

1 RESEARCH

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

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

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

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

7 Keywords: [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

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

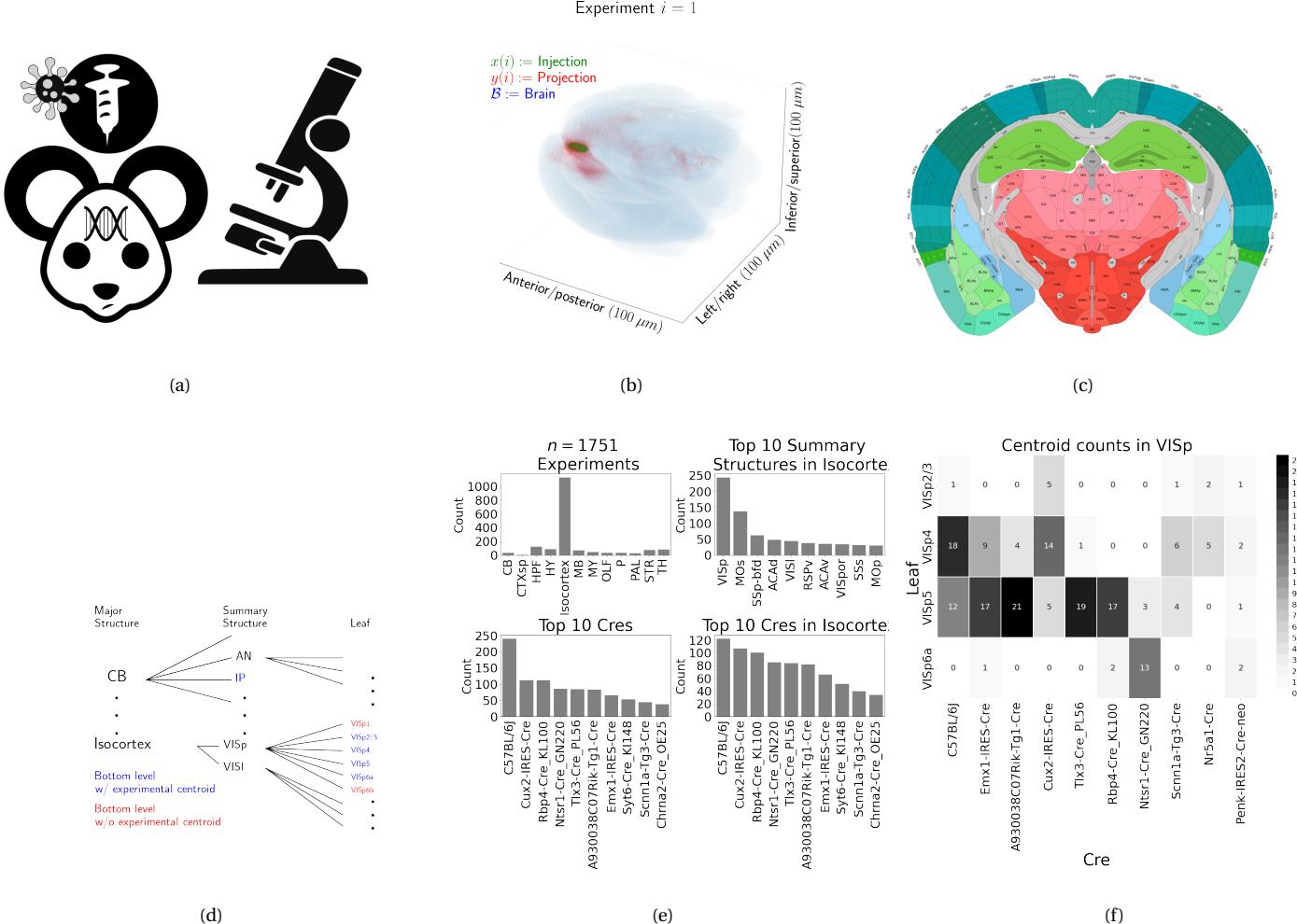


Figure 1: Experimental setting. 1a For each experiment, a potentially Cre-recombinase promoted GFP-expressing transgene cassette is transduced after stereotaxic injection into a Cre-driver mouse, followed by two-photon tomography imaging. 1b An example of the segmentation of projection and injection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

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

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

137 These are the total strength and average strength from source to target regions for each class (SK's
 138 **comment: no density**). Since the normalized strength is computable from the strength via a fixed
 139 normalization, our main statistical goal is to estimate $\mathcal{C}(v, s, t)$ for all v, s and t . We call this estimator
 140 $\hat{\mathcal{C}}$.

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

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

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

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

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

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

161 ***Model evaluation***

162 We select optimum functions from within and between our estimator classes using *leave-one-out*
 163 cross validation, in which the accuracy of the model is assessed by its ability to predict experiments
 164 excluded from the training data. Equation 1 includes a deterministic step f^* included without input
 165 by the data. The performance of $\widehat{\mathcal{C}}(\nu, s, t)$ is thus determined by performance of $\widehat{f}(f_*(\mathcal{D}(\nu, s)))$.
 166 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
 167 location. This is the predictand we evaluate.

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

We use weighted l_2 -loss to evaluate these predictions.

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

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

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

180 ***Connectivity analyses***

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

192 Our algorithm is

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

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

3 RESULTS

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

208 ***Model evaluation***

209 Our EL model generally performs better than the other estimators that we consider. Table 2 contains
210 weighted losses from leave-one-out cross-validation of candidate models, such as the NW Major-WT
211 model from Knox et al. (2019). The EL model combines the good performance of class-specific
212 models like NW Leaf-Cre in regions like Isocortex with the good performance of class-agnostic models
213 in regions like Thalamus. Additional information on model evaluation, including class and structure-
214 specific performance, is given in Appendix 5. In particular, Supplementary Table 4 contains the sizes
215 of these evaluation sets in each major structure, and Supplementary Section 7 contains the structure-
216 and class specific losses.

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

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

217 ***Connectivities***

218 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
 219 to target structures for particular cre-lines v . We confirm the detection of several well-established
 220 connectivities within our tensor, although it is our expectation that additional interesting biological
 221 processes are also manifest. The connectivity tensor and code to reproduce it are available at
 222 https://github.com/AllenInstitute/mouse_connectivity_models/tree/2020.

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

229 Compared with the wild-type specific connectivities in Knox et al. (2019), ours appear more
 230 variable. This is both because of the layer-specific targeting, and also the layer-specificity of the
 231 selected model. Although layer-specificity is a major advantage of including distinct Cre-lines, for
 232 comparison, we also plot coarser projections between summary-structure sources and targets in the
 233 cortex in Figure 2b. These are averages over component layers weighted by layer size. Grossly
 234 congruent with the previous work, our results exhibit a larger range of connectivities than those in
 235 Knox et al. (2019), and therefore appear more dense.

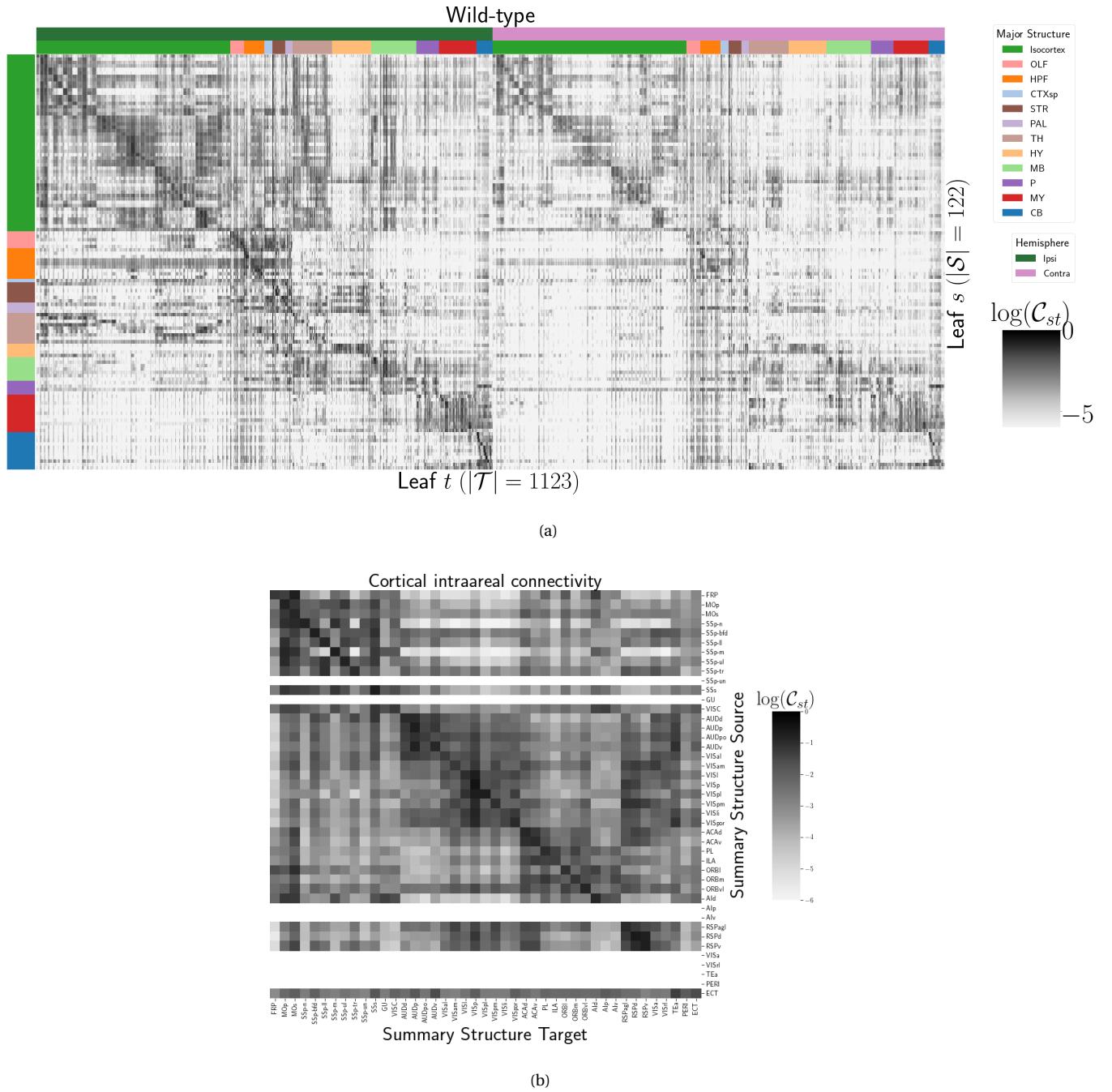
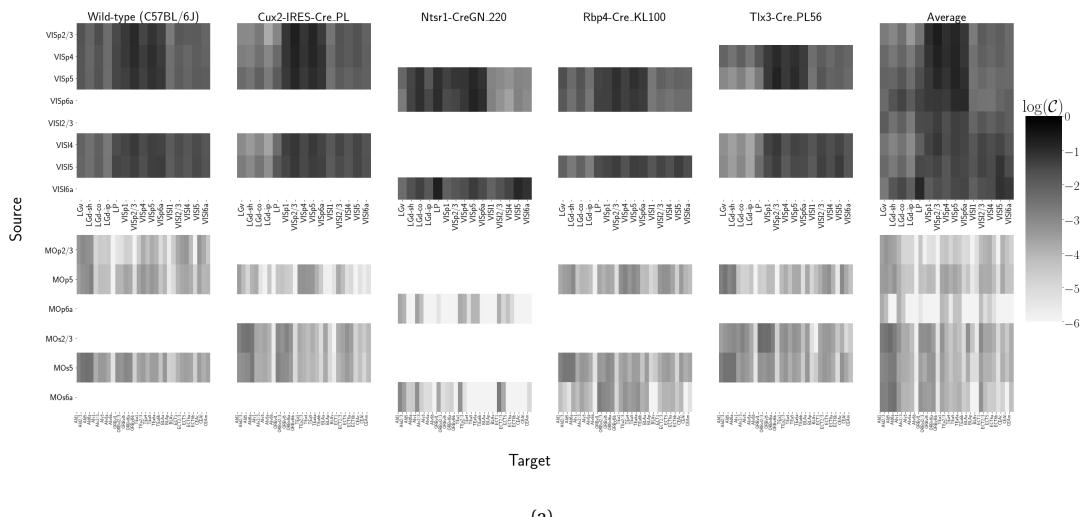


Figure 2: Wild-type connectivities. 2a Log wild-type connectivity matrix $\log \mathcal{C}(s, t, v_{wt})$. 2b Log wild-type intracortical connectivity matrix at the summary structure level.

