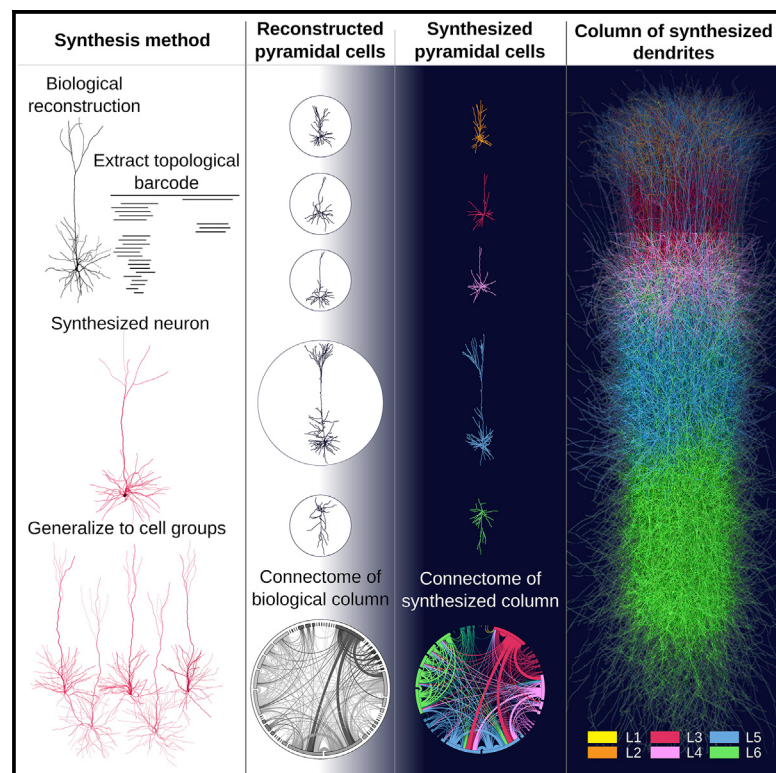


Computational synthesis of cortical dendritic morphologies

Graphical abstract



Authors

Lida Kanari, Hugo Dictus, Athanassia Chalimourda, ..., Julian Shillcock, Kathryn Hess, Henry Markram

Correspondence

lida.kanari@gmail.com

In brief

Kanari et al. present a method to digitally synthesize large numbers of morphologies using topological descriptors of neurons to generate detailed models for healthy and diseased brain regions. This method investigates the links between neuronal morphologies and brain function across a wide range of spatiotemporal scales.

Highlights

- Topological synthesis generates healthy and diseased cortical dendrites
- Synthesized dendrites are indistinguishable from biological reconstructions
- Topological model enables the investigation of the functional roles of cell types
- Links of branching properties and structure of large-scale networks are revealed



Article

Computational synthesis of cortical dendritic morphologies

Lida Kanari,^{1,3,*} Hugo Dictus,¹ Athanassia Chalimourda,¹ Alexis Arnaudon,¹ Werner Van Geit,¹ Benoit Coste,¹ Julian Shillcock,¹ Kathryn Hess,² and Henry Markram¹¹Blue Brain Project, École polytechnique fédérale de Lausanne (EPFL), Campus Biotech, 1202 Geneva, Switzerland²Laboratory for topology and neuroscience, Brain Mind Institute, École polytechnique fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland³Lead contact*Correspondence: lida.kanari@gmail.com<https://doi.org/10.1016/j.celrep.2022.110586>

SUMMARY

Neuronal morphologies provide the foundation for the electrical behavior of neurons, the connectomes they form, and the dynamical properties of the brain. Comprehensive neuron models are essential for defining cell types, discerning their functional roles, and investigating brain-disease-related dendritic alterations. However, a lack of understanding of the principles underlying neuron morphologies has hindered attempts to computationally synthesize morphologies for decades. We introduce a synthesis algorithm based on a topological descriptor of neurons, which enables the rapid digital reconstruction of entire brain regions from few reference cells. This topology-guided synthesis generates dendrites that are statistically similar to biological reconstructions in terms of morpho-electrical and connectivity properties and offers a significant opportunity to investigate the links between neuronal morphology and brain function across different spatiotemporal scales. Synthesized cortical networks based on structurally altered dendrites associated with diverse brain pathologies revealed principles linking branching properties to the structure of large-scale networks.

INTRODUCTION

Neuronal morphologies play a crucial role in governing a neuronal network's dynamical properties, as they influence the function of single neurons (Häusser et al., 2000; Smith et al., 2013; Yi et al., 2017; Schaefer et al., 2003) and constrain the connectivity between neurons (Chklovskii, 2004; Wen et al., 2009; Cuntz et al., 2012). The importance of structural properties of neurons has been shown in multiple fundamental studies that revolutionized our understanding of brain function (y Cajal, 1998; Rall, 1959; Hodgkin and Huxley, 1952). The hand-drawn neuronal morphologies of Ramon y Cajal set the foundations for the first era of *neuronal morphology models*: neuronal shapes represented on paper, initiating the journey to understand the roles of neurons in the brain. A century later, the evidence that the wide variety of neuronal shapes supports the composite functional roles of different cell types is irrefutable, leading to a plethora of projects dedicated to harvesting cellular morphologies from various brain regions and species (Peng et al., 2015; Economo et al., 2016; Gouwens et al., 2019; Benavides-Piccione et al., 2019). A burning challenge of our time is to untangle the still largely unknown roles of distinct morphological cell types.

Recent advances in manual (Economo et al., 2016; Gouwens et al., 2019; Benavides-Piccione et al., 2019) and automatic reconstruction techniques (Peng et al., 2015), as well as the systematic collection and registration of digital morphologies in

standardized databases (Ascoli et al., 2007; Akram et al., 2018; Halavi et al., 2008), have accelerated the acquisition of neuronal reconstructions. These efforts took the study of neuronal morphologies to the next era of *digital reconstructions* (Winnubst et al., 2019; Gouwens et al., 2019). Digital reconstructions of physiologically realistic neuronal networks (Markram et al., 2015; Egger et al., 2014) rely on distinct neuronal shapes of different cell types to approximate biological diversity (Shillcock et al., 2016; Landau et al., 2016; Ramaswamy et al., 2012). We are still far, however, from having enough reconstructions of unique morphologies to populate biologically realistic networks of a whole brain region (1 M neurons for the mouse somatosensory cortex, 10 M neurons for the mouse isocortex) (Ero et al., 2018; Herculano-Houzel et al., 2006), due to the expensive and tedious process of neuronal reconstructions (Farhoodi et al., 2019). This limitation highlights the need to enter a new era for neuronal morphologies, in which *generative models* for digital cells that accurately reproduce the shapes of reconstructed neuronal morphologies should play an essential role.

A crucial obstacle to the computational generation of neurons (Hillman, 1979) is the difficulty in capturing and recreating correlations between morphological features that arise from highly complex intrinsic and environmental factors that influence the propagation and branching of growth cones during neuronal development. Several types of models have been developed in attempts to synthesize cells that reproduce the biological properties of neuronal morphologies. Biophysically accurate models



simulate detailed neural growth by integrating known molecular mechanisms of neuronal development (Zubler and Douglas, 2009). These models focus on the microscopic scale of growth, which is computationally intensive, thereby making difficult the computational synthesis of large numbers of neurons. On the other hand, phenomenological synthesis models are based on either fundamental mathematical principles (Cuntz et al., 2010; Luczak, 2006) or statistical sampling from morphological distributions (Ascoli et al., 2001; Koene et al., 2009). Mathematical models study the effect of different factors on neuronal growth, such as spatial embedding (Luczak, 2006, 2010), minimization of wiring cost (Cuntz et al., 2010), and self-referential forces (Samsonovich and Ascoli, 2003; Memelli et al., 2013). These models require few parameters and provide good intuition about the mechanisms involved in neuronal growth but require manual adjustments for different datasets thereby limiting their ability to generalize to a large number of morphological types. Statistical models (Ascoli et al., 2001; Koene et al., 2009; López-Cruz et al., 2011) generate cells that reproduce distributions of morphological properties from biological reconstructions. However, biologically relevant correlations between morphological features need to be manually identified and implemented in the algorithms (López-Cruz et al., 2011), which renders statistical approaches harder to generalize to diverse datasets. Growth-rule-based morphology models can implicitly account for these correlations, illustrating the need for such rule-based models.

The complications of these models demonstrate the necessity to combine mathematical and statistical properties into a unified synthesis model that circumvents the explicit selection of correlated features while being computationally tractable. Based on principles from algebraic topology and more specifically on the topological morphology descriptor (TMD) (Kanari et al., 2018), we developed a synthesis algorithm, topological neuron synthesis (TNS). Each neuronal tree can be represented with a persistence barcode, which encodes its topological and geometric properties into a mathematical descriptor (see STAR Methods: Topological morphology descriptor). The TMD encodes the start and end of neuronal branches within a tree and provides a simple and elegant definition of branching and termination probabilities that can be used for the generation of artificial neuronal trees (see STAR Methods: Topological neuron synthesis). Complementary features that are not captured in the branching structure, such as soma size, trunk orientation, and dendritic thickness (see supplemental information: Synthesis input), are statistically sampled from biological distributions. Using the TNS algorithm, millions of unique neuronal morphologies were efficiently synthesized (10 M cells in a few hours). A multi-stage validation was then performed to ensure that three modalities of reconstructed neurons are accurately reproduced: the morphological characteristics, the electrical activity of single cells, and the connectivity of the network they form.

The TNS algorithm is sufficiently versatile to be applied to a large variety of cell types without cell-type-specific parameter optimization. Due to the small number of manually chosen input parameters (see STAR Methods: Algorithm inputs parameters; Figure S2), the ranges of these parameters can be chosen consistently across a large range of cell types so that the generated cells are biologically realistic (see Figure S2). We demon-

strated that a small number of exemplar morphologies suffice to capture the diversity of neuronal shapes within a reconstructed population. The adaptability of TNS offers the opportunity to extend the current models of brain regions and study systems that were not accessible so far. As a simple example, we used TNS to generalize the available neuronal reconstructions of the somatosensory cortex (morphological reconstructions published in Markram et al. (2015)) to other regions of the neocortex by varying cortical thicknesses. The synthesized morphologies exhibit structural properties that are in agreement with recent experimental findings (Fletcher and Williams, 2019).

Last but not least, TNS provides a tool to directly investigate the link between local morphological properties and the connectivity of the neuronal network they form. This approach is of particular interest for medical applications, as it enables the investigation of diseases in terms of the emergence of global network pathology from local structural changes in neuron morphologies. We reproduced the effects of dendritic alterations associated to stress disorders (Curran et al., 2017; Dioli et al., 2019; Tornese et al., 2019; Sandini et al., 2020) by implementing two relevant structural changes: the shrinkage of dendritic processes and the loss of dendritic branches, which can be modeled by simple mathematical transformations of the dendritic barcodes. These synthesized networks based on structurally altered dendrites, which are associated with diverse pathologies, revealed principles linking branching properties to the structure of large-scale networks. The TNS algorithm introduces a new era in the study of neuron morphologies, in which millions of biologically accurate synthetic cells can easily be generated and computational models of single cells and networks can be constructed in a principled manner.

RESULTS

The morphological development of neurons in the brain is a complicated process that depends on both genetic and environmental components and differs between species, brain regions, and morphological types. In this study, we focus on the computational synthesis of dendrites and thus will not consider axon path finding. While biological development is not simulated, biological principles of morphological growth inform the design of our computational algorithm, which synthesizes dendritic morphologies.

The three main modalities that we aim to reproduce with synthesized cells are morphology (Scorcioni et al., 2008), electrophysiology (Van Geit et al., 2016), and connectivity (van Pelt et al., 2010; van Pelt and van Ooyen, 2013). To ensure that the synthesized cells recreate all three properties, the microcircuit that was published by Markram et al. (2015) was used as reference. This microcircuit was built based on reconstructions of juvenile rat neurons across all cortical layers for a variety of morphological types. First, we ensure that the morphological and electrical properties of the reconstructed cells are reproduced by neurons whose dendrites have been synthesized with the TNS algorithm. Finally, a network is built from neurons with synthesized dendrites, and compared with the connectivity of the Markram et al. (2015) network (referred to as the reconstructed network).

Algorithm 1. Synthesizer

Input:
 $D_{soma} = \{D_{ss}, D_{nn}, D_{pa}\}$ $\triangleright$ (see [Biological distributions](#))
 $D_{diam} = \{D_{tips}, D_{tr}, D_{rr}, D_{md}\}$ $\triangleright$ (see [Biological distributions](#))
 $Param = \{c, \tau, \rho, step-size\}$ $\triangleright$ (see [Input parameters](#))
 $Barcodes$ $\triangleright$ (see [Biological barcodes](#))
function SAMPLE distr : = draws from input distribution
Generate a Soma and Neurites using (Alg 2, D_{soma}, c) $\triangleright$ (each neurite is initialized with a point on the soma surface, which also defines an initial direction dir_1)
for neurite in Neurites **do**
 $TMD = \text{SAMPLE } Barcodes$
 Sort $bars$ in TMD from longest to shortest
 Initialize first section sec_1 with the longest bar_1
 $Active \leftarrow sec_1$
 while Active sections **do**
 for Section sec_k in Active **do** $\triangleright$ (a section gets a target direction dir_k and a bar bar_k)
 Grow a section using (Alg 3, $dir_k, bar_k = (b_k, d_k, a_k)$)
 Remove bar_k from TMD $\triangleright$ (each bar can be used only once)
 if status = Bifurcate **then**
 Generate children using (Alg 4, TMD, bar_k, dir_k)
 Add children to Active sections
 else if status = Terminate **then**
 Section growth terminates
 Remove current section sec_k from Active
Generate diameters using (D_{diam})
Output:
A neuron: set of points, diameters and their connectivity.

Dendritic synthesis algorithm

Advances in experiments and mathematical models have converged on a set of commonly accepted stages of morphological growth: the initiation of neurites, neurite elongation, axon path finding, and neurite branching (Cuntz et al., 2007; Graham and van Ooyen, 2006). The TNS algorithm (see [Algorithms 1–4](#)) consists of three main components ([Figure 1A](#)): the initiation of dendrites on the soma (see [STAR Methods: I. Initiation of neurites](#)), dendritic branching (see [STAR Methods: II. Branching/termination](#)), and dendritic elongation (see [STAR Methods: III. Elongation of neurites](#)). The neuronal soma is generated based on a radius sampled from a biological distribution. Then, the number of dendrites (basal and apical), is sampled from the biological distribution of the corresponding cell type. Each neurite is

assigned an initial orientation and a barcode based on the data of the respective morphological cell type.

Subsequent steps of the growth take place in a loop. Each branch of the tree is elongated step by step, as a combination of the following components: a random direction, ρ , the target orientation, τ , and a memory of the previous steps, μ ([Figures 1A–1III](#); see [STAR Methods: III. Elongation of neurites](#)). At each step, the growing tip is assigned probabilities to bifurcate, to terminate, or to continue, which depend on the path distance from the soma and are defined by the bars of the sampled barcode ([Figures 1A–1II](#); see [STAR Methods: II. Branching/termination](#)). A new branch is attached to a current branch only if its bar is contained in the other, also known as the Elder rule in topology (see [Kanari et al., 2020](#)). Each bar of the barcode can only be used once. The growth terminates when all the bars of the input barcode have been used. Finally, as an independent step, the diameters of the tree are assigned based on diameter distributions sampled from the reconstructed cells ([Figures 1A–1IV](#); see [STAR Methods: IV. Generation of diameters](#)). The details of the algorithm are described in the [STAR Methods](#).

Single neuron synthesis

First, we focus on single-cell synthesis: starting from a single input morphology we generated synthesized examples. Each cell is validated in terms of morphological and electrical properties with an extensive number of features, examples of which are presented in ([Figure 2](#)). The morphological and electrical properties of synthesized and reconstructed morphologies ([Figures 2C and 2H](#)) are normalized according to [Equation 5](#).

Algorithm 2. Generate soma - neurites initialization

Input:
 $c, D_{soma} = \{D_{ss}, D_{nn}, D_{pa}\}$ $\triangleright$ (see [Biological distributions](#))
function SAMPLE distr : = draws from input distribution
 $d_s = \text{SAMPLED}_{ss}$ $\triangleright$ (soma size)
Soma is a sphere of diameter d_s and center c : $S_{d_s}^c$
 $nn = \text{SAMPLED}_{nn}$ $\triangleright$ (number of neurites)
Sample random unit vector $Vect_1$ of neurite N_1 on $S_{d_s}^c$ surface.
First point of N_1 is $x_1^1 = c + Vect_1$
for Neurite ($N_i | 2 \leq i \leq nn$) **do**
 $Vect_i = Vect_{i-1} + \text{SAMPLE } D_{pa}$
 First point of N_i is $x_i^1 = c + Vect_i$
Output:
A contour of soma points on $S_{d_s}^c$
Initial points x_i^1 for each neurite N_i

Algorithm 3. Elongate section

Input:

```

 $\tau, \rho, dir, bar_k = (b_k, d_k, a_k), x_0$ 
 $\mu = 1 - \tau - \rho$   $\triangleright$  (Normalization of weights to 1)
function PD point : = path distance of point from soma surface
 $n = 1$ 
status = Continue
while status is Continue do
    L: length, R: unit vector randomly sampled
     $M = \sum_{i=1}^5 \exp(1-i) * v_{(k-i)}$ 
     $v_k = \rho * R + \tau * T + \mu * M$ 
     $x_{n+1} = x_n + L * v_k$ 
    rand  $\in [0, 1)$  randomly sampled
    if rand  $\leq \exp(\lambda * (PDx_{n+1} - b_k))$  then
        status = Bifurcate
    else if rand  $\leq \exp(\lambda * (PDx_{n+1} - d_k))$  then
        status = Terminate
    else
        status = Continue

```

Output:

Section points and status: Bifurcate or Terminate.

Morpho-electrical properties

A set of 100 synthesized cells was generated, based on the morphological properties (persistence barcodes and morphometrics) of a selected reconstruction of an L3_TPC (Figure 2A). First, the topology of the synthesized cells was validated by comparing the radial-persistence diagram of the reconstructed to the synthesized cells (Figure 2B). Subsequently, the morphometrics of the input reconstructed cell and a synthesized cell were compared using the Fnorm (see STAR Methods: Single-cell validation) against a reference population of L3_TPC cells (Figure 2C). In addition, the electrical model optimized on a population of L2/3 pyramidal cells (Van Geit et al., 2016) was applied to the reference reconstructed cell and a synthesized cell. Similar to the morphological validation, the Fnorm was used to quantify how well the resulting morpho-electrical combination matches with the statistics of the original experimental data for the 120% threshold current step amplitude (Figure 2F–2H). Their electrical traces in response to the tested stimuli are presented in Figures 2F and 2G and the Fnorm scores of electrical features in Figure 2H.

Due to the stochastic component of the growth process, the synthesized cells are not identical to the reconstructed cell (Figures 2A and 2B), yielding the desired morphological diversity.

Nevertheless, the morphological and electrical properties of reconstructed and synthesized cells are statistically close to each other's, as seen by the normalized differences in Figures 2C and 2H.

Neuronal population synthesis

To reproduce the biological variability, it is essential to generalize the results of the previous section by using as input a large number of biological reconstructions. The algorithm is generalized to a neuronal population by randomly sampling a persistence barcode extracted from the reconstructed population until all dendrites of a cell are grown. The growth of each dendrite is assumed to be independent. We generated synthetic cells for all the rodent cortical cell types reported in previous publications (Markram et al., 2015; Kanari et al., 2019; Marx and Feldmeyer, 2013), using as input a set of reconstructions from the BBP dataset.

