

1 RESEARCH

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

4 Samson Koelle ^{1,2}, **Dana Mastrovito** ¹, **Jennifer D Whitesell** ¹, **Karla E Hirokawa** ¹, **Hongkui Zeng** ¹, **Marina Meila** ², **Julie A**
5 Harris ¹, **Stefan Mihalas** ¹

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

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

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

ABSTRACT

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

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

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

1 INTRODUCTION

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

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

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

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

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

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

2 METHODS

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

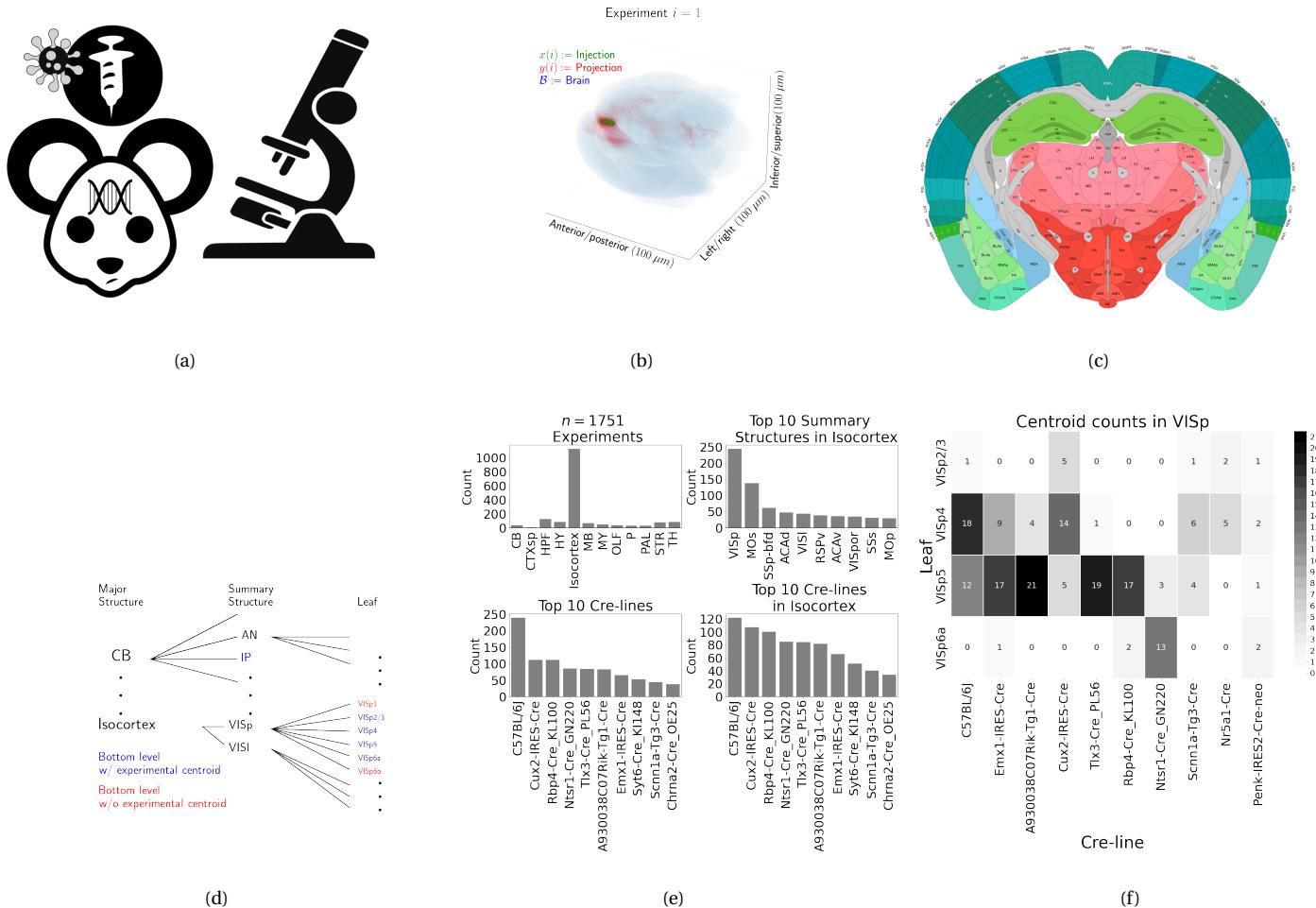


Figure 1: Experimental setting. 1a For each experiment, a Cre-dependent eGFP-expressing transgene cassette is transduced by stereotaxic injection into a Cre-driver mouse, followed by serial two-photon tomography imaging. 1b An example of the segmentation of projection (targets) and injection (source)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

