Neuronal morphologies: the shapes of thoughts

THÈSE N° 8255 (2018)

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Lida KANARI

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2018

To my parents who made me curious and are always by my side.
To Eleftherios for the inspiration and the support.

"The most merciful thing in the world, I think, is the inability of the human mind to correlate all its contents. We live on a placid island of ignorance in the midst of black seas of infinity, and it was not meant that we should voyage far."

– H.P. Lovecraft

Acknowledgements

No scientific discovery can be made without the collective effort of many individuals, and a thesis is certainly no exception. First of all, I would like to thank Henry Markram, my thesis advisor, for giving me the opportunity to work on this interesting topic. During my interview someone asked "how can you do a PhD in neuroscience, you are a mathematician" and Henry replied "of course she can". Thank you Henry for believing in me, for your kind guidance and for challenging me through this long and interesting journey.

I cannot thank enough Julian Shillcock for his infinite support, for being with me at every step of this journey, for believing in my crazy ideas (they were not that crazy in the end, right?), for sending me to great conferences, and of course for transforming all my writings into human-readable text. Thank you Julian for the philosophical discussions over coffee and the horror-movie nights.

I also have to thank Joe Graham for introducing me to neuroscience, for his patience to teach me what a neuron actually is and for convincing me that at some point I might even be able to tell neuron types apart. Thank you so much Joe for your scientific guidance and the fun times in Lausanne! Also many thanks to Guy Atenekeng for the support through all my "impossible" tasks and for his invaluable input and his comments on my work. I also have to infinitely thank Nancy Chalimourda for the long discussions on our quest to find "bad and good" neurons, for her valuable scientific input and most of all for her undivided support at all levels.

I would like to greatly thank Ran Levi and Kathryn Hess (and the awesome AT-team: Pawel, Martina, Kate, Gard) for the interesting discussions, the great ideas, but mostly for reminding me that I am still a mathematician. Thank you so much Kathryn for being an unofficial -but very much significant- supervisor in my thesis. Special thanks to Pawel for the constant support and for always being eager for challenging discussions. Also many thanks to Idan Segev (and of course all the members of his team and especially Yair and Oren) for broadening my neuroscientific horizons, for the inspiration, the opportunity to apply my work on interesting problems and for inviting me to visit his beautiful country.

Abstract

The mammalian brain, one of the most fascinating systems in nature, is a complex biological structure that has kept scientists busy for over a century. Many of the brain's mysteries have been unraveled due to the enormous efforts of the scientific community, but yet many questions remain unsolved. The detailed drawings of Ramón y Cajal revealed the hidden structure of the brain, identifying the neurons as its fundamental structural and functional units. Although a significant amount of experimental reconstructions have been gathered over the past years, neuronal morphologies still remain one of the unsolved riddles of the brain. Why is neuronal diversity important for the functionality of the brain and how do neuronal morphologies "shape" our thoughts?

To address these questions one needs to characterize the various shapes of neuronal morphologies. Traditionally, this task has been performed by using a set of morphological features, such as total length, branch orders and asymmetry. However, these features focus on a specific morphological aspect thereby causing a significant information loss from the original structure. Inspired by algebraic topology, I have conceived a topological descriptor of neuronal trees that couples the topology of a tree with the geometric features of its structure, retaining more details of the original morphology than traditional morphometrics. This descriptor has proved to be very powerful in discriminating several neuronal types into concrete groups based on morphological grounds, and has lead to the discovery of two distinct classes of pyramidal cells in the human cortex. In addition, the Topological Morphology Descriptor is important for the generation of artificial cells whose morphologies remain faithful to the biological ones.

Neurons of the same morphological type have similar topological and geometric characteristics, therefore appearing to be highly structured. However, it is still unknown to what extent the complex neuronal morphology is shaped by the genetic information of an organism and to what extent it arises from stochastic processes. To study the impact of randomness and structure of neuronal morphologies on the connectivity of the network they form, I compared the properties of networks that arise from different artificially generated morphologies, ranging from random walks to constrained branching

structures, against those of biological networks and computational reconstructions built from biological morphologies. Surprisingly, networks that are generated from almost random morphologies share a lot of common properties with biological networks, such as the spatial clustering of connections and the common neighbor effect, indicating that stochastic processes that take place during development, contribute significantly to the observed neuronal shapes.

This thesis resolves a number of the mysteries of neuronal morphologies and questions our beliefs about the role of randomness in the formation of the brain. Thus, it brings us closer to understanding the fundamental differences among morphologies, and how randomness and structure are combined together to generate one of the most complex biological systems.

Keywords: Neuronal morphology, Artificial generation, Neuronal topology, Synthesis, Brain connectivity, Axons, Dendrites, Random walk, Random network, Generative model

Abstract

Le cerveau d'un mammifère, l'un des systèmes les plus fascinants de la nature, est une structure biologique complexe étudiée depuis plus d'un siècle. De nombreux mystères concernant le cerveau ont été élucidés grâce aux énormes efforts de la communauté scientifique, mais beaucoup de questions restent encore sans réponse. Les dessins détaillés de Ramón y Cajal ont révélé la structure cachée du cerveau, identifiant les neurones comme ses unités structurelles et fonctionnelles fondamentales. Malgré la quantité importante de reconstructions expérimentales recueillies au cours des dernières années, il reste encore bon nombre de points obscurs concernant les morphologies neuronales. De quelle maniére la diversité neuronale est-elle importante pour la fonctionnalité du cerveau? Comment les neurones du cerveau forment-ils nos pensées?

Pour répondre à ces questions, il faut d'abord caractériser les formes différentes de morphologies neuronales. Traditionnellement, cette tâche a été effectuée basé sur un ensemble d'attributs morphologiques, tels que la longueur totale, les ordres de branchement et l'asymétrie. Puisque ces attributs ne reflétent que certains aspects morphologiques spécifiques, cela résulte en une perte d'information significative par rapport à la structure originale. Inspiré par la topologie algébrique, j'ai conçu un descripteur topologique d'arbres neuronaux qui couple la topologie d'un arbre aux caractéristiques géométriques de sa structure, conservant plus de détails sur la morphologie originale que les mesures traditionnels. Ce descripteur classifie de façon efficace plusieurs types des neurones en groupes concrets selon leurs motifs morphologiques. Il a également permis la découverte de deux classes distinctes de cellules pyramidales dans le cortex humain. En plus, le Descripteur Topologique des Morphologies joue un rôle crucial dans la synthése de cellules artificielles dont les morphologies restent fidèles à la biologie.

Les neurones du même type morphologique ont des caractéristiques topologiques et géométriques similaires, ce qui implique qu'ils sont très structurés. Cependant, c'est encore inconnu quelle part de la morphologie neuronale est résulte de l'information génétique d'un organisme et quelle part résulte de processus stochastiques. Afin d'étudier l'impact de l'aléatoire et de la structure des morphologies neuronales sur la connectivité du réseau qu'elles forment, j'ai comparé les propriétés des réseaux

de neurones qui sont générés à partir de différentes morphologies artificielles, qui varient de promenades aléatoires à des structures de branchement contraints, aux réseaux de neurones biologiques et aux réseaux computationnels construits à partir de morphologies biologiques. Étonnamment, les réseaux de neurones qui sont générés à partir de morphologies presque aléatoires partagent beaucoup de propriétés avec les réseaux de neurones biologiques, indiquant que les processus stochastiques qui ont lieu au cours du développement contribuent de manière significative aux formes neuronales observées.

Cette thèse éclaircit plusieurs mystères concernant les morphologies neuronales et met en question nos croyances sur le rôle du hasard dans la formation du cerveau. Ainsi, elle nous rapproche de la compréhension des différences fondamentales entre les morphologies neuronales et comment le hasard et la structure se mélangent pour générer l'un des systèmes biologiques les plus complexes.

Table of contents

List of figures

List of tables

1

Introduction

A lot of scientific questions are associated with the brain's structure and functionality, as it is one of the most complicated biological systems known. The composition of the brain is dominated by two cell types: neurons and glia cells. The role of glial cells is not yet well understood. They are considered to participate in homeostatic mechanisms (Sorrentino et al. 2016), and provide physical and metabolic support to neurons (Tasker et al. 2012) while recent studies also suggest their involvement in brain computations (Temburni and Jacob 2001, Perea et al. 2014). Neurons, on the other hand, have been well studied, since they were identified as the fundamental structural and functional units of the nervous system by Ramón y Cajal in the late 19th century. Ramón y Cajal used Golgi's staining method to systematically examine and describe the anatomy of neurons. His highly detailed drawings revealed the hidden structure of the brain, giving him the well deserved title of the *father of neuroscience*. Neurons consist of several components with different anatomical and functional properties: the cell body (soma), the signal receivers (dendrites) and the signal transmitter (axon). They are electrically excitable cells (Rall 1959) that transmit information by electrical and chemical signaling, and communicate with each other through synapses that are formed between the post-synaptic axon and the pre-synaptic dendrites.

It is now well established that neuronal morphology and brain functionality are strongly coupled (Chklovskii 2004, Wen et al. 2009). However, the precise mechanisms through which the neuronal morphology determines the functionality and the connectivity of a neuron are yet to be discovered. Multiple theories have been proposed to explain the neuron's complex shapes. It is tempting to believe that basic biological principles, such as the minimal wiring (Chklovskii 2004, Cuntz et al. 2010) and the synaptic efficacy optimization (Cuntz et al. 2007), are sufficient to explain the branching structure of neuronal trees. However, the great morphological diversity (Ascoli et

al. 2007) suggests that the mechanisms that determine neuronal shape might be much more perplexing.

In an attempt to acquire more information about the systematic roles of neurons in the brain, modern neuro-anatomists observe them under a microscope and visually distinguish them into groups, based on their shapes. This classification method is subject to large variation between experts (DeFelipe et al. 2013). Hence, a digital copy of a cell's structure is generated -the neuronal reconstruction- that corresponds to the digital version of Ramón y Cajal's drawings. The neuronal reconstruction is a mathematical tree that represents the morphology of the neuron and is used for the systematic characterization of its shape.

Morphological features, also known as morphometrics (The Petilla Interneuron Nomenclature Group P 2008), are used to describe and distinguish different neuronal shapes. Standard morphometrics include global measurements such as total length, number of branches and maximum extent, local measurements such as branching angles and tortuosity and topological measurements such as asymmetry (Van Pelt et al. 1991) and branch orders. However, single features that focus on a specific aspect of the neuronal morphology are not appropriate to describe its structure as a whole, as they result in significant information loss. In order to establish a rigorous categorization of morphologies, a set of morphometrics that are indicative of the differences between neuronal shapes and generalizable across brain regions and species is required. However, it is challenging to find the hidden correlation between these morphometrics and combine them into a solid descriptor of the neuronal shape (Lopez-Cruz et al. 2011).

Inspired by Algebraic Topology, I constructed a descriptor (see section 2.2) that encodes the spatial distribution of the neuron's branching structure into a topological representation. The Topological Morphology Descriptor (TMD) couples the topology of a tree with its geometry, providing a link between the local and global properties of a neuron's shape. The TMD was used to distinguish different types of neurons; morphological types of different species and well established classes of rodent cells were successfully retrieved. Based on the topological profiles of rat pyramidal cells, an objective morphological classification scheme was established (see section 2.3). Interestingly, a TMD-based clustering revealed two new classes of human pyramidal cells, that could not be identified with standard morphometrics (see section 2.4). Thus, the TMD is a rigorous mathematical descriptor of branching structures that is fundamental for their objective clustering.

While neuronal structures are biologically interesting themselves, understanding the morphological stages of neuronal growth is of particular interest for various fields

of neuroscience. Advances in the study of neuronal growth (Graham et al. 2006) have provided an insight into the relation between morphology and function and have lead to a better understanding of the observed connectivity patterns in the brain and to algorithms for the automatic generation of their shapes (Burke 1992, Ascoli et al. 2001, Cuntz et al. 2010). To investigate the significance of topology on the neuronal growth, I generated artificial morphologies based on their topological profiles (see Chapter 3). The topological synthesis of neurons accurately recreates the biological shapes of neurons for a large variety of morphological types. More importantly, morphological features that have not been explicitly taken into account are reproduced by this synthesis algorithm. This indicates that the topological morphology descriptor of neurons incorporates feature correlations that are essential for their growth.

The neurons that belong in the same morphological type have similar topological profiles and therefore appear highly structured. However, it is still unknown to what extent the connectivity of a biological network is determined by the genetic information of an organism and to what extent it arises from stochastic processes. To study the impact of randomness and structure on the connectivity of a network, I compared random networks, based on the structural touch points between computationally generated morphologies, with the connectivity of biological networks (See Chapter 4). Initially, these morphologies are simple random paths. Then biological constraints are imposed on the morphologies and more structured networks are generated. Surprisingly, simple mathematical constraints result in networks that succeed in capturing key properties of the biological networks, indicating that simple interactions between growing neurons, such as intersection avoidance, have a significant contribution to their shapes and their connectivity.

References

Ascoli GA, Krichmar JL, Scorcioni R, Nasuto SJ, Senft SL. Computer generation and quantitative morphometric analysis of virtual neurons. (2001). Anat Embryol (Berl). 204(4):283-301.

Ascoli, G.A., Donohue, D., & Halavi, M. Neuromorpho.org:a central resource for neuronal morphologies. (2007). Journal of Neuroscience, 27(35), 9247–51.

Burke RE, Marks WB, Ulfhake B. A parsimonious description of motoneuron dendritic morphology using computer simulation. (1992). J Neurosci. 12(6):2403-16.

Ramón y Cajal, S.(1911). Histologie du Système Nerveux de l'Homme et des Vertèbres. Paris, Maloine.

Chklovskii D.B., Synaptic Connectivity and Neuronal Morphology: Two Sides of the Same Coin, (2004). Neuron, Volume 43, Issue 5, Pages 609-617,

Cuntz H., Borst A., Segev I. Optimization principles of dendritic structure. (2007). Theoretical Biology & Medical Modelling. 4:21.

Cuntz H., Forstner F., Borst A., Hausser M. One Rule to Grow Them All: A General Theory of Neuronal Branching and Its Practical Application. (2010) Morrison A, ed. PLoS Computational Biology. 6(8):e1000877.

DeFelipe, J., Lopez-Cruz, P.L., Benavides-Piccione, R., et al. New insights into the classification and nomenclature of cortical GABAergic interneurons. (2013). Nature Reviews Neuroscience, 14(3), 202–216.

Graham, B.P., van Ooyen, A. Mathematical modelling and numerical simulation of the morphological development of neurons. (2006). BMC Neuroscience. 7(Suppl 1):S9.

Lopez-Cruz P.L., Bielza C., Larrañaga P., Benavides-Piccione R., DeFelipe J. Models and simulation of 3D neuronal dendritic trees using Bayesian networks. (2011). Neuroinformatics. 9(4):347-69.

Perea G., Sur M., Araque A. Neuron-glia networks: integral gear of brain function. (2014) Frontiers in Cellular Neuroscience. 8:378.

Rall W. Electrophysiology of a Dendritic Neuron Model. (1962). Biophysical Journal. 2(2 Pt 2):145-167.

The Petilla Interneuron Nomenclature Group P (2008). Petilla terminology: nomenclature of features of gabaergic interneurons of the cerebral cortex. Nature Reviews Neuroscience, 9(7), 557–568.

Sorrentino F.S., Allkabes M., Salsini G., Bonifazzi C., Perri P. The importance of glial cells in the homeostasis of the retinal microenvironment and their pivotal role in the course of diabetic retinopathy. (2016). Life Sci. 162:54-9.

Tasker J.G., Oliet S.H.R., Bains J.S., Brown C.H., Stern J.E. Glial regulation of neuronal function: from synapse to systems physiology. (2012). Journal of Neuroendocrinology. 24(4):566-576.

Temburni M.K., Jacob MH. New functions for glia in the brain. (2001). Proc Natl Acad Sci U S A. 98(7):3631-2.

Van Pelt, J., Uylings, H.B.M., Verwer, R.W.H., Pentney, R.J., & Woldenberg, M.J. (1991). Tree asymmetry—A sensitive and practical measure for binary topological trees. Bulletin of Mathematical Biology, 54(5), 759–784.

Wen Q., Stepanyants A., Elston G.N., Grosberg A.Y., Chklovskii D.B. Maximization of the connectivity repertoire as a statistical principle governing the shapes of dendritic arbors. (2009). Proc Natl Acad Sci U S A. 2106(30):12536-41.

"Deconstructing" neuronal trees
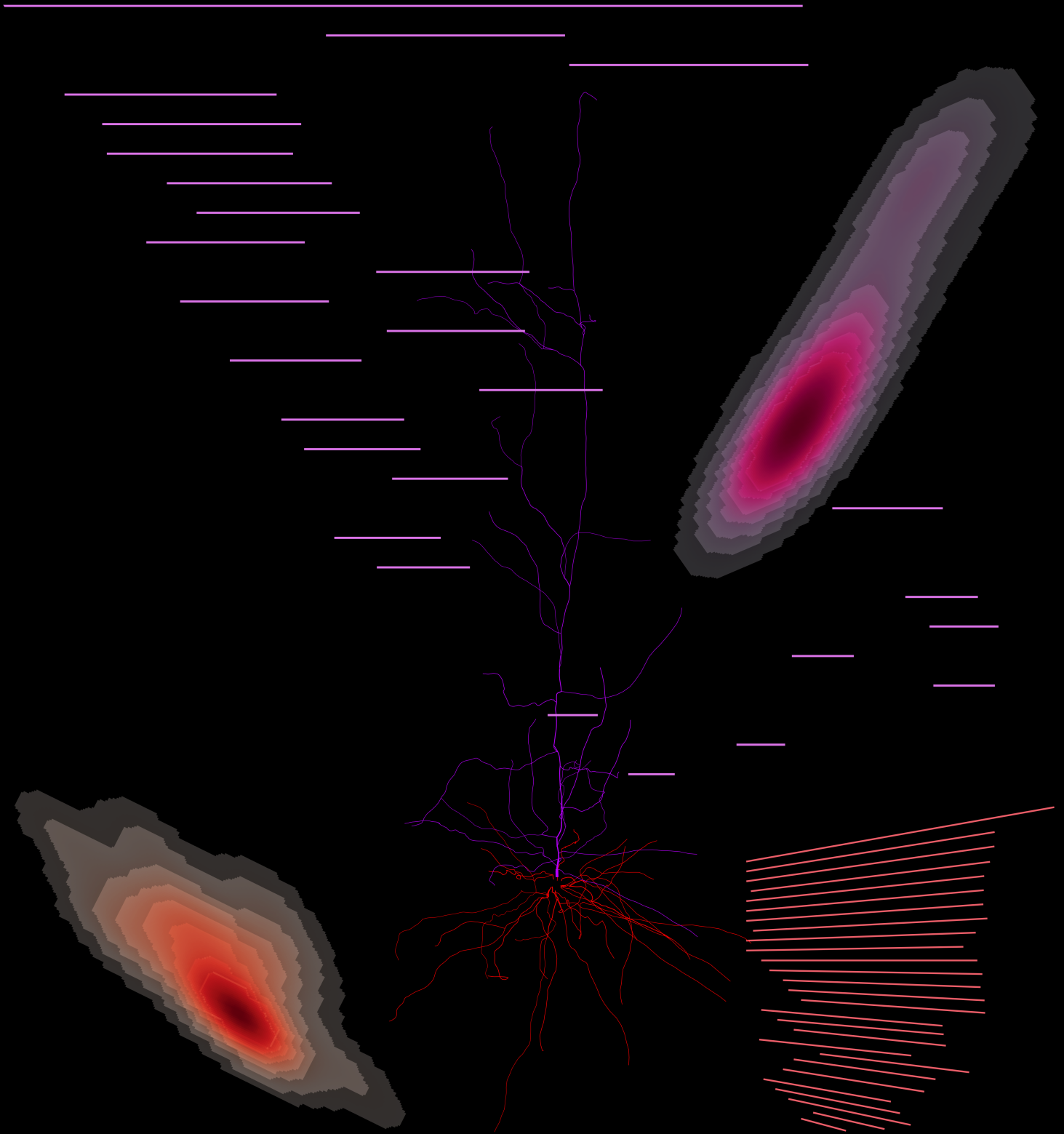
2

Topological analysis of neuronal morphologies

"It would be so nice if something made sense for a change."
                                    – Alice, Alice in Wonderland

2.1 Introduction

Neurons, the fundamental computational units of the brain (Rall 1959), consist of the cell body (soma), and two types of branching trees, the dendrites and axons, collectively referred to as neurites, that transmit electrical signals within the neuronal network. Pyramidal cells, which represent the majority of excitatory cells in the cortex (Lefort et al. 2009), also have a special type of dendrite, the apical dendrite (see Figure 2.1), that is characterized by a unique branching shape. A neuron has a number of dendrites which collect the signal from multiple input sources. The signal is summed at the initial section of the axon, where the action potential is initiated, and is transmitted to other connected neurons. The characteristic branching structures of neurites influence the functional properties of a neuron (Yi et al. 2017). In addition, the position of neurons in the brain tissue and the distribution of their branches in space constrain the appositions, or contact points, between neurons (Peters 1979, Kalisman et al. 2003). As a result, the synaptic distribution depends on the spatial arrangement of neuronal branches.

Inevitably, neuronal morphology and brain functionality are strongly coupled (Yi et al. 2017). As a result, the study of neuronal morphology is important for understanding brain functionality. For this reason, neuro-anatomists examine a cell

under the microscope and generate a digital version of the neuron's morphology, the neuronal reconstruction (see Figure 2.1). As the quality of neuronal staining and imaging improves, more accurate reconstructions are becoming available. The morphological analysis of neurons is also evolving, ranging from simple morphometrics such as the soma size and the number of neurites, to more detailed measurements such as the distributions of section numbers and lengths, local bifurcation angles, radial and path distances, branch orders and tree asymmetry (see Figure 2.1 and Petilla, VanPelt). Neuronal morphologies present a great variety of shapes, even within the same species or brain regions (see Figure 2.2).

A large number of reconstructions is available at NeuroMorpho.Org in a standardized format (Ascoli et al. 2007) as the result of many groups' contributions over many years. An illustrative sample of these morphologies is shown in Figure 2.2 for various rodent (rat and mouse) brain regions. Even though it is not always trivial to identify the morphological differences between these morphologies, it is - in this particular example - possible to distinguish the different shapes by visual inspection. Some morphological differences are more easily identified, such as the size, extent and density of their branches. Others are more difficult to distinguish, such as the tree asymmetry, the branch ordering and the topological complexity of neurons.

More sophisticated morphometrics have been developed to capture various anatomical characteristics that distinguish different shapes of neurons. The spatial density function of branching morphologies is an example that has been studied in Snider et al. (2010) and is useful for the description of distinct neuronal trees, as it represents a universal property of neurons. Similarly, in Samsonovich et al. (2003) and in Fernandez-Gonzalez et al. (2017) universal properties of bifurcation angles across different species and brain areas are examined. Another powerful morphometric is Sholl analysis, which counts the intersections of concentric spheres with the neuronal morphology. A topological description of neurons have been proposed in (Gillette et al. 2015).

Each of these measurements studies a different aspect of the neuronal morphology and therefore they must be used in combination with other morphometrics for the classification of neurons. To avoid over-fitting, which is a result of using a large number of features when few individual cells are available, feature selection is required. However, feature selection is subjective and feature sets proposed by different investigators are mutually inconsistent (DeFelipe et al. 2013). A striking indication of the problem is that experts may assign a different class to a neuron than the term they had chosen in their own original publication for the same neuron (DeFelipe et al. 2013). In fact,
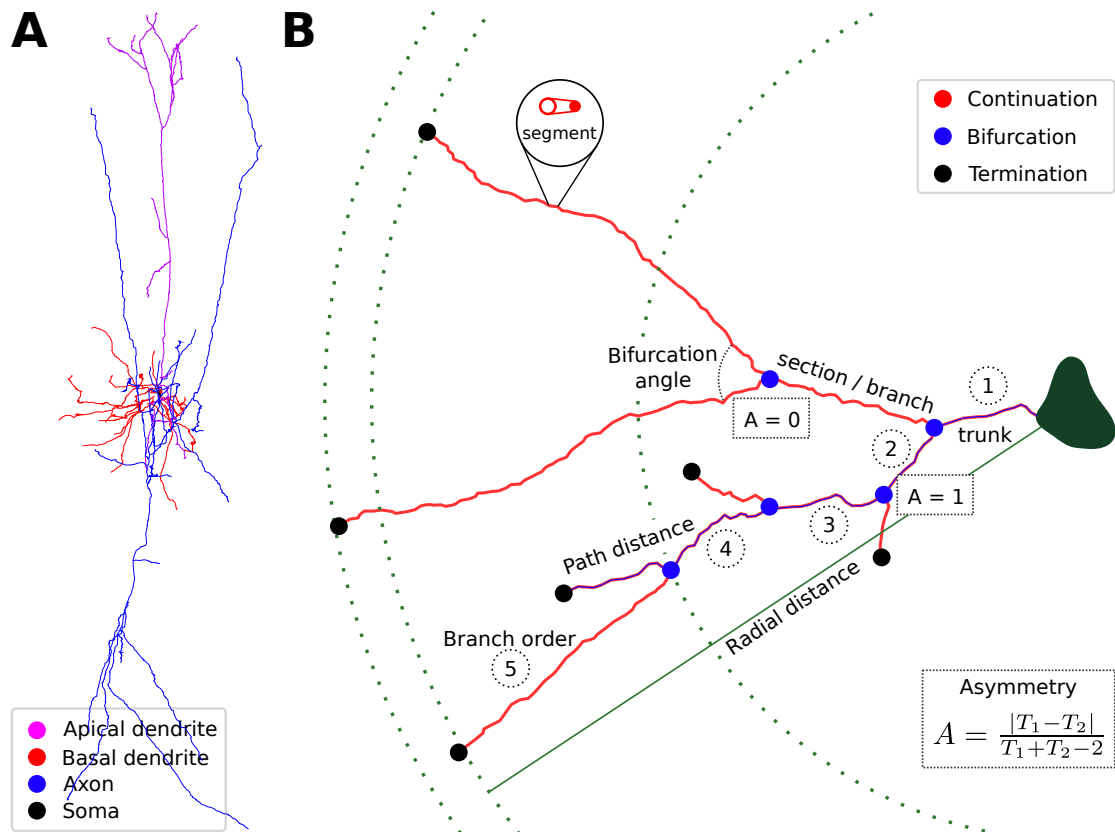
Fig. 2.1 Terminology of neuronal parts and morphometrics. (A) A neuron consists of the cell body (soma, in black), the dendrites (basal, in red, and apical, in purple) that collect the signal from other neurons and the axon (in blue) that sends the signal to other neurons. (B) A neuronal tree consists of segments (or truncated cones) that compose sections, between two bifurcation points, or a bifurcation and a termination. The initial segment of a neurite is called "trunk". Several basic morphometrics are computed from this branching structure, such as section lengths, path and radial distances and bifurcation angles. More complex morphometrics include branch order that counts the number of branches following the path to the soma, and the tree asymmetry A which computes ratio of the number of branches between the two subtrees T_1, T_2 at each bifurcation point. The tree asymmetry varies from $A = 0$ for a perfect binary tree to $A = 1$ for an asymmetric tree.
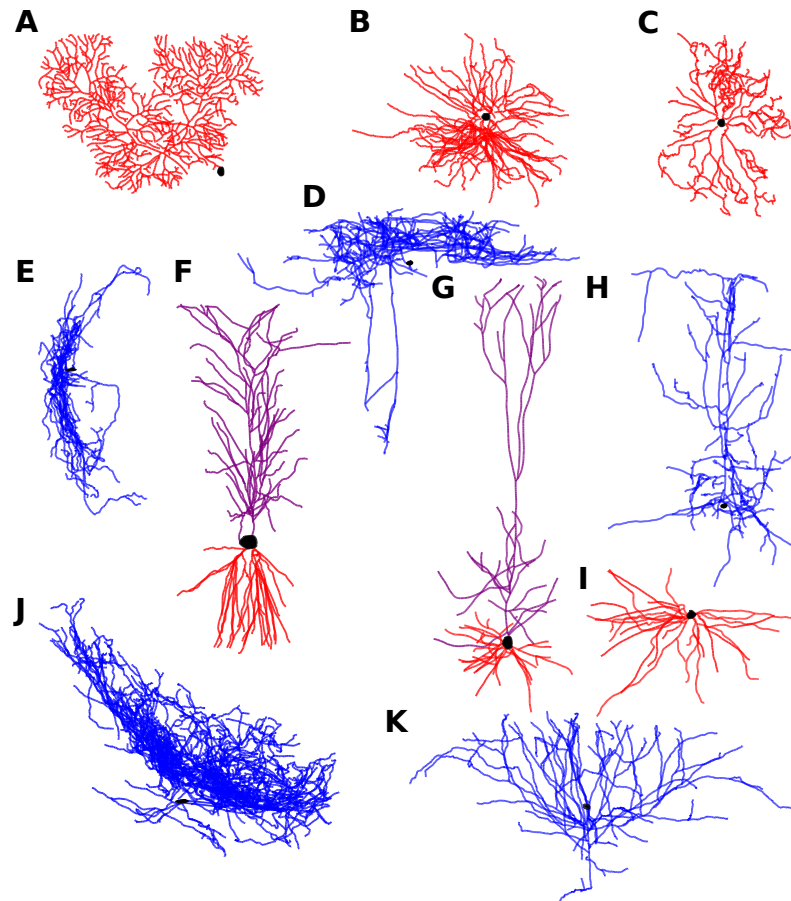
Fig. 2.2 Diversity of rodent morphologies from different brain areas. A: Cerebellum, purkinje cell dendrites (DeSchutter). B: Thalamus, dendrites (Acsady). C: Retina, dendrites (Rodger). D: Cortex, L1 interneuron axon (Markram). E: Hippocampus, axon (Bartos). F: Hippocampus, pyramidal cell dendrites (Amaral). G: Cortex, L5 pyramidal cell dendrites (Markram). H: Cortex, L5 Martinotti cell (Markram). I: Cortex, L5 dendrites (Markram). J: Hippocampus axon (Soltesz). K: Cortex, L2/3 pyramidal cell axon (Markram).
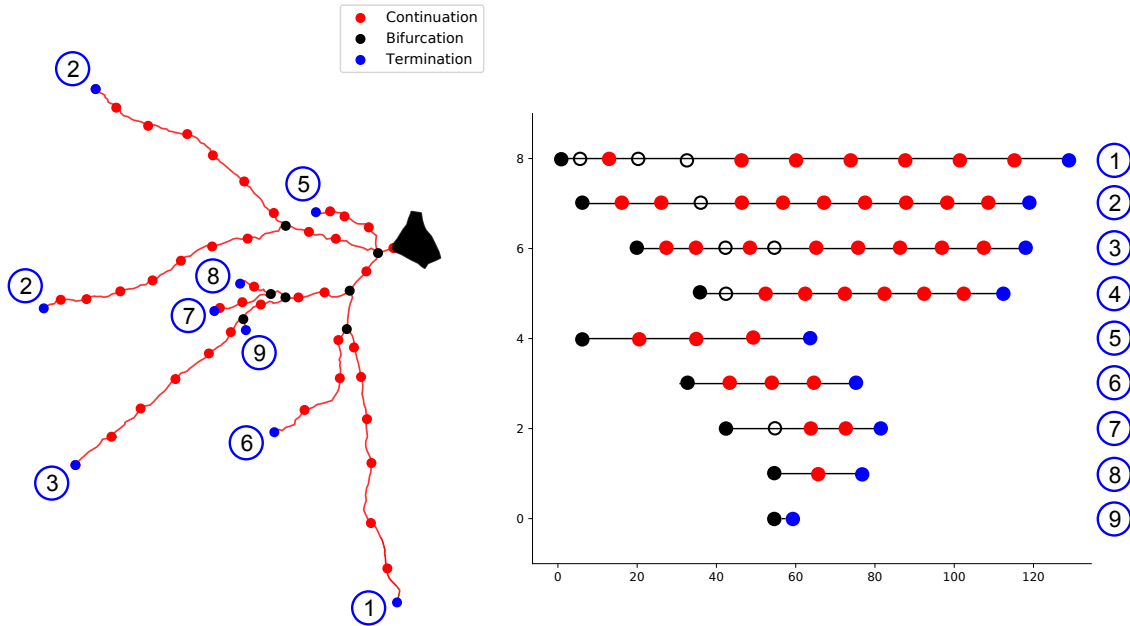
Fig. 2.3 Topological barcode of a branching tree. The branching structure of the tree is encoded in a set of bars. Each bar represents the start and end radial distances from the soma of a component within the tree. The corresponding component starts at a bifurcation (in black) and ends at a termination (in blue, numbered). At each bifurcation point only one, the longest, component persists. The empty nodes in the barcode denote a bifurcation of a longer component to another subtree.

alternative sets of morphometrics result in different classifications (see section 2.2), since statistical features commonly overlap even across markedly different morphological types. As a result, expert-dependent feature selection is not suitable to reliably study complex branching morphologies. In order to establish a rigorous categorization of morphologies, a metric that is indicative of the differences between neuronal shapes and generalizable across brain regions and species is required. For this purpose, I propose an alternative representation of morphologies based on persistent homology.

The Topological Morphology Descriptor (TMD, see section 2.2) algorithm generates a barcode from the neuronal tree, coupling the topology and the geometry of the branching structure, encoding the overall shape of the tree (see Figure 2.3. Local fluctuations with little information content, such as the position of the nodes between branch points, are discarded and thus the computational complexity of a tree is reduced. The TMD is not a complete invariant that fully describes the original tree, but a simplification that retains enough information to perform well in the proposed discrimination tasks, by mapping the tree to a topological representation with less information loss than the usual morphometrics. Correlations between the fundamental morphological features are implicitly captured in the TMD. As a result, computationally

generated random trees that differ in selected morphological features (section length and number, local angles, tortuosity and asymmetry) can be distinguished with the same descriptor. In Figure 2.4, I demonstrate how the diversity of the rodent morphologies of various brain regions Figure 2.2 is reflected into their TMD profiles.

The TMD of a tree provides a topological benchmark for the consistent comparison of different structures by effectively assigning a reliability measure to different groupings of neuronal trees. In addition the TMD can be used for the standardized classification of neurons that cannot be grouped by expert selected morphometrics. This technique can be applied to any rooted tree equipped with a function defined on its nodes. Further biological examples include botanic trees (Lopez et al. 2010), corals (Kruszynski et al. 2007) and roots of plants (Wang et al. 2009). The TMD method is used in different problems to assist the progress of neuronal morphologies grouping.

I demonstrate the discriminative power of this method by applying it to a collection of artificial random trees, whose morphological properties (number of branches, bifurcation angles, branch length, asymmetry) can be precisely modified. Then, the TMD algorithm is used for the grouping of different biological reconstructions of neuronal morphologies. Neurons from different species, pyramidal neurons of the same cortical areas and automatic reconstructions from the BigNeuron project are compared and distinguished according to their TMDs. Based on the topological profiles of rat pyramidal cells, an objective morphological classification scheme has been established (see Section 2.2).

Human cells are rare and difficult to acquire. The analysis of human neurons' morphology can shed light to one of the fundamental questions in neuroscience "what is unique about the human brain". Using the topological profiles of 60 L2 and L3 human pyramidal cells, the TMD-based classification revealed two distinct morphological classes (the "slim-tufted" and the "profuse-tufted" cells) that differ in the shape of their tufts. The two classes could not be distinguished with simple morphological features (see section 2.3). Interestingly, the two classes of pyramidal cells also express different electrical properties as the profuse-tufted cells fire at higher rates. Therefore, the TMD captures basic principles of the underlying structure that are also relevant for the electrical activity of the cells.

The radial distance from the soma is used as the "filtration" function of the TMD that determines the start and end points of the bars. Alternatively, any other function that can be computed on the bifurcations and terminations of the tree could be used. In fact, independent morphometrics, such as the thickness of the processes and the spine density, are not taken into account so trees that differ only on this property cannot be
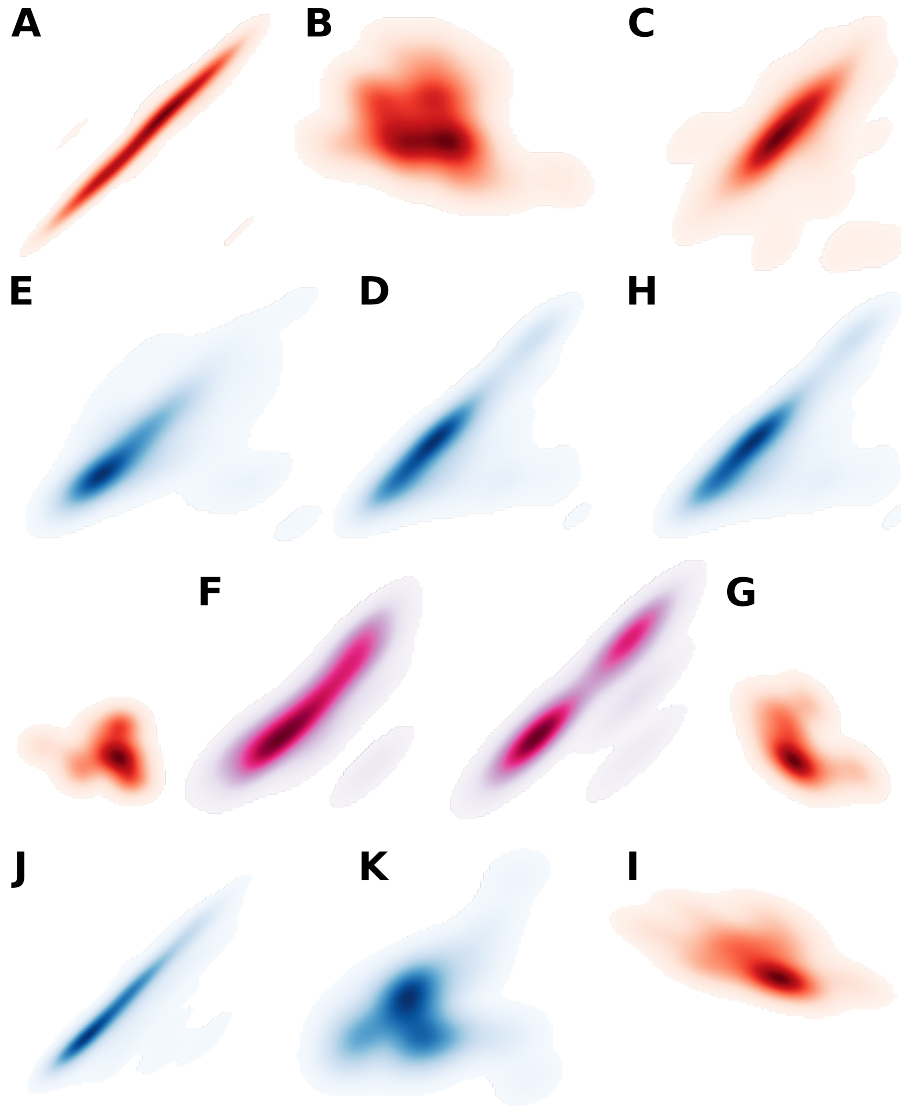
Fig. 2.4 Illustration of rodent morphological diversity using persistence images. The diversity of neuronal morphologies is captured by the respective persistence images of neurons from different brain areas. A: Cerebellum, purkinje cell dendrites (DeSchutter). B: Thalamus, dendrites (Acsady). C: Retina, dendrites (Rodger). D: Cortex, L1 interneuron axon (Markram). E: Hippocampus, axon (Bartos). F: Hippocampus, pyramidal cell dendrites (Amaral). G: Cortex, L5 pyramidal cell dendrites (Markram). H: Cortex, L5 Martinotti cell (Markram). I: Cortex, L5 dendrites (Markram). J: Hippocampus axon (Soltesz). K: Cortex, L2/3 pyramidal cell axon (Markram).

distinguished with the TMD. Another important property that is currently not taken into account is the long range projections of the trees. This is particularly important for axonal trees whose long range projections indicate the brain regions with which they communicate. Nevertheless, the TMD is generalizable to a variety of morphometrics. For instance, a modification of the algorithm that considers the orientation of the tree is used for the analysis and systematic classification of rodent pyramidal cells in the somatosensory cortex (see section 2.4). Therefore the TMD algorithm and its variations is suitable for the supervised and the unsupervised discrimination of neurons into distinct morphological groups.

References

Ascoli, G.A., Donohue, D., & Halavi, M. Neuromorpho.org:a central resource for neuronal morphologies. (2007). Journal of Neuroscience, 27(35), 9247–51.

Fernandez-Gonzalez P., Benavides-Piccione R., Leguey I., Bielza C., Larranaga P., DeFelipe J. Dendritic-branching angles of pyramidal neurons of the human cerebral cortex. (2017). Brain Structure & Function. 222(4):1847-1859.

DeFelipe, J., Lopez-Cruz, P.L., Benavides-Piccione, R., et al. New insights into the classification and nomenclature of cortical GABAergic interneurons. (2013). Nature Reviews Neuroscience, 14(3), 202–216.

Gillette, T., Hosseini, P., & Ascoli, G. (2015). Topological characterization of neuronal arbor morphology via sequence representation: II - global alignment. BMC Bioinformatics, 16, 209.

Kalisman N., Silberberg G., Markram H. Deriving physical connectivity from neuronal morphology. (2003). Biol Cybern. 88(3):210-8.

Kanari L., Dlotko P., Scolamiero M., Levi R., Shillcock J., Hess K., Markram H. A Topological Representation of Branching Neuronal Morphologies. (2017). Neuroinformatics.

Kruszynski, K.J., Kaandorp, J.A., & van Liere, R. (2007). A computational method for quantifying morphological variation in scleractinian corals. Coral Reefs, 26(4), 831–840.

Lefort S., Tomm C., Floyd Sarria J.C., Petersen C.C. The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex. (2009) Neuron. 29;61(2):301-16.

Lopez, L.D., Ding, Y., & Yu, J. (2010). Modeling complex unfoliaged trees from a sparse set of images. Computer Graphics Forum, 29, 2075–2082.

Peters A. (1979). Thalamic input to the cerebral cortex. Trends Neurosci. 2, 1183–1185

Rall W. Electrophysiology of a Dendritic Neuron Model. (1962). Biophysical Journal. 2(2 Pt 2):145-167.

Samsonovich A.V., Ascoli G.A. Statistical morphological analysis of hippocampal principal neurons indicates cell-specific repulsion of dendrites from their own cell. (2003). J Neurosci Res. 15;71(2):173-87.

Snider, J., Pillai, A., & Stevens, C.F. (2010). A universal property of axonal and dendritic arbors. Neuron, 66, 45–56.

Wang, H., Siopongco, J., Wade, L., & Yamauchi, A. (2009). Fractal analysis on root systems of rice plants in response to drought stress. Environmental and Experimental Botany, 65(2–3), 338–344.

Yi G.S., Wang J., Deng B., Wei X.L. Morphology controls how hippocampal CA1 pyramidal neuron responds to uniform electric fields: a biophysical modeling study. (2017). Sci Rep. 7(1):3210.

2.2 A topological representation of branching neuronal morphologies

Lida Kanari, Pawel Dłotko, Martina Scolamiero, Ran Levi, Julian Shillcock
Kathryn Hess and Henry Markram

Abstract: Many biological systems consist of branching structures that exhibit a wide variety of shapes. Our understanding of their systematic roles is hampered from the start by the lack of a fundamental means of standardizing the description of complex branching patterns, such as those of neuronal trees. To solve this problem, we have invented the Topological Morphology Descriptor (TMD), a method for encoding the spatial structure of any tree as a "barcode", a unique topological signature. As opposed to traditional morphometrics, the TMD couples the topology of the branches with their spatial extents by tracking their topological evolution in 3-dimensional space. We prove that neuronal trees, as well as stochastically generated trees, can be accurately categorized based on their TMD profiles. The TMD retains sufficient global and local information to create an unbiased benchmark test for their categorization and is able to quantify and characterize the structural differences between distinct morphological groups. The use of this mathematically rigorous method will advance our understanding of the anatomy and diversity of branching morphologies.

Keywords: topological data analysis, neuronal morphologies, branching morphology, clustering trees

Introduction

The analysis of complex branching structures, such as branched polymers (Alexandrow-icz et al. 1985), viscous fingering (Oded et al. 2002), and fractal trees (Mandelbrot 1983), is essential for understanding a great variety of physical and biological processes. For example, the fundamental units of the nervous system, neurons, possess highly ramified arborizations (Jan et al. 2010) that are thought to reflect their involvement in different computational tasks (Cuntz et al. 2007, Zomorrodi et al. 2010, van Elburg et al. 2010, Ferrante et al. 2013). In order to understand the properties of branching morphologies we must study the differences between distinct arbor types. Much effort has therefore been devoted to grouping morphologies into distinct classes (DeFelipe et al. 2013, Markram et al. 2004, The Petilla Interneuron Nomenclature Group 2008), a categorization process that is important in many fields (Lyons et al. 1999). However, an efficient method for quantitatively analyzing the morphology of such structures has proved difficult to establish.

In general, the properties of branching morphologies have been rigorously studied in two extreme cases: in the limit of the full complexity of the structures (Carlsson 2009), where the entire set of points is used, and in the opposite limit of feature extraction (DeFelipe et al. 2013; Gomez-Gil et al. 2008; Blackman et al. 2014), where a (typically small) number of selected morphometrics (i.e., statistical features) are extracted from the morphology.

Topological data analysis (TDA) has been shown to reliably identify geometric objects based on a sampled point cloud when they are built out of well-understood pieces, such as spheres, cylinders and tori (Carlsson 2009). It suffers, however, from the deficiency that reliable grouping of complex geometric trees by standard TDA methods, such as Rips complexes (Edelsbrunner and Harer 2008), requires thousands of sampled points, which is expensive in terms of both computational complexity and memory requirements.

Feature extraction is thus the only currently feasible solution to establishing a more quantitative approach to analyzing branching morphologies (Scorcioni et al. 2008; Ling et al. 2012; The Petilla Interneuron Nomenclature Group P 2008). While this approach has been efficiently used in specific fields of image recognition (Schurer 1994), the extreme diversity of the branching patterns of neurons (Markram et al. 2004) makes it difficult to identify an optimal set of statistical features that can reliably describe all their shapes. Neuronal classification has traditionally focused on visually distinguishing the shapes observed under a microscope (Masseroli et al. 1993), a method that is subject to large variation between experts (DeFelipe et al. 2013).
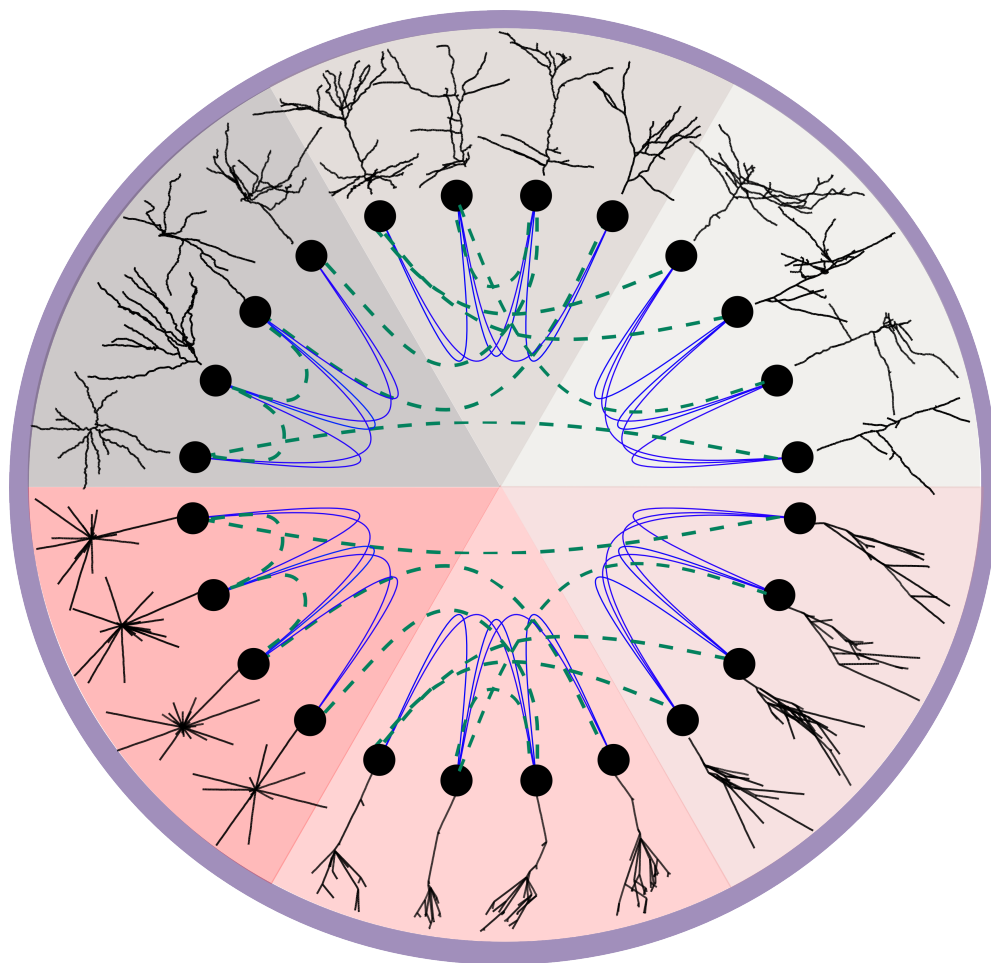
Fig. 2.5 Illustration of the separation of similar tree structures into distinct groups, using topological analysis. The colored pie segments show six distinct tree types: three neuronal types (upper half) and three artificial ones (lower half). The thick blue lines show that our topological analysis can reliably separate similar-looking trees into groups. It is accurate both for artificially-generated trees and neuronal morphologies. The dashed green lines show that classification using an improper set of user-selected features (number of branches, total length) cannot distinguish the correct groups.

For this reason, experts generate a digital version of a cell's structure - a neuronal reconstruction (Dieter 2000) as a set of points in $\mathbb{R}^3$ sampled along each branch, together with edges connecting adjacent pairs of points. This reconstruction is a mathematical tree that represents the neuron's morphology and can be used for the extraction of its morphological properties. To avoid overfitting, which is a result of using a large number of features when few individual cells are available, feature selection is performed by experts who identify the relevant morphometrics for each group of cells. Many sophisticated variants of the standard morphological features have been proposed over the years, such as tree asymmetry (Van Pelt et al. 1991, 2001,

2005), centrifugal ordering (Van Pelt et al. 1989) and Strahler ordering (Strahler 1952; Berry and Bradley 1976; Ledderose et al. 2014), to describe the topology of branching structures. However none of those measurements preserves the correlations between distinct features. In addition, feature selection is subjective, and alternative sets of morphometrics result in different classifications (DeFelipe et al. 2013), as illustrated in Fig 2.5(see also Figs A.1, A.2), since the statistical features commonly overlap even across markedly different morphological types. This is a direct consequence of the significant loss of information introduced by feature selection, as the dimensionality of the data is substantially reduced.

As a result, neither using the full point cloud of the trees nor performing expert-dependent feature selection are suitable to reliably study complex branching morphologies. In order to address this issue, we propose a standardized topological descriptor, the Topological Morphology Descriptor (TMD), of any branching morphology. The TMD algorithm encodes the branching pattern of the morphology by discarding local fluctuations with little information content, such as the position of the nodes between branch points and thus reduces the computational complexity of a tree. The TMD couples the topology of the branching structure with the embedding in the metric space, encoding the overall shape of the tree. Note that the TMD is not a complete invariant that fully describes the original tree, but a simplification that retains enough information to perform well in the proposed discrimination tasks, by mapping the tree to a topological representation with less information loss than the usual morphometrics.

The TMD algorithm takes as input the partially ordered set of branch points (nodes with more than one child) and leaves (nodes with no children) of the tree, where the order is given by the parent-child relation, and produces a multi-set of intervals on the real line known as a *persistence barcode* (Carlsson 2009), Fig 2.6b. Each interval encodes the lifetime of a connected component in the underlying structure (see Glossary), identifying when a branch is first detected (birth) and when it connects to a larger subtree (death). This information can be equivalently represented in a *persistence diagram* (Carlsson 2009), Fig 2.6c in which the pair of birth-death times determines a point in the real plane. Either representation greatly simplifies the mathematical analysis of the trees.

This approach provides a simplified comparison process, since distances inspired by persistent homology theory (Carlsson 2009) can be defined between the outputs of the TMD algorithm (see Appendix A: Distances between persistence diagrams). Existing methods for computing distances between trees, such as the *edit distance* (Bille 2005), the *sequence representation* (Gillette and Ascoli 2015; Gillette et al. 2015), the

blastneuron distance (Wan et al. 2015) and the *functional distortion distance* (Bauer et al. 2014), are in general not universally appropriate, and therefore not biologically useful, and computationally expensive (see Appendix A:Distances between trees).
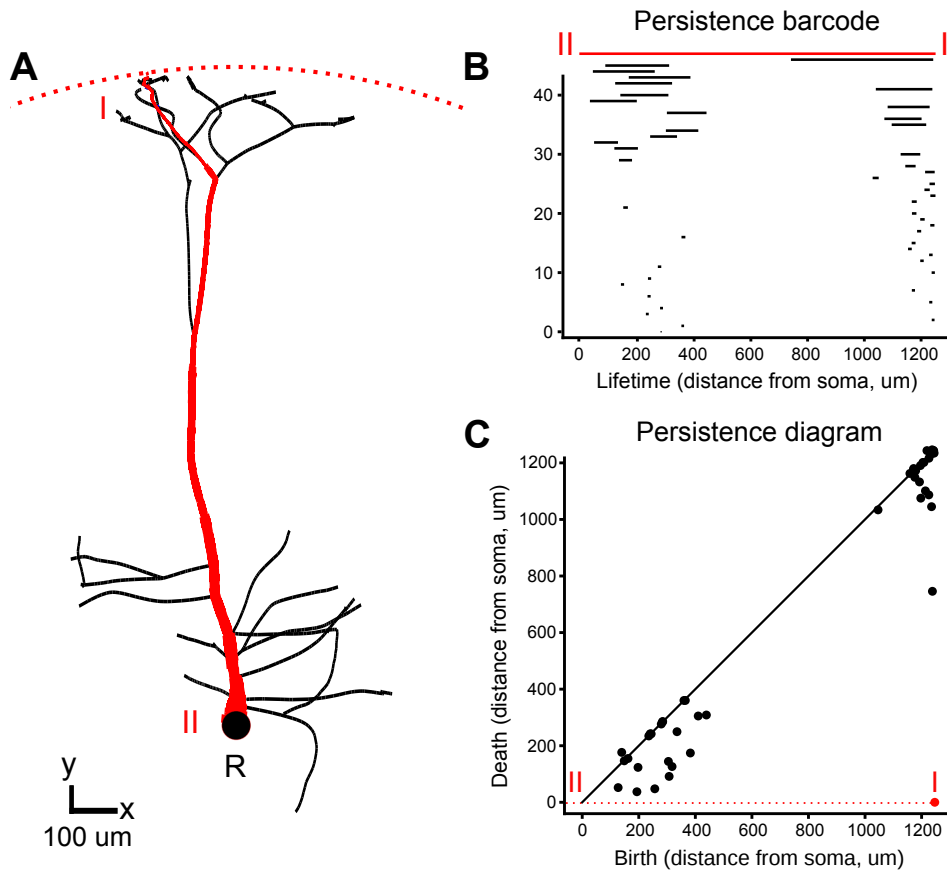
Fig. 2.6 Application of topological analysis to a neuronal tree. (A) showing the largest persistent component (red). The persistence barcode (B) represents each component as a horizontal line whose endpoints mark its birth and death in units that depend on the choice of the function f used for the ordering of the nodes of the tree. In our case, it is radial distance of the nodes from the root (R), so the units are microns. The largest component is again shown in red together with its birth (I) and death (II). This barcode can be equivalently represented as points in a persistence diagram (C) where the birth (I) and death (II) of a component are the X and Y coordinates of a point respectively (in red). The diagonal line is a guide to the eye and marks points with the same birth and death time.

Our method, in contrast, is applicable to any tree-like structure. We demonstrate its generality by applying it first to a collection of artificial random trees, (see Appendix A: Random trees generation), and then to various groups of neuronal trees (see Information Sharing Statement). Our results show that the TMD of tree shapes can be used effectively to assign a reliability measure to different proposed groupings of random and neuronal trees (Fig 2.5). Provided that the available set of morphologies is representative of the biological diversity, we generate a diversity profile (Leinster and

Cobbold 2012) that reflects the abundance of species as well as their differences, in order to further investigate the effects of different classification schemes (see Appendix A: Diversity Index).

Glossary of topological terms

A **component** of T is a path, i.e., a sequence of consecutive edges, in T from a leaf to an internal node, a branch point. The TMD algorithm considers only a subset of all the components.

The **birth** of a component with leaf l and internal node n occurs at time (or radius) $f(l)$ and its **death** occurs at time (or radius) $f(n)$. Note that a component may die before it is born, i.e., it can be true that $f(n) \leq f(l)$. This case is quite usual for biological trees.

The **lifetime** of component with leaf l and internal node n is $[f(n), f(l)]$, when $f(n) \leq f(l)$, and $[f(l), f(n)]$, when $f(n) \geq f(l)$.

In general, a **barcode** is a multiset (i.e., a set with possible repetitions) of closed intervals in the real line. In the special case of the pair (T, f), the barcode consists of all the lifetimes for the components of the tree.

In general, a **persistence diagram** is a multiset of points in the upper righthand quadrant of the real plane. In the special case of the pair (T, f), the persistence diagram consists of all the points $(f(l), f(n))$ and $(f(n), f(l))$.

Let D be the set given by the diagonal in the upper righthand quadrant of the real plane $\{(x, x)|x \geq 0\}$ with infinite multiplicity. A **matching** between two persistence diagrams PD and PD' is a bijection μ between $PD \cup D$ and $PD' \cup D$. Since D, which contains infinitely many points, is included in the bijection, there is always a matching between two persistence diagrams regardless of their number of points.

Given two persistence diagrams PD and PD', the **bottleneck distance** between them is $d_B(PD, PD') = \inf_\mu sup_{x \in PD}||\mu(x) - x||_\infty$, where $||y - z||_\infty = max\{|y_1 - z_1|, |y_2 - z_2|\}$.

A **filtration** of the tree T is an increasing sequence of subgraphs $G_0 \subset G_1 \subset \ldots \subset G_n \subset T$. Note that (sub)graphs may represent a forest, i.e, a disjoint union of trees.

Methods

The extraction of the barcode from an embedded tree T is described by the TMD algorithm. Let T be a rooted, and therefore oriented, tree (Knuth 1998), embedded in $\mathbb{R}^3$. Note that the operation described here is generalizable to trees embedded in any metric space. We denote by $N := B \cup L$ the set of nodes of T, which is the union of the set of branch points B and the set of leaves L. In the case of a neuron, the root R is the node representing the soma. Each node $n \in N$ has references to its parent, i.e., the first node on the path towards the root, and to its children. Nodes with the same parent are called siblings.

Let f be a real-valued function defined on the set of nodes of T. Any function f that is defined on the nodes of T can be used with the TMD algorithm, such as the radial distance, the path distance, the branch length, or the branch order (see Fig A.4). Alternative functions should serve to reveal shape characteristics that are independent from each other and therefore be suitable for different tasks. For the purpose of this study we define f to be the radial distance from the root R. For each $n \in N$, let T_n denote the subtree with root at the node n, and L_n the set of leaves of T_n. We define a function $v : N \to \mathbb{R}$, computed by the TMD algorithm, by $v(n) = \max\{f(x) \mid x \in L_n\}$. An ordering of siblings can then be defined based on v: if $n_1, n_2 \in N$, are siblings and $v(n_1) < v(n_2)$, then n_1 is younger than n_2.

The algorithm is initialized by setting the value of $v(l), l \in L$ equal to the value of $f(l)$. The leaves $l \in L$ are added to a set of nodes, denoted A, which keeps a record of the active nodes. Following the path of each leaf $l \in L$ towards the root R, all but the oldest (with respect to v) siblings are killed, i.e., removed from A, at each branch point. If siblings have the same value v it is equivalent to kill any one of them. For each killed component one interval (birth-death) is added to the persistence barcode (Fig 2.6). The older sibling c_m is replaced by its parent in A and the value $v(p)$ of its parent is set to $f(c_m)$. This operation is applied iteratively to all the nodes until the root R is reached. At this point A contains only one component, the largest one.

When all the branches are outgoing, i.e., the radial distance of the origin of a branch is smaller than the radial distance of its terminal point, the TMD algorithm is equivalent to computing the barcode associated to a filtration of concentric spheres of decreasing radii, centered at R (Fig 2.6). In this case, the birth time of a component is the supremum of the radii of the spheres that do not contain the entire component. The death time is the infimum of the radii of the spheres that contain the branch point at which the component merges with a longer one.

Algorithm 1 TMD algorithm

Require: T with R, B, L, $f : T \to \mathbb{R}$
Ensure: $\text{TMD}(T, f)$, a persistence barcode obtained from a tree T and the function f

1: $\text{TMD}(T, f)$: empty list to contain pairs of real numbers
2: $A \leftarrow L$ ▷ A : set of active nodes
3: **for** every $l \in L$ **do**
4:     $v(l) = f(l)$
5: **while** $R \notin A$ **do**
6:     **for** l in A **do**
7:         p : parent of l
8:         C : children of p
9:         **if** $\forall n \in C, n \in A$ **then**
10:             c_m : randomly choose one of $\{c \mid v(c) = \max_{c'}(v(c'))$ for $c' \in C\}$
11:             Add p to A
12:             **for** c_i in C **do**
13:                 Remove c_i from A
14:                 **if** $c_i \neq c_m$ **then**
15:                     Add $(v(c_i), f(p))$ to $\text{TMD}(T, f)$
16:             $v(p) \leftarrow v(c_m)$
17: Add $(v(R), f(R))$ to $\text{TMD}(T, f)$
18: Return $\text{TMD}(T, f)$

The computational complexity of the TMD algorithm is linear in the number of nodes. Note that the *if* statement in line 9 of the algorithm is critical for the linear complexity. The number of currently active children is saved at each parent node to avoid quadratic complexity.

This process results in a set of intervals on the real line, each of which represents the lifetime of one component of the tree. The TMD algorithm that associates a persistence barcode $\text{TMD}(T, f)$ to a tree T is invariant under rotations and translations, as long as the function f is also. In this paper, f is the radial distance from R and as such it is invariant under rotations about the root and rigid translations of the tree in $\mathbb{R}^3$.

The most common topological metric that is used to compare persistence diagrams is the *bottleneck distance* (Edelsbrunner and Harer 2008), denoted d_B. Given a matching (i.e., a bijection) between two persistence diagrams D_1, D_2, we define the L_∞ distance as the maximum distance between matched points. The bottleneck distance $d_B(D_1, D_2)$ is the infimum over all L_∞ distances for the possible matchings between the two persistence diagrams (Edelsbrunner and Harer 2008).

We prove that TMD: $(T, f) \mapsto \text{TMD}(T, f)$ is stable with respect to the bottleneck distance (see Appendix A: Stability of TMD). For ϵ-small modifications of certain

types in the tree T, the persistence diagram $\mathrm{TMD}(T, f)$ is not modified more than $O(\epsilon)$. In particular, the method is robust with respect to small perturbations in the positions of the nodes and the addition/ deletion of small branches.

However, none of the standard topological distances between persistence diagrams is appropriate for the comparison of neuronal trees. The bottleneck distance as well as distances stable with respect to it, such as the *persistence distortion distance* (Dey et al. 2015) (see Appendix A: Distances between trees) cannot distinguish diagrams that differ in their short components, which are nevertheless important for the distinction of neuronal morphologies.

We therefore define in the space of the barcodes an alternative distance d_{Bar} that we use to compare branching morphologies. For each barcode we generate a density profile as follows: $\forall x \in \mathbb{R}$ the value of the histogram is the number of intervals that contain x, i.e., the number of components alive at that point. The TMD-distance between two barcodes $\mathrm{TMD}(T_1, f)$ and $\mathrm{TMD}(T_2, f)$ is defined as the integral of the absolute differences between the density profiles of the barcodes. This distance is not stable for a large number of ϵ-perturbations of the tree, but it is the only distance we are aware of that succeeds in capturing the differences between the short components of persistence barcodes. This distance is similar to Sholl analysis (Sholl 1953) with a few fundamental differences (see Appendix A: Distances between neurons). However, since this density profile collapses the barcodes into a one-dimensional distribution, it fails to capture the local differences between the branching structures of similar neuronal trees.

For this reason, the persistence diagram was also converted into an *unweighted persistence image*, inspired by persistence images introduced in Adams et al. (2016). We choose to use unweighted persistence images, since points close to the diagonal, which represent short components, are important for the discrimination of the neuronal trees, and these points are ignored in the weighted persistence images. The unweighted persistence image representation allows the construction of an average image for groups of trees, which is useful for quantifying the differences between tree types, since we are not aware of any unambiguous and computationally feasible calculation of an average of persistence barcodes or diagrams. This method is based on the discretization of a sum of Gaussian kernels (Scott 2008), centered at the points of the persistence diagram. This discretization generates a matrix of pixel values, encoding the persistence diagram in a vector, called the unweighted persistence image. Machine learning tools, such as decision trees and support vector machines can then be applied to this vector for the classification of the persistence diagrams. Note that the unweighted persistence images,

unlike the persistence images defined in Adams et al. (2016), do not satisfy stability for the Euclidean distance between their vectors with respect to the perturbations of trees that we consider (see Appendix A: Stability of TMD).

Results

We demonstrate the discriminative power of the TMD by applying it to four examples of increasing complexity. The first application is the grouping of artificial random trees that provide a well-defined test case to explore the method's performance. The random trees are generated by a constrained stochastic algorithm (see Appendix A: Random trees generation) and have properties that can be precisely modified. Next, we have analyzed datasets of more biological relevance: neurons from different species, downloaded from Ascoli et al. (2007), and distinct types of trees obtained from several morphological types of rat cortical pyramidal cells (Romand et al. 2011) (see Information Sharing Statement). This last example is interesting because, although there is biological support for their separation into distinct groups, no rigorous mathematical model has been proposed for their objective classification. Finally, we used the TMD-distance to rank automatic reconstructions from the BigNeuron project (Peng et al. 2015). We thereby illustrate the usefulness of the TMD across non-trivial examples.

Mathematical random trees are defined by a set of parameters that constrain their shape: the tree depth T_d, the branch length B_l, the branch angle B_a, the degree of randomness D_r, and the asymmetry of branches A_b (see Appendix A: Random trees generation). We defined a control group as a set of trees generated with predefined parameters ($T_d = 5$, $B_l = 10$, $B_a = \pi/4$, $D_r = 10\%$, $A_b = 0.0$) and independent random seeds. Each parameter was varied individually to generate groups of trees that differed from the control group in only one property. A tree is assigned to the group which is closer based on the comparison of the distances d_{Bar} between the tree's barcode and the barcodes of the trees in every group. This distance is used to construct a classifier based on a simple hierarchical clustering algorithm (Ward 1963). The accuracy of this classifier is defined as the percentage of successful trials.

We prove that this classifier efficiently separates groups of random trees that differ in their *tree depth* (Fig 2.7), with an accuracy of $96\% \pm 3\%$ (see Appendix A: Random trees grouping). In Figure 2.7 the distance matrix indicates the existence of three distinct groups, and the corresponding clustering. The TMD of random trees generated by varying each of the other parameters B_a, B_l, D_r, A_b are grouped with an accuracy of 88%, 96%, 99% and 100% respectively (see Figures A.7- A.11).
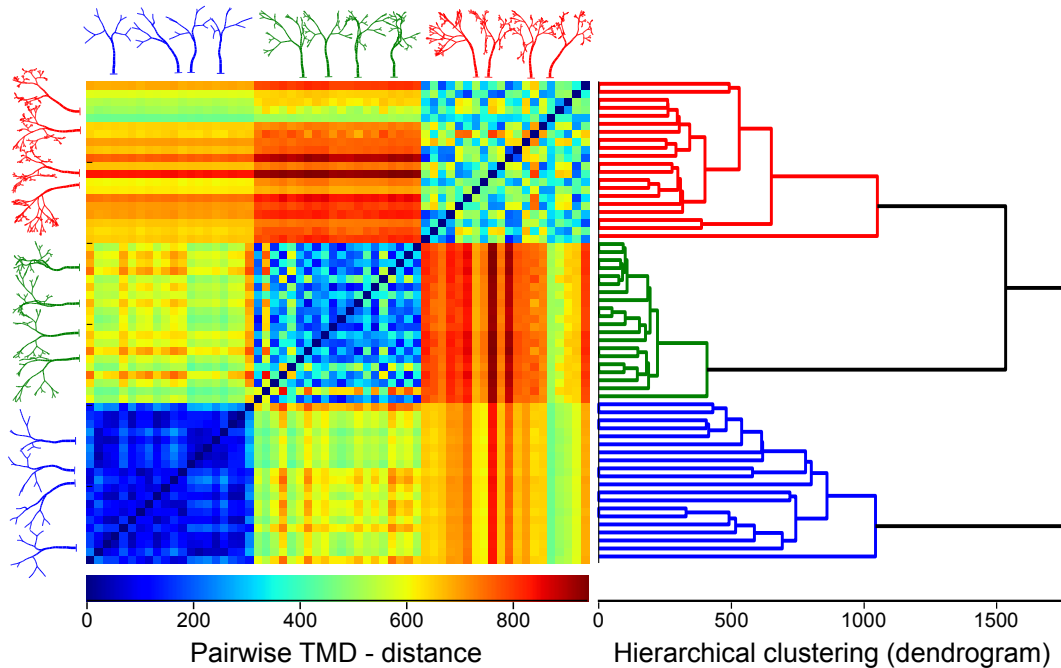
Fig. 2.7 Topological analysis of artificial trees generated using a stochastic process. Three sets of trees are shown (only four individuals out of twenty for clarity). Each group differs from the others only in the tree depth. Each individual of the group is generated using the same tree parameters but a different random number seed. The TMD-distance of the trees allow their accurate separation into groups. The distance matrix indicates the existence of three groups which are identified with high accuracy by a simple dendrogram algorithm.

Next, the TMD is used to quantify differences between neuronal morphologies. Neurons that serve distinct functional purposes exhibit unique branching patterns(Cuntz et al. 2007; Van Elburg and Van Ooyen 2010). In this study, we used cat, dragonfly, fruit fly, mouse and rat neuronal trees. The qualitative differences between the neuronal tree types are evident from the individual geometrical tree shapes (Fig 2.8A) as well as the extracted barcodes (Fig 2.8B). The regions of different branching density are visible in the average unweighted persistence images of each group (Fig 2.8D). Since branching density is thought to be correlated with connection probability (Snider et al. 2010), we can identify the anatomical parts of the trees that are important for the connectivity of different cell types.

