigns a predictive weight to each training point that depends both on its  
<sup>392</sup> centroid-distance and Cre-line. This weight is determined by the expected prediction error of each of  
<sup>393</sup> the two feature types, as determined by cross-validation. For this reason, we call this the Expected  
<sup>394</sup> Loss Estimator. The resulting weights are then utilized in a Nadaraya-Watson estimator in a final  
<sup>395</sup> prediction step.

<sup>396</sup> We formalize Cre-line behavior as the average regionalized projection of a Cre-line in a given  
<sup>397</sup> structure (i.e. leaf). This vectorization of categorical information is known as **target encoding**, and we  
<sup>398</sup> define this as  $\bar{y}_{\mathcal{T},s,v} := \frac{1}{|I_s \cap I_v|} \sum_{i \in (I_s \cap I_v)} y_{\mathcal{T}}(i)$ . We define a **Cre-distance** in a leaf to be the distance  
<sup>399</sup> between the target-encoded projections of two Cre-lines. The relative predictive accuracy of  
<sup>400</sup> Cre-distance and centroid distance is determined by fitting a surface of projection distance as a  
<sup>401</sup> function of Cre-distance and centroid distance.

<sup>402</sup> In mathematical terms, our full feature set consists of the centroid coordinates and the  
<sup>403</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>404</sup>  $f^*$  is defined as in (2). The expected loss estimator is then

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

<sup>405</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>406</sup> and  $s$  is the structure containing  $l$ .

<sup>407</sup> The key step therefore is finding a suitable  $g$  with which to weight the positional and Cre  
<sup>408</sup> information. Note that  $g$  must be a concave, non-decreasing function of its arguments with with  
<sup>409</sup>  $g(0, 0) = 0$ , then  $g$  defines a metric on the product of the metric spaces defined by experiment centroid  
<sup>410</sup> and target-encoded cre-line, and  $\hat{f}_{EL}$  is a Nadaraya-Watson estimator. A derivation of this fact is given

<sup>411</sup> later in this section, and we therefore use shape-constrained B-splines to estimate  $g$ . Similarly to the  
<sup>412</sup> Nadaraya-Watson model, we make the decision to fit a  $g$  separately for each major brain division. We  
<sup>413</sup> can then select  $\hat{\gamma}$  as in 2.

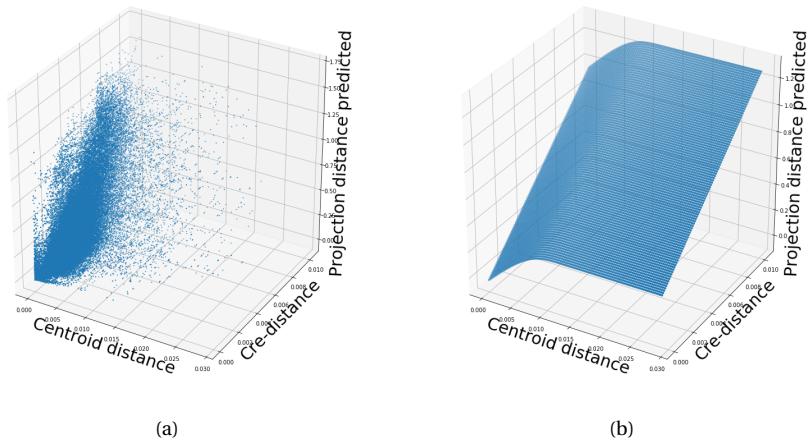


Figure 18: Fitting  $g$ . 18a Distribution of projection errors against centroid distance and cre-distance in Isocortex. 18b  $\hat{g}$  found using B-splines.

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

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

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

- 423    ▪ The first property of a metric is that  $f(x, x') = 0 \iff x = x'$ . The left implication:  
 424        $x = x' \implies f(x_1, x'_1, \dots, x_n, x'_n) = g(0, \dots, 0)$ , since  $d$  are metrics. Then, since  $g(0) = 0$ , we have that  
 425        $f(x, x') = 0$ . The right implication:  $f(x, x') = 0 \implies d = 0 \implies x = x'$  since  $d$  are metrics.
- 426    ▪ The second property of a metric is that  $f(x, x') = f(x', x)$ . This follows immediately from the  
 427       symmetry of the  $d_i$ , i.e.  $f(x, x') = f(x_1, x'_1, \dots, x_n, x'_n) = g(d_1(x_1, x'_1), \dots, d_n(x_n, x'_n)) =$   
 428        $g(d_1(x'_1, x_1), \dots, d_n(x'_n, x_n)) = f(x'_1, x_1, \dots, x'_n, x_n) = f(x', x)$ .
- 429    ▪ The third property of a metric is the triangle inequality:  $f(x, x') \leq f(x, x^*) + f(x^*, x')$ . To show this  
 430       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  
 431       non-decreasing and by the triangle inequality of  $d$ . Then, since  $g$  is concave,  
 432        $g(d(x, x^*) + d(x^*, x')) \leq g(d(x, x^*)) + g(d(x^*, x')) = f(x, x^*) + f(x^*, x')$ .

