Modeling and Simulation of Rat Non-Barrel Somatosensory Cortex.
Part I: Modeling Anatomy

Michael W. Reimann[1,8,†], Sirio Bolaños-Puchet[1,8], Jean-Denis Courcol[1,8], Daniela Egas Santander[1,8], Alexis Arnaudon[1], Benoît Coste[1], Thomas Delemontex[1], Adrien Devresse[1], Hugo Dictus[1], Alexander Dietz[1], András Ecker[1], Cyrille Favreau[1], Gianluca Ficarelli[1], Mike Gevaert[1], Juan B. Hernando[1], Joni Herttuainen[1], James B. Isbister[1], Lida Kanari[1], Daniel Keller[1], James King[1], Pramod Kumbhar[1], Samuel Lapere[1], Jānis Lazovskis[3], Huanxiang Lu[1], Nicolas Ninin[1], Fernando Pereira[1], Judit Planas[1], Christoph Pokorny[1], Juan Luis Riquelme[1], Armando Romani[1], Ying Shi[1], Jason P. Smith[4], Vishal Sood[1], Mohit Srivastava[5], Werner Van Geit[1], Liesbeth Vanherpe[1], Matthias Wolf[1], Ran Levi[6,9], Kathryn Hess[7,9], Felix Schürmann[1,9], Eilif B. Muller[1,9], Srikanth Ramaswamy[1,2,9,†], and Henry Markram[1,9,†]

[1]Blue Brain Project, École polytechnique fédérale de Lausanne (EPFL), Campus Biotech, Geneva, Switzerland
[2]Neural Circuits Laboratory, Newcastle University, UK
[3]Riga Business School, Riga Technical University, Riga, Latvia
[4]Nottingham Trent University, Nottingham, UK
[5]ELKH-University of Debrecen, Neuroscience Research Group, Hungary
[6]University of Aberdeen, Aberdeen, UK
[7]Laboratory for Topology and Neuroscience (UPHESS), Brain Mind Institute,

School of Life Sciences, École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland
[8]Co-lead authors
[9]Co-senior authors
[†]Corresponding authors

August 11, 2022

Abstract

The function of the neocortex is fundamentally determined by its repeating microcircuit motif, but also by its rich, hierarchical, interregional structure with a highly specific laminar architecture. The last decade has seen the emergence of extensive new data sets on anatomy and connectivity at the whole brain scale, providing promising new directions for studies of cortical function that take into account the inseparability of whole-brain and microcircuit architectures. Here, we present a data-driven computational model of the anatomy of non-barrel primary somatosensory cortex of juvenile rat, which integrates whole-brain scale data while providing cellular and subcellular specificity. This multiscale integration was achieved by building the morphologically detailed model of cortical circuitry embedded within a volumetric, digital brain atlas. The model consists of 4.2 million morphologically detailed neurons belonging to 60 different morphological types, placed in the non-barrel subregions of the Paxinos and Watson atlas. They are connected by 13.2 billion synapses determined by axo-dendritic overlap, comprising local connectivity and long-range connectivity defined by topographic mappings between subregions and laminar axonal projection profiles, both parameterized by whole brain data sets. Additionally, we incorporated core- and matrix-type thalamocortical projection systems, associated with sensory and higher-order extrinsic inputs, respectively. An analysis of the modeled synaptic connectivity revealed a highly nonrandom topology with substantial structural differences but also synergy between local and long-range connectivity. Long-range connections featured a more divergent structure with a comparatively small group of neurons serving as hubs to distribute excitation to far away locations. Taken together with analyses at different spatial granularities, these results support the notion that local and interregional connectivity exist on a spectrum of scales, rather than as separate and distinct networks, as is commonly assumed. Finally, we predicted how the emergence of primary sensory cortical maps is constrained by the anatomy of thalamo-cortical projections. A subvolume of the model comprising 211,712 neurons in the front limb, jaw, and dysgranular zone has been made freely and openly available to the community.

1 Introduction

Previously, we introduced a biologically detailed model of neocortical microcircuitry of the juvenile rat (*NMC*, Markram et al. (2015)), which has since been the basis of several studies on cortical structure (Gal et al., 2017), function (Nolte et al., 2019) and their relation (Reimann et al., 2017, 2022; Nolte et al., 2020; Gal et al., 2021). The NMC model comprises over 31,000 neurons of 55 morphological types across six layers, assembled in a volume of $0.29mm^3$. Synapses in this model were instantiated at a subset of axo-dendritic appositions, and models of stochastic synaptic transmission with facilitating, depressing or pseudo-linear short-term dynamics were assigned based on the combination of pre- and post-synaptic neuron types. The electrical behaviors of 207 morpho-electrical types were modeled to match numerous *in vitro* recordings (Ramaswamy et al., 2015), and an electrical type was assigned to each neuron according to a composition recipe derived from biological data.

The defining feature of the NMC model is its unusually high level of biological detail, which allowed its direct parameterization with available experimental data. As a consequence, many measurable physical and biological quantities have direct correspondences in the model. For example, every single synapse has not only physiological parameters associated with it, but also a physical location on the axon and dendrite of its pre- and post-synaptic neuron. This allows for powerful *in silico* experiments, such as comparisons to electron microscopy reconstructions (Markram et al., 2015).

However, this also means that the NMC model represents only what it seeks to model closely: an isolated cortical microcircuit. As such, without the vast majority of its synaptic and neuromodulatory inputs, its activity would not resemble a plausible cortical state (Mei et al., 2022; Lee and Dan, 2012). While we were able to simulate *in vivo*-like (asynchronous sparse) activity through artificial depolarization and mimicking the effects of extracellular calcium (Markram et al., 2015), we cannot expect more complex emergent phenomena that would require, for example, inter-region connectivity and feedback connections (Felleman and Van Essen, 1991; Shepherd and Yamawaki, 2021). Tackling questions related to such phenomena will therefore require a larger model that explicitly contains different regions and the connections between them.

For this reason, we set out to build a new model, based on the NMC model, but at a larger scale, covering the entirety of the non-barrel primary somatosensory cortex (*nbS1*) of juvenile rat. Beyond the increase in scale, this new model includes additional morphological and physiological data, as well as many other additional details and improvements. We decided to exclude the barrel cortex from this model, as its specific anatomical architecture and morphological specificity are not readily modeled with our toolchain.

In this manuscript, we will explain the improvements to the anatomical aspect of the model, including data sources and modeling techniques that we have developed. Upcoming manuscripts will focus on refinements of the physiology of the model and emergent dynamics, and an open web resource providing access to the experimental datasets used and the constituent computational models. First, instead of a simplified hexagonal volume that defined the NMC model, we embedded the nbS1 model within a volumetric atlas, thereby taking into account the actual geometry of the somatosensory regions in the rat neocortex (Bolaños-Puchet and Reimann, prep). Second, we increased the number of morphological neuron reconstructions and validated their expert classification using an objective method. Third, to model synaptic connectivity at the increased spatial scale, we adapted an algorithm for long-range connectivity to place additional synapses (Reimann et al., 2019). Fourth, to model bottom-up synaptic inputs, we placed both *core*- and *matrix*-type thalamo-cortical projections. Such core and matrix-type projections are observed across the cortex, and are exemplified respectively by the inputs into barrel cortex from the ventral posteromedial nucleus (VPM) and posterior medial nucleus (POm) of the thalamus (Harris et al., 2019; Shepherd and Yamawaki, 2021).

By characterizing and contrasting the local and long-range synaptic connectivity, we found distinctly non-random features across both scales. But interestingly, we also found qualitative differences between them: Local connectivity was more homogeneous, while long-range connectivity features a small number of neurons with a high out-degree distributing excitation to distant regions. High-dimensional directed simplices in the local connectivity – a specific type of directed neuron motif (Reimann et al., 2017) – formed a highly distributed skeleton, while long-range connectivity significantly reduced the graph distances between individual motifs. We used thalamo-cortical projections placed in the model to predict from anatomical grounds the assemblies of neurons that are likely to fire together, based on their common innervation by thalamo-cortical afferents. This way, we were able to predict how certain anatomical factors (laminar profiles of innervation, neuronal morphologies and their placement) provide an initial constraint on the emergence of neuronal assemblies in all layers in terms of their scale, and of their spatial separation.

2 Methods

The workflow for modeling the anatomy of juvenile rat nbS1 is based on the work described in Markram et al. (2015), with several additions and refinements. The individual steps and data sources have already been independently described and validated in separate publications (Table 1). Here, we aim to provide an overview of these methods, how we fulfilled their input data requirements, and how we combined them to assemble a single, coherent anatomical circuit model. This is then followed by an analysis of the emergent anatomical structure.

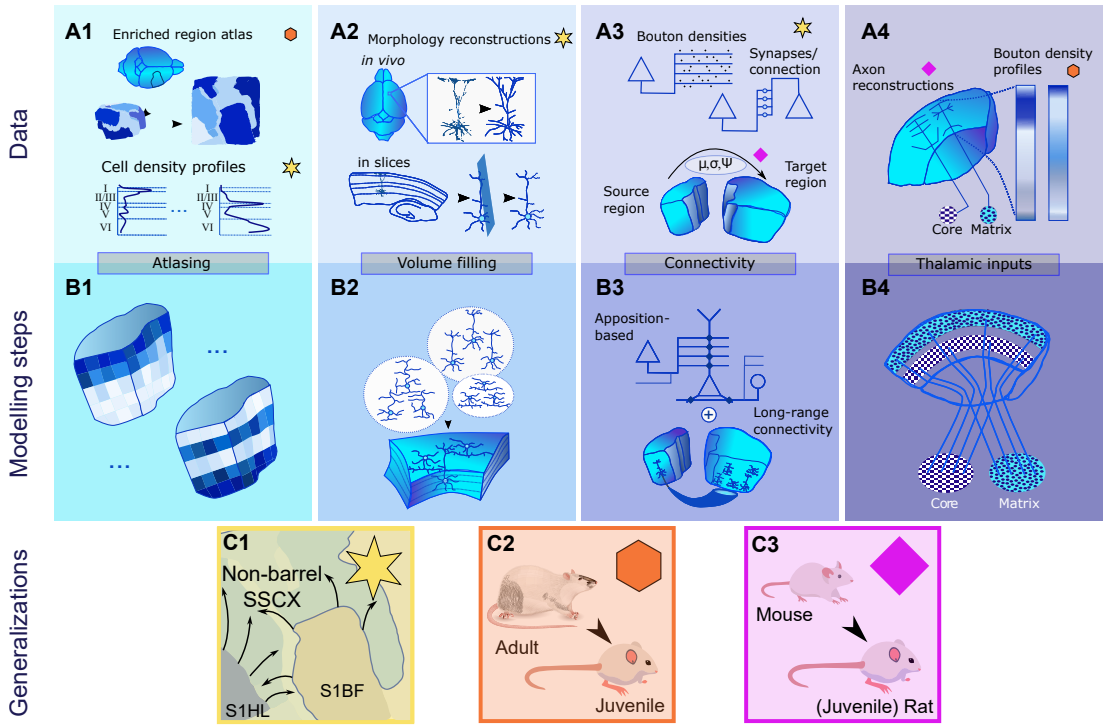

Figure 1: Overview of the model building workflow, showing input data and modeling steps performed at each stage, and data generalization strategies followed. A. Input data: (A1) Digital atlas of rat somatosensory cortex, enhanced with additional spatial information, and cell density layer profiles for all morphological types; (A2) Reconstructed neuronal morphologies from both *in vivo* (top) and slice (bottom) experiments, which undergo a repair process; (A3) Bouton densities and numbers of synapses per connection in local connections, and a parameterization of subregion-to-subregion connectivity in terms of strength, topographical structure and layer profiles; (A4) Reconstructed thalamo-cortical axons, and layer profiles of bouton densities of thalamo-cortical axons. B. Modeling steps: (B1) Generation of cell density atlases; (B2) Morphological classification of neurons, and placement in atlas space according to their densities and under spatial constraints; (B3) Definition of synaptic connectivity as the combination of: local connectivity, based on axonal appositions, and long-range connectivity, parameterized in the form of volumetric targets of individual neurons; (B4) Volumetric placement of thalamo-cortical synapses assigned to individual thalamic fibers. C. Data generalization strategies from: (C1) Hindlimb (S1HL) and barrel field (S1BF) subregions to the whole nbS1, indicated by a star in A; (C2) Adult to juvenile (P14) rat, indicated by a hexagon in A; (C3) Mouse to rat, indicated by a diamond in A.

An overview of the stages in the model building workflow is outlined in Figure 1. We began by preparing a digital atlas of juvenile rat somatosensory cortex, based on the classic work of Paxinos and Watson (2007) (Fig.1 A1). The atlas provides region annotations, i.e., each voxel is labeled by the somatosensory subregion it belongs to. It was further enhanced with spatial data on cortical depth and local orientation towards the cortical surface (Bolaños-Puchet and Reimann,

prep). Using data on layer profiles of cell densities, we generated additional atlas datasets with cell densities for each morphological type (Fig.1 B1).

Input data

Stage	Topic	Reference
Atlasing	Region annotation atlas	Paxinos and Watson (2007)
	Orientation, depth and flat map	Bolaños-Puchet and Reimann (prep)
	Cell density profiles	Keller et al. (2019)
Volume filling	Neuron reconstructions	Markram et al. (2015)
	in vivo neuron reconstructions	*New, original data
Synaptic connectivity	Bouton densities and numbers of synapses per connection	Reimann et al. (2015)
	Pathway strengths, synapse density profiles and topographical mapping	*Reimann et al. (2019)
Thalamic inputs	Bouton density profiles	Meyer et al. (2010)
	Projection axon lengths and widths	Economo et al. (2016)

Modeling methods

Stage	Topic	Reference
Atlasing	Cell density volume generation	Keller et al. (2019)
Volume filling	Neuron classification	Kanari et al. (2019)
	Neuron placement	*Markram et al. (2015)
Synaptic connectivity	Local connectivity	Reimann et al. (2015)
	Inter-region connectivity	Reimann et al. (2019)
Thalamic inputs	Input generation	*Markram et al. (2015)

Table 1: References to publications of input data and methods employed for individual modeling steps. An asterisk next to a reference indicates that substantial adaptations or refinements of the data or methods have been performed that will be explained in this manuscript. In the other cases, a basic summary will be provided.

Next, we filled the modeled volume of the atlas with neurons. Neuronal morphologies were reconstructed, either *in vivo* or in slices (Fig. 1A2), and repaired algorithmically (Anwar et al., 2009; Markram et al., 2015). These morphologies were classified into morphological types, based on expert knowledge and objectively confirmed by topological classification. Then, cell bodies for all morphological types were placed in atlas space according to their prescribed cell densities (Fig. 1B2). At each soma location, a reconstruction of the corresponding morphological type was chosen based on the size and shape of its dendritic and axonal trees. The selection ensured that manually identified features of these trees landed in the correct layers.

Next, the dendritic locations of chemical synapses between neurons were determined using the combined outputs of two connectome modeling algorithms (Fig. 1B3). For local connections with a range of up to approximately $500\mu m$, we used an algorithm that assigns synaptic locations from an anatomically constrained subset of axo-dendritic appositions (Reimann et al., 2015). For long-range connections between subregions, we adapted an algorithm originally developed for mouse whole-isocortex connectivity (Reimann et al., 2019).

Finally, the dendritic locations of synaptic inputs from thalamic sources were determined using experimental data on layer profiles of bouton densities of thalamo-cortical axons, and morphological reconstructions of these types of axons (Fig. 1A4). We modeled two types of thalamic inputs (Fig. 1B4), based on the inputs into barrel cortex from the ventral posteromedial nucleus (VPM) and from the posterior medial nucleus (POm) respectively (Harris et al., 2019; Shepherd and Yamawaki, 2021). While barrel cortex was not a part of the model, we used these projections as examples of a *core*-type projection, providing feed-forward sensory input (VPM) and a *matrix*-type projection, providing higher-order information (POm).

2.1 Data sources and strategies

The somatosensory cortex comprises about $72mm^3$ of the juvenile rat brain, allowing for substantial internal anatomical variability. We based our modeling methods around a high-resolution digital brain atlas, and algorithms that consider individual subregions, voxels or even individual neurons, thus giving us the ability to potentially capture such variability. Fully parameterizing these algorithms would require substantial amounts of biological data that was not always readily available. Because of this, we used data generalization strategies to compensate for missing information (Fig. 1C). For some modeling steps, we pooled data from nearby somatosensory regions, or did a generalization from the previously modeled hindlimb region (Markram et al., 2015) to the entire model (Fig. 1A, C1, stars). Additionally, we decided to use or adapt adult rat data for use in the juvenile rat model (Fig. 1A, C2, hexagons), or to adapt mouse data to the rat model (Fig. 1A, C3, diamonds) due to its higher availability.

In the following sections, we describe our strategies for fulfilling the data requirements of each stage and the generalizations we made. In the future, additional or more specific data can be included in each step.

2.1.1 Data on region geometry

First and foremost, the departure from a simplified hexagonal geometry (as in the NMC model Markram et al. (2015)), required a digital brain atlas defining the anatomy of the region. To that end, we took as starting point the Paxinos and Watson (2007) adult rat brain atlas, processed individual digitized slices, aligned them and labeled them to assemble a smooth three-dimensional volume with region annotations. The resulting atlas was then scaled down from the dimensions of adult rat to juvenile (P14) rat brain by reducing the size of individual voxels from $40\mu m$ to $38.7348\mu m$. The scaling factor was based on the ratio of S1HL thicknesses at those ages ($2082\mu m$ for juvenile vs. $2150\mu m$ for adult). Finally, supplementary atlas datasets were generated as described in Bolaños-Puchet and Reimann (prep). These datasets provide additional spatial information required to ensure biologically accurate placement and orientation of dendritic trees (see below). First, the normalized cortical depth (between 0 and 1) at each point; second, the local orientation as a vector pointing towards the cortical surface at each point; third, the total cortical thickness at each point, i.e., the length of the shortest path from the cortical surface to the bottom of layer 6 passing through that point (Fig. 1A1).

Additionally, we followed Bolaños-Puchet and Reimann (prep) to produce a *flat map* of somatosensory regions, that is, a coordinate transformation associating each voxel with a two-dimensional projection into the plane. This can be used to create a flat view of region annotations that is crucial for the description of the topographical mapping of long-range synaptic connectivity (see below). In short, to produce the flat map, a projection surface was defined by reconstructing a mesh from all points at a relative cortical depth of 0.5. Next, the local orientation field in the supplementary atlas was numerically integrated, yielding streamlines that were used to project each voxel center onto the projection surface. Finally, the projection surface was flattened with an authalic (area-preserving) algorithm. The main property of the resulting flat map is that in any flat view derived from it, each pixel represents a subvolume of cortex that spans all layers, akin to a cortical column.

2.1.2 Data on cell densities

Inhibitory cell densities were constrained following Keller et al. (2019). In brief, several datasets were combined to provide depth profiles of densities for successively more granular neuron classes (Fig. 2). The first dataset consists of neuronal soma density estimations, using antibody stains of neuronal nuclear protein (NeuN) and γ-aminobutyric acid (GABA) from rat neocortex (n = 6, Keller et al. (2019)). Cell counts provided mean densities and a measure of inter-individual variability (Markram et al., 2015). Cell counts were obtained from a single rat in this dataset, and their positions were annotated and divided into 100 equal-width bins extending from the top of layer 1 (L1) to the bottom of layer 6 (L6) (Keller et al., 2019). This provided depth profiles of both

total neuron (from NeuN) and inhibitory neuron densities (from GABA). A similar profile was also estimated from immunostaining with calbindin (CB), calretinin (CR), neuropeptide Y (NPY), parvalbumin (PV), somatostatin (SOM) and vasoactive intestinal peptide (VIP) (at least three slices from at least two rats) (Keller et al., 2019). All stains were corrected for shrinkage (Ghobril, 2015). A single-neuron reverse transcription polymerase chain reaction (RT-PCR) dataset (Toledo-Rodriguez et al., 2005) allowed mapping of biochemical markers to morphological types, by finding spatial distributions of morphological cell types that would reproduce the biochemical marker distribution (Keller et al., 2019). Whenever the classification could not be resolved down to morphological types (i.e., excitatory cells, neurogliaform cells, L1 cells, etc.), an estimation of the fractions of subtypes was used (Muralidhar et al., 2014). The depth profile of excitatory neuron densities was further subdivided into subtypes, maintaining the same layerwise proportions of these types as in previous work (Markram et al., 2015). The final output of this process was a dataset of neuron density as a function of cortical depth for each of the 60 morphological types (Table S1, see Sec. 2.2.2 for details on determining morphological types).

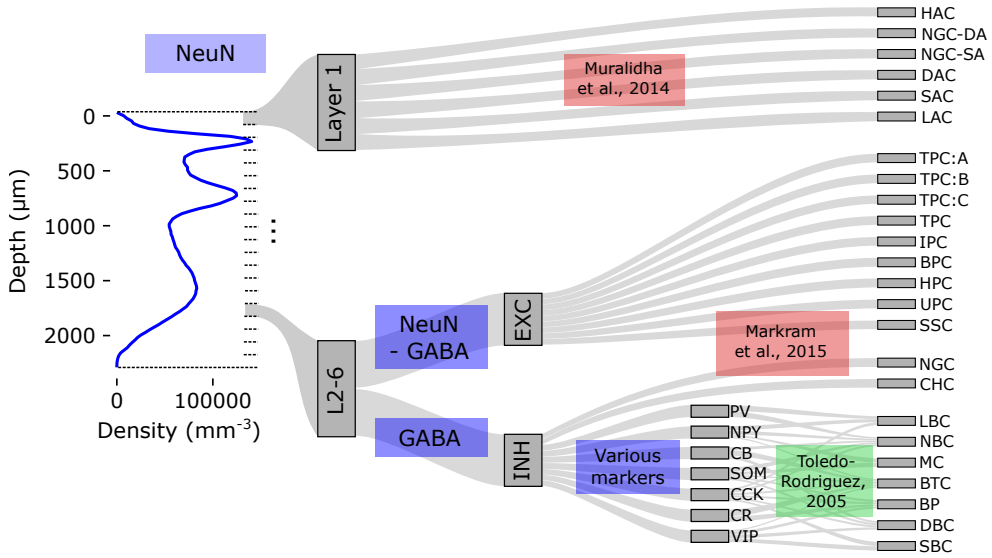

Figure 2: Derivation of neuron density depth profiles. Left: A vertical profile of neuron densities, calculated from antibody stains of neuronal nuclear protein (NeuN). From left to right: Neuron densities in each bin are split into individual morphological types through antibody stains for various markers (blue boxes) and an established mapping of markers to types (green), or by applying established neuronal compositions (red). Information in this flow diagram is illustrative, not quantitative.

2.1.3 Data on neuronal morphologies

The rat nbS1 model consists of approximately 4.2 million neurons. In order to populate this brain region with detailed morphologies, a large number of neuronal reconstructions is required. However, even the most accurate reconstructions from brain slices cannot represent the total axonal mass accurately, as branches outside the slice are lost. By constrast, *in vivo* reconstructions offer a more complete picture of the axons in terms of their total extents and branching structure. For this reason, neuronal reconstructions were collected with two techniques: neurons were either filled with biocytin in a brain slice and reconstructed (Markram et al., 2015), or were filled with a fluorescent dye *in vivo* and reconstructed (Buzás et al. (1998); Karube and Kisvárday (2011), see Supplementary Methods). In total 1,017 unique reconstructions were used, 896 of which were previously used in Markram et al. (2015), 63 were new *in vitro* reconstructions, and 58 were new *in vivo* reconstructions. A topological comparison (Kanari et al., 2018, 2019) between axons and dendrites of neurons in all layers (Fig. S1) revealed that *in vitro* reconstructions from slices could not capture detailed axonal properties beyond 1000μm (Fig. S1A), but could faithfully reproduce

dendritic arborization (Fig. S1B). Thus, the inclusion of *in vivo* reconstructions of axons is crucial for the accurate representation of the neuropil and will consequently provide a more precise approximation of the connectivity of the somatosensory cortex.

