wellcome

Scaling up connectomics

The road to a whole mouse brain connectome

Contents

Authors 4

Executive summary 6

1. Introduction 9

1.1 What is connectomics? 9

1.2 A short history of connectomics 10

1.3 The connectomics pipeline 12

1.4 The connectomics community 14

2. Value proposition and challenge 16

2.1 Value 16

2.2 Challenge 18

3. General strategy: conclusions and recommendations 22

4. Landscape and gap analysis 29

4.1 Sample preparation 30

4.2 Imaging 38

4.3 Data handling 49

4.4 Image alignment 52

4.5 Segmentation 54

4.6 Proofreading 59

4.7 Dissemination, annotation and analysis 62

5. Specific issues for low- and middle-income countries 66

6. Integrative connectomics: adding function to structure 68

6.1 Overlaying data within one specimen: correlative multimodal imaging 69

6.2 Integrating data across specimens: comparative connectomics and beyond 72

6.3 Incorporating preclinical and clinical imaging modalities 76

6.4 Recommendations 77

7. Extended impact 78

7.1 Neuroscience: from the mouse connectome to *connectopathies* *78*

7.2 Research questions that cross scales 79

7.3 High-throughput vEM for reproducible research and mapping morphological variation 79

7.4 Improvements in sample preparation protocols for tissue imaging 80

7.5 Impact on big image data handling and analysis 80

7.6 Recommendations 81

Conclusion 82

References 84

Appendices 95

Appendix A 96

Interviewed experts 96

Appendix B 98

Connectomics software 98

Appendix C 101

Glossary 101

Authors

Gregory Jefferis

Greg Jefferis is an MRC Investigator at the MRC Laboratory of Molecular Biology (LMB), Cambridge. He is also a Director of Research in the University of Cambridge Department of Zoology, where he leads the *Drosophila* Connectomics Group as part of an international collaboration with Oxford University and Howard Hughes Medical Institute Janelia, supported by the Wellcome Trust.

Greg's group studies the neural circuit basis of behaviour using the *Drosophila* (fruit fly) olfactory system as its main model. They were the first to identify widespread sex differences in fly brain neuroanatomy and to uncover a sex-specific switch in connectivity and information flow in an animal brain. Recent work has combined electron microscopy whole-brain connectomics with *in vivo* physiology and behaviour to understand the interaction between learned and innate behaviour. All of this work builds on innovative molecular and computational tools for cellular and synaptic resolution brain mapping.

Greg studied Natural Sciences at Cambridge before obtaining a Neuroscience PhD at Stanford in 2003. He returned to Cambridge as a research fellow of St John's College before taking up his LMB position in 2008. He is a former EMBO Young Investigator and FENS-Kavli Scholar. In 2019 he was awarded the Royal Society's Francis Crick Medal.

Lucy Collinson

Dr Lucy Collinson is an Electron Microscopist with a background in microbiology and cell biology. She has a degree and PhD in Medical Microbiology and carried out postdoctoral research with Professor Colin Hopkins at the Medical Research Council (MRC) Laboratory for Molecular Cell Biology (UCL) and Imperial College London, investigating membrane trafficking pathways in lysosome-related organelles in mammalian cells, using light and electron microscopy as key techniques. She has run a series of biological electron microscopy facilities since 2004, at UCL and then at the Cancer Research UK London Research Institute, which became part of the new Francis Crick Institute in 2015. With a team of electron microscopists and physicists, she oversees more than 100 research projects with over 60 research groups within the Crick, imaging across scales from proteins to whole organisms. Her microscopy and technology development interests include volume EM, correlative imaging techniques, cryomicroscopy, X-ray microscopy, image analysis, and microscope design and prototyping. Her group is using citizen science to gather hundreds of thousands of annotations of EM images to train deep machine-learning algorithms to automatically recognise organelles in EM images through the Etch a Cell project on the Zooniverse platform. She has co-authored more than 100 research and review papers, given more than 70 invited and keynote talks, and has sat on more than 30 international advisory boards, panels and committees in advanced imaging.

Carles Bosch

Carles Bosch is Principal Laboratory Research Scientist of the Sensory Circuits and Neurotechnology Laboratory at the Francis Crick Institute (London). His current research addresses how the mammalian brain detects odours, by integrating insights on the structure and function of the sensory neuronal circuits underlying perception.

Alongside a multidisciplinary team he has developed a correlative multimodal imaging workflow that harnesses the coverage and resolving power of *in vivo* multiphoton microscopy, synchrotron X-ray tomography and volume electron microscopy techniques. This strategy has allowed the targeting of specific regions in the brain to be explored with ultrastructural precision while keeping the big picture that provides context.

After studying Biotechnology at the Autonomous University of Barcelona, he obtained his PhD at the University of Barcelona studying how synaptic plasticity is regulated in the adult mouse brain at the transcriptomic, molecular and structural levels.

More broadly, he is also interested in facilitating scientific discovery through multidisciplinary collaboration. He has participated in, and organised, multiple science-focused events tailored to both specialised and broad audiences.

Marta Costa

Marta Costa is a Senior Research Associate at the Department of Zoology, University of Cambridge and the project lead for the *Drosophila* Connectomics Group directed by Greg Jefferis.

Marta studied Biology at the University of Lisbon, Portugal, followed by an MSc in Neuroscience at UCL and a PhD at the University of Cambridge. She then joined two multidisciplinary projects that collect and integrate genetic and neuronal *Drosophila* data, making it available to the research community via a web-based resource. Since joining Greg's group in 2016, she has worked on understanding the circuit impact of natural variability in neuron shape and connectivity between individuals.

Philipp Schlegel

Philipp Schlegel is a Postdoctoral Researcher at the Department of Zoology of the University of Cambridge and the MRC Laboratory of Molecular Biology (LMB).

Philipp studied Biology at the University of Bonn, Germany, where he also obtained a PhD in Molecular Biomedicine in 2016. During his doctoral studies he was involved in early connectomics work in *Drosophila* larvae before joining Greg Jefferis' group in Cambridge where he uses flies to study how brain wiring differs within and between individuals.

Executive summary

The brain contains a vast network of interconnected neurons. Connectomics is the field that aims to map these brain connections at the resolution of the synapse, the nanometre-scale contacts that allow transfer of information between neurons. These networks have been selected by hundreds of millions of years of evolution, but changes in connectivity over our lifetimes are thought to encode most of our learned experiences. Furthermore, connection differences likely underlie a range of mental health issues. Like genomics, connectomics aims to provide a comprehensive platform for discovery. However, it occupies a crucial mid-level between the developmental instructions encoded in DNA and the dynamic processes of the brain, and in doing so brings us one step closer to understanding the relationship between (brain) structure and function.

Connectomics started in the 1970s with a 15-year effort based in Cambridge (UK) to reconstruct the synaptic wiring diagram of the 302 neurons of the nematode worm, *C. elegans*. This seminal work has supported almost all subsequent investigations of the nematode, one of the most influential genetic model organisms. However, there is a sense in which the worm connectome was ahead of its time. For example, methods to probe the behavioural significance of the connectome by recording and manipulating neuronal activity (e.g. through optogenetics) only came into force in the 2000s and 2010s. Indeed, this early experience has resulted in a widely, although not yet universally, appreciated conclusion that synaptic wiring diagrams are a necessary but not a sufficient resource for understanding brain function.

Recent evidence, especially in the fruit fly, *Drosophila melanogaster*, indicates that whole-brain connectomics can have a transformative impact on neuroscience. Perhaps surprisingly, the structural wiring diagram can even reveal algorithmic principles of dynamic processes such as memory formation and recall or spatial navigation. *C. elegans* neuroscience developed after the worm connectome; in contrast, *Drosophila* neuroscience was already a very active field before connectomes became available, so the impact is more obvious. Furthermore, as brains become larger and numerically more complex, a greater fraction of the behavioural complexity of the animal may reside in the wiring diagram. For these reasons *Drosophila* appears a better guide to the impact of future connectomes in larger animals.

Neurons can extend over distances of centimetres or even metres, but their thin branches and synaptic connections require nanometre-resolution imaging. For this reason, since the beginning, connectomes have been generated by electron microscopy (**EM**). In the last five years, there have been major technical advances in imaging enabling not just whole fly brains but also a cubic millimetre of mouse cortex; furthermore, building on the artificial intelligence (**AI**) revolution, computer vision-based algorithms can now extract the shapes and connections of neurons from the greyscale electron microscopy images. This has reduced human labour by about 50-fold, transforming the 2,000 person-year project of a whole fly brain into an effort of about 40 person-years. These advances, together with the obvious impact of fly connectomics, have set the field dreaming about the possibility of obtaining larger whole-brain connectomes. This has been articulated for example in a 2020 opinion piece (Abbott et al., 2020) as well as a series of workshops (the NIH BRAIN Initiative and the Department of Energy Office of Science) organised in early 2021 by the United States National Institutes of Health (NIH) and Department of Energy (DOE), who, together with the Wellcome Trust, were the largest funders of the Human Genome Project.

However, the road to a mouse connectome is still a long one – over 10,000 times longer than the fly based on the amount of neural wiring in each brain. The mouse brain is a tenth of a teaspoon (500 cubic millimetres) in volume, but this is 500 times what has ever been imaged by electron microscopy; the raw 3D image data for just one mouse brain would fill ~500,000 laptops or the storage servers at the Wellcome Genome Campus. For this reason, it is important to take stock of the scientific, technological and organisational problems that must be solved to prepare for mammalian connectomes, but also to chart what might be valuable stepping stones along the way. Using the goal of a mouse brain connectome as a focal point, this report attempts to do just that.

The report is based on consultation with 50 experts around the globe (**Appendix A**) by a team of five neuroscientists/microscopists based in the UK. Our report reflects a synthesis of these sources and our own interpretation of this research. We first introduce the connectomics field (**Section 1**) and the scale of the challenge, as well as giving a taste of the likely value of this goal (**Section 2**). We then provide an initial synthesis of these consultations by drawing conclusions about high-level strategy and making a matching first set of recommendations (**Section 3**). This is then followed by a detailed consideration of the current state of the art, under three main headings: sample preparation; imaging; data (which includes computer processing and sharing of raw image data as it is transformed into a connectome to be used by neuroscientists). We summarise both the state of the art and gaps that need development and funding (**Section 4**). Finally, we address issues of equitable access both to the ongoing project and the eventual data (**Section 5**) as well as the potential broader impacts across the biomedical sciences (**Sections 6 and 7**). Our recommendations in Sections 4 to 7 indicate their timescale: short (S), medium (M) or long term (L).

Our high-level conclusions and recommendations can be summarised in five main points:

1. A whole mouse brain connectome would have a major impact on basic and translational neuroscience. Today's cutting-edge technology might be able to deliver a connectome, but the timescales and costs remain uncertain. In the short to medium term, targeted investment to develop key technologies and strategies across the connectomics pipeline are still necessary both to de-risk the project and reduce the uncertainty in costs and timelines.
2. A whole mouse connectome should be a flagship goal on a 10–15-year timeline. However, this should be part of a broader effort to develop a connectomics ecosystem. Whole smaller brains (as well as small pieces of mouse brain) should be used as stepping stones; this will drive technology development and demonstration at all stages from sample preparation to analysis and experimental impact.
3. While there will probably only be a handful of centres worldwide that would image a whole mouse brain, later stages in the pipeline could be much more distributed. We will need to develop different modes of collaboration for each. Funders will need to support and shape both, especially the more distributed aspects, to enable equitable global participation and access.
4. To maximise benefit from a reference whole mouse connectome, we will need to enable connectome generation for many smaller subvolumes of mouse brain that have also been subjected to correlative imaging of neuronal activity or molecular labelling. This information can then be overlaid on the reference connectome, a process that we term 'integrative connectomics'. Both correlative and integrative connectomics must be developed as part of the initial project, should target a range of species, and be widely accessible to experimentalists by the time a mouse connectome is available.
5. With today's cutting-edge technology, the most costly step (in time and money) in generating a whole mouse connectome would still be the manual proofreading required to fix errors in the automatic reconstruction of neurons. This must be reduced from the ~100,000 person-years it might take today. Funded initiatives should include an accessible computational platform to lower the barrier of entry to contribute improvements in automatic segmentation methods, more extensive ground truth for existing volumes, as well as improvements in proofreading tools and strategies.

We believe that it is now essentially inevitable that connectomics will have a huge and growing impact across basic and clinical neuroscience, the biomedical sciences more generally and eventually the knowledge economy and wider society. Whether this takes 10, 20 or 30 years will be largely dependent on the effective collaboration between a wide range of scientists and technologists as well as the decisions of public and private funders. Any organisation, or indeed nation, with a serious commitment to basic or clinical neuroscience needs to consider this change in their future strategies. We invite you to join us as we chart our current view of how this long but fruitful road will be travelled.

1. Introduction

1.1 What is connectomics?

The brain contains a vast network of interconnected neurons. These networks have been selected by hundreds of millions of years of evolution, but changes in connectivity are also believed to encode most of our learned experiences; it is increasingly thought that connection differences may underlie a range of mental health issues. Connectomics is the field that aims to map these brain connections and link them to the shape of every neuron. Named by analogy to genomics, connectomics has the long-term goal of being comprehensive – mapping all the connections in a brain or brain region. If the individual bases that make up DNA are the fundamental unit of the genome, then synapses – the individual connections between neurons – are the fundamental unit of the connectome. Although connectomics is sometimes used to refer to lower resolution studies in which projection patterns between different brain regions are revealed, we will focus strictly on approaches where actual connections, i.e. synapses, can be resolved.

Neurons can extend over distances of centimetres, even metres, but synapses and fine neuronal wiring are nanometre-scale structures. Connectome mapping has therefore depended on methods for acquiring and analysing high-resolution large-field-of-view images of the brain, typically using electron microscopy (**EM**). It is now a technologically intensive field with neuroscientists collaborating with developers of high-speed imaging technologies, as well as computer vision experts contributing to extract connectivity information from image data.

1.2 A short history of connectomics

To set the stage, we now present a short history of connectomics. This is necessarily highly selective since we concentrate on studies, especially whole-brain connectomics, that help illuminate the path to a whole mouse connectome.

The worm

The first complete connectome was that of the nematode worm, *C. elegans*. The connectome consisted of the detailed morphology of all 302 neurons in the animal together with synapses between the neurons and muscles. It was obtained at the Medical Research Council (MRC) Laboratory of Molecular Biology in Cambridge (UK) by John White, Eileen Southgate, J N Thomson and Sidney Brenner in over a decade of painstaking work, and finally published in 1986 (White et al., 1986). To achieve this, the group collected serial ultrathin (~50 nm-thick) sections of the metal-stained, resin-embedded nematode, and imaged all the slices in a transmission electron microscope (**TEM**). Multiple nematodes were employed to generate a 'canonical' map extracted from five specimens, which accounted for the total of 302 neurons, classified into 118 classes, and establishing 5,000 chemical synapses, 2,000 neuromuscular junctions and 600 electrical gap junctions. In that first report it is already acknowledged that this structural connectome might not reveal all the routes by which information is exchanged between cells in the nervous system – pointing to other intercellular communication mechanisms such as peptide signalling, humoral signalling or electrical leakage. Still, the wiring diagram reported in this study constituted the first comprehensive representation of a complete nervous system.

The fruit fly

The completion of the *C. elegans* connectome in 1986 signalled the birth of synaptic resolution connectomics. However, the next 20 years turned out to be a relatively fallow period. Perhaps other groups found the ~40 person-year commitment of this heroic first effort a disincentive. Signs of a rebirth in the field include the redevelopment of the serial block-face scanning electron microscopy (SBF-SEM) method (which combined physical sectioning and EM imaging into a single device) (Denk and Horstmann, 2004) and its application to the vertebrate retina (Briggman et al., 2011). In 2006, the Howard Hughes Medical Institute opened their new Janelia Research Campus with a focus on neuroscience and new imaging technologies. This quickly became a hub for work on the nervous system of the fruit fly, *Drosophila*, the most popular invertebrate model organism, with a particular focus on EM connectomics.

