

<sup>1</sup> RESEARCH

<sup>2</sup> **Modeling the cell-type specific murine connectome**

<sup>3</sup> **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>**

<sup>4</sup> <sup>1</sup>Allen Institute for Brain Science, Seattle, WA, USA

<sup>5</sup> <sup>2</sup>Department of Statistics, University of Washington, Seattle, WA, USA

<sup>6</sup> **Keywords:** [a series of capitalized words, separated with commas]

## ABSTRACT

<sup>7</sup> The Allen Brain Connectivity Atlas consists of thousands of labelling experiments targeting  
<sup>8</sup> interrogating diverse structures and classes of projecting neurons. This paper describes the  
<sup>9</sup> conversion of these experiments into class-specific connectivity matrices representing the connection  
<sup>10</sup> between source and target structures. We introduce and validate a novel statistical model for creation  
<sup>11</sup> of connectivity matrices that combines spatial and categorical smoothing to share information  
<sup>12</sup> between similar neuron classes. We then illustrate overall and cell-type specific connectivity patterns  
<sup>13</sup> in the resultant connectivities.

## AUTHOR SUMMARY

## INTRODUCTION

<sup>14</sup> The animal nervous system enables an extraordinary range of natural behaviors, and has inspired  
<sup>15</sup> much of modern artificial intelligence. Neural connectivities - axon-dendrite connections from one  
<sup>16</sup> region to another - form the architecture underlying this capability. These connectivities vary by  
<sup>17</sup> neuron type, as well as axonic source and dendritic target structure. Thus, characterization of the

18 relationship between neuron type and source and target structure is an important step to  
19 understanding the nervous system.

20 Viral tracing experiments - in which a viral vector expressing GFP is transduced into neural cells  
21 through stereotaxic injection - are a useful tool for understanding these connections on the mesoscale  
22 (???). The GFP protein moves from axon to dendrite through the process of anterograde projection, so  
23 neurons 'downstream' of the injection site will also fluoresce. Two-photon tomography imaging can  
24 then determine the location and strength of the fluorescent signals in two-dimensional slices. These  
25 locations can then be mapped back into three-dimensional space, and the signal is partitioned into  
26 the transduced source and merely transfected target regions.

27 The conversion of such experiment-specific signals into an overall estimate of the connectivity  
28 strength of two regions is accomplished by a statistical model. ? and ? describe two such methods.  
29 Intuitively, both of these models provide some improvement over simply averaging the projection  
30 signals of injections in a given region. is another. These models are evaluated based off of their ability  
31 to predict held-out experiments in leave-one-out cross validation. A model that performs well in such  
32 validation experiments is then assumed to generate the most accurate connectivity.

33 Both ? and ? develop models for mostly wild-type mice using a standardized vector over all  
34 experiments. However, recent work (?) has extended these datasets to include viral tracing  
35 experiments inducing cell-type specific fluorescence. This is accomplished by injecting vectors with  
36 Cre-recombinase triggered GFP promoters into transgenic mice with cell-type specific  
37 Cre-recombinase expression Thus, the this paper extends the methodology of ? and ? to deal with the  
38 diverse set of cre-lines described in ?.

39 This extension relies on a to our knowledge novel estimator that takes into account both the spatial  
40 position of the labelled source, as well as the categorical cre-label. This model outperforms the model  
41 of ?, even for wild-type experiments.

42 The resulting cell-type specific connectivity matrices form a multi-way *neural connection tensor* of  
43 information about neural structure. We do not attempt an exhaustive analysis of this data, but do  
44 demonstrate several basic phenomena. First, we verify several cell-type specific patterns found  
45 elsewhere in the literature. Second, we discover cell-type specific signals in the neural connection

<sup>46</sup> tensor. Finally, we decompose the overall (wild-type) connectivity matrix into factors representing  
<sup>47</sup> archetypal connective patterns.

## METHODS

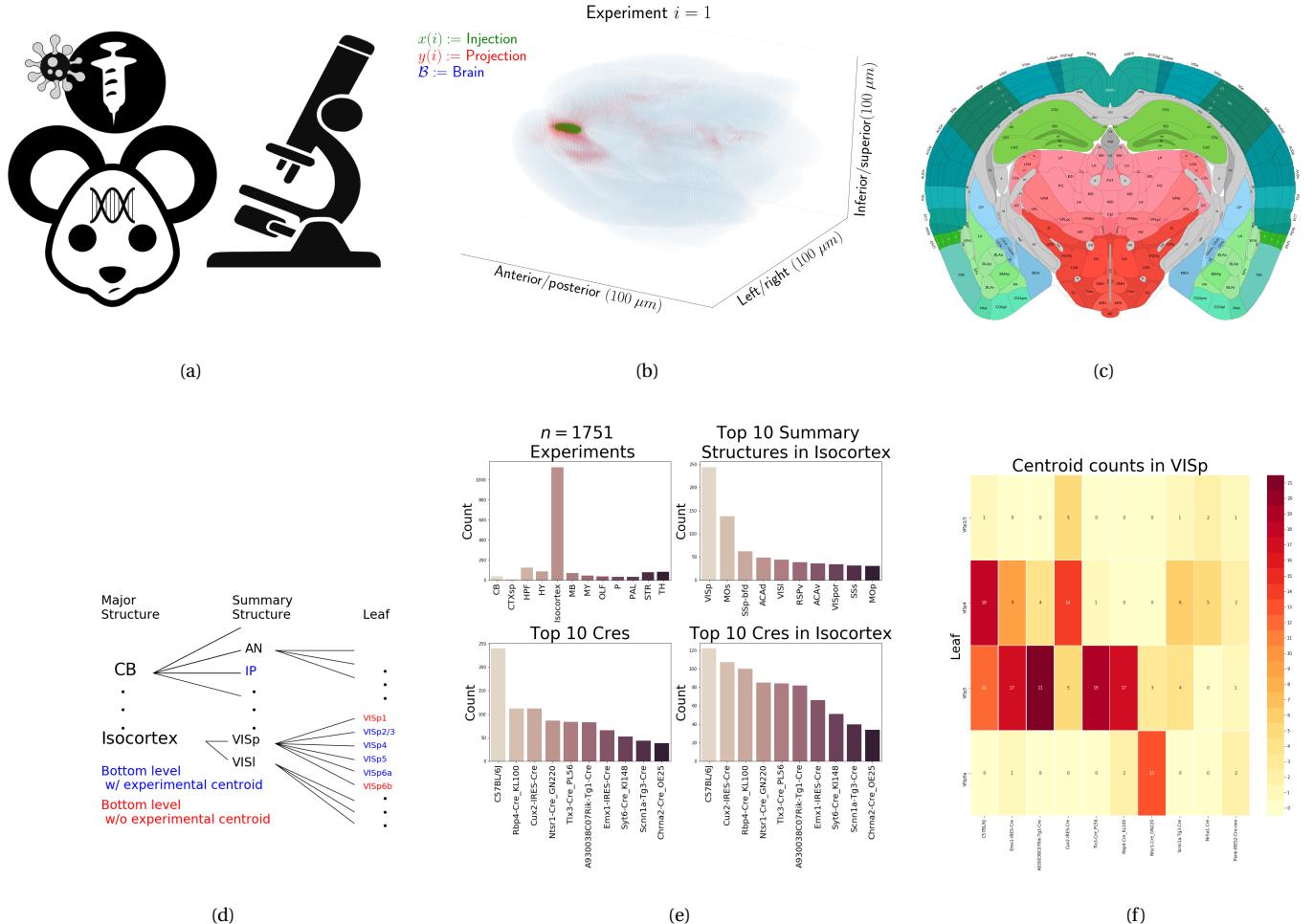


Figure 1: a) Background on histology. a) Within the brain (blue), injection (green) and projection (red) areas are determined via histological analysis and alignment to the Allen Common Coordinate Framework (CCF). b) An example of the segmentation of projection and injection for a single experiment. c) Example of structural segmentation within a horizontal plane. d) Explanation of nested structural ontology highlighting lowest-level and data-relevant structures. e) Abundances of crelines and structural injections. f) Co-occurrence of layer-specific centroids and creline within VISP

48 Our main result is the creation of cell-type specific connectivity matrices using a model trained on  
49 murine viral-tracing experiments. This section first describes the data used to generate the model, the  
50 model itself, evaluation of the model, and the use of the model in creation of the connectivity  
51 matrices. The model we ultimately propose is selected based off it's preferable performance in  
52 creation of connectivity matrices with greater accuracy than alternative approaches. We accompany  
53 these matrices with exploratory analyses of the resulting connectivities that illustrate their key  
54 features.

55 ***Mice***

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

58 ***Data***

59 Our dataset  $\mathcal{D}$  consists of  $n = 1751$  experiments from the Allen Mouse Brain Connectivity Atlas. Figure  
60 2 describes the key features of the dataset. Each experiment is performed by injecting a GFP-labelled  
61 transgene cassette with a potentially cre-specific promoter into a particular location in a cre-driver  
62 mouse. The resultant fluorescent signal is imaged, and aligned into the Allen Common Coordinate  
63 Framework (CCF), a three-dimensional idealized model of the brain that is consistent between  
64 animals.

65 Within the dataset generating the connectivity model reported in this paper, certain structure and  
66 cre-line combinations ( $S, V$ ) appear frequently, while others appear not at all. Since users of the  
67 connectivity matrices may be interested in particular combinations, or interested in the amount of  
68 data used to generate a particular connectivity estimate, we exhaustively present this information  
69 about all experiments in Appendix .

70 ***Data processing***

71 We discretize the fluorescent intensity photographically determined through histological image  
72 analysis at the  $100 \mu\text{m}$  voxel level. Thus, the fluorescence is represented as a tensor  $\mathcal{F} \in \mathbb{R}^B$  where  
73  $B \subset [1 : 132] \times [1 : 80] \times [1 : 104]$  corresponds to the subset of the voxelized  $(1.32 \times 0.8 \times 1.04)$  cm  
74 rectangular space occupied by the standard mouse brain. This fluorescence is segmented into

*75 injection and projection areas corresponding to areas of transduction and transduction/transfection,*  
*76 respectively. For a given experiment, we denote these as  $x(i)$  and  $y(i)$ , respectively. An example of*  
*77 such segmentation areas is given in Figure ??.* In order to relate the regularly discretized 3D space  $B$   
*78 with biologically informative structures such as the cortex, we also apply several levels of*  
*79 regionalization, as shown in Figure 2.* Mathematically, we refer to the regionalization map as  
*80  $r : \mathcal{B} \times \mathbb{R} \rightarrow \mathcal{R} \times \mathbb{R}$ .* Given a vector  $a$ , we also define a normalization map  $n : a \mapsto \frac{a}{\sum_{t \in T} a_t}$ . A detailed  
*81 mathematical description of these data preprocessing steps is given in Appendix .*

## *82 Connectivity*

*83 Our goal is the estimation of structural connectivity from one structure to another. At an essential*  
*84 level, cell-class specific neural connectivity is representable as a function  $f : \mathcal{V} \times \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^+$  giving*  
*85 the connection of a particular cell-class from a source position to a target position. As mentioned in*  
*86 the previous section, our injection and projection intensities are discretized into  $100\mu\text{m}$  cubic voxels.*  
*87 These voxels are contained within structures, so mathematically we can write a region  $R = \{r\}$ . We*  
*88 generically denote source regions  $S = \{s\}$  and target regions  $T = \{t\}$ .*

There are several notions of structural connectivity worth considering based on normalization with respect to the sizes of the source and/or target regions. Given a set of source regions  $\mathcal{S} = \{S\}$ , target regions  $\mathcal{T} = \{T\}$ , and cre-lines  $\mathcal{V}$  we shall estimate the following tensors:

*connectivity strength  $\mathcal{C} \in \mathcal{V} \times \mathcal{S} \times \mathcal{T} \times \mathbb{R}_{\geq 0}$  with  $\mathcal{C}(V, S, T) = \sum_{s \in S} \sum_{t \in T} f(v, s, t)$*

*normalized connectivity strength  $\mathcal{C}^S \in \mathcal{V} \times \mathcal{S} \times \mathcal{T} \times \mathbb{R}_{\geq 0}$  with  $\mathcal{C}^S(V, S, T) = \frac{1}{|S|} \sum_{s \in S} \sum_{t \in T} f(v, s, t)$*

*normalized projection density  $\mathcal{C}^D \in \mathcal{V} \times \mathcal{S} \times \mathcal{T} \times \mathbb{R}_{\geq 0}$  with  $\mathcal{C}^D(V, S, T) = \frac{1}{|S||T|} \sum_{s \in S} \sum_{t \in T} f(v, s, t)$ .*

