Advancing multiscale structural mapping of the brain through fluorescence imaging and analysis across length scales

L. J. Hogstrom, S. M. Guo, K. Murugadoss and M. Bathe

Department of Biological Engineering, Massachusetts Institute of Technology, Building 16, Room 255, Cambridge, MA 02139, USA

Brain function emerges from hierarchical neuronal structure that spans orders of magnitude in length scale, from the nanometre-scale organization of synaptic proteins to the macroscopic wiring of neuronal circuits. Because the synaptic electrochemical signal transmission that drives brain function ultimately relies on the organization of neuronal circuits, understanding brain function requires an understanding of the principles that determine hierarchical neuronal structure in living or intact organisms. Recent advances in fluorescence imaging now enable quantitative characterization of neuronal structure across length scales, ranging from single-molecule localization using super-resolution imaging to whole-brain imaging using light-sheet microscopy on cleared samples. These tools, together with correlative electron microscopy and magnetic resonance imaging at the nanoscopic and macroscopic scales, respectively, now facilitate our ability to probe brain structure across its full range of length scales with cellular and molecular specificity. As these imaging datasets become increasingly accessible to researchers, novel statistical and computational frameworks will play an increasing role in efforts to relate hierarchical brain structure to its function. In this perspective, we discuss several prominent experimental advances that are ushering in a new era of quantitative fluorescence-based imaging in neuroscience along with novel computational and statistical strategies that are helping to distil our understanding of complex brain structure.

1. Introduction

Understanding the closely inter-related structure and function of the human brain presents one of the greatest scientific challenges of this century. Since the seminal discoveries of basic aspects of neuronal morphology and connectivity by Ramon y Cajal over 100 years ago [1,2], our understanding of brain function has remained limited by our inability to efficiently measure brain structure at multiple length scales that range from single synapses to the whole intact organ itself. However, recent advances in fluorescence imaging such as multiplexed super-resolution imaging [3–5] and rapid whole-brain imaging in model organisms using light-sheet microscopy [6] are, together with a rapidly emerging set of genetic engineering tools, increasing our ability to characterize brain structure in a holistic, multiscale manner. As these multi-resolution datasets emerge, new data analysis strategies should, in our view, ideally aim to offer integrative model-based descriptions that facilitate functional interpretation of these datasets across multiple imaging modalities and length scales.

Electron microscopy (EM) is the gold standard for resolving brain structure in utmost detail at the nanometre scale, offering the highest spatial resolution and great promise to generate complete neuronal wiring diagrams, or ‘connectomes’, of the brain. Such efforts are, however, intrinsically limited to fixed, non-living samples and generally lack the ability to resolve specific molecular identities in a multiplexed manner. Moreover, the generation and interpretation of EM datasets present major challenges that typically limit application of this ultra-high-resolution structural procedure to small sample sizes and only a handful of laboratories. In contrast, fluorescence imaging or light microscopy (LM) is a
widely accessible tool that now enables the structural characterization of synaptic molecules up to entire mouse brains in a multiplexed manner using sample clearing techniques [7,8]. While a major strength of fluorescence imaging is its applicability to live organisms, we limit our review here to fixed samples with the aim of quantitatively characterizing hierarchical brain structure across length scales.

Because of the very large size of the datasets produced by both EM and LM imaging modalities, computational tools play an instrumental role in analysing, annotating, organizing and interpreting these datasets. Multiscale approaches such as correlative light and electron microscopy (CLEM) that integrate EM and LM datasets [9–11] are of utmost interest to benefit from these datasets in a complementary manner. This is particularly the case as we attempt to understand not only the connectomic wiring of the brain, but also the synaptic composition of intact brain samples with highly heterogeneous and variable distributions of molecules that ultimately govern synaptic transmission. As novel labelling and imaging techniques rapidly emerge to produce very large datasets consisting of terabytes of LM and EM data, new statistical frameworks and computational tools to characterize the multiscale nature of brain structure are expected to play a leading role in unifying these datasets to extract structural principles. While formal models of multiscale brain structure are currently lacking, we believe that the formulation of model-driven structural analysis techniques is of utmost importance in order to interpret these emerging hierarchical datasets with a view towards their unification and ultimately functional interpretation. These efforts may one day inform in vivo functional studies including electrophysiology [12,13] and whole animal behaviour [14,15], which are not treated in this perspective.