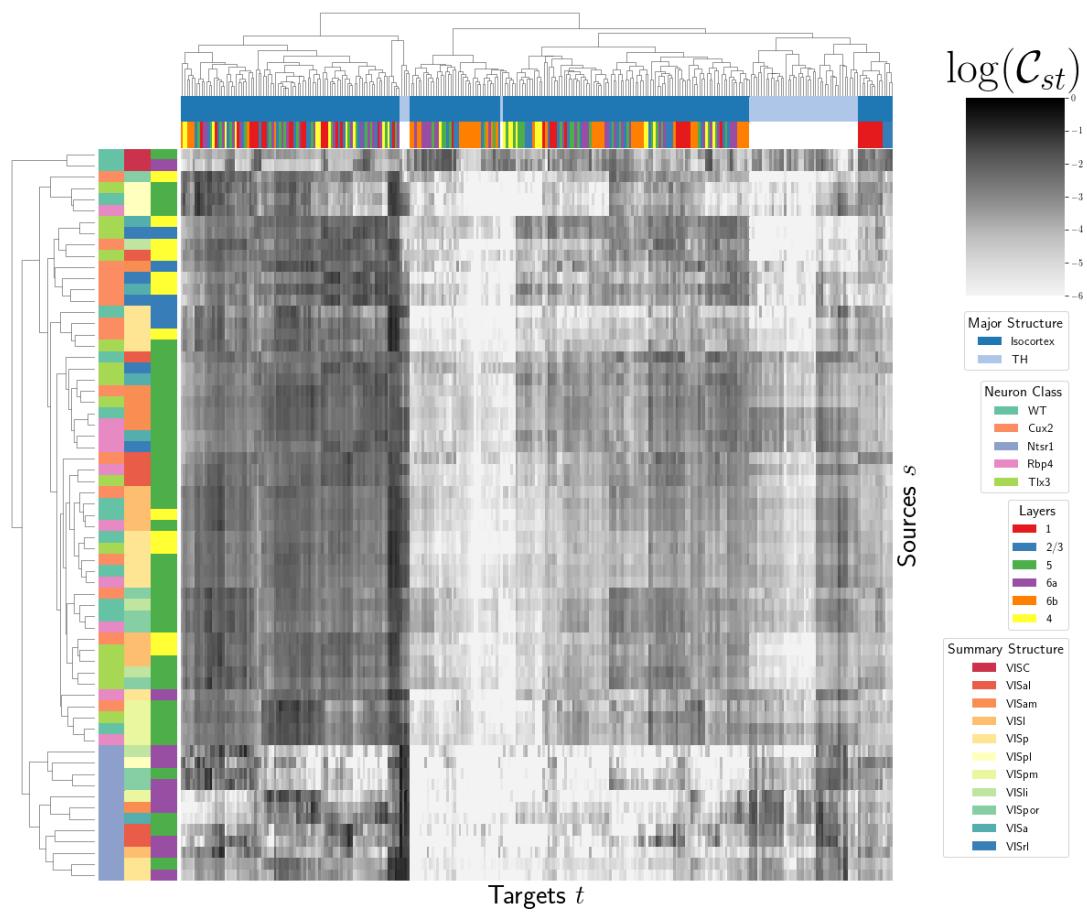
236 *Class-specific connectivities* Source and cell-type combinations which project similarly indicate the
237 network structure underpinning cognition. Our estimates of these class-specific connectivities exhibit
238 certain known behaviors. In Figure 3, we display results for the VISp and MO cortical 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 Figure ??). 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.



(a)



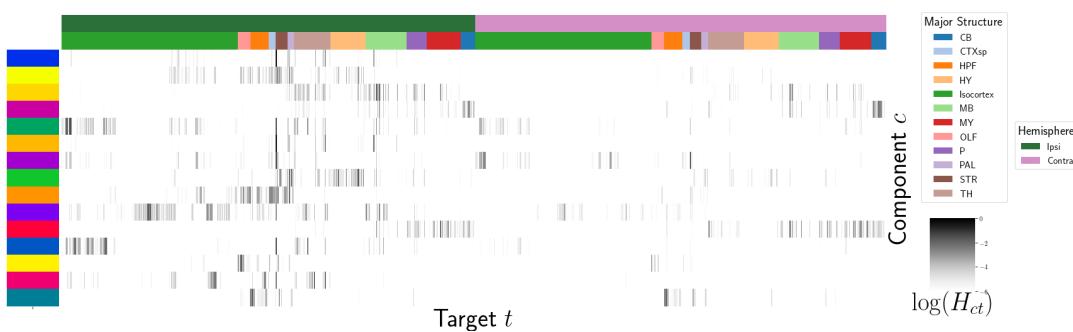
(b)

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 thalymic targets. Cre-line, summary structure, and layer are labelled on the sources. Major brain division and layer are labelled⁸ on the targets.

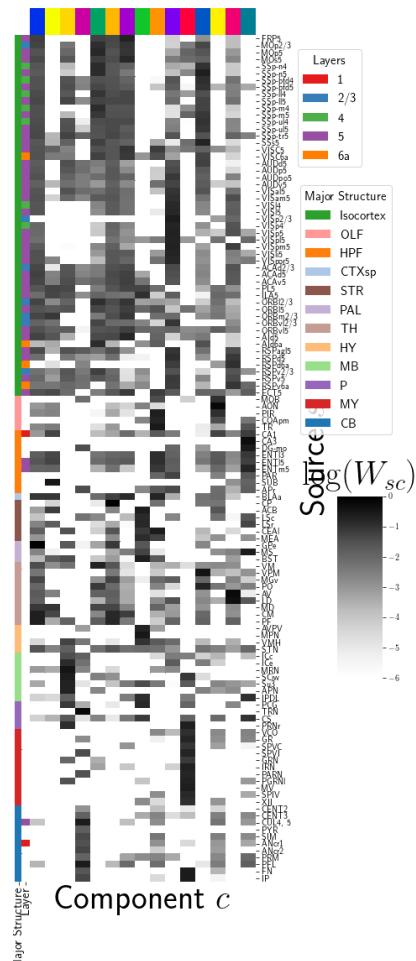
263 *Connectivity Analyses*

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.



(a)



(b)

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

4 DISCUSSION

280 The model presented here is a milestone in characterization of connectomes. It is the first cell-type
281 specific whole brain projectomme for a mammalian species. It can open the door for a large umber of
282 models linking brain structure to computational architectures.

283 The method presented here is also novel. Beyond using a Nadaraya-Watson kernel regression
284 defined in physical space, we define a cell-type space based on similarities of projections, and expand
285 the method to work in this space as well.

286 We see several opportunities for improving on our model. Our particular task of transforming the
287 injection and projection signal depending on cell-type is a non-linear transformation problem with
288 categorical covariate. Model averaging based off of cross-validation has been implemented in Gao,
289 Zhang, Wang, and Zou (2016), but we note that our approach makes use of a non-parametric
290 estimator, rather than an optimization method for selecting the weights (Saul & Roweis, 2003), and is
291 applied specifically to a target-encoded feature space. The properties of this estimator, as well as its
292 relation to estimators fit using an optimization algorithm, are a possible future avenue of research.

293 Therefore, a deep model such as Lotfollahi, Naghipourfar, Theis, and Alexander Wolf (2019) could be
294 appropriate, provided enough data was available. With respect to the model, a Wasserstein-based
295 measure of injection similarity per structure would combine both the physical simplicity of the
296 centroid model while also incorporating structural knowledge. Residual models of the above could
297 also be considered.

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

ACKNOWLEDGMENTS

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

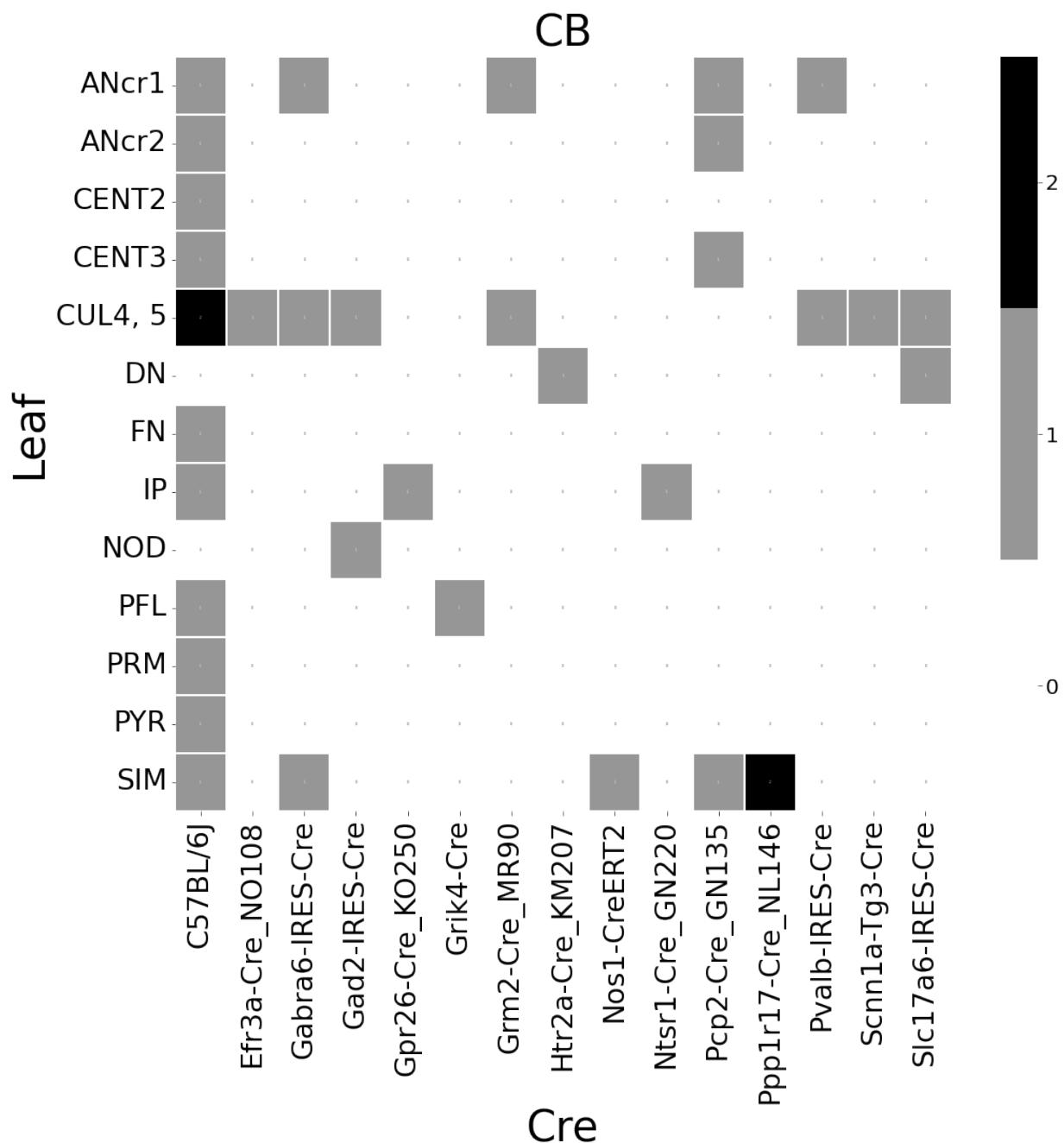
304 This supplement is divided into information about our dataset, supplemental methods, and
305 supplemental results. However, note that there is substantial interconnectedness between these
306 sections. Thus, if a reader is interested in, say, non-negative matrix factorization, they may find
307 relevant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