The performance of a supervised classifier trained on the unweighted persistence images (see Appendix A: Supervised Classification, Classification of neuronal trees) of the TMD results is demonstrated by the grouping of neuronal trees from the different species, shown in Fig 2.8. The neuronal trees of the five different species are accurately (84%) separated into the original groups. We note here that the performance of this
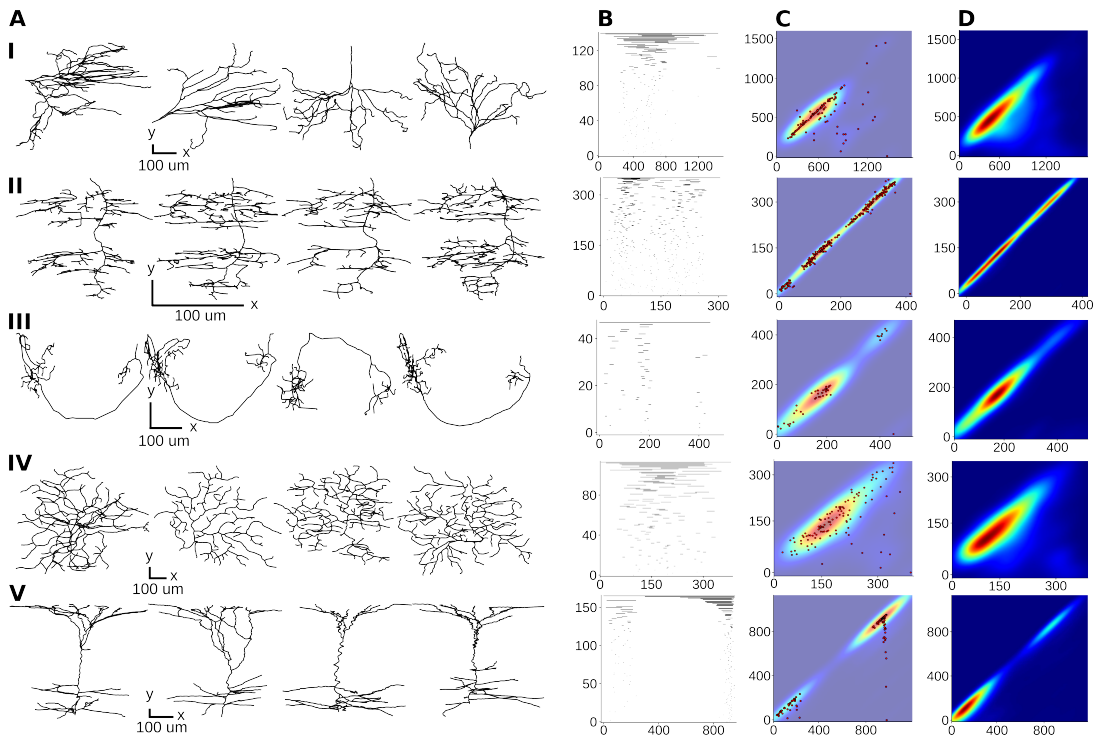
Fig. 2.8 Topological comparison of neurons from different animal species. Each row corresponds to a species: (I) cat, (II) dragonfly, (III) fruit fly, (IV) mouse and (V) rat. Trees from several exemplar cells for each species are shown in the first column (A). Representative persistence barcodes for the cells in A are shown in the second column (B). The structural differences of the trees are clearly evident in these barcodes. II, III and V have clusters of short components, clearly distinct from the largest component, while I and IV have bars of a quasi-continuous distribution of decreasing lengths. Also, barcodes III, and V show empty regions between dense regions of bars, indicating the existence of two clusters in the morphologies, while barcodes I and IV are dense overall. The unweighted persistence image for each representative barcode in B and its superimposed persistence diagram are shown in the third column (C). By combining the persistence diagrams in (C) for several trees we can define an average unweighted persistence image (D) in order to study and quantify the structural differences between distinct morphological groups. The trees in the first row (cat) are more tightly grouped than those in the second row (dragonfly), and two clusters are visible in the dragonfly trees. Considering rows 1 and 4, the extension of the elliptical peak perpendicular to the diagonal line reflects the variance in the length-scale mentioned earlier for a single cell's barcode. Note that the trees, barcodes, and unweighted persistence images are not shown to the same scale for clarity: see the scale bar in each case.

process is reliable ($> 70\%$) even for small training sets that contain only 25% of the whole dataset (see Appendix A: Classification of neuronal trees).

We applied the TMD algorithm to a more challenging use case, because it is difficult for a non-expert to distinguish the different morphologies. While pyramidal cell (PC) morphologies (Fig 2.9A) of the rat appear superficially similar, the unweighted persistence images (2.9B) reveal fundamental morphological differences between them, related to the existence and the shape of the apical tuft. The apical tuft of PCs is
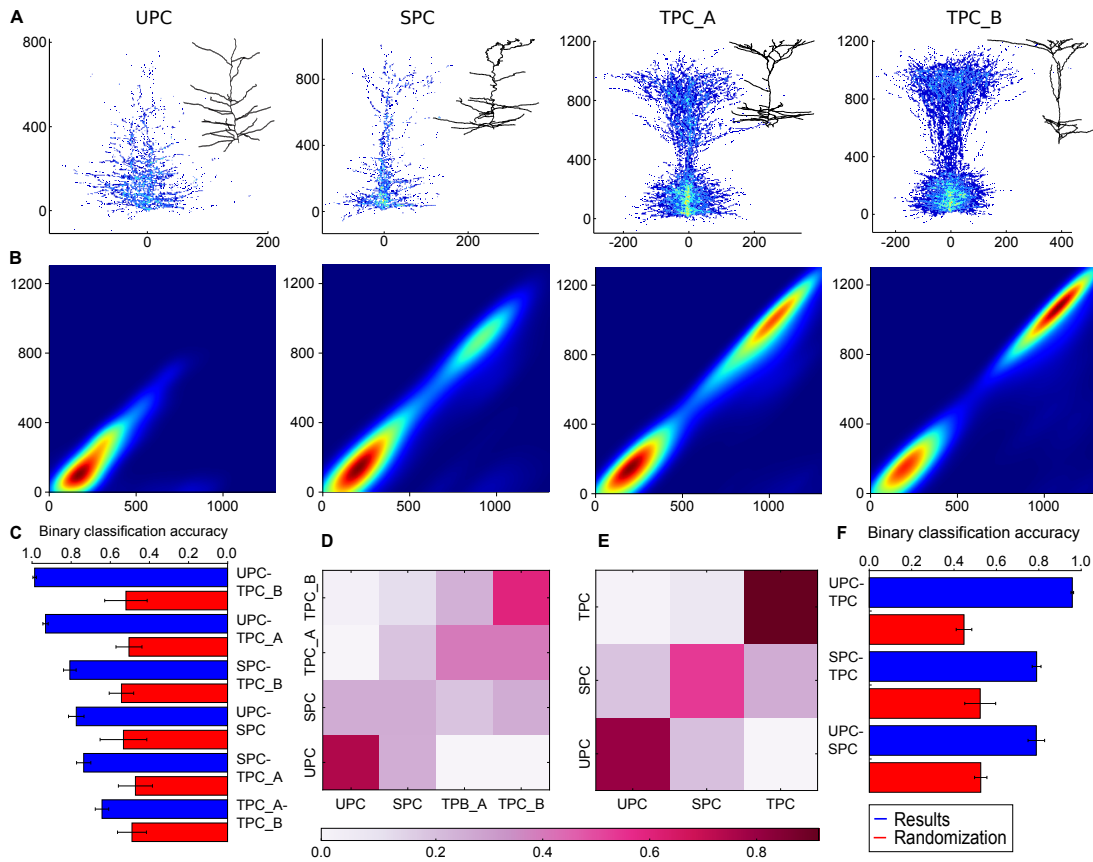
Fig. 2.9 Comparison of the TMD of apical dendrite trees extracted from several types of rat pyramidal neuron. Four cell types are shown in (A): UPC, SPC, TPC-A, TPC-B (left to right). The morphological differences between these cell types are subtle, but the unweighted persistence images (B) clearly reveal them, particularly the presence of two clusters in the TPC-A and TPC-B cell types. From these unweighted persistence images we train a decision tree classifier on the expert-assigned groups of cells. The binary classification (C) and the confusion matrix (D) based on the TMD algorithm shows an overlap of TPC-A and TPC-B trees. When those two classes are merged (E, F) the separation between the remaining types is evident. This result shows that the unweighted persistence images objectively support the expert's classification when the morphological differences between the classes are significant.

known to play a key role in the integration of neuronal inputs through their synapses in higher cortical layers, and is therefore a key indicator for the functional role of the cell.

The separation of the PC trees into four groups cannot be justified based on purely morphological grounds, since there is no coherent difference between the branching patterns of TPC-A and TPC-B (Fig 2.9C, D). On the contrary, the separation in three groups (UPC, SPC and TPC -the superset of TPC-A and TPC-B- Fig 2.9E, F) is supported by TMD-based classifiers, by detecting the fundamental differences between

their branching structures. Therefore, the TMD provides a solid benchmark test to objectively support or disprove proposed classification schemes.

Finally, the TMD algorithm can be used to assess the quality of any manually or automatically reconstructed neuron if a reference morphology is available. The best use case for this application is the datasets of BigNeuron (Peng et al. 2015), a community effort to advance single-neuron automatic reconstruction. The same stack of images of a scanned morphology is used for manual reconstruction (reference morphology) and for automatic reconstructions with a set of algorithms (test set). Due to the large number of reconstructions generated by the BigNeuron project, the analysis of the data requires a high-computational-performance algorithm. The linear complexity of the TMD makes it highly suitable for the analysis of this large dataset.

The automatic reconstructions were ranked based on their TMD-distance from the reference morphology. The TMD was able to accurately assess the quality of the automatic reconstructions, as presented in Fig 2.10, as the similarity of the branching structure of the automatic reconstructions to the reference neuron decreases with the TMD-ranking. The density plot of all the automatic reconstructions Fig 2.10A does not reproduce the shape of the reference morphology, as reconstruction errors are over-represented. On the contrary, the density plot of the ten TMD-best reconstructions closely matches the structure of the reference morphology.

Discussion

The morphological diversity of neurons supports the complex *information-processing* capabilities of the nervous system. A major challenge in neuroscience has therefore been to reliably describe the shape of neurons. We have introduced here the Topological Morphology Descriptor, derived from principles of persistent homology. The TMD of a tree retains enough topological information to allow the systematic comparison between branching morphologies. Therefore, it provides a topological benchmark for the rigorous comparison of different structures and it could advance our understanding of the anatomy and diversity of the neuronal morphologies.

This technique can be applied to any rooted tree equipped with a function defined on its nodes. Further biological examples include botanic trees (Lopez et al. 2010), corals (Kruszynski et al. 2007) and roots of plants (Wang et al. 2009). The method is not restricted to trees in $\mathbb{R}^3$, but can be generalized to any subset T of a metric space M, with a base-point R. A persistence barcode can then be extracted using a filtration by concentric spheres in M centered at R, enabling us to efficiently study the shape of complex multidimensional objects.
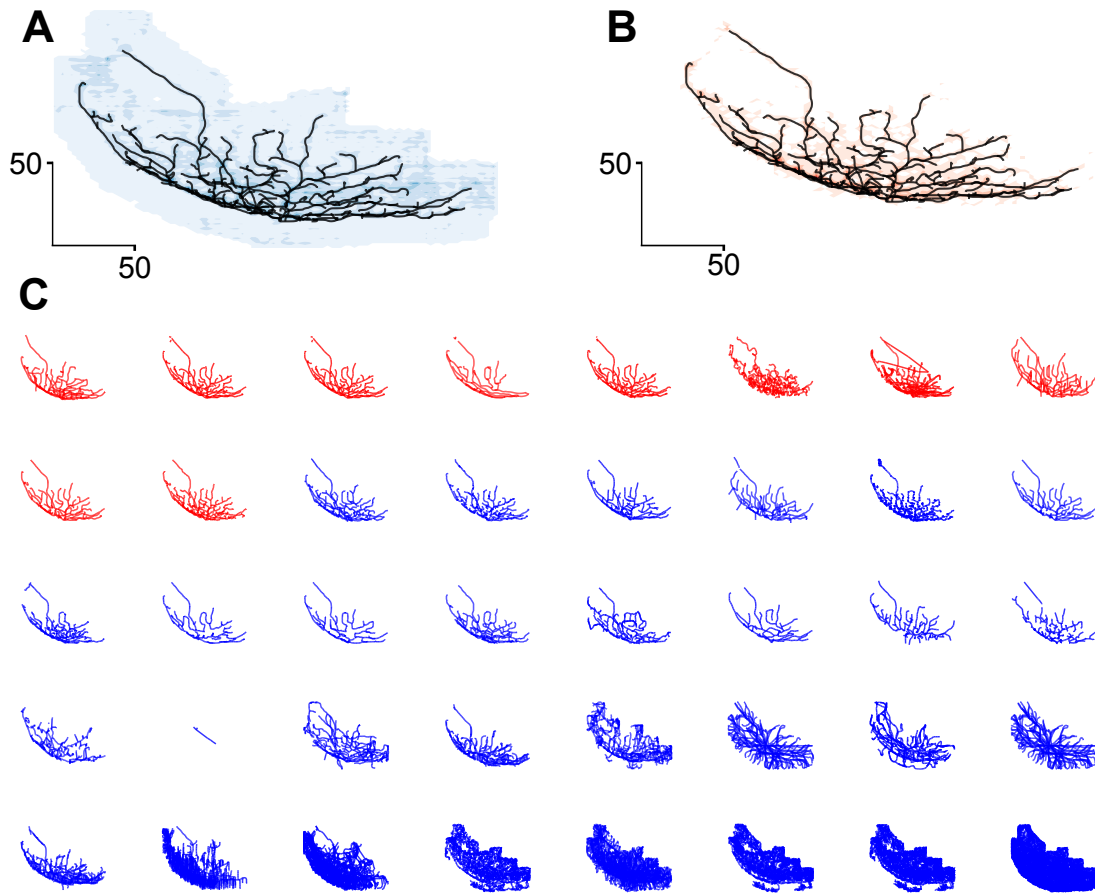
Fig. 2.10 Comparison of the TMD of BigNeuron neuronal morphologies. An image stack is used for the manual reconstruction (reference neuron) and for the automatic reconstructions produced by a variety of community supplied algorithms. The results of each algorithm are illustrated in panel C, from best (top left) to worst (bottom right). The reference neuron (in black) is visualized against the density plot of all the automatically reconstructed neurons (A, in blue), and the density plot of the ten best automatic reconstructions (B, in red), ranked according to their TMD-distance from the reference neuron. A comparison between panels A and B shows that the density plot of the ten highest ranked automatic reconstructions closely matches the structure of the reference morphology.

While the static neuronal structures presented in this paper are biologically interesting themselves, our method could also be generalized to track the morphological evolution of trees. The topological study of the growth of an embedded tree could be addressed through Multidimensional Persistence (Carlsson and Zomorodian 2009), a new area of TDA, for which computational tools are currently being explored (Lesnick and Wright 2015; Gafvert 2016). In this case the spherical filtration identifying relevant topological features of the tree could be enriched with a second filtration representing temporal evolution. This application could be useful in agriculture to study growing

roots (Wang et al. 2009) and trees, and in neuroscience, to study neurons in the developing brain.

Information sharing statement

The artificial random trees used in Fig 2.5 and Fig 2.7 were generated by software developed in BBP. The tree structures can be made available (in hdf5 format) upon request. The biological morphologies used in Fig 2.5, Fig 2.6 and Fig 2.9 are provided from Laboratory of Neural Microcircuitry (LNMC), EPFL (Romand et al. 2011). The biological morphologies used in Fig 2.8 were downloaded from Neuromorpho.org. In particular, cat neurons were provided by Rose et al. (1995), dragonfly neurons by Gonzalez-Bellido et al. (2015), fruit fly neurons by Chiang et al. (2011), mouse neurons by Badea and Nathans (2011) and rat neurons by Romand et al. (2011). The automatic and manual reconstructions used in Fig 2.10 are provided by BigNeuron (Peng et al. 2015).

Acknowledgements

Author contributions

L.K. and K.H. conceived the study. L.K. developed the topological algorithm with the contribution of P.D. P.D., K.H. contributed topological ideas used in the analysis. H.M. suggested the biological datasets analyzed. R.L., M.S. and J.C.S. gave conceptual advice. All authors discussed the results, wrote the main paper and commented on the manuscript at all stages.

References

Adams, H., Chepushtanova, S., Emerson, T., Hanson, E., Kirby, M., Motta, F., Nevill, R., Peterson, C., Shipman, P., & Ziegelmeier, L. (2016). Persistence images: A stable vector representation of persistent homology CoRR arXiv:1507.06217.

Agam, O., Bettelheim, E., Wiegmann, P., & Zabrodin, A. (2002). Viscous fingering and the shape of an electronic droplet in the quantum hall regime.

Alexandrowicz, Z. (1985). Growth and shape of branched polymers, aggregates and percolating clusters. Physics Letters A, 109(4), 169–173.

Ascoli, G.A., Donohue, D., & Halavi, M. (2007). Neuromorpho.org:a central resource for neuronal morphologies. Journal of Neuroscience, 27(35), 9247–51.

Badea, T., & Nathans, J. (2011). Morphologies of mouse retinal ganglion cells expressing transcription factors brn3a, brn3b, and brn3c: Analysis of wild type and mutant cells using genetically-directed sparse labeling. Vision Research, 51(2), 269–279.

Bauer, U., Ge, X., & Wang, Y. (2014). Measuring Distance Between Reeb Graphs, SOCG'14. New York: ACM.

Berry, M., & Bradley, P.M. (1976). The application of network analysis to the study of branching patterns of large dendritic fields. Brain Research, 109(1), 111–132.

Bille, P. (2005). A survey on tree edit distance and related problems. Theoretical Computer Science.

Blackman, A., Grabuschnig, S., Legenstein, R., & Sjostrom, P. (2014). A comparison of manual neuronal reconstruction from biocytin histology or 2-photon imaging: morphometry and computer modeling. Front in Neuroanatomy, 8(65), 65.

Carlsson, G. (2009). Topology and data. Bulletin of American Mathematical Society, 46, 255–308.

Carlsson, G., & Zomorodian, A. (2009). The theory of multidimensional persistence. Discrete & Computational Geometry, 42(1), 71–93.

Chiang, A.S., Lin, C.Y., Chuang, C.C., Chang, H.M., et al. (2011). Three-dimensional reconstruction of brain-wide wiring networks in drosophila at single-cell resolution. Current Biology, 21(1), 1–11.

Cuntz, H., Borst, A., & Segev, I. (2007). Optimization principles of dendritic structure. Theoretical Biology and Medical Model, 4(21), 21.

DeFelipe, J., Lopez-Cruz, P.L., Benavides-Piccione, R., et al. (2013). New insights into the classification and nomenclature of cortical GABAergic interneurons. Nature Reviews Neuroscience, 14(3), 202–216.

Dey, T., Shi D., & Wang, Y. (2015). Comparing graphs via persistence distortion. In SOCG.

Dieter, J. (2000). Accurate Reconstruction of Neuronal Morphology in Computational Neuroscience, Frontiers in Neuroscience. BocaRaton: CRC Press

Edelsbrunner, H., & Harer, J. (2008). Persistent homology—a survey. In Goodman, J.E., & Pach, J. (Eds.) American Mathematical Society, (Vol. 453 pp. 257–282). Providence.

Ferrante, M., Migliore, M., & Ascoli, G. (2013). Functional impact of dendritic branch point morphology. Journal of Neuroscience, 33(5), 2156–65.

Gafvert, O. (2016). Algorithms for multidimensional persistence. Gillette, T.A., & Ascoli, G.A. (2015). Topological characterization of neuronal arbor morphology via sequence representation: I - motif analysis. BMC Bioinformatics, 16, 216.

Gillette, T., Hosseini, P., & Ascoli, G. (2015). Topological characterization of neuronal arbor morphology via sequence representation: II - global alignment. BMC Bioinformatics, 16, 209.

Gomez-Gil, P., Ramirez-Cortes, M., Gonzalez-Bernal, J., Pedrero, A.G., Prieto-Castro, C.I., Valencia, D., Lobato, R., & Alonso, J.E. (2008). A feature extraction method based on morphological operators for automatic classification of leukocytes. pp. 227–232.

Gonzalez-Bellido, P., Peng, H., Yang, J., Georgopoulos, A., & Olberg, R. (2015). Eight pairs of descending visual neurons in the dragonfly give wing motor centers accurate population vector of prey direction. Proceedings of the National Academy of Sciences, 110(2), 696–701.

Jan, Y., & Jan, L. (2010). Branching out: mechanisms of dendritic arborization. Nature reviews. Neuroscience, 11, 316–28.

Knuth, D. (1998). The art of computer programming volume 2: Seminumerical algorithms. Massachusetts: Reading.

Kruszynski, K.J., Kaandorp, J.A., & van Liere, R. (2007). A computational method for quantifying morphological variation in scleractinian corals. Coral Reefs, 26(4), 831–840.

Ledderose, J., Sencion, L., Salgado, H., Arias-Carrion, O., & Trevino, M. (2014). A software tool for the analysis of neuronal morphology data. International Archives of Medicine, 7, 6.

Leinster, T., & Cobbold, C. (2012). Measuring diversity: the importance of species similarity. Ecology, 93(3), 477–489.

Lesnick, M., & Wright, M. (2015). Interactive visualization of 2-d persistence modules. arXiv:1512.00180 [cs, math].

Ling, C., Hendrickson, M., & Kalil, R. (2012). Morphology, classification, and distribution of the projection neurons in the dorsal lateral geniculate nucleus of the rat. PLoS ONE, 7, e49161.

Lopez, L.D., Ding, Y., & Yu, J. (2010). Modeling complex unfoliaged trees from a sparse set of images. Computer Graphics Forum, 29, 2075–2082.

Lyons, M.J., Budynek, J., & Akamatsu, S. (1999). Automatic classification of single facial images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(12), 1357–1362.

Mandelbrot, B., & Freeman, W. (1983). The fractal geometry of nature. Earth Surface Processes and Landforms, 8(4), 406.

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., & Wu, C. (2004). Interneurons of the neocortical inhibitory system. Nature Reviews Neuroscience, 5(10), 793–807.

Masseroli, M., Bollea, A., & Forloni, G. (1993). Quantitative morphology and shape classification of neurons by computerized image analysis. Computer Methods and Programs in Biomedicine, 41(2), 89–99.

Peng, H., Hawrylycz, M., Roskams, J., Hill, S., Spruston, N., Meijering, E., & Ascoli, G.A. (2015). BigNeuron: large-scale 3D neuron reconstruction from optical microscopy images. Neuron, 87, 252–256.

Romand, S., Wang, Y., Toledo-Rodriguez, M., & Markram, H. (2011). Morphological development of thick-tufted layer V pyramidal cells in the rat somatosensory cortex. Frontiers in neuroanatomy, 5, 5.

Rose, P., Jones, T., Nirula, R., & Corneil, T. (1995). Innervation of motoneurons based on dendritic orientation. Journal of Neurophysiology, 73(3), 1319–1322.

Schurer, T. (1994). An experimental comparison of different feature extraction and classification methods for telephone speech. pp. 93–96.

Scorcioni, R., Polavaram, S., & Ascoli, G. (2008). L-measure: a webaccessible tool for the analysis, comparison, and search of digital reconstructions of neuronal morphologies. Nature Protocols, 3, 866–76.

Scott, D. (2008). Kernel density estimators, (pp. 125–193): Wiley. Sholl, D.A. (1953). Dendritic organization in the neurons of the visual and motor cortices of the cat. Journal of Anatomy, 87, 387–406.

Snider, J., Pillai, A., & Stevens, C.F. (2010). A universal property of axonal and dendritic arbors. Neuron, 66, 45–56.

Strahler, A.N. (1952). Hypsometric analysis of erosional topography. Bulletin of the Geological Society of Science, 63, 1117–1142

The Petilla Interneuron Nomenclature Group P (2008). Petilla terminology: nomenclature of features of gabaergic interneurons of the cerebral cortex. Nature Reviews Neuroscience, 9(7), 557–568.

Van Elburg, R., & Van Ooyen, A. (2010). Impact of dendritic size and dendritic topology on burst firing in pyramidal cells. PLoS Computational Biology, 6(5), 1–19.

Van Pelt, J., Verwer, R.W., & Uylings, H.B.M. (1989). Centrifugal-order distributions in binary topological trees. Bulletin of Mathematical Biology, 51(4), 511–536.

Van Pelt, J., Uylings, H.B.M., Verwer, R.W.H., Pentney, R.J., & Woldenberg, M.J. (1991). Tree asymmetry—A sensitive and practical measure for binary topological trees. Bulletin of Mathematical Biology, 54(5), 759–784.

Van Pelt, J., Van Ooyen, A., & Uylings, H.B.M. (2001). Modeling dendritic geometry and the development of nerve connections. In de Schutter, E., & Cannon, R.C. (Eds.) (CD-ROM) Computational Neuroscience: Realistic Modeling for Experimentalist (pp. 179–208). Boca Raton: CRC Press.

Van Pelt, J., Van Ooyen, A., & Uylings, H.B.M. (2005). Natural variability in the geometry of dendritic branching patterns. In Reeke, G.N., Poznanski, R.R., Lindsay, K.A., Rosenberg, J.R., & Sporns, O. (Eds.) Modeling in the Neurosciences: From Biological Systems to Neuromimetic Robotics (pp. 89–115). Boca Raton: CRC Press.

Wan, Y., Long, F., Qu, L., Xiao, H., Hawrylycz, M., Myers, E.W., & Peng, H. (2015). Blastneuron for automated comparison, retrieval and clustering of 3d neuron morphologies. Neuroinformatics, 13(4), 487–499.

Wang, H., Siopongco, J., Wade, L., & Yamauchi, A. (2009). Fractal analysis on root systems of rice plants in response to drought stress. Environmental and Experimental Botany, 65(2–3), 338–344.

Ward, J.H.J. (1963). Hierarchical grouping to optimize an objective function. American Statistics Association, 58(301), 236–244.

Zomorrodi, R., Ferecsko, A., Kovacs, K., Kroger, H., & Timofeev, I. (2010). Analysis of morphological features of thalamocortical neurons from the ventroposterolateral nucleus of the cat. The Journal of Comparative Neurology, 518(17), 3541–3556.

2.3   Objective Classification of Neocortical Pyramidal Cells

Lida Kanari, Srikanth Ramaswamy, Ying Shi, Yun Wang, Julie Meystre, Rodrigo Perin, Julian Shillcock, Kathryn Hess and Henry Markram

Abstract: Objective classification of neuronal morphologies is a difficult task, due to the large diversity of branching shapes and the small number of detailed reconstructions. Pyramidal cells, which are one of the most common type of neuron in the neocortex, have been the focus of extensive studies, yet have not been objectively divided into consistent types based only on local axonal and dendritic features. The topological characterization of dendritic trees shows that the properties of local arbors are sufficient for the objective classification of pyramidal cells and yields a quantitative measure of the uncertainty in the classification. The topological classification reveals the existence of three objectively distinguishable types of pyramidal cells in layer 2, two in layer 3, three in layer 4, three in layer 5, and five types in layer 6. We also present possible subdivisions of these objectively defined classes based on information that is not considered in this topological descriptor. We conclude that local neuronal morphology contains sufficient information for the objective classification of neurons, settling a long-standing debate on whether cell-types are discrete or continuous morphological variations of each other.

Keywords: Neocortex, Microcircuits, Somatosensory Cortex, Pyramidal Cells, Morphology, Dendrites, Axon, Topological Analysis, Persistent homology

Introduction

The mammalian neocortex consists of interneurons, which are predominantly inhibitory, and pyramidal cells, which represent 70-85% of all excitatory neurons in the mammalian cortex (Ramón y Cajal 1911; DeFelipe and Farinas 1992; Spruston 2007; Markram et al. 2015; Ramaswamy and Markram 2015). Pyramidal cells (also termed principal cells) are characterized by a triangular soma, two distinct dendritic domains, both of which exhibit a high density of spines, emanating from the base (basal dendrites) and the apex of the soma (apical dendrites, respectively), and a single axon projecting to long distances that targets other brain regions forming several collateral branches that further bifurcate before leaving the neocortex. Basal dendrites are localized around the soma while apical dendrites ascend toward the pia, typically forming multiple oblique dendrites and terminating in a distinct tuft that is associated with high branching density. Apical dendrite impart unique functional properties to PCs and form the basis for the generation of key synaptic and active events, such as back propagating action potentials, and integration of synaptic inputs from different cortical layers (Larkum et al. 1999; Larkum et al. 2001; Schaefer et al. 2003; Spruston 2008). The unique functional properties of apical dendrites are essential for the integration of top-down (from association areas) and bottom-up streams of input (from primary sensory and motor areas) to the neocortex that shape the output firing pattern of PCs (Markram et al. 1995; Stuart et al. 1997; Larkum et al. 2001).

The characteristic morphological shapes of apical dendrites are associated with their unique functional properties, as objectively defined types of PCs are also associated with unique firing patterns (Deitcher et al. 2017) and form distinct synaptic sub-networks within and across layers (Wang, Markram et al. 2006). Therefore, the branching properties of the apical trees are commonly used for their separation into morphological classes. Visual classification by an expert, although not always trivial, usually makes it possible to distinguish the different shapes of morphologies and group neurons into classes. However, despite the expertise involved, visual inspection is subjective and often results in non-consensual and ambiguous classifications (DeFelipe et al. 2013). A striking indication of this problem, as described previously (DeFelipe et al. 2013), is the fact that experts assign a different class to a neuron than the one they had chosen in their original study for the same neuron, independently of the reconstruction quality (DeFelipe et al. 2013). For this reason, an objective classification scheme is crucial for a consensual and consistent definition of neuronal classes.

Morphological analysis is usually performed on 3D digital reconstructions of neurons. Because standard morphometrics (such as section length, bifurcation angles etc.) focus
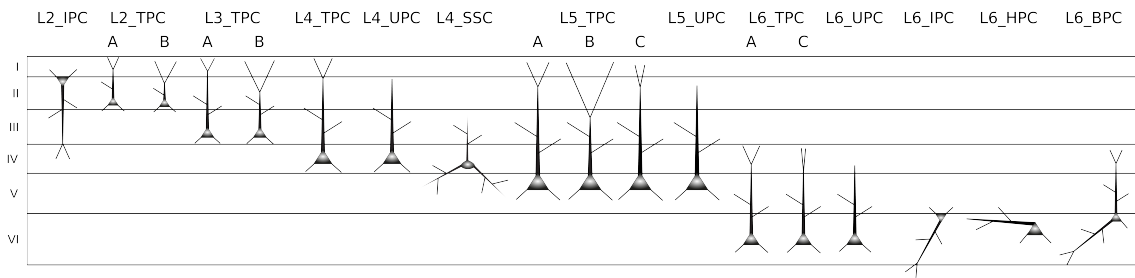
Fig. 2.11 Schematic representation of all PC types/subtypes in layers 2-6. Tufted PCs (TPCs) are found in all layers from 2 to 6 and are identified by the existence of a tuft formation and typically a large number of obliques. TPCs can be separated in the following subtypes: A) early bifurcating, found in layers 2,3,5,6 B) late bifurcating, found in layers 2,3,5 and C) small tufted, found in layers 5, 6. Untufted PCs (UPC) are located in deeper layers (layers 4-6) and are described by the absence of a tuft formation. Inverted PCs (IPC) are vertically oriented in the direction away from the pia (layers 2,6). Bitufted PCs (BPC) are found in layer 6 and are identified by two apical trees that often project to opposite directions (towards the pia and the away from it). Spiny stellate cells (SSC) are found in layer 4 and are described by a short apical dendrite, of length that is comparable to the basal dendrites, the absence of tuft formation and a small number of obliques. Horizontal pyramidal cells (HPC) are only found in layer 6.

on a different local aspect of the neuronal morphology, they must be used in combination with other morphometrics for the classification of neurons. To avoid over-fitting, which is a result of using a large number of features when few individual cells are available, feature selection is required. In order to establish a rigorous categorization of morphologies, it is important to identify the metrics that are indicative of the differences between neuronal shapes and that can be generalized across different brain regions and species. However, feature selection is often subjective and often the feature sets proposed by different experts are inconsistent (DeFelipe et al. 2013). In addition, alternative sets of morphometrics result in different classifications (Kanari et al. 2017), since statistical features commonly overlap even across markedly different morphological types. To avoid this issue, we have developed an alternative representation of morphologies based on persistent homology, that provides a potential standardized quantification of neuronal branching structure.

The Topological Morphology Descriptor algorithm generates a barcode from the neuronal tree, coupling the topology of the branching structure with its geometry, and therefore encoding the overall shape of the tree in a single descriptor(Kanari et al. 2017). The advantage of this metric is that it captures correlations between features overcoming the need for manual feature selection. The TMD is a simplified representation of the original tree that retains enough information to perform well in a discrimination task, by mapping the tree to a topological representation with less information loss than the usual morphometrics. Using this topological representation, we were able

to establish that PCs can be objectively classified based on the branching structure of their apical dendrites. We compared the results of the topological classification to the expert-proposed classes, and illustrated that the majority of subjective classes can be objectively supported, with a single exception of Layer 5 subtypes (TPC:A and TPC:B). The results are also compared to literature to identify the association of local dendritic properties to long-range projections of axons.
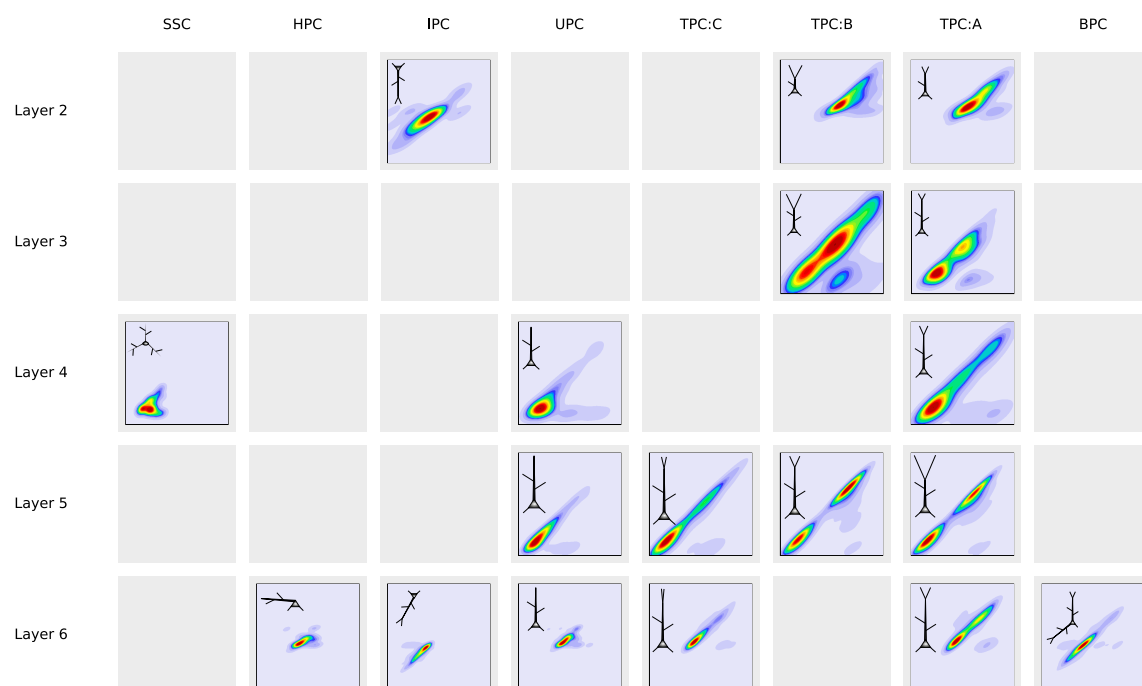
Fig. 2.12 Morphological classification table. Inspired by Ascoli and Wheeler (2016) the classes of cortical pyramidal cells are sorted according to increasing complexity (higher number of bars in the corresponding persistence barcodes, from left to right) and increasing cortical depth (top to bottom) to generate a table of morphological diversity of cortical pyramidal cells.

Methods

Staining and reconstruction techniques

PCs were filled and stained with biocytin and reconstructed in 3D using Neurolucida from $300\mu m$ thick slices of rat somatosensory cortex. The reconstructed PCs from all layers of rat somatosensory cortex were then classified PCs based on their topological profiles.

Topological classification

For the topological classification, we first separated PCs according to the location of their somata into layers. Then, PCs were separated into classes based on the topological morphology descriptor (TMD) of their apical branching patterns (Kanari et al. 2017). The same analysis shows no significant difference on the branching patterns of local axons and basal dendrites of different m-types. The branching pattern of each apical tree is decomposed into a set of components depending on a morphological feature of each branch that is used as a filtration function. In this study we use the radial distance from the soma as the discriminating factor. Because we want to take into account the orientation of the trees, the radial distance is weighted according to the orientation of the tree towards the pia surface. The set of all components is encoded in a topological barcode (Carlsson 2009) using the TMD algorithm that is described in details in (Kanari et al. 2017). The TMD of the tree is then used for the generation of the persistent image (Chepushtanova et al. 2015) of the tree that summarizes the density of components at different radial distances from the soma. The persistence image representation is a vector that can be used as input to different machine learning algorithms.

A supervised classifier is trained on the proposed by the expert classes. Then each neuron is labeled according to its TMD profile. The accuracy of the classification is defined as the ratio of the number of TMD-labels that agree with the expert classification over the total number of cells. The classification is then repeated for a set of randomized labels according to the initial number of classes. If the expert classification accuracy is significantly higher than the randomized classification, the proposed grouping is accepted. If this is not the case the classification cannot be supported by the TMD. Then the classes are redefined according to the TMD profiles of the neurons of the same layer with the objective of an optimal separation between the defined classes.

The TMD classification is unbiased since it is based on a stable topological descriptor of the tree's branching structure and thus it is less prone to user induced biases. There is no need to analyze the tree based on different morphological features in an attempt to combine and use the ones that are significant. This way we avoid over-fitting, i.e., confusing the random noise in the biological structure as a significant discrimination factor, by implicitly accounting for the correlations between features which are incorporated into their TMD profile. Indeed, the feature that is used as a filtration function influences the result of the classification. It is therefore important to select the morphological feature that will serve as the filtration function intelligently

or to combine the TMD profiles for different features. However, the use of alternative features revealed that the use of the radial distance as a filtration function is equivalent to a number of other features (such as path distance, section lengths etc.) and therefore there is no need to combine multiple functions for the study of PCs.

Results

The topological analysis of the branching structure of the apical trees of PCs revealed the existence of three subtypes in Layer 2, two subtypes in Layer 3, three subtypes in Layer 4, three subclasses in Layer 5 and five subclasses in Layer 6 (Fig 2.11). Apical dendrites of PCs in supragranular layers 2 & 3 reach layer 1 and the pia; PCs in layers 4 and 6 often reach the supragranular layers, but not layer 1; major PC subtypes in layer 5 have the longest apical dendrites reaching layer 1 and the pia, and minor PC subtypes in layer 5 tend to extend to the supragranular layers, but not layer 1 (Fig 2.11). The expert analysis of the cells also revealed the existence of two more subtypes: a subtype of TPC cells in Layer 5, and a horizontally oriented cell type in Layer 6. The subtype of horizontally oriented cells in Layer 6 is supported by the main orientation of the apical tree, but the topological profiles of these cells are indistinguishable from the un-tufted pyramidal cells of Layer 6 (L6_UPC). Finally, by combining previous studies mainly on primary sensory cortices, we illustrate the correlation between PCs classified according to the proposed scheme and their long-distance projections, providing further support for the proposed classification scheme. The results of the topological analysis are summarized in the following section organized per layer.

Pyramidal Cells in Layer 2

The TMD-based clustering of Layer 2 apical trees illustrates the existence of two major classes, depending on the direction of the apical tree. The L2_IPCs have apical trees that project in the direction opposite to the pia, therefore generating a higher density of branches in this direction (see Fig 2.13). On the contrary, the L2_TPCs contain apical dendrites that project towards the pia, therefore exhibiting a higher density of branches in this direction. Further analysis of the branching patterns of L2_TPCs results in a separation into two sub-classes (L2_TPC:A and L2_TPC:B) depending on the density of branches on the distal apical dendrites. In fact, L2_TPC:A have a small density of branches within the tuft, while L2_TPC:B do not.

A quantitative analysis based on the morphometrics of 3D reconstructions of the three Layer 2 subtypes of PCs (L2_TPC:A, n = 6; L2_TPC:B, n = 33; L2_IPC,
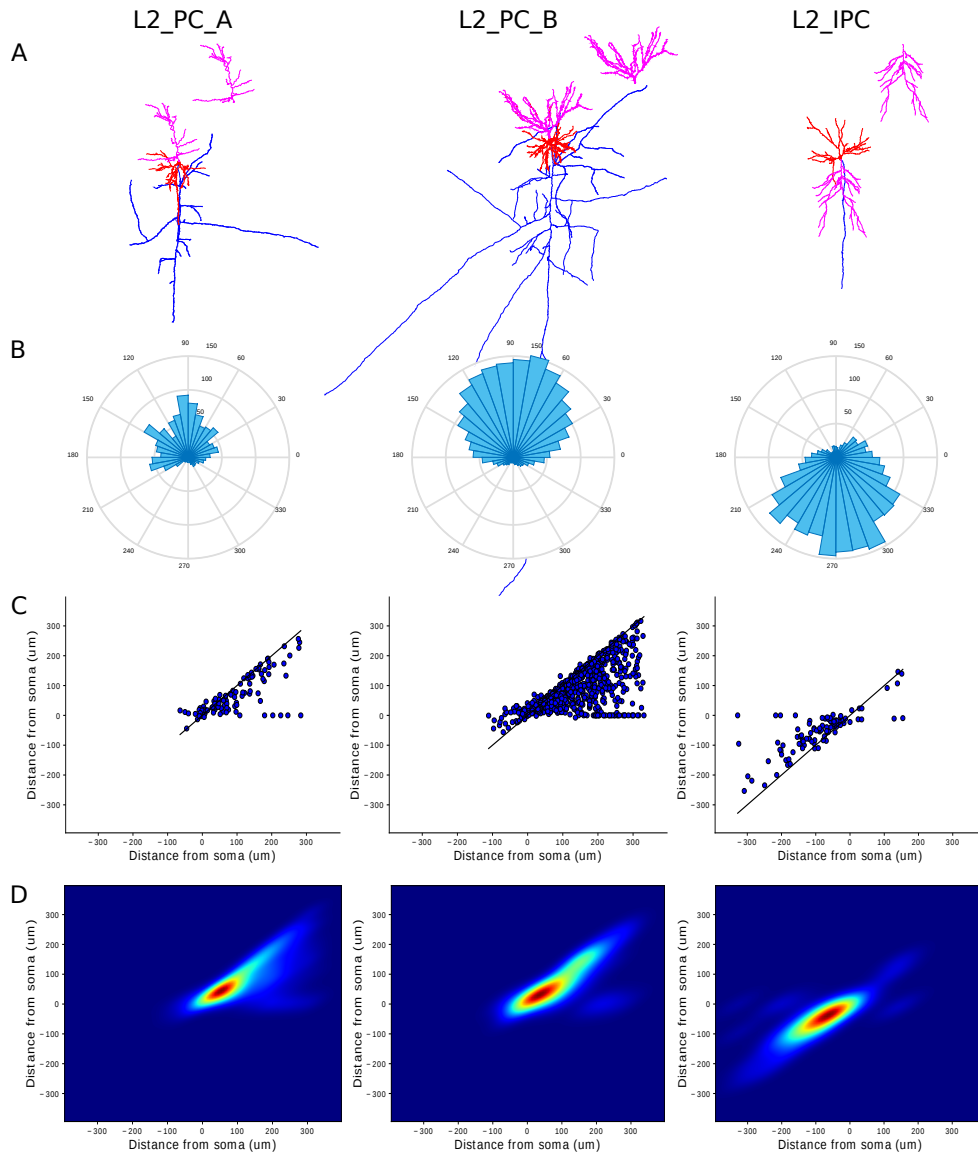
Fig. 2.13 Three PC types/subtypes in layer 2. A. Exemplar 3D reconstructed PCs: apical dendrite in pink, (duplicated at the upper right side for clarity); soma and basal dendrites in red; axons in blue. B. Polar plot analysis of apical dendrites: vertically oriented for L2_TPC:A and L2_TPC:B, vertically oriented in the opposite direction for L2_IPC. C and D. The Topological Morphology Descriptor (TMD) of apical dendrites represents the spatial distribution of branches with respect to the radial distance from the neuronal soma. In the persistence diagram (C) the radial distances of the component's initial (y-axis) and final (x-axis) points are illustrated. The average persistence image (D) of an m-type shows the average density of the tree's components at each radial distance. L2_IPC apicals project on the opposite direction of the pia, L2_TPC apicals project towards the pia; L2_TPC:A apicals have a small density of branches within the tuft compared to the L2_TPC:B apicals.

n = 4) revealed a quantitative difference of soma size, whose surface area is larger for L2_TPC:B compared to L2_TPC:A. The basal dendrites of L2PC subtypes share similar morphological features, and therefore no morphological difference can be quantitatively justified. The total length, surface area and volume of the apical dendrites of L2_TPC:B cells are significantly higher compared to L2_TPC:A cells, reflecting their broad extends. In addition, L2_TPC:B axons extend further, resulting in larger total lengths and surface areas, as well as maximum branch orders, suggesting the formation of dense local axonal clusters.

Expert based observations of the same dataset suggest the existence of three distinct classes: L2_IPC (inverted PC) have a vertically inverted apical dendrite projecting towards deep layers and white matter that forms a proximal or distal extensive tuft formation and multiple oblique dendrites. The apical dendrites of both L2_TPC:A and L2_TPC:B sub-types reach the pia, and differ mainly in the bifurcating point along the apical dendrite where the tufts begin to form; proximal or distal. Therefore, the TMD-based classification supports the subjective observations for Layer 2 PCs, and for consistency we use the expert proposed terminology for those classes.

Pyramidal Cells in Layer 3

The TMD-based clustering of the Layer 3 apical trees (Fig 2.14) illustrates the existence of two sub-classes, depending on the density of branches on both proximal and distal to the soma radial distances. The L3_TPC:A have apical trees with high density of branches close to the soma, but lower density of branches within the tuft. On the contrary, the L3_TPC:B apicals have a smaller density of branches around the soma, but higher density of branches on the tuft.

Quantitative morphological analysis on the two subtypes of layer 3 PCs (L3_TPC:A, n = 35; L3_TPC:B, n = 9) does not reveal any significant difference on the somatic and axonal features of the two sub-types. The differences between the two sub-types are only supported by the morphometrics of the apical dendrites. On average, L3_TPC:A cells have a larger number of oblique dendrites compared to L3_TPC:B cells, which corresponds to the lower densities of the latter observed in their persistence images.

Expert based observations of the same dataset suggest the existence of two distinct classes of layer 3 PCs, both of which are oriented towards the pia: the L3_TPC:A have a vertically projecting apical dendrite, with an often distal (occasionally proximal) onset of tuft formation, which forms a small tuft (occasionally extensive) and multiple oblique dendrites before tuft formation. On the contrary, the L3_TPC:B have a vertically projecting apical dendrite with distal onset of tuft formation, which forms a
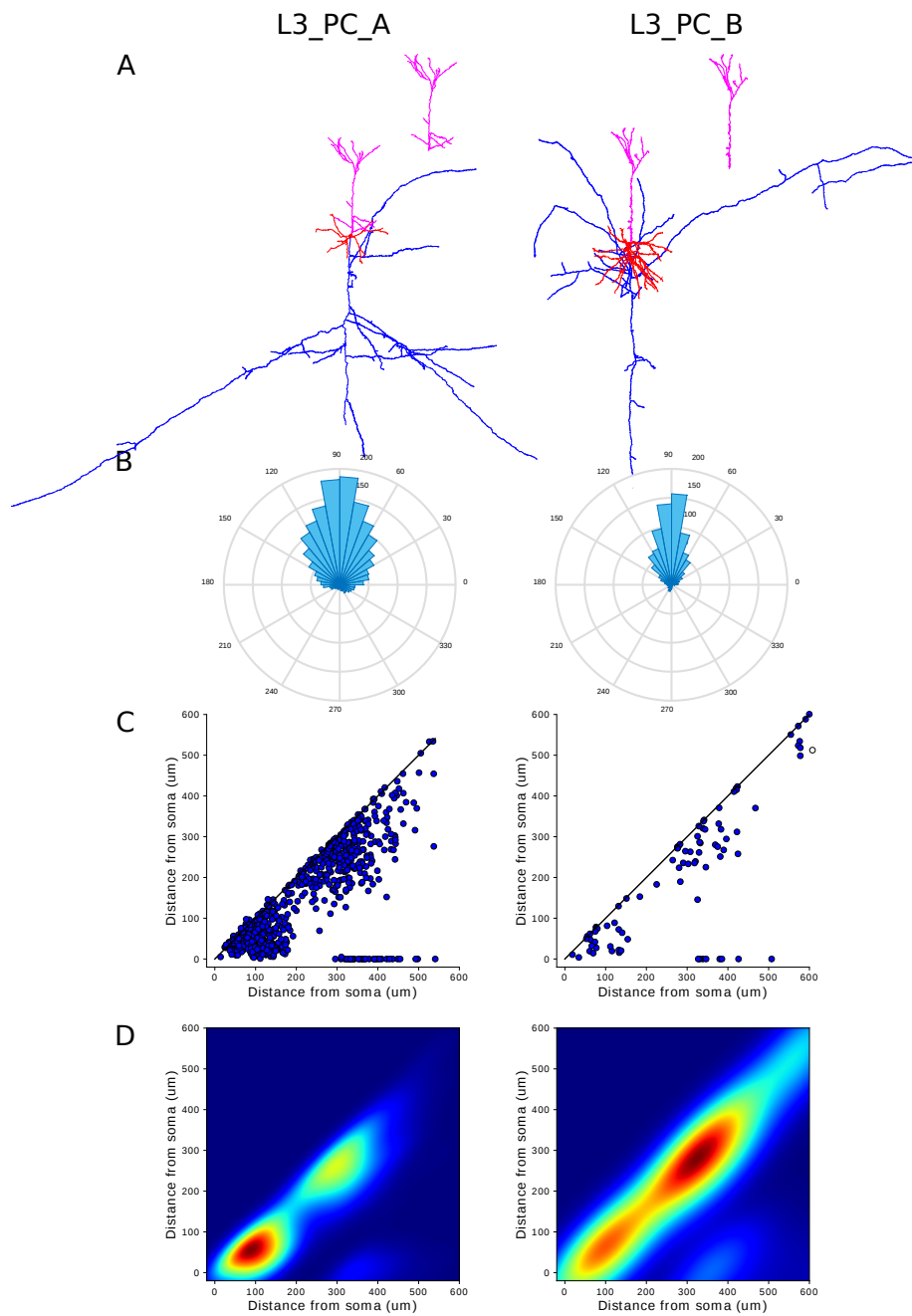
Fig. 2.14 Two PC subtypes in layer 3. A. Representative 3D reconstructions of PCs: The color-coding for different neuronal compartments is the same as in Figure 2.11. B. Polar plot analysis of apical dendrites: vertically oriented for both L3_TPC:A and L3_TPC:B. C and D. Topological Morphology Descriptor (TMD) of L3 apical dendrites. Both L3_TPC sub-types project towards the pia; L3_TPC:A apicals have a high density of branches close to the soma, but lower density of branches within the tuft and L3_TPC:B apicals have a smaller density of branches around the soma, but higher density of branches on the tuft.

small tuft and few oblique dendrites before tuft formation. Therefore, the TMD-based classification supports the subjective observations of two sub-classes in Layer 3 PCs.

Compared to the pyramidal cells of higher layers (Layer 2, Fig 2.13), the PCs of Layer 3 appear to be larger on average, presenting larger extends and higher densities of branches, associated with larger total lengths. However, individual cells of Layer 3 can be smaller than Layer 2 cells, indicating that they cannot be distinguished merely by standard morphometrics. As a result, the information of soma location is essential for the analysis of Layer 2 and Layer 3 pyramidal cells. The axonal bouton density of L2 and L3 PCs is similar, 18 - 21 boutons/100 μm on average. Previous studies examine L2 and L3 PCs together, yielding two subtypes, which primarily differ in axonal morphology (Larsen and Callaway 2006). One subtype of layer 2/3 PCs is sending axonal collaterals into layers 3 and 5 avoiding layer 4. The other subtype is usually located at the border of layer 3 and has significantly more axonal collaterals distributed in layer 4, a subtype that could be associated to the corresponding L3_TPC:B of our study.

Pyramidal Cells in Layer 4

The TMD-based clustering of the Layer 4 apical trees (Fig 2.15) illustrates the existence of three major classes, that differ in the extent and the shapes of their apical trees. The L4_TPCs have a long apical tree that extends to large radial distances and forms a tuft that presents a high density of branches on the distal from the soma radial distances. The L4_UPCs apical trees also extend to large radial distances, but do not form a discrete tuft, as only few branches per tree reach to the maximum radial distances. The apical trees of L4_SSCs present a high density of branches proximal to the soma, but only extend to small radial distances (about half of the radial distances of L4_TPCs).

Quantitative analysis based on the morphometrics of 3D reconstructions of the three subtypes of layer 4 PCs (TPC, n = 44; UPC, n = 33; SSC, n = 12) illustrates that L4_SSC have smaller somata than L4_TPC and L4_UPC. On average, L4_TPCs have a larger number of basal dendrites compared to L4_UPC and L4_SSC, which are also significantly longer. Similarly, L4_TPCs apical trees are bigger (larger total length, areas and volumes) than both other types, even though both L4_TPC and L4_UPC apical extends are significantly longer than L4_SSC. Due to the significant loss of axonal mass, resulting from the slicing preparation, the results concerning the axonal morphometrics are inconclusive. However, the existence of three classes is in agreement with previous studies that used thicker brain slices (500 μm thick), (Staiger,
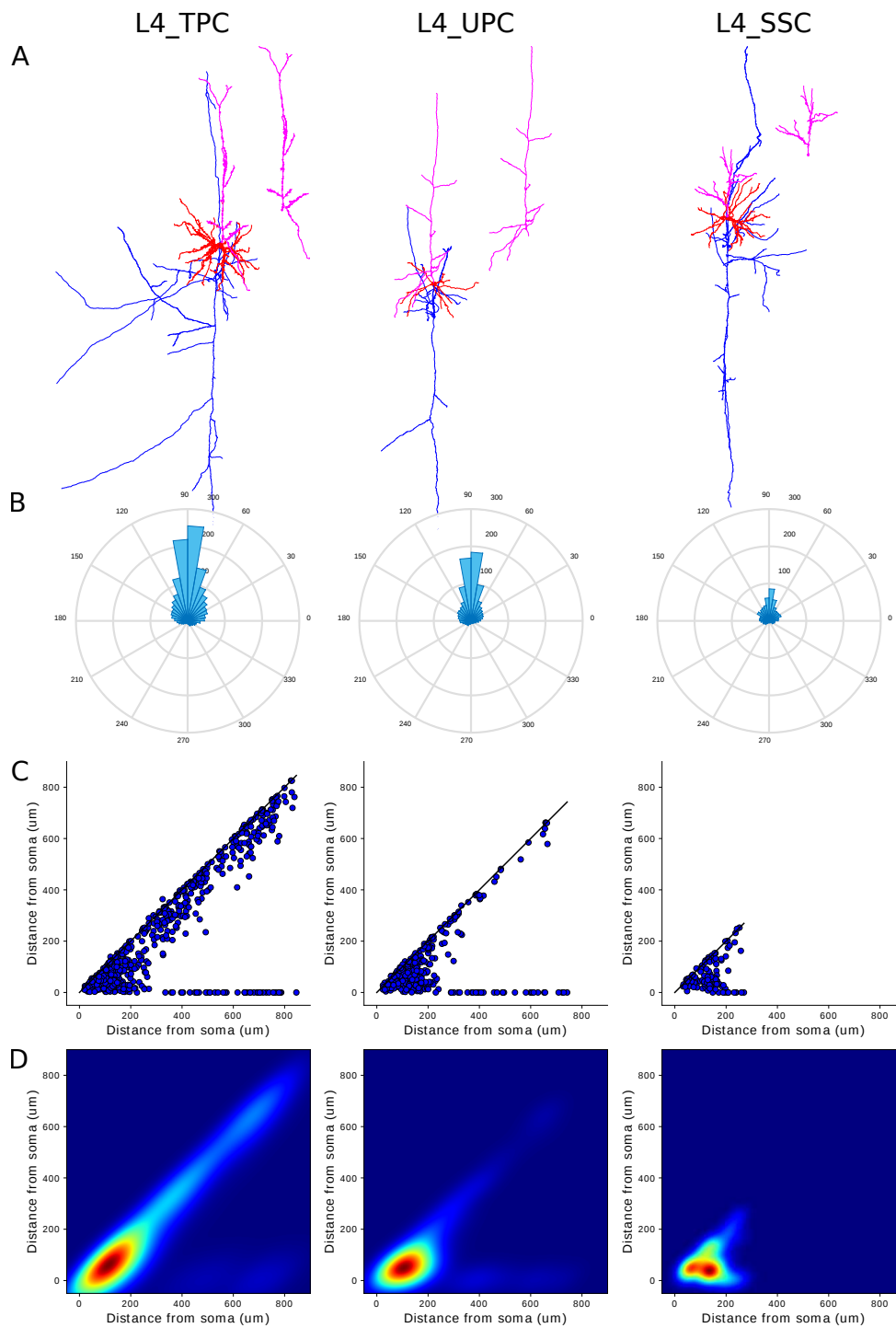
Fig. 2.15 Three PC subtypes in layer 4. A. Representative 3D reconstructions of PCs. B. Polar plot analysis of apical dendrites: vertically oriented for all three subtypes of L4_TPC and L4_UPC and L4_SSC. C and D. Topological Morphology Descriptor (TMD) of L4 apical dendrites. L4_TPC apicals extend to large radial distances and form a tuft of high density of branches on the distal from the soma radial distances. L4_UPC apicals also extend to large radial distances, but do not form a discrete tuft, as only few branches per tree reach to the maximum radial distances. L4_SSC apicals have a high density of branches proximal to the soma, but only extend to small radial distances.

Flagmeyer et al. 2004) and reported three distinct classes based on the axonal patterns of the pyramidal cells of Layer 4. In agreement with this study, the bouton density of L4_UPCs (22 ± 1 boutons/100 μm) is significantly higher than those of L4_TPCs and L4_SSCs (19 ± 1 and 18 ± 1 boutons/100 μm; both P = 0.02).

Expert based observations of the same dataset suggest the existence of three types of layer 4 PCs, based on their apical dendrites. The L4_TPC (tufted pyramidal cells) have a vertically projecting apical dendrite with a distal small tuft and multiple oblique dendrites before tuft formation. The L4_UPC (untufted pyramidal cells) have a vertically projecting apical dendrite without a tuft and multiple oblique dendrites that branch proximally to the soma. The L4_SSC (spiny stellate cells) have a vertically projecting apical dendrite with small radial extends, not much longer than basal dendrites. Typically, the apical dendrites of all PC types in layers 4 do not reach layer 1. Therefore, the TMD-based classification supports the subjective observations of three major classes in Layer 4 PCs.

Pyramidal Cells in Layer 5

The TMD-based clustering of the Layer 5 PCs illustrates the existence of three major classes, that differ in the branching of their apical trees. The L5_TPC:A have a long apical tree that extends to the largest radial distances reaching Layer 1. L5_TPC:A apical trees have two distinct clusters of high density of branches that differ on their radial distance from the soma. The cluster proximal to the soma corresponds to the rich oblique formation, while the distal from the soma region corresponds to the formation of a densely branching tuft. Similarly to L5_TPC:A the apical trees of L5_TPC:C have two distinct clusters of high branching density, one proximal to the soma that corresponds to the obliques and one distal to the soma that corresponds to the tuft. However, the tufts of L5_TPC:C have a lower density of branches even though they reach up to large radial distances. L5_UPC have a single high branching density cluster proximal to the soma, which corresponds to rich oblique formation. The reach of the apical trees of L5_UPC are lower that the rest of L5PCs as the density of branches decreases with the radial distance from the soma, indicating the absence of a tuft.

The quantitative analysis of 3D reconstructions of three subtypes of layer 5 PCs (TPC:A and TPC:B, n = 98; TPC:C, n = 33; UPC, n = 30) showed that L5_TPC:A have significantly bigger somata compared to L5_TPC:C and L5_UPC. The basal dendrites of layer 5 PC extend approximately to the width of a local cortical microcircuit ($300 - 500$ μm), except from L5_UPCs basal dendrites, which are narrower. L5_TPC:A

Fig. 2.16 Four PC subtypes in layer 5. A. Representative 3D reconstructions of PCs. B. Polar plot analysis of apical dendrites: vertically oriented for all four subtypes of L5_TPC:A, L5_TPC:B, L5_TPC:C and L5_UPC. C and D. Topological Morphology Descriptor (TMD) of L5 apical dendrites. L5_TPC:A and L5_TPC:B apicals have two distinct clusters of high branching densities; one proximal to the soma that corresponds to the rich oblique formation, and one distal from the soma that corresponds to the formation of a dense tuft. L5_TPC:C apicals have two distinct clusters of high branching density similarly to L5_TPC:A but present a lower density of branches on the distal cluster indicating a less dense tuft. L5_UPC apicals have a single high branching density cluster proximal to the soma that corresponds to rich oblique formation, but the density of branches decreases with the radial distance from the soma, indicating the absence of a tuft. The expert proposed separation in L5_TPC:A and L5_TPC:B cannot be supported by the TMD classification as no significant differences were found in the topological profiles of those subtypes.

have a significantly larger basal dendritic surface area that enables higher synaptic inputs in comparison to the two sub-types that have longer but thicker basal processes. The morphological properties of L5_TPC:A apical trees confirm the topological results. In addition, L5_TPC:A (15-16 boutons/100 μm) have bouton densities significantly lower than those of L5_TPC:C and L5_UPCs (21 boutons/100 μm). Recent advances in retrograde labeling of single neurons in vivo with recombinant rabies virus (Larsen et al. 2007) resulted in the reconstruction of complete axons of layer 5 PCs, which support the existence of three distinct subtypes based on their axonal properties. The thick-tufted PCs (corresponding to L5_TPC:A) project their local axons within deep cortical layers, while the slender-tufted PCs (L5_TPC:C) and the short untufted PCs (L5_UPCs have extensive projections to superficial layers. The axons of L5_UPCs are relatively columnar, while those of L5_TPC:Cs have extensive laterally spreads within layer 2/3. Compared to in vivo labeling (Larsen et al. 2007, Oberlaender et al. 2011), morphological measurements of axons obtained by in vitro (300 μm thick brain slices) labeling are underestimated, since the laterally spreading axonal processes have been largely severed during the slicing procedure.

Expert based observations of the same dataset suggest the existence of two major classes and four sub-classes. The L5_TPC:A (Thick-tufted PC:A) have a vertically projecting apical dendrite with a distal broad thick tuft and multiple oblique dendrites emerging proximally. The L5_TPC:B (Thick-tufted PC:B) are similar to the L5_TPC:A but further bifurcate into smaller tufts in comparison with L5_TPC:A. The L5_TPC:C (small tufted PC) have a vertically projecting apical dendrite with a distal small tuft and multiple oblique dendrites emerged proximally. The L5_UPC (untufted PC) have a vertically projecting apical dendrite with no tuft formation. The expert classification in L5_TPC:A and L5_TPC:B could not be validated by the TMD classification as no significant differences were found in the topological profiles of those subtypes. Based on the topological profiles of their apical trees, there is a gradient between those two subtypes, defined by experts, rather than a clear separation in two distinct classes. Therefore, the TMD-based classification supports the existence of three major classes of L5_PCS, but not their separation into L5_TPC:A and L5_TPC:B subtypes, as proposed by expert observations.

Pyramidal Cells in Layer 6

The TMD of Layer 6 PCs apical trees indicates the existence of a large diversity of classes, unlike any other layer. TMD-based classification revealed the existence of five distinct types of PCs in Layer 6. L6_BPC are identified by the two vertically

projecting clusters of branching clusters that project to opposite directions. Both of the apical trees of L6_BPC form a small distal tuft, which is indicated by a small distal cluster of branches in the persistence image (Figure 6), and a high density of branches close to the soma. L6_IPC are identified by the orientation of their apical trees, which are directed towards white matter. The low distal branching density of L6_IPC apicals indicates the existence of a small tuft. L6_UPC apicals have a single dense cluster of branches proximal to the soma, which corresponds to a rich oblique formation. L6_UPC have smaller extends than L6_TPC and the density of branches decreases with the radial distance from the soma, indicating the absence of a tuft. L6_TPC, which form a distinct and large tuft, can be separated into two sub-types, as in the case of Layer 5 PCs. The L6_TPC:A have a long apical tree that extends to large radial distances (and reaches Layer 4) and forms two distinct clusters of branches with respect to the radial distance from the soma. The cluster proximal to the soma corresponds to the rich oblique formation, while the distal cluster corresponds to the formation of a densely branching tuft. The L6_TPC:C have also two distinct clusters of branches, one proximal to the soma that corresponds to the obliques and one distal to the soma that corresponds to the tuft. However, the tufts of L6_TPC:C have lower density of branches in their tufts compared to L6_TPC:A. The last subtype of Layer 6 PCs is L6_HPC. The apical dendrites of L6_HPC have similar topological profiles to the L6_UPCs, but appear to have a preferred horizontal orientation, as opposed to all the other Layer 6 PCs.

The quantitative analysis based on the morphometrics of 3D reconstructions of layer 6 PCs (TODO: L6_TPC:A, n = 26; L6_TPC:C, n = 18; L6_UPC, n = 23; L6_IPC, n = 27; L6_BPC, n = 32; L6_HPC, n = 7) illustrates that the somata of L6_HPCs are the biggest in layer 6 compared to other subtypes. L6_TPC:C basal dendrites are the smallest (minimum total length) among all layer 6 PCs, the L6_HPCs basal dendrites have the widest maximum horizontal extent, but the smaller number of dendritic trees. L6_TPC:As and L6_UPCs have greater total dendritic length than all other layer 6 PCs, except HPCs. Quantitative analysis of layer 6 PCs axons demonstrates that they are largely similar, with the exception of L6_TPC:Cs, which have the narrowest axonal trees with the smallest maximum horizontal extent that is approximately equal to the width of a cortical column. In addition, the bouton density of L6_TPC:Cs is the lowest (17 boutons/100 μm) and of L6_HPCs the highest (22 boutons/100 μm). The bouton density is similar between other types/subtypes of layer 6 PCs, ranging from 19 to 20 boutons/100 μm on average. However, axonal collaterals of most PCs filled in slices have been severed to nearly 90% since their axonal clusters

commonly extend across multiple columns, while their dendritic clusters are confined within a column (Boudewijns, Kleele et al. 2011). Due to this technical limitation, the morphometrics of Layer 6 axonal branches will not be discussed further.

Subjective observations suggest the existence of five major types and two subtypes. The L6_TPC:A (tufted PC) have a vertically projecting apical dendrite with a distal small tuft and multiple oblique dendrites. The L6_TPC:C (Narrow PC) have a narrow vertically projecting apical dendrite, with a distal small tuft and often more oblique dendrites than other PC types. The L6_UPC (untufted PC) have a vertically projecting apical dendrite with no tuft formation, but multiple oblique dendrites. The L6_IPC (inverted PC) have a vertically inverted apical dendrite projecting towards the white matter with a distal small tuft and multiple oblique dendrites. The L6_BPC (bitufted PC) have two vertically projecting apical dendrites; one oriented toward the pia with a distal small tuft that forms multiple oblique dendrites and one inverted, projecting towards the white matter with a distal small tuft and multiple oblique dendrites. The L6_HPC (horizontal tufted PC) have a horizontally projecting apical dendrite with a distal small tuft that forms a few oblique dendrites. The apical dendrites of layer 6 PCs often reach layer 4 or supragranular layers, but very rarely reach layer 1. Therefore, the TMD-based classification supports the existence of five subtypes in Layer 6 and an additional class (L6_HPC) can be identified by using the main orientation of the apical tree as a distinctive parameter.

Discussion

Despite the expertise involved, visual inspection is subjective and often results in non-consensual and ambiguous classifications (DeFelipe et al. 2013). In this study, we used a novel metric based on persistent homology (Kanari et al. 2017), which quantifies the branching structure of apical dendrites, to establish an objective standardized classification of pyramidal neuron morphologies in the rat somatosensory cortex. We have demonstrated that the TMD of neuronal trees is not only reliable in validating the quality of the expert classification, but can also propose an alternative separation of cells into groups where the expert classification fails to provide a consensual and consistent definition of neuronal classes.