Here, we highlight recent fluorescence labelling, imaging and analysis approaches that inform our understanding of molecular and cellular organization at these diverse length scales. We posit that increasing our quantitative knowledge of hierarchical brain structure will aid in phenotypic characterization and will likely generate new avenues for understanding a range of human diseases related to molecularly and genetically identifiable targets. Because of inherent limitations of fluorescence imaging, including major challenges that are associated with the high-throughput measurement of molecular numbers and localizations, the quantification of neuronal wiring and connectivity, and whole-brain imaging of living organisms, we additionally discuss how serial EM and whole-brain magnetic resonance imaging (MRI) may be used in conjunction with LM to effectively bridge scales (figure 1). For each approach, we discuss quantitative analysis and modelling strategies that are beginning to enable the integration of neuroanatomy at multiple scales. Finally, we review several examples of neuronal spheroids and cerebral organoids that are offering new opportunities to understand how genetic variation leads to human disease in patient-specific models of brain structure and function.

2. Super-resolution imaging enables multiplexed synaptic protein localization and counting

The link between mutations of synaptic scaffolding and adhesion molecules such as Shank3, neuroligin and neurexin with autism spectrum disorders (ASD) and schizophrenia is well documented and includes associations through rare mutations, de novo copy number variations and chromosomal abnormalities [22–24]. The clinical significance of these genetic aberrations underscores the need to improve our understanding of the relation between genetic variation and synaptic molecular composition. For example, it is plausible that neuronal dysfunction stems from disruptions to pre- and post-synaptic cell adhesion complexes that stabilize synapses during development and are essential to normal synaptic transmission [25]. Understanding how such genetic variation impacts synaptic molecular organization and signal transmission is therefore of central importance in resolving brain structure. Super-resolution fluorescence imaging offers the ability to probe synaptic structure by imaging individual intact synapses with typically 20 nm spatial resolution, in principle across numerous molecular targets in a multiplexed manner [26,27].

Among several distinct super-resolution imaging approaches, localization microscopy enables the highest imaging resolution (approx. 20 nm that is well below the approx. 200 to 500 nm size of an individual synapse) by localizing single fluorophores sequentially. The single-molecule nature of localization microscopy further enables the conversion of observed units of fluorescence into quantitative counts of specific molecules using calibration standards, thereby enabling quantitative in situ proteomics within synapses. For example, recent efforts have used the localization microscopy techniques photo-activated localized microscopy (PALM) and stochastic optical reconstruction microscopy (STORM) to measure the ultrastructure of excitatory and inhibitory synapses [3,4,26,28–30]. Specht et al. [4] used a mouse strain expressing photoconvertible constructs of the inhibitory synaptic scaffolding protein gephyrin together with a photoswitchable dye-labelled antibody to reveal sub-synaptic co-localization of gephyrin molecules with glycine receptors within individual synapses of cultured spinal cord neurons. This result is further supported by the 1:1 stoichiometric ratio of gephyrin to receptor-binding sites measured using a photobleaching-based single-molecule counting approach [4]. Similarly, PALM was used to show that PSD95 forms nanodomains within the post-synaptic density that concentrates AMPA receptors but not NMDA receptors [30]. The recent advent of three-dimensional super-resolution microscopy further enables mapping out the complete three-dimensional organization of synaptic proteins in single synapses. Relative spatial distributions of 10 different synaptic proteins has been determined by performing three-colour three-dimensional STORM using sequential imaging of distinct synaptic proteins using a common molecular reference [3].

While localization microscopy provides high spatial resolution, its low temporal resolution typically limits its application to fixed samples. In contrast, coordinate-targeted super-resolution methods such as stimulated emission depletion (STED) microscopy and reversible saturable optical fluorescence transitions (RESOLFT) have helped to capture the fast dynamics of neuronal structure [5,31–33]. Coordinate-targeted super-resolution microscopy relies on minimizing the illumination volume by selectively ‘turning off’ the fluorescence of a subset of molecules so that the positions of molecules can be determined more accurately. However, high laser power that is required for depleting fluorescence in STED can pose significant phototoxicity to neurons. RESOLFT minimizes the illumination volume by employing reversibly switchable fluorescent proteins whose dark state can be induced with lower laser power, improving spatial resolution by more than threefold in comparison with conventional confocal microscopy. RESOLFT has
been applied to study the morphological dynamics of dendritic spines in living hippocampal brain slices [5]. Three-dimensional super-resolution imaging using selective plane illumination microscopy (SPIM) allows for fast imaging of thick samples with reduced phototoxicity by using a sheet of light to selectively illuminate a thin section of the sample, providing new avenues for performing live-cell in situ super-resolution imaging [34].

The variation in sub-synaptic structures and compositions revealed by super-resolution imaging will provide new data for further characterization of synapse sub-types in addition to excitatory and inhibitory synaptic classes, which will offer crucial structural components to drive modelling of neuronal circuit function. A major future challenge is the application of whole-brain super-resolution imaging to intact samples, given
imunofluorescence for VGluT1 (light blue), glutamine synthetase (orange), synapsin-1 (green) and PSD-95 (red).