5 SUPPLEMENTAL INFORMATION

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

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

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



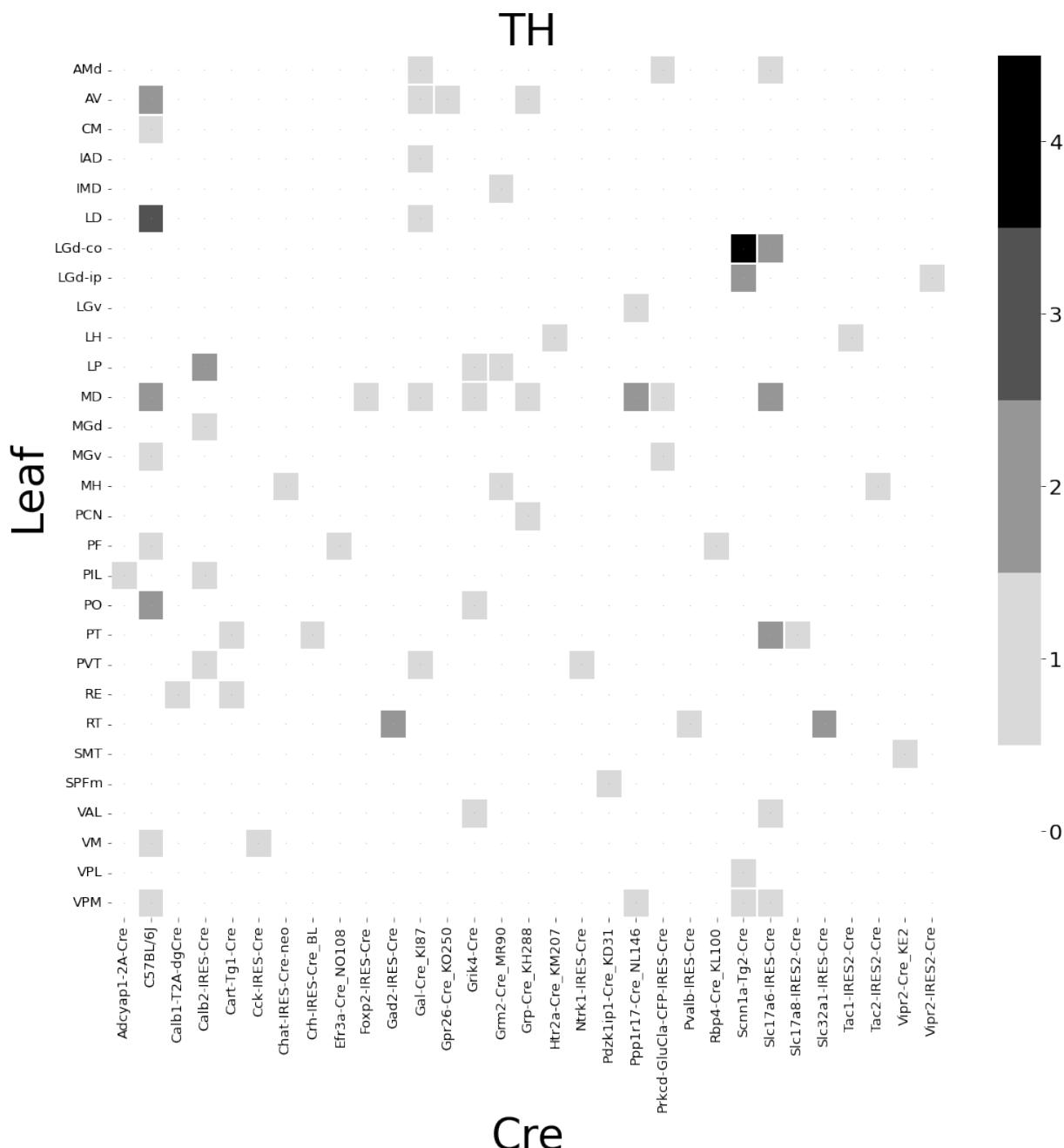


Figure 5: Caption

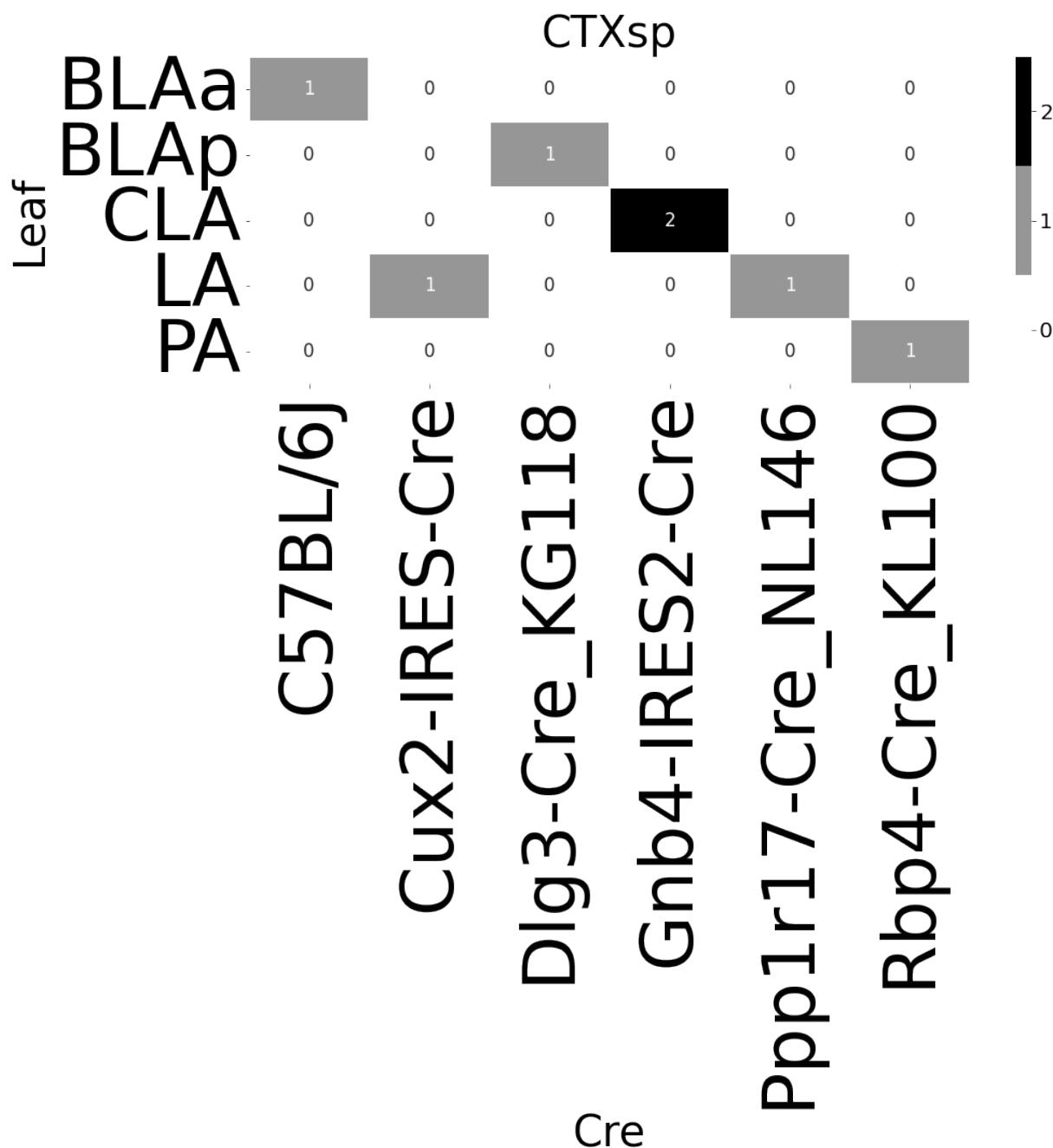
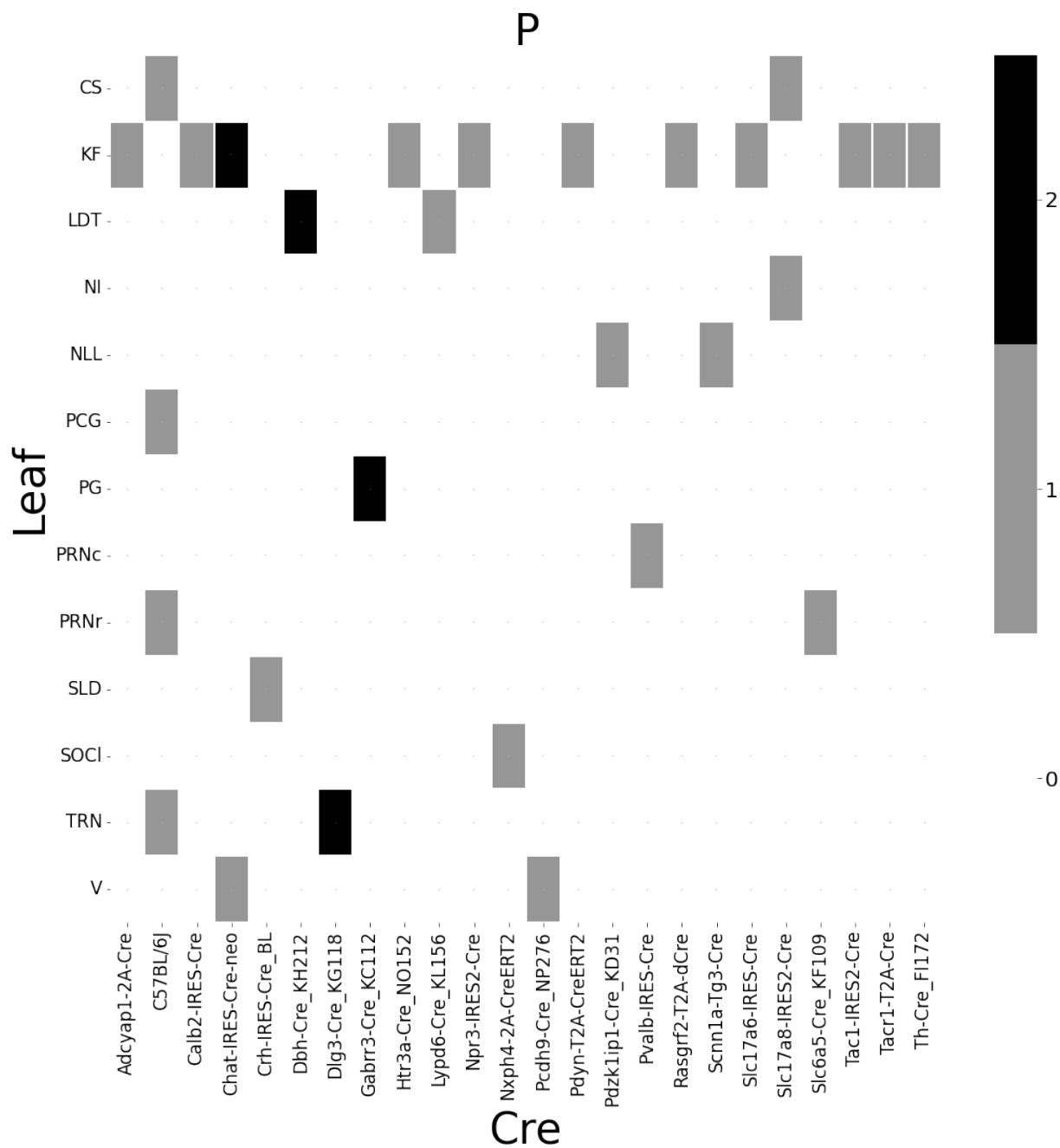
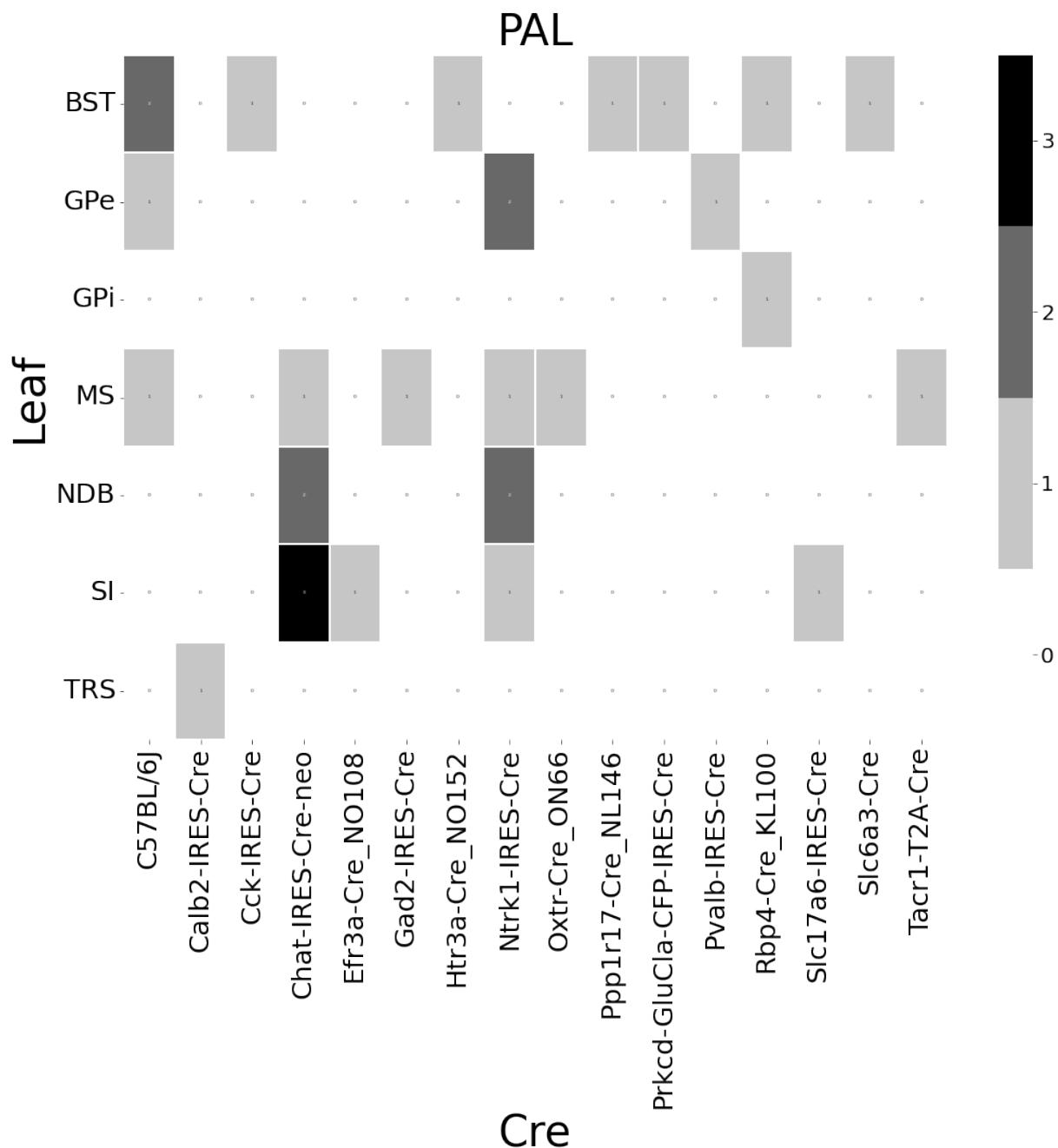
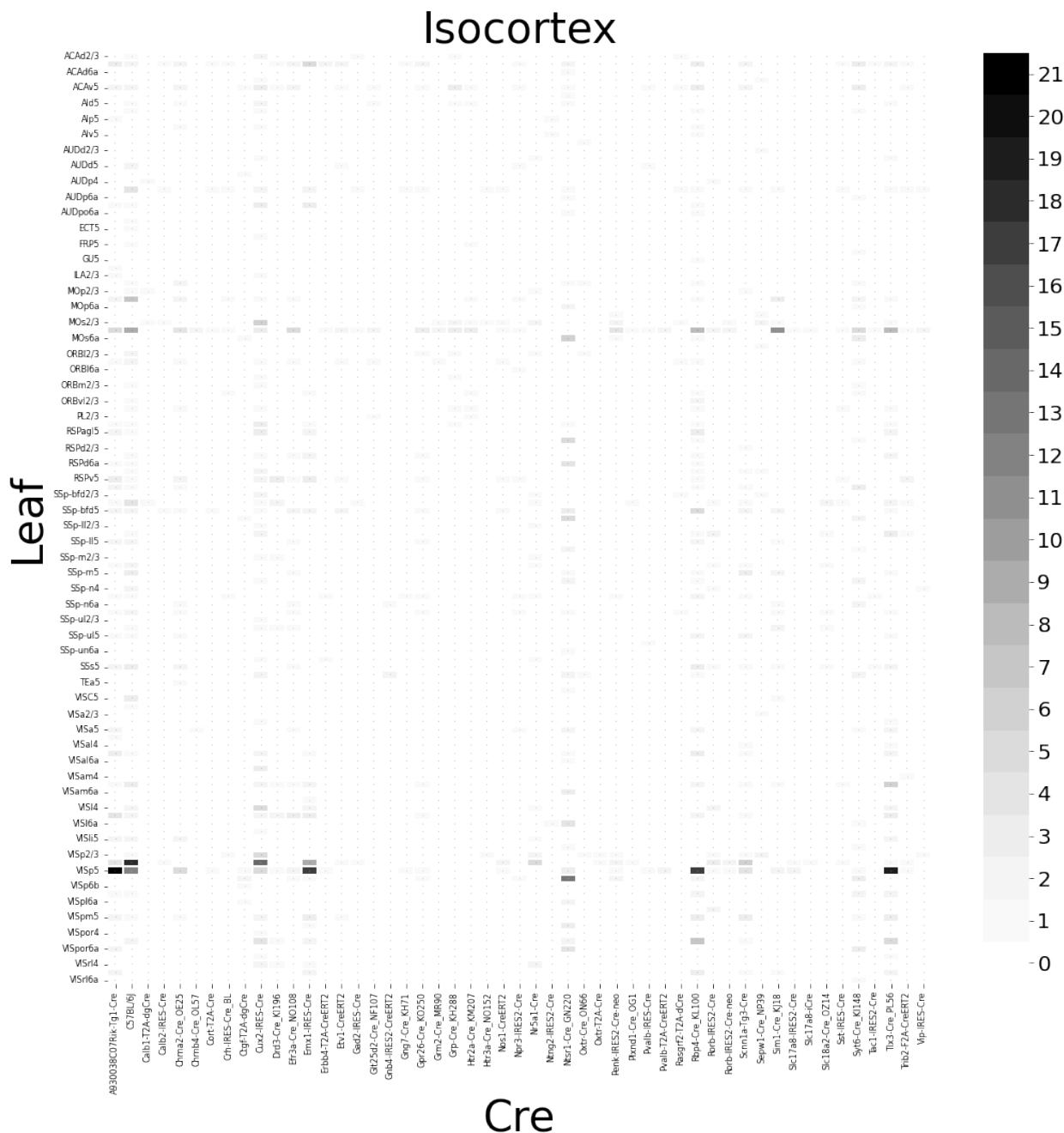
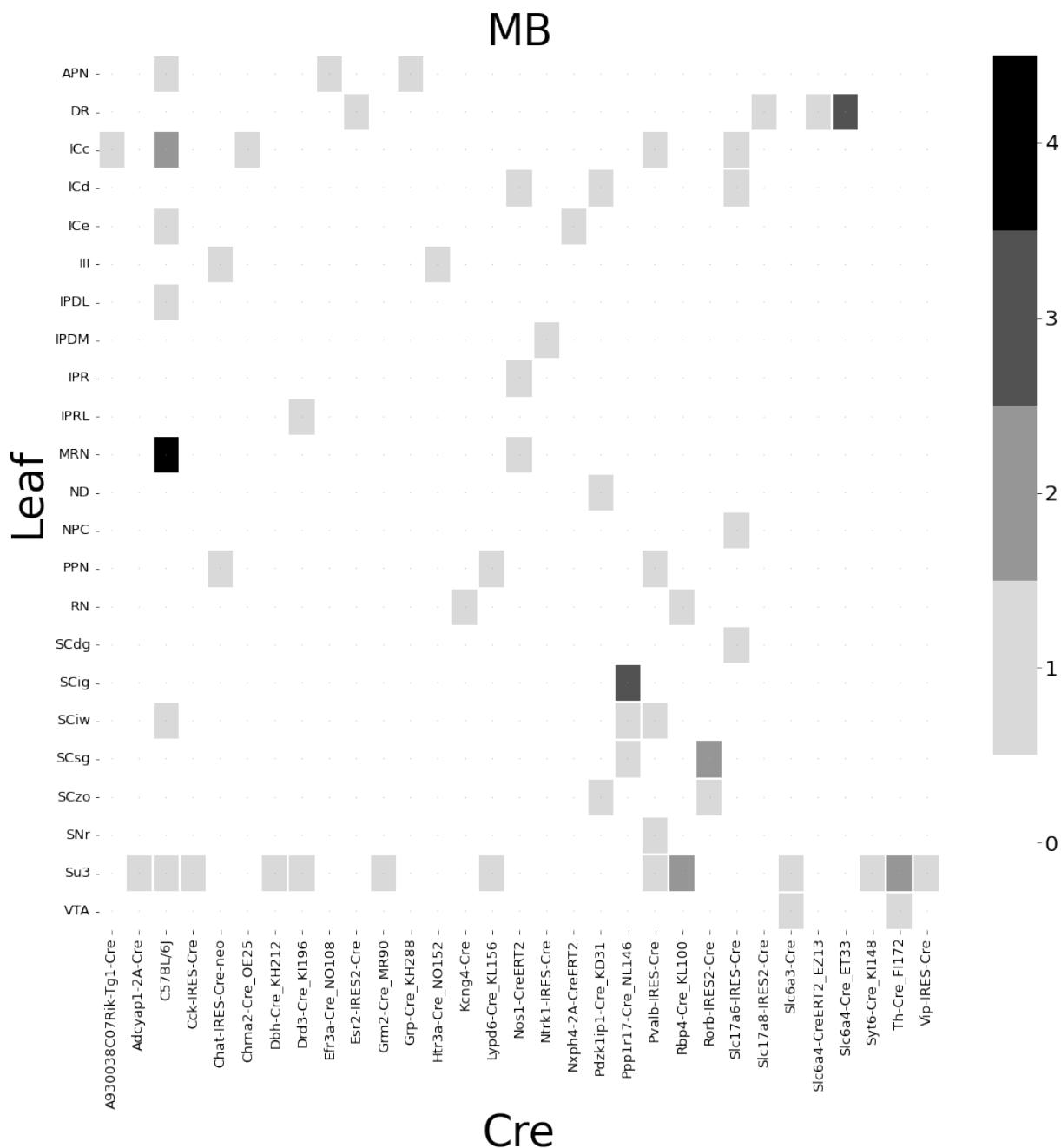