*89 Our goal is to estimate  $\mathcal{C}(V, S, T)$  with data  $\mathcal{D}$ . We call this estimator  $\hat{\mathcal{C}}$ .*

## *90 Modelling connectivity*

Construction of such an estimator raises the important questions of 1) what data to use for estimating which connectivity, 2) how to featurize the dataset, 3) what statistical estimator to use, and 4) how to reconstruct the connectivity using the chosen estimator. Mathematically, we represent these

considerations as

$$\widehat{\mathcal{C}}(V, S, T) = e^*(\widehat{e}(e_*(\mathcal{J}(\mathcal{D}))). \quad (1)$$

This makes explicit the data featurization  $e_*$ , statistical estimator  $\widehat{e}$ , and any potential subsequent transformation  $e^*$  such as averaging over the source region, as well as the fact that different data  $\mathcal{D}$  may be used to estimate different connectivities. For example, a simple model would be to take the mean regionalized projection of all the experiments with injection centroid in a given structure. Table 1 reviews estimators used for this data-type, and explain the intuition behind our new cell-class specific estimator. Additional information is given in Appendix

Model	$e^*$	$\widehat{e}$	$e_*$	Training Data
(?)	$\widehat{e}(S)$	NNLS(X,Y)	$X = r(x(I)), Y = r(y(I))$	$I = I_M$
(?)	$\sum_{s \in S} \widehat{e}(s)$	NW(X,Y)	$X = c(x(I)), Y = r(y(I))$	$I = I_M$
Cre-NW	$\sum_{s \in S} \widehat{e}(s)$	NW(X,Y)	$X = c(x(I)), Y = n(r(y(I)))$	$I = I_S \cap I_V$
Expected-loss	$\sum_{s \in S} \widehat{e}(s)$	EL <sub>S</sub> (X, Y, V)	$X = c(x(I)), Y = n(r(y(I))), V = v(I)$	$I = I_S$

Table 1: Estimation of  $\mathcal{C}$  using connectivity data. The regionalization, estimation, and featurization steps are denoted by  $e^*$ ,  $\widehat{e}$ , and  $e_*$ , respectively. The training data used to fit the model is given by  $I$ . We generically denote the set of experiments used to train a particular model as  $I$ , and experiments from particular major brain divisions, summary structures, and leafs as  $I_M$ ,  $I_U$ , and  $I_L$ , respectively.

Our new methodological contributions in this area - the Cre-NW and Expected-loss models - have several differences from the previous methods. Both the ? non-negative least squares and ? Nadaraya-Watson take into account  $s$  and  $t$ , but not  $v$ . Since our goal is creation of cre-specific connectivities, our new estimators specifically account for this information. The cre-specific Nadaraya-Watson estimator only uses experiments from a particular cre-line to predict cell-class connectivity, while the Lxpected Loss estimator shares information between cre-lines. We also normalize projections by total intensity to account for differences in the cre-driven expression of eGFP via the various transgene promoters.

105 **Evaluating connectivity models**

106 The final modeling question is how to select optimum functions from within and between our  
 107 estimator classes. Examining Equation 1, we can see the equation in 3D coordinates,  
 108  $\hat{f}(v, s, t) = \hat{e}(e_*(\mathcal{J}(\mathcal{D})))$  includes a deterministic step  $e^*$ . Since this step is included without input by  
 109 the data, we can evaluate our model by its ability to predict held-out experiments in cross-validation.  
 110 We use in particular *leave-one-out* cross validation, a simple and effective method for evaluating  
 111 estimator performance. In order to compare between methods, we necessarily restrict to the smallest  
 112 set of evaluation experiments suggested by any of our models. The surface smooth level requires  
 113 computation of a mean for each cre-line. For cross-validation to be possible, two-experiments must  
 114 be present - one with which to compute the mean, and one on which to evaluate the model. This is  
 115 true even if experiments from other cre-lines are present within the structure.

## 116 CONSTRUCTION OF THE EVALUATION SET

If we construct cre-means at the summary-structure level,  
 then even if we smooth at the leaf level, we can evaluate estimator performance on the  
 summary-structure set as long as there at least 1 experiments of any cre-line in that leaf. Predicting an  
 experiment with summary-structure surface and summary-smoothing requires another of the same  
 cre-line in the summary structure, and one of any cre-line in the same summary structure. Predicting  
 an experiment with summary-structure surface and leaf-smoothing requires another of the cre-line in  
 the summary structure, and one of any cre-line in the same leaf. This gives the same evaluation set as  
 summary-surface summary-smooth but with experiments that are the only exemplar of their cre-line  
 in the leaf removed. Predicting an experiment with leaf-structure surface and leaf-smoothing requires  
 another of the same cre-line in the same leaf.

That is,  $E_{sum}^{cre} \cap E_{leaf}^{NW}$ . We note that since the number of parameters fit is quite low relative to the size  
 of the evaluation set, we do not make use of a formal validation-test split. However, evaluating  
 likelihood solely on the training set is trivially a bad idea in Nadaraya-Watson methods.

	Total	Cre-Summary	Cre-Summary, Leaf	Cre-Leaf
0	36	10	9	4
1	7	2	2	2
2	122	79	79	62
3	85	41	41	41
4	1128	838	829	732
<sup>129</sup>	5	68	23	18
6	46	7	7	7
7	35	17	17	17
8	33	8	8	8
9	30	11	11	11
10	78	45	45	45
11	83	29	29	29

<sup>130</sup> Certain aspects of out-of-sample performance are not assessable via LOOCV. In particular, we  
<sup>131</sup> cannot evaluate model performance in regions where there are not enough experiments targeting a  
<sup>132</sup> particular cell-class. This raises the question of whether we should model these regions at all. We  
<sup>133</sup> therefore make a scientifically-motivated distinction between wild-type non-cre injections and  
<sup>134</sup> cell-class specific injections. Since wild-type connectivities are the sum of the component cell-types,  
<sup>135</sup> even if, for example, a summary-structure specific estimator for a particular cre-line with  
<sup>136</sup> leaf-smoothing will use exclusively non-wild type experiments, this will elucidate component  
<sup>137</sup> cell-class connectivities. However, such For the wild-type mice without cre-specific injection, the  
<sup>138</sup> neural connectivity is the sum of the included cell-types, so this is reasonable. However, for cell-class  
<sup>139</sup> specific connectivity, this can lead to predictions of connectivity

LOSS METRICS The loss-function used to evaluate estimator performance on the evaluation set.

We use  $l_2$ -loss and weighted  $l_2$ -loss to evaluate these predictions:

$$\text{l2-loss } l(\hat{f}) = \frac{1}{|I_M|} \sum_{i \in I_M} \|r(y(i)) - \hat{f}(c(i))\|_2^2$$

$$\text{weighted l2-loss } l(\hat{f}) = \frac{1}{|\{S, V\}|} \sum_{s, v \in \{S, V\}} \frac{1}{|I_{s, v}|} \sum_{i \in I_{s, v}} l(r(y(i)), \hat{f}(\mathbb{D} \setminus i))$$

140 As a final modelling step, we establish a lower limit of detection. This is covered in Appendix.

141 ***Connectivity analyses***

142 We quantify and illustrate some of the interesting neuronal processes underlying our estimated  
 143 connectome. First, we cluster projection pattern by cell-class and source structure. This shows that  
 144 cell-class has a dominating effect on projection in certain regions. Second, we extend the  
 145 characterization of ? on structural differences in short-range projections. These are primarily  
 146 assumed to be due to diffusion, and the diffusion-rate helps to characterize the basic structural  
 147 anatomy. Third, since the overall wild-type connectome results from the combination of underlying  
 148 cell-classes, we apply non-negative matrix factorization (NMF) to decompose the observed  
 149 long-range connectivity into *connectivity archetypes* that linearly combine to reproduce the observed  
 150 connectivity. These methods identify structures with both known and plausible biological meaning,  
 151 and simplistically exemplify useful posthoc analyses for data of this type. Technical details of these  
 152 approaches are given in Appendix.

## RESULTS

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

155 ***Model evaluation***

156 Table contains the sizes of these evaluation sets in each major structure. This information may be  
 157 cross-referenced visually with the figures in Our two-stage model generally performs better than the  
 158 cre-line specific NW estimator.

	Estimator	EL	NW	Average	NW	NW-wt
Smoothing		SS	Cre-SS	Cre-SS	SS	M
Target		SS	SS	SS	SS	SS
Structure	# Eval exps					
CB	10	0.044	0.081	0.081	0.058	0.439
CTXsp	2	0.497	0.497	0.497	0.497	0.000
HPF	79	0.122	0.140	0.143	0.155	0.471
HY	41	0.241	0.266	0.269	0.244	1.019
Isocortex	838	0.173	0.195	0.202	0.234	0.404
MB	23	0.151	0.151	0.166	0.139	0.759
MY	7	0.186	0.233	0.233	0.184	0.452
OLF	17	0.069	0.095	0.100	0.073	0.110
P	8	0.236	0.239	0.239	0.264	0.984
PAL	11	0.190	0.198	0.198	0.260	1.401
STR	45	0.084	0.088	0.089	0.097	0.265
TH	29	0.351	0.678	0.678	0.365	1.088

Table 2: Weighted losses with summary structure targets.

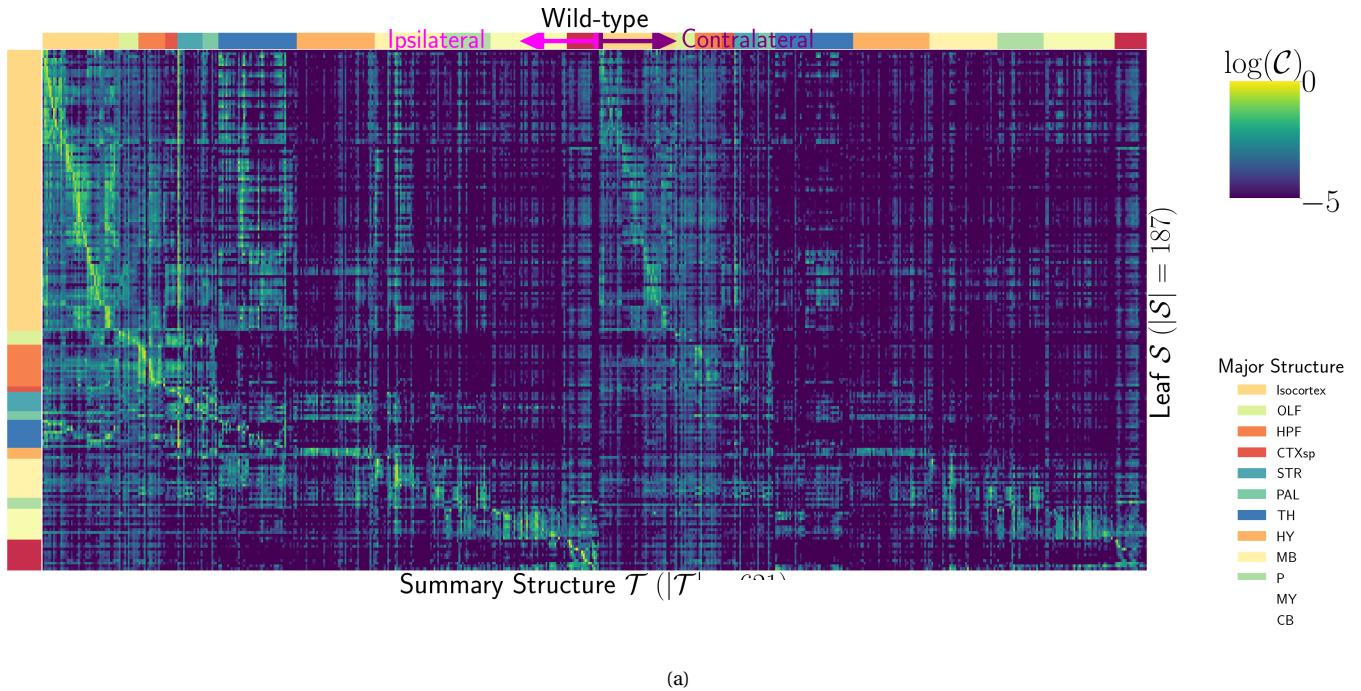
159 **Connectivities**

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

166 [https://github.com/AllenInstitute/mouse\\_connectivity\\_models/tree/2020](https://github.com/AllenInstitute/mouse_connectivity_models/tree/2020).