Embedding and fixation protocols that retain fluorescent protein function have been used for CLEM imaging of genetic sources allows for high-resolution tracing of neuronal processes and ultrastructural characterization with EM provide the most definitive datasets for full neuronal circuit reconstruction. The use of electron beams as illumination sources allows for high-resolution tracing of neuronal processes with unambiguous identification of synaptic contacts. Efforts to make large-scale EM maps of mammalian brain tissue have been undertaken [19,35,36], although these approaches typically do not resolve constituent proteins within cells. Thus, characterization of synapse types with pure EM approaches is challenging because this requires the identification of a number of distinct synaptic targets [37]. Recently, different types of CLEM have been developed to combine the strengths of LM and EM, namely specificity and resolution. Embedding and fixation protocols that retain fluorescent protein function have been used for CLEM imaging of genetically labelled samples [38–40]. Oxygen-radical-generating fluorescent proteins have also been used to enhance EM contrast by locally catalysing formation of polymers that are resolvable by EM [41]. Shu et al. [42] have examined the localization of synaptic cell adhesion molecules SynCAM1 and -2 using both EM and fluorescence microscopy by tagging the target proteins with a fluorescent protein that is engineered to generate singlet oxygen with high yield upon illumination.

In contrast, other correlative methods attempt to combine EM and immunofluorescence, in which imaging is performed separately on the same or adjacent tissues samples. A major challenge of performing EM and immunofluorescence on the same sample arises from their typically mutually exclusive sample preparation procedures—fixation and embedding that is required for EM may disrupt immunoreactivity of the sample. Recent advances have improved tissue preparation procedures to preserve both the immunoreactivity and ultrastructure of thin tissue segments for pairing ultrastructure measurements with the labelling of multiple protein targets [9,43,44]. For example, improved freeze-substitution embedding methods can circumvent the use of osmium tetroxide, a common EM post-fixation step, to improve the preservation of protein epitopes in tissue during antibody labelling. The osmium-tetroxide-free embedding approach has been combined with array tomography (AT) [10,45], a multiplexed immunofluorescence microscopy technique that sequentially stains and strips the same ultrathin tissue sections with different antibodies, enabling high-resolution volumetric imaging of synaptic structures and molecular composition at the same time (figure 2) [11]. New image analysis strategies for EM are also required to improve reconstruction of long-range neuronal

3. Electron microscopy remains the gold standard for connectomics

Despite recent advances in super-resolution imaging, to date, serial tissue sectioning and ultrastructural characterization with EM provide the most definitive datasets for full neuronal circuit reconstruction. The use of electron beams as illumination sources allows for high-resolution tracing of neuronal processes with unambiguous identification of synaptic connections. Efforts to make large-scale EM maps of mammalian brain tissue have been undertaken [19,35,36], although these approaches typically do not resolve constituent proteins within cells. Thus, characterization of synapse types with pure EM approaches is challenging because this requires the identification of a number of distinct synaptic targets [37]. Recently, different types of CLEM have been developed to combine the strengths of LM and EM, namely specificity and resolution. Embedding and fixation protocols that retain fluorescent protein function have been used for CLEM imaging of genetically labelled samples [38–40]. Oxygen-radical-generating fluorescent proteins have also been used to enhance EM contrast by locally catalysing formation of polymers that are resolvable by EM [41]. Shu et al. [42] have examined the localization of synaptic cell adhesion molecules SynCAM1 and -2 using both EM and fluorescence microscopy by tagging the target proteins with a fluorescent protein that is engineered to generate singlet oxygen with high yield upon illumination.

In contrast, other correlative methods attempt to combine EM and immunofluorescence, in which imaging is performed separately on the same or adjacent tissues samples. A major challenge of performing EM and immunofluorescence on the same sample arises from their typically mutually exclusive sample preparation procedures—fixation and embedding that is required for EM may disrupt immunoreactivity of the sample. Recent advances have improved tissue preparation procedures to preserve both the immunoreactivity and ultrastructure of thin tissue segments for pairing ultrastructure measurements with the labelling of multiple protein targets [9,43,44]. For example, improved freeze-substitution embedding methods can circumvent the use of osmium tetroxide, a common EM post-fixation step, to improve the preservation of protein epitopes in tissue during antibody labelling. The osmium-tetroxide-free embedding approach has been combined with array tomography (AT) [10,45], a multiplexed immunofluorescence microscopy technique that sequentially stains and strips the same ultrathin tissue sections with different antibodies, enabling high-resolution volumetric imaging of synaptic structures and molecular composition at the same time (figure 2) [11]. New image analysis strategies for EM are also required to improve reconstruction of long-range neuronal
circuits. While new brain-wide reduced-osmium staining enables the long-range tracing of thin neurite features, including spine necks and small-calibre axons, EM tracing still requires highly involved human-assisted reconstruction [46].