⁸⁶ **Data**

⁸⁷ Our dataset \mathcal{D} consists of $n = 1751$ publicly available murine brain viral tracing experiments from the
⁸⁸ Allen Mouse Brain Connectivity Atlas. Figure 1a summarizes the experimental process used to
⁸⁹ generate this data. In each experiment, a mouse is injected with an adeno-associated virus (AAV)
⁹⁰ encoding green fluorescent protein (GFP) into a single location in the brain. Location of fluorescence
⁹¹ is mediated by the location of the injection, the characteristics of the transgene, and the genotype of
⁹² the mouse. In particular, Cre-driver or, equivalently, Cre-line mice are engineered to express Cre
⁹³ under the control of a specific and single gene promoter. This localizes expression of Cre to regions
⁹⁴ with certain transcriptomic cell-types signatures. In such Cre-driver mice, we used a double-inverted
⁹⁵ floxed AAV to produce eGFP fluorescence that depends on Cre expression in infected cells. To account
⁹⁶ for the complex cell-type targeting induced by a particular combination of Cre-driver genotype and
⁹⁷ GFP promoter, we refer to the combinations of cell-types targeted by a particular combination of AAV
⁹⁸ and Cre-driver mice as cell-classes. For example, we include experiments from Cre-driver lines that
⁹⁹ selectively label cell classes located in distinct cortical layers or other nuclei across the whole brain.
¹⁰⁰ For injections in the wild type mice, we used the Synapsin I promoter (Jackson, Dayton, Deverman, &
¹⁰¹ Klein, 2016; Kögler, Kilic, & Bähr, 2003). For injections into Cre mice, we used the CAG promoter with
¹⁰² a Flex cassette for Cre-mediated recombination control (Saunders, Johnson, & Sabatini, 2012).
¹⁰³ Additional details on are given in J. A. Harris et al. (2019).

¹⁰⁴ For each experiment, the fluorescent signal imaged after injection is aligned into the Allen
¹⁰⁵ Common Coordinate Framework (CCF) v3, a three-dimensional average template brain that is fully
¹⁰⁶ annotated with regional parcellations (Wang et al., 2020). The whole brain imaging and registration
¹⁰⁷ procedures described in detail in Kuan et al. (2015); Oh et al. (2014) produce quantitative metrics of
¹⁰⁸ fluorescence discretized at the $100 \mu\text{m}$ voxel level. Given an experiment, this image was histologically
¹⁰⁹ segmented by an analyst into *injection* and *projection* areas corresponding to areas containing somas,
¹¹⁰ dendrites and axons or exclusively axons of the transfected neurons. An example of a single
¹¹¹ experiment rendered in 3D is given in Figure 1b. Given an experiment i , we represent injections and
¹¹² projections as functions $x(i), y(i) : \mathcal{B} \rightarrow \mathbb{R}_{\geq 0}$, where $\mathcal{B} \subset [1 : 132] \times [1 : 80] \times [1 : 104]$ corresponds to the
¹¹³ subset of the $(1.32 \times 0.8 \times 1.04)$ cm rectangular space occupied by the standard voxelized mouse brain.

¹¹⁴ We also calculate injection centroids $c(i) \in \mathbb{R}^3$ and regionalized projections $y_{\mathcal{T}}(i) \in \mathbb{R}^T$ given by the
¹¹⁵ sum of $y(i)$ in each region. A description of these steps is in Supplemental Section 6.

¹¹⁶ Our goal is the estimation of **regionalized connectivity** from one region to another. A visual
¹¹⁷ depiction of this region parcellation for a two-dimensional slice of the brain is given in Figure 1c. All
¹¹⁸ structures annotated in the CCF belong to a hierarchically ordered ontology, with different areas of the
¹¹⁹ brain are parcellated to differing finer depths within a hierarchical tree. We denote the main levels of
¹²⁰ interest as major structures, summary structures, and layers. Not every summary structure has a layer
¹²¹ decomposition within this ontology, so we typically consider the finest possible regionalization - for
¹²² example, layer within the cortex, and summary structure within the thalamus, and denote these
¹²³ structures as leafs. As indicated in Figure 1d, the dataset used to generate the connectivity model
¹²⁴ reported in this paper contains certain combinations of region and cell class frequently, and others
¹²⁵ not at all. A summary of the most frequently assayed cell classes and structures is given in Figures 1e
¹²⁶ and 1f. Since users of the connectivity matrices may be interested in particular combinations, or
¹²⁷ interested in the amount of data used to generate a particular connectivity estimate, we present this
¹²⁸ information about all experiments in Supplemental Section 5.