167 The connectivity matrix for wild-type connectivities from leaf sources to summary structure targets  
168 is illustrated in Figure ???. The clear intraareal connectivities mirror previous estimates in ? and ? and  
169 descriptive depictions of individual experiments in ?. Compared with ?, our more discretized source  
170 smoothing and greater number of experiments leads to a significantly more discretized connectivity  
171 matrix. This is generally expected - for example, different cortical layers have more substantially  
172 different connectivities.

173 The cell-type specific connectivities that we provide also conform to well-known behaviors.  
174 Examples from the visual processing and motor control regions of the cortex are given in Figure ?? for  
175 both wild type and several cre-lines. Rbp4-Cre and Ntsr1-Cre target layers 5 and 6, respectively. As in  
176 ?, layer 5 projects to anterior basolateral amygdala (BLA) and capsular central amygdala (CEA), while  
177 layer 6 does not.



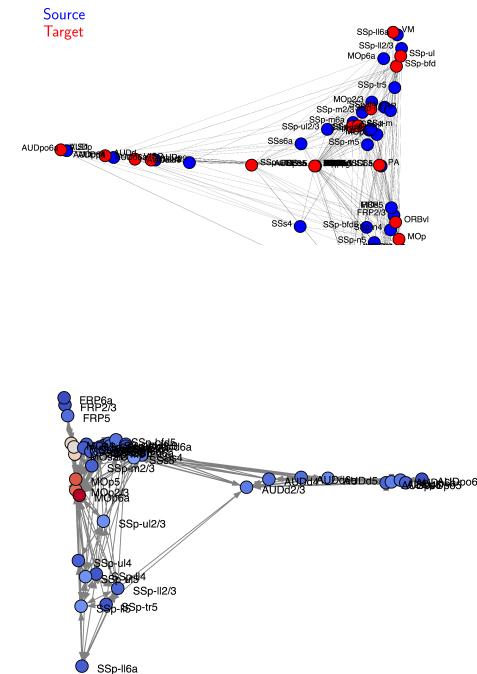
(a)

	# Ipsilateral Leaf Targets	Top Entropy	Bottom Sparsity	Bottom Entropy	Top Sparsity
Isocortex	51	CP	BAC	BAC	ENTI
OLF	11	TMv	III	III	NaN
HPF	15	IG	EPv	PA	NaN
CTXsp	7	TT	FC	APr	TT
STR	14	RPA	ISN	PVR	TU
PAL	9	PG	ACVII	GR	MG
TH	44	NOD	DN	SSp-ll	SCm
HY	44	CLA	SH	LSc	DG
MB	39	NDB	SubG	SGN	SUB
P	26	MT	Acs5	SOC	NDB
MY	43	RT	NaN	OV	EPd
CB	18	ECT	AOB	MOB	GU

(b)

	# Ipsilateral Leaf Targets	Top Entropy	Bottom Sparsity	Bottom Entropy	Top Sparsity
Isocortex	51	CP	BAC	BAC	ENTI
OLF	11	TMv	III	III	NaN
HPF	15	IG	EPv	PA	NaN
CTXsp	7	TT	FC	APr	TT
STR	14	RPA	ISN	PYR	TU
PAL	9	PG	ACVII	GR	MG
TH	44	NOD	DN	SSp-ll	SCm
HY	44	CLA	SH	LSc	DG
MB	39	NDB	SubG	SGN	SUB
P	26	MT	Acs5	SOC	NDB
MY	43	RT	NaN	OV	EPd
CB	18	ECT	AOB	MOB	GU

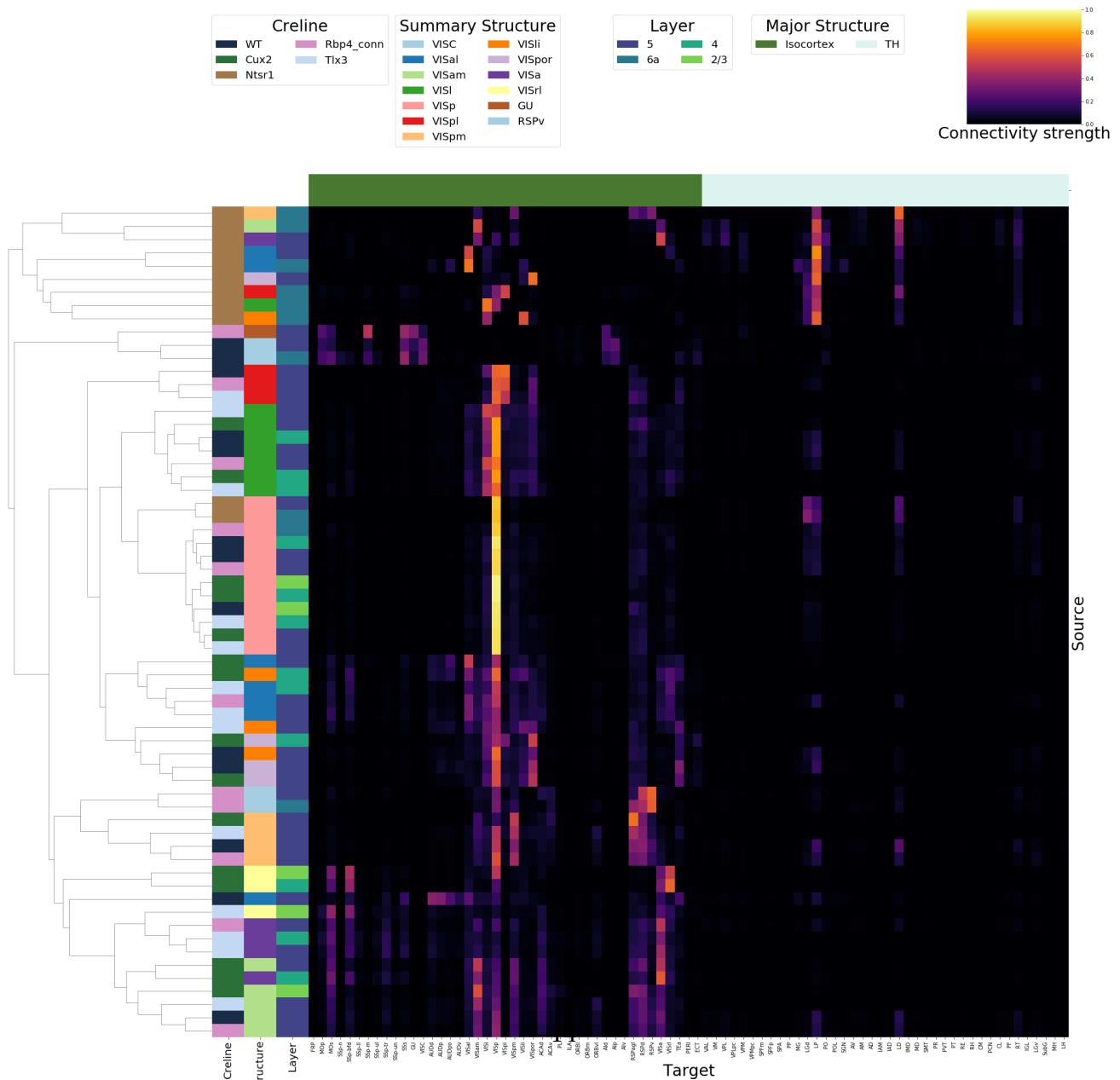
(d)



(e)



(a)



178 **Connectivity Analyses**

179 The connectivity matrix represents a collection of relatively few biological processes. For example,  
180 certain cell-types and layers have a characteristic connectivity pattern, and structures tend to connect  
181 most strongly to the most proximal areas. We elucidate these patterns through two types of analyses.  
182 First, we demonstrate cell-type specific connectivity patterns by hierarchical clustering of  
183 connectivities from multiple cre-lines, and showing that cre-line is a key factor driving the observed  
184 behavior. Then, we perform a different unsupervised analysis - non-negative matrix factorization - of  
185 distal wild-type connectivities, to estimate underlying overall connectivity patterns.

186 Figure ?? shows a collection of connectivity strengths generated using cre-specific models for  
187 wild-type, Cux2, Ntsr1, Rbp4, and Tlx3 cre-lines from visual signal processing leafs in the cortex to  
188 cortical and thalamic nucleii. Heirarchical clustering is applied to sort the different source/cre  
189 combinations by the similarity of their connectivities to summary-structure targets. This analysis  
190 shows that Ntsr1 cre-lines tend to target thalamic nucleii, in particular LP and LD ?. However, with  
191 this exception, for the other plotted cre-lines, connectivity tends to cluster by source structure. That  
192 the tendency for structures to connect to themselves is quite strong emphasizes the special nature of  
193 the Ntsr1-Thalamic connection in this analysis.

194 The overall wild-type connectivity strength matrix also displays an underlying modellable  
195 structure. As discussed in ?, one of the most basic processes underlying the observed connectivity is  
196 the tendency of each source region to predominantly project to proximal regions. The heatmap in ??a)  
197 shows intraregion distances clearly contains an overall pattern reminiscent of the connectivity matrix  
198 in ?. This relationship is plotted in ?? b), showing that there exists substantial variability that would  
199 be impossible to model with low-error in a univariate model, even using the diffusion model  
200 suggested in ?. These connections are biologically meaningful, but also unsurprising, and their  
201 relative strength biases learned latent coordinate representations away from long-range structures.  
202 For this reason, we establish a  $1500\mu m$  'distal' threshold within which to exclude connections for our  
203 analysis. We then apply non-negative matrix factorization (NMF) to decompose the remaining  
204 censored matrix into a relatively small number of distinct projection signals, and apply an

205 unsupervised cross-validation method to select the optimum number of signals ([SK's](#)

206 [comment:Percent error... show reconstruction? log scale?](#)).

## DISCUSSION

207 Flattening  $\mathcal{C}$  prior to unsupervised analysis is not necessarily recommended, but provides an easy  
208 solution for this problem.

209 With respect to the model, a Wasserstein-based measure of injection similarity per structure would  
210 combine both the physical simplicity of the centroid model while also incorporating structural  
211 knowledge.

212 The Nadaraya-Watson weighting procedure introduced here is, to our knowledge, novel. In  
213 particular, our method of utilizing the expected loss to weight points differs from the minimization  
214 task of fitting data to weighted sums of neighbors (?). We make a key assumption: that the additional  
215 statistical accuracy of including more samples makes up for the fact that their expected accuracy is  
216 lower. Note that this assumption can be easily violated, if, for example, the data is distributed on a  
217 circle without error, and only nearest neighbors are most predictive.

218 Model averaging based off of cross-validation has been implemented in ?, but we note that our  
219 approach makes use of a non-parametric estimator, rather than an optimization method for selecting  
220 the weights. ([SK's comment:CITE METHOD THAT SELECTS WEIGHTS IN KERNEL \(has catchy](#)  
221 [name](#))