4. Multiplexed imaging to resolve distinct cellular and molecular sub-types in intact cells and tissues

In addition to increased resolution of synaptic structures, labelling strategies that enable quantification of local protein expression levels and co-localizations provide new data that are likely to prove central to informing functional models based on underlying brain structure that account for both cellular and synaptic identities. While most current imaging modalities are limited to three or four fluorescent channels due to spectral overlap of conventional fluorophores, new multiplexing strategies are leading to compelling advances in our ability to image numerous proteins and cell types simultaneously within the same intact tissue. One such example is the use of diffusible fluorescent probes for localization microscopy, termed points accumulation in nanoscale topography (PAINT), a concept originally introduced by Sharanan and Hochstrasser in 2006 [47]. PAINT employs transiently binding fluorescent probes that interact specifically with molecular targets in order to generate transient target blinking while allowing for probe exchange using wash-out or probe dilution in physiological buffer. Transient binding kinetics enables sequential imaging of arbitrary numbers of molecular targets using a single dye type and laser source, allowing for implementation of this technique on standard fluorescence microscopes. These multiplexed imaging procedures may in principle remove the limit on the number of molecular as well as cellular sub-types that can be resolved within an intact tissue.

A number of variants of PAINT have been developed since the seminal work of Sharanan and Hochstrasser. Universal PAINT (uPAINT) is one alternative super-resolution technique that employs binding and bleaching of high-affinity antibodies to achieve single-molecule imaging and has been used to map out the dynamics and organization of AMPA receptors on the postsynaptic membrane [26]. Other variants of PAINT have employed transiently binding nucleic acid probes that can be exchanged after imaging to enable multiplexed super-resolution microscopy [48], breaking the limit of four targets that can be imaged using conventional fluorescence microscopy due to spectral overlap. Ten-target imaging of DNA nanostructures and four-target cellular imaging were demonstrated using diffusible DNA-based probes and a commercial microscope [27]. However, labelling density is a potential issue for application of DNA-PAINT to highly multiplexed imaging of targets with high density due to the need to use bulky antibodies to label targets in the current implementation. As an alternative, one recent study used protein-fragment-based probes instead of DNA-conjugated antibodies to perform PAINT [49], reporting multiplexed super-resolution imaging of the cytoskeleton and focal adhesion proteins with a considerably higher labelling density, which can similarly be envisioned to be important for imaging the crowded post-synaptic density. However, application of this technique to neurons will require the identification of appropriate protein-binding probes.

Integrating multiplexed in situ measurement of protein localization together with multiplexed measurement of messenger RNAs offers the potential to produce rich phenotypic readouts for characterizing intact brain structure molecularly including both protein and transcript localizations and expression levels [50]. Extending multiplexed imaging to live-cell samples to characterize transcript and protein transport crucial to synapse formation and maturation offers an important yet unmet challenge. Application of model-based fluorescence correlation spectroscopy may offer the ability to quantitatively count protein and transcript copy numbers in situ from fluctuation analysis of transiently binding probes [51,52], and model-based protein and RNA trajectory analysis in live cells would offer insight into local mechanisms of their regulation [53,54]. Combining synaptic protein localization studies together with in situ characterization of amyloid aggregation dynamics is also of interest [52,55]. Resolving RNA and protein localizations is particularly important for complex neuronal structures in which dendritic and axonal processes may span hundreds to thousands of cell bodies away from the soma. Here, local rates of translation and RNA/protein modifications are determined by local signalling mechanisms rather than necessarily the transcriptional state of the cell [53,56,57]. In situ characterization of cell and synapse sub-types should again prove highly useful to informing structure-based models of neuronal circuit function.

5. Quantifying molecular co-localization

Multiplexed super-resolution and conventional widefield imaging generates spatially heterogeneous datasets that report on the levels and localizations of individual proteins and RNAs distributed throughout brain tissues. These datasets offer brain region and cell-type specificity that cannot easily be achieved using bulk biochemical protein–protein association assays such as co-immunoprecipitation, or exogenous association assays such as yeast two-hybrid. Identifying the degree of co-localization of synaptic proteins in a spatially resolved manner that is specific to synapses, cells and brain regions will offer important data to inform the type and function of neuronal synapses and circuits. For example, Dani et al. [3] illustrated super-resolution strategies for calculating co-localization of numerous synaptic proteins axially along pre- and post-synaptic densities, with relevance to understanding neuronal structure that emerges at larger length scales from axonal pathfinding [58].