The first milestone was the imaging and assembly of a complete electron microscopy volume of a *Drosophila* larva central nervous system (completed in 2012 and eventually reported in Ohyama et al., 2015). An international effort has used collaborative web-hosted tools (Schneider-Mizell et al., 2016) to progressively reconstruct the morphology and connectivity of the 2,500 neurons in the first instar larval brain, a process that is now (2022) complete. This ten-year gap emphasises the distinction between completing the imaging of a volume suitable for generating a connectome and the process of extracting and analysing this complete connectivity information. Nevertheless the intervening years saw many studies selectively reconstructing brain-spanning circuits, shedding light on processes including learning and memory (Eichler et al., 2017; Eschbach et al., 2020) and action selection (Jovanic et al., 2016; Ohyama et al., 2015). We will consider later how the model of progressive, collaborative and distributed connectome reconstruction might apply to larger brains, including the mouse.

Two large efforts in adult fly brain connectomics were also taking place at Janelia. The first employed a high-throughput version of the serial section transmission electron microscopy (ssTEM) techniques used to image the worm and fly larva. This resulted in the largest complete EM brain volume to date (Zheng et al., 2018) containing an estimated 130,000 neurons. Another useful quantity in connectomics is the length of the neuronal wires or cable, estimated at about 200 metres in the adult fly brain. Collaborative, manual tracing in this volume proceeded at about 1 mm/person per day with over 10 m of neuronal cable reconstructed by 2020, but complete reconstruction of the whole brain would have taken a prohibitive 2,000 person-years.

A second effort nucleated around the large FlyEM project team (eventually comprising over 50 members) who used a different technique – focused ion beam scanning electron microscopy (FIB-SEM) (Xu et al., 2017), see below – to image roughly half of the central brain of *Drosophila*, containing about 25,000 neurons. In collaboration with Google Research, they deployed a new automated segmentation technique (Januszewski et al., 2018) followed by over 50 person-years of manual proofreading to fix errors. This resulted in the release of the hemibrain, the largest complete connectome to date (Scheffer et al., 2020). This strategy, combining automated segmentation and a large professional proofreading team, meant that less than two years elapsed between completion of the EM volume and public release of a finished connectome. End users can query connectivity between their favourite neurons in seconds over the web, assured that errors are extremely rare. This project also provided an example of a successful partnership with a private company (Google) to deliver a key step in the connectome pipeline.

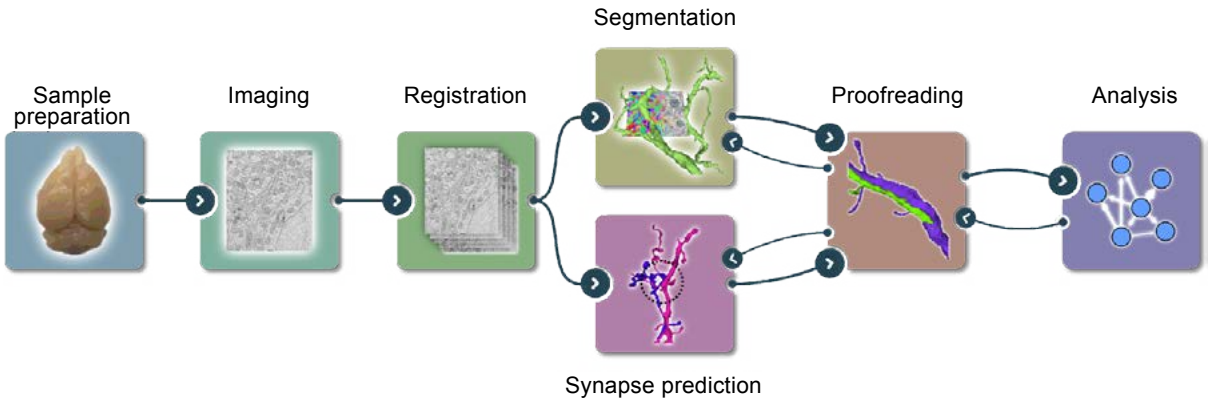
Vertebrate connectomics

Vertebrate connectomics has developed rapidly over the last 15 years, with a major emphasis on small regions of the mouse brain. Early work included studies on sensory structures such as the retinal circuits in the eye (e.g. Briggman et al., 2011), through which information enters the brain, or motor neurons which control the muscles, thereby forming the final pathway for brain output (Lu et al., 2009). Much recent work has been focused on the mouse cortex (Berning et al., 2015; Gour et al., 2021; Kasthuri et al., 2015; Lee et al., 2016; Loomba et al., 2022; Motta et al., 2019a), with some exceptions in other sensory regions (Hua et al., 2021) and also recently in human brain biopsy samples (Shapson-Coe et al., 2021). Important insights have already been extracted from these studies, such as disproving that physical proximity between neurites can predict excitatory connectivity (Kasthuri et al., 2015; Motta et al., 2019b). Long-range projections in these datasets are, however, almost by definition truncated, so these partial connectomes provide limited information about the nature of inputs. Nevertheless, mammalian neuroscientists have successfully combined functional data with EM – for example, overlaying calcium-imaging signals onto individual somata in EM volumes (Bock et al., 2011; Briggman et al., 2011). This powerful correlative multimodal imaging approach, sometimes termed functional connectomics, is an important example of how other information can be layered onto connectomes. However, there are of course limits to the stimulus sets that can be explored, and live imaging methods still have restricted fields of view.

Outside mammals, zebrafish, another influential genetic model organism, is the vertebrate that currently seems to have the best prospects for whole-brain connectomics, albeit in the immature larval stage. The fish larva between 3 and 6 days old already displays a range of innate and some learned behaviours, and has a brain about the size and similar neuron number to adult *Drosophila*. A whole-brain EM volume was reported (Hildebrand et al., 2017), but this does not have the resolution required for tracing connections between neurons, while smaller volumes containing e.g. the olfactory system have been reconstructed to completion (Wanner et al., 2016). A 6-day-old zebrafish larval EM volume has recently been published, and initial indications are that this can be used for reconstruction at synaptic resolution (Svara et al., 2022).

1.3 The connectomics pipeline

Figure 1: Schematic pipeline for whole-brain connectomics



To provide context for the discussion that follows it will be helpful to have an outline understanding of how synaptic resolution connectomes are generated. Although EM connectomics pipelines vary in their details, they can be divided into several key steps with extensive interdependencies (**Figure 1**).

Sample preparation includes extraction and fixation of the brain, heavy-metal staining to provide contrast, and physical sectioning of the specimen to produce a geometry suitable for imaging. Failure modes for this step include loss of material during sectioning and staining artefacts.

The **imaging** modality has to be able to resolve individual synapses and trace membrane-bound neurites which requires very high resolution, typically obtained by voxels 10 nm wide or smaller. Currently, only volume electron microscopy (**vEM**) can achieve this. While the work can be spread across multiple microscopes (if the single specimen can be subdivided), this step is generally slow (weeks to months even for small volumes). The potential pitfalls depend on the exact imaging technique, but loss of tissue sections is always a problem.

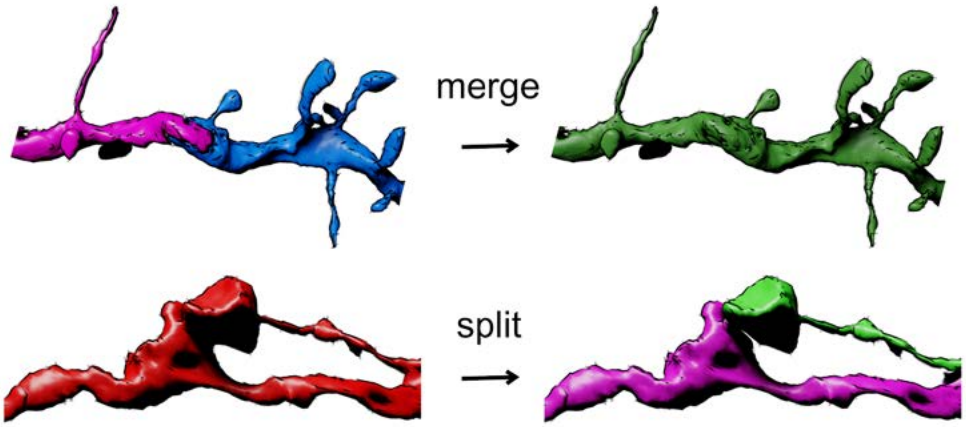
Registration describes the process that takes the small two-dimensional image tiles produced during imaging and aligns them into a cohesive, three-dimensional volume. This process is particularly challenging when thin sections are being imaged, since these can stretch or fold.

From the greyscale image volume, we have to extract the morphology and connectivity of the neurons. Until recently, this was a mostly manual process that involved massive human effort to reconstruct neurons through the sections and identify synapses. Today, machine-learning techniques produce automatic **segmentation** of neurites and predict synaptic contacts.

Even current state-of-the-art segmentation pipelines will not produce out-of-the-box connectomes. Therefore, substantial human **proofreading** effort is still necessary to produce connectomes of sufficient quality for circuit analysis by end users.

Figure 2: Automatic segmentation contains errors that need to be fixed during proofreading

The two basic operations are 'merges', where two fragments that belong to the same objects are merged, and 'splits', where two objects that are incorrectly merged are separated.



Base segmentations contain both 'false splits' (when two parts of the same neuron are separated) and 'false merges' (i.e. when two different neurons are incorrectly joined together) (**Figure 2**).

Finally, the **analysis** stage is a crucial part of maximising the utility of connectomes. For example, expert *annotation* can group neurons with related functions and assign cell types that have previously been reported in the literature, providing an entry point for neuroscientists. Cell typing may also allow molecular or functional data to be layered onto the connectome. Lastly, point-and-click web tools or powerful programmatic interfaces are required to serve different end users.

Critically, the above steps are highly interdependent. For example, a high-quality staining will produce better images which will improve the segmentation, which in turn means less proofreading is required. At the same time, there is also bidirectionality: data analyses can inform proofreading efforts which in turn can help fine-tune segmentation.

1.4 The connectomics community

Stakeholders

The primary stakeholders in the effort to map the mouse brain connectome are the researchers directly involved in the effort to produce the first structural wiring diagram of the mouse brain at synaptic resolution. This community incorporates a range of expertise, including neuroscience, cell biology, chemistry, microscopy, hardware engineering, software engineering, computer vision and artificial intelligence (**AI**). The brain research community will be direct beneficiaries of connectomics efforts, since wiring diagrams will provide an invaluable resource to understand how neuronal circuits operate. The first impact will likely be for research that can clearly link brain function/structure to an input or output (e.g. sensory inputs or motor outputs), whereas processes which involve more complex internal brain circuits (e.g. memory storage and retrieval) may take longer to benefit. However, experience in *Drosophila* suggests that those studying functions located deeper in the brain, such as memory storage and retrieval, will soon benefit from these connectomes. This is both because peripheral neurons can help provide the context to define what information central neurons are processing, and because central circuits appear to have a mix of more stereotyped and more variable nodes. Besides experimental neuroscientists, it is likely that connectomics will usher in a new wave of computational modelling that leverages actual patterns of network connectivity or uses the synaptic location of identified inputs onto individual neurons.

However, while structural wiring diagrams have great intrinsic value, further integration with orthogonal information will greatly assist understanding the relationships between structure and function. Replicates may be necessary to understand circuit stereotypic patterns across individuals, and larger experimental cohorts might be necessary to disentangle the mechanisms underlying the pathologies that depend on circuit deficits – or connectopathies – as well as other disease-focused studies addressing e.g. neurodegeneration, cancer, mental health conditions and ageing. The broader neuroscience community will benefit from technical developments that enable structural analysis of wiring diagrams beyond the brain, including the spinal cord and the peripheral nervous system controlling breathing, heartbeat, blood flow, senses and movement.

Beyond neuroscience, technical developments enabling high-resolution imaging and analysis of tissue volumes will be of interest to the broader life science and clinical research community with a requirement for imaging biological entities that span centimetres with detection of critical features in the 10 nm range; cell atlas consortia (e.g. the Human Cell Atlas project) who wish to increase their depth of information to the organelle and membrane structure level; pathologists, surgeons and patients whose practice and health could be improved by the incorporation of high-speed, high-resolution, large-volume imaging technologies into diagnostic pathways; the physical and material sciences community who could apply the advances made in imaging and analysis in connectomics to their own disciplines;

imaging scientists, engineers, computational scientists, physicists, chemists and research technology professionals, whose expertise will be required to deliver the first mouse brain connectome; microscope and ancillaries manufacturers, and big tech companies, who will contribute to and benefit from advances made in the field; funding bodies and philanthropists, who will support the effort and thus generate new knowledge, ideas and innovation for the public good; and the general public, who will contribute to the project directly through citizen science programmes and who will benefit from discoveries (foreseen and serendipitous) and innovation outputs.

International connectomics activities

Although there is a long and distinguished history of using electron microscopy to examine features of the mammalian brain, the number of groups actively working on large-volume electron microscopy approaches currently suitable for high-throughput connectomics in the mouse is relatively limited. Representation is heavily biased to the USA (Allen Institute, Harvard, Howard Hughes Medical Institute (HHMI) Janelia, Princeton) and Germany (Max Planck Frankfurt, Munich, Bonn). In the UK, the Francis Crick Institute has activity in both volume EM and synchrotron-sourced hard X-ray imaging for connectomics in collaboration with multiple synchrotrons (Diamond Light Source, UK; European Synchrotron Radiation Facility (ESRF), France; and Paul Scherrer Institut (PSI), Switzerland), some of which host dedicated hard X-ray connectomics labs (ESRF, PSI); other groups are active in the Chicago/Argonne National Labs. Large-scale work on zebrafish has also been limited to about three labs (Max Planck Munich, Friedrich Miescher Institute for Biomedical Research Basel, Harvard). Data collection for *Drosophila* connectomics has been driven by groups at HHMI Janelia Research Campus, but the focus of this institute is shifting to cell physiology with a change in leadership; Harvard (US) and Cambridge (UK) are emerging centres. Other centres for invertebrate work include European Molecular Biology Laboratory (EMBL) and Exeter. Compared with biomedical science in general, data acquisition for connectomics appears intensive in capital equipment rather than people. It is therefore quite possible that significant new efforts may appear in locations where institutions or governments are ready to make the necessary infrastructure investments and develop a core of skilled personnel, including in China.

All electron microscopy depends on commercial equipment providers. However, some techniques depend principally on the expertise of small research groups to customise relatively generic instruments (e.g. GridTape-TEM, detailed later). Other promising imaging technologies are currently in the hands of single suppliers (multibeam SEM, Zeiss). In the later stages of the pipeline (segmentation and proofreading) Google Research has been the largest player, but there are also a handful of small specialised commercial entities that work closely with researchers.

2. Value proposition and challenge

2.1 Value

Lessons from invertebrates

The *C. elegans* connectome reported in 1986 (White et al., 1986) has supported almost all subsequent investigations of the nematode, one of the most influential genetic model organisms. This includes both studies of neural function and also development – the availability of the whole animal EM dataset was a critical reference, while the entire cell lineage of the animal was reconstructed (work for which John Sulston shared the 2002 Nobel Prize with Sydney Brenner and Bob Horvitz). However, Brenner's original intention was to help understand the behaviour of the animal – and eventually to model it. With this goal in mind, there is a sense in which the worm connectome was ahead of its time. For example, methods to probe the behavioural significance of the connectome by recording and manipulating neuronal activity (e.g. through optogenetics) only came into force in the 2000s and 2010s. Indeed, this early experience has resulted in a widely accepted, though not yet universally appreciated, conclusion that synaptic wiring diagrams are a necessary, but not sufficient, resource for understanding brain function.

The recent advent of fly connectomes, following several decades of intense work on the structure and function of *Drosophila* brain circuits and behaviour, has clearly had a transformative impact. This process is still ongoing since finished connectomes for the adult brain and nerve cord are only expected over the next year. While many early adopters are already contributing to the completion and analysis of these datasets, the majority are limited to the partial *hemibrain* connectome (Scheffer et al., 2020). Nevertheless, for those in the field, the change is dramatic. The flow of information can be followed brainwide: sensory input > (multisensory) integration > output to locomotor centres (current focus) for a range of complex behaviours from sensory processing, learning and memory, spatial navigation, decision making, aggression, sleep (Galili et al., 2022). This includes cases in which connectomics has defined key algorithmic features of complex dynamic processes such as spatial vector computation (Hardcastle et al., 2021; Hulse et al., 2021; Okubo et al., 2020), providing a circuit basis for previously observed functional and behavioural insights (Galili et al., 2022), or revealed the circuit basis of dynamic cognitive processes such as memory extinction (Felsenberg et al., 2018).