This scheme revealed the existence of a common type of PC in layers 2-6 – the TPC, and those that are uniquely found in specific layers, such as the SSC in layer 4 and the BPC in layer 6. Interestingly, the variability of apical shapes increases with the depth of the cells in the tissue. This indicates that the higher complexity function of deeper cortical layers can be successfully supported by the large morphological variability
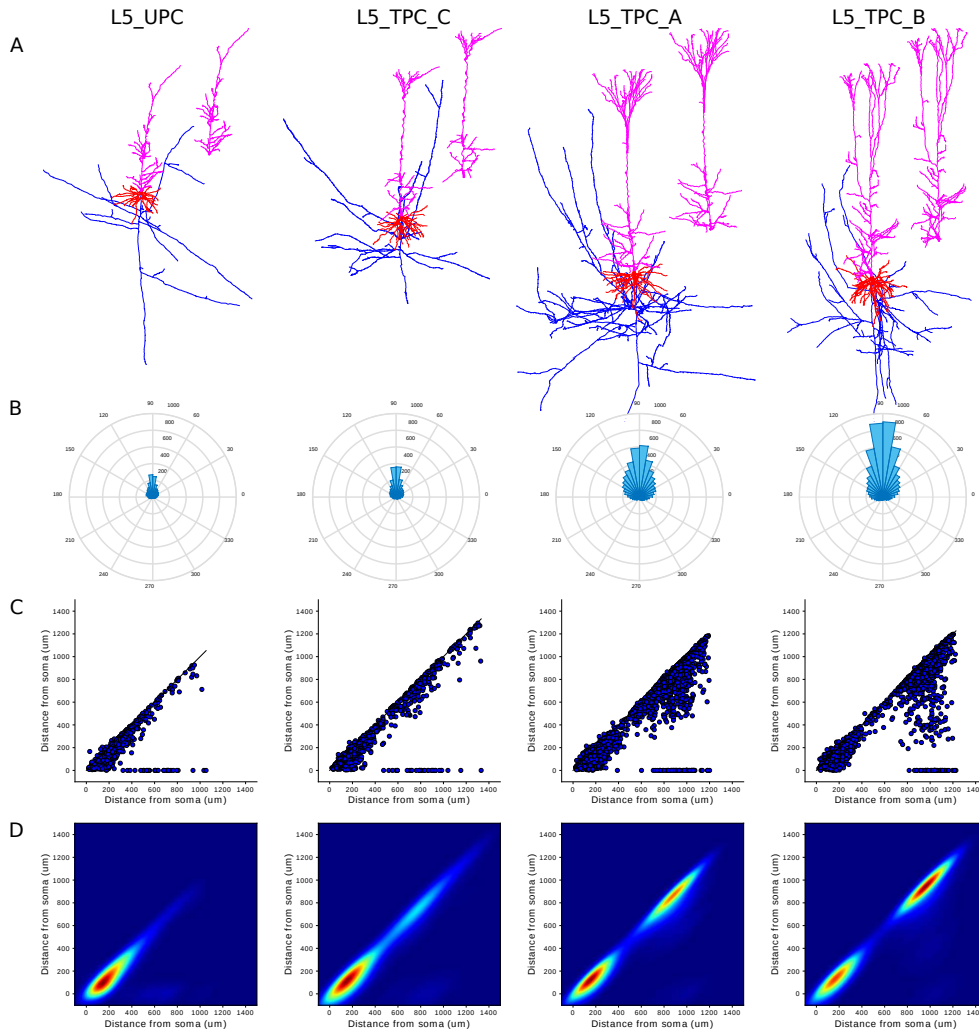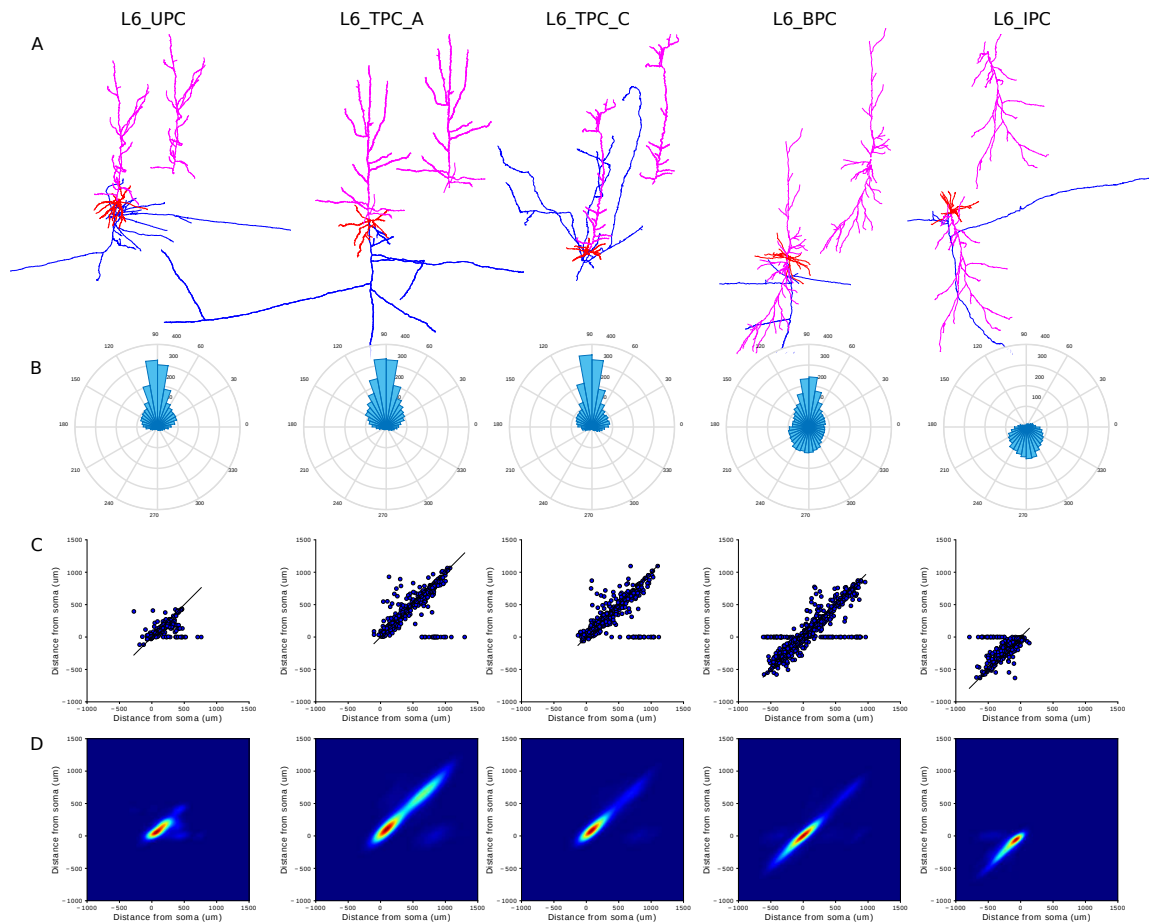
Fig. 2.17 Six PC subtypes in layer 6. A. Representative 3D reconstructions of PCs. B. Polar plot analysis of apical dendrites: The orientation of L6PC apical dendrites varies in different subtypes: vertically oriented towards pial for L6_TPC:A, L6_TPC:C and L6_UPC; vertically oriented in the opposite direction towards white matter for L6_IPC; vertically oriented towards both pial and white matter for L6_BPC; horizontally oriented for L6_HPC. C and D. Topological Morphology Descriptor (TMD) of L6 apical dendrites. L6_BPCs are identified by the two vertically projecting clusters of branching clusters that indicate the existence of two apical trees that project to opposite directions. L6_IPC apicals are directed towards white matter and have a low distal branching density that indicates the existence of a small tuft. L6_UPC apicals have high branching density proximal to the soma that indicates a rich oblique formation and an absence of a tuft. L6_TPC form a distinct and large tuft that can be separated into two sub-types similarly to layer 5; the L6_TPC:C apicals with a small tuft and the L6_TPC:A apicals with a dense tuft. The L6_HPC apicals have similar topological profiles to the L6_UPCs but have a horizontal preferred orientation which is unique among the layer 6 pyramidal cells.
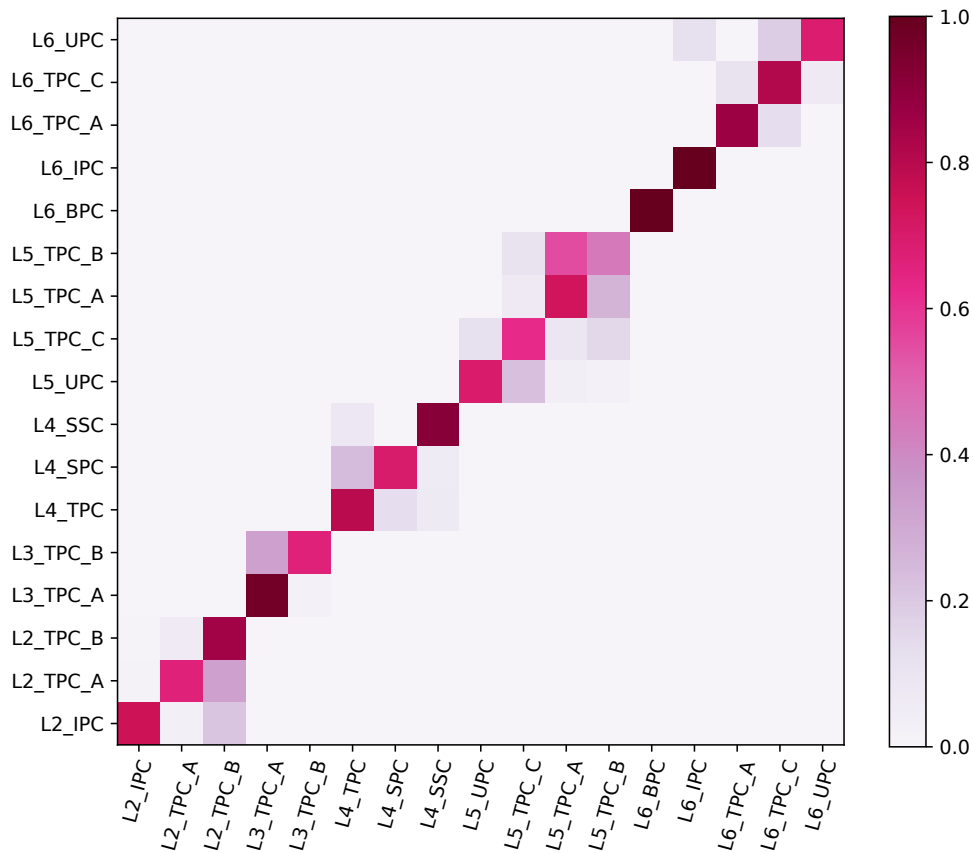
Fig. 2.18 Accuracy of TMD-based classification of cortical pyramidal cells. The colormap indicates the percentage of cells that are labeled as an m-type (ranging from 0.0: no cells, to 1.0 all cells). A perfect separation would result in 1.0 on the diagonal and 0.0 everywhere else. All expert identified classes are supported by the TMD-classification with the exception of L5_TPC_A and L5_TPC_B, which express a gradient of changes rather than a clear separation, and the L6_HPC that requires additional measurements to be identified -not shown here-. Note that all other classes are identified by a significantly high accuracy.

that is present in deeper layers, in agreement with recent observations (Reimann et al. 2017). The TMD-based classification was unable to distinguish few cell types proposed by experts that differ in other morphological characteristics, such as L6_HPC that are pronounced by horizontally oriented dendrites. In this particular case, an additional descriptor, the main orientation of the cell, was used for the objective discrimination of L6_HPC neurons. This demonstrates that expert classification is essential to guide further improvements of the method.

A new subfield of algebraic topology, which studies multidimensional persistence (Carlsson and Zomorodian, 2009), could be used for combining morphological measurements with independent parameters that are currently not considered, into multidimensional barcodes. Using this technique independent characteristics could be combined

into a single topological descriptor to strengthen even further the discriminative power of the TMD descriptor. The cells that differ on parameters that are not currently considered, such as the thickness of the processes and the bouton density, cannot be distinguished with the TMD descriptor. Another important property that is currently not considered, and should ideally be combined in an improved version of the TMD descriptor, is the long-range projections of trees.

This property is particularly important for axonal trees whose long-range projections indicate the brain regions with which they communicate. A significant number of studies suggest a strong association between the locally defined types of PCs with their target distant regions, which are genetically determined early on during differentiation and prior to the migration of the neurons to their destination layers (Larkman and Mason 1990; O'Leary and Koester 1993, Kasper et al. 1994; Franceschetti et al. 1998; Gao and Zheng 2004; Larsen and Callaway 2006; Morishima and Kawaguchi 2006; Kumar and Ohana 2008; Marx and Feldmeyer 2012). Indeed, long-range axonal projection of PCs is an important feature that enables different computational functions and should therefore be taken into account for their classification (Larsen and Callaway 2006; Larsen et al. 2007; Boudewijns et al. 2011).

Due to technical limitations, the long range projections of pyramidal cells are not currently available for a sufficiently large number of cells to allow for their systematic characterization. However, recent advances in optical imaging and long-range axon labeling techniques is enabling a systematic reconstruction of single neurons at the whole brain level (Yuan et al. 2015, Gong et al. 2016). Hopefully, these advances will enable the systematic characterization of whole cells reconstructions in order to quantify their long-range axonal projection properties and associate them to their local properties. Ideally, it will be possible to build a mouse or rat brain atlas at single neuron resolution, which is am essential step for biologically detailed simulations of neuronal microcircuitry, brain regions and the whole brain (Markram 2006).

Author contributions

This study was conceived by HM. Morphological reconstructions were performed by YS, YW. Quantitative morphological analysis and feature extraction was performed by YS. Objective classification and grouping of cells into classes was performed by LK. JS and KH contributed to the development of TMD classification. LK, YS and SR designed and created the figures. LK, YW, SR and HM wrote the manuscript. All authors commented on the manuscript at all stages.

Acknowledgements

References

Ascoli G.A., Wheeler D.W. In search of a periodic table of the neurons: Axonal-dendritic circuitry as the organizing principle: Patterns of axons and dendrites within distinct anatomical parcels provide the blueprint for circuit based neuronal classification. (2016). BioEssays: news and reviews in molecular, cellular and developmental biology. 38(10):969-976.

Boudewijns, Z. S., T. Kleele, et al. (2011). Semi-automated three-dimensional reconstructions of individual neurons reveal cell type-specific circuits in cortex. Commun Integr Biol 4(4): 486-8.

Ramón y Cajal, S.(1911). Histologie du Système Nerveux de l'Homme et des Vertèbres. Paris, Maloine.

Carlsson, G. (2009). Topology and data. Bulletin of American Mathematical Society, 46, 255–308.

Carlsson, G., & Zomorodian, A. (2009). The theory of multidimensional persistence. Discrete & Computational Geometry, 42(1), 71–93.

DeFelipe, J. and I. Farinas (1992). The pyramidal neuron of the cerebral cortex: morphological and chemical characteristics of the synaptic inputs. Prog Neurobiol 39(6): 563-607.

DeFelipe, J., Lopez-Cruz, P.L., Benavides-Piccione, R., et al. (2013). New insights into the classification and nomenclature of cortical GABAergic interneurons. Nature Reviews Neuroscience, 14(3), 202–216.

Franceschetti, S., G. Sancini, et al. (1998). Postnatal differentiation of firing properties and morphological characteristics in layer V pyramidal neurons of the sensorimotor cortex. Neuroscience 83(4): 1013-24.

Gao, W. J. and Z. H. Zheng (2004). Target-specific differences in somatodendritic morphology of layer V pyramidal neurons in rat motor cortex. J Comp Neurol 476(2): 174-85.

Gong, H., D. Xu, et al. (2016). High-throughput dual-colour precision imaging for brain-wide connectome with cytoarchitectonic landmarks at the cellular level. Nat Commun 7: 12142.

Kanari L., Dlotko P., Scolamiero M., Levi R., Shillcock J., Hess K., Markram H. (2017). A Topological Representation of Branching Neuronal Morphologies. Neuroinformatics.

Kasper, E. M., A. U. Larkman, et al. (1994). Pyramidal neurons in layer 5 of the rat visual cortex. I. Correlation among cell morphology, intrinsic electrophysiological properties, and axon targets. J Comp Neurol 339(4): 459-74.

Kumar, P. and O. Ohana (2008). Inter- and intralaminar subcircuits of excitatory and inhibitory neurons in layer 6a of the rat barrel cortex. J Neurophysiol 100(4): 1909-22.

Larkman, A. and A. Mason (1990). Correlations between morphology and electrophysiology of pyramidal neurons in slices of rat visual cortex. I. Establishment of cell classes. J Neurosci 10(5): 1407-14.

Larkum, M. E., J. J. Zhu, et al. (1999). A new cellular mechanism for coupling inputs arriving at different cortical layers. Nature 398(6725): 338-341.

Larkum, M. E., J. J. Zhu, et al. (2001). Dendritic mechanisms underlying the coupling of the dendritic with the axonal action potential initiation zone of adult rat layer 5 pyramidal neurons. 533: 447-466.

Larsen, D. D. and E. M. Callaway (2006). Development of layer-specific axonal arborizations in mouse primary somatosensory cortex. J Comp Neurol 494(3): 398-414.

Larsen, D. D., I. R. Wickersham, et al. (2007). Retrograde tracing with recombinant rabies virus reveals correlations between projection targets and dendritic architecture in layer 5 of mouse barrel cortex. Front Neural Circuits 1: 5.

Markram, H. (2006). The blue brain project. Nat Rev Neurosci 7(2): 153-60.

Markram, H., P. J. Helm, et al. (1995). Dendritic calcium transients evoked by single back-propagating action potentials in rat neocortical pyramidal neurons. J Physiol 485 (Pt 1): 1-20.

Markram, H., E. Muller, et al. (2015). Reconstruction and Simulation of Neocortical Microcircuitry. Cell 163(2): 456-92.

Marx, M. and D. Feldmeyer (2012). Morphology and physiology of excitatory neurons in layer 6b of the somatosensory rat barrel cortex. Cereb Cortex 23(12): 2803-17.

Morishima, M. and Y. Kawaguchi (2006). Recurrent connection patterns of corticostriatal pyramidal cells in frontal cortex. J Neurosci 26(16): 4394-405.

O'Leary, D. D. and S. E. Koester (1993). Development of projection neuron types, axon pathways, and patterned connections of the mammalian cortex. Neuron 10(6): 991-1006.

Oberlaender, M., Z. S. Boudewijns, et al. (2011). Three-dimensional axon morphologies of individual layer 5 neurons indicate cell type-specific intracortical pathways for whisker motion and touch. Proc Natl Acad Sci U S A 108(10): 4188-93.

Ramaswamy, S. and H. Markram (2015). Anatomy and physiology of the thick-tufted layer 5 pyramidal neuron. Front Cell Neurosci 9: 233.

Schaefer, A. T., M. E. Larkum, et al. (2003). Coincidence detection in pyramidal neurons is tuned by their dendritic branching pattern. J Neurophysiol 89(6): 3143-54.

Spruston, N. (2008). Neuroscience: strength in numbers. Nature 452(7186): 420-1.

Staiger, J. F., I. Flagmeyer, et al. (2004). Functional diversity of layer IV spiny neurons in rat somatosensory cortex: quantitative morphology of electrophysiologically characterized and biocytin labeled cells. Cereb Cortex 14(6): 690-701.

Stuart, G., J. Schiller, et al. (1997). Action potential initiation and propagation in rat neocortical pyramidal neurons. J Physiol 505 (Pt 3): 617-32.

Wang, Y., H. Markram, et al. (2006). Heterogeneity in the pyramidal network of the medial prefrontal cortex. Nat Neurosci 9(4): 534-42.

Yuan, J., H. Gong, et al. (2015). Visible rodent brain-wide networks at single-neuron resolution. Front Neuroanat 9: 70.

2.4 Comprehensive morpho-electrotonic analysis shows two distinct classes of L2 and L3 pyramidal neurons in human temporal cortex

Yair Deitcher, Guy Eyal, Lida Kanari, Matthijs B. Verhoog, Guy Antoine Atenekeng Kahou, Huibert D. Mansvelder, Christiaan P.J. De Kock, and Idan Segev

Abstract: There have been few quantitative characterizations of the morphological, biophysical and cable properties of neurons in the human neocortex. We employed feature-based statistical methods on a rare dataset of 60 $3D$ reconstructed pyramidal neurons from L2 and L3 in the human temporal cortex (HL2/L3 PCs) removed after brain surgery. Of these cells, 25 neurons were also characterized physiologically. Thirty-two morphological features were analyzed (e.g., dendritic surface area, $36.333 \pm 18.157 \mu m^2$; number of basal trees, 5.55 ± 1.47; dendritic diameter $0.76 \pm 0.28 \mu m$). Eighteen features showed a significant gradual increase with depth from the pia (e.g., dendritic length, soma radius). The other features showed weak or no correlation with depth (e.g., dendritic diameter). The basal dendritic terminals in HL2/L3 PCs are particularly elongated, enabling multiple nonlinear processing units in these dendrites. Unlike the morphological features, the active biophysical features (e.g., spike shapes and rates) and passive/cable features (e.g., somatic input resistance, $47.68 \pm 15.26 M\Omega$, membrane time constant, $12.03 \pm 1.79 msec$, average dendritic cable length, 0.99 ± 0.24) were depth-independent. A novel descriptor for apical dendritic topology yielded two distinct classes, termed hereby as "slim-tufted" and "profuse-tufted" HL2/L3 PCs; the latter class tends to fire at higher rates. Thus, our morpho-electrotonic analysis shows two distinct classes of HL2/L3 PCs.

Keywords: Dendritic cable properties, Electrical classification, Human pyramidal cells, Morphological classification, Mouse vs. human dendrites

Introduction

The temporal cortex is considered to have important cognitive functions (Mirz et al. 1999; Fortier et al. 2011). In humans, the temporal neocortex is especially thick ($2.773\mu m$) compared to monkeys ($2.300\mu m$) or rodents ($969\mu m$) (see Mohan et al. 2015, and see also DeFelipe, Alonso-Nanclares, and Arellano 2002). Layer 2 and Layer 3 (L2/L3)which receives input from Layer 4 and sends its output to Layer 5 and Layer 6, is considered to play a key role in intrinsic cortical computation (Callaway 2004; Douglas and Martin 2004; Feldmeyer 2012; Constantinople and Bruno 2013; Li et al. 2014). In the human temporal cortex, L2/L3 was found to be particularly thick (average of $949\mu m$, Mohan et al. 2015), which implies that it might endow the human neocortex with enhanced computational capabilities. The major building block of L2/L3 are the pyramidal cells, which in rodents consist of at least $70-80\%$ of its total number of cells (Nieuwenhuys 1994).

L2/L3 pyramidal neurons from the human temporal cortex (HL2/L3 PCs) possess several unique features. Recent fine-scale anatomical studies on these neurons have demonstrated that they are large, in terms of total dendritic length and number of dendritic branches (Mohan et al. 2015). The dendrites of these cells are decorated with a large number of dendritic spines ($25.000-30.000$ spines per neuron, Benavides-Piccione et al. 2013; DeFelipe, Alonso-Nanclares, and Arellano 2002; Defelipe 2011). Dendritic spines are the main targets for excitatory synapses; their large number per neuron implies that L2/3 pyramidal neurons in humans are part of a densely connected network. In addition, dendritic spines are key elements in memory and learning processes (Yuste 2010), suggesting that L2/L3 neurons and the networks that they form are endowed with enhanced memory capacity.

What are the biophysical characteristics of L2/L3 pyramidal neurons from the human temporal cortex? To answer this question, experiments on living human brain tissue are needed; however, this kind of tissue is only available in a few laboratories worldwide and is obtained after brain surgery. Thus, there is scant information about the biophysical properties of human neocortical neurons (Inda et al. 2006; Kohling and Avoli 2006; Szabadics et al. 2006; Molnár et al. 2008; Verhoog et al. 2013; Testa-silva et al. 2014; Tian et al. 2014; Varga et al. 2015). Recently, our team demonstrated that L2/L3 pyramidal neurons in the human temporal cortex have distinctive biophysical features (Eyal et al. 2016) including the fact that their specific membrane capacitance, C_m is $0.5\mu F/cm^2$ is half the conventional value ($1\mu F/cm2$). We showed that such C_m values have important implications for signal transfer and information processing at

both the neuron and the network level. However very little is known either about the anatomy or the physiology of human cortical neurons.

Here, we employed our recently developed feature-based characterization scheme to demonstrate that morphologically, many features of HL2/L3 PCs show a gradual depth dependency (such as a gradual increase with depth in the total dendritic surface area, the number of branches and horizontal range, Figure 2.19). Using a novel topological method, we found HL2/L3 PCs fall into two distinct classes we dub "slim-tufted" and "profuse-tufted" HL2/L3 PCs. These two morphological types are also distinctive in their I/F relationship. Moreover, human L2/L3 express a prominent "sag" in response to hyperpolarizing currents, possibly suggesting that these cells express hyperpolarization-activated cyclic nucleotide-gated (HCN) channels (Magee 1998). We also found that the mean cable length of human neurons is independent of depth, in contrast to the physical mean length. Interestingly, we found a subpopulation of deep neurons that have large diameters and large R_m values, both of which compensate for their long dendrites. We discuss the functional implications of our results and compare our results to those obtained for L2/L3 pyramidal neurons in rodents.

Materials and Methods

Electrical recordings of Human L2/L3 pyramidal cells (acute living slices)

All procedures on human tissue were performed with the approval of the Medical Ethical Committee (METc) of the VU University Medical Centre (VUmc), with written informed consent by patients involved to use brain tissue removed for the treatment of their disease for scientific research, and in accordance with Dutch license procedures and the declaration of Helsinki (VUmc METc approval "kenmerk 2012/362"). Slices of human temporal cortex were cut from neocortical tissue that had to be removed to enable the surgical treatment of deeper brain structures for epilepsy or tumors. In all patients (20–57 years of age), the resected neocortical tissue was located outside the epileptic focus or tumor, and displayed no structural/functional abnormalities in preoperative MRI investigations. After resection, the neocortical tissue was placed within $30s$ in ice-cold artificial cerebrospinal fluid (aCSF) slicing solution which contained in (mM): 110 choline chloride, $26NaHCO_3$, 10 D-glucose, 11.6 sodium ascorbate, $7MgCl_2$, 3.1 sodium pyruvate, $2.5KCl$, $1.25NaH_2PO_4$, and $0.5CaCl_2 - 300mOsm$, saturated with carbogen gas ($95\%O_2/5\%CO_2$) and transported to the neurophysiology laboratory, which is located 500 meters from the operating room. The transition time between resection of the tissue and the start of preparing slices

was less than 15min. Neocortical slices ($350-400\mu m$ thickness) were prepared in ice-cold slicing solution, and were then transferred to holding chambers filled with $aCSF$ containing (in mM): $126NaCl; 3KCl; 1NaH_2PO_4; 1MgSO_4; 2CaCl_2; 26NaHCO_3; 10$ glucose – $300mOsm$, bubbled with carbogen gas ($95\%O2/5\%CO2$). Here, slices were stored for 20 min at 34C, and for at least 30min at room temperature before recording. Whole-cell, patch clamp electrophysiology recordings were then made from human layer 2/3 pyramidal neurons as described previously (Verhoog et al. 2013; Testa-silva et al. 2014). In short, we used standard, uncoated borosilicate glass pipettes (glass thickness 0.64mm) with fire-polished tips ($4.0-6.0MW$ resistance) filled with intracellular solution containing (mM): 110 K-gluconate; 10 KCl; 10 HEPES; 10 K2Phosphocreatine; 4 ATP-Mg; 0.4 GTP, biocytin 5mg ml-1 (pH adjusted with KOH to 7.3; $280-290mOsm$). Recordings were made using a MultiClamp 700B amplifier (Axon Instruments, CA, USA), sampling at $10-50kHz$ and low-pass filtering at $3-30kHz$. Recordings were digitized with an Axon Digidata 1440A and acquired using pClamp software (Axon). Recording aCSF was the same solution as the aCSF in which slices were stored. Recording temperature was $32-35C$.

3D reconstructions of human and mouse L2/L3 pyramidal cells (acute living slices)

Sixty morphologies of human L2/L3 cells, residing at depths of $409-1192\mu m$ below the pia, and 14 morphologies of mouse L2/L3 cells from depths of $222-493\mu m$ were reconstructed in 3D using Neurolucida software (Microbrightfield, Williston, VT, USA), using a 100x oil objective (1.4 N.A.). Dendritic diameters were incorporated into the morphological reconstruction using Neurolucida (Microbrightfield, Williston, VT) by manually setting the diameter of the line segments during reconstruction of the biocytin-filled neurons. Reconstructions were performed with a 100x oil objective (N.A. 1.4) and standardized for both human and mouse reconstructions. The value for dendritic diameter was subsequently extracted for individual segments from the digital files. Dendritic segments are assumed to be truncated cones, with initial and end diameters. The surface area is the area of this dendritic cone, not including the end caps. Additional details regarding the reconstruction methods can be found in Mohan et al. 2015.

For a subset of neurons used in this study, the axons were also reconstructed (Mohan et al. 2015); these axons were not included in the analysis performed in the present study. In this work, dendritic spines were only considered when building cable models of human neurons (Figures 2.24 and 2.25). In these models, the spine membrane area

was incorporated globally using the F factor as in Rapp et al. 1992 (and for human neurons, as in Eyal et al. 2016). In this work, Eyal et al. computed the F factor to be 1.9, based on detailed data from human cingulate cortex and human temporal cortex (Benavides-Piccione et al. 2013). This data is based on two post-mortem samples from two human males (aged 40 and 85) in which the neurons and dendritic spines were reconstructed in $3D$ using high-resolution confocal microscopy. To the best of our knowledge this is the most accurate data about human spines in neocortical pyramidal cells available today. However, we do not yet know whether the density and the size of dendritic spines in human neurons are depth-dependent or if there is any difference in spine density between slim-tufted and profuse-tufted neurons. Thus, for the modeling part of this work, we used $F = 1.9$ for all of our neuron models.

Morphological features

The 32 morphological features used for the present study are listed in Table 2.2. These features are the natural ones to consider when characterizing dendritic morphologies. These 32 features were extracted for each of the sixty human and fourteen mouse L2/L3 pyramidal neurons in our database using both the Pneumatk and NeuroM packages in Python developed by the Blue Brain Project (Juan Palacios, Lida Kanari, Eleftherios Zisis, Mike Gevaert). NeuronM is available in https://github.com/BlueBrain/NeuroM.

Electrophysiology and extraction of biophysical features

Electrical features were extracted from voltage responses to long hyperpolarizing and depolarizing somatic current injections of various amplitudes (Table 2.3). Features 1-8 were extracted from the responses to 150% supra-threshold depolarizing current. Spikes were detected by a crossing of a voltage threshold ($0mV$). We defined the beginning of the spike by detecting the maximum of the second derivative in the rising phase of the spike. The end of the spike was defined as the minimum voltage following the spike. The following features were defined:

1. *Mean AP amplitude*: Mean amplitude of the set of spikes that occurred during the current step. The amplitude of a spike was defined as the difference between the voltage at the beginning and the peak of the spike.

2. *Mean AP half-width*: Mean half-width of the set of spikes that occurred during the current step. The half-width of a spike was defined as the amount of time from the first crossing (in the upward direction) of the half-height voltage value

to the second crossing (in the downward direction) of this value. The half-height voltage is the voltage at the beginning of the spike plus half the spike amplitude.

3. *Mean AP rise time*: The mean rise time of the set of spikes that occurred during the current step. The rise time was defined as the amount of time from the beginning to the peak of the spike.

4. *Mean AHP (after-hyperpolarization) depth*: The mean AHP depth of the set of spikes that occurred during the current step. The AHP depth (relative to rest) was defined as the difference between the voltage at the end of the spike and the resting membrane potential.

5. *First spike latency*: The amount of time from the current onset to the peak of the spike.

6. *Spike frequency*: The firing rate of the neuron during the current step.

7. *ISI-CV*: Coefficient of variation (standard deviation divided by the mean) of the distribution of ISIs (inter-spike interval).

8. *Mean ISI*: The mean of the distribution of ISIs.

9. *Threshold current for spike generation*: The minimal current that elicited a spike.

10. *Membrane time constant, τ_m*: This is estimated through exponential fit to the recovery of the voltage response following astep hyperpolarizing current (Figure 2.23B, inset). To capture the slowest (membrane) time constant, the exponential fit to the voltage trace was calculated after a delay of 10*msec* from the start of voltage recovery.The fitted time constant should be considered as the "effective membrane time constant" as an active Ih current might be involved in this estimate (Figure 2.23B).

11. *Input resistance, R_N*: estimated by the linear fit of the *I/V* curve (Figure 2.23b).

12. *Sag ratio*: defined as $100\frac{V_{ss}-V_{min}}{V_{rmp}-V_{min}}$ where V_{ss} is the voltage at steady state, V_{min} is the minimum value reached after the beginning of the current injection and V_{rmp} is the voltage at the resting membrane potential.

Software

Analysis was carried out by custom software programmed in Matlab. The electrotonic dendrograms in Figure 2.24B1,B2 and the scaled mouse neuron in Figure 2.26D were constructed using the TREES toolbox (Cuntz et al. 2010), and the compartmental modeling simulations of $3D$ reconstructed neurons shown in Figure 2.24C1,C2 were run using NEURON 7.4 (Carnevale and Hines 2006).

Data analysis and Statistics

PCA
Principal component analysis (PCA; Duda et al., 2001) was used to determine the prominent components of the variability in the data by calculating the eigenvectors of the covariance matrix.

Statistical tests
To calculate the correlation between features and depth we used the Pearson correlation coefficient. To correct for multiple correlations the Benjamini & Hochberg procedure (Benjamini and Hochberg 1995) was used to control for the false discovery rate (FDR) of a family of hypothesis tests with a false discovery rate of 0.05. To compare the apical and basal trees and compare between the biophysical features of the two classes (slim-tufted and profuse-tufted) we used the two-sample Kolmogorov-Smirnov test (K-S test), a nonparametric hypothesis test.

Topological Morphology Descriptor (TMD)
We generated the topological profile of a neuron from its branching structure (the detailed method is described in https://arxiv.org/abs/1603.08432). The algorithm takes the branch points and the termination points of a tree as input as well as their connectivity, and produces a set of intervals on the real line known as a persistence diagram (Carlsson 2009). Each interval is a pair of real numbers that encodes the "lifetime" of a single branch in the underlying structure; the first (y-axis) represents the distance from the soma to the starting point of the branch and the second (x-axis) represents the distance from the soma to the end point of the branch. The persistent images (Figure 2.22A,B) are the density plots generated from the persistence diagram, where the intensity of the color corresponds to the density of points in the persistence diagram.

Classification of HL2/L3 PCs

For the classification, we used an open-source tool developed in (Pedregosa et al. 2011). The persistent images (Figure 2.22a,b) were used as input to the classifier. We initially trained the classifier with the persistent images of the 48 apical trees that were clearly distinguishable. Then we assigned the remaining 8 cells to classes according to the assessment of the classifier (Figure 2.22c). We cross-validated the results of the classifier by a leave-one-out method (Evgeniou et al. 2004) based on a Decision Tree classifier. The accuracy of the classifier was measured by the percentage of correct assignments of the persistent images into classes. The accuracy of the classifier, based on the leave-one-out cross-validation, was 90%. To control for the performance of the classifier, we randomized the labels of the cells and repeated the previous experiment. The accuracy of the classifier for the randomized dataset was 50%.

Results

Morphological properties

Depth dependency

The dataset of sixty morphologies from the human neocortex used in this study is shown in Figure 2.19. All the neurons used for the morphological analysis in the present study were human L2/L3 pyramidal neurons from the medial temporal cortex (Brodmann area 21). The considerable variability in dendritic size and shape can be seen in the figure, with a clear increase in the length of the apical tree with increasing distance from the pia.

Table 2.1 summarizes the basic morphological features used for comparing apical to basal dendrites. As shown in Table 2.1, the total surface area (apical and basal) of L2/L3 human dendrites is about $35.000mm^2$ on average. The mean length of the non-terminal branches is much longer for the apical tree ($69.35 \pm 13.62mm$) as compared to the basal tree ($31.06 \pm 5.97mm$, p 0.0001, n = 60, K-S test). By contrast, some morphological features (e.g., the diameter) of the two trees are not significantly different (p = 0.63, n = 60, K-S test).

For a systematic analysis of our morphological database of HL2/L3 PCs, we used the set of 32 features extracted from each of the sixty neurons in our database (Table 2.2). We separated the features into two major groups composed of the features for the apical tree (#1 - #15) and the basal tree (features #16 - #31). We added a single feature related to the soma (feature #32); namely, the mean soma radius.

Interestingly, the majority of the morphological features exhibited a gradual depth-dependent change. Therefore, we sorted the features in descending order according to
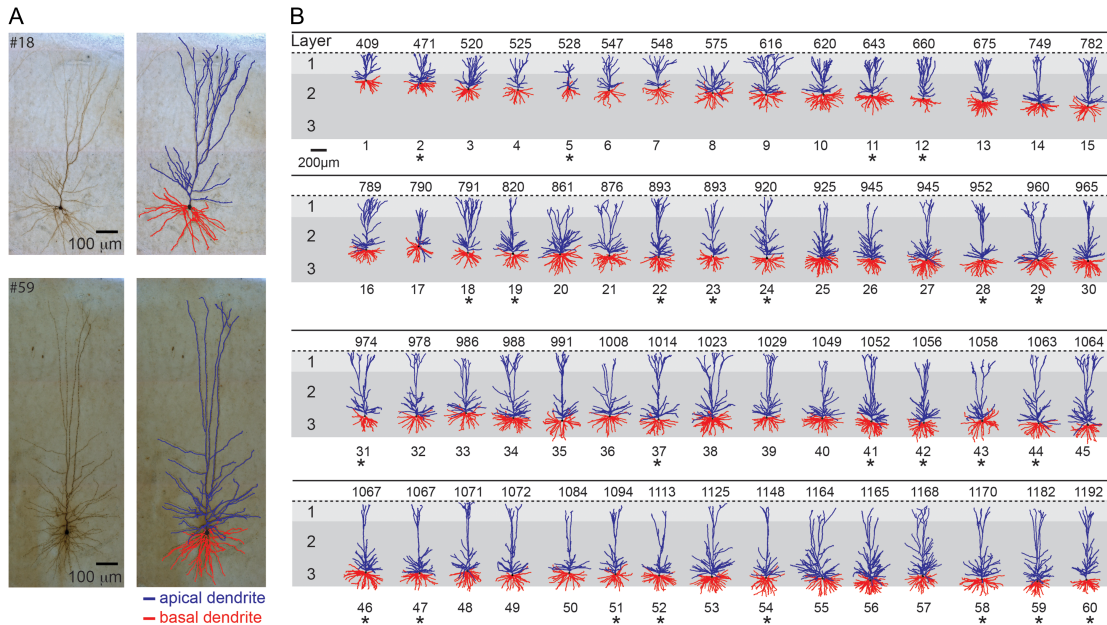
Fig. 2.19 Sixty 3*D* reconstructed Layers 2 and 3 pyramidal cells from the human temporal cortex arranged according to somatic depth with respect to pia surface. (A) Two exemplar cells (top, cell #18, bottom, cell #59; left biocytin filled; right 3*D* reconstruction). Apical dendrites are indicated in blue and basal dendrites in red. (B) Database used for the morphological analysis conducted in the present study. The numbers on top of each cell indicate the depth from the pia in *μm*. The 25 neurons that were also characterized physiologically are marked by an asterisk.

Table 2.1 Representative morphological features and corresponding values for the 60 reconstructed HL2/L3 PCs

Feature	Mean	SD
Features of the apical tree		
Total surface area (μm^2)	20, 125	10, 168
Mean diameter (μm)	0.78	0.29
Mean length of nonterminal branches (μm)	69.35	13.62
Mean length of terminal branches (μm)	162.31	21.97
Features of the basal tree		
Total surface area (μm^2)	20, 125	10, 168
Mean diameter (μm)	0.78	0.29
Mean length of nonterminal branches (μm)	69.35	13.62
Mean length of terminal branches (μm)	162.31	21.97

The complete list of the 32 morphological features that were used in this study are shown in Table 2.2.

their correlation with the cell's depth in the cortex (order of features as in Table 2.2). Note that we sorted the features separately for the apical tree and for the basal tree.

Table 2.2 The 32 features used for the morphological analysis of the 60 reconstructed HL2/L3 PCs

Feature number	Feature description
1	Maximal radial distance of apical tree from soma (μm)
2	Maximal path length of apical tree from soma (μm)
3	Vertical field span of apical tree (μm)
4	Total length of apical tree (μm)
5	Maximal branch order in apical tree
6	Number of branches in apical tree
7	Mean length of nonterminal apical branches (branches between 2 consecutive bifurcations) (μm)
8	Horizontal field span of apical tree (μm)
9	Total surface area of apical tree (μm^2)
10	Total volume of apical tree (μm^3)
11	Trunk diameter of apical tree (μm)
12	Mean length of terminal apical branches (branches between bifurcation and dendritic termination)(μm)
13	Mean diameter of apical tree (μm)
14	Ratio between the horizontal and vertical field span of apical tree
15	Density of apical tree-ratio between the volumes of the apical tree and of its convex hull
16	Total length of basal tree (μm)
17	Number of branches in basal tree
18	Maximal radial distance of basal tree from soma (μm)
19	Mean trunk diameter of basal tree (μm)
20	Total surface area of basal tree (μm^2)
21	Maximal path length of basal tree from soma (μm)
22	Horizontal field span of basal tree (μm)
23	Total volume of basal tree (μm^3)
24	Mean length of terminal basal branches (branches between final bifurcation and dendritic termination) (μm)
25	Maximal branch order in basal tree
26	Number of basal trees
27	Vertical field span of basal tree (μm)
28	Density of basal tree-ratio between the volumes of the basal tree and of its convex hull
29	Ratio between the horizontal and vertical field span of basal tree
30	Mean diameter of basal tree (μm)
31	Mean length of nonterminal basal branches (branches between 2 consecutive bifurcations) (μm)
32	Mean soma radius (μm)

Features related to the apical tree are numbered 1-15; features related to the basal tree are numbered 16-31 and Feature 32 is related to the soma. Radial distance (Feature #1, Feature #18) is the Euclidean distance from the soma to each section terminal. The total length (Feature #4, Feature #16) is the sum of all the section lengths of the neurite. The path length (Feature #2, Feature #21) is the length of the path from a terminal to the soma. The center of the soma is defined as the mean of all the soma points. The mean radius of the soma is defined as the mean distance of all the soma points from the center. For further documentation see https://github.com/BlueBrain/NeuroM.

Figure 2.20A depicts four representative features that show the gradual depth-dependent change. Since all cells reached the pia (Figure 2.19), it is not surprising that the maximal path length of the apical tree was strongly correlated with depth (Figure 2.20A1, feature #2, r = 0.95, p < 0.0005, n = 60). However, other features, perhaps unexpectedly, showed a significant positive correlation with cortical depth, such as the horizontal field span of the apical tree (Figure 2.20A2, feature #8, r = 0.48, p < 0.0005, n = 60), the total length of the basal tree (Figure 2.20A3, feature #16, r = 0.50, p < 0.0005, n = 60) and the mean soma radius (Figure 2.20A4, feature #32, r = 0.35, p < 0.01, n = 60).

Figure 2.20B depicts the correlation coefficients between the feature values and the depths of the cells. Features with a significant correlation coefficient (p < 0.05) are marked with an asterisk. Notably, most of the features displayed a positive correlation with depth. In fact, out of the 32 features used, 18 showed a significant positive correlation with depth and nearly all of the remaining features showed a weak positive correlation. Note that at chance level only 2 features (5% of 32) would have a p-value of less than 0.05. After correcting for multiple correlations (see Methods), 17 features showed a significant positive correlation with depth.

Next, we performed principal component analysis (PCA, see Methods) using all 32 features (Table 2.2) to test the correlation between the linear combination of the features and depth. Figure 2.21A shows the first principal component values for all cells as a function of depth. Note the high correlation between the first principal component score and depth (r = 0.59, p < 0.0001, n = 60), indicating that depth dependency is a major component accounting for the variability in the data. Figure 2.21B shows the fraction of variability represented by each of the first 10 principal components. Over 30% of variance was captured by the first principal component and about 90% of the variance was captured by the first 10 principal components.

The weight of the contribution of each of the 32 features to the first principal component is shown in Figure 2.21C. Features related to the apical dendrite are shown in blue and features related to the basal dendrites are in red; the soma feature is in green (Feature #32). Interestingly, the distribution is broad, with various features carrying substantial weight, both for those related to the apical dendrite and to the basal dendrites (e.g. the total surface area of apical tree, feature #9 and the total surface area of basal tree, feature #20). Together, these results demonstrate that many morphological features contribute strongly to the main source of variability in the data; namely, the depth-dependent change in morphological properties.
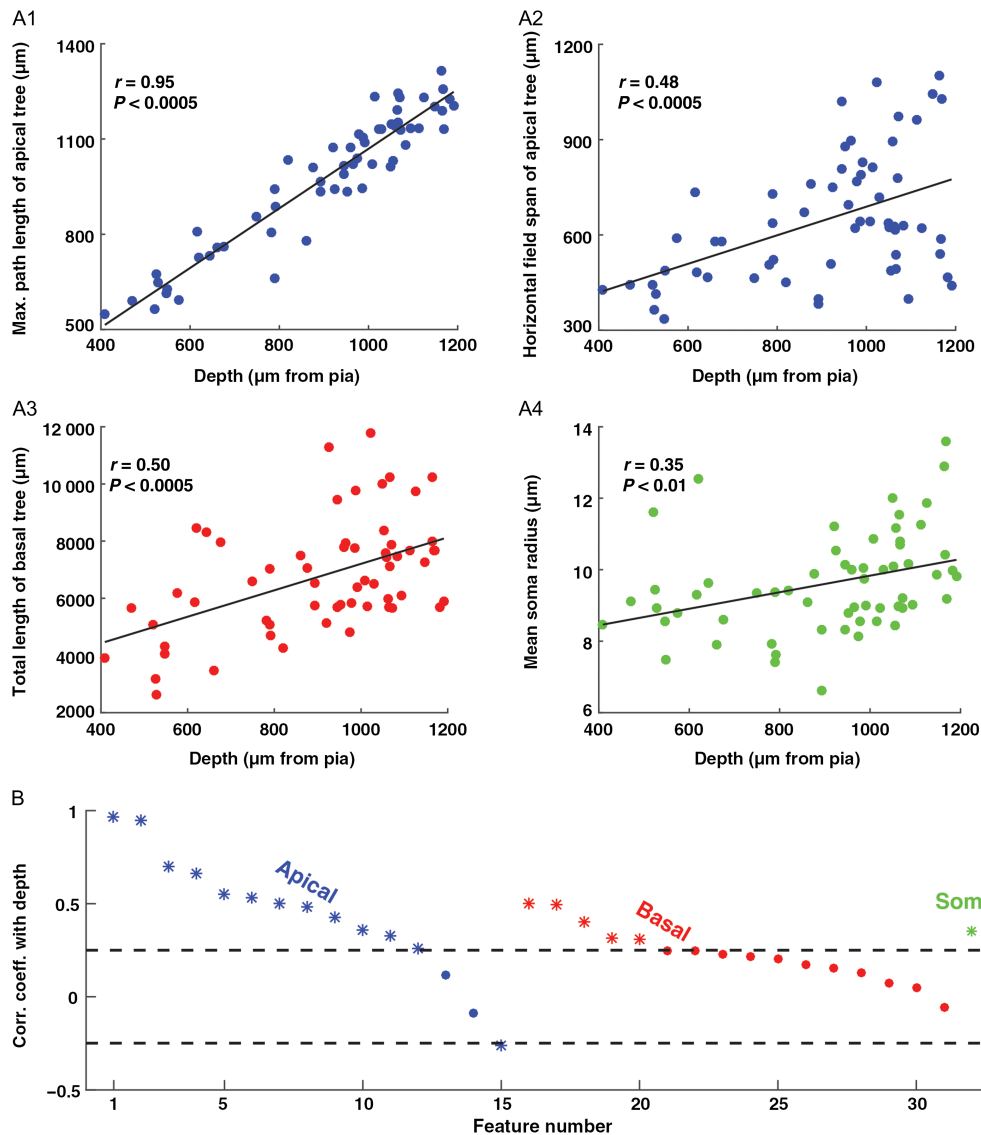
Fig. 2.20 Many of the morphological features of HL2/L3 PCs display a gradual depth-dependent change. (A1–A4) Four representative features that showed a gradual depth-dependent change. The continuous line is the linear regression fit. (B) Correlation coefficient for all 32 features with depth. The features are separated into 3 groups: apical tree features (#1 to #15, blue), basal tree features (#16–#31, red), and a single somatic feature (#32, green); see Table 2.2 for feature definitions. In each group, the features are sorted in descending order according to the correlation coefficient value with depth. Features with a significant correlation coefficient ($P < 0.05$) are marked with an asterisk (outside the dashed lines).
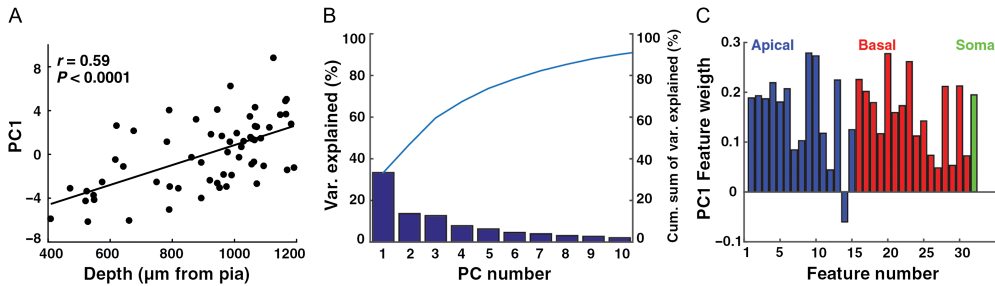
Fig. 2.21 PCA of HL2/L3 PCs morphological features. (A) The first principal component score (PC1, based on 32 morphological features) as a function of the depth from pia. (B) Percentage of variance captured by each of the first ten principal components. The line displays the cumulative sum of the variance captured. (C) The weight of the contribution of each of the 32 features to the first principal component. Apical dendrite features are marked in blue, basal dendrite features in red and the somatic feature in green.

Morphological classification

From work in rodents it is known that neocortical layers can contain multiple sub-types of pyramidal neurons (Hallman et al. 1988; Mason and Larkman 1990). In order to examine the possibility that there were distinct morphological classes in HL2/L3 PCs we employed the Topological Morphology Descriptor (TMD) method for encoding the spatial structure of a branching morphology (see Methods and https://arxiv.org/abs/1603.08432). On the first run of this method we found 4 cells that clearly did not have a tuft that could be distinguished from the oblique branches. Because of the limited group size of these cells, they were excluded from the analysis. Next, the topological analysis of the apical trees of HL2/L3 PCs revealed the existence of two distinct morphological classes which we dubbed the "slim-tufted" (Figure 2.22A) and the "profuse-tufted" (Figure 2.22B) pyramidal cells.

The slim-tufted neurons are characterized by the low density of the tuft branches whereas the profuse-tufted have a significantly higher density of tuft branches. Both types display a high density of oblique branches proximal to the soma. Using the TMD method it was straightforward to distinguish between 24 cells in each class (48 cells altogether); eight additional cells were not obviously belonging to either of these classes. By training a classifier (Decision Tree in Methods) on the two obvious classes, the classifier did label each of the unresolved cells (n = 8) into one of these two classes (Figure 2.22C). The final result thus yielded two distinct groups of 27 slim-tufted, 29 profuse-tufted and n = 4 neurons that did not have a clear tuft (total n = 60). Generally, the profuse-tufted neurons tended to be located more superficially than the slim-tufted neurons. However, the classes were not completely separated by depth (Supplementary Figure 2.19A).
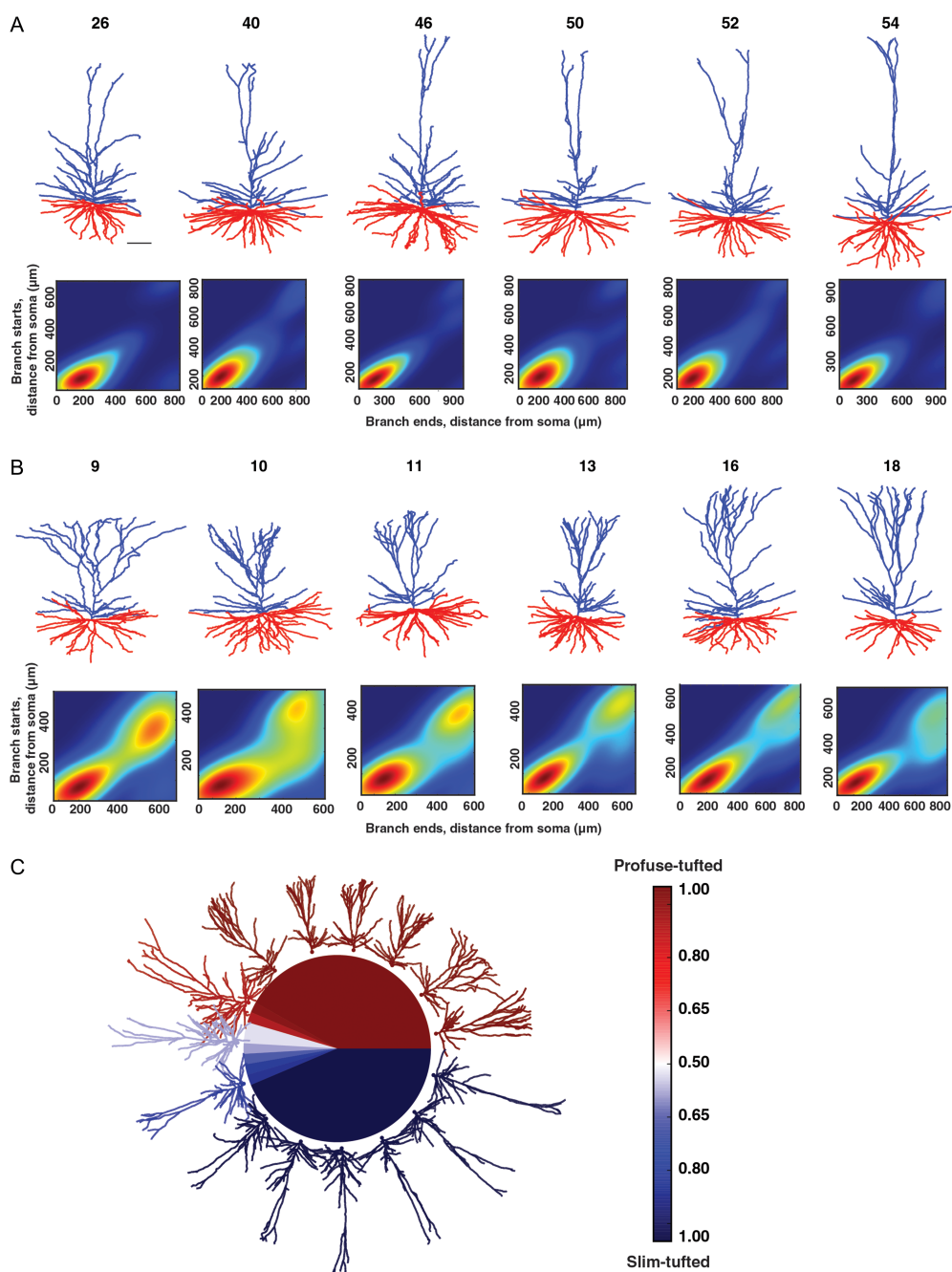
Fig. 2.22 Examples of the 2 classes ("slim-tufted" and "profuse-tufted") in HL2/L3 PCs. (A) Top, six 3D reconstructed exemplars of a slim-tufted HL2/L3 cell, cell numbers are as in Figure 2.19. Scale bar is 100μm. Bottom, density plot of the cells on top (see "Materials and Methods" section). (B) Top, six 3D reconstructed exemplars of the profuse-tufted HL2/L3 cell, cell numbers are as in Figure 2.19. Bottom, density plot of the cells on top. (C) Illustration of the separation of the 2 classes. The colors represent the probability of each neuron to be classified as a certain type: dark blue indicates slim-tufted; dark red indicates profuse-tufted.

To assess the performance of the classifier we cross-validated the proposed grouping using the leave-one-out method (Evgeniou et al. 2004). The proposed grouping was shown to be stable with respect to the Decision Tree classifier with an estimated accuracy of 90%, as opposed to a randomization of the groups, which resulted in 50%accuracy (see Methods).

Finally, we examined if the slim-tufted and the profuse-tufted classes exist in L2/L3 PCs from mouse temporal cortex. Based on our limited data (n = 14) these two classes of neurons could not be found in mouse, suggesting that these two types are unique types in the L2/L3 of the human cortex. Note that both groups of neurons (mouse and human) were sampled across the full L2/L3 range, came from adult subjects and the same methodology was used for both samples, including biocytin filling, histological processing, and reconstruction methods (Mohan et al. 2015).

Biophysical properties

Depth independent

We next examined the biophysical features of the human L2/L3 pyramidal neurons. In this analysis, we used our database of electrical recordings from 25 human neurons, which were included in the morphological analysis above (cells marked with asterisks in Figure 2.19).

To characterize the biophysical characteristics of HL2/L3 PCs, we analyzed the responses of these neurons to hyperpolarizing and depolarizing somatic current injections (Figure 2.23). Typically, a brief high frequency burst of spikes appears at the start of the supra-threshold current pulse; the following spikes appear to be highly regular (Figure 2.23A). A zoom into an individual spike is shown on the right of Figure 2.23a. Figure 2.23B depicts the voltage responses to hyperpolarizing step currents. Inset shows the exponential fit (green dashed line) to the voltage response (black line) following the termination of a current step. For each neuron, the membrane time constant, τ_m (feature #10) was estimated from an exponential fit to the recovery of the voltage response following a step hyperpolarizing current (Figure 2.23B inset and see Methods).

Table 2.3 summarizes the 14 biophysical features used in this study (feature #14, the mean cable length, is analyzed separately in the cable analysis bellow). Note that the comparison between the different spike features was made at 150% threshold current. However, the conclusions drawn below are also valid for 175% threshold current (not shown). The values for these biophysical features in human L2/L3 pyramidal cells (e.g. the mean AP half-width and AP amplitude) are well within the range of
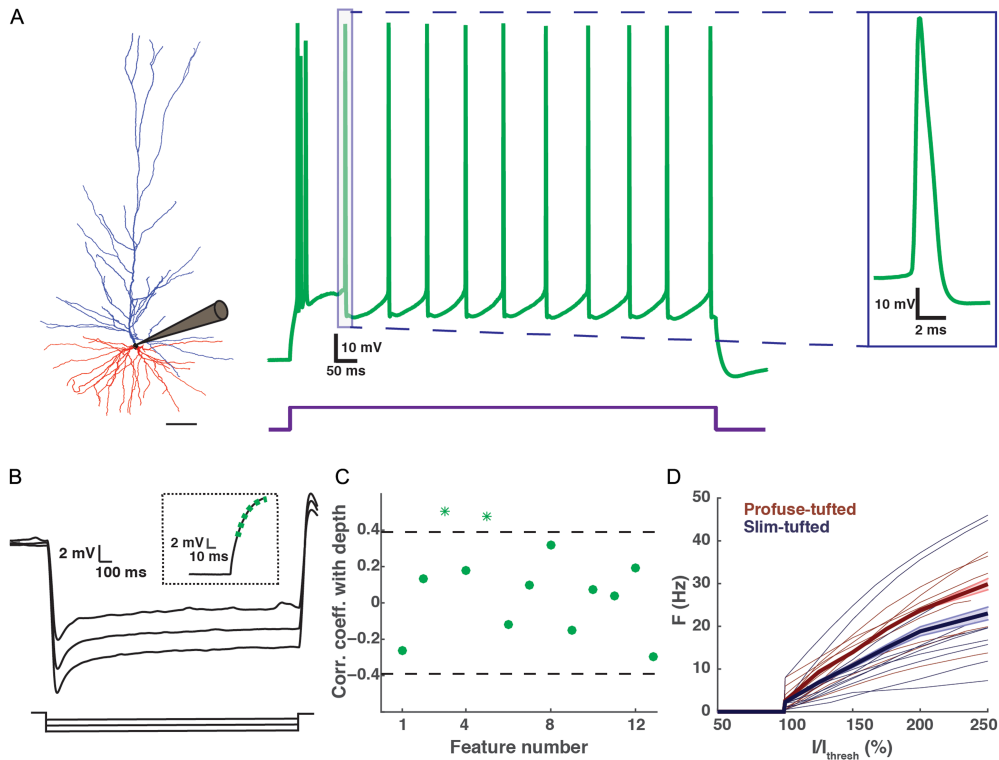
Fig. 2.23 Biophysical features of HL2/L3 PCs. (A) Left, a $3D$ reconstructed exemplar HL2/L3 cell ($1063\mu m$ from pia, cell #44 in Fig. 1). Scale bar is $100\mu m$. Middle, voltage response to a depolarizing step current for the cell at left. The current step ($450pA$) was 150% larger than the threshold current. Right, zoom-in into an individual spike (demarcated by box in middle trace). (B) Voltage responses to hyperpolarizing step currents of -250, -200, -150 pA for cell #59 shown in Figure 2.19. Inset shows the exponential fit (green dashed line) to the voltage response (black line) following the termination of the -250 pA current step. This fit was used to estimate the membrane time constant, τ_m (see "Materials and Methods" section). (C) Correlation coefficient with depth for 13 biophysical features (Table 2.3). Features with a significant correlation coefficient ($P < 0.05$) are marked with an asterisk (outside of the dashed lines). Most biophysical features are depth-independent. (D) Normalized I/F curves for the profuse-tufted (dark red) and slim-tufted (dark blue) pyramidal cells. The thicker lines are the mean of each class respectively; the shaded area is the respective standard error. Note that the profuse class tends to fire at higher rates. Normalization is to the minimal current (I_{thresh}) that generated a spike.

Table 2.3 The 14 features used for the biophysical and cable analysis with their corresponding values

Feature number	Feature description	Mean	SD
1	Mean AP amplitude at 150% threshold current (*mV*)	85.09	5.24
2	Mean AP half-width at 150% threshold current (*ms*)	1.26	0.42
3	Mean AP rise time at 150% threshold current (*ms*)	0.76	0.15
4	Mean AHP depth relative to rest at 150% threshold current (*mV*)	15.39	3.63
5	First spike latency at 150% threshold current (*ms*)	28.85	12.29
6	Spike frequency at 150% threshold current (*Hz*)	12.72	5.62
7	Mean ISI-CV at 150% threshold current	0.37	0.22
8	Mean IS at 150% threshold current (*ms*)	94.64	48.67
9	Threshold current (I_{thresh}) for spike generation (*pA*)	267.20	118.52
10	Membrane time constant, τ_m (*ms*)	12.03	1.79
11	Input resistance, R_N (*mΩ*)	47.68	15.26
12	Sag ratio (%)	16.60	8.16
13	Resting membrane potential (*mV*)	−85.1	3.19
14	Mean cable length, *L*	0.99	0.24

Neurons analyzed are marked in asterisks in Figure1. Features related to AP shape are numbered 1-4; features related to AP firing are numbered 5-9. Feature #14 is analyzed separately in the cable analysis section. See "Materials and Methods" section for further details.

L2/L3 pyramidal neurons in rodents (Staiger et al. 2014). However, one interesting feature worth noting from Table 2.3 is the appearance of a sag in the voltage response to long hyperpolarizing current injections (feature #12). This contrasts with L2/L3 rodent pyramidal neurons in the somatosensory cortex, which show only very small sag (Larkum et al. 2007; but see however Van Aerde and Feldmeyer 2015 who found a subpopulation of L3 pyramidal cells that do display larger sag of 12%). In human L2/L3 pyramidal neurons the sag is prominent, similar to that found in L5 rat pyramidal neurons (Zhu 2000; Larkum et al. 2007; Van Aerde and Feldmeyer 2015). These results suggest that HCN channels might be present in L2/L3 human pyramidal neurons (see Discussion).

In the morphological analysis above we found that many features showed a gradual depth-dependent change. We therefore tested whether this was also the case for the biophysical features. Figure 2.23C presents the correlation coefficient for the feature value and the depth of the cells. In contrast to the morphological features, there was no correlation with depth for most of the biophysical features; only feature #3

(mean AP rise time) and #5 (first spike latency) showed a slight correlation with depth. One unexpected finding was that the input resistance (RN, feature #11) was not correlated with depth. One expects that neurons with smaller surface areas (more superficial neurons) should have larger R_N. However, we found that neurons with smaller surface areas tend to have smaller τ_m (namely, a smaller specific membrane resistivity, R_m, and consequently a smaller R_N than expected if R_m were constant for all cells, Supplementary Figure 2.20A). Other correlations between various biophysical features are depicted in Supplementary Figure 2.20B,C.

Biophysical classification

Next, we compared the biophysical features of the slim-tufted and the profuse-tufted neurons found in the morphological analysis. Figure 2.23D shows the normalized I/F curves of the 24 HL2/L3 PCs profuse-tufted (dark red curves) and slim-tufted (dark blue curves) neurons. The 25th neuron that was measured physiologically did not have a tuft. The threshold current differs among these cells; however, when normalized by the threshold current (I_{thresh}), the average I/F curves clearly distinguish between these two classes (thick dark blue and dark red lines). The profuse-tufted class tends to fire at higher rates than do the neurons belonging to the slim-tufted class. Indeed, at I/I_{thresh} of 125%, 150% and 175%, the firing rate was statistically different between the two classes (K-S test, $p < 0.05$; the number of cells with I/I_{thresh} of 200% was too small for a reliable statistical test). The result of Figure 2.23D suggests that the two morphological classes that were found in this work are also two separate biophysical classes in term of their I/F curves. This assertion should be further validated on a larger data set when it becomes available. Note, however, that other biophysical features (shown in Table 2.3) do not show significant difference between the slim-tufted and the profuse-tufted neurons (K-S test, $p > 0.05$).

Cable length

The cable properties of dendrites determine the integrative properties of the neurons (Rall 1959). The cable properties of neurons are affected by both their morphology and biophysical properties. In this section, we focus on a key cable parameter of HL2/L3 PCs dendrites; namely L, their electrotonic length. L is defined as, $L = x/\lambda$ (in dimensionless units, Rall, 1959), where x is the physical length of the dendritic branch and λ is its space constant $\lambda = \sqrt{dR_m/4R_i}$ where d is the diameter of the dendritic branch, R_m is the specific membrane resistance (in Ωcm^2) and R_i is the specific axial resistivity (in Ωcm).

We define the mean cable length of a neuron as the mean cable length of all (apical and basal) paths from the soma to the dendritic terminals. In our calculation, the axial resistivity, R_i was assumed to be $200\Omega cm$ and R_m was estimated from the experimental membrane time constant, τ_m (Figure 2.23B, inset) using the specific membrane capacitance, $C_m = 0.5\mu F/cm^2$ of human cortical neurons reported before (Eyal et al. 2016).

As in the previous sections we examined hereby the dependence of L on the depth from pia as well as whether it differs between the slim-tufted and the profuse-tufted neurons. We did not find significant differences in the value of L between these two classes (K-S test, p > 0.05) and thus proceeded to explore below its depth-dependency.

Figure 2.24A1 and 6A2 depict the physical dendrograms of a superficial and a deep neuron, respectively. As expected, the deeper neuron was longer in physical units. Surprisingly, when comparing the electrotonic length of both neurons, the physically longer deeper neuron was, on average, shorter in cable units (Figure 2.24B1, 2.24B2, $0.79x/\lambda$ vs. $0.63x/\lambda$ for the superficial and deep neurons, respectively). Moreover, for the cells shown in Figure 2.24, the steady-state voltage attenuation factor from dendrites to soma was larger, on average, for the superficial versus the deep neuron; 54.64 vs. 20.20 respectively (Figure 2.24C1, 2.24C2, respectively). This finding suggests that the electrical compactness (and integration properties) of small and large HL2/L3 PCs is comparable, despite the considerable difference in the physical length of their dendrites (see Discussion).

Figure 2.25A shows the mean physical length of the dendritic tree for the 25 HL2/L3 PCs (marked in asterisks in Figure 2.19) as a function of its depth from the pia. As expected, there was a high correlation between physical length and depth ($n = 25$, $r = 0.62$, $p < 0.001$). However, the mean cable length, L, of these neurons was not correlated with depth (Figure 2.25B, $n = 25$, $r = 0.08$, $p > 0.5$). Two parameters could account for this result: the dendritic diameter might be larger for deep cells, and/ or R_m might be larger for these cells. Figure 2.25C clearly shows that some of the deeper neurons (red circles) that were electronically short indeed had larger mean diameters. For these neurons, their larger diameter partially compensated for their longer dendrites, resulting in relatively small L values. Figure 2.25D demonstrates that several deep neurons also had large R_m values and, consequently, had relatively short electronic lengths. Note that the relative impact of the larger R_m for the deep neurons in scaling their respective L values was smaller than that of the mean diameter (compare the range of the color scale in Figure 2.25C vs. 2.25D). Examination of the top right-most group of cells in Figure 2.25B,C,D shows that there is a population
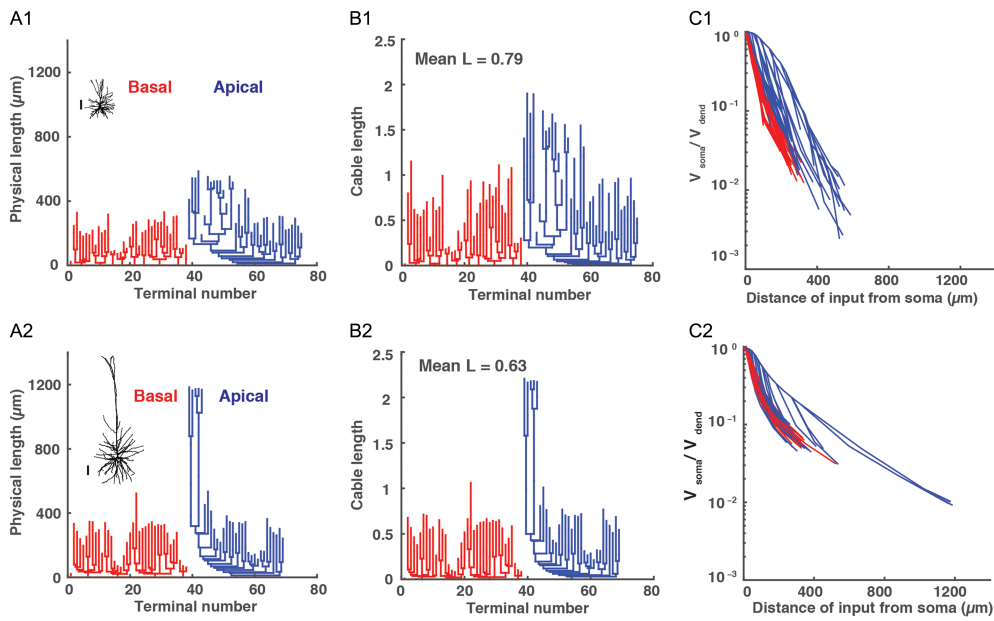
Fig. 2.24 Physical and electrotonic dendritic structure of superficial versus deep HL2/L3 PCs. (A1 and A2) Morphology and dendrogram, in physical units, of HL2/L3 superficial ($471\mu m$ from pia, cell #2 in Fig. 1) and deep ($1148\mu m$ from pia, cell #54 in Fig. 1) exemplar neurons, respectively. Red denotes basal tree and blue the apical tree. Scale bar is $100\mu m$ for both neurons. (B1 and B2) Dendrogram, in cable units, for the neurons in A1 and A2, respectively. These dendograms were calculated assuming $R_i = 200\Omega cm$, $C_m = 0.5\mu F/cm^2$ whereas τm was extracted for each cell from its experimental transients (Fig. 5B, inset). (C1 and C2). Steady voltage attenuation factor from the dendrites (V_{dend}) to the soma (V_{soma}) for the neurons in A1 and A2, respectively.
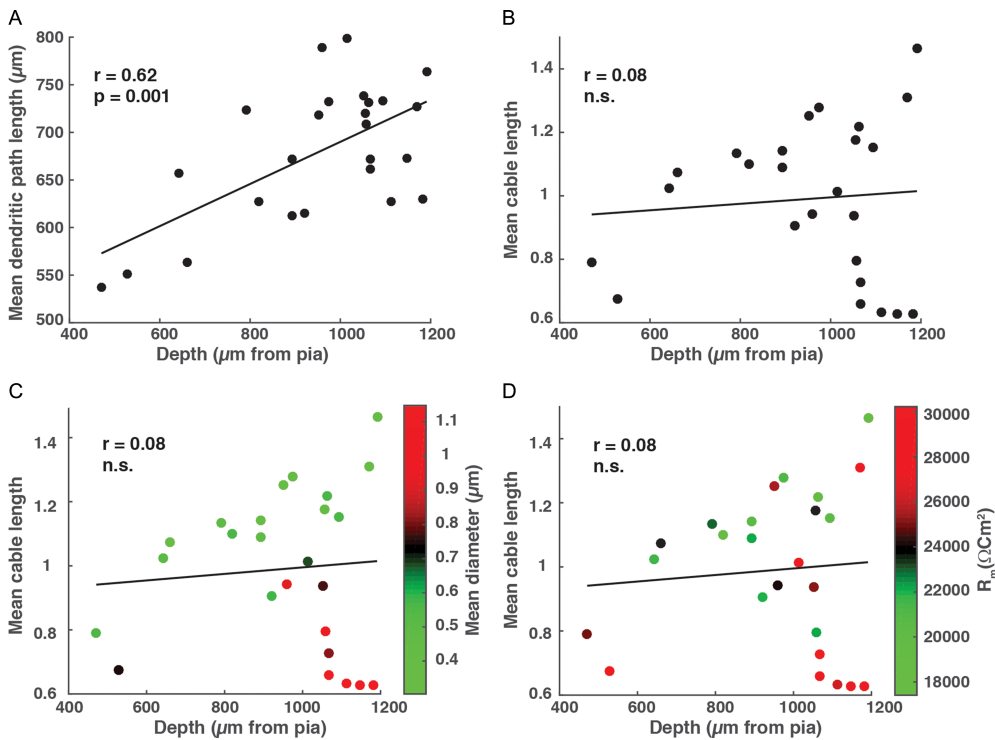
Fig. 2.25 Physical and electronic length of HL2/L3 PCs. (A) The mean physical length of all dendritic paths as a function of depth. (B) The mean cable length of all paths as a function of depth; this length was calculated assuming $R_i = 200\Omega cm$, $C_m = 0.5\mu F/cm^2$ whereas τ_m was extracted for each cell from its experimental transients (Fig. 5B, inset). (C) Same as in B, but neurons are colored according to their mean diameter. Note the group of deep cells with large diameters (red circles). (D) Same as in B but neurons are colored according to their respective R_m values. Note the deep cells with large R_m (red circles).

of deep neurons with particularly large average L values that resulted from small corresponding dendritic diameters and relatively small R_m values.