## ACKNOWLEDGMENTS

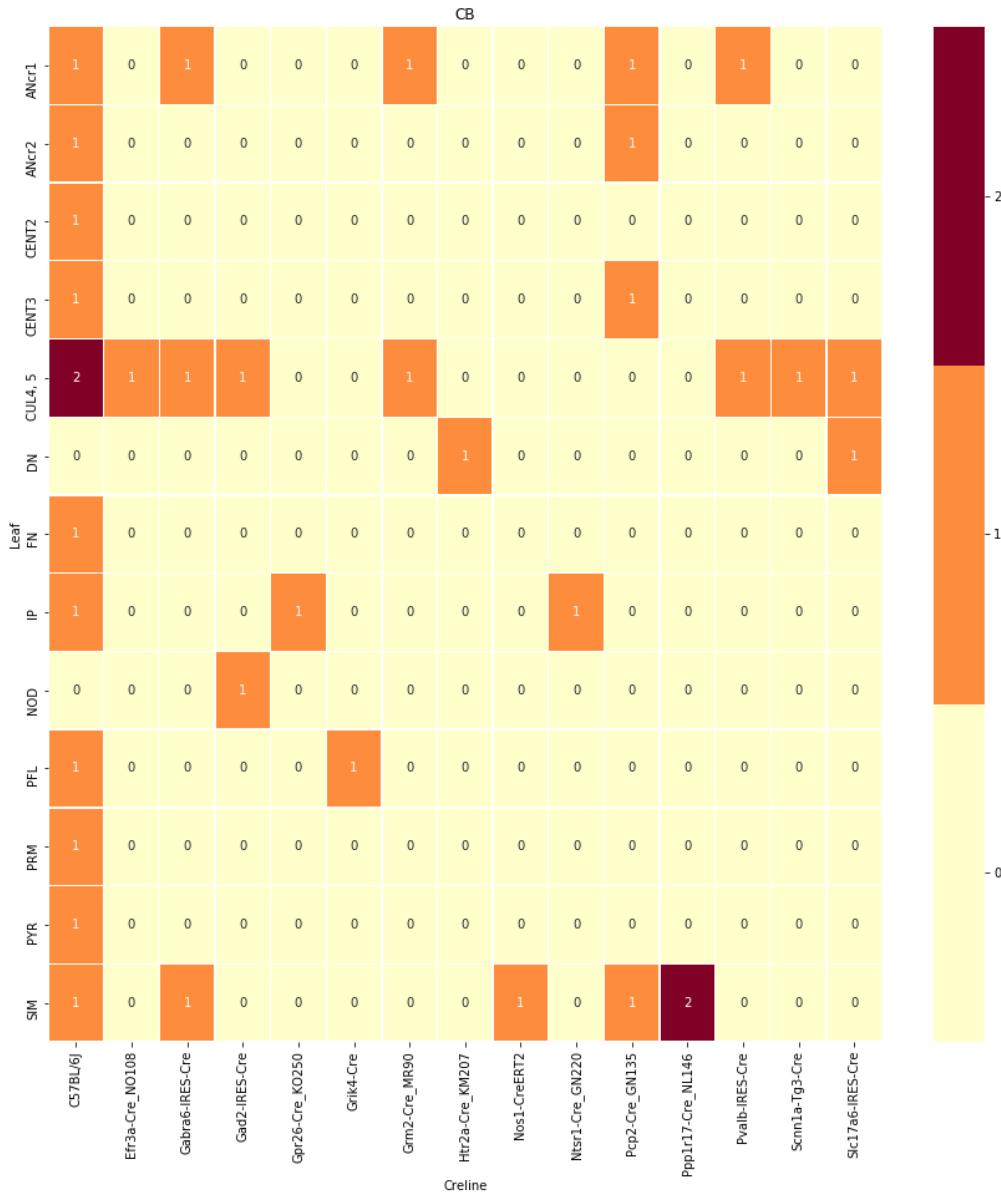
222 The Funder and award ID information you input at submission will be introduced by the publisher  
223 under a Funding Information head during production. Please use this space for any additional  
224 acknowledgements and verbiage required by your funders.

## SUPPORTING INFORMATION

### DATA

225 This section describes the set of leaf and cre-line experimental combinations.

## centroid densityoct12.png



centroid densityoct12.png

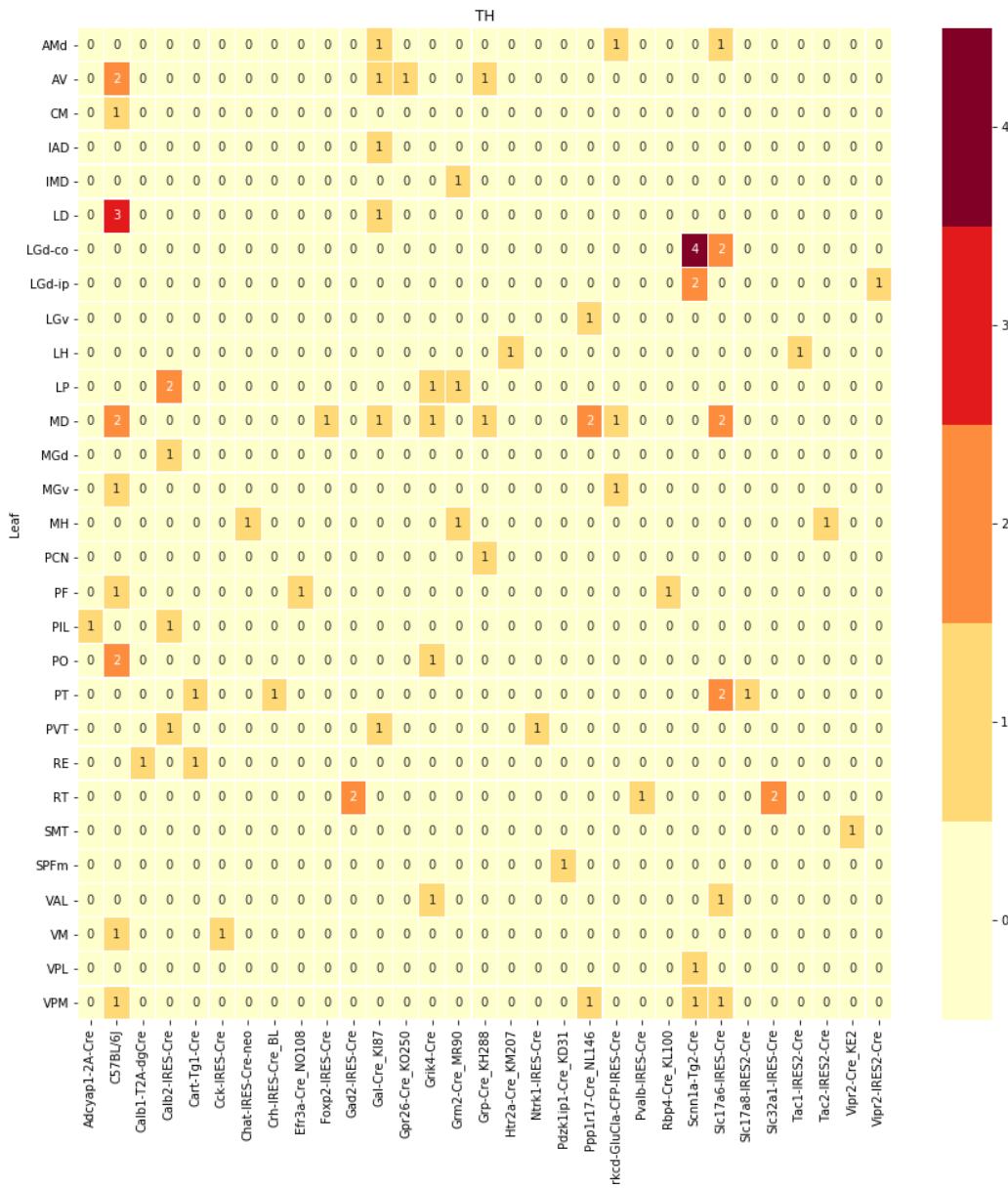


Figure 3: Caption

centroid densityoct12.png

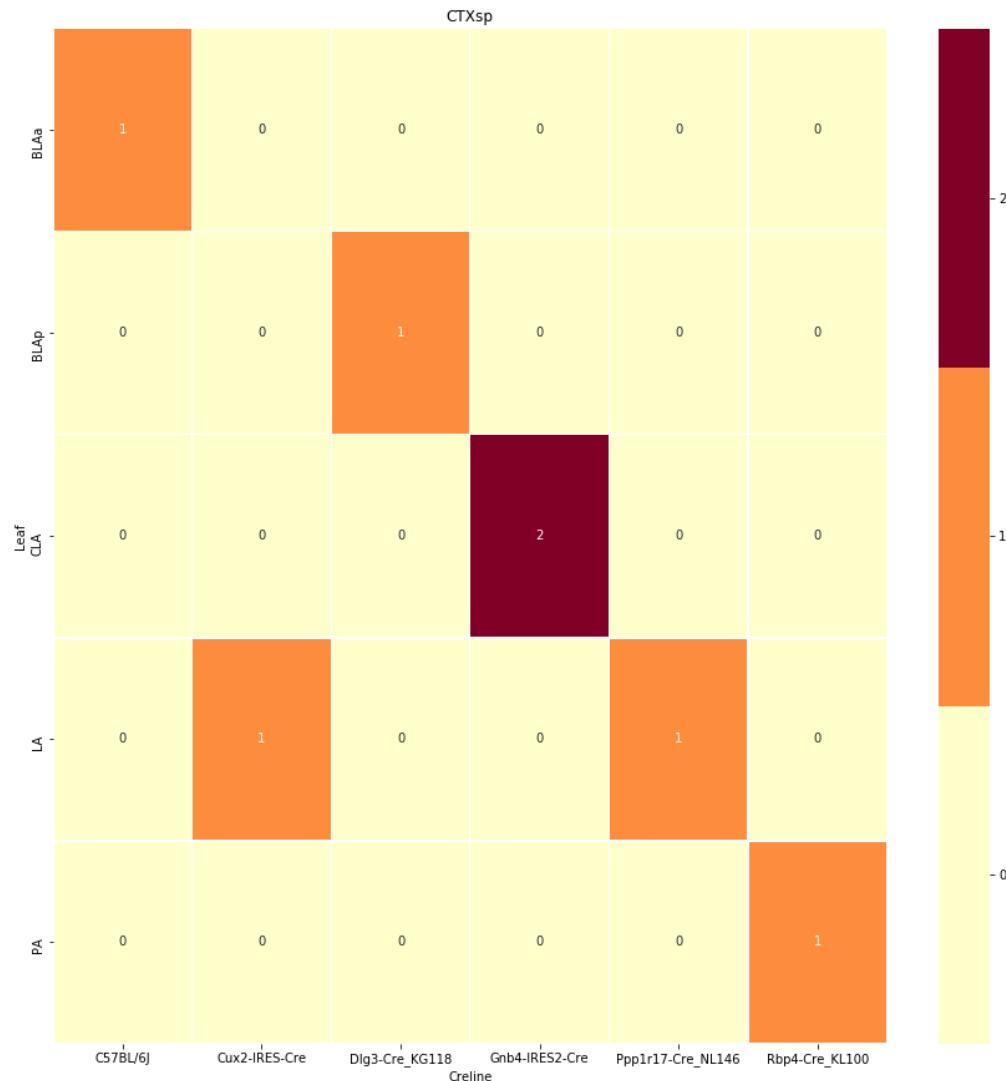
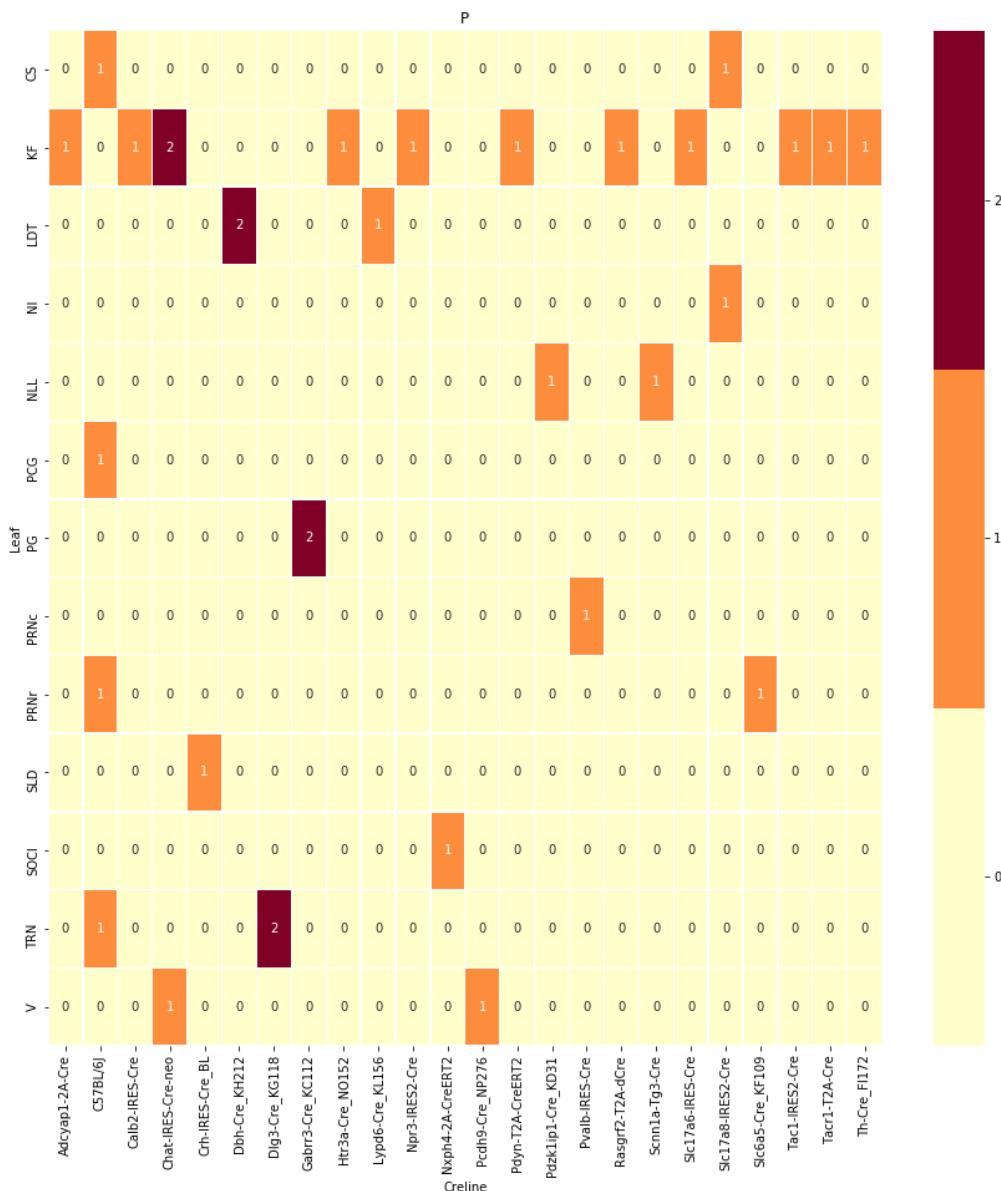
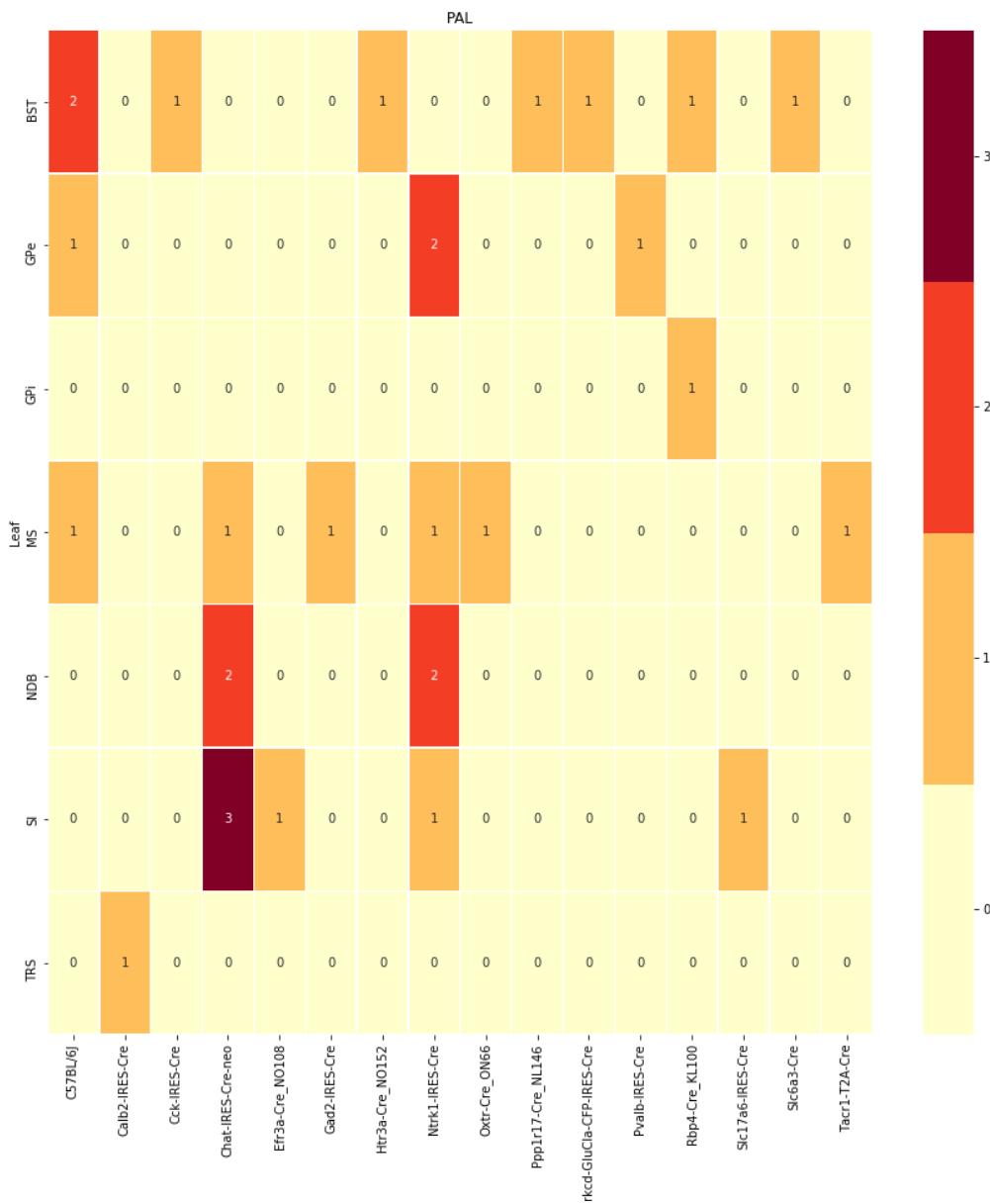