Several crucial observations follow. The value of EM data as a comprehensive cell type catalogue (or neuronal parts list) is very high. The majority of cell types identified in the fly connectome had not previously been reported (Scheffer et al., 2020); even in areas that had been intensively studied, many new cell types were identified (Bates et al., 2020; F. Li et al., 2020). In the worm and the fly, stereotypical neuroanatomy has facilitated classification of neurons into cell types, recognisable between individuals solely on the basis of cell shape and position. This has allowed molecular and functional data acquired across many labs to be layered onto the handful of reference connectome datasets. We refer to this ability to combine data acquired from different individuals as integrative connectomics. Establishing how this will be possible in larger organisms, including mammals, is a key intellectual and practical goal in maximising the value of connectomes. This integrative connectomics approach also facilitates the identification of genetic reagents to label and manipulate neurons of interest, allowing efficient testing of circuit hypotheses.

The mouse brain: a blueprint for mammalian connectomes

The arguments in favour of a whole mouse brain as a target for connectomics have recently been presented in a position paper authored by some of the leaders in the field of connectomics and circuit neuroscience (Abbott et al., 2020). The mouse is the most commonly used animal model in research (Robinson et al., 2019), backed by a huge range of resources for neuroscience. These include transgenic strains and mature viral methods that allow neurons to be labelled based on projections, specific connections and molecular identity (Luo et al., 2018), which will prove invaluable for exploiting a connectome. The mouse also has the most extensive molecular characterisation of cell types among mammals together with molecular brain atlases.

A mouse brain connectome would therefore be an invaluable resource to a wide sector of the research community. As a terrestrial mammal and common disease model, it would be a significant step closer to humans, whose 3,000x larger brains will likely remain off limits to connectomics for several decades. In contrast, the mouse brain, while 500 times the volume imaged with today's state of the art, is both tempting and appears technologically within reach. Therefore the question is not *whether* but *when* it will be achieved. A whole brain brings added value: principal neurons extend their processes through vast regions, making it extremely difficult to understand neuronal circuits entirely from local circuit reconstruction. It is likely therefore that the first complete neuronal circuits may be provided by approaches that map entire brains.

Finally, the technological developments promoted by the quest of obtaining a mouse brain connectome should also accelerate the generation of many smaller connectomes. Altogether, the mouse brain connectome is an inspiring and scientifically relevant goal, capable of attracting the talent this challenge will require.

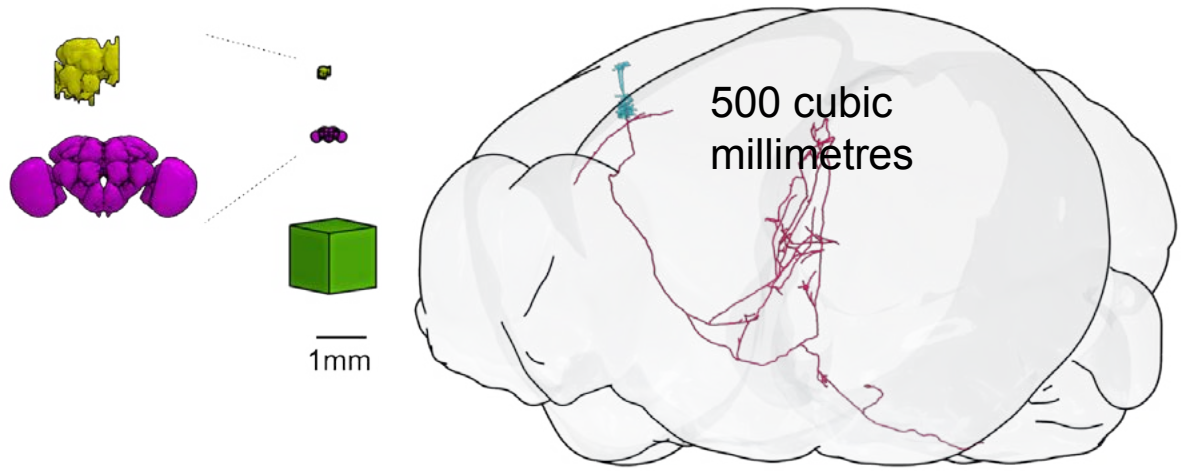
2.2 Challenge

Scale

The fundamental challenge in mapping a whole mouse brain connectome lies in the size of the sample, about 500 mm^3 (Badea et al., 2007); this is almost three orders of magnitude larger than the 1 mm^3 which represents the current demonstrated limit for vEM technology (**Figure 3**). This increased scale then results in many subsidiary challenges through the pipeline. Many of these challenges are relatively obvious after some simple 'back of the envelope' calculations. Others depend on specific technical limitations when moving from a sample that is about 1 mm across to a whole mouse brain, 7–13 mm in its shortest and longest axes (Allen Institute for Brain Science, 2015). Still others are due to complex interrelationships between trade offs that must be made when selecting different sample preparation and imaging strategies. Finally, some challenges, while clearly identified, remain quite poorly constrained at this point: the most obvious being the amount of human labour required to proofread the mouse connectome, with a major impact on the final cost and timeline.

Figure 3: Size comparisons

Frontal view of the *Drosophila* hemibrain dataset (yellow, the largest complete connectome to date), a fly brain (purple, largest synaptic resolution whole-brain EM dataset to date) vs cubic millimetre (largest EM volume to date) vs a whole mouse brain (500 cubic millimetres). A cortical neuron is rendered in the mouse brain with blue dendritic arbour (approximately a millimetre across) and red long-range axonal projection (Janelia MouseLight project).

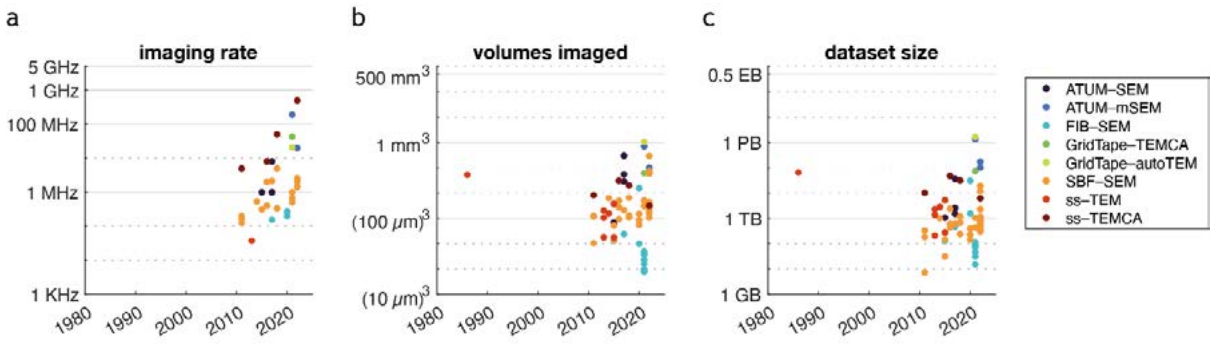


Some brains have finer neuronal processes than others, and different EM imaging technologies have different resolution constraints, but a reasonable estimate is that imaging should have a resolution no coarser than 1,000 nm^3 per voxel: this is equivalent to a cube with a 10 nm side when the imaging resolution is isotropic (i.e. the same in all axes). 1 mm^3 of brain would have 1 mm/10 nm = 100,000 voxels along each edge, making 1,000 teravoxels (1E15 voxels). Assuming just 1 byte of storage for each voxel, 1 mm^3 would occupy 1,000 terabytes = 1 petabyte. A whole mouse brain would be 500 times this (500 petabytes or 0.5 exabytes), similar to 500,000 laptops or the total storage capacity of the Wellcome Genome Campus. Furthermore, as soon as a first mouse brain is completed there would be strong arguments to collect e.g. a different sex or genotype or to look at a connectome after learning. The raw image data will require multiple preprocessing stages before becoming a continuous three-dimensional volume. And then, tracing and annotation efforts will require parallel distributed access in order to be able to complete the task in a reasonable timeframe. Given the scale of the data, there will be major engineering and cost challenges to address.

The cm-scale of the specimen also requires improvements to staining and embedding protocols. Since these processes typically scale with the square of the linear dimensions of the object, a 1 cm brain might take about 100 times as long as a 1 mm tissue block. This year some of our interviewees have shared very promising results on whole-brain staining (Song et al., 2022), but validation in other laboratories is still necessary, as well as work to see how they perform with specific downstream technologies (embedding, sectioning, imaging, segmentation).

A cm-scale object is too large for any existing nm-resolution technique to image in one piece. Physical sectioning is required. Traditional electron microscopy workflows use ultrathin sections (<50 nm thick), but it is not possible to cut a 1 cm x 1 cm x 50 nm section; furthermore, this would imply collecting and preserving 300,000 sections, which is way beyond what has been achieved. New approaches are required. Reliability is also paramount since lost or damaged sections can make it impossible to follow thin neuronal processes.

Figure 4:
Historical overview of published vEM datasets by imaging rate, volume and data size (note that y axes are logarithmic). Imaging rates are based on published information and refer to a single imaging device (normalised when multiple devices were employed in parallel) but must be interpreted with caution since they are often based on burst imaging rates without including downtime. Supporting data available in https://doi.org/10.5281/zenodo.7599974



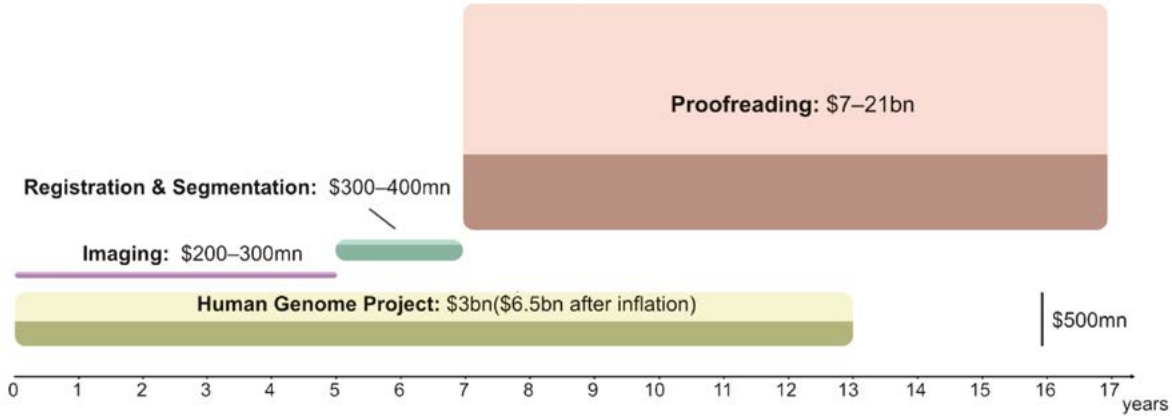
Time and cost

In addition to requiring development of new technology, generating a whole mouse brain connectome would be time- and resource-intensive. The acquisition speed of published volume EM datasets has drastically increased in recent years (**Figure 4** supporting data available in https://doi.org/10.5281/zenodo.7599974). Imaging rates have trended markedly higher over the last decade and are within an order of magnitude of what might be required for a whole mouse brain. In contrast, sample volumes are still almost three orders of magnitude away. Imaging burst rates in the GHz range (i.e. one billion voxels per second) are now possible, but sustained rates across a longer project, given issues such as specimen movement and exchange and machine downtime, mean that 100–200 MHz (i.e. one to two hundred million voxels per second) is currently a more reasonable production benchmark. Acquiring a 500 petavoxel mouse brain volume over five years would require a mean imaging rate of 3.2 GHz; this could presently be reached by about 20 machines operating at realistic rates.

Costs to consider include expert scientific and technical staff (including engineers and computer scientists), equipment and their operating costs (e.g. the microscopes for imaging), compute time (e.g. for registration and segmentation of the image data), data storage and additional staff for proofreading and annotation of the dataset. This is why virtually all large connectomes involve either a consortium of labs or operate as a community effort, and/or are backed by major funding.

Based on extrapolations from existing datasets and numbers provided by expert interviewees, we estimate the costs for a mouse connectome at $7.5–21.7 billion ($7.5–21.7k million) using current state-of-the-art technologies (**Figure 5**, see box for detailed breakdown). In these estimates, by far the largest fraction ($7–21bn) is spent on proofreading, followed by registration and segmentation ($300–400mn) and imaging ($200–300mn). Hence, bringing down proofreading costs would have the biggest impact on the total costs: even a two-fold decrease in the amount of proofreading required would potentially save billions of dollars. Fortunately, substantial improvements in this area seem achievable – and even essential – e.g. through improvements in image or segmentation quality. By contrast, it is less clear whether the costs for imaging and segmentation will go down in the near future given that research and development is primarily focused on improving speed and quality of output, rather than efficiency. (Please see the gap analysis sections for details.) While even our lower end estimate of $7.5bn is very large, it is worth pointing out that the Human Genome Project is estimated to have cost around $3bn or $6.5bn today, after inflation. Note also that these estimates are specific to the generation of a single mouse connectome; they do not include the cost of improving and validating new technology or related work leveraging a mouse connectome.

Figure 5: Estimated cost for a mouse connectome using current technologies



Estimated costs and timeline for a mouse connectome using current technologies

Imaging: A single multibeam SEM can image ~5 mm³ at 10 nm resolution per year. To image the 500 mm³ of the mouse brain in 5 years would require at least 20 such microscopes. At $5mn per machine this incurs an upfront cost of $100mn plus an additional $100–200mn operational costs over the course of 5 years.

Registration and segmentation: Cloud compute costs for image registration and subsequent segmentation are estimated at $300–400mn.

Proofreading: Based on estimates from the fly connectome, proofreading a whole mouse brain would require 0.3–1mn person years. At $20k per year per proofreader this translates into $6–20bn in total. Cloud hosting of the proofreading infrastructure would add up to an additional $1bn over the course of 10 years.

Additional challenges

The scale, time and cost estimates above bring significant challenges of management, funding and collaboration – outside of genomics these are more typical of the physical sciences than biology. Compared with genomics, there are specific challenges related to the nature of the samples. At least initially, mouse connectomes would derive from a single brain; it is not possible to 'clone' a specimen, so sharing it among multiple imaging locations would mean subdividing that single brain. Ensuring that data are released in a timely and efficient fashion, and making these data usable by different kinds of scientists, is obviously vital, and will require both prior planning and long-term support. For a technically intensive project, global equity of participation and access may be challenging. Finally, the large computing requirements mean that sustainability, including energy impacts, should be assessed.

3. General strategy: conclusions and recommendations

We carried out extensive expert interviews (**Appendix A**) as well as our own analysis of the challenges summarised in the previous sections. This leads to many individual conclusions and recommendations. The majority are closely linked to different steps of the connectomics pipeline and technical considerations; we detail these in the landscape and gap analysis that follows in sections ordered from sample preparation, to imaging, to connectome extraction and sharing. In each case, we summarise the current state of the art, as well as our specific conclusions and recommendations. However, some conclusions are more general in nature or represent organisational, political or strategic issues. This section will summarise these initial conclusions since they can be appreciated without a detailed understanding of the technology, and because they have important practical implications for how technology should be deployed.

A whole mouse connectome should be a goal on a 10–15 year timeline. However, it is crucial that this is part of a broader effort to develop a connectomics ecosystem.

There is almost universal agreement that a mouse connectome will be both useful and expensive. As with many things, the debate is principally at what point the utility exceeds the expense. If a mouse connectome could be delivered for a total cost of 20 million dollars over the next 5 years, we should start today. If the price tag were 20 billion dollars, enthusiasm would naturally be muted. The cost of the first mouse connectome will almost certainly be between these two numbers. How much it will cost – and when it will be delivered – will be a function of both technology and politics. If the utility appears to be high, then starting sooner and spending more will be advisable. We therefore recommend that in addition to the necessary technical development, a priority for connectomics research prior to starting a whole mouse connectome should be to generate evidence for the value of whole-brain connectomics.

