

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

5 SUPPLEMENTAL INFORMATION

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

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

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

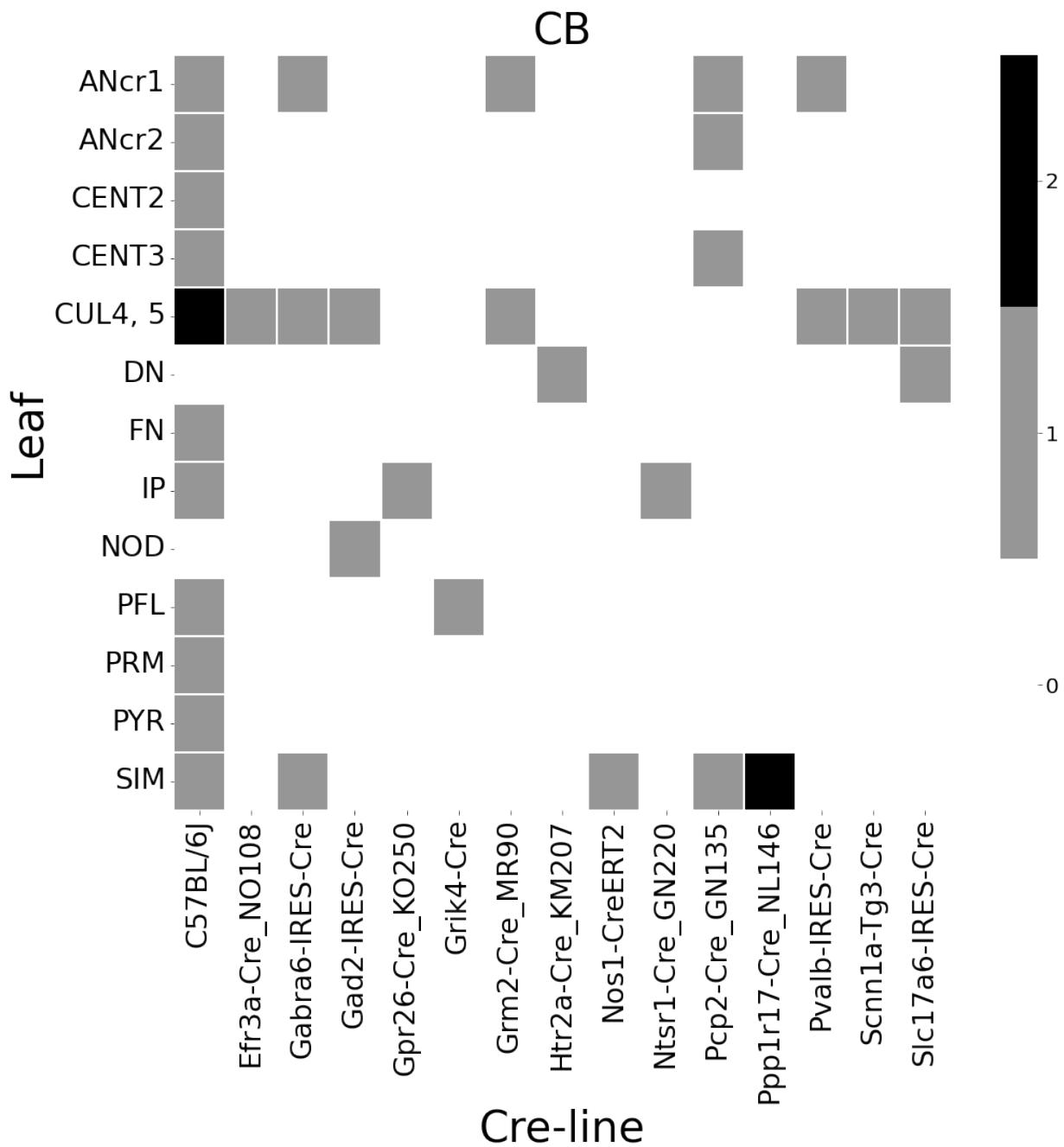


Figure 5: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

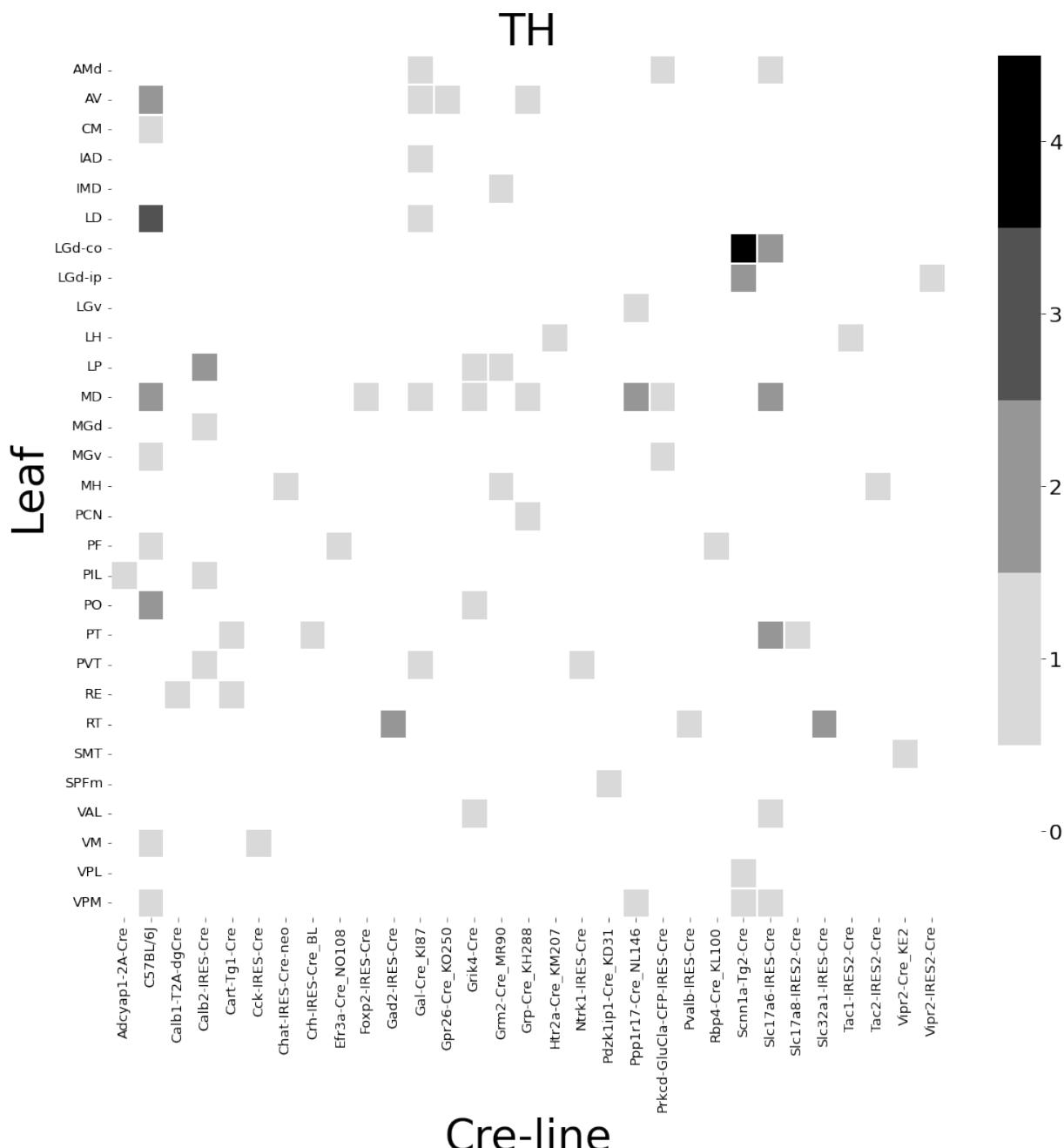


Figure 6: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

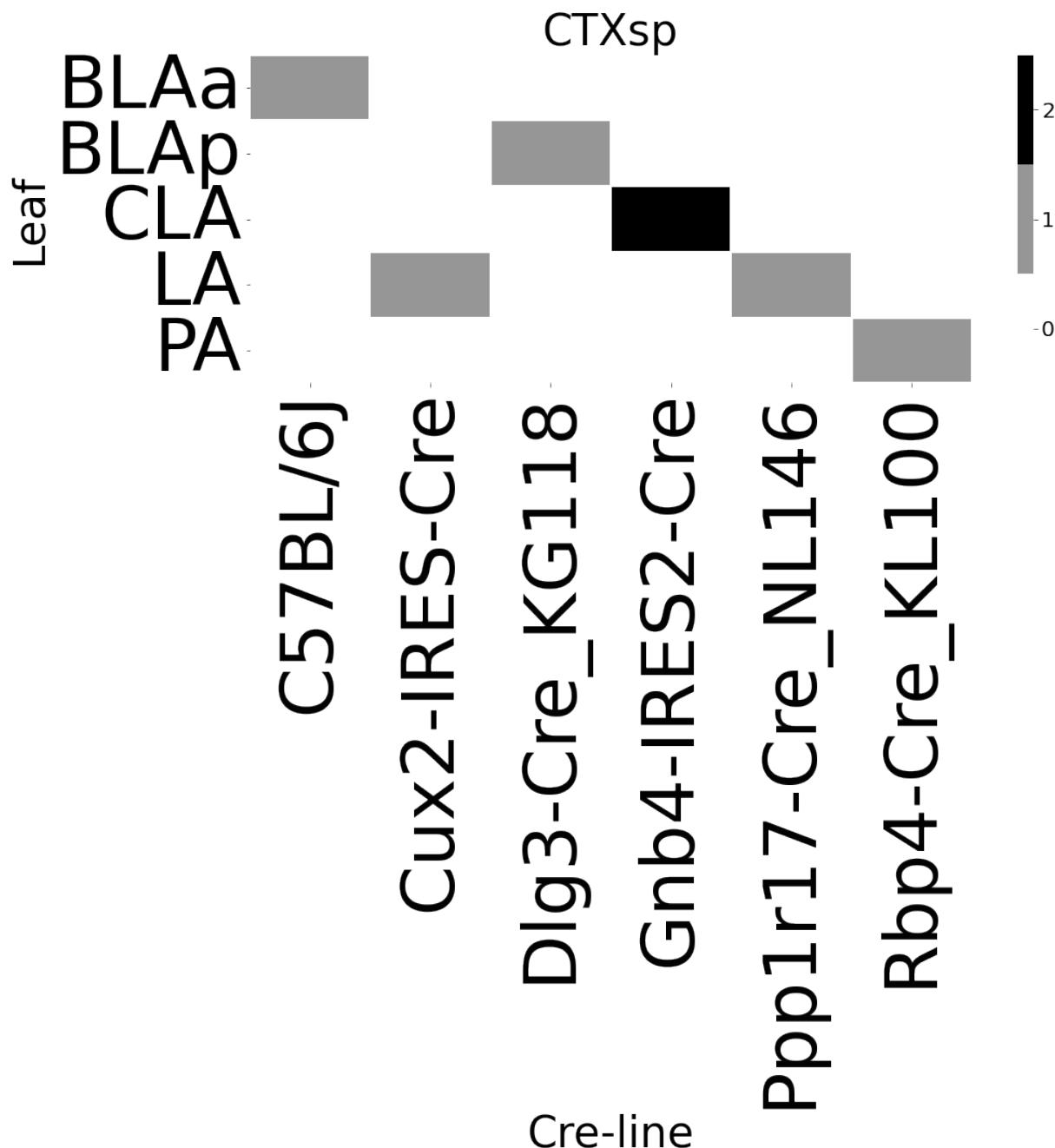


Figure 7: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

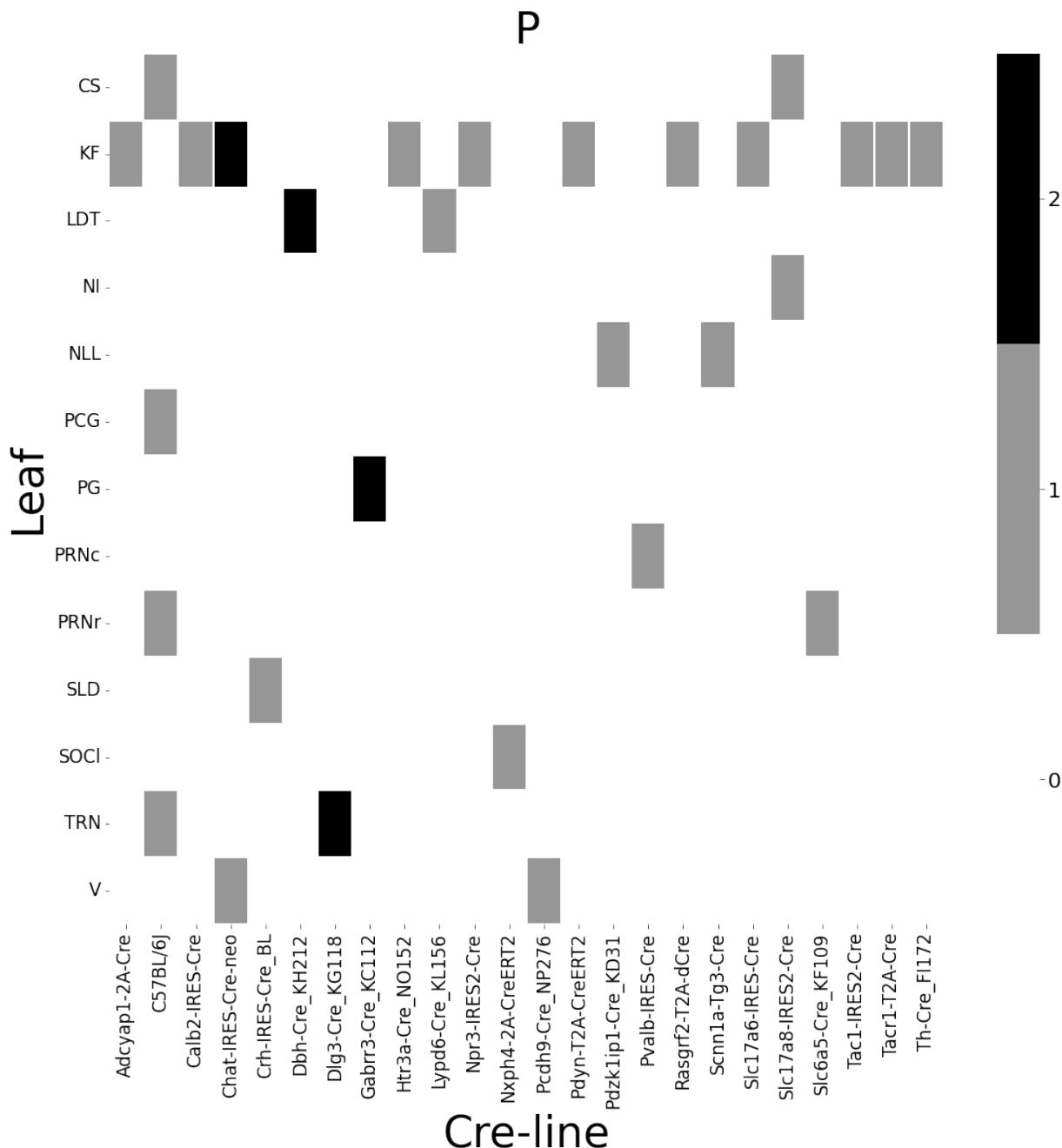


Figure 8: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

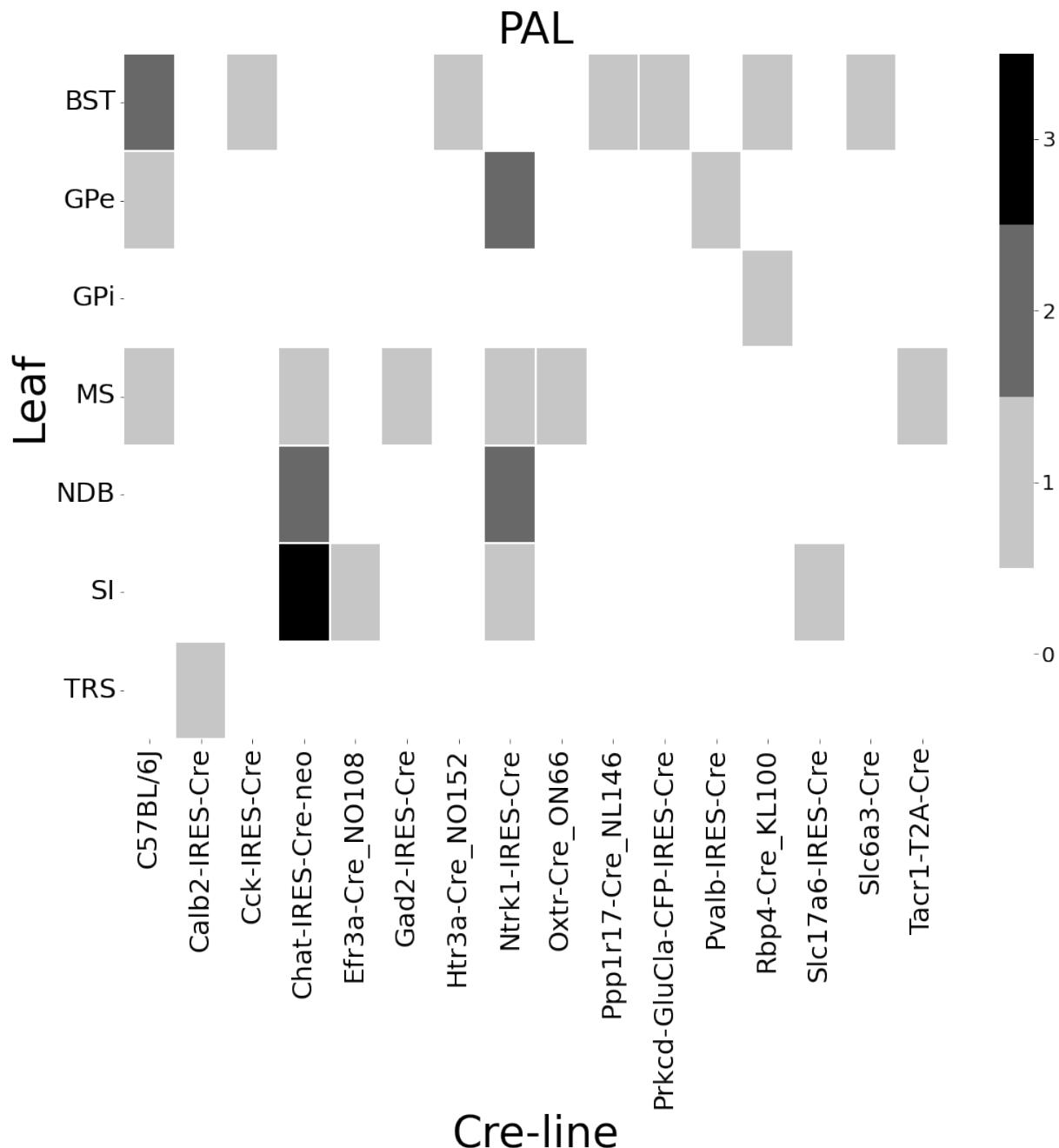


Figure 9: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

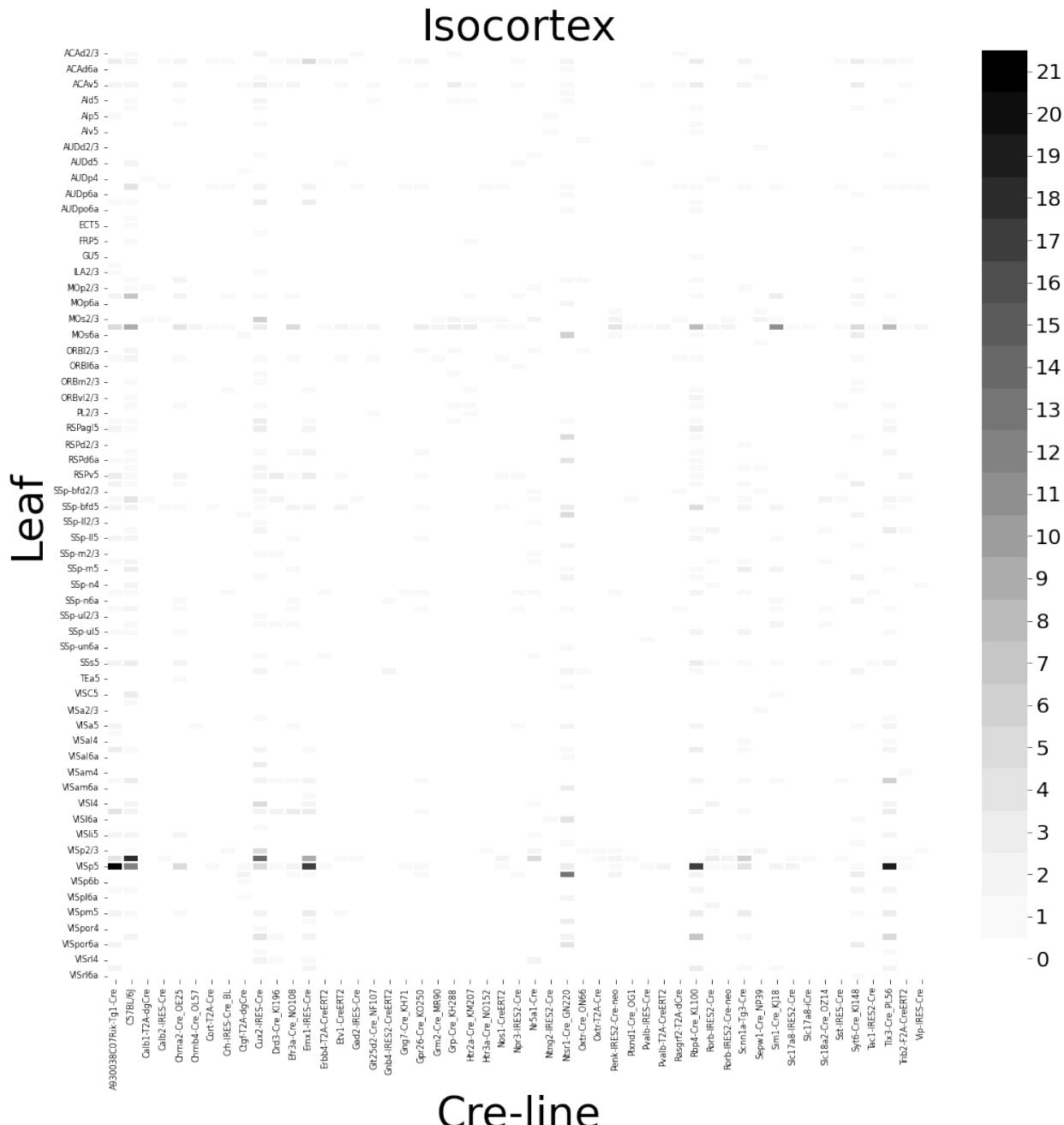


Figure 10: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

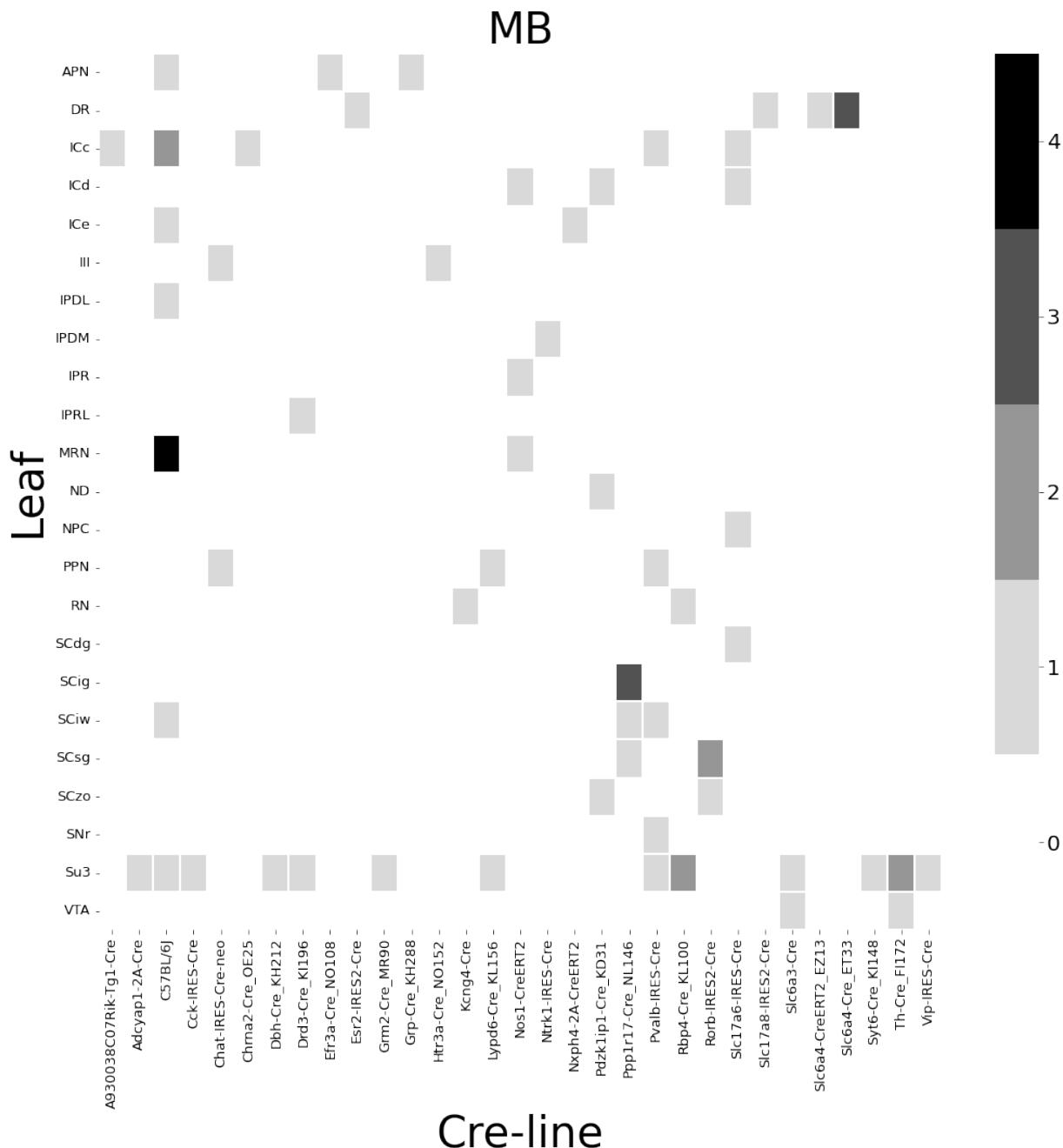


Figure 11: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

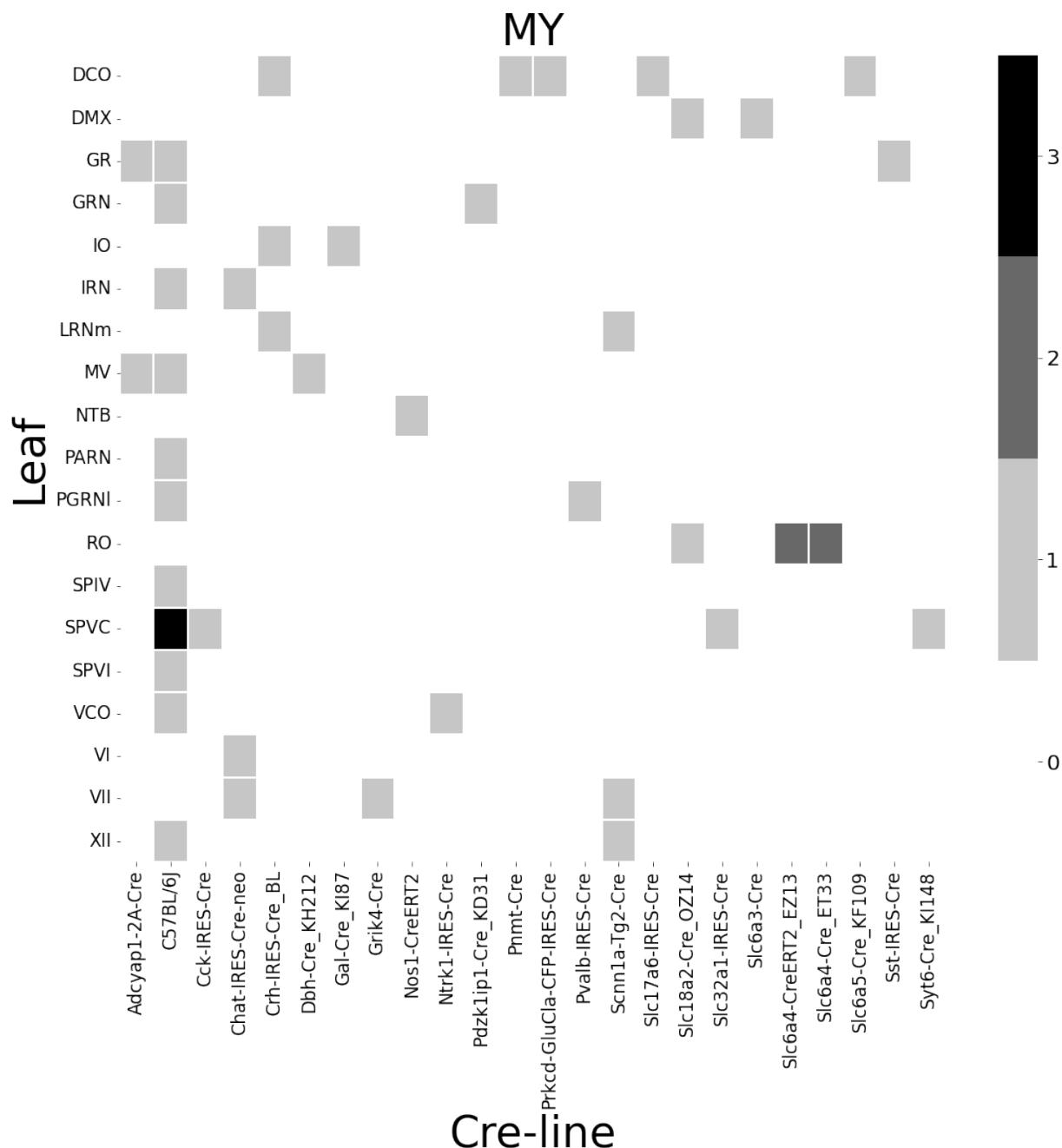


Figure 12: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

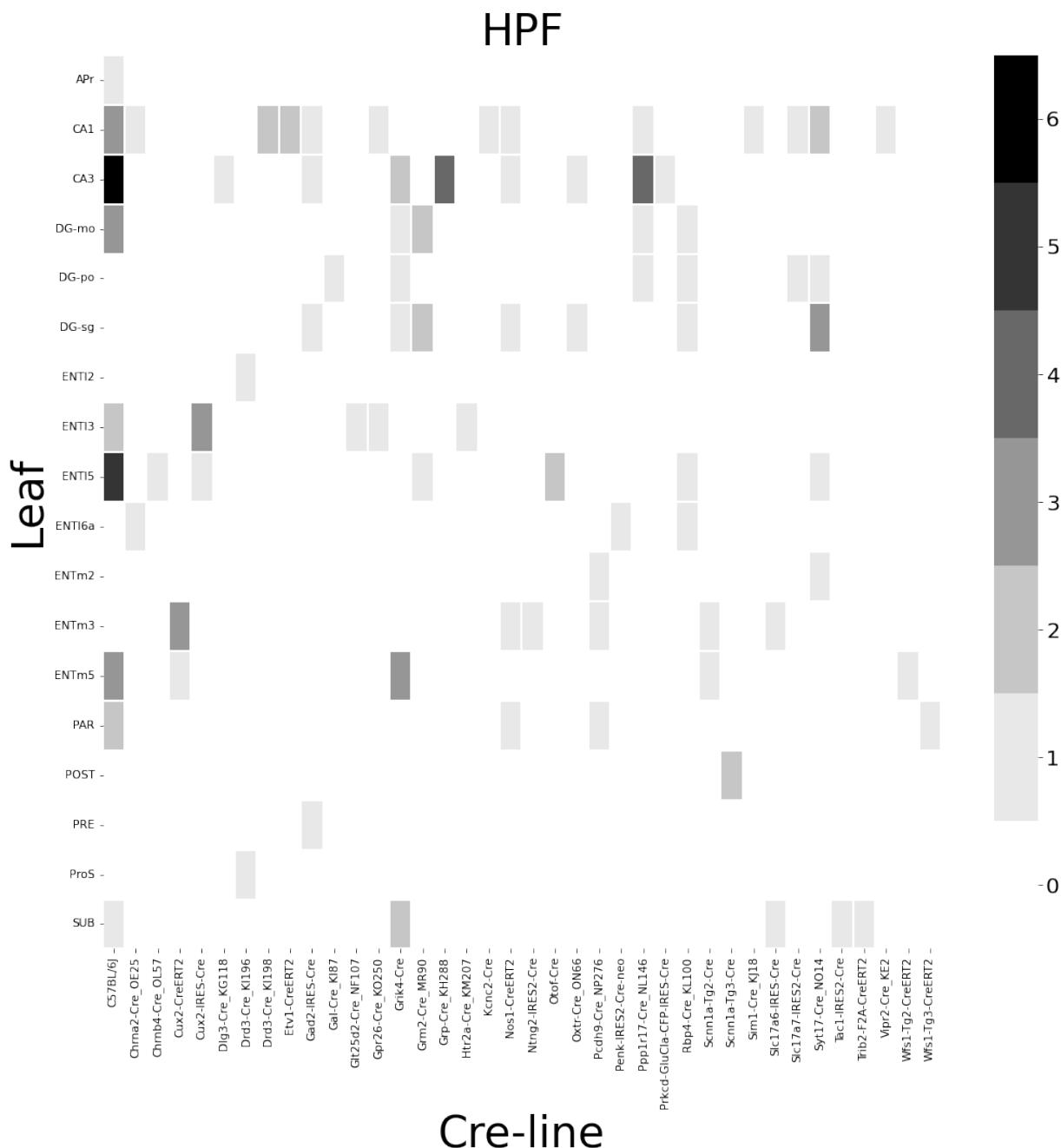


Figure 13: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

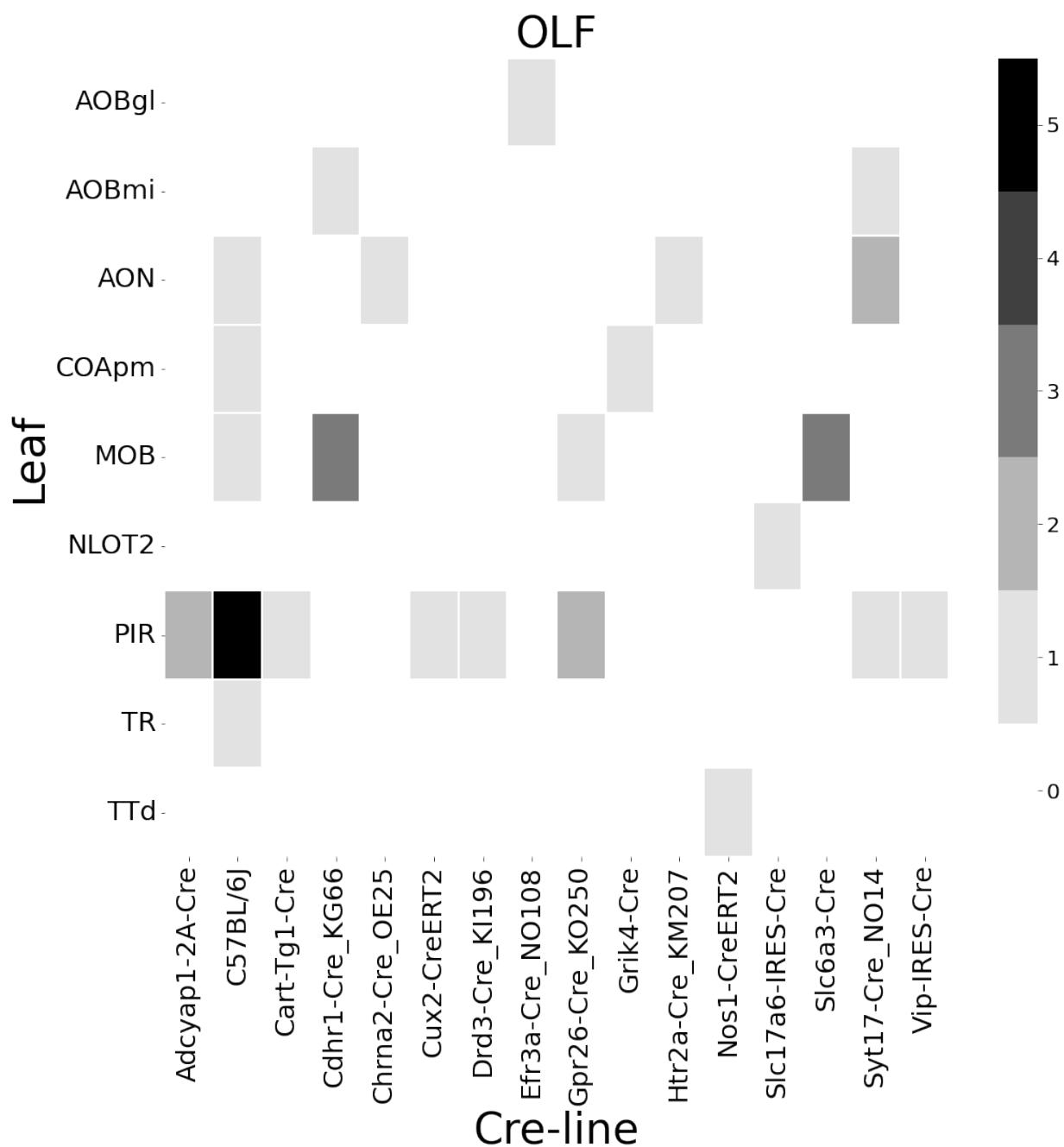