Comparison of reconstructed and synthesized dendritic shapes

Morphologies of the two major types of cortical neurons are generated: excitatory and inhibitory cells. Excitation is mainly mediated by pyramidal cells, with the exception of the spiny stellate cells (L4), and use glutamate as a neurotransmitter. Inhibition is mediated by interneurons, which use GABA as a neurotransmitter. Pyramidal cells (PCs) represent the majority of neurons in the cortex $\approx 85\%$ (Gonchar et al., 2007; Lefort et al., 2009). The wide variety of apical dendrite shapes imparts unique functional properties to PCs and forms the basis for integrating signal inputs from different cortical layers (Larkum et al., 2007, 2009; Spruston, 2008). Eighteen pyramidal cell types from layers 2 to 6 (identified in [Kanari et al., 2019]) were synthesized (Figure 3A, in red) and compared with their biological counterparts (Figure 3A, in blue).

Interneurons represent about 15% (Gonchar et al., 2007; Lefort et al., 2009) of neuronal cells in the cortex. There are many interneuron types, which are distinguished mainly by their axonal shapes, as they do not present large variability in their basal dendrites. Interneurons from different layers were synthesized (Figure 3A, in red) and compared with reconstructed interneurons from the same layers (Figure 3A, in blue). The TNS algorithm generates cells that reproduce the characteristic shapes of pyramidal cell types and the variability of interneuron dendrites.

A cortical column consists of a combination of interneurons and PCs. By putting together the synthesized dendrites of

Algorithm 4. Bifurcate

Input:

```

TMD, dir_k, bar_k = (b_k, d_k, a_k)
function SPLIT vect, a : = returns two unit vectors dir_1, dir_2 according to the input vect and a set of angles a
     $\triangleright$  (see section Branching - Termination)
 $dir_1, dir_2 = \text{SPLIT}(dir_k, a_k)$ 
Find next available indices i in TMD for which  $\min(b_i)$  and  $d_i \leq d_k$ 
Generate child sec_1:  $\leftarrow x_n^k, dir_1, bar_1 = (b_i, d_k, a_i)$ 
Generate child sec_2:  $\leftarrow x_n^k, dir_2, bar_2 = (b_i, d_i, a_i)$ 

```

Output:

Two new sections, each initialized with x_0 , dir and bar.

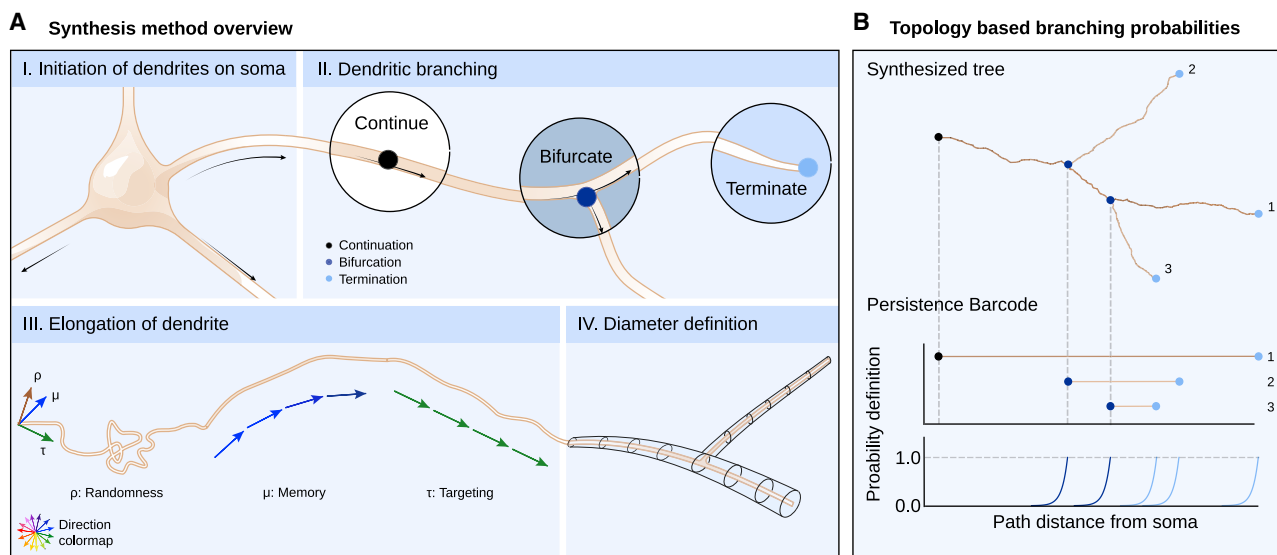


Figure 1. Dendritic synthesis algorithm

(A) Overview of dendritic synthesis based on four stages of growth. (I) Soma generation and initiation of the dendrites on the soma surface. (II) Stochastic definition of bifurcation, termination, and continuation based on topological descriptor (B). (III) Dendritic elongation: during continuation the branch grows based on a segment length and direction. The direction is chosen as a combination of three parameters: randomness, memory (based on the previous directions within a branch), and targeting (based on the initial direction of a branch). (IV) Diameter definition, as a final step, is based on the biological distributions and is subsequent to the branching steps.

(B) Branching based on the topological morphology descriptor (TMD) of a neuronal tree: the probability to bifurcate, terminate, or continue depends on the path distance from the soma and the joint probabilities derived from the TMD of a neuronal morphology. The bifurcation probability increases as the start of the bar in the TMD, and the termination probability increases as the end of the bar is approached. Respective probabilities become one after the start, and end of the bars in TMD. Note that each new bar has to be smaller than its parent, so the growth is performed from larger to smaller branches.

different types with the reconstructed axons, respecting the cell type composition (Figure 3C, only dendrites shown), we generated a rodent cortical column of synthesized neurons that exhibits the basic qualitative features, such as layer structure, that are observed in the reconstructed cortical column in Markram et al. (2015).

Morphological validation of synthesized dendrites

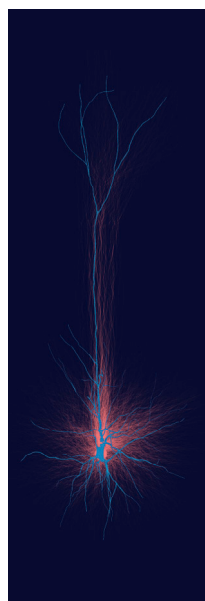
Detailed validations were performed to ensure that the TNS algorithm reproduces the properties of a population of reconstructed cells. As an example, the synthesized cells of type L5_TPC:A were validated against the reconstructed cells using a large set of morphometrics (Ascoli et al., 2008; Scorcioni et al., 2008), a subset of which is presented in Figure 4B for the apical (top) and basal dendrites (bottom). The statistical distributions of the morphological features of the synthesized cells closely match the distributions of the reconstructed cells for both the apical and the basal dendrites.

Reconstructed cells are often not complete and contain artifacts due to the reconstruction techniques (Conde-Sousa et al., 2016; Farhoodi et al., 2019). As a post-processing step, several corrections have been applied before the morphologies can be used by the TNS algorithm, such as unraveling of tortuous branches and repair of out-of-plane cut branches (Markram et al., 2015). Nevertheless, some dendritic trees may still have artifacts not compatible with TNS, such as being too small (i.e., having only one bifurcation point), due to incomplete reconstruc-

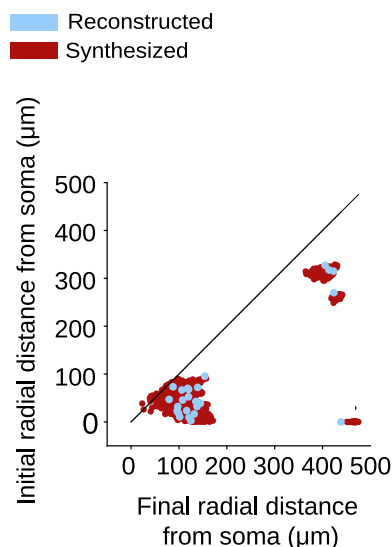
tions that cannot be recovered. Dendrites that do not comply with the quality control measurements (e.g., the number of branches should be larger than 2) are not included in the set of barcodes that are used as input for synthesis, resulting in small discrepancies between reconstructed and synthesized cells. As a result, some morphometrics (such as total number of bifurcations) differ between synthesized and reconstructed cells (see Figure 4B, bottom). For apical dendrites the same morphometrics are well reproduced (see Figure 4B, top), since no apical trees were excluded.

To further assess the statistical agreement of morphological features we perform a detailed validation between the population of reconstructed and synthesized cells (see Figure S8A; Tables S2, S3, and S4) using the DBM/OVS statistical score, MVS for short (see STAR Methods: Population-to-population validation). For each morphological feature, the MVS score provides an estimate of the difference between the median values of two distributions (of reconstructed and synthesized cells), divided by an estimate of the overall spread of the distributions (see Equation 6). For m-types with five or more available reconstructions (see Figure S8A), the algorithm generated cells that are statistically close to the input population. For cell types with few available reconstructions (fewer than five), the small number of input cells did not suffice to judge the quality of the synthesized cells. The minimum requirement for the sample size of the synthesis input population is further investigated in “Versatility of the synthesis algorithm.”

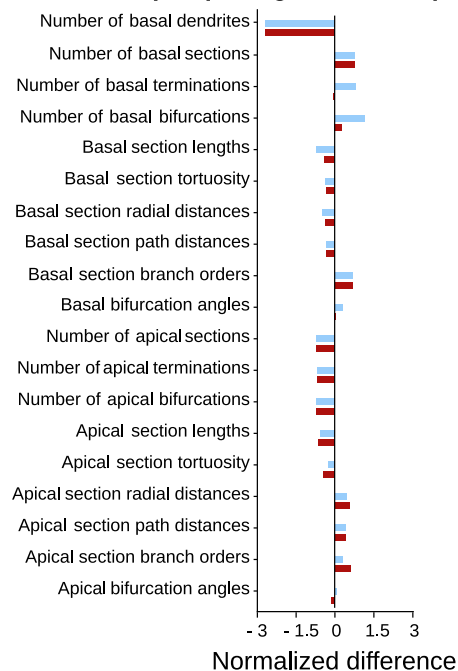
A Layer 3 Tufted Pyramidal Cells



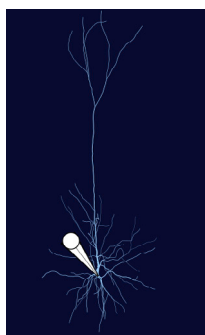
B Persistent diagram



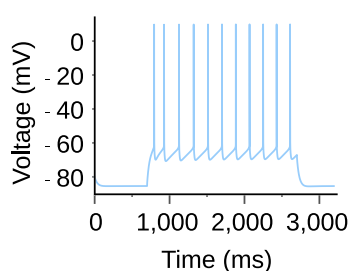
C L3 TPC (morphological features)



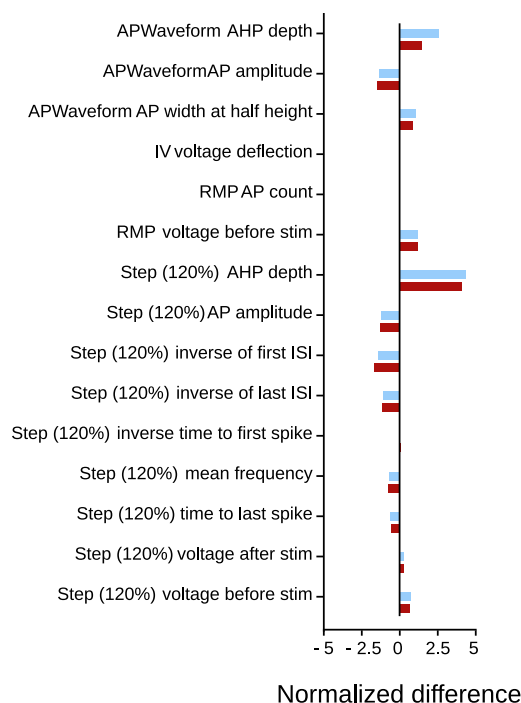
D Reconstructed



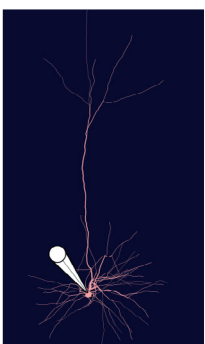
F 120% threshold current



H L3 TPC (electrical features)



E Synthesized



G 120% threshold current

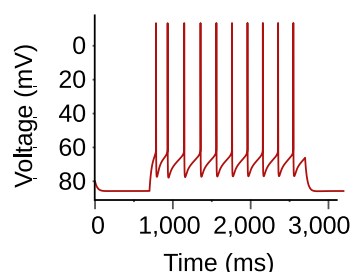


Figure 2. Validation of single-cell morpho-electrical properties

(A) Reconstructed layer 3 tufted pyramidal cells (blue) is used as input for 100 synthesized L3_TPCs (red).

(B) Comparison of topological persistence diagrams of the reconstructed cell and 100 synthesized cells.

(legend continued on next page)

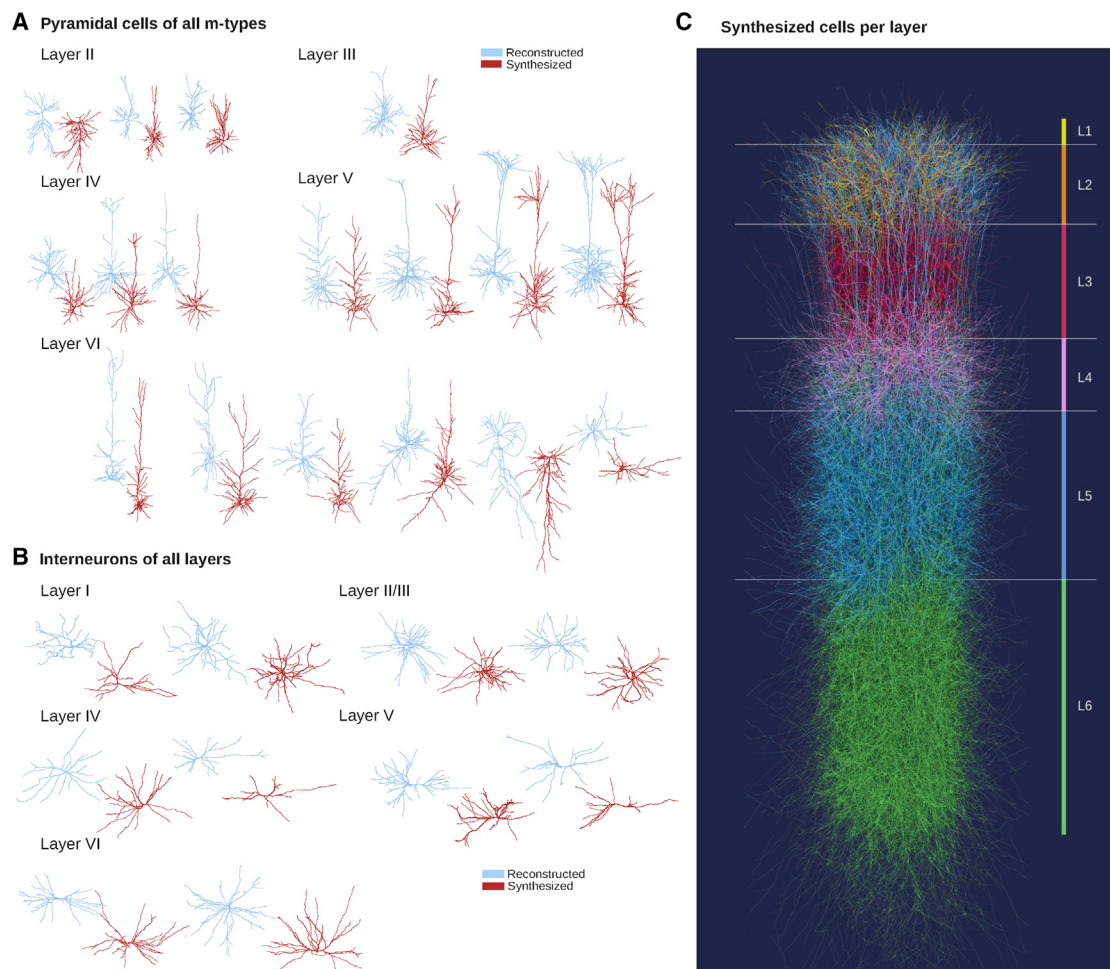


Figure 3. Comparison of reconstructed and synthesized dendritic shapes

(A) Reconstructed (blue) and synthesized (red) pyramidal cell dendrites of all rodent cortical m-types from layers 2 to 6.

(B) Reconstructed (blue) and synthesized (red) dendrites of rodent cortical interneurons of layers 1 to 6. Not all interneuron morphology types are reported, as they differ mainly in their axonal branches and not significantly on the basal dendrites, as illustrated.

(C) A cortical column of synthesized dendrites of all layers, colors correspond to cortical layers from 1 to 6.

Electrical validation of synthesized dendrites

In addition to reproducing the characteristic shapes of biological reconstructions, the synthesized cells are expected to produce similar electrophysiological responses to the reconstructed neurons. This guarantees that electrical models optimized with a single cell (Markram et al., 2015; Van Geit et al., 2016) generalize to synthesized cells as well as generalizing to reconstructed cells of the same m-type. To test the generalization of electrical models on synthesized cells, we used the set of electrical models optimized based on reconstructed cells, according to the principles described in Markram et al. (2015) and Van Geit et al. (2016) for ten different rat electrical types (e-type). The 100 synthesized

cells per m-type that were generated and morphologically validated in the previous section, have been simulated based on the optimized electrical models (Van Geit et al., 2016). The features of the electrical traces produced by these simulations were compared for all reconstructed and synthesized cells for a combination of m-types and e-types (me-types).

The results of this analysis are presented in Figure 4. For layer 5 PCs (Figure 4D, reconstructed; Figure 4F, synthesized) the cADpyr model was used. Two typical traces of layer 5 PCs, the step current and the back-propagating action potentials are shown in Figure 4E for ten randomly selected reconstructed (blue) and synthesized (red) cells. The traces of the cell used to

(C) Comparison of 19 dendritic morphometrics (normalized based on the mean morphological feature values for the L3_TPC population) for a reconstructed and a synthesized cell.

(D and E) The reconstructed (D) and synthesized cell (E) are electrically simulated according to a model optimized on the electrical properties of L3_TPC cells.

(F and G) The electrical response (120% threshold current step) of the reconstructed cell (F) is compared with the synthesized cell's (G).

(H) Comparison of 15 electrical properties of dendrites (normalized based on the mean electrical feature values for the L3_TPC population).