129 ***Modeling Regionalized Connectivity***

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

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

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

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

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

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

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

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

¹⁴¹ available so that its layer-specific behavior is visible, and normalize our data by projection signal in
¹⁴² order to account for differences between cell class. Additional model selection results are given in
¹⁴³ Supplemental Section 5 for alternative normalization strategies, and more detail on estimation is
¹⁴⁴ given in Supplemental Section 6.

Name	f^*	\hat{f}	f_*	$\mathcal{D}(v, s)$
NNLS (Oh et al., 2014)	$\hat{f}(1_s)$	NNLS(X,Y)	$X = x_{\mathcal{S}}, Y = y_{\mathcal{T}}$	I_m/I_m
NW (Knox et al., 2019)	$\sum_{l_s \in s} \hat{f}(l_s)$	NW(X,Y)	$X = l_s, Y = y_{\mathcal{T}}$	I_m/I_m
Cre-NW	$\sum_{l_s \in s} \hat{f}(l_s)$	NW(X,Y)	$X = l_s, Y = y_{\mathcal{T}}$	$(I_l \cap I_v)/I_m$
Expected Loss (EL)	$\sum_{l_s \in s} \hat{f}(s)$	EL(X, Y, v)	$X = l_s, Y = y_{\mathcal{T}}, v$	I_l/I_m

Table 1: Estimation of \mathcal{C} using connectivity data. The regionalization, estimation, and featurization steps are denoted by f^* , \hat{f} , and f_* , respectively. The training data used to fit the model is given by index set I . We denote experiments with centroids in particular major brain divisions and leafs as I_m and I_l , respectively. Data I_l/I_m means that, given a location $l_s \in s \in m$, the model \hat{f} is trained on all of I_m , but only uses I_l for prediction. The non-negative least squares estimator (NNLS) fits a linear model that predicts regionalized projection signal $y_{\mathcal{T}}$ as a function of regionalized injection signal $x_{\mathcal{S}}$. Thus, the regionalization step for a region s is given by applying the learned matrix \hat{f} to the s -th indicator vector. In contrast, the Nadaraya-Watson model (NW) is a local smoothing model that generates a prediction for each voxel within the source structure that are then averaged to create estimate the structure-specific connectivity.

¹⁴⁵ Our contributions - the Cre-NW and Expected Loss (EL) models - have several differences from the
¹⁴⁶ previous methods. In contrast to the non-negative least squares (Oh et al., 2014) and
¹⁴⁷ Nadaraya-Watson (Knox et al., 2019) estimators that account only for source region s , our new
¹⁴⁸ estimators account cell class v , The Cre-NW estimator only uses experiments from a particular class
¹⁴⁹ to predict connectivity for that class, while the EL estimator shares information between classes
¹⁵⁰ within a structure. Both of these estimator take into account both the cell-class and the centroid
¹⁵¹ position of the experimental injection. Like the NW and Cre-NW estimator, the EL estimator generates

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

160 ***Model evaluation***

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

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

We use weighted l_2 -loss to evaluate these predictions.

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

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

180 ***Connectivity analyses***

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

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

199 Our application of NMF to decompose the estimated long-range connectivity is some independent
 200 interest, since we ignore connections between source and target regions less than $1500 \mu m$ apart. This
 201 is because short-range projections resulting from diffusion and traveling fibers dominate the matrices
 202 $\hat{\mathcal{C}}$. Our NMF algorithm solves the following optimization problem

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

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

²⁰⁶ up with a stability analysis via clustering the resultant H from multiple replicates. The medians of the
²⁰⁷ component clusters appearing frequently across NMF replicates are selected as **connectivity**
²⁰⁸ **archetypes**. Details of these approaches are given in Supplementary Sections 6 and 7.