Neuronal morphologies vary significantly across different regions of the neocortex (prefrontal, parietal, occipital), especially with respect to the shape of their axons. However, within the parietal cortex, and more specifically within nbS1, neuronal morphologies share similar properties due to the similarity of layer composition and cortical thickness (Radnikow and Feldmeyer, 2018; Scala et al., 2019). As a result, cell types and neuronal reconstructions extracted from the hindlimb region are suitable to populate the entire nbS1. While there are the known differences in the barrel cortex (Narayanan et al., 2017), this region is not part of the model.

2.1.4 Data on synaptic connectivity

To determine the structure of synaptic connectivity between neurons, data on numbers of synapses and their locations were required. Previous data for mean bouton densities on axons of various neuron types, and number of synapses per connection (Reimann et al., 2015), were used and generalized to all nbS1 subregions. Similar to neuronal morphologies (see above), there is no evidence for anatomical differences in connectivity between nbS1 subregions, with the exception of barrel cortex (not included in the model). However, due to the larger spatial scale of the present model compared to Markram et al. (2015), additional data was required to further constrain synaptic connectivity at a global scale. To that end, we referred to data on relative strengths of synaptic connections from the Allen Mouse Brain Connectivity Atlas (Harris et al., 2019), and generalized it for use with a rat model.

We began by scaling the relative connection densities of Harris et al. (2019) to absolute densities in units of synapses per μm^3. That is, the entire voxel-to-voxel connection matrix of intracortical connectivity was scaled to match the average total density of synapses measured in electron microscopy (Schüz and Palm, 1989).

Next, we summed the densities over voxels belonging to pairs of nbS1 subregions, resulting in a 6 x 6 connection matrix of synapse densities in pathways between and within regions. We mapped this matrix to rat nbS1 by finding corresponding regions in the mouse and rat atlases (Fig. 3A). Here, we assumed that synapse densities in these pathways are comparable between mouse and rat, although the larger dimensions of the rat brain resulted in larger absolute synapse counts (Fig. 3C).

We used the same mapping to generalize the spatial structure of the targeting of connections between regions from mouse to rat. This refers to the question of which specific parts of a region are innervated by individual axons in a different region. Reimann et al. (2019) modeled this innervation as a topographical mapping between pairs of regions in a flat view of neocortex. We used a flat map (see Sec. 2.1.1) to create a flat view of rat nbS1 subregions and recreated a matching topographical mapping between them as follows (see Fig. 3A). First, three points were identified inside the flat view of a mouse region such that the area of the enclosed triangle is maximized. Color labels ("red", "green" and "blue") are arbitrarily assigned to the three points. Next, another three points defining a maximal area triangle were placed in the flat map of the corresponding rat region(s). The same color labels were then manually assigned to these points to best recreate their spatial context in the mouse triangle, e.g., if the point in SSp-n closest to the SSp-bfd border was labeled "red", then the point in S1ULp closest to the S1BF border will also be labeled "red". We then assume that any point in a rat region corresponds to the equally labeled point in the corresponding mouse region, with linear interpolation and extrapolation between points (Fig. 3B). This assumption transplants the existing prediction of the topographical mapping from mouse to rat, resulting in the mapping depicted in Fig. 3D.

Finally, we derived predictions for the relative distributions of synapse locations across cortical layers for connections between subregions. Reimann et al. (2019) provided predictions for all pairs of source and target regions in mouse, which we applied to the corresponding pairs of subregions

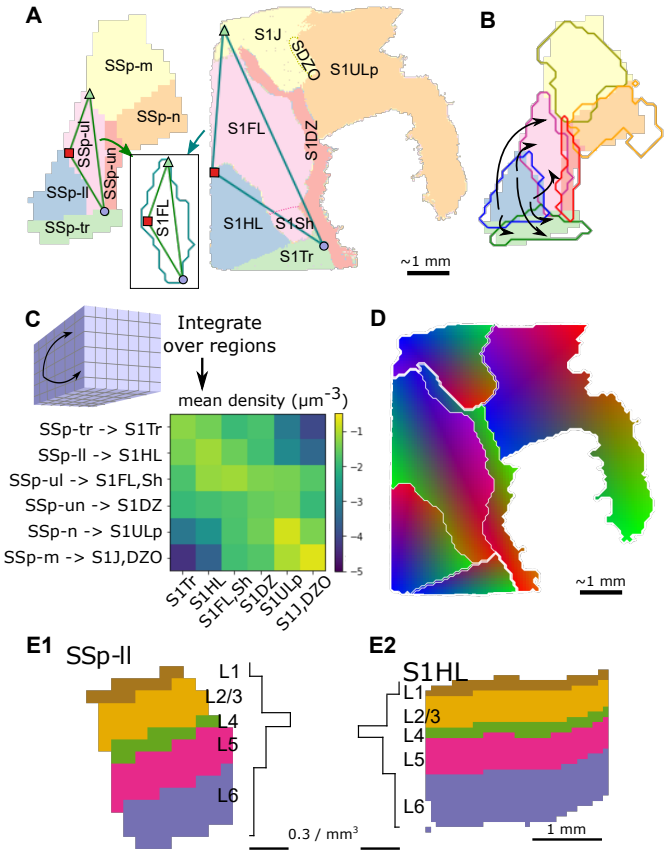

Figure 3: **Data sources for connectivity modeling.** A: Mapping from mouse to equivalent rat subregions. Left: flat view of mouse somatosensory subregions, Right: equivalent flat view for rat. Corresponding subregions are indicated with matching colors. Triangles are drawn on top of each subregion (shown for SSp-ul and S1FL) using manually annotated points that are assumed to correspond to each other. Inset: These triangles define affine transformations between mouse and rat subregions. B: Transformed rat subregions (colored outlines) are shown overlaid to their corresponding mouse subregions (colored areas). The topographical mapping of connections between subregions predicted in Reimann et al. (2019) (black arrows) can thus be generalized to rat connectivity. C: Mean connection densities between rat somatosensory subregions were derived from the Allen Mouse Brain Connectivity Atlas by summing over corresponding mouse somatosensory subregions. Labels on the left side of the connection density matrix describe the mapping applied from mouse to rat subregions. D: Generalized topographical mapping resulting from the process illustrated in B. Equally colored points are predicted to predominately connect to each other. E: Similarly, predicted layer profiles of connections between subregions (Reimann et al., 2019) are generalized from mouse (E1) to rat (E2).

of rat nbS1. That is, we assumed these layer profiles generalize to rat, albeit using the thicker cortical layers of rats (Fig 3E1 vs. E2).

2.1.5 Data on thalamo-cortical inputs

As previously described in Markram et al. (2015), we aimed to add synaptic connections from thalamic regions to the circuit model. While they will ultimately serve as one of the controllable inputs for the simulation of *in vivo*-like experiments, they can also be used to make predictions of innervation strengths in our anatomical model. Specifically, we modeled two such inputs: one "bottom-up" input with a *core*-type laminar profile, and one "top-down" input with a *matrix*-type laminar profile (Harris et al., 2019; Guo et al., 2020; Shepherd and Yamawaki, 2021). In order

to create biologically realistic thalamo-cortical projections with a meaningful distinction between core- and matrix-type inputs, we needed data on the strengths of these pathways and the laminar profiles of their synapses.

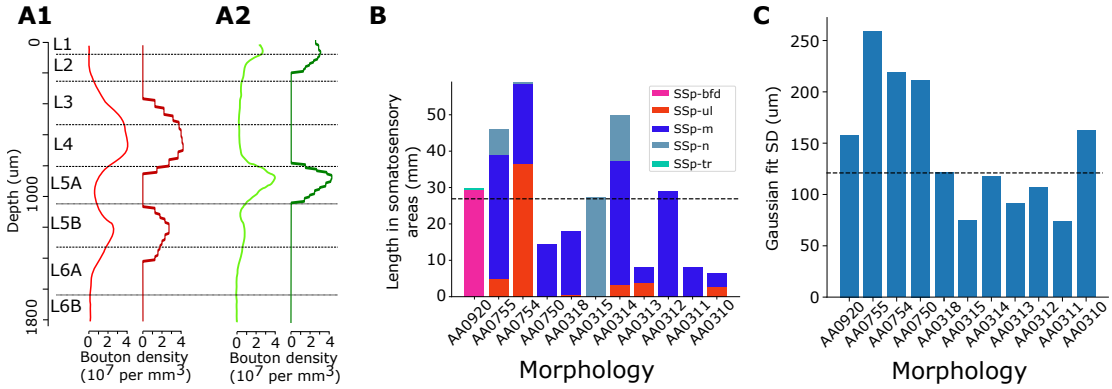

Figure 4: **Input data for the derivation of the locations of thalamic inputs.** A1: Layer profile of bouton densities of VPM axons in the rat barrel field. Light red: Digitized from Meyer et al. (2010). Dark red: Binned into 20 bins per peak with a lower cutoff of $1/mm^3$ applied. A2: Same for POm axons. B: Total length of 11 reconstructed axons of Janelia MouseLight neurons with somas in VPM in various somatosensory subregions. Dashed line: Median C: Standard deviation of a Gaussian fit of the reconstruction points of the same axons around their centroid in somatosensory areas. Dashed line: Median

As in Markram et al. (2015), we used data on the cortical innervation by thalamic sources from Meyer et al. (2010), which yielded information on the depth profiles of thalamo-cortical projections. We considered data for the VPM-S1BF pathway in as representative for core-type inputs, and generalized it to all nbS1 regions (Fig. 4A1, left). Similarly, we considered data for the POm-S1BF pathway as representative of matrix-type inputs (4A2, left), and generalized it to all nbS1 regions. Since the data is reported as absolute volumetric bouton density of thalamo-cortical axons, we were able to derive the total number of synapses to place by assuming one synapse per bouton and summing over the entire innervated volume. The depth profiles for both pathways featured two clearly separated peaks. We digitized the depth profiles and split them into 10 bins per peak. Furthermore, we applied a threshold of $0.01/\mu m^3$ below which values were set to 0 (Fig 4A).

Additionally, to constrain the innervation strength and targeting of individual thalamic axons, we used morphological reconstructions of thalamo-cortical projection neurons from the Janelia MouseLight project (mouselight.janelia.org; Economo et al. (2016)). This is a generalization of mouse data to a rat model, made necessary by the lack of a comparable resource for rat. To parameterize core-type projections, we calculated the total axonal length in somatosensory areas of $n = 11$ reconstructions with somata in VPM and axons reaching the somatosensory areas. We found lengths between 5 and 60 mm (Fig. 4 B) with a median of 27 mm that we combined with an assumed synapse density of $0.2/\mu m$ Reimann et al. (2015) to get an average number of 5400 synapses per projection fiber. To estimate of the lateral spread of the area innervated by individual axons (the vertical component is covered by the layer profiles), we considered the locations of reconstructed axon segments contained within the somatosensory regions. We then fit a Gaussian to the lateral distance of the segments from their center (Fig. 4C), resulting in a median value of $120\mu m$.

Unfortunately, the MouseLight database contained only a single neuron with soma in the POm region and axon reaching somatosensory areas (labeled as AA604), and visual inspection revealed that its axon mostly avoided these areas and to target more medial motor-related regions. As such, to parameterize the matrix-type projections, we instead calculated the lateral spread of the single axon in the motor areas ($300\mu m$ in MOp; $172\mu m$ in MOs; mean: $236\mu m$). Its total length in cortical regions was 28 mm.

2.2 Reconstruction workflow and algorithms

2.2.1 Voxelized neuronal composition

To prepare the modeled volume for cell placement, we first associated each voxel of the atlas with a cortical layer. We assumed layer boundaries to be at the same normalized depth at each point of the region. The depths, derived from Markram et al. (2015), are listed in Table S2. As a voxel atlas of normalized depths was provided as an input, layer identities could be readily looked up from that table. Similarly, neuron densities for each voxel of the atlas were produced from the vertical density profiles that served as inputs (see Sec. 2.1.2) by looking up the corresponding values using the normalized depth.

2.2.2 Neuron morphologies and their classification

All morphological neuron reconstructions (see Section 2.1.3) were curated and repaired to correct reconstruction errors and slicing artifacts, as described in Markram et al. (2015). They were then classified based on the following strategy. Both interneurons and pyramidal cells were first classified by expert reconstructors according to their observed shapes by inspection through the microscope. The expert classification of pyramidal cells, which is based on the shape of the apical dendrites, was then used as input for the training of the algorithms for the objective classification as presented in Kanari et al. (2019). The objective classification was performed based on the *topological morphology descriptor* (TMD) (Kanari et al., 2018), which encodes the branching structure of neuronal trees. The TMD of apical trees was extracted from all morphological reconstructions and was used to train a decision tree classifier for the objective classification of cells into distinct groups. See Fig. S2 for exemplary excitatory morphologies and their features. Conversely, interneurons have been classified by experts based on their axonal shapes and therefore the topological morphology descriptor of dendrites could not be applied for the objective classification of interneurons.

The expert-proposed scheme comprised 60 morphological types (m-types) (18 excitatory and 42 inhibitory). The m-types of pyramidal cells are distinguished first by layer and further by shape of their tuft, such as untufted (UPC) and tufted (TPC) cells; and finally into subclasses (A:large tufted, B:late bifurcating, C:small tufted). The results of the objective classification of pyramidal cells were then used to validate this classification scheme. A classification scheme was deemed valid if the TMD was significantly different between classes. This was the case for classes in all layers except for a distinction between two subgroups in layer 3 (L3_TPC:A vs. L3_TPC:B), which was consequently discarded. Instead we performed an unsupervised clustering of all tufted pyramidal cells in layer 3 based on their TMD and found a different split into two classes, best described as large tufted (L3_TPC:A) and small tufted (L3_TPC:C), which we then used in the model. The resulting list of m-types is found in Table S1.

In the remaining layers the results of the objective classification were used to validate the class assignments of individual pyramidal cells. We found the objective classification to match the expert classification closely (i.e., for 80–90% of the morphologies). Consequently, we considered the expert classification to be sufficiently accurate to build the model. To increase the morphological variability in the model we combined axon reconstructions with soma and dendrite reconstructions from other neurons of the same m-type (*mix & match*; Markram et al. (2015)).

2.2.3 Volume atlasing and neuron placement

Neurons were placed into the volume by first generating soma positions and annotating them with a morphological type according to the voxelized densities generated. Next, for each location we selected a reconstructed morphology from the annotated morphological type. As previously (Markram et al., 2015), we tried to select an appropriate reconstruction taking also the variability within a morphological type into account. For instance, the largest exemplars of layer 5 pyramidal cells cannot be placed close to the top of the layer, as otherwise their tuft would stick out of the top of layer 1. We selected morphologies by scoring them according to how well manually identified features, such as dendritic tufts or horizontal axonal branching, would land in the biologically

appropriate layers, when placed at any given location. The placement rules that were used were the same as in Markram et al. (2015) and are listed in Table S3.

Previously, this process was aided by the use of a simplified geometry where layer boundaries were formed by parallel planes. To execute the algorithm in a realistic brain volume, we used auxiliary voxel atlasses (see Sec. 2.1.1; (Bolanos-Puchet and Reimann, 2022)). The first atlas contained for each voxel the normalized depth of its center point, i.e., a value between 0 and 1 where 0 would indicate its placement at the top of layer 1 and 1 the bottom of layer 6. The second atlas contained for each voxel the total thickness of the cortex at that location. As layer boundaries in the model were always placed at fixed normalized depths (Table S2), we could calculate the absolute distance of the voxel to any layer boundary by subtracting their normalized depths and multiplying the result with the local cortical thickness. Based on this, we calculated the overlap of the dendritic and axonal features of a candidate morphology with the target layer interval (Fig. S3). Morphology selection was then performed as previously (Markram et al., 2015), that is, a morphology was selected randomly from the top 10% scorers for a given position.

2.2.4 Modeling the structure of synaptic connectivity

As mentioned above, the data we used to constrain connectivity comprised data on the structure of local connectivity, such as bouton densities and the mean number of synapses per connection, was previously collected in Reimann et al. (2015); Markram et al. (2015). Larger-scale data were generalized from mouse, including the total mean densities of synapses in connection between regions, their topographical mapping and layer profiles.

We used this data in a previously introduced algorithm (Reimann et al., 2015) that models the structure of synaptic connectivity as a subset of axo-dendritic and axo-axonal appositions, after reconstructed neuron morphologies have been placed in a modeled volume (Markram et al., 2015). This method has been demonstrated to result in nonrandom features matching biology (Reimann et al., 2015; Gal et al., 2017). It explicitly determines the number of synapses in any part of the model by the amount of axonal length present, using biological data on mean bouton densities. As such, it relies on the axons placed in the model being complete and reaching to the biologically accurate locations. While we try to ensure the completeness of axonal arborizations through the use of *in vivo* reconstructions and repair algorithms, this becomes more challenging, the more an axon stretches away from its soma, due to lack of dye penetration and potentially slicing artifacts. We also ensure that local axonal arborizations reach the biologically appropriate layers (see section 2.2.3 above). However, in addition to their local arborization, some cortical neurons feature longer-reaching branches that stretch to neighboring cortical locations with apparent specificity (Livingstone and Hubel (1983), Fig. 5A, red ellipses).

Taken together, this indicates that the apposition-based approach, while valid for local connectivity within a few hundreds of μm laterally, cannot be relied on for more long-ranged connectivity. We therefore decided to model connectivity as the union of synapses placed by two different algorithms: the apposition-based one, and a previously published algorithm (Reimann et al., 2019) that performs better for long-range connections. We will refer to the synapses and connections resulting from either algorithm as the *local connectome* and *long-range connectome* respectively.

Briefly, the algorithm of (Reimann et al., 2019) places long-range synapses according to three biological principles that all need to be separately parameterized: First, connection strength: ensuring that the total number of synapses in a region-to-region pathway matches biology. Second, layer profiles: ensuring that the relative number of synapses in different layers matches biology. Third, topographical mapping: Ensuring that the specific locations within a region targeted by long-range connections of neurons describe a biologically parameterized, topographical mapping.

To parameterize connection strengths, we used data provided by the Allen Institute (Harris et al., 2019), adapted from mouse to rat as described above (see Sec. 2.1.4), yielding expected densities of projection synapses between pairs of regions (Fig. 3C). First, we addressed the question of to what degree the apposition-based algorithm suffices to model the connectivity within a region.

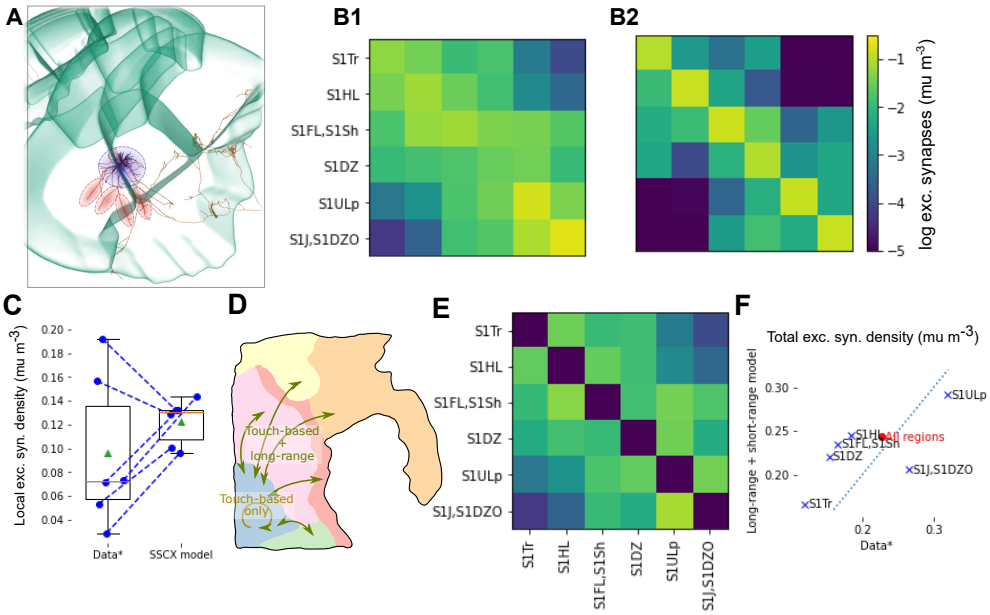

Figure 5: **Connectivity modeled as union of local and long-range connectivity.** A: An exemplary layer 5 PC axon reconstruction shown in the context of mouse somatosensory regions, rendered in the Janelia MouseLight neuron browser (mouselight.janelia.org, Gerfen et al. (2018)). The blue circle highlights local branches all around the soma. Red highlights depict more targeted collaterals into neighboring regions. B1: Predicted pathway strengths as indicated in Figure 3C3. B2: Pathway strength emerging from the application of the apposition-based connectivity algorithm described in Reimann et al. (2015). C: Values of the diagonal of (B1) compared to the diagonal of (B2). D: Schematic of the strategy for connectivity derivation: Within a region only apposition-based connectivity is used; across regions the union of apposition-based and long-range connectivity. E: Connection strength constraints for the long-range connectivity derived by subtracting B2 from B1 and setting elements < 0 to 0 (colors as in B). F: Resulting total density from both types of modeled synaptic connections in individual regions compared to the data in (B1).

To that end, we compared the mean excitatory synapse densities of apposition-based connectivity to the target values adapted from (Reimann et al. (2019), Fig. 5B1 vs. B2; C). We found that the overall average matches the data fairly well, however the variability across regions was lower in the model ($0.123 + -0.017 \mu m^{-}1$; mean +- std in the model vs. $0.097 + -0.06 \mu m^{-}1$). Based on these results, we decided to model connections within a region exclusively based on appositions. The apposition-based synapses created also a number of connections across region borders (Fig. 5B2). Consequently, we parameterized the strengths of additional long-range connections to be placed as the difference between the total strength from the data and the strengths resulting from the apposition-based algorithm, with connection strengths within a region set to zero (Fig. 5E). As a result synaptic connections between neighboring regions will be placed by both algorithms (Fig. 5D), with a split ranging from 20% apposition-based to 70% apposition-based. The lower spread of apposition-based synapse density within a region (see above) will also reduce the variability of combined synapse density from both algorithms (Fig. 5F). While this will halve the coefficient of variation of density across regions from 0.34 to 0.17, the overall mean density over all regions is largely preserved (0.23 data vs. 0.24 for the combined algorithms; Fig. 5F, red dot).

In summary, the anatomy of synaptic connections was constrained by data adapted from mouse as described above. The constraints were then fulfilled by synapses determined by two separate algorithms. First, an algorithm based on axo-dendritic (and axo-somatic) appositions placed about half of them, mostly between nearby pairs within the same region or neighboring regions. Then,

a second algorithm, parameterized according to more abstract biological principles, filled in the remaining density between more distal pairs.

2.2.5 Modeling the structure of thalamo-cortical innervation

We determined dendritic locations of thalamo-cortical input synapses as previously described in Markram et al. (2015), but adapted for the more complex geometry of this model. Briefly, binned depth profiles of densities of thalamo-cortical synapses were used as input (see Sec. 2.1.5, Fig. S6A1, A2). Next, we used the region atlas (see Sec. 2.1.1) to find the corresponding depth bins in the model. This allowed us to find all dendritic sections in the model, whose center point fell within a depth bin. We then performed a random selection of those sections and placed synapses at random locations on each selected section, until the prescribed number of synapses for a given depth bin was reached (Fig. S6A2). Sections were selected with probabilities proportional to their lengths and with replacement.