Early efforts in quantifying protein co-localization relied on global image analysis and intensity-based metrics, such as Pearson correlation, to detect whether two or more proteins occupy a fixed volume. An important alternative approach was introduced by Manders and colleagues in 1996, who formulated co-localization coefficients to quantify the intersection of fluorescence signals from two imaging channels of interest. Given two channels of an image that represent the fluorescence signals from proteins labelled A and B, respectively, the Manders’ overlap coefficients quantify the relative intensity of each channel at their intersection [59].

\[
M_1 = \frac{\sum A_{\text{coloc}}}{\sum A_i} \quad \text{with} \quad A_{\text{coloc}} = A_i \quad \text{if } B_i > 0
\]

and

\[
M_2 = \frac{\sum B_{\text{coloc}}}{\sum B_i} \quad \text{with} \quad B_{\text{coloc}} = B_i \quad \text{if } A_i > 0.
\]
where $A_{\text{coloc}}$ is the pixel intensity from channel $A$ if the value for that pixel in channel $B$ is positive. Similarly, $B_{\text{coloc}}$ is the pixel intensity of channel $B$ if the value for that pixel in channel $A$ is positive. The coefficients $M_1$ and $M_2$ range between 0 and 1 and are proportional to the fluorescence signal in channels $A$ and $B$, respectively, but are restricted only to regions of co-localization. One recent example makes use of the Manders’ overlap coefficients by quantifying the extent to which AMPAR subunits GLUA1 and GLUA2 localize to post-synaptic densities as marked by PSD95 [16]. The authors measured co-localization after glycine stimulation to detect the presence of GLUA2 subunits that render synaptic ion channels impermeable to calcium. By comparing co-localization results to those in cells that did not receive elevated glycine stimulation, the study illustrates how Manders’ coefficients can easily be normalized to control conditions (figure 1a).

Manders’ coefficients and other global image metrics have been improved by new techniques that place better objective bounds on intensity relatedness and statistically evaluate the presence of co-localization events [60,61]. Object-based measures of co-localization and nonlinear correlation metrics such as mutual information that have found widespread application in other fields may also prove valuable for extracting more subtle spatial relationships that are not captured by the linear Pearson and Manders’ metrics. For example, Lagache et al. [62] found that object-based Ripley’s K-function for co-localization showed improved robustness to noise relative to Manders’ and Pearson metrics when tested on dual-channel synthetic images with Gaussian noise. As new experimental techniques enable the recording of numerous molecularly specific channels, the adoption of multiple comparison procedures will be critical to examining many pairwise relationships in these multiplexed images with Gaussian noise. As new experimental techniques showed improved robustness to noise relative to Manders’ and Pearson metrics when tested on dual-channel synthetic images with Gaussian noise. As new experimental techniques enable the recording of numerous molecularly specific channels, the adoption of multiple comparison procedures will be critical to examining many pairwise relationships in these multiplexed images with Gaussian noise.

6. Quantifying dendritic morphology and branching

Independent from molecular-level characterization of synapses, information on cellular geometry and location in cortical layering provides crucial information on neuron type and functional properties. Fluorescently labelled anterograde and retrograde tracers have historically played a central role in labelling neuronal branching morphology in mammalian systems. One example of a recent advance is the conjugation of bright and photostable fluorescent dyes to the conventional cholera toxin subunit B tracer, which has commonly been used in bright-field microscopy to mark multiple long-range neuronal projections [63]. The fluorescent equivalent of this traditional tracer now enables multiplexed imaging of neuronal morphology and integration with alternative fluorescent labels for correlative structural studies with pre- and post-synaptic markers. Similarly, transgenic labelling strategies carry the advantage of marking adjacent neurons with distinct spectral identities (figure 3a) [64]. This ‘Brainbow’ approach has recently been modified to increase colour variety and photostability of probes, as well as used adapted mouse lines where the generation of double transgenics are not required [66].

Once cells are fluorescently labelled and imaged, challenges remain in the quantitative analysis and model-based interpretation of the long-range shape and structure of neurons. Sholl analysis has widely been used as a method for quantifying dendritic size and complexity in studies of neuronal morphology [67], and while not directly model-based, it can in principle be used to drive network-based functional models. In this approach, concentric circles are drawn outward radially at fixed intervals from the soma. Intersection events are defined as the points where a dendrite crosses over the reference circle at a specific distance $r$ from the soma such that the following condition is satisfied [68],

$$
\left[ \sqrt{X_n^2 + Y_n^2} < r \right] \land \left[ \sqrt{X_{n+1}^2 + Y_{n+1}^2} \geq r \right] \lor 
\left[ \sqrt{X_n^2 + Y_n^2} \geq r \right] \land \left[ \sqrt{X_{n+1}^2 + Y_{n+1}^2} < r \right],
$$

where $X_n$ and $Y_n$ are the two-dimensional Cartesian coordinates of a neuritic segment and $X_{n+1}$ and $Y_{n+1}$ are the coordinates of a daughter segment. The principle behind the Sholl equation is that neuritic branching events can be tallied in the concentric circles extending from the soma. In recent work with knockout mice, Sholl profiles of dendritic branching and complexity have been used to measure the morphological impact of synaptic cell adhesion proteins (figure 1b) [18].