3 RESULTS

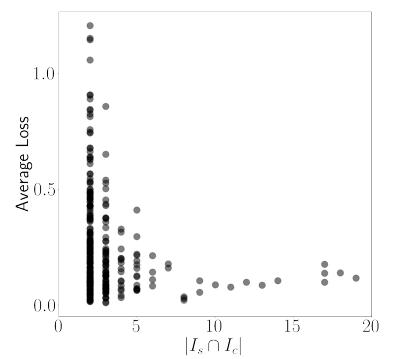
209 The main result of this paper is the creation of cell-type specific connectivity estimates from the Allen
210 Mouse Brain Connectivity Atlas (MCA) experiments. We first establish that our new expected-loss (EL)
211 estimator performs best in validation assays for estimating wild-type and cell-type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

216 ***Model evaluation***

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

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

(a)



(b)

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

225 *Connectivities*

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

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

235 (2014). However, the layer-specific targeting of the different Cre-lines enables our estimated wild-type
236 connectivities to display heterogeneity at the layer level. This contrasts the model in Knox et al.
237 (2019), which is denominated as the NW Major-WT model whose accuracy is evaluated in Table 2a.
238 For comparison, we also plot averages over component layers weighted by layer size projections
239 between summary-structure sources and targets in the cortex in Figure 2b. Importantly, as shown in
240 Table 2a this finer spatial resolution corresponds to the increased accuracy of our EL model over the
241 NW Major-WT model.