After all synapses for a thalamo-cortical projection had been placed, we mapped each of them to a presynaptic thalamic neuron. These neurons were not fully modeled, i.e., they were not assigned a soma position or morphology, the mapping simply allowed us to determine which synapses would be activated together. To parameterize the process, we used an estimate of the number of synapses formed by a single fiber and its horizontal spread (see Sec. 2.1.5). We divided the total number of synapses placed ($590 \cdot 10^6$ core-type; $380 \cdot 10^6$ matrix-type) by the number per fiber to estimate the number of innervating fibers (approximately 100'000 for the "core"-type projection; 73'000 "matrix"-type). These numbers were split between the eight subregions according to their relative volumes (Table S4), with the following steps being executed separately for each of them.

We then abstractly modeled thalamo-cortical afferent axons as lines entering their respective subregion at the bottom of layer 6 with a certain horizontal reach for the formation of synapses. This was done by first randomly distributing locations (one per fiber) across the boundary of layers 4 and 5 (Fig. S6 B) and then moving them $1500\mu m$ along the negative voxel orientation (towards layer 6; Fig. S6 B, black dots). The resulting positions and orientation vectors were used as the starting position of the fibers and their directions, respectively (Fig. S6 B, black arrows). The presynaptic fiber of a synapse was then determined by a stochastic process based on the horizontal reach of individual fibers around their respective location (Fig. S6 B, red areas). This was parameterized as a Gaussian with a $\sigma = 120\mu m$ for core-type, and $\sigma = 235\mu m$ for matrix-type projections (see Sec. 2.1.5). For each placed synapse, its distance to neighboring fibers was calculated and used as inputs into the Gaussian (Fig. S6D). This distance was calculated as the distance between the location of the synapse and the line defined by the fiber's starting point and direction. The probability that any fiber was chosen as innervating fiber of the synapse was then proportional to the values (Fig. S6E).

2.3 Validations

Each individual modeling step has been previously used and extensively validated in independent manuscripts (see Table 1). As such, there was no need to validate the scientific accuracy of the workflow steps themselves again, for example by comparing emerging results to biological data that was not used in modeling. However, this is the first time the methods have been used in conjunction with each other. Therefore, we have to ensure that the algorithms are still able to match the constraints given by their input data, when used in this new context.

2.3.1 Validation of neuronal anatomy

We began by validating that the cortical layers placed in the atlas formed a stack of continuous volumes. To that end, we calculated the fraction of voxels in each layer that were directly adjacent to another layer (Fig. 6A). As expected, adjacent voxels were found only between adjacent layers. Fractions of adjacent voxels decreased in lower layers as they are generally thicker and due to the curved geometry of the volume. Additionally, we tested that layers are spatially continuous by

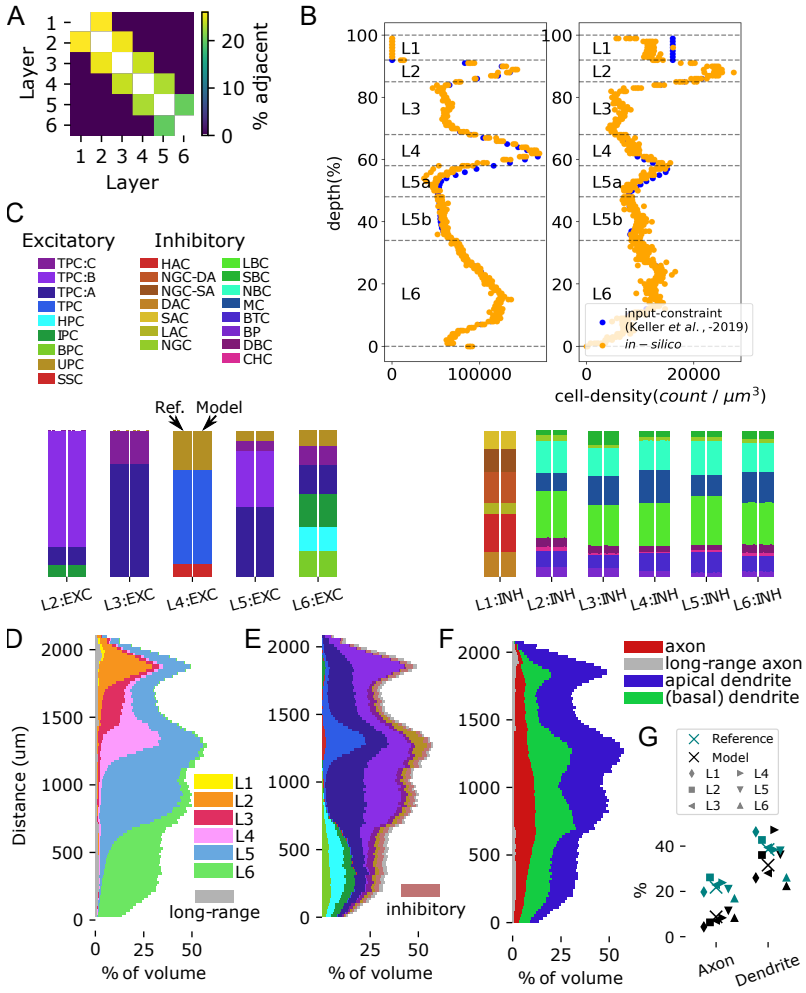

Figure 6: **Validation of neuronal anatomy.** A: Fraction of voxels in the volumetric atlas that are directly adjacent to another layer. B: Densities of excitatory (left) and inhibitory (right) neurons in the model (orange) are compared to the input data constraining the model (blue). C: M-type composition per layer: For each layer, stacked histograms of the relative fractions of each m-type, comparing the model (right bar for each layer) to the input data (left bar). For simplicity, the layer designation is stripped from each m-type. Left: For excitatory types; right: Inhibitory types. D-F: Stacked histograms of the fraction of space filled by neurites at various depths in the model. The y-axis indicates the distance in μm from the bottom of layer 6. D: For neurites of neurons in different layers indicated in different colors. Grey: estimated lower bound for the volume of axons supporting the long-range connectivity. E: For neurites of different m-types. Colors as in C, but inhibitory types are grouped together. F: For different types of neurites. G: Comparing fractions for axons and dendrites to the literature (Santuy et al., 2018). X-marks indicate overall means, other marks means in individual layers; teal: reference, black: model.

confirming that for each voxel there was at least one neighboring voxel in the same layer. This was the case for over 99.99% of the voxels with the handful of violating voxels limited to the periphery of the modeled volume (not shown).

To validate the neuronal composition of the model, we then compared the densities of excitatory and inhibitory neurons placed in the model against the input constraints (Fig. 6B). The vertical density profiles matched the input robustly. Slight differences between the two can be

explained by numerical issues, such as the need to round the number of neurons to place in a voxel to the nearest integer. Considering a more granular classification, we also compared the relative fractions of individual m-types in each layer (Fig. 6C). Once again, we found a perfect match within the expected numerical accuracy.

We validated the placement of reconstructed morphologies within each layer first by visual inspection of the placement of their neurites with respect to layer boundaries (Fig. S4 for examples). For a more quantitative validation, we calculated the fraction of the volume occupied by neurites in 100 depth bins of a cylindrical volume spanning all layers and with a radius of $100\mu m$ (Fig. 6D-F). In addition to the morphologies placed in the model, we also estimated a volume for the axons forming the long-range connectivity. It was estimated by assuming that an axonal segment with a length of $5.4\mu m$ and a diameter of $0.21\mu m$ supported each long-range synapse. The length was based on the inverse of the mean bouton density of excitatory axons in cortex (Reimann et al., 2015); the diameter was the mean diameter of axons in the model. As this excludes parts of the axon not forming boutons, this is merely a lower bound estimate.

We found a clearly layered structure also for the neurite volumes, with apical dendrites contributing substantially to the volume in layers above their somas. The volume was clearly dominated by dendrites, filling between 23% and 47% of the space, compared to 2% to 11% for axons (Fig. 6G). The range of values for dendrites matched literature closely (between 26% and 46%, Santuy et al. (2018)). Conversely, values for axons were lower than literature (between 17% and 24%); however, this is explained by the vast amount of external input axons from non-somatosensory and extracortical sources that are not part of the model.

2.3.2 Validation of synaptic connectivity

The part of the connectome derived with a apposition-based approach was constrained by anatomical data on bouton densities and mean numbers of synapses per connection for different morphological types (Reimann et al., 2015). As such, we validated that the results match these data (Fig. S5A, B). As previously found, the only mismatch lied in the emerging bouton densities of Chandelier Cells. These neurons form synapses only onto the axon initial segment of other neurons, which we model by disregarding appositions on the dendrites. For two Chandelier types, this resulted in an insufficient number of appositions to fulfill bouton density constraints (Fig. S5A, black arrow).

Arguably the most important constraint is the number or density of excitatory synapses in pathways between individual subregions, as it determines the overall excitability of the model and the velocity of the spread of activity. Here, we compared the total number of excitatory synapses, i.e., the union of the output of both algorithms to the data (Fig. S5C), finding a robust qualitative match. Due to the stochastic nature of the connectivity algorithms, an exact match cannot be expected.

Finally, the long-range connectivity algorithm was further constrained by predicted topographical mapping between regions (Fig. S5D) and synapse layer profiles (Sec. 2.1.4E), which we validated against the data.

2.3.3 Validation of thalamo-cortical innervation

One important input used was an estimate of the number of synapses formed in cortex by individual neurons in VPM (as a reference for core projections) and POm (as a reference for matrix projections). Unfortunately, for "matrix"-type projections this was based on only a single axon reconstruction. The input was used to estimate the total number of innervating fibers, and we note that the resulting ratio of 73'000 to 100'000 fibers between core and matrix closely matches the ratio of volumes of the two nuclei used ($1.25mm^3$ for POm to $1.64mm^3$ for VPM; ratio: 0.76).

In silico synapse density profiles of VPM and POm projections were validated against the *in vitro* layer profiles from Meyer et al. (2010) that were used as the input recipes (Fig. S7A, B). To account for non-uniform thicknesses across regions, the depth values of the density profiles were

normalized to the maximum thickness of each of the eight subregions. There is a decent match between the recipe and the actual layer profiles, but with a 15% overshoot at peak densities. This can be explained by the fact that the numbers of synapses to be placed were computed based on region bounding boxes which were larger than the actual volumes. Therefore, the actual densities are slightly higher when computing them on a voxel basis as it was done in this validation.

3 Results

3.1 Cellular anatomy of the model

The volumetric atlas used revealed the curved anatomy of the somatosensory regions (Fig. 7A). It also provided a spatial context for the model, with the non-modeled barrel regions being surrounded on three sides by modeled regions (specifically: S1DZ and S1ULp). The cortical layers (Fig. 7B) formed the expected vertical stack of continuous volumes. The placement of morphological reconstructions matched expectation, showing an appropriately layered structure with only small parts of neurites leaving the modeled volume (Fig. 7C).

We decided to investigate the effect of cortical curvature on the local microcircuit around a given point. To that end, we partitioned the modeled volume into columnar subvolumes with a radius of approximately $230\mu m$ (Fig. 7D, see Sec. 5.3.9). We found that the resulting subvolumes often had conical shapes, depending on the local curvature of the volume (Fig. 7D2). This is because a convex curvature of the cortical surface necessarily results in a contraction towards the lower layers and is independent of the algorithm used to partition the volume. We quantified the conicality of the subvolumes (Fig. 7E1), finding mostly narrowing towards layer 6, but around the S1HL and S1Sh subregions, also expansion. Additionally, we quantified the variability of cortical thickness by calculating the height of the subvolumes (Fig. 7E2). We found that local differences in curvature had significant effects on the neuronal composition (Fig. 7 F). Unsurprisingly it affected the laminar composition, but we also found a weaker, but significant effect on the ratio of excitatory to inhibitory neurons.

3.2 Synaptic connectivity of the model

The construction of a model with detailed neuron-to-neuron connectivity allowed us study the structure of predicted connectivity at large scales stretching up to 6 mm. While the connectivity in the model remains only a prediction, it does take into account known principles of the formation of synaptic connectivity (Reimann et al., 2015). Notably, it takes into account the effect of neuronal morphologies in shaping connectivity on the local scale, as synapses are placed at a subset of axo-dendritic appositions. As this is evaluated in a global coordinate system, connectivity will be affected by the geometry of the brain atlas used. For example, local variation in layer thicknesses, or local curvature may influence how axons and dendrites meet.

Further, the large spatial scale of the model allows us to predict the topological structure of connectivity at different scales. Experimentally, measurements of connectivity happen at large spatial scales, that is, between brain regions or voxels of hundreds of micrometers, using techniques such as DTI tractography or staining of populations of axons (Oh et al., 2014; Bota et al., 2015; Scannell et al., 1995; Scholtens et al., 2014). Connectivity between individual neurons is also measured, albeit only between a comparatively small number of neurons at a time (Perin et al., 2011; Song et al., 2005). Electron-microscopic reconstruction will increase this scale, but at this moment the resulting connectomes are still incomplete, as their spatial scales are not much larger than the dendritic trees of contained neurons. The connectivity between individual neurons is the foundation for connectivity at larger scales. But it remains unclear how the topological structure of connectivity at the various scales relate to each other. For the currently available experimental data, the *resolution* of the larger scale is on the same order of magnitude as the *total volume covered* of the smaller scale; there is a large middle ground between them that we aimed to provide a prediction for using our model.

Consequently, we set out to analyze the emerging connectivity under the following aspects: First, comparing and contrasting the topology of local and long-range connectivity; second analyzing how they complement each other; third, analyzing how the connectivity is structured at increasing spatial resolutions. In this manuscript, the highest resolution we consider is single neuron resolution. Consequently, the following analyses are based on networks where nodes cor-

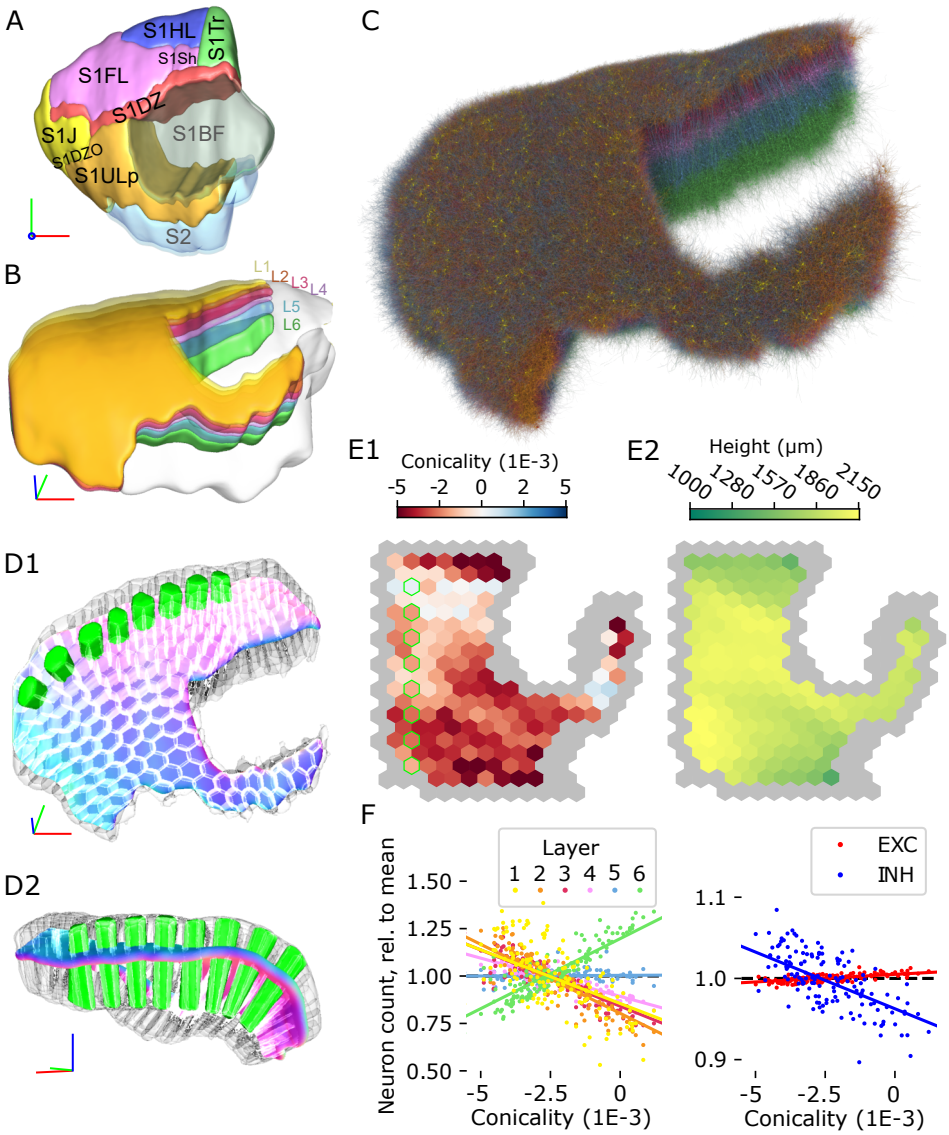

Figure 7: **Cellular anatomy of the model.** A: Modeled brain regions (solid colors) in the context of the non-modeled somatosensory regions (transparent). Derived from (Paxinos and Watson, 2007). B: Modeled cortical layers (solid colors) in the context of non-modeled regions (transparent). C: Rendering of the modeled volume with neuronal morphologies in place. 1% of density shown. D1, D2: Parcellation of the modeled volume into columns with 230μm radius. Exemplary slice of columns highlighted in green. E: Geometrical metrics of column subvolumes in the flat view. Peripheral columns masked out (grey); green outline: highlighted columns in D. E1: Conicality, defined as the slope of a linear fit of depth against column radius. Negative values indicate narrowing towards L6. E2: Column height. F: Conicality of columns against their neuronal composition, normalized against the overall composition of the model. Left: With respect to neurons in different layer. Right: Excitatory vs. inhibitory neurons. Colored lines indicate linear fits.

respond to neurons and edges to the presence of synaptic connections in the indicated direction regardless of their number or dendritic locations.

3.2.1 Global topology of local and long-range connectivity

We began by analyzing the global structure of neuron-to-neuron connectivity of the entire model, separately for apposition-based local connectivity and long-range connectivity. The topology of synaptic connectivity at neuronal resolution has previously been described in terms of the over-expression of directed simplices (Reimann et al., 2017). A *directed n-simplex* is a neuron motif of $n+1$ neurons that are all-to-all connected in feed-forward fashion, such that for any sub-motif consisting of a subset of neurons and all connections among them, there is exactly one source neuron and one sink neuron, where a *source* is a neuron sending connections to all others, and a *sink* is a neuron receiving connections from all others (Fig. S9A; Fig. 8A1, inset); we call n the *dimension* of the simplex. In particular, 0-simplices are the nodes of the network, and 1-simplices are directed edges. Simplex counts of a network of a given density provide a metric of network complexity and can be used to discriminate their underlying structure (Kahle, 2009; Curto et al., 2013; Giusti et al., 2015). Functionally, simplices, especially high-dimensional ones, have been demonstrated to shape the structure of spiking correlation between neurons (Reimann et al., 2017).

In line with previous results (Reimann et al., 2017), we found simplices up to dimension seven in the local connectivity (Fig. 8A1, green). That is, the increased size of the model compared to Markram et al. (2015) did not increase this measure. However, the addition of long-range connectivity drastically changed this picture. In long-range connectivity alone, simplices of dimension up to 15 were observed (Fig. 8A2, orange). This holds true even though local and long-range connectivity have roughly the same number of edges (2.1 billion local, 2.5 billion long-range), indicating that the higher simplex counts are not simply due to a larger number of connections. The combined network of local and long-range connections has simplices up to dimension 18 (Fig. 8A2, gray). Currently, there is no mechanism to determine the simplex counts of the combined network -or even the maximal dimension attained- from the simplex counts of the local and long-range circuits. Note in particular that the simplex counts of the combined circuit are not simply the sum of the local and long-range simplex counts, which are closer to the long-range counts except in dimensions two and three (Fig. 8A2, blue and inset on the right for non logarithmic scale).

We investigated to what degree simple anatomical properties, such as the density of synaptic connections or cortical layers, determine the connectivity structure. To do this, we compared the simplex counts of both the local and the long-range networks of our model to the mean and standard error of the mean in a range of comparable control models (Fig. 8A1,A2). We found that simplex counts have tight expected values for each control model and that these were consistently lower than those of both the local and long-range networks of our model, showing that the connectivity structure is not determined only by the parameters used to generate the controls. Specifically, the control models and the parameters on which they were based were the following:

- *Erdős–Rényi (ER)* used the overall density of connections;
- *stochastic block model (SBM)* used density in m-type-specific pathways;
- *configuration model (CM)* used sequences of in- and out-degrees;
- *distance block model (DBM)* used distance-dependence and neuron locations for individual m-type-specific pathways.

For more details on the models see Sec. 5.3.5. The CM control was the closest control for long-range connectivity, suggesting that the effect of degree is important, which was expected since we found that degree distributions of the long-range connectome were radically long tailed (Fig. S9A). Nonetheless, this control model still largely underestimates the complexity of the long-range network. Simultaneously, the DBM control was the closest control for local connectivity, showing that much of its structure is indeed determined by spatial distance under certain morphological restrictions.