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

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

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

445



***446 Establishing a lower detection limit***

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

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

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

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

458 ***Decomposing the connectivity matrix***

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

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

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

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

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

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

480 *Cross-validating NMF* Cross-validation for NMF is somewhat standard but not entirely well-known,  
 481 and so we review it here. In summary, a NMF model is first fit on a reduced data set, and an evaluation

482 set is held out. After random masking of the evaluation set, the loss of the learned model is then  
 483 evaluated on the basis of successful reconstruction of the held-out values. This procedure is  
 484 performed repeatedly, with replicates of random masks at each tested dimensionality  $q$ . This  
 485 determines the point past which additional hidden units provide no reconstructive value.

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

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

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

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

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

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

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

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

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

491 *Stabilizing NMF* The NMF program is non-convex, and, empirically, individual replicates will not  
 492 converge to the same optima. One solution therefore is to run multiple replicates of the NMF  
 493 algorithm and cluster the resulting vectors. This approach raises the questions of how many clusters  
 494 to use, and how to deal with stochasticity in the clustering algorithm itself. We address this issue  
 495 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

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

<sup>496</sup> A review of this approach is found in von Luxburg (2010b). Intuitively, archetype vectors that cluster  
<sup>497</sup> together frequently over clustering replicates indicate the presence of a stable clustering. For  $d$ , we  
<sup>498</sup> utilize the adjusted Rand Index - a simple dissimilarity measure between clusterings. Note that we  
<sup>499</sup> expect to select slightly more than the  $q$  components suggested by cross-validation, since archetype  
<sup>500</sup> vectors which appear in one NMF replicate generally should appear in others. We then select the  $q$   
<sup>501</sup> clusters with the most archetype vectors - the most stable NMF results - and take the median of each  
<sup>502</sup> cluster to create a sparse representative archetype Kotliar et al. (2019); Wu et al. (2016). We then find  
<sup>503</sup> the according  $H$  using Program 3. Experimental results for these cross-validation and stability  
<sup>504</sup> selection approaches are given in Supplemental Section 7.

## 7 SUPPLEMENTAL EXPERIMENTS

### 505 Establishing a lower limit of detection

506 We give results on the false detection rate at different limits of detection. These conclusively show that  
 507  $10^{-6}$  is the good threshold for our normalized data.

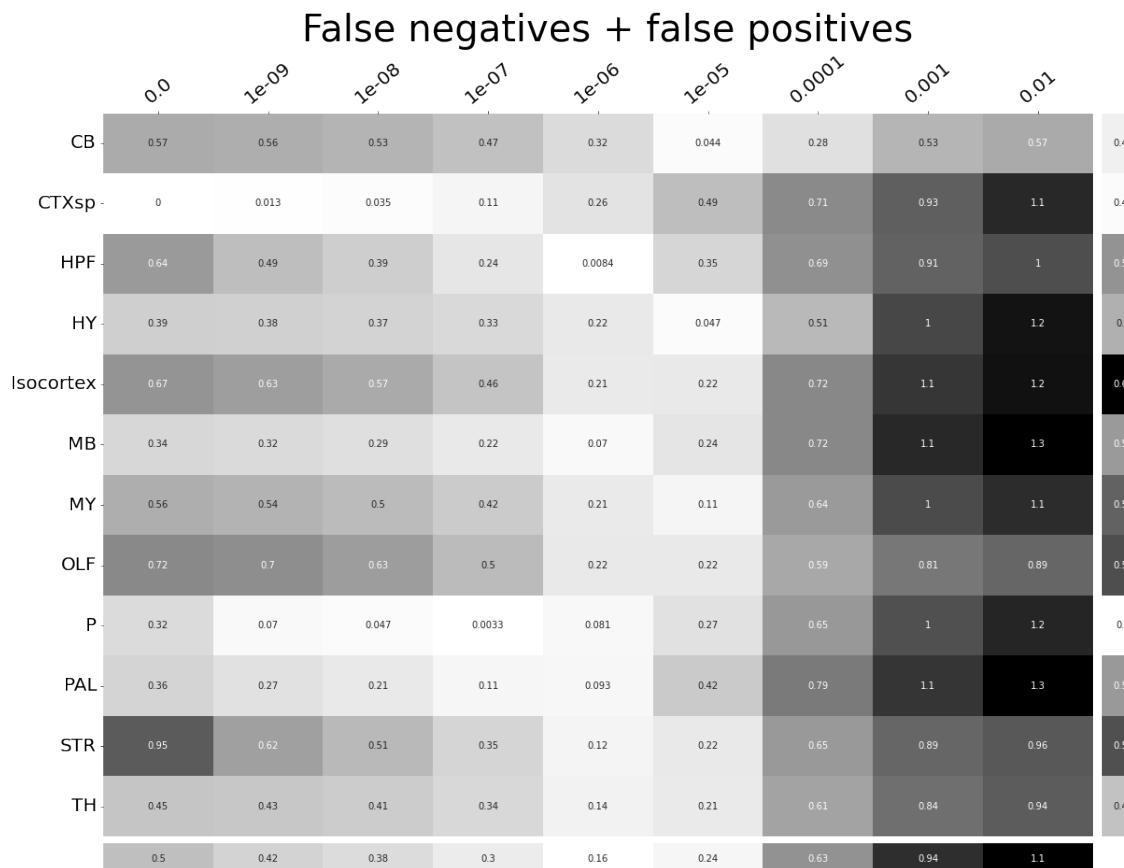


Figure 19:  $\tau$  at different limits of detection in different major structures.  $10^{-6}$  is the optimal detection threshold.