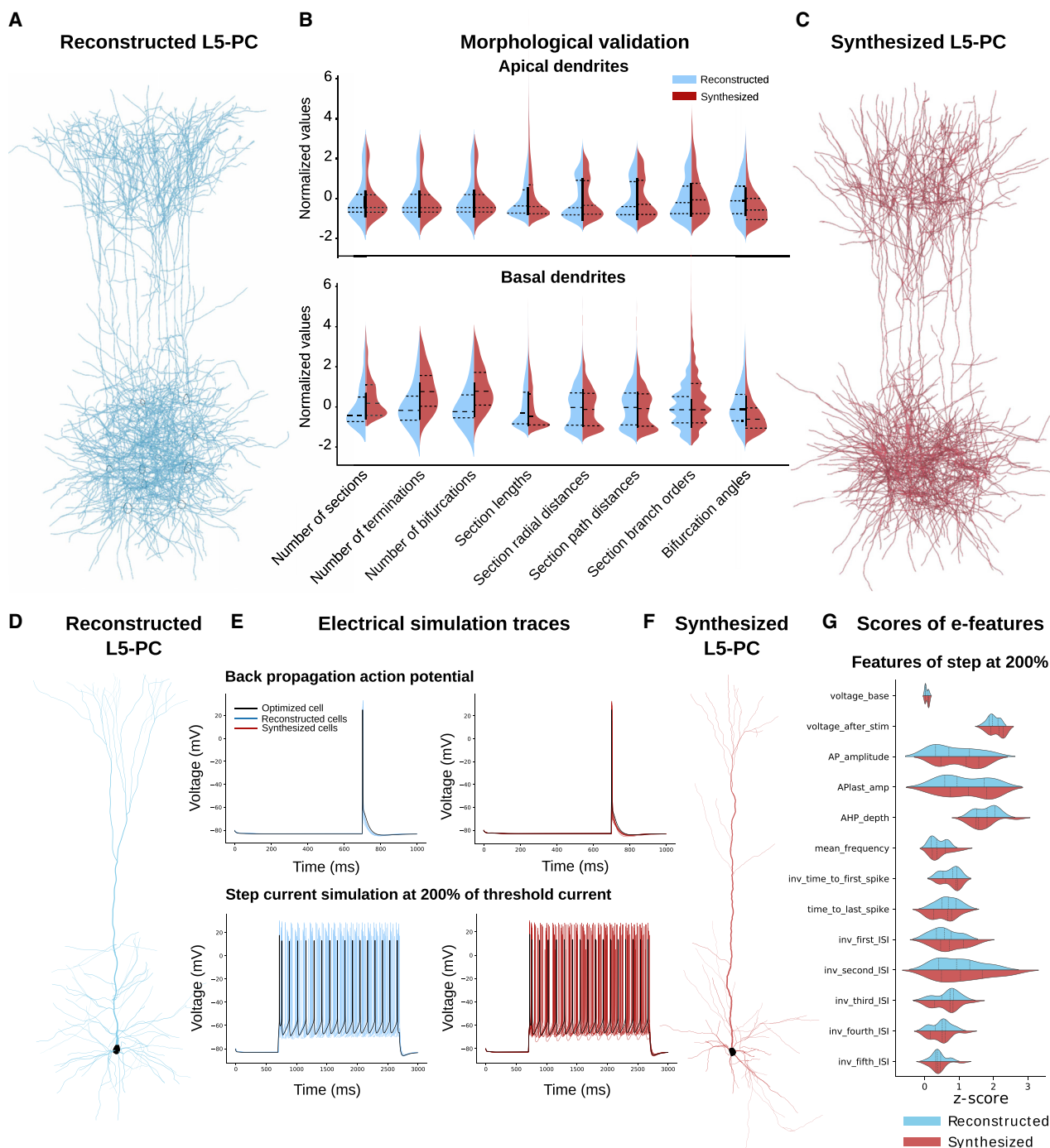


Figure 4. Morphological and electrical validation of synthesized dendrites

(A–F) A set of L5_TPC:C reconstructions (A) (blue, 30 cells) is used as input to generate a population of synthesized cells of the same type (C) (red, 100 cells). The violin plots of morphological properties (B) for apical (top) and basal (bottom) dendrites of the reconstructed cell (in blue) and the synthesized cells (in red) are reported. Electrical traces (E) from simulation of ten reconstructed (D) (blue) and ten synthesized morphologies (F) (red) are compared with the reference trace for the optimized model (black). Step current simulation at 200% of threshold current illustrates similar frequency in firing patterns of synthesized and reconstructed cells. Back-propagation action potential illustrates similar spike shape between reconstructed and synthesized morphologies. (G) Validation of the electrical features extracted from the traces of step current simulation (E) of the reconstructed (blue) and the synthesized (red) cells using Z scores with respect to experimental features.

optimize this electrical model are shown as a reference (in black). Several electrical features are extracted from each trace and compared against a distribution of features from experimental recording using a Z score (Figure 4G). All the Z scores of a given cell obtained under various protocols provide a robust indication of the quality of the electrophysiological behavior of the cell (Markram et al., 2015; Van Geit et al., 2016).

To assess the overall quality of the electrical simulations, the MVS score, as defined for the morphological features (STAR Methods: Population-to-population validation) is applied for the comparison of the electrical feature distributions of reconstructed and synthesized cells (Figure S8B). The synthesized cells perform as well as the reconstructed cells in the electrical simulations. However, some m-types with only few example reconstructions result in several high-scoring features for a subset of e-types (see also comment on morphological validation). The statistical results of the electrical validations are presented in detail in Table S5 and Figure S6A. In addition, we cross-validate the electrical simulations by applying the same e-models on two distinct populations of neurons (reconstructed and synthesized). We demonstrate that the synthesized cells are sufficiently similar to reconstructed cells to reproduce the variance and differences for two e-models (L3 and L5 PC) applied on two distinct populations of L3 and L5 pyramidal cells (see Figure S6B).

Morphological feature correlations

Correlations between morphological features are reportedly essential for any synthesis method (López-Cruz et al., 2011). However, do the inputs to TMD-based synthesis suffice to account for feature correlations, or do we need to define them explicitly? In previous studies, the explicit description of correlated morphometrics was either obtained manually (Koene et al., 2009; van Pelt and van Ooyen, 2013), with the lurking danger that different neuronal datasets might require different feature correlations, or optimized with complex algorithms (López-Cruz et al., 2011), which entails the risk of over-fitting when only a few reconstructions are available. The risk is that, instead of capturing the biological principles of neuronal morphologies, the algorithm might overestimate local and noisy properties.

The inputs for the TNS are defined by the topological barcodes of neurons. Barcodes encode path distances in the neuronal tree, but values such as radial distances, asymmetry, and the precise connectivity of the tree are not encoded in the barcodes. To ensure that the synthesis algorithm reproduces inter-dependencies between features, it is necessary to validate the distributions of the features that have not been explicitly used as inputs for the synthesis algorithm. However, as seen in the thorough statistical validation of morphometrics, synthesized cells do not differ from the biological reconstructions for an extensive list of morphological features, most of which were not direct input to the algorithm.

A necessary condition for this is the coupling between bifurcation and termination probabilities that is encoded in the barcode structure. If bifurcation and termination probabilities in reconstructed cells were in fact independent, this algorithm would suffice to reproduce the branching patterns of the neuronal morphologies (Luczak, 2006; Cuntz et al., 2010). The TNS algorithm

was modified to sample independently bifurcation and termination probabilities and bifurcation angles, instead of the original bars (see Figure S4). The TNS algorithm was modified to investigate the impact of using marginal branching probabilities instead of joint branching probabilities through persistence barcodes. To do this, we use the information encoded in the persistence barcode (b_i , d_i , a_i) as independent variables. Cells were synthesized by decoupled start and end of barcodes (Figure S4III), by decoupling start distance and branching angles (Figure S4IV). As shown in Figure S4, the synthesized cells generated by these versions of modified TNS algorithm are significantly different from the original reconstructions, indicating that the coupling between bifurcation and termination probabilities provides a necessary condition to reproduce correlations between morphological features.

We have thus demonstrated that the persistence barcodes encode the relevant information about the biological branching structures required for the accurate generation of dendritic shapes. In addition, the links between the bifurcation, the termination, and the respective bifurcation angles, are essential for synthesizing biologically accurate cells. The persistence barcodes are therefore a necessary constraint on the synthesis inputs.

Morphological diversity

Another challenge in synthesis is the sparsity of input data for many cell types, which makes it difficult to reproduce the morphological diversity of neurons. If few biological reconstructions are available (fewer than five cells), it is impossible to reproduce the properties of this cell type in its *in vivo* conditions. Using groups of cells with a large number of available reconstructions, such as PCs of layers 3–5, we investigated how the number of input cells influences the diversity of synthesized morphologies in Figure S5.

In particular, we compared the synthesized distributions of fundamental morphological features, such as path distance, branch order, and radial distance for increasingly larger subpopulations of 44 L4_TPC reconstructed cells. We identified the minimum number of cells that is required as input for synthesis to approximate well the morphological features of the original reconstructed population. While a sample size of ≤ 10 cells was not sufficient to approximate the diversity of the reconstructed cells with respect to our choice of morphometrics, with ≥ 15 (one-third of the original dataset), both input and emergent morphometrics were well reproduced (see Figures S9A, S9B, S9D, and S9E). Note that, since morphological features are intrinsically related to each other, we cannot define purely emergent properties, but rather properties that were not directly used as input to the algorithm. In addition, we computed the average KS distance between the distributions of reconstructed and synthesized cells for all features (Figure S9C) to confirm that the KS distance is minimized for both basal and apical features with more than 10–15 input cells.

A generalization of this result to all m-types (see Table S4) showed that PCs require about ≈ 10 –15 samples, while interneurons require ≈ 5 –10 cells to approximate the distributions of their dendritic morphometrics with synthesized cells. Note that this is a generalization as for each specific cell type the

optimal number of reconstructions differs. We computed the statistical score of the difference between reconstructed and synthesized cells for a large number of morphological features for the cells types with more than 15 biological reconstructions by varying the number of input cells: 2, 5, 10, 15, 20, and 30 (Figure S5). These statistical scores are compared against the reference statistical score of synthesized and reconstructed cells, when all the available reconstructions were used as input. The minimum number of cells that is required to approximate the population of reconstructed cells with our synthesis algorithm was identified. Note that, due to the computational complexity of this experiment, we only synthesized a population of ten cells per use case. This leads to the observed differences between the scores reported here and in population validation scores. The optimal number of required reconstructions to approximate a biological population depends on the cell type but, in general, cells with apical dendrites require larger numbers of reconstructions due to the complexity of the biological trees.

Finally, we compared the classification accuracy for three layer 4 pyramidal cell types (L4_TPC, L4_UPC, and L4_SSC) for reconstructed and synthesized cells (Figure S5F). The three cell types were synthesized using the same input parameters, and the persistence barcodes of the three respective groups of PCs. The accuracy was computed by iteratively training a decision tree classifier on all the data except one (leave-one-out [LOO] method) and comparing the original label to the one proposed by the trained classifier (the sk-learn Python package was used [Pedregosa et al., 2011]). The same method (LOO with decision tree classifier) was used for the classification of synthesized cells (Figure S9F, bottom right). The accuracy of the classification of synthesized cells (trained on synthesized cells) was 81% and is therefore higher than the accuracy of reconstructed cells (trained on reconstructed cells, 70%) (Figure S9F, top left). In addition, we performed cross-classifications, the accuracy of synthesized cells (classifier trained on reconstructed cells) was 70%, while the LOO accuracy of reconstructed cells (classifier trained on synthesized cells) was 66%.

Versatility of the synthesis algorithm

As shown above, TMD-based synthesis reproduces the morphological properties of reconstructed cells while preserving the diversity within morphological classes. However, to synthesize large brain regions (Fletcher and Williams, 2019) it is not sufficient to generate cells that reproduce an input population. A simple example to illustrate the limitations of this approach is the generation of neurons within brain regions with varying anatomical properties, such as cortical thickness, which is not constant within the neocortex.

Statistical and biophysical models are often limited to reproducing a specific input population, whereas mathematical models are adaptable to a variety of different conditions. The TNS algorithm, which combines both approaches, can be modified by appropriate mathematical transformations applied to the persistence barcodes. In particular, operations such as scaling, rotation, and sub-sampling of barcodes can be implemented to generate a variety of different dendritic shapes from the original reconstructed morphologies.

To illustrate the versatility of the TNS algorithm, cells for varying cortical thicknesses were synthesized by scaling the persistence barcodes according to different percentages, from 10% to 100% of the original size in Figure 5. The resulting synthesized cells retain the topology of the input cells while respective morphometrics, such as total lengths (Figures 5C–5E), are scaled accordingly.

Connectivity of synthesized and reconstructed networks

In the previous sections, we demonstrated that the topological synthesis generates dendrites that match the morpho-electrical properties of biological reconstructions. To synthesize morphologies for the simulation of realistic digital brain networks, it is essential to validate the connections formed between the dendrites and the axons. For this reason, we generated a synthesized replica of the Markram et al. (2015) digital reconstruction of the rat cortical microcircuit, from now on referred to as a “reconstructed” microcircuit. This somewhat inappropriate term is used for simplicity and consistency throughout the text and is not meant to be confused with an actual reconstructed circuit from biological data. Starting from the initial position and m-type of each cell in the reconstructed microcircuit, the dendrites of all neurons were computationally synthesized and the original axons of the reconstructed microcircuit were used for the definition of appositions, or touch points between dendrites and axons. The connectivity of the synthesized microcircuit was then computed (methods described in Markram et al., 2015) and compared with the connectivity of the reconstructed microcircuit in Figure 6.

The statistical properties of the reconstructed microcircuit (Figure 6A) and the synthesized circuit (Figure 6B) are compared in Figure 6C. For both microcircuits, the connectome of the microcircuit grouped by m-type (1), the connection probabilities (2), and the numbers of synapses per connection (3) are presented in Figure 6, which are in statistical agreement (see Figure S7). Indeed, the differences of these statistical properties shown in Figure 6C are significantly lower than their standard deviations, thus ensuring that the synthesized dendrites do not significantly alter the statistical properties of the network’s connectivity.

Medical applications

As novel techniques (Watts et al., 2013; Sharifi, 2013) emerge for the treatment and analysis of brain disorders (Meng et al., 2018; Heuvel and Sporns, 2019), the need to find new ways to computationally describe these disorders becomes increasingly imperative. As these techniques are frequently applied to animal subjects (Meng et al., 2018), it is important to be able to estimate the effects of these drugs on primates. But any prediction on the effects of complex treatments on humans requires accurate and detailed models of neuronal networks.

With the TNS algorithm we can propose a first step in this direction, by generating unhealthy synthetic neuronal morphologies for which very few or no reconstructions are available. To do so, an appropriate mathematical transformation of barcodes should be defined based on the experimentally observed differences between control and unhealthy dendritic shapes, and

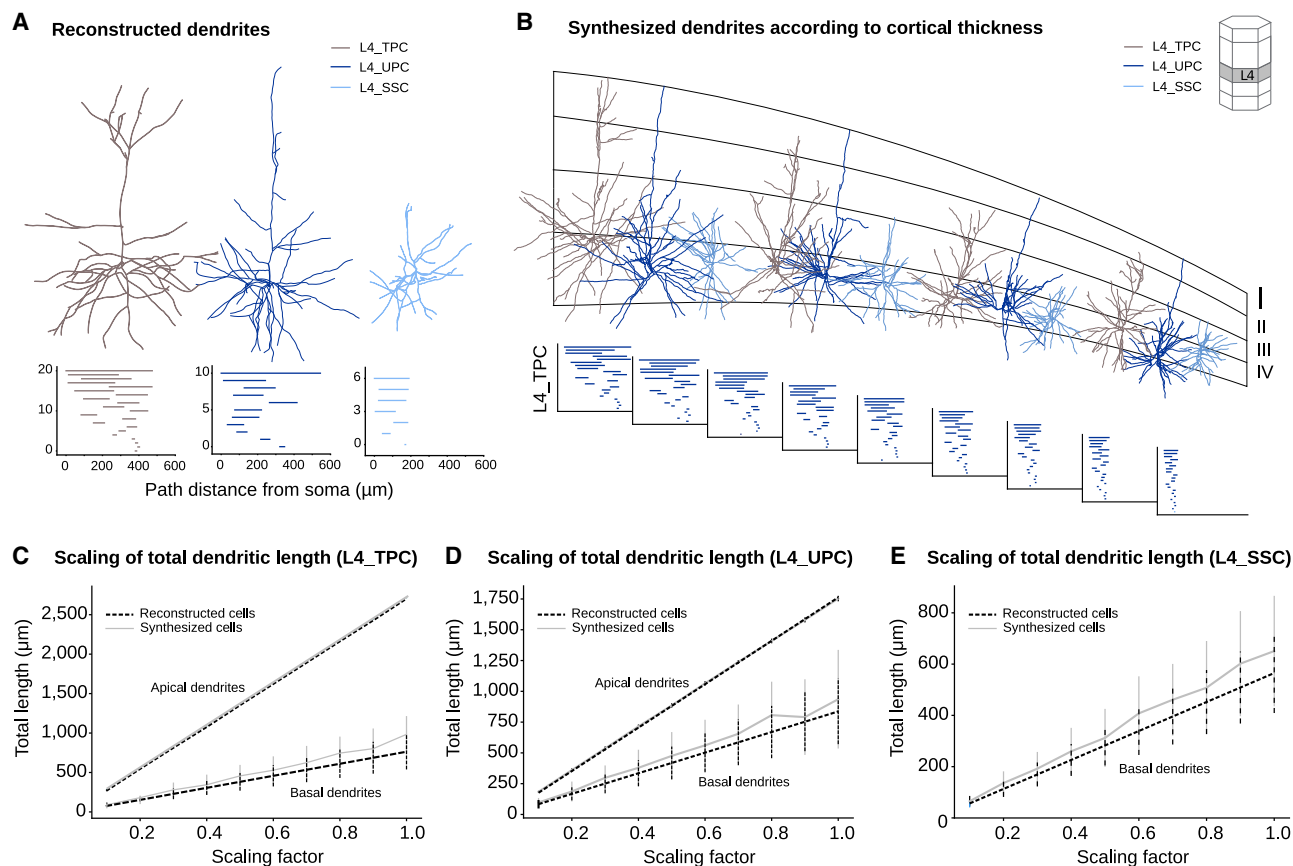


Figure 5. Versatility of the synthesis algorithm

Generalization of topological synthesis for varying cortical thickness.

(A) Exemplar biological reconstructions of three layer 4 pyramidal cell types: L4_TPC (gray), L4_UPC (deep blue), L4_SSC (light blue), and the corresponding persistence barcodes, used as synthesis input.

(B) Scaling of input persistence barcodes and resulting synthesized dendrites ([1.0, 0.8, 0.6, 0.5] of original barcodes). The scaled (from 1.0 to 0.2) barcodes of synthesized L4_TPC apicals presented at the bottom.

(C–E) Total dendritic length of layer 4 cells, as a function of shrinkage factor for basal (bottom) and apical (top) dendrites compared with expected values of scaled biological lengths (black dashed, computed as scaling factor multiplied by total length of reconstructed dendrites) and synthesized (gray continuous) dendrites of L4_TPC (C), L4_UPC (D), and L4_SSC (E). Note that L4_SSC do not have apical dendrites even though they are excitatory cells, therefore only basal dendrite statistics are shown.

applied to the persistence barcodes that are used as synthesis input. This process results in synthesized neuronal morphologies that approximate a target reconstructed population of unhealthy neurons.

As an example, we demonstrate how alterations of dendritic shapes, such as shrinkage of dendritic processes and loss of dendritic branches, that have been associated to mental disorders, such as stress and PTSD (Curran et al., 2017; Dioli et al., 2019; Tornese et al., 2019; Sandini et al., 2020), affect the connectivity of cortical circuits. In general, abnormal dendritic morphology has been linked to brain disorders, such as mental retardation (Kaufmann and Moser, 2000), schizophrenia (Glausier and Lewis, 2013), autism (Phillips and Pozzo-Miller, 2015), and stress disorders (Dioli et al., 2019; Sandini et al., 2020; Shansky and Morrison, 2009). Even though our digital networks do not consist of multiple brain regions, and therefore we cannot reproduce the exact medical conditions reported in the literature, we may still study local ef-

fects of dendritic alterations within a single cortical microcircuit. Such analysis serves as a first step to link local morphological alterations to whole brain networks and study how “small” changes can impact the whole brain functionality.