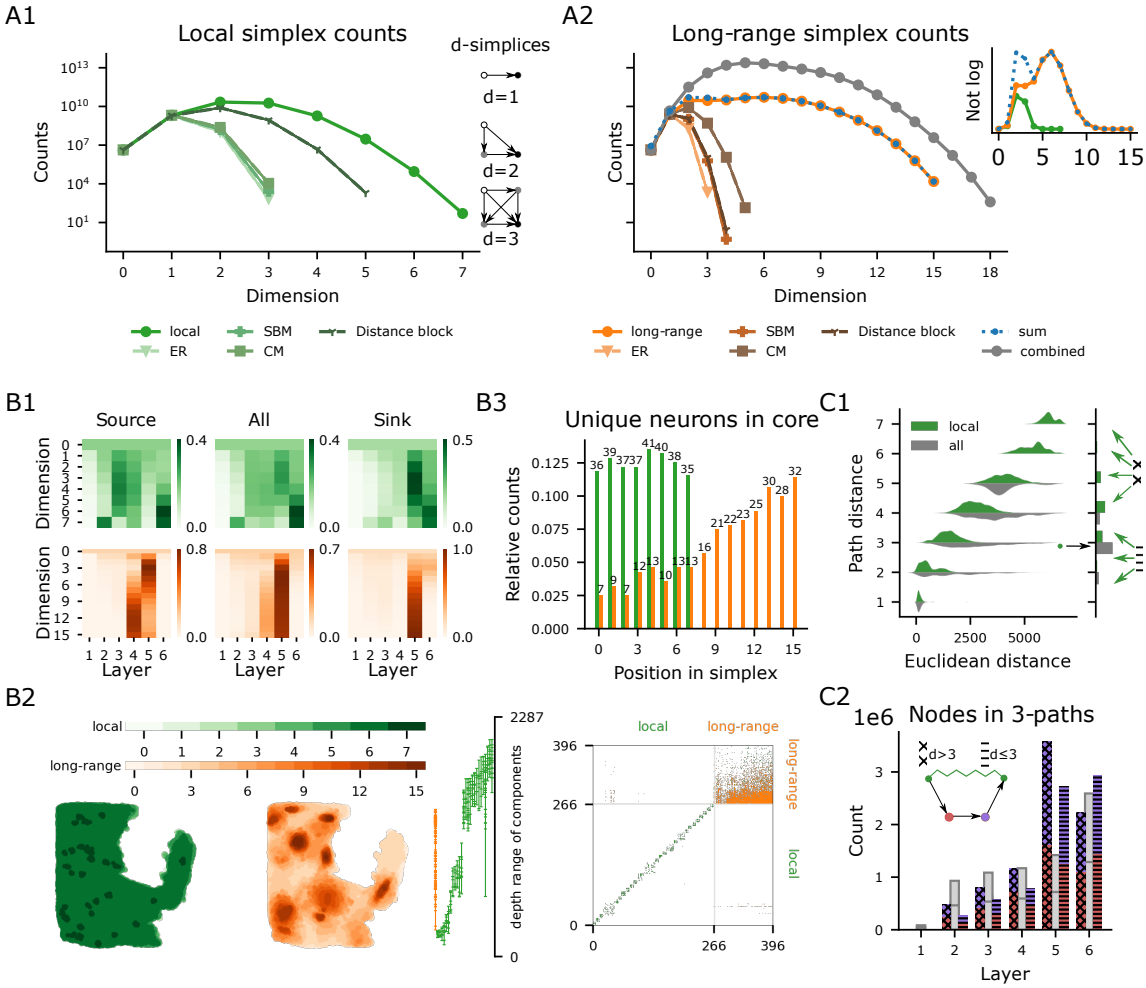

Figure 8: **Global connectivity structure local vs. long-range.** A1: Simplex counts of the local connectivity network and several types of random controls (see text). Examples of d-simplices for $1 \leq d \leq 3$. A2: In orange shades, simplex counts of the long-range connectivity network and several types of random controls. In gray, simplex counts of the combined network and in blue the sum of the local and the long-range simplex counts. Inset: local, long-range and the sum of simplex counts on a linear scale. B1: Normalized node participation per layer top/bottom row local/long-range connectivity network. Left to right: Participation as source, in any position, and as a sink of a simplex. B2: From left to right: spatial location of the cells in the simplicial n-cores in flat coordinates for all n. Depth coordinates for each connected component in the simplicial cores (cells participating in simplices of maximal dimension) each dot marks the depth of a neuron in that component. Adjacency matrix of the simplicial cores with respect to local (green dots) and long-range (orange dots) connections. B3: Number of unique cells in the simplicial cores present in each position of a simplex. C1: Right: Distribution of path distances between pairs of neurons in the local simplicial core; green: along only local edges; grey: along all edges. Left: Euclidean distances between pairs at a given path distance. Arrows: See C2. C2: Total Number of neurons in all paths of length 3 between neurons of the local core which are at distance 3 in the combined circuit (black arrows in C1) split by location. Red/purple: at positions 2 and 3 respectively; grey: expected from randomly assigned m-types while maintaining the global distribution. Cross-/stripe-patterned: For pairs at path distances greater/less-or-equal than 3 along local edges only (green arrows in C1).

Next, we investigated in which layers the neurons forming these high-dimensional structures resided in terms of *node participation*, which is the number of simplices to which a node belongs. Node participation is a measure of a node's centrality in the network (Sizemore et al., 2018). This value can be further split by dimension, giving rise to the notion of n-node participation. For $n = 1$ this metric is the total degree of a node, that is, the sum of its in- and out-degree, so one can think of node participation as a higher dimensional version of degree. Given that the neurons in a directed simplex can be easily sorted from its source to its sink, node participation can be further refined to participation as a source or participation as a sink. We considered n-node participation of neurons in these three different ways, averaged per layer and normalized per dimension (Fig. 8B1).

We found that in local connectivity, most simplices of dimensions 2, 3, and 4 have their source in layer 3 and their sink in layer 5. Yet, for dimensions 5 and above, we found a shift towards layer 6, with both sources and sinks found mostly in that layer. In long-range connectivity the structures are much more concentrated on layer 5, which contains both source and sink. Only for simplices of dimension greater than 8 does layer 4 provide more sources. Taken together, this indicates a robust local flow of information from superficial to deeper layers and within deep layers, with layer 5 forming a backbone of structurally strong long-range connectivity. Neurons in layer 6 form numerous simplices among themselves with no apparent output outside of layer 6, but they are known to be the source of many cortico-thalamic connections (Shepherd and Yamawaki, 2021) that are not part of this model.

3.2.2 Degree distributions and rich-club analysis

A way to understand a network's structure is to study its assortativity, that is, the preference of certain nodes to attach to each other when they have similar properties. The most commonly considered property is node degree, which has an unusual distribution in the model, especially for the long-range network. For instance, long-range connection probability per pathway (Fig. S9B) is characterized by a small number of m-types forming most efferent connections, leading to bimodal distributions of out- and total degrees in the associated DBM and SBM control models (Fig. S9A). However, the actual long-range network has a unimodal, long-tailed degree distribution, similar to biological neuronal networks (Giacopelli et al., 2021). Nonetheless, the connection probabilities still generate part of the structure of the network itself since it imposes the existence of many nodes that are sinks, sources and isolated nodes (see Fig. S9C).

A way to study a network's assortativity by degree is through its *rich-club* curve (Zhou and Mondragon, 2004; van den Heuvel and Sporns, 2011), where the k-rich-club coefficient is the density of the network induced by the nodes with total degree greater than k. The rich-club curve measures whether high degree nodes are more likely to connect together, forming some notion of a "rich-club". This can be generalized by replacing degree with any network metric on each node (Opsahl et al., 2008), such as node participation (see Sec. 5.3.7 for details).

We calculated the rich-club curves for the local and long-range networks and contrasted them with the curves for the corresponding CM controls (Fig. S10B1); dividing one by the other yielded normalized rich-club curves (Fig. S10B2). We observe that the local network has a rich-club effect, since its normalized coefficients are consistently above one. This is expected in a spatially finite model with distance-dependent connectivity, as neurons in the center will have a larger degree than the periphery and also connect preferentially to each other. We confirm this by studying the rich-club curves for the distance dependent control models. Even though the non-normalized rich-club coefficients are smaller for the distance controls, this is simply because their degrees are smaller; the normalized curves show similar behavior, that is, a mostly increasing trend which is consistently above one. On the other hand, the long-range network does not exhibit a rich-club effect, see Figure S10B2 bottom, showing that high degree nodes are not central to the network.

A different picture emerges when looking at simplicial rich-club curves, where for every node we consider its higher dimensional node participation (see Fig. S10B3), instead of simply degree.

We can observe that the rich-club coefficients decrease by degree in the local network. On the other hand, for the long-range network the rich-club coefficients increase with dimension, reaching densities greater than 50 in all dimensions greater than four. We speculate that the long-range network has a high normalized simplicial rich-club coefficient and the local network does not. Unfortunately, this can not be currently verified since generating appropriate random controls for these curves is an open problem currently investigated in the field of random topology (Unger and Krebs, 2022).

These results show that the local network is highly distributed, since its high rich-club coefficients are present only in lower dimensions and are consistent with the behavior of random networks that are distance dependent. On the other hand, the long-range network is highly concentrated, though its central nodes are not given by degree but by node participation in high dimensions. We see more of evidence for this in the next section.

3.2.3 Analyzing the core of the connectivity networks

More complex analyses are hampered by the large sizes of the connectomes. For instance, it is reasonable to hypothesize that the long-range network is small-world, but computing small-world coefficients for a network of this size is infeasible (see Sec. 5.3.8). Therefore, we further explored the difference between the local and long-range networks and the interaction between them by focusing on central sub-networks or *cores*. One way to select such cores is using the *n-core* of a network, which is a sub-network determined by degree. However, from the analysis above, we know that node degree alone does not determine the complex structures of these networks. Thus, instead we consider the *simplicial cores* of the networks, a natural variant of the degree notion of core.

The *n-simplicial core* of a network is the sub-network on the nodes of the network that participate in simplices of dimension n or higher, and the *simplicial core* is the sub-network on the nodes of the network that participate in simplices of maximal dimension, forming a sort of backbone of the network. We will refer to the (n-) simplicial core of the local/long-range network as the *(n-)local/long-range core*, and we will refer to the union of both as *the core*. The general trend observed is that the local network is distributed, while the long-range network is highly localized – simultaneously in terms of connectivity, location of the neurons in space, and layer profiles – which is consistent with the structure of the probability of connectivity per pathway (Fig. S9C) as well as the rich-club analysis.

To see this, we first considered the location of the neurons forming the cores in the flat view, their cortical depth and the connectivity between these (Fig. 8B2). The local core consists of 267 neurons that are fairly evenly distributed throughout the flat projection, forming 26 individual clusters that are disconnected from each other. These clusters tend to be concentrated in higher or lower depths (Fig. 8B2, center). Neurons in n-cores for any n less than seven are also spread out evenly to cover essentially the entire modeled space. On the other hand, the long-range core consists of 129 neurons that are narrowly localized in the flat projection, tend to be concentrated in middle depths, and form a single connected component. Even the neurons in n-cores for smaller n were concentrated around specific locations and covered the entire model only for $n \leq 3$. Curiously, even though the numbers of neurons in the local and long-range cores are in the same order of magnitude, the dimensions and the numbers of simplices they span are radically different, quantifying in a smaller sub-network the difference in complexity of their corresponding structures.

The more distributed nature of the local network is also reflected in the fact that the layer distribution of the neurons in the local core does not differ greatly from the layer distribution of the overall population of neurons. On the other hand, neurons in the long-range core are highly localized in layer five (Fig. 8B1, center column, bottom row). Moreover, there is little overlap between the sets of neurons in the local core forming maximal simplices, while in the long-range core many neurons participate in several simplices (Fig. 8B3). Especially for the source location, we found that just seven neurons were the sources of all of the $15,108$ highest-dimensional long-range simplices. This trend weakens with proximity to the sink, indicating that the long-range

core forms a network mostly characterized by divergence.

Finally, we studied the interaction between the local and long-range cores through their embeddings within the combined network, that is, the network with both local and long-range connections. Even though the sub-network on the nodes in the local core has 26 connected components (Fig. 8B2), there is a directed path in the full local network between any pair of nodes in the local core, through local edges that may pass through nodes outside the core. We computed the lengths of the paths in the full local network for all pairs of nodes of the local core (Fig. 8C1, green), obtaining path distances between between one and seven, with a median of four. The path distances depend strongly on the euclidean distance between the neurons (Fig. 8C1, green), consistent with the local core being distributed throughout space and local connectivity being strongly distance dependent. When the long-range connections are added, the maximum path distance drops to five, with a negligible number of pairs at the maximum path distance. The median path distance drops to three, and the dependence of the path distance on the euclidean distance of the pairs nearly disappears (Fig. 8C1, grey).

We studied in more detail pairs of nodes at the most common path distance of three within the combined circuit, finding all paths of length three between them. We labeled the neurons along the paths from position 1 (start-node) to position 4 (end-node). We found that layer 5 (the layer with the highest outgoing edge probability, see Fig. S9B) is over-represented in nodes in positions 2 and 3 compared to a random assignment of m-types (Fig. 8C2). Moreover, the over-representation depends on the path distance between the start-node and end-node within the local network: when the value is larger than three (bars hatched by crosses), the effect is stronger than for pairs at a distance three or less (bars hatched by horizontal bars). This demonstrates that neurons in layer 5 act as "highway hubs" providing shortcuts between neurons that are far away from each other in the local circuit.

3.2.4 Connectivity at different scales

Connectivity, both functional and anatomical, is often described at lower resolutions than the neuron-to-neuron level (Oh et al., 2014; Bota et al., 2015; Scannell et al., 1995; Scholtens et al., 2014). However, doing so requires an often arbitrary choice of parcellation, limiting the strength of conclusions that can be drawn (de Reus and van den Heuvel, 2013). We used our model to predict how connectivities at different scales are related to each other, by generating and analyzing matrices of directed connectivity at reduced spatial resolution. We partitioned the neurons based on their flattened spatial locations into groups that we will refer to as *nodes*, then counted the numbers of synaptic contacts between pairs of nodes. By varying the number of neurons per node, we can analyze and compare connectivity at different resolutions. Since higher resolutions result in larger connectivity matrices, making comparisons difficult, we chose to increasingly limit the spatial scale for higher resolutions, considering only groups of neurons within a given spatial neighborhood. We developed a way to systematically change the scale of connectivity matrices along with their resolution, based on a partitioning of the neurons using a KD-Tree, and resulting in a number of 512×512 connectivity matrices at each resolution (Maneewongvatana and Mount (1999), Fig. 9A; for details, see Sec. 5.3.10). At the highest resolution, this corresponded to connectivity of individual neurons in a small number of minicolumns (Mountcastle (1998); Markram et al. (2015), Fig. 9B); at the lowest, it corresponds to connectivity between populations slightly smaller than a cortical column as defined by Markram et al. (2015).

We first investigated how synaptic connections were distributed across various spatial scales. As neuron pairs at distances exceeding the scale were excluded, the number of pairs considered was drastically reduced along with the scale (Fig. S8A). Specifically, the number of pairs was halved at each step (Fig. S8B, red). By contrast, the number of synaptic connections as a fraction of the total number of connections in the full model, fell much more slowly for the local connectivity (Fig. S8B1, blue), reflecting that connections are more prevalent at smaller distances. This led to an increase in the relative density of connections by over two orders of magnitude at lower scales (Fig. 9C, red). For long-range connectivity on the other hand, the fraction of connections

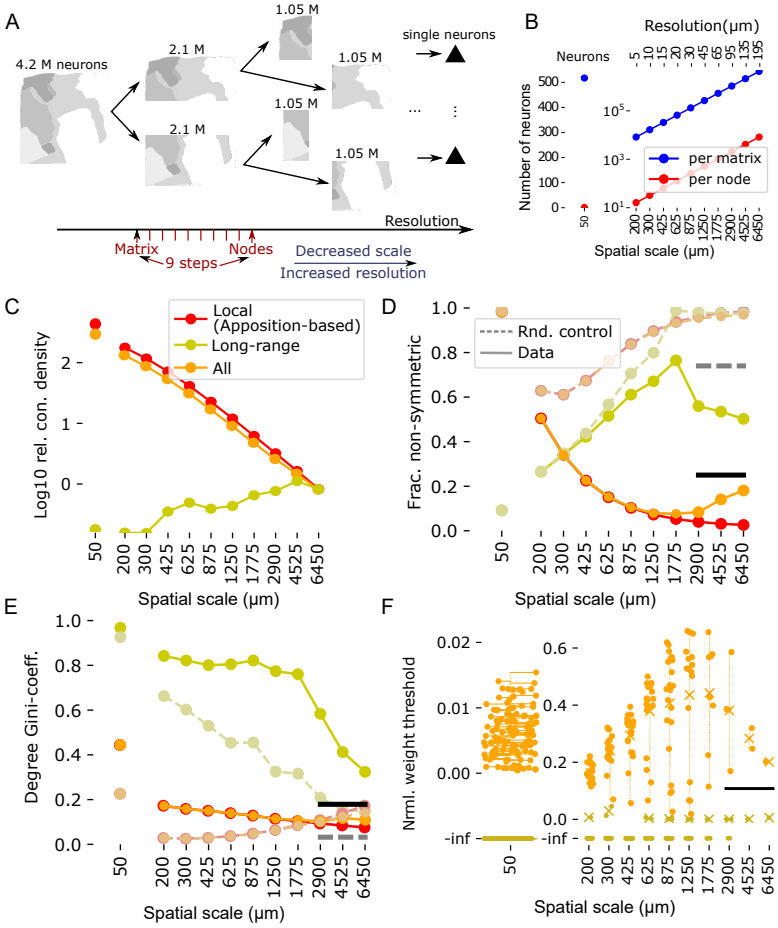

Figure 9: **Structure of connectivity at different resolutions.** A: At each step, a population is split into two equally sized ones according to their 2d locations, leading to a hierarchy of populations at decreasing scale. For a given population, we consider the matrix of connection counts between sub-populations contained within it at a scale 9 steps lower. B: Resulting number of neurons considered per connectivity matrix (blue) and per node of the matrix (red) at different scales / resolutions. C: Number of connections divided by number of neuron pairs considered in connectivity matrices at different scales, for apposition- (red), long-range- (yellow) and all connections (orange). Normalized to the values at the largest scale. D: Fraction of non-symmetric connections, i.e. directed connections that do not have a corresponding connection in the other direction. E: Gini-coefficient of node degrees at various scales. F: Lowest normalized (to the maximum) threshold weight such that the sub-network of stronger connections is fully connected. "-inf" indicates the full network is not fully connected. dots: samples; x-marks: mean of non-inf values. Results for local connectivity not shown, but almost identical to all connections. Black bars in D-F: Comparable data point from the AIBS voxelized mouse connectome. Grey bars: Comparable controls based on the AIBS connectome.

fell more proportionally to the number of pairs (Fig. S8B2), indicating that it is more evenly distributed along scales (Fig. 9C, yellow).

Interestingly, for local connectivity at larger scales and consequently lower resolutions, a large part of synaptic connections were within the population represented by a node, i.e., in entries of the main diagonal of the connectivity matrix. When we excluded those entries from the count (Fig. S8B1, teal), we found a drop in the fraction of connections considered at scales above $1775\mu m$. This was not found for long-range connectivity, but persisted when both were analyzed together (Fig. S8B2, B3).

Over-expression of bidirectional connectivity has been found both on a neuron-to-neuron level (Perin et al., 2011; Song et al., 2005) and whole-brain level (Oh et al., 2014; van den Heuvel et al., 2016). Thus, we next analyzed the symmetry of connectivity matrices at different scales by measuring the fraction of connections that do *not* have a corresponding connection in the opposite direction (Fig. 9D). As this measurement depends on the number of bidirectional connections arising by chance and thus the overall distribution of values in a matrix, we compared the results to a control with shuffled matrices (Fig. 9D, desaturated colors with dashed lines). We found generally increasing asymmetry for long-range and decreasing asymmetry for local connectivity, although both values were less than that of the random control, with the difference increasing at larger scales. Curiously, when both connectomes were combined and at scales over $1775\mu m$, a reversal was observed that was not present in the individual connectomes: asymmetry increased and the measure approached that of the random control.

As a reference point, we also calculated this measure directly for the primary somatosensory regions of the AIBS voxelized mouse connectome, which we reduced to a comparable 518×518 connectivity matrix (for details, see Sec. 5.3.11). In terms of organization, this would correspond to the largest spatial scale, but due to the smaller size of the mouse brain relative to the rat brain, the actual spatial scale is closer to $2500\mu m$. We found that the result matched the one for all connections at the largest scale, although a larger fraction of bidirectional connections is present in the random control (Fig. 9D, black and grey bars).

Another measure of connectivity matrices often employed is the distribution of node degrees (Gal et al., 2017; Nigam et al., 2016; Bonifazi et al., 2009). Here, we calculated the Gini-coefficient, measuring the inequality of degree distributions, at various scales (Fig. 9E). For all connectomes we found more equal distributions at larger scales, although the drop in Gini-coefficient was more pronounced for long-range connectivity. Overall, distributions were more unequal than the random controls, except for local connectivity and all connections at scales over $2900\mu m$. This is in contrast to the AIBS somatosensory regions where the Gini-coefficient was larger than the control (Fig. 9E, black and grey bars), but the overall value is close to the results at the largest scale.

Finally, we considered the robustness of connectivity at all scales, in terms of the following question: What is the largest weight threshold, such that the sub-network containing only edges stronger than the threshold has a single strongly connected component, i.e. any neuron can be reached from any other neuron along the remaining edges. We report this threshold, normalized to the maximum weight in the network, or negative infinity if the network with all edges is not connected (Fig. 9F). First, the long-range connections had in most cases more than one connected component, except at the largest scales. Even then, the weight threshold was close to zero indicating that all connections were required to form a single component. For all connections results were increasingly more robust for larger scales until a scale of $1775\mu m$ where only connections stronger than half the maximum were required. After that point the trend reversed to a threshold of 0.18 at full scale – close to the value found for the AIBS (Fig. 9F, black bar). Results for only local connections were almost identical to all connections (not shown).

Taking all of this together, we found for all measurements a vast range of different values at different scales. As such, the close match between results at the largest scale and the comparable scale in the raw AIBS mouse connectome is significant. Even though the AIBS connectome was used as a data source for model building, it serves to validate the complex workflow of adapting data from mouse to rat and combining two connectome algorithms. Additionally, the gradual rather than sudden shift of values for increasing scales when all connections were considered indicates that connectivity exists along a spectrum of scales, instead of as two separate (local and long-range) networks.

One key observation for almost all measurements was a reversal of trends at a scale of around $1775\mu m$ – or at the corresponding resolution of $65\mu m$. At that point, matrices specify the strengths of connections between 512 groups of around 1000 neurons each. The groups are almost exclusively driven by symmetrical connections to / from the other groups that robustly tie them together into a single component. Conversely, at full scale (512 groups of around 8000 neurons), groups

are more driven by internal connections and between-node connections with lower symmetry. At that scale, we were arguably analyzing connectivity between individual cortical columns, in that the number of neurons per node was close to that characterized by Markram et al. (2015). Consequently, greater importance of internal connections was expected, as columns are thought to act as independent computational units. The decrease in symmetry can then be interpreted as directionality of information flow emerging at that scale. Conversely, the asymmetry at the smallest scales would relate to the information flow within such a computational unit.

3.3 Features of thalamic innervation

The locations of synaptic contacts from thalamo-cortical projections were directly prescribed in the form of a layer profile and a horizontal range (see Sec. 2.1.5). However, the locations and identity of neurons innervated by these synapses is the result of a number of additional factors. We assume that any dendritic segment in a given depth bin is a potential target of innervation with no specific preference (see Sec. 2.2.5). As such, neuronal innervation is also the result of dendritic morphology and placement.

We counted the number of thalamic fibers innervating neurons at various depths and compared the result to the prescribed synapse densities (Fig. 10A). We found that for both core- and matrix-type projections, the pattern of peaks and troughs is informed by the pattern of synapse densities, but with the locations of peaks shifted downwards. In fact, most peaks were located in depth bins where the corresponding synapse density was close to zero, indicating that synapse density profiles alone can be misleading about the location of innervated neurons. On a per-neuron basis, the number of thalamic inputs varied drastically, even within the same layer (Fig. 10B). Overall, the matrix-type projection innervated neurons in layers 1 and 2 more strongly than the core-type one, while in layers 3, 4, and 6, the roles were reversed. Neurons in layer 5 were innervated on average equally strongly by both projections.

In the model, each thalamic fiber innervates only a given subvolume of the somatosensory cortex. This is based on the observation that the axons have a limited horizontal extent after entering the region (see Sec. 2.1.5), which leads to nearby pairs of cortical neurons being innervated by strongly overlapping sets of fibers. As pairs with many common inputs are likely to have similar stimulus preferences, the magnitude and range of this effect has consequences for the emergence of functional assemblies of neurons. While the model describes a stochastic state that is unaffected by plasticity, we can still assess the bounds of such an effect. We calculated the common thalamic innervation (CTI) for pairs as the intersection over the union of the sets of thalamic fibers innervating each of them.

As expected, the CTI was distance dependent (Fig. 10C, D), albeit with strong variability. Even directly neighboring pairs might not share a single thalamic afferent, leading to a sparse spatial distribution of pairs with strong overlap. We fit Gaussian profiles of the distance dependence of CTI for both types of thalamic inputs. Both core- and matrix-type projections produced roughly equally strong overlapping innervation (Fig. 10E1) that increases for lower layers in the case of core-type and decreases for matrix-type projections, although layer 5 is an outlier with equally strong CTI for both input streams. The horizontal range of common innervation was stronger for matrix-type projections in all layers (Fig. 10E2).