Figure 4: Caption

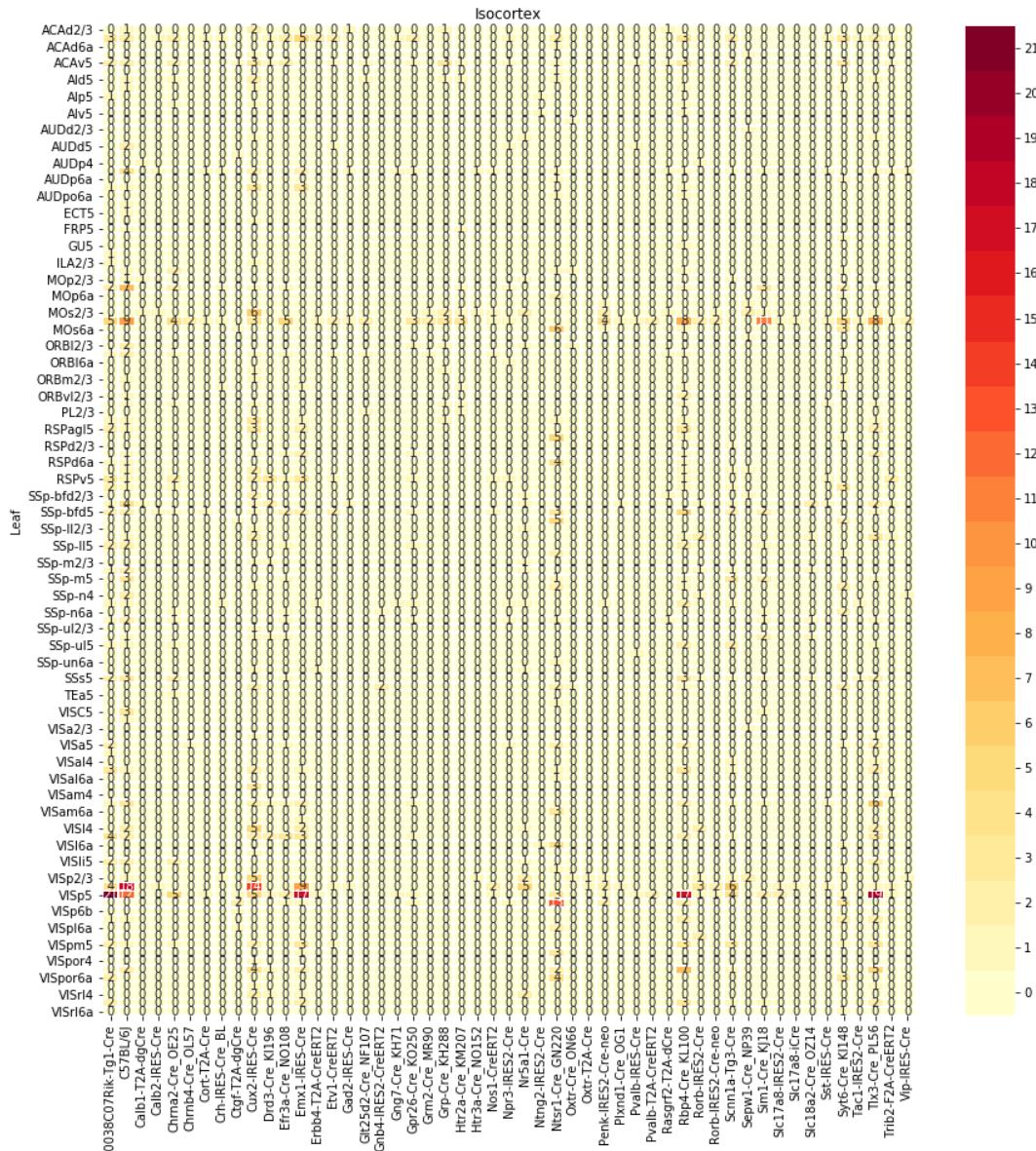
## centroid densityoct12.png



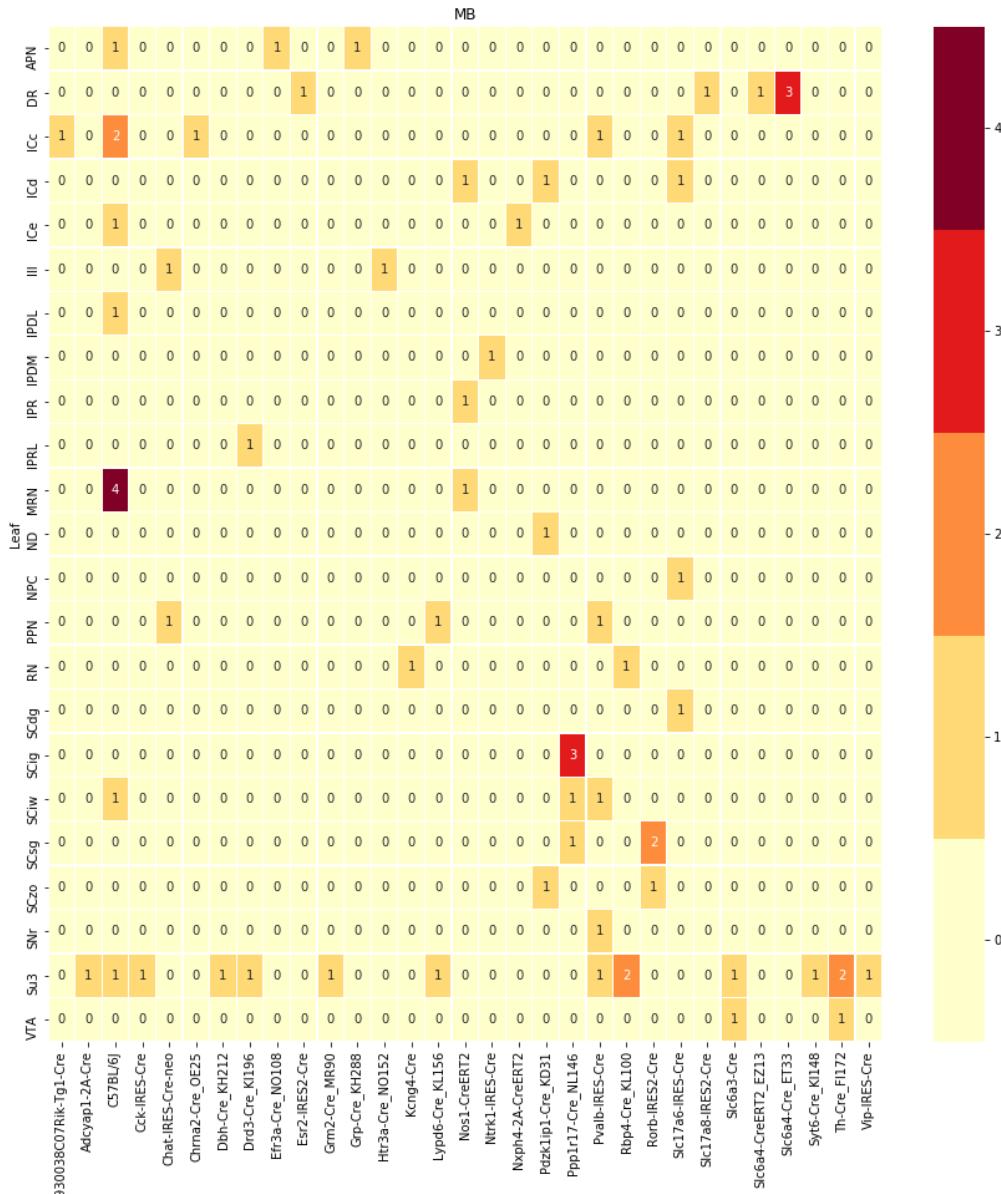
## centroid densityoct12.png



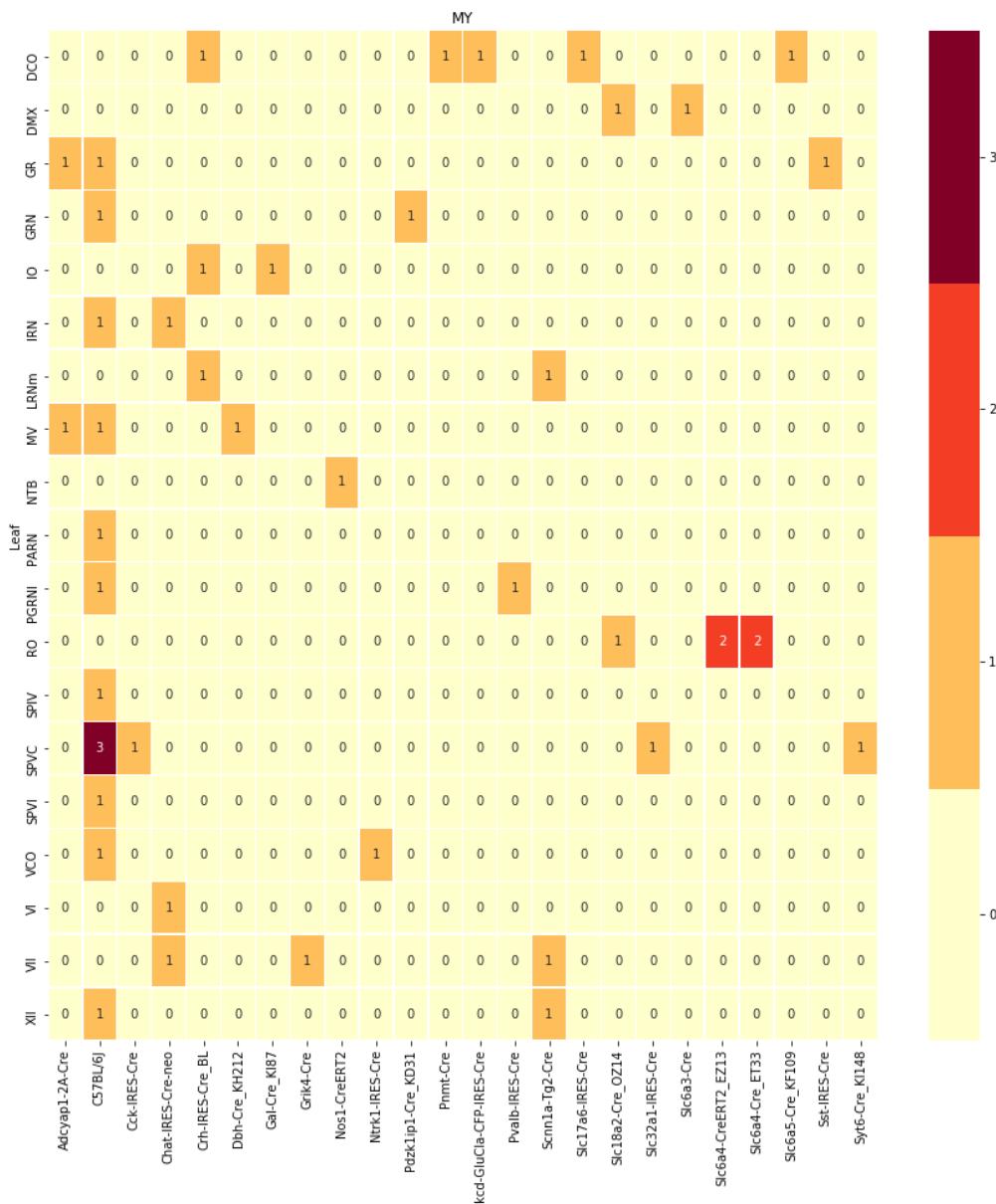
centroid densityoct12.png



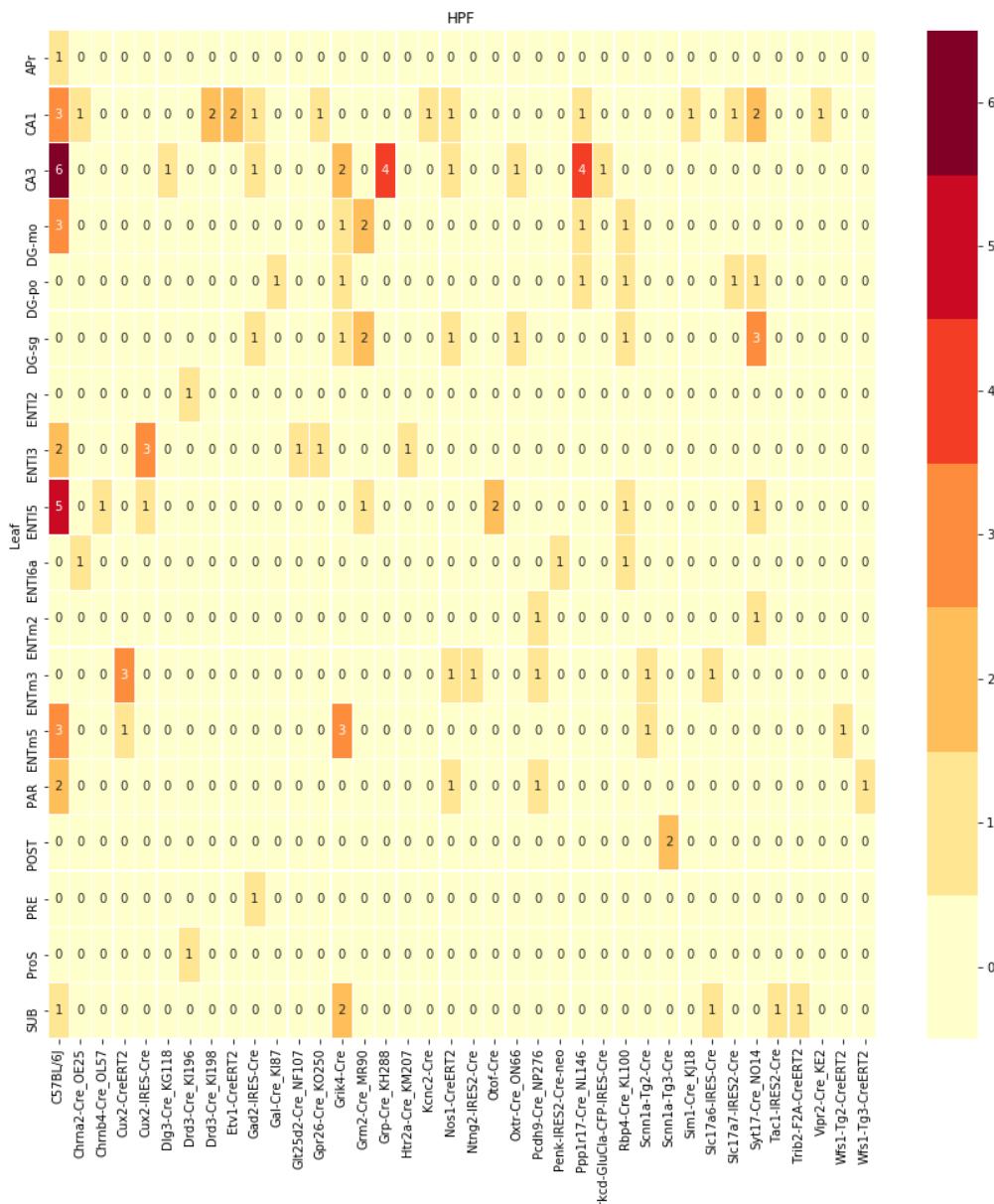
centroid densityoct12.png



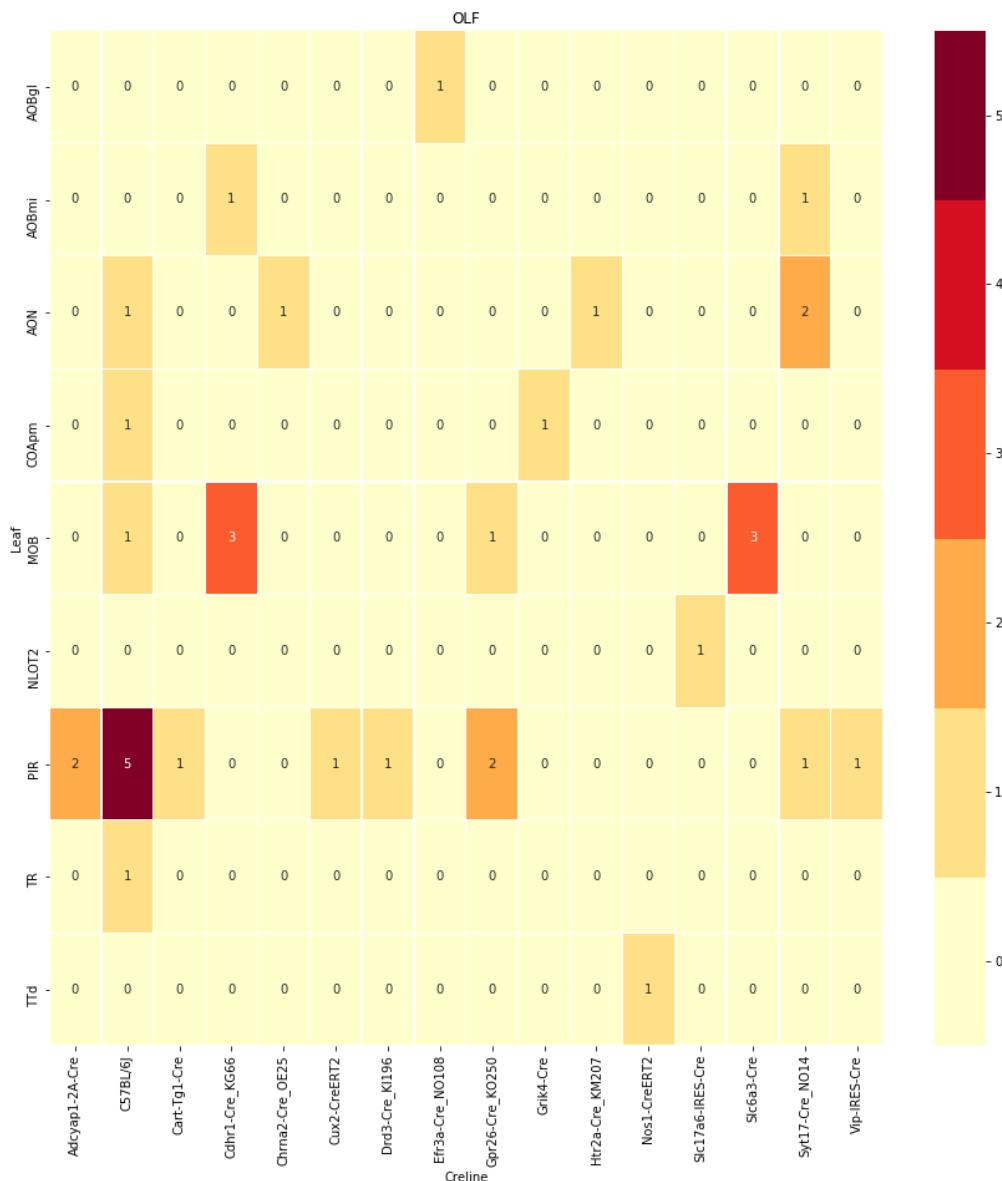
## centroid densityoct12.png



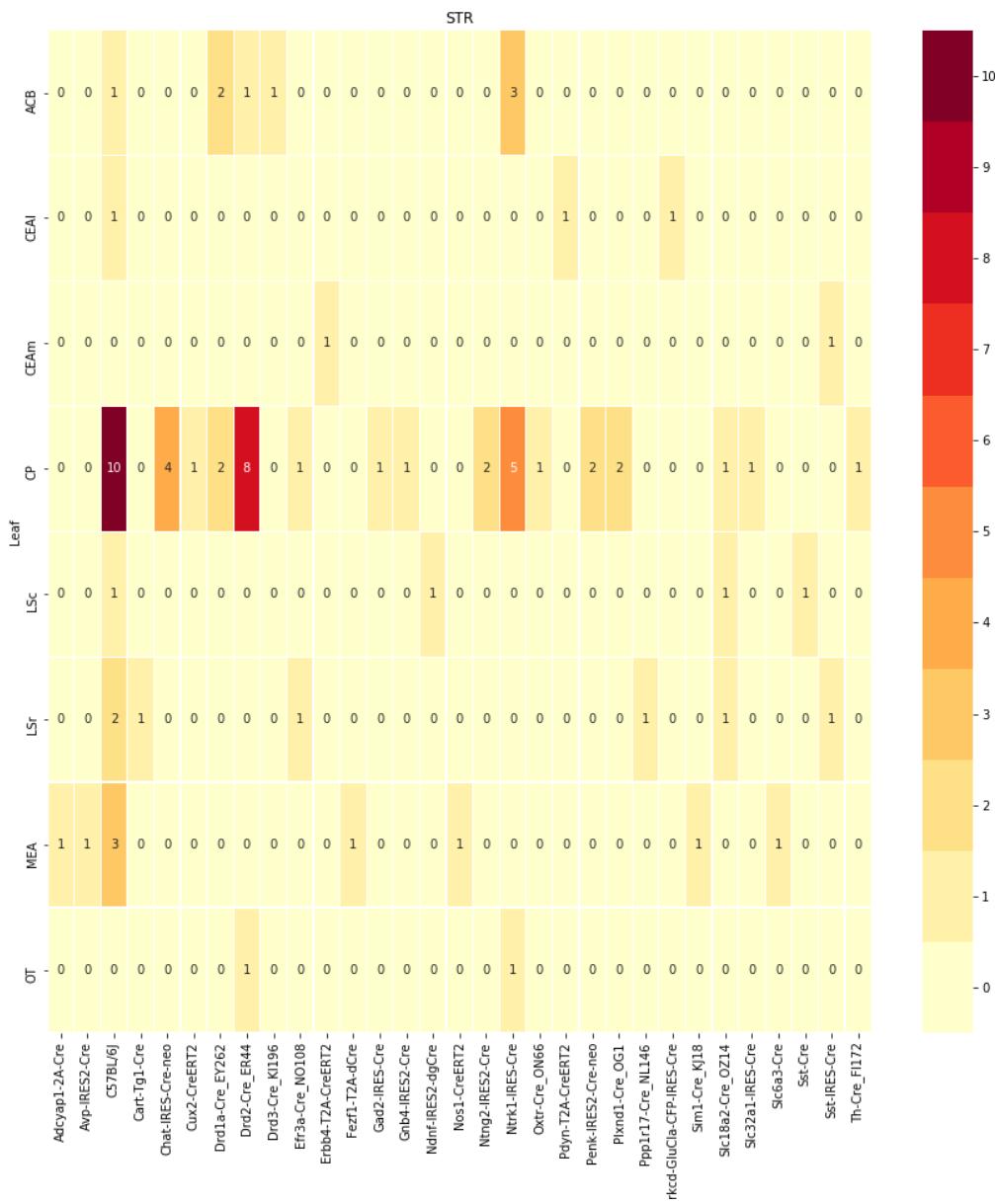
## centroid densityoct12.png



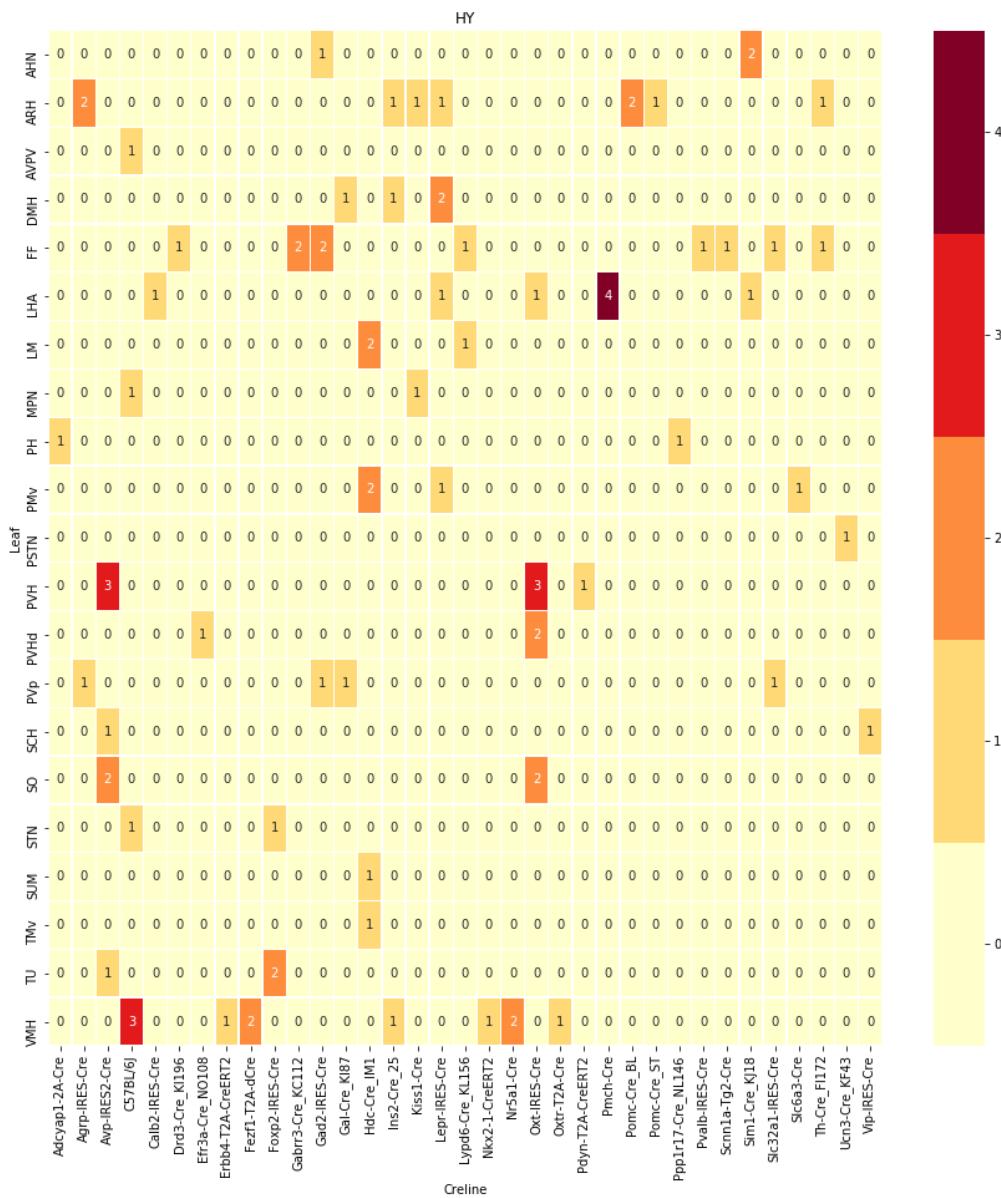
## centroid densityoct12.png



## centroid densityoct12.png



## centroid densityoct12.png



## DATA PREPROCESSING

226 Several data preprocessing steps take place prior to evaluations of the connectivity matrices.  
 227 Injections and projections were downloaded using the Allen SDK. These were originally annotated  
 228 manually. The injection and projection vectors are Hadamard multiplied by a data quality matrix. We  
 229 also have a map  $A : \mathbb{R} \rightarrow \mathbb{R}^{|S|}$  where  $S$  is the number of structures that takes the average value for voxels  
 230 in that structure

231 Data-quality censor  $y_M(i) = M(i) \odot y(i)$  (also  $x_M(i) = M(i) \odot x(i)$  for linear model)

232 The data-quality censor is established by (SK's comment:fill) The injection fraction accounts for the  
 233 relatively coarse graining of the voxel grid compared with the histological analysis used to establish  
 234 the injection region. In particular, certain voxels are only partially contained within the injection  
 235 region.

236 *Normalization* One basic but significant methodological change from ? is the normalization of  
 237 projection vectors.

238 The loss function in ? is

$$\frac{\|y - \hat{y}\|}{\|y\| \|\hat{y}\|}$$

239 *Estimators*

240 Our estimators span a range of training and featurization methods. One commonality is that they  
 241 model a connectivity vector  $f(\mathcal{D}, v, s) \in \mathbb{R}^T$ , and so we may write

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

242 Thus, for the remainder of this section, we will discuss only  $f(s, v)$ .

243 REGIONALIZED NON-NEGATIVE LEAST SQUARES In the non-negative least squares approach of ?, the  