We implemented two structural changes: shrinkage of dendritic processes (by reducing the lengths of all bars within the input barcodes) and loss of dendritic branches (by removing bars from smaller to larger ones; Figures 7B and 7C). The modified input barcodes were used to synthesize dendrites that populate networks representing different states. Surprisingly, these two types of local dendritic alterations have distinct effects on the cortical networks they form. Neuronal networks formed from dendrites that are gradually shortened lose connections almost linearly (Figure 7D, red), thus higher-order connections collapse rapidly (Figure 7E, red) as a function of the total lost dendritic extent. The topological analysis suggests that networks based on dendrites with lost branches are

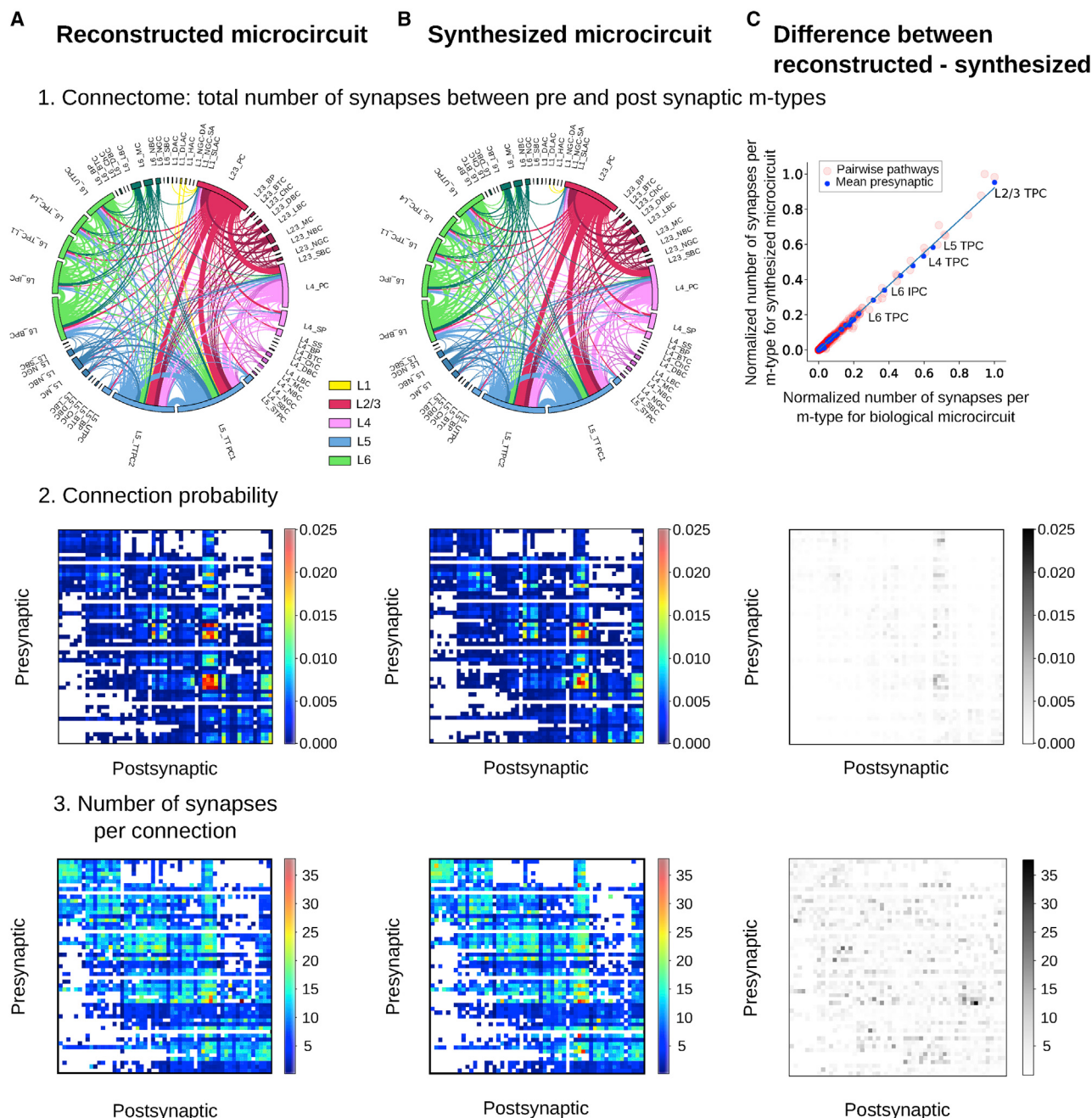


Figure 6. Connectivity of synthesized and reconstructed networks

(A) The connectivity properties of a reconstructed microcircuit (Markram et al., 2015).

(B) The connectivity properties of a microcircuit of fully synthesized dendrites, and reconstructed axons.

(C) Difference between reconstructed and synthesized microcircuits. (1) The connectomes of the microcircuits grouped by m-type. Colors group m-types by layer and correspond to axonal outputs. The thickness of ribbons is proportional to the total number of synapses. (2) Connection probability. A matrix of average connection probability per pathway (350 μ m, central micro-column, 10 K pairs). (3) Synapses per connection. A matrix of the average number of synapses per connection for multi-synapse connections formed between the 55 m-types (10 K pairs).

more resilient to connectivity loss. In fact, the effect of these local dendritic changes on the resulting network is observable only when larger branches (≈ 200 μ m) are lost (Figures 7D and 7E, blue).

This method enables the investigation of the impact of structural neuronal abnormalities and could lead to more advanced diagnostic or treatment techniques. However, this analysis is only a first step toward a more thorough investigation of the links

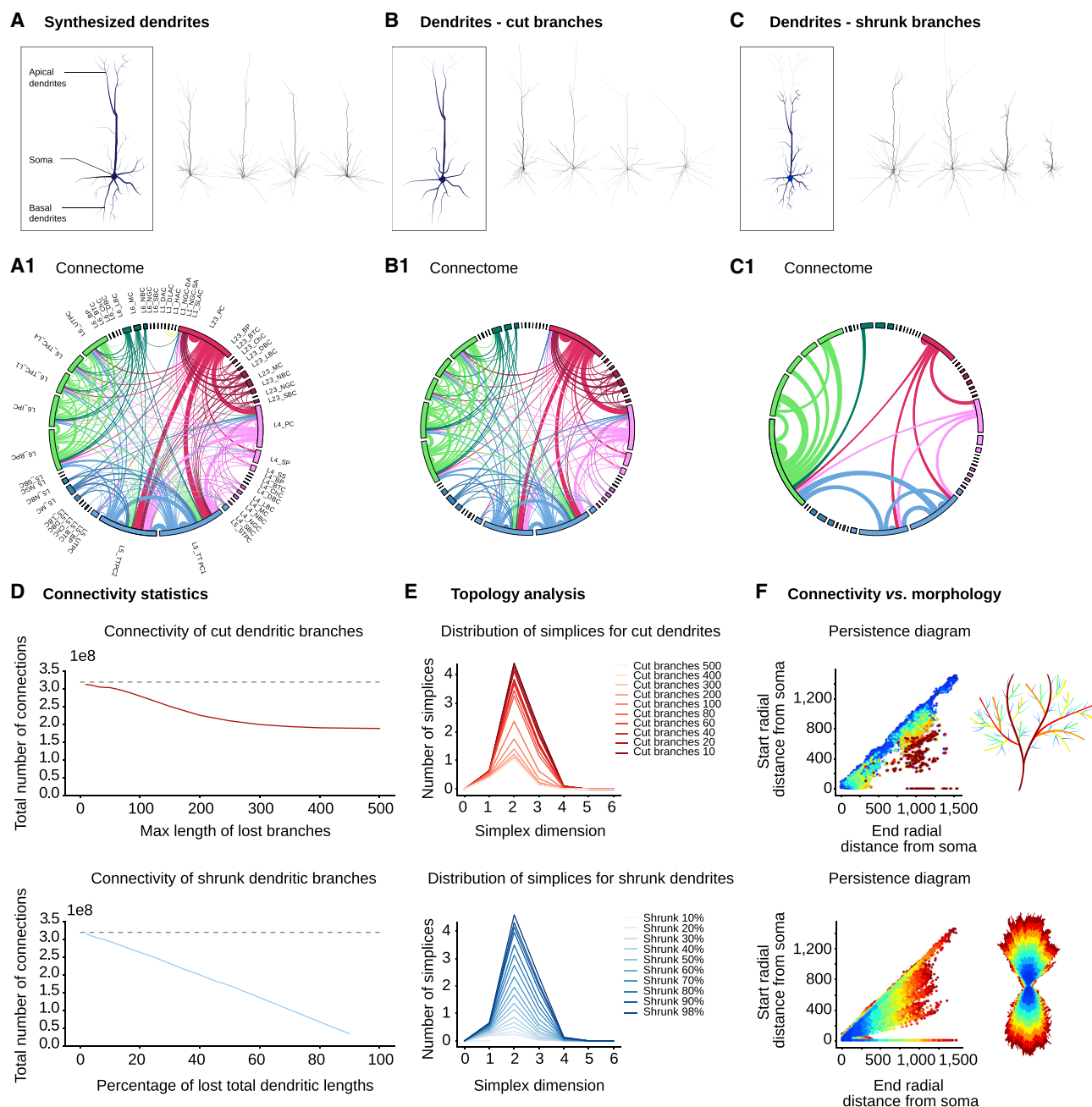


Figure 7. Medical applications

(A–C) Connectivity of synthesized networks based on structural alterations of dendritic morphologies. Schematic representation and examples of layer 5 synthesized pyramidal cells (A), in comparison with cut dendritic branches (B) (lengths above 10, 100, 200, and 400 μm), and shrunk dendrites (C) (98%, 90%, 60%, and 30%). Connectome (presented in subpanel 1) of each synthesized microcircuit: (A) synthesized, (B) cut branches of lengths above 400 μm , (C) shrunk dendrites 10% (connections above 150 K shown, line thickness corresponds to connection strength).

(D) Total number of connections for alterations of type B (red) and C (blue) compared with synthesized network A (black).

(E) Topological analysis of corresponding networks; distribution of directed simplices for alterations of type B (red, top) and C (blue, bottom).

(F) Morphological characteristics and connectivity with respect to alterations of type B (top) and C (bottom). The main branches form the majority of connections (top) and larger dendritic extents (bottom) form more connections. Colormap corresponds to normalized number of connections: from maximum number of connections (3.5×10^8 in red) to minimum (10^7 in blue).

between structural dendritic alterations and the functionality of the corresponding neuronal networks. For the simulation of brain diseases it is essential to generalize these results to whole brain regions and to study how different combinations of neuronal deformations can lead to disruptions of brain networks both structurally and functionally.

DISCUSSION

Since the beginning of the study of the microscopic structure of the brain in the late 19th century by anatomists, such as Deiters, His, and Golgi, which led to the beautiful and detailed formal descriptions of neuronal morphologies by Ramon y Cajal, there have been numerous astounding breakthroughs in our understanding of the brain. Discoveries range through multiple spatial scales from the dynamical properties of ion channels (Ranjan et al., 2019) and single-cell morphologies, physiology, and transcriptomics (Winnubst et al., 2019; Gouwens et al., 2019) to the structural and functional connectivity of the whole brain (Wang et al., 2015; Hahn et al., 2019). Due to rapid experimental progress, the burning challenge of our time is to assemble all the gathered data into a realistic description of the brain. A fundamental step toward understanding brain function is to elucidate the roles of its fundamental cellular components, primarily the neurons. However, much remains unknown concerning the structural properties of neurons and how their morphologies may influence the structural and functional properties of brain networks. A promising approach to discerning the roles of individual neurons in the brain consists of computationally recreating them, i.e., synthesizing them, to study their behavior within digital brain networks.

Due to the complex biological growth mechanisms of neurons that result in intricate branching structures, the highly correlated morphological features of neurons are difficult to reproduce. In this study, the computational synthesis of neuronal morphologies was based on the TMD (Kanari et al., 2018), which retains sufficient information about both the topology and the geometry of a neuronal tree to reproduce the shapes of reconstructed neurons. We demonstrated that this topology-based synthesis algorithm preserves correlations between morphological features. Another challenge for neuronal synthesis is the sparsity of data for many cell types in the cortex, which makes it difficult to approximate the expected biological diversity. As we demonstrated, the minimum number of reconstructed neurons sufficient to achieve biologically relevant diversity with TNS is around 15 to 20 cells. This is a relatively small number compared with the thousands of morphologies that are expressed for each neuron type within a brain region. Our algorithm thus overcomes major limitations of previous synthesis techniques, enabling the large-scale reconstructions of biophysically detailed neuronal networks with unique neuronal morphologies.

Taking this work a step further, we generalized the TNS algorithm to reproduce cells with altered structural properties by applying appropriate transformations to the TMD of reconstructed neurons to study brain pathologies. For example, by simulating the effects of stress on single neurons, we demonstrated that the degree and type of degeneration of dendrites influences the nature of global defects exhibited by cortical microcircuits. The

next step toward the simulation of brain diseases is to generalize these results to whole brain regions, e.g., the neocortex, thalamus, and hippocampus, and to study how different combinations of neuronal deformations can lead to disruptions of brain networks both structurally and functionally. Recent datasets that record the transcriptomics, the electrical properties, and morphologies of cells (Hodge et al., 2019; Gouwens et al., 2019) will be essential for this effort, by enabling the reconstruction of whole brain areas based on their genetic profiles.

Limitations of the study

The TNS is a data-driven synthesis approach. As such it is important to emphasize that the computationally generated neuronal shapes reflect the quality of the original reconstructions that were used as input to the algorithm. Thus, an original set of reconstructed neurons that are representative of the biological population of a selected cell type is required to generate neurons that reflect the biological diversity.

While the synthesis of dendrites is already an important step toward the digital reconstruction of more realistic brain networks, the ability to synthesize axonal trees is the next challenge that should be addressed. The current method does not address the generation of axonal morphologies. Axons will be challenging to grow but necessary to allow the computational modeling of brain networks (Wang et al., 2015) as their branching structure is an essential determinant of the connectivity of the network (van Pelt et al., 2010). In addition, few complete axonal reconstructions are available because of their highly complex branching structures, because the experimental reconstruction of axons requires considerably more effort than dendrites.

Another important limitation of the current synthesis technique is that it does not take into account environmental factors during the growth process. Spatially embedded synthesis algorithms will be crucial for the generation of more complex branching patterns of interneurons, PCs, or long-range projecting axons, such as nigrostriatal dopaminergic neurons (Matsuda et al., 2009) and densely connected claustrum cells (Torgerson et al., 2015).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Definition of morpho-electrical terms
 - Definition of morphological terms
 - Morphometrics glossary
 - Electrical features glossary
 - Synthesis input
 - Biological barcodes
 - Biological distributions
 - Soma parameters

- Diameter parameters
- Input parameters
- **METHOD DETAILS**
 - Topological morphology descriptor
 - Topological neuron synthesis
 - Dendritic synthesis algorithm
 - Connectivity between branches
 - Quantification and Statistical Analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2022.110586>.

ACKNOWLEDGMENTS

The morphological reconstructions used for this study are part of a publication that is currently in preparation; we would like to acknowledge the work of Ying Shi, Thomas Berger, Shruti Muralidhar, Rodrigo de Campos Perin, and Zoltan Kisvarday for the collection of these reconstructions and we refer the reader to <https://bbp.epfl.ch/sscx-portal/experimental-data/neuron-morphology/> for a proper citation of the morphological reconstructions. We acknowledge Joseph Graham and Guy Atenekeg for crucial input on the synthesis algorithm, at the early stages of this work. We thank Jay Coggan for helpful conversations in various stages of this research and for a thorough read of the manuscript. We thank Michael Reimann for his input on the connectivity analysis, Rembrandt Bakker (Bakker et al., 2017) for visualization tools, Nicolas Antille and Cyrille Favreau for visualizations, Caitlin Monney for figure editing, and Michael Gevaert and Adrien Berchet for software engineering support. This study was supported by funding to the Blue Brain Project, a research center of the École polytechnique fédérale de Lausanne, from the Swiss government's ETH Board of the Swiss Federal Institutes of Technology.

AUTHOR CONTRIBUTIONS

Conceptualization, L.K.; methodology, L.K.; investigation, L.K., H.D., A.A., J.C., and K.H.; validation, L.K., A.C., A.A., and H.D.; software, L.K., A.A., H.D., and B.C.; writing – original draft, L.K.; writing – review & editing, H.D., A.C., A.A., W.V.G., B.C., J.S., K.H., and H.M.; funding acquisition, H.M.; supervision, K.H. and H.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 15, 2021

Revised: July 22, 2021

Accepted: March 8, 2022

Published: April 5, 2022

REFERENCES

Akram, M.A., Nanda, S., Maraver, P., Armañanzas, R., and Ascoli, G. (2018). An open repository for single-cell reconstructions of the brain forest. *Sci. Data* 5, 180006.

Ascoli, G., Donohue, D.E., and Halavi, M. (2007). Neuromorpho.org: A central resource for neuronal morphologies. *J. Neurosci.* 27, 9247–9251.

Ascoli, G., Alonso-Nanclares, L., Anderson, S., Barrionuevo, G., and Benavides-Piccione, R. (2008). Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nat. Rev. Neurosci.* 9, 557–568.

Ascoli, G., Krichmar, J., Scorcioni, R., Nasuto, S.J., Senft, S., and Krichmar, G.L. (2001). Computer generation and quantitative morphometric analysis of virtual neurons. *Anat. Embryol.* 204, 283–301.

Aslangul, C., Pottier, N., Chvosta, P., and D Saint-James, D. (1993). Directed random walk with spatially correlated random transfer rates. *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.* 47, 1610–1617.

Bakker, R., Garcia-Amado, M., Evangelio, M., Clasca, F., and Tiesinga, P. (2017). Workflow, data format and tools to register neuron morphologies to a reference brain atlas. 26th annual computational neuroscience meeting (cns*2017): Part 3. *BMC Neurosci.* 18, 203–208.

Benavides-Piccione, R., Regalado-Reyes, M., Fernaud-Espinosa, I., Kastanauskaitė, A., Tapia-González, S., León-Espinosa, G., Rojo, C., Insausti, R., Segev, I., and DeFelipe, J. (2019). Differential structure of hippocampal ca1 pyramidal neurons in the human and mouse. *Cereb. Cortex* 30, 730–752.

Bird, A.D., and Cuntz, H. (2016). Optimal current transfer in dendrites. *PLoS Comput. Biol.* 12, e1004897.

Burke, R., Marks, W.B., and Ulfhake, B. (1992). A parsimonious description of motoneuron dendritic morphology using computer simulation. *J. Neurosci.* 12, 2403–2416.

Cajal, R.Y. (1998). *Histologie du système nerveux de l'homme et des vertébrés*. *J. Neuropathol. Exp. Neurol.* 57, 883.

Carlsson, G.E. (2009). Topology and data. *Bull. Am. Math. Soc.* 46, 255–308.

Chklovskii, D. (2004). Synaptic connectivity and neuronal morphology two sides of the same coin. *Neuron* 43, 609–617.

Conde-Sousa, E., Szűcs, P., Peng, H., and Aguiar, P. (2016). N3dfix: an algorithm for automatic removal of swelling artifacts in neuronal reconstructions. *Neuroinformatics* 15, 5–12.

Cuntz, H., Borst, A., and Segev, I. (2007). Optimization principles of dendritic structure. *Theor. Biol. Med. Model.* 4, 21.

Cuntz, H., Forstner, F., Borst, A., and Häusser, M. (2010). One rule to grow them all: a general theory of neuronal branching and its practical application. *PLoS Comput. Biol.* 6, e1000877.

Cuntz, H., Mathy, A., and Häusser, M. (2012). A scaling law derived from optimal dendritic wiring. *Proc. Natl. Acad. Sci. U S A* 109, 11014–11018.

Curran, M.M., Sandman, C., Davis, E.P., Glynn, L., and Baram, T. (2017). Abnormal dendritic maturation of developing cortical neurons exposed to corticotropin releasing hormone (crh): insights into effects of prenatal adversity? *PLoS One* 12, e0180311.

Deitcher, Y., Eyal, G., Kanari, L., Verhoog, M.B., Atenekeg Kahou, G.A., Mansvelder, H.D., de Kock, C.P.J., and Segev, I. (2017). Comprehensive morpho-electronic analysis shows 2 distinct classes of I2 and I3 pyramidal neurons in human temporal cortex. *Cereb. Cortex* 27, 5398–5414.

Dioli, C., Patricio, P., Sousa, N., Kokras, N., Dalla, C., Guerreiro, S., Santos-Silva, Miguel A., Rego, A., Pinto, L., Ferreira, E., and Sotiropoulos, I. (2019). Chronic stress triggers divergent dendritic alterations in immature neurons of the adult hippocampus, depending on their ultimate terminal fields. *Translational Psychiatry* 9, 143.

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

Egger, R., Derksen, V.J., Udvar, D., Hege, H., and Oberlaender, M. (2014). Generation of dense statistical connectomes from sparse morphological data. *Front. Neuroanat.* 8, 129.

van Elburg, R., and van Ooyen, A. (2010). Impact of dendritic size and dendritic topology on burst firing in pyramidal cells. *Plos Comput. Biol.* 6, e1000781.

Ero, C., Gewaltig, M.O., Keller, D., and Markram, H. (2018). A cell atlas for the mouse brain. *Front. Neuroinf.* 12, 84.

