Dissecting origins of wiring specificity in dense cortical connectomes

Philipp Harth[1], Daniel Udvary[2], Jan Boelts[3], Daniel Baum[1], Jakob H. Macke[3], Hans-Christian Hege[1], Marcel Oberlaender[2,4]

[1]Department of Visual and Data-centric Computing, Zuse Institute Berlin, Germany; [2]In Silico Brain Sciences, Max Planck Institute for Neurobiology of Behavior – caesar, Bonn; [3]Machine Learning in Science, University of Tübingen and Tübingen AI Center, Germany; [4]Department of Integrative Neurophysiology, Center for Neurogenomics and Cognitive Research, VU Amsterdam, The Netherlands

Wiring specificity in the cortex is observed across scales from the subcellular to the network level. It describes the deviations of connectivity patterns from those expected in randomly connected networks. Understanding the origins of wiring specificity in neural networks remains difficult as a variety of generative mechanisms could have contributed to the observed connectome. To take a step forward, we propose a generative modeling framework that operates directly on dense connectome data as provided by saturated reconstructions of neural tissue. The computational framework allows testing different assumptions of synaptic specificity while accounting for anatomical constraints posed by neuron morphology, which is a known confounding source of wiring specificity. We evaluated the framework on dense reconstructions of the mouse visual and the human temporal cortex. Our template model incorporates assumptions of synaptic specificity based on cell type, single-cell identity, and subcellular compartment. Combinations of these assumptions were sufficient to model various connectivity patterns that are indicative of wiring specificity. Moreover, the identified synaptic specificity parameters showed interesting similarities between both datasets, motivating further analysis of wiring specificity across species.

Introduction

Cortical connectomes are among the most complex neural networks in nature, exemplified by the human cortex, which contains billions of neurons interconnected by trillions of synapses (Sporns and Tononi, 2005). Due to the enormous number of neurons and synapses in cortical brain networks, the synaptic wiring diagram cannot be explicitly passed on through the genome, which has been described as the genomic bottleneck (Zador, 2019). Instead, the genome encodes a developmental growth process that brings the presynaptic and postsynaptic structures of neurons into spatial proximity (Agi et al., 2020), followed by local synapse formation and refinement throughout the organism's lifetime (Südhof, 2018). While the structures of any two neurons must be in spatial proximity to establish synaptic connectivity, this process frequently results in highly targeted connections, a phenomenon known as synaptic specificity (Sanes & Yamagata, 2009). Examples of intricate biological mechanisms implementing synaptic specificity include cell surface recognition molecules and molecular cues expressed by guidepost cells (Sanes & Zipursky, 2020; Valdes-Aleman et al., 2021; Yogev & Shen, 2014). In combination with neurite-growth-dependent positional strategies that shape the overall morphology of neurons (Agi et al., 2020; Kaiser, 2017), local synaptic specificity gives rise to non-random connectivity patterns between neurons, which is what we collectively refer to as wiring specificity.

Evidence for wiring specificity can be observed at different levels in brain networks. At the network-topological level, it has been found that certain connectivity motifs between three or more neurons

occur more often than in equivalent random networks (Song et al., 2005). On the cellular level, it has been observed that connection probabilities between subtypes of pyramidal neurons deviate from those predicted by a null model based on the overlap of pre- and postsynaptic structures, indicating synaptic specificity based on cell type (Brown & Hestrin, 2009). At the subcellular level, neurons can form multisynaptic connections, and such synapse clusters do occur more often than expected under a null model that assumes locally random synaptic connectivity based on the number of available contact sites (Kasthuri et al., 2015). Synapses can also be formed at specific subcellular compartments of the postsynaptic neuron, for example, it has been found that interneurons selectively target the dendrites, soma, or axon initial segments of pyramidal neurons depending on the subtype of the interneuron (Favuzzi et al., 2019). To discover such instances of wiring specificity in neural networks and identify the underlying origins, a combination of experimentally acquired connectome data at synaptic resolution and computational modeling is required.

Fortunately, large-scale saturated reconstructions of neural tissue using electron microscopy (EM) are increasingly available (MICrONS Consortium et al., 2021; Shapson-Coe et al., 2024), which can be attributed to technological advances in both image acquisition (Kornfeld & Denk, 2018) and automated segmentation (Macrina et al., 2021). Several studies have confirmed the presence of wiring specificity in densely reconstructed connectomes and have computationally modeled assumptions about the potential origins of such patterns (Karimi et al., 2020; Motta et al., 2019; Schneider-Mizell et al., 2021). It has been shown, for example, that only a certain fraction of strong synaptic inputs could be explained by Hebbian learning (Motta et al., 2019) and that certain morphological features of pyramidal neurons are correlated with inhibitory input onto their axon initial segments (Schneider-Mizell et al., 2021). However, a constant challenge in such data-driven analyses of dense connectomes is the systematic selection and evaluation of the modeled wiring principles that could have generated the observed connectivity patterns. It also remains difficult to disentangle the influence of morphological features, such as spatial proximity of pre- and postsynaptic contact sites (Udvary et al., 2022), from wiring preferences that are present at the synaptic level.

To address these challenges, we present a generative modeling approach (Betzel & Bassett, 2017) that allows us to statistically infer which wiring principles are necessary, sufficient, and best-suited to explain wiring specificity at different network levels. Based on the dense anatomical data, our method generates an anatomically constrained ensemble of connectomes which contains all connectomes that are anatomically plausible (based on the presence of pre- and postsynaptic contact sites in a certain volume) without making additional assumptions about local synaptic specificity. Thus, investigating the wiring patterns predicted by the anatomically constrained ensemble of connectomes without any further parameters and assumptions allows to investigate the impact of morphology as a baseline. In essence, this anatomically constrained ensemble of connectomes represents a null model (Váša Františekand Mišić, 2022) that allows testing which connections can and cannot be accounted for by neuron morphology alone. Hence, the null model allows identifying those neurons and their connections that require additional assumptions about synaptic specificity to be explained.

The framework provides a generic interface for defining biologically interpretable assumptions about synaptic specificity, for example, wiring preferences based on cell type. Using Simulation-based inference (Boelts et al., 2023) the method then identifies those synaptic preference parameters that could best reproduce the empirically observed wiring patterns, which are not explained by neuron morphology (i.e., the null hypothesis). Synaptic specificity parameters can be determined for individual neurons at subcellular resolution. On the synaptic level, the method therefore predicts which types of biological mechanisms could be principally involved between which cells to explain the

observed connectivity. As a template for illustrating the framework, we defined three categories of synaptic specificity assumptions based on (1) cell type, (2) single cell identity, and (3) subcellular compartment of the postsynaptic cell (i.e., soma, dendrite, axon initial segment).

We apply the framework to dense reconstructions of the mouse visual cortex (Turner et al., 2022) and the human temporal cortex (Shapson-Coe et al., 2024). We find that morphological overlap alone is always insufficient to predict synaptic connectivity as empirically observed at nanoscale resolution. Our analyses also confirm the presence of wiring specificity at multiple levels, as reflected in the number of multisynaptic connections between neurons and the occurrence frequencies of triplet motifs. Introducing assumptions of local synaptic specificity into the model allows us to better approximate these notable non-random connectivity patterns. Surprisingly, our approach yields qualitatively similar synaptic specificity parameters in the mouse visual and human temporal cortex, setting the stage for further cross-species investigations into the origins of wiring specificity in cortical connectomes.

Results

Framework for dissecting origins of wiring specificity in dense cortical connectomes

The proposed framework for dissecting origins of wiring specificity operates on dense reconstructions of neural tissue. The reconstructed tissue volume is partitioned into local overlap volumes (**Fig. 1A**). All synapses in the dense reconstruction are split into their pre- and postsynaptic contact sites and are assigned to overlap volumes. To define a null model of connectivity, we assume that two neurons sharing at least one pre- and postsynaptic contact site in an overlap volume can in principle be connected. Specifically, we assume that connections between all pre- and postsynaptic contact sites in an overlap volume are equally likely (see methods). Thus, the null model provides a baseline prediction of connectivity in the absence of local synaptic specificity and solely informed by overlap of pre- and postsynaptic contact sites. To demonstrate the utility of the framework for studying wiring specificity, we analyze three connectivity patterns at different levels (**Fig. 1B**): (1) the number of synaptic connections between neurons grouped by cell type and/or postsynaptic target site; (2) the occurrences of synaptic clusters (i.e., multisynaptic connections); and (3) the occurrence frequencies of triplet motifs. Conceptually, all deviations of these connectivity patterns from those expected in a random network are what we refer to as evidence of wiring specificity.

Observations of connectivity patterns that are sufficiently predicted by the null model do not require additional assumptions about synaptic specificity, indicating that morphological overlap alone can account for these patterns. For connectivity patterns not fully explained by morphology, we define different levels of assumptions to test the influence of local synaptic specificity (**Fig. 1C**). For instance, at the population level, we investigate whether preferential connections between excitatory and inhibitory neurons can account for unexplained connectivity patterns. At the cellular level, we examine correlations in wiring between individual neurons, and at the subcellular level, we assess whether targeting of specific neuronal segments such as the soma, dendrites, or axon initial segments could be involved. However, our algorithmic framework is generic and provides the flexibility to incorporate user-defined specificity models by parameterizing positive and negative synaptic preferences that are conditional on any of the neuronal properties and annotations available in the dataset.

To summarize, our approach starts by generating an ensemble of anatomically plausible connectomes solely based on the mutual presence of the pre- and postsynaptic contact sites of neurons in overlap volumes that partition the reconstructed tissue volume. Therefore, by accounting

for anatomical constraints imposed by neurite morphology, our approach inherently rules out many wiring diagrams that might be deemed possible in other generative modeling approaches. Nevertheless, the method still captures a large space of structurally possible connectomes because all pre- and postsynaptic contact sites within an overlap volume can be connected (**Fig. 1D**). We tested the robustness of the method under different overlap volume sizes (**Fig. S1**). By assessing which empirically observed wiring patterns at different scales are sufficiently accounted for by anatomy and identifying those that are not, we determine the most likely configurations based on morphology alone and evaluate how likely the empirically observed connectome is without additional assumptions. Then, by systematically testing different synaptic specificity assumptions individually and in combination, we identify which assumptions are necessary, sufficient, and best-suited to account for the connectivity patterns that are not explained by morphology. On a technical level, our framework computes the specificity parameters at subcellular resolution (see methods), thereby inferring the best assumptions that could explain every empirically observed connection in the dataset (**Fig. 1E**). Thus, our methodology not only allows quantifying and visualizing synaptic specificity at nanoscale resolution but also facilitates comparisons across different datasets and specificity-based neuron classification schemes, offering a comprehensive workflow for analyzing and understanding the origins of wiring specificity (**Fig. 1F**).

Application of framework to a dense reconstruction of the mouse visual cortex

We applied our approach to a dense reconstruction from layer 2/3 of the mouse visual cortex (Turner et al., 2022). The reconstruction covers a volume of 250 x 140 x 90 mm^3 in V1 and contains 417 excitatory (Pyramidal) neurons and 34 interneurons with cell bodies inside the reconstructed volume; the interneurons are assigned to the following subclasses: Bipolar (12), Basket (4), Chandelier (2), Martinotti (1), Neurogliaform (1), Unclassified (14) (Turner et al., 2022). After correcting for false-positive synapses and integrating additional synapse annotations (see methods), the final dataset includes 3.2 million synapses, most of which are between unknown neurites that cannot be traced back to a cell body inside the volume. In our analyses, we systematically account for the presence of these background synapses in addition to the selected synapses that involve known and annotated neurons as pre- or postsynaptic partners (see methods). To apply our method, we spatially partitioned the reconstructed volume into 32,441 local overlap volumes using a uniform 5 µm grid (**Fig. 2A**) but also tested the robustness of the analyses with different grid sizes (**Fig. S1**). Typically, the pre- and postsynaptic processes (i.e., axon and dendrites) of individual cells innervate many different overlap volumes (**Fig. 2B**). For two neurons to be connected, their pre- and postsynaptic processes must be jointly present in at least one of the overlap volumes (**Fig. 2C**). By accounting for these anatomical constraints, we obtain an ensemble of structurally possible connectomes from the dataset, which contains the empirically observed connectome as one of its instances (**Fig. 2D**). Naturally, only a small fraction of the structurally possible connections will be realized in the empirically observed connectome (**Fig. 2E**). Structurally possible and empirically observed connections between cells can be arbitrarily aggregated to obtain summary statistics or disaggregated to inspect, e.g., the subcellular location of synapses on specific compartments (**Fig. 2F**).

The first step in our analysis is the application of the null hypothesis to determine whether additional assumptions are necessary and, if so, for which connectivity features. Therefore, without any further assumptions, we applied the null model in the local overlap volumes and derived the most likely connectivity patterns across the ensemble of connectomes. We first grouped neurons by their cell type and computed the aggregate synapse counts for all cell type pairings (**Fig. 3A**). This simple pairwise statistic was surprisingly well captured by the null model (with the largest deviations observed for excitatory to inhibitory connections). However, when examining whether these synapses are

formed on the soma, dendrite, or axon initial segment (AIS), more significant deviations became apparent (**Fig. 3B**): while the anatomy could well account for the number of synapses on the dendrite, it overestimated excitatory synapses to the soma and AIS of other excitatory neurons and underestimated inhibitory synapses to the soma and AIS of excitatory neurons. This suggests that to account for these observations, additional assumptions of synaptic specificity are needed. Furthermore, we analyzed the occurrences of synapse clusters (**Fig. 3C**). While multisynaptic connections are predicted even by the null model, the occurrences of such synapse clusters are systematically underestimated, especially for clusters consisting of more than three synapses. Finally, we computed the frequencies of triplet motifs (**Fig. 3D**) and found that the null model provides good approximations for some motifs but fails to predict others. These findings confirm that synaptic specificity is involved in creating some of the more subtle network features, such as targeted synapsing onto subcellular compartments like the soma or the axon initial segment (Schneider-Mizell et al., 2021), the presence of multisynaptic connections (Fares & Stepanyants, 2009; Kasthuri et al., 2015), and certain network motifs (Song et al., 2005).