Although a whole mouse brain connectome would have a major impact on basic and translational neuroscience, the utility will only be clearly demonstrable after the fact – another parallel with genomics (Lichtman and Sanes, 2008). Indeed, it is the personal experience of some of the authors (GJ, MC, PS, working in *Drosophila*) that even those closely involved with generating whole-brain connectomes have trouble predicting the full impact until the final product is in their hands. Related to this, perhaps surprisingly, several mouse connectomics experts said that the best guide to future impact of a whole-brain mouse connectome is the fly. In other words, if we want to think about how a *whole-brain* mouse connectome might impact mouse neuroscience, we should look at how *whole Drosophila* connectomes are impacting fly neuroscience. While this view makes sense to us, we would be surprised if all neuroscientists agreed. Bringing these threads together, the scientific and technological basis as well as the political case for a whole mouse connectome should be strengthened by:

1. Continuing to develop the scale, diversity and impact of whole-brain connectomics, starting with the current state of the art (i.e. *Drosophila*).
2. Generating connectomes in vertebrates that more sceptical neuroscientists could agree to be more representative of a mouse (and ultimately a human). Image volumes would clearly have to be smaller than a whole mouse brain, and should be selected to maximise scientific value rather than just being based on progressively larger imaging volumes.

Both whole small brains as well as small volumes of mouse brain should be used as stepping stones

A stepping-stone approach is a logical response to the huge difference in scale between connectomics today and where we need to be for a mouse brain. The largest finished connectome (the fly hemibrain) is about 50,000 times smaller by volume. The largest imaged volumes (of order 1 mm^3) are admittedly only 500 times smaller, but it is clear that existing automated segmentation methods produce a poor-quality outcome without proofreading. For example, in the MICrONS dataset fewer than 1% of neurons (n=601) have been proofread, but an average of 140 edits per neuron (evenly balanced between corrections of false merges to other cells and missed continuations) was required to recover good-quality morphology and connectivity; the unmyelinated axonal arbours of principal neurons were particularly challenging (MICrONS Consortium et al., 2021). With these current limitations, it is important to think carefully about how many steps we might need, and what brains/regions would make useful intermediate targets.

Based on our community consultation, there are multiple reasons why we feel that a mixed stepping-stone approach is important. First, some technical problems relevant to mouse connectomics can only be solved in the mouse, and therefore connectomes of mouse brain regions of increasing sizes will need to be acquired to secure the path to obtaining a full mouse brain connectome. Funding agencies such as NIH have already acknowledged this need (see below). However, other issues can only be solved by imaging complete circuits in whole brains. For example, no mammalian principal neuron has ever been reconstructed in its entirety by EM – so to understand how to study long-range circuits it might be beneficial to work in a smaller organism. Second, work in invertebrates suggests that there is a huge non-linear increase in the value of connectomes as they become more complete. In fact there are two distinct kinds of completeness to consider: one is the fraction of the brain that is present within the imaged volume; the other is the fraction of neurons within a volume that has been successfully extracted from the image data. Both must be pushed to high levels for maximum impact. However, it is worth noting that, for end users, it is much easier to interpret connectomes that have been densely segmented – i.e. in which all neurons present have been reconstructed. The *Drosophila* hemibrain (Scheffer et al., 2020) is an example in which dense segmentation of a very large, but still incomplete, part of a brain was initially more useful for end users than a sparse reconstruction of a full female brain (Dorkenwald et al., 2022b; Zheng et al., 2018).

Related to this there is currently much high-impact work in connectomics being carried out in species other than mice. There is a potential danger that too early or exclusive a focus on the road to a mouse connectome could leave this work undersupported, whereas until capacity increases it may actually provide the best scientific value for the available investment and the best justification for the eventual investment required to obtain a whole mouse connectome. Finally, although a mouse connectome is probably still a decade off, we should think carefully about what we will do when we have one. Data analysis will be a major challenge. It therefore makes sense to solve some of the intellectual problems that we can expect to find while thinking about related but smaller brains, as well as frontloading development of the required analysis infrastructure.

Today's cutting-edge technology might be able to deliver a mouse connectome, but the timescales and costs remain uncertain. In the short to medium term, targeted investment to develop key technologies and strategies across the connectomics pipeline are still necessary both to de-risk the project and to reduce the uncertainty in costs and timelines

There are many specific aspects of technology development that are considered in detail in Section 4, 'Landscape and gap analysis'. However, we can signal right away that specific technology development efforts are necessary to bridge the gap between mm- and cm-scale samples, where more is actually different. These include whole-brain staining and physical sectioning of a brain into sizes suitable for microscopy. At this point there are multiple strategies that might be able to deliver a mouse connectome. However, even if some strategies look more likely to succeed than others, there is no clear winner. So we recommend that a broad portfolio of candidate technologies should be supported in the short to medium term. Furthermore, it is not a given that a single technology will fit all applications. For example we propose that 'integrative connectomics' may depend on imaging subvolumes of mouse brain; this might be better suited to cheaper/more robust technology, especially with equity of access in mind. Therefore these are relevant factors for funders to consider alongside absolute performance.

Our cost estimates in **Figure 5** are based on data collated from interviews and the recent literature. It is still too early to produce time and cost calculations based on validated performance of methods that are demonstrated to scale to a whole mouse brain. In the interim, we can see that our estimates ($7.5–21.7 billion) are almost certainly too large for a consensus to emerge that a whole mouse connectome is worth starting now. Instead, we would currently recommend pushing technology development using smaller stepping-stone targets. Ultimately, we believe that technical improvements across the whole pipeline will dramatically reduce the proofreading costs, which form the bulk of that figure. It would be much better to go into a whole mouse brain project with clear evidence from a stepping-stone volume that proofreading costs of <$100mn could be expected.

Another way of addressing this issue is to consider what thresholds in time or money must be met to give the go ahead on a first whole mouse brain collection project. Technology development could then focus on achieving those threshold values in addition to pass/fail assessments of feasibility. Deciding when to commit to imaging and segmenting a whole mouse brain will be a critical transition. For example, an objective assessment could find that $300mn is a trivially justifiable expenditure on a mouse connectome because it would supply answers to many questions of network anatomy that could otherwise only be answered by expensive experimental work: e.g. using viral labelling and whole-brain imaging.

Besides cost, we also feel that close attention must be paid to timescales. While a 17-year data-acquisition, processing and analysis timescale (as shown in **Figure 5**) might be possible, there would be huge challenges with personnel, technology changes, etc. Trying to get a project from sample preparation to initial data release on a five year timescale would be more motivating and easier to manage, so long as the technical groundwork had already been laid.

It would therefore be prudent to try to reduce the imaging time to below two years rather than five, and to have as much as possible of the image processing happen in tandem with the imaging. As was the case for genomics, we expect technology to improve rapidly with appropriate investment. However, unlike for genomics, each connectome must derive from a single physical brain which cannot be duplicated. This has a secondary consequence that it will be difficult to switch imaging technology for a given specimen. A five year imaging run raises the probability that a disruptive change in technology will become available in the middle of a run, precipitating a severe sunk cost dilemma.

Establish a few large centres worldwide as knowledge hubs

There was strong agreement among our interviewees that the early stages of obtaining a whole mouse connectome dataset (sample preparation and imaging) would need to be centralised. Sample preparation must be optimised for, and validated by, the selected imaging systems in tandem with other expensive equipment such as lab-source X-ray μCT. Electron microscopes are large and sensitive instruments, and it is very likely that the machines used for a mouse connectome would either be the latest commercial instruments or heavily customised. This will require significant local expertise and infrastructure for development and maintenance.

There was also consensus that such centres would need strong end-to-end pipeline expertise in order to guarantee that data were fit for purpose. For example, raw images should be validated immediately after acquisition by downstream stages of the image-processing pipeline, such as registration and segmentation. Note that it is still possible, even likely, that key stages of the image-processing pipeline might be driven by experts at other locations in collaboration with local staff. Later still in the pipeline, annotation and analysis often feed back into quality control of segmentation; neurobiologists should be included throughout the process to ensure biological validity. If substantial amounts of human proofreading are required to correct errors in automated segmentation then some centralised quality control will be required.

We discussed the ideal number of imaging centres with expert interviewees. Responses mostly favoured a handful of such centres. Arguments in favour of a single centre include the requirements for highly consistent imaging and the difficulty of dividing a single brain for imaging across multiple centres (though this should be possible if it is prioritised). Arguments for multiple centres include the difficulty of assembling global talent in a single location, the utility of 'friendly competition' in large projects, avoiding groupthink within a single institution and support from an international panel of funders. These centres could also act as training hubs as well as infrastructure that would allow experimental labs to conduct correlative multimodal connectomics experiments in which e.g. functional imaging of neuronal activity is combined with connectomics targeted at a subregion of the brain. This has similarities to the idea of brain observatories proposed for large-scale functional characterisation of brain areas (Koch et al., 2022).

There was less consensus about whether it would be advisable to develop new institutions to house these centres. Some advised setting up dedicated centres (aka Brenner Institutes). Others advised that new institutions are difficult to create from scratch (both in terms of people and expertise) and that it might be better to co-locate with existing centres of expertise. Yet others pointed out that the infrastructure requirements to house EM instruments are significant, and unless machines get smaller or throughput goes up markedly there might be relatively few locations that could host 5–20 instruments without a building programme. It was pointed out that national facilities such as synchrotrons might have suitable infrastructure, even if biological laboratories might not.

Ensure that post-imaging steps of the pipeline can support globally distributed teams

Although imaging will likely be centralised, once the mouse brain has been transformed into digital form, later stages in the pipeline could be much more distributed. However the raw image data will be so large that they cannot simply be downloaded at will. Serious computation on the whole image dataset may need to happen on compute infrastructure co-located with the data. But continuing along the pipeline will result in massive data compression – the final connectivity graph should fit on one of today's laptops. There is still scope in this model for collaborative and distributed research across the pipeline, so long as it is prioritised by scientists and supported by funding agencies e.g. through access to shared computational resources (whether dedicated or commercial offerings in the cloud). This is particularly important when considering equitable global access to connectomics projects, which should include participation during the project, not just the ability to use data when completed.

Incentivise collaborative culture

It is essential to incentivise collaborative culture to speed up the technology development and validation phases. As discussed above it is quite likely that during scale-up multiple centres will need to collaborate. By analogy with genomics projects it has been suggested that, once in the scale-up phase, formalised collaborations should be in place so that 'everyone is a winner', even if e.g. one centre's sample ends up being selected for imaging and segmentation. We would recommend funding small programmes that actively support technology and data sharing across centres, likely with dedicated staff. Specific ideas could include help for packaging up protocols, documenting code, specific funding for inter-site visits, or a globe-trotting research specialist who keeps collaborating groups up to date. There may be lessons to learn from e.g. the International Brain Laboratory (Abbott et al., 2017) as well as large genomics projects (Birney, 2012).

Notwithstanding the point above that all centres should be part of the same collaborative team, it may also be possible to embrace the fact that it is hard to cut up and share a brain. For example, three centres could image different specimens, such as two female brains and one male brain. The race to finish the first brain could produce 'friendly competition'. However, an opposite sex brain would have major interest. Similarly a second same-sex brain could help establish the nature of inter vs intra individual variation. In this set up there would be incentives to move quickly, but there would still be a unique pay off for each individual team. Such a set up could also provide redundancy should any individual specimen fail. Although it would require creativity and good faith to come up with effective scenarios, it seems that it would be worth thinking through such issues from an early stage. Funders could likely help with both planning and implementation of such arrangements.

It may well be necessary to set up a framework for coordination e.g. to identify early where the critical steps in shared pipelines are. This will be an ongoing process to ensure that everything fits together. It will be linked to the chosen collaborative approaches: not only sharing knowledge but also working with a common goal in mind, which requires project management. Options could include an overarching scientific advisory board or perhaps coordination through funders themselves.

Personnel and funding issues

Connectomics research currently appears biased away from universities towards research institutes, often with core funding. This is unlikely to be an accident, but may reflect factors such as the strong technological development slant which favours long-term research programmes with corresponding patience for long-term pay off, the presence of support facilities such as electrical/mechanical workshops that have been phased out in many universities and a greater enthusiasm for team science. This mix also raises recruitment challenges. Hardware and software methods development depends on engineers and computer scientists who can often earn considerably more in the private sector. Making an attractive long-term career structure for such individuals may be a significant factor in the long-term rate of progress in connectomics. Related issues occur for other technologists such as those with specialisms in sample preparation.

This is typically not regarded as a particularly glamorous endeavour, but improved sample preparation could literally save hundreds of millions of dollars in downstream costs by permitting faster imaging, improved segmentation and reduced proofreading. Cooperation between public and private sector funders could help address these issues, while a flagship goal such as the mouse connectome could provide a useful motivator.

More generally, there will be many areas where funders should have opportunities for complementarity in funding actions. Currently the largest investments in high-throughput mammalian connectomics include recurrent funding from the Max Planck Society and the NIH BRAIN (Brain Research through Advancing Innovative Neurotechnologies) capital initiative programmes, including the NIH BRAIN CONNECTS call (https://grants.nih.gov/grants/guide/rfa-files/RFA-NS-22-048.html) which is committing about $30mn per year under three headings: 1) scaling up mouse connectomics; 2) technical demonstration and development projects; 3) lower resolution projectomics mapping of primates (including human). Although NIH certainly supports connectomics work through other calls, these amounts are probably insufficient to support the full range of plausible imaging technology options, meaning that some important contenders will probably get dropped; this is likely to delay the point at which viable options are identified. Interviewees consistently argued that more funding of technology development and demonstration would have a significant impact. While it is not exactly surprising that investigators in a field would welcome more funding, we do believe that investment now could have a significant impact on when large-scale mammalian connectomics will be possible.

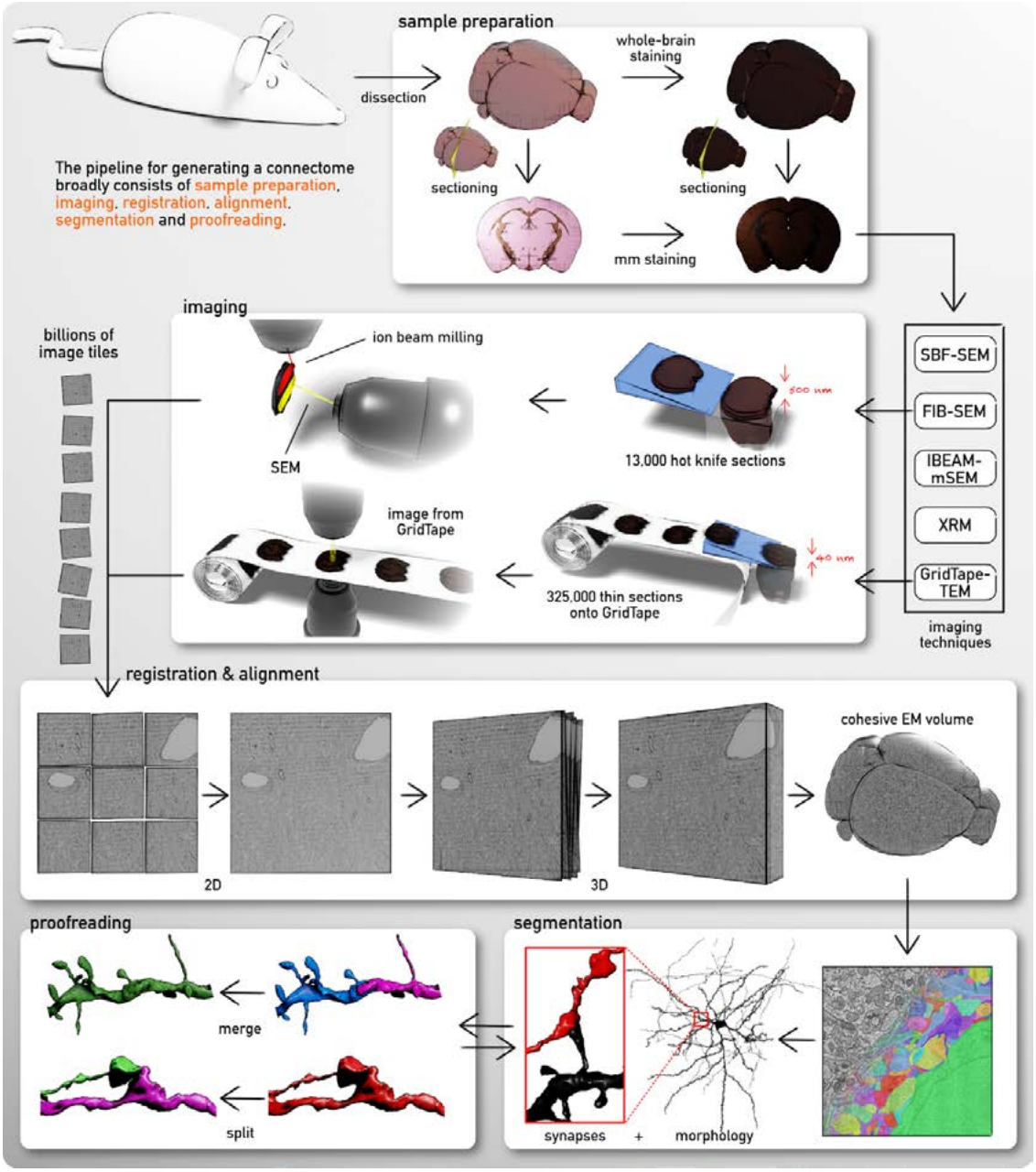
Beyond the monetary value, each funder will have constraints on where and what they can fund. For example, although NIH is open to funding research group leaders based outside the USA, empirically this seems to be rare and nearly always with US-based lead investigators. Other major funders such as Wellcome and Max Planck have international programmes, but still concentrate funds in their home countries. Besides geographic issues, it is clear that not all funders will have equal enthusiasm for all topics and funding mechanisms. This could be a concern, but also creates an opportunity for complementarity if funders can coordinate with a common goal in mind. For example, the NIH BRAIN CONNECTS programme is clearly focused on the mouse and on primates. Our consultation suggests this leaves an important gap in smaller vertebrates or invertebrates. Conversely, interviewees identified competitions in the area of image segmentation (see **Section 4.5**) as a potentially impactful funding intervention – these initiatives might be more compatible with private foundations than public funders.