Figure 14: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

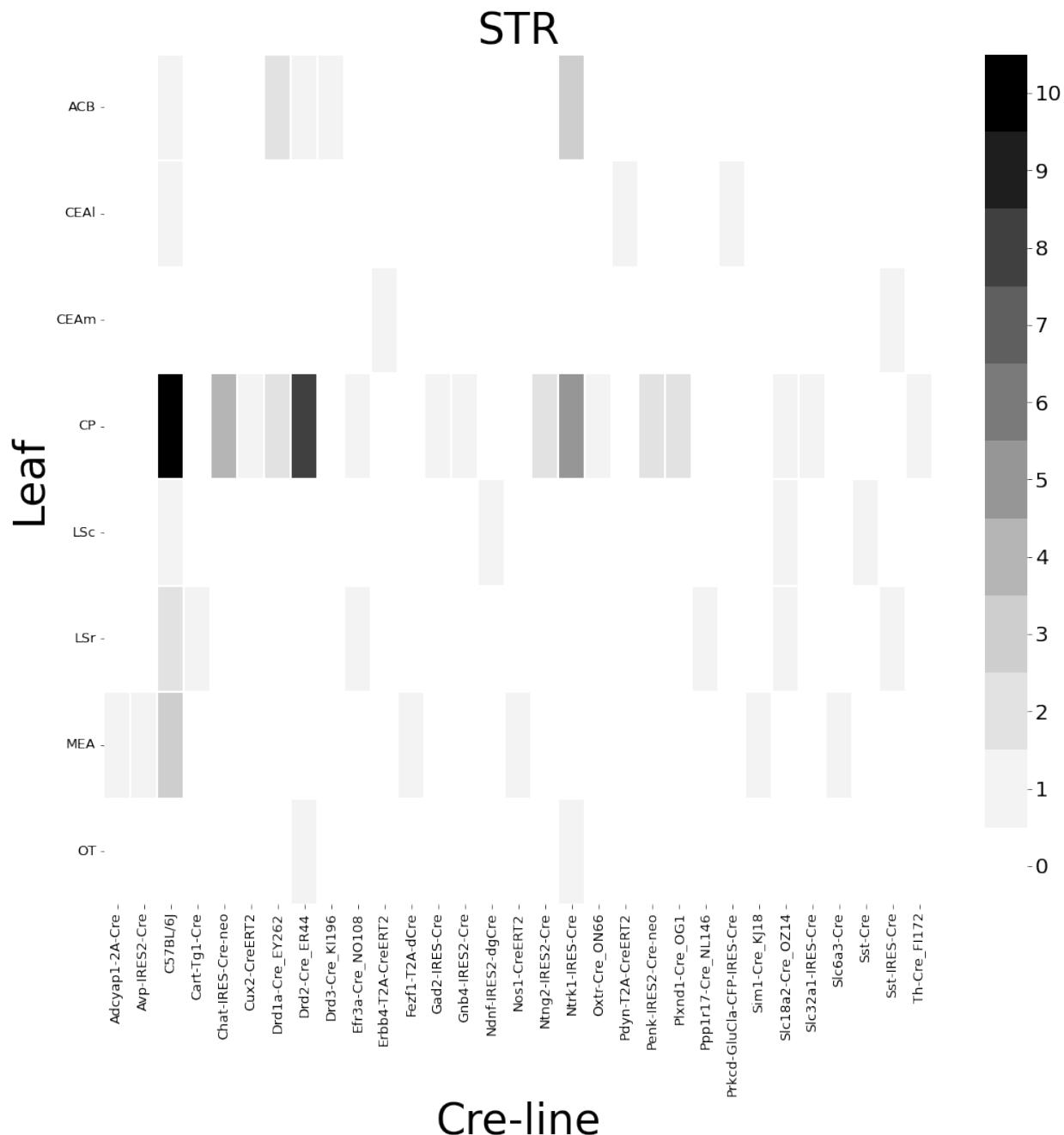


Figure 15: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

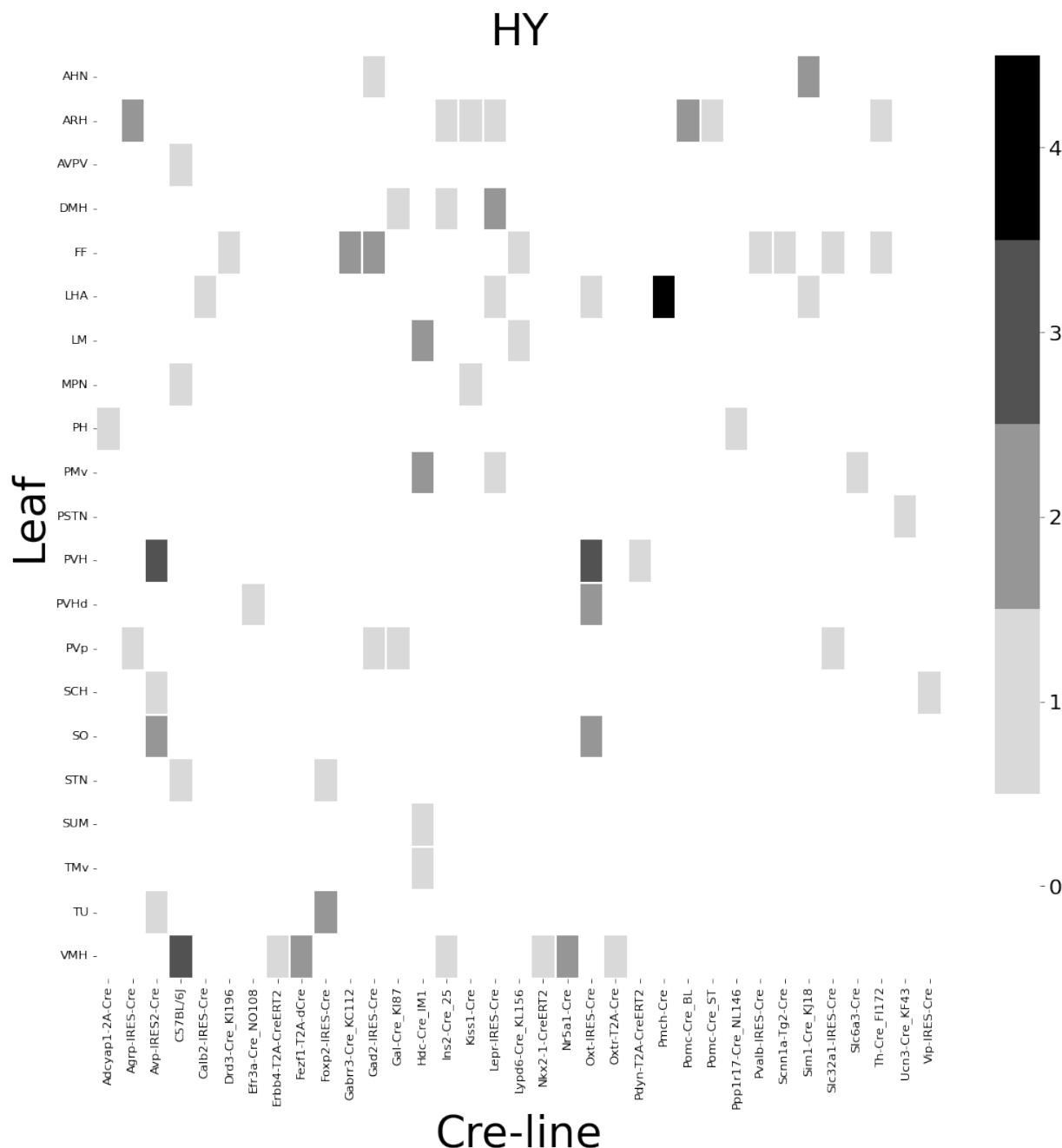


Figure 16: Frequencies of Cre-line and leaf-centroid combinations in our dataset.

401 ***Distances between structures***

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

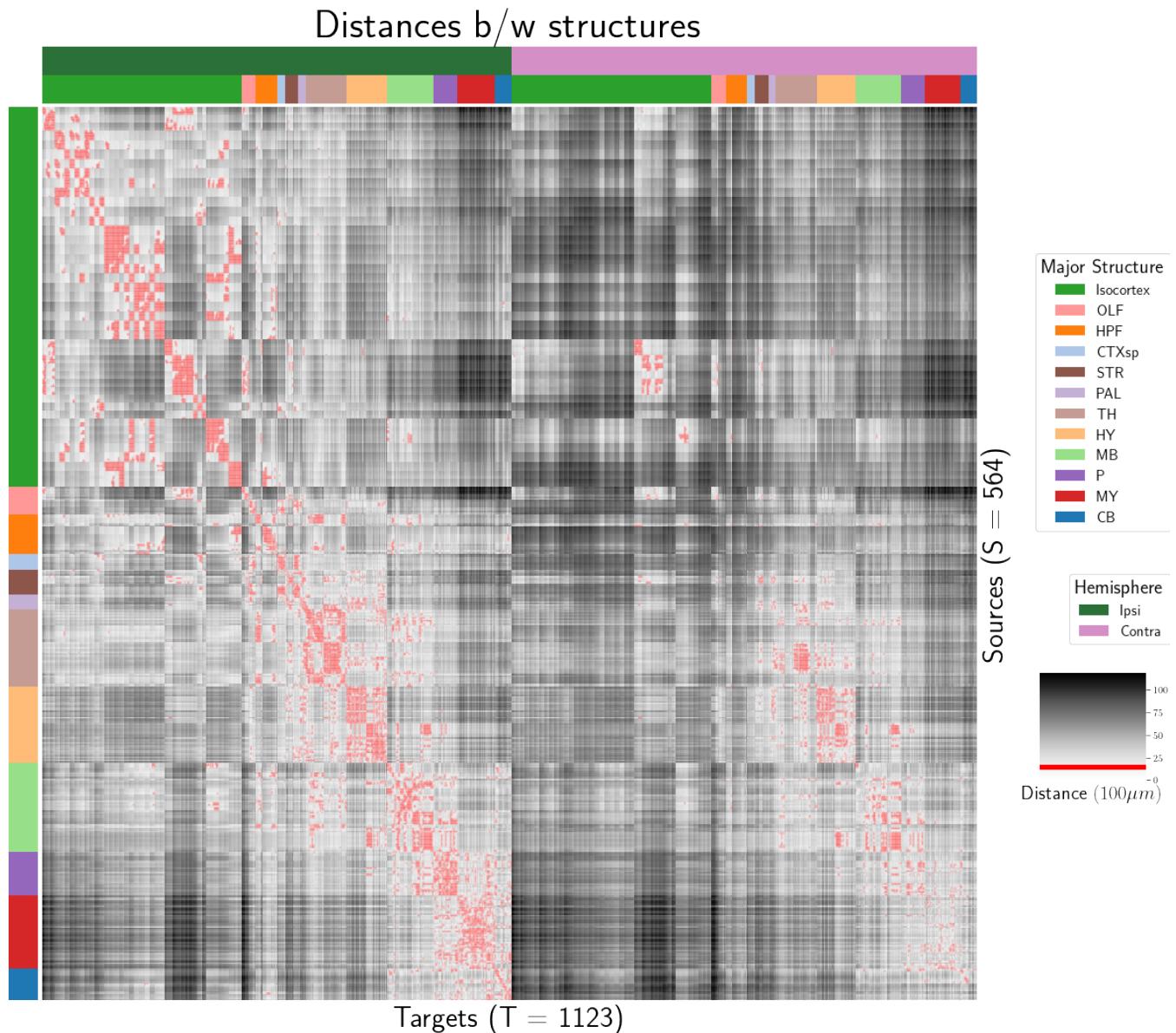


Figure 17: Distance between structures. Short-range connections are masked in red.

405 ***Model evaluation***

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

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

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

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

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

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

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

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

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

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

6 SUPPLEMENTAL METHODS

⁴²² This section consists of additional information on preprocessing of the neural connectivity data,