On human versus mouse L2/L3 pyramidal cells from temporal cortex

We utilized the same 32 features as in Table 2.1 to compare human and mouse L2/L3 pyramidal neurons in the temporal cortex. As expected, the dendritic tree of the HL2/L3 PCs was significantly larger than that of the mouse. However, the dendritic diameter was similar in mouse and human cells (Supplementary Figure 2.21A). Unlike in L2/L3 in the human temporal cortex (as well as in L2/L3 pyramidal neurons in rat barrel cortex, Staiger et al. 2014), mouse L2/L3 neurons from the temporal cortex do not show a significant gradual change with depth in most of their morphological features (Supplementary Figure 2.21B). We further applied Principal Component

Analysis and found a clear separation between human and mouse neurons based on all 32 morphological features. As shown in Figure 2.26A, the first principal component, separated both groups well. Over 40% of the variance was captured by the first principal component and many features carried substantial weight in the first principal component (Figure 2.26B). Note that a complete separation between the mouse and human neurons was found when using a spectral k-means algorithm (Uw et al. 2001) (not shown).

We also examined whether human neurons were morphologically just a "scaled" version of mouse neurons; this does not seem to be the case (Figure 2.26D). First, HL2/L3 PCs are not just longer, but they have significantly more branches in both the apical and basal trees (Mohan et al. 2015). Moreover, interestingly, we found that human terminal branches were particularly long compared to the terminal branches in mouse. Yet, the non-terminal branches were similar in human and mouse cells (Figure 2.26C). From this result, it is clear that human L2/L3 neurons are not a simple linear stretch of mouse L2/L3 neurons. Further analyses should be conducted to determine the structural rule for "transforming" HL2/L3 PCs into the respective mouse neurons. It is important to note that this comparison was based on the morphological properties of the neurons. As demonstrated above in the cable analysis, from a functional viewpoint, the morphological properties only provide a partial perspective thus making it crucial to compare the biophysical and cable properties of both groups. In any case, many elongated, thin basal dendrites as found here in human L2/L3 pyramidal neurons constitute independent computational "subunits", and have been argued to enhance the computation repertoire of the neuron (Poirazi and Mel 2001; Polsky et al. 2004).

Discussion

Using several cluster analysis methods, as well as cable theory, we analyzed a large dataset of human L2/L3 pyramidal cells from the temporal cortex. This yielded a systematic description of the morphological, biophysical and cable properties of HL2/L3 PCs. Since these data were taken from tissue removed after brain surgery (treatment of deep tumors and/or epileptic seizures) there is ample justification in inquiring the extent to which these cells were healthy. First, the neocortical tissue we study is always well away from the epileptic focus or tumour, so never part of the disease. Microglia and other inflammatory markers are at normal levels in this tissue, and the cytoarchitecture of the tissue is normal. Pathologists of our hospital label this tissue as "normal tissue". Moreover, we compare parameters in patient groups with different
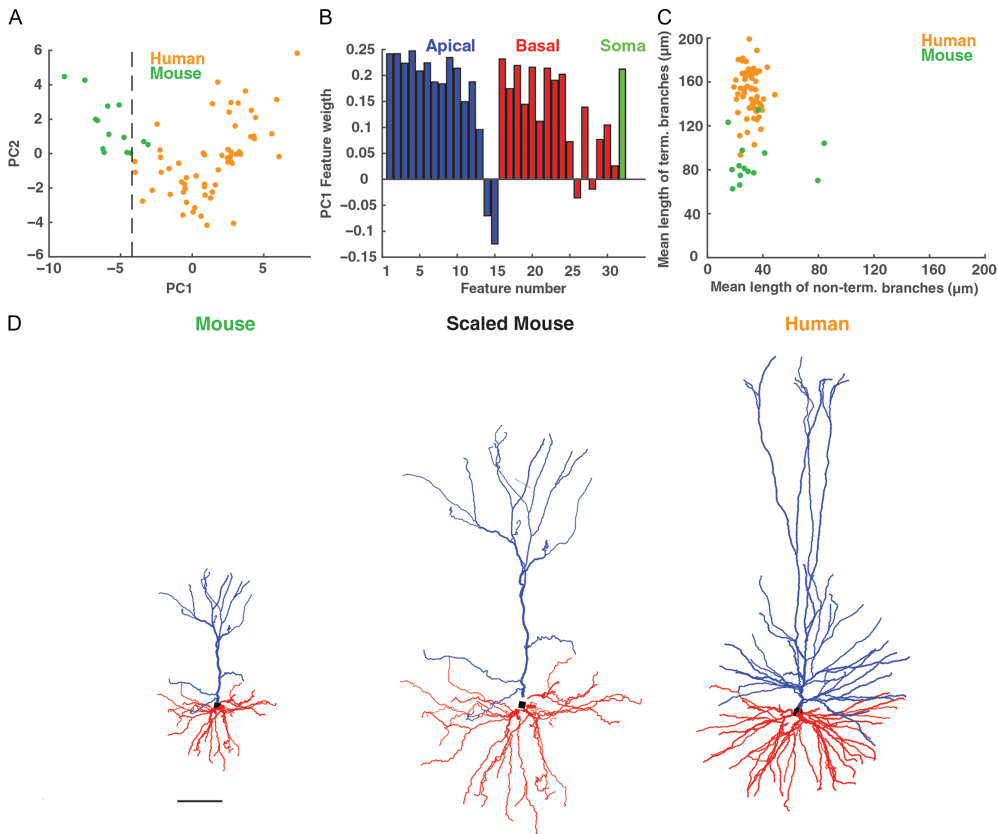
Fig. 2.26 Human versus mouse L2/L3 pyramidal cells from the temporal cortex. (A) PCA of morphological features of human (n = 60, orange) as well as mouse (n = 14, green) cells; the first principal component (PC1) and the second principal component (PC2) are based on the 32 morphological features for human and mouse cells combined (Table 2.2). Black dashed line separates the 2 cell populations well. (B) The weight of the contribution of each of the 32 features to the first principal component. Apical dendrite features are marked in blue, basal dendrite features in red and the somatic feature in green. (C) Plot of the nonterminal branches versus the terminal branches of the basal tree for human (orange) and mouse (green) neurons. Note the distinctly elongated terminal branches in human neurons. (D) Visual comparison of original exemplar mouse neuron (left), whose dendritic branches are stretched by a factor of 2 (middle), and an original human (right) L2/L3 exemplar pyramidal neuron (cell #41 in Fig. 1). Scale bar 100 μm.

disease backgrounds (sclerosis, glioma, meningitis, cavernoma). These groups receive very different medications (see for instance Table 2.1 in Mohan et al. 2015). When the parameter overlaps between different groups, we conclude that it generalizes across disease backgrounds. Finally, if the epilepsy has a role in the parameter, one would expect that the severity of the disease would play a role. Therefore, we quantify the potential influence of the disease history on the parameter by correlating the parameter against the number of years of epilepsy of the patient and the seizure frequency. These numbers widely vary between patients, from 1 to 50 years of epilepsy and from 1 to 400 seizures per month (see Figure 4 in Mohan et al., 2015). When there is no correlation between disease severity and the parameter, we conclude that the epilepsy has little to do with the parameter value. Further details are presented in Mohan et al. 2015.

In this study, we found that throughout the particularly thick L2/L3 ($949 \pm 179 \mu m$) in the already thick human temporal cortex ($2.773 \mu m$), cell bodies that are close to the pia have small apical trees and overall, about half of the dendritic length and surface area as compared to the deep neurons; these measurements were obtained for the five most superficial neurons and the five deepest neurons in Figure 2.19. When averaging the dendritic length of the entire 60 HL2/L3 PCs cells in our dataset, the average length was $14.793 \mu m$; hence, almost three times larger than the L2/L3 pyramidal cells from the rodent temporal cortex, and actually closer to the value found in L5 rodent pyramidal cells from the somatosensory cortex (an average of $12.758 \mu m$) (Hay et al. 2013). Despite the considerable differences between superficial and deep HL2/L3 PCs, our key morphological finding is that most of the 32 morphological features used here (Table 2.2) show a gradual depth-dependent change (Figure 2.20 and Figure 2.21).

In general, the apical tree of cortical pyramidal cells reaches the pia, enabling these cells at all depths to receive inputs that target layer 1 (mostly cortico-cortical bundles and input from secondary thalamic nuclei). Inputs to layer 1 were recently shown to regulate robustness to sensory inputs (Egger et al. 2015). Furthermore, the dependence of various features on depth (e.g., of total dendritic length, number of branches, horizontal filed span of dendrites) might suggest that L2/L3 cells at different depths sample differently the input sources arriving to these layers. Also, deeper cells (with longer total dendritic length and surface area) receive more synapses (about twice assuming that the synaptic density/ unit length is identical at all depths) and are more likely to serve as "hub neurons" as compared to superficial neurons (Gal et al. 2017). Interestingly, it was also shown that morphological variability of the dendritic tree in the cortical tissue ensures that the average synaptic properties are robust to changes at the local network level as compared to the case where all neurons are similar to each

other (Ramaswamy et al. 2012). This might be another functional consequence of the large morphological variability of HL2/L3 PCs (see also Reimann et al. 2017).

By implementing a new method to characterize the global topology of neurons we identified two distinct morphological types of L2/L3 neurons in the human temporal cortex which we dubbed "profuse-tufted" and "slim-tufted" pyramidal neurons (Figure 2.22). These two M-types also show a significant difference in their I/F curves; the profuse-tufted fire at higher rates (Figure 2.23D). In contrast, based on our limited data (n = 14) of L2/L3 PCs from mouse temporal cortex, these two classes of neurons could not be found (but see Van Aerde and Feldmeyer 2015 showing that in L3 in rat medial prefrontal cortex there are several electro-morphological subtypes).

We would like to emphasize that our approach revealed two morphologically different cell-types and in parallel found that the I/F curves were cell-type specific. The finding of these two cell-types could have implications beyond the I/F curves. For example, the excitability of the apical tufts could differ between these classes (perhaps showing e.g., NMDA spikes in one class and not/less so in the other class), and/or in the properties of the back and/or forward propagation of electrical signals along the dendrites. Other features that might also differ among these classes could be their spine density, wiring diagrams between cell-types, etc. These are important questions for future studies.

It is tempting to speculate that these two cell types in HL2/L3 PCs are comparable to the two excitatory cell types in layer 5 of the rodents; i.e., the slender and thick tufted pyramidal cells (Hallman et al. 1988; Mason and Larkman 1990), which have been found in the somatosensory, visual, auditory, motor and prefrontal cortices (Hübener et al. 1990; Gao and Zheng 2004; Morishima and Kawaguchi 2006; Larsen et al. 2007; Sakata and Harris 2009; Meyer et al. 2010; Oberlaender et al. 2012; Van Aerde and Feldmeyer 2015). In rodents, these two types of neurons are thought to be the main output cells of these cortices, and presumably project to different regions (Alloway 2008; Groh et al. 2008; Aronoff et al. 2010; Meyer et al. 2010; Oberlaender et al. 2011; Narayanan et al. 2015). Studies have also shown that thick and thin tufted neurons differentially increase their firing activity depending on the behavioral state of the animal (de Kock et al. 2007; de Kock and Sakmann 2009). Whether the slim-tufted and the profuse-tufted neurons that we found in human L2/L3 also project to different regions should be tested, perhaps using the whole tissue taken out during surgery, which includes subcortical regions as well. We stress that we gave new names to these human L2/L3 pyramidal cells to avoid potential confusion with the two L5 pyramidal cell types found in rodents.

We further analyzed whether the patient parameters were different between the two groups. Four parameters were compared: the age of epilepsy onset, seizure frequency, the total number of seizures and years of epilepsy. None of the four patient parameters corresponded to the clustering of individual morphologies into slim-tufted and profuse-tufted pyramidal cells (Supplementary Figure 2.19B-F). Thus, the two groups of cells are not related to the patient parameters.

L2/L3 neurons using 14 biophysical features, both passive and active (see Table 2.3). In contrast to the morphological features, human L2/L3 pyramidal neurons did not show gradual depth-dependent changes for most of the biophysical features (Figure 2.23C). Interestingly and counterintuitively, we found that the input resistance of the neurons did not correlate with the surface area of the cells. However, the membrane time constant did correlate (but not strongly) with the surface area (Supplementary Figure 2.20A).

We also analyzed more fully than ever before the active biophysical properties of HL2/L3 PCs (Table 2.3). The properties of individual spikes for HL2/L3 PCs (mean half width of 1.26 ± 0.42, mean spike amplitude of 85.09 ± 5.24, Table 2.3) were typical of the cortical pyramidal cells of rodents (Staiger et al. 2014). Interestingly, HL2/L3 PCs show a prominent sag in the voltage response to long current injections (Table 2.3, feature #12), unlike the L2/L3 pyramidal neurons from sensory cortices in rodents (Larkum et al. 2007; but see however Van Aerde and Feldmeyer 2015 who found a subpopulation of L3 pyramidal cells that do display larger sag of 12%). The sag voltage is an indication of the presence of HCN channels. In L5 pyramidal cells HCN channels are located almost exclusively in the apical dendritic tree (Williams and Stuart 2000; Berger et al. 2001; Lörincz et al. 2002; Kole et al. 2006; Harnett et al. 2015). These channels are active at resting membrane potentials, further activated with hyperpolarization and deactivated with depolarization. The reverse voltage-dependence of HCN channels suggests a regulating role in which its main function is to oppose changes in membrane potential (Wahl-Schott and Biel 2009). Moreover, HCN channels play an important role in various dendritic computations and temporal summation and act as a spatial filter that preferentially dampens distal inputs (Magee 1998; Williams and Stuart 2000; Harnett et al. 2015). HCN channels were also shown to have a behavioral role in controlling spatial working memory (Wang et al. 2007) and long term synaptic plasticity (Nolan et al. 2004). Hence, HCN channels may endow human L2/L3 neurons with potentially greater computation properties than in the corresponding neurons in the mouse.

We also computed the cable length, **L**, of HL2/L3 PCs, which is a feature that combines both morphological and biophysical properties. Unlike the physical length, which correlated with depth, the mean cable length of the neurons was depth independent. Some of the deeper neurons had a larger diameter and a larger specific membrane resistivity, both of which can be interpreted as compensating for the expected marked voltage attenuation in long dendrites (Figure 2.24 and Figure 2.25). This compensation resulted in a comparable voltage attenuation along the dendritic tree in both the small (near pia) and large (deeper) HL2/L3 PCs. Our passive cable analysis showed very large voltage attenuations from distal dendrites in the entire HL2/L3 PCs population (Figure 2.24 C1 and C2); we therefore predict that active mechanisms (e.g., dendritic Ca_2+ spikes and NMDA spikes) operate in these cells in order to compensate for this large voltage attenuation, as is found in rodents (Larkum et al. 2009; Chen et al. 2011; Xu et al. 2012; Major et al. 2013; Smith et al. 2013; Grienberger et al. 2015; Takahashi et al. 2016). The characterization of the membrane properties of human dendrites constitutes a crucial experimental challenge for the near future.

Our study provides a first systematic multi-feature analysis of the morphological, biophysical and passive cable properties of human Layer 2 and Layer 3 pyramidal neurons. We found two district morpho-electrotonic types within this population, which we termed "slim-tufted" and "profuse-tufted" pyramidal neurons. It would also be important to further characterize additional building blocks (the various neuron types, Wang et al. 2015) composing the human cortex which, in many ways, enable the unique cognitive capabilities found in humans.

Author contributions

Funding

Acknowledgements

References

Alloway KD. 2008. Information Processing Streams in Rodent Barrel Cortex: The Differential Functions of Barrel and Septal Circuits. Cereb Cortex. 18:979–989.

Aronoff R, Matyas F, Mateo C, Ciron C, Schneider B, Petersen CCH. 2010. Long-range connectivity of mouse primary somatosensory barrel cortex. Eur J Neurosci. 31:2221–2233.

Benavides-Piccione R, Fernaud-Espinosa I, Robles V, Yuste R, DeFelipe J. 2013. Age-based comparison of human dendritic spine structure using complete three-dimensional reconstructions. Cereb Cortex. 23:1798–1810.

Benjamini Y, Hochberg Y. 1995. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B. 57:289–300.

Berger T, Larkum ME, Lüscher HR. 2001. High I(h) channel density in the distal apical dendrite of layer V pyramidal cells increases bidirectional attenuation of EPSPs. J Neurophysiol. 85:855–868.

Callaway EM. 2004. Feedforward, feedback and inhibitory connections in primate visual cortex. Neural Networks. 17:625–632.

Carlsson G. 2009. TOPOLOGY AND DATA. Bull New Ser Am Math Soc. 46:255–308.

Carnevale NT, Hines ML. 2006. The NEURON Book. Cambridge, UK: Cambridge University Press.

Chen X, Leischner U, Rochefort NL, Nelken I, Konnerth A. 2011. Functional mapping of single spines in cortical neurons in vivo. Nature. 475:501–505.

Constantinople CM, Bruno RM. 2013. Deep cortical layers are activated directly by thalamus. Science. 340:1591–1594.

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

de Kock CPJ, Bruno RM, Spors H, Sakmann B. 2007. Layer- and cell-type-specific suprathreshold stimulus representation in rat primary somatosensory cortex. J Physiol. 581:139–154.

de Kock CPJ, Sakmann B. 2009. Spiking in primary somatosensory cortex during natural whisking in awake head-restrained rats is cell-type specific. Proc Natl Acad Sci U S A. 106:16446–16450.

Defelipe J. 2011. The evolution of the brain , the human nature of cortical circuits, and intellectual creativity. 5:1–17.

DeFelipe J, Alonso-Nanclares L, Arellano JI. 2002. Microstructure of the neocortex: Comparative aspects. J Neurocytol. 31:299–316.

Douglas RJ, Martin KAC. 2004. NEURONAL CIRCUITS OF THE NEOCORTEX. Annu Rev Neurosci. 27:419–451.

Duda RO, Hart PE, Stork DG. 2001. Pattern Classification. New York John Wiley, Sect.

Egger R, Schmitt AC, Wallace DJ, Sakmann B, Oberlaender M, Kerr JND. 2015. Robustness of sensory-evoked excitation is increased by inhibitory inputs to distal apical tuft dendrites. Proc Natl Acad Sci. 112:14072–14077.

Elston GN, Benavides-Piccione R, Defelipe J. 2005. A study of pyramidal cell structure in the cingulate cortex of the macaque monkey with comparative notes on inferotemporal and primary visual cortex. Cereb Cortex. 15:64–73.

Elston GN, Benavides-Piccione R, DeFelipe J. 2001. The pyramidal cell in cognition: a comparative study in human and monkey. J Neurosci. 21:RC163.

Evgeniou T, Pontil M, Elisseeff A. 2004. Leave One Out Error, Stability, and Generalization of Voting Combinations of Classifiers. Mach Learn. 55:71–97.

Eyal G, Verhoog MB, Testa-Silva G, Deitcher Y, Lodder JC, Benavides-Piccione R, Morales J, DeFelipe J, de Kock CP, Mansvelder HD, Segev I. 2016. Unique membrane properties and enhanced signal processing in human neocortical neurons. Elife. 5:e16553.

Feldmeyer D. 2012. Excitatory neuronal connectivity in the barrel cortex. Front Neuroanat. 6:24.

Fernandez-Gonzalez P, Benavides-Piccione R, Leguey I, Bielza C, Larrañaga P, DeFelipe J. 2017. Dendritic-branching angles of pyramidal neurons of the human cerebral cortex. Brain Struct Funct. 222:1847–1859.

Fortier É, Noreau A, Lepore F, Boivin M, Pérusse D, Rouleau GA, Beauregard M. 2011. Early influence of the rs4675690 on the neural substrates of sadness. J Affect Disord. 135:336–340.

Gal E, London M, Globerson A, Ramaswamy S, Reimann MW, Muller E, Markram H, Segev I. 2017. Rich cell-type-specific network topology in neocortical microcircuitry. Nat Neurosci. 20:1004–1013.

Gao W-J, Zheng Z-H. 2004. Target-specific differences in somatodendritic morphology of layer V pyramidal neurons in rat motor cortex. J Comp Neurol. 476:174–185.

Grienberger C, Chen X, Konnerth A. 2015. Dendritic function in vivo. Trends Neurosci. 38:45–54.

Groh A, de Kock CPJ, Wimmer VC, Sakmann B, Kuner T. 2008. Driver or Coincidence Detector: Modal Switch of a Corticothalamic Giant Synapse Controlled by Spontaneous Activity and Short-Term Depression. J Neurosci. 28.

Hallman LE, Schofield BR, Lin C-S. 1988. Dendritic morphology and axon collaterals of corticotectal, corticopontine, and callosal neurons in layer V of primary visual cortex of the hooded rat. J Comp Neurol. 272:149–160.

Harnett MT, Magee JC, Williams SR, MT H, JC M, SR W. 2015. Distribution and Function of HCN Channels in the Apical Dendritic Tuft of Neocortical Pyramidal Neurons. J Neurosci. 35:1024–1037.

Hay E, Schurmann F, Markram H, Segev I. 2013. Preserving axosomatic spiking features despite diverse dendritic morphology. J Neurophysiol. 109:2972–2981.

Hübener M, Schwarz C, Bolz J. 1990. Morphological types of projection neurons in layer 5 of cat visual cortex. J Comp Neurol. 301:655–674.

Inda MC, DeFelipe J, Muñoz A. 2006. Voltage-gated ion channels in the axon initial segment of human cortical pyramidal cells and their relationship with chandelier cells. Proc Natl Acad Sci U S A. 103:2920–2925.

Köhling R, Avoli M. 2006. Methodological approaches to exploring epileptic disorders in the human brain in vitro. J Neurosci Methods. 155:1–19.

Kole MHP, Hallermann S, Stuart GJ. 2006. Single Ih channels in pyramidal neuron dendrites: properties, distribution, and impact on action potential output. J Neurosci. 26:1677–1687.

Larkum ME, Nevian T, Sandler M, Polsky A, Schiller J. 2009. Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: a new unifying principle. Science. 325:756–760.

Larkum ME, Waters J, Sakmann B, Helmchen F. 2007. Dendritic Spikes in Apical Dendrites of Neocortical Layer 2/3 Pyramidal Neurons. J Neurosci. 27:8999–9008.

Larsen DD, Wickersham IR, Callaway EM. 2007. Retrograde tracing with recombinant rabies virus reveals correlations between projection targets and dendritic architecture in layer 5 of mouse barrel cortex. Front Neural Circuits. 1:5.

Li L, Ji X, Liang F, Li Y, Xiao Z, Tao HW, Zhang LI. 2014. A Feedforward Inhibitory Circuit Mediates Lateral Refinement of Sensory Representation in Upper Layer 2 / 3 of Mouse Primary Auditory Cortex. 34:13670–13683.

Lörincz A, Notomi T, Tamás G, Shigemoto R, Nusser Z. 2002. Polarized and compartment-dependent distribution of HCN1 in pyramidal cell dendrites. Nat Neurosci. 5:1185–1193.

Magee JC. 1998. Dendritic hyperpolarization-activated currents modify the integrative properties of hippocampal CA1 pyramidal neurons. J Neurosci. 18:7613–7624.

Major G, Larkum ME, Schiller J. 2013. Active properties of neocortical pyramidal neuron dendrites. Annu Rev Neurosci. 36:1–24.

Marx M, Feldmeyer D. 2013. Morphology and physiology of excitatory neurons in layer 6b of the somatosensory rat barrel cortex. Cereb Cortex. 23:2803–2817.

Mason A, Larkman A. 1990. Correlations between morphology and electrophysiology of pyramidal neurons in slices of rat visual cortex. II. Electrophysiology. J Neurosci. 10:1415–1428.

Meyer HS, Wimmer VC, Oberlaender M, De Kock CPJ, Sakmann B, Helmstaedter M. 2010. Number and laminar distribution of neurons in a thalamocortical projection column of rat vibrissal cortex. Cereb Cortex.

Mirz F, Ovesen T, Ishizu K, Johannsen P, Madsen S, Gjedde A, Pedersen CB. 1999. Stimulus-dependent central processing of auditory stimuli: a PET study. Scand Audiol. 28:161–169.

Mohan H, Verhoog MB, Doreswamy KK, Eyal G, Aardse R, Lodder BN, Goriounova N a., Asamoah B, B. Brakspear a. BC, Groot C, van der Sluis S, Testa-Silva G, Obermayer J, Boudewijns ZSRM, Narayanan RT, Baayen JC, Segev I, Mansvelder HD, de Kock CPJ. 2015. Dendritic and Axonal Architecture of Individual Pyramidal Neurons across Layers of Adult Human Neocortex. Cereb Cortex. bhv188.

Molnár G, Oláh S, Komlosi G, Füle M, Szabadics J, Varga C, Barzo P, Tamás G. 2008. Complex Events Initiated by Individual Spikes in the Human Cerebral Cortex. PLoS Biol. 6:e222.

Morishima M, Kawaguchi Y. 2006. Recurrent Connection Patterns of Corticostriatal Pyramidal Cells in Frontal Cortex. J Neurosci. 26.

Narayanan RT, Egger R, Johnson a. S, Mansvelder HD, Sakmann B, de Kock CPJ, Oberlaender M. 2015. Beyond Columnar Organization: Cell Type- and Target

Layer-Specific Principles of Horizontal Axon Projection Patterns in Rat Vibrissal Cortex. Cereb Cortex. bhv053-.

Nieuwenhuys R. 1994. The neocortex. An overview of its evolutionary development, structural organization and synaptology. Anat Embryol (Berl). 190:307–337.

Nolan MF, Malleret G, Dudman JT, Buhl DL, Santoro B, Gibbs E, Vronskaya S, Buzsáki G, Siegelbaum SA, Kandel ER, Morozov A. 2004. A behavioral role for dendritic integration: HCN1 channels constrain spatial memory and plasticity at inputs to distal dendrites of CA1 pyramidal neurons. Cell. 119:719–732.

Oberlaender M, Boudewijns ZSRM, Kleele T, Mansvelder HD, Sakmann B, de Kock CPJ. 2011. Three-dimensional axon morphologies of individual layer 5 neurons indicate cell type-specific intracortical pathways for whisker motion and touch. Proc Natl Acad Sci. 108:4188–4193.

Oberlaender M, de Kock CPJ, Bruno RM, Ramirez A, Meyer HS, Dercksen VJ, Helmstaedter M, Sakmann B. 2012. Cell Type-Specific Three-Dimensional Structure of Thalamocortical Circuits in a Column of Rat Vibrissal Cortex. Cereb Cortex. 22:2375–2391.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. 2011. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 12:2825–2830.

Poirazi P, Mel BW. 2001. Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. Neuron. 29:779–796.

Polsky A, Mel BW, Schiller J. 2004. Computational subunits in thin dendrites of pyramidal cells. Nat Neurosci. 7:621–627.

Qi G, Radnikow G, Feldmeyer D. 2015. Electrophysiological and morphological characterization of neuronal microcircuits in acute brain slices using paired patch-clamp recordings. J Vis Exp. 52358.

Rall W. 1959. Branching dendritic trees and motoneuron membrane resistivity. Exp Neurol. 1:491–527.

Ramaswamy S, Hill SL, King JG, Schürmann F, Wang Y, 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.

Reimann MW, Horlemann A-L, Ramaswamy S, Muller EB, Markram H. 2017. Morphological Diversity Strongly Constrains Synaptic Connectivity and Plasticity. Cereb Cortex. 8:1–16.

Sakata S, Harris KD. 2009. Laminar structure of spontaneous and sensory-evoked population activity in auditory cortex. Neuron. 64:404–418.

Smith SL, Smith IT, Branco T, Häusser M. 2013. Dendritic spikes enhance stimulus selectivity in cortical neurons in vivo. Nature. 503:115–120.

Staiger JF, Bojak I, Miceli S, Schubert D. 2014. A gradual depth-dependent change in connectivity features of supragranular pyramidal cells in rat barrel cortex. Brain Struct Funct. 220:1–21.

Szabadics J, Varga C, Molnár G, Oláh S, Barzo P, Tamás G. 2006. Excitatory effect of GABAergic axo-axonic cells in cortical microcircuits. Science. 311:233–235.
Takahashi N, Oertner TG, Hegemann P, Larkum ME. 2016. Active cortical dendrites modulate perception. Science (80-). 354.

Testa-silva G, Verhoog MB, Linaro D, Kock CPJ De, Baayen JC, Meredith RM, Zeeuw CI De, Giugliano M, Mansvelder HD. 2014. High Bandwidth Synaptic Communication and Frequency Tracking in Human Neocortex. 12.

Tian C, Wang K, Ke W, Guo H, Shu Y. 2014. Molecular identity of axonal sodium channels in human cortical pyramidal cells. Front Cell Neurosci. 8:297.

Uw S, Ng AY, Jordan MI, Weiss Y. 2001. On spectral clustering: Analysis and an algorithm. Adv Neural Inf Process Syst 14. 849–856.

Van Aerde KI, Feldmeyer D. 2015. Morphological and physiological characterization of pyramidal neuron subtypes in rat medial prefrontal cortex. Cereb Cortex. 25:788–805.

Varga C, Tamas G, Barzo P, Olah S, Somogyi P. 2015. Molecular and Electrophysiological Characterization of GABAergic Interneurons Expressing the Transcription Factor COUP-TFII in the Adult Human Temporal Cortex. Cereb Cortex. 25:4430–4449.

Verhoog MB, Goriounova NA, Obermayer J, Stroeder J, Hjorth JJJ, Testa-silva G, Baayen JC, Kock CPJ De, Meredith RM, Mansvelder HD. 2013. Mechanisms Underlying the Rules for Associative Plasticity at Adult Human Neocortical Synapses. 33:17197–17208.

Wahl-Schott C, Biel M. 2009. HCN channels: Structure, cellular regulation and physiological function. Cell Mol Life Sci.

Wang B, Ke W, Guang J, Chen G, Yin L, Deng S, He Q, Liu Y, He T, Zheng R, Jiang Y, Zhang X, Li T, Luan G, Lu HD, Zhang M, Zhang X, Shu Y. 2016. Firing Frequency Maxima of Fast-Spiking Neurons in Human, Monkey, and Mouse Neocortex. Front Cell Neurosci. 10:239.

Wang B, Yin L, Zou X, Ye M, Liu Y, He T, Deng S, Jiang Y, Zheng R, Wang Y, Yang M, Lu H, Wu S, Shu Y. 2015. A subtype of inhibitory interneuron with intrinsic persistent activity in human and monkey neocortex. Cell Rep. 10:1450–1458.

Wang M, Ramos BP, Paspalas CD, Shu Y, Simen A, Duque A, Vijayraghavan S, Brennan A, Dudley A, Nou E, Mazer J a., McCormick D a., Arnsten AFT. 2007. α 2A-Adrenoceptors Strengthen Working Memory Networks by Inhibiting cAMP-HCN Channel Signaling in Prefrontal Cortex. Cell. 129:397–410.

Williams SR, Stuart GJ. 2000. Site independence of EPSP time course is mediated by dendritic I(h) in neocortical pyramidal neurons. J Neurophysiol. 83:3177–3182.

Xu N, Harnett MT, Williams SR, Huber D, O'Connor DH, Svoboda K, Magee JC. 2012. Nonlinear dendritic integration of sensory and motor input during an active sensing task. Nature. 492:247–251.

Yuste R. 2010. Dendritic spines. MIT Press.

Zhu JJ. 2000. Maturation of layer 5 neocortical pyramidal neurons: amplifying salient layer 1 and layer 4 inputs by Ca 2+ action potentials in adult rat tuft dendrites. J Physiol. 526:571–587.

Forest of artificial neurons
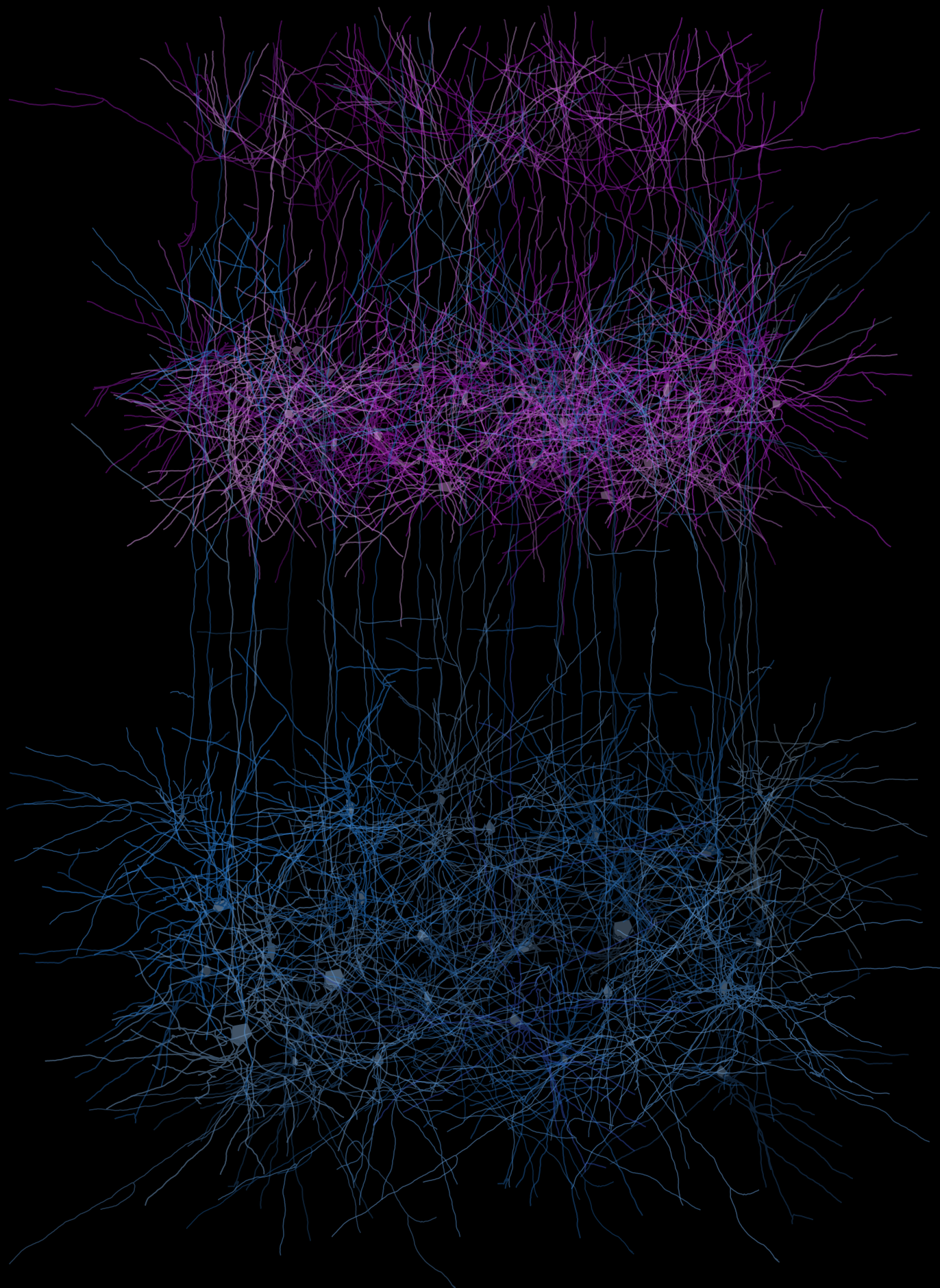
3

Generating neuronal morphologies

"What I cannot create, I do not understand"
– Richard Feyman

3.1 Introduction

Recent advances in neuroimaging techniques (Peng 2008, Haberl et al. 2015, Economo et al. 2016) and open source platforms for automatic reconstructions algorithms (Peng et al. 2015) are rapidly increasing the number of available morphological reconstructions (Ascoli et al. 2007). However, the number of such biological reconstructions of morphologies within a brain region are far from being sufficient to populate digital reconstructions of large brain networks that consist of tens of thousands cells (Markram et al. 2015, Egger et al. 2014). It has been found (Landau et al. 2016, Ramaswamy et al. 2012) that morphological variability influences the functionality of computational brain networks. As a result, these computational models require "clones", i.e., noisy copies, of the original morphologies to increase the morphological variability of the derived networks.However, the clones only differ from the original cells in local fluctuations and therefore cannot capture the biological variability that is required for the simulation of digital reconstructions. Therefore, an algorithm to generate large numbers of distinct artificial morphologies that are statistically indistinguishable from the biological neurons is required.

A large variety of synthesis models have been proposed to simulate the growth of neuronal morphologies (Hillman 1979, Burke et al. 1992, Ascoli et al. 2001). These range from highly-detailed models that simulate the molecular mechanisms of neuronal growth, to simplified models that simulate phenomenological growth mechanisms based

on either basic mathematical rules or statistical sampling of morphological features (see Figure 3.1). Biophysically accurate models, that simulate the detailed neural growth (Zubler and Douglas 2009), focus on the details of the microscopic growth in order to understand the biological mechanisms that govern neuronal development. As a result, they are not optimized for the efficient generation of large numbers of neurons. On the other hand, mathematical models implement simplified growth mechanisms to study the effect of basic mathematical principles on neuronal growth. Different growth models focus on different geometric and physical aspects of growth. The effect of spatial boundaries and spatial embedding (Luczak 2006, Luczak 2010), the optimization of local dendritic arborization properties (such as material cost and conduction time, Cuntz et al. 2010) and the self-referential dendritic forces (Samsonovich and Ascoli 2003, Memelli et al. 2013) are just a few of the mathematical models studied. These models provide a better intuition about the different mechanisms involved in neuronal growth. Statistical models use a set of morphological properties extracted from the original cells (Ascoli et al. 2001, Koene et al. 2009, Lopez-Cruz et al. 2011) to constrain the growth process.

To study the effect of topological constraints on neuronal growth, I have developed a synthesis algorithm based on the topological profile of branching morphologies (see Chapter 2.1 for more details on the method). The topological neuron synthesis (TNS) method is a combination of the mathematical and statistical approaches. A limited number of features, extracted from the biological population are used to constrain the neuronal growth, such as the soma size and the thickness of the branches of the tree. The TMD of a biological neuron (see Chapter 2) is used to guide the neuronal growth. Each neurite grows independently from the others taking into account a persistence barcode extracted from a neurite of a reconstructed morphology. This persistence barcode regulates the probability of each branch to bifurcate and terminate, as a function of its radial distance from the soma. Each branch is generated as a directed random walk, i.e., the direction of the next step within a branch is a combination of a random direction, defined by a unit vector, a target direction, defined by the bifurcation angle at the beginning of the branch, and the history of the previous directions of the branch.

The thickness of the neuronal branches is as important as its branching structure for its functionality (Cuntz et al. 2007, van Elburg and van Ooyen 2010, Bird and Cuntz 2016). Unfortunately, despite the great progress in imaging techniques, technical limitations often result in a poor diameter estimation and a few methods have been developed to correct these artifacts (Conde-Sousa et al. 2017). As a result, a method to

A I. Sample points according to spatial distribution. II. Connect points according to algorithm. III. Add details to make the tree more realistic.

B I. Grow tree point by point. II. Sampling biological distributions. III. Add details to make the tree more realistic.

$$x_{n+1} = x_n + \sum_i (\vec{F}_i)$$

II. Or... according to a mathematical model.
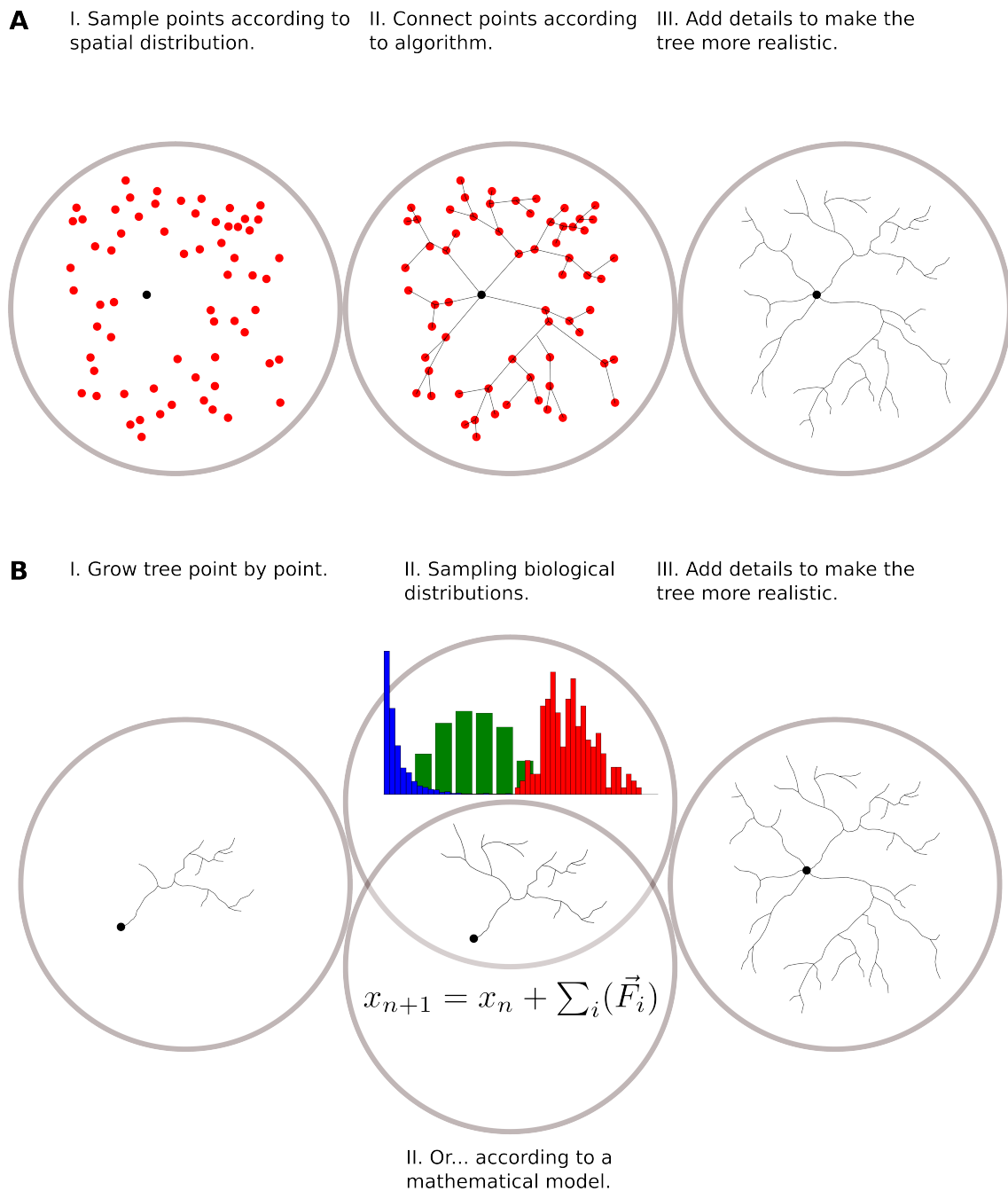
Fig. 3.1 Phenomenological synthesis methods. Two types of synthesis models have been developed: the generalized and the local models. A. The generalized schemes generate the key points of the tree and then connect them according to a mathematical rule, which corresponds to an optimization process, such as the minimization of wiring, and describes the properties of biological morphologies. B. The local growth models generate neuronal trees segment by segment either respecting a set of statistical distributions, such as section lengths (in blue), number of sections (in green) and bifurcation angles (in red), or according to a mathematical principle, such as the sum of the forces that are applied on the growing tip. Both models need a final step to improve independent properties, such as the thickness of the trees.

computationally assign accurate diameters from sparse biological data is required. The algorithm to assign accurate diameters that I propose (see SI, TNS paper: Algorithm 5) traverses the tree from the tips to the root and increases the diameters according to the biologically sampled taper rate (Scorcioni et al. 2008, Koene et al. 2009). At each bifurcation the parent diameter is computed from the rall ratio n (Rall 1962) and the diameters of its children $D^n = d_1^n + d_2^n + \ldots$. This algorithm generates diameters that monotonically decrease with the path distance from the soma, and are statistically close to the input neuronal branch diameters.

In order to ensure that the synthesis algorithm reproduces the statistical properties of the biological neurons' morphologies a number of independent validations are performed (see section Validation). Initially, the topological profiles of the synthesized neurons are compared to those of the biological neurons to ensure that the stochastic algorithm generates cells with similar topology. Then, each synthesized neuron is compared to the biological dataset, with respect to a large set of morphological features: the "input" features, that are used as input in the algorithm, and the "emergent" features that are not used as input. A cell is included in the synthesized population if its morphometrics are statistically close to the biological features. Then the synthesized population, which consists of cells that have passed the single-cell validations, is compared against the biological dataset taking into account an extensive set of morphological features (see Appendix C for details). This last step ensures that the TNS algorithm reproduces the variability of the input biological morphologies.

The TNS algorithm is used to generate dendrites of interneurons and pyramidal cells of different cortical morphological cell types. To validate the accuracy of this algorithm, a large number of cells of the most abundant morphological types in the cortex (L2/3 and L5 pyramidal cells) are synthesized and validated (see Results, Validation). The TNS algorithm generates cells that are statistically close to the biological cells and reproduces a large variety of dendritic shapes without the need for manual fine tunning of the input parameters. Figures [?] and [?] illustrate that the biological diversity of different interneuron and pyramidal cell types is reflected in the synthesized morphologies. Cells that are synthesized with the same input parameters but without taking into account the correlations between the bifurcation and termination probabilities that are introduced by their topological profiles, are of very poor quality. Therefore correlations between morphological features are essential for the generation of biologically accurate cells, and the topological profiles of neurons capture these correlations, without the need to explicitly use them.

3.2 Generative model of cortical dendritic morphologies based on the topology of their branching architecture

Lida Kanari, Athanassia Chalimourda, Guy Atenekeng, Joe Graham,
Julian Shillcock, Kathryn Hess, and Henry Markram

Abstract: Neuronal morphologies "shape" the dynamical properties of the brain. As a result, the generation of digital morphologies that reproduce the branching structures of neurons is a vital step towards the reconstruction and simulation of physiologically realistic brain networks. However, the principles that define how dendritic and axonal arbors take shape are still largely unknown. In a recent study, we introduced a topological descriptor of branching morphologies that is able to reliably categorize neurons into morphologically distinct groups. Here, we use this descriptor and a small set of morphometrics to generate virtual dendritic morphologies. Each generated morphology is validated against biological neurons, based on a large number of morphological features. The synthesis algorithm driven by the topological architecture of dendrites generates realistic morphologies for various distinct neuronal types. Our results demonstrate that a topology-based generative model of neurons implicitly captures correlations of features within a growing shape, without the need for explicit identification of dependencies between features.

Keywords: Dendritic morphology; Virtual neuron; Topological generation of cells; Artificial neuron; Topological Morphology Descriptor, Morphological synthesis

Introduction

The dynamical properties of a neuronal network depend crucially on the branching of neuronal morphologies, which affects both the functionality (Yi et al. 2017) and connectivity of a neuron (Chklovskii 2004, Wen et al. 2009). Cajal argued that the shape of neurons reflects the communication between them (Cajal 1899). It is now well established that the anatomy of neurons regulates the wiring of the brain (Chklovskii 2004). More specifically, dendritic arborizations determine the integration of input signals (Larkum et al. 2009, Yi et al. 2017), while axonal projections govern the propagation of signals to different brain regions (Wang et al. 2015). The wide variety of neuronal shapes supports the composite functional roles of different cell types, though the precise role of each neuronal type is not yet fully understood due to the complexity of biological branching patterns.

A digital reconstruction of a physiologically realistic network, such as the Blue Brain Project (Markram et al. 2015, Egger et al. 2014), requires a large number of detailed neuronal morphologies (Shillcock et al. 2016). Recent advances in automatic reconstruction (Peng et al. 2015) and the systematic registration of reconstructions in standardized databases (Ascoli et al. 2007) are a long way from making available sufficient numbers of unique morphologies to populate biologically realistic networks of a brain region. However, the variability of neuronal shapes is essential for the functionality of computational models (Landau et al. 2016, Ramaswamy et al. 2012). Thus an algorithm that can accurately reproduce the diversity of biological morphologies for all cell types is necessary for the reconstruction and simulation of biophysically accurate computational models of the brain (Shillcock et al. 2016).

The fundamental problem of neuronal synthesis (Hillman 1979) is the difficulty of capturing and recreating the correlations between morphological features, from the few available reconstructions of a morphological type. These correlations arise from highly complicated developmental processes, which take place over many length and time scales. Previous synthesis models have addressed this problem in different ways. Detailed models operate at the molecular scale in an attempt to retain as many details as possible in order to capture all the correlations in the cellular growth. Phenomenological models are based on fundamental mathematical principles (Luczak 2006, Cuntz et al. 2010), or on statistical sampling of the morphological distributions (Ascoli et al. 2001, Koene et al. 2009). These unfortunately tend to disregard most correlations as the sampling is usually performed independently, and even if not, correlations must be explicitly identified (Lopez-Cruz et al. 2011).

Biophysically accurate models simulate detailed neural growth by taking into account the known molecular mechanisms that contribute to the development of neurons (Zubler and Douglas 2009). While these models are very important for understanding the biological mechanisms that govern neuronal development, they focus on the microscopic scale of growth and have a large number of parameters. As a result, they are not appropriate for the generation of large numbers of neurons at whole brain length-scales.

Mathematical models with few parameters, on the other hand, focus on a specific growth mechanism to study the effect of different factors on neuronal growth. The impact of spatial boundaries and embedding has been studied by Luczak (2006, 2010), the minimization of wiring cost by Cuntz (2010) and self-referential forces by Samsonovich and Ascoli (2003), and Memelli (2013), to name a few. These models provide good intuition about the selected mechanisms involved in neuronal growth. However, due to the small number of parameters, they fail to capture the full complexity of neuronal arborizations for a wide variety of morphological types (m-types) without appropriate adjustments to the algorithms for specific branching shapes. Statistical models are based on sampling from a set of morphological properties (Ascoli et al. 2001, Koene et al. 2009, Lopez-Cruz et al. 2011) that guide the growth process. Such statistical models can produce cells of specific morphological types with high accuracy (Koene et al. 2009) but cannot capture the correlations that have not been explicitly defined.

The limitations of previous synthesis methods applied to large numbers of morphologically diverse cells imply that it is necessary to combine the mathematical and statistical models into a unified synthesis method that circumvents the explicit selection of correlated features, while also being computationally tractable. We developed a synthesis algorithm based on the topological profile of a branching morphology (Kanari et al. 2017) to investigate the significance of topology on neuronal shapes. The Topological Morphology Descriptor (TMD) of a tree encodes its branching structure in a "barcode", i.e., a set of closed intervals, called "bars", in the real line. Each bar encodes the starting and ending radial distance from the soma of a component in the underlying branching structure (see Appendix C: Topological morphology descriptor algorithm). Since the TMD couples the topological and geometric properties of a tree into a single descriptor, it is effective for the discrimination of tree types. Thus, this topological descriptor that encodes the relevant correlations between morphological features is also appropriate for the generation of artificial neurons.

We therefore use a TMD profile of a morphological type to define the branching and termination probabilities of a neuron. Each branch is simulated as a directed random

walk, and the local directionality is defined by the biological branching angles. The degree of randomness of each branch defines the tortuosity of the neuronal morphology and is an independent input parameter. The topological profiles of the neuronal reconstructions and the statistical distributions of a small set of morphological features (soma size, thickness of branches) are used as input to the algorithm.

The topological neuron synthesis (TNS) algorithm (see also Appendix C: Algorithm 2) is used for the generation of artificial rodent dendrites of a large variety of m-types (see Figures 3.3-3.4). We demonstrate that the qualitative (Figure 3.5-3.6) and quantitative (Figures 3.7-3.8) validations of all dendritic types agree with the biological data. The TNS-generated neurons are compared to neurons generated by the same input parameters but assuming that the bifurcation and termination probabilities are independent. The algorithm that does not take into account the topological profiles of the biological trees fails to generate accurate neuronal trees (see Figure C.8). Therefore, the topology of a neuron is not only relevant for the generation of accurate dendritic trees, as it implicitly captures morphological correlations, but is also crucial for reproducing the large variety of shapes that populate the rodent cortex.

A consistent comparison of the artificial morphologies generated by different synthesis models has not yet been performed to our knowledge, since each synthesis scheme proposes a different type of morphological validation (Ascoli et al. 2001, Koene et al. 2009, Cuntz et al. 2010, Lopez-Cruz et al. 2011). In order to ensure that our synthesis algorithm reproduces the statistical properties and the variability of the biological neurons, a two-stage validation is performed, based on a wide range of morphometrics. Each virtual morphology is validated against a biological set of neurons (single-cell validation). The "input features" used for the validation ensure the reliability of the algorithm with respect to the input distributions. The "emergent features", which have not been used as input, provide a measure of the effectiveness of the algorithm to capture implicit correlations that were not used in the modeling. The morphologies that pass the single-cell quality control compose the synthesized population, which is further validated against the biological dataset in a population-to-population comparison. The validation of the synthesized cells as a population ensures that they recreate the statistical properties of the original population while capturing the biological variability. The details of the validation framework are described in Appendix C: Validation of synthesized basal and apical dendrites.

Methods

The morphological development of neurons in the brain is a complicated process that involves a large number of molecules and depends on both genetic and environmental components. The molecular processes that contribute to neuronal growth differ between species, brain regions, and morphological types. Advances in experiments and mathematical and computational models have converged to 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). These growth stages are useful for the computational modeling of the generation of artificial neurons. In this study we focus on the generation of artificial dendrites and thus we will not consider the axon path-finding.The biological development is not simulated, but information from the biological principles of morphological growth are used to design a computational algorithm that generates accurate artificial dendritic morphologies.

The TNS algorithm, which respects these biological stages of growth, consists of three main components (Figure 3.2): the initiation, elongation and branching of neurites. As illustrated in Figure 3.2, the first part of a neuron to be generated is the cell body (or soma), which is modeled as a sphere (Fig 3.2A), whose radius is sampled from a biological distribution (see Appendix C: Algorithm 3). Then, the number of neurites is sampled from the biological distribution according to its cell type. Each neurite is initialized with a trunk, the initial branch of the tree (Fig 3.2A) and a barcode sampled from the biological set of trees. 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 Appendix C: Algorithm 4, Fig 3.2B). At each step a growing tip is assigned probabilities to bifurcate and terminate that depend on the Euclidean distance from the soma and are defined by the bars of the topological profile of the associated m-type (Fig 3.2C, see Appendix C: Algorithm 5). Once a bar is used, it is removed from the barcode. The growth terminates when all the bars of the input barcode have been used.

A. Initiation of neurites

Previous studies have disregarded the direction of the neurite protrusion from the soma despite its importance (Graham and van Ooyen 2006). For some neurites the initial direction is trivially defined; for example cortical apical dendrites typically grow towards the pia. By contrast, the outgrowth direction of basal dendrites superficially appears random and is usually assumed to be so. An in-depth analysis indicates, however,

that this assumption is inaccurate, since the orientations of a neuron's processes are correlated (Figure C.2). This correlation is captured in the pairwise trunk-angle distribution, which depends on the morphological type, and is used for the generation of trunks (see Appendix C, Algorithm 3).
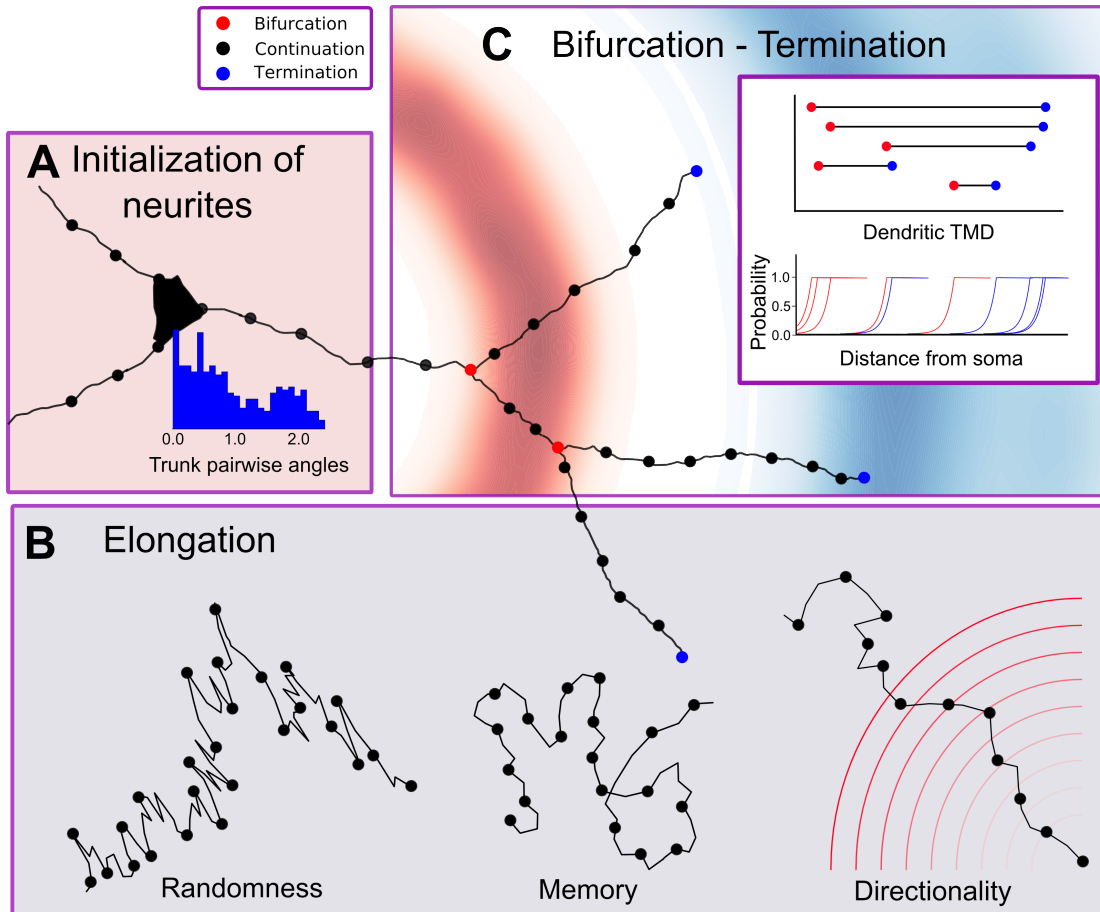
Fig. 3.2 Method of Topological Neuron Synthesis. (A) Soma generation and initiation of the neurites on the soma surface. The number and the position of the neurites on the soma define its shape. Each neurite is initialized with a point on the soma surface, an initial direction, and a diameter. The positions of the neurites on the soma depend on the pairwise trunk-angle distribution. (B) Each branch of the neurites is elongated segment by segment; a segment is defined by a length and a direction. There are three parameters associated with the elongation of neurites and govern the direction of a segment: history, targeting, randomness and history. The history takes into account the previous directions of the branch. The targeting depends on the branching angle that is defined at each bifurcation. (C) The probability of a growing tip to bifurcate, terminate or continue depends on the Euclidean distance from the soma and the topological profile (TMD) of the morphological type.

First, the initial point of a single neurite is randomly sampled on the soma surface, then the other neurites are added successively in places that respect the pairwise trunk-angle distribution. Each neurite trunk consists of a point on the soma surface, an initial direction that is normal to the soma surface, and an initial diameter. Note that

the diameters are independently corrected in the final step of the synthesis algorithm (see section D). The positions of the trunks define the soma shape; the pyramidal soma of excitatory cells originates from the apical dendrite that points towards the pia, while the spherical soma of interneurons arises from the homogeneous positioning of trunks on the surface of their cell bodies.

B. Elongation

A segment is defined as a pair of two consecutive points in the neuronal tree that has a length L and a direction $D_{segment}$, defined by a unit vector. Each virtual neurite is grown segment by segment. The segment length is taken to be constant and equal to one micron $L = 1\mu m$. 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 (Koene et al. 2009). The memory term is a weighted sum of the previous directions of the branch with the weights decreasing with distance from the tip. Different weight functions were examined, but as long as the memory function decreases- faster than linearly - with the distance from the growing tip, its form is not significant. The target vector is defined at the beginning of each branch and depends on the biological branch angles (see Appendix C, Algorithm 4). The random component is a vector of fixed length sampled uniformly from three-dimensional space at each step. For computational efficiency the growth of each branch is independent of other branches. The tortuosity of the path is defined by three parameters:

$$D_{segment} = \rho R + \tau T + \mu M,$$

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

An increase of the randomness weight ρ results in a highly tortuous branch, approaching the limit of a simple random walk when $\rho = 1$ (Pearson et al. 1905). If the targeting weight $\tau = 1$, the branch will be a straight line in the target direction (see Figure C.3). Different combinations of the three parameters (τ, ρ, μ) can generate more or less meandering branches and can reproduce the large variability of dendrites (see Appendix C, Figure C.3).

C. Branching-Termination

The neuronal branching pattern is generated as a Galton-Watson tree (Galton and Watson 1875), which is a discrete random tree generated as follows. At each growing

tip a number of offspring is independently sampled from a distribution. A neuronal tree consists only of bifurcations and terminations so the accepted values for the number of offspring are: zero (a termination), one (a continuation) or two (a bifurcation). Since the Galton-Watson tree only generates the branching structure and ignores the embedding in space, we modify the traditional process to introduce a dependency of the neuronal growth on the embedding, so that the bifurcation/termination probabilities depend on the Euclidean distance of the growing tip from the soma.

Each growing tip is assigned a bar, sampled from the barcode, that includes a starting radial distance B, an ending radial distance T and a bifurcation angle A (see Appendix C). At each step the growing tip first checks the probability to bifurcate. If a bifurcation does not occur, then the growing tip checks the probability to terminate. If the growing tip does not bifurcate or terminate, then the branch continues to elongate (see section B). The probability to bifurcate depends on the starting radial distance B. As the growing tip gets closer to the radial distance B the probability to bifurcate increases exponentially until it reaches the highest possible value (1.0). Similarly, the probability to terminate depends exponentially on the ending radial distance T.

The probabilities to bifurcate and terminate are taken from an exponential distribution $e^{-\lambda x}$, whose free parameter λ should be wisely chosen. A very steep exponential distribution (high value of λ) will result in cells that are very close to the biological input and thus will reduce the variability of the synthesized cells. On the other hand, a very low value of λ will result in cells that are almost random, since the dependence on the input persistence barcodes will be decreased significantly. The value of the parameter λ should be of the order of the step size (see Appendix C: Branching-Termination). As a result, we select a critical correlation length $\lambda = 2$, so that the bifurcation and termination points are stochastically chosen but depend on the biological persistence barcodes (See Appendix C: Description of synthesis steps).

Previous 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 are captured in the structure of the barcode. When the growing tip bifurcates, 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 biological branching structure. In the event of a termination, the growing tip is deactivated and the bar that corresponds to this termination point is similarly removed from the input TMD.

In the event of a bifurcation, two new branches are generated (Appendix C, Algorithm 5) and the directions of the daughter branches depend on the bifurcation

angle A. Three branching methods were examined (symmetric, biased and composite, see Appendix C, Fig [**?**]). The symmetric method assumes that the two daughter branches split symmetrically with respect to their parent branch's direction. The biased method assumes that one of the daughter branches inherits the parent direction and therefore the split is asymmetric. The composite method assumes a combination of the symmetric and the biased methods; for this approach two biological angles are required. We found that the basal dendrites cannot be accurately generated by the symmetric or the biased methods (see Appendix C, Fig [**?**]) but the composite method recreates the biological branching angles. The apical dendrite requires a combination of methods: the composite method for the tuft and the biased method for the obliques. The selection of the branching method is the only difference between the synthesis algorithms of the basal and apical dendrites of cortical pyramidal cells.

D. Tapering

The thickness of a neuron's branches should also be accurately reproduced (Koene et al. 2009), as it is equally important to the branching structure (Cuntz et al. 2007, van Elburg and van Ooyen 2010, Bird and Cuntz 2016). Despite the great progress in imaging techniques that enables the generation of large numbers of reconstructions (Peng 2008, Haberl et al. 2015, Economo et al. 2016), their resolution is still too limited to allow for accurate determination of diameters, which are on the order of a few microns. As a result, accurate diameters must be computationally inferred from sparse datasets of biological reconstructions. Conde-Sousa proposed a method to correct the swelling of the reconstructed diameters (Conde-Sousa et al. 2017) that usually results in lower mean diameters.

In the absence of a curated dataset, the original diameters of the reconstructed cells are used as input for the synthesis algorithm. The reconstructions are analyzed with NeuroM (github.com/BlueBrain/NeuroM) to extract the taper rate (TR, Scorcioni et al. 2008) within a branch, the Rall ratio (RR, Rall 1962) at the bifurcation points and the termination (TD) and maximum diameters (MD) of a tree. These values are used to assign diameters independently to each synthesized dendrite.

The algorithm (see Appendix C: Algorithm 6) starts from the tips of the tree and assigns diameters to the termination points sampled from the biological distribution TD. Then, the tree is traversed from the tips to the root (post-order), and the diameters are increased according to the biologically sampled taper rate TR, as long as the new diameter is less than a sampled maximum diameter MD. When the diameters of all

the children of a section have been computed, the new section is assigned a diameter according to the sampled Rall ratio RR:

$$d_{parent} = (d_1^{RR} + d_2^{RR} + \dots)^{1/RR}$$

This algorithm results in diameters that follow the original distribution. The synthesized diameters monotonically decrease with distance from the soma, a property that ensures that basic physical principles (Cuntz et al. 2007) are taken into account. Note that the swelling of the dendritic trees, resulting from staining artifacts, is not compensated for and therefore the diameters of the synthesized cells might be overestimated.

E. Validation method

Single-cell validation

In order to identify and remove the poor quality cells from the synthesized population, the distributions of key features of each cell are compared against a set of reconstructed biological cells. To measure a cell's difference from the biological dataset, a statistical distance, defined as the absolute "Difference Between the Medians" over the "Overall Visible Spread" (DBM/OVS), is used. Intuitively, this measure quantifies the difference between the medians of the two distributions with respect to their joint dispersion (See Appendix C: Validation of synthesized basal and apical dendrites). A cell is declared an outlier when at least one of the key-features, such as total length, number of bifurcations, terminations, surpasses the feature-specific thresholds, which are chosen so that the reconstructed biological cells that represent the "gold standard" are not considered as outliers. The percentage of the detected outliers in the set of computationally generated cells illustrates the accuracy of the synthesis process. The cells that successfully pass the single cell validation compose the population of synthesized cells.

Population to population validation

The synthesized population is validated by comparing the distributions of a large number of morphological features to those of the biological reconstructions. Essential features, such as the Sholl analysis, the degree of the dendritic tree (number of terminations), the branch orders and the number of sections, the total length per neurite and the path length are included in the validation and shown in Figures 3.7-3.8. Summary statistics such as mean, standard deviation, median of morphological features are reported in Table TBD. Each of the morphological distributions is compared to the biological one,

using the Kolmogorov-Smirnov distance, which quantifies the dissimilarity between two distributions. The K-S metric measures the maximum distance between two cumulative distributions, ranging from 0 for identical distributions to 1 in the case of maximal difference between them.

F. Synthesis input

The neuronal reconstructions of different morphological types were used as input to the synthesis algorithm. Few modifications were performed on the original reconstructions to compensate for reconstruction artifacts. For example, the slicing of the brain tissue and the filling of the cells with biocytin (Horikawa and Armstrong 1988) in order to retrieve and reconstruct the original morphologies, results in their shrinkage. This issue affects the tortuosity of the reconstructions (as cells appear more tortuous than they originally were) and the extent of their processes decreases. To compensate for those 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). Because repair modifies the branching properties of the tree, only cells that have been unraveled, but not repaired are used as input to the TNS algorithm for the current study. To compensate for the loss of arborizations, trees that contain less than three sections are considered cut and are thus discarded from the synthesis input at the beginning of the process.

Results

There are two major types of cortical cell, which are distinguished based on their functional roles: the excitatory cells and the inhibitory cells. Excitation is mainly mediated by the pyramidal cells, with the exception of the spiny stellate cells of layer 4, and use glutamate as a neurotransmitter. Inhibition is mediated by the interneurons, which use GABA as a neurotransmitter to regulate the cortical activity. The various types of interneurons, which also play different functional roles, are distinguished by their axonal branching patterns. An interneuron's dendrites (see Figure 3.3A), which are basal dendrites that emanate from the base of the cell body and are localized mainly around the soma, are less complex than the apical dendrites of pyramidal cells (Figure 3.4A), which ascend towards the pia and present a wider diversity of shapes. The TNS

algorithm is used for the generation of artificial dendrites of both interneurons and pyramidal cells of a large variety of morphological types.
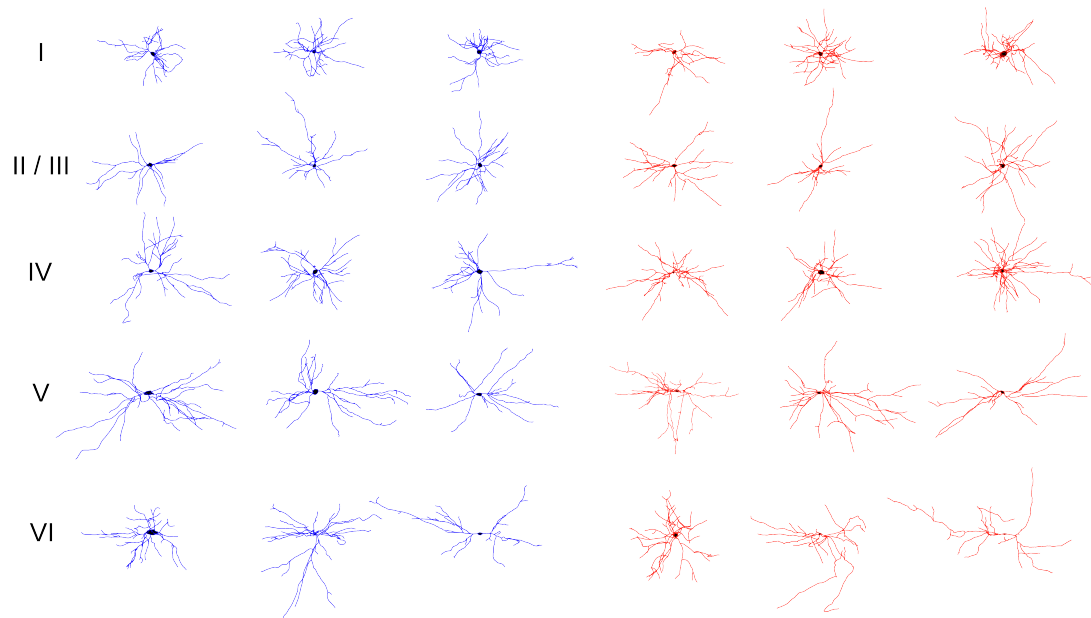
Fig. 3.3 Visual comparison of biological and virtual dendrites of rodent cortical layer 1 to 6 interneurons. The biological dendrites of interneurons (left, in blue) are compared to the corresponding synthesized dendrites of the same m-type (right, in red).