Farhoodi, R., Lansdell, B., and Kording, K. (2019). Quantifying how staining methods bias measurements of neuron morphologies. *Front. Neuroinf.* 13, 36.

Fletcher, L.N., and Williams, S. (2019). Neocortical topology governs the dendritic integrative capacity of layer 5 pyramidal neurons. *Neuron* 101, 76–90.e4.

Galton, F., and Watson, H.W. (1875). On the probability of the extinction of families. *J. Anthropol. Inst. G. B. Ireland* 4, 399–406.

Van Geit, W., Gevaert, M., Chindemi, G., Rössert, C., Courcol, J.D., Müller, E., Schürmann, F., Segev, I., and Markram, H. (2016). Bluepyopt: leveraging open

- source software and cloud infrastructure to optimise model parameters in neuroscience. *Front. Neuroinf.* 10, 17.
- Glausier, J., and Lewis, D. (2013). Dendritic spine pathology in schizophrenia. *Neuroscience* 251, 90–107.
- Gonchar, Y., Wang, Q., and Burkhalter, A. (2007). Multiple distinct subtypes of gabaergic neurons in mouse visual cortex identified by triple immunostaining. *Front. Neuroanat.* 1, 3.
- Gouwens, N.W., Sorensen, S., Berg, J., Lee, C., Jarsky, T., Ting, J., Sunkin, S., Feng, D., Anastassiou, C.A., Barkan, E., et al. (2019). Classification of electrophysiological and morphological neuron types in the mouse visual cortex. *Nat. Neurosci.* 22, 1182–1195.
- Graham, B., and van Ooyen, A. (2006). Mathematical modelling and numerical simulation of the morphological development of neurons. *BMC Neurosci.* 7, S9.
- Haberl, M., da Silva, S.V., Guest, J.M., Ginger, M., Ghanem, A., Mulle, C., Oberlaender, M., Conzelmann, K., and Frick, A. (2014). An anterograde rabies virus vector for high-resolution large-scale reconstruction of 3d neuron morphology. *Brain Struct. Funct.* 220, 1369–1379.
- Hahn, G., Skeide, M.A., Mantini, D., Ganzetti, M., Destexhe, A., Friederici, A., and Deco, G. (2019). A new computational approach to estimate whole-brain effective connectivity from functional and structural mri, applied to language development. *Sci. Rep.* 9, 8479.
- Halavi, M., Polavaram, S., Donohue, D.E., Hamilton, G., Hoyt, J., Smith, K., and Ascoli, G. (2008). Neuromorpho.org implementation of digital neuroscience: dense coverage and integration with the nif. *Neuroinformatics* 6, 241–252.
- Häusser, M., Spruston, N., and Stuart, G. (2000). Diversity and dynamics of dendritic signaling. *Science* 290, 739–744.
- Hedges, Larry V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *J. Educ. Stat.* 6, 107–128.
- Herculano-Houzel, S., Mota, B., and Lent, R. (2006). Cellular scaling rules for rodent brains. *Proc. Natl. Acad. Sci. U S A* 103, 12138–12143.
- Heuvel, M.P., and Sporns, O. (2019). A cross-disorder connectome landscape of brain dysconnectivity. *Nat. Rev. Neurosci.* 20, 435–446.
- Hillman, D.E. (1979). Neuronal shape parameters and substructures as a basis of neuronal form. *The Neurosciences: Fourth Study Program (The MIT Press)*, pp. 477–498.
- Hodge, R., Bakken, T.E., Miller, J., Smith, K., Barkan, E., Graybiel, L.T., Close, J., Long, B.R., Johansen, N., Penn, O., et al. (2019). Conserved cell types with divergent features in human versus mouse cortex. *Nature* 573, 61–68.
- Hodgkin, A., and Huxley, A. (1952). Propagation of electrical signals along giant nerve fibres. *Proc. R. Soc. Lond. Ser. B - Biol. Sci.* 140, 177–183.
- 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. *Cereb. Cortex* 29, 1719–1735.
- Kanari, L., Diotko, P., Scolamiero, M., Levi, R., Shillcock, J., Hess, K., and Markram, H. (2018). A topological representation of branching neuronal morphologies. *Neuroinformatics* 16, 3–13.
- Kanari, Lida, Garin, Ad'elie, and Hess, Kathryn (2020). From trees to barcodes and back again: theoretical and statistical perspectives. *Algorithms* 13, 335.
- Kaufmann, W., and Moser, H. (2000). Dendritic anomalies in disorders associated with mental retardation. *Cereb. Cortex* 10, 981–991.
- Koene, R., Tijms, B., van Hees, P., Postma, F., Ridder, A., Ramakers, G., van Pelt, J., and Netmorph, A. van Ooyen. (2009). A framework for the stochastic generation of large scale neuronal networks with realistic neuron morphologies. *Neuroinformatics* 7, 195–210.
- Landau, I., Egger, R., Dercksen, V.J., Oberlaender, M., and Sompolsky, H. (2016). The impact of structural heterogeneity on excitation-inhibition balance in cortical networks. *Neuron* 92, 1106–1121.
- Larkum, M., Waters, J., Sakmann, B., and Helmchen, F. (2007). Dendritic spines in apical dendrites of neocortical layer 2/3 pyramidal neurons. *J. Neurosci.* 27, 8999–9008.
- Larkum, M., Nevian, T., Sandler, M., Polsky, A., and Schiller, J. (2009). Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: a new unifying principle. *Science* 325, 756–760.
- Ledda, F., and Paratcha, G. (2017). Mechanisms regulating dendritic arbor patterning. *Cell Mol. Life Sci.* 74, 4511–4537.
- Lefort, S.a., Tomm, C., Sarria, J.F., and Petersen, C. (2009). The excitatory neuronal network of the c2 barrel column in mouse primary somatosensory cortex. *Neuron* 61, 301–316.
- López-Cruz, P.L., Bielza, C., Larrañaga, P., Benavides-Piccione, R., and De-Felipe, J. (2011). Models and simulation of 3d neuronal dendritic trees using bayesian networks. *Neuroinformatics* 9, 347–369.
- Luczak, A. (2006). Spatial embedding of neuronal trees modeled by diffusive growth. *J. Neurosci. Methods* 157, 132–141.
- Luczak, A. (2010). Measuring neuronal branching patterns using model-based approach. *Front. Comput. Neurosci.* 4, 135.
- Markram, H., Muller, E., Ramaswamy, S., Reimann, M.W., Abdellah, M., Sanchez, C.A., Ailamaki, A., Alonso-Nanclares, L., Antille, N., Arsever, S., et al. (2015). Reconstruction and simulation of neocortical microcircuitry. *Cell* 163, 456–492.
- Marx, M., and Feldmeyer, D. (2013). Morphology and physiology of excitatory neurons in layer 6b of the somatosensory rat barrel cortex. *Cereb. Cortex* 23, 2803–2817.
- Matsuda, W., Furuta, T., Nakamura, K., Hioki, H., Fujiyama, F., Arai, R., and Kaneko, T. (2009). Single nigrostriatal dopaminergic neurons form widely spread and highly dense axonal arborizations in the neostriatum. *J. Neurosci.* 29, 444–453.
- Memelli, H., Torben-Nielsen, B., and Kozloski, J. (2013). Self-referential forces are sufficient to explain different dendritic morphologies. *Front. Neuroinf.* 7, 1.
- Meng, C., Zhou, J., Papaneri, A.B., Peddada, T., and Cui, G. (2018). Spectrally resolved fiber photometry for multi-component analysis of brain circuits. *Neuron* 98, 707–717.e4.
- Pearson, K. (1905). The problem of the random walk. *Nature* 72, 294.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.
- van Pelt, J., and van Ooyen, A. (2013). Estimating neuronal connectivity from axonal and dendritic density fields. *Front. Comput. Neurosci.* 7, 160.
- van Pelt, J., Carnell, A., Ridder, S.D., Mansvelder, H., and van Ooyen, A. (2010). An algorithm for finding candidate synaptic sites in computer generated networks of neurons with realistic morphologies. *Front. Comput. Neurosci.* 4, 148.
- Peng, H. (2008). Bioimage informatics: a new area of engineering biology. *Bioinformatics* 24, 1827–1836.
- Peng, H., Hawrylycz, M., Roskams, J., Hill, S., Spruston, N., Meijering, E., and Bigneuron, G. Ascoli. (2015). Large-scale 3d neuron reconstruction from optical microscopy images. *Neuron* 87, 252–256.
- Phillips, M., and Pozzo-Miller, L. (2015). Dendritic spine dysgenesis in autism related disorders. *Neurosci. Lett.* 601, 30–40.
- Rall, W. (1959). Branching dendritic trees and motoneuron membrane resistivity. *Exp. Neurol.* 1, 491–527.
- Ramaswamy, S., Courcol, J.-D., et al. (2015). The neocortical microcircuit collaboration portal: a resource for rat somatosensory cortex. *Front. Neural Circuits* 9, 44.
- Ramaswamy, S., Hill, S.L., King, J.G., Schürmann, F., Wang, Y., and Markram, H. (2012). Intrinsic morphological diversity of thick-tufted layer 5 pyramidal neurons ensures robust and invariant properties of in silico synaptic connections. *J. Physiol.* 590, 737–752.
- Ranjan, R., Logette, E., Marani, M., Herzog, M., Tâche, V., Scantamburlo, E., Buchillier, V., and Markram, H. (2019). A kinetic map of the homomeric voltage-gated potassium channel (kv) family. *Front. Cell Neurosci.* 13, 358.

- Samsonovich, A., and Ascoli, G. (2003). Statistical morphological analysis of hippocampal principal neurons indicates cell-specific repulsion of dendrites from their own cell. *J. Neurosci. Res.* **71**, 173–187.
- Sandini, C., Chambaz, M., Schneider, M., and Eliez, S. (2020). Pituitary dysmaturation affects psychopathology and neurodevelopment in 22q11.2 deletion syndrome. *Psychoneuroendocrinology* **113**, 104540.
- Schaefer, A., Larkum, M., Sakmann, B., and Roth, A. (2003). Coincidence detection in pyramidal neurons is tuned by their dendritic branching pattern. *J. Neurophysiol.* **89**, 3143–3154.
- Scorcioni, R., Polavaram, S., and Ascoli, G. (2008). L-measure: a web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies. *Nat. Protoc.* **3**, 866–876.
- Shansky, R., and Morrison, J. (2009). Stress-induced dendritic remodeling in the medial prefrontal cortex: effects of circuit, hormones and rest. *Brain Res.* **1293**, 108–113.
- Sharifi, M.S. (2013). Treatment of neurological and psychiatric disorders with deep brain stimulation; raising hopes and future challenges. *Basic Clin. Neurosci.* **4**, 266–270.
- Shillcock, J., Hawrylycz, M., Hill, S.L., and Peng, H. (2016). Reconstructing the brain: from image stacks to neuron synthesis. *Brain Inform.* **3**, 205–209.
- Smith, S., Smith, I.T., Branco, T., and Häusser, M. (2013). Dendritic spikes enhance stimulus selectivity in cortical neurons in vivo. *Nature* **503**, 115–120.
- Spruston, N. (2008). Neuroscience: strength in numbers. *Nature* **452**, 420–421.
- Torgerson, C., Irimia, A., Goh, S.M., and Van Horn, J.V. (2015). The dti connectivity of the human claustrum. *Hum. Brain Mapp.* **36**, 827–838.
- Tornese, P., Sala, N., Bonini, D., Bonifacino, T., La Via, L., Milanese, M., Trecani, G., Seguni, M., Ieraci, A., Mingardi, J., et al. (2019). Chronic mild stress induces anhedonic behavior and changes in glutamate release, bdnf trafficking and dendrite morphology only in stress vulnerable rats. the rapid restorative action of ketamine. *Neurobiol. Stress* **10**, 100160.
- Wang, D., He, X., Zhao, Z., Feng, Q., Lin, R., Sun, Y., Ding, T., Xu, F., Luo, M., and Zhan, C. (2015). Whole-brain mapping of the direct inputs and axonal projections of pomc and agrp neurons. *Front. Neuroanat.* **9**, 40.
- Watts, B.V., Schnurr, P., Mayo, L., Young-Xu, Y., Weeks, W., and Friedman, M. (2013). Meta-analysis of the efficacy of treatments for posttraumatic stress disorder. *J. Clin. Psychiatry* **74**, e541–e550.
- Wen, Q., Stepanyants, A., Elston, G., Grosberg, A.Y., and Chklovskii, D. (2009). Maximization of the connectivity repertoire as a statistical principle governing the shapes of dendritic arbors. *Proc. Natl. Acad. Sci. U S A* **106**, 12536–12541.
- Winnubst, J., Bas, E., Ferreira, T.A., Wu, Z., Economo, M.N., Edson, P., Arthur, B.J., Bruns, C., Rokicki, K., Schauder, D.M., et al. (2019). Reconstruction of 1,000 Projection Neurons Reveals New Cell Types and Organization of Long-Range Connectivity in the Mouse Brain. *Cell* **179** (1), 268–281.
- Yi, G.S., Wang, J., Deng, B., and Wei, X. (2017). Morphology controls how hippocampal ca1 pyramidal neuron responds to uniform electric fields: a biophysical modeling study. *Sci. Rep.* **7**, 3210.
- Zubler, F., and Douglas, R. (2009). A framework for modeling the growth and development of neurons and networks. *Front. Comput. Neurosci.* **3**, 25.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Experimental reconstructions	Markram et al., 2015; Ramaswamy and Courcol, 2015	http://microcircuits.epfl.ch/#/article/article_3_mph
Experimental reconstructions (additional data)	This paper	https://doi.org/10.5281/zenodo.5909613
Software and algorithms		
Python version >3.7	Python Software Foundation	https://www.python.org
Sklearn	Pedregosa et al., 2011	https://scikit-learn.org
Neuronal Morphology Analysis	Blue Brain Project	https://github.com/BlueBrain/NeuroM
Topological Morphology Descriptor (TMD)	Kanari et al., 2018	https://github.com/BlueBrain/TMD
Topological Neuron Synthesis (NeuroTS)	This paper https://github.com/BlueBrain/NeuroTS	https://doi.org/10.5281/zenodo.6319675
Diameter correction on dendrites (diameter-synthesis)	This paper https://github.com/BlueBrain/diameter-synthesis	https://doi.org/10.5281/zenodo.6244166
Scripts for analysis	This paper	https://doi.org/10.5281/zenodo.5909613
Others		
Computationally synthesized neuronal morphologies	This paper	https://doi.org/10.5281/zenodo.5909613

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and code should be directed to and will be fulfilled by the lead contact, Lida Kanari (lida.kanari@gmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All original data as well as code for individual experiments have been deposited at Zenodo and will be publicly available as of the date of publication. The DOI is listed in the [Key resources table](#).
- The code for the computational generation of cells is available in GitHub, <https://github.com/BlueBrain/NeuroTS>, as well as the code for the denition of dendritic diameter generation, <https://github.com/BlueBrain/diameter-synthesis>. All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the [Key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Definition of morpho-electrical terms

As a guide to the reader, the main definitions of morphological terms that will be used through this paper are summarized in the following table. Note that these terms might have different meanings elsewhere in the literature. Since morphological terminology is often not consistent through the literature, we clarify the meaning of each term used in this paper.

Definition of morphological terms

- Soma: the cell body is described as a sphere $S_{d_s}^c$ of diameter d_s and center c_s . The contour of a soma consists of a set of points on the x-y plane around the soma surface.
- Neurite: A neuronal tree; a dendrite or an axon.
- Neurite point: (x, y, z, d) , where (x, y, z) are the 3D coordinates and d is the dendritic diameter that represents the thickness of the neurite at that point.

- Branchpoint: a point in the tree with two (or more) children, also known as a bifurcation point (or multibifurcation if it has more than two children).
- Neurite tip: a point in the tree without any children, also known as a termination point.
- Neurite section: a set of points of a neurite, either between two branch points or between a branchpoint and a termination point.
- Neurite trunk: the first section of a neurite, starting from the soma and ending at the first branch point.
- Vect: A unit vector in 3D space, which is equivalently represented by a pair of angles, and defines a direction, or orientation, in 3D space.

For the definition of morphological features, we use as guide the Petilla terminology paper (features 15, [Ascoli et al. \(2008\)](#)), as well as some additional features (6–9).

Morphometrics glossary

1. Thickness: the dendritic thickness in micrometers.
2. Taper: percentage of thickness narrowing per unit distance.
3. Bifurcation angle: angle between two daughter branches at a branch point (bifurcation).
4. Tortuosity: the straight-line distance between two consecutive branch points divided by the length of the neuronal path between those points.
5. Partition asymmetry: Ratio of the absolute value of the difference and the sum of the number of bifurcations in the two daughter subtrees at a branch point.
6. Radial distance: the end-to-end straight-line distance between the branchpoint of a section and the soma surface.
7. Path distance: the distance along the neuronal path, between the branch point of a section and the soma surface.
8. Branch order: the number of bifurcations between current position and the root.
9. Section length: total length of a section.

For more details about the electrical features, we refer to [Van Geit et al., 2016](#). Basic definitions are summarized in the table below.

Electrical features glossary

1. AP: action potential.
2. RMP: resting membrane potential.
3. IV: a step protocol used for studying the current-voltage relationship. The currents are kept low to avoid action potential generation.
4. APWaveform: Short step protocol used for studying the action potential shape.
5. Step (120%): a longer (2 seconds) step stimulus, the current amplitude is 120% of the current amplitude that is necessary to trigger one action potential.
6. AHP: after-hyperpolarization after an action potential.
7. ISI: inter-spike interval.

Synthesis input

Neuronal reconstructions of a large number of cortical morphological classes were used as input to the synthesis algorithm. Biological reconstructions from the BBP dataset of all rodent cortical cell types reported in previous publications ([Markram et al., 2015](#); [Kanari et al., 2019](#); [Ramaswamy and Courcol, 2015](#)) were used as synthesis input. The BBP dataset was chosen as input because of the consistency of the quality of the input morphologies, as they were all generated using the same reconstruction protocol. In addition, the electrical models that have been generated for all cell types ([Markram et al., 2015](#); [Van Geit et al., 2016](#)) make the comparison of multiple properties (morphological, electrical) feasible.

The input consists of morphology files (in one of the following formats: neuroLucida ASC, SWC, or H5), which contain the three dimensional positions (x, y, z) and the thickness (d) of the neuronal nodes and their adjacency relations, to form a tree. A few modifications were performed on the original reconstructions to compensate for reconstruction artifacts. The slicing of the brain tissue and the filling of the cells with biocytin (Horikawa and Armstrong, 1988) result in their shrinkage. This effect modifies the tortuosity of the reconstructions (as cells appear more tortuous than they originally were), and the extent of their processes decreases. To compensate for these artifacts, the cells that are used as input for synthesis are initially “unraveled”, as described in [Markram et al. \(2015\)](#). Another important artifact is the loss of arborization, due to slicing of the tissue during the reconstruction process. This error is compensated for with a “repair” process described in [Markram et al. \(2015\)](#). To compensate for the loss of arborizations, trees that contain fewer than three sections are considered cut and thus discarded from the synthesis input during preprocessing of the input data.

Biological barcodes

The algorithm that extracts a persistence barcode [Kanari et al., 2018](#) from a neuronal tree is described in the previous section. The barcodes that are used as input for synthesis are enhanced with the bifurcation angles of their corresponding components, i.e., at each bifurcation point, where a new branch emerges, the bifurcation angle α from its parent is recorded. The i^{th} bar in TMD (tree),

denoted bar_i , corresponds to the i^{th} branch of the tree and is encoded as a triplet (b_i, d_i, a_i) , where b_i is the initial radial distance of the branch, d_i is the terminal radial distance, and a_i is the bifurcation angle at which the component emerges from its parent. Note that b_i and d_i are real numbers, while a_i is a set of four real numbers, which represent the following angles.