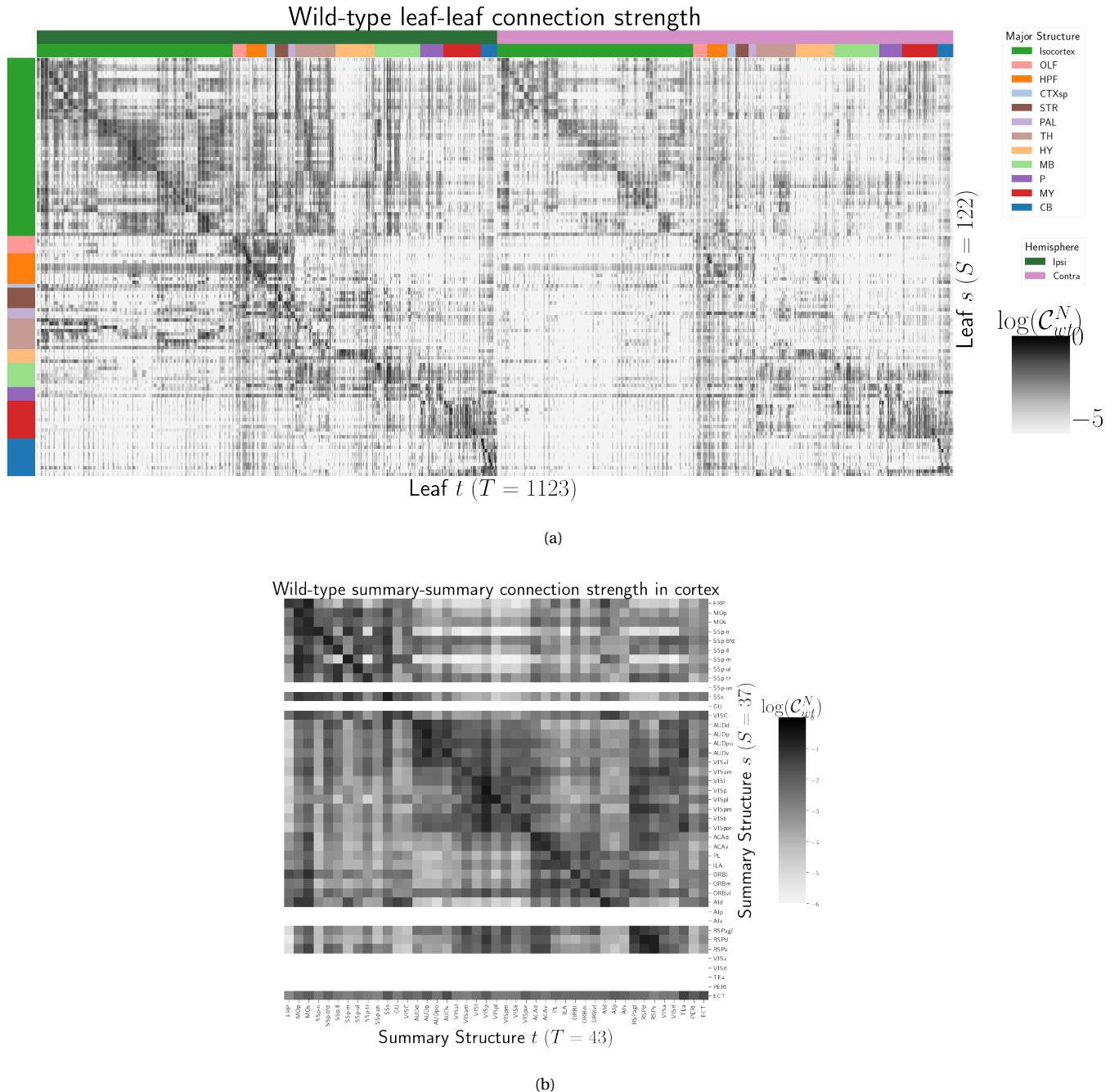


Figure 2: Wild-type connectivities. 2a Log wild-type leaf-to-leaf connectivity matrix $\log \mathcal{C}(s, t, v_{wt})$. Sources and target major brain division structure are shown. 2b Log wild-type intracortical connectivity matrix at the summary structure level. Summary structures without an injection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

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

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

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

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

284 DR PROJECTIONS IN THE SLC6A4-CRE'ET33 MODEL. Serotonergic projections from cells in the
 285 dorsal raphe (DR) are widely distributed and innervate many forebrain structures including isocortex
 286 and amygdala. In the Allen MCA, serotonergic neurons were labeled using Slc6a4-Cre_ET33 and
 287 Slc6a4-CreERT2_EZ13 mice. This small nucleus appears to contain a complex mix of molecularly
 288 distinct serotonergic neuron subtypes with some hints of subtype-specific projection patterns (Huang
 289 et al., 2019; Ren et al., 2018, 2019). We expect that the Cre-lines used here in the Allen MCA, which use
 290 the serotonin transporter promoter (Slc6a4-Cre and -CreERT2), will lead to expression of tracer in all
 291 the serotonergic subtypes recently described in an unbiased way, but this assumption has not been
 292 tested directly. We compared our model data to a single cell reconstruction dataset consisting of n=50
 293 serotonergic cells with somas in the DR that also had bulk tracer validation 3c. After processing our
 294 data to match the target structure ontology level across studies, we identified 37 targets from the DR
 295 with weights above the 90th percentile, whereas Ren et al. (2019) listed 55 targets across the single cell
 296 reconstructions. Twenty seven of these targets where shared.