508 ***Loss subsets***

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



Figure 20: Weighted loss for cre-leaf combinations in CB. Missing values are omitted. Row and column averages are also plotted.

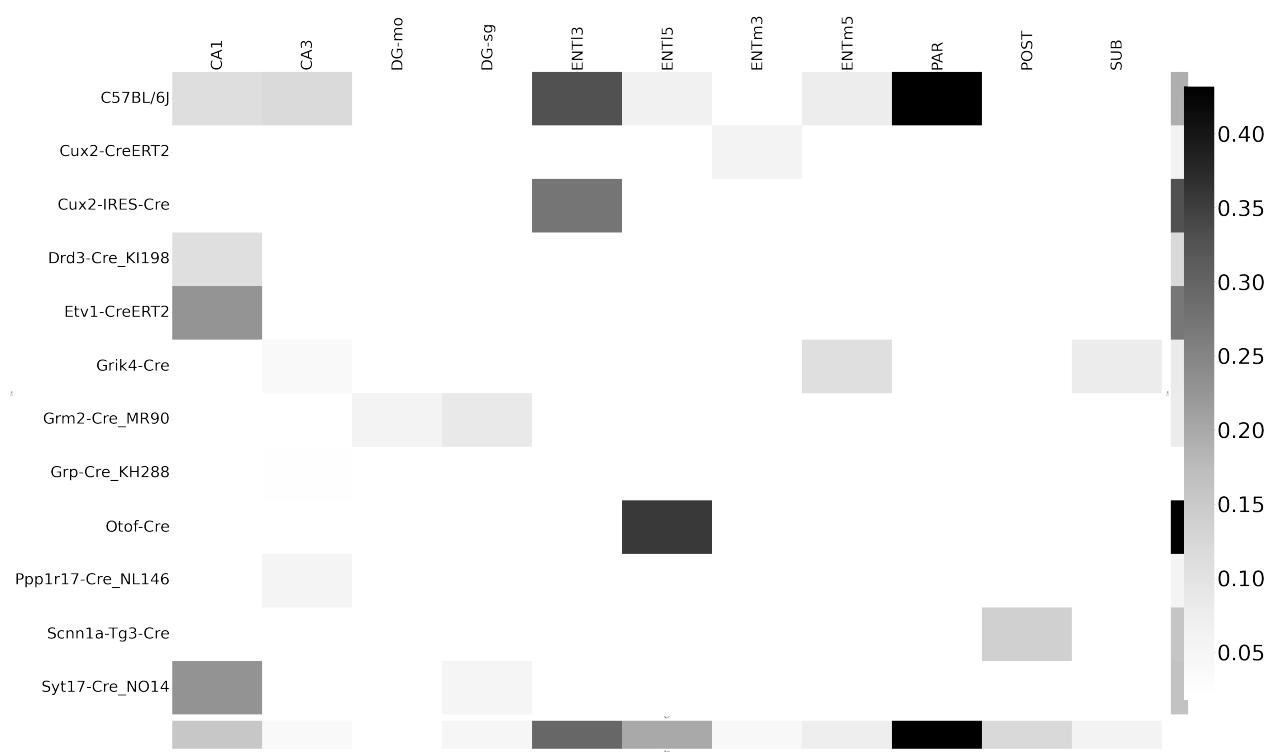


Figure 21: Weighted loss for cre-leaf combinations in HPF. Missing values are omitted. Row and column averages are also plotted.

