

# Phenotypic variation of transcriptomic cell types in mouse motor cortex

<https://doi.org/10.1038/s41586-020-2907-3>

Received: 5 February 2020

Accepted: 16 October 2020

Published online: 12 November 2020

Open access

 Check for updates

Federico Scala<sup>1,2,12</sup>, Dmitry Kobak<sup>3,12</sup>, Matteo Bernabucci<sup>1,2</sup>, Yves Bornaerts<sup>3,4</sup>, Cathryn René Cadwell<sup>5</sup>, Jesus Ramon Castro<sup>1,2</sup>, Leonard Hartmanis<sup>6</sup>, Xiaolong Jiang<sup>1,2,7</sup>, Sophie Laternus<sup>8</sup>, Elaine Miranda<sup>1,2</sup>, Shalaka Mulherkar<sup>2</sup>, Zheng Huan Tan<sup>1,2</sup>, Zizhen Yao<sup>8</sup>, Hongkui Zeng<sup>8</sup>, Rickard Sandberg<sup>6</sup>, Philipp Berens<sup>3,9,10,11</sup> & Andreas S. Tolias<sup>1,2</sup>✉

Cortical neurons exhibit extreme diversity in gene expression as well as in morphological and electrophysiological properties<sup>1,2</sup>. Most existing neural taxonomies are based on either transcriptomic<sup>3,4</sup> or morpho-electric<sup>5,6</sup> criteria, as it has been technically challenging to study both aspects of neuronal diversity in the same set of cells<sup>7</sup>. Here we used Patch-seq<sup>8</sup> to combine patch-clamp recording, biocytin staining, and single-cell RNA sequencing of more than 1,300 neurons in adult mouse primary motor cortex, providing a morpho-electric annotation of almost all transcriptomically defined neural cell types. We found that, although broad families of transcriptomic types (those expressing *Vip*, *Pvalb*, *Sst* and so on) had distinct and essentially non-overlapping morpho-electric phenotypes, individual transcriptomic types within the same family were not well separated in the morpho-electric space. Instead, there was a continuum of variability in morphology and electrophysiology, with neighbouring transcriptomic cell types showing similar morpho-electric features, often without clear boundaries between them. Our results suggest that neuronal types in the neocortex do not always form discrete entities. Instead, neurons form a hierarchy that consists of distinct non-overlapping branches at the level of families, but can form continuous and correlated transcriptomic and morpho-electrical landscapes within families.

As animals can be grouped into species and assembled into a hierarchy of phylogenetic relationships to form the ‘tree of life’, neurons in the brain are thought to form discrete cell types, which in turn can be cast in a hierarchy of neuronal families and classes. The current view is that a neuronal cell type is characterized by a common genetic profile that gives rise to distinct physiological and anatomical properties, including patterns of connectivity<sup>7,9</sup>. A comprehensive multi-modal taxonomy of neurons would resemble a ‘parts list’ of the brain, helping us to decipher its bewildering complexity<sup>10,11</sup>.

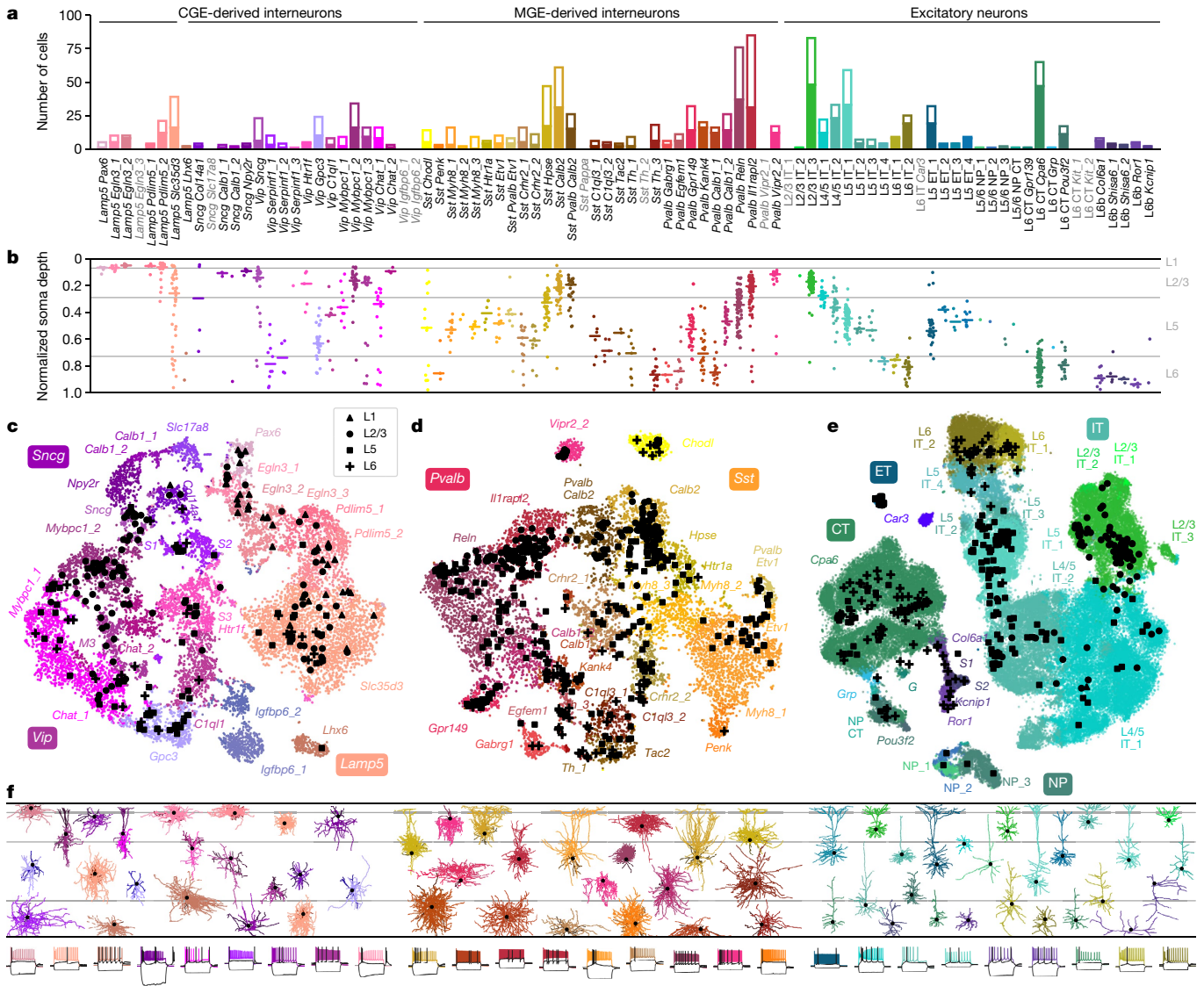
For decades, neurons have been classified into types by their anatomical and physiological characteristics, and more recently by molecular markers<sup>1,2,12,13</sup>. High-throughput single-cell sequencing techniques have identified dozens of types of neuron on the basis of their transcriptional profiles<sup>3,4,14,15</sup>, but linking transcriptomically defined cell types (t-types) to their phenotypes has remained a major challenge<sup>16</sup>. However, to understand the roles of t-types in cortical computations, it is necessary to know their anatomy, connectivity, and electrophysiology<sup>7</sup>.

Our work is part of the BRAIN initiative cell census network (BICCN) effort to fully characterize the cellular taxonomy of neurons in mouse primary motor cortex (MOp). We used the Patch-seq technique<sup>8,17–19</sup> to describe the morpho-electric phenotypes for most of the t-types in MOp<sup>20</sup>. Our analysis suggests that, in both excitatory and inhibitory classes of neurons, broad transcriptomic families (also known as ‘sub-classes’<sup>20</sup>) have largely distinct phenotypes, but uncovers continuous morpho-electric variation within most of these families.

## Patch-seq of mouse primary motor cortex

We sampled neurons across all layers (L) of adult mouse MOp (median postnatal day (P) 75) using various Cre driver lines to cover as diverse a population of neurons as possible. Neurons in acute slices were patch-clamped and stimulated with brief current pulses to record their electrophysiological activity at room temperature and then filled with biocytin for subsequent morphological recovery and reconstruction, and their RNA was extracted and sequenced using the Smart-seq2

<sup>1</sup>Center for Neuroscience and Artificial Intelligence, Baylor College of Medicine, Houston, TX, USA. <sup>2</sup>Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA. <sup>3</sup>Institute for Ophthalmic Research, University of Tübingen, Tübingen, Germany. <sup>4</sup>International Max Planck Research School for Intelligent Systems, Tübingen, Germany. <sup>5</sup>Department of Pathology, University of California San Francisco, San Francisco, CA, USA. <sup>6</sup>Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm, Sweden. <sup>7</sup>Jan and Dan Duncan Neurological Research Institute, Houston, TX, USA. <sup>8</sup>Allen Institute for Brain Science, Seattle, WA, USA. <sup>9</sup>Center for Integrative Neuroscience, University of Tübingen, Tübingen, Germany. <sup>10</sup>Institute for Bioinformatics and Medical Informatics, University of Tübingen, Tübingen, Germany. <sup>11</sup>Bernstein Center for Computational Neuroscience, University of Tübingen, Tübingen, Germany. <sup>12</sup>These authors contributed equally: Federico Scala, Dmitry Kobak. ✉e-mail: philipp.berens@uni-tuebingen.de; astolias@bcm.edu



**Fig. 1 | Transcriptomic coverage.** **a**, Number of Patch-seq cells assigned to each of the neural transcriptomic types (t-types)<sup>20</sup>. Colours and the order of types are taken from the original publication<sup>20</sup>. The filled part of each bar shows the number of morphologically reconstructed neurons. Grey labels, t-types with no cells. Total number of neurons, 1,227. **b**, Normalized soma depths of all neurons of each t-type. For t-types with at least three cells, horizontal lines show medians. Soma depths were normalized by the cortical thickness in each slice (0, pia; 1, white matter). Grey horizontal lines, approximate layer boundaries identified by Nissl staining (L1, 0.07; L2/3, 0.29; L5, 0.73). Total number of neurons, 1,187 (for some cells soma depth could not be measured owing to failed staining). **c**, t-SNE representation of CGE-derived interneurons from the single-cell 10x v2 reference data set ( $n = 15,511$ ;

perplexity, 30). t-Type names are shortened by omitting the first word; some are abbreviated. Patch-seq cells from the *Vip*, *Sncg*, and *Lamp5* subclasses were positioned on this t-SNE atlas<sup>23</sup> (black symbols). **d**, As in **c** but for MGE-derived interneurons ( $n = 12,083$ ; perplexity, 30). **e**, As in **c** but for excitatory neurons ( $n = 93,829$ ; perplexity, 100). **f**, Example morphologies coloured by t-type. For interneurons, dendrites are shown in darker colours. For excitatory neurons, only dendrites are shown. Black dots mark soma locations. Three voltage traces are shown below for some exemplary cells: the hyperpolarization trace obtained with the smallest current stimulation, the first depolarization trace that elicited at least one action potential, and the depolarization trace showing maximal firing rate. Stimulation length, 600 ms.

protocol<sup>21</sup> (Extended Data Fig. 1). In total, we performed whole-cell recordings from more than 2,000 cells, of which 1,329 cells (from 266 mice) passed initial quality control. The mRNA of these 1,329 cells was sequenced, yielding on average 1.3 million exonic and 0.7 million intronic reads (medians; mean  $\pm$  s.d. on a log<sub>10</sub> scale:  $6.0 \pm 0.6$  and  $5.6 \pm 0.8$ , respectively) and  $9,100 \pm 3,500$  (mean  $\pm$  s.d.) detected genes per cell (Extended Data Fig. 2). Of these neurons, 646 had sufficient staining for their morphologies to be reconstructed.

Using the gene expression profiles, we mapped all sequenced neurons to the transcriptomic cell types (t-types) that have been identified using dissociated cells in a parallel study within the BICCN consortium<sup>20</sup>. To

assign cell types, we used a nearest centroid classifier with Pearson correlation of log-expression across the most variable genes as a distance metric (Extended Data Fig. 1). Bootstrapping over genes was used to assess mapping confidence. The mapping was done separately using each of the seven reference data sets obtained with different sequencing technologies, including single-cell and single-nucleus Smart-seq2 and 10x sequencing<sup>20</sup>. We found that Patch-seq expression profiles were most similar to the single-nucleus Smart-seq2 data (Extended Data Fig. 2g, h). At the same time, there was good agreement between t-type assignments based on Smart-seq2 and those based on 10x reference data (Extended Data Fig. 2i), so consensus t-type assignment over all

seven reference data sets was used for all subsequent analysis. Cells that showed poor mapping (owing to a low read count or excessive RNA contamination) were excluded (Extended Data Fig. 2), leaving 1,227 neurons for further analysis (817 inhibitory, 410 excitatory; 369 and 269 with morphological reconstructions, respectively).

The resulting data set covered 77 out of the 90 neuronal t-types (Fig. 1a), with 73 t-types having at least one morphologically reconstructed neuron. The coverage was good for interneurons derived from the caudal and medial ganglionic eminences (CGE and MGE) and for excitatory neurons. Within-type distributions of soma depths (Fig. 1b) agreed well with previous data<sup>4</sup> and with the layer-specific nomenclature of excitatory t-types, confirming the validity of our t-type assignment. Positioning all cells on reference maps made with *t*-distributed stochastic neighbour embedding (t-SNE)<sup>22,23</sup> also showed good overall coverage (Fig. 1c–e) with only few uncovered regions.

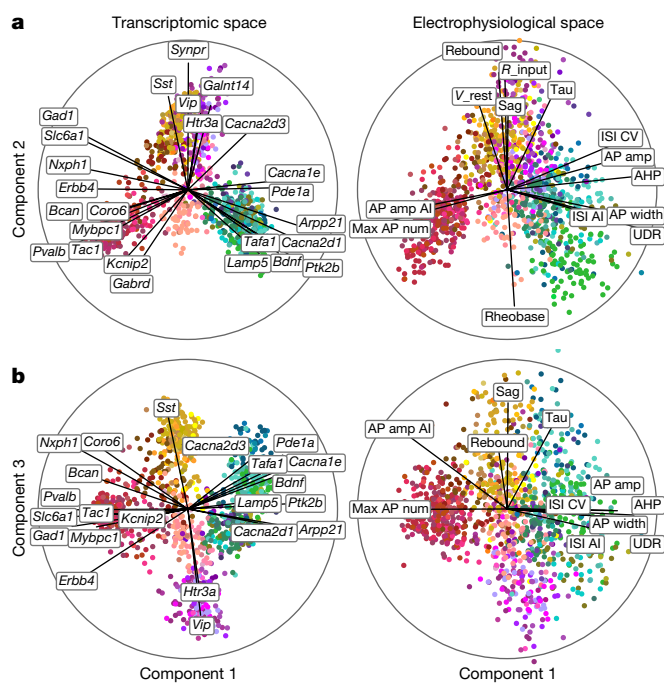
The observed phenotypes included most of the morphological and electrophysiological types of cortical neurons that have been described previously in mice and rats<sup>5,6,24</sup>, allowing us to link transcriptomic and morpho-electric descriptions (Extended Data Fig. 3, Supplementary File 1).

A detailed description of all t-types is provided in Extended Data Tables 1, 2. One interesting case was the transcriptomically isolated *Lamp5 Lhx6* type, which consists of deep L5/L6 neurogliaform cells (NGCs). This type, unlike all other *Lamp5* types, is putatively MGE-derived<sup>4</sup>, so its identity was an open question<sup>16</sup>. Our results suggest that although all deep NGCs belong to the *Lamp5* subclass, some are derived from the CGE and some from the MGE, as in the hippocampus<sup>25–27</sup>. Another finding was that the *Sst Pvalb Calb2* type, which is transcriptomically in between the *Sst* and *Pvalb* subclasses, was also in between these subclasses in terms of its morpho-electric phenotype<sup>28</sup>. Furthermore, we confirmed that chandelier cells from both superficial and deep layers belonged to transcriptomically isolated *Pvalb Vipr2* types. We also showed that three previously described morphological types of L5 *Pvalb* cells<sup>5</sup>, as well as two morphological types of L5 Martinotti cells<sup>29,30</sup>, corresponded to different t-types. We were also able to identify a t-type, L4/5 IT\_1, that was located on the boundary between L2/3 and L5 and probably corresponds to the quasi-L4 neurons described previously in motor cortex<sup>31</sup>.

### Distinct phenotypes of major families

We next asked to what extent the morpho-electric phenotype could be predicted by gene expression across the entire data set. To obtain quantitative characterizations of the morpho-electric phenotypes, we extracted 29 electrophysiological (Extended Data Fig. 4, Supplementary File 2) and about 50 morphological features for each cell. We first focused on 17 electrophysiological features and used sparse reduced-rank regression<sup>32</sup>, a technique that predicts the firing properties on the basis of a low-dimensional latent space representation computed from a sparse selection of genes. We used cross-validation to tune the regularization strength (Extended Data Fig. 5). The selected model used 25 genes with a 5-dimensional latent space and achieved a cross-validated  $R^2$  of 0.38. To visualize the structure of the latent space, we projected gene expression and electrophysiological properties onto the latent dimensions (Fig. 2). The cross-validated correlations between the first three pairs of projections were 0.90, 0.74, and 0.67, respectively.

These first three components clearly separated five major groups of neurons: the *Pvalb*, *Sst*, *Vip*, and *Lamp5* interneuron subclasses, and the excitatory neuron class (Fig. 2). These groups had distinct electrophysiological properties: for example, as expected, *Pvalb* neurons were characterized by high firing rates while *Sst* neurons had high values of the hyperpolarization sag and rebound (Fig. 2, right). Some of the genes selected by the model were prominent marker genes, such as the pan-inhibitory markers *Gad1* and *Slc6a1* (related to GABA



**Fig. 2 | Sparse reduced-rank regression.** **a, b**, A sparse reduced-rank regression (RRR) model<sup>32</sup> to predict combined electrophysiological features from gene expression. Transcriptomic data are linearly projected to a low-dimensional space that allows reconstruction of electrophysiological data; components 1 and 2 (**a**) and 1 and 3 (**b**) of rank-5 model are shown.  $n = 1,219$ . Colour corresponds to t-type. The model selected 25 genes (left). Each panel is a biplot, showing correlations of original features with both components; the circle corresponds to correlation 1. Only features with average correlation above 0.4 are shown. Labels were automatically positioned to reduce overlap. AI, adaptation index; AP, action potential; CV, coefficient of variation; ISI, interspike interval;  $R_{input}$ , input resistance;  $V_{rest}$ , resting potential; UDR, upstroke-to-downstroke ratio.

( $\gamma$ -aminobutyric acid) processing), or the more specific inhibitory markers *Sst*, *Vip*, *Pvalb*, *Tac1*, and *Htr3a*. Notably, some other selected genes were more directly related to electrophysiological properties, such as the calcium channel subunit genes *Cacna1e* and *Cacna2d3* or the potassium channel-interacting protein gene *Kcnp2*, which can modulate firing properties in individual families. A reduced-rank regression model restricted to using only ion channel genes (Extended Data Fig. 5) did not perform much worse than the full model (cross-validated  $R^2 = 0.33$  and correlations 0.86, 0.71, and 0.56, respectively, with regularization set to select 25 genes). Reduced-rank regression analysis using morphological features supported the separation of major families (Extended Data Fig. 5).

Similarly, a 2D t-SNE embedding of Patch-seq cells based on the same electrophysiological features showed that the major transcriptomic families have distinct electrophysiological properties (Fig. 3a): the *Pvalb*, *Lamp5*, *Sst*, *Vip*, CT (corticothalamic), IT (intratelencephalic), and ET (extratelencephalic) subclasses were mostly well separated from each other. We quantified this separation using a confusion matrix for *k*-nearest neighbours (kNN) classification of cells into families: it was mostly diagonal, with only the ET and IT subclasses strongly overlapping (Fig. 3d). We confirmed the electrophysiological overlap between IT and ET neurons in follow-up experiments at 34 °C (Extended Data Fig. 6).