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

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

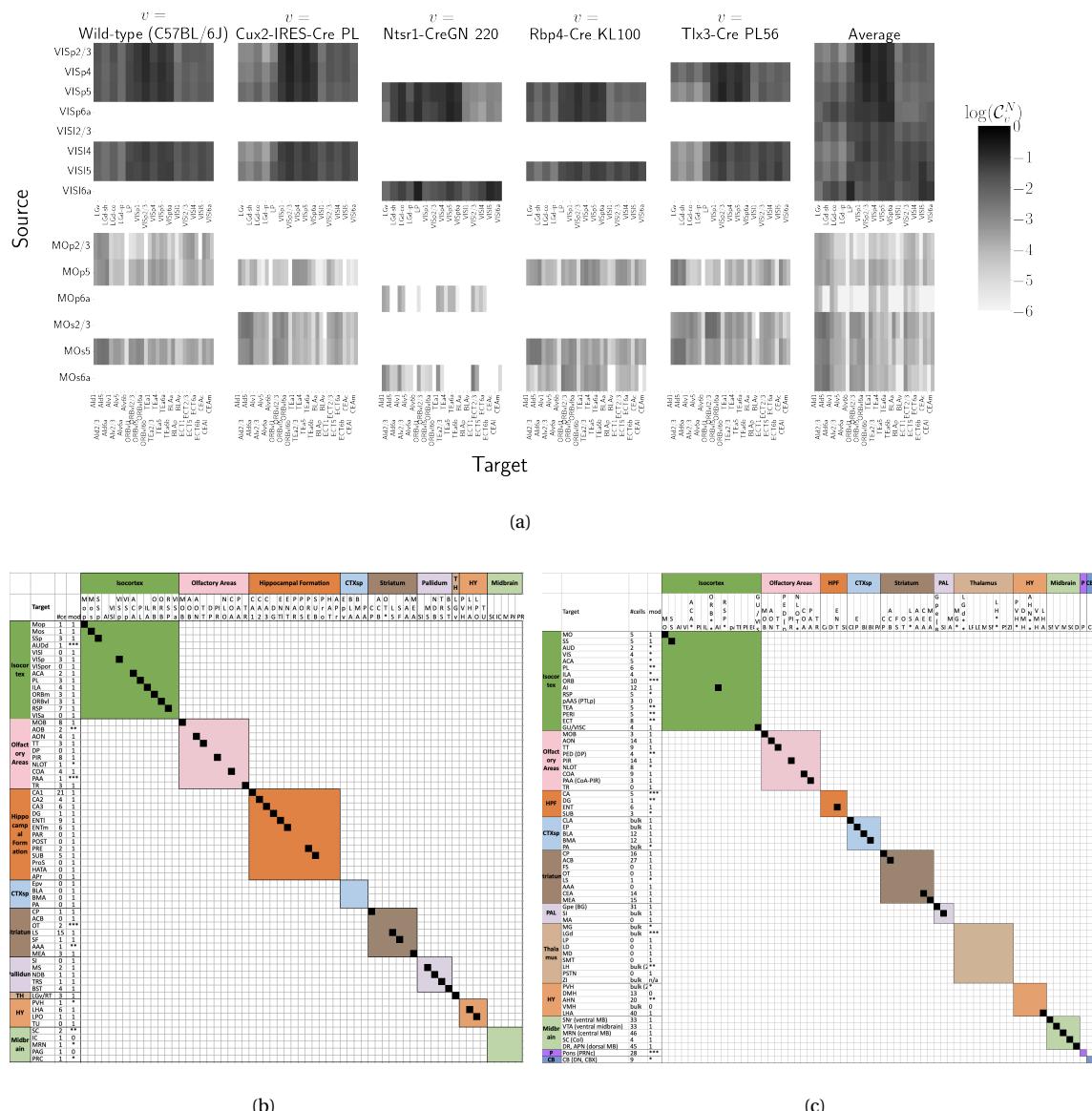


Figure 3: Cell-class specificity. 3a Selected cell-class and layer specific connectivities from two visual and two motor areas. Sources without an injection of a Cre-line are not estimated due to lack of data. 3b Targets reported for cholinergic cells in MS/NDB in Li et al. (2018) and above the 90th percentile in our Chat-IRES-Cre-neo model. A black box along the diagonal indicates the target was identified in both studies ($n = 35$). The number of single cells (out of 50) with a projection to each target is shown in the first column after the target acronym and the next column shows whether a target appeared in the 90% thresholded model weights (1= present, 0=absent). Some of these targets only appeared at lower thresholds as indicated by asterisks, * * * > 85th%, ** > 75th%, * > 50th%. 3c Targets reported for serotonergic cells in DR in Ren et al. (2018, 2019) and above the 90th percentile in our Slc6a4-Cre_ET33 model.