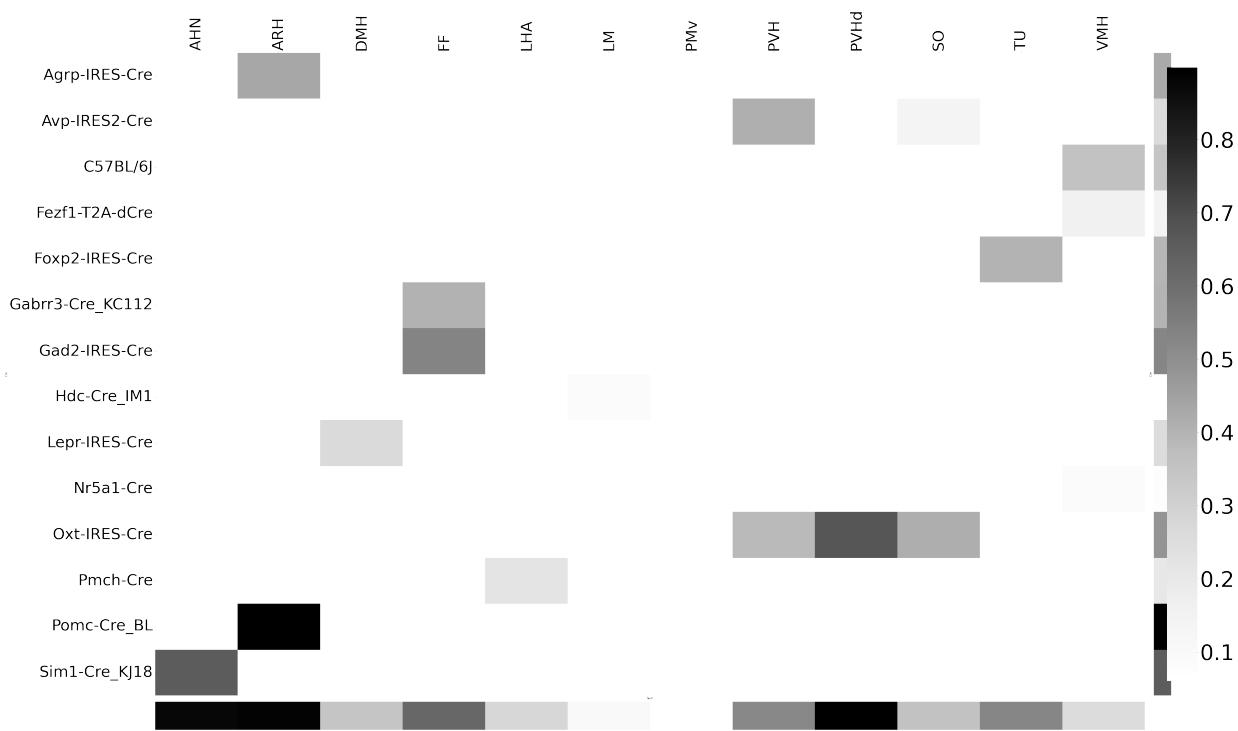


Figure 22: Weighted loss for cre-leaf combinations in HY. Missing values are omitted. Row and column averages are also plotted.

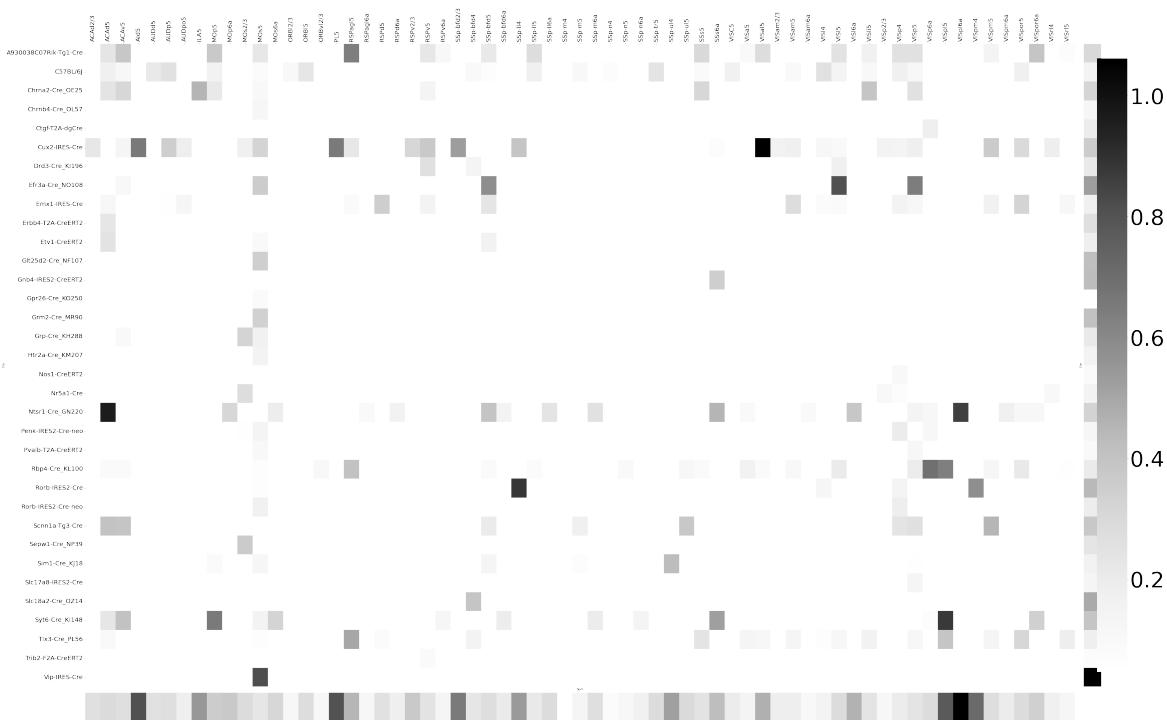


Figure 23: Weighted loss for cre-leaf combinations in Isocortex. Missing values are omitted. Row and column averages are also plotted.



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.