Now that we have identified wiring patterns which cannot be explained by anatomy alone, what synaptic specificity assumptions are necessary to explain them? We first tested assumptions at the population level, namely whether positive or negative wiring preferences between excitatory and inhibitory cell types exist. When we included synaptic specificity based on cell type into the model, we indeed found that the pairwise connectivity between cell type populations is now perfectly explained (**Fig. 4A**). Cell type specificity also adjusts the predicted connectivity on the dendrites (when synapse counts are disaggregated by postsynaptic target site) but does not explain targeted synapsing on the soma and axon initial segment (**Fig. 4B**). This is resolved by additionally introducing synaptic specificity based on the subcellular compartment into the model. Including these specificity sources improves the prediction of synapse clusters to some degree, however, specificity based on single cell identity is needed for accurate prediction of such multisynaptic connections (**Fig. 4C**). Interestingly, a good approximation for the occurrences of triplet motifs is already achieved through cell type specificity (**Fig. 4D**). We conclude that not all empirically observed wiring specificity, such as synapse counts on dendrites or certain network motifs require the full set of synaptic specificity assumptions. This raises the question of how to select and systematically evaluate different assumptions that model synaptic specificity.

Quantifying model accuracy at cellular and subcellular resolution

To quantify the accuracy of synaptic specificity models more analytically than through the deviations of aggregate connectivity statistics that can be chosen at will, we defined a metric (see methods) – in the following referred to as loss function. The loss function can be visualized and quantified at the same resolution level as the connectome (**Fig. 5A**). Essentially, for every synaptic connection and every possible synaptic connection we can identify which assumption leads to an improvement compared to not having this assumption. Thus, even for an individual neuron we can now measure which assumptions are necessary, sufficient, and best-suited to explain its synaptic connections (**Fig. 5B**). This also allows comparing the model loss (see methods) for connections of the same neuron to different postsynaptic partners, where negative outlier edges may indicate the presence of synaptic specificity mechanisms that are not yet well captured by the model (**Fig. 5C**). As became already apparent by looking at aggregate connectivity statistics, introducing synaptic specificity based on cell type and subcellular compartment alone can significantly improve the model predictions. This is also reflected by the reduction in loss (**Fig. 5D**). Interestingly, the reduction in loss is more pronounced when the presynaptic neuron is inhibitory, which may indicate a higher synaptic specificity of such cells.

So far, we have shown that the tested synaptic specificity models can explain aggregate connectivity statistics and identified the necessary assumptions for the selected statistics. To test the generalizability of our analyses, we introduced a final control stage into the analysis workflow, where we use the specificity parameters identified by the framework to generate discrete realizations of the connectome that can then be compared to the empirically observed connectome. Specifically, we generated 1000 connectome realizations for all model configurations and measure how often the generated connectomes match observed or unobserved connections (**Fig. 6A**). If the model capacity was perfect, then from 1000 realizations, we would see 1000 times exactly the empirically observed connectome. If the model capacity was completely inadequate, we would never see the empirically observed connectome in the realizations. Therefore, our metric captures how many times out of the 1000 realizations is any connection that is observed empirically (and any connection that is not observed empirically) generated by the model, when we apply the specificity parameters to the dense anatomy. As with our loss function, we can visualize this metric at the resolution of individual neurons and their connections (**Fig. 6B**). As one would expect, individual connections that have a poor model loss do also far worse in this metric (**Fig. 6C**), indicating that more assumptions may be required to explain this specific connection. When we compared the metric for different models, we found that the application of no parameters (i.e., the null model) generates networks that share some of the aggregate statistics but virtually never generate the empirically observed connectome in detail. Thus, we conclude that the anatomy is in general insufficient to reproduce the connectome, but it will generate networks which share some of its statistics. When we increased the model complexity step by step, we could observe how the fraction of accurately predicted connections improves (**Fig. 6D**). If we apply the most complex model in which all synaptic specificity assumptions are combined, 80% of all empirically observed connections are found in the model-generated realizations at least 70.6% of the time.

Comparative analysis in the human temporal cortex

As we have seen, the specificity parameters identified in the mouse visual cortex (Turner et al., 2022) can generate connectomes that match the empirically observed connectome. But what is the significance of the identified synaptic specificity parameters in comparison with other dense reconstructions of cortical tissue? To test this, we repeated the analyses in a dense reconstruction of the human temporal cortex (Shapson-Coe et al., 2024). The human cortex dataset is significantly larger and covers complete dendritic trees across the cortical layers in a volume of 1 mm^3 (**Fig. 7A**), containing 16,087 neurons with 133.7 million synapses (Shapson-Coe et al., 2024). For our analysis, we selected 200 strongly connected excitatory and inhibitory neurons with cell bodies in layer 3 and divided the tissue volume innervated by the axons and dendrites of these cells into 9184 local overlap volumes using a 25 μm grid (see methods). We then computed the same aggregate connectivity statistics as in the mouse visual cortex for the null model and the synaptic specificity models (**Fig. S5**). We observed surprising similarities to the mouse visual cortex: Aggregate synapse counts grouped by cell type are well approximated by the null model except for excitatory to inhibitory synapses (**Fig. 7B**). When synapse counts are disaggregated by postsynaptic compartment, inhibitory synapses onto the soma and the axon initial segment of excitatory neurons are underestimated by the cell type specificity model (**Fig. 7C**). The null model underestimates the presence of multisynaptic connections (**Fig. 7D**). These similarities are reflected in the synaptic specificity parameters identified by the framework, which show similar patterns between the mouse visual and human temporal cortex (**Fig. 7E**). A caveat to biological interpretations of this finding is that due to the limited tissue volume of the mouse visual cortex and the proofreading state of the human

temporal cortex dataset, most synapses in both datasets involve unknown neurites that cannot be traced back to a cell body.

Classification of neurons based on their synaptic specificity

Can the synaptic specificity parameters determined by the framework be used to group neurons into different classes? We tested this question using the mouse visual cortex dataset (Turner et al., 2022) given its better proofreading state and extensive cell type annotations. First, we performed a PCA embedding of neurons (of known cell type) solely based on the synaptic specificity of their axons (see methods). Excitatory neurons form a cluster while most inhibitory neurons are separated from the excitatory cluster, which underscores differences in the synaptic specificity of their axons (**Fig. 8A**). Next, we performed a hierarchical clustering of interneurons with at least three outgoing synapses. Remarkably, the clusters based on synaptic specificity correlate with interneuron subtype annotations that were provided with the dataset (**Fig. 8B**). Interneurons with unknown subtypes are assigned to different clusters of annotated neurons, which suggests similarity in synaptic wiring preference and may aid their further classification. Hence, we conclude that the synaptic specificity parameters determined with our framework may add additional quantitative information for classifying the cell type of axons innervating the reconstructed volume, in addition to the morphological features and observed synaptic connections of these neurons.

Discussion

We have presented a generative modeling approach that can be directly applied to large-scale dense reconstructions of neural tissue for dissecting the origins of wiring specificity. Our method allows us to systematically differentiate origins of wiring specificity rooted in morphology (Udvary et al., 2022) from origins that are based on synaptic specificity mechanisms at the subcellular level (Sanes & Yamagata, 2009). Our algorithmic framework not only yields the anatomically constrained ensemble of connectomes that can be derived from the reconstructed data but also the synaptic specificity parameters that may explain empirically observed connectivity patterns that cannot be explained by overlap of pre- and postsynaptic structures. To assess the quality of our models, we introduced a metric (loss function) that shows how well the tested assumptions of synaptic specificity can reproduce the empirically observed connectome. On a technical level, we utilize simulation-based inference to determine all synaptic specificity parameters that are compatible with the observed empirical data (Boelts et al., 2023; Tejero-Cantero et al., 2020). Linking the generative model with anatomical data allows us to compute and inspect the probabilities of empirically observed connectivity patterns at cellular resolution, i.e., at the level of individual neurons with their synaptic connections. We can therefore make experimentally testable predictions about biologically interpretable assumptions about synaptic specificity at the level of individual neurons.

We demonstrated the applicability of our method on dense reconstructions of the mouse visual (Turner et al., 2022) and human temporal cortex (Shapson-Coe et al., 2024). Despite the limitations of these datasets, this analysis already provided some remarkable insights into the origins of wiring specificity in cortical connectomes. For example, it has been a long-standing question to what degree neural morphology controls connectivity between neurons (Braitenberg & Schüz, 1991; Rees et al., 2017). Consistent with previous findings (Udvary et al., 2022), we show in the mouse and human cortex that a null model that makes no assumptions about synaptic specificity leads to non-random connectivity patterns (i.e., triplet motif distributions). However, while morphology influences such statistical network properties, it does not fully determine the empirically observed connectome. This indicates that neuron morphology is a confounding factor to various aggregate statistics of the wiring

diagram, however, overlap of pre- and postsynaptic structures alone can virtually never predict the nanoscale connectome that was observed experimentally.

When including assumptions about synaptic specificity, we were able to better predict aggregate connectivity statistics (e.g., triplet motif distributions or the number of multisynaptic connections). In our template model we tested tested three assumptions of synaptic specificity based on: (1) cell type level, (2) single cell identity, and (3) subcellular compartment (i.e., soma, dendrite, axon initial segment). These three assumptions, despite dataset limitations, provided consistent predictions about synaptic specificity in both the mouse and the human cortex. However, as both datasets continue to be proofread (many neurites in the reconstructed volumes have no cell type labels and cannot be traced back to a soma), the specificity parameter values may change accordingly. Nonetheless, it is tempting to speculate that very similar mechanisms of wiring are preserved across species. In any case, our template analyses illustrate how our method can provide concrete predictions for each neuron and what kind of specificity could be involved to account for its empirically observed synaptic connectivity.

Therefore, our method can help to clarify the role of morphology in constraining neural network configurations while indicating where additional assumptions of local synaptic specificity are necessary to account for the empirically observed connectivity patterns. Our results indicate that anatomy alone is a confounding factor to some connections and aggregate statistics, while the presence (or absence) of other connections requires further assumptions about synaptic specificity, which suggests a complex interplay of morphology and synaptic specificity (Rees et al., 2017) for generating the connectome. Our approach thus provides a robust framework for analyzing wiring specificity, offering insights into how anatomical and other factors together shape the connectome across species.

From a technical perspective, the presented method falls into the category of generative network models, which have significantly advanced our understanding of connectomes (Betzel & Bassett, 2017). Generative models of varying complexity can be used to test the statistical significance of empirically observed connectivity patterns and to identify the minimal set of generative rules that can explain certain characteristics of the connectome (Váša Františekand Mišić, 2022). When applied to macroscale brain networks, the nodes in such network models typically represent entire brain regions and edges represent structural or functional connections between these regions. (Fornito et al., 2016). Generative models applied at the macroscale can successfully predict architectural features of the connectome that closely match empirical observations (Betzel et al., 2016; Beul et al., 2017). When applied at the microscale, the nodes in network models typically refer to neurons and the edges to synaptic connections. At this level, it has been shown that using only combinations of simple structural features, maximum-entropy models can generate realistic connectomes across species (Haber et al., 2023). In contrast to ‘single shot’ models that assign connectivity in one step, growth and developmental models aim to explicitly model the temporal evolution of the neural network (Betzel & Bassett, 2017). Following the above nomenclature, our method is a single shot generative model operating at the subcellular scale, where it predicts synaptic connectivity inside overlap volumes.

In summary, the presented method enables us to generate anatomically constrained ensembles of connectomes directly from dense reconstructions of neural tissue and infer quantitative information about synaptic specificity in a robust and well-defined manner. The framework, which is publicly available as a Python library, provides users with generic interfaces to define their own assumptions

of synaptic specificity and identify the necessary, sufficient, and best-suited parameters for a given connectome dataset. It can therefore be tested which types of synaptic specificity may or may not generalize across datasets. This could facilitate tracing and proof-editing efforts in the future as well as comparative approaches across species and developmental stages. Our method may therefore facilitate ongoing efforts (Loomba et al., 2022) that seek to understand cortical cell types through joint analysis of connectivity, morphology, and the underlying synaptic specificity profiles. This sets the stage for future experiments to test what kind of molecular mechanisms could be at work and at which time in neural development.

Acknowledgements

This work was supported by the German Research Foundation (DFG) through SPP 2041 “Computational Connectomics” and the German Federal Ministry of Education and Research (BMBF) through the Tübingen AI Center, FKZ: 01IS18039A. J.H.M. is a member of the MLCoE, EXC number 2064/1 – project number 390727645. We declare that we have no competing interests.

Author contributions

M.O. and P.H. conceived the study; P.H. developed the computational framework and performed the analyses under supervision of D.B. and H.H.; D.U. contributed to the null model of synapse formation and the definition of connectivity statistics; J.B. and J.H.M. developed the simulation-based Bayesian inference framework; M.O. and P.H. wrote the paper.