 244 injection is considered only through its centroid, while the projection is considered regionalized. The  
 245 prediction for a given region is then given by the integral of predictions over that region, which is  
 246 computed as a sum over voxels.

247 CENTROID-BASED NADARAYA-WATSON In the Nadaraya-Watson approach of ?, the injection is  
 248 considered only through its centroid, while the projection is considered regionalized. The prediction  
 249 for a given region is then given by the integral of predictions over that region, which is computed as a  
 250 sum over voxels. That is,

$$f_*(\mathcal{D}_i) = \{c(x_i), r(y_i)\}.$$

251 Since the injection is considered only by its centroid, this model only generates predictions for  
 252 particular locations  $c$ . The prediction for a structure  $s$  is given by integrating over locations within the  
 253 structure. That is,

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

254 Here, we set  $\hat{f}$  to be the so-called Nadaraya-Watson estimator

$$\hat{f}_{NW}(c(x_{1:n}), r(y_{1:n}))(c, \nu) = \sum_{i \in I} \frac{\omega_{c(x_i)c}}{\sum_{i \in I} \omega_{c(x_i)c}} r(y_i)$$

255 with  $\omega_{c(x_i)c} = \exp(-\gamma d(c, c(x_i))^2)$  where  $d$  is the Euclidean distance between centroid  $c(x_i)$  and voxel  $c$ .

256 Several facets of the estimator are visible here.  $f_{NW}^\gamma$  is the Nadaraya-Watson estimator with  
 257 smoothing given by inverse-bandwidth  $\gamma$ . A smaller  $\gamma$  corresponds to a greater amount of smoothing,  
 258 and index set  $I \subseteq \{1 : n\}$  indicates which experiments to use to generate the prediction. Fitting  $\gamma$  via  
 259 empirical risk minimization therefore bridges between 1-nearest neighbor prediction and averaging  
 260 of all experiments in  $I$ . In ?,  $I$  consisted of experiments sharing the same brain division. This model is  
 261 easily extensible to a cre-specific model by restricting the index set to only include experiments with  
 262 the same cre-line.

263 THE EXPECTED-LOSS ESTIMATOR The response induced by each of the cre-lines is effected by both  
 264 the injection location and the targeted cell types. Cre-lines that target similar cell types are therefore  