To estimate the possible implications of CTI on the formation of assemblies, we used the matrix of CTI of pairs of neurons to split them into 100 clusters (see Supplementary methods). The resulting clusters are spatially overlapping (Fig. 10F). To estimate to what degree neuron location determines cluster membership, we calculated the silhouette score with respect to neuron distances. A low score would indicate strongly spatially overlapping clusters. We found consistently lower scores for matrix- than for core-type projections in all layers (Fig. 10G). Especially in layer 6, the location of a neuron seems to have little impact on its cluster based on thalamic innervation. One reason may be the presence of pyramidal cells with horizontal dendrites in that layer that consequently collect and integrate input from many different fibers.

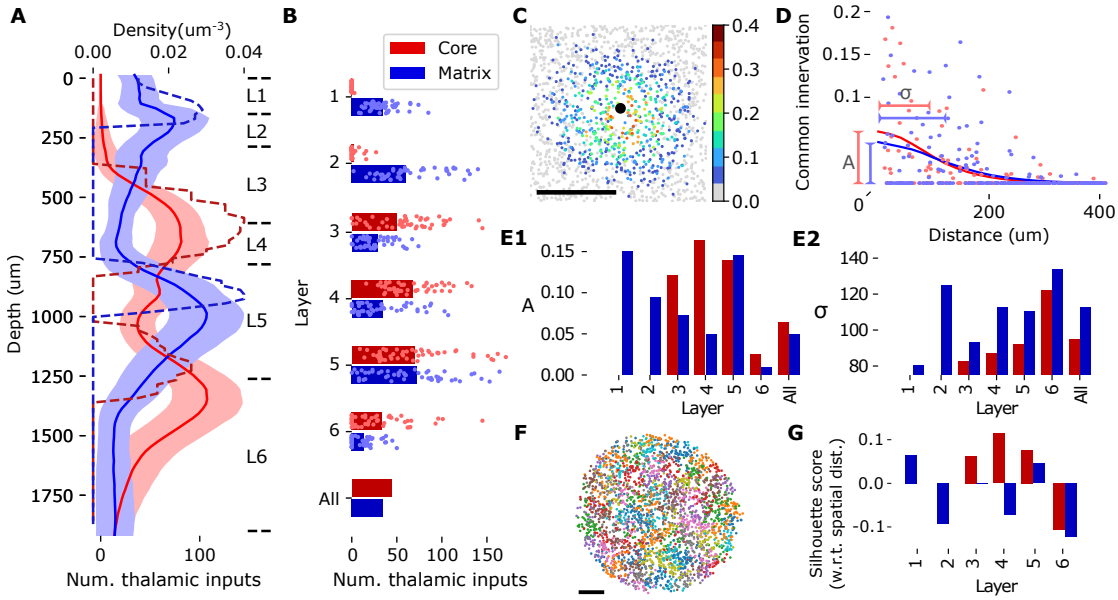

Figure 10: **Anatomy of thalamic innervation.** A: Depth profiles of synapse densities (dashed lines) and mean number of thalamic inputs per neuron (solid lines) for core (red) and matrix-type (blue) thalamo-cortical projections. Shaded area indicates the standard error of mean. B: Mean number of thalamic inputs for neurons in individual layers or all neurons. Dots indicate values for 50 randomly picked neurons of the indicated layers. Colors as in A. C: Common thalamic innervation (CTI) of an exemplary neuron (black dot) and neurons surrounding it, calculated as the intersection over union of the sets of thalamic fibers innervating each of them. Scale bar: $200\mu m$. D: CTI of pairs of neurons at various horizontal distances. Dots indicate values for 125 randomly picked pairs; lines indicate a sliding average with a window size of $40\mu m$. We perform a Gaussian fit to the data, extracting the amplitude at $0\mu m$ (A) and the standard deviation (σ). E: Values of A and σ for pairs in the individual layers or all pairs. Colors as in A. F: Clusters based on CTI of 3000 randomly selected neurons in a circular region of layer 4 by the matrix-type projection. Scale bar: $200\mu m$. G: Silhouette score of clusters in individual layers, evaluated against the matrix of spatial distances of the neurons.

4 Discussion

We have presented a model of the non-barrel primary somatosensory cortex of the juvenile rat that represents its neuronal and synaptic anatomy in high detail. The model contains biologically validated features at multiple scales, ranging from subcellular, such as predicted dendritic locations of synapses, to regional, such as its neuronal composition. While the modeling process was based on a previously published pipeline (Markram et al., 2015), many refinements have been and published since then (see Table 1). This manuscript integrated all these improvements into a single model. The model was also built on a much increased spatial scale, covering $48.7mm^3$, compared to the previously built $0.29mm^3$.

Improvements came in the form of integrating additional data, such as anatomical atlases (Paxinos and Watson, 2007; Bolaños-Puchet and Reimann, prep), cell density profiles (Keller et al., 2019), *in-vivo* neuron reconstructions, predicted strengths and structure of synaptic pathways (Reimann et al., 2019), the synapse density profile of a second type of thalamic input (Meyer et al., 2010), and the morphology of thalamo-cortical axons (Economo et al., 2016). Existing modeling methods were also improved by validating the classification of neuronal morphologies objectively (Kanari et al., 2019). The most complex changes, however, were a result of the increased spatial scale requiring the use of new modeling methods.

While local cortical curvature and other anatomical variability could be considered to be negligible at the previous scale, these had to be taken into account for this model. This required an enrichment of the anatomical atlas with data on local orientations, i.e., the direction towards the cortical surface. Other methods had then to be adapted to make use of that information, such as the transformation of layer density profiles to density atlases, and the algorithm ensuring proper placement of dendritic and axonal features. Further, while at the smaller scale synaptic connectivity could be modeled purely based on axonal and dendritic morphologies, this method is limited to connectivity on a local scale up to a few hundreds of μm. We found that relying on that method alone would have underestimated the connectivity between somatosensory sub-regions, requiring us to fill in the missing connectivity by adapting an algorithm and its associated data from mouse to juvenile rat.

Yet, these challenges also provided us with opportunities for anatomical analyses and predictions leading to novel insights with respect to the structure of synaptic connectivity. First, we were able to predict the effect of cortical curvature on neuronal composition by increasing or decreasing the space available for individual layers. If we assume that each layer has a given computational purpose (Felleman and Van Essen, 1991), then this may have functional consequences. However, there may be mechanisms compensating for this, which the model is not capturing.
Second, we predicted the topology of synaptic connectivity with neuronal resolution at an unprecedented scale. We were also able to compare and contrast the topology of the local and long-range connectivity. We found that the structure of neither can be explained by connection probabilities, degree-distributions, or distance-dependence alone, not even when individual pathways formed between morphological types are taken into account. Further, we predict that the long-distance connectivity forms a strong structural backbone distributing information from a small number of hubs. The paths within the long-range network strongly rely on neurons in layer 5 and form shortcut paths for neurons further away than around $2mm$; for smaller distances, local connectivity provides equivalent or shorter paths.

Third, we were able to generate lower-resolution versions of the predicted connectome, allowing us to compare a biologically realistic meso-scale connectome to its underlying neuron-to-neuron micro-scale connectome. The techniques we developed to that end resulted in comparable connectivity matrices at eleven levels of resolution that could be analyzed with the same approaches. We found that they are drastically different at different resolutions for most analyses, highlighting the difficulty of predicting the micro-scale from meso-level data.

Fourth, we predicted strong constraints on the emergence of cortical maps from the anatomy of thalamo-cortical innervation, i.e., from the combination of: shape and placement of cortical dendrites, the specific layer pattern formed by thalamo-cortical axons, and their horizontal reach. We demonstrated how differences in these parameters may lead to stronger or weaker topographical mapping with more or less spatially overlapping neuronal clusters. Although many more mechanisms will ultimately affect this – chief among them synaptic plasticity – this shows that the mere anatomy can already have an effect on the outcome.

Taken together, the above points demonstrate how novel insights that are not readily apparent in disparate data and individual models can be gained when they are combined in a way that creates a coherent whole. For example, our predictions on the structure of common thalamic innervation is on the surface a consequence of the specific layer profiles of the innervations. However, as we have demonstrated, the layer profiles alone can be misleading, as neurons – especially pyramidal cells – can gather inputs from outside their home layer, leading to a shift of the profiles. As such, the result was obtainable only by combining layer profiles with dendritic morphologies, and doing so within the same spatial context, thereby creating a coherent model volume.

Our results also demonstrate the value of modeling at large scales. The increased spatial scale of this work compared against (Markram et al., 2015) allowed us to construct a meso-scale connectome from the model with complete knowledge about its predicted implementation in the form of billions of individual synaptic connections. This allowed us to perform comparable analyses of the structure of the connectome at scales spanning three orders of magnitude, revealing a drastic shift in terms of its relative connection density, and a curious reversal at scales around $1500\mu m$ in terms of its symmetry and robustness of connectedness. This is a first attempt at contrasting quantitatively the connectivity between multiple microcircuits and within the very same microcircuits.

This iteration of the model is far from complete in several ways. Most of them stem from compromises and generalizations that were made to be able to parameterize the process with biological data. For example, for a subset of input data, we made no distinction between individual subregions. This means that the model is not capturing the known differences between the disgranular (S1DZ, S1DZO) and granular zones. Should improved and more specific data sources, such as EM-based connectomes or whole-brain neuron reconstructions, become available in the future, our modeling pipepline is designed to readily utilize them for refinement. There are also known limitations to the actual model building steps. While we enforce cell density profiles along the depth-dimension, densities are assumed to be otherwise uniform within a given subregion. For additional structure, the algorithms will need to be updated; this will be required for example, for the modeling of the barrel field. Similarly, during volume filling, orientation and placement of neuron morphologies is considered only along the depth-axis. modeling of the barrel system will also require the consideration of orientation and location with respect to nearby barrel centers.

The functional and computational implications of our findings are unfortunately hard to predict without also considering data on neuron activity. Yet the anatomical model presented here provides a solid basis for large-scale electrical simulations of neuron activity. A future manuscript will describe refinements in single-cell electrical modeling, synaptic physiology, and compensation for missing extrinsic inputs from regions outwith the somatosensory area.

References

Abdellah, M., Hernando, J., Eilemann, S., Lapere, S., Antille, N., Markram, H., and Schürmann, F. (2018). NeuroMorphoVis: A collaborative framework for analysis and visualization of neuronal morphology skeletons reconstructed from microscopy stacks. *Bioinformatics*, 34(13):i574–i582.

Anwar, H., Riachi, I., Hill, S., Schürmann, F., and Markram, H. (2009). An Approach to Capturing Neuron Morphological Diversity. In De Schutter, E., editor, *Computational Modeling Methods for Neuroscientists*, pages 211–232. The MIT Press.

Bolaños-Puchet, S. and Reimann, M. W. (prep). Flattening of enhanced cortical atlases opens up new possibilities for data-driven modeling and data visualization. *In preparation*.

Bonifazi, P., Goldin, M., Picardo, M. A., Jorquera, I., Cattani, A., Bianconi, G., Represa, A., Ben-Ari, Y., and Cossart, R. (2009). GABAergic Hub Neurons Orchestrate Synchrony in Developing Hippocampal Networks. *Science*, 326(5958):1419–1424.

Bota, M., Sporns, O., and Swanson, L. W. (2015). Architecture of the cerebral cortical association connectome underlying cognition. *Proceedings of the National Academy of Sciences*, 112(16).

Buzás, P., Eysel, U. T., and Kisvárday, Z. F. (1998). Functional topography of single cortical cells: an intracellular approach combined with optical imaging. *Brain research. Brain research protocols*, 3 2:199–208.

Colizza, V., Flammini, A., Serrano, M. A., and Vespignani, A. (2006). Detecting rich-club ordering in complex networks. *Nature physics*, 2(2):110–115.

Curto, C., Giusti, C., Marku, K., Pastalkova, E., and Itskov, V. (2013). Pairwise correlation graphs from hippocampal population activity have highly non-random, low-dimensional clique topology. *BMC neuroscience*, 14(1):1–2.

de Reus, M. A. and van den Heuvel, M. P. (2013). The parcellation-based connectome: Limitations and extensions. *NeuroImage*, 80:397–404.

Diestel, R. (2005). Graph theory 3rd ed. *Graduate texts in mathematics*, 173:33.

Economo, M. N., Clack, N. G., Lavis, L. D., Gerfen, C. R., Svoboda, K., Myers, E. W., and Chandrashekar, J. (2016). A platform for brain-wide imaging and reconstruction of individual neurons. *eLife*, 5:e10566.

Felleman, D. J. and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, 1(1):1–47.

Gal, E., Amsalem, O., Schindel, A., London, M., Schürmann, F., Markram, H., and Segev, I. (2021). The Role of Hub Neurons in Modulating Cortical Dynamics. *Frontiers in Neural Circuits*, 15(718270).

Gal, E., London, M., Globerson, A., Ramaswamy, S., Reimann, M. W., Muller, E., Markram, H., and Segev, I. (2017). Rich cell-type-specific network topology in neocortical microcircuitry. *Nature Neuroscience*, 20(7):1004–1013.

Gerfen, C. R., Economo, M. N., and Chandrashekar, J. (2018). Long distance projections of cortical pyramidal neurons. *Journal of Neuroscience Research*, 96(9):1467–1475.

Ghobril, J.-P. (2015). *Doctoral Dissertation: Large Volume Imaging of Rodent Brain Anatomy with Emphasis on Selective Plane Illumination Microscopy*. École polytechnique fédérale de Lausanne.

Giacopelli, G., Tegolo, D., Spera, E., and Migliore, M. (2021). On the structural connectivity of large-scale models of brain networks at cellular level. *Scientific Reports*, 11(1):4345.

Giusti, C., Pastalkova, E., Curto, C., and Itskov, V. (2015). Clique topology reveals intrinsic geometric structure in neural correlations. *Proceedings of the National Academy of Sciences*, 112(44):13455–13460.

Guo, K. H., Yamawaki, N., Barrett, J. M., Tapies, M., and Shepherd, G. M. (2020). Cortico-thalamo-cortical circuits of mouse forelimb S1 are organized primarily as recurrent loops. *Journal of Neuroscience*, 40(14):2849–2858.

Harris, J. A., Mihalas, S., et al. (2019). Hierarchical organization of cortical and thalamic connectivity. *Nature*, 575(7781):195–202.

Kahle, M. (2009). Topology of random clique complexes. *Discrete mathematics*, 309(6):1658–1671.

Kahle, M. et al. (2014). Topology of random simplicial complexes: a survey. *AMS Contemp. Math*, 620:201–222.

Kanari, L., Dłotko, P., Scolamiero, M., Levi, R., Shillcock, J., Hess, K., and Markram, H. (2018). A topological representation of branching neuronal morphologies. *Neuroinformatics*, 16(1):3–13.

Kanari, L., Ramaswamy, S., Shi, Y., Morand, S., Meystre, J., Perin, R., Abdellah, M., Wang, Y., Hess, K., and Markram, H. (2019). Objective Morphological Classification of Neocortical Pyramidal Cells. *Cerebral Cortex*, 29(4):1719–1735.

Karube, F. and Kisvárday, Z. F. (2011). Axon topography of layer iv spiny cells to orientation map in the cat primary visual cortex (area 18). *Cerebral cortex*, 21 6:1443–58.

Keller, D., Meystre, J., Veettil, R. V., Burri, O., Guiet, R., Schürmann, F., and Markram, H. (2019). A Derived Positional Mapping of Inhibitory Subtypes in the Somatosensory Cortex. *Frontiers in Neuroanatomy*, 13(78).

Lee, S.-H. and Dan, Y. (2012). Neuromodulation of Brain States. *Neuron*, 76(1):209–222.

Livingstone, M. S. and Hubel, D. H. (1983). Specificity of cortico-cortical connections in monkey visual system. *Nature*, 304(5926):531–534.

Lütgehetmann, D., Govc, D., Smith, J. P., and Levi, R. (2020). Computing Persistent Homology of Directed Flag Complexes. *Algorithms*, 13(1):19.

Maneewongvatana, S. and Mount, D. M. (1999). It's okay to be skinny, if your friends are fat. In *Center for geometric computing 4th annual workshop on computational geometry*, volume 2, pages 1–8.

Markram, H., Muller, E. B., Ramaswamy, S., Reimann, M. W., et al. (2015). Reconstruction and Simulation of Neocortical Microcircuitry. *Cell*, 163:456–492.

Mei, J., Muller, E., and Ramaswamy, S. (2022). Informing deep neural networks by multiscale principles of neuromodulatory systems. *Trends in Neurosciences*, 45(3):237–250.

Meyer, H. S., Wimmer, V. C., Hemberger, M., Bruno, R. M., De Kock, C. P., Frick, A., Sakmann, B., and Helmstaedter, M. (2010). Cell type-specific thalamic innervation in a column of rat vibrissal cortex. *Cerebral Cortex*, 20(10):2287–2303.

Mountcastle, V. B. (1998). *Perceptual neuroscience: the cerebral cortex*. Harvard University Press.

Muralidhar, S., Wang, Y., and Markram, H. (2014). Synaptic and cellular organization of layer 1 of the developing rat somatosensory cortex. *Frontiers in Neuroanatomy*, 7.

Narayanan, R. T., Udvary, D., and Oberlaender, M. (2017). Cell type-specific structural organization of the six layers in rat barrel cortex. *Frontiers in Neuroanatomy*, 11.

Newman, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Review*, 45(2):167–256.

Nigam, S., Shimono, M., Ito, S., Yeh, F.-C., Timme, N., Myroshnychenko, M., Lapish, C. C., Tosi, Z., Hottowy, P., Smith, W. C., Masmanidis, S. C., Litke, A. M., Sporns, O., and Beggs, J. M. (2016). Rich-Club Organization in Effective Connectivity among Cortical Neurons. *The Journal of Neuroscience*, 36(3):670–684.

Nolte, M., Gal, E., Markram, H., and Reimann, M. W. (2020). Impact of higher order network structure on emergent cortical activity. *Network Neuroscience*, 4(1):292–314.

Nolte, M., Reimann, M. W., King, J. G., Markram, H., and Muller, E. B. (2019). Cortical reliability amid noise and chaos. *Nature Communications*, 10(3792).

Oh, S. W., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A. M., Mortrud, M. T., Ouellette, B., Nguyen, T. N., Sorensen, S. A., Slaughterbeck, C. R., Wakeman, W., Li, Y., Feng, D., Ho, A., Nicholas, E., Hirokawa, K. E., Bohn, P., Joines, K. M., Peng, H., Hawrylycz, M. J., Phillips, J. W., Hohmann, J. G., Wohnoutka, P., Gerfen, C. R., Koch, C., Bernard, A., Dang, C., Jones, A. R., and Zeng, H. (2014). A mesoscale connectome of the mouse brain. *Nature*, 508(7495):207–214.

Opsahl, T., Colizza, V., Panzarasa, P., and Ramasco, J. J. (2008). Prominence and control: the weighted rich-club effect. *Physical review letters*, 101(16):168702.

Paxinos, G. and Watson, C. (2007). *The Rat Brain in Stereotaxic Coordinates (Sixth Edition)*. Academic Press, London.

Perin, R., Berger, T. K., and Markram, H. (2011). A synaptic organizing principle for cortical neuronal groups. *Proceedings of the National Academy of Sciences*, 108(13):5419–5424.

Radnikow, G. and Feldmeyer, D. (2018). Layer- and cell type-specific modulation of excitatory neuronal activity in the neocortex. *Frontiers in Neuroanatomy*, 12.

Ramaswamy, S., Courcol, J.-D., et al. (2015). The neocortical microcircuit collaboration portal: a resource for rat somatosensory cortex. *Frontiers in Neural Circuits*, 9(44).

Reimann, M. W., Gevaert, M., Shi, Y., Lu, H., Markram, H., and Muller, E. (2019). A null model of the mouse whole-neocortex micro-connectome. *Nature Communications*, 10(3903).

Reimann, M. W., King, J. G., Muller, E. B., Ramaswamy, S., and Markram, H. (2015). An algorithm to predict the connectome of neural microcircuits. *Frontiers in computational neuroscience*, 9(120).

Reimann, M. W., Nolte, M., Scolamiero, M., Turner, K., Perin, R., Chindemi, G., Dłotko, P., Levi, R., Hess, K., and Markram, H. (2017). Cliques of neurons bound into cavities provide a missing link between structure and function. *Frontiers in Computational Neuroscience*, 11(48).

Reimann, M. W., Riihimäki, H., Smith, J. P., Lazovskis, J., Pokorny, C., and Levi, R. (2022). Topology of synaptic connectivity constrains neuronal stimulus representation, predicting two complementary coding strategies. *PLoS ONE*, 17(1):e0261702.

Santuy, A., Turégano-López, M., Rodríguez, J. R., Alonso-Nanclares, L., DeFelipe, J., and Merchán-Pérez, A. (2018). A Quantitative Study on the Distribution of Mitochondria in the Neuropil of the Juvenile Rat Somatosensory Cortex. *Cerebral Cortex*, 28(10):3673–3684.

Scala, F., Kobak, D., Shan, S., Bernaerts, Y., Laturnus, S., Cadwell, C. R., Hartmanis, L., Froudarakis, E., Castro, J. R., Tan, Z. H., Papadopoulos, S., Patel, S. S., Sandberg, R., Berens, P., Jiang, X., and Tolias, A. S. (2019). Layer 4 of mouse neocortex differs in cell types and circuit organization between sensory areas. *Nature Communications*, 10.

Scannell, J., Blakemore, C., and Young, M. (1995). Analysis of connectivity in the cat cerebral cortex. *The Journal of Neuroscience*, 15(2):1463–1483.

Scholtens, L. H., Schmidt, R., de Reus, M. A., and van den Heuvel, M. P. (2014). Linking Macroscale Graph Analytical Organization to Microscale Neuroarchitectonics in the Macaque Connectome. *Journal of Neuroscience*, 34(36):12192–12205.

Schüz, A. and Palm, G. (1989). Density of neurons and synapses in the cerebral cortex of the mouse: NEURONS AND SYNAPSES IN THE MOUSE CORTEX. *Journal of Comparative Neurology*, 286(4):442–455.

Shepherd, G. M. and Yamawaki, N. (2021). Untangling the cortico-thalamo-cortical loop: cellular pieces of a knotty circuit puzzle. *Nature Reviews Neuroscience*, 22(7):389–406.

Sizemore, A. E., Giusti, C., Kahn, A., Vettel, J. M., Betzel, R. F., and Bassett, D. S. (2018). Cliques and cavities in the human connectome. *Journal of Computational Neuroscience*, 44(1):115–145.

Somogyi, P. and Freund, T. F. (1989). Immunocytochemistry and Synaptic Relationships of Physiologically Characterized HRP-Filled Neurons. In Heimer, L. and Záborszky, L., editors, *Neuroanatomical Tract-Tracing Methods 2*, pages 239–264. Springer US, Boston, MA.

Song, S., Sjöström, P. J., Reigl, M., Nelson, S., and Chklovskii, D. B. (2005). Highly Nonrandom Features of Synaptic Connectivity in Local Cortical Circuits. *PLoS Biology*, 3(3):e68.

Toledo-Rodriguez, M., Goodman, P., Illic, M., Wu, C., and Markram, H. (2005). Neuropeptide and calcium-binding protein gene expression profiles predict neuronal anatomical type in the juvenile rat. *Journal of Physiology*, 567(2):401–413.