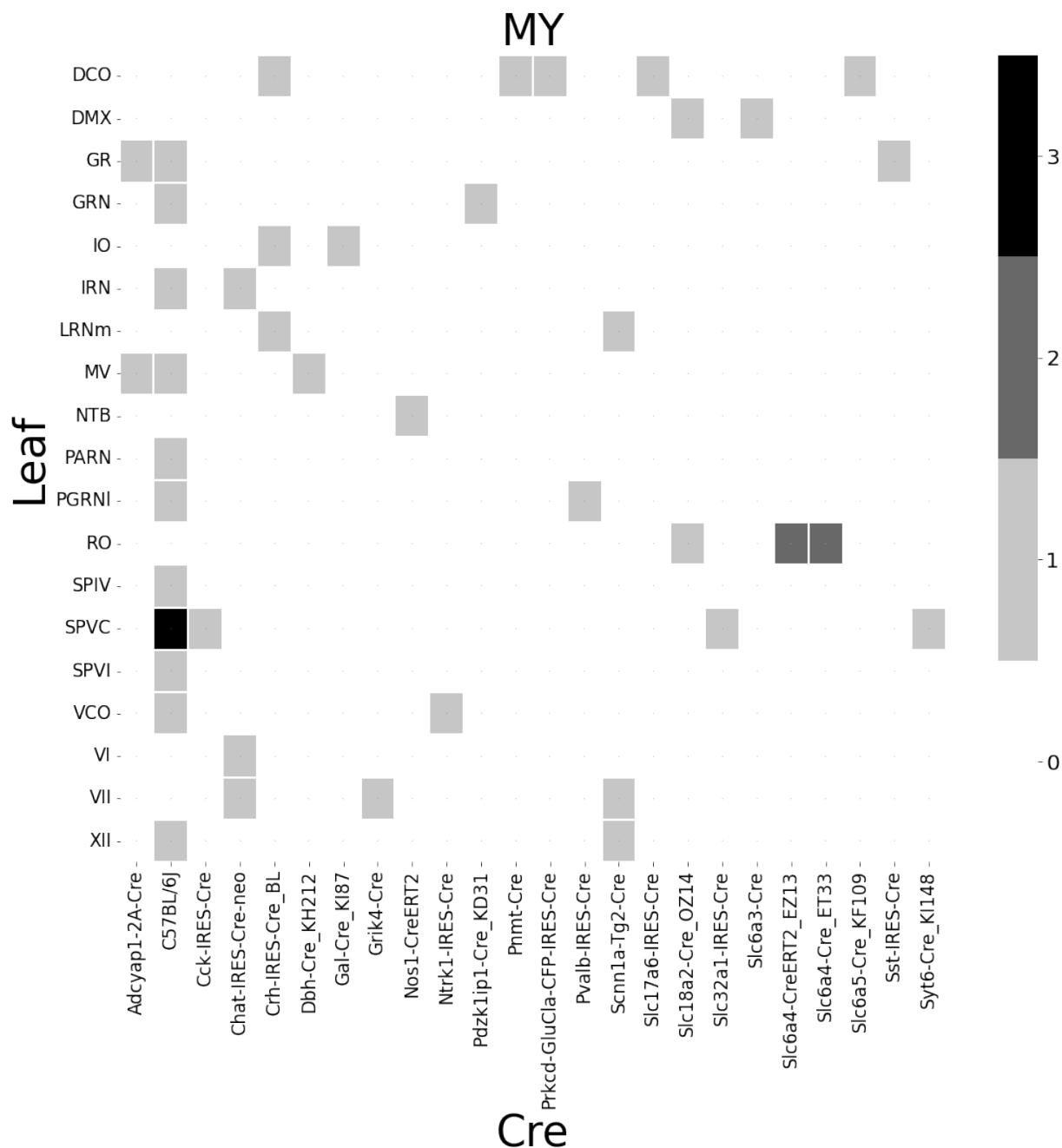
Figure 6: Caption

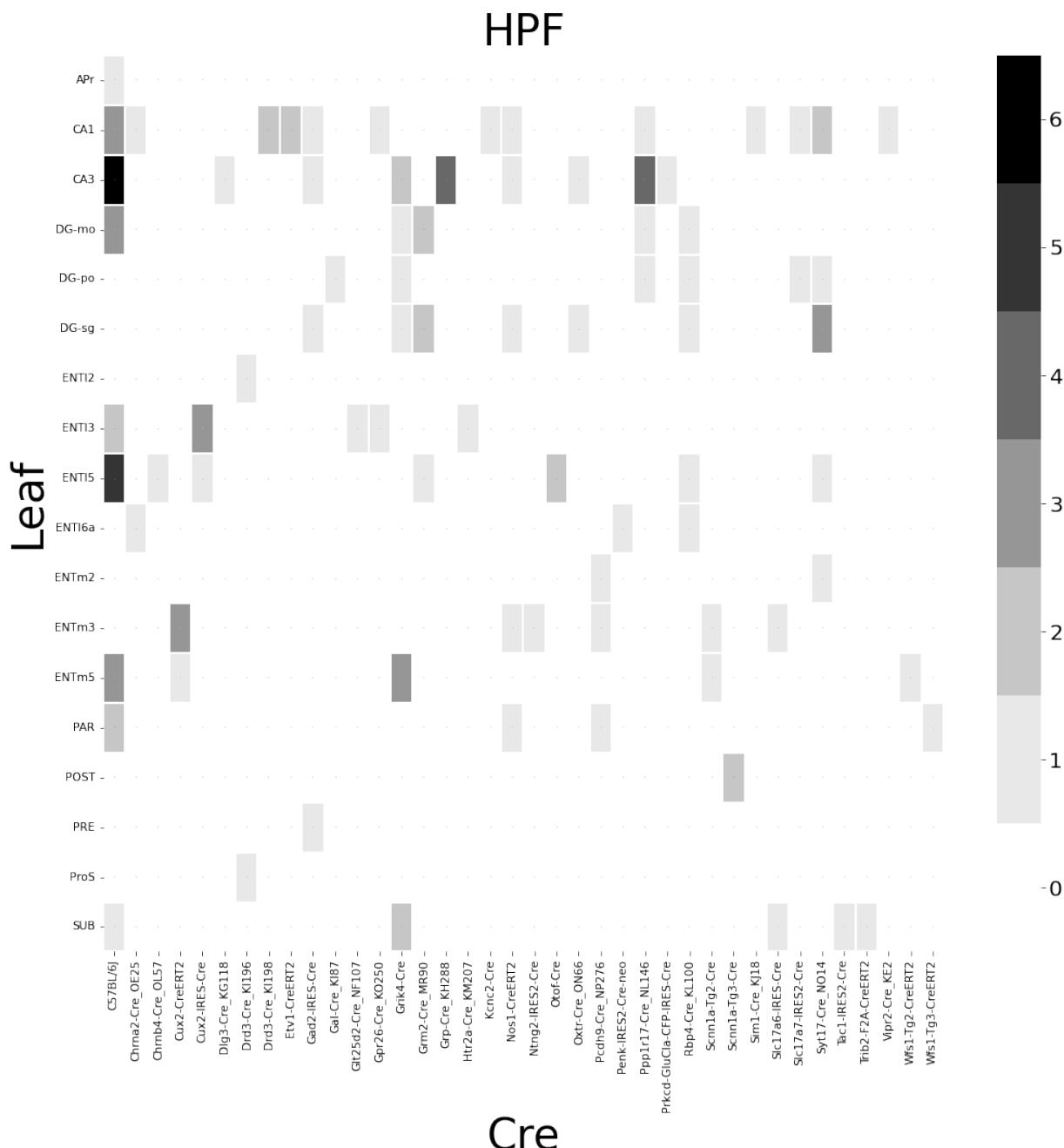


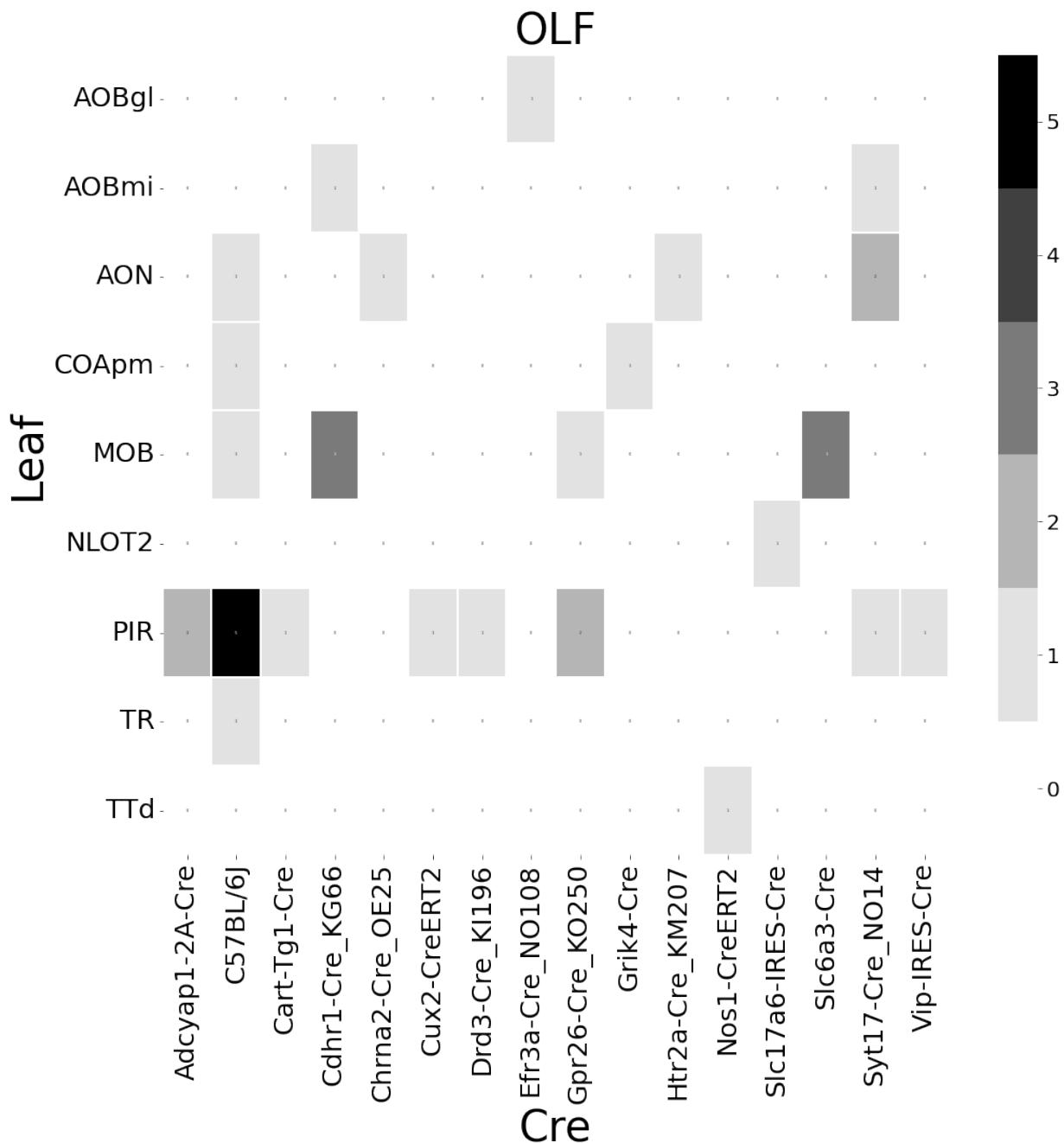


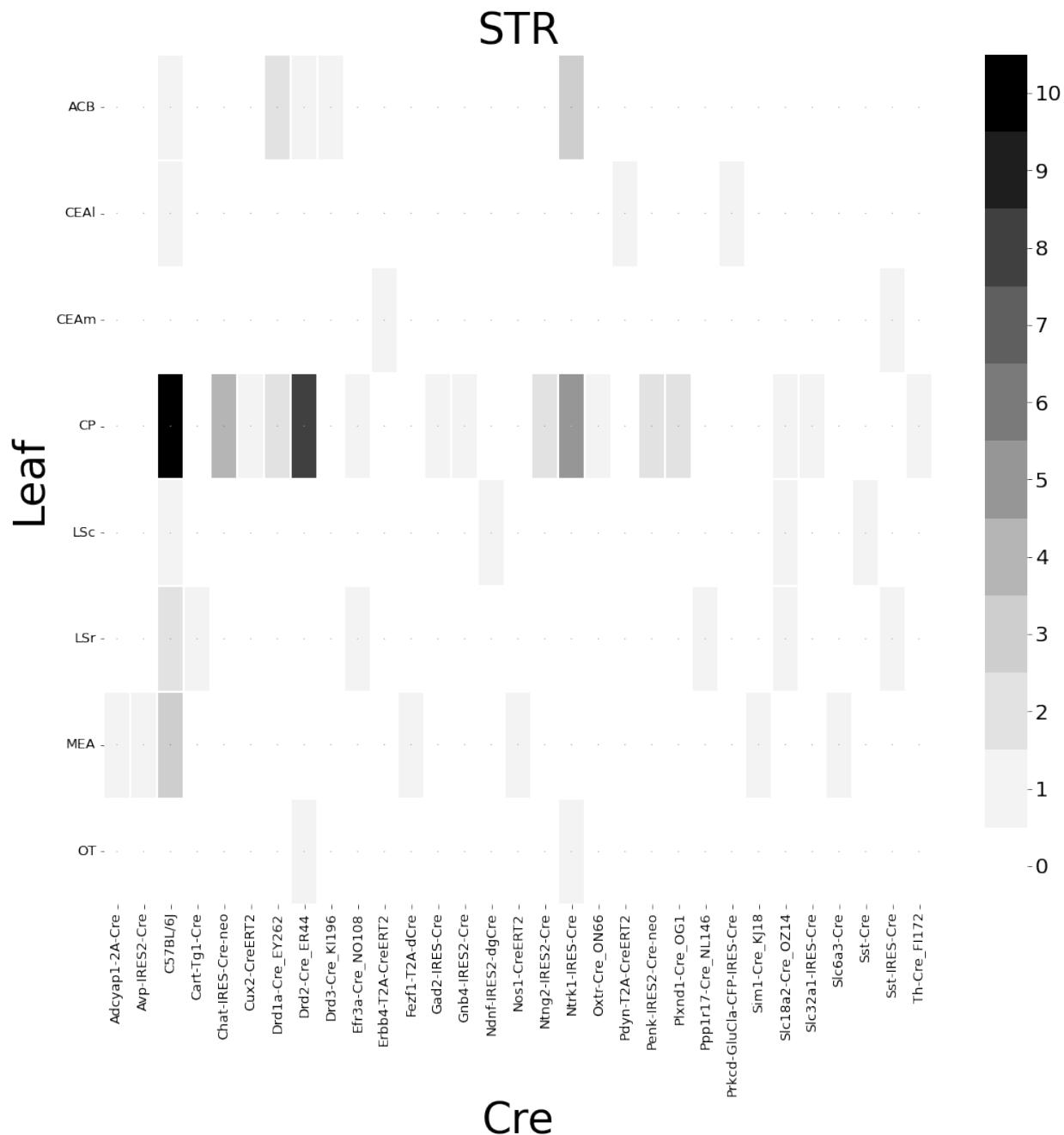


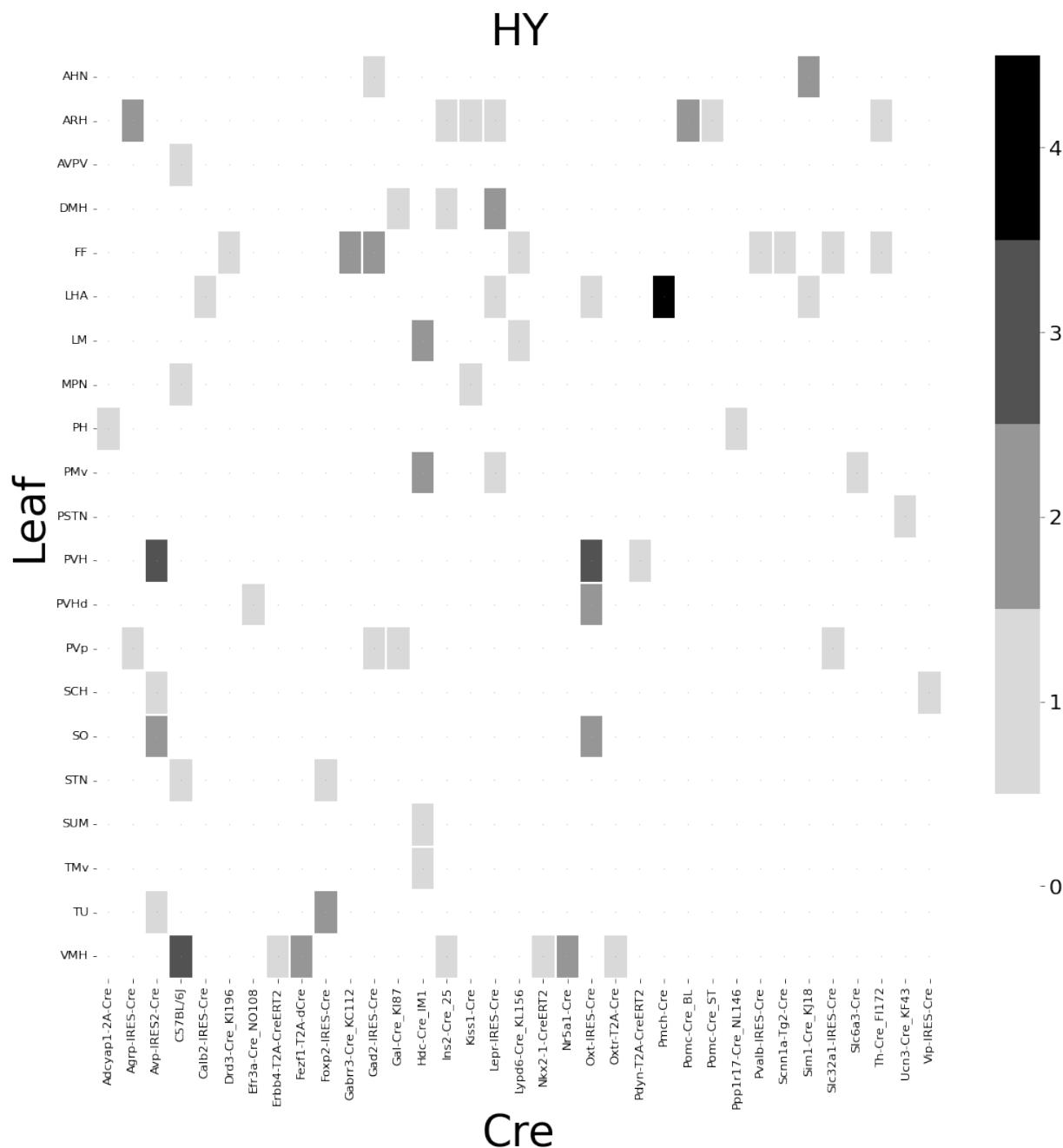












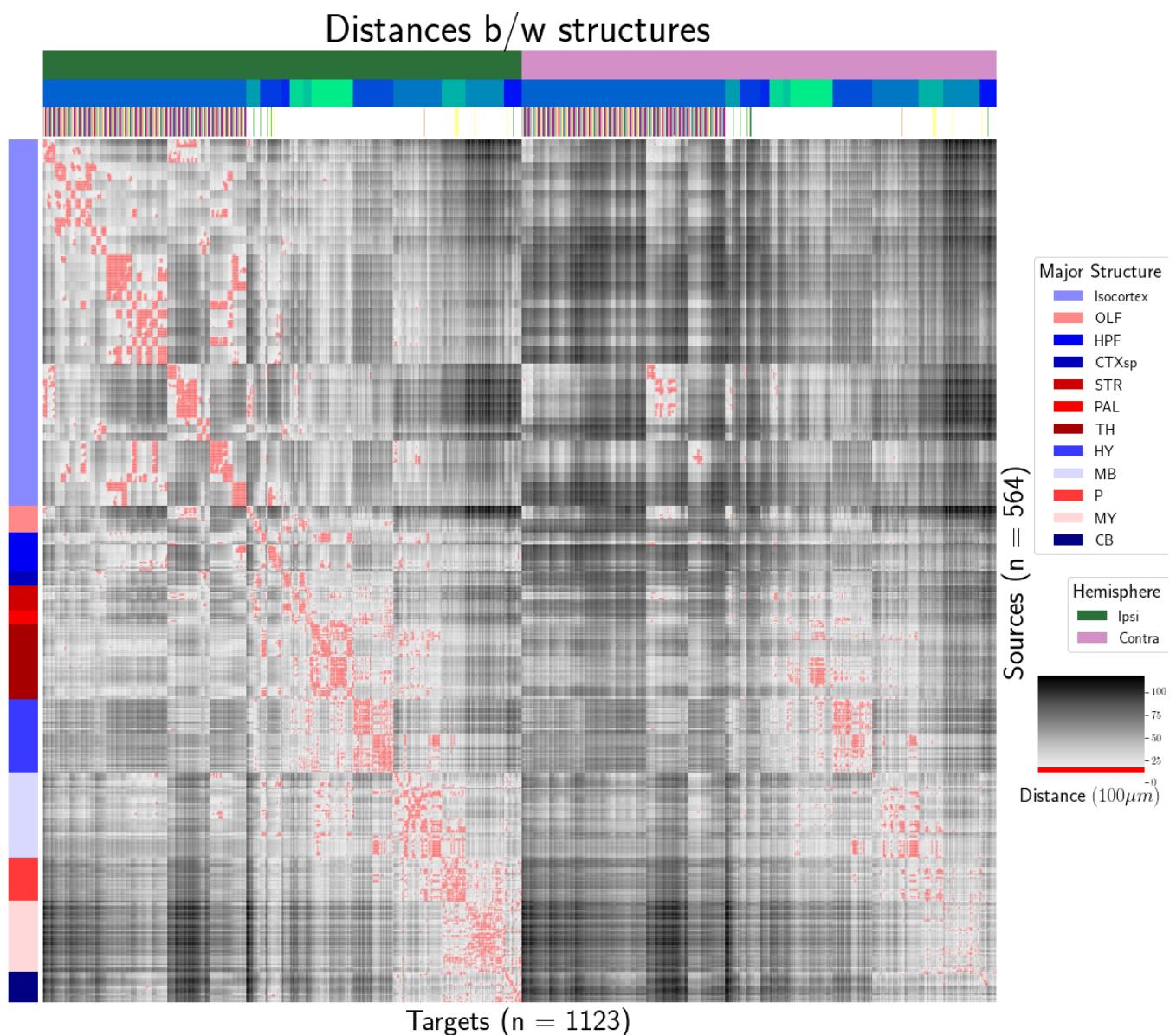
314 ***Distances between structures***

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

315 ***Model evaluation***

	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.