Figures and Figure Legends

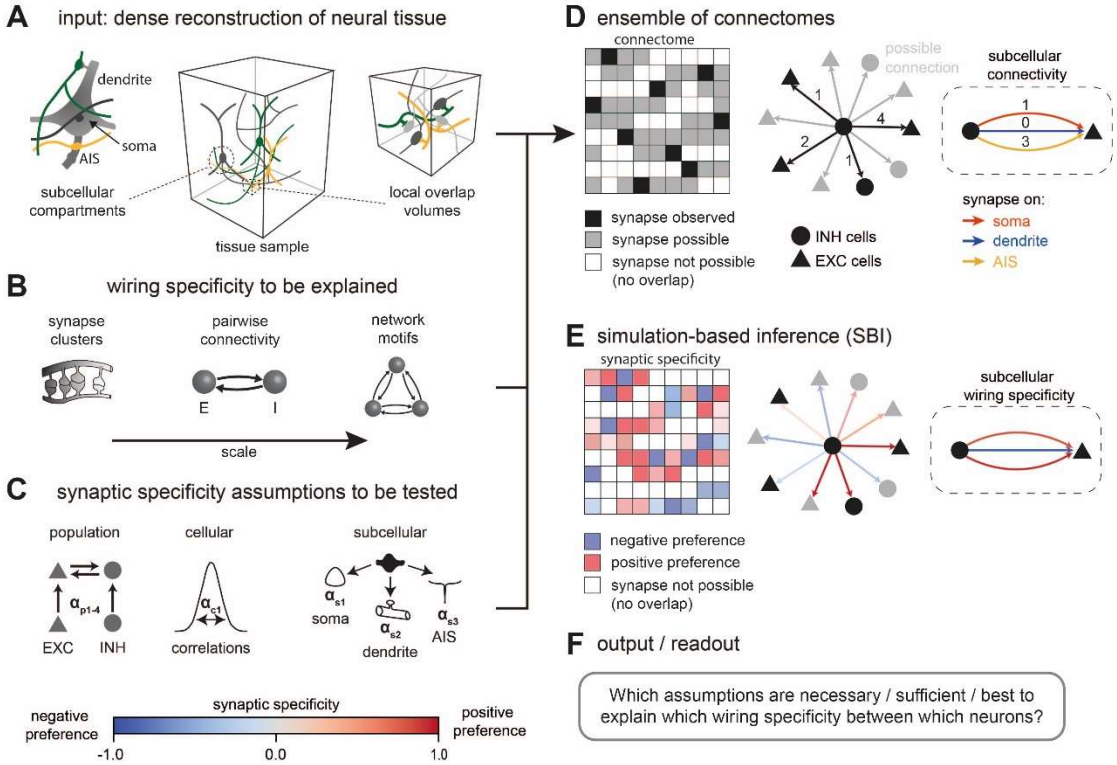

Figure 1: Framework for dissecting origins of wiring specificity in dense reconstructions. (A) Input to the model: dense reconstruction of a neural tissue volume (middle) providing ground truth synaptic connectivity down to subcellular detail (left), differentiating synaptic connectivity onto, e.g., soma, dendritic spines, and axon initial segment (AIS). For analysis, the reconstructed volume is uniformly partitioned into local overlap volumes (right). **(B)** Examples of connectivity statistics that

may indicate wiring specificity (i.e., deviations from a null model): presence of multisynaptic connections (i.e., synapse clusters), aggregate synapse counts (e.g., between excitatory (E) and inhibitory (I) neuron populations), and frequency of certain network motifs. **(C)** Examples of synaptic specificity assumptions that can be tested with the framework: cell type-dependent specificity, specificity based on single cell identity (correlated wiring), and specificity based on subcellular target domain. Each synaptic specificity mechanism is modeled through negative ($\alpha_i < 0$) or positive ($\alpha_i > 0$) preference values in reference to a null model ($\alpha_i = 0$) of locally random synapse formation. **(D)** Left: ensemble of anatomically possible connectomes visualized as potential connectivity matrix. Middle: node link diagram showing observed (black) and potential connections (gray) of an individual neuron. Right: expanded edge with synapse counts onto soma, dendrite, and AIS. **(E)** Simulation-based Bayesian inference (SBI) identifies synaptic specificity parameters α_i that could have generated the empirically observed connectome. The inferred parameters can be visualized as synaptic specificity matrix (left) and as node link diagram (middle) down to subcellular resolution (right). **(F)** Overall, the modeling approach allows linking local synaptic specificity to empirically observed connectivity statistics at different scales.

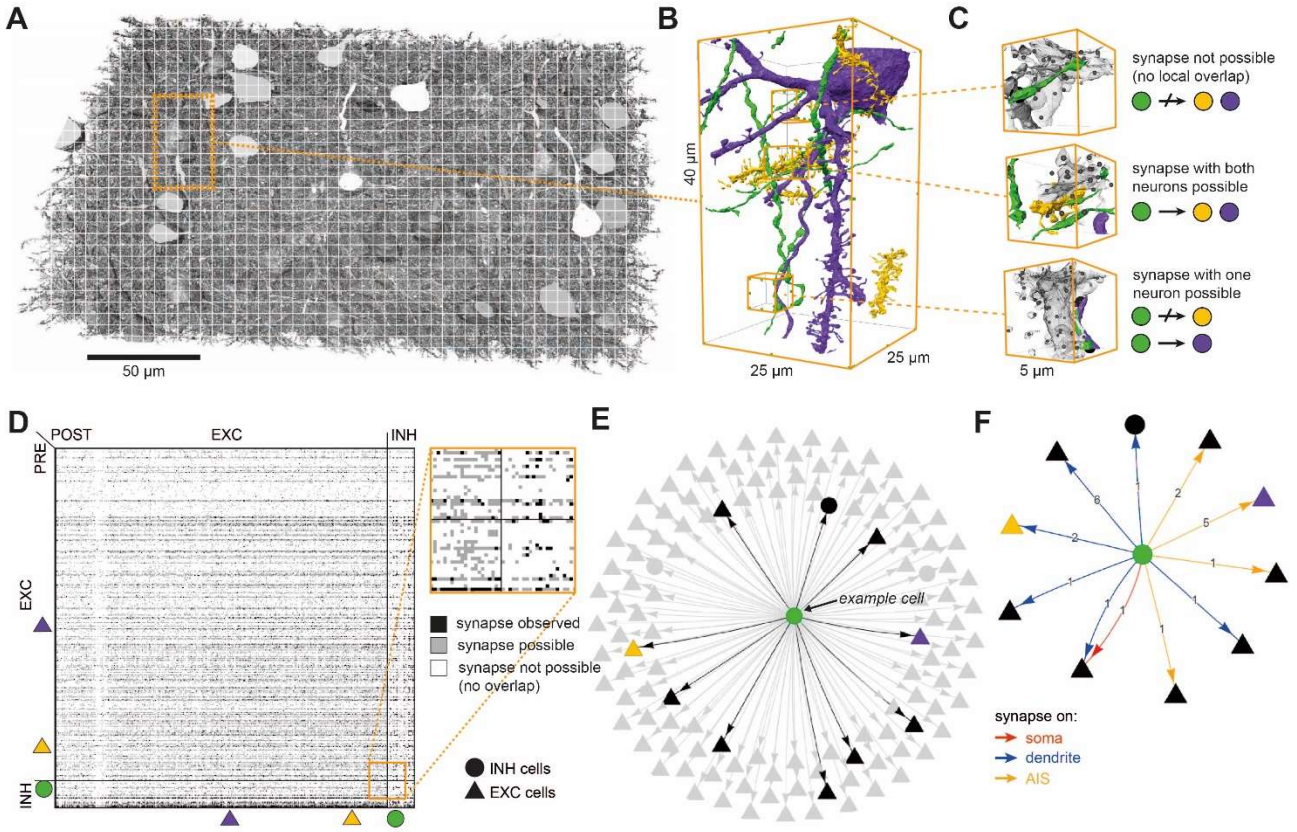

Figure 2: Generating an ensemble of structurally possible connectomes from data. (A) Dense reconstruction of a volume in layer 2/3 of the mouse visual cortex (Turner et al., 2022). The entire tissue volume is uniformly partitioned into local overlap volumes (indicated by the white grid). **(B)** Manually selected subvolume of the dataset covering axon segments of one interneuron (green), dendrite segments of an excitatory neuron (yellow), and soma, dendrite, and axon initial segments of another excitatory neuron (purple). **(C)** Selected local overlap volumes from panel (B) based on a $5\,\mu m$ grid showing different configurations of innervating neurites. The presence of pre- and postsynaptic sites in at least one shared local overlap volume is required for synaptic connectivity between two neurons. **(D)** Potential connectivity matrix representing ensemble of anatomically feasible connectomes. Rows correspond to presynaptic neurons and columns to postsynaptic

neurons. Matrix cells indicate if a synapse between two neurons is possible but absent (gray), possible and empirically observed (black), and not possible due to missing overlap (white). The corresponding rows and columns of the three highlighted neurons (green, yellow, purple) are marked with symbols. **(E)** Node link diagram showing all potential (gray) and empirically observed (black) connections of the highlighted interneuron (center, green). **(F)** Edges in the node link diagram can be expanded to distinguish synaptic connections onto soma, dendrite, and axon initial segment (AIS). Edge labels show the number of synapses.

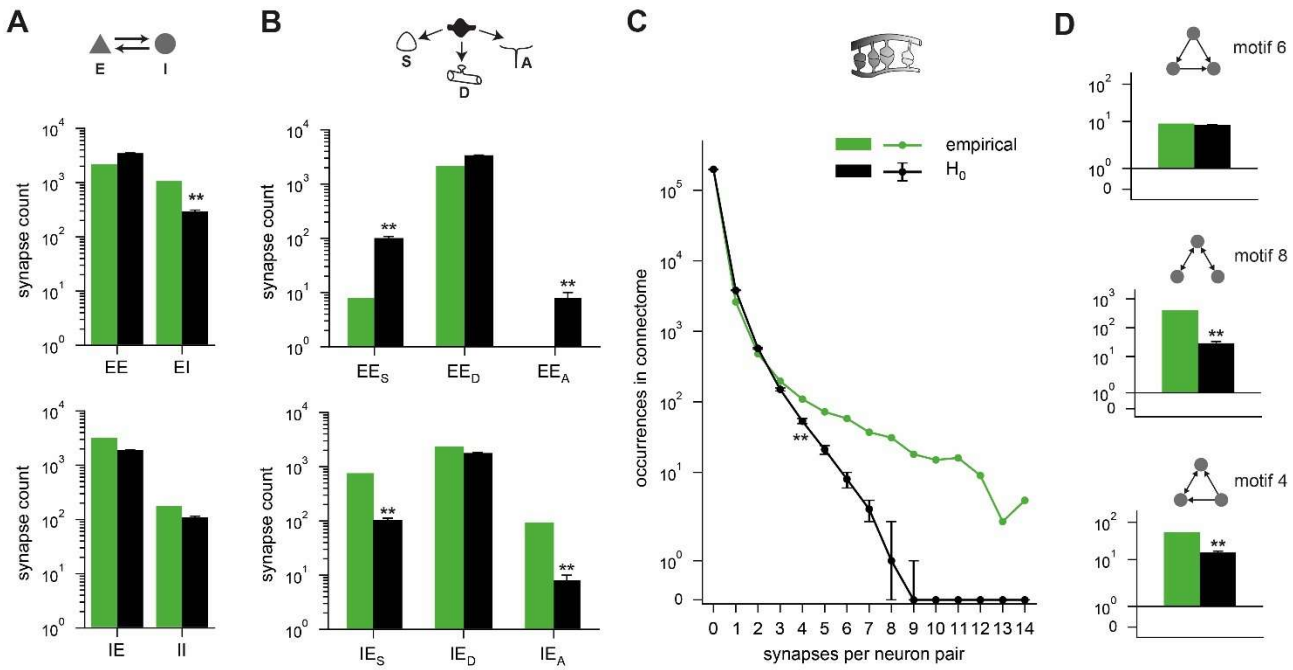

Figure 3: Evidence for wiring specificity in the mouse visual cortex (Turner et al., 2022). Comparison of empirically observed connectivity statistics (green) with null model (H_0, black) that assumes the absence of synaptic specificity. **(A)** Connectivity between excitatory (E) and inhibitory (I) neuron populations is qualitatively well approximated by the null model; the largest deviation is observed in the number of excitatory to inhibitory synapses (EI). **(B)** When synapse counts are disaggregated by the target domain of the postsynaptic neuron (S: soma, D: dendrite, A: axon initial segment), significant deviations are observed for connections on the soma and the axon initial segment but less so for connections on the dendrite. **(C)** The null model underestimates the presence of multisynaptic connections (i.e., synapse clusters), particularly, when such connections consist of more than three synapses. **(D)** Some triplet motifs are already well-approximated by the null model (e.g., motif 6) while significant deviations for others exist (e.g., motif 8 and 4); the full motif spectrum and additional statistics are given in Fig. S2.