We also constructed a 2D t-SNE embedding based on the morphological features (Fig. 3b). We used only dendritic features for the excitatory cells, but both axonal and dendritic features for the inhibitory cells, leading to a strong separation between these two



major classes. Within each class, cells were strongly segregated by the soma depth, with excitatory cells forming mostly a one-dimensional manifold. The separability between inhibitory families was weaker than with electrophysiological features (Fig. 3d). The between-family separability was strongest when we had combined electrophysiological and morphological features into a joint representation (Fig. 3c, d), showing that these sets of properties are not redundant. The ellipses in Fig. 3c highlight prominent t-types and groups of t-types with similar morpho-electric properties.

In summary, different transcriptomic families had largely distinct morpho-electric phenotypes, in agreement with them being well separated in the transcriptomic space<sup>4</sup>.

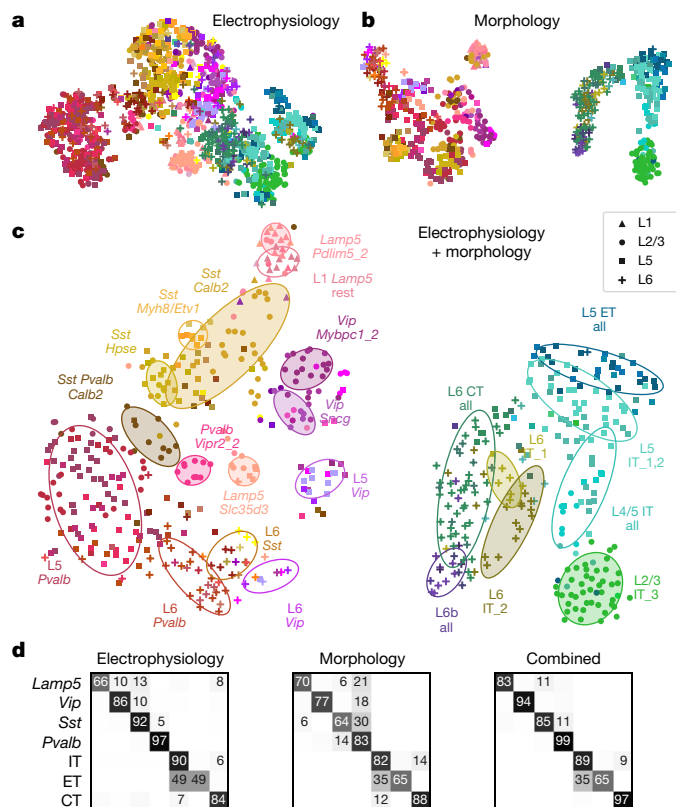
### Continuous phenotypic variation

Within individual transcriptomic families, morpho-electric phenotypes rarely formed isolated clusters (Fig. 3). Moreover, we often found that morpho-electric phenotypes varied continuously from one t-type to another (Fig. 4). For example, electrophysiological properties of the t-types within the *Vip* subclass varied continuously across the transcriptomic landscape; the membrane time constant, for instance, had its largest values close to the *Sncg* subclass and gradually decreased towards *Vip Gpc3* (Fig. 4a). We observed the same in the *Sst* subclass, which is known to be transcriptomically<sup>4</sup> and morpho-electrically<sup>29,30,33</sup> diverse in L5. Here we also found that morpho-electric properties varied continuously across the transcriptomic landscape, with neighbouring t-types consistently showing similar morphologies and similar rebound values (Fig. 4b). We confirmed this effect in follow-up experiments at physiological temperature (Extended Data Fig. 6).

To quantify this effect, for each pair of t-types within each family we computed the transcriptomic distance (correlation distance between average log-counts in the reference data) and the electrophysiological distance (Euclidean distance between average feature vectors) between them. Pooling the pairs across all families, we found that these two distance measures were correlated, with  $r = 0.60$  (Fig. 4c,  $n = 200$  pairs; Extended Data Fig. 7). The correlation was also observed within multiple individual families and for many individual electrophysiological features (Extended Data Fig. 7).

The IT subclass provides an example of a similar phenomenon in another data modality (Fig. 4d). IT neurons span all layers from L2/3 to L6, and IT t-types are largely layer-restricted<sup>4</sup>. However, we found that IT t-types did not form distinct groups for each cortical layer; instead, the soma depth and RNA expression varied continuously along a one-dimensional manifold (Fig. 4d), in agreement with parallel findings based on a spatial transcriptomics approach<sup>34</sup>. For example, L4/5 and L5 IT t-types that were transcriptomically close to the L2/3 IT t-types were located at the top of L5 close to the border between L2/3 and L5, whereas L5 IT t-types that were transcriptomically close to L6 IT t-types were located at the bottom of L5 close to the border with L6. Transcriptomic distances between t-types were strongly correlated with the average soma depth differences ( $r = 0.70$ ; Fig. 4d).

Finally, the *Pvalb* subclass is usually understood as electrophysiologically homogenous (all neurons are fast spiking) but has been described as morphologically diverse, in particular in L5<sup>5</sup>. However, it was previously unclear whether different morphologies such as shrub-like or horizontally elongated correspond to different t-types<sup>5</sup>. While we found that different t-types had different preferred morphologies (Extended Data Table 1), they showed substantial overlap, in agreement with the L5 *Pvalb* t-types themselves not having clear boundaries<sup>4</sup> (Fig. 1d). The shape of the axonal arbor showed continuous changes across the transcriptomic landscape (Fig. 4e): small shrub-like basket cells, horizontally elongated basket cells, and vertically elongated classical basket cells were located in different corners of the t-SNE embedding, with intermediate morphologies in between.



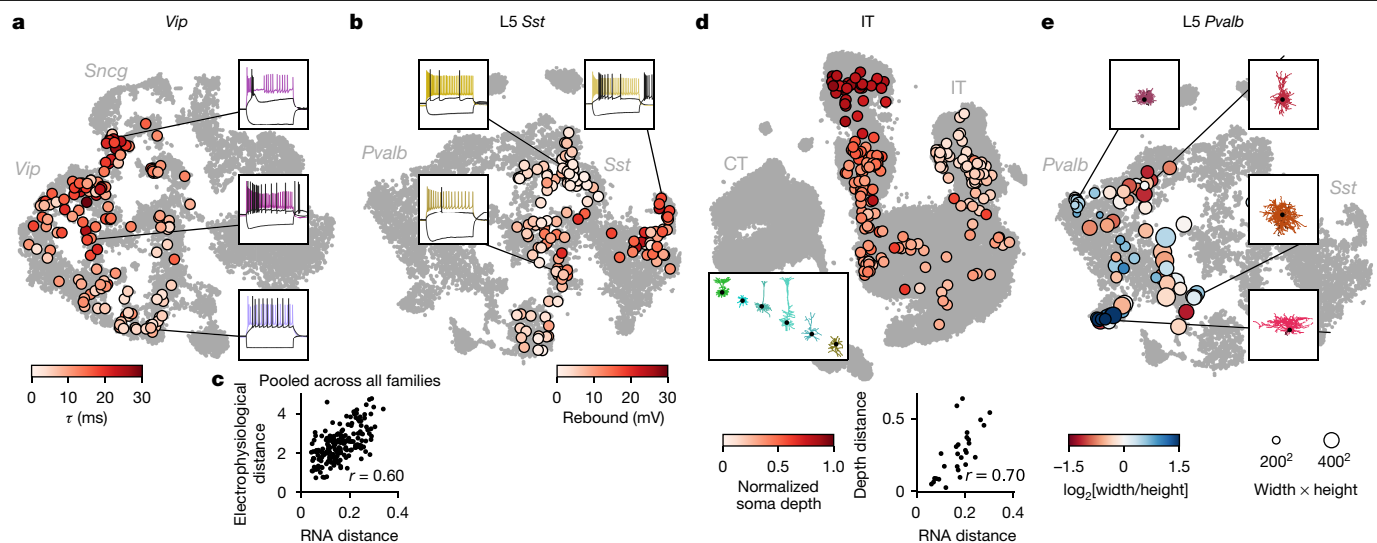
**Fig. 3 | Morpho-electric t-SNE embeddings.** **a**, t-SNE embedding constructed using electrophysiological features. Colour corresponds to t-type.  $n = 1,320$  cells used to construct the embedding, 1,219 cells with t-type labels shown. Perplexity, 30. **b**, t-SNE embedding constructed using combined morphometric features and z-profiles.  $n = 636$  cells. Perplexity, 30. **c**, t-SNE embedding constructed using combined electrophysiological and morphological features.  $n = 628$  cells. Perplexity, 30. Ellipses show 80% coverage ellipses for the most prominent t-types (shaded) and for some groups of related t-types and some layer-restricted families (unshaded). We chose these groups to reduce the overlap between ellipses. **d**, Confusion matrices for classifying cells into seven transcriptomic families using kNN classifier ( $k = 10$ ) and three feature sets. Each row shows what fraction of cells from a given family is classified in each of the seven families. The values in each row sum to 100% but only values above 5% are shown.

In summary, within major transcriptomic families, morpho-electric phenotypes and/or soma depth often varied smoothly across neighbouring t-types, indicating that transcriptomic neighbourhood relationships in many cases corresponded to similarities in other modalities.

### Variability in individual t-types

To study the morpho-electric phenotypes of individual t-types, we measured how consistently they conformed to their respective transcriptomic families (Fig. 5a) and how variable they were within a t-type (Fig. 5b). First, we used a kNN classifier to classify cells from each t-type with at least ten cells into transcriptomic families, using electrophysiological features. Most t-types could be unambiguously placed into the correct family (Fig. 5a), but some t-types were in between two families. For example, many *Sst Pvalb Calb2* neurons were classified as belonging to the *Pvalb* subclass on the basis of electrophysiology. Similarly, *Lamp5 Egl3\_1* neurons had rather *Vip*- and *Sst*-like firing instead of the typical *Lamp5* electrophysiology, and *Vip Mybpc1* neurons often had *Sst*-like firing. Thus, while overall transcriptomic family was highly predictive of the cell phenotype, some t-types exhibited properties similar to those of another transcriptomic family.





**Fig. 4 | Phenotypic variability within transcriptomic families.** **a**, *Vip* neurons mapped to the reference t-SNE embedding from Fig. 1c, coloured by membrane time constant ( $\tau$ ). Insets, example firing traces. **b**, *Sst* neurons from layer 5 (excluding *Sst Chodl* t-type) mapped to the reference t-SNE embedding from Fig. 1d, coloured by rebound value. **c**, Correlation between transcriptomic distances and electrophysiological distances across all 200 pairs of t-types from the same family (for 50 t-types with at least 5 cells), pooled across all families. Transcriptomic distance was computed using the reference 10x data as the correlation between average log-expression across most variable genes. Electrophysiological distance is Euclidean distance between the average

feature vectors. **d**, IT neurons mapped to the reference t-SNE embedding from Fig. 1e, coloured by normalized soma depth. Inset, examples of IT neurons at different depths, coloured by t-type. Scatter plot used eight t-types with at least five cells and shows correlation between transcriptomic distances and cortical depth distances. Cortical depth distance is Euclidean distance between the average normalized soma depths. **e**, *Pvalb* neurons from layer 5 mapped to the reference t-SNE embedding from Fig. 1d, coloured by axonal width/height log-ratio. Circle area corresponds to the width  $\times$  height product. Insets, example morphologies.

Next, we measured the normalized total variance of each t-type using electrophysiological features and compared it to the normalized total variance of phenotype clusters derived by *k*-means clustering (with *k* set to the number of t-types). The rationale here was that the variance of the *k*-means clusters would reflect the minimal possible variance obtainable in our data set. Values much above the cluster variances indicate non-trivial phenotypic variability within a t-type.

We found that many t-types had total variance substantially above the variances of the *k*-means clusters (Fig. 5b) and an alternative analysis using entropies of Leiden clustering<sup>35</sup> often highlighted the same t-types as variable (Extended Data Fig. 8). Not all t-types showed high variability: some of them, such as *Pvalb Vipr2\_2* (chandelier cells), appeared morpho-electrically homogeneous. By contrast, *Vip Mybpc1\_2* was marked as having high electrophysiological variability and indeed had high variance in input resistance, membrane time constant, and rebound (Extended Data Fig. 4). This variability was not random: overlaying the rebound values on the t-SNE embedding (Fig. 5c) showed that cells with low rebound were located close to the boundary with the low-rebound *Vip Sncg* type. Similarly, *Sst Pvalb Calb2* cells had high variability in terms of the maximum firing rate, but high-firing cells were mostly grouped in one part of the transcriptomic landscape (Fig. 5d).

We found similar examples in the morphological modality (Extended Data Fig. 8). Together, these examples suggest that within-t-type morpho-electric variability can in some cases be related to the underlying transcriptomic variability. This is in agreement with the idea that on a fine within-family scale, both transcriptomic and morpho-electric landscapes are continuous rather than discrete.

## Discussion

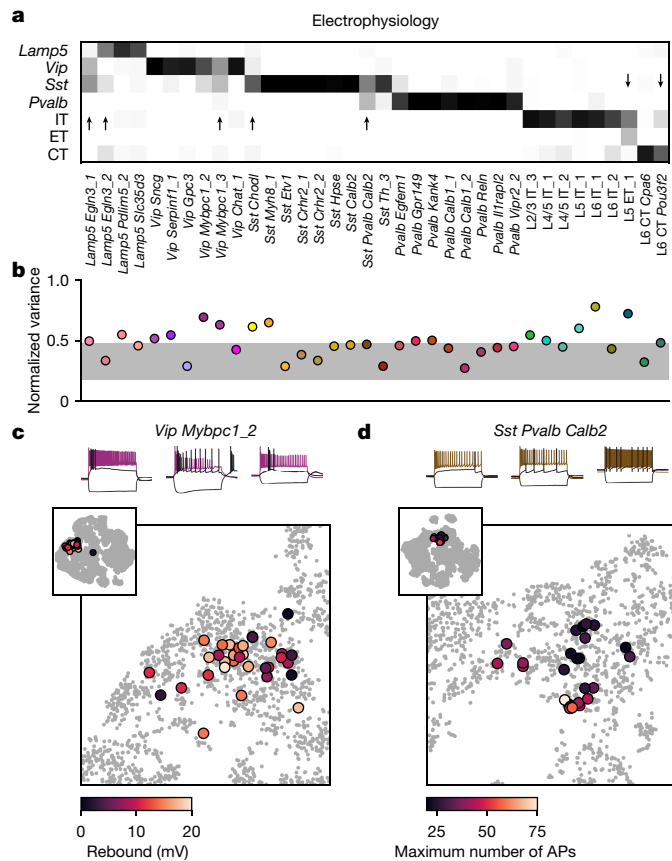
We used Patch-seq to provide the missing link between transcriptomic and morpho-electric descriptions of neurons in adult mouse motor cortex. Broad transcriptomic families were mostly well separated in their

morpho-electric properties. Previous studies using transgenic lines had shown that morpho-electric properties within these families can be highly variable<sup>5,24</sup>. We found that this variation is structured across the transcriptomic landscape, such that the morpho-electric distance between t-types within a family is correlated with their transcriptomic distance. Furthermore, we found non-trivial morpho-electric variability within multiple t-types. Although we cannot fully exclude the possibility that some of this variability can be attributed to technical challenges of Patch-seq or to factors such as the exact spatial location of the cell within motor cortex<sup>36</sup>, there are clear cases in our data that suggest that within-type morpho-electric variability is related to within-type transcriptomic variability.

We therefore suggest that the ‘tree of cortical cell types’ may look more like a banana tree with a few large leaves, rather than an olive tree with many small ones. In this metaphor, neurons follow a hierarchy consisting of distinct, non-overlapping branches at the level of families (large leaves), but with a spectrum of cells forming continuous and correlated transcriptomic and morpho-electric landscapes within each leaf.

This is at odds with the notion that t-types are discrete entities, an implicit assumption behind any cluster analysis. Consistent with our interpretation, recent transcriptomic and anatomical studies have argued that neurons in hippocampus, striatum, and cerebellum can be better described as forming partially continuous manifolds<sup>27,37–39</sup>. Similarly, cortical studies have identified many intermediate cells with uncertain t-type assignments<sup>3,4</sup>. Thus, the goal to assemble an exhaustive inventory of neural cell types might be unattainable if the types, unlike the chemical elements in the periodic table, are not discrete entities. We believe that there is an urgent need for theoretical work on how to conceptualize and model hierarchical discrete/continuous cell variability in a principled way<sup>7</sup>.

Developmentally, it is thought that neural diversity is generated through a combination of intrinsic genetic programs in progenitor cells, and activity-dependent and environmental factors<sup>40–44</sup>. It remains



**Fig. 5 | Phenotypic variability of individual t-types.** **a**, Confusion matrix for classifying cells from each t-type into seven transcriptomic families using electrophysiological features. Only t-types with at least ten cells are shown. Values in each column sum to 1. Arrows mark t-types that are classified into wrong families more than 25% of the time. We used a kNN-based classifier with  $k = 10$ . **b**, Normalized total variance of features in each t-type. Higher values correspond to t-types with more variable phenotypes. Horizontal grey band, minimum to maximum normalized variances of  $k$ -means clusters. **c**, Three exemplary traces from *Vip Mybpc1\_2* cells (all with confidence  $\geq 95\%$ ) and t-SNE overlay coloured by rebound. Inset, the same t-SNE embedding as in Fig. 1. Main plot, magnification. **d**, Three exemplary traces from *Sst Pvalb Calb2* cells (confidence  $\geq 95\%$ ) and t-SNE overlay coloured by maximum firing rate.

unclear to what extent the interplay between hard-wired genetic programs and extrinsic cues might explain our observations.

Our study has several limitations. First, some t-types were covered only sparsely or not at all. Additional experiments with more specific Cre lines could fill some of the gaps, but some very rare putative t-types might not be amenable to Patch-seq study. Second, as the RNA extraction process may have interfered with biocytin diffusion<sup>17</sup> and as MOP is quite thick, it was difficult to recover complete morphologies of some groups of neurons, such as deep L5 Martinotti cells with thin long axons that reach all the way to L1.

A parallel Patch-seq study of the inhibitory neurons in the mouse visual cortex<sup>45</sup> focused on isolating multimodal neural types ('met-types') but also often observed continuous variation. Our data sets are overall in good agreement (Extended Data Fig. 9) and together offer an unprecedented view of cell type variability in the neocortex. Future studies will need to add additional modalities, such as long-range projections, local connectivity, and in vivo functional characterization.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information,

acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2907-3>.