Unger, F. and Krebs, J. (2022). Mcmc sampling of directed flag complexes. *In preparation*.

van den Heuvel, M. P., Bullmore, E. T., and Sporns, O. (2016). Comparative Connectomics. *Trends in Cognitive Sciences*, 20(5):345–361.

van den Heuvel, M. P. and Sporns, O. (2011). Rich-Club Organization of the Human Connectome. *Journal of Neuroscience*, 31(44):15775–15786.

Young, J.-G., Petri, G., Vaccarino, F., and Patania, A. (2017). Construction of and efficient sampling from the simplicial configuration model. *Physical Review E*, 96(3):032312.

Zhou, S. and Mondragon, R. (2004). The Rich-Club Phenomenon in the Internet Topology. *IEEE Communications Letters*, 8(3):180–182.

Author contributions

- Conceptualization: H.M., E.B.M., S.R., M.W.R.
- Data curation: Y.S., W.V.G., S.B.P., H.L., D.K., L.K.
- Formal analysis: L.K., S.B.P., A.E., D.E.S., M.W.R.
- Funding acquisition: H.M.
- Investigation: D.E.S., M.W.R., J.P.S., M.S.
- Methodology: E.B.M, L.K., W.V.G, S.R., S.B.P., A.R., D.E.S., N.N., M.W.R., K.H., J.L., R.L., J.P.S., D.K.
- Project administration: H.M., F.S., E.B.M, S.R., A.R., M.W.R., K.H., R.L., S.L., J.P.
- Resources: F.S., J.G.K., P.K., F.P., J.-D.C.
- Software: A.A., H.D., V.S., S.B.P, A.E., M.W.R., H.L., J.L., J.P.S., F.D., A.De., J.G.K., P.K., F.P., M.W., B.C., J.-D.C., T.D., G.F., M.G., J.B.H., L.V., J.H., J.L.R.
- Supervision: H.M., F.S., E.B.M, S.R., A.R., M.W.R., K.H., R.L., F.D., J.G.K., J.-D.C.
- Validation: L.K., W.V.G, H.D., V.S., S.B.P, A.E., C.P., M.W.R., A.Di., M.G.
- Visualization: L.K., S.B.P, A.E., D.E.S., J.L., C.P., M.W.R., C.F.
- Writing - original draft: S.B.P, D.E.S., L.K., C.P., J.P.S., M.W.R., D.K., M.S.
- Writing - review & editing: A.A., S.B.P., S.R., A.E., J.I., M.W.R., K.H., R.L.

Acknowledgements

The authors would like to thank Giuseppe Chindemi, Javier DeFelipe and Rajnish Ranjan for help with the scientific development of the model; Tristan Carel, James Dynes, Stefan Eilemann, Bruno Magalhães and Arseny Povolotsky for contributions and support to the engineering challenges; the BBP Core Services team for responding to IT requests and services surrounding the research; Marwan Abdellah, Elvis Boci and Nadir Roman Guerrero for help with visualizations of the model; Zoltán Kisvárday for supervision of morphology reconstruction efforts; Eva Kenny, Silvia Scarabelli and Riccardo Sinsi for help with project management; and Karin Holm, Akiko Sato and Georges Khazen for support of manuscript development and helpful discussions.

Funding

This study was supported by funding to the Blue Brain Project, a research center of the École polytechnique fédérale de Lausanne (EPFL), from the Swiss government's ETH Board of the Swiss Federal Institutes of Technology. RL, JPS and JL were supported by EPSRC under grant number EP/P025072/1. RL was supported by a collaboration grant from EPFL.

Data and code availability

A subvolume of the model containing around 210,000 neurons, their connectivity and thalamic inputs is freely available in the Sonata format on Zenodo at the following link:

`https://zenodo.org/record/6906785`

It is also available via its digital object identifier:

https://doi.org/10.5281/zenodo.6906785

Code packages:

- FLAGSER-COUNT: Computes the number of directed cliques in a directed network, available from https://github.com/JasonPSmith/flagser-count
- PATHFINDER: Finds all directed paths between nodes in a directed network, available from https://github.com/JasonPSmith/pathfinder
- CONNECTOME_ANALYSIS: Contains various functions for the analysis of connectomes, including: control models generation, degree distributions, rich-club computations, available from https://github.com/danielaegassan/connectome_analysis

5 Supplementary Material

5.1 List of assumptions made in model building

Inherited assumptions
We assume that the published data sources and modelling steps (Table 1) are correct and consequently inherit their assumptions. Below, we will only list the assumptions made on top of that.

Data assumptions
We make a number assumptions related to generalizing input data

- We assume the juvenile rat SSCx volume is a uniformly scaled-down version of the adult SSCx.
- We assume no significant variability between nbS1 regions in terms of: Cell density layer profiles, neuronal morphologies, bouton densities, number of synapses per connection in pathways.
- We assume the relative strengths of connectivity and structure axonal targeting is comparable in SSCx regions between mouse and rat.
- We assume that the VPM to barrel cortex projection in adult rat is representative for a general core-type, feedforward thalamic input in juvenile rat. Similarly, POm to barrel cortex for matrix-type, feedback input.
- We assume mouse thalamo-cortical projection axons are informative for the structure of the rat thalamo-cortical projection system.

Structuring assumptions

- We assume that grouping of cortical neurons into 60 morphological classes is also useful for the description of the neuronal composition (cell density profiles) and the synaptic connectivity between them.
- We assume that the eight established subregions provide a parcellation that is useful for the description of long-range connectivity in the model.

Modeling assumption

- We assume that purely vertical cell density profiles capture all relevant details of neuronal composition.
- We assume that the union of synaptic connections from two separate algorithms accurately describes the connectivity at all scales that are relevant in the model. Specifically, we assume there is no "midrange gap" remaining between the two algorithms.
- We assume that thalamo-cortical axons do not target specific classes of neurons, beyond what is given by their layer profiles.

- We assume the horizontal spread of thalamo-cortical axons can be captured by a Gaussian profile for the synapses formed.

5.2 Supplementary Tables and Figures

L1_DAC	Layer 1, descending axon cell	L1_HAC	Layer 1, horizontal axon cell
L1_LAC		L1_NGC-DA	
L1_NGC-SA		L1_SAC	
L23_BP	Layer 2, bipolar cell	L23_BTC	Layer 2/3, bitufted cell
L23_CHC	Layer 2/3, chandelier cell	L23_DBC	Layer 2/3, double bouquet cell
L23_LBC	Layer 2/3, large basket cell	L23_MC	Layer 2/3, martinotti cell
L23_NBC	Layer 2/3, nest basket cell	L23_NGC	Layer 2/3, neurogliaform cell
L23_SBC	Layer 2/3, small basket cell	L2_IPC	Layer 2, inverted pyramidal cell
L2_TPC:A	Layer 2, large tufted pyramidal cell	L2_TPC:B	Layer 2, early bifurcating pyramidal cell
L3_TPC:A	Layer 3, large tufted pyramidal cell	L3_TPC:C	Layer 3, small tufted pyramidal cell
L4_BP	Layer 4, bipolar cell	L4_BTC	Layer 4, bitufted cell
L4_CHC	Layer 4, chandelier cell	L4_DBC	Layer 4, double bouquet cell
L4_LBC	Layer 4, large basket cell	L4_MC	Layer 4, martinotti cell
L4_NBC	Layer 4, nest basket cell	L4_NGC	Layer 4, neurogliaform cell
L4_SBC	Layer 4, small basket cell	L4_SSC	Layer 4, spiny stellate cell
L4_TPC	Layer 4, tufted pyramidal cell	L4_UPC	Layer 4, untufted pyramidal cell
L5_BP	Layer 5, bipolar cell	L5_BTC	Layer 5, bitufted cell
L5_CHC	Layer 5, chandelier cell	L5_DBC	Layer 5, double bouquet cell
L5_LBC	Layer 5, large basket cell	L5_MC	Layer 5, martinotti cell
L5_NBC	Layer 5, nest basket cell	L5_NGC	Layer 5, neurogliaform cell
L5_SBC	Layer 5, small basket cell	L5_TPC:A	Layer 5, large tufted pyramidal cell
L5_TPC:B	Layer 5, early bifurcating pyramidal cell	L5_TPC:C	Layer 5, small tufted pyramidal cell
L5_UPC	Layer 5, untufted pyramidal cell	L6_BP	Layer 6, bipolar cell
L6_BPC	Layer 6 bipolar cell	L6_BTC	Layer 6, bitufted cell
L6_CHC	Layer 6, chandelier cell	L6_DBC	Layer 6, double bouquet cell
L6_HPC	Layer 6, horizontal pyramidal cell	L6_IPC	Layer 6, inverted pyramidal cell
L6_LBC	Layer 6, large baskte cell	L6_MC	Layer 6, martinotti cell
L6_NBC	Layer 6, nest basket cell	L6_NGC	Layer 6, neurogliaform cell
L6_SBC	Layer 6, small basket cell	L6_TPC:A	Layer 6, large tufted pyramidal cell
L6_TPC:C	Layer 6, small tufted pyramidal cell	L6_UPC	Layer 6 untufted pyramidal cell

Table S1: Morphological types (m-types) used in the model

Boundary		nrmlz. depth
Layer 1	top	0
	bottom	0.079
Layer 2	top	
	bottom	0.151
Layer 3	top	
	bottom	0.32
Layer 4	top	
	bottom	0.411
Layer 5	top	
	bottom	0.664
Layer 6	top	
	bottom	1.0

Table S2: Normalized depths used for layer boundaries

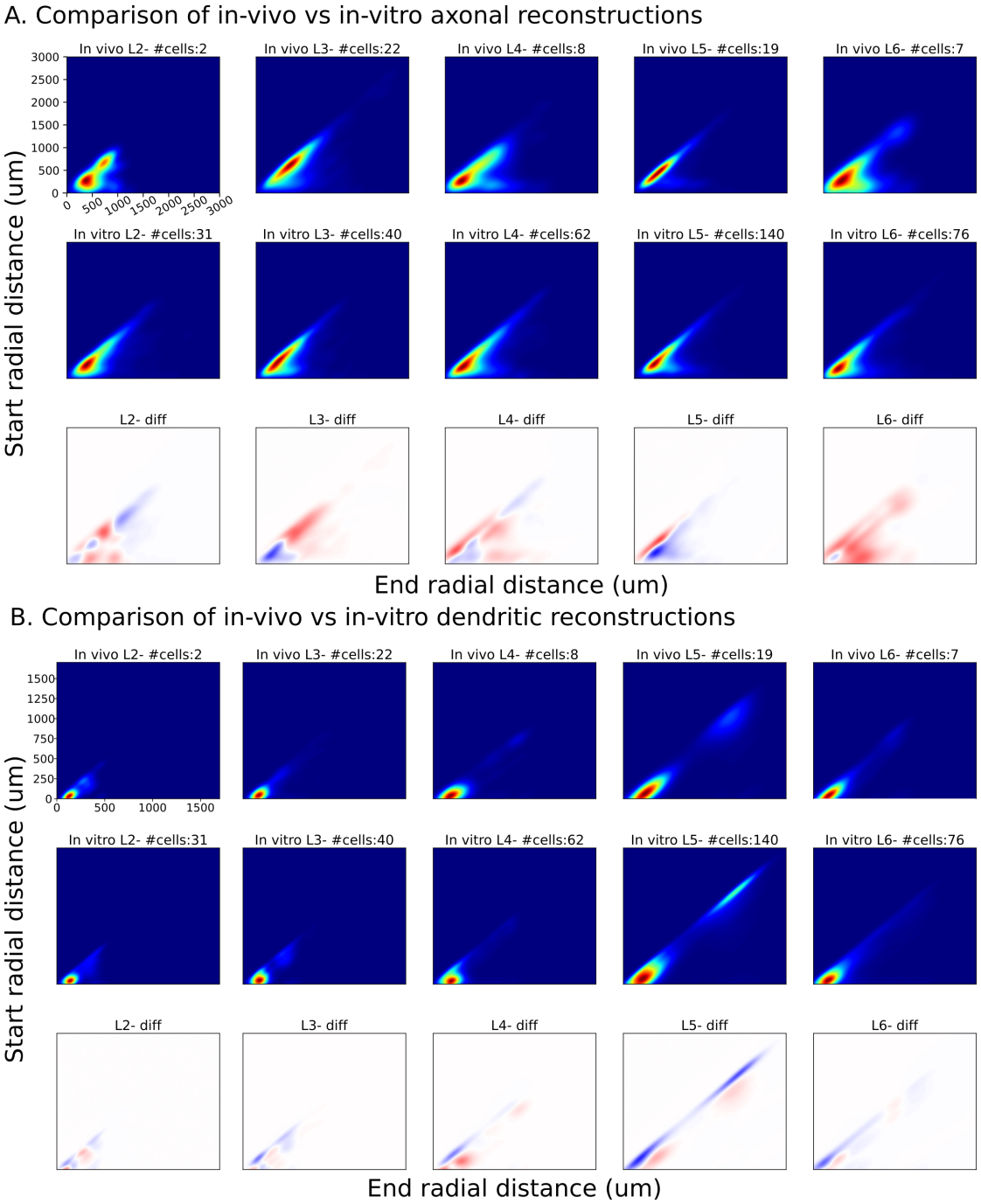

Figure S1: Topological comparison as in (Kanari et al., 2018, 2019) of in-vivo and in-vitro (i.e., from slices) axonal (A) and dendritic (B) reconstructions of rat somatosensory cortex of pyramidal cells from layers 2-6. Top row presents the topological profiles of in-vivo reconstructions, second row presents the topological profiles of in-vitro reconstructions and bottom row the difference between them (red: in-vivo, blue: in-vitro). Number of cells per layer are reported in individual pannels.

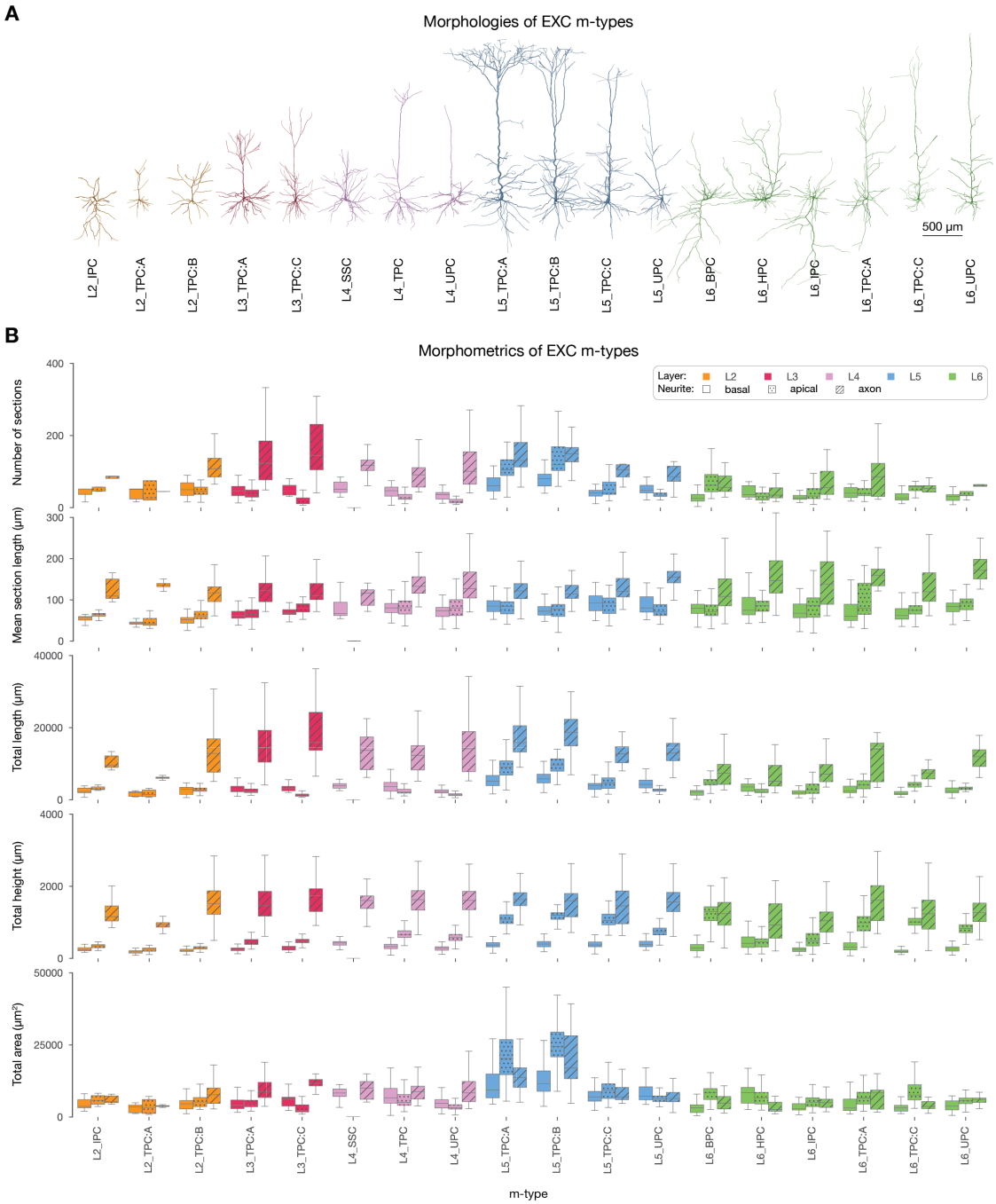

Figure S2: **Classification and morphometrics of excitatory morphologies.** A: Exemplar 3D reconstructions of the 18 excitatory m-types. Rendering and visualization was done in NeuroMorphoVis (Abdellah et al., 2018). Dendritic diameters are scaled (x3) for better resolution. B: Morphometrics of all morphologies belonging to the 18 excitatory m-types.

L1_HAC	axon	[I:0.00; I:1.00]
L23_MC	axon	[I:0.00; I:1.00]
L2_TPC	axon	[V:0.00; V:1.00]
		[IV:0.00; IV:1.00]
	dendrite	[I:0.00; I:1.00]
L2_TPC:A	axon	[V:0.00; V:1.00]
		[IV:0.00; IV:1.00]
	dendrite	[I:0.00; I:1.00]
L2_TPC:B	axon	[V:0.00; V:1.00]
		[IV:0.00; IV:1.00]
	dendrite	[I:0.00; I:1.00]
L3_TPC	axon	[V:0.00; V:1.00]
		[IV:0.00; IV:1.00]
	dendrite	[I:0.00; I:1.00]
L3_TPC:A	axon	[V:0.00; V:1.00]
		[IV:0.00; IV:1.00]
	dendrite	[I:0.00; I:1.00]
L3_TPC:B	axon	[V:0.00; V:1.00]
		[IV:0.00; IV:1.00]
	dendrite	[I:0.00; I:1.00]
L3_TPC:C	axon	[V:0.00; V:1.00]
		[IV:0.00; IV:1.00]
	dendrite	[I:0.00; I:1.00]
L4_MC	axon	[I:0.00; I:1.00]
L4_SSC	axon	[II:0.00; I:1.00]
L4_TPC	dendrite	[II:0.00; I:1.00]
L4_UPC	dendrite	[II:0.00; I:1.00]
L5_MC	axon	[I:0.00; I:1.00]
		[IV:0.00; IV:1.00]
L5_TPC	axon	[I:0.00; I:1.00]
	dendrite	[I:0.00; I:1.00]
L5_TPC:A	axon	[I:0.00; I:1.00]
	dendrite	[I:0.00; I:1.00]
L5_TPC:B	axon	[I:0.00; I:1.00]
	dendrite	[I:0.00; I:1.00]
L5_TPC:C	axon	[I:0.00; I:1.00]
	dendrite	[I:0.00; I:1.00]
L5_UPC	axon	[I:0.00; I:1.00]
L6_BPC	dendrite	[V:0.50; III:0.10]
L6_HPC	dendrite	[VI:0.00; VI:1.00]
L6_IPC	dendrite	[VI:0.00; V:0.20]
L6_MC	axon	[II:0.00; I:1.00]
		[IV:0.00; IV:1.00]
L6_TPC	dendrite	[V:0.80; III:0.50]
L6_TPC:A	dendrite	[V:0.80; III:0.50]
L6_TPC:C	dendrite	[V:0.80; III:0.50]
L6_UPC	dendrite	[V:0.70; III:0.50]

Table S3: Placement features used for modeling. The right column denotes the vertical target interval by specifying a layer and a relative offset within the layer, with 0.0 indication the bottom and 1.0 the top of the layer.

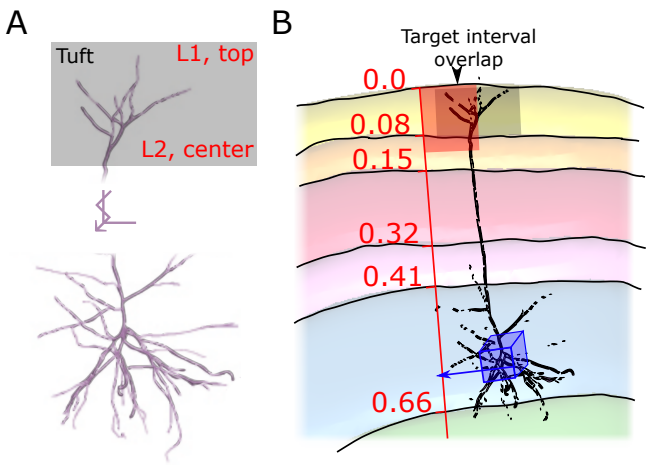

Figure S3: **Scoring the placement of a neuron morphology for a voxel.** A: Neurite features, here the apical tuft, were manually given a vertical *annotation interval* (grey) and assigned a *target interval*, expressed as a layer interval (red). (Note: parts of the apical dendrite visually shortened.) B: Then the placement of a neuron morphology in a given voxel (blue) is scored as follows: Normalized depth values of the *target interval* are calculated (red); the normalized depth of the voxel is looked up in an atlas (blue); it is used to calculate the normalized depth of the *annotation interval* when the morphology is placed in the voxel (grey); a score is calculated as in (Markram et al., 2015) based on the overlap of the intervals.