316 NUMBER OF EXPERIMENTS IN EVALUATION SETS

317 INJECTION-NORMALIZED LOSSES

\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	0.396
OLF	0.499	0.504	0.494	0.441	0.543	0.543	0.437
HPF	0.336	0.483	0.332	0.444	0.501	0.501	0.321
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	0.322
PAL	0.519	0.497	0.519	0.436	0.459	0.459	0.434
TH	0.769	0.767	0.769	0.514	0.539	0.539	0.556
HY	0.414	0.439	0.414	0.441	0.452	0.452	0.399
MB	0.459	0.396	0.397	0.358	0.324	0.324	0.403
P	0.562	0.562	0.562	0.758	0.764	0.764	0.562
MY	0.699	0.552	0.621	0.439	0.578	0.578	0.439
CB	0.849	0.689	0.849	0.500	0.615	0.615	0.495

Table 4: Injection-normalized losses.

6 SUPPLEMENTAL METHODS

³¹⁸ This section consists of additional information on preprocessing of the neural connectivity data,
³¹⁹ estimation of connectivity, and matrix factorization.

³²⁰ ***Data preprocessing***

³²¹ Several data preprocessing steps take place prior to evaluations of the connectivity matrices. These
³²² steps are described in Algorithm PREPROCESS. The arguments of this normalization process - injection
³²³ signals $x(i)$, projection signals $y(i)$, injection fraction $F(i)$, and data quality mask $q(i)$ - were
³²⁴ downloaded using the Allen SDK. The injections and projection signals $\mathcal{B} \rightarrow [0, 1]$ were segmented
³²⁵ manually in histological analysis. The projection signal gives the proportion of pixels within the voxel
³²⁶ displaying fluorescence, and the injection signal gives the proportion of pixels within the
³²⁷ histologically-selected injection subset displaying fluorescence. The injection fraction $\mathcal{B} \rightarrow [0, 1]$ gives
³²⁸ the proportion of pixels within each voxel in the injection subset. Finally, the data quality mask
³²⁹ $\mathcal{B} \rightarrow \{0, 1\}$ gives the voxels that have valid data.

³³⁰ Our preprocessing makes use of the above ingredients, as well as several other essential steps. First,
³³¹ we compute the weighted injection centroid

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

³³² where $x(i)(l)$ is the injection density at location $l \in \mathbb{R}^3$. Given a regionalization \mathcal{R} from the Allen SDK,
³³³ we can also access regionalization map $R: \mathcal{B} \rightarrow \mathcal{R}$. This induces a functional of connectivities from
³³⁴ the space of maps $\{\mathcal{X} = x: \mathcal{B} \rightarrow [0, 1]$

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

³³⁵ We also can restrict a signal to a individual structure as

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

³³⁶ Finally, given a vector or array $a \in \mathbb{R}^T$, we have the $L1$ normalization map

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

³³⁷ We define these objects as functions and functionals, but this is for notational convenience and
³³⁸ non-essential. A function $x(i) : \mathcal{B} \rightarrow [0, 1]$ is mathematically equivalent to the graph
³³⁹ $\mathcal{G}(x(i)) \in \mathcal{B} \times [0, 1]$. As an abuse of notation, we define $x \odot x' := z$ such that $z(l) = x(l)x'(l)$ for all $l \in \mathcal{B}$.
³⁴⁰ Also, denote $m(i)$ as the major structure containing experiment i . We then can write the
³⁴¹ 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 $\tilde{y}_{\mathcal{T}}(i), c(i)$

342 **Estimators**

343 As mentioned previously, we can consider our estimators as modelling a connectivity vector
 344 $f_{\mathcal{T}}(\nu, s) \in \mathbb{R}_{\geq 0}^T$. Thus, for the remainder of this section, we will discuss only $f(\nu, s)$. We review the
 345 Nadaraya-Watson estimator from Knox et al. (2019), and describe its conversion into our cell-class
 346 specific Expected Loss estimator.

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

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

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

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

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

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

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

355 Several facets of the estimator are visible here. A smaller γ corresponds to a greater amount of
 356 smoothing, and the index set $I \subseteq \{1 : n\}$ generally depends on s and ν . Fitting γ via empirical risk
 357 minimization therefore bridges between 1-nearest neighbor prediction and averaging of all
 358 experiments in I . In Knox et al. (2019), I consisted of experiments sharing the same brain division, i.e.
 359 $I = I_m$, while restricting of index set to only include experiments with the same cell class gives the
 360 class-specific Cre-NW model. Despite this restriction, we fit γ for each m rather than a smaller subset
 361 like s or ν . That is,

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

³⁶² *The Expected-Loss estimator* Besides the injection location, the targeted cell class also influences
³⁶³ projection. Since Cre-lines that target similar classes are induce similar projections, and including
³⁶⁴ similar Cre-lines in the Nadaraya-Watson estimator increases effective sample size, we introduce an
³⁶⁵ estimator that assigns a predictive weight to each training point that depends both on its
³⁶⁶ centroid-distance and Cre-line. This weight is determined by the expected prediction error of each of
³⁶⁷ the two feature types, as determined by cross-validation. For this reason, we call this the Expected
³⁶⁸ Loss Estimator. The resulting weights are then utilized in a Nadaraya-Watson estimator in a final
³⁶⁹ prediction step.

³⁷⁰ We formalize Cre-line behavior as the average regionalized projection of a Cre-line in a given
³⁷¹ structure (i.e. leaf). This vectorization of categorical information is known as **target encoding**, and we
³⁷² 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
³⁷³ between the target-encoded projections of two Cre-lines. The relative predictive accuracy of
³⁷⁴ Cre-distance and centroid distance is determined by fitting a surface of projection distance as a
³⁷⁵ function of Cre-distance and centroid distance.

³⁷⁶ In mathematical terms, our full feature set consists of the centroid coordinates and the
³⁷⁷ target-encoded means of the combinations of virus type and injection-centroid structure. That is,

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

³⁷⁸ f^* is defined as in (2). The expected loss estimator is then

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

³⁷⁹ where

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

³⁸⁰ and s is the structure containing l .

³⁸¹ The key step therefore is finding a suitable g with which to weight the positional and Cre
³⁸² information. Note that g must be a concave, non-decreasing function of its arguments with with
³⁸³ $g(0, 0) = 0$, then g defines a metric on the product of the metric spaces defined by experiment centroid
³⁸⁴ and target-encoded cre-line, and \hat{f}_{EL} is a Nadaraya-Watson estimator. A derivation of this fact is given

³⁸⁵ later in this section, and we therefore use shape-constrained B-splines to estimate g . Similarly to the
³⁸⁶ Nadaraya-Watson model, we make the decision to fit a g separately for each major brain division. We
³⁸⁷ can then select $\hat{\gamma}$ as in 2.

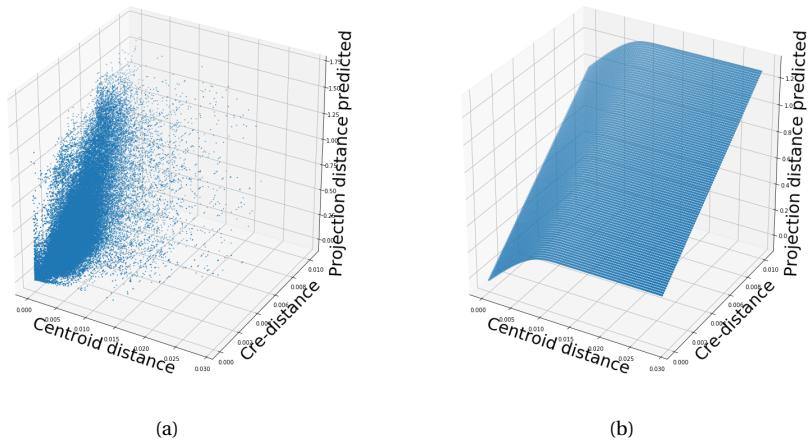


Figure 8: Fitting g . 8a Distribution of projection errors against centroid distance and cre-distance in Isocortex. 8b \hat{g} found using B-splines.

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

393 **Proposition 1.** *Given a collection of metric spaces X_1, \dots, X_n with metrics d_1, \dots, d_n (e.g. $d_{centroid}, d_{cre}$),
 394 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
 395 concave, non-decreasing and $g(d) = 0 \iff d = 0$.*

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

- 397 ▪ The first property of a metric is that $f(x, x') = 0 \iff x = x'$. The left implication:
 398 $x = x' \implies f(x_1, x'_1, \dots, x_n, x'_n) = g(0, \dots, 0)$, since d are metrics. Then, since $g(0) = 0$, we have that
 399 $f(x, x') = 0$. The right implication: $f(x, x') = 0 \implies d = 0 \implies x = x'$ since d are metrics.
- 400 ▪ The second property of a metric is that $f(x, x') = f(x', x)$. This follows immediately from the
 401 symmetry of the d_i , i.e. $f(x, x') = f(x_1, x'_1, \dots, x_n, x'_n) = g(d_1(x_1, x'_1), \dots, d_n(x_n, x'_n)) =$
 402 $g(d_1(x'_1, x_1), \dots, d_n(x'_n, x_n)) = f(x'_1, x_1, \dots, x'_n, x_n) = f(x', x)$.
- 403 ▪ The third property of a metric is the triangle inequality: $f(x, x') \leq f(x, x^*) + f(x^*, x')$. To show this
 404 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
 405 non-decreasing and by the triangle inequality of d . Then, since g is concave,
 406 $g(d(x, x^*) + d(x^*, x')) \leq g(d(x, x^*)) + g(d(x^*, x')) = f(x, x^*) + f(x^*, x')$.

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

- 408 ▪ The first property is that $g(d) = 0 \iff d = 0$. We first show the right implication: $g(d) = 0$, and
 409 $g(d) = f(x, x')$, so $x = x'$ (since f is a metric), so $d = 0$. We then show the left implication:
 410 $d = 0 \implies x = x'$, since d is a metric, so $f(x, x') = 0$, since f is a metric, and thus $g(d) = 0$.
- 411 ▪ The second property is that g is non-decreasing. We proceed by contradiction. Suppose g is
 412 decreasing in argument d_1 in some region $[l, u]$ with $0 < l < u$. Then
 413 $g(d_1(0, l), 0) \geq g(d_1(0, 0), 0) + g(d_1(0, u), 0) = g(d_1(0, u), 0)$, which violates the triangle inequality on
 414 f . Thus, decreasing g means that f is not a metric, so f a metric implies non-decreasing g .

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

419

□

420 Establishing a lower detection limit

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

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

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

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

432 ***Decomposing the connectivity matrix***

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

438 *Non-negative matrix factorization* As discussed in Knox et al. (2019), one of the most basic processes
 439 underlying the observed connectivity is the tendency of each source region to predominantly project
 440 to proximal regions. For example, the heatmap in Supplemental Figure 7 shows infrastructure
 441 distances clearly contains a diagonal pattern resembling the connectivity matrix in 2. These
 442 connections are biologically meaningful, but also unsurprising, and their relative strength biases
 443 learned latent coordinate representations away from long-range structures. For this reason, we
 444 establish a $1500\mu m$ 'distal' threshold within which to exclude connections for our analysis.

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

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

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

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

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

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

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

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

464 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

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

465 *Stabilizing NMF* The NMF program is non-convex, and, empirically, individual replicates will not
 466 converge to the same optima. One solution therefore is to run multiple replicates of the NMF
 467 algorithm and cluster the resulting vectors. This approach raises the questions of how many clusters
 468 to use, and how to deal with stochasticity in the clustering algorithm itself. We address this issue
 469 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

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

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

7 SUPPLEMENTAL EXPERIMENTS

479 *Establishing a lower limit of detection*

480 We give results on the false detection rate at different limits of detection. These conclusively show that
 481 10^{-6} is the good threshold for our normalized data.