Recently, exciting efforts have extended morphological analysis to three dimensions and allowed for disease modelling in mammalian tissue. For example, one recent study used juxtasomal biocytin injection and dimensionality reduction to describe the trans-columnar neuronal morphology in the somatosensory cortex (vS1) of rats [69]. After neuronal skeleton reconstruction, principal component analysis was performed to identify clusters of morphological patterns and assign cell types. Importantly, the authors were able to analyse the distribution of bouton density along different axonal projections for each cell type. This analysis offers new insight into how distinct cell types, characterized by overall morphology, impact columnar versus horizontal connections in the vS1. Using a similar approach, Oberlaender et al. [70] reconstructed the three-dimensional circuit of excitatory neurons in a single cortical column of the rat vibrissal cortex. Topological properties of the neurons including the distribution of somata, dendrites and thalamocortical synapses were correlated with electrophysiological recordings. These results provided an improved map of the vibrissal cortex, as well as a new understanding of how spontaneous and evoked neural spike recordings relate to cell morphology. Future progress in this area may continue to relate features of cellular shape and position to functional recordings of neuronal activity in model organisms. A major challenge for multiscale and multi-modal imaging is to integrate synaptic molecular composition with dendritic branching to relate their structure to neuronal function. The design of new labelling strategies that are amenable to high-throughput manipulations, such as enhanced adeno-associated viral vectors, will also enhance efforts to describe long-range neuronal paths that complement descriptions from MRI diffusion data.

7. Model-based whole-brain circuit reconstruction

Neuronal circuit reconstruction begins with the identification of synapses or inter-neuronal contacts and follows...
the trace of individual neurites back to their soma. Recent serial block-face EM studies, for example, track the frequency of synaptic and incidental cell–cell overlaps [19,35]. Binning cell inter-connections according to surface areas helps to identify sites of contact that are most likely to be involved in synaptic transmission. This node-path detection of brain structure is ideally suited to graph-based modelling that attempts to distil governing structural principles from complex networks. Graph-theoretic approaches can be used to quantify differences in neuroanatomical connectivity patterns, as well as attempt to infer their functional consequences. For example, in 2014, Oh et al. [20] at the Allen Brain Institute released data for a whole-brain connectivity matrix based on stereotaxic injection and neuronal labelling of hundreds of mouse brain regions (figure 1d). In this and related studies, summary network measures have been used to analyse long-range regional connectivity data, as well as guide the generation of new connectomic datasets.

Standard results from graph theory illustrate how structural networks can be evaluated by counting the total number of connections observed in a given system. For example, in the directed graph \( G = (V,E) \), where \( V \) and \( E \) are the set of all nodes (or vertices) and edges, respectively, the degree sum shows the relatedness between the number of incoming connections \( \deg^+(v) \) and outgoing connections \( \deg^-(v) \) at a given node. Across a complete graph, the sum of incoming node degrees should equal the sum of outgoing node degrees and match the total number of edges [72],

\[
\sum_{v \in V} \deg^+(v) = \sum_{v \in V} \deg^-(v) = |E|,
\]

where \( \deg^+(v) \) represents the incoming connections at a specific node, \( \deg^-(v) \) represents outgoing connections and \( |E| \) is the total number of edges in the network system. Importantly, this and related network formulations may not be valid in cases where the complete map of the network system is not available. The violation of such a common graphical principle in the context of missing data points to the need for researchers to evaluate the assumptions and limitations of working with subsampled data from small segments of brain tissue or subsampled regional connectivity data.

Node degree distributions play an important role in describing the topological properties of a network, which can give rise to differential ability to effectively communicate long-range electrochemical signals in neuronal networks. The node degree of large, randomly generated networks, for example, follows a Poisson distribution [72], and can serve as a control profile with which to compare empirical neuronal data, as random graphs do not confer any regional specialization. Recent EM and neuronal tracer data have provided node degree distributions and offered insight into the complex neural networks of the brain.
8. Quantifying modularity of neuronal wiring diagrams

In addition to node degree distributions in neuronal connectomic data, other graph-based metrics offer insight into structural and functional properties of neuronal networks. Seminal studies have identified features of hierarchical and small-world architecture through the laminar patterns of connections between brain regions in macaque [74] and cat [75]. More recently, researchers have described structural hierarchy at multiple scales and in multiple model systems by quantifying the density of links within and between neuronal sub-structures. For example, Bassett et al. [21] applied Louvain community detection to the wiring diagrams of Caenorhabditis elegans as well as human structural and diffusion MRI data (figure 1c). The authors used modularity, a measure of to what extent like is connected to like, to describe the organizational hierarchy of multiple model organisms. From this analysis, they observed multiple examples of densely inter-connected modules with sparse intra-module connectivity. Similarly, Teller et al. [76] assessed calcium signalling dynamics between neurons that are organized into compact clusters in cell culture, finding significant modularity that represents a preference in the connectivity of neurons to those that are most similar in signalling activity.