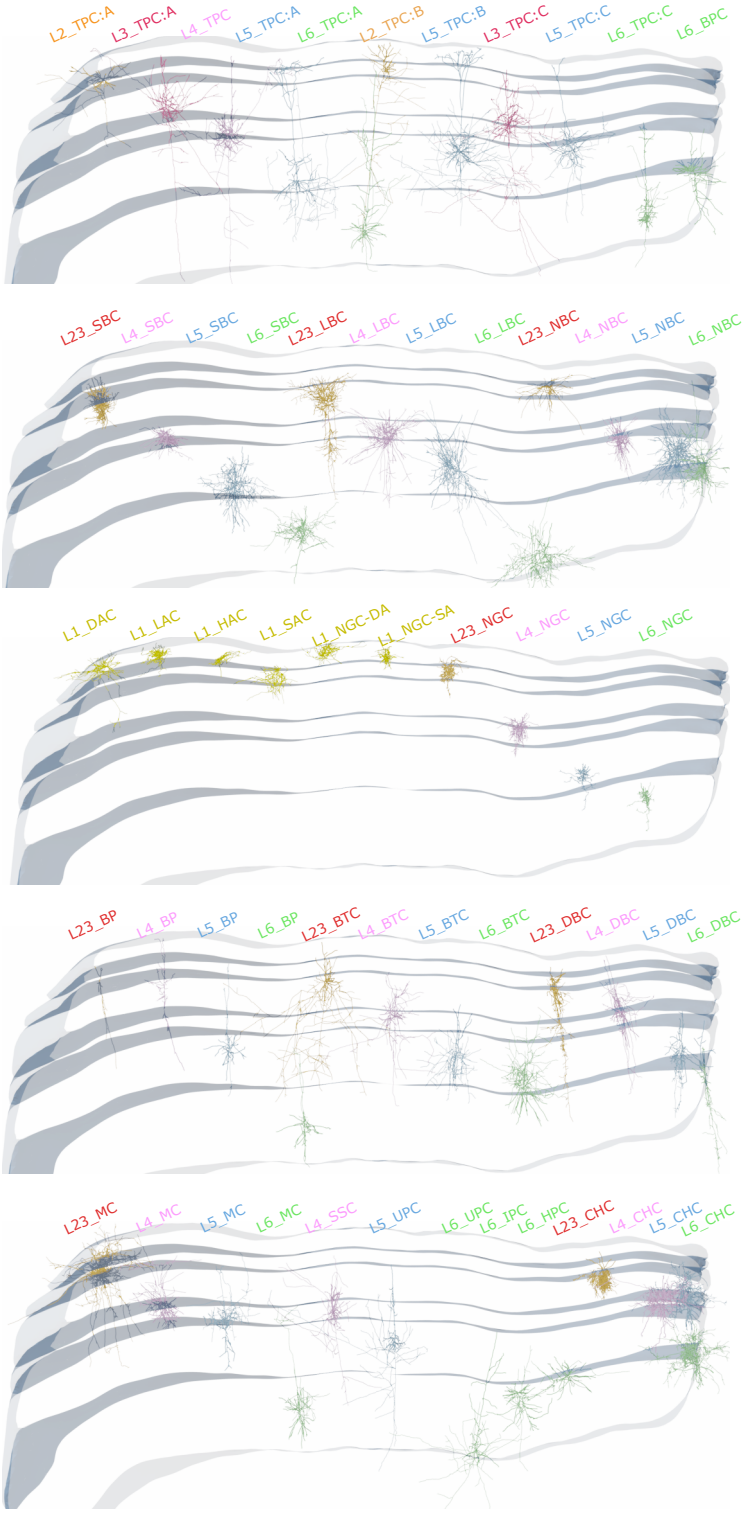

Figure S4: Exemplary neurons in the model, one per m-type, rendered in the context of a slice spanning all cortical layers (grey borders).

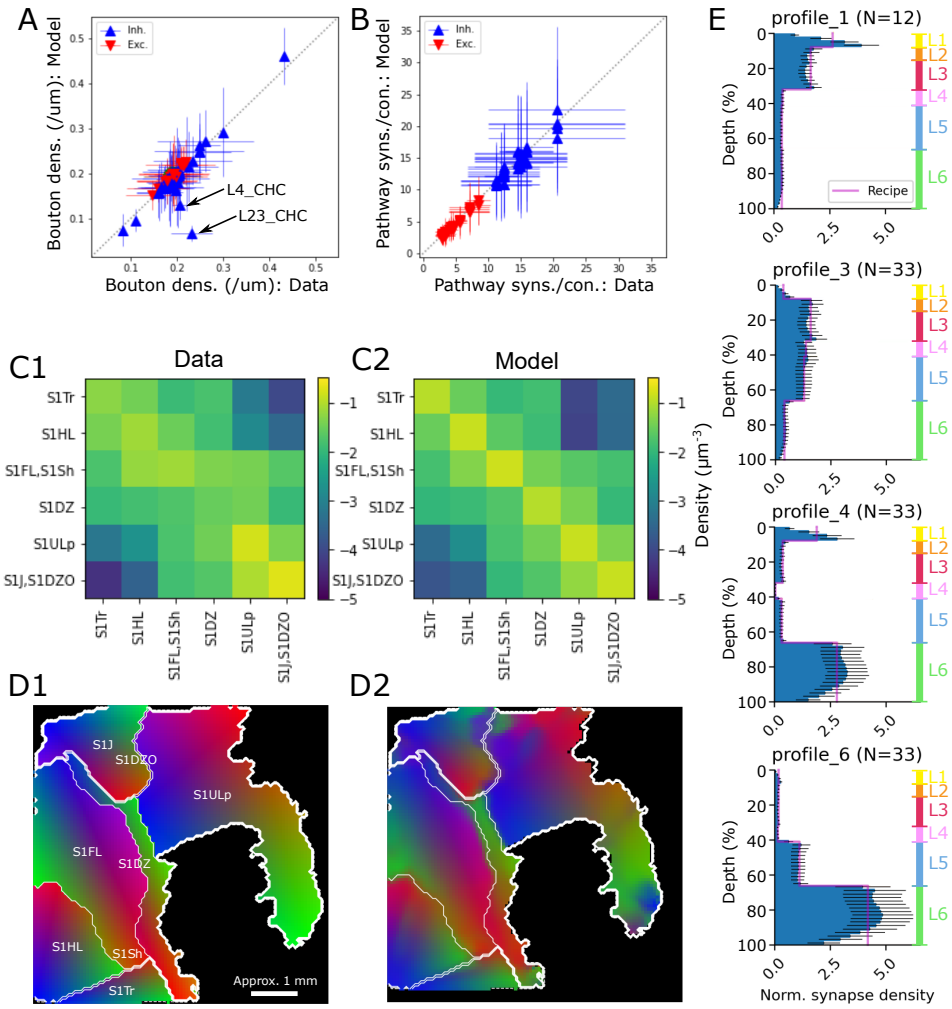

Figure S5: **Validation of modeled connectivity.** A, B: Touch connectivity is validated by comparing emerging bouton densities (A) and mean numbers of synapses per connection (B) to the target values from the data. Each data point depicts a single morphological type (A) or type-specific pathway (B). Black arrows: Bouton density for Chandelier Cells (ChC) in layers 2, 3 and 4. C: Mean densities of synapses from the union of touch connectivity and long-range connectivity in pathways within and between regions (C2) compared to that target values from the data (C1). D: Structure of the topographical mapping of connections; each part of the model predominately connects to neurons at equally colored locations. (D1) Target mapping from the data. (D2) Analyzed in the union of touch connectivity and long-range connectivity. E: Layer profiles of synapses placed in long-range connections between regions (blue bars) compared to the predicted target profiles from (Reimann et al. (2019), pink lines). Depicted is the density of synapses in a depth bin relative to the overall mean density over the entire depth.

Subregion	Number of thalamic input fibers	
	Core-type	Matrix-type
All	100000	72950
S1ULp	29701	21667
S1FL	25561	18647
S1J	15895	11596
S1HL	12154	8866
S1DZ	7962	5808
S1Tr	5326	3885
S1Sh	1991	1452
S1DZO	1410	1029

Table S4: Number of thalamic fibers providing inputs to the model and each of its subregions.

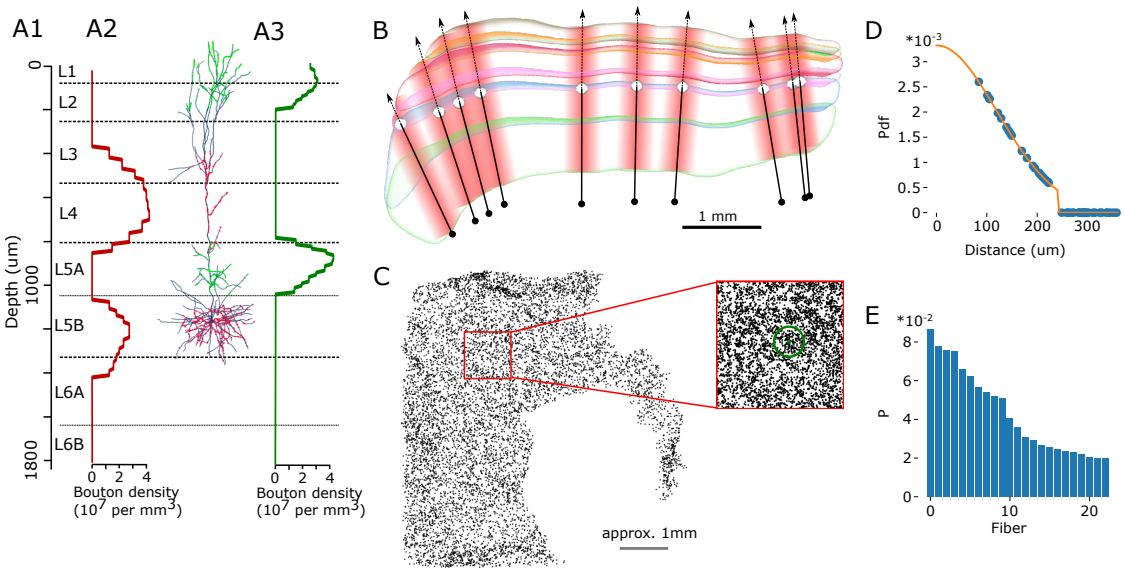

Figure S6: **Modeling thalamo-cortical synaptic inputs.** A: Input vertical profiles of thalamo-cortical synapses from core-type projections (A1) and matrix-type projections (A3). A2: Exemplary model neuron with projection synapses placed on dendrites according to the prescribed densities. B: Modeling of afferent thalamic fiber locations: Depicted is an exemplary slice of the model. For each fiber a location at the border between layers 4 and 5 is randomly chosen (10 examples shown; white spheres). A point $1500\mu m$ towards the bottom of layer 6 is chosen as the starting point of each fiber (black dots). A domain of influence is then assigned around a line starting at that location with the indicated direction (black arrows and red areas). Influence weakens with distance from the line with a Gaussian profile; for details, see D, E. C: Locations of fibers in the flat view. 10% of the full density shown. Inset: Locations of approximately $1mm^2$ shown in full density. The green circle indicates a single standard deviation of the Gaussian of influence strength used for the core style projections (i.e., $120\mu m$). D: For an exemplary synapse its distances to surrounding fibers is measured and the value of the Gaussian for those distances calculated. It is cut off at two standard deviations. E: The probability that a given fiber is chosen as innervator of a synaptic location (blue bar) is proportional to its value in D (shown only for fibers with nonzero probability).

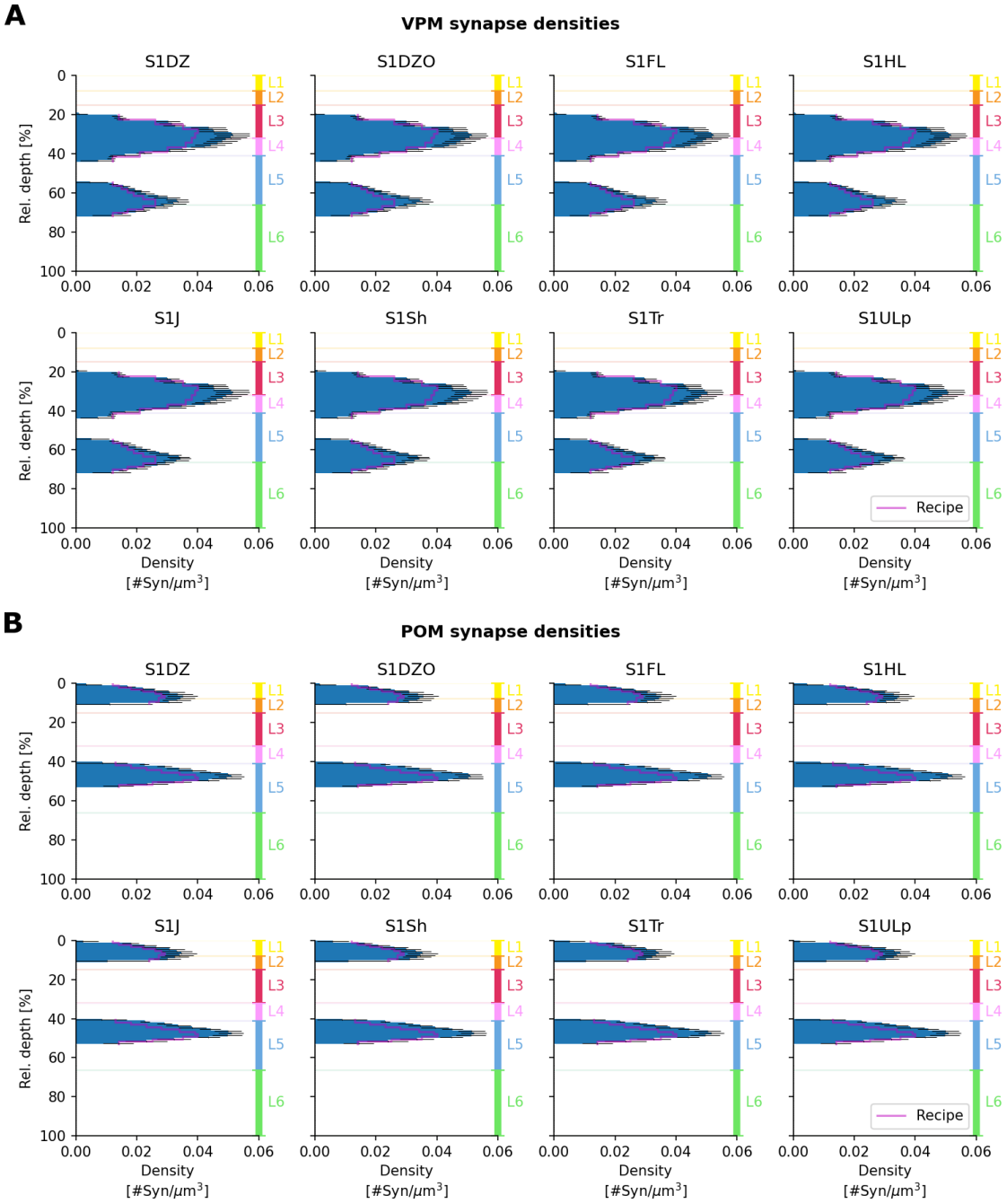

Figure S7: **Validation of thalamo-cortical density profiles.** Synapse density profiles of thalamo-cortical synapses from core-type (VPM) projections (A) and matrix-type (POm) projections (B), compared with the target profiles from Meyer et al. (2010) (pink lines). Depicted are the mean densities (black lines: ± SD) over voxels within depth bins relative to the respective total cortical thickness in each of the eight sub-regions.

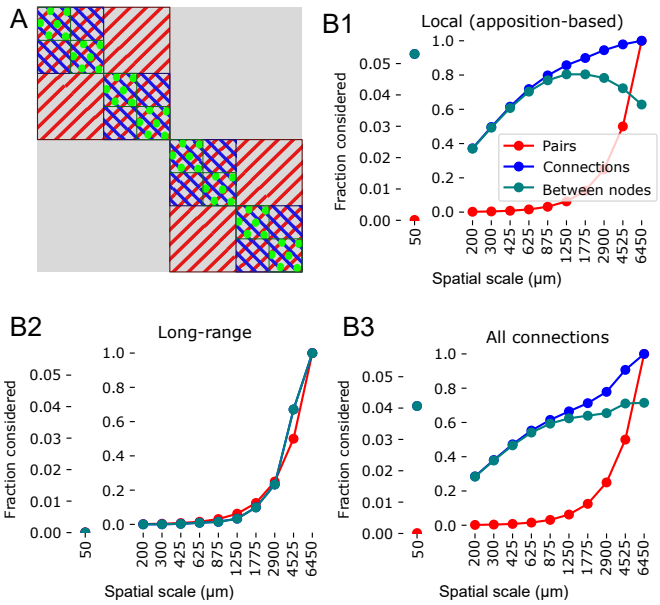

Figure S8: A: Reduction of neuron pairs considered at higher resolutions: At the largest scale, the entire connectivity matrix of the model is considered, albeit at lowered resolution (grey). However, at the next scale, the model is partitioned into two subvolumes whose submatrices are analyzed separately, considering only half of the pairs (red stripes). This continues at each further step (blue stripes; green dots). Neurons within a matrix are partitioned into 512 groups (or *nodes*) and the number of synaptic connections between them are counted. B: Fraction of neuron pairs (red) and connections (blue) considered when analyzing connectivity at various scales, i.e., the area of the submatrices in A compared to the normalized sum of entries within them. Teal: As blue, but only considering connections between nodes, i.e. off the main diagonal of a matrix. Offset to the left: At single neuron resolution (one neuron per nodes); right: at lower resolutions with at least 128 neurons per node. The x-axis specifies the approximate spatial scale over which neurons in a matrix extent. B1-B3: For local connectivity, long-range connectivity and all connections.

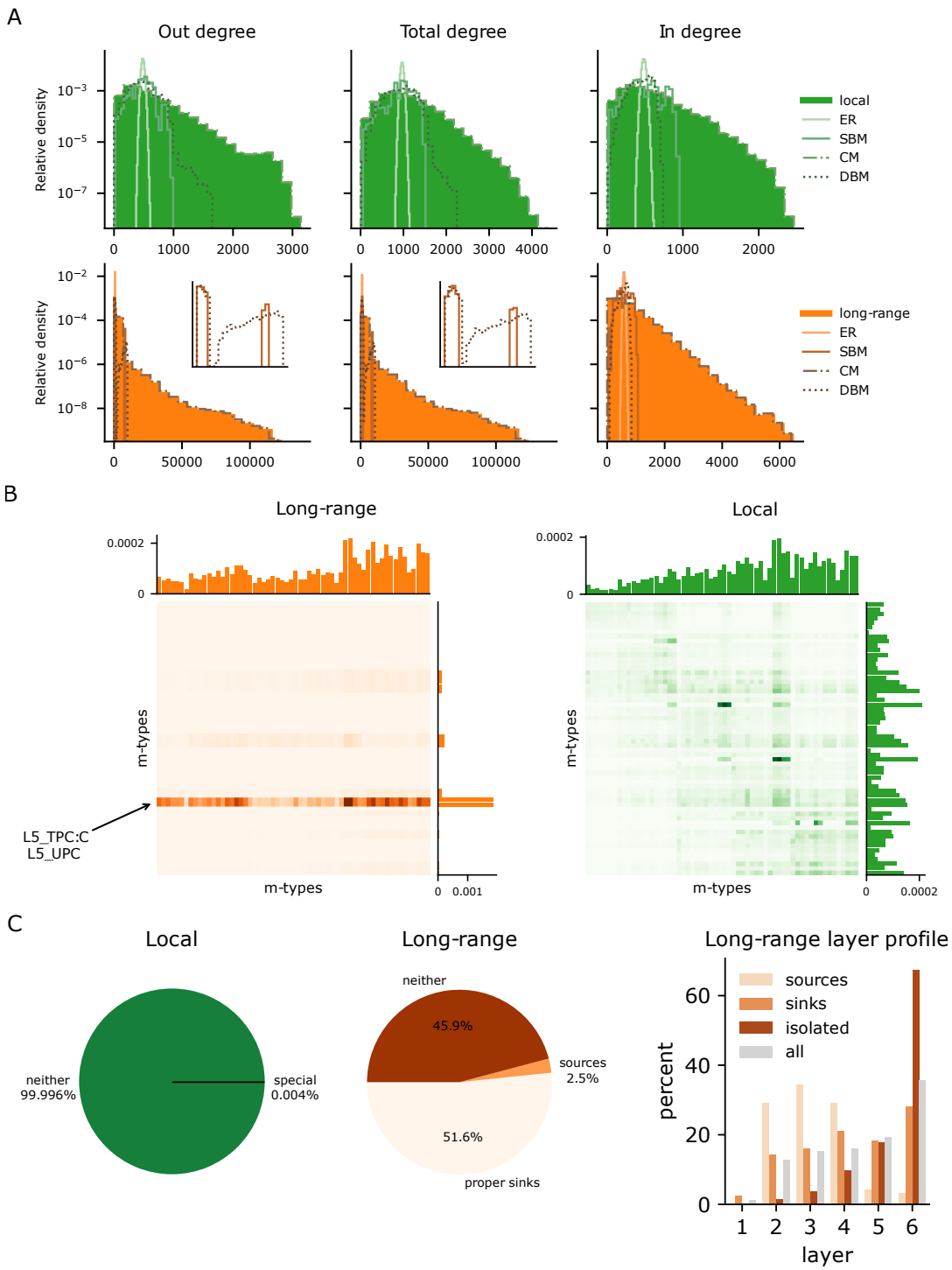

Figure S9: **Basic properties of network connectivity.** A:Top/bottom degree distributions of the local/long-range networks and their corresponding control models. On the second row additional inserts are provided showing the bi-modality of the out and total degree distributions of the SBM and DBM controls. B: Matrices showing the probability of connection between given pre/post-synpatic m-types. Vertically at the right of each matrix the probability of connection for a fixed pre-synaptic m-type. Horizontally on top of each matrix, the probability of connection for a fixed post-synaptic m-type. C: Left and center, percentage of special nodes in both networks, sources (nodes of in-degree 0) and proper sinks (nodes of out-degree 0 that are not isolated nodes). Left: layer distribution of the special nodes of the long-range network.

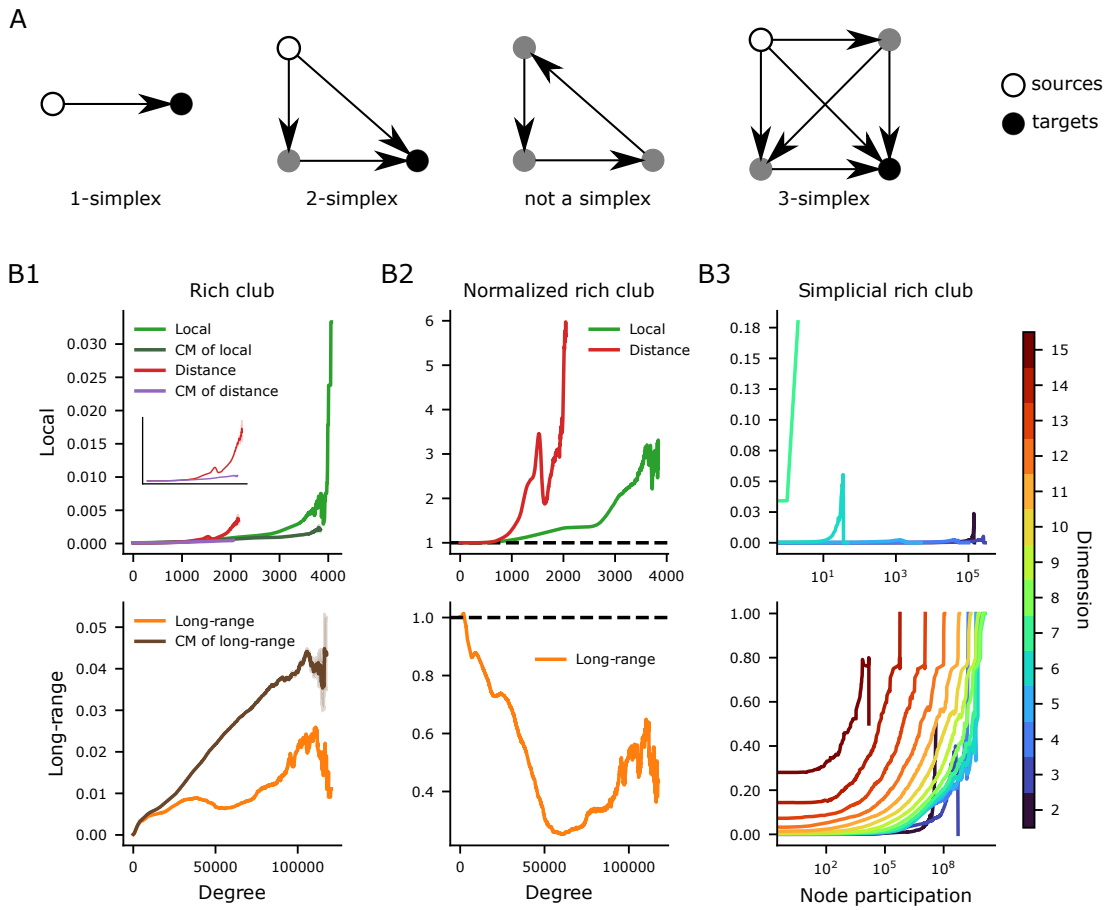

Figure S10: **Rich club analysis** A: Depiction of simplices and non-simplices in various dimensions with distinction between the source and target nodes. B:Rich-club analysis, top/bottom rows for the local/long-range networks. Left column, rich-club curves for the original networks as well as their CM controls. Top left additionally shows the rich-club curves of the DBM controls of the local-networks as well as their corresponding CM controls. Middle column shows normalized rich club curves obtained by dividing the rich-club curves for the original networks by the mean of the rich-club curves of their corresponding controls. Right column shows the simplicial rich club curves for color-coded dimensions ≥ 2.