- Harris, K. D. & Shepherd, G. M. The neocortical circuit: themes and variations. *Nat. Neurosci.* **18**, 170–181 (2015).
- Tremblay, R., Lee, S. & Rudy, B. GABAergic interneurons in the neocortex: from cellular properties to circuits. *Neuron* **91**, 260–292 (2016).
- Tasic, B. et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016).
- Tasic, B. et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).
- Jiang, X. et al. Principles of connectivity among morphologically defined cell types in adult neocortex. *Science* **350**, aac9462 (2015).
- Markram, H. et al. Reconstruction and simulation of neocortical microcircuitry. *Cell* **163**, 456–492 (2015).
- Zeng, H. & Sanes, J. R. Neuronal cell-type classification: challenges, opportunities and the path forward. *Nat. Rev. Neurosci.* **18**, 530–546 (2017).
- Cadwell, C. R. et al. Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat. Biotechnol.* **34**, 199–203 (2016).
- Maslund, R. H. Neuronal cell types. *Curr. Biol.* **14**, R497–R500 (2004).
- Ecker, J. R. et al. The BRAIN Initiative Cell Census Consortium: lessons learned toward generating a comprehensive brain cell atlas. *Neuron* **96**, 542–557 (2017).
- Mukamel, E. A. & Ngai, J. Perspectives on defining cell types in the brain. *Curr. Opin. Neurobiol.* **56**, 61–68 (2019).
- Kepecs, A. & Fishell, G. Interneuron cell types are fit to function. *Nature* **505**, 318–326 (2014).
- Rudy, B., Fishell, G., Lee, S. & Hjerling-Lefler, J. Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Dev. Neurobiol.* **71**, 45–61 (2011).
- Zeisel, A. et al. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).
- Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014. e22 (2018).
- Huang, Z. J. & Paul, A. The diversity of GABAergic neurons and neural communication elements. *Nat. Rev. Neurosci.* **20**, 563–572 (2019).
- Cadwell, C. R. et al. Multimodal profiling of single-cell morphology, electrophysiology, and gene expression using Patch-seq. *Nat. Protocols* **12**, 2531 (2017).
- Fuzik, J. et al. Integration of electrophysiological recordings with single-cell RNA-seq data identifies neuronal subtypes. *Nat. Biotechnol.* **34**, 175–183 (2016).
- Földy, C. et al. Single-cell RNAseq reveals cell adhesion molecule profiles in electrophysiologically defined neurons. *Proc. Natl Acad. Sci. USA* **113**, E5222–E5231 (2016).
- Yao, Z. et al. A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature* <https://doi.org/10.1038/s41586-021-03500-8> (2021).
- Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).
- van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
- Kobak, D. & Berens, P. The art of using t-SNE for single-cell transcriptomics. *Nat. Commun.* **5416**, 10 (2019).
- Gouwens, N. W. et al. Classification of electrophysiological and morphological neuron types in the mouse visual cortex. *Nat. Neurosci.* **22**, 1182–1195 (2019).
- Tricoire, L. et al. Common origins of hippocampal Ivy and nitric oxide synthase expressing neurogliaform cells. *J. Neurosci.* **30**, 2165–2176 (2010).
- Pelkey, K. A. et al. Hippocampal GABAergic inhibitory interneurons. *Physiol. Rev.* **97**, 1619–1747 (2017).
- Harris, K. D. et al. Classes and continua of hippocampal CA1 inhibitory neurons revealed by single-cell transcriptomics. *PLoS Biol.* **16**, e2006387 (2018).
- Scala, F. et al. Layer 4 of mouse neocortex differs in cell types and circuit organization between sensory areas. *Nat. Commun.* **10**, 4174 (2019).
- Muñoz, W., Tremblay, R., Levenstein, D. & Rudy, B. Layer-specific modulation of neocortical dendritic inhibition during active wakefulness. *Science* **355**, 954–959 (2017).
- Nigro, M. J., Hashikawa-Yamasaki, Y. & Rudy, B. Diversity and connectivity of layer 5 somatostatin-expressing interneurons in the mouse barrel cortex. *J. Neurosci.* **38**, 1622–1633 (2018).
- Yamawaki, N., Borges, K., Suter, B. A., Harris, K. D. & Shepherd, G. M. A genuine layer 4 in motor cortex with prototypical synaptic circuit connectivity. *eLife* **3**, e05422 (2014).
- Kobak, D. et al. Sparse reduced-rank regression for exploratory visualization of paired multivariate data sets. *R. Stat. Soc.* **70**, 980–1000 (2021).
- Naka, A. et al. Complementary networks of cortical somatostatin interneurons enforce layer specific control. *eLife* **8**, e43696 (2019).
- Zhang, M. et al. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* <https://doi.org/10.1038/s41586-021-03705-x> (2021).
- Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
- Cembrowski, M. S. et al. Spatial gene-expression gradients underlie prominent heterogeneity of ca1 pyramidal neurons. *Neuron* **89**, 351–368 (2016).
- Muñoz-Manchado, A. B. et al. Diversity of interneurons in the dorsal striatum revealed by single-cell RNA sequencing and PatchSeq. *Cell Rep.* **24**, 2179–2190.e7 (2018).
- Stanley, G. et al. Continuous and discrete neuron types of the adult murine striatum. *Neuron* **105**, 688–699.e8 (2019).

39. Wang, W. X. & Lefebvre, J. L. Morphological pseudotime ordering and fate mapping reveals diversification of cerebellar inhibitory interneurons. Preprint at <https://doi.org/10.1101/2020.02.29.971366> (2020).
40. De Marco García, N. V., Karayannis, T. & Fishell, G. Neuronal activity is required for the development of specific cortical interneuron subtypes. *Nature* **472**, 351–355 (2011).
41. Dehorter, N. et al. Tuning of fast-spiking interneuron properties by an activity-dependent transcriptional switch. *Science* **349**, 1216–1220 (2015).
42. Wamsley, B. & Fishell, G. Genetic and activity-dependent mechanisms underlying interneuron diversity. *Nat. Rev. Neurosci.* **18**, 299–309 (2017).
43. Lim, L., Mi, D., Llorca, A. & Marín, O. Development and functional diversification of cortical interneurons. *Neuron* **100**, 294–313 (2018).
44. Cadwell, C. R. et al. Development and arealization of the cerebral cortex. *Neuron* **103**, 980–1004 (2019).
45. Gouwens, N. W. et al. Integrated morphoelectric and transcriptomic classification of cortical GABAergic cells. *Cell* **183**, 1–19 (2020).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020



## Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment, unless otherwise stated.

### Animals

Experiments on adult male and female mice ( $n = 266$ ; median age 75 days, interquartile range 64–100, full range 35–245 days, Extended Data Fig. 2a) were performed on wild-type C57Bl/6 ( $n = 27$ ), *Viaat*-Cre/Ai9 (vesicular inhibitory amino acid transporter, encoded by the *Slc32a1* gene,  $n = 24$ ), *Sst*-Cre/Ai9 (somatostatin,  $n = 75$ ), *Vip*-Cre/Ai9 (vasoactive intestinal polypeptide,  $n = 46$ ), *Pvalb*-Cre/Ai9 (parvalbumin,  $n = 76$ ), *Npy*-Cre/Ai9 (neuropeptide Y,  $n = 2$ ), *Vipr2*-Cre/Ai9 (vasoactive intestinal peptide receptor 2,  $n = 7$ ), *Scl17a8*-Cre/Ai9 (VGLUT3, vesicular glutamate transporter 3,  $n = 6$ ), *Gnb4*-Cre/Ai9 ( $n = 1$ ), and *Slc17a8*-iCre/Ai9 ( $n = 2$ ) mice. Numbers above refer to mice from which sequencing data were successfully obtained. Several more animals were used for measuring layer boundaries and follow-up experiments at physiological temperature (see below). Mice were co-housed with littermates (2–5 per cage) in a controlled environment at 22–24 °C and 30–70% humidity. Mice were maintained with unrestricted access to food and water on a 12-h light/dark cycle. Procedures for mouse maintenance and surgeries were performed according to protocols approved by the Institutional Animal Care and Use Committee (IACUC) of Baylor College of Medicine.

The *Viaat*-Cre line was generously donated by Huda Zoghbi (Baylor College of Medicine), the *Slc17a8*-iCre line by Rebecca Seal (University of Pittsburgh). The *Gnb4*-Cre line was from the Allen Institute for Brain Science. The other Cre and reporter lines were purchased from the Jackson Laboratory: *Sst*-Cre (stock no. 013044), *Vip*-Cre (stock no. 010908), *Pvalb*-Cre (stock no. 008069), *Vipr2*-Cre (stock no. 031332), *Slc17a8*-Cre (stock no. 028534), *Npy*-Cre (stock no. 027851), Ai9 reporter (stock no. 007909).

We were unable to find any labelled cells in MOP in the *Gnb4*-Cre mice: all labelled cells were far outside of MOP and close to the claustrum<sup>46</sup>. For this reason, the data set does not include any *Gnb4*-positive cells.

### Slice preparation

The MOP brain slices were obtained following previously described protocols<sup>5,28</sup>. In brief, the animals were deeply anaesthetized using 3% isoflurane and decapitated. The brain was rapidly removed and collected into cold (0–4 °C) oxygenated NMDG (*N*-methyl-D-glucamine) solution containing 93 mM NMDG, 93 mM HCl, 2.5 mM KCl, 1.2 mM NaH<sub>2</sub>PO<sub>4</sub>, 30 mM NaHCO<sub>3</sub>, 20 mM HEPES, 25 mM glucose, 5 mM sodium ascorbate, 2 mM thiourea, 3 mM sodium pyruvate, 10 mM MgSO<sub>4</sub> and 0.5 mM CaCl<sub>2</sub>, pH 7.35 (all from Sigma-Aldrich). We cut 300- $\mu$ m-thick coronal slices using a Leica VT1200 microtome following coordinates provided in the Allen Brain Atlas for adult mouse (<http://atlas.brain-map.org>). The slices were subsequently incubated at  $34.0 \pm 0.5$  °C in oxygenated NMDG solution for 10–15 min before being transferred to the artificial cerebrospinal fluid (ACSF) solution containing: 125 mM NaCl, 2.5 mM KCl, 1.25 mM NaH<sub>2</sub>PO<sub>4</sub>, 25 mM NaHCO<sub>3</sub>, 1 mM MgCl<sub>2</sub>, 11.1 mM glucose and 2 mM CaCl<sub>2</sub>, pH 7.4 (all from Sigma-Aldrich) for about 1 h. The slices were allowed to recover in ACSF equilibrated with CO<sub>2</sub>/O<sub>2</sub> gas mixture (5% CO<sub>2</sub>, 95% O<sub>2</sub>), at room temperature (approximately 25 °C) for 1 h before experiments. During the recordings, slices were submerged in a customized chamber continuously perfused with oxygenated physiological solution. Recorded cells were generally located 15–60  $\mu$ m deep under the slice surface.

### Patch-seq recording procedures

In order to simultaneously obtain electrophysiological, morphological and transcriptomic data from the same neurons, we applied our recently developed Patch-seq protocol<sup>17</sup>, with some modifications. In particular,

changes were made to the internal solution to optimize its osmolarity in order to improve staining quality. RNase-free intracellular solution was prepared as follows: we dissolved 111 mM potassium gluconate, 4 mM KCl, 10 mM HEPES and 0.2 mM EGTA in RNase-free water in a 125-ml Erlenmeyer flask. We then covered the solution with aluminium foil and autoclaved it. After the solution was cooled down to room temperature, we added 4 mM MgATP, 0.3 mM Na<sub>3</sub>GTP, 5 mM sodium phosphocreatine, and 13.4 mM biocytin (all from Sigma-Aldrich). The pH was adjusted to 7.25 with RNase-free 0.5 M KOH using a dedicated pH meter (cleaned with RNase Zap and RNase-free water before each use). RNase-free water was then added to the solution in order to obtain the desired volume. After carefully checking its osmolarity (approximately 235–240 mOSM) the solution was stored at –20 °C and used for no longer than 3 weeks.

Before each experiment, we combined 494  $\mu$ l internal solution with 6  $\mu$ l recombinant RNase inhibitor (1 U/ $\mu$ l, Takara) to increase RNA yield. The addition of the inhibitor resulted in an increase in osmolarity to the desired value of 315–320 mOSM without a further dilution<sup>17</sup>. The osmolarity of the ACSF was monitored before each experiment and adjusted to be 18–20 mOSM lower than the internal solution. In particular, when the ACSF osmolarity was too low, we added a small amount of sucrose to ACSF to increase its osmolarity and bring it to the desired range. This osmolarity difference between ACSF and the internal solution is important to obtain slight swelling of the cell during the recording session, which improves the diffusion of biocytin in the neuronal processes. All glassware, spatulas, stir bars, counters, and anything else that may come into contact with the reagents or solution were cleaned thoroughly with RNase Zap before use.

Recording pipettes (B200-116-10; Sutter Instrument) of –3–7 M $\Omega$  resistance were filled with 0.1–0.3  $\mu$ l RNase-free intracellular solution. The size of the pipette tip was chosen according to the target neuron size: 3–4-M $\Omega$  pipettes were used to record large neurons (for example, L5 ET excitatory neurons) and 6–7-M $\Omega$  pipettes were used to record small cells such as L1 or *Vip* interneurons.

The PatchMaster software (HEKA Elektronik) and custom Matlab scripts were used to operate the Quadro EPC 10 amplifiers and to perform online and offline data analysis. We used the following quality control criteria: (1) seal resistance value >1 G $\Omega$  before achieving whole-cell configuration; (2) access resistance <30 M $\Omega$ . Each neuron was injected with 600-ms-long current pulses starting from –200 pA and up to 1,380 pA with 20-pA increment steps (in some cases stimulation was stopped before reaching 1,380 pA). There were 1.3- or 1.4-s intervals between successive current pulses, depending on the used setup. For most neurons, the stimulation was then repeated multiple times from the beginning. Electrophysiological traces used for the analysis were acquired between 3 and 15 min after achieving the whole-cell configuration. Recordings were performed at room temperature (25 °C), as opposed to physiological temperature (34 °C), in order to keep the cells alive for longer. We performed control experiments at physiological temperature as well (see below).

Typically, excitatory neurons were recorded for 5–20 min while interneurons were recorded for 20–50 min in order to allow biocytin to diffuse into distal axonal segments. During the recording, the access resistance was checked every three minutes in order to maintain a stable seal that would ensure successful biocytin diffusion. The resulting cDNA yield was not correlated with the hold time (Spearman correlation –0.01).

### Experiments at physiological temperature

A subset of electrophysiological recordings was performed at 34 °C in the presence of fast glutamatergic and GABAergic synaptic transmission blockers, 1 mM kynurenic acid (Sigma-Aldrich) and 0.1 mM picrotoxin (Tocris), respectively. The temperature was maintained stable, and constantly monitored using the temperature controller TCO7 (Luigs and Neumann). In this set of experiments, the morphologies were not recovered and multiple neurons were recorded in each slice. The soma

# Article

depth and the slice thickness were measured before each recording using Linlab2 software (Scientifica). Intrinsic electrophysiological recordings were obtained using the same stimulation paradigm as described above.

In these experiments, we targeted L5 *Sst* and excitatory neurons (Extended Data Fig. 6). We sequenced in total 185 neurons, obtained from 8 adult mice (7 *Sst*-Cre/Ai9 and 1 *Pvalb*-Cre/Ai9), of which 177 neurons passed the transcriptomic quality control and got a t-type assignment (see below). One hundred and ten cells mapped to the *Sst* subclass, 43 to IT, 12 to ET, 10 to *Pvalb*, and 2 to NP. 175 cells were assigned to L5 in the post hoc analysis (see below). We obtained high-quality electrophysiological recordings and extracted electrophysiological features of 184 cells.

## RNA sequencing of patched cells

At the end of the recording session, cell contents were aspirated into the glass pipette by applying a gentle negative pressure (0.7–1.5 pounds per square inch) for 1–5 min until the size of the cell body was visibly reduced. In most cases, the cell nucleus was visibly attached to the pipette tip and extracted from the cell body. We avoided complete nucleus aspiration, because it can lead to the collapse of the soma structure and of the nearby neurites, resulting in lower staining quality and stronger background staining. During the aspiration process, the cell body structure and access resistance were constantly monitored. Special care was taken to ensure that the seal between the pipette and the cell membrane remained intact to reduce contamination from the extracellular environment. After aspiration, the contents of the pipette were immediately ejected into a 0.2-ml PCR tube containing 4  $\mu$ l lysis buffer (with ERCC spike-ins), and RNA was subsequently converted into cDNA using a Smart-seq2-based protocol<sup>21</sup> as described previously<sup>17</sup>. The resulting cDNA libraries were screened using an Agilent Bioanalyzer 2100. Samples containing less than around 1 ng total cDNA (in the 15  $\mu$ l final volume) or with an average size less than 1,500 bp were typically not sequenced (with some occasional exceptions). The cDNA libraries were then frozen and sent for sequencing in 12 separate batches.

The cDNA libraries derived from each neuron were purified and 0.2 ng of the purified cDNA was tagmented using the Illumina Nextera XT Library Preparation with one-fifth of the volumes stated in the manufacturer's recommendation. Custom 8-bp index primers were used at a final concentration of 0.1  $\mu$ M. The resulting cDNA library of each batch was sequenced on an Illumina NextSeq500 instrument with a sequencing setup of 75-bp single-end reads and 8-bp index reads. The investigators were blinded to the cell type of each sample during library construction and sequencing.

The sequencing data were processed using the zUMIs 2.5.6b pipeline with default settings<sup>47</sup>. Sequencing reads were aligned to the mm10 mouse reference genome using STAR version 2.5.4b<sup>48</sup> and transcript assignment performed with Gencode transcript annotations, version M23. A substantial portion of the RNA extracted from the neurons was nascent and contained intronic sequences. To accommodate this, gene expression counts were separately calculated using reads mapping to annotated intronic and exonic regions. We detected 42,466 genes, including pseudogenes and annotated non-coding segments, in at least one cell. The resulting exonic and intronic read count data were used for all transcriptomic analyses presented here. To quantify gene expression, we typically normalized exon and intron counts by exonic and intronic gene lengths in kilobases and added normalized counts together to obtain normalized exonic + intronic expression levels. See below for more details. Throughout the manuscript, 'detected gene' refers to a gene with a non-zero exonic or intronic count.

## Biocytin staining and morphological reconstructions

Morphological recovery was carried out as previously described<sup>5,17,28</sup>. In brief, after the recordings, the slices were immersed in freshly prepared 2.5% glutaraldehyde, 4% paraformaldehyde solution in 0.1 M PBS at

4 °C for at least 48 h. The slices were subsequently processed with the avidin-biotin-peroxidase method to reveal the morphology of the neurons. As previously described, we took several steps to improve the staining quality of the fine axonal branches of interneurons<sup>5,17</sup>. First, we used a high biocytin concentration (0.5 g/100 ml). Second, we incubated with avidin-biotin complex and detergents at a high concentration (Triton X-100, 5%) for at least 24 h before staining with 3,3'-diaminobenzidine (DAB).

Recovered cells were manually reconstructed using a 100 $\times$  oil-immersion lens and a camera lucida system (MicroBrightField). We aimed to reconstruct all cells that had staining of sufficient quality (axons and dendrites for the inhibitory neurons; only dendrites for the excitatory neurons), and obtained 646 reconstructions in total. In addition, we reconstructed the dendrites of 30 neurons from the *Vip* and *Scng* subclasses that lacked sufficient axonal staining. *Vip* neurons are traditionally classified on the basis of dendritic morphology, so these reconstructions can inform t-type characterizations. These additional 30 reconstructions are shown, together with the main 646 reconstructions, in Supplementary File 1.

Forty-five sequenced cells were mistakenly recorded using a solution with a much smaller concentration of biocytin, and their morphologies could not be recovered. We made sure that the measured electrophysiological properties of these cells were not systematically different from those of the other sequenced cells.

Inevitably, neuronal structures can be severed as a result of the slicing procedure. We took special care to exclude reconstructions of all neurons that showed any signs of damage, lack of contrast, or poor overall staining. Consistently with previous studies, tissue shrinkage due to the fixation and staining procedures was about 10–20%<sup>5,28,49</sup>. This shrinkage was not compensated for in our analysis.

## Cortical thickness normalization and layer assignment

Nissl-stained slices ( $n = 15$  from two wild-type adult mice) were used to measure normalized layer boundaries in MOp. The Nissl staining protocol was adapted from ref.<sup>50</sup>. In brief, brain slices were mounted on slides and allowed to dry. The sections were then demyelinated, stained with 0.1% cresyl violet-acetate (C5042, Sigma) for 30 min at 60 °C and further destained. The sections were then coverslipped in Cytoseal 60 (Richard Allan Scientific). For each slice we measured total thickness from pia to white matter and the depths of the three between-layer boundaries (L1 to L2/3, L2/3 to L5, L5 to L6), based on the cortical cytoarchitecture, using a NeuroLucida system with 10 $\times$  or 20 $\times$  magnification. All measurements were normalized by the respective slice thickness, and the averages over all  $n = 15$  slices were used as the normalized layer boundaries (Extended Data Fig. 2b).

For the Patch-seq neurons, we measured soma depth and the cortical thickness of the slice using a NeuroLucida system. We took their ratio as the normalized soma depth, and assigned each neuron to a layer (L1, L2/3, L5, or L6) based on the Nissl-determined layer boundaries (Extended Data Fig. 2b). We obtained soma depth information for 1,284 neurons out of 1,329 (45 neurons were mistakenly recorded using a solution with insufficient biocytin concentration, and we could measure soma depths for only 2 of those; for 2 other neurons the measurements could not be carried out because the slices were lost). For the 45 neurons with missing soma depth measurements, we used the layer targeted during the recording for all layer-based analyses and visualizations (marker shapes in Figs. 1c–e, 3a–c, layer-restricted analysis in Fig. 4, Extended Data Fig. 8).