314 Connectivity Analyses

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

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

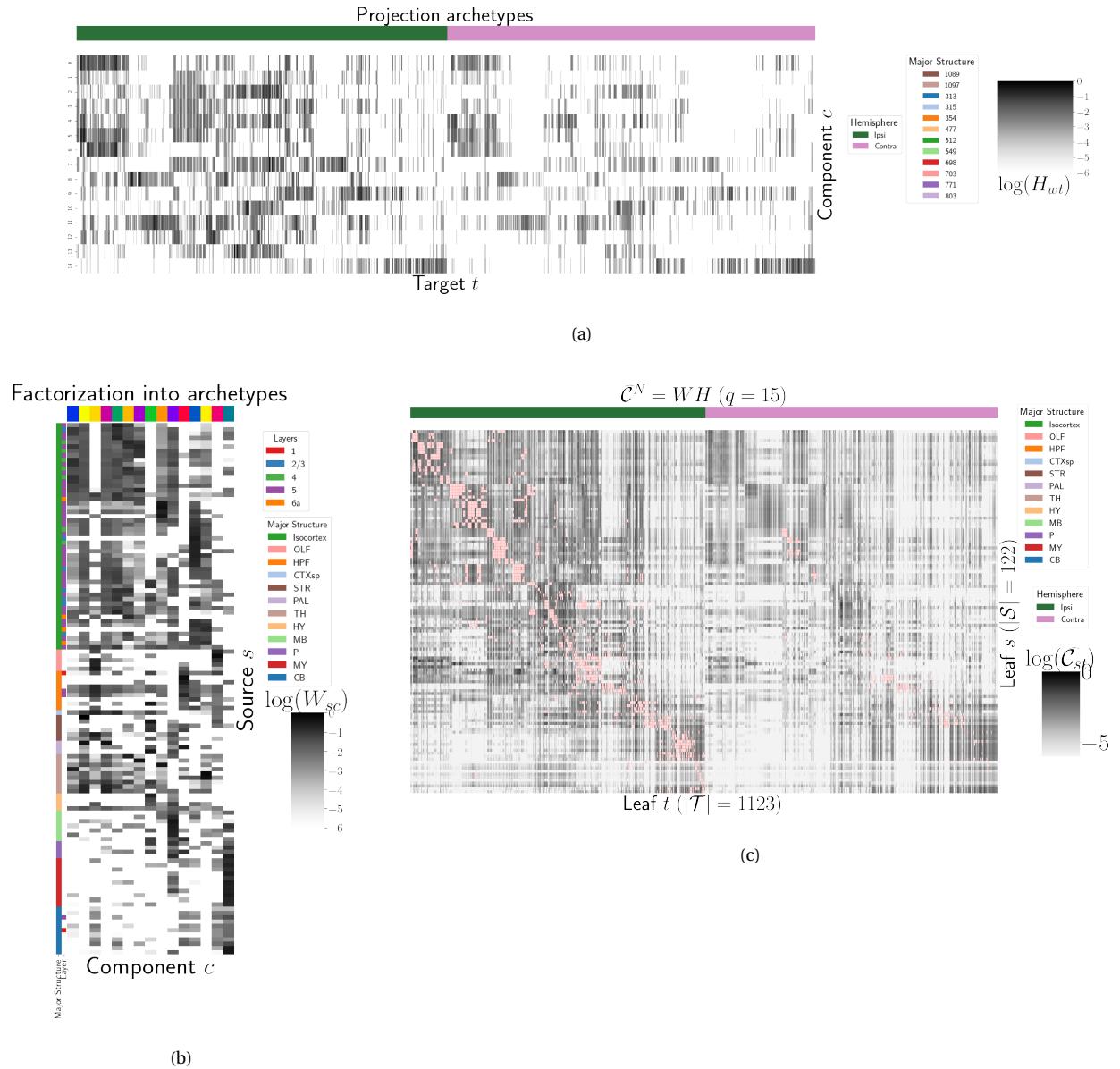


Figure 4: Non-negative matrix factorization results $\mathcal{C}_{wt}^N = WH$ for $q = 15$ components. 4a Latent space coordinates H of \mathcal{C} . Target major structure and hemisphere are plotted. 4b Loading matrix W . Source major structure and layer are plotted. 4c Reconstruction of the normalized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

4 DISCUSSION

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

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

³⁶⁶ similarity per structure could naturally combine both the physical simplicity of the centroid model
³⁶⁷ while also incorporating the full distribution of the injection signal.

³⁶⁸ The factorization of the connectivity matrix could also be improved. Non-linear data
³⁶⁹ transformations or matrix decompositions, or tensor factorizations that account for correlations
³⁷⁰ between cell-types could better capture the true nature of latent neural connections (K. D. Harris et
³⁷¹ al., 2016). Cre- or layer-specific signal recovery as performed here could be used to evaluate a range of
³⁷² matrix decompositions. This could help for example to understand the influence of traveling fibers on
³⁷³ the observed connectivity (Llano & Sherman, 2008). From a statistical perspective, stability-based
³⁷⁴ method for establishing archetypal connectivities in NMF is similar to those applied to genomic data
³⁷⁵ Kotliar et al. (2019); Wu et al. (2016). Regardless of statistical approach, as in genomics, latent
³⁷⁶ low-dimensional organization in connectivity should inspire search for similarly parsimonious
³⁷⁷ biological correlates.

ACKNOWLEDGMENTS

³⁷⁸ We thank the Allen Institute for Brain Science founder, Paul G. Allen, for his vision, encouragement,
³⁷⁹ and support.