⁴²³ estimation of connectivity, and matrix factorization.

⁴²⁴ ***Data preprocessing***

⁴²⁵ Several data preprocessing steps take place prior to evaluations of the connectivity matrices. These
⁴²⁶ steps are described in Algorithm PREPROCESS. The arguments of this normalization process - injection
⁴²⁷ signals $x(i)$, projection signals $y(i)$, injection fraction $F(i)$, and data quality mask $q(i)$ - were
⁴²⁸ downloaded using the Allen SDK, a programmatic interface to the brain connectivity data. The
⁴²⁹ injections and projection signals $\mathcal{B} \rightarrow [0, 1]$ were segmented manually in histological analysis. The
⁴³⁰ projection signal gives the proportion of pixels within the voxel displaying fluorescence, and the
⁴³¹ injection signal gives the proportion of pixels within the histologically-selected injection subset
⁴³² displaying fluorescence. The injection fraction $F(i) : \mathcal{B} \rightarrow [0, 1]$ gives the proportion of pixels within
⁴³³ each voxel in the injection subset. Finally, the data quality mask $q(i) : \mathcal{B} \rightarrow \{0, 1\}$ gives the voxels that
⁴³⁴ have valid data.

⁴³⁵ Our preprocessing makes use of the above ingredients, as well as several other essential steps. First,
⁴³⁶ we compute the weighted injection centroid

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

⁴³⁷ where $x(i)|_l$ is the injection density at location $l \in \mathbb{R}^3$. Given a regionalization \mathcal{R} from the Allen SDK,
⁴³⁸ we can also access regionalization map $R : \mathcal{B} \rightarrow \mathcal{R}$. This induces a functional of connectivities from
⁴³⁹ the space of maps $\{\mathcal{X} = x : \mathcal{B} \rightarrow [0, 1]\}$

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

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

⁴⁴⁰ We also can restrict a signal to a individual structure as

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

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

⁴⁴¹ Finally, given a vector or array $a \in \mathbb{R}^T$, we have the $l1$ normalization map

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

⁴⁴² Denote m as the major structure containing an experiment, and define \odot for maps $\mathcal{B} \rightarrow [0, 1]$ by e.g.

⁴⁴³ $(y(i) \odot q(i))|_l := (y(i)|_l)(q(i)|_l)$. We then can write the preprocessing algorithm.

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

Injection fraction $x_F \leftarrow x \odot F$

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

Restrict injection $x_m = 1|_m x_q$.