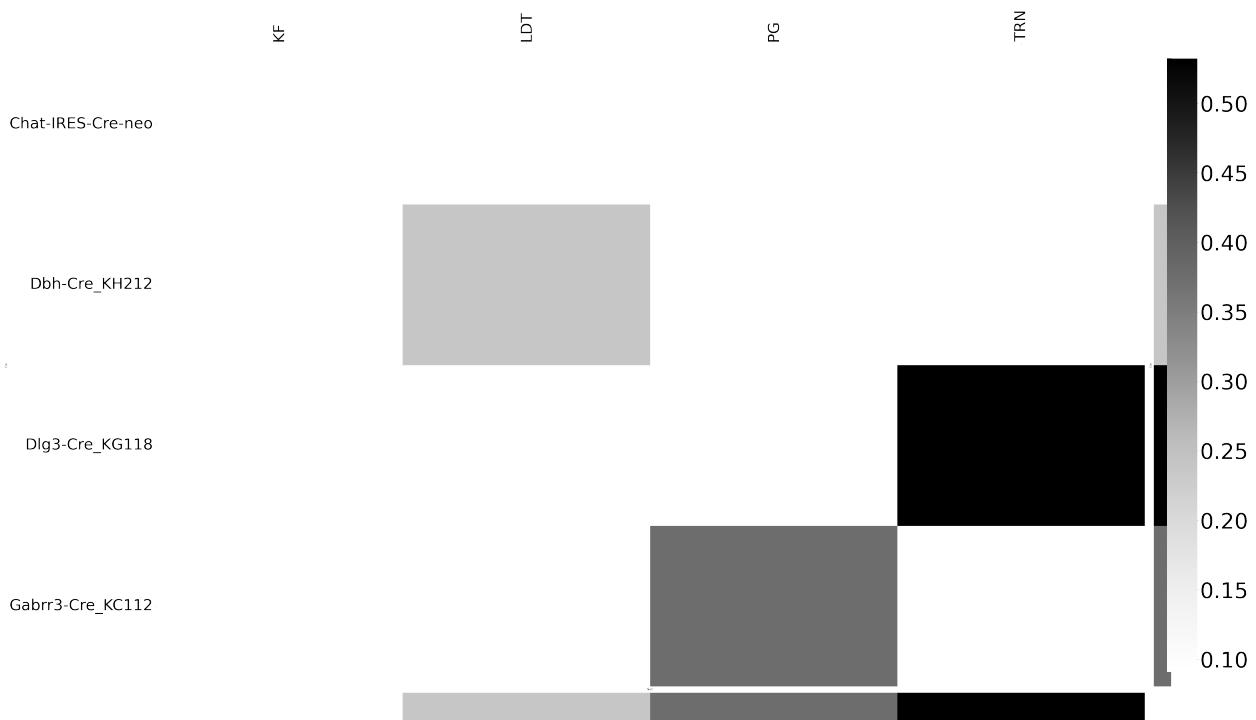


Figure 27: Weighted loss for cre-leaf combinations in P. Missing values are omitted. Row and column averages are also plotted.

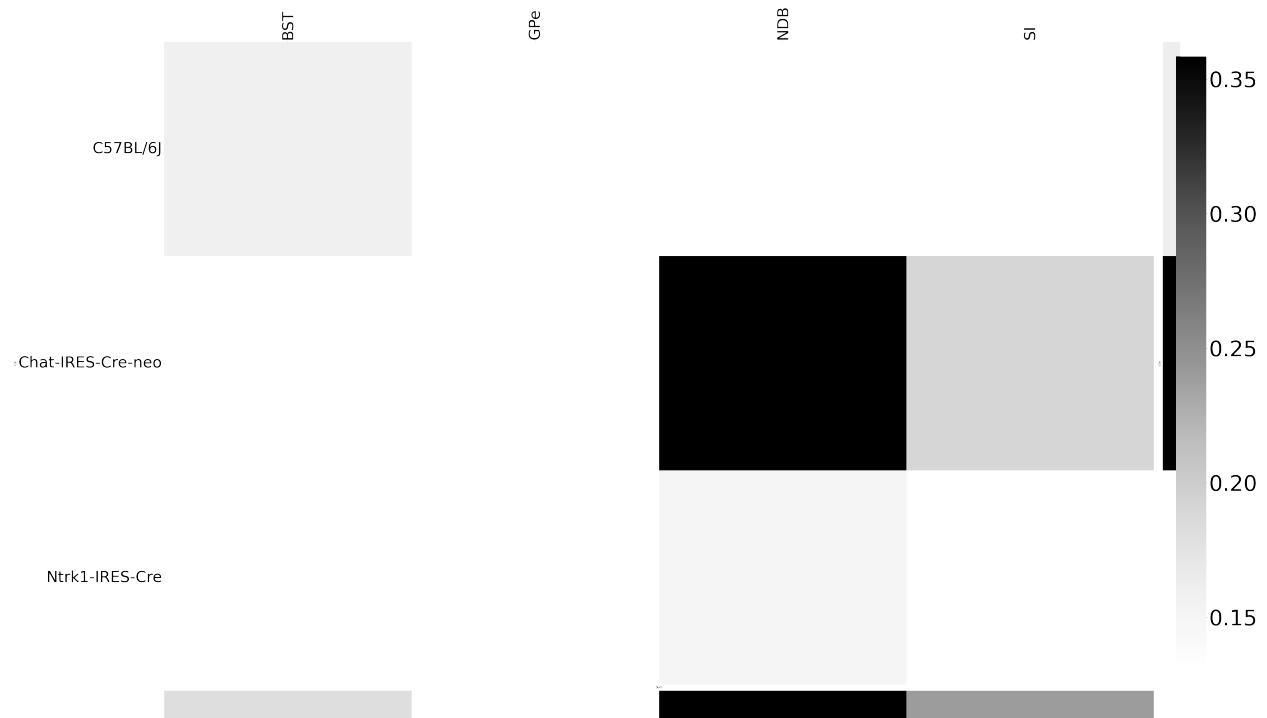


Figure 28: Weighted loss for cre-leaf combinations in PAL. Missing values are omitted. Row and column averages are also plotted.