In highly modular systems, the density of links inside sub-communities is much higher than the links between communities. Given a series of neuronal communities (or node types) labelled for each vertex \( c_1, c_2, \ldots, c_m \), the total number of edges that runs between nodes of the same type can be summed as follows [72]:

\[
\frac{1}{2} \sum_{ij} A_{ij} \delta(c_i, c_j),
\]

where \( A_{ij} \) is a symmetric and binarized connectivity matrix containing the value 1 for entries where nodes \( i \) and \( j \) are connected and 0 otherwise. The Kronecker delta \( \delta(c_i, c_j) \) is a piecewise function and is set to 1 whenever two nodes belong to the same node type. The factor of 1/2 accounts for the fact that every connection is counted twice in the binarized connectivity matrix \( A_{ij} \). To compare the sum of within-community connections observed in real tissues, it is helpful to subtract off the number of such connections that would occur by chance. In a network with \( m \) total edges, each node \( i \) in the system can be labelled as having degree \( k_i \). Because the total number of edge ends is \( 2m \), the probability that node \( i \) will connect to any other node randomly drawn from the graph is \( k_i / 2m \). The expected number of connections between nodes \( i \) and \( j \) in a random graph is then \( k_i k_j / 2m \). Subtracting this random connection component from the sum of within-community connections allows us to define the modularity quantity [72],

\[
Q = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j).
\]

Modularity is an important aspect of hierarchical network structure and has been used to evaluate the cost-efficiency of mapping complex neuronal systems that are densely packed into physical space [21].

By quantitatively measuring community structure through modularity, researchers can evaluate the extent to which processing networks are supported by sparsely inter-connected modules that are each composed of highly intra-connected neurons. Such high-level network views of neuronal systems are likely to become increasingly powerful as graphical models are built on more accurate and comprehensive EM wiring diagrams and inter-region tracer data. At present, graph-theoretic approaches can serve as a quantitative analysis framework on which functional models are based and can be applied at multiple scales to convey important structural insight even when based on down-sampled connectivity data. As discussed earlier, researchers applying network models to EM, fluorescence imaging and MRI have illustrated such an approach. In the future, such frameworks will likely be amenable to direct integration and across-scale comparisons of fluorescence and diffusion-based MRI studies [77,78].

9. Neuronal spheroids and organoids as models for human disease

Given challenges in obtaining human brain tissues for phenotypic profiling and structural characterization, assaying brain structure and function in patient-derived model tissues and organs may play a central role in relating genetic variation present in neurodevelopmental and degenerative diseases with human neuronal disorders. Human-derived somatic cells are now routinely induced into pluripotent stem cells (iPSCs) that are subsequently differentiated into neuronal aggregates known as spheroids, which serve as tissue models. Such models are known to take on physiological features not present in two-dimensional cell cultures. For example, a recent study established a three-dimensional neuronal sprouting assay for peripheral nerve regeneration and showed that neurite length is notably higher in spheroids as compared with two-dimensional co-cultures [79]. This
and related observations support the notion that three-dimensional cellular organization accounts for physiological interactions that are absent in two-dimensional counterparts. A similar example is the human-based tissue modelling of the congenital condition microcephaly. Lancaster et al. [80] highlight premature neural differentiation in microcephaly-affected patient-derived tissues, which results in a reduced expansion of radial glial cells, implying a markedly smaller size (figure 4a–c).

The accumulation of amyloid-β, which has been related to tauopathy in Alzheimer’s patients, has also been modelled in three-dimensional culture, enabling the investigation of pathogenic mechanisms (figure 4d) [81]. With improved measures of cell fate, spheroids can be used in existing high-throughput screening and imaging modalities and high content analysis pipelines. This rise has been accompanied by complementary progress in sample preparation methods. As the study of such systems with intact and whole-brain imaging techniques becomes more broadly accessible, an important limitation of spheroids is their highly scattering nature due to their large size and lipid composition, which renders conventional LM applied to native samples ineffective.

To overcome this obstacle and fully realize the utility of structural and molecular characterization of these three-dimensional samples, effective clearing techniques are proving indispensable to improve optical transparency and permeability to labelling reagents such as antibodies and nanobodies [82,83]. Optical clearing extracts lipid structures from deep within the tissue to eliminate scattering, rendering it impossible to study cellular ultrastructure using EM or lipophilic dyes (figure 3b). While a wide variety of clearing protocols have been developed, the underlying objective of attaining optically clear specimens remains similar [7]. One approach, three-dimensional imaging of solvent-cleared organs (3DISCO), is a fast procedure based on sequential incubation with organic solvents that offers the capability to image neuronal connections in several millimetres of depth with high resolution [8]. Passive clarity technique (PACT) is an alternative approach for three-dimensional imaging that is slower than 3DISCO but uses a solvent that is compatible with commercial light-sheet instruments and offers improved diffusion of antibodies and small molecules, thereby enabling sub-cellular level molecular interrogation of the entire cleared tissue (figure 3c) [7].