Compute centroid c from x_m

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

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

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

444 Estimators

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

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

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

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

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

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

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

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

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

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

⁴⁶⁴ *The Expected-Loss estimator* Besides location of the injection centroid, cell class also influences
⁴⁶⁵ projection. Thus, we introduce method for estimating the effect of Cre-distance, which we define as
⁴⁶⁶ the distance between the projections of the mean experiment of one (Cre,leaf) pair with another.
⁴⁶⁷ Equivalently, relatively small Cre-distance defines what we call similar cell classes. This method
⁴⁶⁸ assigns a predictive weight to each pair of training points that depends both on their
⁴⁶⁹ centroid-distance and Cre-distance. This weight is determined by the expected prediction error of
⁴⁷⁰ each of the two feature types

⁴⁷¹ We define Cre-line behavior as the average regionalized projection of a Cre-line in a given structure
⁴⁷² (i.e. leaf). The vectorization of categorical information is known as **target encoding**

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

⁴⁷³ We then define a **Cre-distance** in a leaf to be the distance between the target-encoded projections of
⁴⁷⁴ two Cre-lines. The relative predictive accuracy of Cre-distance and centroid distance is determined by
⁴⁷⁵ fitting a surface of projection distance as a function of Cre-distance and centroid distance. For this
⁴⁷⁶ reason, we call this the Expected Loss Estimator. When we use shape-constrained B-splines to
⁴⁷⁷ estimate this weight, the weights then may be said to be used in a Nadaraya-Watson estimator. The
⁴⁷⁸ resulting weights are then utilized in a Nadaraya-Watson estimator in a final prediction step.

⁴⁷⁹ In mathematical terms, our full feature set consists of the centroid coordinates and the
⁴⁸⁰ target-encoded means of the combinations of virus type and injection-centroid structure. That is,

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

⁴⁸¹ f^* is defined as in (2). The expected loss estimator is then

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

⁴⁸² where

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

⁴⁸³ and s is the structure containing l .

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

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

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

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

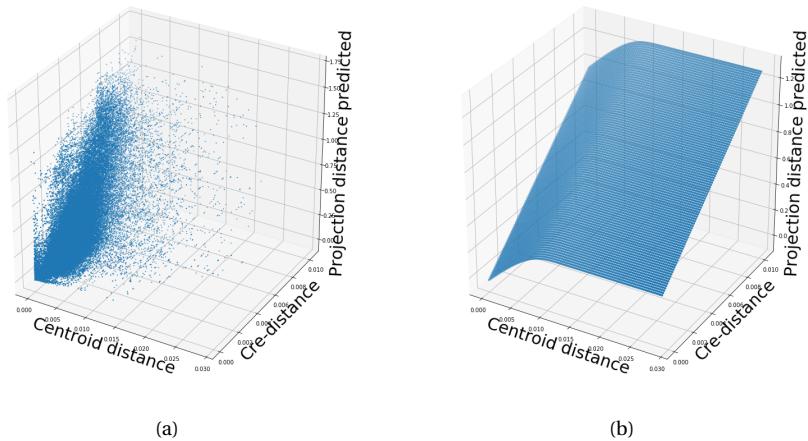


Figure 18: Fitting g . 18a Distribution of projection errors against centroid distance and cre-distance in Isocortex. 18b estimated \hat{g} using B-splines. Projection distance is $\|y_{\mathcal{T}}(i) - y_{\mathcal{T}}(i')\|_2$, Cre-distance is $\|\bar{y}_{\mathcal{T},s,v} - \bar{y}_{\mathcal{T},s,v}(i)\|_2$,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

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

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

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

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

521 ***Setting a lower detection threshold***

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

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

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

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

532 ***Decomposing the connectivity matrix***

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

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

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

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

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

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

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

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

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

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

We therefore generate uniformly random masks $1_{M(p)} \in \mathbb{R}^{S \times T}$ where

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

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

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

where $1_{M(p)_r^c}$ is the binary complement of $1_{M(p)_r}$ and R is a number of replicates. Theoretically, the optimum number of components is then

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

566 *Stabilizing NMF* The NMF program is non-convex, and, empirically, individual replicates will not
 567 converge to the same optima. One solution therefore is to run multiple replicates of the NMF
 568 algorithm and cluster the resulting vectors. This approach raises the questions of how many clusters
 569 to use, and how to deal with stochasticity in the clustering algorithm itself. We address this issue
 570 through the notion of clustering stability (von Luxburg, 2010a).

The clustering stability approach is to generate L replicas of k-cluster partitions $\{C_{kl} : l \in 1 \dots L\}$ and then compute the average dissimilarity between clusterings

$$\xi_k = \frac{2}{L(L-1)} \sum_{l=1}^L \sum_{l'=1}^l d(C_{kl}, C_{kl'}).$$

Then, the optimum number of clusters is

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

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

7 SUPPLEMENTAL EXPERIMENTS

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

583 *Setting detection threshold τ*

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

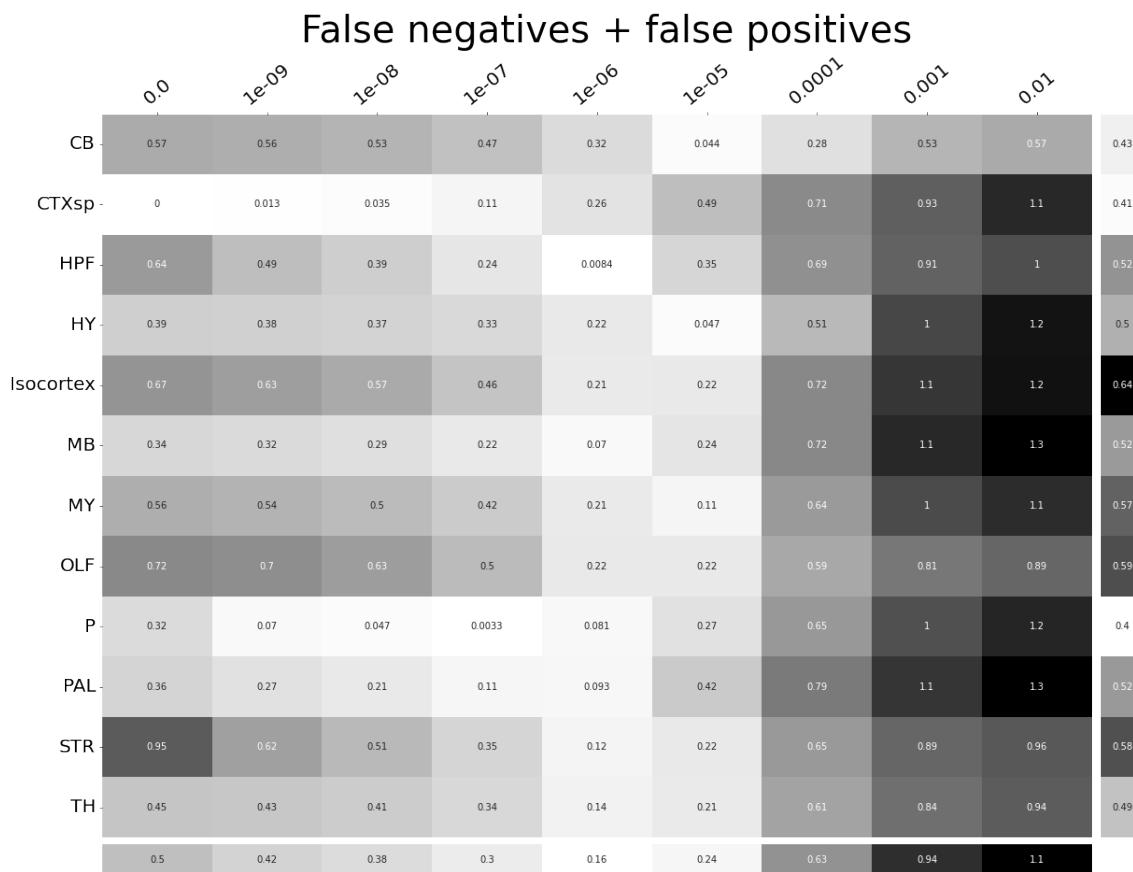


Figure 19: τ at different limits of detection in different major structures. 10^{-6} is the optimal detection threshold.

586 ***Loss subsets***

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

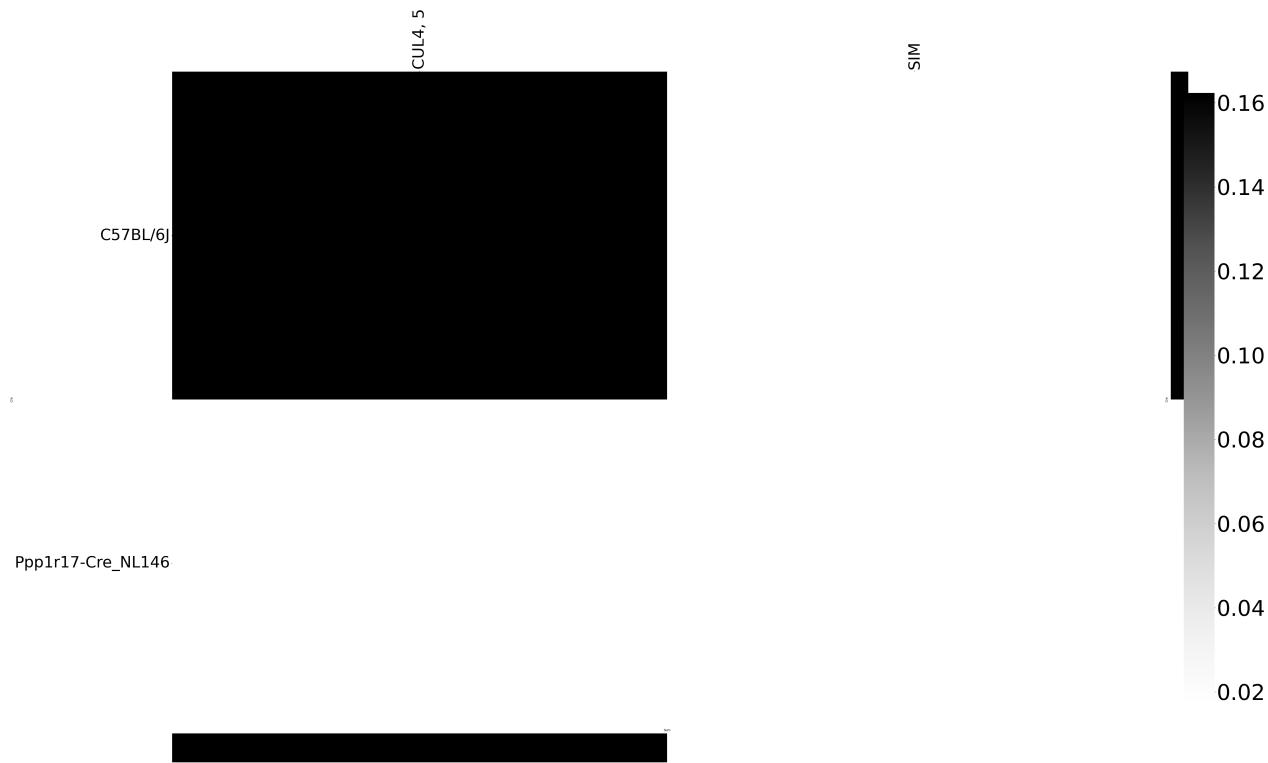


Figure 20: Weighted loss for Cre-leaf combinations in CB. Missing values are omitted. For example, this figure has one present and three missing values. Row and column averages are also plotted.

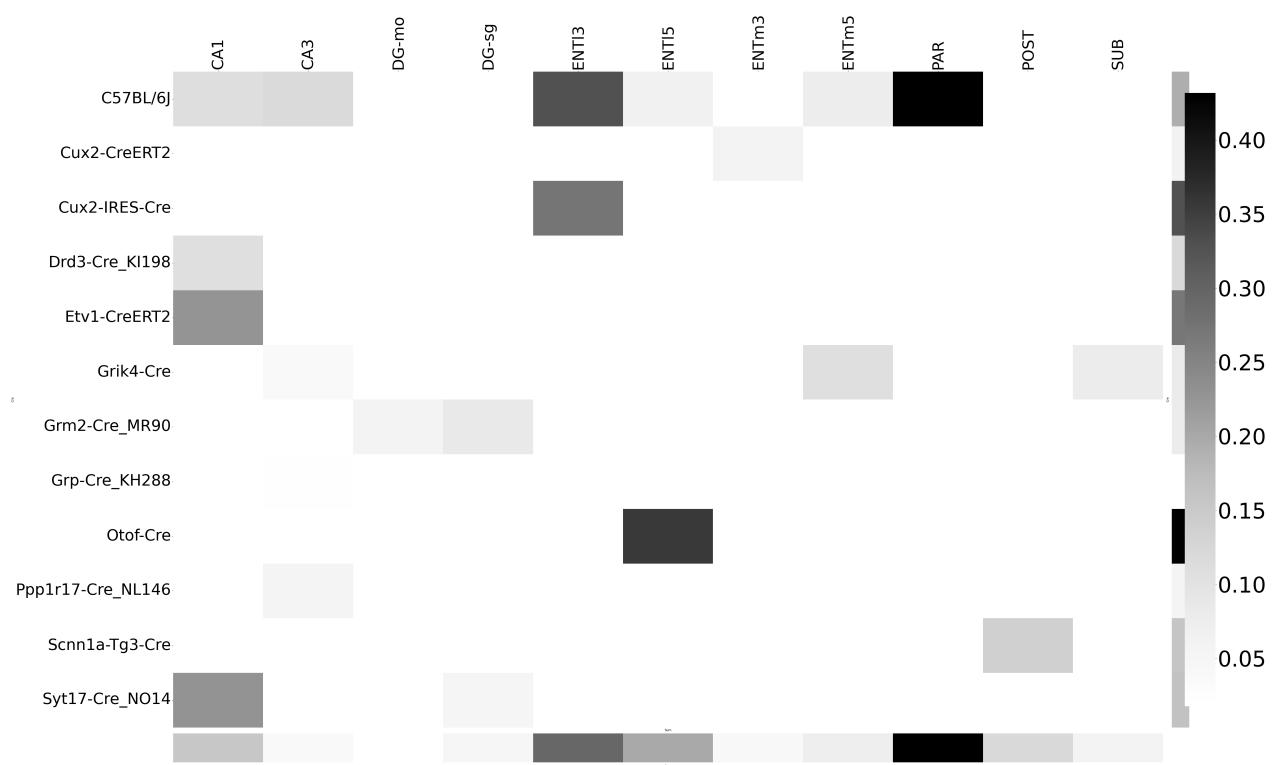


Figure 21: Weighted loss for Cre-leaf combinations in HPF. Missing values are omitted. Row and column averages are also plotted.