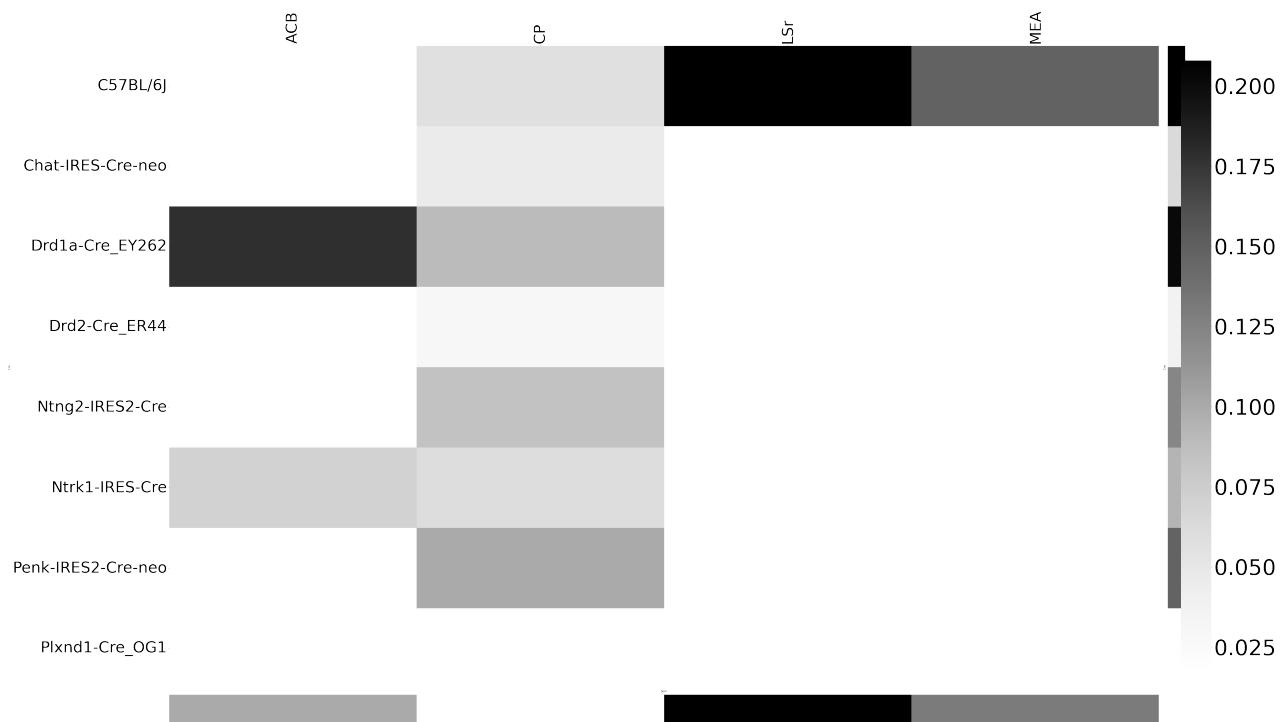


Figure 29: Weighted loss for cre-leaf combinations in STR. Missing values are omitted. Row and column averages are also plotted.