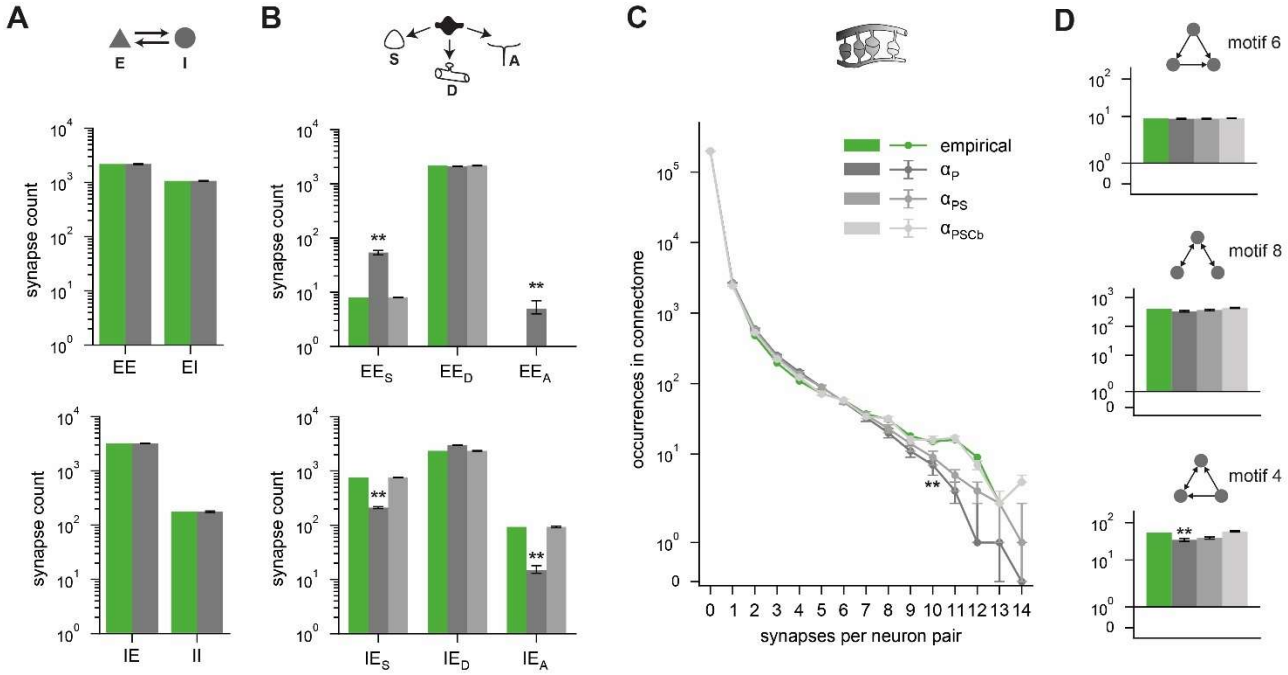

Figure 4: Introducing synaptic specificity into the model leads to better approximations of connectivity statistics. (A) Aggregate synapse counts between excitatory (E) and inhibitory (I) neuron populations. Using population specificity α_P, the connectivity between excitatory (E) and inhibitory (I) neuron populations is very well approximated. **(B)** Synapse counts disaggregated by the subcellular target domain of the postsynaptic neurons (S: soma, D: dendrite, A: axon initial segment). Synapse counts are still disaggregated by the target domain of the postsynaptic neuron when only using population specificity α_P. Adding subcellular specificity accounts for this. **(C)** The α_P and α_{PS} models still underestimate the number of multisynaptic connections (i.e., synapse clusters) between neurons; however, they both significantly improve on the null model H_0 shown in Fig. 3C. Additionally including synaptic specificity based on single cell identity (α_{PSCb}) leads to a close prediction of multisynaptic connections. **(D)** Triplet motif 6 is already well approximated by the null model (see Fig. 3D) and hence also by the specificity models; the α_P and α_{PS} models improve the fit of motif 8 and 4 compared to the null model to varying degrees. For exact match of all motifs, the most expressive α_{PSCb} model is required. Refer to Fig. S2 for the full motif spectrum and additional statistics.

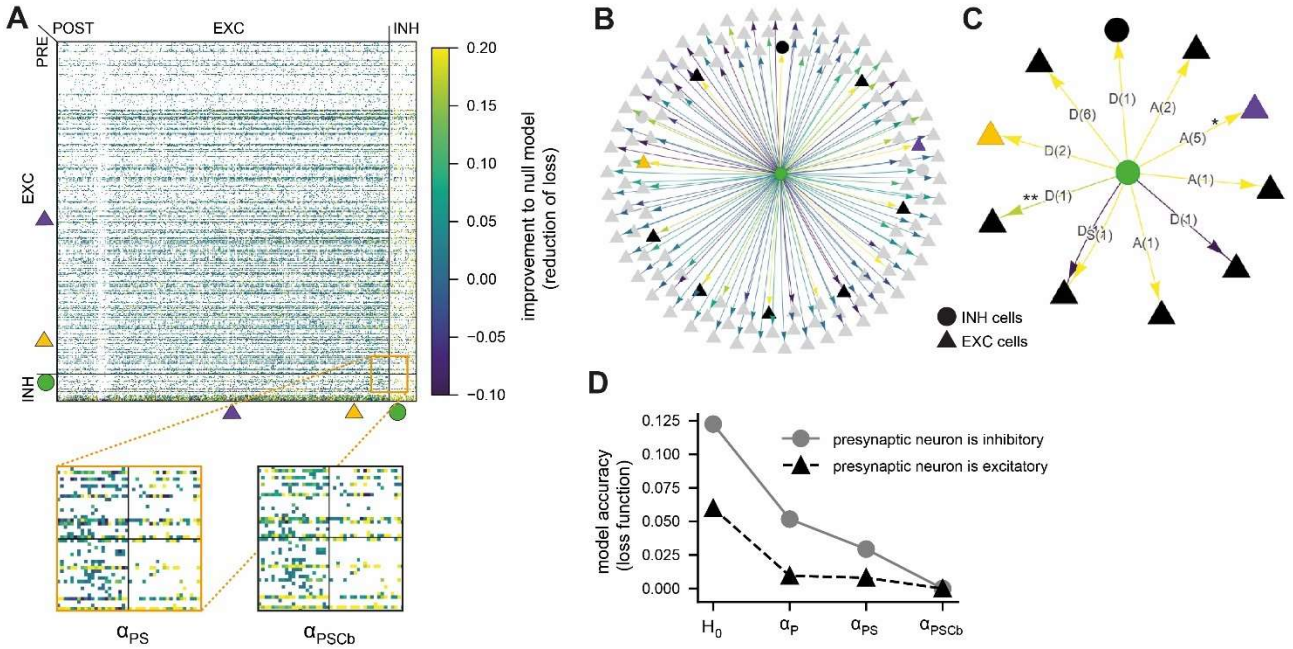

Figure 5: Quantifying the accuracy of models at cellular and subcellular resolution with a loss function. (A) Delta loss matrix showing the improvement (i.e., reduction in loss function) of the α_{PS} model in reference to the null model H_0 for all neuron pairs. Matrix rows correspond to presynaptic neurons and matrix columns to postsynaptic neurons; brighter colors indicate a higher reduction in loss. Bottom: zoomed-in matrix areas show comparison of α_{PS} with α_{PSCb} model; the more expressive α_{PSCb} model results in a higher reduction in loss compared to the null model H_0 for the same neuron pairs. **(B)** Node link diagram showing one presynaptic interneuron (green) and its postsynaptic partners (compare to Fig. 2E). Edge colors indicate the model improvement analogue to the matrix in panel A. **(C)** Expanded edges show the model improvement separately for different postsynaptic compartments (S: soma, D: dendrite, A: axon initial segment). The edge marked with ** shows an excitatory neuron for which the model improvement is smaller than for other excitatory neurons (e.g., the edge marked with *). **(D)** With increasing model complexity (i.e., when additional synaptic specificity sources are introduced) the absolute loss converges to zero.

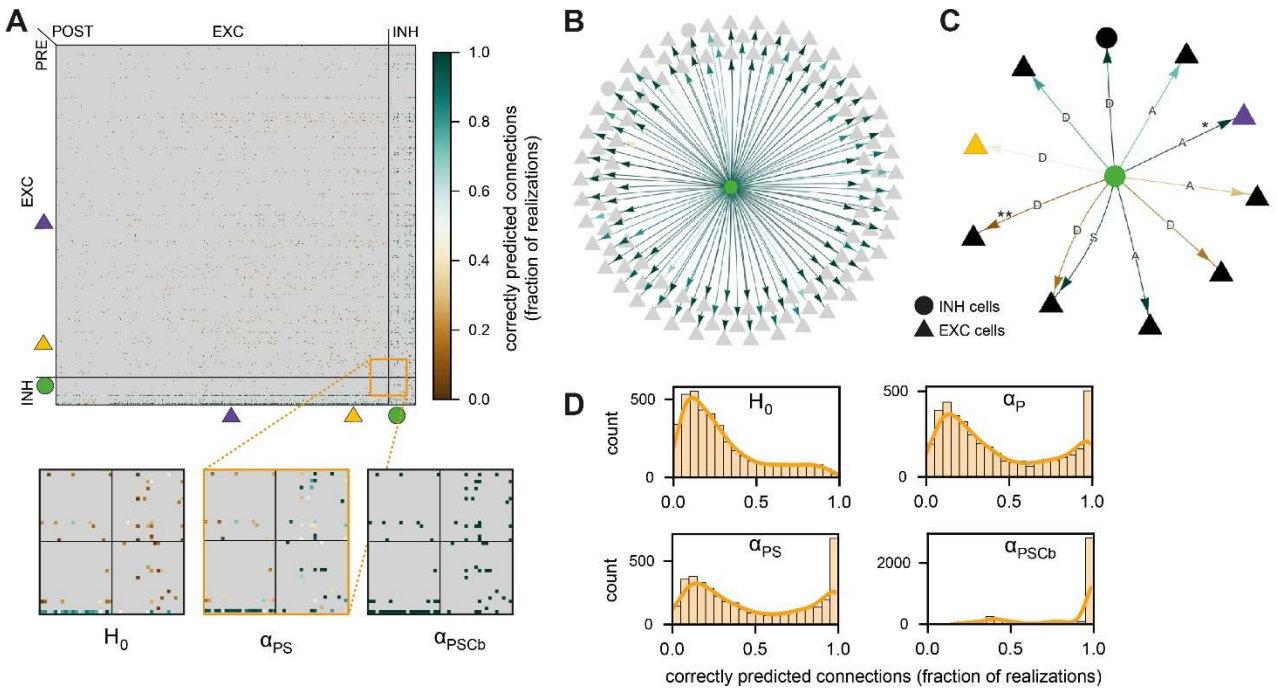

Figure 6: Generating discrete realizations of the connectome using different models. (A) Matrix showing fraction of correctly predicted connections in n=1000 connectome realizations based on the α_{PS} model. Brown colors indicate a prediction accuracy below 50% and green colors above 50% (i.e., if two neurons are connected in the empirically observed connectome, the same neurons are connected in more than 50% of the model-generated connectomes). Bottom: zoomed-in matrix area shows a comparison between null model H_0, model α_{PS}, and model α_{PSCb}. With increasing model complexity, the model accuracy increases. **(B)** Node link diagram showing one presynaptic inhibitory neuron (compare to Fig. 5B) and its potential postsynaptic partner cells to which it is *not* connected empirically. Edge colors indicate the fraction of model-generated connectomes in which such connections are also absent (i.e., the accuracy of predicting unconnected neurons). **(C)** The same neuron from panel B with the postynaptic partners to which it *is* empirically connected. Edges are expanded to differentiate the postynaptic compartment (S: soma, D: dendrite, A: axon initial segment). Edge colors indicate the prediction accuracy analogous to the matrix in panel A. The edge marked with ** shows a connection that is poorly predicted in contrast to the edge marked with * that has a very high prediction accuracy (compare to Fig. 5C). **(D)** Distribition of correctly predicted connections under different models.

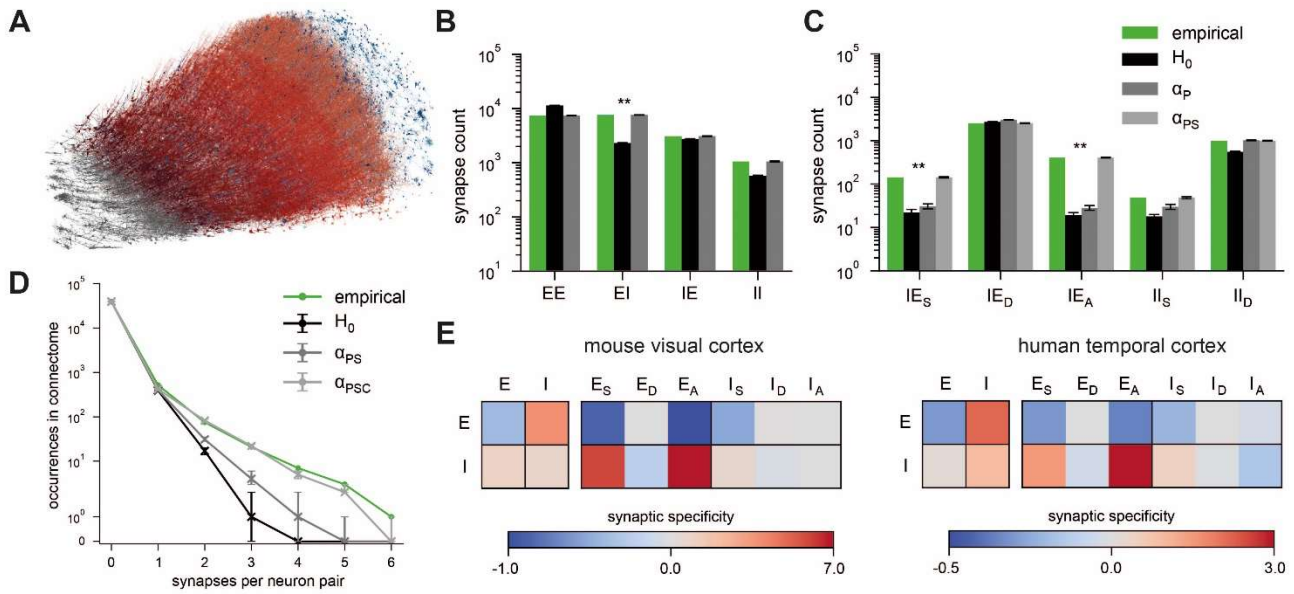

Figure 7: Applying the framework to a dense reconstruction of the human temporal cortex (Shapson-Coe et al., 2024)**. (A)** Visualization of the human temporal cortex dataset showing excitatory neuron morphologies in red (shaded by cortical layer) and inhibitory neurons in blue. **(B)** Aggregate synapse counts between excitatory (E) and inhibitory (I) neurons. Like in the mouse visual cortex (see Fig. 3A) the null model H_0 underestimates the number of excitatory to inhibitory synapses (EI). Introducing synaptic specificity based on cell type (α_P) adjusts for these deviations. **(C)** Synapse counts from inhibitory neurons disaggregated by the postsynaptic target domain (S: soma, D: dendrite, A: axon initial segment). **(D)** Number of synapses per neuron pair. Values larger than 1 represent multisynaptic connections (i.e., synapse clusters). **(E)** Left: synaptic specificity parameters in the mouse visual cortex for the α_P model (2×2 table) and α_{PS} model (2×6 table). Right: the corresponding synaptic specificity parameters in the human temporal cortex.