- a_i^1 : angle between child and parent in x-y plane
- a_i^2 : angle between child and parent in x-z plane
- a_i^3 : angle between two children in x-y plane
- a_i^4 : angle between two children in x-z plane

In order to recreate a population of biological reconstructions, we need to generate the persistence barcodes for each tree of the reconstructed morphologies, a process that we refer to as barcode extraction. The complete set of barcodes that will be used as input to the synthesis algorithm is denoted as

$$\text{Barcodes} = \{TMD_j | 1 \leq j \leq n_t\},$$

where $n_t = \#$ of neuronal trees in the biological population, where for the j^{th} tree

$$TMD_j = \{bar_i = (b_i, d_i, a_i) | 1 \leq i \leq n_j\},$$

where $n_j = \#$ of components (bars) of TMD_j .

Barcodes are randomly and independently sampled during the computational growth of a neuron from the set of input barcodes. Each synthesized neurite (dendrite) is generated by a single barcode TMD , which corresponds to a unique biological tree.

Biological distributions

In addition to the persistence barcodes, which encode the topology of the neurites, a number of independent morphometrics are measured. The first is the size of the cell body; the soma is considered to be a sphere and therefore a center and a diameter are sufficient to describe it. The center, or origin of the cell, is a user-defined parameter, while the diameter is sampled from the corresponding biological distribution D_{ss} . The number of dendrites D_{nn} and their relative pairwise angles D_{pa} are also sampled from the biological distributions, and define the number of dendrites that will be generated as well as their initial outward directions. Finally, a set of distributions that describe the thickness of the dendrites is also extracted, sampled from the biological reconstructions. All the morphological features that are used as input to synthesis are summarized in the following table.

Soma parameters

- D_{ss} : Soma diameter.
- D_{nn} : Number of neurites of a specific type within a neuron; basal or apical dendrite.
- D_{pa} : Pairwise angles between neurites, as they emerge from the soma.

Diameter parameters

- D_{tips} : Diameters of the tips, or terminations of the neurite.
- D_{tr} : Taper rates, define the tapering, i.e., the difference between the diameters at the beginning and end of a section, normalized by the length: $(D_{final} - D_{initial})/length$. The distribution of taper rates is computed for each neuronal section.
- D_{rr} : Rall ratio, which describes the relation between the diameter of a parent and those of its children at a branchpoint. The value of the Rall ratio is the exponent n such that $D^n = d_1^n + d_2^n + \dots$ at a branchpoint, where d_i is the diameter of the i^{th} child and D is the parent diameter. The distribution of Rall ratios is computed for each bifurcation.
- D_{md} : Maximum diameter of each neurite, per type. Under the assumption that dendritic diameter decreases with distance from the soma, this value corresponds to the diameter of the trunk.

Input parameters

- τ : targeting weight that defines how “straight” the generated section will be.
- ρ : randomness weight that defines how “tortuous” the generated section will be.
- μ : memory weight that defines the independence from previous steps. It is defined by $\mu = 1 - \tau - \rho$.
- c : defines the center of the soma, and the starting point for the growth of the neuron.
- $step\text{-}size$: defines the length of each step during the growth. We represent the step size as a normal distribution with mean 1 μm and standard deviation 0.2 μm .

Note that both the input distributions and the persistence barcodes are automatically computed from the population of reconstructed cells. The growth process is also controlled by a set of parameters that are defined by the user. The first two parameters, τ and ρ , characterize the elongation of a branch. The origin of the morphology, i.e., the center c of the soma, is an input parameter that allows the user to control the initial position of the cell. This is important for the generation of cells in place within a brain region, in which case this parameter is chosen by complex placement algorithms. Last, the step size is selected by the user and defines the

frequency of points on the morphology. For neurons this parameter is set to 1 μm which is a reasonable frequency of points that corresponds to the mean frequency of points on the biological reconstructions.

METHOD DETAILS

Topological morphology descriptor

The Topological Morphology Descriptor (Kanari et al., 2018) encodes the topological and geometric properties of a neuronal tree into a single mathematical object: the persistence barcode. The persistence barcode of a neuronal tree represents the start and end distances (i.e., radial or path distance from the soma) of all branches within the tree as pairs of real numbers, or bars. Note that a branch can consist of multiple bifurcations along the tree path until a terminal point is reached. The barcodes are used to define the bifurcation and termination probabilities during synthesis; the coupling between bifurcation and termination probabilities is necessary to reproduce correlations between morphological features (see Figure S7). TMD has been successfully used for the classification (Kanari et al., 2018, 2019) and the clustering (Deitcher et al., 2017) of neuronal morphologies due to its ability to capture a combination of morphological characteristics that are important for the grouping of cortical morphologies.

We define a tree as an embedding in 3D-space of a connected, acyclic, directed graph, with a defined root, i.e. the neuronal soma. The details of the algorithm are given in Kanari et al., 2018. The TMD algorithm (Figure S1) takes as input a set of branch points, or bifurcations, i.e., nodes with more than one child, and leaves, or termination points, i.e., nodes with no children, of the tree and produces a persistence barcode TMD (tree), which is a multi-set of intervals, called bars, on the real line Figure S1B. Each bar encodes the lifetime of a component in the underlying structure, identifying the distance at which a branch is first detected, emerging from a larger subtree (birth, b_i) and the distance at which the branch terminates (death, d_i). Equivalently, the persistence diagram (Figure S1C, Carlsson, 2009) represents the pair of birth-death times of each component as a point in the 2D plane. Note that all distances are computed in μm and for clarity the projection of radial distances is presented in figure Figure S1.

Topological neuron synthesis

The morphological development of neurons in the brain is a complicated process that depends on both genetic and environmental components (Ledda and Paratcha, 2017). The processes that contribute to neuronal growth differ between species, brain regions, and morphological types. Advances in experiments and mathematical models have converged on a set of commonly accepted stages of morphological growth: the initiation of neurites, neurite elongation, axon path-finding and neurite branching (Graham and van Ooyen, 2006). While biological development is not simulated, biological principles of morphological growth inform the design of our computational algorithm to synthesize dendritic morphologies. In this study, we focus on the computational synthesis of dendrites and thus axon path-finding will not be considered.

Dendritic synthesis algorithm

The TNS algorithm consists of three main components (Figure 1A): the initiation (section I. Initiation of neurites), branching (section II. Branching/termination), and elongation (section III. Elongation of neurites) of neurites. The first part of a neuron to be generated is the cell body, i.e., the soma, which is modeled as a sphere (Figures 1A–1I), whose radius is sampled from a biological distribution (see SI:Topological neuron synthesis algorithms, Algorithm 2). The number of neurites is then sampled from the biological distribution of the corresponding cell type. Each neurite is initialized with a “trunk”, the initial branch of the tree (Figure 1A) and a barcode, which is randomly selected from the set of barcodes derived from the reconstructed dendrites of the corresponding m-type.

Subsequent steps of the growth take place in a loop. Each branch of the tree is elongated as a directed random walk (Aslangul et al., 1993) with memory (see SI:Topological neuron synthesis algorithms, Algorithm 3, Figures 1A–1III). At each step, a growing tip is assigned probabilities to bifurcate, terminate, or continue that depend on the path distance from the soma and are defined by the bars of the selected barcode (Figures 1A–1III, see SI:Topological neuron synthesis algorithms, Algorithm 3). Once a bar is used, it is removed from the barcode and the growth terminates when all the bars of the input barcode have been used. As an independent step, the diameters of the tree are assigned based on diameter distributions sampled from the biological reconstructions (Figures 1A–1IV, section IV. Generation of diameters).

Initiation of neurites

The direction of neurite protrusion is particularly important for the generation of dendrites that present an orientation preference Graham and van Ooyen (2006). For some neurites, such as cortical apical dendrites, the initial direction is trivially defined as apicals typically grow toward the pia. By contrast, the outgrowth direction of basal dendrites superficially appears random and is frequently assumed to be so. An in-depth analysis reveals, however, that this assumption is inaccurate, since the orientations of a neuron’s processes are correlated. This correlation is captured in the pairwise trunk-angle distribution, which depends on the morphological type, and is used for the initiation of neurites on the soma surface (see Topological neuron synthesis algorithms, Algorithm 3). The basal dendrites’ positions on the soma are sampled randomly so that the pairwise angles between them are respected. Each neurite trunk consists of a point on the soma surface and an initial direction that is normal to the soma surface.

Branching/termination

The neuronal branching process in the TNS algorithm is based on the concept of a Galton-Watson tree (Galton and Watson, 1875), which is a discrete random tree generated by the following process. At each step, a number of offspring is independently sampled

from a distribution. A neuronal tree consists only of bifurcations, terminations, and continuations, so the accepted values for the number of offspring are: zero (termination), one (continuation), or two (bifurcation). Since the Galton-Watson tree is combinatorial, i.e., is not embedded in space, we introduced a dependency of the neuronal growth on the embedding by choosing the bifurcation/termination probabilities to depend on the path distance of the growing tip from the soma.

These probabilities are defined by the barcode assigned to the neurite, depending on its type and the respective population of reconstructed cells. Barcodes are sampled independently -with replacement- from the input population. Each growing tip of the neurite is assigned a bar bar_i from the barcode, that includes a start path distance b_i , an end path distance d_i and a bifurcation angle a_i (see SI: [Topological neuron synthesis algorithms](#)). The start (respectively end) path distance defines the probability to bifurcate (terminate) as an exponential distribution $e^{-\lambda x}$ with a free parameter λ .

$$P_B(\text{bifurcation} | d_{\text{tip}}) = \exp(\lambda(d_{\text{tip}} - b_i)) \quad (\text{Equation 1})$$

$$P_T(\text{termination} | d_{\text{tip}}) = \exp(\lambda(d_{\text{tip}} - d_i)). \quad (\text{Equation 2})$$

To check if the branch bifurcates (or terminates), a number, r is randomly sampled in the uniform distribution $U(0, 1)$ and compared to the probability of [Equations 1](#) and [2](#). If the random number r is smaller than this probability, i.e.,

$$r \leq \exp(\lambda(d_{\text{tip}} - b_i)), \quad (\text{Equation 3})$$

then a bifurcation occurs and similarly for the termination probability. If a branch does not bifurcate or terminate it continues to elongate (see [III. Elongation of neurites](#)).

If $d_{\text{tip}} < b_i$, the probability to bifurcate P_B is less than 1 and increases as the target value b_i is approached. Once $d_{\text{tip}} = b_i$, $P_B = 1$ and therefore the branch will necessarily bifurcate. The same process holds for the termination of a branch. If d_{tip} is much smaller than b_i , the probability to bifurcate is very low, as P_B is close to 0. Quantitatively, this property depends on the selection of the parameter λ of the exponential probability function. The parameter λ controls the slope of the probability distribution for bifurcation and termination (see [Figure S2B](#)).

The exponential probability of the branching and termination events implicates that they will occur before or marginally at the target positions of b_i and d_i . As a result, dendrites will bifurcate and terminate before or at the target positions, a property that introduces a small bias toward the center. This bias is very small or insignificant for the appropriate choice of randomness and targeting parameters (as indeed seen in [Figure S2A](#)). The choice of the probability function as an exponential is influenced by the need to confine dendrites within a target brain region while growing within a spatial context. If we have chosen a gaussian probability around the target start, end distances we could end up with neurons that are larger than the original and they would expand beyond the selected brain region.

The λ parameter needs to be wisely chosen in order to generate cells that are neither identical to the original barcode (high λ value) nor random and completely independent from the input (low λ value). A very steep exponential distribution (high value of λ) generates cells that are very close to the biological input and thus the variability of the synthesized cells is reduced (right panel [Figure S2B](#)). On the other hand, a very low value for λ generates cells that are almost random, since the probabilities to bifurcate and terminate are high long before the target values are approached (left panel [Figure S2B](#)). In practice, we found that if λ is chosen to be at the order of the step size $\approx 1 \mu\text{m}$, the bifurcation and termination points are stochastically chosen but are still strongly correlated with the input persistence barcodes. As a result, the generated shapes will not be identical to the input tree, thereby increasing the variability of synthesized cells, but will preserve the overall shape of the input tree, generating biologically acceptable branching structures.

Practically, even though the synthesized barcodes are statistically similar to the biological reconstructions' barcodes, the barcode loses information from the original structure. Connectivity between branches is a loss that we cannot avoid, however for a given barcode only a subset of possible connectivities can be realised (see [Connectivity between branches](#) for more information) which are consistent with biological structures. In addition, barcodes do not retain information about diameters and angles, properties that we also include in the distributions we extract from the biological reconstructions to account for this information loss.

Other synthesis algorithms ([Burke et al., 1992](#); [Koene et al., 2009](#)) sample the branching and termination probabilities from independent distributions. In TNS, the correlation of these probabilities is captured in the structure of the barcode. Because each bar encodes the start and the end path distances, the bifurcation and the termination probabilities are linked accordingly. In addition, when the growing tip bifurcates (respectively terminates), the corresponding bar is removed from the input TMD to exclude re-sampling of the same conditional probability. This keeps a record of the neuronal growth history and is essential for reproducing the branching structure of biological reconstructions.

At each bifurcation, two new branches are initiated (see [Topological neuron synthesis algorithms, Algorithm 4](#)) with initial directions depending on the bifurcation angle a_i . Depending on the neurite type, the bifurcation angles are used differently.

For basal dendrites, the optimal rule for bifurcation is to use the corresponding bifurcation angle a_i as encoded in the barcode. On the other hand, the apical tree is separated into two parts: the apical tuft, which is the densely branched subtree close to the cortical surface, and the obliques, which are the shorter branches near the soma. The apical tuft is separated from the obliques by the "apical point". This point can be accurately identified based on the persistence barcode of the apical tree, as the distance that maximizes the separation between the two modes of the bars distribution, which is bimodal if the apical tree has a tuft. The apical point distance is

computed from the persistence barcode, as the point at which a tuft is first encountered. This is a parameter that is not controlled by the user. This distance is used to modify the angles of the bifurcations during the growth of an apical dendrite. For the apical dendrites, different branching behaviors need to be adopted for the tuft and the obliques. Before the apical point, one of the branches, the major branch, which corresponds to the branch of the longest bar, follows the targeting direction (usually the orientation toward the pia). The other branch follows the angle defined by a_i . Once the apical point is reached, the apical tufts bifurcate according to the bifurcation angle a_i just like the basal dendrite.

Connectivity between branches

The decomposition of a tree into a barcode results in loss of connectivity between the branches that correspond to the bars within the barcode. In a recent publication [Kanari et al. \(2020\)](#) we explain in depth the mathematics of this process. Here we mention the basic rules of the process that defines the connectivity rules.

To recreate a tree that is close to TMD-equivalent to the original, the branch corresponding to a particular bar (b_i, d_i) in the barcode can be attached only to branches corresponding to bars (b_j, d_j), $0 < i < j$ such that:

$$d_j < d_i$$

and

$$b_j > b_i$$

This rule ensures that at each bifurcation branches can only be attached to larger branches. In topology this refers to the Elder rule (see [\(Kanari et al., 2020\)](#)) that holds in the *tmd* transformation. As a result, only a subset of trees with n branches can be generated by the TNS from a given barcode with n bars.

This rule is described in [Algorithm 4](#), in which an active branch, i.e. a branch that has not yet terminated, bifurcates into two smaller branches. One of the new branches inherits the termination distance from the bar of the active bar k (sec_1) while the second branch takes a new bar from all the available branches i that have not been used yet during the growth. The new bar i is chosen so that the bifurcation distance b_i is minimum among the available branches and in addition $d_i \leq d_k$.

Elongation of neurites

A segment is a pair of consecutive points that is determined by a vector of length l and unit vector direction $\mathbf{D}_{segment}$. Each branch grows segment by segment and independently of all other branches. The direction of the segment is a weighted sum of three unit vector terms: the cumulative memory of the directions of previous segments within a branch $\mathbf{M}$, a target vector $\mathbf{T}$, and a random vector $\mathbf{R}$ ([Koene et al., 2009](#)).

$$\mathbf{D}_{segment} = \rho \mathbf{R} + \tau \mathbf{T} + \mu \mathbf{M},$$

where $\rho + \tau + \mu = 1$.

The cumulative memory $\mathbf{M}$ is a weighted sum of the previous directions of the branch, with the weights decreasing with distance from the tip. If the current segment of the growing branch is the k^{th} segment, then the five previous steps contribute to the memory as follows:

$$\mathbf{M} = \sum_{i=1}^5 \exp(1-i) \mathbf{V}_{(k-i)}. \quad (\text{Equation 4})$$

Different weight functions were tested, but as long as the memory function decreases- faster than linearly - with the distance from the growing tip, its exact form is not relevant. The target vector $\mathbf{T}$ is defined at the beginning of each branch, for example the main apical trunk grows towards the pia (see [Topological neuron synthesis algorithms, Algorithm 3](#)). The random vector $\mathbf{R}$ component is sampled uniformly from three-dimensional space.

The tortuosity of the path is defined by these three parameters. A large randomness weight ρ results in a highly tortuous branch, approaching the limit of a simple random walk when $\rho = 1$ ([Pearson, 1905](#)). On the contrary, if the targeting weight $\tau = 1$, the branch will be a straight line in the target direction. Different combinations of the three parameters (τ, ρ, μ) can generate more or less meandering branches and can reproduce the large diversity of dendritic sections (see [Figure S2A](#)).

For a dendrite, the effects of randomness and targeting are more complex. Synthesized cells for a set of varying combinations of randomness and targeting are presented in [Figure S2C](#). For smaller targeting ($\tau \approx 0.1-0.2$) and larger randomness ($\rho \approx 0.4$) the synthesized dendrites are too short to reach the original dendritic extent, as demonstrated in the corresponding persistence diagrams of radial distances ([Figure S2C](#)). For a broad range of parameters ($\rho \leq 0.3$ and $0.1 \leq \tau$), the radial distances of synthesized trees ([Figure S2C](#), red) approximate accurately the reconstructed cells ([Figure S2C](#), blue).

The segment length is drawn from a normal distribution with parameters $L_{mean} = 1 \mu\text{m}$, $L_{std} = 0.2 \mu\text{m}$. The weights are normalized so that $\tau + \rho + \mu = 1$. Therefore, only two of the input parameters need to be chosen by the user. The different shapes that can be generated by combinations of the weights are presented in [Figure S2C](#).

Generation of diameters

In addition to the topology of a neuron, the thickness of neuronal branches must be assigned accurately to ensure a valid functional role of the cell ([Cuntz et al., 2007](#); [Koene et al., 2009](#); [van Elburg and van Ooyen, 2010](#); [Bird and Cuntz, 2016](#)). Despite recent progresses in imaging techniques enabling the generation of a large number of reconstructions ([Peng, 2008](#); [Haberl et al., 2014](#);

Economo et al., 2016), their resolution is still too limited to allow for precise measurement of diameters that are typically on the order of a few microns. As a result, we observe a large variability in thickness profiles, even among reconstructions of cells from the same m-type. In addition, swelling of diameters may occur that is usually considered as a staining artifact (Conde-Sousa et al., 2016).

To faithfully reproduce diameters while filtering out any experimental artifacts, we first extract relevant properties from a population of cells for each m-type, and use these distributions to generate new diameters. We extract distributions for the trunk and the tip diameters, the taper rates, the Rall and sibling ratios at each bifurcation (see [Morphometrics glossary](#)). These distributions are truncated to the 5% and 95% percentiles to remove outliers. For the taper rate, which is particularly noisy, we allowed larger truncation. Indeed taper rates are very sensitive to artefacts such as swelling, and present a very broad distribution of values that is not compatible with dendritic thickness without artefacts.

The process of diameter synthesis starts from the root and proceeds towards the tips of each neurite. First, a trunk diameter is sampled from the biological distribution. Then, within a section, the diameters are assigned according to a sampled taper rate. At each bifurcation, the two daughter diameters are assigned according to sampled Rall and sibling ratios. If the diameters become smaller than a sampled tip diameter, the diameter is set to the tip diameter. Once all the diameters have been assigned a check is performed and if any of the obtained tip diameters is larger than the biological tip diameters, the process restarts by reducing the trunk diameter, until the tip diameters are small enough to be consistent with the biological distribution. This restart is rare in practice but ensures that the diameters of a neurite will be proportional to its size, as smaller neurites need to have smaller trunk diameters to achieve an appropriate distribution of tip diameters. In the future this could be achieved by introducing a relevant feature correlation.