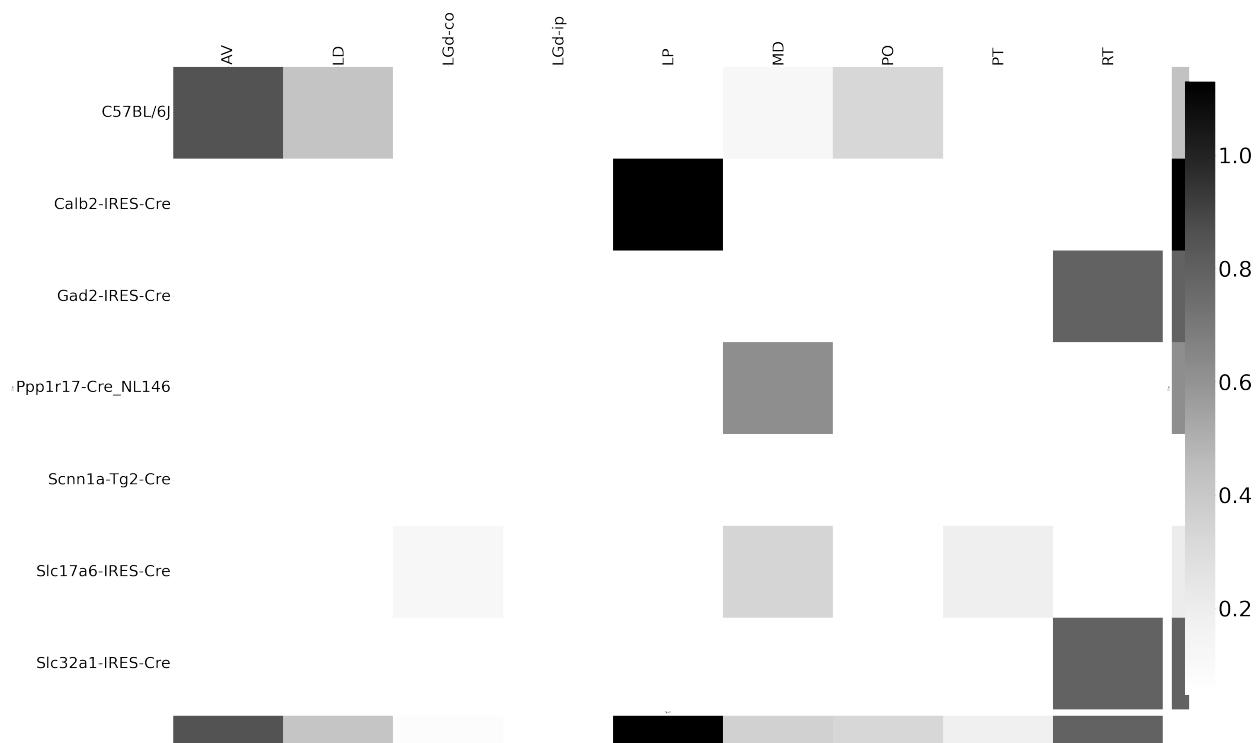


Figure 30: Weighted loss for cre-leaf combinations in TH. Missing values are omitted. Row and column averages are also plotted.

512 **Matrix Factorization**

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

516 *Cross-validation* We set  $\alpha = 0.002$  and run Program 2 on  $\mathcal{C}_{wt}$ . We use a random mask with  $p = .3$  to  
 517 evaluate prediction accuracy of models trained on the unmasked data on the masked data. To  
 518 account for stochasticity in the NMF algorithm, we run  $R = 8$  replicates at each potential dimension  $q$ .  
 519 This selects  $\hat{q} = 60$ .

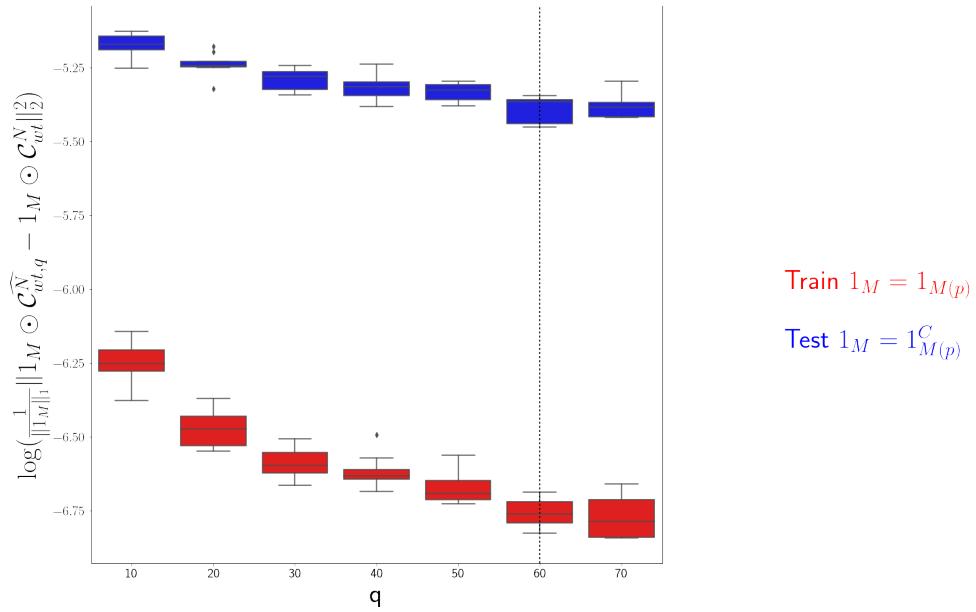


Figure 31: Train and test error using NMF decomposition.