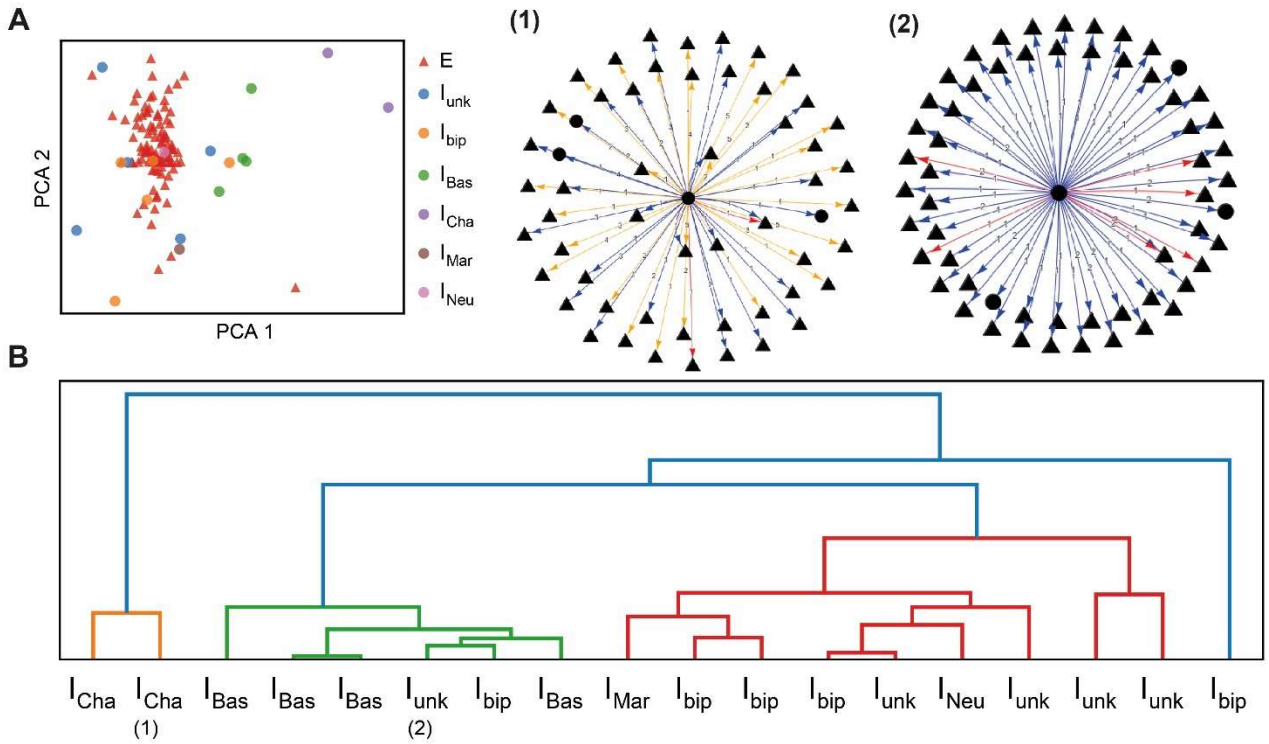

Figure 8: Clustering neurons based on their synaptic specificity. (A) PCA-embedding of excitatory and inhibitory neurons based on their synaptic wiring specificity. The embedding separates the inhibitory subtypes from the excitatory neurons (red triangles), which are clustered together. **(B)** Hierarchical clustering of manually labelled interneurons (Cha: Chandelier, Bas: Basket, bip: bipolar, Mar: Martinotti Neu: Neurogliaform, unk: unknown) based on the synaptic specificity of their axons. The automated clustering based on synaptic specificity reflects the manually assigned classes (e.g., both Chandelier cells are assigned to the orange cluster and the Basket cells to the green cluster). **(1)** Chandelier cell targeting axon initial segments. **(2)** An interneuron with unknown subtype that was assigned to the green cluster consisting of four Basket cells and one bipolar cell.

Methods

Code availability

The framework is available at: https://github.com/zibneuro/connectomedf

Dataset preprocessing

As raw input the method requires a csv-file with all identified synapses in the dense reconstruction. The required fields are the x/y/z-coordinates of the synapses (in discrete units, e.g., nanometer) and the IDs of the pre- and postsynaptic neurons. Additional meta information can be provided in extra columns. In our analyses we used 3 additional columns based on annotations from the datasets: presynaptic cell type, postsynaptic cell type, and postsynaptic compartment (i.e., soma, dendrite, axon initial segment). The only requirement for extra columns is that they are integer numbers, which allows including custom cell type classification schemes and other categorical labels from the dataset. The framework ‘splits’ the synapses from the csv-file into pre- and postsynaptic contact sites and assigns these potential contact sites to local overlap volumes that spatially partition entire volume of the dataset. Within the overlap volumes contact sites with the same annotations are aggregated and their count is recorded. For example, if the axon of a neuron with cell type ‘E’ and neuron ID ‘9123’ makes two synapses in an overlap volume, the overlap volume internally records these presynaptic contact sites in a table using the neuron properties as row index and the number of contact sites as

value (E, 9123): 2. Consequently, the exact x/y/z-locations of synapses within the overlap volumes are discarded. Synapses from or onto orphaned neurites (i.e., axons or dendrites that are not traced back to a cell body inside the reconstructed volume) and that may therefore not have meaningful cell type or neuron ID annotations, can be assigned the default annotation -1. Such background synapses from orphaned neurites are then aggregated into one group of unspecified contact sites, e.g., (-1, -1): 10237. Postsynaptic contact sites additionally contain the postsynaptic compartment as attribute (S: soma, D: dendrite, A: axon initial segment). Assume that the dendrite ('D') of an excitatory ('E') neuron with neuron ID '9123' receives 3 synapses in an overlap volume. Then the table of postsynaptic contact sites will contain the following row: (E, 9123, D): 3.

To compile the input csv-file of synapses for the mouse visual cortex dataset (Turner et al., 2022), we obtained the raw data with 3.2 million synapses from https://doi.org/10.5281/zenodo.3710458 . We additionally included annotations of synapses located on axon initial segments (Schneider-Mizell et al., 2021) and assigned 'soma' compartment labels to synapses within a distance of $15\ \mu m$ to the cell body center of the respective postsynaptic neuron. We assigned 450 annotated cells with cell body inside the reconstructed volume as 'selected' neurons and set the neuron IDs orphaned neurites to -1. We partitioned the entire volume using 5 / 10 / 25 μm sized uniform grids resulting in 32441 / 361 / 4431 local overlap volumes.

Our analyses are based on the C3 stage segmentation of the human temporal cortex (Shapson-Coe et al., 2024) with 133.7 million synapses. A common challenge in dense reconstructions of neural tissue based on electron microscopy are incompletely traced neurites, especially axons (Wei et al., 2021). Connectivity between neurons at the cellular level is therefore underestimated in the currently available large-scale datasets. To compensate for this effect, we identified 200 strongly connected cells in layer 3, which we denote as 'selected' neurons. We refer to pyramidal cells as excitatory and to interneurons as inhibitory. Using a $25\ \mu m$ grid we partitioned volume innervated by axons or dendrites of the selected neurons into 9184 local overlap volumes containing 36.8 million synapses.

Model for synaptic specificity

A specificity model f_α computes the expected number of synapses $\lambda_{a,b,k}$ between groups of pre- and postsynaptic contact sites a, b in an overlap volume k: $\lambda_{a,b,k}(\alpha) := f_\alpha(\lambda^*_{a,b,k})$. Models can be concatenated such that one model operates on the predicted number of synapses $\lambda^*_{a,b,k}$ from a previous model. Starting from a null model H_0, we defined a chain of models that represents increasing levels of synaptic specificity: $f_{H_0} \rightarrow f_{\alpha_P} \rightarrow f_{\alpha_{PS}} \rightarrow f_{\alpha_{PSC}}$.

Presynaptic contact sites have the attributes cell type population P (e.g., $P = \{E, I\}$) and single cell identity $C = \{id_1, id_2, \ldots\}$. Postsynaptic contact sites have an additional subcellular compartment attribute $S = \{S, D, A\}$ where S denotes soma, D dendrite, and A axon initial segment. Hence, the possible groupings of presynaptic contact sites a and postsynaptic contact sites b are the respective cartesian products of these attributes, i.e., $a \in P \times C$ and $b \in P \times C \times S$.

The parameter-free null model f_{H_0} computes synapse counts under the assumption that each connection inside the overlap volume is equally likely. The expected synapse counts under the null model are $f_{H_0}: \lambda_{a,b,k} = |a| \cdot |b| \cdot n_k^{-1}$ where $|a|$ and $|b|$ denote the number of pre- and postsynaptic contact sites for the respective group (e.g., $|a| = |(\mathrm{I}, 812)| = 2$ and $|b| = |(\mathrm{E}, 9123, \mathrm{D})| = 3$) and n_k is the total synapse count in the overlap volume k, which is always preserved under the null model: $\sum_{a,b} \lambda_{a,b,k} = n_k$.

The specificity model based on cell type population f_{α_P} introduces specificity parameters α_P for every cell type pair. E.g., assume that the cell type populations are $P = \{E, I\}$, then $\alpha_P = (\alpha_{EE}, \alpha_{EI}, \alpha_{IE}, \alpha_{II})$.

The cell type specificity parameters α_{xy} adjust the respective synapse counts from the null model: f_{α_P}: $\lambda_{a,b,k} = \alpha_{xy} \cdot \lambda_{a,b,k}^{H_0}$ where $a \in \{x\} \times C$ and $b \in \{y\} \times C \times S$.
The specificity model based on subcellular compartment $f_{\alpha_{PS}}$ introduces additional parameters based on the target site of the postsynaptic neuron $\alpha_S = (\alpha_S, \alpha_D, \alpha_A)$ that adjust the synapse counts from the preceding model: $f_{\alpha_{PS}}$: $\lambda_{a,b,k} = \alpha_x \cdot \lambda_{a,b,k}^{\alpha_P}$ where $a \in P \times C$ and $b \in P \times C \times \{x\}$.
The specificity model based on single cell identity $f_{\alpha_{PSC}}$ introduces specificity parameters based on the identity of the postsynaptic neuron $\alpha_C = (\alpha_{id1}, \alpha_{id2}, \dots)$ that are applied to the preceding model: $f_{\alpha_{PSC}}$: $\lambda_{a,b,k} = \alpha_x \cdot \lambda_{a,b,k}^{\alpha_{PS}}$ where $a \in P \times C$ and $b \in P \times \{x\} \times S$.
The domain of specificity parameters α_i is $\mathbb{R}_0^+$. Specificity parameters $\alpha_i < 1$ can be associated with a negative wiring preference because the model-predicted synapse count is scaled down compared to the reference synapse count; likewise, parameters $\alpha_i > 1$ can be associated with a positive wiring preference. For improved clarity, we visualize the specificity parameters decremented by 1 as $\alpha_i' := \alpha_i - 1$. Then $-1 < \alpha_i' < 0$ denotes a negative preference, $\alpha_i' = 0$ no preference, and $\alpha_i' > 0$ a positive preference.
To generate discrete realizations of the connectome from the model-predicted synapse counts $\lambda_{a,b,k}$ in the overlap volumes, we sample discrete synapse counts $n_{a,b,k}$ from a Poisson distribution (i.e., $n_{a,b,k} \sim Poisson(\lambda_{a,b,k})$) unless $\lambda_{a,b,k}$ is itself discrete in cases where no local combinatorial variability is left.