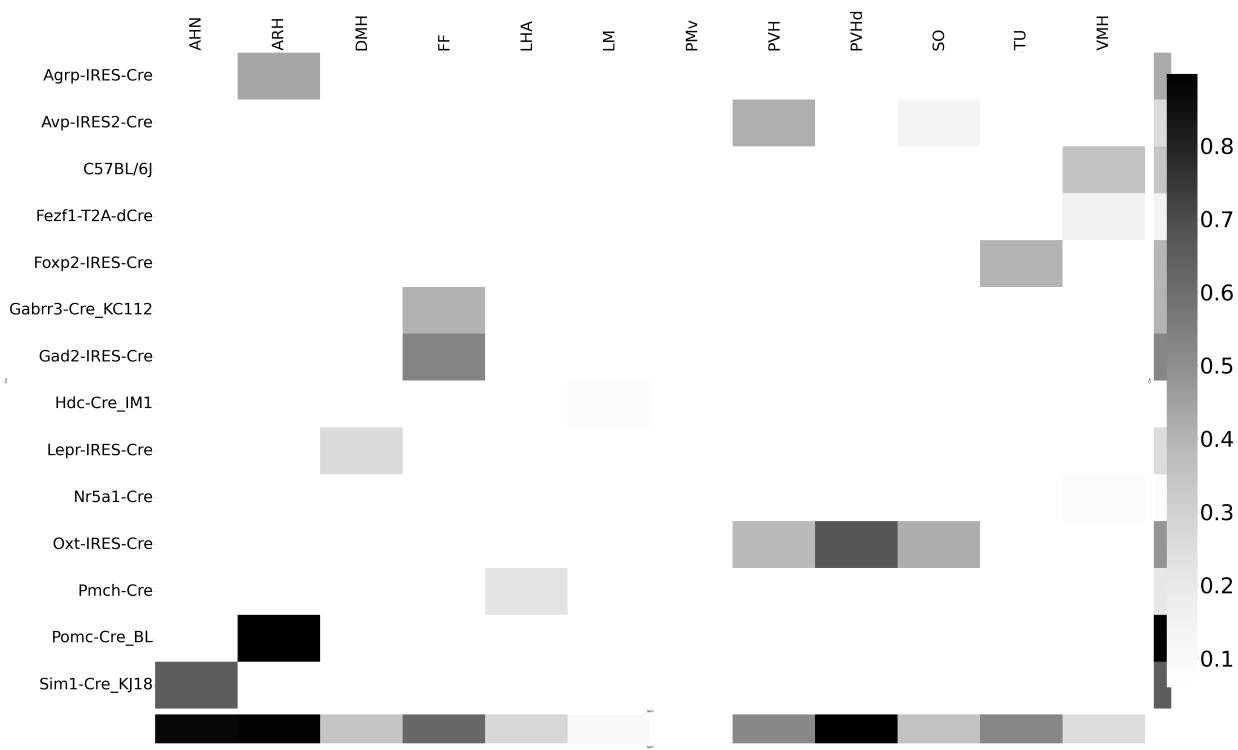


Figure 22: Weighted loss for Cre-leaf combinations in HY. Missing values are omitted. Row and column averages are also plotted.

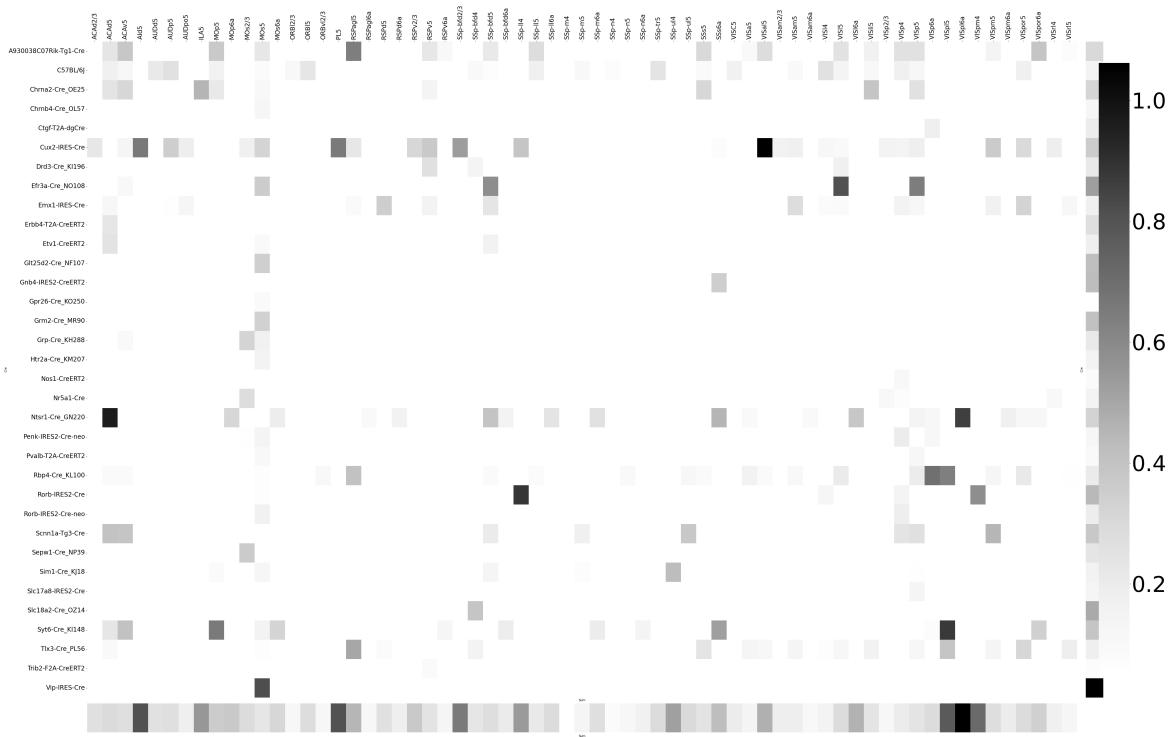


Figure 23: Weighted loss for Cre-leaf combinations in Isocortex. Missing values are omitted. Row and column averages are also plotted.

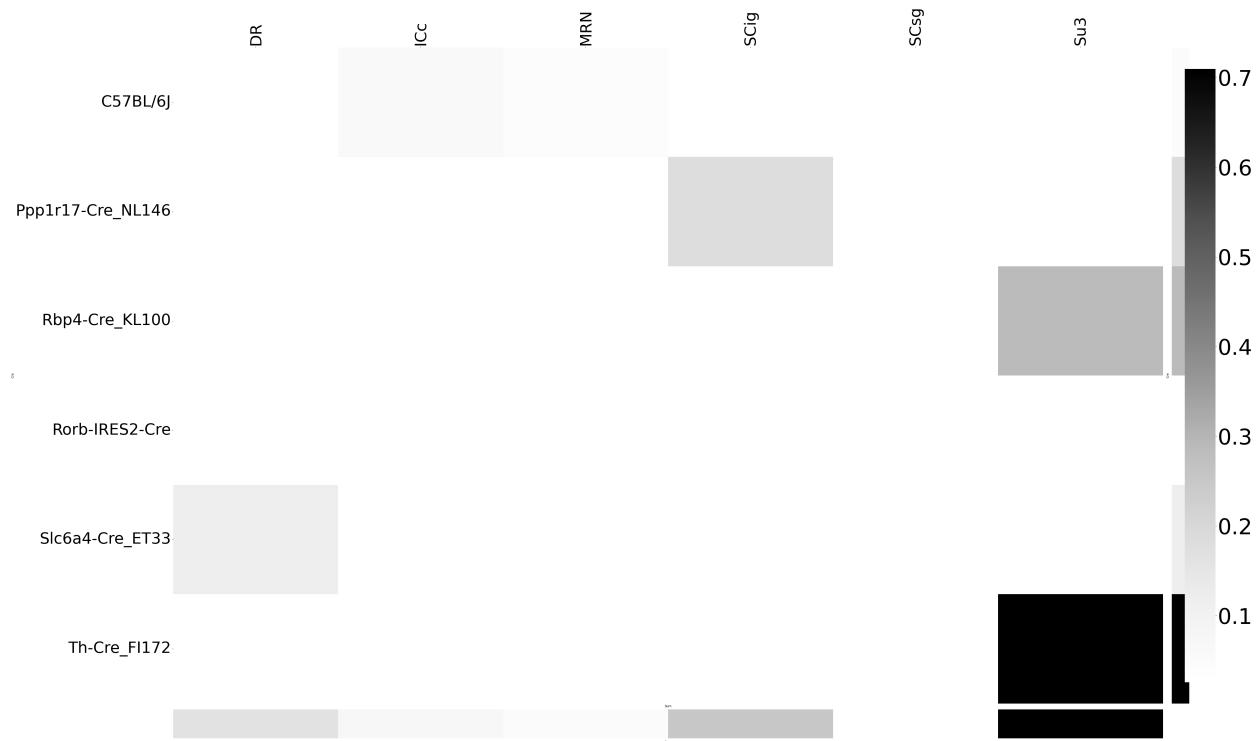


Figure 24: Weighted loss for Cre-leaf combinations in MB. Missing values are omitted. Row and column averages are also plotted.

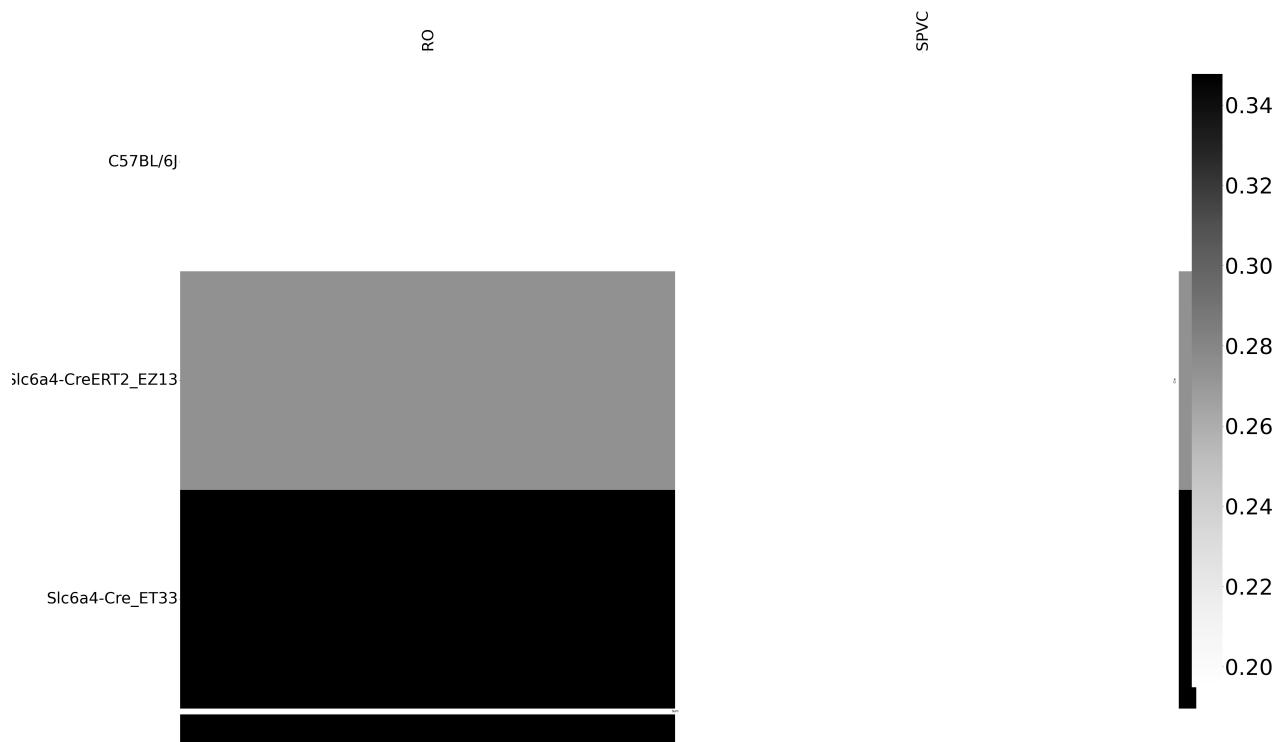


Figure 25: Weighted loss for Cre-leaf combinations in MY. Missing values are omitted. Row and column averages are also plotted.