5.3 Supplementary Methods

5.3.1 Animal surgery

Normal C57BL/6 mice (P52-60) were used which were bred and maintained in the animal house facility of Department of Anatomy, Physiology and Embryology under appropriately controlled conditions (approval of Local Ethics Committee for Animal Research Studies at the University of Debrecen in line with European Union guidelines for the care of laboratory animals, Directive 2010/63/EU).

For initial anesthesia, animals were injected with pentobarbital (0.15ml (6mg/ml), i.p.). Prolonged anesthesia was achieved by injecting 0.05 ml every 60-90 minutes depending on the reaction to the toe-test. Head restraining was used and craniotomy performed in both hemispheres at coordinates ML 1.5 - 2 mm ; AP 0.5 - 1 mm (Paxinos and Watson, 2007) in order to expose the primary somatosensory cortex at the representation of the (hind-limb and fore-limb). We applied the topical anesthetic Lidocaine (Xylocaine gel, Egis Gyógyszergyár ZRT, Hungary) to all surgical wounds and pressure points. A custom-made plastic chamber (prepared by cutting a ring from a 1 ml plastic syringe) was mounted on the skull using super glue and was surrounded by 4% agar (VWR International Kft, Hungary). Then, the cortex was exposed by making a slit on the dura mater with the bent tip of a 14 gauge hypodermic needle.

5.3.2 Single neuron labelling

For this purpose, borosilicate glass pipettes (GB150F-8P, Science Products GmbH, Germany) were pulled (Model P-97, Sutter Instrument Co., USA) with a resistance in the range of 60-80MΩ (bevelled on a BV-10M beveler, Sutter Instrument Co., USA) and filled with 1M KCl containing 2% biocytin (Sigma-Aldrich Chemie GmbH, Germany). The microelectrode was attached to a hydraulic microdrive (MHW-4 Narishige, Japan) and the tip guided in the cortex under the guidance of a surgical microscope (OP-MI, Zeiss). Then the chamber was then filled with 4% agar (VWR International Kft, Hungary) for better stability of the micropipette. Neuronal activity was recorded and amplified with AxoClamp-2A (Axon Instruments Inc., USA). After filtering, the signal was displayed on an oscilloscope and an audio monitor to aid and control intracellular penetration. Successful penetration of the cell membrane was indicated by a sudden drop of the resting membrane potential (below -40mV) while applying 0.05 nA in the step-current mode (put here stimulus configuration from Master-8). Then, biocytin was injected with +2 nA using a duty cycle of 400 msec on and 200 msec off typically for 20 min. In each hemisphere 1-3 penetrations were made with approximately 0.3 mm spacing from each other. Neuronal activity was searched blindly across the entire cortical depth.

5.3.3 Histology

The animals received a lethal dose of anesthetics and were perfused transcardially first with the washing medium (oxygenated Tyrode's solution) for 2 min or until the blood showed clearing and then with a fixative (approx. 100 ml) containing 2% paraformaldehyde (VWR International Kft, Hungary) and 1% glutaraldehyde (Sigma-Aldrich Chemie Gmbh, Germany) in 0.1 M phosphate buffer (PB, pH 7.4) for 50 minutes. Next, the brain was removed from the skull and tissue blocks containing the region of interest were dissected. A series of 60-80 um thick vibratome (Leica VT1000S, Leica Biosystem) sections were cut in the coronal plane and collected in 5x10 lots in glass vials. The sections were washed in 0.05M Tris-buffer saline (TBS, 10 min) and 0.05M TBS containing 0.1% Triton X-100 (2 x 10 min) and incubated at 4°C in avidin-biotin complexed-HRP (ABC-Elite kit, Vector Laboratories, Inc., USA), diluted 1:200 in 0.05M TBS containing 0.1% Triton X-100 (Sigma-Aldrich Chemie Gmbh, Germany) for overnight. Then the sections were washed in TBS for 2 x 10 min and in TB for 10 min. They were treated with 0.05% DAB (3.3' diaminobenzidine-4HCl, Sigma-Aldrich Chemie Gmbh, Germany) diluted in 0.05 M TB containing 0.0025% CoCl2 for 30 min while agitating on an electric shaker. Finally the labelling was visualized in the presence of 0.1% H2O2 (5 min to 10 mins). After washing the sections in 0.05M TBS (3 x 10 min) and 10 min in 0.1 M PB the quality of the DAB reaction and the presence

of intracellularly labelled cells were inspected while wet under a light microscope (x10 objective). All sections of blocks containing strongly labelled neurons underwent osmification, dehydration and resin embedding in order to retain the 3D structure of their axons and dendrites. Accordingly, sections were treated with 1% OsO4 (osmium tetroxide, PI Chem Supplies, USA) diluted in 0.1% PB for 15 minutes. After rinsing in 0.1M PB for 3x10 min, they were dehydrated in ascending series of ethanol (50, 70, 90, 95 per cent and abs ethanol), propylene oxide, each step for 2x10 min, and submerged in resin (Durcupan, Sigma-Aldrich Chemie GmbH, Germany) overnight at room temperature. Finally, the sections were mounted on glass slides, coversliped and cured at 56oC for 24 hours (Somogyi and Freund, 1989).

5.3.4 Morphological neuron reconstruction

Labelled neurons were reconstructed in 3D using the Neurolucida neuron reconstruction system running on Windows XP platform (Neurolucida v.8.23, MicroBrightField Inc., Williston, USA). For this purpose a Leica DMRB microscope (x100 objective) was coupled to a motorized XY-stage and a z-motor via a stage controller (Märzhäuser Wetzlar GmbH & Co. KG, Wetzlar, Germany). Each neuron was reconstructed from 20-32 adjoining sections. Neighboring sections were aligned using the 3-point alignment and the least-squares algorithm for the cut ends of labelled processes and fiducial landmarks such as the contour of cut blood vessels. The cell body, dendrites, axons and axon terminals were reconstructed together with their thickness value.

5.3.5 Generation of random connectivity control models

This section contains definitions of the random connectivity control models and explains how these were generated, presented in weakly increasing order of complexity.

The *Erdős–Rényi* controls (ER) are random directed networks where the edges are added with a fixed probability independently at random. The controls were constructed by taking each ordered pair of nodes (i, j) and adding an edge from i to j at random with probability $p = \frac{E}{N(N-1)}$, where E (respectively N) is the number of edges (respectively nodes) in the original network.

The *stochastic block model* controls (SBM) are random networks where the edges are added independently at random with a fixed probability dependent on the m-type pathway of the pre- and post-synaptic neurons. These controls were built as for the ER-controls, but with a different probability for each pathway. More precisely, if (i, j) are two neurons of m-types A and B, then their probability of connection is

$$p_{AB} = \begin{cases} \frac{E_{AB}}{(N_{AB})^2} & \text{if } A \neq B, \\ \frac{E_{AA}}{(N_{AA})(N_{AA}-1)} & \text{if } A = B, \end{cases}$$

where N_{AB} is the number of neurons of m-type A and B and E_{AB} is the number of edges from neurons of type A to neurons of type B in the original network.

The *directed configuration model* controls (CM) are random networks that closely approximate a given in/out-degree sequence, i.e., a vector of length the number of nodes, whose entries are their in/out-degrees. To build these controls, we encode the edges of the original matrix by two vectors, *sources* and *targets*, such that $(sources[i], targets[i])$ is a directed edge of the matrix (this corresponds to a binary matrix in coordinate format, useful when working with sparse matrices). Then, we shuffle the entries of both vectors *sources* and *targets* independently, which gives a new directed network, with the same degree sequence as the original. This new network might contain loops or parallel edges, so we remove them, thus this construction only approximates the original degree sequence. The density of loops and parallel edges tends to decrease as the number of nodes increases (Newman, 2003), so the approximation is good. Indeed, in our controls for local connectivity, $235,574.0 \pm 376.0$ (mean $\pm$ std) connections out of $2,050,028,490$ were missing, which is less than 0.012% of all connections. For long-range connectivity the number was $629,493.4 \pm 2,228.0$ out of $2,482,296,102$, which is less that 0.026% of all connections.

Finally, the distance block model controls (DBM) are random networks where the edges are added independently at random with a probability $p_A(d) = \alpha_A \cdot e^{-\beta_A \cdot d}$, that is, exponentially decreasing with distance d, where α_A is the probability at distance zero and β_A the decay constant, both depending only on the pre-synaptic cell m-type A. For local connectivity, the distance between a pair of neurons used was the Euclidean distance of their position in space (in μm). For long-range connectivity, the specific method used to originally construct the connectome (Reimann et al., 2019) was taken into account, preferentially connecting neurons at one location within a subregion A to neurons in another subregion B by first parameterizing linear transformations of the flattened coordinates of the two regions. The resulting *virtual coordinate system* $V_{A,B}$ is then used to connect pairs of neurons from A and B in a distance-dependent way. Therefore, the distance considered in this control between pairs of neurons was the Euclidean distance between these virtual coordinates.

For computational reasons, the model coefficients α_A and β_A were estimated by randomly sub-sampling (up to) 100 000 pre-synaptic neurons of m-type A and 100 000 post-synaptic neurons of any type and computing their pairwise distances. For local connectivity, distances from 0 to $1000\mu m$ were then divided into 20 evenly spaced $50\mu m$ bins. For long-range connectivity, distances from 0 to 400 ($a.u.$) were divided into 20 evenly spaced 20 ($a.u.$) bins. Connection probabilities were estimated in each bin by dividing the number of existing connections by the number of possible connections between all pairs of neurons within that bin (excluding connections at distance zero, that is, between one neuron with itself). Model coefficients were then determined by fitting the exponential probability function $p(d)$ to these data points. The whole procedure was repeated ten times with different random sub-samples of neurons, and the averaged model coefficients over these ten estimates were used to build the controls. Overall, the relative standard errors of the mean over the ten estimates were on average across all m-types less than 1%. For local connectivity, model coefficients of all 60 pre-synaptic m-types were found to be on average $\overline{\alpha} = 0.138 \pm 0.102$ (SD) and $\overline{\beta} = 0.0096 \pm 0.002$ (SD) respectively. For long-range connectivity, the model coefficients of all 18 m-types that had any outgoing long-range connections were on average $\overline{\alpha} = 0.0006 \pm 0.001$ (SD) and $\overline{\beta} = 0.012 \pm 0.001$ (SD) respectively.

The code used to generate all the control models is available in the CONNECTOME_ANALYSIS package, and is a combination of python and C++ code.

5.3.6 Finding and counting simplices in connectivity networks

Directed simplices were computed using the custom-made C++ package FLAGSER-COUNT. This code is a variation of the FLAGSER package (Lütgehetmann et al., 2020). The code takes as input the adjacency matrix of a directed network in compressed sparse row (or column) format, outputs the number of directed cliques in the network, and includes the option to print all simplices. There is also the option to output node participation, so for every node v and every dimension d a value is given for the number of d-simplices that contain v. The algorithm considers each node independently as a source node and then performs a depth-first search on that node to find all simplices, where at each step it creates a new simplex by adding any node that is an out-neighbor of all nodes in the current simplex.

Since each source node is considered independently, the computations can be easily parallelized. The simplex count computations were conducted on the Blue Brain high performance computing system. Each computation was run on two Intel Xeon Gold 6248 CPUs using a total of 40 physical cores. To compute the simplex counts for the local network took 3 hours and 15 minutes using 8.2GB of memory, the long-range network took 14 hours and 44 minutes using 9.8GB of memory and the combined network took 115 days and 20 hours using 69.5GB of memory. The transpose adjacency matrices of the long-range and combined networks were used, which have the same simplex counts, because these were significantly faster to compute due to the degree distributions. The computation for the combined network was attempted on the non-transposed matrix and ran for 121 days, and computed partial counts, but encountered 898 nodes for which the counts could not be computed within 24 hours, such nodes were deemed too computational expensive and skipped.

5.3.7 Rich-club and generalized rich-club

The *k-rich-club coefficient* of an undirected network is the density of the subnetwork on the nodes of degree greater than k, and is given by the formula

$$\phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k} - 1)},$$

where $N_{>k}$ is the number of nodes of degree greater than k, and $E_{>k}$ is the number of edges between the nodes in $N_{>k}$. We call the function ϕ the *rich-club curve*. For a directed network we obtain three variations of this curve depending on whether we consider in-degree, out-degree, or total-degree (the sum of in- and out-degree), and we remove the coefficient 2 due to the fact that the maximum number of edges is now $N(N-1)$ instead of $N(N-1)/2$. This gives us three formulas

$$\phi^{total}(k) = \frac{E^{total}_{>k}}{N^{total}_{>k}(N^{total}_{>k} - 1)}, \quad \phi^{in}(k) = \frac{E^{in}_{>k}}{N^{in}_{>k}(N^{in}_{>k} - 1)}, \quad \phi^{out}(k) = \frac{E^{out}_{>k}}{N^{out}_{>k}(N^{out}_{>k} - 1)}, \tag{1}$$

where $N^{total}_{>k}$ (resp. $E^{total}_{>k}$) is equivalent to $N_{>k}$ (resp. $E_{>k}$) but restricted to the total degree, and similarly for in and out.

The rich-club coefficient can be naturally generalized by replacing the degree with any network metric (Opsahl et al., 2008). In particular, we define the *simplicial rich-club coefficient* by

$$\phi^{d}(k) = \frac{E^{d}_{>k}}{N^{d}_{>k}(N^{d}_{>k} - 1)}, \tag{2}$$

where $N^{d}_{>k}$ is the number of nodes with d-node participation at least k, and $E^{d}_{>k}$ is the number of edges between them. Note that $\phi^1 = \phi^{total}$ since a 1-simplex is simply an edge of the network.

A large rich-club coefficient value is said to indicate a preference of high degree nodes to connect together. However, it is shown in Colizza et al. (2006) that the rich-club coefficient is increasing even in Erdős–Rényi networks, and is actually a consequence of higher degree nodes naturally being more likely to connect. Therefore, it is important to normalize the rich-club coefficient by dividing by the rich-club of an appropriate control, that is, a control with the same degree distribution such as a configuration model, see Section 5.3.5. Hence we define the *normalized rich-club coefficient* as

$$\rho^{total}(k) = \frac{\phi^{total}(k)}{\phi^{total}_{rand}(k)}, \tag{3}$$

where $\phi^{total}_{rand}(k)$ is the rich-club coefficient of a configuration model control of the original network, and we similarly define $\rho^{in}(k)$ and $\rho^{out}(k)$. A network is said to have a *rich-club effect* if $\rho^{total}(k) > 1$, as this indicates the high degree nodes connect together more than is "expected". The importance of this normalization can be seen in Figure S9, where B1 seems to indicate the presence of a rich-club effect for the long-range network, which disappears when normalized in B2.

Normalizing the simplicial rich-club is difficult as it requires a control model on directed flag complexes that fixes node participation. Randomly sampling simplicial complexes with fixed simplex counts or other high dimensional network properties is an active area of research (Kahle et al., 2014; Young et al., 2017; Unger and Krebs, 2022), but as of yet there is no known appropriate control model for the simplicial rich-club, and it is a topic that warrants further investigation.

5.3.8 Paths and path distances

Let G be a directed network on n nodes, with adjacency matrix A. A *path* of length k (or k-path) from node u to node v is a sequence of nodes $u = x_1, \ldots, x_{k+1} = v$ such that (x_i, x_{i+1}) is an edge of G for all $i = 1, \ldots, k$. The *path distance* between u and v is the length of the shortest path from u to v. A path from u to v is *geodesic* if its length is the path distance from u to v.

The number of k-paths between all nodes in G can be easily computed using matrix multiplication, since the number of k-paths between u and v is given by $A^k_{u,v}$, where A is the adjacency

matrix of G (Diestel, 2005). Consequently, the number of k-paths from the nodes $i_1, \ldots, i_t$ to all other nodes in G is given by

$$A^k_{[i_1,\ldots,i_t]} = A_{[i_1,\ldots,i_t]}A^{k-1} = A^{k-1}_{[i_1,\ldots,i_t]}A,$$

where $A_{[i_1,\ldots,i_t]}$ is the $t \times n$ matrix obtained from A by taking only the rows $i_1, \ldots, i_t$. Note that the smallest k for which entry i, j is nonzero is the path distance from i to j.

Using this approach we were able to compute the path distances in the combined network between pairs of all 396 nodes in the local and long-range cores. We ran the computation in parallel by partitioning the nodes into 13 sets $C_1, \ldots, C_{13}$ of approximately 23 nodes each and computing

$$A^k_{[C_i]} = (\ldots(A_{[C_i]}\underbrace{A)A)\ldots A}_{\times k-1}).$$

Note that the order of the brackets ensures that at each step a 23×4234929 matrix is computed, rather than a 4234929×4234929 matrix.

Once we had the path distances between the nodes in the local and long-range cores, we computed the geodesic paths between them. This was done using the custom C++ package PATHFINDER. This code functions in a similar way to FLAGSER-COUNT, except that at each step it creates a new path by considering any node that is an out-neighbor of the final node in the current path. At input the start-nodes, end-nodes, and path distances can be specified. To compute all geodesic paths of length at most three between nodes within the local and long-range cores took 8 hours and 47 minutes, using 51.9GB of memory. Between these 396 nodes there are $6,383$ geodesic 1-paths, $1,039,814$ geodesic 2-paths, and $21,701,345$ geodesic 3-paths. To compute all geodesic 4-paths was attempted, but after running for 24 hours failed to finish, and is likely to be computationally infeasible due to the exponential growth of the number of paths.

5.3.9 Generation of columnar subvolumes

We partitioned the modeled volume into subvolumes with shapes approximating hexagonal prisms. To that end, we represented each voxel by the flattened location of its center. In the resulting two-dimensional coordinate system, we distributed seed points in a hexagonal grid with a distance of $460\mu m$ (based on the $230\mu m$ column radius of Markram et al. (2015)) between neighboring points. Next, for each voxel we determined to which seed point it was closest. All voxels that were closest to the same seed point were then grouped together as parts of the same subvolume. As subvolumes near the periphery of the model could be incomplete, i.e., not form complete hexagonal prisms, we excluded them from further analysis if either of the following two conditions was met: 1. The subvolume had fewer than six neighbors in the hexagonal grid; 2. Its volume was less than $0.166mm^3$. That threshold is based on the volume of a circular prism with a radius of $230\mu m$ and a height of $1000\mu m$.

5.3.10 Analyzing connectivity at different spatial scales

To generate connectivity matrices at various spatial scales we first considered the coordinates of all neurons in the model in a flattened, two-dimensional coordinate system, constructed as described in (Bolaños-Puchet and Reimann, prep). We then fed the coordinates into a KD-Tree (Maneewongvatana and Mount, 1999) using the scipy package in python 3.8. This generated a binary tree where each node is associated with a population of neurons as follows: 1. The root is associated with the entire population. 2. The populations associated with the two children of a node N are of a partition of the population associated with N into two groups of roughly equal size. 3. This partition is performed by splitting the population along either their flat x-coordinate or their flat y-coordinate, depending on which has larger variance over the population. 4. This partitioning is continued, growing the tree until populations of only single neurons are reached at its leaves.

As a result of this construction, the set of nodes at a distance n from the root defines a partition of the neuron population into 2^n groups, such that each group is spatially continuous

in the flattened coordinate system and stretches along comparable distances of the flat x- and y-coordinates. We define a distance n_{scale} that sets the spatial scale of connectivity to be analyzed. Each group of neurons defined by n_{scale} will be compiled into a separate connectivity matrix, i.e., connections between groups defined at that scale will not be taken into account.

A connectivity matrix is then generated as follows. For a given node at a distance n_{scale} from the root, we consider the subtree rooted at that node. Within that subtree we find all nodes at a distance of 9 from the root. This defines a partition into $2^9 = 512$ groups of neurons. We then generate the 512 x 512 matrix of the numbers of synaptic connections between these groups. All $2^{n_{scale}}$ matrices at a given scale are then analyzed using the metrics explained in the main text. For large n, we skipped the scales with more than one but fewer than 16 neurons per node, where it is unlikely that neurons from all cortical layers are included.

We estimate the spatial scale and resolution associated to a given scale as follows. We calculate the area of the axis-aligned bounding box of the flat coordinates of all neurons considered in one of the matrices at that scale. We consider its square root to be an estimate of the spatial scale analyzed. In the manuscript, we report the mean of this measure over all matrices at a given n_{scale}, rounded to the nearest multiple of $5\mu m$. The spatial resolution is calculated similarly, but considering instead the bounding boxes of neurons belonging to the same row or column of a matrix.

5.3.11 Calculation of structural measures in the Allen voxelized connectome

We calculated structural metrics (symmetry, degree distributions, maximum weight cutoff that keeps a fully connected core) for the AIBS voxelized connectome in a way that makes the results comparable to our analysis of connectivity at different scales (see above). The following points were to be considered. First, our analysis at different scales split the model along the horizontal dimensions, consequently grouping together neurons from all layers. Second, our analysis resulted in connectivity matrices with a constant size of 512 x 512.

To that end, we applied the flat-mapping procedure of (Bolaños-Puchet and Reimann, prep) to the cortical regions of the AIBS v3 mouse brain atlas with a resolution of $100\mu m$. This projected the voxel into a two-dimensional coordinate system along paths that are roughly orthogonal to layer boundaries. We then discretized those flat coordinates with a resolution of $100\mu m$ and grouped together voxels of the non-barrel somatosensory regions with the same discrete flat coordinate. This resulted in 518 groups, close to the target value of 512. The number of voxels per group was 17.4 ∓ 6.1. We considered the matrix of mean connection strengths between the groups of voxels. The equivalent resolution of this matrix was given by the resolution of the discretization step, $100\mu m$. The equivalent scale was $2600\mu m$.

5.3.12 Generation of tentative clusters based on thalamic innervation

We considered neurons in a central cylindrical volume of the model with a radius of $1000\mu m$. For a given cortical layer and thalamic input source (core-type or matrix-type), we identified for each neuron in that layer the set of input fibers innervating it with at least one synapse. For each pair of neurons, we then calculated the amount of overlap in the thalamic innervation as the size of the intersection, divided by the size of the union of their sets of innervating fibers.

We converted the matrix of pairwise common innervations of a layer into a distance matrix, by taking the negative base 10 logarithm of the measure. Where the common innervation was zero, the distance was set to five. We then clustered the neurons based on this distance matrix using Ward linkage, as implemented in scipy. The maximum number of clusters was set to 100. To evaluate the spatial overlap or separation of the resulting clusters, we also calculated the matrix of pairwise distances of soma locations and then calculated the silhouette score of the clustering based on that distance matrix.

As different layers have very different neuron densities, it would be hard to compare the results across layers. Therefore, we performed the clustering described above on 3000 randomly selected neurons within the layer of interest. That number was selected based on the approximate number of neurons within the $1000\mu m$ region of interest in the least dense layer (layer 1).