Interneurons represent about $\approx 15\%$ (Gonchar et al. 2006, Lefort et al. 2009) of neuronal cells in the cortex. There are many interneuron types, which mainly differ in their axonal branching patterns. In general, the basal dendrites of cortical interneurons do not vary greatly for different morphological types. A sample of interneuron basal dendrites of all layers is presented in Figure 3.3, (left, in blue). Artificially generated dendrites of the same types (Figure 3.3, (right, in red) are generated based on the corresponding topological profiles of the biological reconstructions. Due to the low diversity of branching patterns of interneurons, the synthesized cells were validated against the biological ones as a single population, without separating them into individual morphological types. The TNS algorithm generates basal dendrites that reproduce the characteristic shapes of different m-types of interneurons.

Pyramidal cells (PCs) represent the majority of neurons in the cortex ($\approx 85\%$, Gonchar et al. 2006, Lefort et al. 2009). There is a large variety of apical dendrite shapes that impart unique functional properties to PCs and form the basis for integrating signal inputs from different cortical layers (Larkum et al. 2007, 2009, Spruston 2008). As a result, a digital reconstruction of a cortical network requires the generation of large numbers of unique morphologies of pyramidal cells in order to reproduce the biological

variability that is observed to influence the functionality of biological networks (Landau et al. 2016, Ramaswamy et al. 2012). The great diversity of apical shapes of pyramidal cells reflects the variety of their functional roles. As a result, the morphological types of pyramidal shapes are defined based on the branching shapes of their apical trees (Marx and Feldmeyer 2013). In Figure 3.4A we illustrate a sample of reconstructed morphologies of different m-types of pyramidal cells from layers 2 to 5. Artificially generated dendrites based on the topological profiles of the same types of pyramidal cells are illustrated in Figure 3.4B. The TNS algorithm generates cells that reproduce the characteristic shapes of the reconstructed morphologies of the different m-types of pyramidal cells.
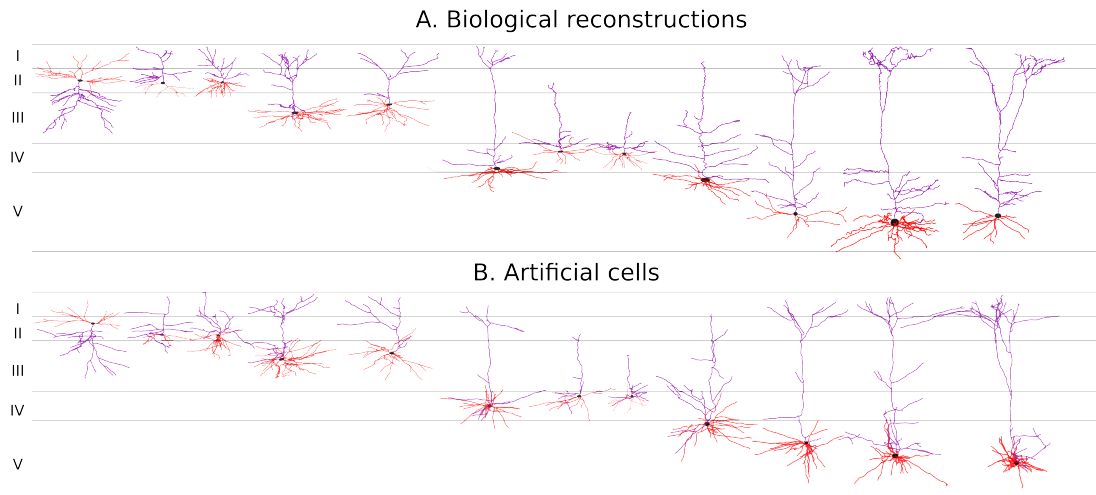
Fig. 3.4 Visual comparison of biological and virtual dendrites of rodent cortical layer 2 to 5 pyramidal cells. The biological reconstructions of dendrites of different pyramidal cell types (A) are compared to the corresponding artificial cells (B) of the same m-type.

In order to ensure that the topological profiles of the original reconstructions are reproduced by the TNS algorithm, the topological distances (Kanari et al. 2017) between the biological and the synthesized cells are computed and compared to the topological distances within the biological population. Note that if the exact TMD of the biological cells were recreated by the TNS algorithm, this distance would be zero. The mean topological distance among the dendrites of biological interneurons is $\approx 9,800$ while the maximum distance is $\approx 60,000$. Respectively, the mean distance between the dendrites of synthesized and biological cells is $\approx 9,000$ and the maximum distance is $\approx 57,000$ (See Figure C.6). Since we aim to reproduce the variability of the input morphologies, it is important that the topological profiles of the synthesized cells do not exactly match the biological ones, otherwise the artificially created neurons

will differ from their input cells only in local properties. This agreement of topological distances reflects the fact that the TMD captures the variability of the input biological cells.
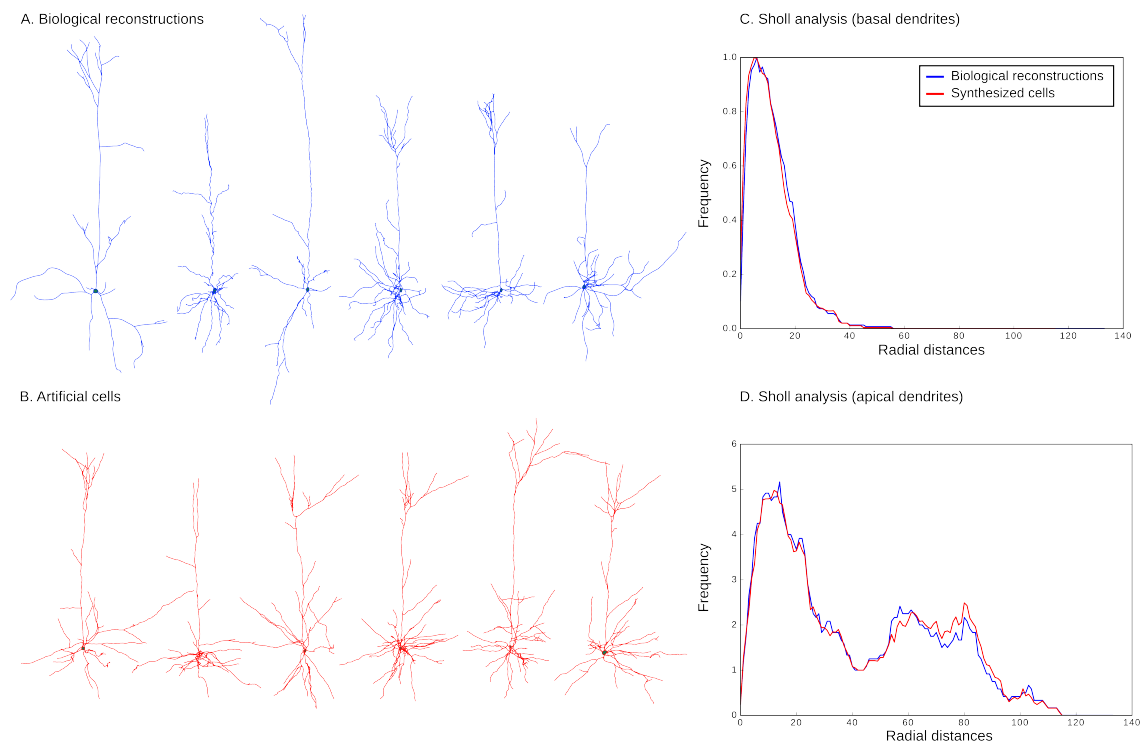
Fig. 3.5 Comparison of biological and virtual dendrites of L5-SPC cells. A selected sample of dendrites from biological reconstructions (A, in blue) are compared to the corresponding artificial dendrites of the same type (B, in red). The Sholl frequency is also presented for the two populations for the comparison of the basal dendrites (C) and the apical dendrites (D).

The unique structures of pyramidal cells require an m-type specific validation. Dendritic trees (basals and apicals) of an example population of the most representative morphological type (Layer 5 pyramidal cells) have been generated and validated against the corresponding biological datasets (Figures 3.5 to 3.8 respectively). The Sholl analysis of the basal dendrites (Fig 3.5C) and the apical dendrites (Fig 3.5D) illustrates the matching between the biological reconstructions and the artificially generated cells. However, since the Sholl analysis represents a projection of the TMD, which was used as input to the algorithm, more detailed validations are required in order to ensure the good quality of the synthesized cells. For this reason, the artificial cells of L5SPC are validated against the biological population using a large set of morphometrics. A sample of those morphometrics is presented in Figure 3.7 (for the basal dendrites) and in Figure 3.8 (for the apical dendrites): three global features (number of terminations, number of sections, section branch orders), three length-related features (total length

per neurite, section termination lengths, section bifurcation lengths) and three more
detailed features (remote bifurcation angles, section path distances, section radial
distances). The statistical distributions of the morphological features of the synthesized
cells closely match the distributions of the biological cells (see Appendix C: Distances
of morphometrics). Due to the exclusion of biological trees that contain less than three
sections from the synthesis input (see section F) the distributions that correspond
to single neurite morphometrics (number of terminations, sections and total length)
were computed for the filtered population rather that the whole set of biological
reconstructions.

The TNS algorithm generates artificial dendrites that reproduce the statistical
properties of the input reconstruction morphologies. An essential question is whether
accurate synthesis requires the correlations encoded in the TMD of trees, or whether the
independent bifurcation and termination probabilities suffice to describe and reproduce
the branching patterns of the neuronal morphologies, as in the mathematical models
proposed by Luczak (2006) and Cuntz (2010). To answer this question we generated
artificial cells that do not take into account the connectivity of the trees encoded
in their TMD profiles (See Appendix C: Synthesis without correlation of bifurcation
/ termination). Instead of using the joint probability distribution to bifurcate and
terminate, as encoded in the persistence barcodes of the biological dendrites, the
marginal probabilities were used. This method results in surprisingly poor quality
of synthesized cells (see Appendix C: Synthesis without correlation of bifurcation /
termination), indicating that the correlations encoded in the TMD of dendrites are
essential for the accurate generation of artificial cells.

Correlations between morphological features have been found to be important for
any synthesis method (Lopez-Cruz et al. 2011). However, an explicit description of
correlated morphometrics has either to be identified manually (Koene et al. 2009)
or optimized using complex automatic algorithms (Lopez-Cruz et al. 2011). The
manual identification of feature correlations is problematic, as different experts disagree
on the optimal set of features that describes neuronal morphologies (DeFelipe et al.
2013). The set of optimal morphometrics may also not generalize across different
m-types. On the other hand, the complicated machine learning techniques that infer
feature correlations have the risk of over-fitting when only a few biological examples
are available. In this case, instead of reproducing the biological principles of neuronal
morphologies, the algorithm will overestimate local properties and assume complicated
correlations that reproduce the noise of the input cells.
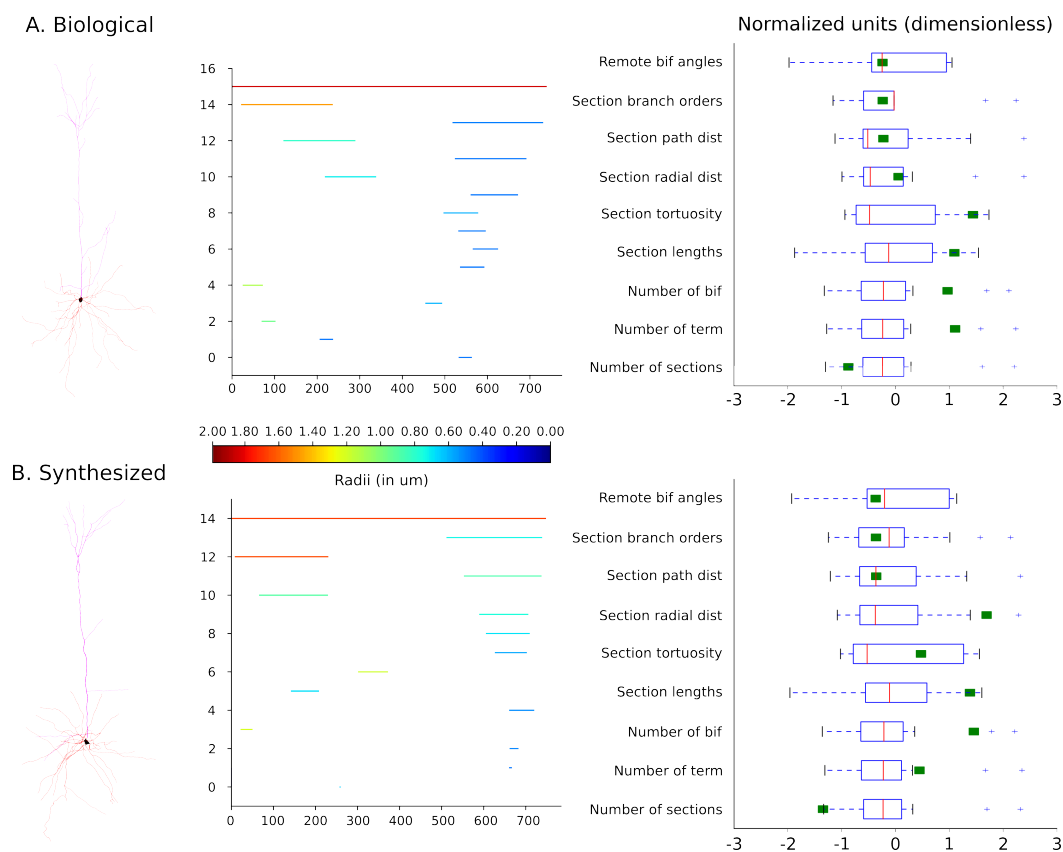
Fig. 3.6 Single-cell validation. A. A biological reconstruction of layer 5 pyramidal cell with accurate thickness. The characteristic persistence barcode of the apical dendrite is colored according to the thickness of the dendrite. Note that the thickness decreases with the distance from the soma. The mean values of the morphometrics of the biological cell (green square) are compared to the morphological features of the population of biological reconstructions (boxplots on top, right). B. An artificially generated layer 5 pyramidal cell with accurate thickness. The characteristic persistence barcode of the apical dendrite is colored according to the thickness of the dendrite that also decreases with the distance from the soma respecting this biological principle. The mean values of the morphometrics of the synthesized cell (green square) are compared to the morphological features of the population of biological reconstructions (boxplots on top, right). Note that all the features of the synthesized cell fall within the limits of the biological distributions.

By contrast, the TNS algorithm implicitly considers the correlations of individual features, as the TMD captures the topology along with the geometry of neuronal trees. The TNS method generalizes to different m-types without repeated optimization of the parameters that define the branching structure. The only parameters that need to be adjusted are the local parameters of targeting and randomness that define the tortuosity of sections. As a result, this algorithm is appropriate for the generation of numerous dendrites that will populate the digital reconstructions of large brain regions.
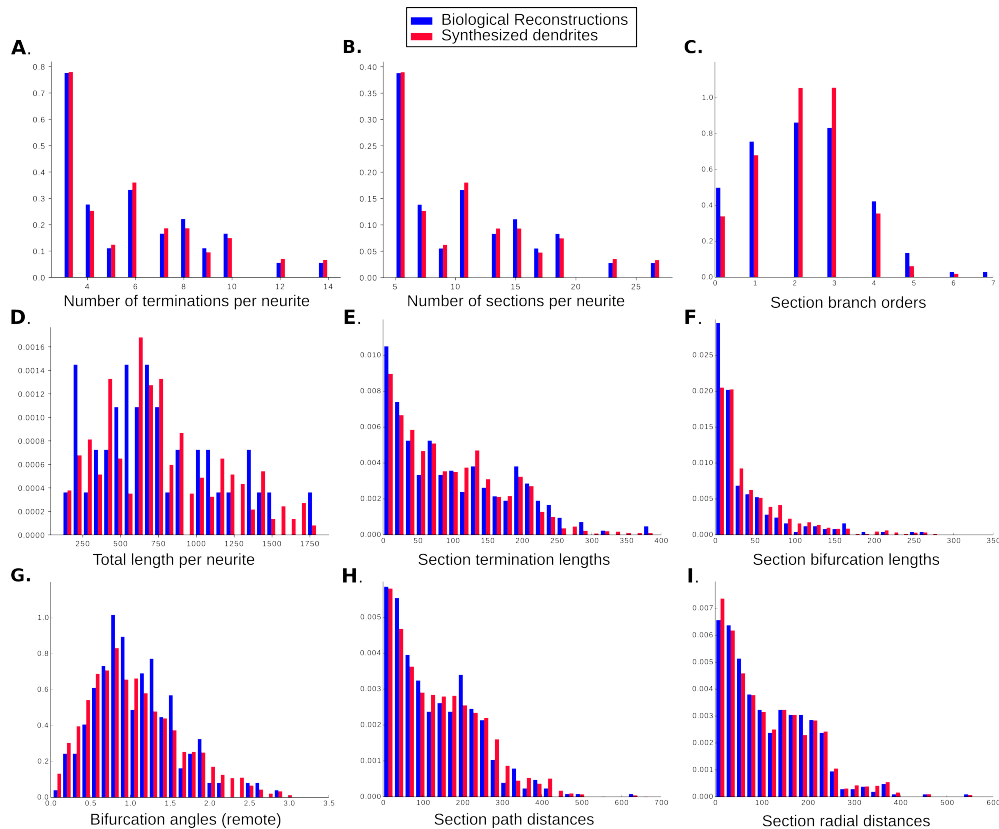
Fig. 3.7 Validation of a virtual population of L5_SPC cells (basal dendrites). The artificially generated population of Layer 5 SPC basal dendrites (red) is compared to the biological reconstructions of the same type (blue). A subset of the features used for the validation is presented: A. number of terminations, B. number of sections, C. section branch orders, D. total length per neurite, E. section termination lengths, F. section bifurcation lengths, G. remote bifurcation angles, H. section path distances, I. section radial distances.

Discussion

The generation of artificial neuronal morphologies that are statistically similar to a biological population of reconstructed neurons is a challenging problem that is nevertheless essential for increasing the variability of morphological shapes in digital reconstructions. There are two major problems faced in any attempt to generate artificial neurons. First, the complex biological growth mechanisms of neurons result in branching structures with highly correlated morphological features that are difficult to identify and reproduce. Second, the sparsity of data on many cell types in the cortex makes it difficult to recreate the biological diversity.

The topological profiles of cortical neurites were successfully used for the generation of artificial dendrites of a wide variety of cortical cell types. The artificially generated dendrites of interneurons and pyramidal cells accurately reproduced the morphometrics
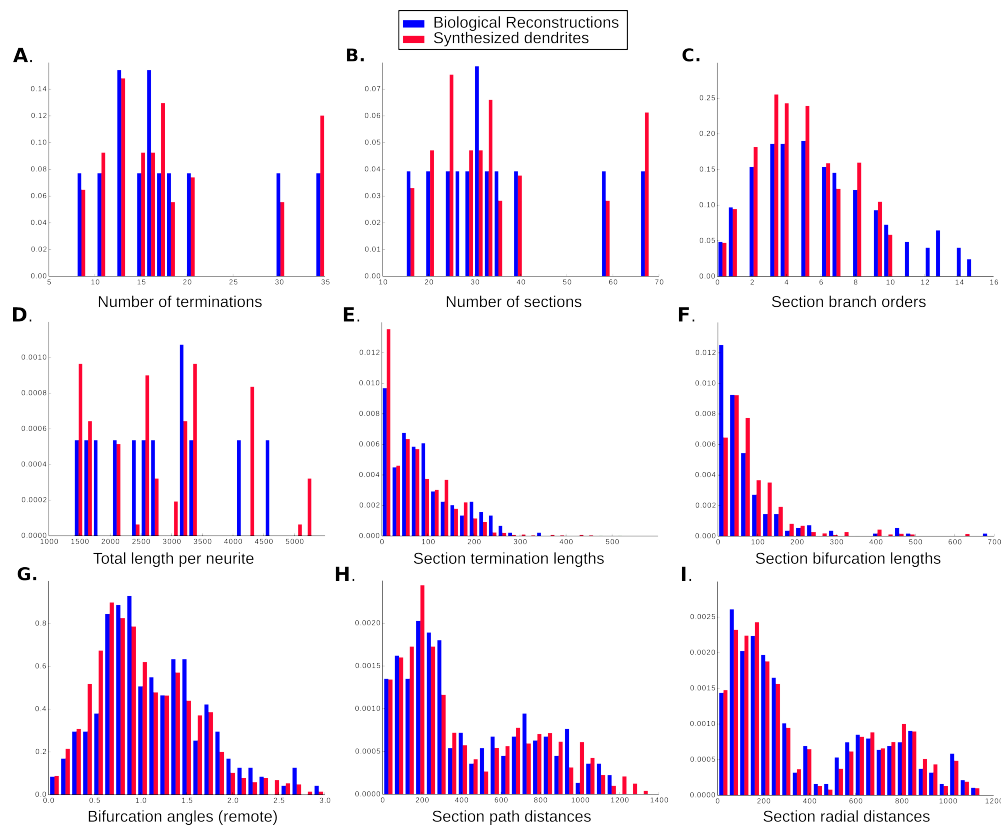
Fig. 3.8 Validation of a virtual population of L5_SPC cells (apical dendrites). The artificially generated population of Layer 5 SPC basal dendrites (red) is compared to the biological reconstructions of the same type (blue). A subset of the features used for the validation is presented: A. number of terminations, B. number of sections, C. section branch orders, D. total length per neurite, E. section termination lengths, F. section bifurcation lengths, G. remote bifurcation angles, H. section path distances, I. section radial distances.

and the topology of the original cells. The TMDs of dendritic trees thus retain sufficient information of the original structure to faithfully and effectively reproduce the statistical properties of neuronal morphologies, overcoming the need to explicitly describe the correlations between morphological features.

While the generation of dendritic trees is already an important step towards the digital reconstruction of more realistic brain networks, it is also crucial to generate artificial axonal trees. This problem is of particular interest for the computational modeling of brain networks for a number of reasons. First, because of their highly complex branching structures, the reconstruction of axons requires considerably more effort and time than dendrites. As a result only a small number of intact (not cut) axonal reconstructions are available. In addition, the branching structure of axonal morphologies is an essential determinant of the functionality of a network, as it provides the contact points between neurons and thus defines the connectivity of the network.

For the generation of accurate axonal shapes, the current synthesis scheme should be extended to take into account environmental influences. These include long-range targeting (Wang et al. 2015) and intersection avoidance between neighboring cells (Grueber and Sagasti 2010). This will improve the generation of neurites with complex branching patterns, such as cortical axons of both interneurons and pyramidal cells, glial cells and long range projecting cells, such as nigrostriatal dopaminergic neurons (Matsuda et al. 2009) and densely connected claustrum cells (Torgerson et al. 2015) and thus allow the digital reconstruction of brain areas that extend beyond a local region, such as the somatosensory cortex.

The small variability of the available biological reconstructions is the second major challenge faced in neuronal synthesis. Due to the high degree of similarity between the TMDs of neurons of a single morphological type, it is tempting to conjecture that a small number of biological examples of a cell type suffice to synthesize a large number of unique morphologies with the same characteristics. A stochastic algorithm to generate unique persistence barcodes that obey the principles of the biological diagrams from sparse biological examples would resolve the problem of the low variability of the synthesis input. An algorithm for the generation of persistence diagrams based on a single instance of an original dataset has in fact been proposed in (Adler et al. 2017) and could be used in order to increase the variability of the topological profiles of neurons when the available biological reconstructions are not sufficient. This technique exploits the power of randomness to generate unique copies of neuronal morphologies while ensuring that they all share the statistical properties of the original cells.

Discussion

Author contributions

L.K. conceived and implemented the topological synthesis algorithm and generated the synthesized populations. A.C. implemented and performed the statistical validations of the synthesized cells. G.A., J.G. and J.S. contributed to the development of the synthesis algorithms. K.H. contributed to the development of the topological synthesis and supervised this work. H.M. supervised the development of the biological synthesis at all stages of this work. All authors discussed the results, wrote the paper and commented on the manuscript at all stages.

Acknowledgments

References

Adler, R. J., Agami S., Pranav P. Modeling and replicating statistical topology, and evidence for cmb non-homogeneity. (2017).

Ascoli G.A., Krichmar J.L., Scorcioni R., Nasuto S.J., Senft S.L. Computer generation and quantitative morphometric analysis of virtual neurons. (2001). Anat Embryol (Berl). 204(4):283-301.

Ascoli, G.A., Donohue, D., & Halavi, M. Neuromorpho.org:a central resource for neuronal morphologies. (2007). Journal of Neuroscience, 27(35), 9247–51.

Aslangul C., Pottier N., Chvosta, P., Saint-James D., Directed random walk with spatially correlated random transfer rates (1993), Phys. Rev. E. (American Physical Society), 47(3), 1610-1617.

Bird A.D., Cuntz H. Optimal Current Transfer in Dendrites. (2016) van Rossum MCW, ed. PLoS Computational Biology. 12(5):e1004897.

Rembrandt Bakker R., Garcia-Amado M.,Evangelio M., Clasca F., Tiesinga P. Workflow, data format and tools to register neuron morphologies to a reference brain atlas (2017). 26th Annual Computational Neuroscience Meeting (CNS*2017): Part 3 (P271)

Burke R.E., Marks W.B., Ulfhake B. A parsimonious description of motoneuron dendritic morphology using computer simulation. (1992). J Neurosci. 12(6):2403-16. PubMed PMID: 1607948.

Chklovskii D.B., Synaptic Connectivity and Neuronal Morphology: Two Sides of the Same Coin, (2004). Neuron, Volume 43, Issue 5, Pages 609-617, ISSN 0896-6273, https://doi.org/10.1016/j.neuron.2004.08.012.

Conde-Sousa E., Szücs P., Peng H., Aguiar P. N3DFix: an Algorithm for Automatic Removal of Swelling Artifacts in Neuronal Reconstructions. (2017). Neuroinformatics. 15(1):5-12.

Cuntz H., Borst A., Segev I. Optimization principles of dendritic structure. (2007). Theoretical Biology & Medical Modelling. 4:21.

Cuntz H., Forstner F., Borst A., Häusser M. One Rule to Grow Them All: A General Theory of Neuronal Branching and Its Practical Application. (2010) Morrison A, ed. PLoS Computational Biology. 6(8):e1000877.

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

Egger R., Dercksen V.J., Udvary D., Hege H.C., Oberlaender M. Generation of dense statistical connectomes from sparse morphological data. (2014). Frontiers in Neuroanatomy Volume 8, ISSN 1662-5129.

Galton, F. and Watson, H. W. (1875). On the probability of the extinction of families. Journal of the Royal Anthropological Institute, 4, 138–144

Gonchar Y., Wang Q., Burkhalter A. Multiple distinct subtypes of GABAergic neurons in mouse visual cortex identified by triple immunostaining. (2008). Front Neuroanat. 28;1:3.

Graham, B.P., van Ooyen, A. Mathematical modelling and numerical simulation of the morphological development of neurons. (2006). BMC Neuroscience. 7(Suppl 1):S9.

Grueber W.B., Sagasti A. Self-avoidance and Tiling: Mechanisms of Dendrite and Axon Spacing. (2010). Cold Spring Harbor Perspectives in Biology. 2(9):a001750.

Haberl, M.G., Viana da Silva, S., Guest, J.M., Ginger, M., Ghanem, A., Mulle, C., Oberlaender, M., Conzelmann, K., Frick, A. An anterograde rabies virus vector for high-resolution large-scale reconstruction of 3D neuron morphology. (2015). Brain Structure and Function 220, 1369-1379.

Hillman D.E. Neuronal shape parameters and substructures as a basis of neuronal form. (1979). Schmitt FO, Worden FG, editors. The Neurosciences, 4th Study program. Cambridge: MIT Press. pp. 477–498.

Horikawa, K. Armstrong, W.E. A versatile means of intracellular labeling: injection of biocytin and its detection with avidin conjugates. (1988). J. Neurosci. Methods 25, 1–11.

Koene R.A., Tijms B., van Hees P., Postma F., de Ridder A., Ramakers G.J., van Pelt J., van Ooyen A. NETMORPH: a framework for the stochastic generation of large scale neuronal networks with realistic neuron morphologies. (2009). Neuroinformatics. 7(3):195-210.

Landau I.D., Egger R., Dercksen V.J., Oberlaender M. and Sompolinsky H. The Impact of Structural Heterogeneity on Excitation-Inhibition Balance in Cortical Networks (2016) . Neuron 92(5) 1106–1121.

Larkum M.E., Waters J., Sakmann B., Helmchen F. Dendritic Spikes in Apical Dendrites of Neocortical Layer 2/3 Pyramidal Neuron (2007) Journal of Neuroscience 27 (34) 8999-9008.

Larkum, M.E., Nevian T., et al. Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: a new unifying principle. (2009). Science 325(5941): 756-60.

Lefort S., Tomm C., Floyd Sarria J.C., Petersen C.C. The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex. (2009) Neuron. 29;61(2):301-16.

Lopez-Cruz P.L., Bielza C., Larrañaga P., Benavides-Piccione R., DeFelipe J. Models and simulation of 3D neuronal dendritic trees using Bayesian networks. (2011). Neuroinformatics. 9(4):347-69.

Luczak A. Spatial embedding of neuronal trees modeled by diffusive growth. (2006). J Neurosci Methods. Oct 15;157(1):132-41.

Luczak A. Measuring Neuronal Branching Patterns Using Model-Based Approach. Frontiers in Computational Neuroscience. 2010;4:135. doi:10.3389/fncom.2010.00135.

Markram H., Muller E., Ramaswamy S., Reimann M.W. et al., Reconstruction and Simulation of Neocortical Microcircuitry (2015). Cell, Volume 163, Issue 2, Pages 456-492, ISSN 0092-8674,

Marx M., Feldmeyer D. Morphology and Physiology of Excitatory Neurons in Layer 6b of the Somatosensory Rat Barrel Cortex. (2013). Cerebral Cortex (New York, NY). 23(12):2803-2817.

Matsuda W., Furuta T., Nakamura K. C., Hioki H., Fujiyama F., Arai R., Kaneko T. Single Nigrostriatal Dopaminergic Neurons Form Widely Spread and Highly Dense Axonal Arborizations in the Neostriatum. (2009). Journal of Neuroscience 29(2) 444-453

Memelli H., Torben-Nielsen B., Kozloski J. Self-referential forces are sufficient to explain different dendritic morphologies. (2013). Frontiers in Neuroinformatics. 7:1.

Pearson K. The Problem of the Random Walk. (1905). Nature 72(1865):294.

Peng H. Bioimage informatics: a new area of engineering biology (2008). Bioinformatics, Volume 24, Issue 17, Pages 1827–1836.

Peng H., Hawrylycz M., Roskams J., Hill S., Spruston N., Meijering E., Ascoli G.A., BigNeuron: Large-Scale 3D Neuron Reconstruction from Optical Microscopy Images (2015). Neuron, Volume 87, Issue 2, Pages 252-256, ISSN 0896-6273.

Ramaswamy S., Hill S.L., King J.G., Schürmann F., Wang Y., Markram H. Intrinsic morphological diversity of thick-tufted layer 5 pyramidal neurons ensures robust and invariant properties of in silico synaptic connections. (2012). J Physiol. 15;590(4):737-52.

Samsonovich A.V., Ascoli G.A. Statistical morphological analysis of hippocampal principal neurons indicates cell-specific repulsion of dendrites from their own cell. (2003). J Neurosci Res. 15;71(2):173-87.

Shillcock J.C., Hawrylycz M., Hill S., Peng H. Reconstructing the brain: from image stacks to neuron synthesis. (2016) Brain Inform. 3(4):205-209.

Spruston, N. Neuroscience: strength in numbers. (2008). Nature 452(7186): 420-1.

Torgerson C.M., Irimia A., Goh S.Y., Van Horn J.D. The DTI connectivity of the human claustrum. (2051). Hum Brain Mapp. 36(3):827-38.

van Elburg R.A., van Ooyen A. Impact of dendritic size and dendritic topology on burst firing in pyramidal cells. (2010). PLoS Comput Biol. 6(5):e1000781.

Wang D., He X., Zhao Z., et al. Whole-brain mapping of the direct inputs and axonal projections of POMC and AgRP neurons. (2015). Frontiers in Neuroanatomy. 9:40.

Wen Q., Stepanyants A., Elston G.N., Grosberg A.Y., Chklovskii D.B. Maximization of the connectivity repertoire as a statistical principle governing the shapes of dendritic arbors. (2009). Proc Natl Acad Sci U S A. 2106(30):12536-41.

Yi G.S., Wang J., Deng B., Wei X.L. Morphology controls how hippocampal CA1 pyramidal neuron responds to uniform electric fields: a biophysical modeling study. (2017). Sci Rep. 7(1):3210.

Zubler F., Douglas R. A Framework for Modeling the Growth and Development of Neurons and Networks. (2009) Frontiers in Computational Neuroscience. 3:25.

Network of random walks
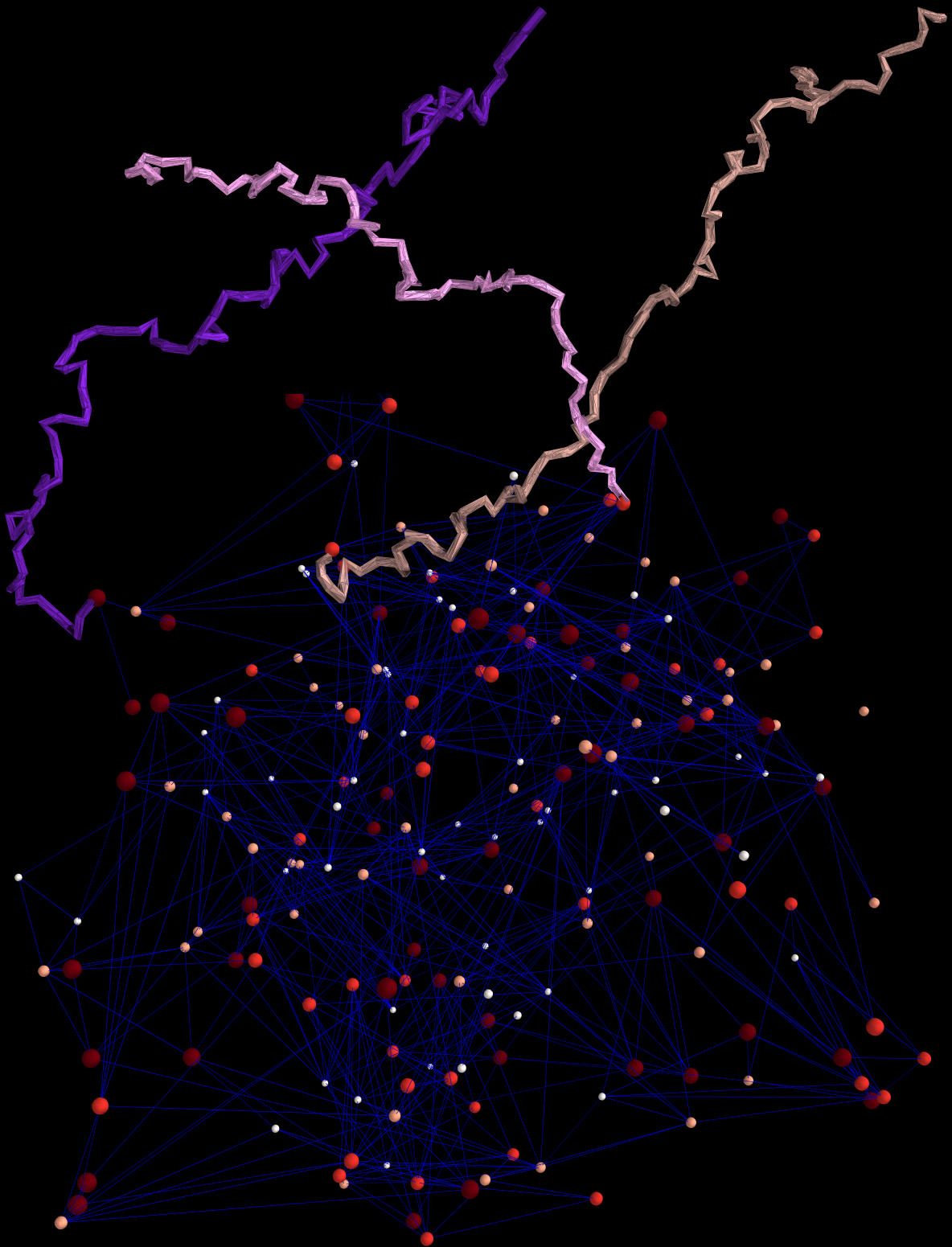
4

Modeling multi-neuron growth via random walks

> *"What we call chaos is just patterns we haven't recognized. What we call random is just patterns we can't decipher."*
>
> – Chuck Palahniuk, Survivor

4.1 Introduction

The rodent cerebral cortex is a complex ensemble of neuronal morphologies whose highly ramified arbors meander through $3D$ space. The physical contacts between two neurons are formed between the pre-synaptic axonal arbor and the post-synaptic dendritic span (Peters 1979). The shape of neuronal morphologies and their relative positions in space are two key geometrical factors that govern the formation of a physical connection, i.e., an apposition (Peters 1979, Kalisman et al. 2003). From these appositions between neurons the functional connections are formed, which establish the synaptic distribution of the brain. We define as a *neuronal network* the network that consists of morphologically detailed neurons, where the connectivity is defined by the physical appositions between them. Since the functionality of a neuronal network depends on its synaptic distribution, an accurate computational model of the brain should faithfully reproduce its synaptic connectivity.

A variety of novel imaging techniques (Shi and Toga 2017, Craddock et al. 2013), which are rapidly advancing, allow accurate mapping of the brain connectivity across different length scales. As a result, an increasing number of connectivity matrices of different species becoming available (Human Connectome Project) and new methods

have been developed to analyze them (Rubinov and Sporns 2010, Reimann et al. 2017). These connectivity matrices represent a snapshot of development at some point in the animal's life. It is still not clear how the interactions between neurons during the developmental process lead to the complex connectivity patterns that are observed. One of the fundamental questions is to what extent the structure of the neuronal network is encoded in the genetic information of an organism and to what extent the connectivity patterns stochastically emerge from interactions between growing structures.

Studying the common characteristics, or invariants, between the neuronal networks of different organisms can shed light on this question. Connectivity patterns within neuronal networks that are commonly observed between different individuals and species are usually considered to be the outcome of precise growth mechanisms (Graham and van Ooyen 2006) that shape the neuropil into functional brain networks. However, this approach underestimates stochastic processes that influence the neuronal growth. The contributions of randomness to the neuronal connectivity are to a large extent still unknown. These contributions cannot however be ignored, as natural patterns often occur as the sum of stochastic forces. For example, the random motion of particles suspended in a fluid generates trajectories that are statistically similar to each other. The average characteristics of these trajectories are precisely defined, even though they have been generated by random processes. Contrary to intuition, recent studies suggest that stochastic components of interacting morphologies contribute to the generation of invariants in the resulting networks, and therefore cannot be ignored (Tekin et al. 2016, Weigand et al. 2016).

Another approach to investigate neuronal network formation is to disentangle the roles of randomness and structure in artificially generated networks. The idea of generating random point-neuron networks (termed here as artificial neuronal networks, ANN) and studying their properties was introduced by Erdős and Renyi (1959). The Erdős-Renyi (ER) graphs have a fixed number of vertices, connected randomly by edges assigned between pairs of vertices with equal probabilities. A lot of variations to this basic model (Meghanathan 2015) have been proposed. In these models the connection probabilities of the random graphs are sampled from different statistical distributions. The generated ANNs do not consider the space in which the neurons are embedded, thereby disregarding the physical dimensions of the individual morphologies.

In this study, we propose a model that differs from the classical random ANN generative models in one important aspect: the neuronal network is generated by morphologies that are embedded in the $3D$ space. As a result, the actual paths that

connect two morphologies are preserved and the positions of their connections are known. The idea is based on the following question; how does the increasing structural complexity of individual morphologies affect the connectivity of the resulting neuronal network? This concept can be studied by starting from purely stochastic processes in $3D$ (i.e., random walks in space, Pearson 1905) and adding more and more constraints, increasing the complexity of the morphologies to make them more biologically accurate.

A simple mathematical model based on random walks is designed to study the effect of different interactions between growing morphologies on the neuronal network that they generate. The initial positions of a set of morphologies are placed within a $3D$ bounding box. Then the morphologies are grown in place according to a growth model described by a set of mathematical rules. The contact points of these morphologies are interpreted as appositions, or potential connections (Reimann et al. 2015). The neuronal network formed by these appositions represents the connectivity between the morphologies. A sequence of different growth models of increasing complexity are studied. Initially, the morphologies are simple paths (straight lines, random walks). Then, structural constraints are imposed on the morphologies to reproduce the morphological properties of biological reconstructions. The last step consists of simulating the interactions between the morphologies (avoidance, targeting, optimal path finding). The connectivity matrices that are generated from each growth model -of increasing complexity- are compared to the connectivity of the digital reconstruction of BBP (Markram et al. 2015).

Through this sequence of models, we are able to distinguish the connectivity patterns that can be reproduced by stochastic processes from these that also require complex interactions between biologically accurate morphologies. Interestingly, many structural and connectivity properties of biological systems can be reproduced by networks of simple random morphologies. Contrary to intuition, but in agreement with recent studies, we provide evidence in favor of the involvement of stochastic interactions to the generation of connectivity patterns that reproduce the local connectivity of biological neuronal networks. The complicated processes that take place in brain development involve a great variety of biochemical interactions that influence the shape of the neurons and the connectivity of the resulting neuronal network (Scott and Luo 2001). Since the proposed generative model represents a simplification of this highly elegant process, we cannot conclude that the same rules govern the actual growth of neurons. We can however, propose that basic principles, derived from the fundamental mathematical and physical properties of interacting morphologies, are crucial in the formation of these patterns.

4.2   Methods

In order to study the effect of the probabilistic interactions between growing neurons, a set of morphologies is generated within a cuboid domain (L_x, L_y, L_z) (Figure 4.1). The artificially generated network (AGN) consists of R morphologies each of which represents a neuron and has one or more emerging paths or neurites, that correspond to neuronal trees. The step size D is fixed and is used for the generation of points at a distance D from each other. Each morphology consists of T points in total and its total path length is $T * D$. The growth of each path starts from the root of the morphology that corresponds to the soma of a neuron. The direction of the next point in a path is given by a "growth model", and all the points of the morphologies are restricted to lie within the bounding box (L_x, L_y, L_z). Different growth models are illustrated in Figure 4.2 and include simple and non-intersecting random walks and simple branching tree structures. Since the morphologies start growing from the soma, R initial positions are distributed within the bounding box (see Figure 4.1B). From these R positions, the morphologies grow until they reach their target length T or until they stop at the boundaries of the domain. For periodic boundary conditions, the morphologies do not terminate once they reach the boundaries of the domain but continue their growth at the other side. For computational efficiency the domain is decomposed into a three dimensional grid of voxels.

The growth-model independent parameters used for the generation of the random walk networks are the dimensions of the bounding box (L_x, L_y, L_z), the number of morphologies R, the total length of each morphology T and the step size D. In order to use dimensionless parameters we define the step length as $D = 1\mu m$, and express all other lengths in units of D. When a branching growth model is selected, the number of sections, or branches, N_{sec} per neurite must also be defined. For the branching morphologies, there is also the possibility to select different numbers of steps for each section, in which case T is a list of size N_{sec}. The actual thickness of the morphologies is not considered in the simulations in order to reduce the computational cost and each morphology is registered in the grid voxels.

The paths of two morphologies that cross each other generate intersections that cannot physically occur in biological systems. In fact, the geometric intersections that are present in computationally generated networks introduce considerable artifacts such as the unrealistic correlations between the connections of the generated network. Morphologies that grow independently, i.e. not taking into account other paths, in the same space have a high probability to intersect with each other even at low densities. This probability increases with the morphology density. A number of growth models

prohibit computational intersections, so that only two paths at most can occupy the same voxel in space. In this case the two morphologies, whose paths pass through the same voxel, are considered to form a connection. The touch points between a pair of morphologies are defined by the voxels that contain more than one paths and are considered as potential connections, or "appositions".

A. Bounding Box   **B.** Soma Positions   **C.** Random walks   **D.** Connections
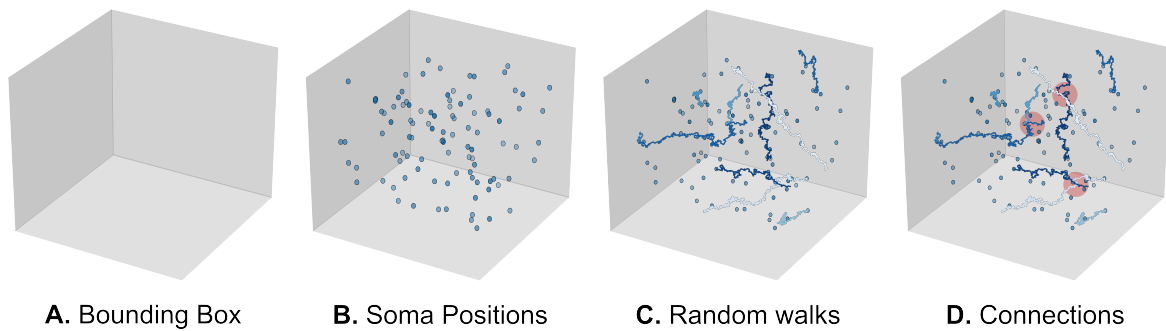
Fig. 4.1 Artificially generated network process. A) The bounding box defines the domain in which the morphologies grow. B) R points are distributed within the bounding box and define the roots of morphologies. C) Random paths are generated according to a set of mathematical rules until their target length T is reached. In this illustration ten directed random paths are presented. D) The appositions of the morphologies define the potential connections from which the structural connectivity matrix of the network is constructed.

The different growth models define the trajectory of each path of a morphology in $3D$ space (see Figure 4.2). A set of different growth-models of increasing complexity are defined:

1. Straight line walk (SLW)

2. Simple random walk (SRW)

3. Biased random walk (BRW)

4. Self avoiding random walk (SAW)

5. Directional self-avoiding random walk (DSAW)

6. Directional intersection-avoiding random walk (DIAW)

7. Star tree directional self-avoiding random walk (STRW)

8. Branching tree directional self-avoiding random walk (BTRW): symmetric or asymmetric

The simplest possible growth model is the generation of a morphology as a straight line (SLW). An initial direction is randomly sampled on the unit sphere and every point in the morphology follows the same direction. At the other extreme, the simple random walk model (SRW, Pearson 1905) selects the next point of a morphology in a random direction on the unit sphere independently of the previous directions. A random walk that is biased towards an initial direction (BRW) is a combination of the previous two growth-rules; the next point is computed as a weighted sum of the initial vector and a random vector. The weight of the initial direction is the "bias" of the random walk, and for normalization purposes the sum of the weights is one.

A. **S**traight **L**ine **W**alk

B. **S**imple **R**andom **W**alk

C. **B**iased **R**andom **W**alk

D. **S**elf-**A**voiding Random **W**alk

E. **D**irectional **S**elf-**A**voiding Random **W**alk

F. **D**irectional **I**ntersection-**A**voiding Random **W**alk

G. **S**tar **T**ree Directional Self-Avoiding **R**andom **W**alk

H. **B**ranching **T**ree Directional Self-Avoiding **R**andom **W**alk (symmetric branching)

I. **B**ranching **T**ree Directional Self-Avoiding **R**andom **W**alk (asymmetric branching)

Fig. 4.2 Growth models of morphologies of increasing complexity. Simple growth models include the straight line walk (A), the simple random walk (B) and the biased random walk (C). More complex models consider self-interactions, such as the self avoiding random walk (D) and the directional self-avoiding random walk (E) or interactions with other morphologies, such as the directional intersection-avoiding random walk (F). Models that generate branching structures include the star tree directional random walk (G) and the symmetric (H) or asymmetric (J) branching tree directional random walk.

Growth models of higher complexity include more advanced interactions with their own history or with other paths. A random walk that rejects new points that

would result in self-intersections is called a self-avoiding random walk (SAW). To avoid potential intersections the SAW re-samples a new random point. A directional self-avoiding walk (DSAW) is a combination of a SAW and a BRW. The next point is chosen randomly, as in the simple random walk, unless an intersection is detected. To avoid a potential intersection the DSAW grows towards the initial direction instead. Since the number of self-intersections for a random walk in $3D$ cannot be predicted exactly the "bias" of the DSAW can only be computationally approximated. A variation of this model is the directional intersection-avoiding random walk (DIAW). The next point of the DIAW is sampled randomly unless an intersection with its own or any other path is detected. It this case, the DIAW follows the initial direction to avoid intersections.

The models described so far generate morphologies containing a single path. These shapes are obviously simpler than the biologically observed shapes of neurons. For this reason, growth models of higher branching complexity are also studied. The first improvement to the previous models is the generation of multiple paths from the initial point of the morphology. In this case a star morphology is generated (SRW). A BTRW growth model implements a split of the morphology into branches to create a tree structure. A symmetric or an asymmetric tree can be generated, according to the mathematical rule used for the definition of the branching process. These growth-models are in fact a collection of many possible models, since many different values of asymmetry can be chosen.

The generation of R morphologies of T steps that grow independently has a computational cost of $O(R * T)$. For the generation of the same R morphologies when interactions between them are accounted for, each new point needs to check its distance from all the existing points. As a result the computational cost increases to $O(R^2 * T^2)$. In order to localize the intersection detection and reduce the computational cost, each new point of a path is registered into a voxel in space. The domain (L_x, L_y, L_z) is discretized with a grid of resolution similar to the step size D so that only the neighboring voxels need to be checked for intersections. The computational cost is reduced to $O(27 * R * T * T_{actual})$, where 27 is the number of neighboring voxels in a $3D$ regular lattice, and T_{actual} is the expected number of points contained in each voxel, which is $O(1)$. This way, the computational cost is reduced to linear on the total number of points $O(R * T)$. This is crucial to extend this growth scheme to large scale models of growth.

4.3   Analysis of artificially generated networks

For each growth model, a set of morphologies is generated and a number of measurements are computed to quantify the statistical properties of the model. We define as spatial observables the properties that describe the distribution of morphology points in space and as connectivity measurements the properties related to the connectivity of the resulting network. The spatial observables and the connectivity measurements are then compared to those of biological networks. Improvements of the growth model can then be proposed in order to generate more and more realistic networks.

The morphologies are generated point-by-point and registered in the grid voxels of the three dimensional space. Since all the information about the positions of each morphology are saved in the grid structure, spatial observables of the resulting network, such as the density, the pair correlation function and more advanced measures of the three dimensional morphology points in space (see Spatial observables) can be easily measured. For an appropriate selection of the parameters of the model (such as number and total length of morphologies) the generated network can be compared to the biological packing of neuronal processes, i.e., the neuropil.

The connectivity matrix of the generated network is computed from the touch points, i.e., appositions, of the morphologies, when two morphologies pass through the same voxel. Because of the density of connections in a brain region (1 connection per μm^3, Braitenberg and Schuz 1998, Kasthuri et al. 2015, Anton-Sanchez et al. 2014) the voxel size is chosen to be of the order of $1\mu m^3$, which also corresponds to the chosen growth step $D = 1\mu m$. The connectivity graph C of a network is an $R \times R$ matrix whose element C_{ij} is the *number* of connections between a pair of morphologies i, j. Since the morphologies do not take into account the differences between axonal or dendritic processes, the connectivity matrix is symmetric. The corresponding binary connectivity C_{bin} is defined by computing the presence or absence of a connection between two cells, without taking into account the number of connections.

From the connectivity matrix of a network a variety of measurements can be computed. These include standard graph measurements, such as the degree distribution, the shortest path length and the distribution of common neighbors (see Connectivity measurements, Rubinov and Sporns 2010); and topological measurements, such as the clustering properties and the distribution of cliques within the connectivity graph (Reimann et al. 2017). Biological data acquired through years of meticulous studies have revealed the structure of local cortical circuits in the rodent brain (Perin et al. 2011, Kasthuri et al. 2015). However, due to the effort required for the collection of this data, the information acquired from biological systems is limited to very small parts of

the brain (a cubic box of a few microns, Kasthuri et al. 2015). Due to these limitations, the connectivity measurements of the random walk models are also compared against detailed computational models that approximate the brain networks (Markram et al. 2015, Reimann et al. 2017).

The BBP circuit is reconstructed with a wide variety of morphological types of neurons that are placed within the cortical microcircuit (of $0.29mm^3$ volume). These morphologies correspond to biological reconstructions that are curated in order to correct experimental artifacts and cloned to increase the variability of the morphological diversity of the computational model. The structural connectivity is acquired from the appositions between morphologies. The structural appositions are transformed into functional connections with an algorithm developed in BBP (Reimann et al. 2015) that takes into account biological constraints. As a result, the simple random walk models cannot be compared to the final state of the BBP circuit connectivity. The artificially generated networks (AGN) based on different random walk models are compared to the initial state of the BBP circuit, i.e., the structural connectivity.

In order to acquire a dataset of homogeneous connectivity avoiding artifacts related to the boundary conditions and the size of the cortical column, a symmetric bounding box that is located in the middle of the column is examined. However, the computationally intensive calculations that are required for the collection of information from this very densely packed area, restrict us to work with a small bounding box of around $100\mu m \times 100\mu m \times 100\mu m$. For this reason the AGN models are also constructed in a domain of the same size.

Spatial observables

The spatial structure of both the morphologies and their connections are measured from the distribution of points in the gridded domain. A variety of measurements can be computed for a set of points embedded in a metric space. The relative positions of the points are studied with the distribution of distances between them and the pair correlation function. The spatial observables are compared to the properties of the neuropil, and the connectivity measurements to the connectome of a brain region. The combination of these measurements reveals interesting properties about the generated network that are discussed in detail in Section 4.4.

For a network of R morphologies of T points each, the expected density, i.e. the density excluding intersections of points in space, is the number of generated points divided by the total number of voxels in space.

$$Dens = \frac{RT}{L_x L_y L_z} \tag{4.1}$$

The observed density, however, is the number of voxels that are occupied by at least one path divided by the total number of voxels in the domain. As a result, the observed density depends on the number of paths that are contained within each voxel. For intersecting morphologies this density is lower than the expected density, since two or more paths can occupy the same voxel. Similarly, the density of connections in space is measured as the number of voxels that contain at least two paths divided by the total number of voxels in space.

The distribution of morphology points in space is described by their pair-correlation function, $g(r)$. This function measures the number of pairs separated by Euclidean distance r in space.

$$g(r) = \frac{1}{4\pi r^2} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{k \neq i}^{N} \delta(r - |r_k - r_i|) \tag{4.2}$$

The pair correlation function is the probability to find a point in space at a certain distance from another point. For homogeneous processes, the pair correlation function (PCF) converges to a constant value due to the regularity of the distribution of points in space.

We define as "failure rate" the number of morphologies that fail to grow without intersections within a domain of a fixed density. This measurement is not only an indication of the limitations of the computational method but also a fundamental property of packing objects in space. This is similar to the problem of packing spheres within a box (Sloane 2002), where the maximum density of spheres that can fit in the domain is determined

Connectivity measurements

An element c_{ij} in the connectivity graph C is the number of connections between the morphologies i and j. A variety of measurements that characterize the connectivity of the system can be computed from this graph.

The *connection probability* (Rubinov and Sporns 2010) is measured as the total number of connections divided by the maximum number of all possible connections in the graph which is $R(R-1)$ for an $R \times R$ connectivity matrix. The *degree* of a node in the graph counts the total number of connections that this morphology forms.

For a binary connectivity matrix C_{bin} the degree of a node i counts the number of morphologies that are connected to this node. The degree k_i of a node i is computed:

$$k_i = \sum_{\{j \leq R | j \neq i\}} c_{ij} \tag{4.3}$$

Another interesting measurement is the *number of connections N_c* between a pair of connected morphologies. This can be measured from the connectivity matrix C if the number of connections is encoded in the connectivity matrix. The *common neighbors CN_{ij}* of a pair of morphologies (Perin et al. 2011) are the morphologies that are connected to both of them. The number of common neighbors between a pair of neurons is a determinant of the connection probability between a pair of neurons P_c. Neurons that share more common neighbors are known to be connected with higher probabilities than neurons that do not share any common neighbors (Perin et al. 2011). To examine if this principle also holds for the model of random walks, the conditional connection probabilities are computed as a function of the number of common neighbors that a pair of morphologies shares.

$$P_c(\text{connected}|CN) = \sum_{j \neq i | \text{ share } CN_{ij}}^{R} \frac{\# \text{ connected cells}}{\# \text{ cells}} \tag{4.4}$$

The *shortest path distance s_{ij}*, (Rubinov and Sporns 2010) between two nodes i, j measures the minimum number of edges of the graph that need to be traversed to reach node j starting from node i. Because the connectivity matrix C is undirected, the direction of the shortest path distance is not relevant and hence $s_{ij} = s_{ji}$. The average shortest path distance L_s of the network reflects how well connected a graph is. The shortest path distances are small for a highly connected graph, and high for a sparsely connected graph. Note that the shortest path distance can be computed only for connected graphs.

The *betweenness centrality bc* (Rubinov and Sporns 2010) of a graph measures the distribution of nodes within the shortest paths of the network. A node with high betweenness centrality will have more control over the information flow as it participates in more shortest path connections. The betweenness centrality bc_v of a node v is defined as the fraction of all shortest paths in the network that pass through the given node v.

$$bc_v = \sum_{(i,j) \in C, \quad i \neq j \neq v} \frac{ns_{ij}(v)}{ns_{ij}}, \tag{4.5}$$

where ns_{ij} is the number of shortest paths between i, j and $ns_{ij}(V)$ is the number of shortest paths between i, j that pass through the node v.

The *clustering coefficient* (Rubinov and Sporns 2010) measures whether the nodes of a graph tend to cluster together or not. The local clustering coefficient CC_i of a node i is the proportion of existing connections between the neighbors of the node i divided by all the possible connections that they could form among themselves. For a node i of degree k_i the number of possible connections of its neighbors is $k_i(k_i - 1)/2$. Therefore the clustering coefficient CC_i of a node i is:

$$CC_i = \frac{2}{k_i(k_i - 1)} \sum_{k,l: \text{ neighbors of node } i} c_{kl} \qquad (4.6)$$

A *small world network* (Watts and Strogatz 1998) is a graph that is highly clustered but has a small average shortest path length. Small world networks are characterized by few long range connections, while the local clustering still remains high because neighbors tend to connect to each other. The small worldness σ of a graph is computed from the ratio of the average clustering coefficient CC and the average shortest path L_s of a graph:

$$\sigma = \frac{CC/CC_r}{L_s/L_r}, \qquad (4.7)$$

where CC_r and L_r are the average clustering coefficient and the average shortest path length of a random graph. If $\sigma > 1$ the network is considered to be a "small-world" network, so this practically requires $CC \gg CC_r$ and $L_s \approx L_r$.

4.4   Results

The impact of soma positioning and branching

The first step is to position the somata of the morphologies in space. To explore the effect of the soma positioning on the connectivity of a generated network, we have constructed two types of artificially generated networks with the same parameters ($LX = LY = LZ = 200, R = 125, T = \{100, \ldots, 8000\}$). The first type of networks were generated from regularly positioned roots of morphologies on a three-dimensional lattice (REG-network) (Figure 4.3A). The second type of networks were generated from randomly positioned roots of morphologies (RAND-network) in the three-dimensional space of the bounding box (Figure 4.3B). The morphologies were initially simulated by the DSAW growth model. This experiment did not reveal any statistically significant

structural differences between the number of connections formed within the REG and RAND networks, despite the significantly different effect of the two models on the spatial distribution of their points in space (Figure 4.3). Then, each morphology was generated by STRW growth model with six independent paths. The RAND-network of STRW morphologies generated significantly more connections compared to the regular network with the same properties (Figure 4.3). The results presented in Figure 4.3 correspond to the statistical averages and the standard deviations from 50 repetitions.

For a network of STRW morphologies of total length $T = 1600$, which corresponds to a density of morphology points $\approx 2.5\%$, the number of connections increases from $\approx 5,000$ for the REG-network to $\approx 8,500$ for the RAND-network. The corresponding AGN with the same input parameters generated from DSAW morphologies resulted in $\approx 2,000$ connections. In addition, the mean degree of the STRW network's nodes increases from 21 for the REG-network to 31 for the RAND-network, while the corresponding DSAW networks (REG and RAND) had a degree of ≈ 12. The average clustering coefficient of the STRW networks increases from 0.31 for the REG-network to 0.37 for the RAND-network, while both DSAW networks have a clustering coefficient around 0.12. The shortest path distance of the STRW networks decreases from 1.95 for the REG-network to 1.75 for the RAND-network, while the DSAW networks have shortest path distances around 2.2.

The previous results demonstrate that the random positioning of the somata of STRW morphologies generates more connections in space, but also increases the mean number of morphologies that each cell is connected to, , while the single path morphologies generate consistently a smaller number of connections. The RAND-network is not only more connected but also the connections within the network are more clustered compared to the REG-network for the branching random walks. For further evaluation of this result, we have also generated two types of REG and RAND networks of (STRW and DSAW) random walks for different morphology densities ($T = 8000$), using the same properties (somata positions, number of cells, box size) that were used for the generation of the DSAW and STRW networks.

Indeed for a network of STRW morphologies of total length $T = 8000$, which corresponds to a density of morphology points $\approx 11\%$, the number of connections increases from $\approx 55,000$ for the REG-network to $\approx 57,000$ for the RAND-network. The corresponding AGN with the same input parameters generated from DSAW morphologies resulted in $\approx 53,000$ connections. The differences in the number of connections of the networks are much smaller compared to the lower density networks but they are statistically significant. However, the connectivity properties of the

A1. Regular somata   **A2.** Morphologies   **A3.** Regular connections

B1. Random somata   **B2.** Morphologies   **B3.** Disordered connections

C. Statistical results
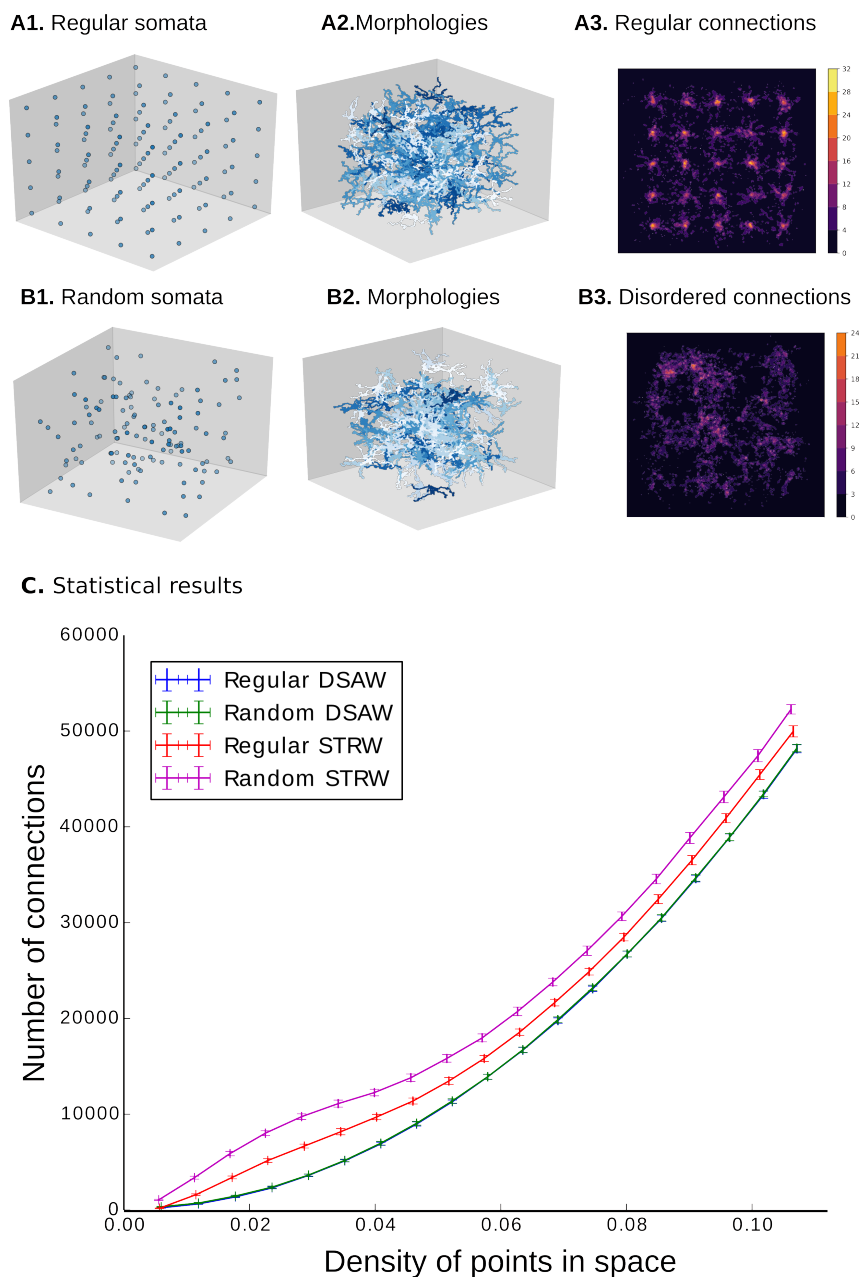
Fig. 4.3 The effect of soma positioning and branching on the network connectivity. A. Regular soma positions are distributed on a three-dimensional lattice (example shown in A1) and two types of morphologies, DSAW (not shown) and STRW (example shown in A2), are generated from these initial positions. Both morphological types generate a distribution of points in space that corresponds to a regular connectivity in space (STRW, A3). B. Random placement of soma positions (B1) for the same two growth models (STRW, B2) generates disordered connections in space (STRW, B3). C. The number of connections as a function of the density of morphology points in space for regular DSAW (blue), random DSAW (green), regular STRW (red) and random STRW (purple) illustrates that STRW form more connections than the DSAW for the same densities, and that the randomly positioned STRW form more connections compared to the regularly positioned STRW. The results shown correspond to the average curves and the standard deviations of 50 repetitions of artificially generated neuronal networks.

resulting networks are not statistically distinguishable. The average degree of all networks is around $112 - 114$, the clustering coefficient around $0.91 - 0.92$ and the shortest path distance around $1.06 - 1.08$. This result indicates that for highly connected networks 50% the effect of somata positioning and branching is not as effective.

As a result, a natural mechanism to take advantage of the properties of randomness in space, in order to increase the number of connections and the degree of connectivity is to generate multiple branches from the root of the morphology. In this case, for the same wiring length, the local space around a morphology is optimally sampled in all directions and more connections with a larger number of cells are generated. For sparsely connected networks this approach also generates more clustered networks with shortest path distances between their nodes.

Statistical properties of intersections

To study if the intersection avoidance has a significant contribution on the connectivity of computational models of detailed neuronal networks, the effect of intersections on the spatial distribution of structural connections is examined. The frequency of intersections between the structural connections of the BBP circuit (Markram et al. 2015) is measured by registering the connections in a gridded domain and counting the number of connections that belong to each voxel. In a $80\mu m \times 80\mu m \times 80\mu m$ domain of which 80% is occupied by neuronal processes (60% axons, 20% dendrites), about $400,000$ appositions are generated between axons and dendrites, $250,000$ among dendrites and $450,000$ among axons. From those, 24% of the dendritic appositions intersect with each other, i.e., are registered at the same voxel, which corresponds to 6% of the voxels being occupied by more that one connections. Similarly, 26% of the axon-axonic appositions intersect with each other, which corresponds to 12% of the voxels being occupied by more that one connection. Respectively, 38% of the axon-dendritic appositions intersects with each other, which corresponds to 16% of the voxels being occupied by more that one connections. As a result, the intersections among the connections of the network are too frequent to be ignored, and the impact of the intersections on the resulting connectivity needs to be evaluated.

The intersections of morphologies in a computational model can either be biologically relevant or represent an artifact arising from the placement of morphologies in space. To test if the intersections in the BBP circuit can be reproduced by random morphologies, we compared the intersections present in the BBP circuit to the intersections of an artificially generated network (AGN) of DSAW morphologies. A low density (20%) AGN of DSAWs reproduces the appositions between dendrites and a high density (60%)
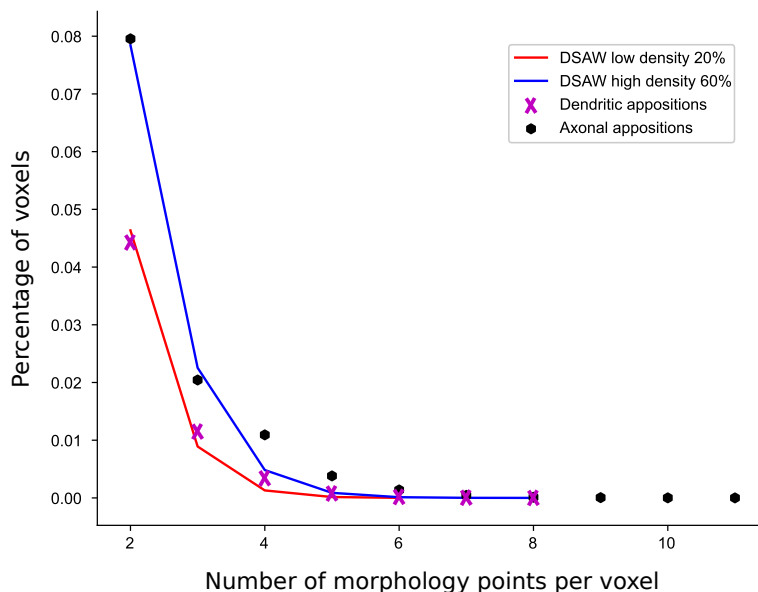
Fig. 4.4 Comparison of appositions in space. The percentage of voxels that contain a certain number of morphology points is presented for the BBP model of dendrites (purple) and axons (black). This property is compared to networks of artificially generated morphologies based on the DSAW growth model of similar densities (red, 20% to simulate dendrites and blue, 60% density to simulate axons).