All reconstructed morphologies were normalized by the cortical thickness of the respective slice to make it possible to display several morphologies next to each other, as in Extended Data Fig. 3.

## t-Type assignment

The t-type assignment procedure was done in two rounds. The first round was for quality control and initial assignment to one of the three

large transcriptomic groups (CGE-derived interneurons, MGE-derived interneurons, and excitatory neurons) that are perfectly separated from each other with no transcriptomically intermediate cells<sup>4</sup>. The second round was done to assign the cells to specific t-types.

In the first round, we mapped each Patch-seq cell to a large annotated Smart-seq2 reference data set from adult mouse cortex<sup>4</sup>, using a procedure similar to the one described in ref.<sup>28</sup>. Specifically, using the exon count matrix of the reference data set, we selected the 3,000 most variable genes (see below). We then normalized all exon counts by exonic gene lengths in kilobases, all intron counts by intronic gene lengths in kilobases (plus  $10^{-6}$ , to avoid division by zero) and added normalized counts together to obtain normalized exonic + intronic expression levels. We log-transformed these values using  $\log_2(x+1)$  transformation and averaged the log-transformed values across all cells in each of the 133 t-types, to obtain reference transcriptomic profiles of each t-type ( $133 \times 3,000$  matrix). Out of these 3,000 genes, 2,666 were present in the genome annotation that we used and were detected in our data set. We applied the same normalization and log-transformation procedure to the exonic and intronic read counts of our cells, and for each cell computed Pearson correlation across the 2,666 genes with each of the 133 t-types. Each cell was assigned to the t-type to which it had the highest correlation (Extended Data Fig. 1d).

Cells meeting any of the following exclusion criteria were declared low quality and did not get a t-type assignment (Extended Data Fig. 2e): cells with the highest correlation below 0.4 (78 cells); cells that would be assigned to non-neural t-types, presumably owing to RNA contamination<sup>51</sup> (14 cells; see also Extended Data Fig. 2j–n); cells with the highest correlation less than 0.02 above the maximal correlation in one of the other two large transcriptomic groups (5 cells). The remaining 1,232 cells passed quality control and entered the second round.

In the second round, cells were independently mapped to the seven transcriptomic data sets obtained from mouse MOp<sup>20</sup>. The mapping was done only to the t-types from the transcriptomic group identified in the first round, using the 500 most variable genes in that data set for that transcriptomic group (so using  $7 \times 3 = 21$  sets of 500 most variable genes). Gene selection was performed as described below, and t-type assignment was done exactly as described above. Across the 21 reference subsets, 421–494 most variable genes were present in our data set, and were used for the t-type assignment (Extended Data Fig. 1e). When mapping to the Smart-seq2 reference data sets, we used normalized intronic and exonic reference counts, as above. When mapping to the UMI-based reference data sets, we used the unique molecular identifier (UMI) counts directly, without gene length normalization.

We used bootstrapping over genes to assess the confidence of each t-type assignment. For each cell and for each of the seven reference data sets, we repeatedly selected a bootstrap sample of genes (that is, the same number of genes, selected randomly with repetitions) and repeated the mapping. This was done 100 times and the fraction of times the cell mapped to each t-type was taken as the t-type assignment confidence for that t-type (Extended Data Fig. 1f). The confidences obtained with seven reference data sets agreed well with each other (Extended Data Fig. 2i) and were averaged to obtain the consensus confidence. Finally, the cell was assigned to the t-type with the highest consensus confidence.

Four cells were assigned to an excitatory t-type, despite having clearly inhibitory firing, morphology, and/or soma depth location (such as L1). The most likely cause of this was RNA contamination from excitatory cells, which are much more abundant than inhibitory cells in the mouse cortex (Extended Data Fig. 2). These four cells were excluded from all analyses and visualizations (as if they did not pass the transcriptomic quality control). In addition, one cell was probably located outside MOp, based on the slice anatomy, and was excluded as well. The final number of cells with t-type assignment was 1,227.

### Selection of most variable genes

Several steps of our analysis required selecting a set of the most variable genes in a given transcriptomic data set. We always selected a fixed predefined number of genes (such as 500, 1,000, or 3,000).

To select the most variable genes, we found genes that had, at the same time, high non-zero expression and a high probability of near-zero expression<sup>52</sup>. Our procedure is described in more detail elsewhere<sup>23</sup>. Specifically, we excluded all genes that had counts of at least  $c_{\min}$  (for Patch-seq and Smart-seq2:  $c_{\min} = 32$ ; for 10x:  $c_{\min} = 0$ ) in fewer than 10 cells. For each remaining gene we computed the mean  $\log_2$  count across all counts that were larger than  $c_{\min}$  (non-zero expression,  $\mu$ ) and the fraction of counts that were smaller than or equal to  $c_{\min}$  (probability of near-zero expression,  $\tau$ ). Across genes, there was a clear inverse relationship between  $\mu$  and  $\tau$ , that roughly followed the exponential law:  $\tau \approx \exp(-1.5 \times \mu + a)$  for some horizontal offset  $a$ . Using a binary search, we found a value  $b$  of this offset that yielded the desired number of genes with  $\tau > \exp(-1.5 \times \mu + b) + 0.002$ .

For Smart-seq2 and Patch-seq data sets, we used only exonic counts to perform gene selection.

### t-SNE visualization of the transcriptomic data

t-SNE embeddings<sup>22</sup> of the three subsets of the single-cell 10x v2 data set<sup>20</sup> (Fig. 1c–e) were constructed using the same 500 most variable genes that were used for t-type assignment (see above). The UMI counts were normalized by each cell's sequencing depth (sum of counts), multiplied by the median sequencing depth across all cells,  $\log_2(x+1)$ -transformed, and reduced to 50 principal components. The resulting  $n \times 50$  matrix was used as input to t-SNE. We used Fit-SNE 1.2.1<sup>53</sup> with default parameters (including learning rate  $n/12$  and scaled principal component analysis (PCA) initialization<sup>23</sup>). Perplexity was left at the default value of 30 for both inhibitory subsets and increased to 100 for the excitatory subset.

To position Patch-seq cells on a reference t-SNE embedding, we used a published procedure<sup>23</sup>. In brief, each cell was positioned at the median embedding location of its ten nearest neighbours, based on Pearson correlation distance in the high-dimensional space. As above, we used the sum of the normalized exonic and intronic counts for Patch-seq cells, and raw UMI counts for the reference cells. All values were  $\log_2(x+1)$ -transformed and correlations were computed across the same genes that were used for t-type assignments (see above).

### Extraction of electrophysiological features

Twenty-nine electrophysiological properties of the neurons were automatically extracted based on the raw membrane voltage traces (Extended Data Fig. 4) using Python scripts from the Allen Software Development Kit (SDK) (<https://github.com/AllenInstitute/AllenSDK>) with some modifications to account for our experimental paradigm (<https://github.com/berenslab/EphysExtraction>).

For each hyperpolarizing current injection, the resting membrane potential was computed as the mean membrane voltage during 100 ms before stimulation onset and the input resistance as the difference between the steady state voltage and the resting membrane potential, divided by the injected current value (we took the average voltage of the last 100 ms before stimulus offset as steady state). The median of these values over all hyperpolarizing traces was taken as the final resting membrane potential and input resistance, respectively.

To estimate the rheobase (the minimum current needed to elicit any spikes), we used robust regression (random sample consensus algorithm, as implemented in `sklearn.linear_model.RANSACRegressor`) of the spiking frequency onto the injected current using the five lowest depolarizing currents with non-zero spike count (if there were fewer than five, we used those available). The point at which the regression line crossed the x-axis gave the rheobase estimate (Extended Data Fig. 4). We restricted it to be between the highest injected current that



# Article

elicited no spikes and the lowest injected current that elicited at least one spike. If the regression line crossed the  $x$ -axis outside this interval, the first current step that elicited at least one spike was used.

The action potential (AP) threshold, AP amplitude, AP width, afterhyperpolarization (AHP), afterdepolarization (ADP), the first AP latency, and the upstroke-to-downstroke ratio (UDR) were computed as illustrated in Extended Data Fig. 4, using the first AP fired by the neuron. AP width was computed at the AP half-height. UDR refers to the ratio of the maximal membrane voltage derivative during the AP upstroke to the maximal absolute value of the membrane voltage derivative during the AP downstroke. We also computed the first AP latency at 20 pA current above the smallest current stimulation value that elicited a spike.

The interspike interval (ISI) adaptation index for each trace was defined as the ratio of the second ISI to the first one. The ISI average adaptation index was defined as the mean of ISI ratios corresponding to all consecutive pairs of ISIs in that trace. For both quantities we took the median over the five lowest depolarizing currents that elicited at least three spikes (if fewer than five were available, we used all of them). AP amplitude adaptation index and AP amplitude average adaptation index were defined analogously to the two ISI adaptation indices, but using the ratios of consecutive AP amplitudes (and using the median over the five lowest depolarizing currents that elicited at least two spikes).

The maximum number of APs refers to the number of APs emitted during the 600-ms stimulation window of the highest firing trace. The spike frequency adaptation (SFA) denotes the ratio of the number of APs in the second half of the stimulation window to the number of APs in the first half of the stimulation window of the highest firing trace. If the highest firing trace had fewer than five APs, SFA was not defined. Here and below the highest firing trace corresponds to the first depolarizing current step that showed the maximum number of APs during the current stimulation window (after excluding all stimulation currents for which at least one AP was observed in 100 ms before or in 200 ms after the stimulation window; see below).

The membrane time constant ( $\tau$ ) was computed as the time constant of the exponential fit to the membrane voltage from the stimulation onset to the first local minimum (we took the median over all hyperpolarizing traces). Three further features described the sag of the first (the lowest) hyperpolarization trace. The sag ratio was defined as the difference between the sag trough voltage (average voltage in a 5-ms window around the sag trough) and the resting membrane potential, divided by the steady state membrane voltage difference from the resting membrane potential. The sag time was defined as the time period between the first and the second moments at which the membrane voltage crossed the steady-state value after the stimulation onset. The sag area refers to the absolute value of the integral of the membrane voltage minus the steady-state voltage during the sag time period (Extended Data Fig. 4). If the sag trough voltage and the steady-state voltage differed by less than 4 mV, the sag time and sag area were set to zero.

The rebound was defined as the voltage difference between the resting membrane potential and the average voltage over 150 ms (or whatever time remained until 300 ms after the stimulation offset) after rebound onset, which we identified as the time point after stimulation offset at which the membrane voltage reached the value of the resting membrane potential. If the membrane voltage never reached the resting membrane potential during the 300 ms after the stimulation offset, the rebound was set to zero. The rebound number of APs was defined as the number of APs emitted during the same period of time. Both rebound features were computed using the lowest hyperpolarization trace.

The ISI coefficient of variation (CV) refers to the standard deviation divided by the mean of all ISIs in the highest firing trace. Note that a Poisson firing neuron would have ISI CV equal to one. The ISI Fano factor refers to the variance divided by the mean of all ISIs in the highest firing

rate. The AP CV and AP Fano factor refer to the CV and the Fano factor of the AP amplitudes in the highest firing trace, respectively.

The burstiness was defined as the difference between the inverse of the smallest ISI within a detected burst and the inverse of the smallest ISI outside bursts, divided by their sum. We took the median over the first five depolarizing traces. We relied on the Allen SDK code to detect the bursts. In brief, within that code a burst onset was identified whenever a 'detour' ISI was followed by a 'direct' ISI. Detour ISIs are ISIs with a non-zero ADP or a drop of at least 0.5 mV of the membrane voltage after the first AP terminates and before the next one is elicited. Direct ISIs are ISIs with no ADP and no such drop of membrane voltage before the second AP. A burst offset was identified whenever a direct ISI was followed by a detour ISI. Additionally, bursts were required to contain no 'pauselike' ISIs, defined as unusually long ISIs for that trace (see Allen SDK for the implementation details).

Some neurons (in particular neurogliaform cells) started to emit APs before and after the current stimulation window, after the stimulation currents exceeded a certain amount. To quantify this effect, we defined wildness as the difference in the number of APs between the highest firing trace (possibly showing APs before or after the stimulation window) and the highest firing trace as defined above (without any APs outside the stimulation window). For most neurons, wildness was equal to zero.

For all statistical analysis we used 17 features out of the extracted 29, excluding features that were equal to zero for many cells (afterdepolarization, burstiness, rebound number of APs, sag area, sag time, wildness), two Fano factor features that were highly correlated with the corresponding coefficient of variation features (AP Fano factor, ISI Fano factor) and another measure of latency that was highly correlated with the latency itself, features that had very skewed distributions (AP amplitude average adaptation index, ISI average adaptation index), and features that were undefined for some of the cells (spike frequency adaptation). Four features were log-transformed to make their distribution more Gaussian-like: AP coefficient of variation, ISI coefficient of variation, ISI adaptation index, and latency.

## Extraction of morphological features

Reconstructed morphologies were converted into the SWC format using NLMorphologyConverter 0.9.0 (<http://neuronland.org>) and further analysed using MorphoPy (<https://github.com/berenslab/MorphoPy>, version 0.6)<sup>54</sup>. Each cell was soma-centred in the  $x$  (slice width) and  $y$  (slice depth) dimensions, and aligned to pia in the  $z$  (cortical depth) dimension so that  $z = 0$  corresponded to pia. All neurites were smoothed in the slice depth dimension ( $y$ ) using a Savitzky-Golay filter of order 3 and window length 21, after resampling points to have maximally 1  $\mu\text{m}$  spacing. For further analysis we computed two different feature representations of each cell: the normalized  $z$ -profile and a set of morphometric statistics<sup>24,28,55</sup>.

To compute the normalized  $z$ -profile, we divided all the coordinates of the neuronal point cloud by the thickness of the respective cortical slice, so that  $z = 1$  corresponded to the white matter border. We projected this point cloud onto the  $z$ -axis and binned it into 20 equal-sized bins spanning  $[0, 1]$ . The resulting histogram describes a neuron's normalized depth profile perpendicular to the pia. For the purposes of downstream analysis, we treated this as a set of 20 features. The  $z$ -profiles were separately computed for axons and dendrites.

Morphometric statistics were separately computed for the dendritic and axonal neurites to quantify their arborization shape and branching patterns. For the excitatory neurons, several additional morphometric statistics were computed for the apical dendrites, where apical dendrite was operationally defined as the dendrite with the longest total path length. We further used two 'somatic' features: normalized soma depth and soma radius. We did not use any features measuring

morphological properties in the slice depth ( $y$ ) direction because of possible slice shrinkage artefacts. We did not use any axonal features for the excitatory cells because only a small part of the axon could typically be reconstructed. For the inhibitory cells, where dendrite and axon could both be fully recovered, we included some measures of dendritic and axonal overlap. The full list of morphometric statistics is given in Supplementary File 3.

We extracted a set of 75 features, of which 40 were defined for excitatory neurons and 62 for inhibitory neurons, and processed the data for excitatory and inhibitory neurons separately. In each case, we excluded features with coefficient of variation below 0.25 (among the features with only positive values). This procedure excluded five features for the excitatory and nine features for the inhibitory cells. The distributions of the remaining features were visually checked for outliers and for meaningful variation between transcriptomic types, leading to a further exclusion of three features for the inhibitory cells. The full list of excluded features is given in Supplementary File 3. The resulting set of morphometric statistics used for further analysis consisted of 35 features defined for the excitatory neurons and 50 features defined for the inhibitory neurons.

### Reduced-rank regression

For the RRR analysis<sup>32</sup> we used 17 electrophysiological features and all 1,219 cells for which values for all 17 features and a t-type assignment could be computed. Electrophysiological features were standardized. Exon counts and intron counts were normalized by the exon/intron gene lengths as described above, summed together, converted to CPM,  $\log_2(x+1)$ -transformed, and then standardized. We selected the 1,000 most variable genes (using raw exonic counts) and used only those for the RRR analysis.

In brief, RRR finds a linear mapping of gene expression levels to a low-dimensional latent representation, from which the electrophysiological features are then predicted with another linear transformation (for mathematical details, see ref.<sup>32</sup>). The model uses sparsity constraints in the form of elastic net penalty to select only a small number of genes. For Fig. 2 we used a model with rank  $r=5$ , zero ridge penalty ( $\alpha=1$ ), and lasso penalty tuned to yield a selection of 25 genes ( $\lambda=0.5$ ). Cross-validation (Extended Data Fig. 5) was done using 10 folds, elastic net  $\alpha$ -values 0.5, 0.75, and 1.0, and  $\lambda$ -values from 0.2 to 6.0.

The plots shown in Fig. 2a, b are called biplots because they combine two biplots: the left biplot shows a mapping of gene expression levels onto the two latent dimensions; the right biplot shows the same mapping of electrophysiological features. To illustrate the meaning of the latent dimensions, each biplot combines the resulting scatter plots with lines showing how original features are related to the latent dimensions. Specifically, we computed the correlations of individual genes or electrophysiological properties with the latent dimensions and visualized these correlations as lines on the biplot. The circle shows the maximal possible correlation; only lines longer than 0.4 times the circle radius are shown in Fig. 2. Label positions were automatically adjusted by simulating repulsive forces between all overlapping pairs of labels, until there was no overlap.

For the model based on ion channel genes, we obtained the list of 328 ion channel genes from <https://www.genenames.org/data/genegroup/#!/group/177> and used all 307 of them that had non-zero expression in at least 10 of our cells. We used rank  $r=5$ ,  $\alpha=1$ , and  $\lambda$  tuned to yield 25 genes ( $\lambda=0.303$ ), as above.

### t-SNE visualization of the morpho-electric phenotypes

For the t-SNE visualization<sup>22</sup> of the electrophysiological phenotypes, we used 17 features as described above and all  $n=1,320$  cells that had values for all 17 features. All features were standardized across this set of cells and transformed with PCA into a set of 17 PCs. We scaled the PCs by the standard deviation of PC1. We used the t-SNE implementation

from scikit-learn Python library with the default perplexity (30), early exaggeration 4 (the default value 12 can be too large for small data sets), and scaled PCA initialization<sup>23</sup>. Fig. 3a shows  $n=1,219$  cells that had a t-type assignment.

For the t-SNE visualization of the morphological phenotypes, we combined morphometric statistics with the normalized z-profiles. The pre-processing, including PCA, was done separately for the excitatory and inhibitory neurons because they used different sets of morphometric statistics (see above). Only neurons with assigned t-types were used for this analysis. Two inhibitory neurons were left out because some of the morphometric statistics could not be extracted owing to insufficient dendritic recovery; this left 367 inhibitory neurons (with 50 morphometric features) and 269 excitatory neurons (with 35 morphometric features). All features were standardized and each set was reduced to 20 PCs. We scaled the PCs by the standard deviation of the respective PC1, to make the inhibitory and the excitatory PCs have comparable variances.

We used dendritic z-profiles for the excitatory neurons and axonal z-profiles for the inhibitory neurons. We reduced each set to five PCs, discarded PC1 (it was strongly correlated with the normalized soma depth and made the resulting embedding strongly influenced by the soma depth), and scaled the PCs by the standard deviation of the respective PC2. We stacked the 20 scaled morphometric PCs and the 4 scaled z-profile PCs to get a combined 24-dimensional representation, separately for the excitatory and for the inhibitory neurons. We then combined these representations into one block-diagonal 48-dimensional matrix. This procedure makes the excitatory and the inhibitory populations both have zero mean. To prevent overlap between these two populations, we added a small constant value of 0.25 to the excitatory block-diagonal block, leading to the strong excitatory–inhibitory separation in Fig. 3b. The t-SNE was performed exactly as described above.

For the t-SNE visualization of the morpho-electrical landscape, we stacked together the 48-dimensional morphological representation and the 16-dimensional electrophysiological representation obtained above, using only cells that had all morphological and all electrophysiological features ( $n=628$ ). We multiplied the electrophysiological block by  $\sqrt{2}$  to put its total variance on a similar scale (it only consisted of one set of scaled PCs, whereas the morphological representation consisted of two sets of scaled PCs: morphometrics and z-profiles). The resulting 64-dimensional morpho-electrical representation was used for t-SNE, exactly as described above.