Other areas where flexible funding could make a difference are in the recruitment and retention of technologists, as mentioned above. Indeed there may be areas of technology development where flexible funding might be important. Significant aspects of technology development in both hardware and software occur in the private sector, and there are numerous private/public interactions. These interactions are positive and important. However, our interviews also lead us to emphasise that in a project at the scale of a whole mouse brain, which may also have significant commercial implications, dependence on single private entities at any stage of the project should be avoided if at all possible.

4. Landscape and gap analysis

We now present a detailed landscape and gap analysis organised around the typical connectomics pipeline from sample preparation and imaging to data processing, analysis and connectome dissemination (**Figure 6**).

Figure 6: Pipeline overview
This figure illustrates the basic steps required for generating a mouse connectome. Importantly, the steps are highly interdependent: e.g. improvements in imaging will likely lead to higher quality alignment and better segmentation.



4.1 Sample preparation

Current state of the art

Optimal sample preparation is critical to the effort of mapping the mouse brain connectome, to produce high-quality raw image data and thus to ensure faithful digital reproduction of circuits. Significant progress has been made in the last ten years, enabling reliable fixation, staining and embedding of soft-tissue samples up to 1 mm thick in the shortest dimension. While available protocols share a similar high-level structure, details such as the choice of reagents, concentrations, incubation times and temperatures often vary. Next-generation protocols that have the potential to reliably prepare samples at the centimetre scale are now being proposed, a critical advance that would be necessary to map the connectome of a whole mouse brain. The following sections cover the essential steps in the sample preparation protocol.

Tissue extraction

The mouse brain must first be removed before subsequent processing steps can proceed. There are two options at this point – perfusion fixation of the mouse prior to brain removal, or brain removal followed by immersion fixation. In both cases, it is critical to consider changes to the brain ultrastructure occurring during anaesthesia, post-mortem and tissue handling.

It is important to plan for the datasets to be reused in multiple downstream studies, thereby contributing to a reduction in the number of animals required for the project, in line with the 3Rs in animal research (Replacement, Reduction, Refinement). Maximal value could be extracted from each animal by routinely retaining additional samples from each specimen, in the form of other tissues and organs fixed for high-resolution imaging, DNA extraction and genome sequencing, or (spatial) mRNA and protein expression profiling.

Fixation

The aim of fixation is to preserve the sample in as close to its native state as possible for further processing to a state compatible with electron microscopy. The affordable time lapse between a healthy living tissue and it being fixed without substantial loss in ultrastructure due to degradation is limited to a few minutes, especially for fat-rich tissues such as the brain. The state of the art in fixation of biological samples for EM is vitrification, a process in which a cryogen is applied to the sample at high speed and/or under pressure, causing the water to form an amorphous glass instead of ice crystals (which would expand and destroy the ultrastructure). However, the upper limit of sample thickness for vitrification is 200 µm (using high-pressure freezing), 35x thinner than the smallest dimension of the mouse brain. The state of the art for fixing mouse tissue specimens is therefore chemical fixation, usually by intracardiac perfusion rather than immersion in fixative, to minimise changes to the tissue during dissection and to enable fixation of arbitrarily large specimens. Perfusion fixation minimises the interlude between physiological state and fixed state down to minutes across all tissues in the organism, and since it employs the circulatory system to deliver the fixative, the diffusion distance will always be related to the density of blood vessels in the tissue/organism, regardless of its overall size. Variables for perfusion fixation include the composition of the fixative (usually a mix of (para)formaldehyde and glutaraldehyde in a buffer), the duration of fixation, perfusion pressure, and perfusion method (syringes,

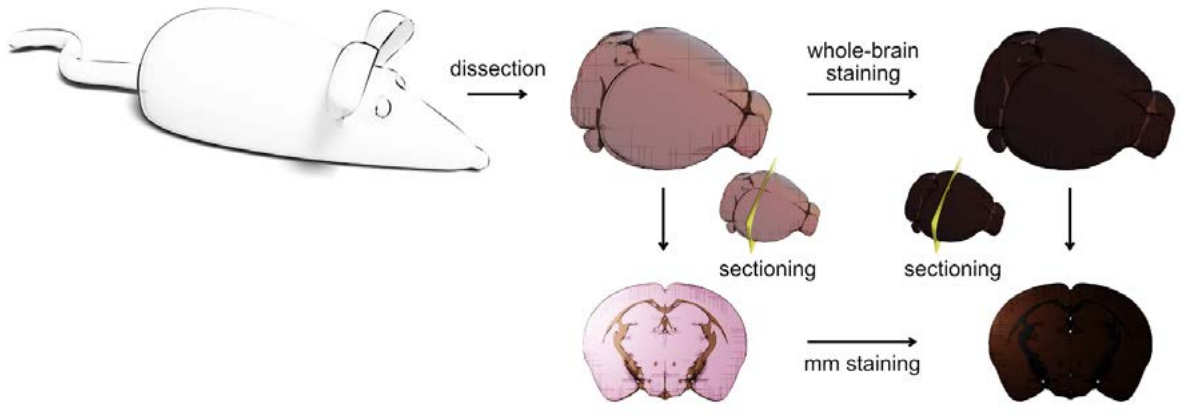
bottles at height, perfusion pumps, and the use of vessel cannulation tools). However, results are highly variable, with artefacts including: anoxia leading to cell death and changes in the morphology of mitochondria; loss of extracellular space and synaptic connections; osmotic effects; and gross disruption due to mismatched pressure. Moreover, perfusion is not an option for some specimens – such as extracted tissue (biopsies), post-mortem tissue or certain animal models (e.g. fish).

Staining

Soft biological tissues are mainly composed of light elements (O, C, H, N, Ca, P) that suffer from poor contrast in EM. Traditionally, heavy metals that bind to lipids in cell membranes have been used to increase contrast. Staining protocols have been optimised for vEM, and especially for serial block-face imaging in the SEM, where additional layers of heavy metals are required, both to provide sufficient contrast and to suppress charging artefacts (caused by electrons accumulating at the surface of the non-conductive resin-embedded sample during imaging with the electron beam, which disrupts imaging and slicing) (**Figure 7**). Combinations of osmium, uranium and lead salts are popular, but parameters including the choice of the bridging agent between successive osmium steps, reagent concentrations, buffers, timings and temperature are highly variable between protocols (Briggman et al., 2011; Genoud et al., 2018; Hua et al., 2015; Mikula and Denk, 2015; Pallotto et al., 2015; Tapia et al., 2012; Wanner et al., 2016; Wilke et al., 2013; Zhang et al., 2022). Homogeneous penetration of stains is also a major issue, leading to gradients of contrast through the tissue volume and associated challenges for automated segmentation algorithms further down the pipeline. Finally, diffusion rates depend quadratically on the diffusion distance (staining a 10x thicker tissue takes 100x longer) (Dempster, 1960; Hagstroem and Bahr, 1960; Ströh et al., 2022). Consequently, staining protocols designed for whole mouse brains that rely on the passive diffusion of chemicals suffer from incubations lasting for hours or days, leading to whole-protocol durations in the month-long scale (Mikula et al., 2012; Mikula and Denk, 2015; Song et al., 2022). Protocol optimisation is therefore a lengthy and arduous task.

Figure 7: Sample preparation for whole mouse brain connectomics

Two main approaches can be devised, consisting of staining the brain before or after slicing it into smaller pieces losslessly. Both approaches will require specific developments related to lossless slicing of the cm-scale specimen.



Dehydration

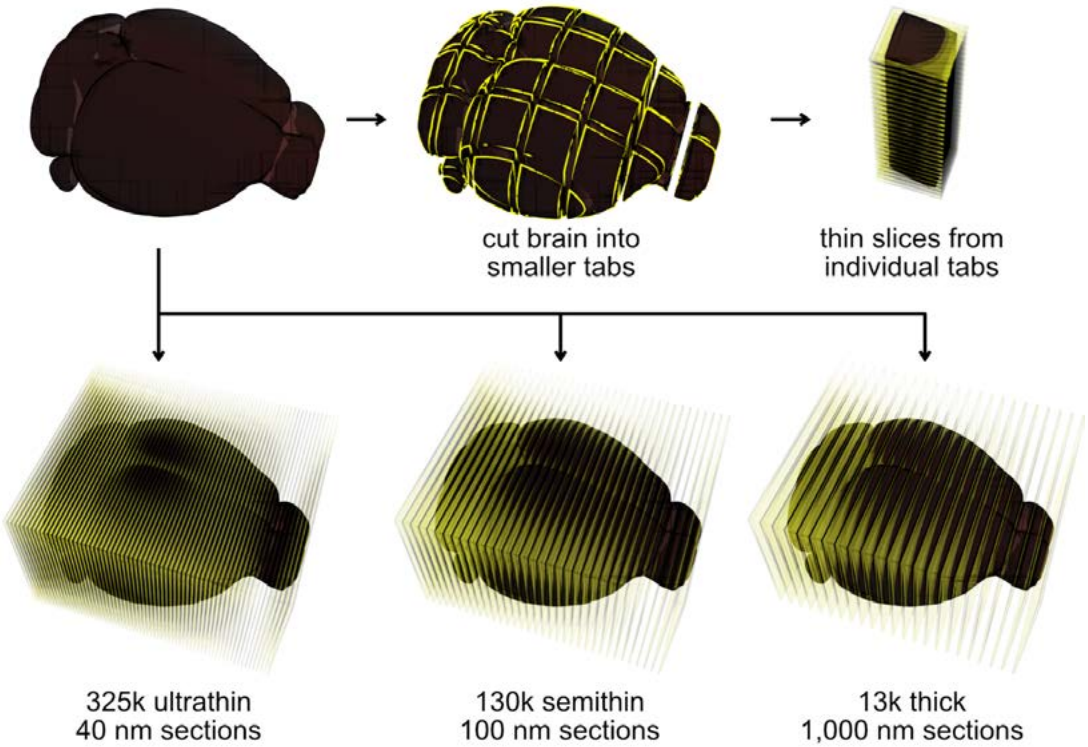
This step replaces the water in the sample with solvents such as ethanol or acetone, which are miscible with the liquid epoxy resins used later in the embedding step. Solvent, timings and dehydration gradients are highly variable between labs. Shrinkage, warping and extraction of soluble components are major artefacts induced by dehydration, which can be minimised by dehydrating at low temperature.

Resin embedding

To transition the sample into a solid state, the solvent is first replaced by a liquid resin, which is then polymerised using heat or UV light. The state of the art for connectomics is embedding in epoxy resins, which perform well for both knife sectioning and ion milling, and tolerate electron beam exposure during imaging. New 'conductive' resins have been reported that help to suppress charging and therefore improve image quality. Infiltration gradients, timings, resin formulations and polymerisation parameters vary between labs. As described for fixation and staining, resin embedding relies on passive diffusion of the (often highly viscous) non-polymerised resin components through the stained tissue, and therefore scaling this step from the mm to the cm scale presents challenges.

Figure 8: Sample sectioning

To be compatible with the forecast imaging technologies, the brain will have to be subdivided into a vast number of smaller pieces. Critically, this has to be done without losing too much tissue in the process. In particular, loss of entire sections could be detrimental for reconstructing the wiring diagram across the complete volume. The three thicknesses correspond to three distinct mechanical properties of the respective slices, and therefore specific slicing and manipulation constraints apply.



Trimming and sectioning

Trimming is the process of mechanically sculpting the specimen to a target geometry optimised for the subsequent step, which is often serial slicing into ultrathin sections (**Figure 8**). Trimming employs a diverse set of tools of increasing precision, from hand-held rotary tools and blades, to glass and diamond knives mounted on an ultramicrotome.

The majority of currently available ultrathin sectioning methods require the specimen to be <1 mm thick in one dimension. For a sample >1 mm thick in all dimensions, like the mouse brain, the sample must be divided into 1 mm-thick specimens without any loss of material. Each specimen must then be imaged and the image sets stitched into a continuous volume. Lossless cutting of thick (1–20 µm) serial sections has been demonstrated on stained and epoxy-embedded samples by locally raising the temperature of the resin (hot knife, Hayworth et al., 2015). However, scaling this technique from the current total thickness of 100 µm to the cm-scale will require changes to staining protocols, which currently render the tissue brittle due to the high metal content. Alternatively, lossless sectioning of the unstained, fixed tissue has been proposed, but has not been demonstrated to be viable to date (**Figure 7**).

Once a specimen is <1 mm in one dimension it will typically become accessible to most imaging modalities. For array tomography techniques that require collection of a series of ultrathin (30–50 nm) sections prior to imaging, automated systems have streamlined the process using, for example, collection of sections onto tape for SEM (ATUM, Schalek et al., 2011) and TEM (GridTape, Phelps et al., 2021) and magnetic-driven dense section collection onto wafers for SEM (MagC, Templier, 2019). For block-face imaging techniques, the ultrathin slicing takes place within the SEM chamber under vacuum, coupled with the imaging step through an automated cycle. Cutting is performed by an ion beam (FIB-SEM, Knott et al., 2011) or a diamond knife (SBEM, Denk and Horstmann, 2004), with both imaging techniques imposing specific requirements on the block-trimming step.

Finally, in studies aiming to target only a specific subvolume that is enclosed inside a larger specimen (such as a brain region inside a mouse brain), the trimming step becomes one essential non-reversible decision point that might permanently disrupt or preserve the precious specimen. These studies therefore benefit from correlative multimodal imaging approaches that enable informed decision making for trimming (Bosch et al., 2021). Those approaches rely on conserved landmarks between imaging modalities to overlay datasets of the same specimen obtained using different methods. The nature of the landmarks may vary in each type of experiment, requiring ad hoc development. Laboratory-based X-ray computed tomography (**CT**) and synchrotron hosted X-ray microscopy (**XRM**) are compatible with mm-scale stained and resin-embedded samples and provide invaluable resources for identifying experimentally relevant features and eventually targeting the trimming strategy for subsequent vEM (Bosch et al., 2022; Bushong et al., 2015; Dyer et al., 2017; Kuan et al., 2020; Meechan et al., 2022). These and other synchrotron X-ray microscopy modalities specifically designed for larger specimens (Walsh et al., 2021) might together enable non-destructive access to the relevant features on a stained whole mouse brain.

Hardware

Hardware is an integral part of sample preparation for EM and must be considered in the context of the mouse brain connectome. Sample processing units may be used to improve reproducibility and automated reagent exchange. Custom microwave units may be used to speed up processing of large samples and to improve penetration of reagents into the sample. Freeze substitution units may be used to minimise dehydration artefacts using a technique called 'progressive lowering of temperature' (PLT). Vibratomes and ultramicrotomes are used for tissue and resin sectioning respectively, alongside the semi-automated section retrieval systems mentioned above. Though basic versions of these systems are ubiquitous in life sciences EM facilities, none are optimised for samples above 1 mm^3.