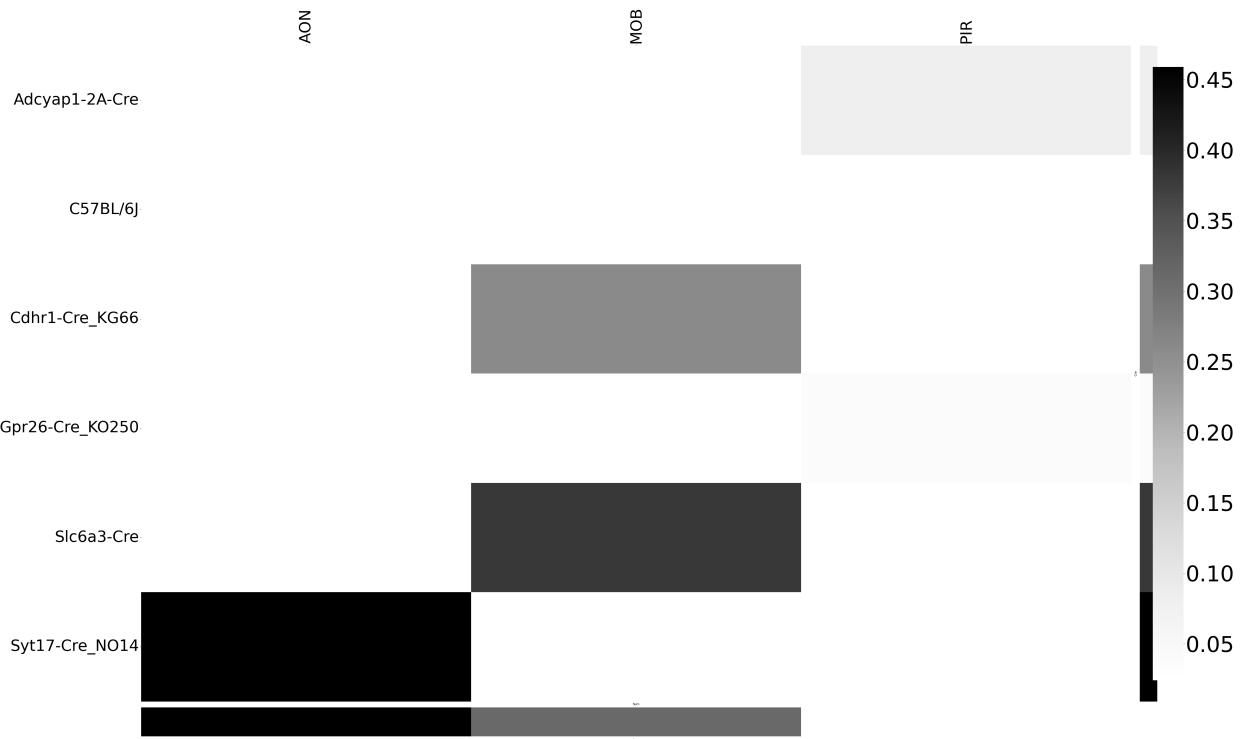


Figure 26: Weighted loss for Cre-leaf combinations in OLF. Missing values are omitted. Row and column averages are also plotted.

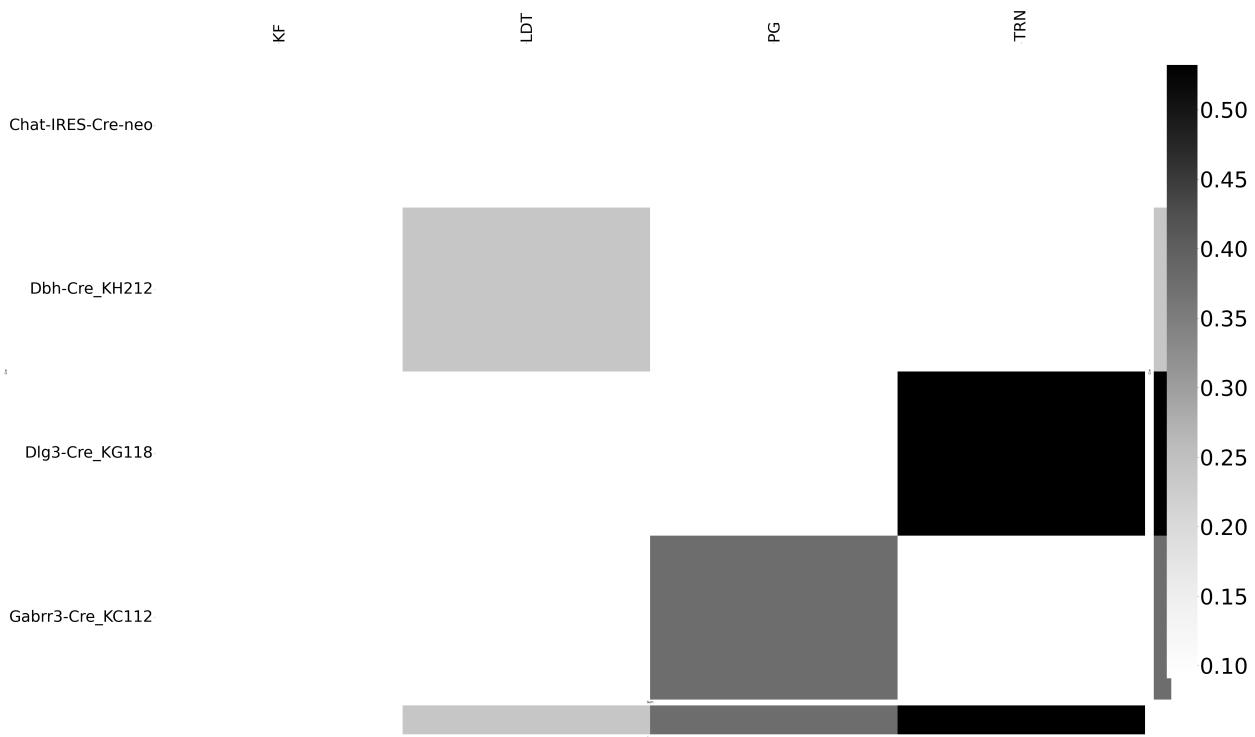


Figure 27: Weighted loss for Cre-leaf combinations in P. Missing values are omitted. Row and column averages are also plotted.

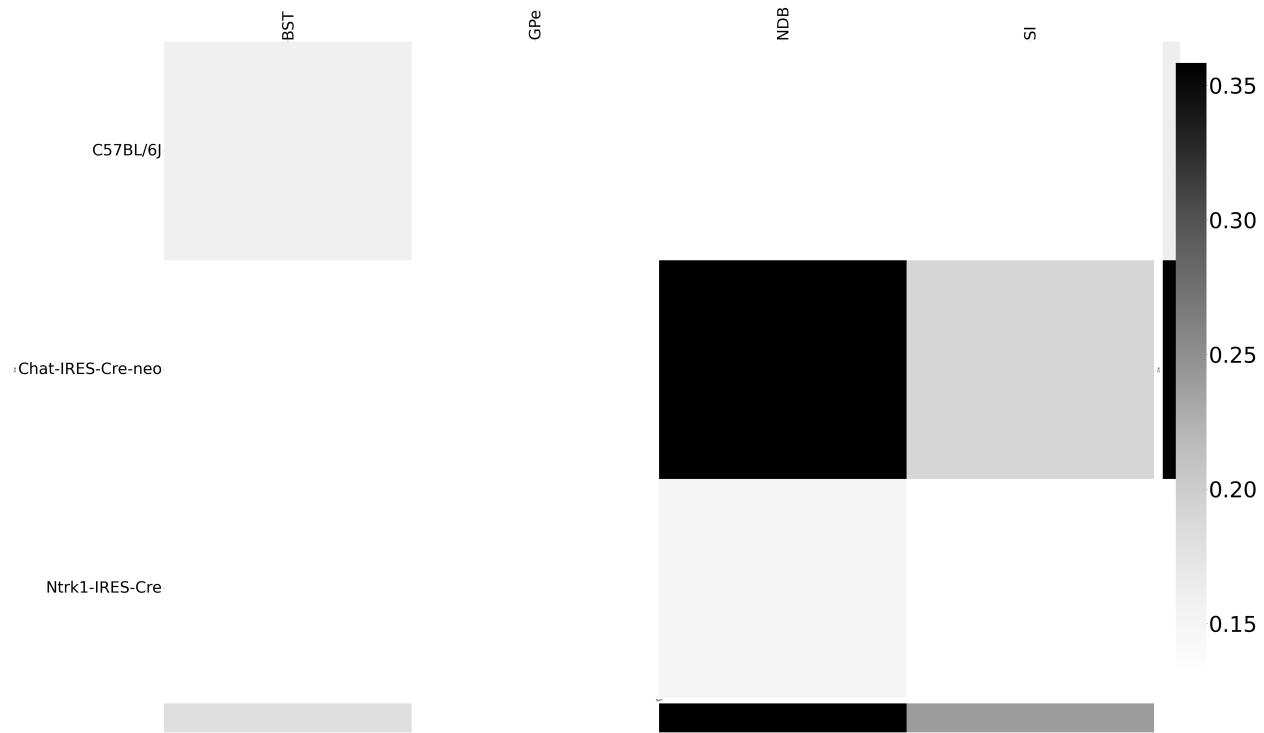


Figure 28: Weighted loss for Cre-leaf combinations in PAL. Missing values are omitted. Row and column averages are also plotted.

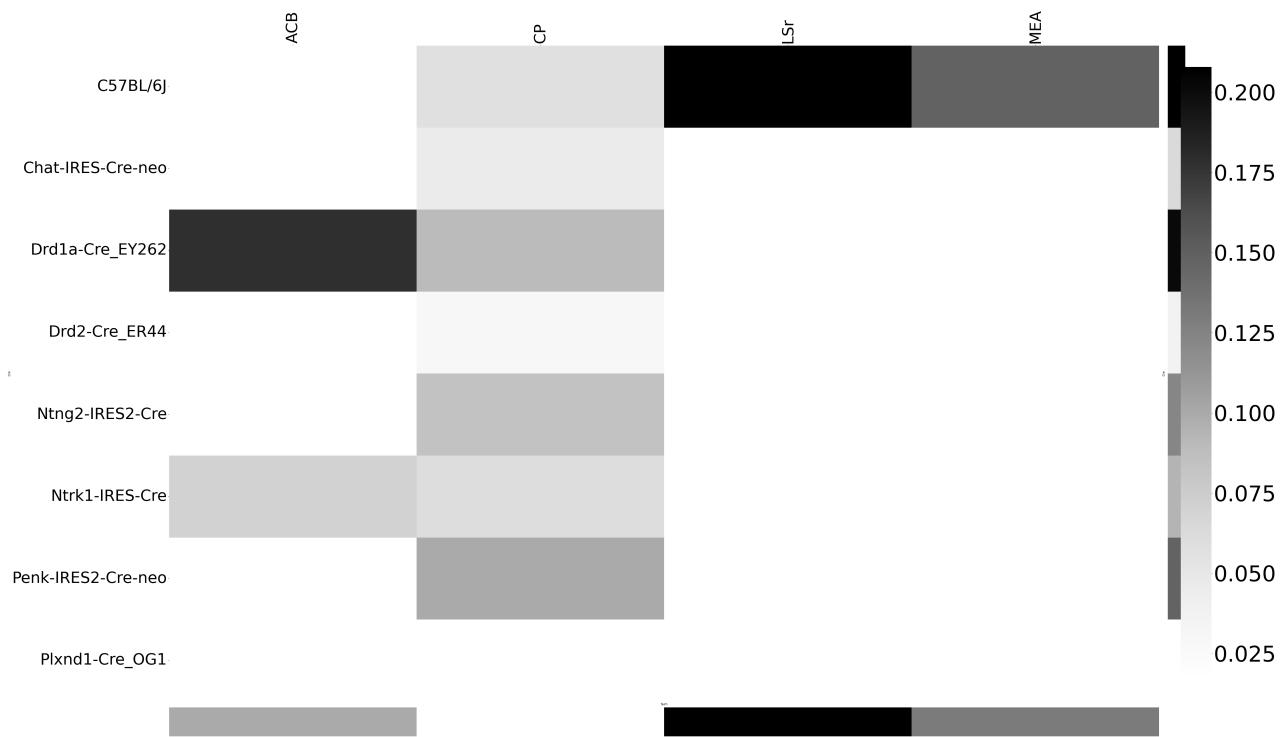


Figure 29: Weighted loss for Cre-leaf combinations in STR. Missing values are omitted. Row and column averages are also plotted.