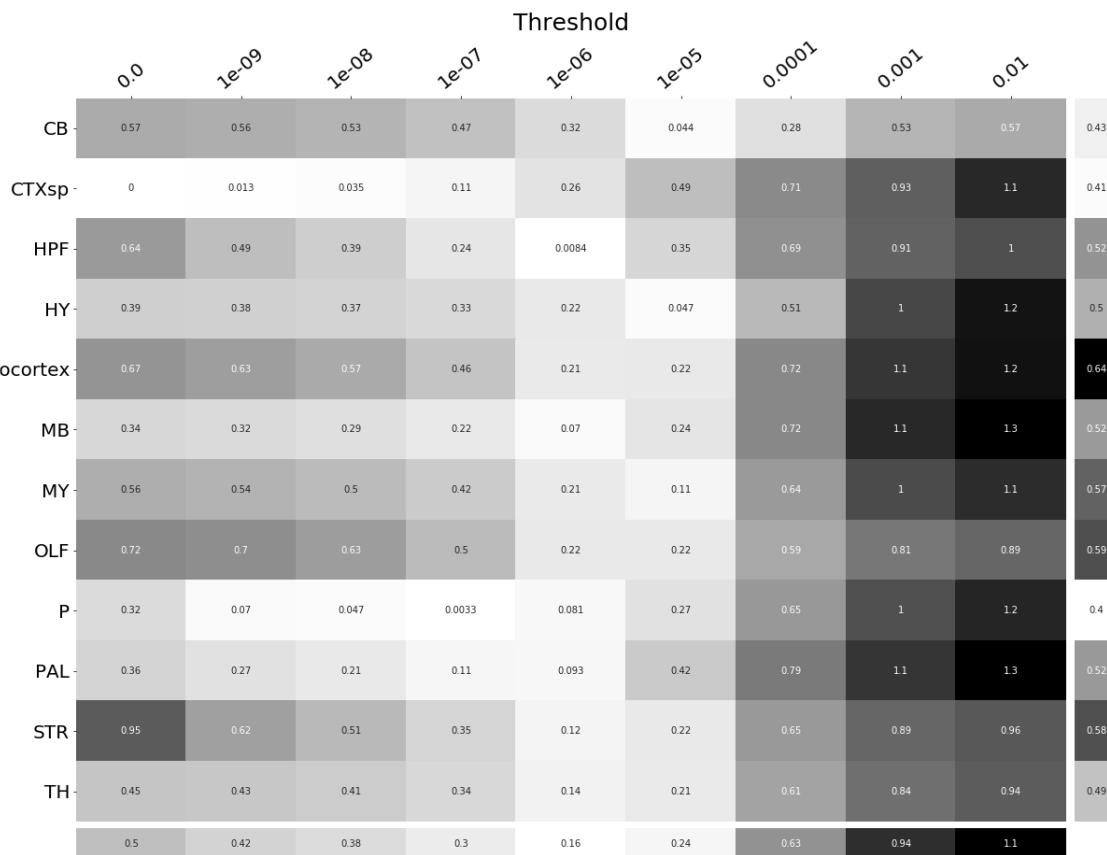


Figure 9: τ at different limits of detection in different major structures. 10^{-6} is clearly the optimal detection threshold.

482 **Loss subsets**

483 We report model accuracies for our *EL* model by neuron class and structure. These expand upon the
484 results in Table 2 and give more specific information about the quality of our estimates.