 265 expected to induce similar projections, and including similar cre-lines in our estimator thus increases  
 266 the effective sample size. In order to leverage this fact in a data-driven way, we introduce an estimator  
 267 that assigns a predictive weight to each training point that depends both on its centroid-distance and  
 268 cre-line. This weight is determined by the expected prediction error of each of the two feature types,

<sup>269</sup> as determined by cross-validation. These weights are then utilized in a Nadaraya-Watson estimator in  
<sup>270</sup> a final prediction step.

<sup>271</sup> We formalize cre-line behavior as the average regionalized projection of a cre-line in a given leaf.  
<sup>272</sup> This vectorization of categorical information is known as target encoding. We define a *cre*-distance in  
<sup>273</sup> a leaf to be the distance between the target-encoded projections of two cre-lines. The relative  
<sup>274</sup> predictive accuracy of *cre*-distance and centroid distance is determined by fitting a surface of  
<sup>275</sup> projection distance as a function of *cre*-distance and centroid distance.

<sup>276</sup> In mathematical terms, our full feature set consists of the centroid coordinates and the  
<sup>277</sup> target-encoded means of the combinations of virus type and injection-centroid structure. That is,

$$f_*(\mathcal{D}_i) = \{c(x_i), \bar{r}(y_{I_v}), r(y_i)\}.$$

<sup>278</sup>  $f^*$  is defined as in (2). The expected loss estimator is then

$$\hat{f}_{EL}(c, c(x_i), v, r(y_{I_v})) = \sum_{i \in I} \frac{\nu(c(x_i), c, v_i, v)}{\sum_{i \in I} \nu(c(x_i), c, v_i, v)} r(y_i)$$

<sup>279</sup> where

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

<sup>280</sup> Note that  $g$  must be a concave, non-decreasing function of its arguments with  $g(0, 0) = 0$ , then  $g$   
<sup>281</sup> defines a metric on the product of the metric spaces defined by experiment centroid and  
<sup>282</sup> target-encoded cre-line, and  $\hat{f}_{EL}$  is a Nadaraya-Watson estimator. A derivation of this fact is given in  
<sup>283</sup> Appendix , and we therefore use shape-constrained B-splines to estimate  $g$ .