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

524 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

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

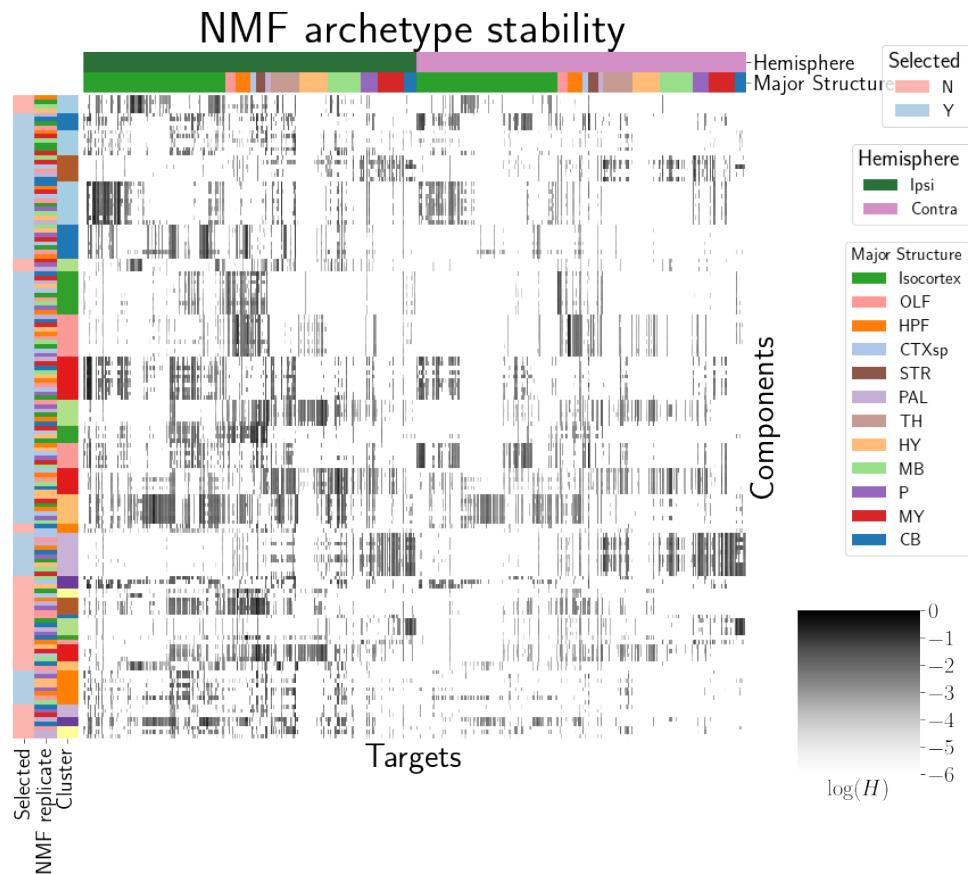


Figure 32: Stability of NMF results across replicates. Replicate and NMF component are shown on rows. Components that are in the top 15 are also indicated.

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

## 8 COMPETING INTERESTS

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

530

531

532

## REFERENCES

- 533 Brunet, J.-P., Tamayo, P., Golub, T. R., & Mesirov, J. P. (2004). Metagenes and molecular pattern discovery using matrix  
534 factorization. *Proc. Natl. Acad. Sci. U. S. A.*, 101(12), 4164–4169.
- 535 Chamberlin, N. L., Du, B., de Lacalle, S., & Saper, C. B. (1998). Recombinant adeno-associated virus vector: use for  
536 transgene expression and anterograde tract tracing in the CNS. *Brain Res.*, 793(1-2), 169–175.
- 537 Daigle, T. L., Madisen, L., Hage, T. A., Valley, M. T., Knoblich, U., Larsen, R. S., ... Zeng, H. (2018). A suite of transgenic  
538 driver and reporter mouse lines with enhanced Brain-Cell-Type targeting and functionality. *Cell*, 174(2), 465–480.e22.
- 539 Gao, Y., Zhang, X., Wang, S., & Zou, G. (2016). Model averaging based on leave-subject-out cross-validation. *J. Econom.*,  
540 192(1), 139–151.
- 541 Harris, J. A., Mihalas, S., Hirokawa, K. E., Whitesell, J. D., Choi, H., Bernard, A., ... Zeng, H. (2019). Hierarchical  
542 organization of cortical and thalamic connectivity. *Nature*, 575(7781), 195–202.
- 543 Harris, J. A., Oh, S. W., & Zeng, H. (2012). Adeno-associated viral vectors for anterograde axonal tracing with fluorescent  
544 proteins in nontransgenic and cre driver mice. *Curr. Protoc. Neurosci., Chapter 1, Unit 1.20.1–18*.
- 545 Harris, K. D., Mihalas, S., & Shea-Brown, E. (2016). Nonnegative spline regression of incomplete tracing data reveals high  
546 resolution neural connectivity.
- 547 Jeong, M., Kim, Y., Kim, J., Ferrante, D. D., Mitra, P. P., Osten, P., & Kim, D. (2016). Comparative three-dimensional  
548 connectome map of motor cortical projections in the mouse brain. *Sci. Rep.*, 6, 20072.
- 549 Knox, J. E., Harris, K. D., Graddis, N., Whitesell, J. D., Zeng, H., Harris, J. A., ... Mihalas, S. (2019). High-resolution  
550 data-driven model of the mouse connectome. *Netw Neurosci*, 3(1), 217–236.
- 551 Kotliar, D., Veres, A., Nagy, M. A., Tabrizi, S., Hodis, E., Melton, D. A., & Sabeti, P. C. (2019). Identifying gene expression  
552 programs of cell-type identity and cellular activity with single-cell RNA-Seq. *Elife*, 8.
- 553 Lotfollahi, M., Naghipourfar, M., Theis, F. J., & Alexander Wolf, F. (2019). Conditional out-of-sample generation for  
554 unpaired data using trVAE.

- 555 Oh, S. W., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., ... Zeng, H. (2014). A mesoscale connectome of the  
556 mouse brain. *Nature*, 508(7495), 207–214.
- 557 Saul, L. K., & Roweis, S. T. (2003). Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *J.*  
558 *Mach. Learn. Res.*, 4(Jun), 119–155.
- 559 von Luxburg, U. (2010a). Clustering stability: An overview.
- 560 von Luxburg, U. (2010b). Clustering stability: An overview.
- 561 Wu, S., Joseph, A., Hammonds, A. S., Celtniker, S. E., Yu, B., & Frise, E. (2016). Stability-driven nonnegative matrix  
562 factorization to interpret spatial gene expression and build local gene networks. *Proc. Natl. Acad. Sci. U. S. A.*, 113(16),  
563 4290–4295.

## 9 TECHNICAL TERMS

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

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

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

569 **Structural connectivities** connectivity between structures

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

571 **Structural connection tensor** Connectivities between structures given a neuron class

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

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

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

576 **Connectivity archetypes** Typical connectivity patterns

577 **Expected loss** Our new estimator that weights different features by their estimated predictive  
578 power.