This algorithm results in dendritic thickness of synthesized cells that is statistically similar to the reconstructed cells, minimizes experimental artifacts, and creates thicknesses that monotonically decrease with distance from the soma, a property ensuring that biophysical principles of dendrites (Cuntz et al., 2007) are reproduced.

Topological neuron synthesis algorithms

The TNS algorithm consists of three main components: the initiation ([Algorithm 2](#)), the elongation (see [Algorithm 3](#)) and the branching ([Algorithm 4](#)) of neurites.

The first part of a neuron to be generated is the cell body, i.e., the soma (see Soma generation: initiation of neurites), whose radius is sampled from a biological distribution. The number and the orientation of the neurites are sampled from the biological distributions.

A segment is defined by a length L and a direction v_k , specified by a unit vector. The direction of the segment is a weighted sum of three unit vector terms: the cumulative memory of the directions of previous segments within a branch M , a target vector T , and a random vector R .

Each growing tip is assigned a bar bar_i , sampled from the barcode, that includes a starting path distance b_i , an ending path distance d_i and a bifurcation angle a_i .

At a bifurcation, two new branches are generated, and the directions of the daughter branches depend on the bifurcation angle a_i .

Quantification and Statistical Analysis

Single-cell validation

In order to validate the quality of single cells and identify individuals of poor quality within the synthesized population, the distributions of key features F of each cell are compared against a set of reconstructed cells. To measure a cell's difference from the reconstructed cells, we compute a statistical score that corresponds to a normalized distance, which is the difference between the median value of the test cell T and the median value of the reconstructed population P_r , divided by the standard deviation of the reconstructed population P_r :

$$F_{norm} = \frac{\tilde{F}_{P_r} - \tilde{F}_T}{\sigma(F_{P_r})}. \quad (\text{Equations 5})$$

Identifying outliers of synthesized cells

As an additional validation of synthesized cells, we developed a method to identify outliers in the population of synthesized neurons, by comparing each synthesized cell to the population of reconstructed cells. The percentage of detected outliers in the synthesized population gives a measure of accuracy of the synthesis process (see [Table S1](#)).

To identify outliers, we compare the distributions of key features of each neuron, such as section lengths, bifurcation angles etc, to the reconstructed population. We quantify the distance between the distributions with the ratio of the absolute value of the Difference Between the Medians divided the Overall Visible Spread (DBM/OVS, (see [Figure S3A](#))).

The MVS score is defined as follows:

$$MVS(F) = \frac{|\tilde{F}_{P_r} - \tilde{F}_{P_s}|}{OVS(F_{P_s}, F_{P_b})} \quad (\text{Equation 6})$$

This is an intuitive measure of the difference between the medians of the two distributions with respect to their joint dispersion. The Overall Visible Spread is usually defined as the range between the minimum 25th percentile of the two distributions and the maximum 75th percentile of the distributions. Its minimum is 0 when the medians of the distributions coincide, and the two distributions are very close to each other. Its maximal value is 1 in the special case where the smaller median coincides with the smaller 25th percentile and the larger median with the largest 75th percentile. The closer the DBM/OVS gets to 1, the larger is the difference of medians with respect to the overall spread of the distributions, making it likely to reject the neuron in question. We modify this definition due to the large variance of features within the set of reconstructed cells, and use the OVS defined as the range between the 10th to 90th percentile of the population.

The DBM/OVS measure works better than other measures of standardized difference of means, such as [Hedges \(1981\)](#)'s, for non-symmetric distributions (see two last references). A cell is considered an outlier when at least one of the key features mentioned above is outside of certain feature-specific thresholds. We choose the thresholds so that the reconstructed biological cells are not rejected as outliers, since they represent the gold standard. One can imagine the thresholds as defining a hypercube in the space of features. If the feature vector of a synthesized cell falls out of this cube, it is rejected as an outlier (see [Figure S3B](#)).

As an example, we provide the results for a zero-outlier m-type (L5_UPC) and for one with outliers (L5_CHC) in [Figure S3B](#). The large number (28) of examples for layer 5 UPC cells results in the DBM/OVS of synthesized cells to be within the limits. On the contrary, layer 5 CHC consists of only three reconstructed cells, and thus the number of outliers for this m-type is much larger. The spread of DBM/OVS for the L5_UPC is much smaller than the L5_CHC. A large number of available reconstructions is clearly essential for the accurate definition of the synthesis input. Detailed results are given in [Table S1](#).

Population-to-population validation

Synthesis is validated at a population level by comparing the distributions of a large number of morphological features (see [Table S2](#)) to those of the reconstructed cells per morphological type. Essential features F , such as the degree of the dendritic tree (number of terminations), the branch orders, the number of sections, the total length per neurite, the radial and path distances from the soma shown in [Figure 4](#). Each of the distributions of morphological features F is compared between a population of synthesized cells P_c and of reconstructed cells P_r with the MVS score 6. The MVS score can be used as a validation score between two populations for each morphological feature F .

Electrical simulation of synthesized cells

A biophysically detailed electrical model (e-model) for a L3_TPC was applied in the synthesized morphologies to assess how well the electrical behavior generated by the synthesized morphologies compares with their reconstructed counterparts. The original e-model was obtained by applying a multi-objective optimization of the electrical parameters as described in [Markram et al., 2015](#) to a reconstructed morphology. The e-model has 31 parameters that are used to control the maximal conductance of the ion channels in four morphological areas (somatic, axonal, basal, and apical) and the calcium dynamics and the decay constant of Na channels along the dendrite. Note that in the case of synthesized cells, only dendrites are computationally generated; the axonal morphology is copied from a reconstructed cell. The e-model consists of Hodgkin-Huxley-based channel models for persistent and transient Na/K, high- and low-voltage activated Ca, Kv3.1, Ih and SK calcium-activated potassium channels. When the e-model is instantiated, the axon is replaced by a shorter axon initial segment with diameters based on the original morphology. The constraints consist of electrical features extracted (eFEL) from somatic whole-cell current clamp recordings and dendritic back-propagating action potential features obtained from literature. The stimulation currents used in the experiments and models are scaled by the spiking threshold currents of the cells. The e-model was applied to the synthesized morphologies and, as in the morphological validation, the Fnorm was used to quantify how well the resulting morpho-electrical combination matches with the statistics of the original experimental data.

Cell Reports, Volume 39

Supplemental information

Computational synthesis of cortical dendritic morphologies

Lida Kanari, Hugo Dictus, Athanassia Chalimourda, Alexis Arnaudon, Werner Van Geit, Benoit Coste, Julian Shillcock, Kathryn Hess, and Henry Markram

M-type (IN)	Outliers	Cell count	M-type (PC)	Outliers	Cell count
L1_DAC	5	16	L2_IPC	6	4
L1_HAC	4	27	L2_TPC:A	0	4
L1_LAC	2	10	L2_TPC:B	0	33
L1_NGC-DA	0	19	L3_TPC:A	0	46
L1_NGC-SA	7	19	L3_TPC:B	0	17
L1_SAC	2	14	L4_SSC	0	10
L23_BTC	0	14	L4_TPC	0	37
L23_CHC	0	10	L4_UPC	0	33
L23_DBC	1	18	L5_TPC:A	0	72
L23_LBC	0	41	L5_TPC:B	0	42
L23_MC	3	29	L5_TPC:C	0	25
L23_NBC	0	18	L5_UPC	0	28
L23_NGC	1	4	L6_BPC	0	25
L23_SBC	0	22	L6_HPC	0	23
L4_BTC	4	7	L6_IPC	1	27
L4_DBC	4	7	L6_TPC:A	0	25
L4_LBC	0	31	L6_TPC:C	0	19
L4_MC	0	17	L6_UPC	0	17
L4_NBC	0	25			
L4_SBC	0	14			
L5_BP	0	3			
L5_BTC	1	11			
L5_CHC	16	3			
L5_DBC	3	9			
L5_LBC	1	21			
L5_MC	0	38			
L5_NBC	0	22			
L5_SBC	4	4			
L6_LBC	0	24			
L6_MC	0	20			
L6_NBC	0	11			
L6_SBC	2	5			

Table S1: Outlier analysis. Number of outliers (and input cells) according to statistical score, per interneuron and pyramidal cell type. Cell types with 1 or 2 reconstructions are not reported. 100 cells were synthesized for each m-type and the number of outliers based on the DBM/OVS for the 10 – 90% percentile of the input population was computed. Related to Figure 4.

Mtype	Cell number	2 cells	5 cells	10 cells	15 cells	All
L1_DAC	16	0.22	0.20	0.18	0.17	0.14
L1_HAC	27	0.16	0.20	0.16	0.16	0.16
L1_NGC-DA	19	0.17	0.19	0.12	0.13	0.22
L1_NGC-SA	19	0.14	0.13	0.15	0.13	0.13
L23_DBC	18	0.20	0.16	0.12	0.13	0.16
L23_LBC	41	0.21	0.18	0.19	0.16	0.11
L23_MC	29	0.26	0.18	0.18	0.14	0.30
L23_NBC	18	0.22	0.24	0.24	0.21	0.22
L23_SBC	22	0.23	0.23	0.18	0.26	0.26
L2_TPC:B	33	0.21	0.18	0.18	0.17	0.15
L3_TPC:A	51	0.21	0.15	0.15	0.12	0.12
L4_LBC	31	0.15	0.14	0.19	0.14	0.14
L4_MC	17	0.28	0.19	0.16	0.20	0.21
L4_NBC	25	0.26	0.13	0.17	0.16	0.12
L4_TPC	36	0.22	0.16	0.13	0.13	0.09
L4_UPC	34	0.30	0.18	0.14	0.13	0.11
L5_LBC	21	0.20	0.11	0.11	0.20	0.12
L5_MC	38	0.22	0.16	0.21	0.13	0.12
L5_NBC	22	0.21	0.13	0.18	0.16	0.17
L5_TPC:A	64	0.19	0.18	0.15	0.14	0.19
L5_TPC:B	38	0.19	0.17	0.17	0.17	0.16
L5_TPC:C	30	0.19	0.19	0.12	0.14	0.13
L5_UPC	27	0.14	0.16	0.13	0.13	0.13
L6_BPC	28	0.20	0.16	0.13	0.14	0.20
L6_HPC	21	0.23	0.18	0.19	0.16	0.15
L6_IPC	26	0.22	0.17	0.15	0.18	0.27
L6_LBC	24	0.20	0.14	0.14	0.16	0.22
L6_MC	20	0.19	0.14	0.18	0.17	0.19
L6_TPC:A	20	0.15	0.18	0.17	0.12	0.13
L6_TPC:C	18	0.26	0.18	0.20	0.20	0.18
L6_UPC	17	0.21	0.16	0.16	0.17	0.25

Table S4: Number of input cells statistical analysis. Statistical score per mtype for different number of input cells from the population of the respective reconstructed cells. Related to Figure 4.

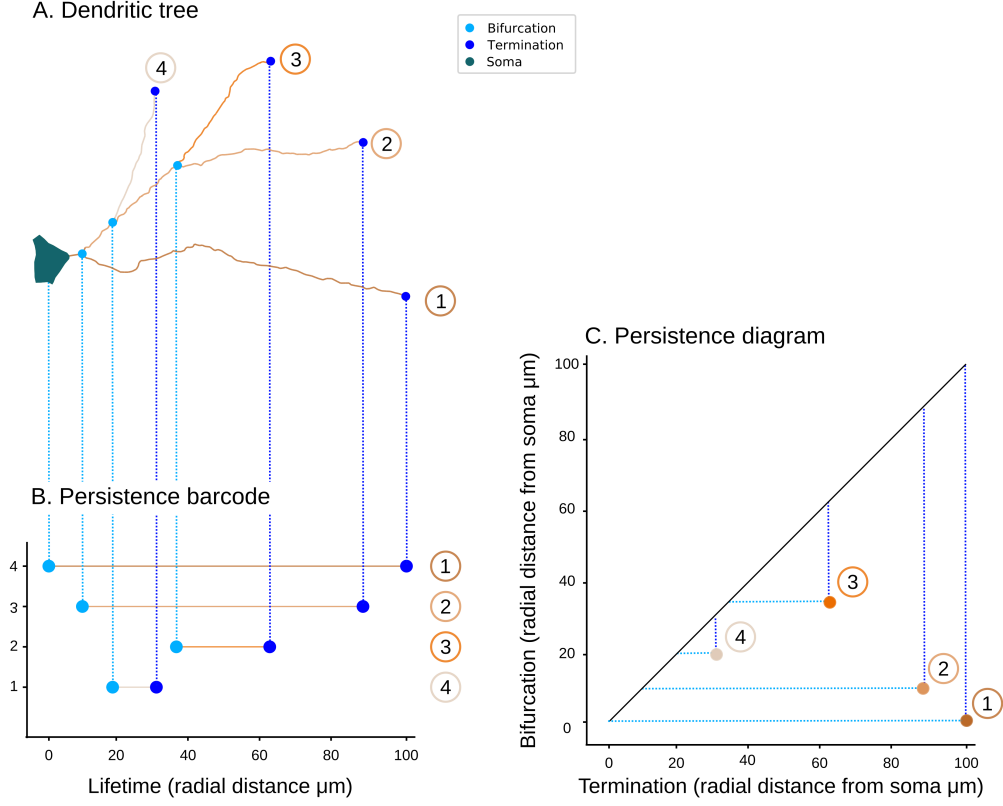


Figure S1: Illustration of the TMD algorithm. A rooted tree (A) is transformed with the TMD algorithm into the corresponding persistence barcode (B) and persistence diagram (C). The root or soma, colored in blue-green, defines the reference of zero distance (for example, for radial distances). Continuations are presented in shades of beige, bifurcations in light blue and terminations in dark blue. The correspondence between the tree (A) and its extracted barcode (B) is indicated by the assignment of the same node numbers for each termination in tree A and each bar in the barcode B. Each bar represents the lifetime of a branch within the tree, defined by a bifurcation and termination radial distance from the soma. We take as an example branches #2 and #3, which are siblings. The branch with the largest radial distance from the soma, here branch #2, persists, while the smallest sibling #3 “dies” generating a bar within the persistence barcode which encodes the initial and final radial distances (b_i, d_i). The longest bar (1) corresponds to the branch within the tree with the largest termination radial distance. Equivalent to the persistence barcode (B), the persistence diagram (C) represents the start and end radial distances in a 2D plane. All radial distances are measured in μm . Related to Figure 1.

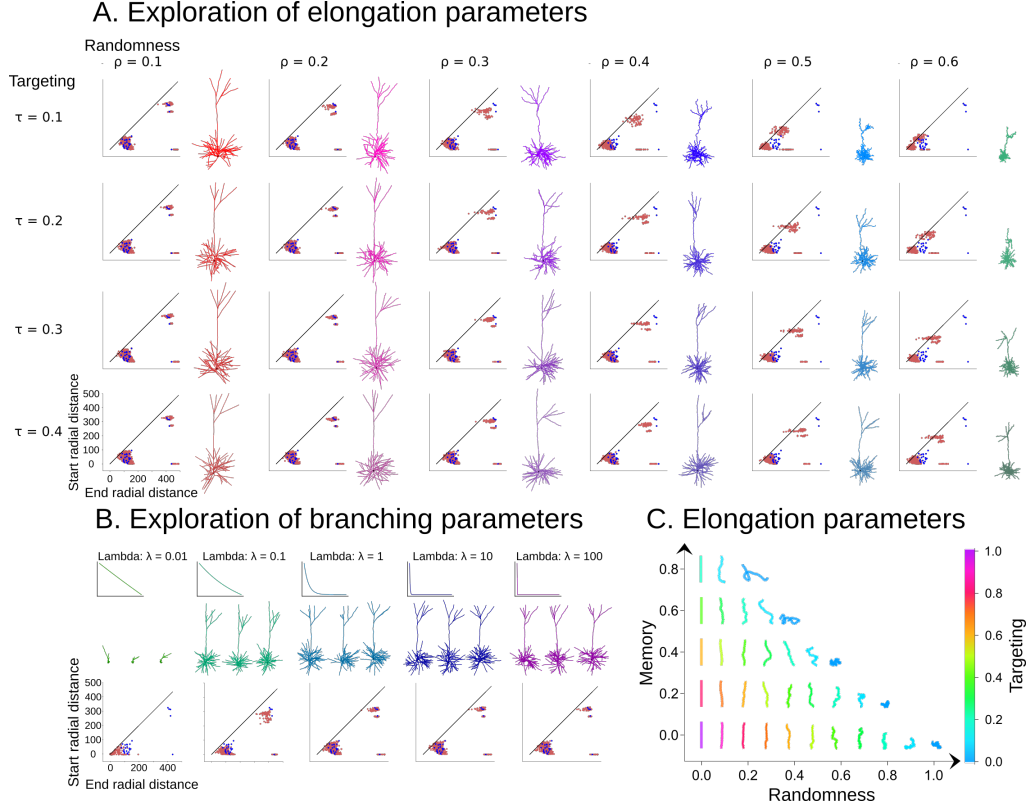


Figure S2: Parameter sensitivity analysis. A. Sensitivity analysis for elongation parameters. Synthesized dendrites with different input parameters, and the corresponding radial distance persistence diagrams for each tree. Blue diagrams correspond to the target reconstructed cell, red diagrams correspond to synthesized trees for different parameters (x -axis corresponds to randomness weight ρ , y -axis corresponds to targeting weight τ). B. Sensitivity analysis for the bifurcation/termination parameter λ . Synthesized dendrites with different input parameters λ , and the corresponding radial distance persistence diagrams for each tree. Blue diagrams correspond to the target reconstructed cell, red diagrams correspond to synthesized dendrites for $\lambda : 0.01 - 100$ from left to right. C. Synthesized branches for different input parameters. The y -axis represents the memory weight μ , the x -axis represents the randomness weight ρ , and the colorbar corresponds to the targeting weight τ . Related to Figure 1.

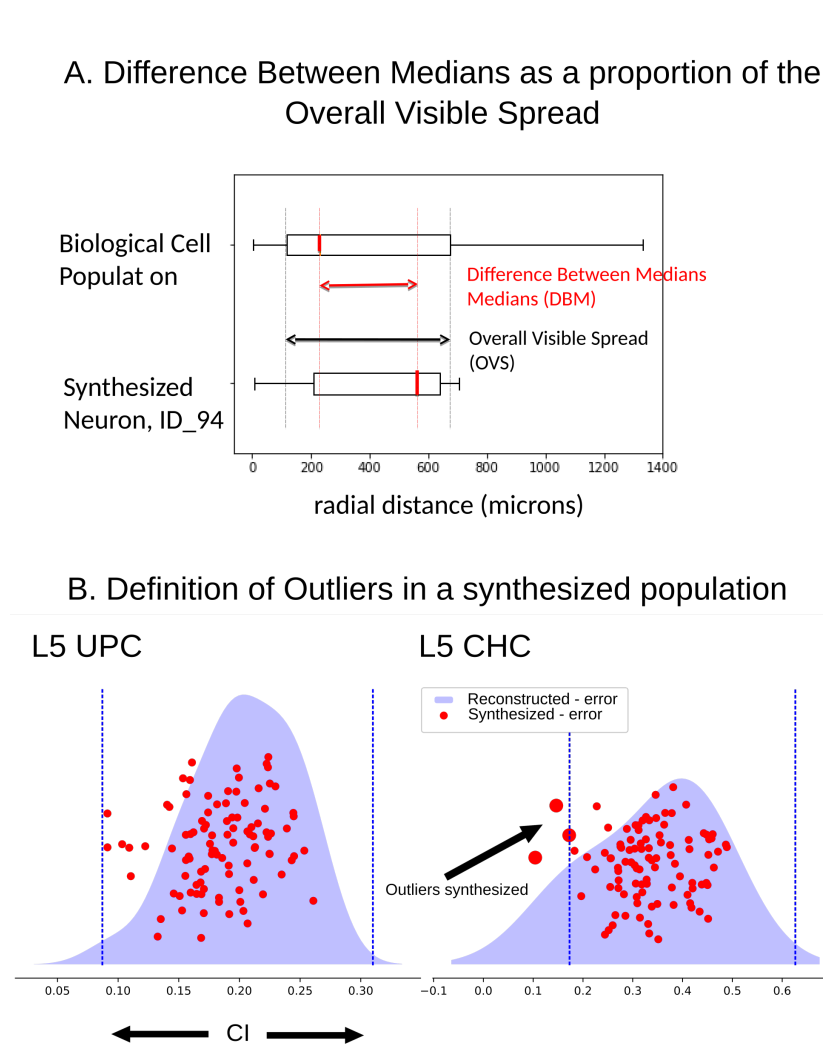


Figure S3: Validation of synthesized dendrites. A. Statistical measurement for the identification of outliers: the absolute value of the Difference Between the Medians divided by the Overall Visible Spread (DBM/OVS). B. Detection of outliers in synthesized cells. For two population of cells (layer 5 UPC and CHC), the distribution of DBM/OVS for reconstructed cells is shown in blue, and the corresponding DBM/OVS of synthesized cells in red dots. The dotted lines define the OVS (for 90th percentile) of the distribution. Anything that is outside the OVS is considered an outlier. The large number (28) of examples for layer 5 UPC cells results in the DBM/OVS of synthesized cells to be within the limits. On the contrary, layer 5 CHC consists of only three reconstructed cells, and thus the number of outliers for this m-type is much larger. The spread of DBM/OVS for the L5_UPC is much smaller than the L5_CHC. Related to Figure 4.