Parameter inference

We can analytically determine specificity parameters α_x in all overlap volumes, which provides us with a vector/distribution $[\alpha_x]$ of locally optimal specificity parameters as a result. We do so by solving $\lambda_{a,b,k}^E = \alpha_x \cdot \lambda_{a,b,k}^{prec}$ for α_x where $\lambda_{a,b,k}^E$ is the known empirical synapse count (for the grouping of pre- and postsynaptic contact sites a, b in overlap volume k) and $\lambda_{a,b,k}^{prec}$ the synapse count predicted by the preceding model in the chain of models.
Generally, user-defined synaptic specificity models may have parameters that are not analytically tractable. Simulation-based Bayesian inference (SBI) (Cranmer et al., 2020; Greenberg et al., 2019; Papamakarios & Murray, 2016) allows inferring suitable parameters for analytically intractable models from a set of simulation runs with sampled parameters. The prerequisites for applying SBI are (1) a stochastic simulator $p(x|\alpha)$ that takes parameters α as input and generates observable statistics x, (2) access to ground truth observed statistics x_0, and (3) a prior distribution over model parameters $p(\alpha)$. SBI uses a set of parameters sampled from the prior and corresponding simulated data $\{\alpha_i, x_i\}_{i=1}^N$ to train an artificial neural network-based density estimator to approximate the posterior distribution. After training, the density estimator is applied to the observed data x_0 to obtain the posterior distribution over model parameters $p(\alpha|x_0)$. Importantly, the inferred posterior distribution gives us access to all data-compatible parameters and their covariance structure, enabling us to analyze parameter uncertainties and interactions (see Boelts et al., 2023; Gonçalves et al., 2020for details).
We observed that the distribution of analytically determined specificity parameters (i.e., calculated in each overlap volume) is compatible with specificity parameters that are inferred globally (**Fig. S6**).
For extended usage examples of the SBI toolkit (Tejero-Cantero et al., 2020) refer to sbi-dev/sbi: Simulation-based inference toolkit.

Loss function

Let $n_{a,b}$ be the empirically observed number of synapses from a presynaptic neuron a onto the specific compartment of a postsynaptic neuron b and let $\lambda_{a,b} = \sum_k \lambda_{a,b,k}$ be the corresponding model-predicted

synapse count summed over all overlap volumes k. Assuming a Poisson random process, the probability that a connects with at least one synapse to b is $p_{a,b} = 1 - \exp(-\lambda_{a,b})$ and $\bar{p}_{a,b} = 1 - p_{a,b}$. We define the model loss as:

$$loss(\lambda_{a,b}) = (-1) \cdot \begin{cases} \log(\bar{p}_{a,b}), & n_{a,b} = 0 \\ \log(p_{a,b}), & n_{a,b} > 0 \end{cases}$$

We measure the improvement of a specificity model α to the null model H_0 by computing the delta loss, i.e., $\Delta\text{loss}_{a,b}(\alpha, H_0) = loss(\lambda_{a,b}^{H_0}) - loss(\lambda_{a,b}^{\alpha})$.

Connectivity statistics

The connectivity statistics were computed based on discrete realizations of the model-generated connectomes. Aggregate synapse counts were directly obtained by grouping synapses according to the respective attributes (e.g., pre- and postsynaptic cell type). Synapse clusters refer to connections between selected neurons (see data preprocessing) in both datasets. We count all occurrences of the 16 canonical triplet motifs between the selected neurons using a GPU-based implementation and normalize these occurrence counts by an equivalent random network to obtain the relative frequencies.

Clustering based on synaptic specificity

We analyzed annotated neurons (see data preprocessing) in the mouse visual cortex with at least 3 outgoing synapses to other annotated cells, resulting in n=127 neurons (109 excitatory, 18 inhibitory). We defined a specificity model that is agnostic to the (known) presynaptic cell type, thereby simulating a scenario in which these neurons are either unclassified or represent incompletely traced axon fragments. Based on the synaptic specificity parameters determined by the framework, we defined a $X := n \times 5$ feature matrix in which the rows represent neurons and the columns their specificity values onto: (1) excitatory cells, (2) inhibitory cells, (3) soma, (4) dendrite, and (5) axon initial segment. Using scikit-learn we performed a standard scaling of the features and generated the dendrograms with the linkage function from scipy (ward method with optimal leaf ordering). The PCA-embedding was performed with sklearn. After computing the dendrogram and embedding, the known cell type labels from the dataset were added for reference.

Visualization and statistical analysis

Bar charts and line charts show the median of the underlying distribution of connectivity statistics, and error bars indicate the 25th and 75th percentiles, respectively. The colored specificity matrices show the mean values of the respective parameter distributions, mapped to the diverging blue/red color scale with the TwoSlopeNorm function from matplotlib. All charts were generated with matplotlib in combination with the numpy and sciPy libraries. Anatomical renderings of cells from the mouse visual cortex and the human temporal cortex were generated with Amira (Stalling et al., 2005).

References

Agi, E., Kulkarni, A., & Hiesinger, P. R. (2020). Neuronal strategies for meeting the right partner during brain wiring. *Current Opinion in Neurobiology*, *63*, 1–8. https://doi.org/10.1016/j.conb.2020.01.002

Betzel, R. F., Avena-Koenigsberger, A., Goñi, J., He, Y., de Reus, M. A., Griffa, A., Vértes, P. E., Mišic, B., Thiran, J. P., Hagmann, P., van den Heuvel, M., Zuo, X. N., Bullmore, E. T., & Sporns, O. (2016). Generative models of the human connectome. *NeuroImage*, *124*, 1054–1064. https://doi.org/10.1016/j.neuroimage.2015.09.041

Betzel, R. F., & Bassett, D. S. (2017). Generative models for network neuroscience: Prospects and promise. *Journal of the Royal Society Interface*, *14*(136). https://doi.org/10.1098/rsif.2017.0623

Beul, S. F., Barbas, H., & Hilgetag, C. C. (2017). A Predictive Structural Model of the Primate Connectome. *Scientific Reports*, *7*(March), 1–12. https://doi.org/10.1038/srep43176

Boelts, J., Harth, P., Gao, R., Udvary, D., Yáñez, F., Baum, D., Hege, H.-C., Oberlaender, M., & Macke, J. H. (2023). Simulation-based inference for efficient identification of generative models in computational connectomics. *PLOS Computational Biology*, *19*(9), e1011406. https://doi.org/10.1371/journal.pcbi.1011406

Braitenberg, V., & Schüz, A. (1991). *Peters' Rule and White's Exceptions* (pp. 109–112). https://doi.org/10.1007/978-3-662-02728-8_21

Brown, S. P., & Hestrin, S. (2009). Intracortical circuits of pyramidal neurons reflect their long-range axonal targets. *Nature*, *457*(7233), 1133–1136. https://doi.org/10.1038/nature07658

Cranmer, K., Brehmer, J., & Louppe, G. (2020). The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, *117*(48), 30055–30062. https://doi.org/10.1073/pnas.1912789117

Fares, T., & Stepanyants, A. (2009). Cooperative synapse formation in the neocortex. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(38), 16463–16468. https://doi.org/10.1073/pnas.0813265106

Favuzzi, E., Deogracias, R., Marques-Smith, A., Maeso, P., Jezequel, J., Exposito-Alonso, D., Balia, M., Kroon, T., Hinojosa, A. J., F. Maraver, E., & Rico, B. (2019). Distinct molecular programs regulate synapse specificity in cortical inhibitory circuits. *Science*, *363*(6425), 413–417. https://doi.org/10.1126/science.aau8977

Fornito, A., Zalesky, A., & Bullmore, E. T. (2016). *Fundamentals of brain network analysis*. Elsevier Academic Press. https://doi.org/10.1016/C2012-0-06036-X

Gonçalves, P. J., Lueckmann, J.-M., Deistler, M., Nonnenmacher, M., Öcal, K., Bassetto, G., Chintaluri, C., Podlaski, W. F., Haddad, S. A., Vogels, T. P., Greenberg, D. S., & Macke, J. H. (2020). Training deep neural density estimators to identify mechanistic models of neural dynamics. *ELife*, *9*. https://doi.org/10.7554/eLife.56261

Greenberg, D., Nonnenmacher, M., & Macke, J. (2019). Automatic Posterior Transformation for Likelihood-Free Inference. In K. Chaudhuri & R. Salakhutdinov (Eds.), *Proceedings of the 36th International Conference on Machine Learning* (Vol. 97, pp. 2404–2414). PMLR. https://proceedings.mlr.press/v97/greenberg19a.html

Haber, A., Wanner, A. A., Friedrich, R. W., & Schneidman, E. (2023). The structure and function of neural connectomes are shaped by a small number of design principles. *BioRxiv*, 2023.03.15.532611. https://doi.org/10.1101/2023.03.15.532611

Kaiser, M. (2017). Mechanisms of Connectome Development. In *Trends in Cognitive Sciences* (Vol. 21, Issue 9, pp. 703–717). Elsevier Ltd. https://doi.org/10.1016/j.tics.2017.05.010

Karimi, A., Odenthal, J., Drawitsch, F., Boergens, K. M., & Helmstaedter, M. (2020). Cell-type specific innervation of cortical pyramidal cells at their apical dendrites. *ELife*, *9*. https://doi.org/10.7554/eLife.46876

Kasthuri, N., Hayworth, K. J., Berger, D. R., Schalek, R. L., Conchello, J. A., Knowles-Barley, S., Lee, D., Vázquez-Reina, A., Kaynig, V., Jones, T. R., Roberts, M., Morgan, J. L., Tapia, J. C., Seung, H. S., Roncal, W. G., Vogelstein, J. T., Burns, R., Sussman, D. L., Priebe, C. E., … Lichtman, J. W. (2015). Saturated Reconstruction of a Volume of Neocortex. *Cell*, *162*(3), 648–661. https://doi.org/10.1016/j.cell.2015.06.054

Kornfeld, J., & Denk, W. (2018). Progress and remaining challenges in high-throughput volume electron microscopy. In *Current Opinion in Neurobiology*. https://doi.org/10.1016/j.conb.2018.04.030

Loomba, S., Straehle, J., Gangadharan, V., Heike, N., Khalifa, A., Motta, A., Ju, N., Sievers, M., Gempt, J., Meyer, H. S., & Helmstaedter, M. (2022). Connectomic comparison of mouse and human cortex. *Science*, *377*(6602), eabo0924. https://doi.org/10.1126/science.abo0924

Macrina, T., Lee, K., Lu, R., Turner, N. L., Wu, J., Popovych, S., Silversmith, W., Kemnitz, N., Bae, J. A., Castro, M. A., Dorkenwald, S., Halageri, A., Jia, Z., Jordan, C., Li, K., Mitchell, E., Mondal, S. S., Mu, S., Nehoran, B., … Seung, H. S. (2021). Petascale neural circuit reconstruction: automated methods. *BioRxiv*. https://doi.org/10.1101/2021.08.04.455162

MICrONS Consortium, Bae, J. A., Baptiste, M., Bodor, A. L., Brittain, D., Buchanan, J., Bumbarger, D. J., Castro, M. A., Celii, B., Cobos, E., Collman, F., da Costa, N. M., Dorkenwald, S., Elabbady, L., Fahey, P. G., Fliss, T., Froudarakis, E., Gager, J., Gamlin, C., … Yu, S. (2021). Functional connectomics spanning multiple areas of mouse visual cortex. *BioRxiv*. https://doi.org/10.1101/2021.07.28.454025

Motta, A., Berning, M., Boergens, K. M., Staffler, B., Beining, M., Loomba, S., Hennig, P., Wissler, H., & Helmstaedter, M. (2019). Dense connectomic reconstruction in layer 4 of the somatosensory cortex. *Science*, *366*(6469). https://doi.org/10.1126/science.aay3134

Papamakarios, G., & Murray, I. (2016). Fast ε-Free Inference of Simulation Models with Bayesian Conditional Density Estimation. *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 1036–1044.

Rees, C. L., Moradi, K., & Ascoli, G. A. (2017). Weighing the Evidence in Peters' Rule: Does Neuronal Morphology Predict Connectivity? In *Trends in Neurosciences* (Vol. 40, Issue 2, pp. 63–71). Elsevier Ltd. https://doi.org/10.1016/j.tins.2016.11.007

Sanes, J. R., & Yamagata, M. (2009). Many Paths to Synaptic Specificity. *Annual Review of Cell and Developmental Biology*, *25*(1), 161–195. https://doi.org/10.1146/annurev.cellbio.24.110707.175402

Sanes, J. R., & Zipursky, S. L. (2020). Synaptic Specificity, Recognition Molecules, and Assembly of Neural Circuits. *Cell*, *181*(3), 536–556. https://doi.org/10.1016/j.cell.2020.04.008

Schneider-Mizell, C. M., Bodor, A. L., Collman, F., Brittain, D., Bleckert, A., Dorkenwald, S., Turner, N. L., Macrina, T., Lee, K., Lu, R., Wu, J., Zhuang, J., Nandi, A., Hu, B., Buchanan, J., Takeno, M. M., Torres, R., Mahalingam, G., Bumbarger, D. J., … Costa, N. M. da. (2021). Structure and function of axo-axonic inhibition. *ELife*, *10*. https://doi.org/10.7554/elife.73783

Shapson-Coe, A., Januszewski, M., Berger, D. R., Pope, A., Wu, Y., Blakely, T., Schalek, R. L., Li, P. H., Wang, S., Maitin-Shepard, J., Karlupia, N., Dorkenwald, S., Sjostedt, E., Leavitt, L., Lee, D., Troidl, J., Collman, F., Bailey, L., Fitzmaurice, A., … Lichtman, J. W. (2024). A petavoxel fragment of human cerebral cortex reconstructed at nanoscale resolution. *Science*, *384*(6696), eadk4858. https://doi.org/10.1126/science.adk4858

Song, S., Sjöström, P. J., Reigl, M., Nelson, S., & Chklovskii, D. B. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biology*, *3*(3), 0507–0519. https://doi.org/10.1371/journal.pbio.0030068

Sporns Olaf AND Tononi, G. A. N. D. K. R. (2005). The Human Connectome: A Structural Description of the Human Brain. *PLOS Comput. Biol.*, *1*(4), e42. https://doi.org/10.1371/journal.pcbi.0010042