### kNN classification of transcriptomic families

To classify neurons into transcriptomic families on the basis of electrophysiological, morphological, or combined features (Figs. 3d, 5a, Extended Data Fig. 8a), we used a kNN classifier with  $k=10$  and Euclidean distance metric (taking the majority family among the  $k$  nearest neighbours). This is effectively a leave-one-out cross-validation procedure. For each data modality we took the exact same data representation that was used for computing t-SNE embeddings (Fig. 3a–c; see above). Note that the t-SNE algorithm is also based on nearest neighbours and makes all close neighbours attract each other in the embedding. We chose the kNN classifier as a simple but versatile non-parametric classifier that is directly related to the t-SNE embeddings. We did not use the *Sncg* and NP families owing to insufficient coverage in our data set (Fig. 1).

Fig. 3d shows the fraction of cells from each family that was classified into each family. Fig. 5a and Extended Data Fig. 8a show fractions of cells from each t-type that were classified into each family. For morphological and combined features, Extended Data Fig. 8a shows fractions of cells from the majority layer of each t-type. For example, the *Pvalb Reln* type occurred most often in L5, so only cells from that layer were taken for that type. Only t-types with at least ten cells (or at least ten layer-restricted cells) are shown.

## Within-family analysis

To study the relationship between transcriptomic and electrophysiological distances between pairs of t-types (Fig. 4c, d, Extended Data Figs. 6, 7), we took all t-types with five or more cells assigned to them (for Extended Data Fig. 7a: with ten or more). For each pair of t-types, transcriptomic distance was computed as the Pearson correlation between the average  $\log_2(x+1)$ -transformed UMI counts in the single-cell 10x v2 data<sup>20</sup>. The 1,000 most variable genes across all neural types were used for Fig. 4c and Extended Data Fig. 7a, b, and the 500 most variable genes across the respective transcriptomic group (see above) were used for Fig. 4d and Extended Data Figs. 6i, j and 7c–n. Electrophysiological distance was computed as the Euclidean distance between the average feature vectors. Fig. 4d used the soma depth distance, computed as the absolute value of the difference between the average normalized soma depths.

## T-type variability analysis

The normalized total variance in Fig. 5b and Extended Data Fig. 8b was computed as follows. For each modality, we took the exact same data representation that was used for computing t-SNE embeddings (Fig. 3a–c; see above). For each t-type (or layer-restricted t-type; see above), we took the sum of its variances in all dimensions as the total variance and divided by the sum of variances in all dimensions across the whole data set:

$$\frac{\sum_j \frac{1}{|T|} \sum_{i \in T} \left( X_{ij} - \frac{1}{|T|} \sum_{i \in T} X_{ij} \right)^2}{\sum_j \frac{1}{n} \sum_i \left( X_{ij} - \frac{1}{n} \sum_i X_{ij} \right)^2},$$

where  $X_{ij}$  is a value of feature  $j$  of cell  $i$ ,  $n$  is the total number of cells, and  $T$  is the set of cell numbers belonging to the given t-type. The value 0 indicates that all cells from this t-type have exactly identical features. The value 1 indicates that there is as much variance in this one t-type as in the whole data set. Only t-types with at least ten cells (or at least ten layer-restricted cells) are shown in Fig. 5b and Extended Data Fig. 8b.

To provide a sensible baseline for the range of possible normalized total variances in a population of morpho-electrically homogeneous types, we used a clustering analysis. For the cells of all the  $K$  t-types (or layer-restricted t-types) with at least ten cells in a given panel, we used the  $k$ -means algorithm to cluster them into  $K$  clusters, reasoning that these clusters should be as homogeneous as possible given the variability in our data set. We used the  $k$ -means implementation from scikit-learn with default parameters. We then computed the normalized total variance of each cluster as described above. Grey shading in Fig. 5b and Extended Data Fig. 8b shows the interval between the minimum and the maximum cluster variances. Note that the  $k$ -means algorithm directly minimizes within-cluster total variances.

We used the entropies of a Leiden clustering<sup>35</sup> as an alternative way to approach the same question. For each modality, using the exact same data representation as above, we constructed its kNN graph with  $k=10$  and clustered it using the Leiden algorithm as implemented in the Python package `leidenalg` with `RBConfigurationVertexPartition` quality function and resolution parameter manually tuned to produce roughly the same number of clusters for each modality as in ref.<sup>24</sup>. (Extended Data Fig. 8). For each t-type (or layer-restricted t-type), we then measured the entropy of the distribution of electrophysiological or morphological cluster IDs, after randomly subsampling the t-type to ten cells. Subsampling was done to eliminate a possible bias due to the t-type abundance. The whole procedure was repeated 100 times with different random seeds for the Leiden clustering and for the subsampling.

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

All preprocessed data (gene counts, electrophysiological and morphological features) and meta data are available at <https://github.com/berenslab/mini-atlas>, together with direct links to the raw data. Electrophysiological recordings are available at <https://dandiarchive.org/dandiset/000008> (main data set) and <https://dandiarchive.org/dandiset/000035> (physiological temperature) in NWB format. Sequencing data are available at <http://data.nemoarchive.org/biccn/grant/zeng/tolias> in FASTQ format. Morphological reconstructions are available at <https://download.brainimagelibrary.org/3a/88/3a88a7687ab66069/> in SWC format.

## Code availability

The analysis code in Python is available at <https://github.com/berenslab/mini-atlas>.

- Peng, H. et al. Morphological diversity of single neurons in molecularly defined cell types. *Nature* <https://doi.org/10.1038/s41586-021-03941-1> (2021).
- Parekh, S., Ziegenhain, C., Vieth, B., Enard, W. & Hellmann, I. zUMIs - A fast and flexible pipeline to process RNA sequencing data with UMIs. *Gigascience* **7**, giy059 (2018).
- Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- Markram, H., Lübke, J., Frotscher, M., Roth, A. & Sakmann, B. Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *J. Physiol. (Lond.)* **500**, 409–440 (1997).
- Paul, C. A., Beltz, B. & Berger-Sweeney, J. The nissl stain: a stain for cell bodies in brain sections. *Cold Spring Harb. Protoc.* (2008).
- Tripathy, S. J. et al. Assessing transcriptome quality in patch-seq datasets. *Front. Mol. Neurosci.* **11**, 363 (2018).
- Andrews, T. S. & Hemberg, M. M3Drop: dropout-based feature selection for scRNASeq. *Bioinformatics* **35**, 2865–2867 (2019).
- Linderman, G. C. et al. Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data. *Nat. Methods* **16**, 243–245 (2019).
- Laturnus, S. MorphoPy: A python package for feature extraction of neural morphologies. *J. Open Sci. Software* **5**, 2339 (2020).
- Laturnus, S., Kobak, D. & Berens, P. A systematic evaluation of interneuron morphology representations for cell type discrimination. *Neuroinformatics* **18**, 591–609 (2020).
- Schuman, B. et al. Four unique interneuron populations reside in neocortical layer 1. *J. Neurosci.* **39**, 125–139 (2019).
- Prönneke, A. et al. Characterizing VIP neurons in the barrel cortex of VIPcre/tomato mice reveals layer-specific differences. *Cereb. Cortex* **25**, 4854–4868 (2015).
- Perrenoud, Q., Rossier, J., Geoffroy, H., Vitalis, T. & Gallopin, T. Diversity of GABAergic interneurons in layer VIa and VIb of mouse barrel cortex. *Cereb. Cortex* **23**, 423–441 (2013).
- Urban-Ciecko, J. & Barth, A. L. Somatostatin-expressing neurons in cortical networks. *Nat. Rev. Neurosci.* **17**, 401–409 (2016).
- Ascoli, G. A. et al. Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nat. Rev. Neurosci.* **9**, 557–568 (2008).
- Bortone, D. S., Olsen, S. R. & Scanziani, M. Translaminar inhibitory cells recruited by layer 6 corticothalamic neurons suppress visual cortex. *Neuron* **82**, 474–485 (2014).
- Oberlaender, M. et al. Cell type-specific three-dimensional structure of thalamocortical circuits in a column of rat vibrissal cortex. *Cereb. Cortex* **22**, 2375–2391 (2012).
- Marx, M. & Feldmeyer, D. Morphology and physiology of excitatory neurons in layer 6b of the somatosensory rat barrel cortex. *Cereb. Cortex* **23**, 2803–2817 (2013).
- Wang, Y., Ye, M., Kuang, X., Li, Y. & Hu, S. A simplified morphological classification scheme for pyramidal cells in six layers of primary somatosensory cortex of juvenile rats. *IBRO Rep.* **5**, 74–90 (2018).
- Kanari, L. et al. Objective morphological classification of neocortical pyramidal cells. *Cereb. Cortex* **29**, 1719–1735 (2019).
- Zhang, Z.-W. & Deschênes, M. Intracortical axonal projections of lamina VI cells of the primary somatosensory cortex in the rat: a single-cell labeling study. *J. Neurosci.* **17**, 6365–6379 (1997).
- Oswald, M. J. et al. Diversity of layer 5 projection neurons in the mouse motor cortex. *Front. Cell. Neurosci.* **7**, 174 (2013).
- Ramaswamy, S. & Markram, H. Anatomy and physiology of the thick-tufted layer 5 pyramidal neuron. *Front. Cell. Neurosci.* **9**, 233 (2015).
- Economo, M. N. et al. Distinct descending motor cortex pathways and their roles in movement. *Nature* **563**, 79–84 (2018).
- Thomson, A. M. Neocortical layer 6, a review. *Front. Neuroanat.* **4**, 13 (2010).

**Acknowledgements** We thank B. Dichter for help with converting electrophysiological recordings into the NWB format. This work was supported by the National Institute of Mental Health and the National Eye Institute of the National Institutes of Health (NIH) under Award Numbers U19MH114830 (A.S.T., R.S., P.B., H.Z.), R01MH109556 (A.S.T.), EY002520 (A.S.T.), T32EY07001, the Deutsche Forschungsgemeinschaft through a Heisenberg Professorship (BE5601/4-1; P.B.), the Cluster of Excellence “Machine Learning — New Perspectives for Science” (EXC 2064, project number 390727645; P.B.), the Collaborative Research Center 1233 “Robust Vision” (project number 276693517; P.B.), the German Federal Ministry of



Education and Research (FKZ O1GQ1601 and O1IS18039A; P.B.), and the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior/Interior Business Center (DoI/IBC), contract number D16PC00003 (A.S.T.). The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of NIH, IARPA, DoI/IBC, or the US Government.

**Author contributions** F.S., D.K., R.S., P.B. and A.S.T. designed the study; F.S. and M.B. performed Patch-seq experiments; S.M. performed Nissl stainings; F.S. and E.M. performed neuronal reconstructions aided by J.R.C.; J.R.C. prepared cDNA libraries and performed DAB stainings; L.H. performed sequencing and initial bioinformatic analysis under the supervision of R.S.; Z.Y. and H.Z. shared the MOp reference scRNA-seq datasets and transcriptomic taxonomy results; C.R.C. and X.J. helped to adjust the Patch-seq procedure; Y.B. analysed electrophysiological data; S.L. analysed morphological data; D.K. analysed transcriptomic data; D.K. and P.B. supervised data analysis; J.R.C. and Z.H.T. sustained animal colonies and

provided experimental support; F.S., D.K., Y.B., L.H., S.L., R.S., P.B. and A.S.T. discussed the results; D.K. prepared the figures; D.K. and P.B. wrote the manuscript with input from F.S., Y.B., S.L., L.H., C.R.C., R.S., H.Z., and A.S.T. P.B. and A.S.T. supervised all stages of the project. M.B., Y.B., C.R.C., J.R.C., L.H., X.J., S.L., E.M., S.M., Z.H.T., Z.Y., H.Z. are listed in the author list alphabetically.

**Competing interests** The authors declare no competing interests.

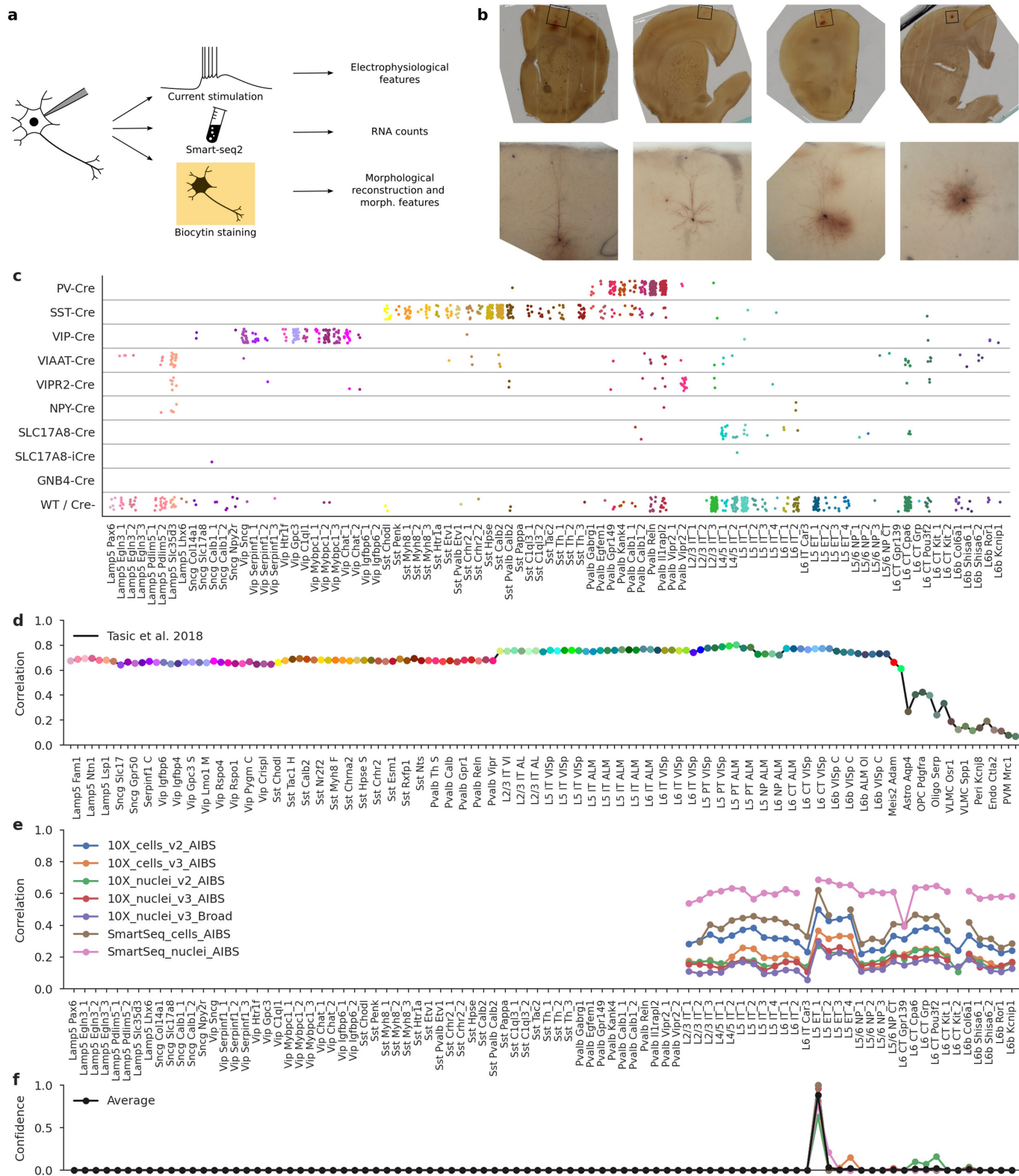
**Additional information**

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2907-3>.

**Correspondence and requests for materials** should be addressed to P.B. or A.S.T.

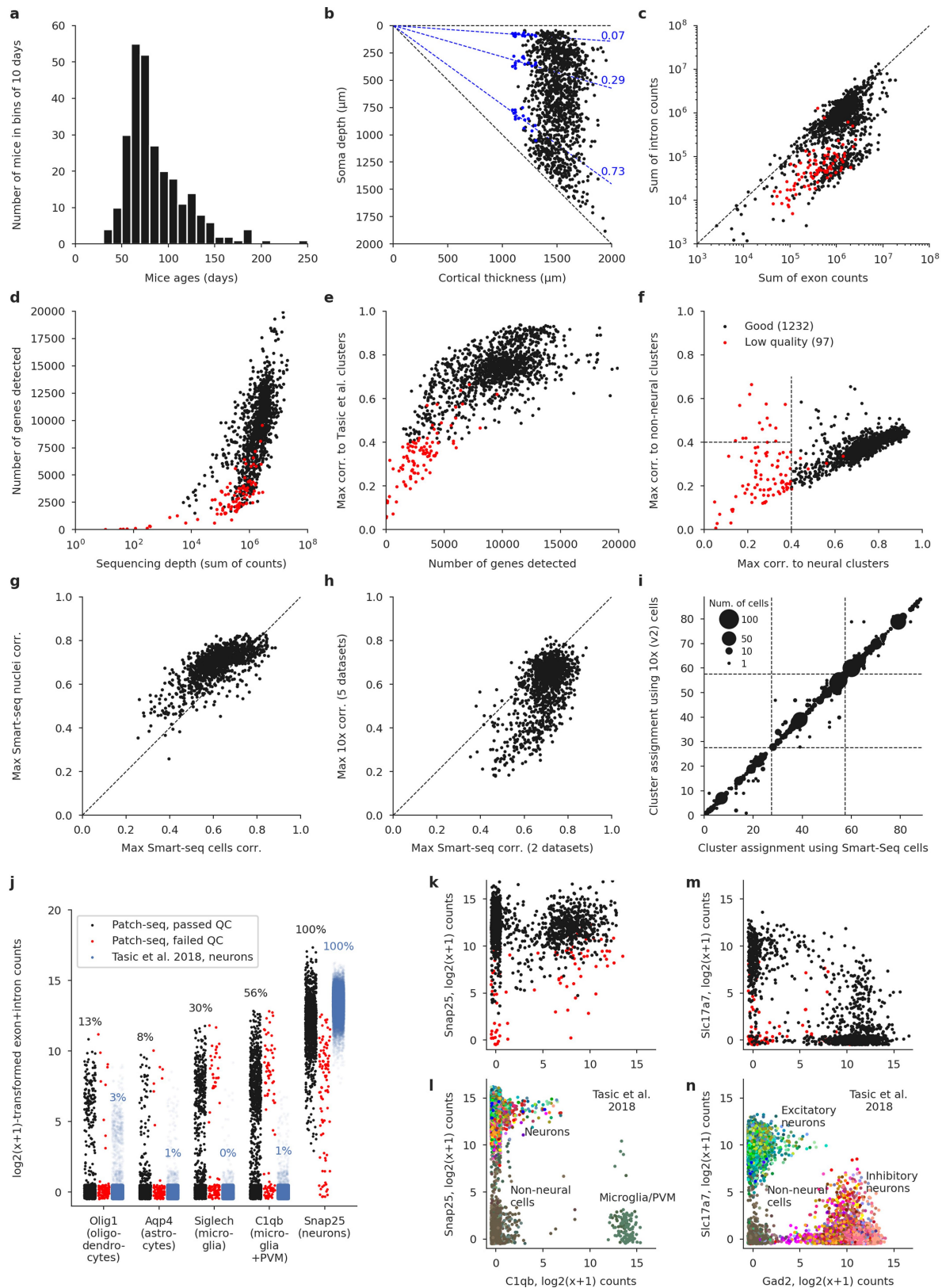
**Peer review information** *Nature* thanks Shreejoy Tripathy and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



**Extended Data Fig. 1 | Patch-seq protocol, mouse Cre lines, and t-type assignment.** **a**, Patch-seq combines electrophysiological recordings, RNA sequencing using Smart-seq2, and biocytin staining in the same cell. **b**, Four exemplary slice images. Top: an image of the whole slice using 4x magnification. Bottom: a flattened 3D image stack using 20x magnification. From left to right: L5 ET neuron, L2/3 IT neuron, L5 Sst neuron, L5 Pvalb neuron. **c**, t-Types assigned to cells collected in mice from different Cre lines. ‘WT/Cre-’ stands for cells from any Cre line that were not labelled with a fluorescent indicator, or for the cells patched in wild type mice. 1,227 cells shown. **d**, t-Type

assignment procedure for one example cell (**d-f**). Correlations to the mean log expression of all t-types from ref. <sup>4</sup>, using 3,000 most variable genes. Maximum correlation is to the excitatory neurons. t-Type names are shortened, and every second one is omitted for compactness. **e**, Correlations to all excitatory t-types from ref. <sup>20</sup> using all seven reference data sets and 500 most variable genes. **f**, t-Type assignment confidences for all seven data sets, obtained via bootstrapping over genes. The average confidence is shown in black. The mode of the average confidence was taken as the final t-type.