Figure 10

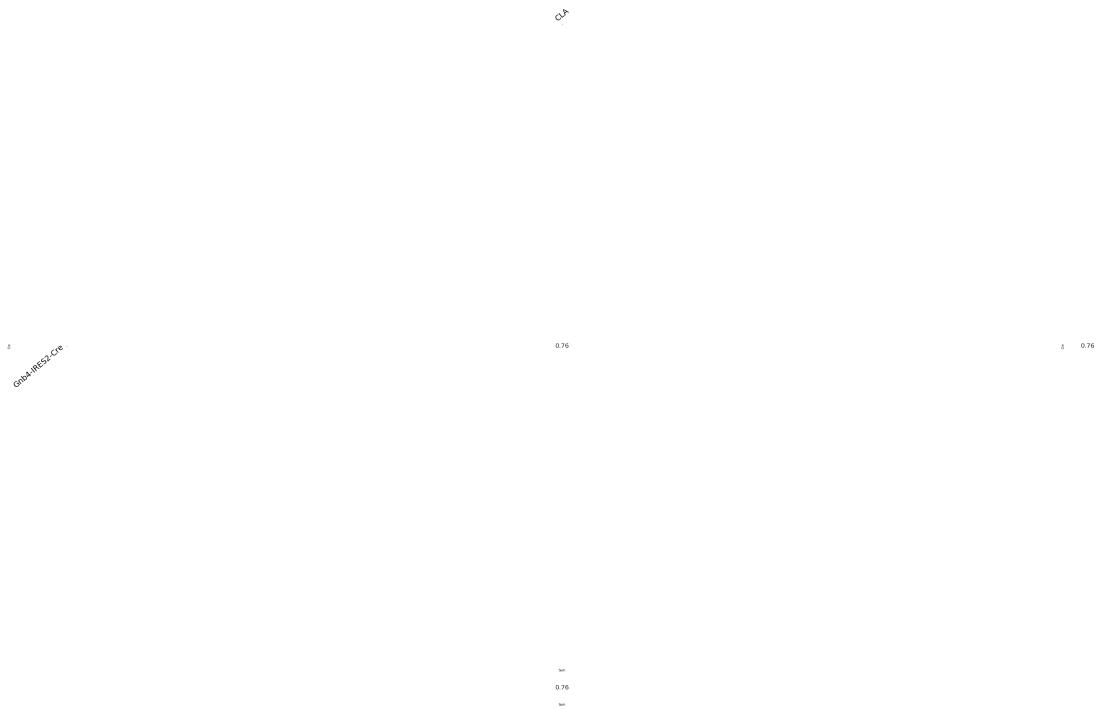


Figure 11

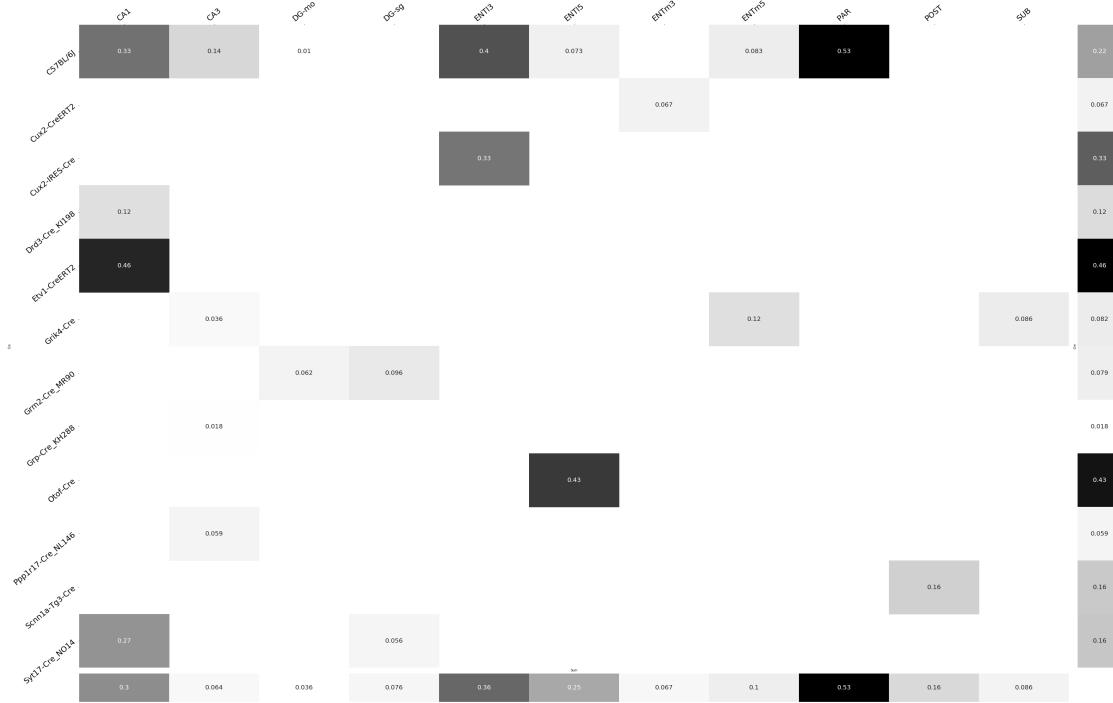


Figure 12

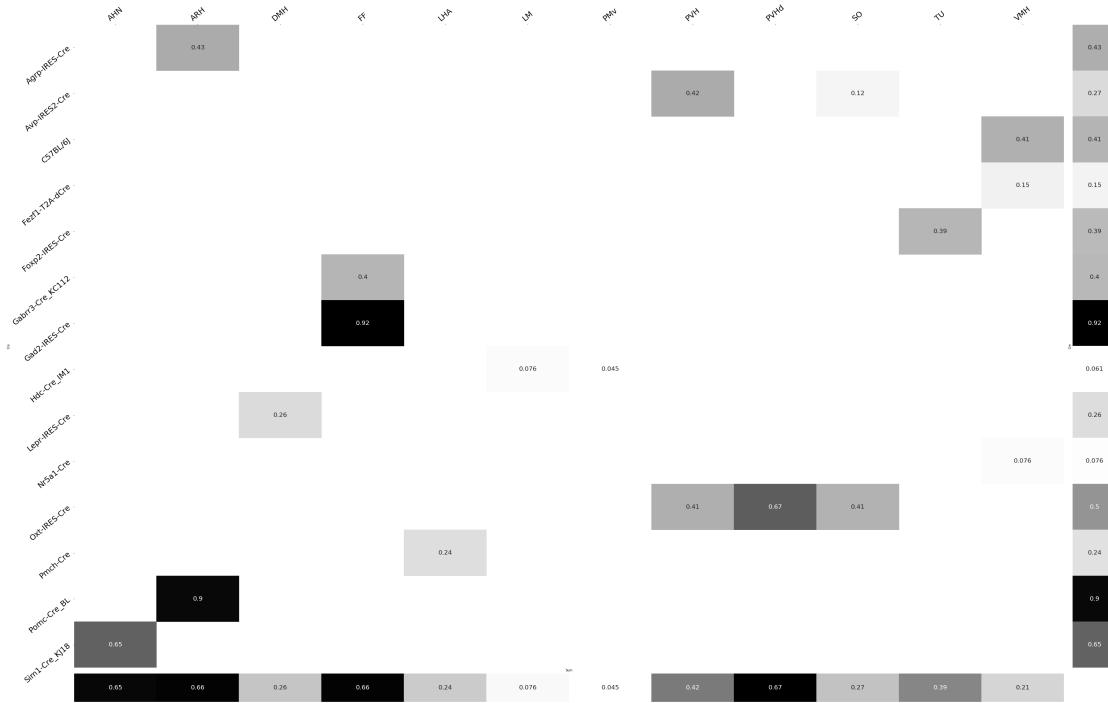


Figure 13

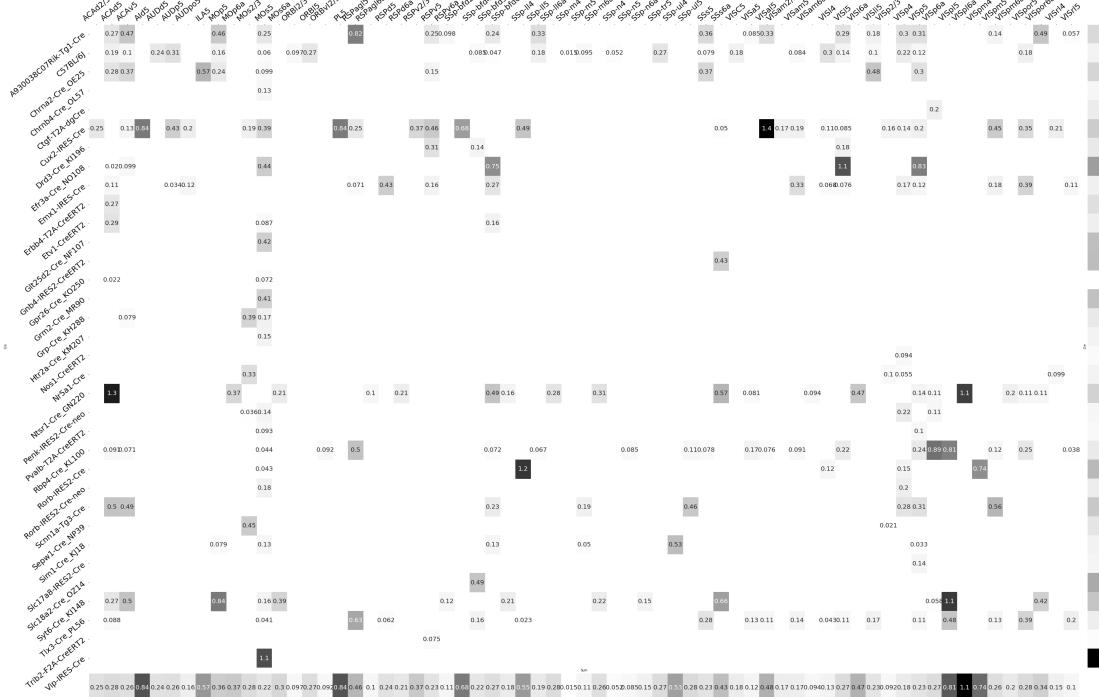


Figure 14



Figure 15



Figure 16



Figure 17

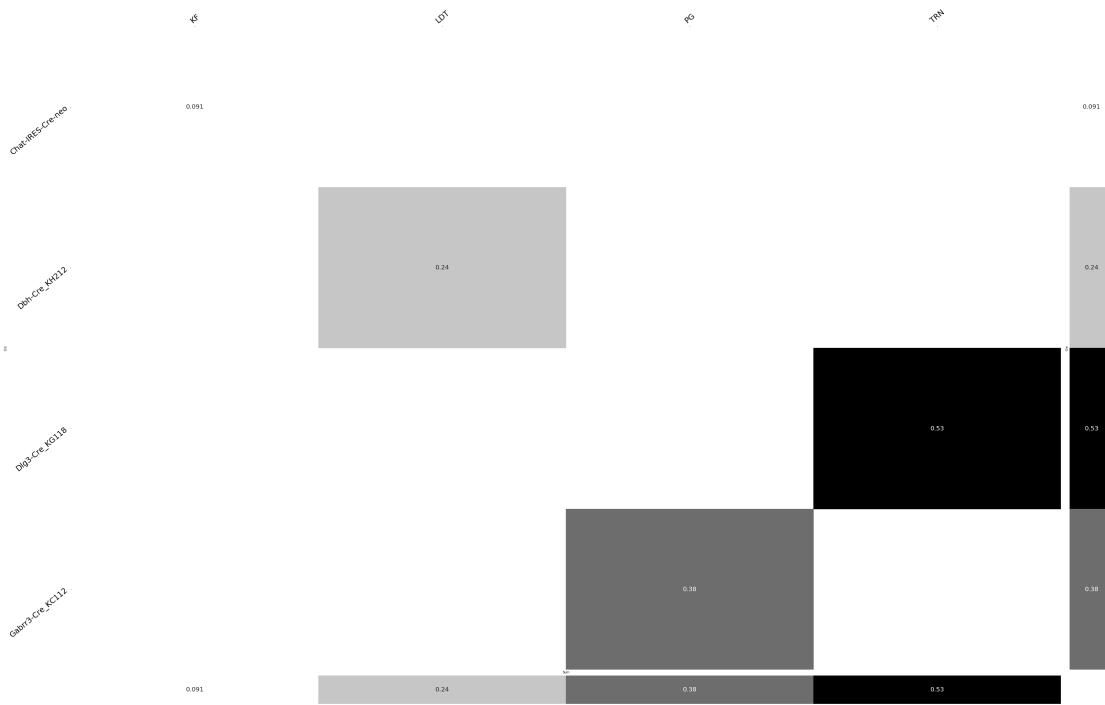


Figure 18

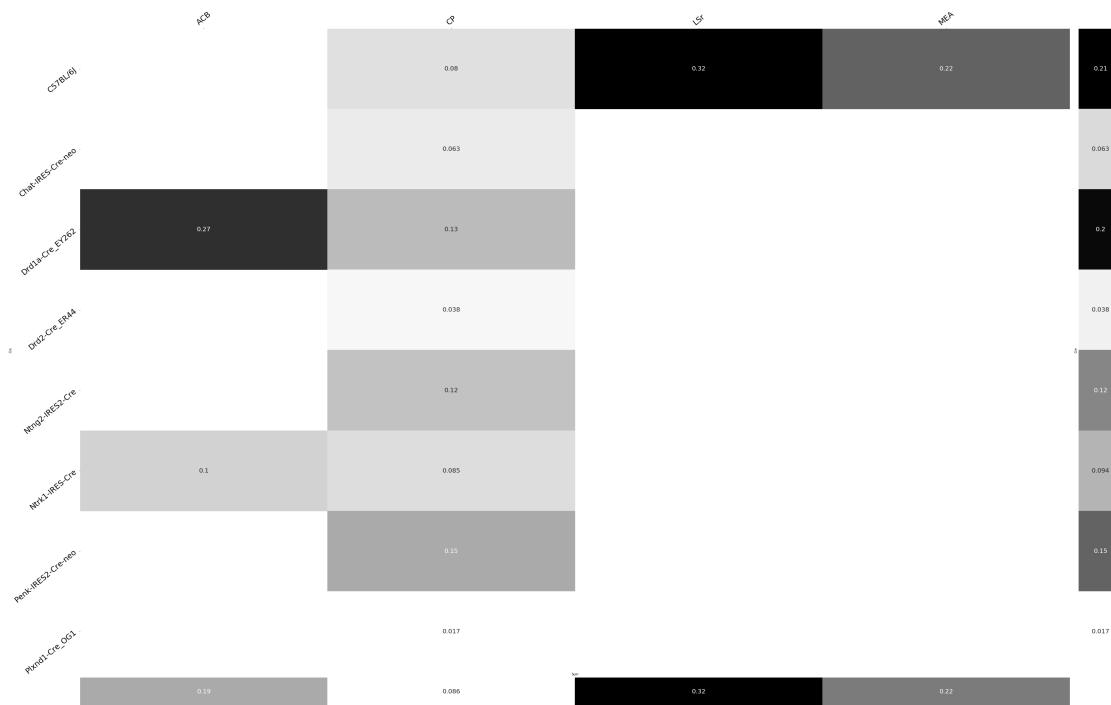


Figure 19

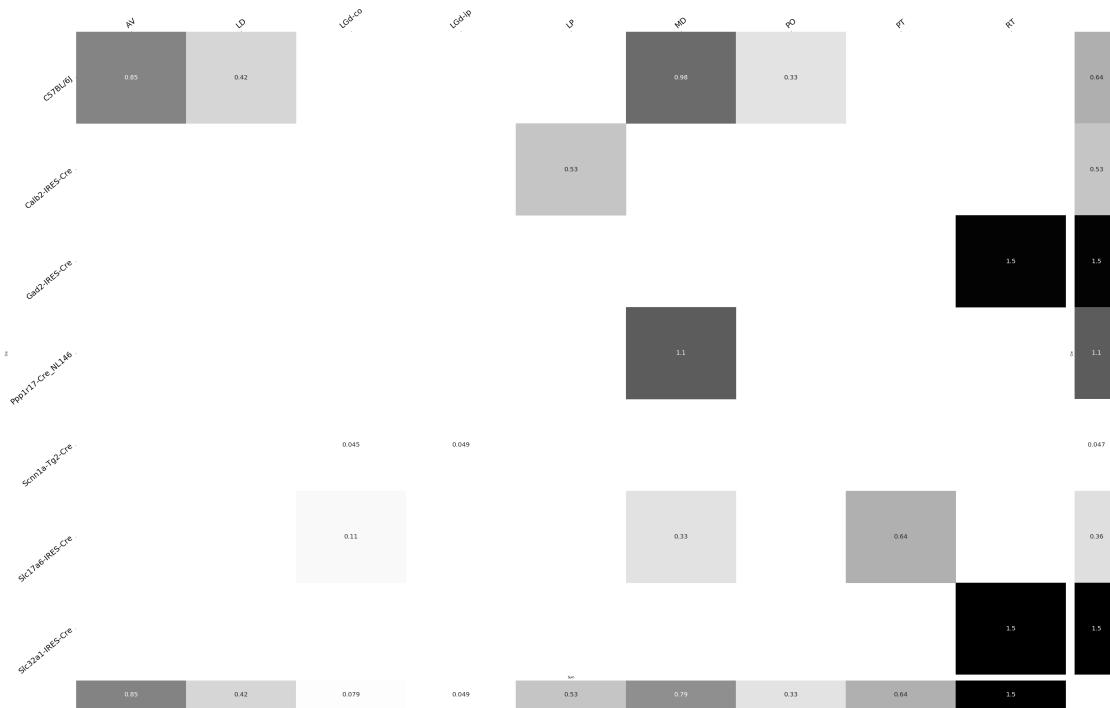


Figure 20

485 **Matrix Factorization**

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

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

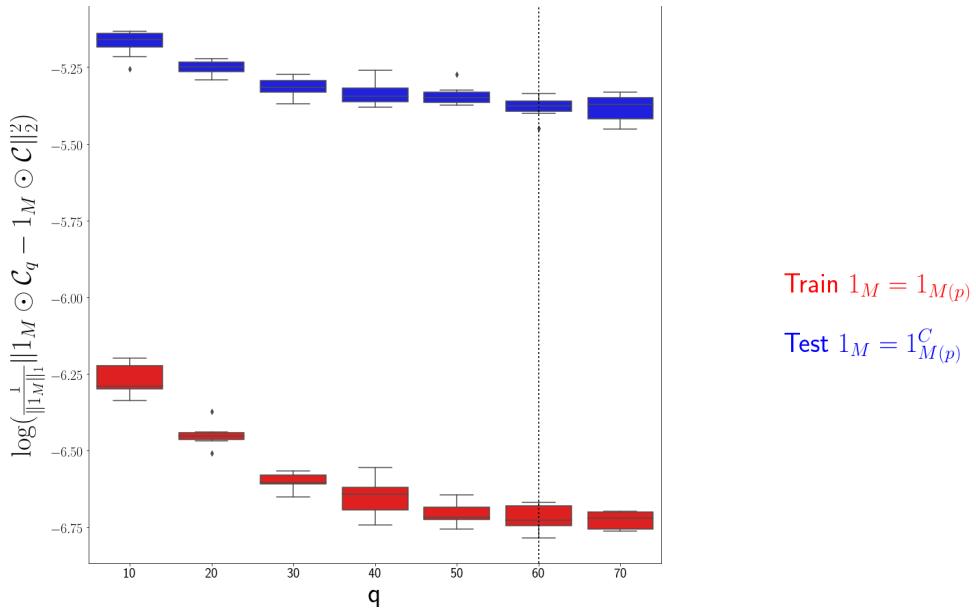


Figure 21: Train and test error using NMF decomposition.

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

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

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

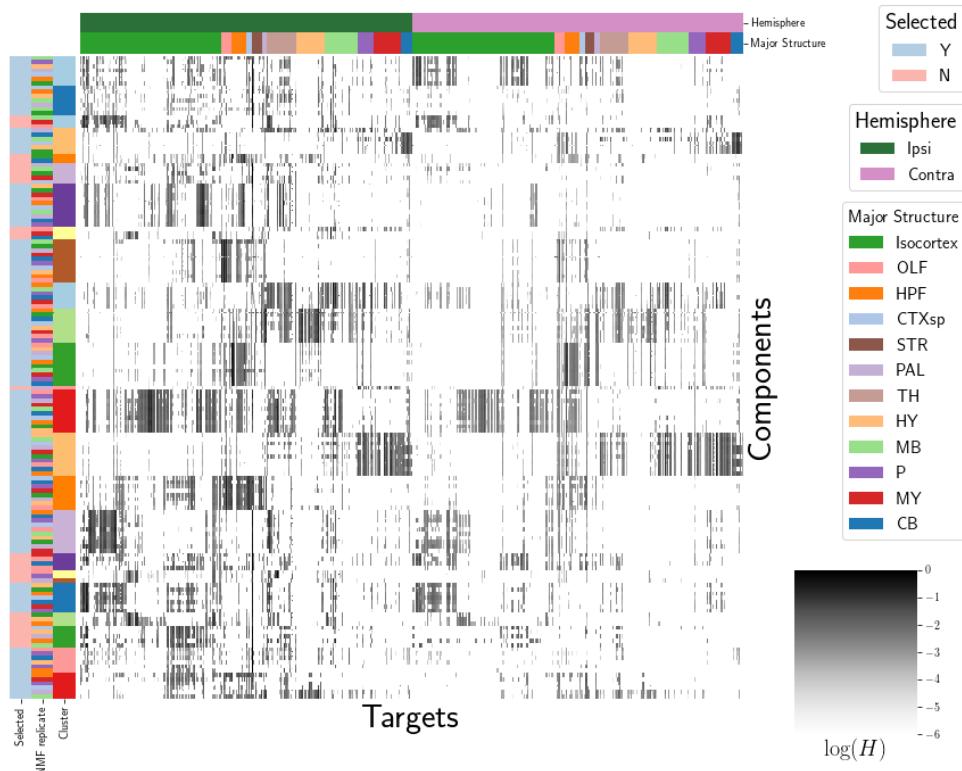


Figure 22: Stability of NMF results across replicates. Replicate and NMF component are shown on rows. Components that are in the top 15 are also indicated.

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

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

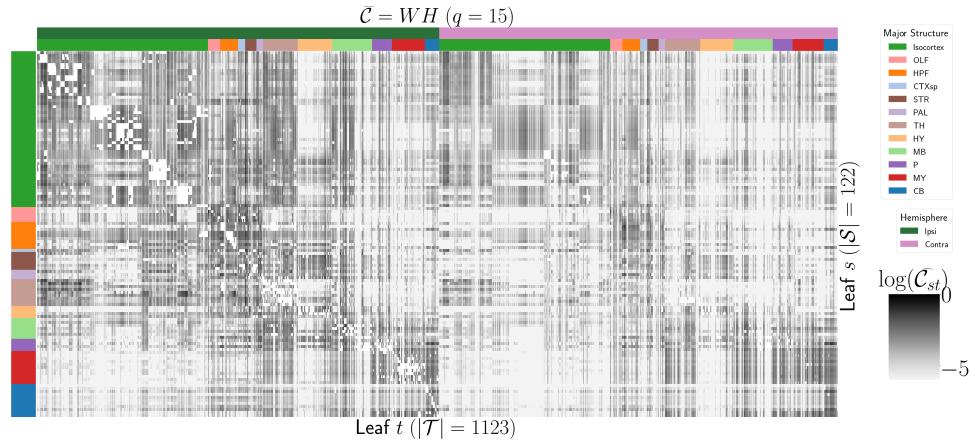


Figure 23: Reconstruction of \mathcal{C} from H and W with $q = 15$ in Figure 4.

8 GLOSSARY OF SYMBOLS

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

9 COMPETING INTERESTS

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

505

506 REFERENCES

507

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

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

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

514 Gao, Y., Zhang, X., Wang, S., & Zou, G. (2016). Model averaging based on leave-subject-out cross-validation. *J. Econom.*,
515 192(1), 139–151.

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

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

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

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

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

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

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

- 530 Oh, S. W., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., ... Zeng, H. (2014). A mesoscale connectome of the
531 mouse brain. *Nature*, 508(7495), 207–214.
- 532 Saul, L. K., & Roweis, S. T. (2003). Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *J.
533 Mach. Learn. Res.*, 4(Jun), 119–155.
- 534 von Luxburg, U. (2010a). Clustering stability: An overview.
- 535 von Luxburg, U. (2010b). Clustering stability: An overview.
- 536 Wu, S., Joseph, A., Hammonds, A. S., Celtniker, S. E., Yu, B., & Frise, E. (2016). Stability-driven nonnegative matrix
537 factorization to interpret spatial gene expression and build local gene networks. *Proc. Natl. Acad. Sci. U. S. A.*, 113(16),
538 4290–4295.

10 TECHNICAL TERMS

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

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

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

544 **structural connectivities** connectivity between structures

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

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

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

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

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

551 **connectivity archetypes** Typical connectivity patterns

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