Stalling, D., Westerhoff, M., & Hege, H.-C. (2005). amira: A Highly Interactive System for Visual Data Analysis. In *Visualization Handbook* (pp. 749–767). Elsevier. https://doi.org/10.1016/B978-012387582-2/50040-X

Südhof, T. C. (2018). Towards an Understanding of Synapse Formation. *Neuron*, *100*(2), 276–293. https://doi.org/10.1016/j.neuron.2018.09.040

Tejero-Cantero, A., Boelts, J., Deistler, M., Lueckmann, J.-M., Durkan, C., Gonçalves, P., Greenberg, D., & Macke, J. (2020). sbi: A toolkit for simulation-based inference. *Journal of Open Source Software*, *5*(52), 2505. https://doi.org/10.21105/joss.02505

Turner, N. L., Macrina, T., Bae, J. A., Yang, R., Wilson, A. M., Schneider-Mizell, C., Lee, K., Lu, R., Wu, J., Bodor, A. L., Bleckert, A. A., Brittain, D., Froudarakis, E., Dorkenwald, S., Collman, F., Kemnitz, N., Ih, D., Silversmith, W. M., Zung, J., … Seung, H. S. (2022). Reconstruction of neocortex: Organelles, compartments, cells, circuits, and activity. *Cell*, *185*(6), 1082-1100.e24. https://doi.org/https://doi.org/10.1016/j.cell.2022.01.023

Udvary, D., Harth, P., Macke, J. H., Hege, H.-C., de Kock, C. P. J., Sakmann, B., & Oberlaender, M. (2022). The impact of neuron morphology on cortical network architecture. *Cell Reports*, *39*(2). https://doi.org/10.1016/j.celrep.2022.110677

Valdes-Aleman, J., Fetter, R. D., Sales, E. C., Heckman, E. L., Venkatasubramanian, L., Doe, C. Q., Landgraf, M., Cardona, A., & Zlatic, M. (2021). Comparative Connectomics Reveals How Partner Identity, Location, and Activity Specify Synaptic Connectivity in Drosophila. *Neuron*, *109*(1), 105-122.e7. https://doi.org/10.1016/j.neuron.2020.10.004

Váša František and Mišić, B. (2022). Null models in network neuroscience. *Nature Reviews Neuroscience*. https://doi.org/10.1038/s41583-022-00601-9

Wei, D., Lee, K., Li, H., Lu, R., Bae, J. A., Liu, Z., Zhang, L., dos Santos, M., Lin, Z., Uram, T., Wang, X., Arganda-Carreras, I., Matejek, B., Kasthuri, N., Lichtman, J., & Pfister, H. (2021). AxonEM Dataset: 3D Axon Instance Segmentation of Brain Cortical Regions. In M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, & C. Essert (Eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021* (pp. 175–185). Springer International Publishing.

Yogev, S., & Shen, K. (2014). Cellular and molecular mechanisms of synaptic specificity. *Annual Review of Cell and Developmental Biology*, *30*, 417–437. https://doi.org/10.1146/annurev-cellbio-100913-012953

Supplementary Figures

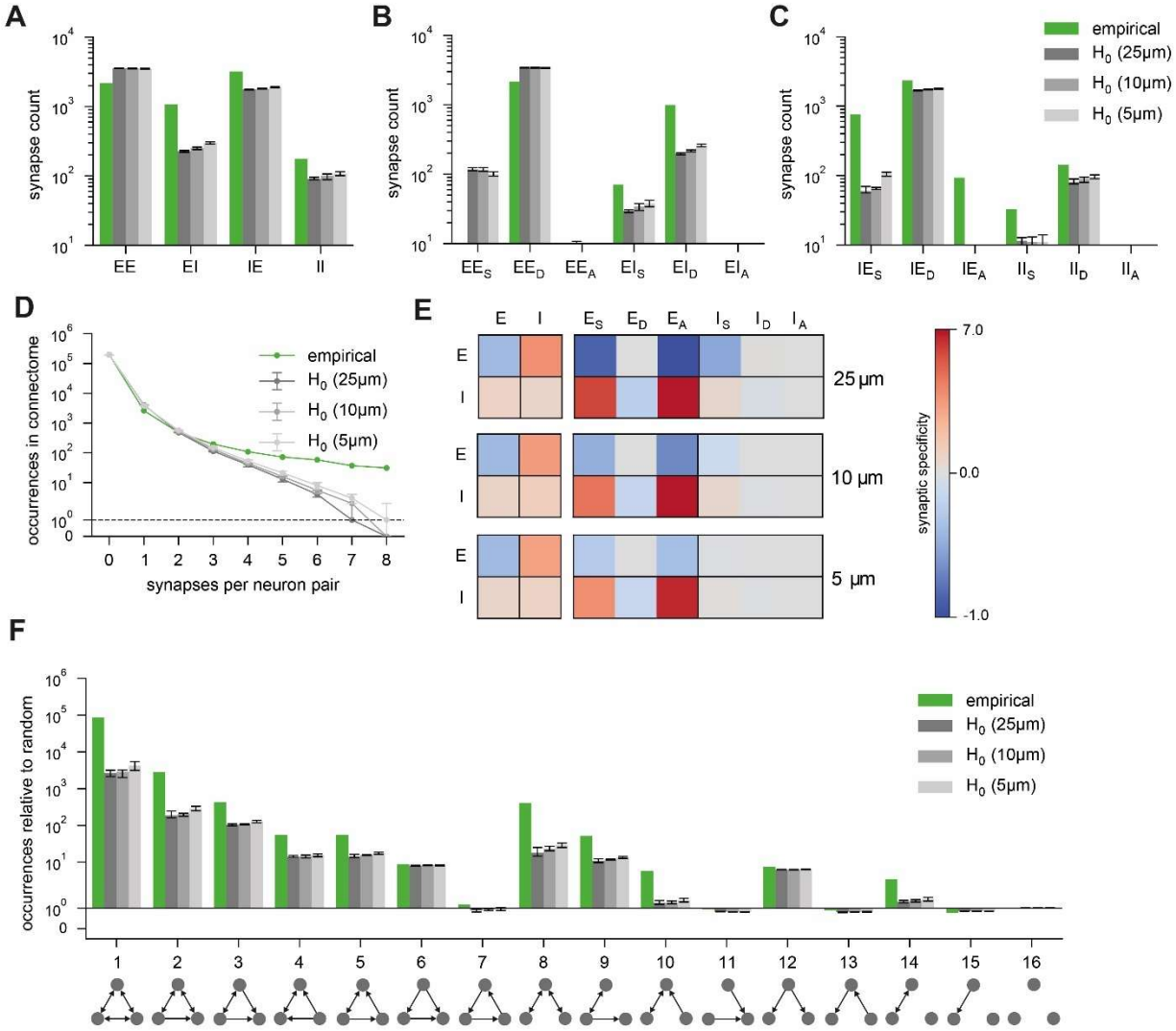

Figure S1: Effect of overlap volume size on connectivity statistics and synaptic specificity in the mouse visual cortex (Turner et al., 2022)**. (A)** Aggregate synapse counts between excitatory (E) and inhibitory (I) neuron populations. **(B)** Synapse counts from presynaptic excitatory neurons disaggregated by postsynaptic cell type and compartment (S: soma, D: dendrite, A: axon initial segment). **(C)** Synapse counts from inhibitory neurons disaggregated by postsynaptic cell type and compartment. **(D)** Number of synapses per neuron pair and multisynaptic connections (i.e., synapse clusters). **(E)** Resulting synaptic specificity parameters determined with different overlap volume sizes. **(F)** Triplet motif occurrences as predicted by the null model based on different overlap volume sizes.

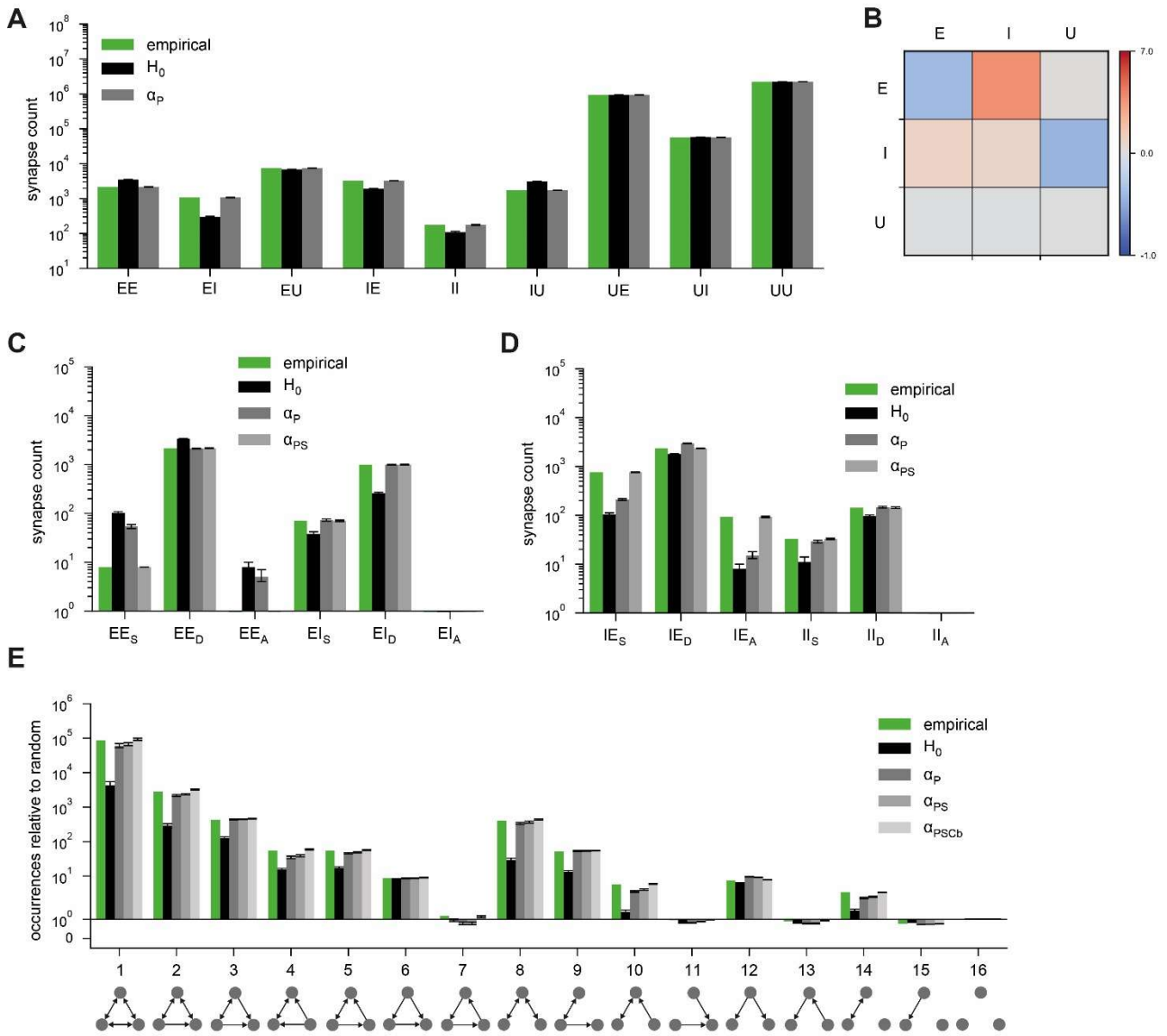

Figure S2: Extended connectivity statistics and model predictions in the mouse visual cortex (Turner et al., 2022)**. (A)** Aggregate synapse counts between neuron populations (E: excitatory, I: inhibitory, U: unknown). **(B)** Synaptic specificity parameters based on the α_P model. **(C)** Synapse counts from presynaptic excitatory neurons disaggregated by the subcellular compartment of the postsynaptic neuron (S: soma, D: dendrite, A: axon initial segment). **(D)** Synapse counts from presynaptic inhibitory neurons disaggregated by the postsynaptic compartment. **(E)** Occurrences of triplet motifs comparing empirically observed frequencies to the predictions of the $H_0, \alpha_P, \alpha_{PS}, \alpha_{PSCb}$ models.

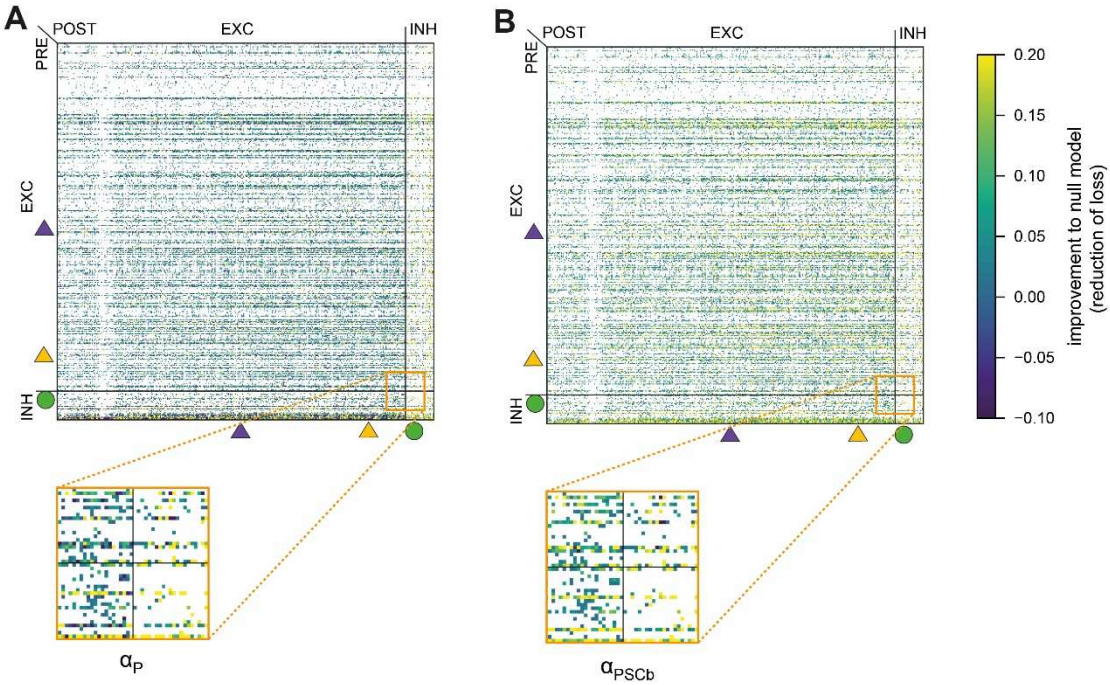

Figure S3: Model improvement compared to null model at cellular resolution. (A) Loss delta-matrix showing the reduction in loss (i.e., model improvement) of the α_P model in reference to the null model H_0. Matrix rows correspond to presynaptic neurons and matrix columns to postsynaptic neurons; brighter colors indicate a higher reduction in loss. **(B)** Loss-delta matrix for the α_{PSCb} model.