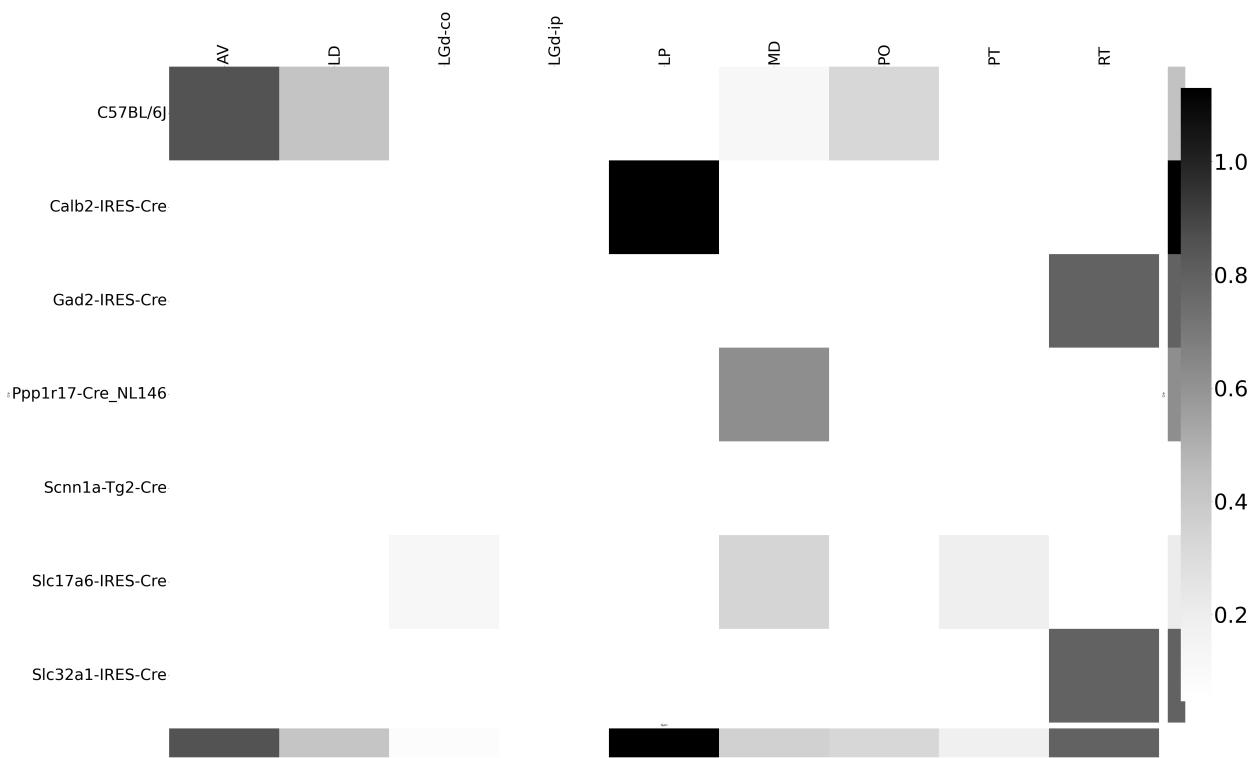


Figure 30: Weighted loss for Cre-leaf combinations in TH. Missing values are omitted. Row and column averages are also plotted.

590 ***Cell-type specificity***

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

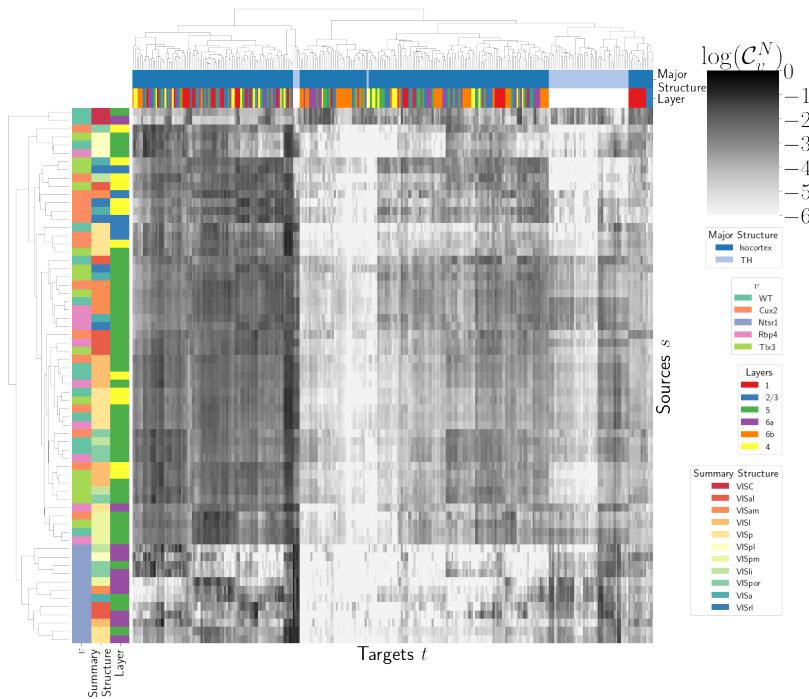


Figure 31: Hierarchical clustering of connectivity strengths from visual cortex cell-types to cortical and thalamic targets. Cre-line, summary structure, and layer are labelled on the sources. Major brain division and layer are labelled on the targets.

595 **Matrix Factorization**

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

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

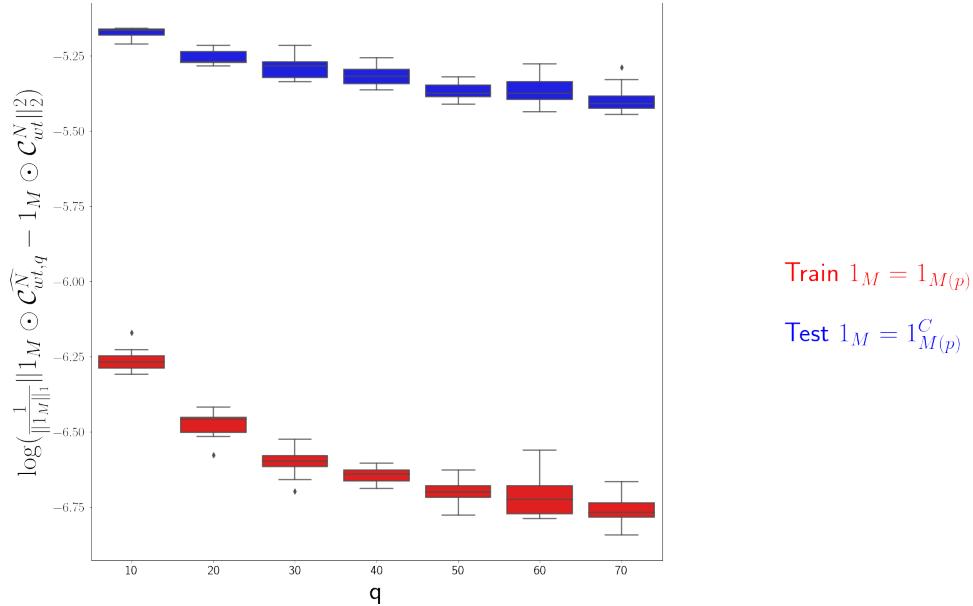


Figure 32: Train and test error using NMF decomposition.

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

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

q	10.000000	20.000000	30.000000	40.000000	50.000000
Rand index	0.772544	0.844981	0.932957	0.929827	0.885862

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

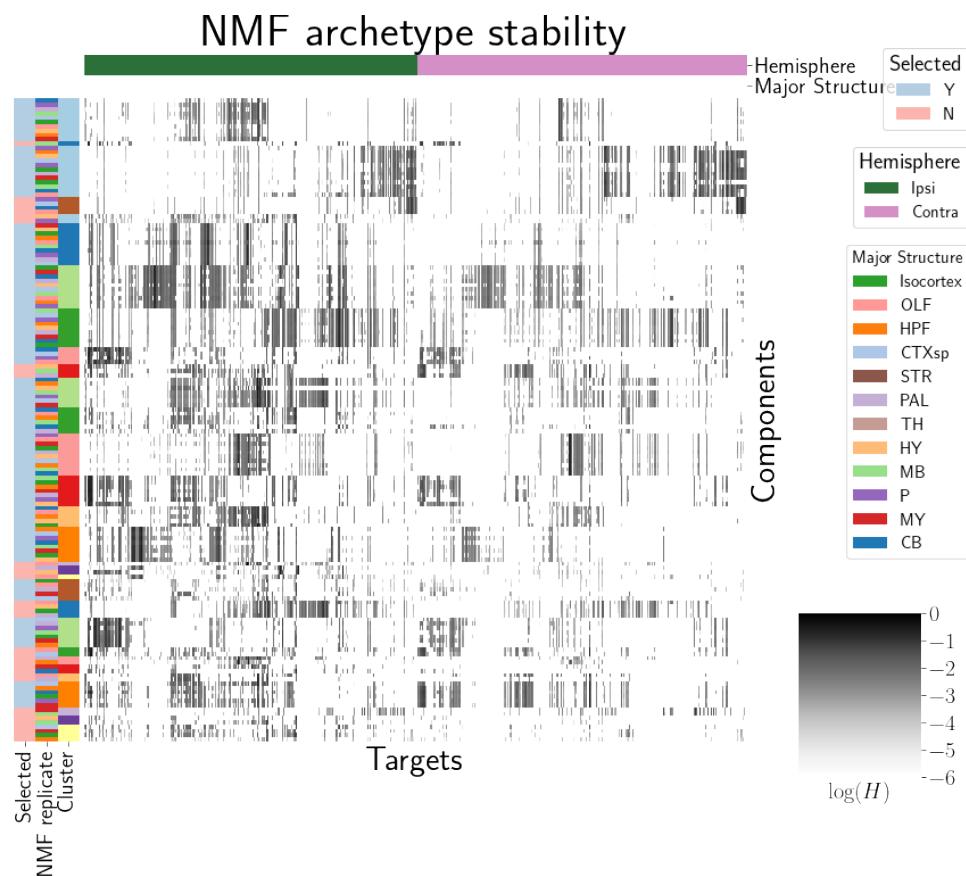


Figure 33: Stability of NMF results across replicates. Replicate and NMF component are shown on rows. Components that are in the top 15 are also indicated.

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

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

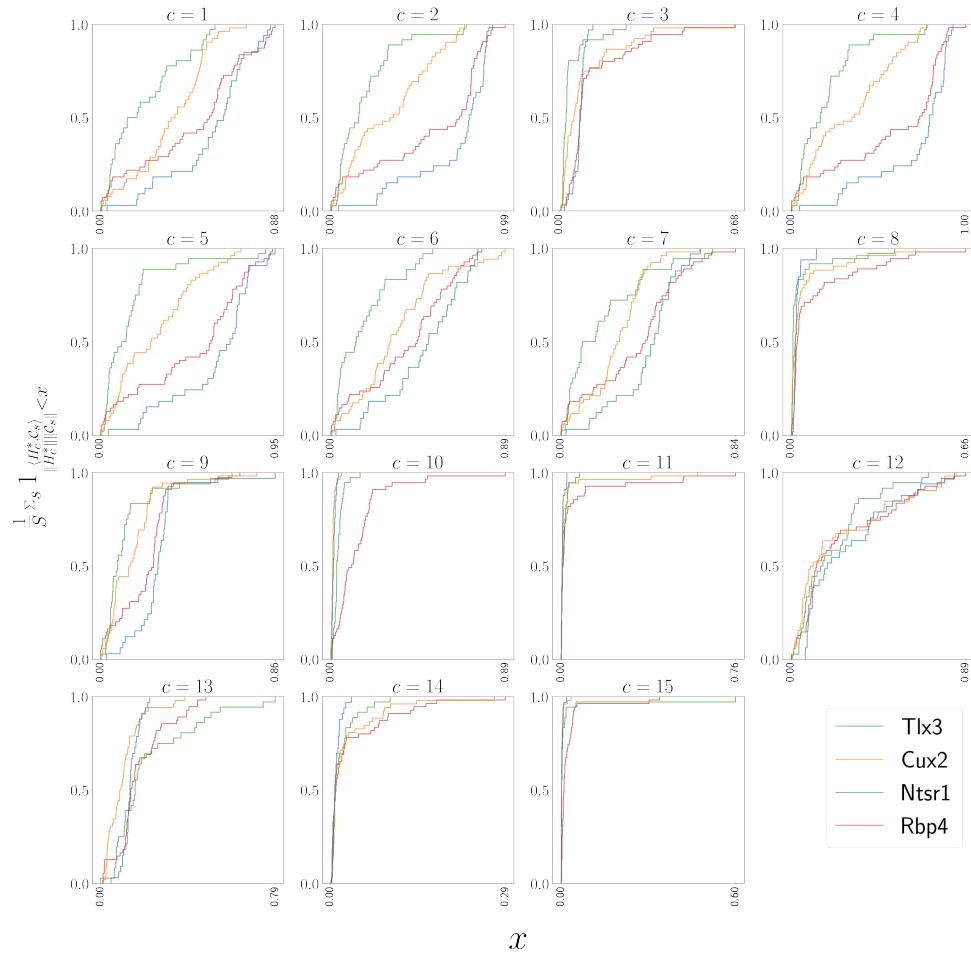


Figure 34: Empirical cumulative distributions of cosine similarities between source structures and connectivity components for four different Cre-lines.

619

620 **REFERENCES**

621

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

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

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

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

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

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

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

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

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

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

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

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

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

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