<sup>284</sup> This contrasts with the model is ?, where  $\hat{f}(c)$  does not depend on  $v$ , and ?, where connectivity was  
<sup>285</sup> directly estimated by  $\hat{f}$  a function of  $S$  without an integral. Estimating  $\hat{f}(v, c)$  shares the advantage of  
<sup>286</sup> fine-scale spatial resolution with ?, but in addition enables us to model a particular virus-type  $v$ , and,  
<sup>287</sup> as we will see, make use of experimental data in our estimator.

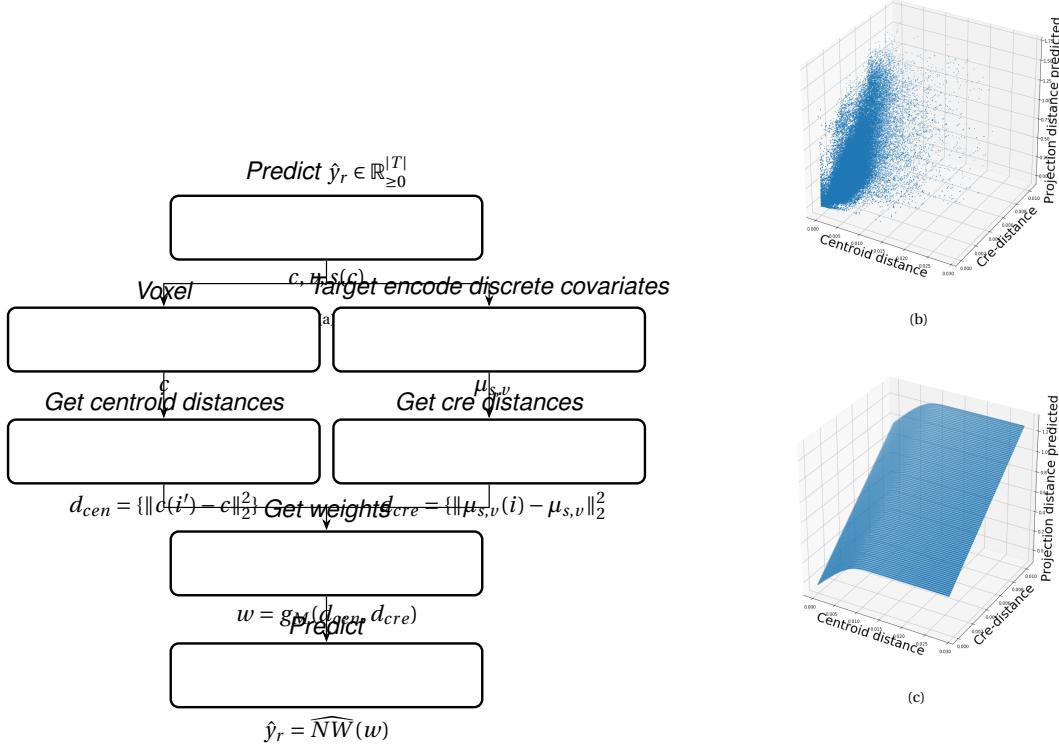


Figure 5: The Expected-Loss estimator

## THE EXPECTED-LOSS ESTIMATOR

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

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

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

- 297    ▪ The first property of a metric is that  $f(x, x') = 0 \iff x = x'$ . The left implication:  
 298        $x = x' \implies f(x_1, x'_1, \dots, x_n, x'_n) = g(0, \dots, 0)$ , since  $d$  are metrics. Then, since  $g(0) = 0$ , we have that  
 299        $f(x, x') = 0$ . The right implication:  $f(x, x') = 0 \implies d = 0 \implies x = x'$  since  $d$  are metrics.
- 300    ▪ The second property of a metric is that  $f(x, x') = f(x', x)$ . This follows immediately from the  
 301       symmetry of the  $d_i$ , i.e.  $f(x, x') = f(x_1, x'_1, \dots, x_n, x'_n) = g(d_1(x_1, x'_1), \dots, d_n(x_n, x'_n)) =$   
 302        $g(d_1(x'_1, x_1), \dots, d_n(x'_n, x_n)) = f(x'_1, x_1, \dots, x'_n, x_n) = f(x', x)$ .
- 303    ▪ The third property of a metric is the triangle inequality:  $f(x, x') \leq f(x, x^*) + f(x^*, x')$ . To show this  
 304       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  
 305       non-decreasing and by the triangle inequality of  $d$ . Then, since  $g$  is concave,  
 306        $g(d(x, x^*) + d(x^*, x')) \leq g(d(x, x^*)) + g(d(x^*, x')) = f(x, x^*) + f(x^*, x')$ .

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

- 308    ▪ The first property is that  $g(d) = 0 \iff d = 0$ . We first show the right implication:  $g(d) = 0$ , and  
 309        $g(d) = f(x, x')$ , so  $x = x'$  (since  $f$  is a metric), so  $d = 0$ . We then show the left implication:  
 310        $d = 0 \implies x = x'$ , since  $d$  is a metric, so  $f(x, x') = 0$ , since  $f$  is a metric, and thus  $g(d) = 0$ .
- 311    ▪ The second property is that  $g$  is non-decreasing. We proceed by contradiction. Suppose  $g$  is  
 312       decreasing in argument  $d_1$  in some region  $[l, u]$  with  $0 < l < u$ . Then  
 313        $g(d_1(0, l), 0) \geq g(d_1(0, 0), 0) + g(d_1(0, u), 0) = g(d_1(0, u), 0)$ , which violates the triangle inequality on  
 314        $f$ . Thus, decreasing  $g$  means that  $f$  is not a metric, so  $f$  a metric implies non-decreasing  $g$ .
- 315    ▪ The final property is that  $g$  is concave. We proceed by contradiction. Suppose  $g$  is strictly convex.  
 316       Then there exist vectors  $d, d'$  such that  $g(d + d') < g(d) + g(d')$ . Assume that  $d$  and  $d'$  only are  
 317       non-zero in the first position, and  $d = d(0, x), d' = d(0, x')$ . Then,  $f(0, x) + f(0, x') < f(0, x + x')$ ,  
 318       which violates the triangle inequality on  $f$ . Therefore,  $g$  must be concave.

## DECOMPOSING THE CONNECTIVITY MATRIX

- 319    We utilize non-negative matrix factorization (NMF) to analyze the principal signals in our  
 320    connectivity matrix. Here, we review this approach as applied to decomposition of the distal elements  
 321    of the estimated connectivity matrix  $\hat{\mathcal{C}}$  to identify  $q$  connectivity archetypes. Aside from the NMF

322 program itself, the key elements are selection of the number of archetypes  $q$  and stabilization of the  
 323 tendency of NMF to give random results over different initialization.

324 ***Non-negative matrix factorization***

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

$$NMF(X, q) = \arg \min_{W \in \mathbb{R}_{\geq 0}^{a \times q}, H \in \mathbb{R}_{\geq 0}^{q \times b}} \| (X - WH) \|_2^2.$$

327 NMF creates a useful decomposition since  $X$  is in the positive orthant, and PCA cannot not apply.  
 328 There is no orthogonality without sparsity.

329 We note the existence of NMF with alternative norms for certain marginal distributions, but leave  
 330 utilization of this approach for future work (?). We can also apply a mask  $1_M \in \mathbb{R}^{S \times T}$  of ones and zeros  
 331 and solve

$$\arg \min_{W \in \mathbb{R}_{\geq 0}, H \in \mathbb{R}_{\geq 0}} \| 1_M \odot ((\hat{\mathcal{C}} - WH)) \|_2^2$$

332 For us, such a mask serves for two purposes. First, it enables computation of the NMF objective while  
 333 excluding self and nearby connections. These connections are both strong and linearly independent,  
 334 and so would dominate the *NMF* reconstruction error. Long range connections are more biologically  
 335 interesting or cell-type dependent. Second, it enables cross-validation based selection of the number  
 336 of retained components.

337 ***Cross-validating NMF***

338 Perhaps surprisingly, cross-validation techniques may also be applied to unsupervised learning  
 339 problems. These techniques are somewhat standard, but not entirely well-known, so we review them  
 340 here, in particular as they apply to the NMF problem. A NMF model is first fit on a reduced data set,  
 341 and an evaluation set is held out. After random masking of the evaluation set, the loss of the learned  
 342 model is then evaluated on the basis of successful reconstruction of the held-out values. This  
 343 procedure is performed repeatedly, with different held out regions and random mask at different  
 344 dimensionalities  $l$ , to determine to point past which additional hidden units provide no  
 345 reconstructive value.

That is, given a matrix  $X \in \mathbb{R}^{S \times T}$  we can decompose  $X \sim d(e(X))$  where  $e(X)$  is some map that encodes  $X$  in a learned representation, and  $d$  is the decoding reconstruction map. In our case,  $d$  is simply left multiplication by  $W$ , and  $e$  is the solution of a regularized non-negative least squares optimization problem

$$H := e_W(X) = \arg \min_{\beta} \|X - W\beta\|_2^2.$$

- <sup>346</sup> The form of this solution particularly motivates our cross-validation estimator.

Recall that in supervised learning, the learned model is  $Y \sim f(X)$ . Standard cross-validation removes elements of  $X$ , fits  $f$ , and then uses the  $f$  learned from part of the data to predict  $Y$ . A good  $f$  will have low error on the training data, and also low error on the test data, indicating that it has not overfit. Although there is no assumed dichotomy between  $X$  and  $Y$  in unsupervised learning, for techniques like autoencoders, the above paradigm still applies, i.e., one can still hold out values of  $X$ . We can then estimate

$$\arg \min_{d,e} \widehat{E}(l(X, d_{XC}(e_{XC}(X)))) = \sum_{r=1}^R l(X_r, d_{XC_r}(e_{XC_r}(X_r)))$$

over  $R$  random samples of rows of  $X$ . However, in our setting, since computing  $e(X)$  on the test rows amounts to fitting a non-negative least squares w.r.t.  $W$ , so the negative effects of an overfit model can simply be optimized away from. Thus, the standard solution is to generate uniformly random masks  $1_{M(p)} \in \mathbb{R}^{S \times T}$  where

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

Our cross-validation error is then

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

where

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

The optimum number of components is then

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

347 ***Stabilizing NMF***

348 The NMF program is non-convex, and, empirically, individual replicates will not converge to the same  
 349 optima. One solution therefore is to run multiple replicates of the NMF algorithm, cluster the  
 350 resulting vectors. This approach raises the questions of how many clusters to use, and how to deal  
 351 with stochasticity in the clustering algorithm itself. We address this issue through the notion of  
 352 clustering stability (?).

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

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

353 A review of this approach is found in ?. Intuitively, archetype vectors that cluster together frequently  
 354 over clustering replicates indicate the presence of a stable clustering. For  $d$ , we utilize the adjusted  
 355 Rand Index - a simple dissimilarity measure between clusterings. Note that we expect to select slightly  
 356 more than the  $q$  components suggested by cross-validation, since archetype vectors which appear in  
 357 one NMF replicate generally should appear in others. We then select the  $q$  clusters with the most  
 358 archetype vectors - the most stable NMF results - and take the median of each cluster to create a  
 359 sparse representative archetype.

## ESTABLISHING A LOWER DETECTION LIMIT

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

<sup>367</sup> threshold to zero. This should make our estimated connectivities more accurate, in particular in the  
<sup>368</sup> biologically-important sense of sparsity.

<sup>369</sup> The limit-of-detection problem is common in a variety of scientific fields, but statistical  
<sup>370</sup> methodology is not We utilize the zero-inflated adjusted Kendall- $\tau$  statistic

$$\tau_0 = p_{11}^2 \tau_{11} + 2(p_{00}p_{11} - p_{10}p_{01}),$$

<sup>371</sup> where [\(??\)](#)

## COMPETING INTERESTS

<sup>372</sup> This is an optional section. If you declared a conflict of interest when you submitted your manuscript,  
<sup>373</sup> please use this space to provide details about this conflict.

## TECHNICAL TERMS

<sup>374</sup> All NETN article types require Technical Terms.

<sup>375</sup> Identify approximately 10 key terms that are mentioned in your article and whose usage and  
<sup>376</sup> definition may not be familiar across the broad readership of the journal. Provide brief (20-word or  
<sup>377</sup> less) definitions for each term, avoiding in these definitions the use of jargon, or highly technical or  
<sup>378</sup> specialized language. When the article is typeset, the Technical Terms will appear in the margins at or  
<sup>379</sup> near their first mention in the text.

<sup>380</sup> In your manuscript, bold the first occurrence of each **Technical Term** and then provide a list of the  
<sup>381</sup> terms and their definitions at the end of the manuscript after the references.

<sup>382</sup> **Technical Term** a key term that is mentioned in an NETN article and whose usage and definition  
<sup>383</sup> may not be familiar across the broad readership of the journal.