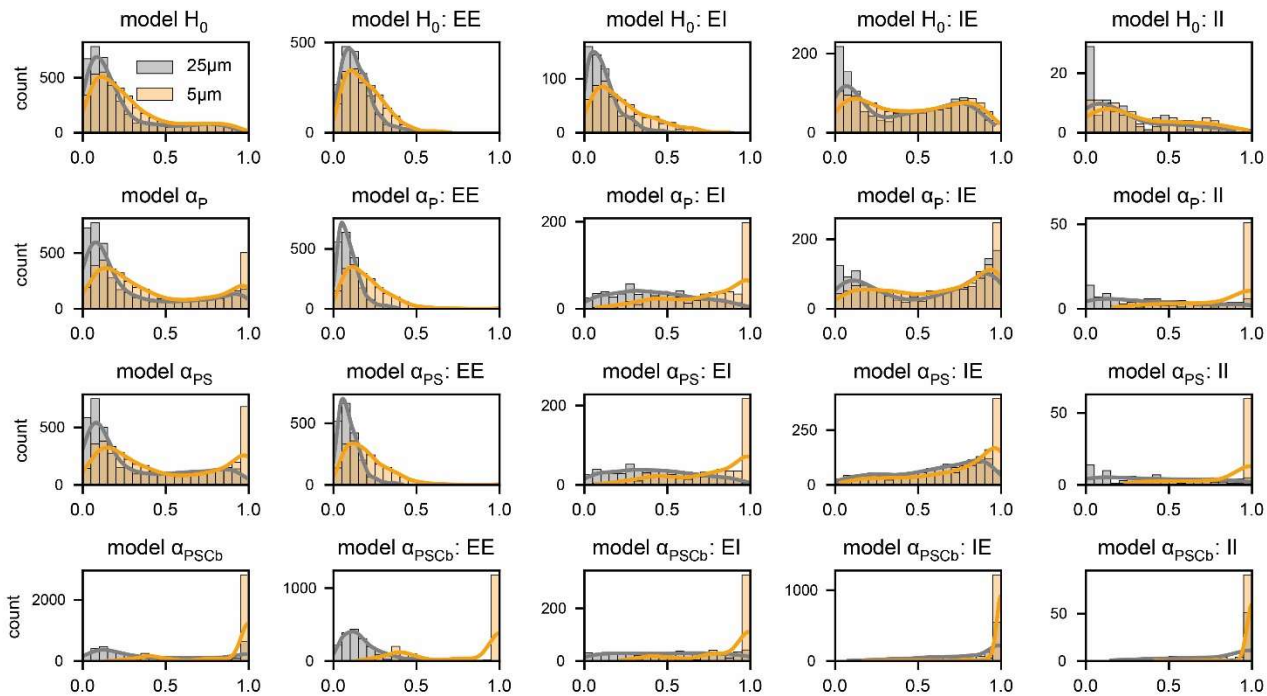

Figure S4: Distributions of connection prediction accurracies comparing $25\mu m$ (gray) and $5\mu m$ (orange) overlap volume sizes. The distributions show the fractions of correctly predicted neuron-to-neuron connections (i.e., a neuron pair connected in the ground truth connectome is also connected in the model) in n=1000 realizations of the connectome under the given model. The distributions in the leftmost columns represent all neuron pairs; in the four columns on the right, the neuron pairs are separated by cell type (EE: excitatory-excitatory, EI: excitatory-inhibitory, IE:

inhibitory-excitatory, II: inhibitory-inhibitory). Model complexity increases by row ($H_0, \alpha_P, \alpha_{PS}, \alpha_{PSCb}$) and, as a result, the masses of the distributions move to the right (i.e., towards higher accuracy).

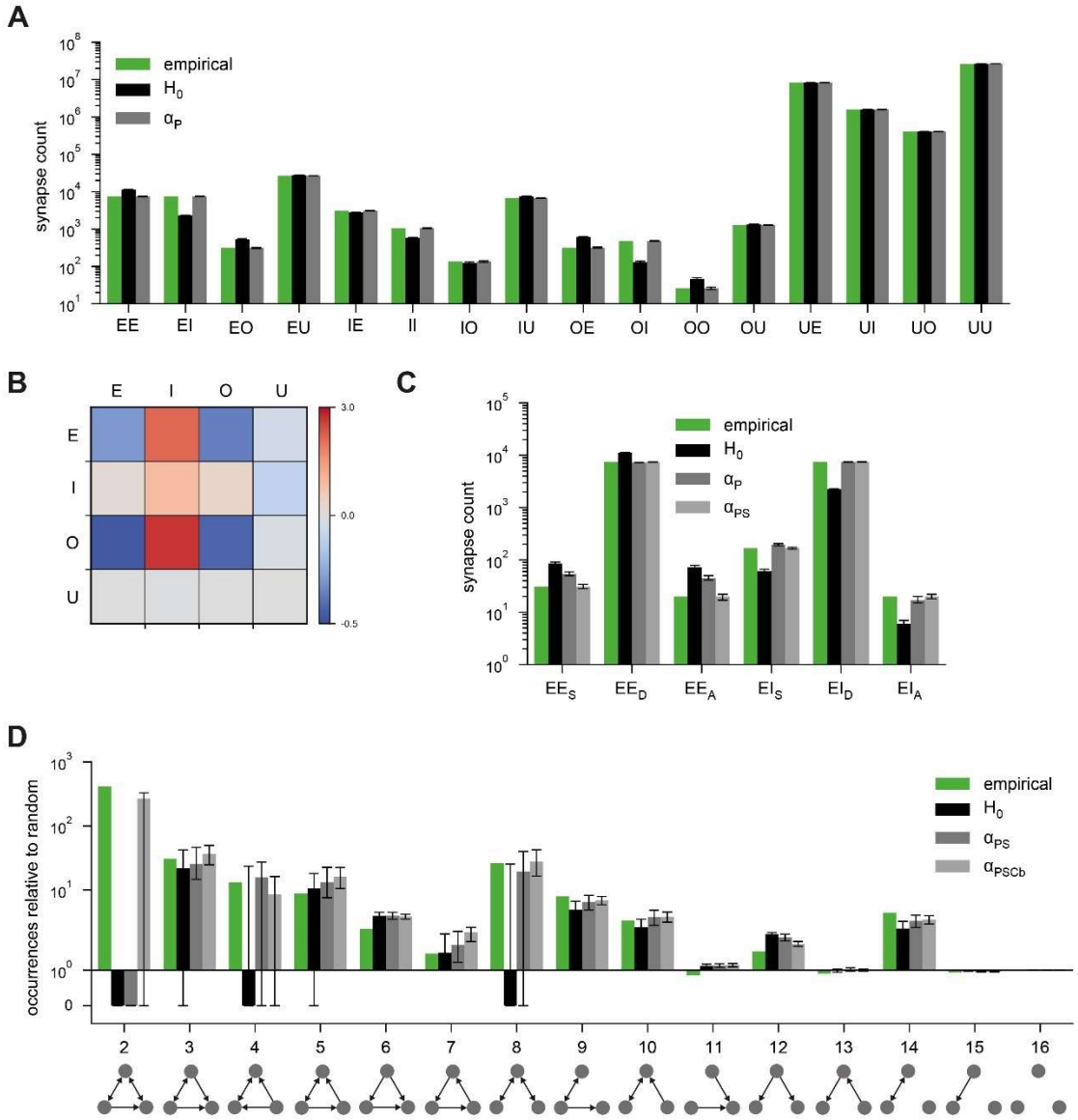

Figure S5: Extended connectivity statistics and model predictions in the human temporal cortex (Shapson-Coe et al., 2024)**. (A)** Aggregate synapse counts between neuron populations grouped by cell type (E: excitatory, I: inhibitory, O: other, U: unknown). **(B)** Synaptic specificity values based on the α_P model. **(C)** Synapse counts from presynaptic excitatory neurons disaggregated by postsynaptic cell type and subcellular compartment (S: soma, D: dendrite, A: axon initial segment). **(D)** Triplet motif occurrences among 200 neurons from layer 3 (see methods).

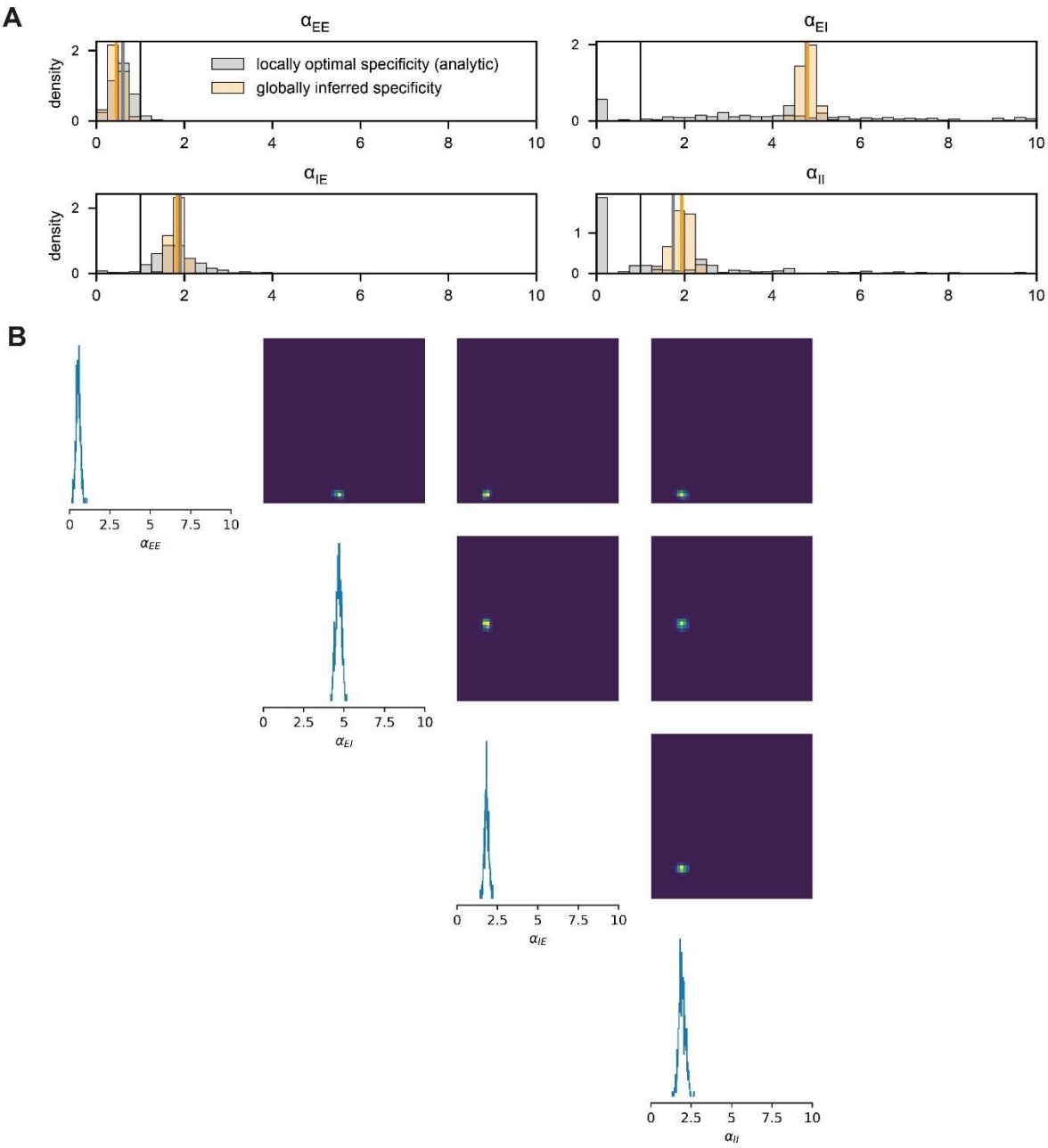

Figure S6: Inference of specificity parameters. (A) Comparison of locally optimal specificity values determined in overlap volumes (gray) and inferred values based on global aggregate statistics (orange). Gray and orange vertical bars show the mean values of the respective distributions of parameters. **(B)** Posterior distribution of inferred specificity parameters. Values smaller than one indicate a negative preference and values larger than one a positive preference (see methods).