The challenge of delivering sample preparation hardware for the centimetre scale is most apparent when considering sample slicing (**Figure 8**). For techniques that require ultrathin sections to be collected before imaging, the challenge involves slicing, storing and transferring 325k ultrathin (40 nm) sections. While this number could in theory decrease to 'only' 12k slices for techniques that can image 1,000 nm-thick sections, the burden of losing a single section becomes intolerable, since it would no longer be possible to connect neurites across the gap. In both cases, the slice counts provided are based on whole-brain coronal sections; the slice count would be much higher if cut from multiple chunks losslessly partitioned from a whole brain. For the process to be feasible, suitable hardware must be robust and resilient at all stages – from batch staining to automated trimming and automated slice collection.

Artefacts

Sample preparation for electron microscopy typically involves fixation, heavy-metal staining, dehydration and embedding in resin. Artefacts are introduced at every stage. Some artefacts have been well described, while others are poorly characterised. Some are associated with the sample preparation technique, and some vary according to tissue type. When considering artefacts, the critical question is whether they seriously affect our ability to use and interpret the resulting images with respect to the target biological question. Some of the main artefact classes affecting connectomics are: gross changes in ultrastructure due to anoxia, poor fixation or tissue handling; loss of extracellular space; changes in synapse morphology; loss of other cell types, including glia and astrocytes; stain precipitation or gradients of staining across the tissue; poor infiltration and polymerisation of resins; sectioning artefacts, including chatter, knife marks and compression. Taking into consideration the investment that will be put into the chosen brain, a suitable quality-control process will need to be developed to ensure the specimen will perform to set standards throughout the pipeline. For example, recent work has shown that monitoring the dynamics and patterns of stain penetration using X-ray CT can allow modelling of specific staining steps (Ströh et al., 2022) as well as design of quality controls (Mikula and Denk, 2015; Zhang et al., 2022).

Gap analysis

The critical gap in sample preparation for the mouse brain connectome is our ability to prepare a sample (up to the stage it can be ingested by the imaging technique) at the 500 mm^3 scale, 1,000-fold larger than current techniques allow. Doing so will require advances at every step of the sample preparation protocol, though it seems likely that these advances will be based on known chemistries and technologies, rather than being disruptive to the field. Community consensus on the image data parameters required to successfully segment and reconstruct the mouse brain connectome would inform the type and severity of sample preparation artefacts that can be tolerated, and thus the constraints on the design of new protocols. Early decisions should be taken on whether to engineer in the resolution to enable subsequent deeper mining of the data at the cytoskeleton and membrane bilayer level, which could be important for interpretation of the connectome and could help to reduce the number of animals required to output maximum information. To date, there has been a lack of dedicated funding for this type of work, and a lack of engagement from outside a limited set of labs and technical platforms, meaning that EM protocols have been relatively neglected by the commercial sector compared to, for example, sample preparation for genomics and proteomics. It may be useful to promote broader engagement by framing the problem as a new field of tissue engineering or architecture.

To address this gap, three approaches can be considered:

A first approach would be to slice the brain after perfusion fixation and before staining (**Figure 7**). Assuming a brain size of 7 x 10 x 13 mm (Allen Institute for Brain Science, 2015), the brain could be vibratome-cut into 13 x 1 mm-thick slices, each with a surface area of 7 x 10 mm^2, which could then be stained using current protocols. The main challenge of this approach is likely to be warping, compression and tissue loss at the cutting interface during vibratomy, leading to broken neural paths in the image segmentation and reconstruction step. A new method to losslessly section hydrated tissue would therefore be required.

A second approach would be to stain and resin embed the whole mouse brain and later losslessly divide it into smaller chunks that could be cut using current ultrathin sectioning techniques. This approach would first require development of new tools and techniques for staining and embedding the whole intact mouse brain. Though there has been progress made towards this goal, the field has yet to achieve reproducible, homogeneous, high-quality preservation, and whole-brain staining and embedding. New approaches are required to deliver fast and even penetration of reagents into the centre of the intact brain, which may require multidisciplinary collaborations across the fields of chemistry, physics and engineering. Finally, this approach will also require lossless slicing techniques to operate reliably at the cm-scale for heavy-metal-stained and resin-embedded tissues.

A third approach would be cryofixation of the specimen to embed it in vitreous ice. While this approach could be disruptive in terms of bypassing artefacts caused by chemical fixation, dehydration and resin embedding, it presents major challenges due to the scale of the specimen: vitreous ice only forms upon fast freezing, while crystalline ice, which disrupts tissue ultrastructure, originates otherwise. While the process of ice crystallisation can be slowed under high-pressure environments inside efficient heat-transferring media, heat-diffusion kinetics ultimately limit the thickest dimension to ~200 μm. Therefore, this approach would only be compatible with whole mouse brain samples if lossless hydrated sample slicing is developed. As in previous approaches, it would also require a coordinated effort harnessing expertise in chemistry, physics, engineering and biology.

In all cases, the field would benefit from:

- 'Ground truth' ultrastructure for different regions of the mouse brain, against which to measure the success of new sample preparation protocols.
- A method for assessing the quality and consistency of commercial reagents.
- Full parameterisation of the sample preparation space, followed by wet lab development using a screening pipeline that incorporates imaging as a readout.
- A better understanding of perfusion fixation to better engineer reproducible methods optimised for e.g. individual variations in blood pressure.
- Improved reagents with lower toxicity.
- Lower viscosity resins with improved conductivity.
- Optimised lossless thick sectioning of hydrated and/or resin-embedded tissue.
- Lossless ultrathin sectioning without artefacts (folds, tears, compression, knife marks, chatter).
- Benchmark samples for optimisation and comparison of downstream imaging modalities.
- Automation to improve quality and reproducibility, likely requiring expertise in robotics, micromanipulation and liquid handling. Though this is unlikely to be an area with high potential for commercial profit, automation could reap financial benefits in terms of minimising the time lost in processing poorly prepared samples through the pipeline and could also benefit pathology labs.

Additional considerations:

- Current protocols have been optimised for the parts of the brain that have been imaged to date, but might need specific adjustments to be suitable for other brain regions, other tissues beyond the brain, and other species.
- Other life science fields studying biological soft-tissue systems whose logic is defined by intercellular contacts within a mm^3–cm^3 neighbourhood are likely to benefit from developments made by the mouse brain connectome project. These fields may include the study of tumour microenvironments, immunological processes in lymph nodes, or the embryonic development of multicellular organisms, to name a few.
- Research technical professionals will play a critical role in the effort to optimise EM protocols, especially where the task is an unsuitable topic for a PhD or postdoctoral research.
- This should be an international effort.

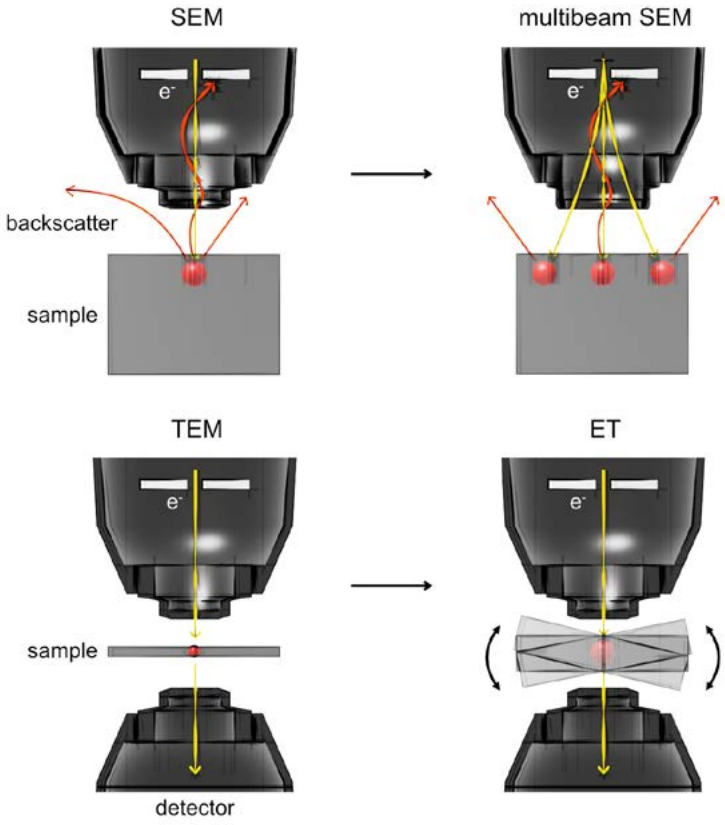
Recommendations

All these are [SML] unless noted:

- Promote efficient knowledge transfer in sample preparation by organising a workshop for research technology professionals involved in the development of protocols, to review the state of the art and especially to share experiences of what has been tested but found not to work, since much of this information is not in the public domain but could speed up the development process and lead to new innovations. [SM]
- Provide access to X-ray CT imaging for teams developing and using sample preparation protocols for large-scale connectomics for routine quality control of processed samples. [SM]
- Fund multidisciplinary collaborations in the sample preparation domain, including expertise in chemistry, to develop new staining reagents and protocols. [SM]
- Fund pipeline-wide optimisations that specifically consider internal feedback loops and controls within the connectomics pipeline, covering sample preparation protocol development, imaging, registration, segmentation and proofreading.

4.2 Imaging

Figure 9: EM techniques
Scanning electron microscopes (SEM) image only the top layer of the sample, which is why they need to be combined with a cutting or milling step (e.g. via a focused ion beam, FIB-SEM) to produce 3D image volumes. Modern SEMs utilise multiple electron beams to improve scanning times. Transmission electron microscopes (TEM) offer high resolution and fast imaging rates. However, electrons need to penetrate the sample which limits the thickness to <100 nm. Collection and mounting of hundreds of thousands of such ultrathin sections is a major challenge. Modern electron tomography (ET) microscopes try to overcome these limitations by imaging the sample from multiple angles which allows for thicker samples.



The fundamental requirements for an imaging modality to be able to deliver the first mouse brain connectome may be expressed as: **imaging** of 500 mm^3 of tissue with **synaptic and membrane resolution** (typically obtained in EM with voxel sizes ranging (8–40 nm)^3), in a **reasonable timeframe**, and with a **signal-to-noise ratio (SNR) good enough for automated image analysis** algorithms to segment features of interest (neurites and synaptic contacts). To understand the landscape and the potential of current tools, this expression bears further examination.

The distance between sequential images (or slice thickness, which defines the axial resolution) should be set as the largest distance over which continuity of neural tracing can be guaranteed. The main parameter affecting continuity will be the smallest diameter of the smallest feature of interest (neurites or synapses) within the mouse brain. A secondary consideration will be the size of any additional features required to build a structural wiring diagram – for example, protein scaffolds revealed as local electron densities and vesicles at synapses (Montero-Crespo et al., 2020; Santuy et al., 2018). Current estimates put the required imaging interval, or section thickness, somewhere between 8 and 40 nm (Briggman and Bock, 2012; Helmstaedter et al., 2008; Kornfeld and Denk, 2018).

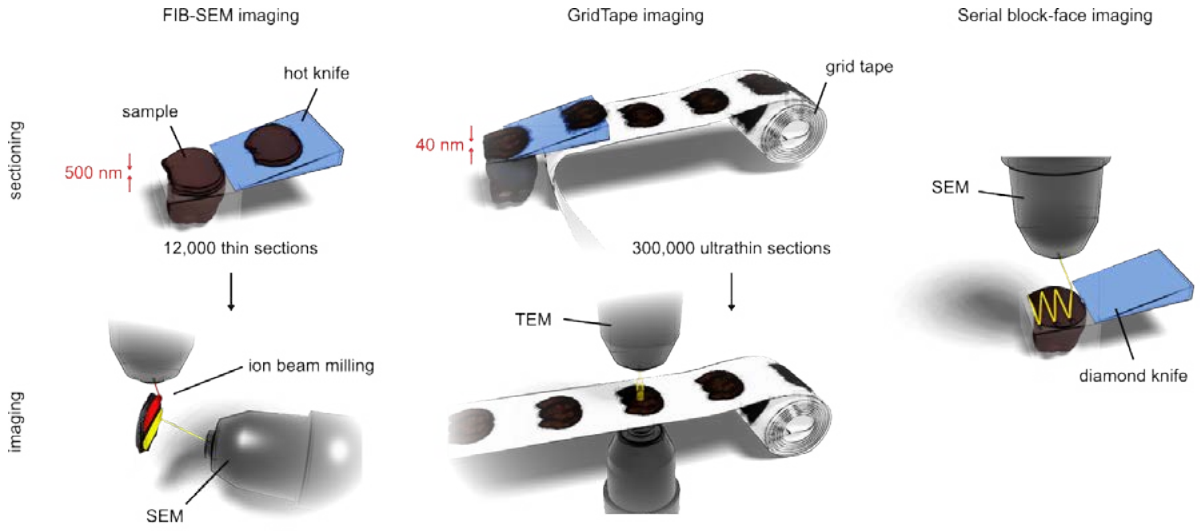
Lateral resolution of the image depends on a number of factors, including the sample preparation process, contrast generation mechanism, microscope optics, beam diameter in a scanning microscope, pixel size of the detector and the image reconstruction method (where applicable). Current estimates put the required lateral resolution for the mouse brain connectome between 8 and 40 nm, which is generally sufficient to resolve the plasma membrane of neurons and most subcellular organelles, though not usually cytoskeletal elements, ribosomes and macromolecular complexes. Ideally, the axial and lateral resolution will be the same, giving isotropic voxels, so that the resulting image data can be digitally resliced in any direction, providing associated benefits for image segmentation. However, other approaches have resolved connectomes at a satisfactory confidence with coarser axial resolutions, with slice thicknesses in the range of an ultrathin section: 30–50 nm (Berning et al., 2015; Gour et al., 2021; Loomba et al., 2022; MICrONS Consortium et al., 2021; Motta et al., 2019a; Shapson-Coe et al., 2021). Allowing this coarser resolution in the axial dimension enables the use of diamond knife sectioning approaches, broadening the range of available imaging technologies and thereby enhancing the feasibility of the overall project.

A ‘reasonable timeframe’ is a concept that requires discussion within the community. The first connectome acquired from the worm *C. elegans* took a decade to complete, and experts have speculated that the mouse brain connectome will also be a decadal effort (Abbott et al., 2020). Given that imaging is only one step in the pipeline, and that the data analysis phase is likely to be lengthy, one might speculate that a reasonable timeframe for image acquisition might be two to five years for the first connectome. One might also speculate that the knowledge gained from the effort to map the first mouse brain connectome will lead to a rapid reduction in the time taken to image subsequent connectomes, as was seen for the generation of genomes in the Human Genome Project.

An ‘SNR good enough for automated image analysis’ is a concept that will likely require further experimental effort to define, taking into account both the contrast profile of the specimen and the signal-to-noise ratios associated with different imaging modalities. The field of automated image analysis is evolving rapidly and, over time, segmentation algorithms produced by new machine-learning methods may be able to produce high-quality segmentations from lower quality image data.

Finally, community consensus should be sought on the 'features of interest' required to produce the mouse brain connectome, since these will inform: the sample preparation artefacts that can be tolerated; the resolution, contrast and SNR required to generate image data suitable for segmentation; and the design of quality-control metrics by which the success of image acquisition can be judged.

Figure 10: Three of the most commonly used imaging pipelines for vEM
FIB-SEM combines hot knife sectioning with focused ion beam milling during SEM imaging. For GridTape imaging, ultrathin sections are mounted on a tape that has a trough, which can then be efficiently imaged by a TEM. Serial block-face imaging consists of alternating SEM imaging and sectioning using a diamond knife.



Current state of the art

Several imaging modalities show potential to deliver the first mouse brain connectome (**Table 1**). Though none of these imaging modalities is currently capable of delivering an image volume of 500 mm^3 at 8 nm^3 voxels in a reasonable timeframe, it seems likely that the first wiring diagram will be delivered by an evolution of one of these systems, rather than an entirely new disruptive imaging modality.