AGN of DSAWs reproduced the appositions between axons, as illustrated in Figure 4.4. The percentage of voxels that contain a given number of connections of low density DSAW morphologies matches the distribution of dendritic connections, while the curve for the high density DSAW morphologies follow the distribution of axonal connections. In both cases, the exact BBP distributions are not reproduced precisely, but they are very close. It is possible that the inconsistency of the AGN and BBP curves is introduced by the difference of the algorithms that calculate the appositions. In the AGN, appositions are counted from the paths that are registered in the same voxel, while the BBP circuit appositions are measured from the morphologies whose paths come closer than a few microns. It is also interesting to note that for both axonal and dendritic intersections the maximum number of morphology points contained in a voxel are much higher than their equivalent random walk models. This effect may be caused by the duplication of the same morphologies in the digital reconstruction, but this artifact occurs very rarely.

Biological reconstructions were also registered in the bounding box, starting from random positions for their roots, until the selected density was reached. The "growth model" that corresponds to biological reconstruction morphologies registered in space will be named "rNeurons", for random neurons. This is useful for the comparison of
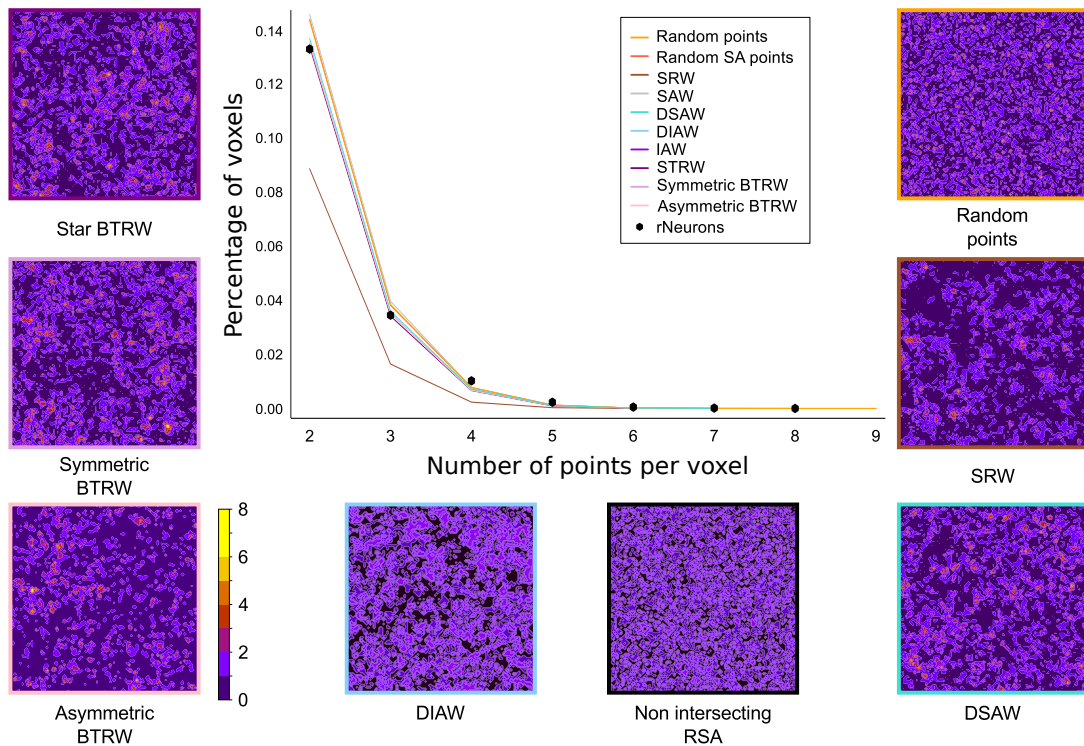
Fig. 4.5 Comparison of distribution of points per voxel for different growth models. The percentage of voxels that contain a certain number of morphology points is presented for different growth rules. The DSAW growth model is the simplest model that can reproduce the properties of neuronal morphologies. A $x - y$ slice of the corresponding network, where the colormap represents the number of points per voxel, is presented for a selected number of growth models. The non-intersecting model of DSAW generates a distribution of points in space that is similar to the three-dimensional distribution of connections in space, according to Anton-Sanchez et al. (2014).

the AGN intersection peaks to the neuronal ones computed with the same method of intersection detection. For this experiment the target density of occupied positions in space was chosen to be 80%. The intersection peaks for a number of different growth models are presented in Figure 4.5. The DSAW growth model is the simplest model that can reproduce the distribution of neuronal intersections in space. Random walks have consistently lower numbers of voxels with a certain number of points compared to all other models. On the contrary the random points in space have higher numbers of voxels with a certain number of points, compared to all other models. Interestingly, more complex branching patterns have intersection peaks that are very similar to the DSAW model and the rNeurons.

The numbers of intersections per voxel are presented in a slice of the corresponding box for different growth models. Random points appear to be distributed homogeneously in space, while SRW, DSAW, STRW and BTRW show a correlation of intersection peaks

in space (for the detailed analysis of correlations see Spatial correlations of connectivity). Due to the continuity of paths in space, the voxels that contain the maximum number of points, named peaks, are surrounded by voxels which also contain a large number of points. As a result, this peak-effect would make any attempt to remove the intersections after the construction of the circuit computationally intractable.

In a recent paper (Anton-Sanchez et al. 2014) found that the distribution of neuronal synapses in space is reproduced by a random sequential adsorption (RSA) without intersections. This model generates a homogeneous distribution of synapses in space, which captures the properties of the three dimensional distribution of cortical synapses. Even though the quantitative features of this model will be difficult to reproduce with a set of random walk morphologies, the qualitative properties of this model are effectively recreated by the DIAW growth model. The intersection avoidance between paths, that is used in the growth models such as the DIAW, seems to be sufficient for the generation of a spatial distribution of connections that is in agreement with the model proposed by Anton-Sanchez et al. (2014). Intersection avoidance and tiling are in fact mechanisms that are biologically important for the generation of specific connectivity patterns in visual cortex and in networks of glia cells (Grueber et al. 2010).

Limitations of non-intersecting growth models

The constraint of intersection avoidance is sufficient to approximate the spatial distribution of synapses in space proposed in the phenomenological model of Anton-Sanchez et al. (2014). A natural question is whether this model can generate morphologies that achieve the high density of the brain tissue. The mean density of matter in cortical areas is estimated to be around 60% (Braitenberg and Schuz 1998). However, the local density can be higher in dense areas, such as layers 4 and 5. To check if morphologies that avoid intersections can grow up to the high densities that are observed in biology, the failure rate of a non-intersecting DIAW is measured. Multiple simple random walk networks are generated for different densities (from 0% to 100%, see Figure 4.6). For each network, the failure rate is measured as the percentage of DIAW morphologies that fail to grow to their full length (400 for a box of size $100 \times 100 \times 100$).

As expected, the mean failure rate is zero for very low densities and increases to one for densities above 90% where it is not possible any more to grow non-intersecting structures. However, it is important to note that almost no morphologies fail to grow for densities below 50%, while even at the biological densities $60 - 70\%$ only a small fraction of morphologies $10 - 20\%$ fail to grow their full length. This result suggests
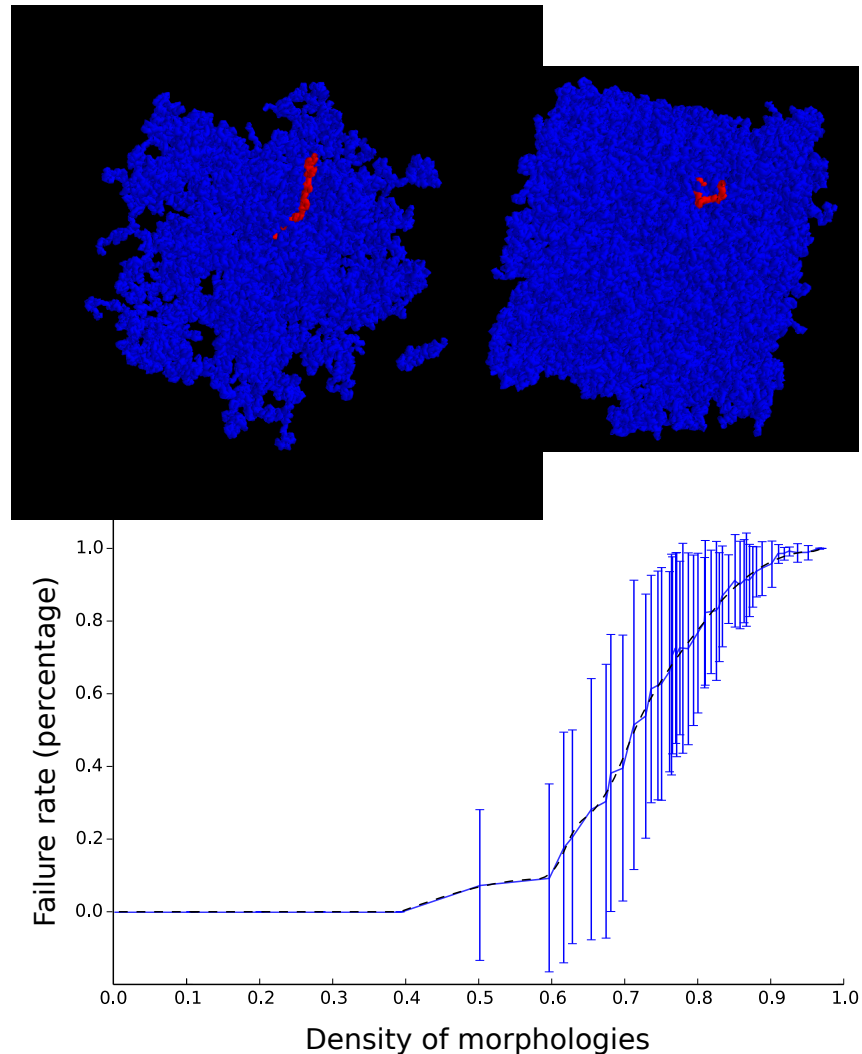
Fig. 4.6 Distributions of failed attempts to grow. A. An example of 20% dense space of random walks, within which a non-intersecting path can fully grow (in red). B. An example of 80% dense space of random walks, within which a non-intersecting path is blocked, due to the difficulty to find non occupied positions in space to grow to (in red). C. The failure rate of non-intersecting morphology growth, as a function of the density of points in space reveals that random morphologies can grow up to the biological densities $60 - 70\%$ without significant failure rate.

that even a simple rule of intersection avoidance can generate dense morphological systems. Therefore, it is possible to generate intersection-avoiding morphologically detailed neuronal networks that are as dense as observed biological systems (Grueber et al. 2010).
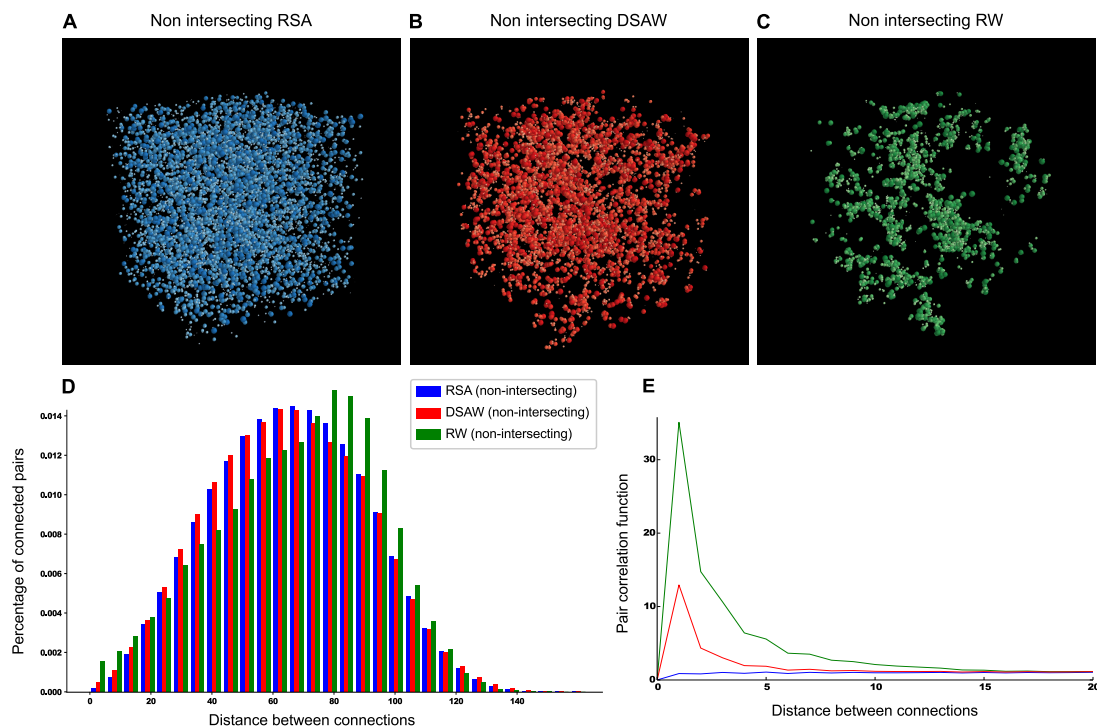
Fig. 4.7 Distributions of appositions in space. The spatial distribution of appositions in space are presented for three different growth models of non-intersecting morphologies: (A) random points in space, (B) directional self avoiding random walks and (C) random walks. The distribution of distances between the appositions of the three models (D) shows that DSAW appositions are similar to the RSA. The pair correlation function of appositions (E) indicates that the DSAW model is a mixture of the random walk and the phenomenological models. For short length scales ($< 5\mu m$) the DSAW connections are highly correlated, while this correlation disappears for longer length scales ($> 5\mu m$).

Spatial correlation of connections

The previous statistical analysis revealed that the three-dimensional distribution of connections of a non-intersecting set of DIAW morphologies is similar to the phenomenological model, proposed by Anton-Sanchez et al. (2014), that approximates the spatial distribution of cortical synapses. In order to quantify this result, the pair correlation function of connections in space is computed, which quantifies the probability to find a connection at a certain distance from another one in space. Three types of non-intersecting morphological systems were generated for this study: a set of random points, a set of random walks and a set of directional self avoiding random walks (Figure 4.7). The roots of the morphologies were randomly distributed in space for all the growth models.

The phenomenological model of synaptic distribution corresponds to a random distribution of points in space with intersection avoidance, as proposed by Anton-
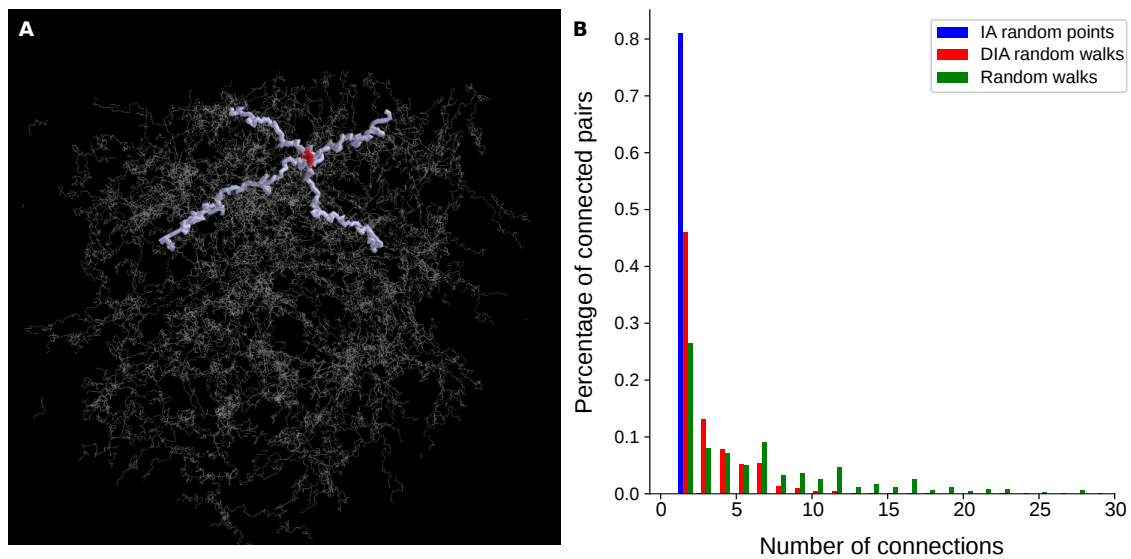
Fig. 4.8 Clustering of connections between DIAW morphologies. A. Example of a connected pair of morphologies (bold white), within a network of non-intersecting DIAW morphologies (light gray). The two morphologies share a large number of connections, clustered in space (in red). B. The number of connections between a pair of connected morphologies is low (≈ 1) for the random points model and goes up to large values (≈ 30) for the network of random walks. The DIAW morphologies have approximately between $1-10$ connections.

Sanchez et al. (2014) and can be reproduced by the DSAW model for large length scales. The resulting networks of the RSA and DSAW are not only visually similar (Figure 4.7A. blue, B. red respectively) compared to the random walk network (Figure 4.7C, green), but also have the same distribution of distances among the positions of their connection (Figure 4.7D). Interestingly the pair correlation function (see equation 4.2) of appositions (Figure 4.7E) indicates that the DSAW model behaves as intermediate between the random walk and the phenomenological models. For short length scales ($< 5\mu m$) the DSAW connections are highly correlated, while this correlation disappears for longer length scales ($> 5\mu m$).

The spatial correlation of connections between a pair of DIAW morphologies is a fundamental property that arises from the continuity of the morphological paths in the embedded space. In addition, the clustering of connections in space is supported by a number of fundamental biological observations. First, the number of connections between a pair of connected neurons ranges from one to multiple connections (3-8 synapses, Fauth et al. 2015), a property that is important for the functionality of brain networks. In fact the probability for a pair of connected morphologies to have more than one connection is significantly large for morphologies in cortical brain regions compared to equivalent random networks (Kasthuri et al. 2015, Markram et al. 2015).

The high "clustering" of connectivity in biological networks is present even in networks generated within a set of cultured neurons (Downes et al. 2012).

The spatial correlation of synapses is also in agreement with the optimization of information processing by dendrites as proposed in the computational model of Poirazi and Mel (2001). This computational model (Poirazi and Mel, 2001) of structural plasticity suggests that non-linear integration of signals by dendrites supports larger storage capacities. Dendritic branches that act as neuronal subunits, in which connections are clustered together in space, integrate signals optimally. The mechanism they propose can be used by a structural learning rule that combines random synapse formation with activity-dependent stabilization or elimination. This result supports the existence of spatial clustering of connections as a mechanism for learning optimization.

Therefore, the spatial correlation of connections seems to play be important for the functionality of dendrites. Consequently, even though this effect is not supported by the RSA model, there is strong biological and computational evidence in favor of this property. The spatial clustering of local connections that arises due to the continuity of paths in space, could indeed be a useful property that biological networks take advantage of during their development.

Connection probability depends on the number of common neighbors

Once the statistical and spatial properties of connections are approximated by a computational model, the connectivity graph, which encodes the connection probability between any pair of morphologies, need to be examined. A recent study (Perin et al. 2011) revealed one of the fundamental principles of cortical networks; the connection probability and the synaptic weights of cortical neurons depend on the number of common neighbors. A pair of neurons has a higher probability to be connected, even for higher intersomatic distances, if they have at least one common neighbor. This probability also increases further with the number of common neighbors that they share. It is interesting to note that Erdős-Renyi networks do not reproduce the common neighbor effect of biological systems, even when modified accordingly to be distant dependent.

However, the common-neighbor effect is reproduced by several AGN growth models. For this experiment, sparse networks of the same morphological density (6%) were generated and the connection probability was computed as a function of the number of common neighbors (see equation 4.6). For random walks this probability increases
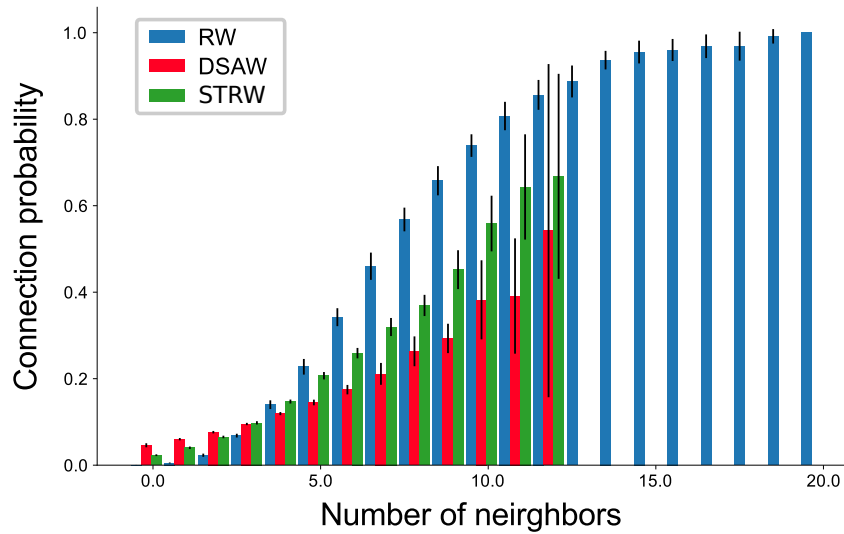
Fig. 4.9 Common neighbor effect. Connection probability between two cells depends on the number of common neighbors that they share. For different types of AGN (random walks: blue, directional self avoiding walks: red and branching DSAWs: green) the connection probability increases with the number of common neighbors.

quickly and reaches up to one. This property can be explained by the local high density of random walk networks, which also results in pairs of connected morphologies that share a large number of connections (see also Figure 4.7). Similarly, the connection probability of DSAW and STRW networks also increase with the number of common neighbors. However, the connection probabilities remain much lower in the case of DSAW and STRW morphologies, as they do not posses such a high local density of branches.

A number of connectivity measurements (number of connections, mean degree, average shortest path distance and average clustering coefficient) of these three growth models (RW, DSAW, STRW) are compared to the connectivity properties of a random Erdős Renyi graph (ER, Erdős and Renyi 1959). For the same connection probability (number of nodes and number of edges preserved) the DSAW have $\approx +20\%$ higher clustering coefficient (0.120 ± 0.002) compared to ER random networks (0.095 ± 0.002). The STRW networks have even higher ($\approx +30\%$) clustering coefficient (0.142 ± 0.003) compared to the ER random networks. This indicates that small world networks can be generated by DSAW and STRW morphologies. Those morphologies generate not only spatially clustered connections, but also highly clustered connectivity networks.
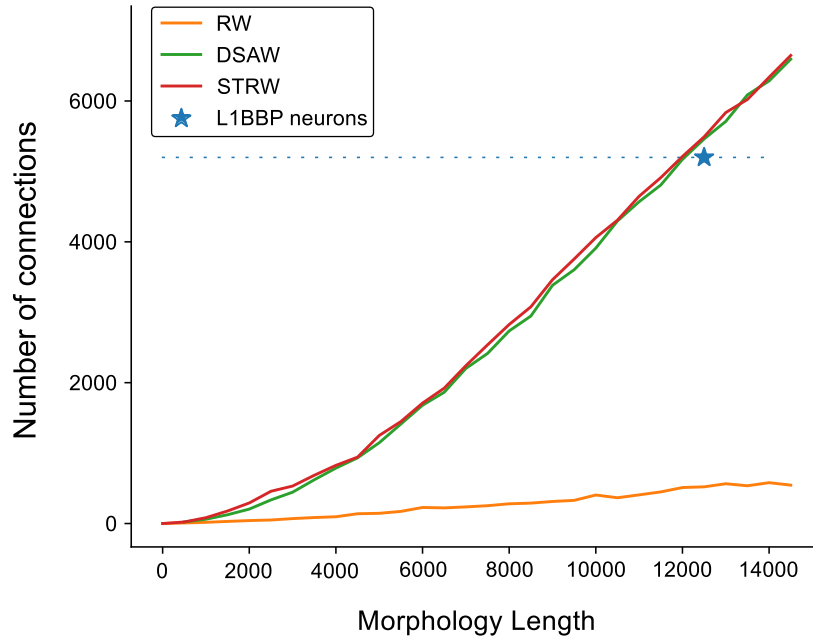
Fig. 4.10 Number of connections as a function of morphology length. For different types of AGN (RW: orange, DSAW: green and STRW: red). The equivalent biological mean length is presented with a blue star, and the equivalent dotted line is shown for guidance to the eye.

Connectivity of different growth models

In section "The impact of soma positioning and branching" we demonstrated the significance of the placement of the morphologies' roots for the connectivity of a network. In order to compare the connectivity of artificially generated networks to the connectivity of the BBP circuit, the original positions of the BBP morphologies should be used as the roots of the artificial morphologies. For this study, we used a small subset of the densely connected structural BBP network, which corresponds to the connectivity of neurons in the superficial cortical layer (Layer 1).

First, we examined the dependence of the number of connections in the connectivity matrix C as a function of the morphological length T. The RW cannot reproduce the properties of the BBP circuit, as the connection probability remains very low, due to the localization of the morphologies around the root, which in turn minimizes the radial extents of the RW morphologies. The RW network requires more morphologies or morphologies of higher lengths to reach the same morphology density and therefore the same number of connections. The DSAW and the STRW reach the BBP number of connections for a total length that is close to the mean length of axonal trees ($\approx 12500 \mu m$) of biological cells in Layer 1 (Figure 4.10).

A. BBP circuit

B. Non-intersecting STRW circuit

C. Degree of nodes

D. Shortest path length

E. Betweenness centrality

F. Clustering coefficient
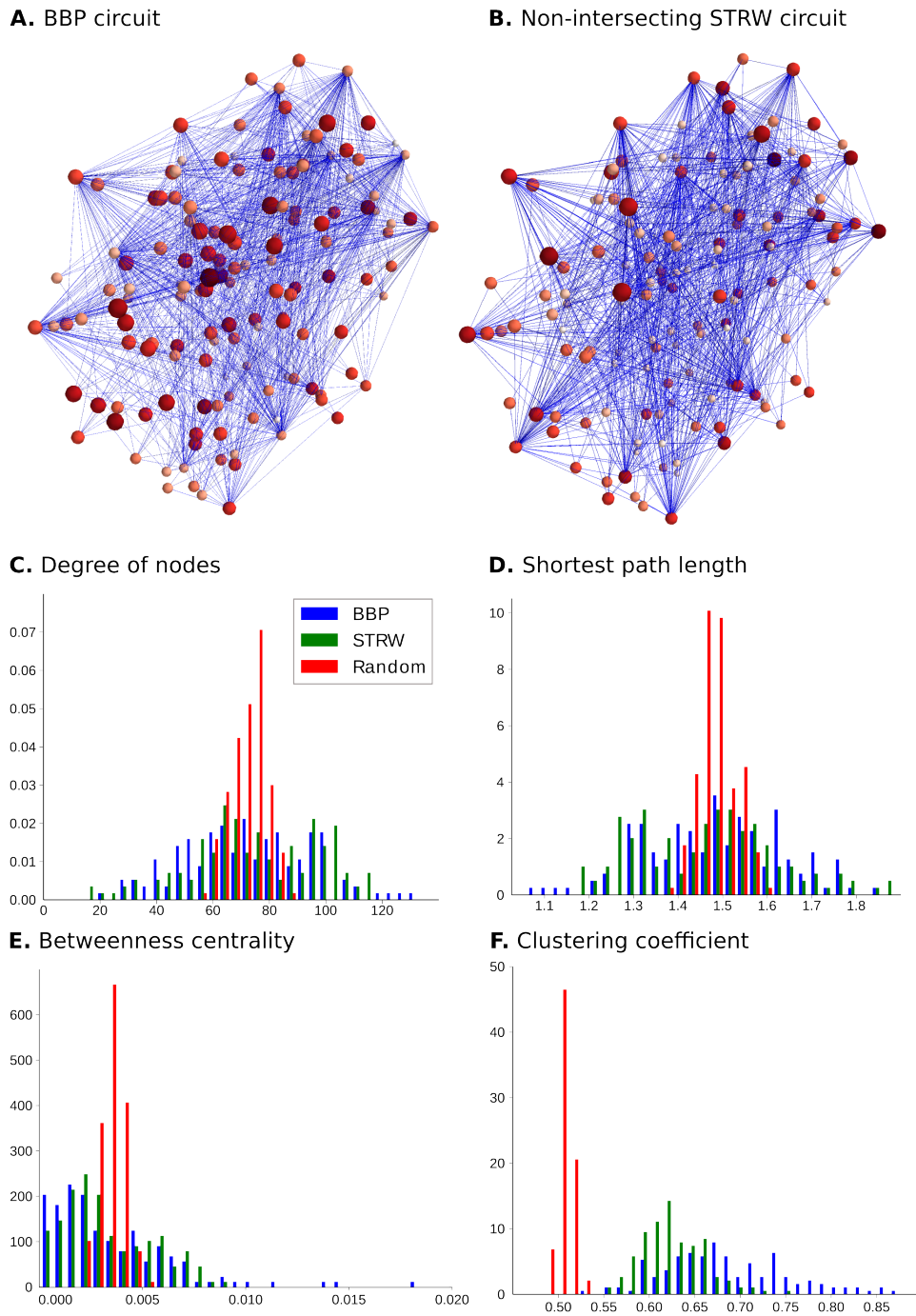
Fig. 4.11 Comparison of AGN and BBP networks' connectivity. Instances of the BBP (A) and the STRW (B) networks with randomly chosen 10% connections, starting from the same somata positions. The colormap and the size of the somata is proportional to the number of connections of the morphologies. The connectivity measurements distributions (C-F) are shown for the BBP network (in blue), the STRW network (in green) and an equivalent (same number of nodes and edges) Erdős-Renyi network (in red). C. Degree of nodes, D. Betweenness centrality, E. Shortest path lengths, F. Clustering coefficient.

The same morphological length is also retrieved by other branching growth models (BTRW, symmetric and asymmetric, not shown). There seem to be no significant difference in the number of connections, as a function of the length of the morphologies, among the networks generated by different branching rules (see Figure 4.10). This is expected by the mathematical properties of the "random walk"-like shapes, as morphologies of the same (significantly large) length that are restricted within the same space, are expected to make the same number of intersections. Note that even though the total number of connections does not depend on the branching shape for very long morphologies, the percentage of connections and the number of morphologies that a cell connects to increases with the complexity of the branching structure (see section "The impact of soma positioning and branching", Figure 4.3).

The importance of the total length of a morphology (equivalent to the volume density of the morphology) for the connectivity of a network is a well known fact (Kalisman et al. 2003). It is interesting to observe that even morphologies with simple stochastic shapes make the same number of connections that complex biological shapes generate. However, the statistical properties of the connectivity of different networks demonstrate the effect of different growth models on the connectivity of the network. Even though morphologies of length equal to the average biological length have a similar average degree, the distribution of degrees of the DSAW network's nodes is very sharp and does not correspond to the distribution of the BBP circuit (not shown). This is a result of the lack of variability of lengths in the artificially generated morphologies (Figure 4.11).

For this reason, a network of STRW morphologies of a wide distribution of lengths (from $800\mu m$ to $24,000\mu m$) with average length of $\approx 12,500\mu m$ is generated. This network has the same average degree as the BBP circuit, but it also closely approximates the broad distribution of degrees observed in the BBP network (see Figure 4.11A). The morphologies of larger lengths generate more connections with a larger number of neurons, a property that results in higher degrees. On the other hand, shorter morphologies have smaller degrees. Therefore, an increase in the variability of morphological lengths results in a network that closely approximates the biological variance of degrees. This network also reproduces a number of other connectivity measurements, such as the distribution of average shortest path distances (Figure 4.11B), the distribution of betweenness centralities (Figure 4.11C) and the distribution of intersomatic distances between connected morphologies (not shown). However, the clustering of the DSAW network is significantly lower (Figure 4.11D) compared to the BBP circuit, indicating

that more complex morphologies are required for the recreation of the high clustering that is observed in biological networks.

4.5 Conclusions

Stochastic processes that influence neuronal growth have been underestimated in the generation of morphologically detailed computational networks, as the contribution of randomness to biological neuronal connectivity is to a large extent unknown. In this study, we have demonstrated that networks built from random morphologies can reproduce a number of structural and connectivity properties of biological networks. The most important morphological determinants for the connectivity of a network are predicted to be the somata positions, the total branch length, the targeting of individual morphologies, the intersection avoidance between the branches of different morphologies and the number of trees that emanate from the root. Therefore, contrary to common belief, stochastic components of neuronal interactions significantly affect the generation of morphologically detailed networks and should not be ignored.

A network of directed random walks with the same total path length as a set of biological morphologies can reproduce their number of connections in space. The three-dimensional distribution of synapses in space, described in DeFelipe et al. (2013), is best approximated by enforcing intersection avoidance between growing morphologies. In addition, star-branched directional self-avoiding random walks recreate a number of connectivity measurements, such as connectivity probability, average shortest path distance and the degree distribution of the highly detailed computational model of BBP (Markram et al. 2015). Interestingly, artificially generated morphologies whose initial positions are unstructured appear to maximize the total number of connections and the number of cells a morphology connects to, compared to the equivalent structured initial positioning of somata.

These observations imply that there is a strong connection between the randomness of morphologies and the observed structure in the resulting network. It is, therefore, possible that the complicated natural mechanisms that result in the formation of efficient brain networks take advantage of the properties of randomness to generate efficient brain networks. An example supporting this suggestion is the formation of spatially correlated connections in networks of directional non-intersecting random walks, a property that is hypothesized to optimize the learning capacity of dendrites (Poirazi and Mel 2001).

It would be wrong to assume that biological neuronal networks are nothing more than the result of randomness, partly because this analysis studies only the behavior of the initial state of structural connections, and not the functional synapses of the network. In fact, plasticity mechanisms that define which connections are activated (Ramirez-Amaya 2007) are particularly unlikely to be reproduced by random processes, as they are the outcome of complex learning processes that associate the connectivity of the network to external inputs. The results of this study suggest that the brains of rodents have a structural connectivity that is initially largely random, thereby enabling a large number of functional connectivity patterns to be realized. Thus, we propose a simple mechanism to incorporate stochastic processes in the generation of morphologically detailed networks in order to reproduce a number of fundamental principles of biological networks.

References

Anton-Sanchez L., Bielza C., Merchan-Perez A., Rodriguez J.-R., DeFelipe J., Larranaga P. Three-dimensional distribution of cortical synapses: a replicated point pattern-based analysis. (2014) Frontiers in Neuroanatomy. 8:85.

Braitenberg, V. Brain Size and Number of Neurons: An Exercise in Synthetic Neuroanatomy (2001) J Comput Neurosci 10: 71.

Braitenberg V., Schuz A Cortex: Statistic and Geometry of Neuronal Connectivity (1998) (2nd ed.). Springer-Verlag, New York. (revised edition of Anatomy of the Cortex: Statistics and Geometry, Springer-Verlag, 1991)

Chklovskii D.B., Synaptic Connectivity and Neuronal Morphology: Two Sides of the Same Coin, (2004). Neuron, Volume 43, Issue 5, Pages 609-617, ISSN 0896-6273, https://doi.org/10.1016/j.neuron.2004.08.012.

Craddock R.C., Jbabdi S., Yan C.G., Vogelstein J.T., Castellanos F.X., Di Martino A., Kelly C., Heberlein K., Colcombe S., Milham M.P. Imaging human connectomes at the macroscale. (2013). Nat Methods. 10(6):524-39.

Downes J.H., Hammond M.W., Xydas D., Spencer M.C., Becerra V.M., Warwick K., Whalley B.J., Nasuto S.J. Emergence of a small-world functional network in cultured neurons. (2012). PLoS Comput Biol. 8(5):e1002522.

Egger R., Dercksen V.J., Udvary D., Hege H.C., Oberlaender M. Generation of dense statistical connectomes from sparse morphological data. (2014). Frontiers in Neuroanatomy Volume 8, ISSN 1662-5129.

Erdős, P., Renyi, A. On Random Graphs. I (1959). Publicationes Mathematicae. 6: 290–297.

Fauth M., Worgotter F., Tetzlaff C. The Formation of Multi-synaptic Connections by the Interaction of Synaptic and Structural Plasticity and Their Functional Consequences. (2015). Beck J, ed. PLoS Computational Biology. 2015;11(1):e1004031.

Graham, B.P., van Ooyen, A. Mathematical modelling and numerical simulation of the morphological development of neurons. (2006). BMC Neuroscience. 7(Suppl 1):S9.

Grueber W.B., Sagasti A. Self-avoidance and Tiling: Mechanisms of Dendrite and Axon Spacing. (2010). Cold Spring Harbor Perspectives in Biology. 2(9):a001750.

Kalisman N., Silberberg G., Markram H. Deriving physical connectivity from neuronal morphology. (2003). Biol Cybern. 88(3):210-8.

Kasthuri N., Hayworth K.J., Berger D.R., Schalek R.L., Conchello J.A., Knowles-Barley S., Lee D., Vazquez-Reina A., Kaynig V., Jones T.R., Roberts M., Morgan J.L., Tapia J.C., Seung H.S., Roncal W.G., Vogelstein J.T., Burns R., Sussman D.L., Priebe C.E., Pfister H., Lichtman J.W. Saturated Reconstruction of a Volume of Neocortex. (2015) Cell. 162(3):648-61.

Lotwick, H. W., and B. W. Silverman. Methods for Analysing Spatial Processes of Several Types of Points. (1982) Journal of the Royal Statistical Society. Series B (Methodological), 44(3), 1982, pp. 406–413.

Markram H., Muller E., Ramaswamy S., Reimann M.W. et al., Reconstruction and Simulation of Neocortical Microcircuitry (2015). Cell, Volume 163, Issue 2, Pages 456-492, ISSN 0092-8674,

Meghanathan N., A Random Network Model with High Clustering Coefficient and Variation in Node Degree, (2015). 8th International Conference on Control and Automation (CA), pp. 54-57.

Pearson K. The Problem of the Random Walk. (1905). Nature 72(1865):294.

Perin, R., Berger, T. K., & Markram, H. A synaptic organizing principle for cortical neuronal groups. (2011). Proceedings of the National Academy of Sciences of the United States of America, 108, 5419-5424.

Peters A. Thalamic input to the cerebral cortex.(1979). Trends Neurosci. 2, 1183–1185

Poirazi P., Mel B.W., Impact of Active Dendrites and Structural Plasticity on the Memory Capacity of Neural Tissue. (2001) In Neuron, Volume 29, Issue 3, Pages 779-796, ISSN 0896-6273.

Ramirez-Amaya V. Molecular Mechanisms of Synaptic Plasticity Underlying Long-Term Memory Formation. (2007). In: Bermúdez-Rattoni F, editor. Neural Plasticity

and Memory: From Genes to Brain Imaging. Boca Raton (FL): CRC Press/Taylor & Francis; Chapter 3.

Reimann, M.W., Muller, E.B., Ramaswamy, S., and Markram, H. (2015). An algorithm to predict the connectome of neural microcircuits. (2015). Front. Comput. Neurosci. 9,28.

Reimann M.W., Nolte M., Scolamiero M., Turner K., Perin R., Chindemi G., Dłotko P., Levi R., Hess K., Markram H. Cliques of Neurons Bound into Cavities Provide a Missing Link between Structure and Function. (2017). Front Comput Neurosci. 11:48.

Rubinov M., Sporns O. Complex network measures of brain connectivity: uses and interpretations. (2010). Neuroimage. 52(3):1059-69.

Scott E.K., Luo L. How do dendrites take their shape? (2001). Nat Neurosci. 4(4):359-65. Review. PubMed PMID: 11276225.

Shi Y., Toga A.W. Connectome imaging for mapping human brain pathways. (2017). Molecular Psychiatry. 22(9):1230-1240.

Sloane, N.J.A. The Sphere-Packing Problem (2002) ArXiv Mathematics e-prints

Tekin E., Hunt D., Newberry M.G., Savage V.M. Do Vascular Networks Branch Optimally or Randomly across Spatial Scales? (2016) PLoS Comput Biol. 12(11):e1005223.

Watts D.J., Strogatz S.H. Collective dynamics of small-world networks. (1998) Nature. 393(6684):440-2.

Weigand M., Sartori F., Cuntz H. Universal transition from unstructured to structured neural maps. (2017). Proceedings of the National Academy of Sciences of the United States of America. 114(20):E4057-E4064.

5

Future directions

The purpose of this work was to study the properties of neuronal morphologies and investigate how they contribute to the spatial structure and the connectivity of the brain. While bringing us closer to understanding the fundamental differences between morphologies, and how randomness and structure are combined to generate one of the most fascinating biological systems, many questions remain unsolved. Some of these problems could be addressed with small extensions to the techniques that have been developed and described in this thesis, other problems are more speculative. In this thesis I focused on the development of mathematical methods for neuroscience. Future applications can be pursued in several directions including those outside the field of neuroscience. First, general improvements to the TMD algorithm will benefit multiple applications followed by new applications to neuronal axon synthesis among others. Second, extensions to the TMD based synthesis will improve the quality of artificial morphologies, contributing to better digital reconstructions of larger brain regions. The TMD has been developed and applied in the field of neuron morphological analysis and synthesis, but its general nature suggests that it will find applications in many other fields, and will be extended in ways that have not even been described here.

General Extensions to Topological Morphology Descriptor

The topological morphology descriptor (TMD) algorithm is useful for the description and the discrimination of the branching structure of morphologies. However, it does not encode all the information that describe the structural aspects of a neuron. Several properties of the branching structure, such as the diameters of the branches or the

bouton densities, are not represented in the persistence barcode based on the radial distances of the tree's branches from the soma. As a result, the grouping of cells based on this barcode ignores information about the original morphology that is important for the functional role of the cells.

In order to establish a more complete classification scheme, the extraction of the TMD should be extended to take into account additional morphological properties that are currently not considered, as it has been implemented for the classification of the apical trees of pyramidal cells by accounting for the direction of the trees in space (see Chapter 2.3). This idea can be taken a step further by combining multiple morphometrics into a single descriptor. This could be achieved using the newly developed tools of Multidimensional Persistence (Carlsson and Zomorodian 2009), which combine different filtration functions into a single metric.

This approach would revolutionize the way we think about neuronal trees as it will allow us to study the correlations not only between static morphological features, but also between the spatial and temporal properties of neurons. A concrete example of this idea is the study of the structural properties of growing neurons during the different stages of their morphological development, such as the datasets published in (Goncalves et al. 2016). This application is not limited to the study of growing neurons but could be useful for studying the morphological growth of other branching structures, such as corals and botanic roots.

Applications of Topological Morphology Descriptor

In Chapter 2, the TMD was used for the characterization of neuronal trees. However, there is an abundance of biological trees that can be studied with the same method. Some biological examples include botanic trees, roots and corals. A particularly challenging application of the TMD is the study of vascular systems. These are networks of veins and arteries that carry blood, such as the cerebral and pulmonary vascular systems. A topological descriptor that would distinguish healthy from diseased vascular systems would be very useful for identification and treatment of blood flow diseases. In order to apply the TMD algorithm to vascular systems we need to adapt the algorithm accordingly to extract the barcode of graphs as opposed to trees.

Improvements of Topological Neuron Synthesis

The generation of artificial neuronal morphologies that are statistically close to a biological population of reconstructed neurons is a challenging problem that is nevertheless essential for increasing the morphological variability of digital reconstructions. In Chapter 3, I have used the TMDs of cortical neurites to generate artificial dendrites, of a wide variety of cell types, and demonstrated that they closely match the morphometrics and the topology of the original cells, thereby improving the quality of the synthesized cells and increasing the limited biological variability.

The thickness of dendritic branches is an important determinant of their functionality (Cuntz et al. 2007), as it directly impacts the density and distribution of ion channels on their surface and consequently the way that dendrites integrate signals. In Chapter 3, I introduced a simplified algorithm to generate diameters for the branches of the synthesized trees. This algorithm reproduces the distribution of diameters on the biological reconstructions available from the labs. However, it does not reproduce the branch diameters of biological cells as reconstructions are subject to often significant diameter errors. Ideally I would like to use High-Resolution Cryo-EM data to estimate and correct the reconstruction error in order to use more accurate thickness data on the modeling of branch diameters.

Another important point that needs further investigation is the simulation of the branching angles of the growing morphologies. Currently, the branching angles are sampled using a composite method that considers the biological angles between siblings as well as the angles between parent-children and works well for the generation of dendritic branches. However, this method is limited by the quality of the input dataset. A more realistic approach for the definition of branching angles would be more effective for the generalization of the topological synthesis to a variety of different neuronal morphologies. An example of a more sophisticated bifurcation algorithm, which is based on self-referential forces, is proposed in Memelli at al. (2013).

The current synthesis technique increases the limited variability of the sparse biological reconstructions of certain m-types. However, because the topological profiles of neurons of the same m-type are similar, it is tempting to conjecture that a small number of biological reconstructions of a cell type could suffice to synthesize a large number of unique morphologies with similar characteristics. A stochastic algorithm that creates unique persistence barcodes from a small number of exemplar barcodes generated from the sparse biological reconstructions of a cell type, would significantly increase the variability for the synthesis input. If this assumption is true, synthesis would be essential for the generation of large numbers of unique morphologies from

few examples of biological cells, to populate computational models that extend beyond the scale of a few microcircuits.

Applications of Topological Neuron Synthesis

The next logical step is to use the TNS algorithm to synthesize neurons of different brain areas and species. Preliminary results suggest that cell types with more regular spatial conformations, such as Purkinje cells that are almost planar, require additional constraints in order to be accurately reproduced. The invariance of the TMD to rotations is useful for the generation of spatially symmetric trees, such as the cortical dendrites, but is insufficient to capture the preferred orientations of more stereotypical cell structures. It is thus, important to identify the limitations of the TNS algorithm for different cell types and propose improvements for its generalization.

One of the most challenging problems in computational neuroscience is the generation of artificial axonal trees, and especially long range axonal projections, which are rarely included in biological reconstructions. The generation of artificial axons is of particular interest for the computational modeling of brain networks for a number of reasons. First, because of their highly complex branching structures, the reconstruction of axons requires considerable effort and time. As a result only a small number of intact (not cut) axonal reconstructions are available. In addition, the branching structure of axonal morphologies is an essential determinant of the functionality of a network, as it provides the contact points between neurons and thus defines the connectivity of the network.

The cortex is curved in shape, with curvatures ranging from place to place. In order to synthesize axonal trees it is crucial to couple the TNS algorithm with environmental cues derived from a curved space, a property that is not currently taken into account. This will also contribute to the generation of artificial cells in place, that could then directly populate curved brain regions. Therefore, the current synthesis scheme should be extended to take into account environmental influences, such as long-range targeting and intersection avoidance between neighboring cells (Grueber and Sagasti 2010). This approach will also improve the generation of neurites of complex branching patterns, such as cortical axons of both interneurons and pyramidal cells, glial cells and long range projecting cells, such as nigrostriatal dopaminergic neurons (Matsuda et al. 2009) and densely connected claustrum cells (Torgerson et al. 2015) and thus allow the digital reconstruction of brain areas that extend beyond a single microcircuit.

Extensions to random walk models for synthesis in 3d space

The artificially generated networks (AGN) of random walks, described in Chapter 4, were developed with the intention of studying the interactions between growing structures and identifying their impact on the resulting networks. As a result the significance of randomness and intersection avoidance on the resulting networks have been demonstrated. However, the growth-models that were used produced simplified morphologies that do not correspond to the biological cells that are found in the brain. A natural next step is to add stricter constraints on the growth-models to generate realistic morphologies and examine the impact of specific morphological characteristics on the resulting network. One of the most important morphological features of a neuron is considered to be the asymmetry of a neuronal tree (Van Pelt et al. 1992), which depends on the topology of its branching structure. With this model we can study if symmetric trees generate different connectivity patterns compared to asymmetric trees. Another feature that is considered to be essential for the morphology of a neuron (Fernandez-Gonzalez et al. 2017) is the branching angles. The AGN model will allow us to study whether branching angles have an impact on the connectivity of a network or not.

An improvement of the growth models could be introduced by considering interactions between the growing morphologies and the embedded space. For instance, a gradient field that would guide the individual morphologies to preferred targets in space would be a meaningful approximation for the growth of long range axons. This way, the complex chemical and molecular guidance cues that have been ignored in the synthesis algorithm described in Chapter 3, could be approximated. Applying environmental constraints to the simplified structures will make the implementation of more complex interactions feasible in order to evaluate their impact on artificially generated networks.

Apart from the effect of environmental cues on the growth of morphologies, the interactions between morphologies is one of the most interesting problems that can be studied with this model. As illustrated in Chapter 4, intersections between morphologies have a significant impact on the spatial distribution of synapses, and therefore should be removed from a digital reconstruction in order to accurately approximate the biological properties of brain networks. However, the elimination of the intersections at the final stage of the digital reconstruction, i.e., when all full-grown morphologies are placed in space, is computationally intractable. As a result, the optimal strategy would be to

synthesize all neurons simultaneously in space. However, this approach will require the modeling of more complex forces, such as attraction and repulsion, between the growing morphologies. Different types of interaction could be studied with the model proposed in Chapter 4, in order to identify their limitations and their impact on artificially generated networks.

References

Carlsson, G., & Zomorodian, A. (2009). The theory of multidimensional persistence. Discrete & Computational Geometry, 42(1), 71–93.

Cuntz H., Borst A., Segev I. Optimization principles of dendritic structure. (2007). Theoretical Biology & Medical Modelling. 4:21.

Fernandez-Gonzalez P., Benavides-Piccione R., Leguey I., Bielza C., Larranaga P., DeFelipe J. Dendritic-branching angles of pyramidal neurons of the human cerebral cortex. (2017). Brain Structure & Function. 222(4):1847-1859.

Goncalves J.T., Bloyd C.W., Shtrahman M., Johnston S.T., Schafer S.T., Parylak S.L., Tran T., Chang T., Gage F.H. In vivo imaging of dendritic pruning in dentate granule cells. (2016). Nat Neurosci. 19(6):788-91.

Grueber W.B., Sagasti A. Self-avoidance and Tiling: Mechanisms of Dendrite and Axon Spacing. (2010). Cold Spring Harbor Perspectives in Biology. 2(9):a001750.

Matsuda W., Furuta T., Nakamura K. C., Hioki H., Fujiyama F., Arai R., Kaneko T. Single Nigrostriatal Dopaminergic Neurons Form Widely Spread and Highly Dense Axonal Arborizations in the Neostriatum. (2009). Journal of Neuroscience 29(2) 444-453

Memelli H., Torben-Nielsen B., Kozloski J. Self-referential forces are sufficient to explain different dendritic morphologies. (2013). Frontiers in Neuroinformatics. 7:1.

Torgerson C.M., Irimia A., Goh S.Y., Van Horn J.D. The DTI connectivity of the human claustrum. (2051). Hum Brain Mapp. 36(3):827-38.

Van Pelt J., Uylings H.B., Verwer R.W., Pentney R.J., Woldenberg M.J. Tree asymmetry–a sensitive and practical measure for binary topological trees. (1992). Bull Math Biol. 54(5):759-84.

Appendix A

SI: A topological representation of branching neuronal morphologies

Morphological clustering

Traditionally, the different morphological shapes of neurons have been qualitatively described based on visual inspection and quantitatively described based on morphometric parameters. Feature extraction results in significant loss of information, as the dimensionality of the data is significantly reduced. As a result, a limited set of selected features is not sufficient to capture the full complexity of the neuronal shapes. On the other hand, a large number of features will result in overfitting, since the correlated features are accounted for multiple times. In fact, the feature-based classification of neuronal trees strongly depends on the set of morphometrics that are used as input. Alternative sets of morphometrics result in different classifications (DeFelipe et al. 2013) for the same set of cells.

In this section we illustrate the problems of this method with a simple example. In Fig 2.5 we present the results of the feature classification for a set of neuronal trees that belong in three distinct groups (axons, basal and apical dendrites). The data used for this grouping are given in Tables A.1, A.2, A.3. The visual separation of the trees into three groups is presented in Fig A.1A. Even though two of the most important anatomical features, i.e., the total length and the total number of branches of the tree, are used, the resulting clustering does not correspond to the biological grouping (colormap in Fig A.1).
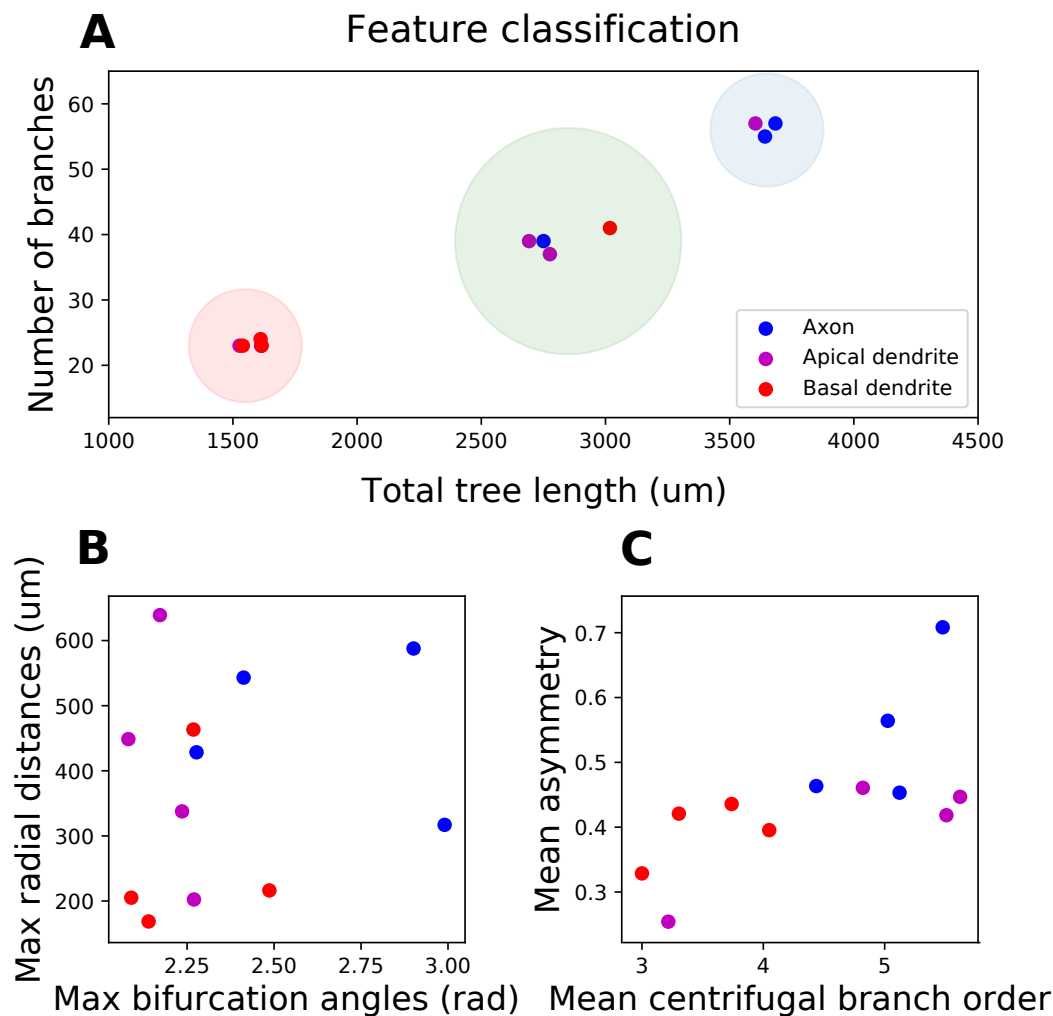
Fig. A.1 Visual separation of trees into groups according to selected morphological features. Three instances of the clustering are presented using different pairs of features (A: total tree length - number of branches, B: maximum branch angles - maximum radial distances, C: average branch order - average asymmetry); the results are not consistent. Inappropriate feature selection will therefore result in a grouping that does not correspond to the biological role of the three tree types (axons, apical and basal dendrites) as shown in all three cases.

Alternative feature pairs (maximum branch angles - maximum radial distances Fig A.1B, average centrifugal branch order (Van Pelt and Verwer 1989)- average asymmetry (Van Pelt and Verwer1983) Fig A.1C) cannot reproduce the biological grouping or the initial classification (Fig A.1A). The results of the principal component analysis (PCA-decomposition) based on the six morphological features are presented in Fig A.2. We present the three first components with explained variance 96%. Even though six of the most significant morphological features were used for this analysis, the three biological types of trees cannot be retrieved. This observation indicates

Table A.1 Morphological features (Number of branches, Total length) extracted from the trees (axons, apicals, basals) presented in Fig A.1A.

Tree ID	Num branches	Total length(um)
Axon 1	57	3684.24
Axon 2	55	3642.24
Axon 3	39	2750.47
Axon 4	23	1614.04
Apical 1	57	3603.75
Apical 2	37	2776.20
Apical 3	39	2692.38
Apical 4	23	1526.87
Basal 1	41	3017.89
Basal 2	24	1611.49
Basal 3	23	1539.61
Basal 4	23	1615.44

Table A.2 Morphological features (Max branch angles, Max radial distances) extracted from the trees (axons, apicals, basals) presented in Fig A.1B.

Tree ID	Max branch angles	Max radial distances
Axon 1	2.99	316.97
Axon 2	2.28	428.41
Axon 3	2.90	587.82
Axon 4	2.41	543.14
Apical 1	2.08	448.66
Apical 2	2.17	639.10
Apical 3	2.24	337.68
Apical 4	2.27	202.35
Basal 1	2.27	463.33
Basal 2	2.09	205.11
Basal 3	2.14	168.58
Basal 4	2.49	216.26

that the feature based classification of neuronal trees is very sensitive to the selected features. As a result, a feature based classification is not reliable, as the appropriate feature set cannot be generalized across different groups.

Table A.3 Morphological features (Mean branch orders, Mean asymmetry) extracted from the trees (axons, apicals, basals) presented in Fig A.1C.

Tree ID	Mean centrifugal branch orders	Mean asymmetry
Axon 1	5.12	0.45
Axon 2	4.44	0.46
Axon 3	5.03	0.56
Axon 4	5.48	0.71
Apical 1	5.51	0.42
Apical 2	5.62	0.45
Apical 3	4.82	0.46
Apical 4	3.22	0.25
Basal 1	4.05	0.40
Basal 2	3.00	0.33
Basal 3	3.74	0.44
Basal 4	3.30	0.42
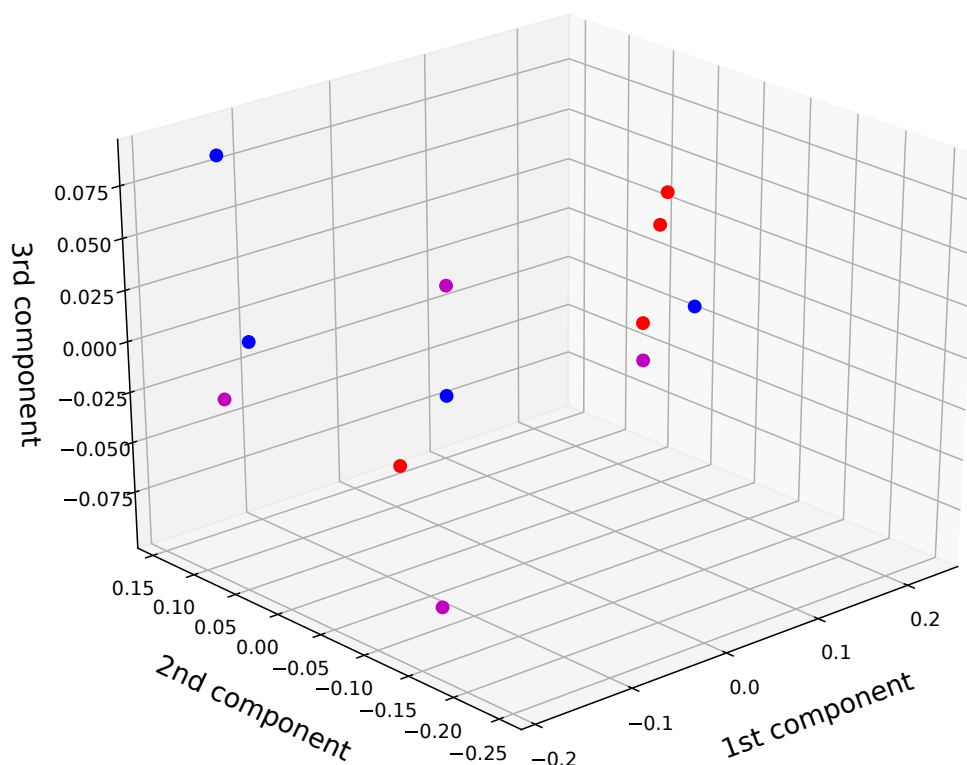
Fig. A.2 Principal component analysis on the normalized values of the six selected morphometrics of Fig A.1 (total tree length, number of branches, maximum branch angles, maximum radial distances, average branch order, average asymmetry). The PCA of the selected feature set is not able to distinguish the correct groups that correspond to the biological role of each tree, despite the different branching patterns that they present. Hence, inappropriate feature selection can indeed result in misclassification.

Demonstration of the TMD algorithm

The idea of the TMD algorithm is presented in Figs A.3A. The input of the TMD algorithm is a rooted tree with a function f defined on the set of nodes. In this example, the function f is the radial distance. The root, denoted by R, is shown in red, while the other nodes of the tree are labeled $a - i$. Note that the set of nodes consists of the branch points and the leaves. During the initialization of the algorithm, the leaves (a, c, e, g, h) are inserted into the list of active nodes, A. The algorithm then iterates over the members of A. The order of this process is not significant. Recall that the function v assigns to a node n of the tree the largest value of the function f on the leaves of the subtree with root at n (see Methods).
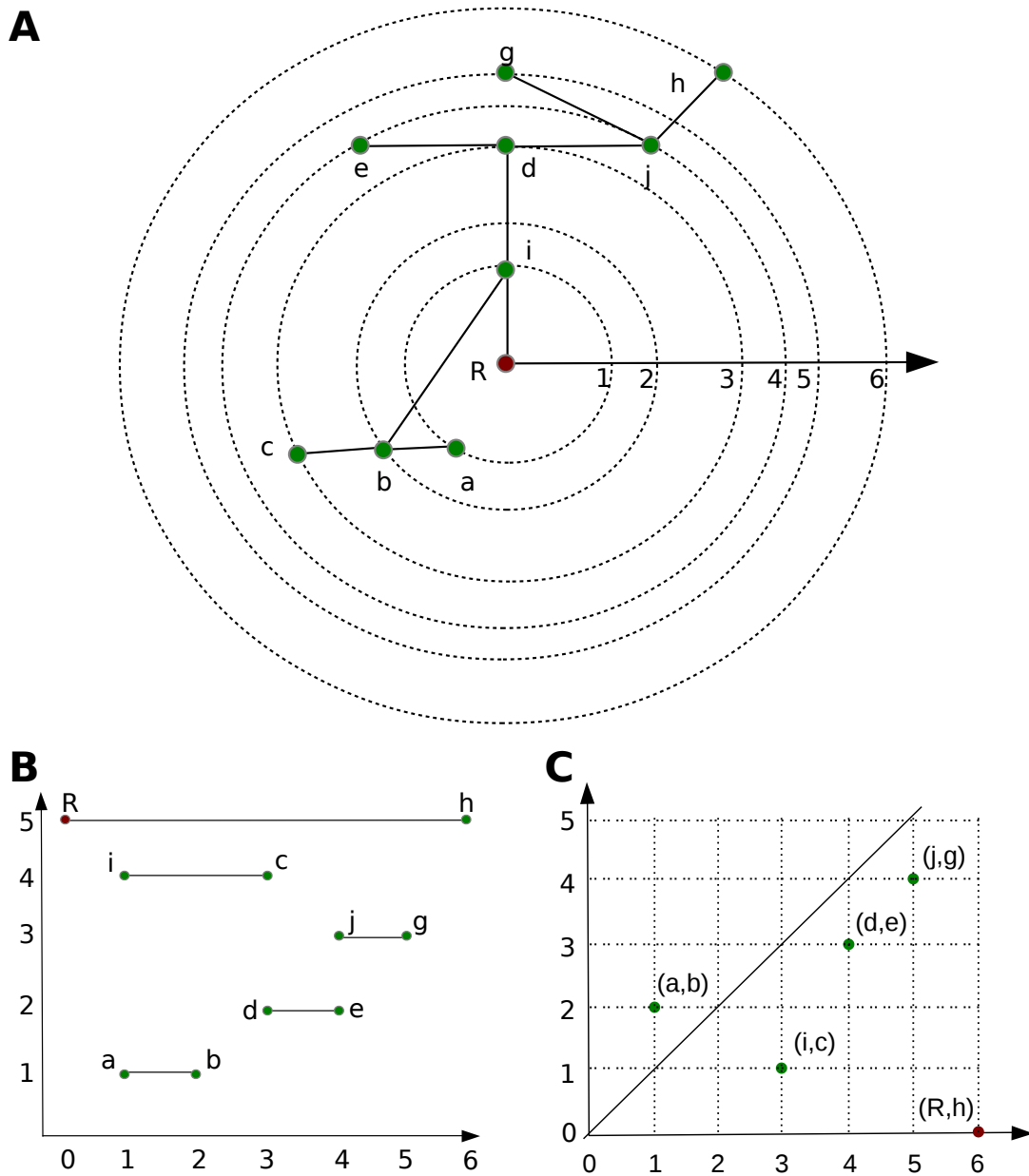
Fig. A.3 Demonstration of the TMD algorithm: A simple embedded rooted tree (A) is transformed with the TMD algorithm into the corresponding persistence barcode (B) and the equivalent persistence diagram (C). The root (R) is colored red, while the branch points and leaves are shown in green. The edges connecting corresponding pairs of points are presented by straight lines. The dashed circles are provided as a guide to the eye to indicate different levels of radial distances. The correspondence between the tree (A) and its extracted barcode (B) and its diagram (C) is given by the notation of the same nodes in both figures. Each bar in (B) represents the lifetime of a component. The positions of x-axis correspond to the circles in (A) while y-axis represents individual components, ordered according to their length. In (C) each point represents the birth and death time of a branch component in A.

The algorithm assigns the values of v on the leaves: $v(a) = f(a) = 1, v(c) = f(c) = 3, v(e) = f(e) = 4, v(g) = f(g) = 5, v(h) = f(h) = 6$. Consider the node a as the first element of the list A. The parent of a is b and its only other child is c. Since both a and c are in A, the algorithm orders the siblings according to the values of function v. The older sibling is c and therefore $v(b) = v(c) = 3$. The interval $[1, 2]$ is added to the persistence barcode (Fig A.3B) $\text{TMD}(T, f)$ representing the lifetime of the node a. This interval is equivalently represented as a point ab on the persistence diagram (Fig A.3C). Nodes a and c are removed from A, and b is added to A. The next vertex in the list A is e. The algorithm finds its parent, d, but this node is not processed further at this stage, since j, the sibling of d, is not in A. The next node to be processed is g. Both children g and h of j are in A. The oldest child is h and therefore $v(j) = v(h) = 6$. The interval $[5, 4]$, representing the lifetime of node g, is added to $\text{TMD}(T, f)$. The node j is added to A, and both g and h are removed from A. The list of alive components then consists of b, e, j. The node b cannot be processed since its sibling d is not in A. The next node to be processed is therefore e, whose parent d has all of its children in A. In this case, the node with highest value of v is j, and therefore $v(d) = v(j) = 6$. The interval $[4, 3]$ is added to $\text{TMD}(T, f)$. Then nodes e and j are removed from A, and d is added to A. The next node in A is b, whose parent is i. Since both children of i are now in A, the algorithm finds the older sibling, d and assigns $v(i) = v(d) = 6$. The interval $[3, 1]$ is added to $\text{TMD}(T, f)$ and d, b are removed from A, while i is added to A. The only alive node is now i whose parent is the root R. The algorithm computes $v(R) = v(i) = 6$, i is removed from A and R is added in A. Since only the root R is alive, the while loop in the algorithm terminates. The last step adds the interval $[6, 0]$ to $\text{TMD}(T, f)$, which represents the largest component of the tree.

Using alternative functions for the TMD algorithm

In the previous section, we applied the TMD algorithm with the radial distance as the filtration function f. Any alternative function f can be used, such as the path distance from the root, which should serve to reveal shape characteristics that are independent of the radial distance and thus not captured by this approach. The constraint of rotational invariance could also be relaxed by projecting the radial distance to a selected axis, to map the spherical filtration into an ellipsoidal one, in order to study the relation of a tree's spatial density to its embedding space.

Depending on the classification problem, alternative morphometrics could be more appropriate for the separation of trees in classes. For instance, the path distance Fig A.4A would be more appropriate to capture the differences between tortuous and straight trees, while the projected radial distance Fig A.4C can discriminate trees with different spatial distribution of branches. In Fig A.4 we present four variations of the TMD using different morphometrics (radial distance from the soma (A), path distance from the soma (B), projected to the axis towards the pia radial distance (C), branch orders). Each morphometric captures different properties of the branching structure. Those and other morphometrics could be combined in a multidimensional persistence diagram for the better discrimination of trees.
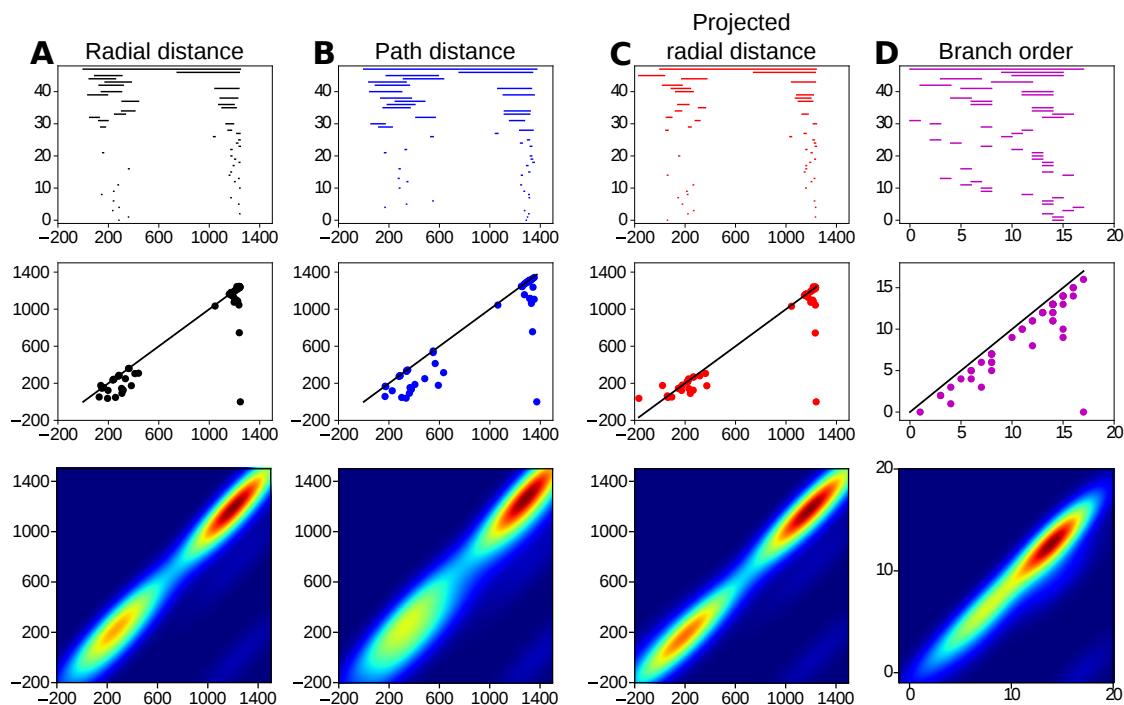
Fig. A.4 Demonstration of TMD algorithm for different morphological features. A. Radial distance from the soma. B. Path distance from the soma. C. Projected radial distance from the soma to the axis normal to the pia; this measurement can discriminate trees with different spatial distributions. D. Branch order; this measurement does not take into account the embedding in space, only the combinatorial branching patterns of the tree. Note the similarity among the three first morphometrics.