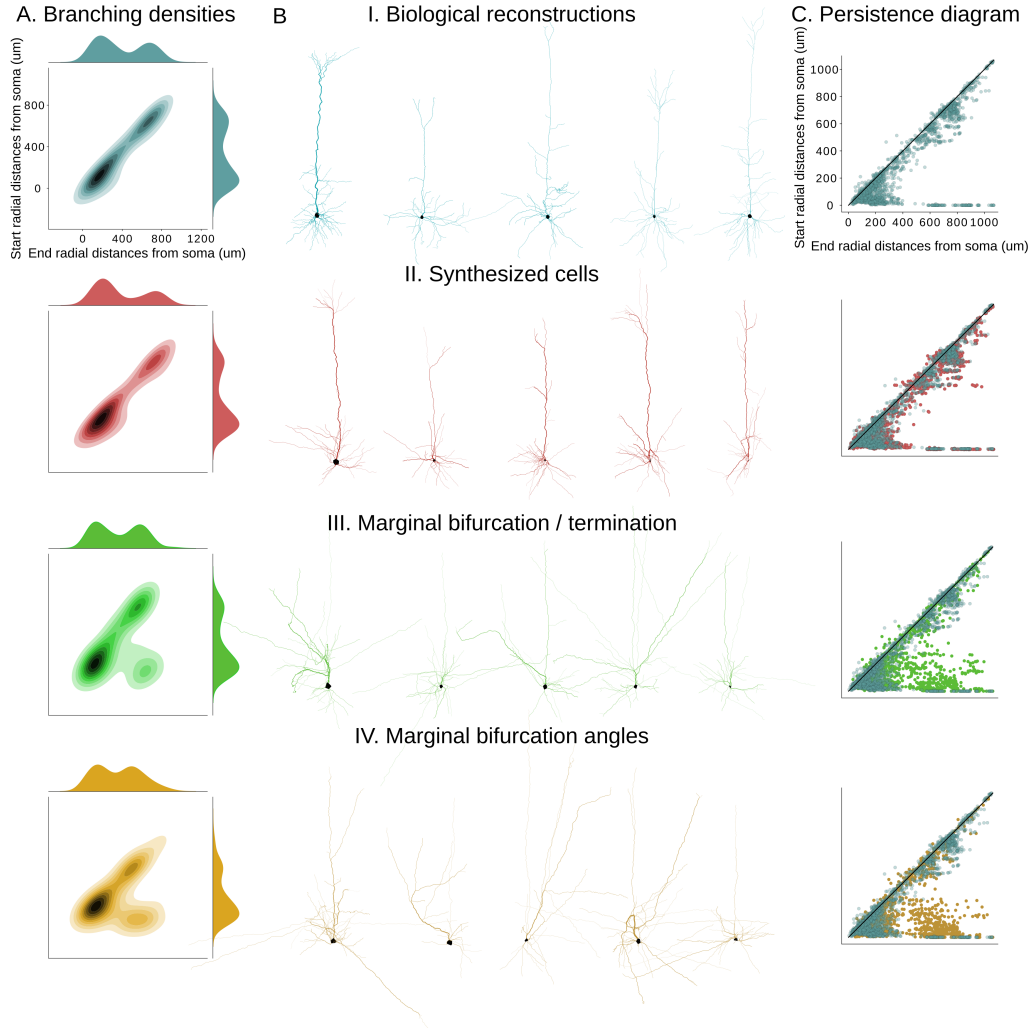


Figure S4: Comparison of synthesis methods. Comparison of synthesized cells for different synthesis methods. A. Density and marginal projections of persistence diagrams for reconstructed cells (I), synthesized cells (II), synthesized without correlation of branching / termination (III), and synthesized without correlation between branching and bifurcation angles (IV). B. Examples for the same data. C. Respective persistence diagrams. Related to Figure 3.

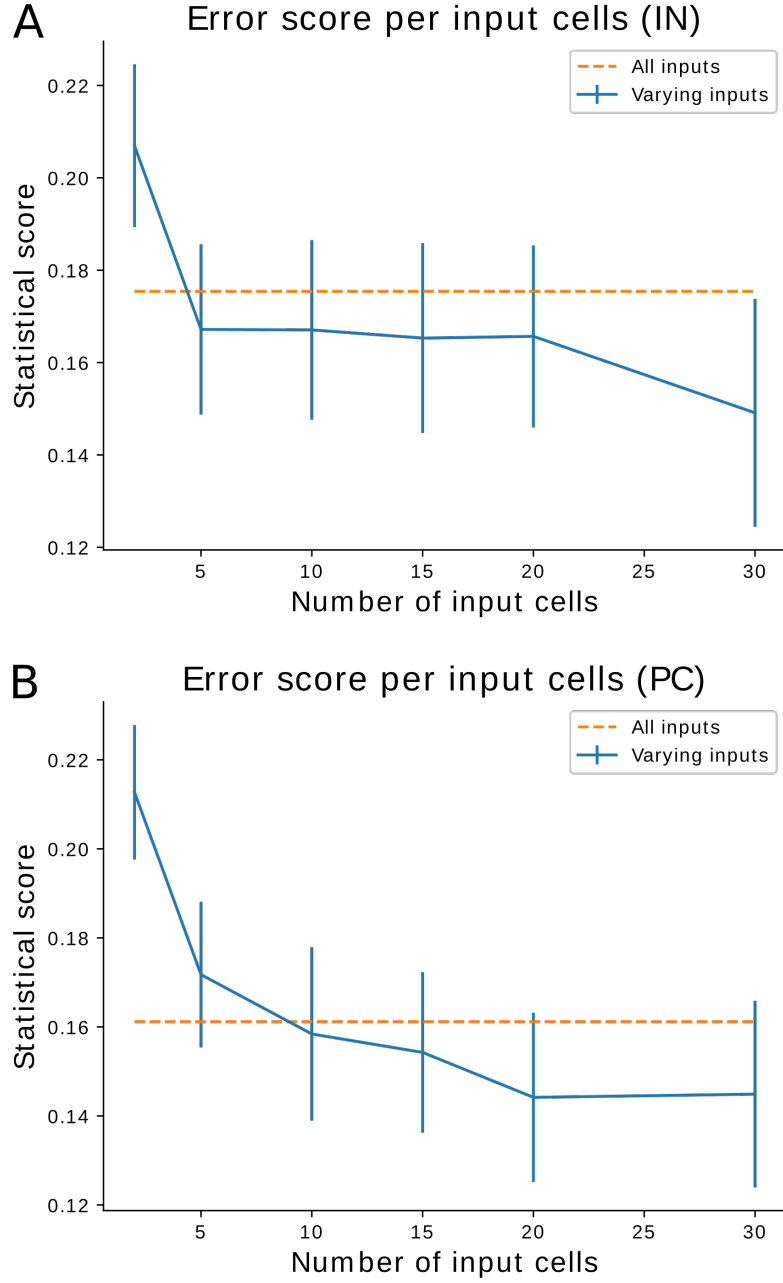


Figure S5: Outlier analysis. Comparison of synthesized cells for a varying number of input cells (2, 5, 10, 15, 20, 30). A. Statistical scores between synthesized and reconstructed cells for all interneuron cell types with more than 15 reconstructions. B. Statistical scores between synthesized and reconstructed cells for all pyramidal cell types with more than 15 reconstructions. Related to Figure 4.

A. Transfer Impedance profiles on dendrites B. Cross-validation of electrical properties

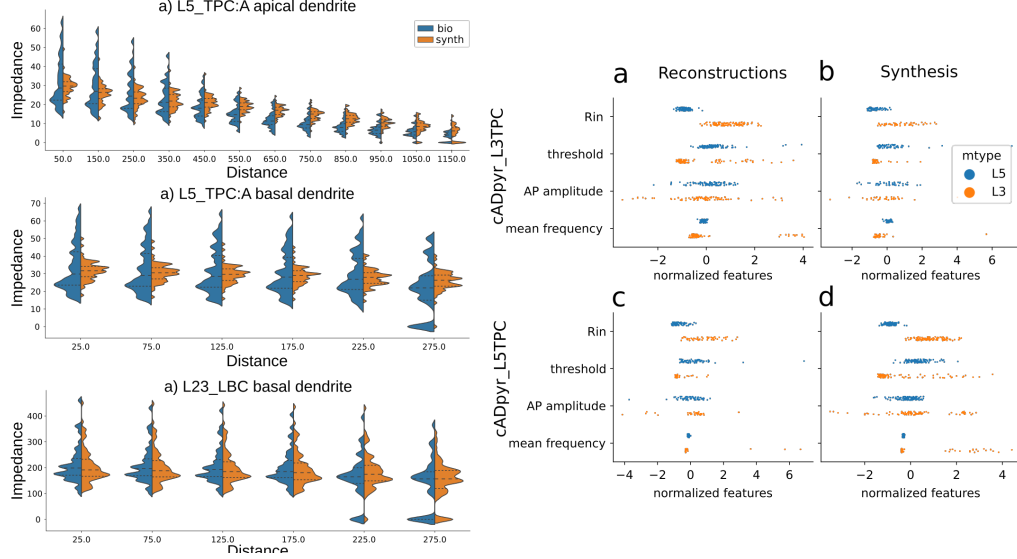


Figure S6: Electrical Validations. A. Transfer impedance profiles of (a) L5_TPC:A apical, (b) L5_TPC:A basal with cAD etype and (c) L23_LBC basal with etype cAC. Transfer impedance profiles are obtained via the computation of transfer impedance between all segments and soma at 10Hz with NEURON simulator, and averaged in path distance bins (50micron for basal and 100micron for apical) for each cell. We then compare the histogram of these average transfer impedance across the two populations of reconstruction and synthesized cells. We observe that the measurement of impedance is very sensitive to diameters. As a result the overall standard deviation is smaller for synthesized cells, mostly due to the more consistent (smooth) diameters, and a small deviation for apical dendrites near the middle section, also due to small differences in trunk diameters. B. Cross validation of electrical models with two distinct neuronal populations to test if synthesized morphologies are sufficiently different across morphological types to produce different electrophysiological signatures under distinct e-models. a. Electrical features of cADpyr_L3TPC emodel applied on L3 and L5 pyramidal cells of reconstructed morphologies. b. Electrical features of cADpyr_L3TPC emodel applied on L3 and L5 pyramidal cells of synthesized morphologies. c. Electrical features of cADpyr_L5TPC emodel applied on L3 and L5 pyramidal cells of reconstructed morphologies. d. Electrical features of cADpyr_L5TPC emodel applied on L3 and L5 pyramidal cells of synthesized morphologies. Electrical features show distinct properties that depend on the neuronal morphologies for both reconstructed and synthesized cells. Related to Figure 4.

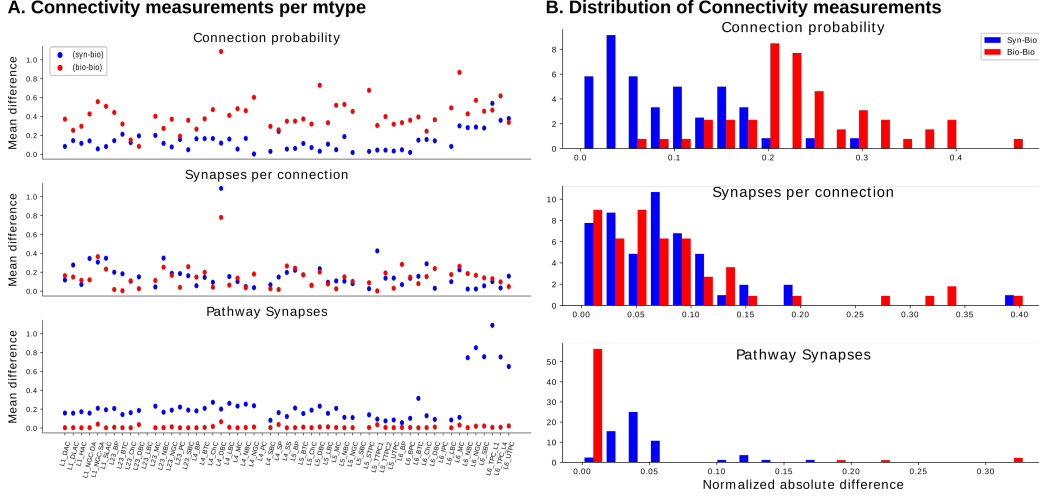


Figure S7: Statistical results for connectivity. A. Absolute difference between the computed connectivity measurements between synthesized (blue) and reconstructed microcircuits (red). For all three measurements: connection probability, synapses per connection and pathway synapses the difference between the mean values is less than one standard deviation, in which case the normalized difference would be one. For reference, the difference between two instances of reconstructed microcircuits is shown in red. B. Absolute difference between the computed connectivity measurements between synthesized and reconstructed microcircuits divided by the mean reconstructed value for each respective mtype. For all three measurements: connection probability, synapses per connection and Synapses per pathway the difference between the mean values is less than one standard deviation. For reference, the difference between two instances of reconstructed microcircuits is shown in red. Related to Figure 6.

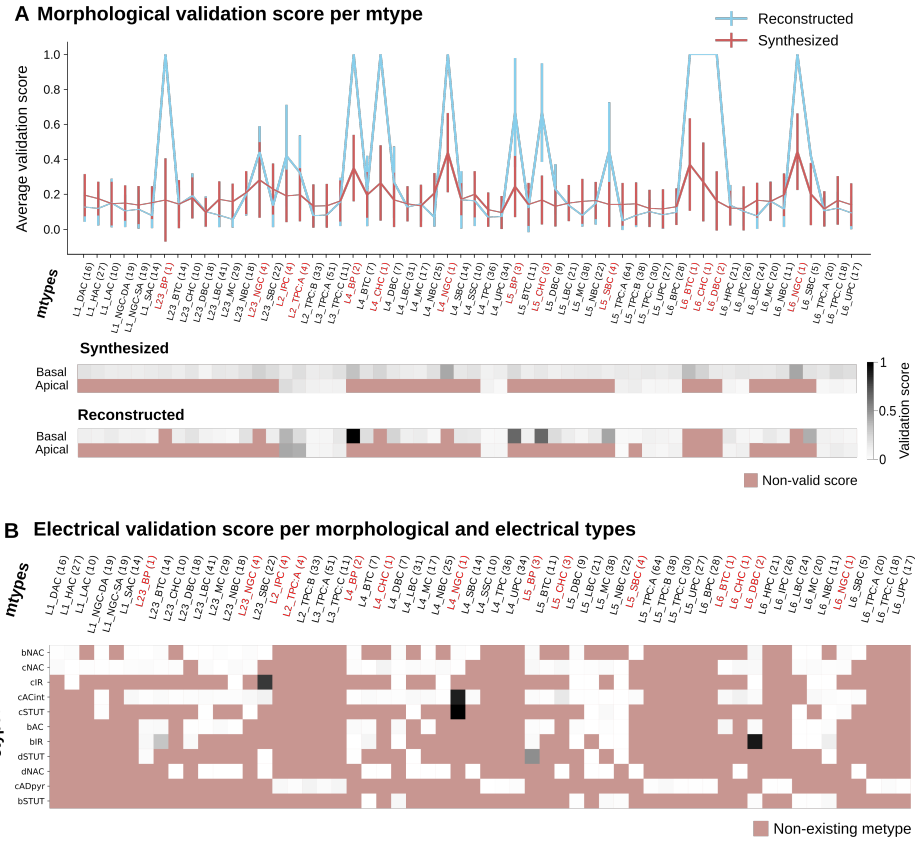


Figure S8: Morphoelectrical validation. A. Morphological validation of synthesized cells per mtype (the sample size is shown in parenthesis, mtypes with fewer than 5 exemplars are annotated in red). Validation (MVS) scores averaged over all features for each mtype (top) for reconstructed (blue) and synthesized (red) cells. Comparison of average (MVS) scores between reconstructed - synthesized cells (top) and within reconstructed cells (bottom) for basal and apical dendrites. B. MVS scores between z-scores of reconstructed and synthesized cells' electrical features for each morpho-electrical type. Related to Figure 4.

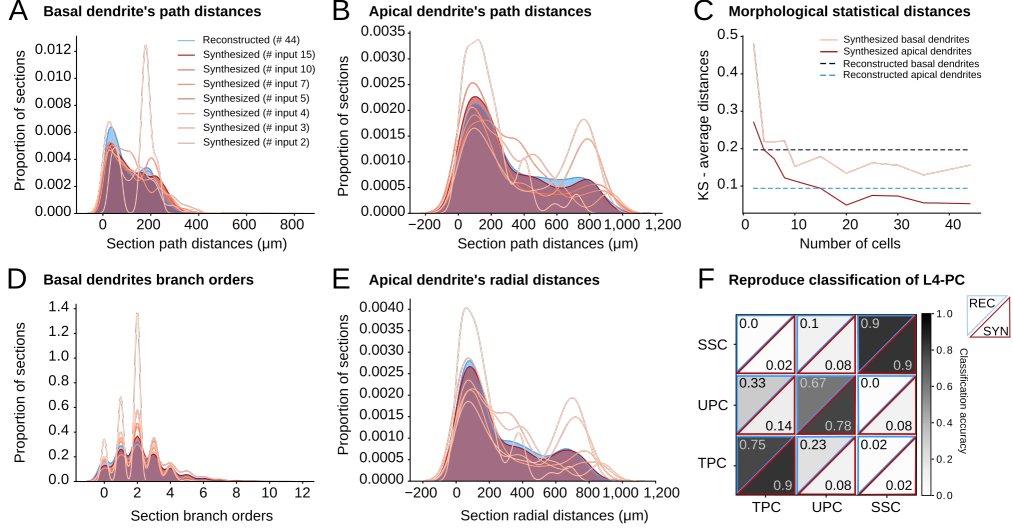


Figure S9: Morphological diversity. Comparison of dendrites from 44 reconstructed L4_TPC cells (in blue) to synthesized dendrites (based on subsets of increasing numbers of cells from the original population used as inputs: from 2 to 15, red shades from lighter to darker). Comparison of path distance (A, direct input) and branch order (B, emergent property) for basal dendrites. Comparison of path distances (D, direct input) and radial distance (E, emergent property) for apical dendrites. The original distributions are well approximated by a subset of input cells (15 out of 44). C. Average statistical (Kolmogorov-Smirnov) distance for numerous morphometrics, within reconstructed cells (in blue) and between reconstructed and synthesized cells (in red) as a function of increasing synthesis inputs. F. TMD based classification of three L4_PC types for reconstructed (top left, blue) and synthesized (bottom right, red) cells. Classification accuracy is same or higher for the synthesized population. Related to Figure 5.