Figure 4. Tissue and disease modelling with cerebral organoids. (a) Brain region identity is captured using human cerebral organoids; fluorescence imaging for PAX6 (forebrain marker), KROX20 (hindbrain marker) and PAX2 (hindbrain marker); hippocampal markers NRP2, FZD9, PROX1 [80]. (b) Control-derived spheroids form thick fluid-filled cortical tissues. Microcephaly patient-derived tissues (line 14B) display neuroepithelium and outgrowth of neuronal morphology as compared to control [80]. (c) Modelling of microcephaly through cerebral organoids; day 22 staining indicates higher number of neurons (TUJ1, arrows) in patient-derived tissue [80]. (d) Comparison between six-week differentiated control (ReN-G) and familial Alzheimer’s disease (FAD) ReN cells (ReN-GA, ReN-mGAP). Detection of amyloid plaques in ReN-mGAP cells with Amylo-Glo (green—GFP, red—3D6 anti-amyloid-β, blue—Amylo-Glo, arrows—Amylo-Glo-positive aggregates) [81]. (Online version in colour.)
is highly relevant as an imaging modality to combine with clearing approaches due to its deep tissue penetration, multi-view imaging and reconstruction [84], and fast acquisition rate compared with scanning confocal or two-photon microscopy. Light-sheet imaging additionally reduces phototoxicity dramatically in living specimens due to selective illumination in comparison with widefield or confocal imaging. The technique has been used to image optically cleared whole mouse brains in which EGFP was used to target hippocampal pyramidal neurons [85]. A similar study used tissue clearing and SPIM to create three-dimensional models of dendritic spines and trees in the CA1 neurons in mouse hippocampi [86]. Work by Niedworok et al. [6] employed two-colour light-sheet microscopy to identify connectivity maps in neural circuits marked using virus-mediated trans-synaptic tracing. Combining multiplexed imaging of synaptic composition and structure with light-sheet microscopy, clearing and modelling of human disease using iPSCs offers important potential for understanding the impact of genetic variation on brain structure and function, as well as identifying therapeutic targets [87]. The use of physical expansion of brain tissues to simultaneously increase optical resolution and clear tissue samples also offers great promise for in situ characterization and hierarchical neuronal structure [88]. Bridging multiscale imaging of cerebral organoid structure with quantitative characterization that drives hierarchical functional modelling is in our view an important long-term aim of multiscale modelling research into the genetic basis of human brain function and disease.

10. Outlook and future challenges

In neurological disorders such as Alzheimer’s disease, neurodevelopmental disorders such as ASD, and psychiatric diseases including schizophrenia and depression, polygenic variation contributes to disease burden and can impact high-level brain function including memory and cognition. As genetic studies of human patient populations increasingly reveal genetic variants involving synaptic proteins, the need to relate synaptic protein levels and localizations to circuit function and organization becomes increasingly pressing. Characterization of neuronal structure in a hierarchical manner is now enabled by recent advances in multi-resolution imaging and structural analysis highlighted in this perspective. Important future potential lies ahead in leveraging the convergence of imaging, labelling, and genetic engineering technologies to better understand the mechanistic basis for currently untreatable and debilitating diseases of the brain. For LM, major aims include increased capacity for multiplexed imaging and localization using advanced molecular probes, as well as developing new strategies to employ LM and EM on the same sample. Computational tools that improve our ability to localize and count synaptic molecules and labelling strategies that allow for resolution of dendritic morphology in the same sample may offer new insight into integrated network-based models of brain structure that leverage powerful graph-theoretic approaches. Model ‘organoid’ systems that enable the ability to model human disease and characterize neuronal structure, identity and function in situ offer an important new opportunity to potentially diagnose, prognose and also treat major classes of neuronal disease. Because highly heterogeneous neuronal structures, from hundreds of individual molecules expressed within single synapses to complete neuronal circuit architectures, are at the root of all brain function, multiscale EM and LM imaging, analysis and structure-based modelling of brain and organoid function is certain to play a leading role in this area.

References

2. Cajal SR. 1906 In Nobel Lectures, physiology or medicine 1901 – 1921.

Authors’ contributions. L.H. and M.B. conceived the article. L.H., S.G., K.M. and M.B. wrote the article.

Competing interests. We declare we have no competing interests.

Funding. This research was funded by the National Institute of Health (NIH U01-MH106011-02) to M.B. and the National Science Foundation (NSF PHY 1305372) to M.B.

Acknowledgements. The authors are grateful to Simon Gordonov for useful discussions.