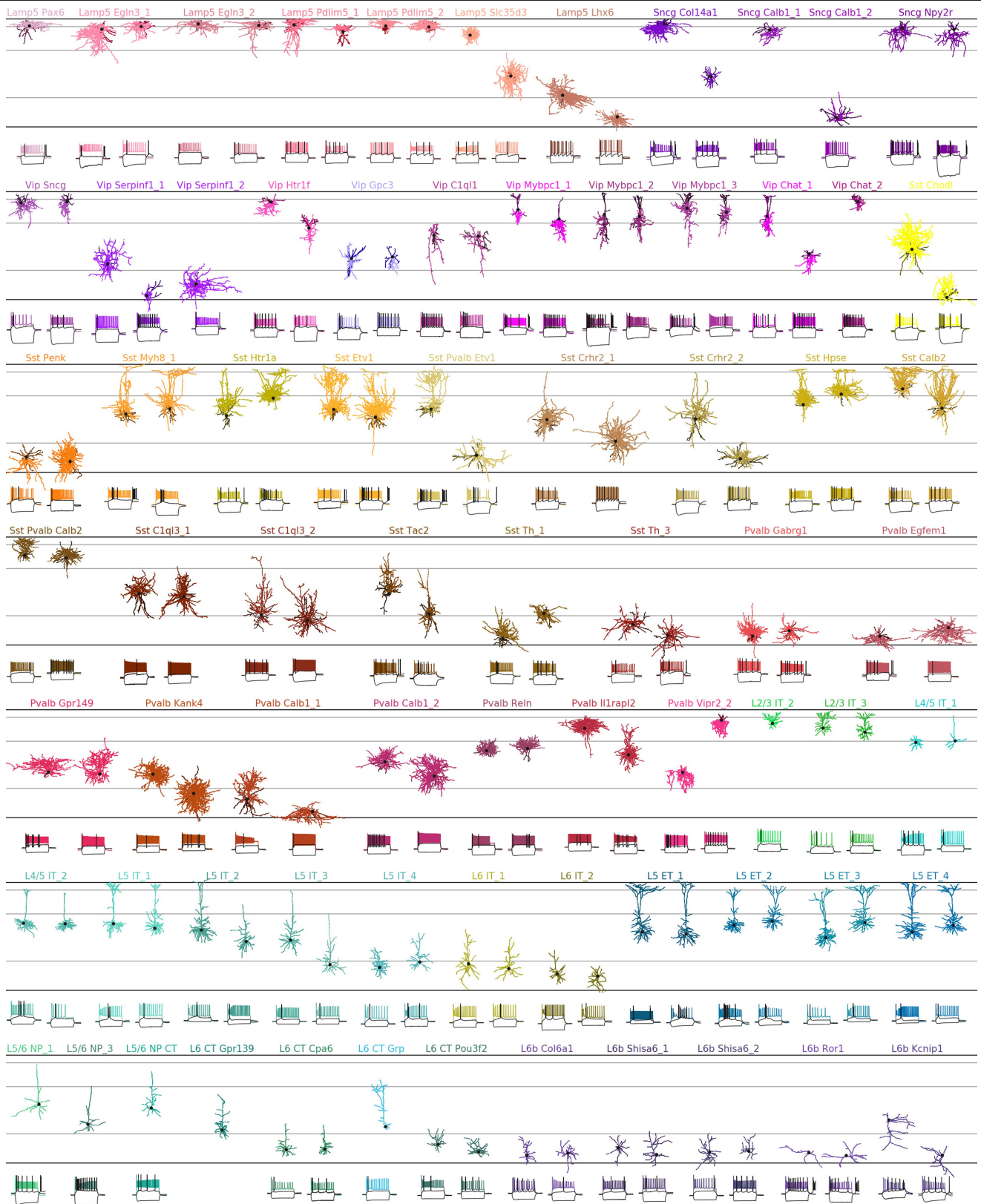
**Extended Data Fig. 2** | See next page for caption.

# Article

**Extended Data Fig. 2 | Quality control.** **a**, Age distribution of the mice used in the experiments. Median: 75 days. **b**, Soma depths of all cells and cortical thickness of the corresponding slices. Dashed lines show layer boundaries, based on the Nissl-stained slices (measured layer boundaries shown as blue points). All soma depths were normalized by dividing them by the cortical thickness. **c**, Relationship between the number of exonic and intronic counts. The apparent bimodality could be explained by whether the nucleus was extracted or not during Patch-seq aspiration. Whenever the nucleus was not extracted, low amount of nonspliced RNA led to low intronic counts; otherwise, the number of intronic and exonic counts was almost the same. Red: cells eventually failing quality control. **d**, Relationship between sequencing depth (total number of reads) and the number of detected genes (number of genes with non-zero counts). **e**, Relationship between the number of detected genes and the maximal correlation to clusters from ref.<sup>4</sup>. Cells with maximal correlation below 0.4 were declared low quality. **f**, Relationship between the maximal correlation across neural clusters and the maximal correlation across non-neural clusters from ref.<sup>4</sup>. Cells with maximal neural correlation below 0.4 were declared low quality. See Methods for additional QC criteria. **g**, Maximal correlations using single-cell and single-nucleus Smart-seq2 reference data sets<sup>20</sup>. **h**, Maximal correlations using Smart-seq2 reference data sets (maximum

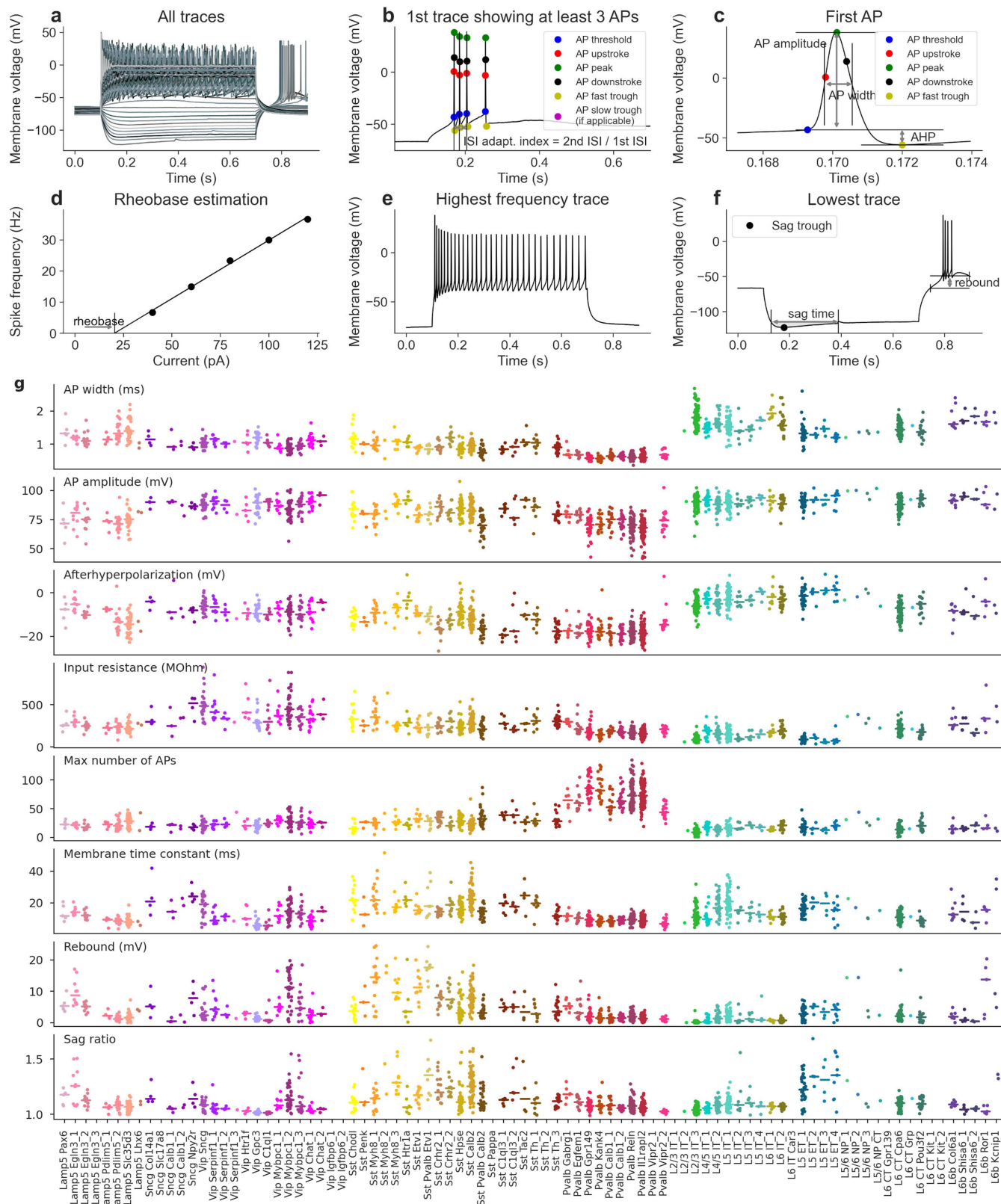
across cell types and across two data sets) and using 10x reference data sets (maximum across cell types and across five data sets). **i**, t-Type assignment using single-cell Smart-seq2 reference data set and using single-cell 10x v2 reference data set. All points are on the integer grid; marker size shows the number of cells at the corresponding location. Dashed lines separate CGE-derived interneurons, MGE-derived interneurons, and excitatory neurons. The mapping was done within each order, so there cannot be any cells outside of the diagonal blocks. **j**, Expression of several prominent markers of non-neural cells, in comparison to the Smart-seq2 data set from ref.<sup>4</sup>. The values are  $\log_2(x+1)$ -transformed sums of exonic and intronic counts, shown with random  $U\left(-\frac{1}{2}, \frac{1}{2}\right)$  jitter. Percentage values refer to the fraction of cells with non-zero counts. PVM stands for perivascular macrophages. We selected these markers because they have very low expression in neural cells. A neuronal marker *Snap25* is shown for comparison. Cells from the reference data set are shown with the alpha-level set to the ratio of our data set size to that data set size (0.06), to make the dot plots more comparable. **k**, **l**, Neural and glial expression in our data set (**k**) and in the FACS-sorted data set<sup>4</sup> (**l**) (plotted using the colours from the original publication, without transparency). **m**, **n**, The same using the excitatory marker *Slc17a7* and the inhibitory marker *Gad2*.





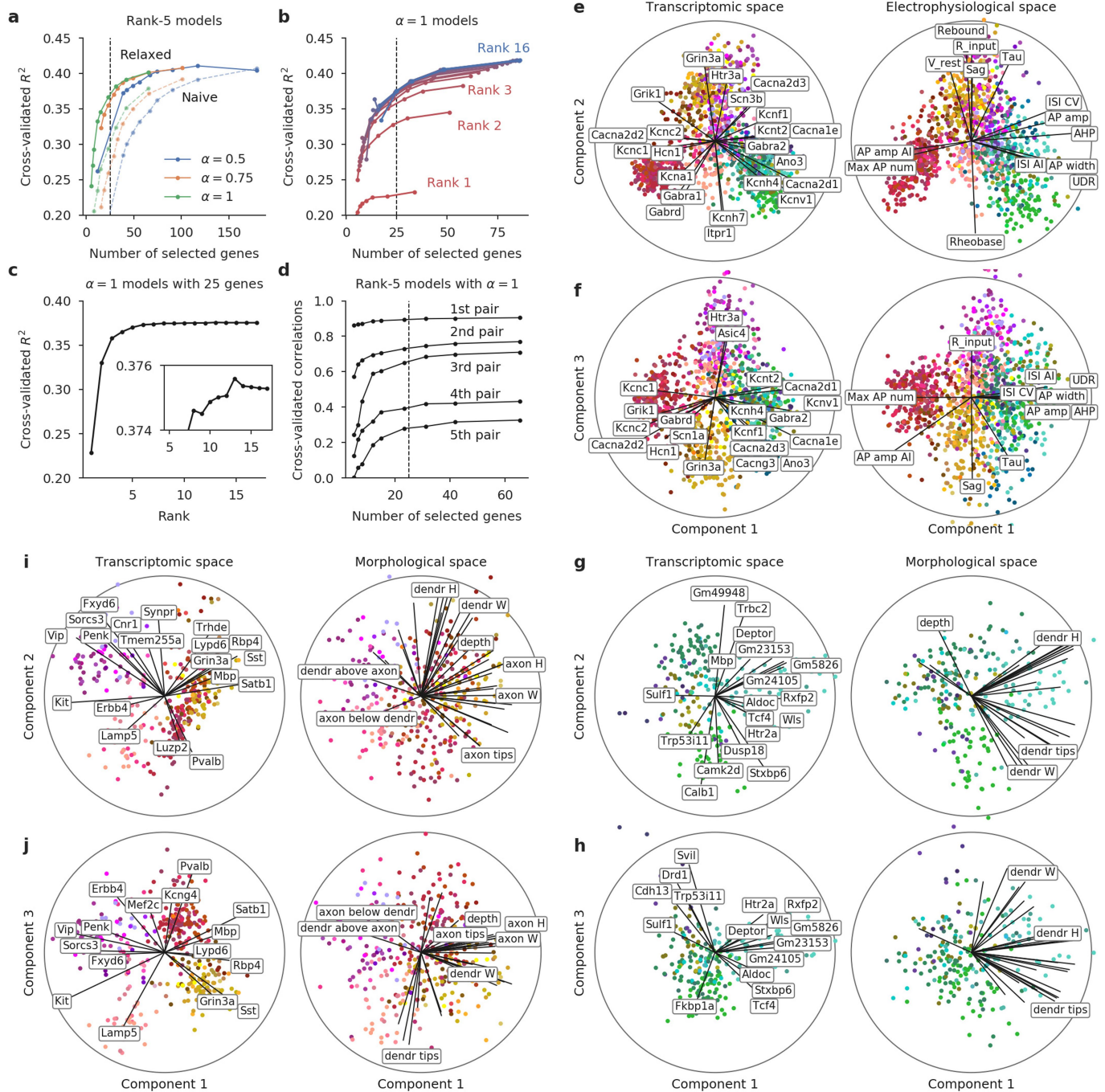
**Extended Data Fig. 3 | Diversity of mouse cortical neurons.** Two representative examples per t-type, or one if only one reconstruction was available. In total 135 neurons in 73 t-types. For interneurons, dendrites are shown in darker colours. For excitatory neurons, only dendrites are shown. Black dots mark soma locations. Horizontal grey lines show approximate layer boundaries. Three voltage traces are shown for each neuron: the

hyperpolarization trace obtained with the smallest current stimulation, the first depolarization trace eliciting at least one action potential, and the depolarization trace showing maximal firing rate. Stimulation length: 600 ms. The length of the shown voltage traces: 900 ms. Electrophysiological recording for one neuron did not pass quality control and is not shown.



**Extended Data Fig. 4 | Extraction and distribution of electrophysiological features.** Panels **a–f** show data from the same exemplary cell. **a**, Membrane potential responses to the consecutive step current injections. Hyperpolarizing currents were used to compute the input resistance (274.80 MOhm) and membrane time constant tau (21.95 ms). **b**, The first five traces showing spikes were used to compute ISI adaptation index (1.26), ISI average adaptation index (1.15), AP amplitude adaptation index (0.91) and AP amplitude average adaptation index (0.99). **c**, The first AP elicited in this neuron. It was used to compute AP threshold (–40.18 mV), AP amplitude (81.17 mV), AP width (0.80 ms), AHP (–12.60 mV),

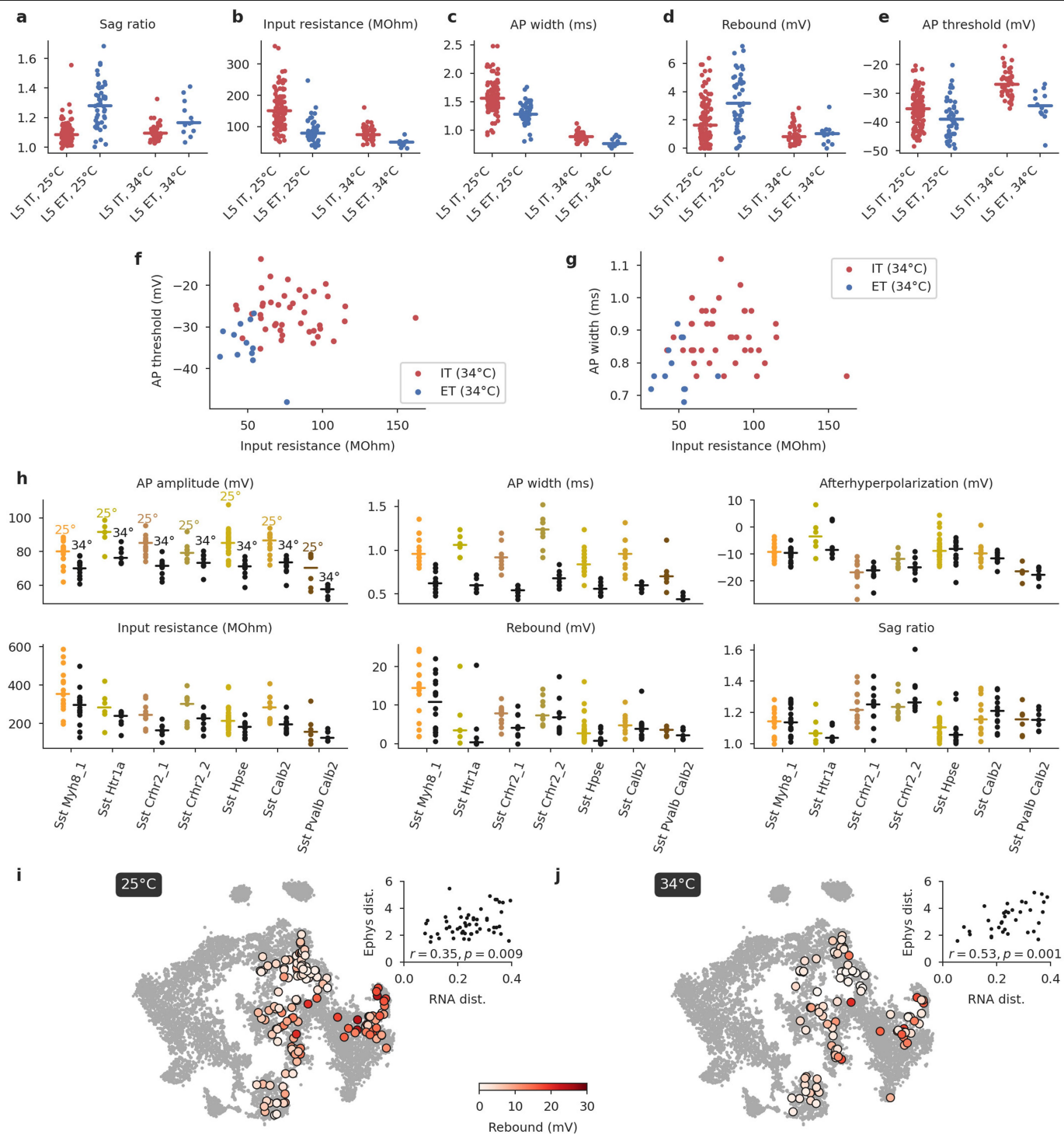
ADP (0 mV), UDR (1.62) and latency of the first spike (69.28 ms). **d**, Regression line gives the rheobase estimate (20.44 pA). **e**, The highest firing trace with 32 APs. This trace was used to estimate the ISI CV (0.27), ISI Fano factor (0.0014 ms), AP CV (0.17) and AP Fano factor (1.32 mV). **f**, The lowest hyperpolarizing trace was used to compute the sag ratio (1.17), sag time (0.26 ms), sag area (31.16 mV·ms) and rebound (17.84 mV). **g**, Eight important electrophysiological features are shown for all cells across all t-types. For t-types with at least three cells, horizontal lines show median values. See Supplementary File 2 for all electrophysiological features.