Table 1: Current and prospective (yellow) vEM technologies

Imaging		Slicing: low-throughput, knife	Slicing: low-throughput, ion	Slicing: high-throughput, knife	Slicing: high-throughput, ion
low	TEM	ss-TEM		GridTape-TEM	
low	SEM	SB-SEM	FIB-SEM	ATUM-SEM	IBEAM-SEM
high-throughput	FASTEM	ss-FASTEM			
high-throughput	TEMCA	ss-TEMCA		GridTape-TEMCA	
high-throughput	mSEM	~~SB-mSEM~~	~~FIB-mSEM~~	ATUM-mSEM	IBEAM-mSEM

ultrathin section collection dependence
> 100 MHz
unknown imaging speed on experimental data

The group of imaging modalities showing the most promise for mapping the mouse brain connectome is volume EM (Peddie et al., 2022). vEM describes a set of electron-imaging modalities that combine a contrast mechanism that allows visualisation of all membranes within the sample, with the resolution to detect features of interest smaller than 10 nm, and the potential to image large volumes via a serial imaging regime. Different vEM techniques have different potentials for scaling to an image volume of 500 mm^3 with (8 nm)^^{3} voxels. vEM imaging technologies involve two main processes: sample slicing (providing the axial resolution) and image acquisition (providing the lateral resolution).

The slicing and imaging processes can be implemented independently using a method called array tomography. This requires sample slicing using a diamond knife and ultramicrotome, and section collection on grids, wafers or tape to give a library of ultrathin sections spanning the volume of interest. Imaging can then be performed in transmission mode in a TEM, or in scanning mode in an SEM (**Figure 9**). Array tomography allows for independent optimisation of slicing and imaging, and provides a means for permanent storage of the imaged substrate, but suffers from artefacts derived from the production of ultrathin sections such as folds, cracks, compression, debris and loss of slices.

Alternatively, slicing and imaging can be combined into one instrument, based in the SEM since the chamber is large enough to hold the slicing mechanism, which may be based on a diamond knife, focused ion beam or plasma milling (**Figure 10**). The block face is imaged and then a slice cut and discarded, a process that is repeated sequentially in an automated cycle. Block-face-imaging approaches bypass most sectioning artefacts, thereby simplifying subsequent image registration steps. However, they usually involve destruction of the sample, and therefore are prone to catastrophic errors if undetected malfunctions occur through the slicing-imaging cycle.

A wide variety of vEM approaches have been recently developed, each employing a specific combination of the aforementioned parameters (**Figure 10**). Higher-throughput variants of both slicing and imaging processes have been proposed that could push volume imaging rates well beyond 100 MHz (a rate equivalent to imaging (0.1 mm)^2 with 10 nm voxels per second).

Current vEM technologies employed in connectomics are:

- **Serial section transmission electron microscopy (ssTEM)**

This is the first vEM technique developed at scale to deliver the pioneering *C. elegans* connectome (White et al., 1986). Slicing requires manual collection of ultrathin (40–60 nm) serial sections on 3 mm diameter grids, followed by post-embedding staining and imaging using a transmission electron microscope. Though manual, slow and prone to errors, this method requires little specialist hardware beyond what is available in most electron microscopy facilities, and has few constraints on the type of sample that can be imaged.

- **Serial block-face scanning electron microscopy (SBF-SEM)**

This technique, prototyped in 1981 (Leighton, 1981) and fully developed for connectomics research in the early 2000s (Denk and Horstmann, 2004), provided unprecedented access to landscapes of biological tissue spanning hundreds of micrometres. It consists of an ultramicrotome inside the vacuum chamber of a scanning electron microscope, equipped with a backscattered electron detector. This technique relies on an oscillating diamond knife for slicing serial sections from the block face and sequentially imaging the newly exposed block face. It benefits from a low image distortion compared to array tomography methods. On the other hand, it is a destructive technique and is sensitive to charging artefacts (derived from electrons of the incident beam building up at the specimen surface and interfering with the imaging process), although developments have been proposed to mitigate this (Deerinck et al., 2018; Titze and Denk, 2013). The area of block face that can be cut and imaged is restricted to less than 1 mm^2 due to limits on the size and reach of the diamond knife.

- **Focused ion beam scanning electron microscopy (FIB-SEM)**

This technique relies on a microscope containing both a scanning electron source and a focused ion beam source. The ion source provides a nano-focused beam of gallium ions that mills the surface of the sample at single-digit nm increments, with the newly exposed surface being sequentially imaged with the scanning electron beam and a backscattered electron detector (Knott et al., 2008). This technique allows fields of view of up to several hundred microns wide to be imaged, but loss of ion beam energy and curtaining artefacts limit the depth of high-quality milling to around 10 µm, rising to 100 µm upon careful system optimisation. The number of slices in an uninterrupted milling run is limited by the stock of Ga^{2+} ions, though next-generation enhanced FIB-SEMs have delivered stable systems able to losslessly reset the milling plane after refreshing the ion beam (Xu et al., 2017). This technique is destructive.

Next-generation vEM technologies with potential to deliver connectomes at >100 MHz rate include:

- **GridTape section collection with TEM camera array (GridTape-TEMCA)**

This technique automates two steps of ssTEM (Phelps et al., 2021). Serial sectioning is automated by collection of sections onto a tape with preallocated, barcoded, electron transparent apertures. The tape is then loaded onto an upgraded TEM capable of automatically detecting and imaging the sections. Increasing the distance between the sample and the detector enables the use of large area detectors. When combined with transmission imaging mode, the large area detector provides high imaging speeds. The technique is non-destructive but depends on ultrathin section collection, and is therefore susceptible to sectioning artefacts and section loss, and z-resolution is limited to 40–50 nm.

- **Automated tape collection multibeam SEM (ATUM-mSEM)**

This technique incorporates improvements in both slicing and imaging. First, serial sectioning is automated and sections collected on a tape compatible with scanning electron imaging (Hayworth et al., 2014; Schalek et al., 2011). The sections are then imaged using a multibeam SEM, which contains a beam splitter enabling multithreaded simultaneous acquisition of images, currently employing up to 91 beamlets (Eberle et al., 2015). ATUM-mSEM thus benefits from high imaging speed. The technique is non-destructive but depends on ultrathin section collection, and is therefore susceptible to sectioning artefacts and section loss, and z-resolution is limited to 40–50 nm. mSEM also utilises secondary electron detectors (rather than backscattered electron detectors), which are more susceptible to charging artefacts.

- **Ion beam etching and milling with multibeam scanning electron microscopy (IBEAM-mSEM)**

This novel technique, only presented in conference reports at the time of preparing this landscaping project, benefits from the high imaging rates of mSEM and high isotropic resolution achieved with focused ion beams, while overcoming the field-of-view limitations of gallium FIB-SEM and of its first upgrade, gas cluster ion beam (Hayworth et al., 2020). IBEAM instead relies on argon beam milling, which in principle allows an arbitrarily large sample (into the centimetre scale) and unlimited continuity of the milling and imaging cycle. Large sample surfaces can be evenly milled through sample rotation under an inclined incident Ar beam, and the Ar gas can be provided from a circuit instead of an enclosed stock, removing the need to stop the scan for gas refilling, thus allowing continuous acquisition over longer periods. If successful, this technique could produce datasets with 10 nm^3 voxels through cm-scale volumes at acceptable imaging rates. On the other hand, it is a destructive technique, and mSEM utilises lower energy secondary electron detectors which are more susceptible to charging artefacts.

Imaging technologies other than electron microscopy

Beyond vEM, X-ray microscopy is also a strong contender for mapping the first mouse brain connectome. Though laboratory-based microcomputed tomography (**CT**) systems and synchrotron-hosted XRM beamlines have been used to image hard biological tissues like bone for many years, it is only recently that optimisation of sample preparation procedures and imaging systems has begun for soft-tissue specimens. The current state of the art leverages heavy-metal staining and resin-embedding protocols used for vEM to generate contrast in soft tissue for XRM. Since the contrast generation agents are the same as for vEM, the nano- and microanatomy landscapes provided by vEM and XRM are in essence equivalent, and images acquired at similar resolutions using the two different imaging modalities could eventually become indistinguishable. With some hard X-ray imaging modalities enabling sub-100 nm resolution, such as nanoholotomography (Kuan et al., 2020) and ptychography (Holler et al., 2017; Shahmoradian et al., 2017), subcellular compartments such as dendrites, nuclei, organelles and even synaptic contacts have been detected. Moreover, for specimens prepared using vEM protocols, hard X-ray imaging is a complementary technique that can add robustness to the connectomics pipeline, by providing non-destructive internal views of the subcellular structure of the specimen before it is sliced into smaller portions compatible with subsequent vEM steps. Considering future avenues for optimisation, it is possible that minimising heavy-metal staining might allow larger specimens to be imaged in transmission mode with hard X-rays while maintaining resolution, leveraging scattering rather than absorption contrast. Altogether, with improvements to synchrotron X-ray sources, stages, detectors and software continuing at pace, and with optimally developed staining methods, it is possible that XRM will match the resolution and contrast profiles of vEM, at which point it could become the imaging modality of choice for connectomics due to the increased speed of imaging per unit volume.

Light microscopy has undergone revolutionary advances in the last decade, both in terms of resolution and volume (Velicky et al., 2022). However, there are few techniques that have the potential to combine the field of view (at the centimetre scale) with the resolution (< 10nm) required for mouse brain connectomics. A technique of potential interest is the combination of tissue expansion with light sheet microscopy, though this has only reached volumes in the cubic micron regime with resolutions in the 100 nm range. Some critical improvements are needed for this technique to provide connectomes at the relevant scales, including preservation of sample integrity (the expansion and clearing processes can disrupt membranes and synapses within the sample), dense labelling of all relevant lipids and proteins required to produce a connectome (progress has been made in this area by tagging all protein components (M'Saad and Bewersdorf, 2020; M'Saad et al., 2022) and lipids (Karagiannis et al., 2019), and processing of samples with volumes of multiple mm^3, before expansion.

Finally, a few other emerging technologies are gathering growing interest for their potential connectomics applications. By mRNA barcoding, the identity of the neurons extending any neurites into a specific brain region could be obtained by leveraging advances in genetic engineering, viral targeting and genome sequencing technologies (Huang et al., 2020). Another imaging technique raising interest in the community is stimulated Raman scattering microscopy (Hu et al., 2019), which can map chemical identity across the specimen. While the fields of view at reach with this technique when imaging scattering specimens (such as mouse brain tissue) are currently below the size of a whole mouse brain, and the spatial resolution is coarser than the detail required for connectomics, future developments could boost the latter by means of e.g. super-resolution approaches, and lossless fresh specimen slicing could eventually address the issues with the former.

Gap analysis

Rigorous assessment of potential imaging modalities

To date, individual connectomics labs have tended to select and develop a single imaging modality to deliver connectomics datasets. For example, the worm connectome was mapped using ssTEM (White et al., 1986); the fly hemibrain was imaged using FIB-SEM (Scheffer et al., 2020); and the recent larval zebrafish connectome was mapped using SBF-SEM (Svara et al., 2022). This means that there is a lack of rigorous comparison of the performance of different imaging modalities using carefully controlled benchmark samples and metrics in the public domain. It is therefore challenging to make evidence-based quantitative predictions of the potential of different imaging modalities to deliver the mouse brain connectome in a reasonable timeframe. The required experiment to address this gap may be framed as: imaging of a limited volume from the same region of the same benchmark sample using different imaging modalities, to determine which modality gives the highest imaging speed for a given resolution and SNR. This project would require: expertise in connectomics, sample preparation, imaging and analysis; access to GridTape-TEMCA, ATUM-mSEM, IBEAM-mSEM and synchrotron XRM; and rigorous planning to ensure that appropriate parameters are selected and metadata captured at different sites. The cost of each solution should also be considered, for individual microscopes and for running multiple microscopes, in parallel where needed, to scale to 500 mm^3 of tissue.

Quality control and standardisation

There is a general lack of standards for quality control across the life sciences vEM and XRM domains, but these are essential to ensure that microscopes are properly aligned and operating at peak performance. The Quality Assessment and Reproducibility for Instruments and Images in Light Microscopy (QUAREP-LiMi; https://quarep.org/) consortium is addressing this challenge in light microscopy, with working groups focusing on illumination power and uniformity, detectors, aberrations, resolution, data provenance and quality control metadata. This effort could be used as a model to develop a similar consortium for vEM and XRM, which would support the mouse brain connectome and provide fundamental tools across this imaging domain.

Assuming the imaging system is operating correctly, measuring the quality of resultant image volumes is still a challenge. Currently the state of the art seems to be spot checks by experts, followed by a test segmentation and estimates of how much proofreading would be required to produce a connectome with desired quality metrics. These metrics are usually based on how many synapses are associated with each segmented neurite in the dataset, but the interpretation is not straightforward. Altogether this situation is distinct from genomics where we can have a useful summary of per base sequencing accuracy without having to assemble a genome, let alone find genes. The assessment proposed in the previous section could be used as the basis for generating a more rapid and standardised approach to assessing image quality.

Improvements to imaging hardware

In considering which imaging modality may be capable of delivering the mouse brain connectome, the community should discuss and propose potential hardware upgrades that could deliver step changes in speed, resolution, volume and SNR. For example, in the single particle cryo-EM field, direct electron detectors played a critical role in increasing imaging speed and sensitivity to deliver near-atomic resolution for structural biology. Improvements in SEM sources, optics and detectors have already kick-started the vEM field. It will be important to understand whether there are step-change improvements still to be made in vEM and XRM hardware, or whether the technology is nearing maturity. A small group of technologies described above currently provide imaging rates above the 100 MHz mark. However, there remain multiple trade offs affecting each technology. This suggests that the most productive scenario and the one worth nurturing over at least the next five years is an ecosystem of connectome-delivering technologies. Doing so should both deliver results in the short term as well as enable informed crosstalk between alternatives to select optimal solutions for whole-brain imaging.

Full automation of image acquisition

There is potential for further automation of the imaging process in most of the vEM and XRM techniques that show potential for mapping the mouse brain connectome. Advances could be made at all stages of the process, from robotic loading of samples to detection of the sample position within the microscope, alignment of the beams or knife to the sample, adjustment of imaging parameters, detection and resolution of issues during the run, and post-acquisition alignment, processing and storage of the image data. Open software initiatives applicable to multiple imaging systems could benefit from faster improvements by integrating feedback from a wider user base (Titze et al., 2018). Conversely, imaging systems will likely mature faster with more robust performance and lower cost as a function of their applicability beyond connectomics or even the life sciences. This may be a factor worth considering when evaluating the long-term prospects for a given method. For example, there may be synergistic development of the multibeam systems for applications in the semiconductor industry. Finally, for maximum speed, efficiency and stability, human intervention must be minimised or removed from the data-acquisition process, as has happened to a large extent in genomics.

Smart imaging to increase imaging speed

Taking inspiration from the correlative imaging concept of data reduction outside ROIs (regions of interest), it may be possible to increase the speed of imaging for connectomics by developing smart imaging workflows compatible with reconstruction of the wiring diagram. Unlike correlative imaging, it would not be possible to sacrifice large regions of the tissue since connectomics requires dense reconstruction throughout the brain volume. However, there might be subcellular regions that are not required for reconstruction of the wiring diagram, such as the interior of the cell nuclei. Such features would not be known *a priori* however, and would need to be distinct from those necessary to improve segmentation. For data reduction and increased speed, these regions would need to be automatically detected during imaging, perhaps in a low-resolution map of the area, to inform faster low-resolution imaging in these regions, bringing AI close to the edge during data acquisition. Given the uncertainties about what image detail might be needed for improvement of segmentation, this may be a strategy that could be designed post hoc, i.e. after a first connectome has been successfully obtained, to obtain new connectomes more rapidly. Note also that this approach is only likely to be relevant to beam scanning imaging technologies (including multibeam SEM) where the scan pattern can be modified, rather than TEM imaging which acquires a large field of view simultaneously. Finally, using low-resolution images to inform higher resolution imaging is a more general class of problem that will occur regularly. It might be particularly relevant e.g. to situations where it is decided to image large parts of a specimen at low resolution and follow specific anatomical regions or long-range projections at high resolution. This could be both challenging but impactful for destructive imaging techniques such as IBEAM-mSEM.