Definition of distances

In order to establish the comparison with the current literature we need to define a notion of distance between trees equipped with a real-valued function on their nodes, as well as a notion of distance between persistence diagrams.

Distances between persistence diagrams

Below we recall various representations of persistence diagrams and some notions of distance between them. We also provide a reference to software that computes the distances considered, when available. All of the metrics summarized below can be applied directly to the output of the TMD algorithm.

The most classical distances used in topological data analysis are the *bottleneck* and *Wasserstein* distances. Given a persistence diagram D, the points in the diagonal are "virtual" points, which have birth time equal to their death time. Therefore, we assume without loss of generality that a persistence diagram contain points in the diagonal with infinite multiplicity. Given two persistence diagrams D_1 and D_2, we construct a matching (i.e., a bijection) $\phi : D_1 \to D_2$ and define two numbers

$$B_\phi = \sup_{x \in D_1} d(x, \phi(x))$$

and

$$W_\phi^p = \left(\sum_{x \in D_1} d(x, \phi(x))^p \right)^{\frac{1}{p}},$$

where d is the standard Euclidean distance in $\mathbb{R}^2$. Note that B_ϕ is simply the longest distance that ϕ shifts a point in D_1, while $(W_\phi^p)^p$ is a sum of p-th powers of lengths of the line segments joining x and $\phi(x)$, for all x. The infimum of B_ϕ over all possible matchings is the *bottleneck distance* between D_1 and D_2. The infimum of W_ϕ^p over all possible matchings is the *p−Wasserstein distance* between D_1 and D_2. Given two persistence diagrams D_1 and D_2, their bottleneck distance will be denoted by $d_B(D_1, D_2)$ and their p−Wasserstein distance by $W_p(D_1, D_2)$. One implementation of these distances is given in Morozov (2016)and a faster approximation in Kerber et al. (2016).

A persistence diagram can also be represented by a persistence landscape, i.e., a piecewise linear function $\mathcal{L} : \mathbb{R} \times \mathbb{N} \to \mathbb{R}$. Given two persistence landscapes, we

can compute the distance between them in $\mathcal{L}^p$ space (Bubenik et al. 2015). The implementation is described in Dłotko et al. (2015).

One can also encode persistence diagrams by unweighted persistence images as described in the main text. The idea is to apply a smoothing function, i.e., a Gaussian kernel, at every point of the diagram and then to discretize the distribution obtained into a pixel-based image. It is then straightforward to compute a distance between two unweighted persistence images, using common image-recognition techniques. A simplified version of this representation is used in the classification of morphological types of neurons in the experimental section of this paper. We are not aware of a publicly available implementation of this approach. An implementation is provided with the software of this paper.

Distances between trees

A classic metric to compare trees, the *edit distance* (Bille 2005), is based on the transformation of one tree T_1 into another T_2 by a sequence of operations (deletion and insertion of vertices), each of which has a non-negative cost. The edit distance (Bille 2005) between T_1 and T_2 is defined to be the infimum of the total cost of all possible transformations from T_1 to T_2. However, the edit distance is not relevant to our problem, since it does not involve geometric information about the tree structure and is known to be NP-complete (Shapira et al. 2011).

An important notion of distance is the one between merge trees as defined in (Beketayev et al. 2014) and (Morozov et al. 2013). This distance is applied to merge trees of sublevel sets of functions. For a function $f : X \to \mathbb{R}$, where X is a metric space, the sublevel set at level $a \in \mathbb{R}$ is $\{x \in X | f(x) \leq a\}$. The differences captured by merge trees are considerably more subtle than the differences captured by the persistent homology of the function's sublevel sets. The authors of (Beketayev et al. 2014) and (Morozov et al. 2013) provide examples of pairs of simple merge trees T and T' that have the same persistence diagrams, but that are a nonzero distance apart. It is clear that in this particular case, the TMD would provide the persistent homology of the sublevel sets of the function. Therefore, by rescaling T and T', the difference between the distances used in those papers and the distances used in the current paper can get arbitrarily large.

Another relevant metric is the *persistence distortion distance* (Dey et al. 2015) between two graphs G_1 and G_2. To compute this distance one must calculate the shortest path distance from a fixed point to any other point in the tree for all $v_1 \in G_1$, denoted $P(G_1, v_1)$, and all $v_2 \in G_2$, denoted $P(G_2, v_2)$. Given the shortest paths,

the persistence distortion is defined as the minimal bottleneck distance between the persistence diagrams in dimension zero of the superlevel sets of the distance functions $P(G_1, v_1)$ and $P(G_2, v_2)$. The persistence diagrams obtained in the process are conceptually very close to the diagrams we get from the TMD algorithm. In our case, we obtain a significant computational advantage from working with rooted trees, since there is always a unique path between every pair of vertices. Moreover a reasonable choice of initial vertices v_1 and v_2 from which to compute shortest paths is to take the root of the trees considered, given that this is the computational center of the neuron. In this case, the persistence diagram arising when computing the persistence distortion distance is the one we would get from the TMD algorithm when the function f is the path distance from the root. The computational cost of the distortion distance is considerable in the general case, but linear in our case. However, since the persistence distortion distance is based on the bottleneck distance, it suffers from that metric's limitations, i.e., the shortest components, which are important for the neuronal morphologies, are not taken into account. The code to compute persistence distortion distance is available here (Dey and Shi 2016).

Distances between neurons

Strahler ordering (Strahler et al. 1952 , Ledderose et al. 2014), a metric introduced for the study of a river's branching patterns, assigns a number to each branch of the tree, starting from the terminal branches (order 1) and increasing the ordering when branches of the same order merge. Strahler ordering analysis is similar to the TMD-algorithm because it starts from the terminal branches of the tree and proceeds from the outer branches towards the root. However, since the embedding of the tree is not considered, branches of different lengths are treated equally and their spatial distribution cannot be studied. The advantage of Strahler ordering is that the overall branching topology of the tree is captured in a single value and hence the comparison between trees is straightforward. However, depending on the branching structure, very complicated neuronal trees can be assigned low Strahler orders (for example a Hippocampus pyramidal cell can be of Strahler order 4, (see Ledderose et al. 2014, Figure 3) so they are inseparable from simpler structures. This is once again due to the significant information loss of this analysis.

Sholl analysis (Sholl 1953) is a typical measurement used to study neuronal morphologies. It counts the number of segments that intersect with a set of equidistant spheres $\{S_0, S_1, S_2, ..., S_s\}$ of increasing diameters $\{0, sd, 2sd, ...\}$. Because of the high frequency of local fluctuations, the choice of the diameter step sd has a significant

impact on the result of this analysis. While the Sholl analysis counts the number of components at each level, the persistence diagram of a tree T tracks the evolution of those components in space. As a result, the persistence diagram of a tree contains strictly more information than the Sholl analysis. In fact, the Sholl analysis can be retrieved from the TMD of a tree using a discretized version of distance d_{Bar}, which is defined in Methods. Similarly to the Sholl analysis the d_{Bar} distance encodes the number of components of the tree for a set of spheres of increasing diameters with a few significant differences. First, d_{Bar} does not depend on a choice of diameter step, so it is not subject to local fluctuations. In addition, the distance d_{Bar} counts the number of intersections of the branches of a tree with a sphere, as opposed to the segments that are counted in Sholl analysis. As a result, this distance is equivalent to a continuous version of Sholl analysis that processes the branches of the tree. This distance collapses the barcode structure into one dimension which results in significant information loss. As a result, it is not appropriate to distinguish subtypes of trees that express similar branching structures, such as subtypes of pyramidal cells (Fig 2.9).

A novel metric that is useful for distinguishing neuronal trees was proposed in (Wan at al. 2015). *Blastneuron* focuses on the comparison of neurons based on the alignment of the branches by topology and path shapes after first defining similar neurons on the basis of their morphometrics. A set of morphological features is extracted from the trees, and the initial estimation of the distance between them is defined by the distance between the extracted features. An alignment algorithm is then applied to pairs of trees in order to identify local similarities. The local alignment requires the comparison of all pairs of branches, making the computation very expensive. This method is designed for the efficient matching of trees with highly similar structures, but the high variability within the groups of rat cortical neurons does not allow similar trees to be grouped together by local alignment, since local structures are often altered, depending on the location of the cells in the tissue.

The most recent advance in the field was made by *sequence representation* (Gillette et al. 2015), an original encoding of trees as sequences of characters 'ACT' representing the local topology. Bifurcations are encoded on the basis of whether their children branch or terminate. Arborizing bifurcations (in which both child branches bifurcate) are encoded with the letter 'A', bifurcations with one bifurcating child and one terminating are encoded as 'C' and terminating bifurcations (with two terminating children) as 'T'. This method enables us to align different trees via Sequence-based Tree Alignment, which can be used for the assignment of a similarity score between trees, using cluster analysis. Furthermore, this method is useful for the generation of a

consensus representation (Gillette et al. 2015) from a group of neurons that reveals the conserved structural properties of the corresponding trees. This technique is the closest existing method to the proposed TMD, since it reveals the topological properties that are persistent throughout a group of trees. However, the TMD takes into account the embedding of the tree in space preserving the relation between the short and long components of the tree. Furthermore, the TMD algorithm has a computational advantage over the highly computationally demanding sequence alignment techniques.

Stability of TMD

Let T denote a finite rooted tree with vertex set N containing a distinguished root R, which endows each edge of T with a natural orientation away from the R. Let $f : N \to \mathbb{R}$ be any function satisfying $f(n) > f(R)$ for all $n \neq R$ in N, i.e., f takes its lowest value at the root R. A pair (T, f), where T is a rooted tree (not assumed to be embedded in any ambient space) and f is a function satisfying the condition above, is referred to as a *TMD-pair*.

In this section we prove that the *TMD* algorithm that associates with a TMD-pair (T, f) a persistence diagram $\text{TMD}(T, f)$ is robust under the type of perturbations of the tree T and the function f that are most likely to arise in the *reconstruction process*, i.e., the transformation of a physical tree-like object, such as a neuron, into input data for the TMD algorithm. We consider two types of reconstruction errors:

E1. error in measuring the exact placement of a node, and

E2. omission or addition of a small branch.

Errors of type **E2** may have the effect of changing the tree considered, which implies that the function f defined on its nodes takes on new values or loses a few of its previous values. Errors of type **E1** may affect the values of the function f on the nodes of the tree, though the abstract graph underlying the tree remains the same.

We now define four types of perturbations of TMD-pairs that will be considered admissible for our purposes. If T is a tree, then by *"adding a branch"* to T, we mean attaching a new branch to any node of T or adding a node to the interior of an existing branch of T and attaching a new branch to that node.

Definition 1 *Fix a TMD-pair (T, f) and a real number $\epsilon > 0$. An elementary ϵ-perturbation of (T, f) is a TMD-pair (T', f') obtained from (T, f) by one of the following operations.*

T1. $T = T'$, $f(R) = f'(R)$, and for all $n \neq R$, $|f'(n) - f(n)| < \epsilon$.

T2. T' is obtained from T by adding a branch at a node n of T, $f'(n) = f(n)$, and $|f'(n') - f(n)| < \epsilon$, where n' is the added leaf (univalent node). The restriction of f' to the nodes of T is equal to f.

T3. T' is obtained from T by adding an internal node n' on an existing edge in T, with incident nodes u and v, and a branch at n' with leaf n'', such that $|f'(n') - f'(n'')| < \epsilon$, while $f'(n')$ lies between $f(u)$ and $f(v)$, or $|f'(n') - f(u)| < \epsilon$, or $|f'(n') - f(v)| < \epsilon$. The restriction of f' to the nodes of T is equal to f.

T4. T' is obtained from T by removing a branch with incident nodes n', n'', where n'' is a leaf, such that $|f(n') - f(n'')| < \epsilon$. The function f' is the restriction of f to T'.

A TMD-pair (T', f') is said to be an ϵ-perturbation of (T, f) if (T', f') is obtained from (T, f) by

i) performing operations of type **T1** on a subset of the set of nodes of T, and then

ii) performing a finite number of operations of types **T2**, **T3**, and **T4** on the resulting tree, such that every branch that is present in T' but not in T is a leaf, and the following condition holds.

— If nodes $\{v_i\}_{i=1}^t$ are added via operations of type **T3** to a branch in T with incident nodes u and v, then the deviation from linear order of the values $f'(v_i)$ according to the position of the v_i on the branch is smaller than ϵ for every pair of adjacent nodes.

Let $P_\epsilon(T, f) = \{(T', f') \mid (T', f') \text{ is an } \epsilon\text{-perturbation of } (T, f)\}$

Example 1 Let T be a rooted tree embedded in $\mathbb{R}^3$, and let f be the real-valued function that assigns to a node n in T its Euclidean distance to the root R. An elementary ϵ-perturbation of type **T1** corresponds to moving nodes in space by at most ϵ. Elementary perturbations of types **T2**, **T3** and **T4** correspond to removing branches from or adding branches to T, such that the distance between their nodes is at most ϵ. At the end of this section we observe that in fact any TMD-pair can be thought of as arising in this way.

The following definition is standard in the literature.

Definition 2 *Let T be a rooted tree with root R. The depth of a node n in T is the number of edges in the unique path connecting n to R. The depth of a tree T is the maximum depth of a node in T.*

A tree of depth 1 is said to be a *corolla*. Let T be a corolla with root R and leaves $l_1, \ldots l_m$. Let m_i denote the multiplicity of the value $f(l_i)$, for $1 \le i \le k$, where $l_1, ..l_k$ are the leaves on the function f assumes distinct values. The persistence diagram associated to (T, f) through the TMD algorithm has the form

$$TMD(T, f) = \left\{ (f(l_i), f(R))^{m_i} \mid 1 \le i \le k \right\} \cup D,$$

where $(x, y)^j$ denotes the point (x, y) with multiplicity j, and D is the diagonal. If the values of function f are non-negative, then it suffices to consider D as the first quadrant diagonal.
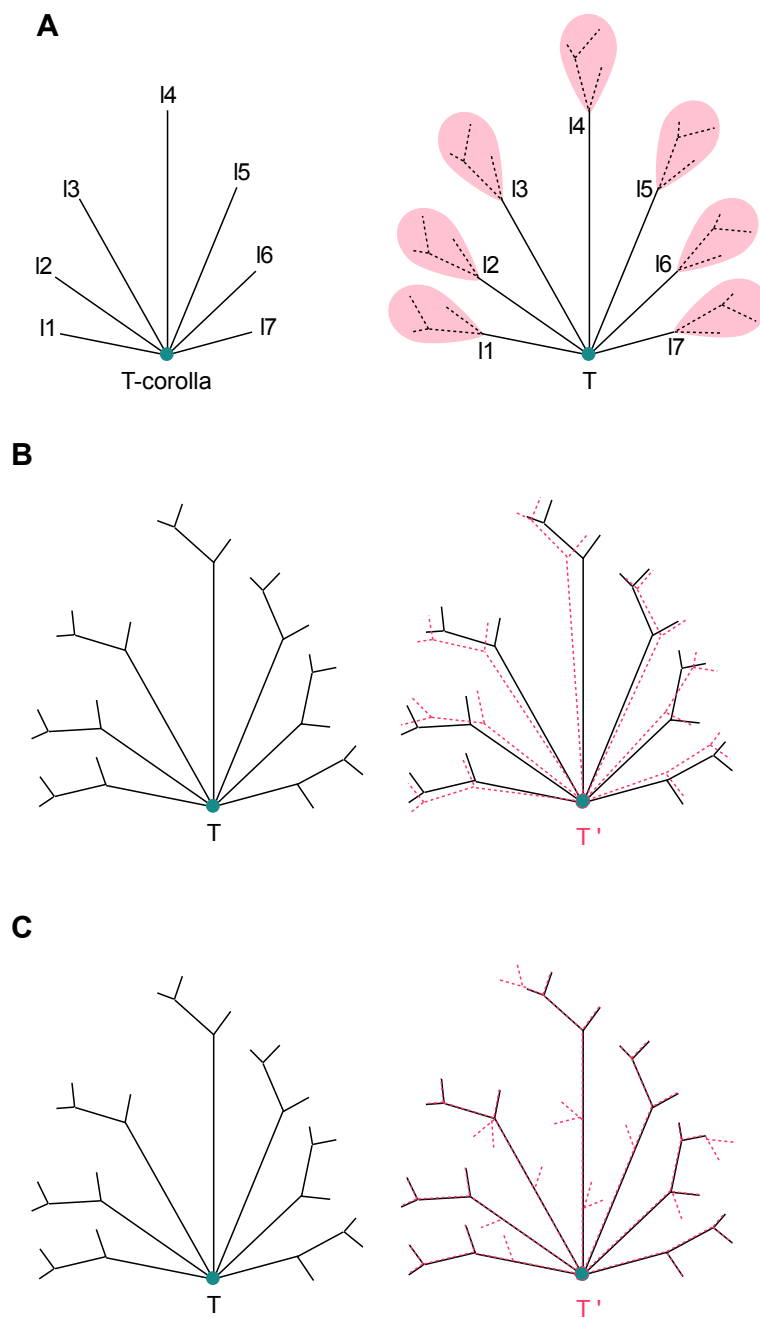
Fig. A.5 Proof of stability illustration. A. Illustration of a T-corolla tree and an associated tree with basis the corolla tree. B.- C. Small perturbations in the structure of the tree: B) type E1: the node's h position has been moved by ϵ to the node h'. This figure illustrates one of the possible perturbations, that will be used for the proof of stability of the TMD algorithm. C) type E2: ϵ small branches have been added to the tree. This figure illustrates one of the possible perturbations, that will be used for the proof of stability of the TMD algorithm.

Let T be any tree of depth h and root R. For a node $n \in T$, recall that T_n denotes the subtree of T starting at n, considered as a rooted tree with root n. In particular $T_R = T$ and for any $n \neq R$, T_n is a tree of depth strictly less than h.

Let $n_1, \ldots n_m$ be the nodes in T of depth 1 (i.e., the children of R). For every i, let

$$b_i = \max\{x \mid (x, y) \in TMD(T_{n_i}, f)\}$$

(i.e., b_i is the largest value of f on a node of T_{n_i}). Then for each i the point $(b_i, f(n_i))$ is in the persistence diagram $\mathrm{TMD}(T_{n_i}, f)$, and one easily observes that

$$TMD(T, f) = \{(b_i, f(R)) \mid 1 \leq i \leq m\} \sqcup \coprod_{i=1}^{m} TMD(T_{n_i}, f) \setminus \{(b_i, f(n_i)) \mid 1 \leq i \leq m\}$$

We can now establish the stability of the TMD algorithm with respect to bottleneck distance under ϵ-perturbations of TMD-pairs.

Theorem 1 *Let (T, f) be a TMD-pair, and let $\epsilon > 0$. If (T', f') is an ϵ-perturbation of (T, f), then*
$$d_B(TMD(T, f), TMD(T', f')) \leq 3\epsilon.$$

Proof 1 *The proof proceeds by induction on the depth of T, separating the cases in which T' is obtained from T through operations of type **T1**, **T2**, or **T3**. Since any set of operations of type **T4** reverses a corresponding set of operations of types **T2** and **T3**, and since bottleneck distance is a metric (and hence symmetric), the effect of perturbations of type **T4** will be discussed only briefly.*

Perturbations of type **T1**.

*If (T', f') is a TMD-pair obtained from (T, f) by perturbations of type **T1**, then the depth of T is equal to the depth of T'. For every node n in T, we denote by n' the corresponding node in T'. To compute an upper bound on the bottleneck distance between TMD(T, f) and TMD(T', f'), we construct a specific type of matching between their persistence diagrams. Recall that*

$$TMD(T, f) = \{(b_i, f(R)) \mid 1 \leq i \leq m\} \sqcup \coprod_{i=1}^{m} TMD(T_{n_i}, f) \setminus \{(b_i, f(n_i)) \mid 1 \leq i \leq m\}$$

and

$$TMD(T', f') = \{(b'_i, f'(R')) \mid 1 \le i \le m\} \sqcup \coprod_{i=1}^{m} TMD(T'_{n'_i}, f') \setminus \{(b'_i, f'(n'_i)) \mid 1 \le i \le m\}$$

We show by induction on the depth of T that there exists a matching between $TMD(T, f)$ and $TMD(T', f')$ such that $(b_i, f(R))$ is matched with $(b'_i, f'(R'))$ for every i and such that the L_∞-distance between each pair of matched points is less than ϵ, from which we deduce that the bottleneck distance bewteen $TMD(T, f)$ and $TMD(T', f')$ is also less than ϵ.

For the base step of the induction we consider a corolla T, with root R and leaves $l_1, \ldots, l_m$, whence T' is also a corolla with root R' and leaves $l'_1, \ldots, l'_m$. It follows that

$$TMD(T, f) = \left\{ (u_i, f(R))^{m_i} \mid 1 \le i \le k \right\} \cup D,$$

and

$$TMD(T', f') = \left\{ (u'_i, f'(R'))^{m'_i} \mid 1 \le i \le k' \right\} \cup D,$$

where $\{u_i \mid 1 \le i \le k\}$ is the set of values of f on the nodes of T (other than the root R), and m_i is the multiplicity of u_i, for $1 \le i \le k$, while $\{u'_i \mid 1 \le i \le k'\}$ is the set of values of f' on the nodes of T' (other than the root R'), and m'_i denotes the multiplicity of the value u'_i, for $1 \le i \le k'$.

Condition $\mathbf{T1}$ implies that $|f(l_i) - f'(l'_i)| < \epsilon$ for all $1 \le i \le m$ and $|f(R) - f'(R')| < \epsilon$ and thus the L_∞-distance between the points $(f(l_i), f(R))$ and $(f'(l'_i), f'(R'))$ is less than ϵ. Matching $(f(l_i), f(R))$ with $(f'(l'_i), f'(R'))$ for every $1 \le i \le m$, we see that ϵ is an upper bound on the bottleneck distance between $TMD(T, f)$ and $TMD(T', f')$ in this case, i.e.,

$$d_B(TMD(T, f), TMD(T', f')) < \epsilon.$$

The constructed matching is of the desired type.

Suppose now that the inductive hypothesis holds for all TMD-pairs (T, f), where T is a tree of depth less than h, and all $(T', f') \in P_\epsilon(T, f)$ obtained by perturbations

of type **T1***. Let* (T, f) *be a TMD-pair where* T *is a tree of depth* h*. Assume that* $(T', f') \in P_\epsilon(T, f)$ *is obtained by perturbations of type* **T1** *from* (T, f)*.*

For each $1 \leq i \leq m$*, let*

$$C = \{(b_i, f(R)) \mid 1 \leq i \leq m\},$$

$$C' = \{(b'_i, f'(R')) \mid 1 \leq i \leq m\},$$

$$D_i = TMD(T_{n_i}, f) \setminus \{(b_i, f(n_i))\}, \ \text{and}$$

$$D'_i = TMD(T'_{n'_i}, f') \setminus \{(b'_i, f'(n'_i))\}.$$

Matchings between D_i *and* D'_i *for every* i *and a matching between* $C \cup D$ *and* $C' \cup D$ *together give rise to a matching between* $TMD(T, f)$ *and* $TMD(T', f')$*, from which we can compute an upper bound on* $d_B(TMD(T, f), TMD(T', f'))$*.*

Since $(T'_{n'_i}, f')$ *is an* ϵ*-perturbation of* (T_{n_i}, f) *of type* **T1** *for all* i*, and each* T_i *is of depth less than* h*, the inductive hypothesis implies that for all* i*, there is a matching between* $TMD(T_{n_i}, f)$ *and* $TMD(T'_{n'_i}, f')$ *such that* $(b_i, f(n_i))$ *is matched with* $(b'_i, f'(n'_i))$ *and such that the* L_∞*-distance between each pair of matched points is less than* ϵ*. By removing the matched pairs of points* $(b_i, f(n_i))$ *and* $(b'_i, f'(n'_i))$*, we obtain a matching between* D_i *and* D'_i *such that the* L_∞*-distance between every pair of matched points is less than* ϵ*. Moreover, the argument for the corolla case shows that there is a matching between* $C \cup D$ *and* $C' \cup D$ *that matches* $(b_i, f(R))$ *with* $(b'_i, f'(R'))$ *for every* i *and such that the* L_∞*-distance between every pair of matched points is less than* ϵ*. The union of these two matchings gives rise to the desired matching between* $TMD(T, f)$ *and* $TMD(T', f')$ *that satisfies the inductive hypothesis. In particular,*

$$d_B(TMD(T, f), TMD(T', f')) \leq \max\left(\{d_B(D_i, D'_i) \mid 1 \leq i \leq m\} \cup \{d_B(C, C')\}\right) < \epsilon.$$

Perturbations of type **T2**.

Let $(T', f') \in P_\epsilon(T, f)$ *be a TMD-pair obtained from* (T, f) *by perturbations of type* **T2***. To set our notation, let* $\{n_i\}_{i=1}^m$ *denote the set of all nodes in* T *different from the root* R*. Let* $\{n_i\}_{i=1}^r$ *($r \leq m$) denote the nodes where new branches were added. For each* $1 \leq i \leq r$*, let* $\{u_{i,k}\}_{k=1}^{q_i}$ *denote the new nodes resulting from adding new branches at the node* n_i*. Finally let,* $\{z_s\}_{s=1}^n$ *denote the nodes added to* T' *as a result of adding branches at the root* R*. Thus the nodes in* T' *are*

$$\{R, n_1, \ldots, n_r, n_{r+1}, \ldots, n_m\} \cup \{u_{i,k} \mid 1 \le k \le q_i, 1 \le i \le r\} \cup \{z_s \mid 1 \le s \le n\}.$$

With this notation, Condition **T2** *ensures that $f(R) = f'(R)$, and*

- *for all $1 \le i \le m$, $f'(n_i) = f(n_i)$, and for all $1 \le i \le r$ and $1 \le k \le q_i$, $|f'(u_{i,k}) - f'(n_i)| < \epsilon$, and*

- *for all $1 \le s \le n$, $|f'(R) - f'(z_s)| < \epsilon$.*

As in the previous case, the proof is carried out by induction: we prove the statement first in the case where T is a corolla, and then move on to the general case.

Assume T is a corolla. The persistence diagram for (T, f) has the form:

$$TMD(T, f) = \coprod_{i=1}^{m} (f(n_i), f(R)) \cup D,$$

where D is the diagonal. On the other hand, the persistence diagram for (T', f') has the form:

$$TMD(T', f') = \coprod_{i=r+1}^{m} \left(f'(n_i), f'(R) \right) \sqcup$$

$$\coprod_{i=1}^{r} \left(f'(u_{i,k_i}), f'(R) \right) \sqcup L \sqcup \coprod_{s=1}^{n} \left(f'(z_s), f'(R) \right) \cup D,$$

where for each $1 \le i \le r$, u_{i,k_i} is a node on which f' obtains a maximal value among all nodes $\{u_{i,k}\}_{k=1}^{q_i}$, and L is a collection of points of the form $(n_i, u_{i,j})$ for those $j \ne k_i$ such that $f'(n_i) > f'(u_{i,j})$, and $(u_{i,j}, n_i)$ for $j \ne k_i$ such that $f'(n_i) < f'(u_{i,j})$. There is an obvious matching between the sets

$\coprod_{i=1}^{m}(f(n_i), f(R))$ *in TMD(T, f) and*

$\coprod_{i=r+1}^{m} \left(f'(n_i), f'(R) \right) \sqcup \coprod_{i=1}^{r} \left(f'(u_{i,k_i}), f'(R) \right)$ *in TMD(T', f'),*

and the distance between any pair in this matching is bounded above by ϵ by Condition **T2**. All other points are at L_∞-distance at most ϵ from the diagonal. Hence matching those points to the diagonal gives an upper bound of ϵ on the bottleneck distance in this case.

For the induction step, let T be a tree of depth h with root R, where the nodes of depth 1 are denoted $l_1, \ldots, l_m$. For each $1 \leq i \leq m$, let T_{l_i} denote the subtree of T with root l_i. Let $x_i = \underset{T_{l_i}}{\operatorname{argmax}} f$ for each i. Let $TMD_0(T_{l_i}, f)$ denote the sub-diagram of $TMD(T_{l_i}, f)$ consisting of all points except the unique one with $f(l_i)$ as its y-coordinate. The persistence diagram for T is of the form

$$TMD(T, f) = \coprod_{i=1}^{m} TMD_0(T_{l_i}, f) \sqcup \coprod_{i=1}^{m} (f(x_i), f(R)) \cup D.$$

Let T' be a rooted tree obtained from T with operations of type **T2**. For each i, let T'_{l_i} denote the subtree of T' with root l_i. As above,

$$
\begin{aligned}
TMD(T', f') = \coprod_{i=1}^{m} TMD_0(T'_{l_i}, f') \sqcup \coprod_{i=1}^{m} (f'(y_i), f'(R)) \sqcup \\
\coprod_{s=1}^{n} (f'(z_s), f'(R)) \cup D,
\end{aligned}
$$

where $y_i = \underset{T'_{l_i}}{\operatorname{argmax}}(f')$ for each $1 \leq i \leq m$.

Notice that if D_1, D'_1, D_2, D'_2 are persistence diagrams such that $d_B(D_i, D'_i) \leq \delta$ for some $\delta > 0$ and for $i = 1, 2$, then $d_B(D_1 \sqcup D_2, D'_1 \sqcup D'_2) \leq \delta$. This observation and the induction hypothesis together show that

$$d_B\left(\coprod_{i=1}^{m} TMD_0(T_{l_i}, f), \coprod_{i=1}^{m} TMD_0(T'_{l_i}, f')\right) \leq \epsilon.$$

Clearly, $y_i - x_i \leq \epsilon$ for each $1 \leq i \leq m$. Thus it follows that matching the points $(x_i, f(R))$ and $(y_i, f'(R))$ for each $1 \leq i \leq m$ does not increase the distance between the corresponding sub-diagrams. Finally notice that each point of the form $(f'(z_s), f'(R))$ is of L_∞-distance at most ϵ from the diagonal. Putting these observations together we conclude that

$$d_B(TMD(T, f), TMD(T', f')) \leq \epsilon,$$

as claimed.

Perturbations of type **T3**.

Let $(T', f') \in P_\epsilon(T, f)$ be a TMD-pair obtained from (T, f) by perturbations of type **T3**. *To set our notation for this case, let $\{v_j\}_{j=1}^t$ denote the new (internal) nodes added to T, i.e., the v_j are the nodes in T' where a branching point occurs that is not present in T. For each $1 \leq j \leq t$, let $\{w_{j,l}\}_{l=1}^{p_j}$ denote the new nodes resulting from adding branches at v_j.*

Condition **T3** *ensures that $f(R) = f'(R)$ and that the following statements hold.*

- *For all $1 \leq j \leq t$, $f'(v_j)$ is either an intermediate value between the values of f on the nodes incident to the edge along which v_j was added, or $f'(v_j)$ is no more than ϵ away from the value of f on at least one of those nodes.*

- *For all $1 \leq j \leq t$, and all $1 \leq l \leq p_j$, $|f'(w_{j,l}) - f'(v_j)| < \epsilon$.*

Notice also that the values of f' on new nodes added on a single branch in T satisfy the extra linear ordering condition in Definition 1.

Once more, we start by assuming T is a corolla. As before, in this case,

$$TMD(T, f) = \coprod_{i=1}^m (f(n_i), f(R)) \cup D,$$

where D is the diagonal. For each $1 \leq i \leq m$, let e_i denote the i-th branch in T, and let e'_i denote the branch of T' corresponding to e_i. It follows that e'_i either is identical to e_i or contains one or more new branching points. Notice that $TMD(T, f) = \coprod_{i=1}^m TMD(e_i, f|_{e_i}) \cup D$, and similarly that $TMD(T', f') = \coprod_{i=1}^m TMD(e'_i, f|_{e'_i}) \cup D$. Hence it suffices to prove the claim for $m = 1$, i.e., when T is a corolla with exactly one leaf.

*Let T consist of the root R and a node n with a single edge between them. Let $\{v_j\}_{j=1}^{t}$ denote the internal nodes added in T', and let $\{w_{j,l}\}_{l=1}^{p_j}$ denote the leaves added at v_j. For nodes v_j such that $f'(v_j)$ is not intermediate between $f(n)$ and $f(R)$, condition **T3** guarantees that the value of f' on those nodes and their branches is at most 2ϵ away from $f(n)$. Indeed, notice first that $f'(v_j)$ cannot be smaller than $f(R)$, by hypothesis. Hence the only way for $f'(v_j)$ not to be intermediate is to have $f'(v_j) > f(n)$. If this is the case, then $|f'(w_{j,l}) - f(n)| < 2\epsilon$. On the other hand, for nodes v_j such that $f'(v_j)$ is an intermediate value between $f(n)$ and $f(R)$, the contribution of an added leaf with end node $w_{j,l}$ to TMD(T', f') is easily seen to be L_∞-distance at most ϵ from the diagonal. Thus, let u be a node in T' such that $f'(u)$ is maximal (possibly $u = n$). Then the point $(f'(u), f'(R))$ can be matched with $(f(n), f(R))$. It is now easy to observe that all remaining points in TMD(T', f') are of L_∞-distance at most ϵ from the diagonal, and hence can be matched with diagonal points, so that the claim for the corolla follows.*

*The induction step now follows very similarly to the case of perturbations of type **T2**. Hence in this case we obtain once more*

$$d_B(TMD(T, f), TMD(T', f')) < 2\epsilon.$$

Perturbations of type **T4**.

*Let $(T', f') \in P_\epsilon(T, f)$ be a TMD-pair obtained from (T, f) by perturbations of type **T4**. Reversing the roles, (T, f) is a TMD-pair obtained from (T', f') by perturbations of type **T2** and **T3**. Hence by the discussion of perturbations of these types*

$$d_B(TMD(T, f), TMD(T', f')) < 2\epsilon.$$

Conclusion.

*We are now ready to complete the proof of the theorem. Notice that perturbations of types **T2**, **T3**, and **T4** can be performed on a TMD-pair (T, f) simultaneously without any complications. Hence if (T', f') is a TMD-pair that results from (T, f) by applying these operations, then the bottleneck distance between the corresponding persistence diagrams is bounded above by 2ϵ. Adding perturbations of type **T1** and using the triangle inequality for bottleneck distance, we conclude that*

$$d_B(TMD(T, f), TMD(T', f')) \leq 3\epsilon$$

for all $(T', f') \in P_\epsilon(T, f)$.

Note that errors in the connectivity of the tree are not considered. When the connectivity is modified, the new tree T^* will have a different topology. Therefore it is not possible to ensure that T and T^* will be ϵ-close and as a consequence $TMD(T)$ and $TMD(T^*)$ are not restricted to be ϵ-close.

A geometric interpretation.

We finish the section by observing that Example 1 is generic, in the sense that every TMD-pair (T, f) can be thought of as a rooted tree embedded in $\mathbb{R}^3$, with f the function given by radial distance from the root. Indeed, since we assume that $f(R)$ is the absolute minimal value of f on the nodes of T, there is no loss of generality in assuming that $f(R) = 0$. Since the set of nodes of the T is finite, the function f takes on finitely many values $0 < a_1 < \cdots < a_r$. Identify R with the origin in $\mathbb{R}^3$, and embed the set $f^{-1}(a_i)$ into the sphere of radius a_i about the origin for each i. Connect by a straight line each pair of points corresponding to nodes in T that are connected by an edge. Compactness of a finite union of line segments allows the transformation of this into an embedding by small perturbations, without moving points off of the sphere of radius a_i. The function f is now given by radial distance from the origin.
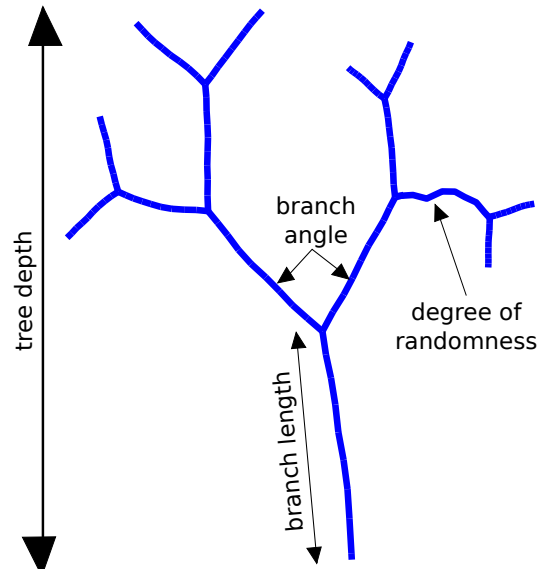
Random trees

Random tree generation

Fig. A.6 Random tree generation. Definition of growth parameters of artificial random trees: each tree is a perfect binary tree, which consists of branch points and leaves. A random walk defines the edges that connect pairs of points on the tree. The order of a branch is defined as the number of bifurcations between the branch point (or leaf) and the root. The tree depth is the maximum branch order of a tree. The branch length is the length of each edge. The branch angle defines the bifurcation angle between two children of a branch point. The degree of randomness indicates if the edge is a straight line or a simple random walk.

The random trees that were used for testing the TMD algorithm's performance were generated with software developed within the Blue Brain Project (BBP). Each tree consists of branches, i.e., paths between two branch points, which are generated based on a simple random walk, SRW, (Pearson 1905) in $\mathbb{R}^3$. The position of the walk at each step is given as a weighted sum of a predefined direction d_n and a simple random walk Ψ:

$$X_{n+1} = X_n + w_s \cdot ((1 - D_r) \cdot d_n + D_r \cdot \Psi),$$

where w_s is the step size, and D_r defines the randomness of each step and Ψ is a random vector in $\mathbb{R}^3$ sampled from a uniform distribution. For $D_r = 0$ the branch is a straight line, while for $D_r = 1$ the branch is a SRW. The number of steps is given by the preselected branch length B_l. Once the number of steps is reached, the tree bifurcates, i.e., two new branches are created. The angle between the initial points

of the branches is defined by the branch angle B_a. The tree generated this way is a perfect binary tree, i.e., every leaf has the same depth, since new branches are added at every branch point until the preselected tree depth T_d (i.e., the maximum number of edges in the unique path from a leaf to the root R) is reached. The total number of branches in the tree is then $2^{T_d} - 1$. For example, the tree in Fig A.6 has $T_d = 4$ and consists of $2^4 - 1 = 15$ branches.

This set of parameters $\{T_d, B_l, B_a, D_r\}$ defines the global properties of the tree. Random trees that are generated with the same set of parameters share common morphometric properties, but have unique spatial structures, due to the stochastic component of the growth. This allowed us to check the effectiveness of the algorithm at identifying sets of trees that have been generated with the same input parameters $\{T_d, B_l, B_a, D_r\}$ and that differ only in the random seed. Random trees constructed with the described algorithm can intersect geometrically, even though the probability of this event is very low. However, for the random tree generation, the connectivity is obtained from the branches of the tree and therefore even if branches intersect geometrically, no cycle will be created in the tree.

Grouping random trees

We defined a control group as a set of trees generated with fixed parameters ($T_d = 5$, $B_l = 10$, $B_a = \pi/4$, $D_r = 10\%$) but independent random seeds. Then, we varied each parameter individually to generate groups of trees that differed from the control group in only one property. For all trees we extracted the persistence barcode using the TMD algorithm. The assignment of a tree to a group based on the comparison of the distances d_{Bar} between the tree's barcode and the barcodes of the trees in every group constitutes one trial. The trial is successful if the tree is correctly assigned to its original group. The performance of the TMD-based classifier in separating groups of trees generated with different values for each of the described parameters is summarized in Table A.4. We cross-validate our method by generating 100 trees for each group, divided into 5 subsets of 20 trees. The standard deviation in Table A.4 shows the statistical significance of our results.

Table A.4 Summary of accuracy results for the classification of random trees.

$T_d : (4, 6, 8)$	$B_a : (\frac{\pi}{4}, \frac{\pi}{2}, \pi)$	$B_l : (5, 10, 30)$	$D_r : (0.1, 0.5, 0.8)$	$A_b : (0.0, 0.3, 0.9)$
$96 \pm 3\%$	$88 \pm 9\%$	$96 \pm 4\%$	$99 \pm 1\%$	$100 \pm 0\%$

The variation of the previous parameters includes only quantitative morphological features. All the generated trees are binary trees. In order to assess the performance of the TMD algorithm at grouping trees with different asymmetries, we generated trees with the same morphological features: number of terminal branches (16), branch lengths(100um), branch angles($\pi/3$) and degree of randomness(0.1) but different degree of asymmetry as defined in (Van Pelt and Verwer 1983). Trees with different degree of asymmetry express different topology of their branching patterns, the probabilities of which are described in (Van Pelt and Verwer 1983). The results of this analysis are presented in Fig A.11 for asymmetries of $0.0, 0.3, 0.9$.

The influence of each morphological feature (tree depth, branch length, branch angles, degree of randomness and degree of asymmetry) on the corresponding persistence barcode is described in detailed in Figures A.7 - A.11. The TMD-based classifier is able to distinguish the variation of all five parameters with significantly high accuracy. This indicates that the TMD-based distance is effectively separating artificial random trees that differ in one of the described morphological properties.
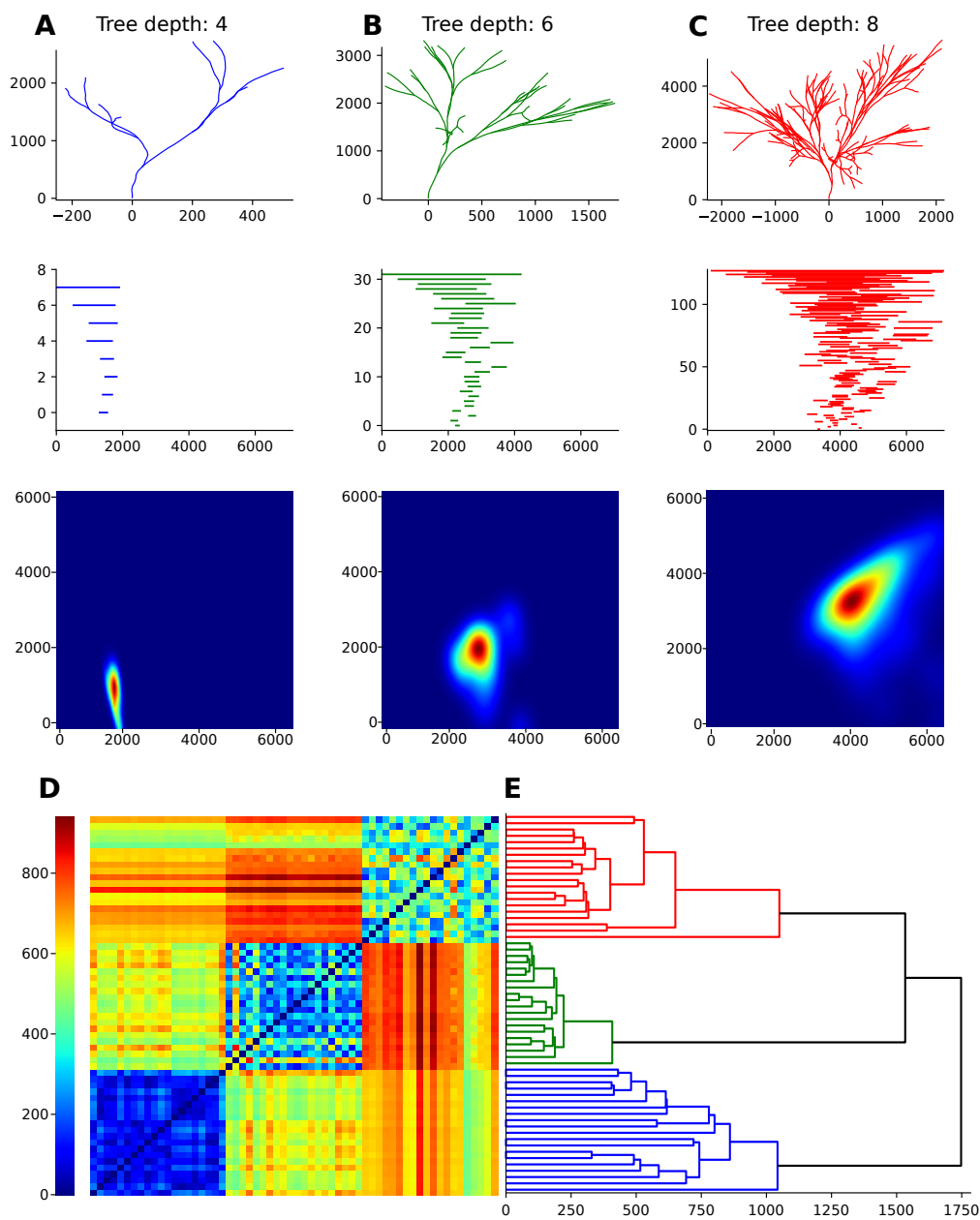
Fig. A.7 Groups of trees with different tree depths (4(A), 6(B), 8(C)) can be effectively separated. Larger tree depths result in larger number of branches on the tree ($N_{branches} = 2^{T_d} - 1$). As a result the density of branches increases with the tree depth, and a larger number of topological components is generated in the respective persistence barcodes. The distance matrix (D) indicates the existence of three groups that are identified with high accuracy by a simple dendrogram (E).
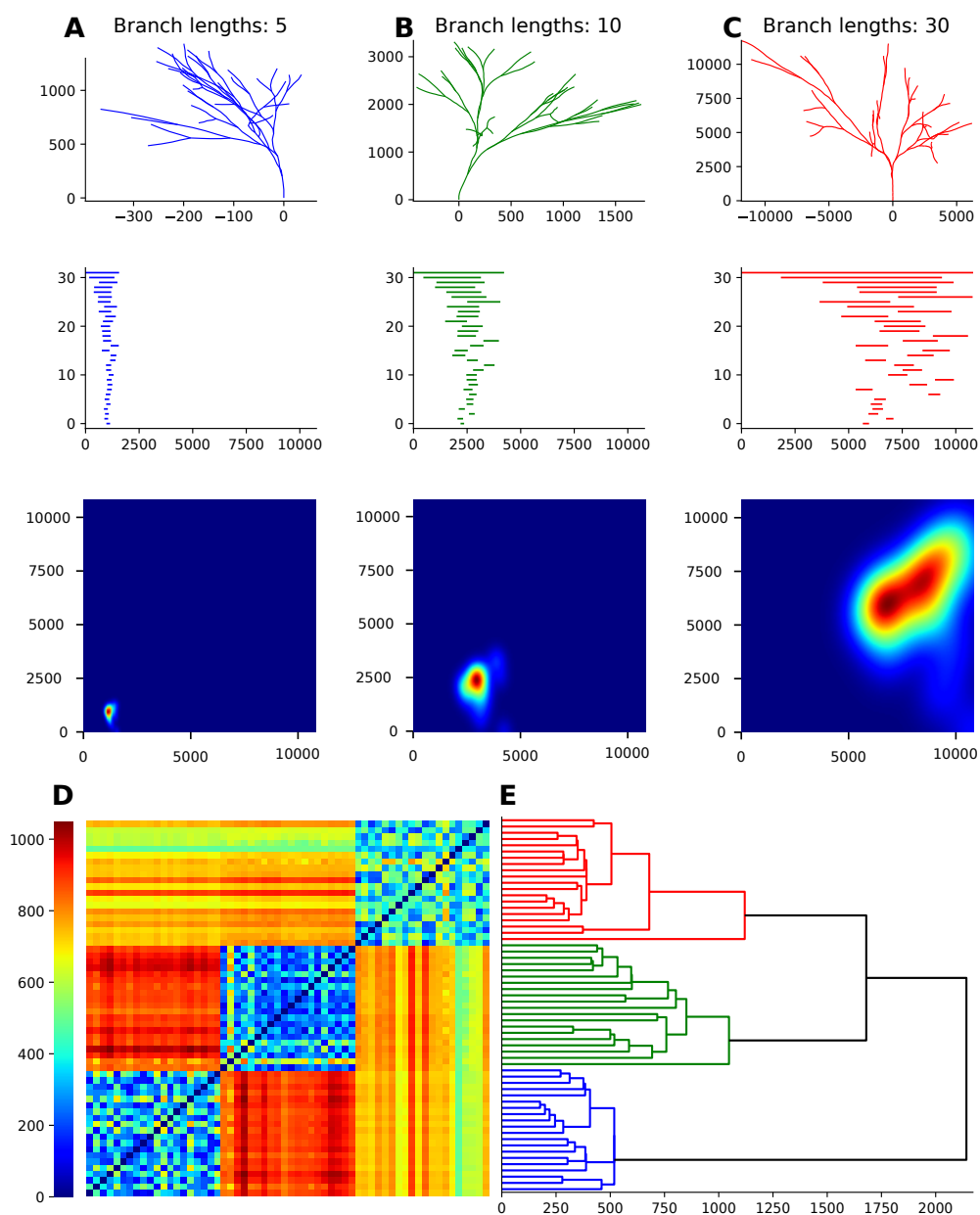
Fig. A.8 Groups of trees with different constant branch lengths (5(A), 10(B), 30(C)) can be effectively separated. The length of the branches is reflected in the lengths of the topological components in the respective persistence barcodes. The increasing branch length results in the presence of bars at larger radial distances. The distance matrix (D) indicates the existence of three groups that are identified with high accuracy by a simple dendrogram (E).
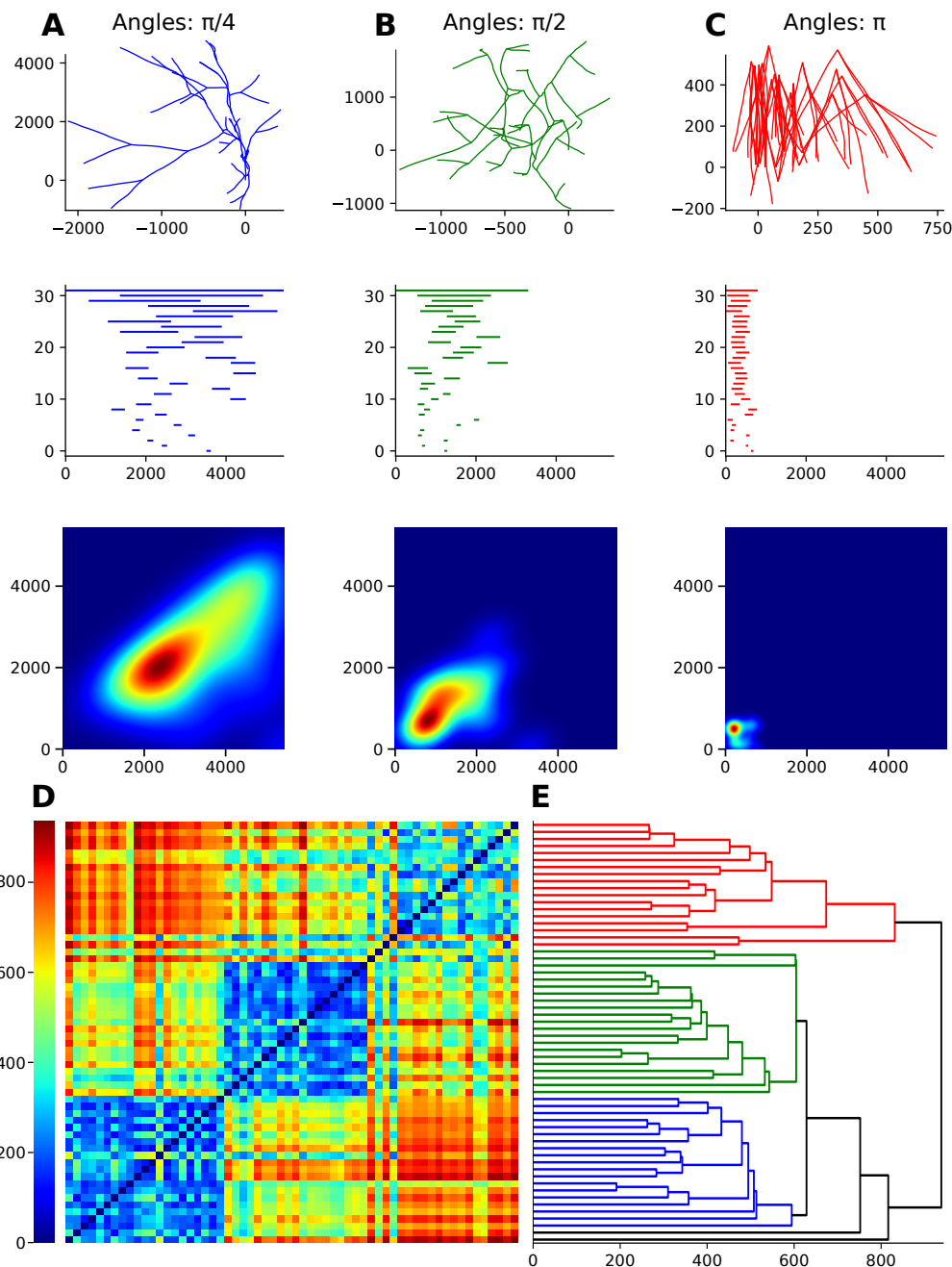
Fig. A.9 Groups of trees with different constant branch angles on the $x - y$ plane ($\pi/4$(A), $\pi/2$(B), π(C)) can be effectively separated. The branch angles influence the radial distances of the branches and as a result their respective persistence barcodes. For smaller branch angles the branches of the trees extend to larger radial distances, resulting in longer bars. The distance matrix (D) indicates the existence of three groups that are identified with high accuracy by a simple dendrogram (E). A few mis-classifications are present in the dendrogram (denoted in black). This fact indicates that this distance is not appropriate for 100% accurate separation of branch angles since the branch angles are not directly accounted for in the TMD algorithm. However, the secondary effects of the branch angles can distinguish the trees with very high accuracy (97%).
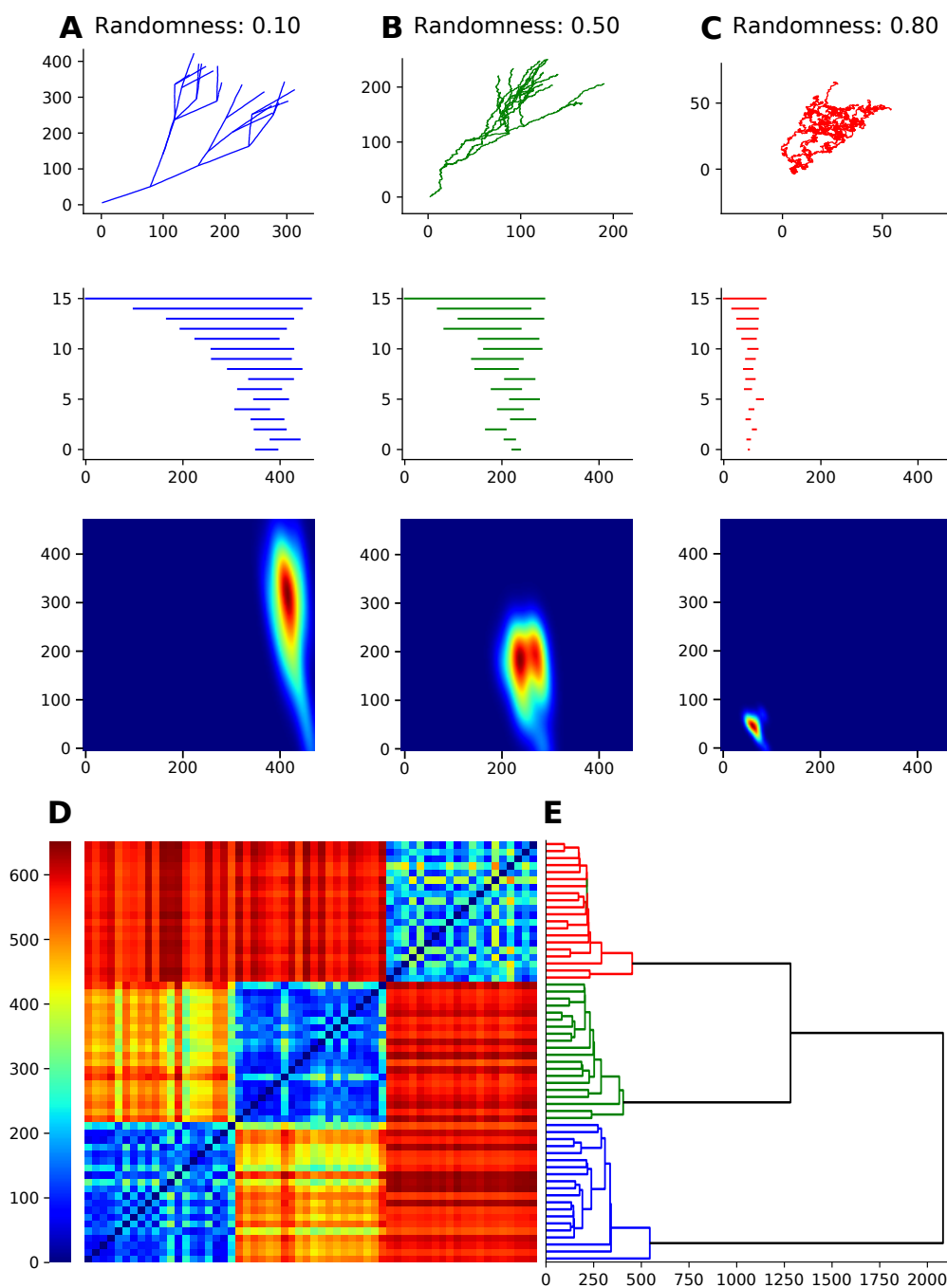
Fig. A.10 Groups of trees with different degrees of randomness (0.10(A), 0.50(B), 0.80(C)) can be effectively separated. The degree of randomness influences the extent of individual branches on the trees. For lower values of randomness the trees are less tortuous and extend to larger radial distances. As a result, the trees with smaller degree of randomness generate longer bars in their respective persistence barcodes. The distance matrix (D) indicates the existence of three groups that are identified with high accuracy by a simple dendrogram (E).

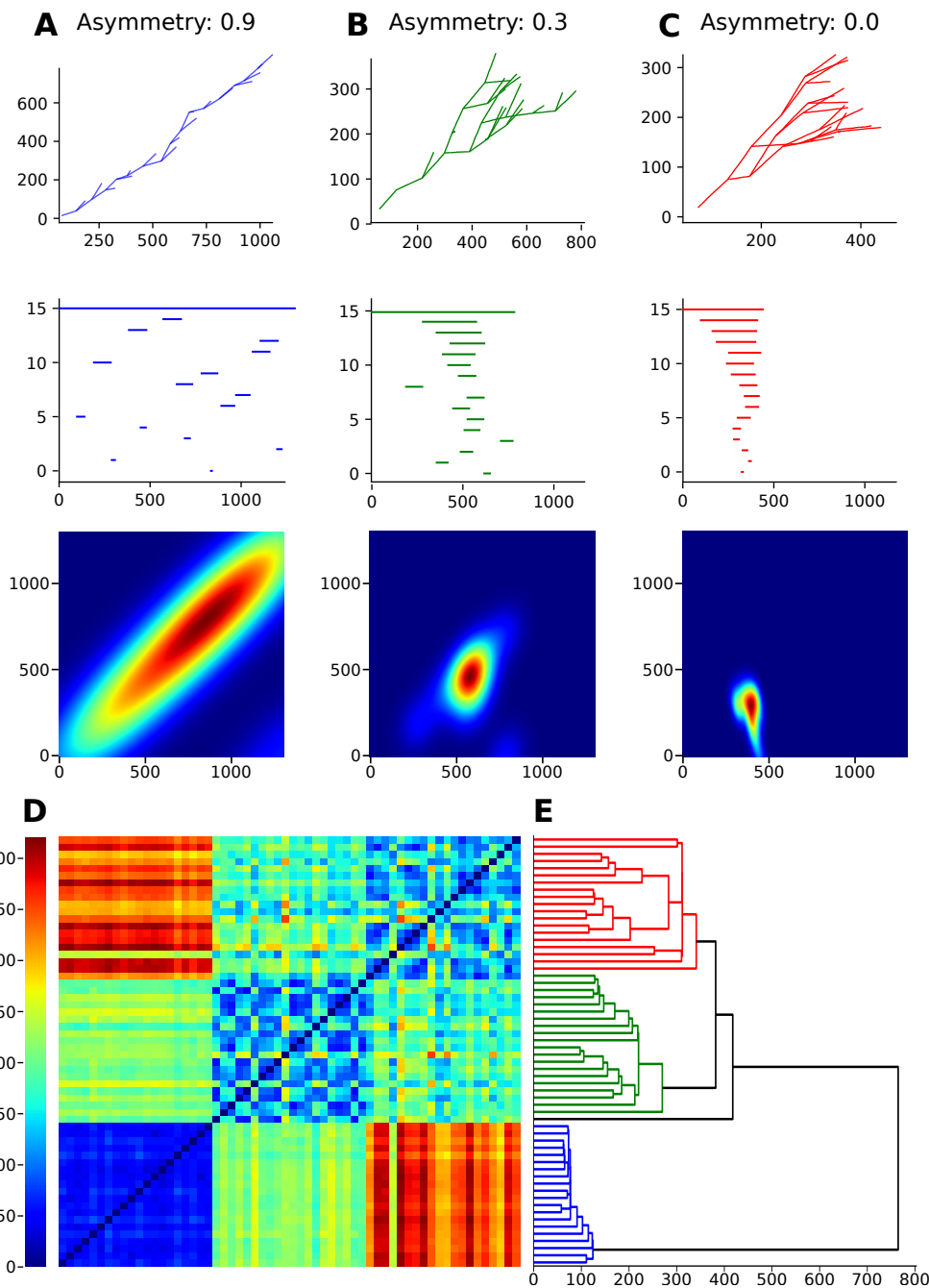Fig. A.11 Groups of trees with different topological patterns that result in different degrees of asymmetry (0.9(A), 0.3(B), 0.0(C)) can be effectively separated. The asymmetry of the branching structure generates distinct patterns in the respective persistence barcodes. Interestingly, the more asymmetric trees (A) result in a more homogeneous distribution of branches in space along the path of the main branch. As a result, the corresponding pesristence images are more symmetric around the diagonal. The asymmetry of the trees is reflected in the barcodes by an inverse relation, as the more symmetric trees are encoded in more skewed barcodes. The distance matrix (D) indicates the existence of three groups that are identified with high accuracy by a simple dendrogram (E).

Supervised Classification

Supervised classification is a machine learning technique in which a sample dataset (training set) is presented to the algorithm, which then predicts the labels of the individuals that have not been presented (test set).
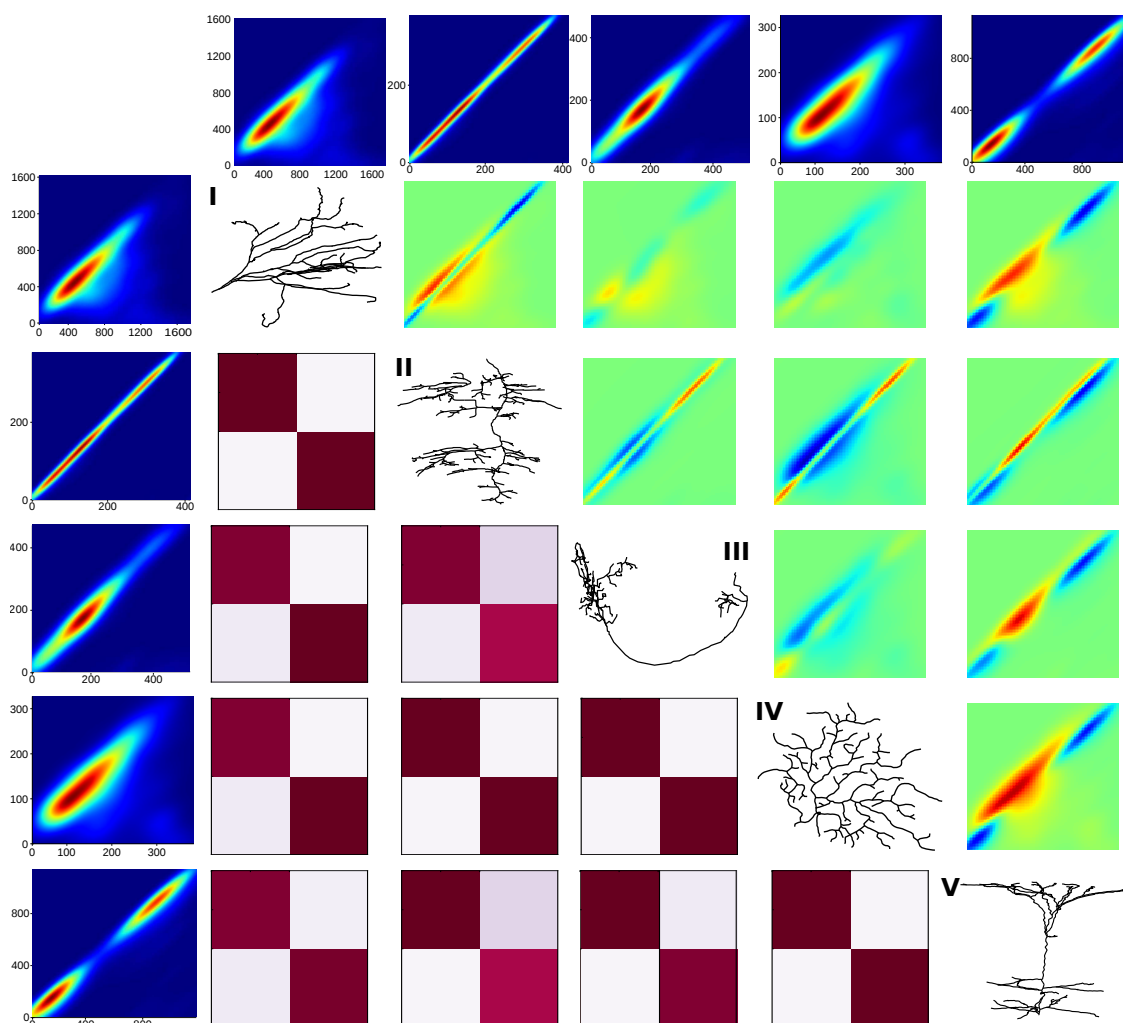

Fig. A.12 Results of supervised classification on the dataset of Fig 4 based on the average unweighted persistence images of neuronal morphologies from different species: (I) cat, (II) dragonfly, (III) fruit fly, (IV) mouse and (V) rat. Traditional classification methods measure the degree of separation between two classes, as opposed to the TMD which also reveals the structural principles that differentiate distinct morphological groups. Below the diagonal we illustrate the separation of each pair of groups by presenting the confusion matrices (color-scale from 0 to 1) for the binary classification of the two groups in question. Above the diagonal we present the structural differences between the two groups, as they are revealed by subtracting their unweighted persistence images. Note that since we are studying the structure and not the size differences, the data are not normalized according to the size of the neurons. As a result, the structural differences are unscaled and the relative sizes are presented in the average unweighted persistence images, on the left.

In this section we present the results of the supervised classification that was performed on the trees of the five groups of neurons from different species that are shown in Fig 2.8. A supervised classification algorithm (Decision Tree) is trained on the unweighted persistence images. The trained classifier is used to predict the class of trees of the test set. The accuracy of the classifier is defined by the number of the correct predictions divided by the total number of predictions.

The results of the classification are presented with the overall accuracy (percentage) and the confusion matrix. The confusion matrix represents the performance of the classification: true positives are presented in the diagonal, where false positives are presented in non-diagonal elements. A perfect classification would result in ones on the diagonal and zero values everywhere else.
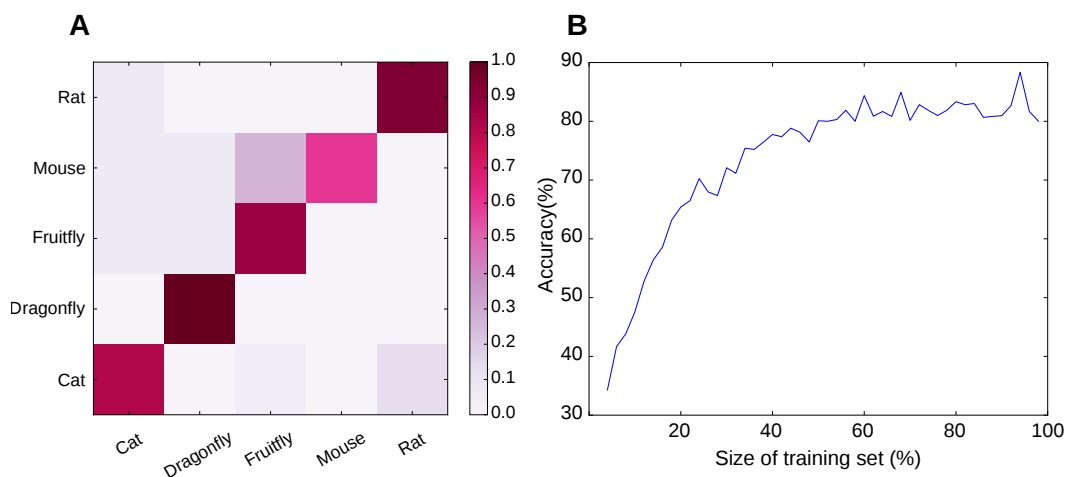

Fig. A.13 Supervised classification of neuronal species. A. Results of supervised classification, trained on the unweighted persistence images of the five groups of neuronal trees presented in Fig 2.8. The confusion matrix represents the performance of the classification: true positives are presented in the diagonal, where false positives are presented in non-diagonal elements. Intense red indicates high fraction of data and white shades indicate small fraction of data that correspond to each element of the matrix. The fact that the diagonal is represented in intense red indicates that in most of the cases the classifier accurately predicts the initial group of the neuronal trees. B. For the same dataset (Fig 2.8) we quantify the accuracy of the supervised classification as the number of correctly predicted labels. The classifier is trained with a subset of the data, as shown in x-axis. As the number of samples that are used for the training is increased the accuracy increases. Note that the accuracy reaches 70% when one fourth of the data (25%, 20 individuals) is used for the training. As a result, a relatively small subset of the data is needed in order to achieve very high accuracy.