**Extended Data Fig. 5 | Additional reduced-rank regression analysis and cross-validation.** **a**, Cross-validated  $R^2$  of 'naive' and 'relaxed' sparse RRR solutions<sup>32</sup> for various elastic net penalties ( $\alpha$  and  $\lambda$ ). 'Relaxed' means that the model was re-fit without a lasso penalty using only the selected genes; 'naive' means that it was not re-fit. Vertical dashed lines at 25 genes corresponds to the choice made for Fig. 2. The best performance is around ~100 genes, but we chose 25 for the sake of interpretability. The subsequent panels only show results for the 'relaxed' models. **b**, Cross-validated  $R^2$  using  $\alpha = 1$  for different ranks from rank 1 to rank 16 (full rank). **c**, Cross-validated  $R^2$  using  $\alpha = 1$  and  $\lambda$  needed to obtain 25 genes for different ranks. The peak performance is achieved with rank -13 (inset), but rank-5 model used in the main text is almost as good. **d**, Cross-validated correlations between sequential projections of the transcriptomic and electrophysiological data sets (rank-5 models with  $\alpha = 1$ ).

For any given number of selected genes, correlations decrease monotonically for higher components. **e, f**, Reduced-rank regression model using only ion channel genes. A full analogue of Fig. 2 but using only 328 ion channel genes (see Methods), of which 307 were detected in our data set in at least 10 cells. **g-j**, Reduced-rank regression model predicting morphological features. An analogue of Fig. 2 but using morphological, instead of electrophysiological features. The analysis was done separately for the excitatory (**g-h**) and for the inhibitory (**i-j**) neurons because different sets of morphological features were computed for these sets of neurons. Excitatory neurons: 269 cells, 35 features. Rank-5 model,  $\lambda = 0.59$ , adjusted to yield 25 genes. Only a subset of morphological features are labelled to reduce the clutter (abbreviations: "W" – width, "H" – height). Inhibitory neurons: 367 cells, 50 features,  $\lambda = 0.49$ .

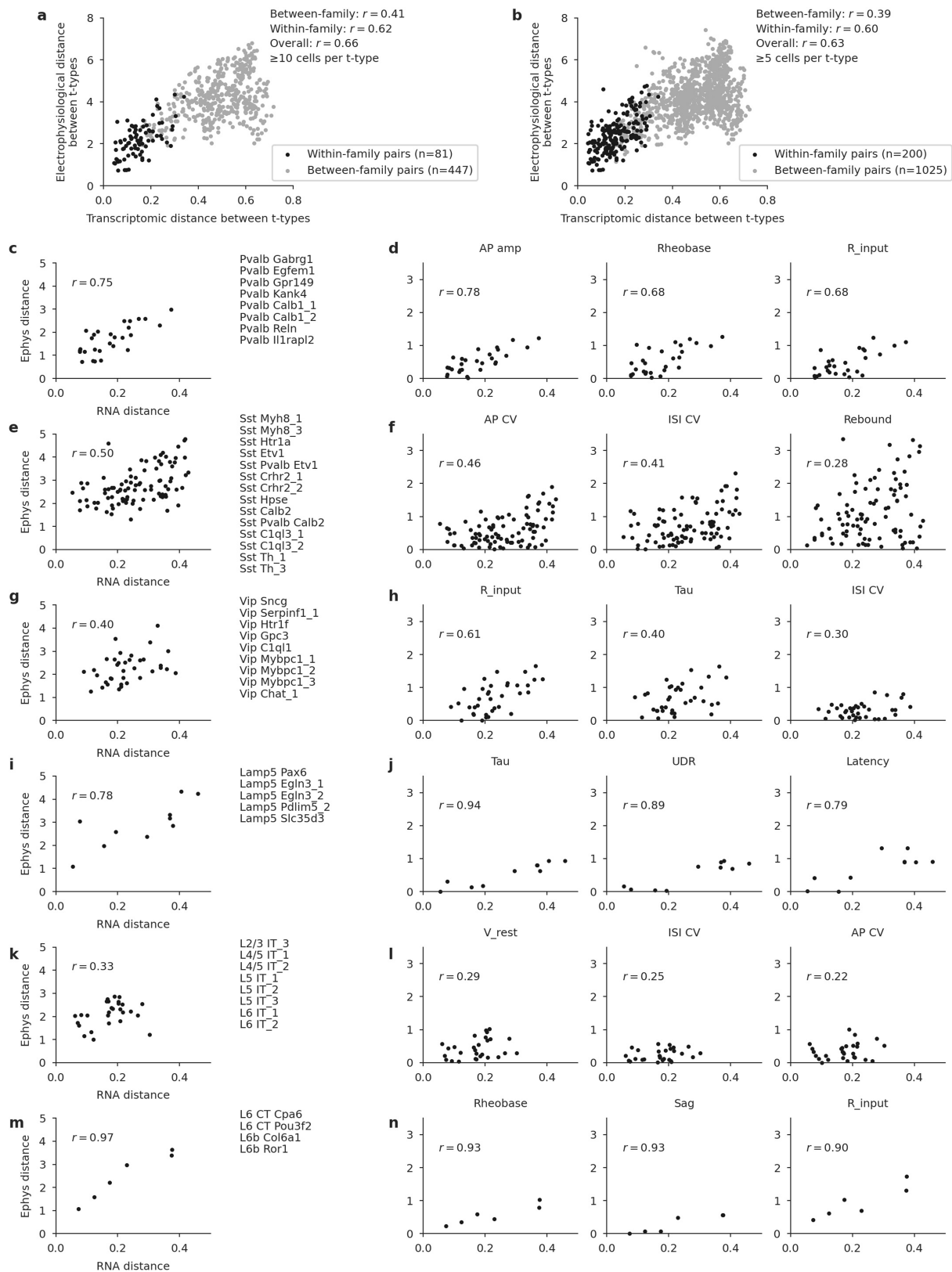




**Extended Data Fig. 6 | Electrophysiological properties of IT, ET, and Sst neurons in Layer 5 at physiological temperature.** **a–e**, Each panel shows a comparison between L5 neurons from the IT and the ET subclasses (pooled across all t-types within each subclass). The main set of experiments was done at room temperature (25°C). Follow-up experiments were done at physiological temperature (34°C), in the presence of 1mM kynurenic acid and 0.1mM picrotoxin in order to block fast glutamatergic and GABAergic synaptic transmission. Horizontal lines show median values. The first four panels correspond to features showing the largest IT/ET differences at room temperature, according to the two-sided Wilcoxon–Mann–Whitney test statistic (and omitting several features that are very correlated with the shown ones: upstroke-to-downstroke ratio, sag time, and sag area). The last panel additionally shows one feature that showed prominent difference at 34°C. **f, g**, IT and ET neurons recorded at 34°C in two-dimensional representations using the features with highest separability. **h**, The change of

electrophysiological properties between room temperature (25°C) and physiological temperature (34°C) for various t-types from the Sst subclass. Only L5 neurons are shown. Only t-types with  $\geq 5$  cells in both conditions are shown. Horizontal lines denote median values. AP amplitude and AP width changed the most between conditions, but the relative differences between t-types stayed roughly the same. The other four shown features did not change much, and the relative differences between t-types stayed the same. **i**, Overlay of the L5 Sst cells over the reference t-SNE embedding, coloured by rebound, as in Fig. 4b. The inset shows the correlation between transcriptomic distances and electrophysiological differences between all pairs of Sst t-types (only for t-types with at least 5 cells, and excluding *Sst Chodl*), together with its *p*-value. **j**, The same analysis as in (c) but using the experiments performed at physiological temperature. No corrections for multiple comparisons were applied.

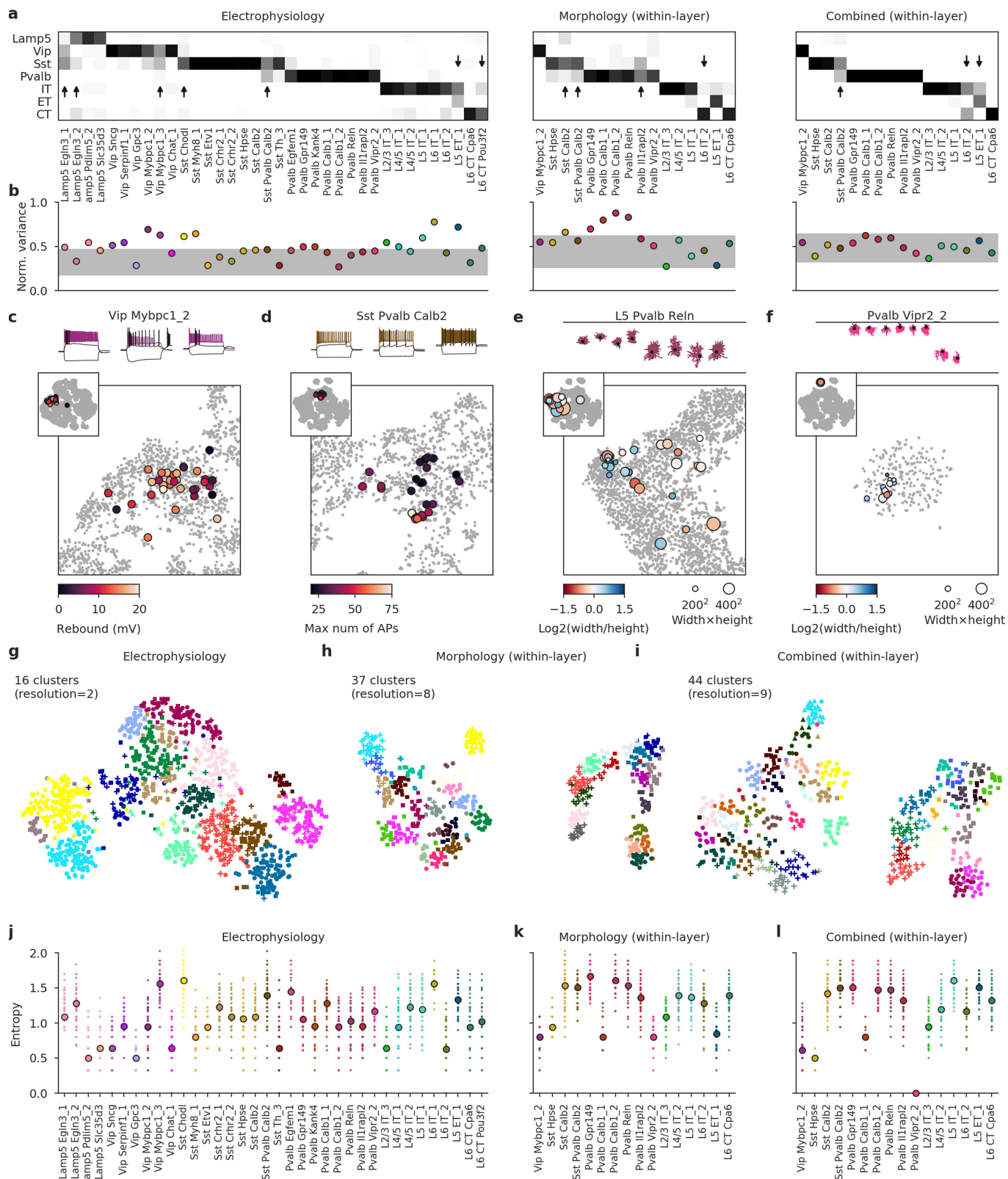




**Extended Data Fig. 7 | Transcriptomic and electrophysiological distances within individual families.** **a, b**, Pooled within-family analysis. The same analysis as in Fig. 4c but showing within-family as well as between-family pairs of t-types. Using a cutoff of at least 10 neurons per t-type (a) and a cutoff of at least 5 neurons per t-type (b). **c–n**, Transcriptomic and electrophysiological

distances within individual families. Only t-types with  $\geq 5$  neurons are used for this analysis (used t-types are listed in the second column). Transcriptomically well-isolated *Sst Chodl* and *Pvalb Vipr2\_2* were excluded. Three electrophysiological features with the highest correlation to the transcriptomic distance are shown on the right, for each family.

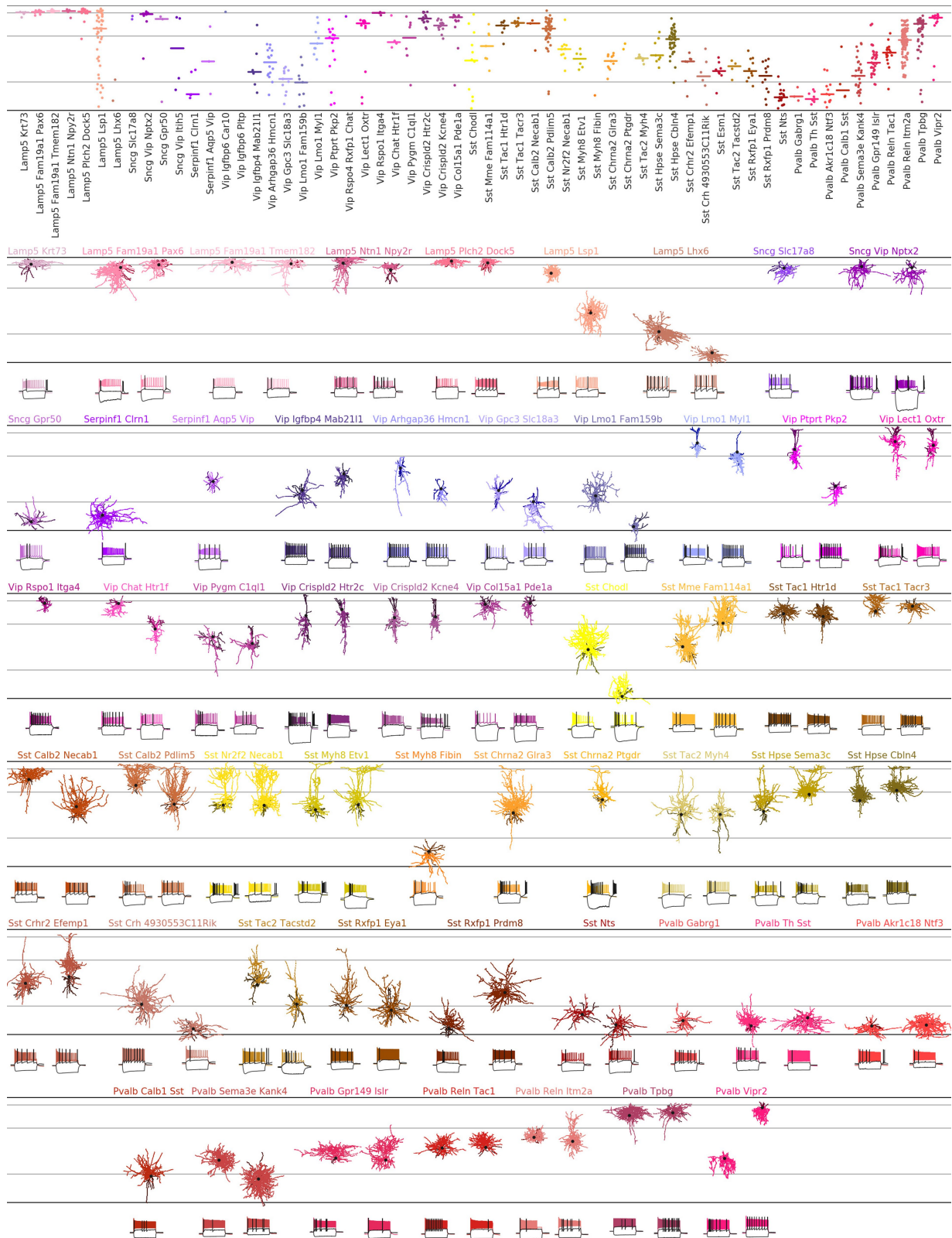
# Article



Extended Data Fig. 8 | See next page for caption.

**Extended Data Fig. 8 | Phenotypic variability of individual t-types.** The extended version of Fig. 5. **a**, Confusion matrices for classifying cells from each t-type into seven transcriptomic families, using electrophysiological, morphological, and combined features. Only t-types with at least 10 cells are shown. For morphological and combined features we only took cells from one cortical layer. Values in each column sum to 1. Arrows mark t-types that are classified into wrong families more often than 25% of the time. We used kNN-based classifier with  $k = 10$ . **b**, Normalized total variance of features in each t-type. Higher values correspond to t-types with more variable phenotypes. Horizontal grey band shows the min/max normalized variances of  $k$ -means clusters. **c**, Three exemplary traces of cells from the *Vip Mybp1\_2* type (all with confidence  $\geq 95\%$ ) and t-SNE overlay coloured by the rebound. Inset: the same t-SNE embedding as in Fig. 1. Main plot: zoom-in. **d**, Three exemplary traces of cells from the *Sst Pvalb Calb2* (confidence  $\geq 95\%$ ) and t-SNE overlay coloured by the maximum firing rate. **e**, Exemplary morphologies of L5 cells from the *Pvalb Reln* type and t-SNE overlay coloured by the axonal width/height log-ratio as in

Fig. 4e. **f**, Exemplary morphologies of *Pvalb Vipr2\_2* chandelier neurons and t-SNE overlay coloured by the axonal width/height log-ratio as in Fig. 4e. **g-i**, We used Leiden clustering<sup>35</sup> to cluster the cells based on electrophysiological, morphological, and combined features. The clustering resolution was adjusted to roughly match the number of e-types, m-types, and em-types from ref.<sup>24</sup>. The cluster colours in these panels are arbitrary and not the same as the colours used for t-types. **j-l**, For each t-type with at least 10 cells, we measured the entropy of the cluster assignments. Entropy zero corresponds to all cells getting into one cluster. Higher entropies mean that cells get distributed across many clusters. We repeated the clustering 100 times with different random seeds, and for each of them, subsampled each t-type to 10 cells to measure the entropy. Points show 100 repetitions, big markers show medians. When using morphological and combined features, all t-types were layer-restricted, as above. The t-type colours do not correspond to the colours in panels (j-l).



**Extended Data Fig. 9 | Interneurons assigned to the Tasic et al. <sup>4</sup> t-types.** This is an exact analogue of Fig. 1b and Extended Data Fig. 3 using inhibitory t-types from ref. <sup>4</sup>. It allows the direct comparison with the results from ref. <sup>45</sup>. We used

the same neurons as in Extended Data Fig. 3 whenever possible. 99 neurons in 55 t-types.



## Extended Data Table 1 | Description of the inhibitory t-types

All marker genes are given based on Ref. [20].

### Lamp5 subclass

<i>Lamp5 Pax6</i>	L1
<i>Lamp5 Egl3_1</i>	Alpha7 cells in L1 [56]: <i>Ndnf</i> <sup>-</sup> <i>Chrna7</i> <sup>+</sup> . Compared to <i>Egl3_2</i> , larger hyperpolarization sag, stronger bursts, and rebound firing.
<i>Lamp5 Egl3_2</i>	Canopy cells in L1 [56]: <i>Ndnf</i> <sup>+</sup> <i>Npy</i> <sup>-</sup> .
<i>Lamp5 Egl3_3</i>	-
<i>Lamp5 Pdlim5_1</i>	L1
<i>Lamp5 Pdlim5_2</i>	<i>Ndnf</i> <sup>+</sup> <i>Npy</i> <sup>+</sup> . Late-spiking neurogliaform cells (NGCs) in L1 with wide asymmetric action potentials (APs) and deep afterhyperpolarization (AHP).
<i>Lamp5 Slc35d3</i>	<i>Ndnf</i> <sup>-</sup> . NGCs in L2/3, L5, and L6, showing layer-adapting axonal morphology.
<i>Lamp5 Lhx6</i>	Deep L5/L6 neurogliaform-like cells with NGC morphology and deep AHP but narrow APs. Putatively MGE-derived [4], suggesting that although all deep NGCs belong to the <i>Lamp5</i> subclass, some are CGE- and some are MGE-derived (as is the case in hippocampus [25, 26, 27]).