Classification of neuronal trees

The average unweighted persistence images were used for the efficient separation of different morphological classes. The hierarchical clustering (Fig A.12) as well as the supervised classification (Fig A.13A) illustrate the clear separation between the neurons

of different species. In addition, by subtracting the persistence images of two groups we can identify the nature of their structural differences, as opposed to traditional methods. Note that the average persistence images have been scaled according to the largest processes for each species in order to illustrate the scale invariant branching properties of each neuronal type.

For example in Fig A.12, we illustrate the spatial differences of the branching patterns of neuronal trees from the different species of Fig 2.8. The dragonfly neurons consist of much smaller processes that generate a high concentration of branches around the diagonal, which are not present in other species. Mouse neurons present a wide variety of branch lengths which result in a wider distribution of points around the diagonal compared to all the other species. The rat pyramidal neurons present a tuft at larger radial distances that differentiates them from the other species.

The results of supervised classification, trained on the unweighted persistence images of the five groups of neuronal trees of Fig 2.8 are shown in Fig A.13A. The higher values in the diagonal of the confusion matrix (true positives) as opposed to small values at the rest of the cases (false positives) indicates that the classifier predicts the actual group of the neuronal trees with high accuracy.

The performance of the classifier as a function of the size of the training set is presented in Fig A.13B for the same dataset (Fig 2.8). As the number of samples that are used for the training increases the accuracy increases accordingly. Note that the accuracy reaches 70% when one fourth of the data (25%, 20 individuals) is used for the training. As a result, a relatively small subset of the data is needed in order to achieve very high accuracy. The classifier based on the unweighted persistence images is capable to predict the class of neuronal trees even when it is trained with very small datasets. This property is very useful for the classification of neuronal trees, where usually only few data of each class are available.

Diversity Index

The diversity index of a community is a quantitative measure that reflects how many different types are present in the dataset and how evenly they are distributed. The diversity increases with the number of types. For a given number of groups, the diversity index is maximized when all groups are equally represented in the dataset. However, most diversity indices behave as if different species had nothing in common.

An alternative method for the characterization of the diversity of a community has been proposed in (Leinster and Cobbold 2012). The diversity profile, i.e., the

graph of the diversity index versus a sensitivity parameter q, describes the shape of the community as the perceived diversity changes with respect to the richness (rare species are influencing the graph for small q) and the dominance of the species (common species almost exclusively define the graph for large q). Therefore, the parameter q represents the inverse of the sensitivity to rare species. The density profile takes into account the *actual similarity* between different groups, as opposed to classical measurements that use the *naive similarity*, i.e., the identity matrix, assuming that different species are completely independent. Based on this method, we generate the diversity profiles of the biological datasets that have been studied in this paper: neurons of five different species (Fig A.14A) and layer five pyramidal cells (Fig A.14B).
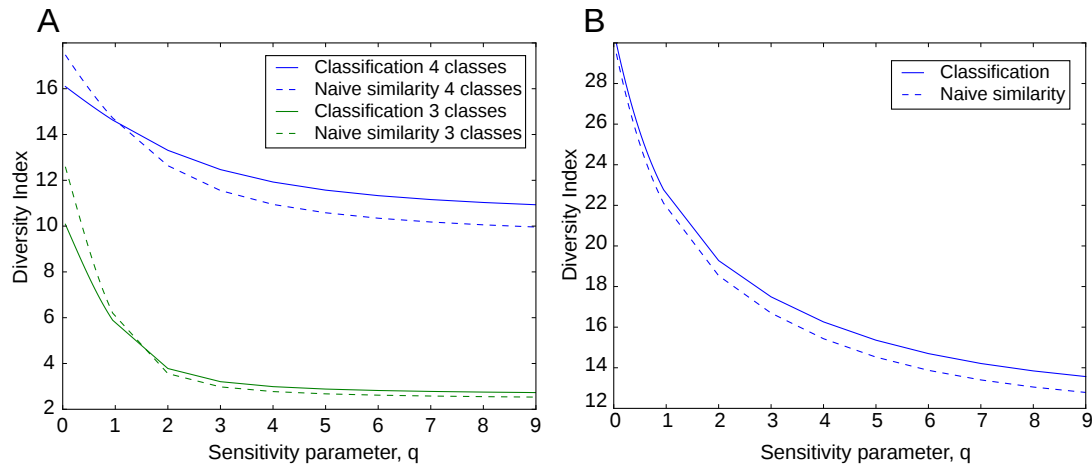

Fig. A.14 A. Diversity indices for the species (A) and the L5 pyramidal cells (B).

For the neurons of different species (Fig A.14 A) the perceived diversity does not change significantly, when we use the actual similarity matrix (solid line) compared to the naive similarity matrix (dashed line). This is due to the fact that neurons of different species are very distinct and therefore their similarity matrix is very close to the identity matrix. It is however interesting to notice that the values of diversity index are much higher in this example compared to the ones of the layer 5 pyramidal cells (Fig A.14 B), indicating that this dataset is indeed more diverse, as expected from visual examination of the neurons.

On the contrary, the diversity profile of layer 5 pyramidal cells (Fig A.14 B) is strongly influenced by the similarity matrix in the case of four classes, while this effect is highly reduced in the case of three classes. This indicates that the classification of the neurons in three classes is much more robust. In this case the classes are more distinct and the similarity matrix is closer to the identity matrix.

References

Beketayev K, Yeliussizov D, Morozov D, Weber G, Hamann B (2014) Topological Methods in Data Analysis and Visualization III: Theory, Algorithms, and Applications, eds.

Bille P (2005) A survey on tree edit distance and related problems. Theoretical Computer Science.

Bremer PT, Hotz I, Pascucci V, Peikert R. (Springer International Publishing, Cham), pp. 151–165.

Bubenik P (2015) Statistical topological data analysis using persistence landscapes. J. Mach. Learn. Res. 16(1):77–102.

Bubenik P, Dłotko P (2015) A persistence landscapes toolbox for topological statistics. Journal of Symbolic Computations. abs/1501.00179.

Dey T, Shi D, Wang Y (2015) Comparing graphs via persistence distortion. SOCG.

Dey T, Shi D (2016) "PersistenceDistortion" software.

DeFelipe J, Lpez-Cruz PL, Benavides-Piccione R, et al. (2013) New insights into the classification and nomenclature of cortical GABAergic interneurons. Nat Rev Neurosci 14(3):202–216.

Gillette TA, Ascoli GA (2015) Topological characterization of neuronal arbor morphology via sequence representation: I - motif analysis. BMC Bioinformatics 16:216.

Gillette T, Hosseini P, Ascoli G (2015) Topological characterization of neuronal arbor morphology via sequence representation: II - global alignment. BMC Bioinformatics 16:209.

Kerber M, Morozov D, Nigmetov A (2016) Geometry Helps to Compare Persistence Diagrams. pp. 103–112.

Ledderose J, Sencion L, Salgado H, Arias-Carrion O, Trevino M, (2014) A software tool for the analysis of neuronal morphology data. Int Arch Med. 7:6.

Leinster T, Cobbold C (2012) Measuring diversity: the importance of species similarity. Ecology 93(3):477–489.

Morozov D, Beketayev K, Weber G (2013) Interleaving distance between merge trees. Discrete and Computational Geometry 49:22–45.

Morozov D (2016) Dionysus, http://mrzv.org/software/dionysus/.

Pearson K (1905) The Problem of the Random Walk. Nature 72(1865):294.

Shapira D, Storer JA (2011) Edit Distance with Block Deletions (Algorithms), 4(1) pp. 40–60.

Sholl DA, (1953) Dendritic organization in the neurons of the visual and motor cortices of the cat. J Anat. 87: 387–406.

Strahler AN, (1952) Hypsometric analysis of erosional topography. Bull Geol Soc Am 63:1117–1142.

Van Pelt J, Verwer RW (1983) The exact probabilities of branching patterns under terminal and segmental growth hypotheses. Bull Math Biol. 45(2):269–85.

Van Pelt J, Verwer RW, Uylings HBM (1989) Centrifugal-order distributions in binary topological trees. Bull Math Biol. 51(4):511–536.

Wan Y, Long F, Qu L, Xiao H, Hawrylycz M, Myers EW, Peng H (2015) Blast-neuron for automated comparison, retrieval and clustering of 3d neuron morphologies. Neuroinformatics 13(4):487–499.

Appendix B

Of mouse and human cells

The topological profiles of mouse cells
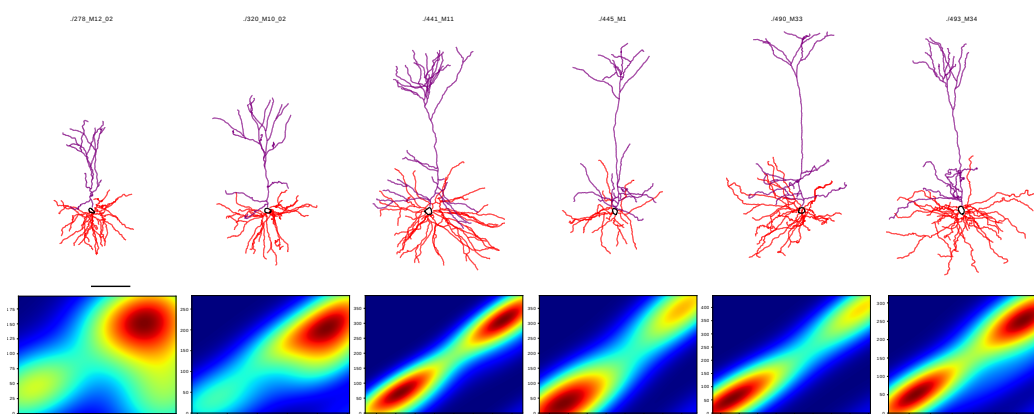
Fig. B.1 Examples of mouse L2/3 pyramidal cells. According to the topological profiles of the mouse cells (and the TMD analysis as described for the human pyramidal cells) we do not find sufficient evidence to support the existence of two classes in the mouse dataset. In agreement with literature data the L2/3 mouse pyramidal cells seem to express a single morphological type, that consists of few obliques and a dense tuft. The variability of quantitative features within this type is large: for example the branches of superficial cells extend to smaller distances from the soma (note the different scales on the persistent images) and the location of the tuft varies according to the depth of the cell in the tissue.

Comparison between mouse and human cells
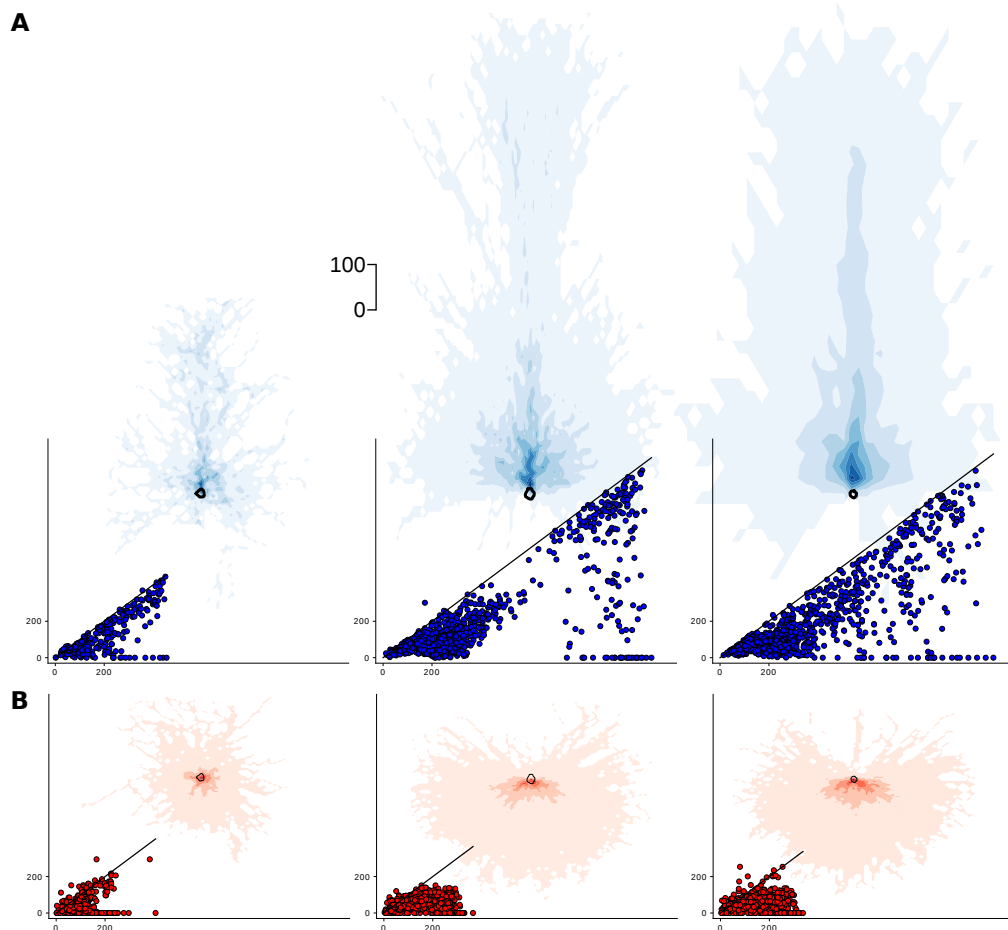
Fig. B.2 Comparison between mouse pyramidal cells and human pyramidal cells of two types. A. Compared to the human pyramidal cells, the mouse apicals have in principle a smaller number of obliques. The slim tufted human cells possess a smaller number of tuft branches than the mouse cells, while the profuse-tufted possess a larger number of tuft branches. It is interesting to note that the human apical dendrites of both classes present a larger variability of branch lengths compared to the mouse apical dendrites. B. There is no evidence for morphological differences between the basal dendrites of the two human pyramidal cell types. The basal dendrites of human cells are typically fewer in number compared to the mouse cells, but present a larger variability of lengths, with larger branches, and comparable extends from the soma. Note that the mouse cells appear more symmetric while the human cells appear to have an orientation preference towards the direction opposite to the pia, a property that we cannot justify at the moment.

Appendix C

Topological synthesis of cortical pyramidal cells

Topological morphology descriptor algorithm

We briefly explain the TMD algorithm that generates a persistence barcode *Barcode* from a tree T; the details of the algorithm are given in (Kanari et al. 2017). The TMD algorithm encodes the branching pattern of the morphology into a topological representation. The local fluctuations with little information content, such as the position of the nodes between branch points, are discarded, and thus the computational complexity of the tree is significantly reduced. The TMD algorithm couples the topology of the branching structure with the geometric properties of the tree (in this case the radial distance from the soma), encoding its over-all shape into a single descriptor. The algorithm takes as input the set of branch points (nodes with more than one child) and leaves (nodes with no children) of the tree and produces a multi-set of intervals, i.e., bars, on the real line known as a persistence barcode (Carlsson 2009), Fig. C.1B. Each bar encodes the lifetime of a component in the underlying structure, identifying when a branch is first detected (birth) and when it connects to a larger subtree (death). Equivalently the persistence diagram (Carlsson 2009), Fig. C.1C, represents the bars of birth-death times of each component as a point in the real plane. These representations greatly simplify the mathematical complexity of the tree.

The main concept of the TMD algorithm is presented in Fig C.1. The TMD algorithm takes as input a rooted tree T with a function f defined on the set of the tree's nodes. In this example, the function f is the radial distance from the soma. The root, denoted by R, is shown in red, and the rest of the tree's nodes (branch points and terminations, or leaves) that are labeled $a - i$, are shown in green. All the leaves
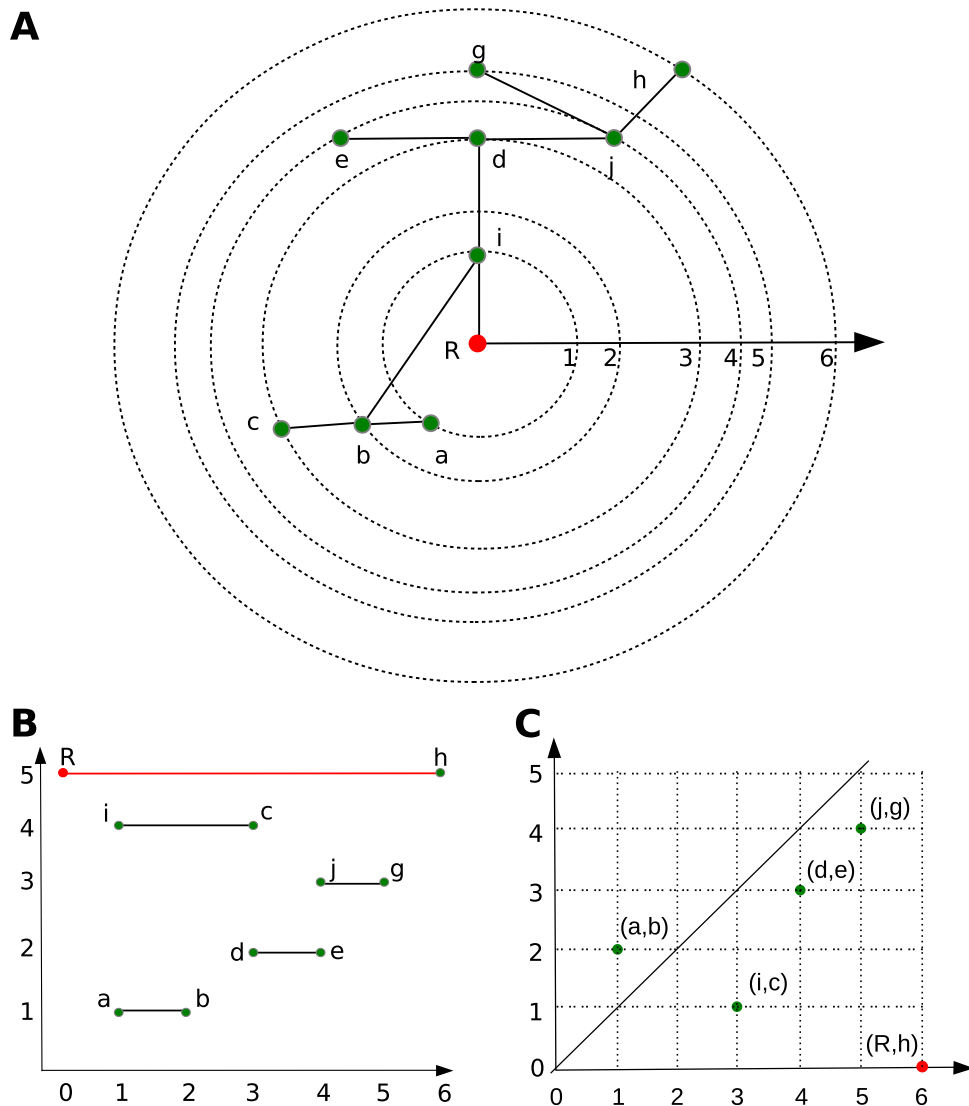
Fig. C.1 Demonstration of the TMD algorithm (Figure published in the supplementary information of "A topological representation of branching neuronal morphologies", Kanari et al. 2017, Neuroinformatics. A rooted tree (A) is transformed with the TMD algorithm into the corresponding persistence barcode (B) and the equivalent persistence diagram (C). The root R is colored red, while the branch points and leaves are shown in green. The edges connecting corresponding pairs of points are presented by straight lines. The dashed circles are provided as a guide to the eye to indicate different levels of radial distances. The correspondence between the tree (A) and its extracted barcode (B) and its diagram (C) is given by the notation of the same nodes in both figures. Each bar in (B) represents the lifetime of a component. The positions of x-axis correspond to the circles in (A) while y-axis represents individual components, ordered according to their length. In (C) each point represents the start and end radial distance of a branch component in A. The longest component is shown in red (B,C).

(a, c, e, g, h) are initially inserted into the list of active nodes, A. The algorithm then iterates over the active nodes A and traverses the tree towards the root. At each branch

point, for example at node b, the children nodes, (a, c), are checked. The node of the minimum value f, in this case radial distance from the soma, forms a bar (a, b) that is added to the *Barcode*. Each bar encodes the starting radial distance of the component a and the ending radial distance b. This process is iteratively performed until the root R is reached. The component that survives until the root is the longest component of the tree (shown in red).

Synthesis input

A set of neuronal reconstructions is used as input to the synthesis algorithm. From the input neuronal trees, the persistence barcodes are generated. Along with the topological profiles of the neurons, a set of basic morphometrics, related to the features of the soma and the thickness of the tree, are also extracted. Those morphometrics are summarized in the "Algorithm input distributions". Apart from the biological inputs a set of user-defined parameters, described in "Algorithm input parameters", are also used as input for the TNS algorithm.

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 in other references, but the morphological terminology is often not consistent through the literature.

> ### Definition of morphological terms.
>
> Soma: the cell body is described as a sphere $S_{d_s}^c$ of diameter d_s and center c.
>
> ---
>
> Neurite: A neuronal tree.
>
> ---
>
> Neurite point: (x, y, z, d), where (x, y, z) are the coordinates in $3D$ space and d is the diameter that represents the thickness of the neurite at that point.
>
> ---
>
> Neurite section: a list of points in the neurite, that are between two branch points of between a branch point and a termination. A section can also be referred to as a branch.
>
> ---
>
> Neurite tips: The collection of termination points of a neurite.
>
> ---
>
> Neurite trunk: the initial section of a neurite, as it emerges from the soma.
>
> ---
>
> *Vect*: A spherical unit vector, which is equivalently represented by a pair of angles and defines a direction, or orientation, in $3D$ space.

Biological persistence barcodes

The algorithm that extracts a persistence barcode 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. At the point where the component terminates the bifurcation angle A with its parent is encoded. Therefore for each component of the tree T, a $Bar = (B, T, A)$ is defined, where B is the initial radial distance of the component, T is the terminal radial distance and A is the bifurcation angle at which the component emerges from its parent. In order to use a population of neurons as input for synthesis, the barcodes of all the biological trees are extracted.

$$BARcodes = \{Barcode_j | 1 \le j \le n\},$$

where $n = \#$ of neuronal trees in the biological population.

This distribution of persistence barcodes is sampled during the artificial growth of a neuron and a single barcode, which corresponds to a biological tree is used for the generation of a synthesized neuronal tree.

$$Barcode_j = \{Bar_i = (B_i, T_i, A_i) | 1 \leq i \leq b_j\},$$

where $b_j = \#$ of components (bars) of $Barcode_j$

Biological distributions of morphological features

Along with the persistence barcodes, which encode the topology of the neurites, a number of independent morphometrics also need to be measured. The first morphometrics that will be used as input for synthesis are the ones that define the size of the cell body, or soma. The soma is initially considered to be a sphere and therefore a center and a diameter are sufficient to describe it. The center is passed as a user defined input, while the diameter is sampled from the corresponding biological distribution SS.

Algorithm input distributions.
Soma parameters
SS: Distribution of soma diameters.
NN: Distribution of number of neurites of a specific type within a neuron.
PA: Distribution of unit vectors that define pairwise angles between neurites.
Diameter parameters
TD: Distribution of the diameters of the tips, or terminations, of the neurite.
TR: Distribution of taper rates that define the tapering, i.e., the difference of the diameters normalized by the length, within a section of the neurite: $(D_{final} - D_{initial})/length$. This value is actually corrected so that the mean value of the diameters within a section is preserved.
RR: Distribution of Rall ratios n that define the ratio between the diameter of a parent and its children. The Rall ratio n is the exponent for which $D^n = d_1^n + d_2^n + \ldots$ at a branch point, where d_i are the diameters of the children and D is the parent diameter.
MD: Distribution of maximum diameters of each neurite within a neurite.

Input parameters

A few input parameters are used to define the properties that are not measured as input from the biological dataset. The first two parameters τ, ρ define the properties of the elongation within a branch. The center of the soma c is an input parameter to allow the user to control the initial position of the cell. The apical point distance D_{apical} is initially a user-defined parameter. However, we should compute this from the biological barcode in future versions of the synthesis algorithm.

Algorithm inputs parameters.
τ: weight of the targeting bias that is used for the elongation of a section.
ρ: the weight of the random component that is used for the elongation of a section.
c: defines the center of the soma, and the starting point for the growth of the neuron.
D_{apical}: radial distance of the apical point from the soma. This distance is used to modify the oblique branching method used during the growth of an apical tree to the tuft branching method.

Topological Neuron Synthesis algorithms

Algorithm 2 Synthesizer

Input:

  $Bio = \{SS, NN, PA, TD, TR, RR, MD\}$ ▷ (see Biological distributions)

  $Param = \{c, \tau, \rho\}$ ▷ (see Input Parameters)

  $BARcodes$ ▷ (see Biological barcodes)

  function Sample(distr) := draws from input distribution

  Generate a Soma and Neurites using (Alg 2, Bio, c)   ▷ (each neurite is initialized with a point on the soma surface, which defines a direction dir_1)

  for neurite in Neurites **do**

      $Active \leftarrow$ neurite's initial point

      $Barcode = $ Sample($BARcodes$)

      Sort $Bars$ in $Barcode$ from longest to shortest

      Initialize first section S_1 with the longest Bar_1

      while $Active$ sections **do**

         for Section S_k in $Active$ **do**   ▷ (a section gets a target direction dir_k and a bar Bar_k)

            Grow a section using (Alg 3, dir_k, $Bar_k = (B_k, T_k, A_k)$)

            Remove Bar_k from $Barcode$ ▷ (each Bar can be used only once)

            if $status = Bifurcate$ **then**

               Generate children using (Alg 4, $Barcode, Bar_k, dir_k$)

               Add children to $Active$ sections

            else if $status = Terminate$ **then**

               Section growth terminates

            Remove current section S_k from $Active$

  Generate accurate diameters using (Alg 5, Bio)

Output:

  A neuron: a set of points and their connectivity.

Algorithm 3 Generate soma - neurites initialization

Input:

   $Bio = \{SS, NN, PA, c\}$ ▷ (see Biological distributions)

   function SAMPLE(distr) := draws from input distribution

   $d_s = $ SAMPLE(SS)

   Soma is a sphere of diameter d_s and center c: $S_{d_s}^c$

   $\#neurites = $ SAMPLE(NN)

   Create first neurite N_1 on $S_{d_s}^c$ surface at random direction, $Vect_1$

   The first point of the neurite P_1^1 is on $S_{d_s}^c$ surface

   for Neurite $(N_i | 2 \leq i \leq \#neurites)$ **do**

      $Vect_i = Vect_{i-1} + $ SAMPLE(PA)

      First point of N_i is P_i^1 which corresponds to $Vect_i$

Output:

   A soma $S_{d_s}^c$ and the initial points of each neurite N_i

---

Algorithm 4 Elongate section

Input:

   $\tau, \rho, dir, Bar_k = (B_k, T_k, A_k), x_0$

   $\mu = 1 - \tau - \rho$ ▷ (Normalization of weights to 1)

   function RD(point) := radial distance of point from soma

   $n = 1$

   $status = Continue$

   while $status$ is $Continue$ **do**

      $random = $ random direction sampled uniformly in a unit sphere

      $memory = $ direction from the weighted sum of previous directions $x_{i \leq n}$

      $x_{n+1} = x_n + \rho * random + \tau * dir + \mu * memory$

      if Check $\Pr(Bifurcate \mid $ RD$(x_{n+1}), B_k)$ **then**

         $status = Bifurcate$

      else if Check $\Pr(Terminate \mid $ RD$(x_{n+1}), T_k)$ **then**

         $status = Terminate$

      else

         $status = Continue$

Output:

   A section and a status which is either a bifurcation or a termination.

Algorithm 5 Bifurcate

Input:

  $Barcode$, $Bar_k = (B_k, T_k, A_k)$

  function SPLIT(vect) := returns two new unit vectors dir_1, dir_2 from the input unit vector

  $dir_1, dir_2 = $ SPLIT(A_k)

  Find next available index i in $Barcode$ for which $min(B_i)$

  Generate child section 1: $\leftarrow dir_1, Bar_1 = (B_i, T_k, A_i)$

  Generate child section 2: $\leftarrow dir_2, Bar_2 = (B_{i+1}, T_{i+1}, A_{i+1})$

Output:

  Two new sections, each initialized with a direction dir and a Bar.

Algorithm 6 Diametrizer

Input:

  $Bio : TD, TR, RR, MD$ (see Biological distributions)

  function SAMPLE(distr) := draws number from distribution

  for all Neurite tips **do** $d_{tip} = $ SAMPLE(TD)

  $Active \leftarrow tips$

  while Active **do**

    for Section in Active **do**

      $taper = $ SAMPLE(TR)

      for Point in Section **do**   ▷ From termination to the root

        $d_{new} = d_{n+1} + taper * length$

        if $d_{new} \leq MD$ **then**

          $d_n = d_{new}$

        else

          $d_n = d_{n+1}$

      Remove Section from Active

      if all siblings $1, 2, \ldots$ computed **then**

        $n = $ SAMPLE(RR)

        $D_{parent} = (d_1^n + d_2^n + \ldots)^{(1/n)}$

        Add parent Section to Active

Output:

  Assigns new values to the diameters of the neuron.

Description of synthesis steps

Soma generation: initiation of neurites

The direction of the neurite protrusion is particularly important for the generation of neurites have preferred orientations (Graham et al. 2006). For some neurites the initial direction is well defined; for example cortical apical dendrites typically grow towards the pia. By contrast, the outgrowth direction of basal dendrites superficially appears random. In Figure C.2A the directions of the biological (top) and of the synthesized (bottom) basal dendrites are illustrated. Even though the directions visually appear to be random, a careful analysis of the pair angles between the biological neurites reveals that they are in fact correlated. This orientation preference is illustrated on the top row of Figure C.2B, where examples of the biological dendritic directions are shown. If the correlations between the dendritic orientations of a neuron are disregarded (Figure C.2B, median row) the generated morphologies could result in dendrites that are closer than the biologically accepted distances. On the contrary, if the pairwise angles distribution is used (see Algorithm 2), the synthesized dendrites respect the biologically acceptable distances (Figure C.2B, bottom row).
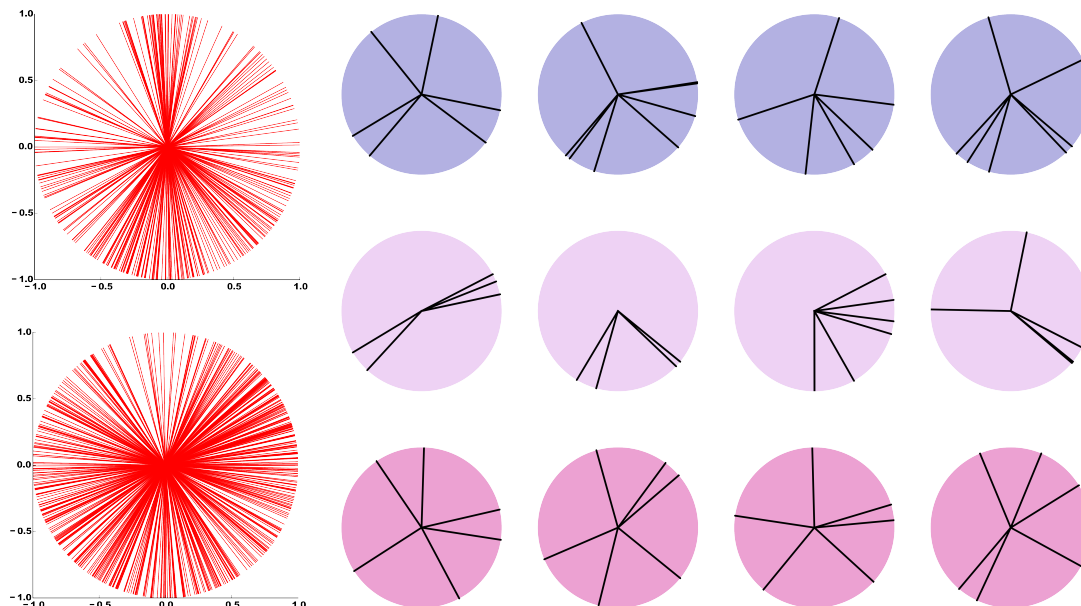

Fig. C.2 Illustration of dendritic orientations.

Section elongation

A synthesized neurite is grown segment-by-segment, where each segment has a length L and a direction $D_{segment}$ (see Algorithm 3). The direction of a new segment is defined by a unit vector as a weighted sum of three components: the cumulative memory M of the directions of the previous segments within a branch, the target vector T, which is chosen at the beginning of a branch, and the randomly sampled vector from the unit sphere R (Koene et al. 2009). The direction of the new segment is:

$$D_{segment} = \rho * R + \tau * T + \mu * M.$$

Since the segment length is constant and equal to one micron $L = 1\mu m$, the weights that define the components of the next point are also normalized to one $\tau + \rho + \mu = 1$. As a result, only two of the input parameters need to be defined. The different shapes that can be generated by combinations of the weights are presented in Figure C.3.
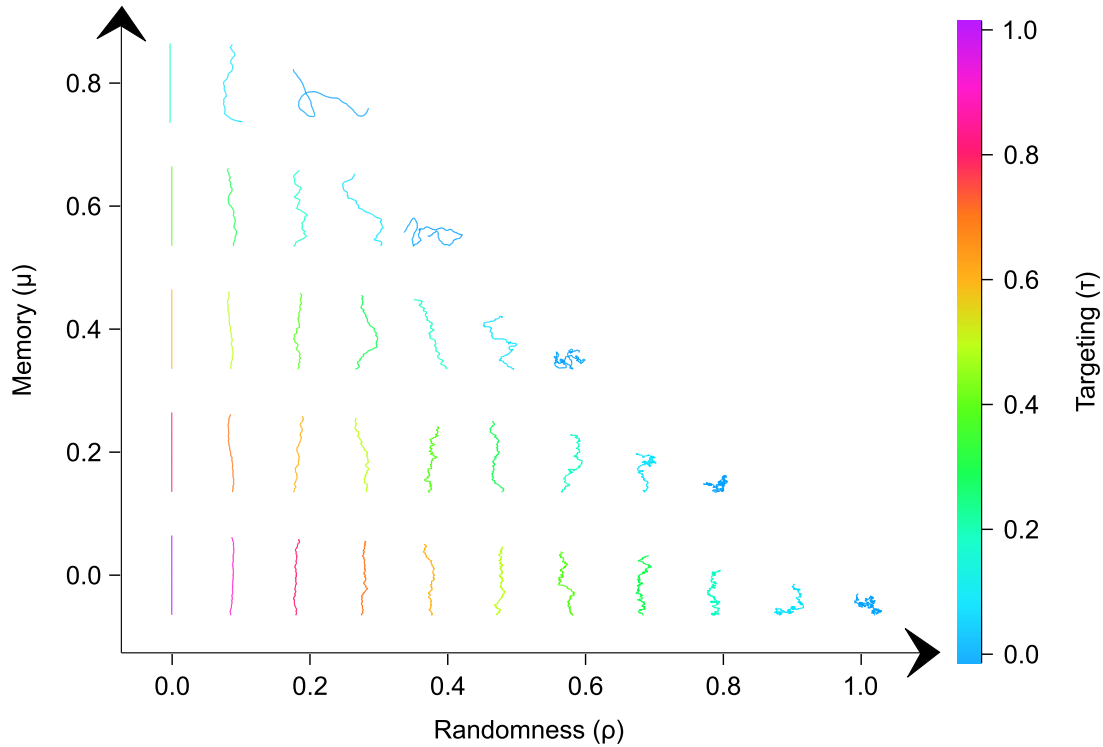

Fig. C.3 Illustration of artificially generated sections with different input parameters.

As the randomness weight ρ is increased the branch becomes more tortuous, approaching the limit of a simple random walk when $\rho = 1$. On the contrary, an increase of the targeting weight τ results in more and more straight branches towards

the target direction (purple straight line in C.3). The memory μ has a more complicated effect on the generated branch. For high values of the targeting parameter τ the line is already straight so the effect of larger memory weights is not significant. For lower values of the targeting parameter τ the memory generates shapes of larger curvature but lower local randomness as the correlations between segment directions are preserved for longer distances.

Branching - Termination

Each growing tip is assigned a barcode that includes a starting radial distance B, an ending radial distance T and a bifurcation angle A (see TMD definition). The active tip first checks the probability to bifurcate. If a bifurcation does not occur, then the active tip checks the probability to terminate. If the active tip does not bifurcate or terminate, then the branch continues to elongate. The probability to bifurcate depends on the starting radial distance B and the probability to terminate depends exponentially on the ending radial distance T.

As the active tip gets closer to the radial distance B the probability to bifurcate increases according to an exponential distribution $e^{-\lambda x}$ until it reaches the highest possible value 1.0 when the tip exceeds the target radial distance. The rate parameter λ of the exponential distribution $e^{-\lambda x}$ controls the probability to bifurcate and terminate. The parameter λ should be wisely chosen in order to generate biologically relevant neurites (see Figure C.4).

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 (top row of Figure C.4). On the other hand, a very low value of λ generates cells that are almost random, since the dependence on the input persistence barcodes will be decreased significantly. If the value of parameter λ is of the order of the step size, the bifurcation will occur within a few steps from the target radial distance. As a result, the generated shapes will differ from the input branching structure, introducing the necessary variability, but will preserve the overall shape of the input tree, generating biologically acceptable structures.

In the event of a bifurcation, two new branches are generated (see Algorithm 4) and the directions of the daughter branches depend on the input bifurcation angle A. Three branching methods were examined: the symmetric, the biased and the composite. The "symmetric method" assumes that the two daughter branches emerge at the same angles from their parent branch. The bifurcation angle is split in two and equally distributed among the daughter branches. The "biased method" assumes that one of the
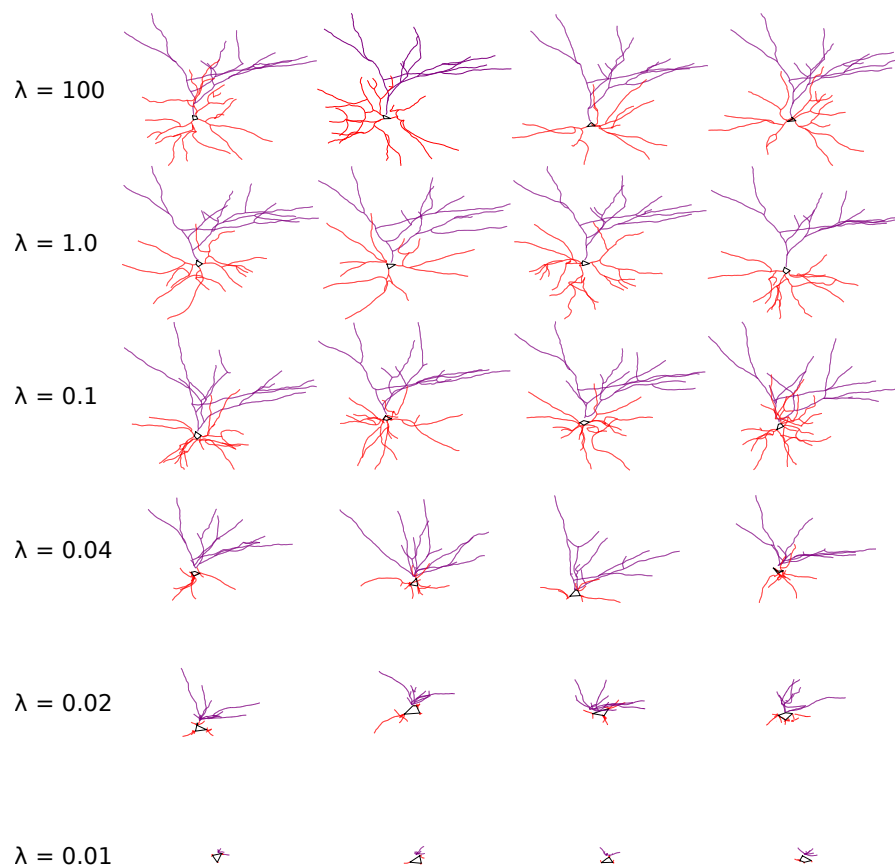
Fig. C.4 Illustration of artificially generated neuronal trees with different branching probabilities.

daughter branches continues to grow towards the direction of its parent and therefore the split is asymmetric. The "composite method" assumes that a combination of the symmetric and the biased methods is required. For this approach two biological angles are required, the bifurcation angle *A* which defines the angle between the daughter branches, and the parent-daughter angle which defines the angle between the parent and one of the daughter branches.

Validation of synthesized basal and apical dendrites

Identifying Outliers in the Synthesized Cells Population

We developed a method to identify outliers in the synthesized neurons in order to remove them and improve the statistics of the population. Furthermore, the percentage of the detected outliers in the synthesized population gives an idea of the accuracy of the synthesis process.
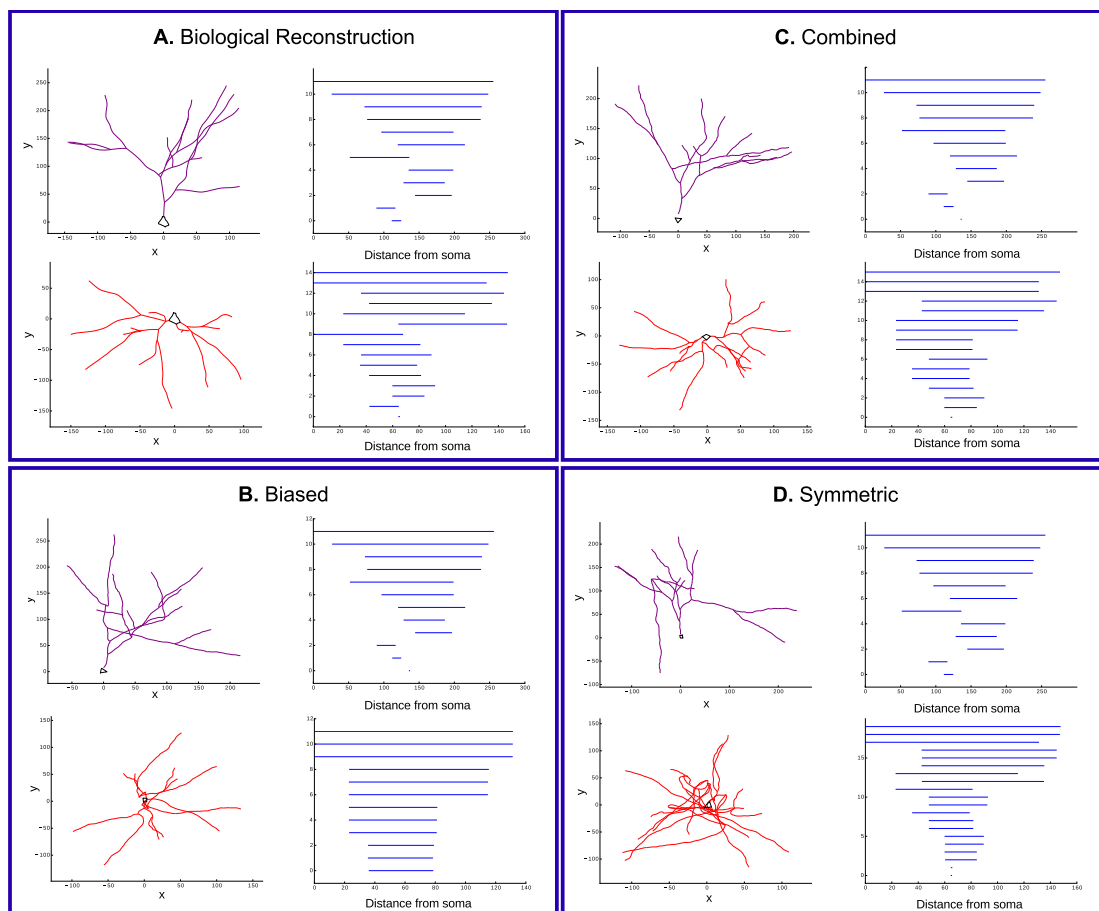
Fig. C.5 Different branching methods: symmetric, biased and combined.

For identifying outliers, we compare the distributions of key features of each neuron to the reconstructed biological cell population. We quantify the distance between the distributions with the absolute Difference Between the Medians as a proportion of the Overall Visible Spread (DBM/OVS). This is an intuitive measure of quantifying the difference between the medians of the two distributions with respect to their joined dispersion. The Overall Visible Spread is 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. 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. This is a very special case, though; in most cases the DBM/OVS measure is below 1. The closest the DBM/OVS gets to 1 the largest is the difference of medians with respect to the overall spread of the distributions making likely to reject the neuron in

question. For an illustration, see figure in the supplementary and the first reference for a thorough discussion.

The DBM/OVS measure works better than other measures of standardized difference of means like Hedges' g, for example, 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 above certain feature-specific thresholds. We choose the thresholds so that the reconstructed biological cells are not sorted out as outliers, since they represent the golden standard. One can imagine the thresholds defining a hypercube in the space of features. If a cell falls out of this cube, it is rejected as an outlier.

Population to Population Validations

We validate the synthesized neurons by comparing the distributions of their morphometrics features to those of the biological cell reconstructions. In the following figures, we show the comparisons for the Sholl analysis and further key features like the degree of the dendritic tree (number of terminations), branch orders, number of sections per neurite, total length, section intermediate and section terminating lengths and path length to the terminal tips. We report summary statistics like means, standard deviations, medians and sample sizes. The Kolmogorov-Smirnov distance quantifies the dissimilarity between the distributions. Concretely, it measures the maximum distance between the two empirical cumulative distributions. It ranges from 0 in the case of identical distributions to 1 in the case of maximal difference, for example when the distributions are completely shifted. In the case of discrete distributions like the branch orders, we use the adapted version of the Kolmogorov-Smirnov test as described in Arnold and Emerson (see reference).

Additional features comparisons between the synthesized and the reconstructed cell populations are given in the following table. These include the number of neurites, number of sections per neurite, number of bifurcations, radial (Euclidean) distance, bifurcation angles, tortuosity and partition as an expression of the asymmetry of the cells.

Results

Topological distances of synthesized dendrites

In order to ensure that the topological profiles of the original reconstructions are reproduced by the TNS algorithm, the topological distances between the biological and

the synthesized cells are computed and compared to the topological distances within the biological population. The mean topological distance among the dendrites of biological interneurons is $\approx 9,800$ while the maximum distance is $\approx 60,000$. Respectively, the mean distance between the dendrites of synthesized and biological cells is $\approx 9,000$ and the maximum distance is $\approx 57,000$ (Figure C.6). Note that if the exact TMD of the biological cells were recreated by the TNS algorithm, this distance would be zero. However, since we aim to increase the morphological variability of the digitally reconstructed network, it is important that the topological profiles of the synthesized cells do not exactly match the biological ones, otherwise the artificially created neurons will differ from their input cells only in local properties.
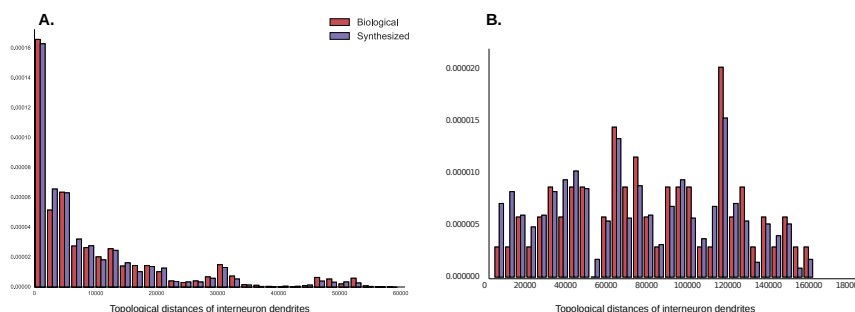

Fig. C.6 Comparison between biological and TMD synthesized densrites. The topological distances between biological dendrites (in red) are compared against the topological distances between synthesized and biological dendrites (in blue). A. Interneuron distances. B. Pyramidal cells distances.

Distances of Morphometrics

In order to ensure that the statistical properties of the original reconstructions' morphometrics are reproduced by the TNS algorithm, we also compute the statistical distance between the biological and the synthesized distributions of each morphological feature that was used for the validation of the synthesized cells (Figures 3.7-3.8). The statistical distance that we will use for this test is the KS-distance (see Population to Population Validations). A comparative value is computed for the corresponding distances of the morphometrics of cells within the biological population. The mean values of the biological distances are presented in Figure C.7, in blue, and compared to the distances between synthesized and biological cells, in red. The distance between synthesized and biological cells are typically smaller that the corresponding distances within the biological population. This indicates that the TNS algorithm faithfully reproduces the morphometrics of the input population.
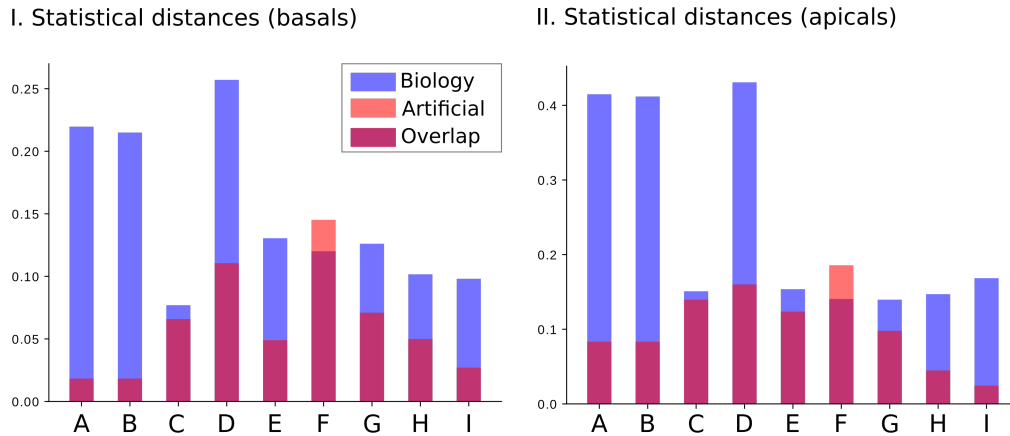
Fig. C.7 Comparison between biological and TMD synthesized dendrites morphometrics. The statistical distances between biological dendrites (I. basals, II. apicals) are computed using bootstraping (in blue) and they are compared to the statistical distances between synthesized and biological dendrites (in red). The statistical distances are defined as the KS-distances between the corresponding distributions. The overlap is shown in purple. The distances between synthesized and biological cells are consistently smaller than the distances within the biological population. The morphological features (A.-H.) correspond to the validation features presented in Figures 3.7-3.8: A. number of terminations, B. number of sections, C. section branch orders, D. total length per neurite, E. section termination lengths, F. section bifurcation lengths, G. remote bifurcation angles, H. section path distances, I. section radial distances.

Synthesis without correlation of bifurcation / termination

To study whether accurate synthesis requires the correlations encoded in the TMD of trees, or whether the independent bifurcation and termination probabilities suffice to describe and reproduce the branching patterns of the neuronal morphologies, we have generated un-correlated artificial dendrites of L5SPC (Figure C.8C). Instead of using the joint probability distribution to bifurcate and terminate, as encoded in the persistence barcodes of the biological dendrites, the marginal probabilities were used. This method results in surprisingly poor quality of synthesized cells, indicating that the information encoded in the TMD of dendrites is essential for the accurate generation of artificial cells.

The validation of the artificial dendrites synthesized without taking account correlations is summarized in Figures C.9- C.10. The basal dendrites appear to have morphometrics that closely match the biological, similarly to the TMD dendrites. However, the more complex apical structures cannot be retrieved by this simplified rule. A few of the features closely match the distributions of the input cells, such as the number of terminations and sections. However, the rest of the features fail to reproduce the morphometrics of the input cells. The synthesized cells appear to be longer, with
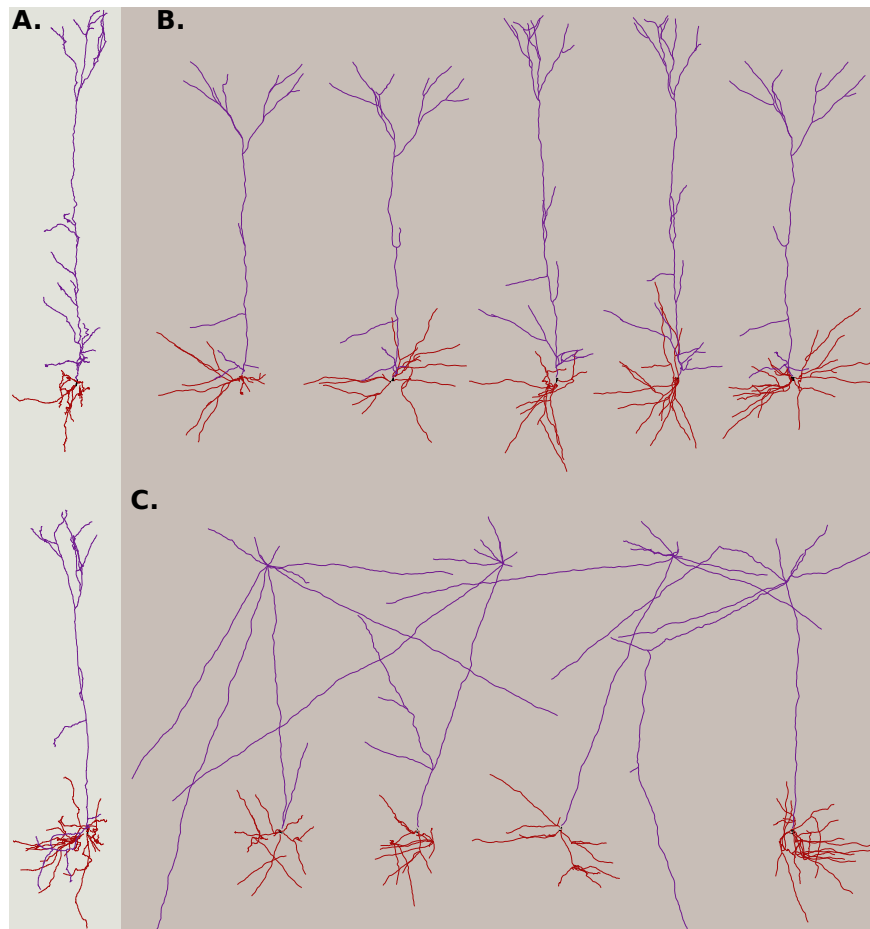
Fig. C.8 Comparison between random and TMD synthesized cells. A. Biological reconstructions. B. Artificial cells synthesized with the TMD algorithm. C. Artificial cells synthesized with the non-correlated probabilities of branching and termination. The TMD generated cells are visually close to the biological reconstructions compared to the non-correlated cells.

smaller branch orders, a behavior that indicates that this synthesis algorithm does not respect basic morphological principles of the original reconstructions, as the correlations between features are not considered.

References

Carlsson, G. (2009). Topology and data. Bulletin of American Mathematical Society, 46, 255–308.

Graham, B.P., van Ooyen, A. Mathematical modelling and numerical simulation of the morphological development of neurons. (2006). BMC Neuroscience. 7(Suppl 1):S9.
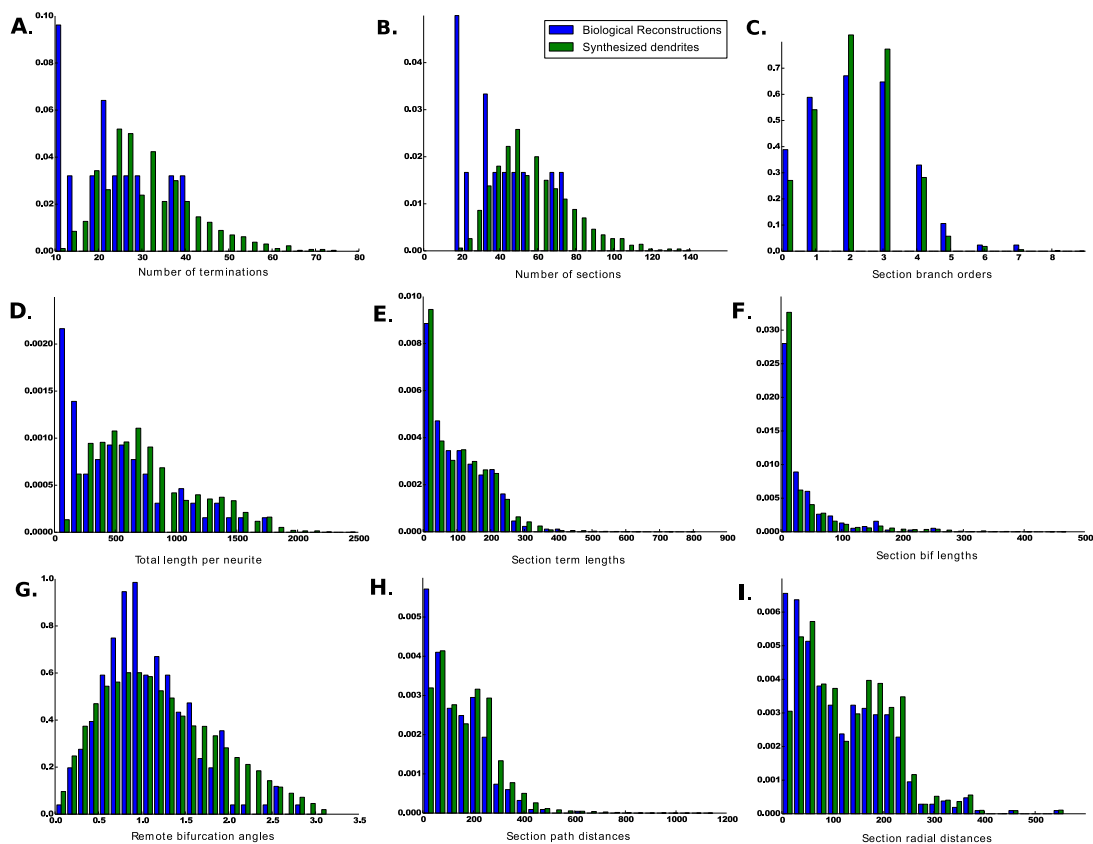
Fig. C.9 Validation of a virtual population of L5_SPC cells (basal dendrites), synthesized without correlations between bifurcation / termination. The artificially generated population of Layer 5 SPC basal dendrites (green) is compared to the biological reconstructions of the same type (blue). A subset of the features used for the validation is presented: A. number of terminations, B. number of sections, C. section branch orders, D. total length per neurite, E. section termination lengths, F. section bifurcation lengths, G. remote bifurcation angles, H. section path distances, I. section radial distances.

Kanari, L., Dłotko, P., Scolamiero, M. Levi R., Shillcock J., Hess K., Markram H. A Topological Representation of Branching Neuronal Morphologies (2017). Neuroinformatics.

Koene R.A., Tijms B., van Hees P., Postma F., de Ridder A., Ramakers G.J., van Pelt J., van Ooyen A. NETMORPH: a framework for the stochastic generation of large scale neuronal networks with realistic neuron morphologies. (2009). Neuroinformatics. 7(3):195-210.

Wild C.J., Pfannkuch M., Regan M. and Horton N.J. (2011). Towards more accessible conceptions of statistical Inference. J. R. Statist. Soc. A (2011) 174, Part 2, pp. 247–295
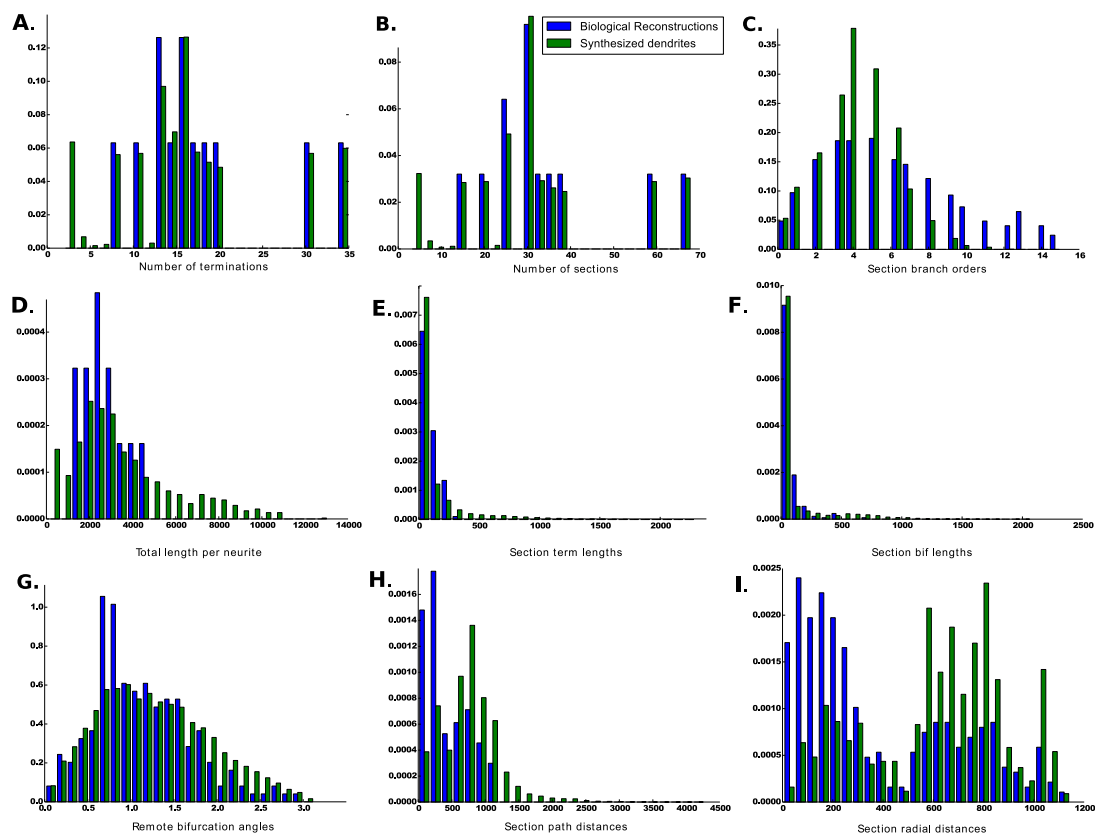
Fig. C.10 Validation of a virtual population of L5_SPC cells (apical dendrites), synthesized without correlations between bifurcation / termination. The artificially generated population of Layer 5 SPC basal dendrites (green) is compared to the biological reconstructions of the same type (blue). A subset of the features used for the validation is presented: A. number of terminations, B. number of sections, C. section branch orders, D. total length per neurite, E. section termination lengths, F. section bifurcation lengths, G. remote bifurcation angles, H. section path distances, I. section radial distances.

Nakagawa S., Cuthill I.C. (2007). Effect size, confidence intervals and statistical significance: a practical guide for biologists. Biological reviews of the Cambridge Philosophical Society. 82. 591-605.

Larry V. Hedges (1981). Distribution theory for Glass' estimator of effect size and related estimators. Journal of Educational Statistics. 6 (2): 107–128.

Appendix D

List of publications

Complete list of Journal Publications

2018   *Generative model of cortical dendritic morphologies based on the topology of their branching architecture* Kanari L., Chalimourda A., Atenekeng G., Graham J.W., Shillcock J., Hess K., Markram H. To be submitted

2018   *Objective classes of neocortical pyramidal cells.* Kanari L., Ramaswamy S., Shi Y., Wang Y., Meystre J., Perin R., Shillcock J., Hess K., Markram H. To be submitted

2017 Nov. *A Topological representation of branching neuronal morphologies.* Kanari L., Dlotko P., Scolamiero M., Levi R., Shillcock J., Hess K., Markram H. Neuroinformatics

2017 Sep. *Comprehensive morpho-electrotonic analysis shows 2 distinct classes of L2 and L3 pyramidal neurons in human temporal cortex.* Deitcher Y., Eyal G., Kanari L., Verhoog M.B., Atenekeng G.K., Mansvelder H.D., de Kock C.P.J., Segev I. Cerebral Cortex

2016 Aug. *Framework for efficient synthesis of spatially embedded morphologies.* Vanherpe L., Kanari L., Atenekeng G., Palacios J., Shillcock J. Phys. Rev. E

2015 Dec. *In situ synthesis and simulation of polydisperse amphiphilic membranes.* Vanherpe L., Kanari L., Atenekeng G., Palacios J., Shillcock J. Int J Adv Eng Sci Appl Math

2015 Oct. Reconstruction and simulation of neocortical microcircuitry. Markram H., Muller E., Ramaswamy S., Reimann M.W. et al., Cell

2015 Oct. The neocortical microcircuit collaboration portal: a resource for rat somatosensory cortex. Ramaswamy S., Courcol J.D. et al, Front. Neural Circuits

Complete list of Conferences

2017 SfN Society for Neuroscience *Topological classification of pyramidal cells in juvenile rat somatosensory cortex* Kanari L., Wang Y., Shi Y., Ramaswamy S., Perin R., Shillcock J., Hess K., Markram H.

Washington DC, US

2017 CNS Organization for Computational Neuroscience *Topological representation of neuronal morphologies.* Kanari L., Dlotko P., Scolamiero M., Levi R., Shillcock J., de Kock C.P.J. , Hess K., Markram H.

Antwerp, Belgium

2017 ELSC Edmond and Lily Safra Center for Brain Sciences retreat *Topological representation of neuronal morphologies.* Kanari L., Dlotko P., Scolamiero M., Levi R., Shillcock J., Hess K., Markram H.

Ein Gedi, Israel

2017 AAT Applied Algebraic Topology *Topological representation of neuronal morphologies.* Kanari L., Dlotko P., Scolamiero M., Levi R., Shillcock J., Hess K., Markram H.

Sapporo, Jappan

2016 SfN Society for Neuroscience *Quantitative topological analysis of neuronal trees.* Kanari L., Dlotko P., Scolamiero M., Levi R., Shillcock J., Hess K., Markram H.

San Diego, US

2016 HBP Human Brain Project Summit *Quantitative analysis of neuronal morphologies.* Kanari L., Zisis E., Vanherpe L., Shillcock J., Palacios J., Markram H.

Florence, Italy

2016 BNH BigNeuron Hackathon *Analysis, validation and synthesis of neuronal morphologies.* Kanari L.

London, UK

2016 ASMP Applications and Statistics of Multidimensional Persistence *Topological representation of branching neuronal morphologies.* Kanari L., Dlotko P., Scolamiero M., Levi R., Shillcock J., Hess K., Markram H.

Lausanne, Switzerland

2016 FENS Federation of European Neuroscience Societies *Synthesis of spatially-embedded neuronal morphologies* Vanherpe L., Romani A., Kanari L., Atenekeng G., Palacios J., Chalimourda A., Shillcock J., Markram H.

Copenhagen, Denmark

2015 BIH International Conference on Brain Informatics and Health *Automated reconstruction, analysis, classification and synthesis of neuronal morphologies.* Kanari L., Atenekeng G., Vanherpe L., Shillcock J., Markram H.

London, UK

2015 BNH BigNeuron Hackathon *Automatic analysis of neuronal morphologies.* Kanari L.

Cambridge, UK

2015 HBP Human Brain Project Summit *Automated reconstruction, analysis, classification and synthesis of neuronal morphologies.* Kanari L., Atenekeng G., Vanherpe L., Shillcock J., Markram H. Madrid, Spain

2014 HBP Human Brain Project Summit *Scalable brain synthesis* Kanari L., Atenekeng G., Vanherpe L., Shillcock J., Markram H.

Heidelberg, Germany

2014 SfN Society for Neuroscience *Morphological analysis and classification of neuronal types in the neocortical microcircuit* Kanari L., Atenekeng G., Graham J.W., Hateland R., Shi Y., Shillcock J., Schuermann F., Markram H.

Washington DC, US

2013 EPMN European PhD Meeting of Neuroscience *PNEUMATK: A Python tool kit for the analysis, comparison, and classification of neuronal morphologies* Kanari L., Graham J.W., Atenekeng G., Shillcock J., Schuermann F., Markram H.

Grenoble, France

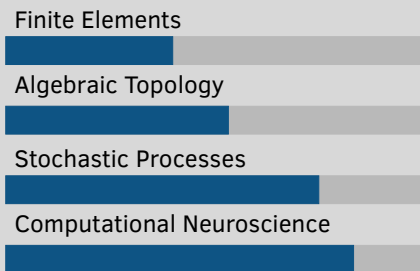# Appendix E

Curriculum Vitae (CV)

Lida Kanari

- 🛈 17 October 1988
- 🌐 From Athens, Greece
- ✉ Lives in Lausanne, Switzerland
- @ lida.kanari@gmail.com

About me ——————

Lida is a trained mathematician and physicist that found herself in the world of neuroscience, during her PhD. She managed to survive by incorporating as much mathematics as possible into this world, giving rise to interesting concepts about neuronal trees that proved useful for neuroscientists. Lida's interests include mathematical modeling of randomness and structure in cerebral and vascular systems.

Skills ——————

Finite Elements

Algebraic Topology

Stochastic Processes

Computational Neuroscience

Programming ——————

C++, C

Fortran

Python

(*)[From Basic to Major experience.]

Languages ——————
Greek, English, French

Education

2012-2017	Ph.D. candidate in Computational Neuroscience *Neuronal morphologies: the shapes of thoughts.*	EPFL, Lausanne
2011-2012	M.Sc. Computational Mechanics Majoring in Fluid mechanics and Biomechanics	NTUA, Greece
2006-2011	M.Sc. Applied Mathematics and Physics Majoring in Theoretical and Particle Physics	NTUA, Greece

Awards

2017	Organization for Computational Neuroscience Best poster award
2011-2012	Limmat Stiftung Scholarship. Highest GPA in MsC, Computational Mechanics
	State Foundation of Fellowships, Greece. State Fellowship for MsC, Computational Mechanics
2006-2011	State Foundation of Fellowships, Greece. Highest GPA in MsC, Applied Mathematics and Physics
2009-2010	16th International Mathematical Competition, Budapest. Honorable Mention
2003-2005	Hellenic Mathematical Society. Distinctions & Awards

Teaching

2016-2017	Symmetry and Conservation in the Cell by Dr. J. Shillcock	EPFL, Lausanne
2015-2016	Simulation Neuroscience by Prof. H. Markram, Prof. I. Segev	MOOK, online
2013-2014	Analysis II & Numerical Analysis by Prof. F. Margot by Prof. S. Deparis	EPFL, Lausanne
2012-2013	Analysis II & Advanced Analysis II by Prof. R. Hans-jörg	EPFL, Lausanne

Publications

2017	A Topological Representation of Branching Neuronal Morphologies. Kanari, Dlotko, Scolamiero et al.	*Neuroinformatics*
2017	Comprehensive Morpho-Electrotonic Analysis Shows 2 Distinct Classes of L2 and L3 Pyramidal Neurons in Human Temporal Cortex. Deitcher, Eyal, Kanari et al.	*Cerebral Cortex*
2016	Framework for efficient synthesis of spatially embedded morphologies. Vanherpe, Kanari, Atenekeng et al.	*Phys. Rev. E*
2015	Reconstruction and Simulation of Neocortical Microcircuitry. Markram, Muller, Ramaswamy et al.	*Cell*

Conferences

2017	SfN, CNS, ELSC Retreat, Applied Algebraic Topology
2016	SfN, BigNeuron, HBP Summit, Applications and Statistics of Multidimensional Persistence
2015	BIH, BigNeuron, HBP Summit
2014	SfN, HBP Summit

Other interests

Apart from scientific research, Lida is interested in theater, music, movies, photographs and books. When she does not find herself closed within four walls - which is quite a rare occurence- she also enjoys hiking, biking and kayaking.