### Sncg subclass

<i>Sncg Col14a1</i>	} The <i>Sncg</i> subclass (mostly <i>Vip</i> <sup>-</sup> and strongly <i>Cck</i> <sup>+</sup> ) proved difficult to sample due to the lack of specialized Cre lines. We found them in all layers from L1 to L6 (preferentially in the upper L2/3) with diverse morphologies; they mostly showed irregular firing, sometimes with a strong rebound. Several cells in the upper L2/3 had large axonal morphologies, likely corresponding to the 'large <i>Cck</i> basket cells' [2].
<i>Sncg Slc17a8</i>	
<i>Sncg Calb1_1</i>	
<i>Sncg Calb1_2</i>	
<i>Sncg Npy2r</i>	

### Vip subclass

<i>Vip Sncg</i>	Abundant in L2/3 [57], strong <i>Cck</i> <sup>+</sup> expression, local dendritic and axonal morphologies ⇒ 'small <i>Cck</i> basket cells' [2]. High input resistance [2].
<i>Vip Serpinf1_1</i>	} <i>Cck</i> <sup>+</sup> , deep L5 and L6, large axonal arborization.
<i>Vip Serpinf1_2</i>	
<i>Vip Serpinf1_3</i>	} L2/3, upper L5
<i>Vip Htr1f</i>	
<i>Vip Gpc3</i>	} L5, mostly local morphology [57], some cells with deep-projecting axons [5][57]. In <i>Vip Mybpc1_1</i> ( <i>Calb2</i> <sup>+</sup> ) and <i>Vip Chat_1</i> ( <i>Calb2</i> <sup>+</sup> <i>Chat</i> <sup>+</sup> ) also some bipolar cells in L2/3 and upper L5. <i>Vip</i> neurons are traditionally characterized by high input resistance [2], but some of these types, especially <i>Vip Gpc3</i> , showed only moderate input resistance, comparable to the <i>Sst</i> subclass. This type also had particularly low resting membrane potential.
<i>Vip C1ql1</i>	
<i>Vip Mybpc1_1</i>	
<i>Vip Chat_1</i>	
<i>Vip Mybpc1_2</i>	} Abundant in L2/3 [57]. Vertically oriented morphologies, sometimes with bipolar dendritic structure. Diverse firing patterns with some neurons exhibiting large membrane time constant, hyperpolarization sag and strong rebound firing. High input resistance [2].
<i>Vip Mybpc1_3</i>	
<i>Vip Chat_2</i>	
<i>Vip Igfbp6_1</i>	} - [We did not obtain any cells from these types, presumably due to their very weak <i>Vip</i> expression [4].]
<i>Vip Igfbp6_2</i>	

### Sst subclass

<i>Sst Chodl</i>	All layers from upper L2/3 down to the bottom of L6. This type is thought to have long-range projections [4, 24] and for two cells in L6 we could indeed see an axon disappearing into the white matter. Low rebound potential and low hyperpolarization sag. High variability in membrane time constant.
<i>Sst Penk</i>	L6. Mostly local axonal arborization within L6 [58].
<i>Sst Myh8_1</i>	L5. 'T-shaped' Martinotti morphologies [29, 30] and strong rebound firing.
<i>Sst Myh8_2</i>	L5. Strong rebound firing.
<i>Sst Myh8_3</i>	L5. Sometimes rebound firing. Strong hyperpolarization sag.
<i>Sst Htr1a</i>	L5. Similar to transcriptomically neighboring <i>Sst Hpse</i> .
<i>Sst Etv1</i>	L5. 'T-shaped' Martinotti morphologies [29, 30] and strong rebound firing.
<i>Sst Pvalb Etv1</i>	L5. 'T-shaped' Martinotti morphologies [29, 30] and strong rebound firing. Strong hyperpolarization sag.
<i>Sst Crhr2_1</i>	} Lower L5. Mostly local axonal arbor but with some sparse ascending axons.
<i>Sst Crhr2_2</i>	
<i>Sst Hpse</i>	
<i>Sst Calb2</i>	Abundant in L2/3, Martinotti morphology, adapting firing pattern [2][59]. 'Fanning-out' Martinotti morphology [29, 30] in upper L5. Non-zero ADP.
<i>Sst Pvalb Calb2</i>	L2/3. Lower AP width and higher firing rate than typical for the <i>Sst</i> subclass [28]. Some cells have Martinotti, some — basket cell morphology.
<i>Sst Pappa</i>	-
<i>Sst C1ql3_1</i>	} Lower L5. Non-Martinotti morphology without ascending axons [2, 24]. Deep AHP.
<i>Sst C1ql3_2</i>	
<i>Sst Tac2</i>	L5.
<i>Sst Th_1</i>	Lower L5 and L6.
<i>Sst Th_2</i>	-
<i>Sst Th_3</i>	L6. Mostly local axonal arborization within L6 [58].

### Pvalb subclass

<i>Pvalb Gabrg1</i>	} L6, fast spiking. Mostly local axons [58]. Some neurons exhibited a horizontally elongated or downward projecting axon mostly innervating L6b [24].
<i>Pvalb Egfem1</i>	
<i>Pvalb Gpr149</i>	Larger hyperpolarization sag and rebound potential compared to the other FS neurons.
<i>Pvalb Kank4</i>	Middle L5, fast spiking. Preferentially horizontally elongated basket cells [5].
<i>Pvalb Calb1_1</i>	Lower L5 and L6, fast spiking. A variety of axonal morphologies, including some with large local arborization with dense spherical shape.
<i>Pvalb Calb1_2</i>	L6, fast spiking. Mostly local axons [58]. Some neurons exhibited a horizontally elongated or downward projecting axon mostly innervating L6b [24].
<i>Pvalb ReIn</i>	Upper L5, fast spiking. A variety of axonal morphologies, including some with large local arborization with dense spherical shape.
<i>Pvalb Il1rapl2</i>	Upper L5, fast spiking. Preferentially small (or shrub) basket cells [5].
<i>Pvalb Vipr2_1</i>	Abundant in L2/3, fast spiking. Strongly layer-adapting morphology: classical L2/3 basket morphology in L2/3 [5], mostly large basket cell morphology in upper L5 [5]. Some neurons showed delayed firing [60] with large latency.
<i>Pvalb Vipr2_2</i>	-
	Chandelier cells, recognizable by the straight terminal axonal 'cartridges' [2]. Mostly upper L2/3, some in deep L5. Fast spiking, but lower firing rate compared to basket cells, practically absent hyperpolarization sag.

Note: we did not encounter double-bouquet basket cells, previously described in L2/3 of mouse V1 [5], or L6 FS cells with translaminal axons reaching up to L1, that were also reported in V1 [24][61].

References<sup>56-61</sup> are cited in this table.

# Article

## Extended Data Table 2 | Description of the excitatory t-types

Transcriptomically, cortical excitatory neurons are classified into the well-separated intertelencephalic (IT), extratelencephalic (ET, also called PT for pyramidal tract), corticothalamic (CT), and near-projecting (NP) subclasses [4]. Morphologically, they have been classified into big-tufted, small-tufted, untufted neurons, depending on the shape of the apical dendrite tuft, stellate neurons without an apical dendrite, and horizontal/inverted neurons in L6 [62, 63, 64, 65].

### IT subclass

<i>L2/3 IT_1</i>	–
<i>L2/3 IT_2</i>	–
<i>L2/3 IT_3</i>	Abundant in L2/3. Tufted pyramidal cells with high rheobase.
<i>L4/5 IT_1</i>	Located on the boundary between L2/3 and L5, likely corresponded to the quasi-L4 neurons described previously in motor cortex [31]. Diverse morphologies with some pyramidal and some stellate cells.
<i>L4/5 IT_2</i>	Upper L5. Thin untufted apical dendrite.
<i>L5 IT_1</i>	} L5. Large tufted pyramidal neurons. Most belonged to <i>L5 IT_1</i> .
<i>L5 IT_2</i>	
<i>L5 IT_3</i>	
<i>L5 IT_4</i>	Located at the boundary between L5 and L6. Short and untufted pyramidal cells.
<i>L6 IT_1</i>	Upper L6. Short and untufted pyramidal cells. Broad APs.
<i>L6 IT_2</i>	L6. Often stellate or inverted [66].
<i>L6 IT Car3</i>	– [Using Gnb4-Cre mouse line, we did not find any labeled cells in MOp, but only near the claustrum [46]. Those all mapped to this type.]

### ET subclass

<i>L5 ET_1</i>	} Large big-tufted cells with the apical dendrite often bifurcating close to the soma, suggesting that these were corticospinal cells [67, 68]. They had bigger hyperpolarization sag, lower input resistance, and smaller AP width, compared to the L5 IT neurons (we confirmed these differences in follow-up experiments at 34 °C, Extended Data Fig. E6). We did not observe consistent morpho-electric differences between the four ET types, apart from <i>L5 ET_1</i> being located deeper than the rest (Fig. 1b; consistent with Ref. [34]), but they may have different projection targets [4][69].
<i>L5 ET_2</i>	
<i>L5 ET_3</i>	
<i>L5 ET_4</i>	

### NP subclass

<i>L5/6 NP_1</i>	} NP neurons proved very difficult to obtain without a specialized Cre driver line. The few neurons in our dataset were all untufted, with sparse basal dendrites, in agreement with prior literature [24].
<i>L5/6 NP_2</i>	
<i>L5/6 NP_3</i>	
<i>L5/6 NP CT</i>	

### CT subclass

<i>L6 CT Gpr139</i>	} CT types were located in L6 and lower L5 and had mostly untufted apical dendrites [66]. In line with previous literature [70], they could be distinguished from L6 IT neurons by a lower inter-spike interval adaptation index (Supplementary File 2). The most abundant type was <i>L6 CT Cpa6</i> .
<i>L6 CT Cpa6</i>	
<i>L6 CT Grp</i>	
<i>L6 CT Pou3f2</i>	
<i>L6 CT Kit_1</i>	–
<i>L6 CT Kit_2</i>	–
<i>L6b Col6a1</i>	} L6b types, transcriptomically related to the CT subclass (Fig. 1e), were all stellate, inverted, or horizontal, located preferentially in the bottom of L6. The <i>L6b Ror1</i> type stood out, having horizontal dendritic morphology and showing strong rebound firing.
<i>L6b Shisa6_1</i>	
<i>L6b Shisa6_2</i>	
<i>L6b Ror1</i>	
<i>L6b Kcnip1</i>	

References<sup>62-70</sup> are cited in this table.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

**Data collection** Electrophysiological recordings: HEKA patch master v2x65 and v2x90.3; morphological reconstructions: neuroleucida and neuroleucida explorer 11.04. In a subset of data the soma position was evaluated in slices during recording using Linlab2 1.0.

**Data analysis** The sequencing data were processed using the zUMIs 2.5.6b pipeline with default settings; sequencing reads were aligned to the mouse reference genome using STAR version 2.5.4b; reconstructed morphologies were converted into the SWC format using NLMorphologyConverter 0.9.0 and further analyzed using MorphoPy 0.6; we used custom Python scripts to perform the data analysis. The analysis code can be found at <https://github.com/berenslab/mini-atlas>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Electrophysiological recordings are available at <https://dandiarchive.org/dandiset/000008> in NWB for- mat.

Sequencing data are available at <http://data.nemoarchive.org/biccn/grant/zeng/tolia> in FASTQ format.

Morphological reconstructions are available at <ftp://download.brainimagelibrary.org:8811/3a/88/3a88a7687ab66069> in SWC format.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sampling strategy was determined using pre-existing knowledge of the transcriptional diversity of the mouse cortex (Tasic et al., 2018; Yao et al., 2020) and based also on the variability of morphological and electrophysiological types predicted by existing literature (Jiang et al., 2015, Gouwens et al. 2019, Scala et al., 2019).
Data exclusions	Cells meeting any of the exclusion criteria described in the following were declared low quality and did not get a t-type assignment: cells with the highest correlation below 0.4 (78 cells); cells that would be assigned to non-neural t-types, presumably due to RNA contamination (14 cells); cells with the highest correlation less than 0.02 above the maximal correlation in one of the other two transcriptomic orders (5 cells). Four cells were assigned to an excitatory t-type, despite having clearly inhibitory firing, morphology, and/or soma depth location (such as L1). The most likely cause was RNA contamination from excitatory cells that are much more abundant in the mouse cortex. These four cells were excluded from all analyses and visualizations (as if they did not pass the transcriptomic quality control). In addition, one cell was likely located outside of MOp, based on the slice anatomy, and was excluded as well. For the electrophysiology, the cells were not recorded or included when seal resistance values were <1 G $\Omega$ before achieving whole-cell configuration and/or initial access resistance was >30 M $\Omega$ . Cells were excluded from morphological analysis when the staining quality did not match pre-established criteria for inclusion. Cells that showed low staining quality such as poor fill, excessive background, dendritic or axonal truncation were not reconstructed and not included in the dataset.
Replication	The results of this study were not directly replicated. However, all the results were collected from multiple animals from multiple litters per wild-type and transgenic lines.
Randomization	There was no randomization performed as the study does not involve multiple study groups.
Blinding	There was no blinding performed as the study does not involve multiple study groups. However, for these study most of the neuronal electrophysiological proprieties and morphological reconstructions were obtained without having informations about the molecular finger print of the neuron.

## Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	<i>Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).</i>
Research sample	<i>State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.</i>
Sampling strategy	<i>Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.</i>
Data collection	<i>Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.</i>
Timing	<i>Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.</i>
Data exclusions	<i>If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.</i>
Non-participation	<i>State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.</i>



## Randomization

If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.

## Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

## Study description

Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.

## Research sample

Describe the research sample (e.g. a group of tagged *Passer domesticus*, all *Stenocereus thurberi* within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.

## Sampling strategy

Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.

## Data collection

Describe the data collection procedure, including who recorded the data and how.

## Timing and spatial scale

Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken

## Data exclusions

If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

## Reproducibility

Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.

## Randomization

Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.

## Blinding

Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Did the study involve field work?  Yes  No

## Field work, collection and transport

## Field conditions

Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).

## Location

State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).

## Access &amp; import/export

Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).

## Disturbance

Describe any disturbance caused by the study and how it was minimized.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials &amp; experimental systems

## Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used *Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*

Validation *Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

## Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s) *State the source of each cell line used.*

Authentication *Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*

Mycoplasma contamination *Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*

Commonly misidentified lines (See [ICLAC](#) register) *Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

## Palaeontology and Archaeology

Specimen provenance *Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).*

Specimen deposition *Indicate where the specimens have been deposited to permit free access by other researchers.*

Dating methods *If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight *Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals *Male and Female mice (median age 75 days, interquartile range 64-100, full range 35-245 days) were used in this study. Specific information about every single animal can be found in <https://github.com/berenslab/mini-atlas>. In particular, we used C57Bl/6 Wild type, Viat-Cre/Ai9 mice, SOM-Cre/Ai9, VIPCre/Ai9, PV-Cre/Ai9, NPY-Cre/Ai9, Scl17a8-Cre/Ai9, Scl17a8-iCre/Ai9, Vipr2-Cre/Ai9 and Gnb4-Cre/Ai9. Detailed information about the origin of each single Cre line reported here can be find in the main text.*

Wild animals *This study did not involve wild animals.*

Field-collected samples *The study did not involve samples collected from the field.*

Ethics oversight *Procedures for mouse maintenance and mouse surgeries were performed according to protocols approved by the Institutional Animal Care and Use Committee (IACUC) of Baylor College of Medicine.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics *Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*

Recruitment *Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*

Ethics oversight *Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	<input type="text" value="Provide the trial registration number from ClinicalTrials.gov or an equivalent agency."/>
Study protocol	<input type="text" value="Note where the full trial protocol can be accessed OR if not available, explain why."/>
Data collection	<input type="text" value="Describe the settings and locales of data collection, noting the time periods of recruitment and data collection."/>
Outcomes	<input type="text" value="Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures."/>

## Dual use research of concern

Policy information about [dual use research of concern](#)

### Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
<input type="checkbox"/>	<input type="checkbox"/>	Public health
<input type="checkbox"/>	<input type="checkbox"/>	National security
<input type="checkbox"/>	<input type="checkbox"/>	Crops and/or livestock
<input type="checkbox"/>	<input type="checkbox"/>	Ecosystems
<input type="checkbox"/>	<input type="checkbox"/>	Any other significant area

### Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
<input type="checkbox"/>	<input type="checkbox"/>	Demonstrate how to render a vaccine ineffective
<input type="checkbox"/>	<input type="checkbox"/>	Confer resistance to therapeutically useful antibiotics or antiviral agents
<input type="checkbox"/>	<input type="checkbox"/>	Enhance the virulence of a pathogen or render a nonpathogen virulent
<input type="checkbox"/>	<input type="checkbox"/>	Increase transmissibility of a pathogen
<input type="checkbox"/>	<input type="checkbox"/>	Alter the host range of a pathogen
<input type="checkbox"/>	<input type="checkbox"/>	Enable evasion of diagnostic/detection modalities
<input type="checkbox"/>	<input type="checkbox"/>	Enable the weaponization of a biological agent or toxin
<input type="checkbox"/>	<input type="checkbox"/>	Any other potentially harmful combination of experiments and agents

## ChIP-seq

### Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links <i>May remain private before publication.</i>	<input type="text" value="For 'Initial submission' or 'Revised version' documents, provide reviewer access links. For your 'Final submission' document, provide a link to the deposited data."/>
Files in database submission	<input type="text" value="Provide a list of all files available in the database submission."/>
Genome browser session (e.g. <a href="#">UCSC</a> )	<input type="text" value="Provide a link to an anonymized genome browser session for 'Initial submission' and 'Revised version' documents only, to enable peer review. Write 'no longer applicable' for 'Final submission' documents."/>

### Methodology

Replicates	<input type="text" value="Describe the experimental replicates, specifying number, type and replicate agreement."/>
Sequencing depth	<input type="text" value="Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and"/>

Sequencing depth	<i>whether they were paired- or single-end.</i>
Antibodies	<i>Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.</i>
Peak calling parameters	<i>Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.</i>
Data quality	<i>Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.</i>
Software	<i>Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.</i>

## Flow Cytometry

### Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

Sample preparation	<i>Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.</i>
Instrument	<i>Identify the instrument used for data collection, specifying make and model number.</i>
Software	<i>Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.</i>
Cell population abundance	<i>Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.</i>
Gating strategy	<i>Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.</i>
<input type="checkbox"/> Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.	

## Magnetic resonance imaging

### Experimental design

Design type	<i>Indicate task or resting state; event-related or block design.</i>
Design specifications	<i>Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.</i>
Behavioral performance measures	<i>State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).</i>

### Acquisition

Imaging type(s)	<i>Specify: functional, structural, diffusion, perfusion.</i>
Field strength	<i>Specify in Tesla</i>
Sequence & imaging parameters	<i>Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.</i>
Area of acquisition	<i>State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.</i>
Diffusion MRI	<input type="checkbox"/> Used <input type="checkbox"/> Not used

## Preprocessing

Preprocessing software	<i>Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).</i>
Normalization	<i>If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.</i>
Normalization template	<i>Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.</i>
Noise and artifact removal	<i>Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).</i>
Volume censoring	<i>Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.</i>

## Statistical modeling & inference

Model type and settings	<i>Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).</i>
Effect(s) tested	<i>Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.</i>
Specify type of analysis:	<input type="checkbox"/> Whole brain <input type="checkbox"/> ROI-based <input type="checkbox"/> Both
Statistic type for inference (See <a href="#">Eklund et al. 2016</a> )	<i>Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.</i>
Correction	<i>Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).</i>

## Models & analysis

n/a	Involvement in the study
<input type="checkbox"/>	<input type="checkbox"/> Functional and/or effective connectivity
<input type="checkbox"/>	<input type="checkbox"/> Graph analysis
<input type="checkbox"/>	<input type="checkbox"/> Multivariate modeling or predictive analysis
Functional and/or effective connectivity	<i>Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).</i>
Graph analysis	<i>Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).</i>
Multivariate modeling and predictive analysis	<i>Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.</i>