Recommendations

All these are [SML]:

- For whole mouse brain connectomics, the primary goal should be to obtain high-quality structural connectomes from single specimens without attempting to overlay functional or molecular data (see also **Section 6.1**).
- Enable a quantitative, evidence-based assessment of the rate at which key imaging modalities could map the first mouse brain connectome, by funding a consortium of expert labs to collect a limited volume of image data from a carefully controlled benchmark mouse brain sample with set parameters (e.g. resolution, SNR) using e.g. GridTape-TEMCA, ATUM-mSEM, IBEAM-mSEM and synchrotron XRM. The resulting datasets will also provide the first direct comparison of image data collected from different imaging systems on the same sample, and will provide a valuable resource for the image analysis community to develop, validate and compare the performance of segmentation algorithms on similar data from different imaging modalities.
- Promote the development of technologies currently capable of imaging at rates beyond 100 MHz: GridTape-TEMCA, ATUM-mSEM, IBEAM-mSEM.
- Develop quality assessment and reproducibility metrics, standards and protocols for imaging mouse brain tissue using vEM and synchrotron XRM. This exercise, while focused on the mouse brain, will provide the know-how for development of similar processes across the life science vEM and XRM sector.
- Promote initiatives aimed at improving the reliability of the imaging pipeline.
- Promote initiatives exploring and developing synchrotron hard X-ray techniques for connectomics.
- Stimulate the development of emerging technologies with a potential impact in connectomics, including stimulated Raman microscopy, expansion microscopy and neuronal barcoding.
- Organise a workshop for technology developers to discuss and propose imaging hardware upgrades and advances that could improve the speed and quality of image data generation for the mouse brain connectome.
- Develop a call for smart imaging strategies, to fund research and development into: increasing the rate of acquisition of images suitable for reconstruction of the wiring diagram of the mouse brain; full automation of the image acquisition process, thereby minimising or removing human intervention for long-term imaging runs; incorporating data-reduction techniques to mitigate the burden on data processing, storage and sharing; development and testing of combined or integrated imaging solutions to increase data-acquisition rates.

4.3 Data handling

As soon as raw digital image data starts coming off the microscope there is a data-handling problem. Although those working in connectomics have already encountered, and to a large extent solved, such issues at the cubic millimetre/petabyte scale, a mouse brain connectome would push into the exabyte scale. This is the largest range for biological data. There will undoubtedly be lessons to learn from other data-intensive scientific projects (Birney, 2012) and institutions such as national genomics centres (who currently handle data on this scale, albeit for a whole community rather than a single research project). Our expert interviews did result in a number of conclusions that we feel will be important for the connectomics community. Furthermore, issues such as the location of archival vs working image data storage and the process of versioning and release of data have strategic implications that funders likely need to consider.

Gap analysis

Data engineering

Data handling and engineering at this stage will be a major challenge in and of itself. These challenges are even more significant in a multicentre project and must be mapped out with detailed solutions before acquisition begins. Experts emphasised the significance not just of common file formats and shared APIs (application programming interfaces) for online data access, but also data modelling – i.e. taking a high-level view of the process and describing in detail the information that must be collected and preserved across the pipeline. This process also needs to consider who will use different categories of data and when and how they will be shared.

Communication bandwidth

The path from the microscope to the raw data store must operate at the speed of the microscope (which may exceed 3GB/s per instrument or 20GB/s per centre) to avoid a data pile-up. This is a significant challenge if data storage is remote to the collection site; this would be expected if multiple centres are working on the same specimen, or if remote collaborators are involved in steps such as image segmentation, which has usually been the case until now. In the UK, most universities and research organisations are connected to the JANET academic network. Only a handful of institutions have the highest link speed, operating at 25 GB/s (more commonly quoted as 200 gigabits/s). In other words, 20 GB/s of raw connectome data alone would almost saturate the existing network connection of a research university. JANET currently carries a total of 3 petabytes of traffic per day. This is approximately the amount of raw image data that would be generated per day if a mouse connectome project were targeting complete imaging of a brain in one year.

Image compression and archive storage

The raw image data will be by far the largest part of the project's data-storage requirements. Compression will be essential from the beginning to reduce the storage space occupied. Furthermore, given the considerations about upload bandwidth it may be necessary to pre-compress raw image data before it leaves the acquisition site. EM and XRM image data is rich in detail with very little blank space (e.g. in contrast to fluorescence image data or astronomy). Lossless compression (in which every byte can be decompressed to its original value) may achieve reductions of the order 3x,

but to achieve higher reductions, some data loss by removing noisy (and poorly compressible) pixels is essential. The state of the art is currently 15x lossy compression of connectomics data without measurable impact on automatic segmentation of neurites (Minnen et al., 2021). However, it is possible that future improvements in segmentation could have increased reliance on ultrastructural information inside the cell.

For a range of reasons, including the possibility that lossy compression could remove information needed for improved segmentation, it is likely that a losslessly compressed archival version will be required. This will be large and expensive. At present, working datasets for connectomics are mostly stored with commercial cloud providers. If we assume 250 petabytes after this compression, cloud storage costs today would be at least $1mn a month with cheaper commercial providers. Even after lossy compression, we would expect the dataset to be 50 petabytes. This is such a large volume that moving the data around becomes a major issue for those who need to do computation at whole-volume scale. For example, it will likely require extremely careful optimisation to allow image data to be uploaded to cloud storage at the rate that it is produced by multiple imaging systems. We can therefore imagine that it might take 250/50 = 1/5 of the imaging time to download the whole dataset, i.e. months to years. Therefore, data-storage issues also have significant implications for the location of compute infrastructure. This also means that there would be strong vendor lock-in after selecting a cloud provider. Time will erode these costs and increase transfer rates, but whether it will do so soon enough to obviate these concerns must be kept under review.

Other connectome data

The above discussion is focused on whole dataset issues. Fortunately, as we proceed further down the processing pipeline, data sizes become smaller – effectively data processing is a giant exercise in compression. At the final stage, the connectome graph – which records only the number of synapses between neurons – should come in at around 1 terabyte or approximately 1 millionth of the raw image data. Other objects, such as 3D skeletons or surface meshes defining neuronal structure, will be intermediate in size, but will typically only be requested in small numbers by end users. API endpoints allow users to search for and retrieve such data. The state of the art for dissemination (see **Section 4.7**) includes both web tools and programmatic access to such APIs, but data exploration and analysis is an important form of quality control, even while data acquisition is still in progress.

Preparing to share

Connectomics data is already being richly and effectively shared to end users. However, these mechanisms are still evolving, and in many cases guarantees of data longevity or timelines for data release are quite informal. Most connectome projects have involved some kind of embargo period – either by restricting data access to project collaborators or by placing restrictions on publication.

A project on the scale of the mouse connectome, especially if it receives significant public or charitable funding, will require a more formalised process that may be skewed towards early data sharing. Data-sharing policies could be adapted from the genome project: the Bermuda Principles set out rules for the rapid and public release of raw DNA sequence data;

the Fort Lauderdale Agreement set out expectations for data reuse and the role of funders in guaranteeing this. However, we note that raw image data may not be the equivalent of raw sequence data – it may only be objects such as the segmentation or extracted neuron morphologies and synapses that actually constitute usable biological information. The data-sharing plans for CERN (European Organization for Nuclear Research) experiments (opendata.cern.ch) provide another interesting case, carefully defining retention and sharing of different levels of data and allowing exclusive data access to the CERN team during a short embargo period.

Recommendations

Data policy [SML]

Prepare a detailed policy for data preservation, access and reuse before data collection begins. This must define the different outputs that are expected and the timeline and mechanisms for sharing, as well as the data-licensing rules. Funders should help define, as well as enforce, the data-sharing policy. This policy must be prepared early, since it will shape data-engineering challenges for the project.

Points to consider include:

- Think carefully about whether archival storage of a dataset of the value of a mouse connectome raw image volume should be on public or private infrastructure.
- A 'working copy' of the image data after lossy compression might be 5x smaller, and other outputs will be smaller again – but still the decision to use a single commercial partner to host a dataset must be carefully considered.
    - As datasets get smaller along the pipeline and the number of users goes up, commercial providers will likely have the edge as they normally offer optimised content delivery to any location on the web.
- If data is hosted at no or reduced cost by a commercial partner, appropriate guarantees about longevity and egress must be in place.

Data engineering [SML]

A mouse connectome will be a major data-engineering challenge. Early planning will be critical and should consider the following points:

- Data pipelines must be engineered with regular versioning, eventual data release and long-term retention in mind. Considering the data volumes in question, there will need to be dedicated staff in these roles.
- High-quality metadata must be retained throughout the pipeline.
- Data quality control is essential across the project. The most effective way to do this is likely to be running the downstream stages of the image-processing pipeline as soon as possible. There will likely be dedicated staff in these roles.
- We recommend that the APIs that will eventually be used for connectome analysis by end users should be put in place as early as possible for quality control.
- Given the long-term costs and strategic implementations, it would make sense for funders to support these discussions – for example, by helping steer a data-management working group.

4.4 Image alignment

Current state of the art

The output of EM imaging will be billions of small, partially overlapping two-dimensional image tiles – unless microscope fields of view change markedly, we can expect in the order of 10–100 billion tiles for a mouse brain. These tiles have to be aligned and stitched into a cohesive three-dimensional volume before neurons and synapses can be extracted from the image data. Depending on the exact imaging modality, this is typically a two-step process where individual tiles are first stitched into two-dimensional sections which are then aligned along the third dimension (**Figure 11**).

Figure 11: Registration and alignment

The imaging pipeline produces billions of small two-dimensional tiles which have to be registered and aligned into a cohesive volume.

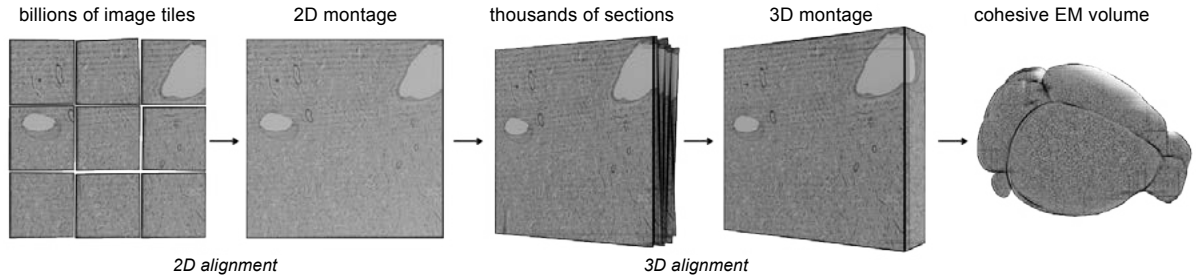


Image alignment (also termed image registration) is typically based on automatically finding corresponding features in neighbouring tiles/sections and turning these landmarks into a non-rigid transformation that aligns the images (Khairy et al., 2018; Saalfeld et al., 2012). Recently, neural network approaches have been used for fine-scale alignment at the end of the process, with significant improvements in section-to-section 3D registration quality (Popovych et al., 2022); this combined approach represents the current state of the art for datasets based on thin sections, which due to their fragile nature are often subject to significant physical deformation (e.g. tears and folds). In contrast, block-face strategies which repeatedly image and ablate the surface of a large specimen, typically start out with a better alignment between consecutive sections (because the imaging target was the surface of a large rigid block rather than a thin and deformable section).

Alignment quality is crucial for subsequent automated segmentation as demonstrated by the FAFB (full adult fly brain) image dataset (Zheng et al., 2018). Initially, reconstruction was purely manual: human tracers were quite good at coping with the many consecutive sections which were misaligned. Subsequently there have been three generations of automated segmentation, each of which had to realign the volume. The first used local reregistration only (PH Li et al., 2020); the second used a full global 3D reregistration to produce the volume used in the FlyWire project, with significantly more intact neurons (FAFB v14.1: Dorkenwald et al., 2022b); a further registration has been generated with improvements in both the initial 2D alignment and the full 3D registration (FAFB v15: Popovych et al., 2022). However, this highlights another problem: FAFB v15 has yet to be deployed, even though it has substantially better automatic segmentation than v14.1, because of the difficulty of mapping the large amount of proofreading from one assembly to the next.

So far alignment algorithms seem to be quite portable across datasets/species. The same registration approach used for FAFB v15 was applied to the cubic millimetre mouse MICrONS dataset (Popovych et al., 2022). Reflecting the significance of registration for subsequent image processing, both human and mouse cubic millimetre datasets included quality-control steps for registration directly into the imaging pipeline (Shapson-Coe et al., 2021; Yin et al., 2020).

Gap analysis

- Alignment of a whole mouse brain does not introduce any fundamentally new problems. However, there will be a scale difference of two to three orders of magnitude, and this may precipitate unforeseen issues. In the transition from 1 terabyte larval *Drosophila* brain to 100 terabyte adult brain, significant registration issues were encountered. Arguing against this concern, modern registration pipelines are already designed to work as distributed processes (e.g. on a CPU/GPU cluster) which in theory allows them to scale to very large image volumes. The pipeline used for initial alignment of the MICrONS cortical mm^3 (2 petabytes), for example, is thought to scale to exabyte datasets which would in effect cover a mouse brain (Mahalingam et al., 2022). In contrast, it is less clear if the fine-scale 3D registration approaches would have scaling issues.
- New imaging strategies could introduce new alignment problems. For example, any strategy that involves subdividing the brain into chunks will likely generate interfaces that are harder to align than consecutive serial ultrathin (<50 nm) sections. This was certainly true for the *Drosophila* hemibrain in which multiple 20 µm slabs had to be coaligned (Scheffer et al., 2020). The same concern may hold for imaging strategies that use semithin sections in the 0.5–2 µm range; even then those interfaces should be cleaner than those between 20 µm slabs, though they would be extremely numerous (order 5,000–20,000 sections).
- Most alignment pipelines in current use have been released in open source form, including basic documentation (e.g. through GitHub). However, similar to segmentation, these pipelines are very complex and often highly adapted for a given computational platform. This can make it challenging for third-party labs to redeploy them for their own dataset without additional technical support.

Recommendations

All these are [SML]:

- Host workshops on alignment targeted at developers as well as those who might be tasked with deploying or adapting registration pipelines at individual centres.
- Funding for (post-publication) technical support of code bases (bug fixing, deployment, etc).
- Investigate options for alignment-as-a-service either through partnership with specialist private entities or by supporting preconfigured cloud environments.

4.5 Segmentation

Current state of the art

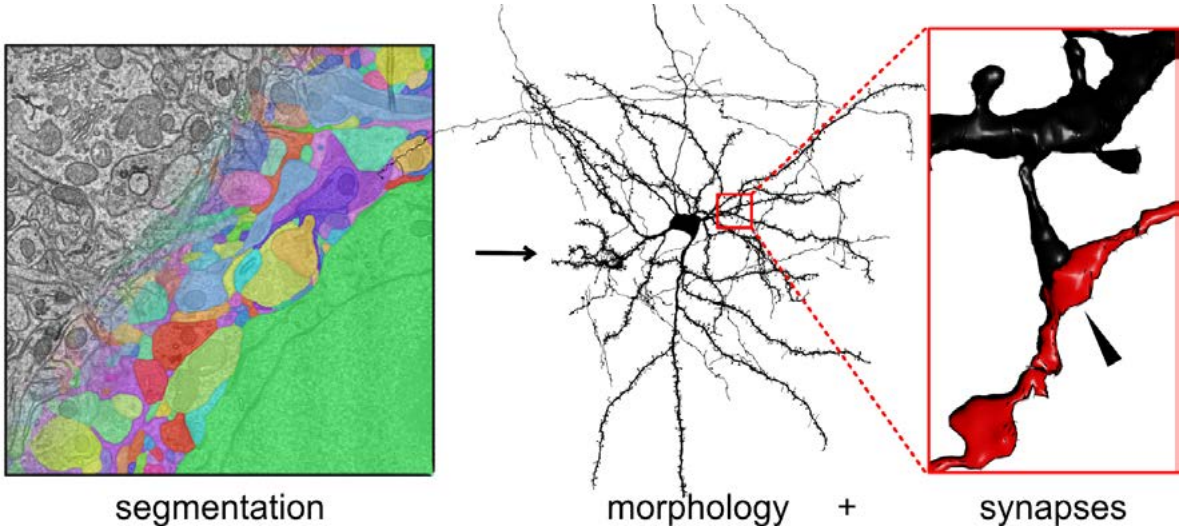
An EM image volume, however large, is not a connectome. Segmentation is necessary to extract neuronal morphology and synaptic contacts from the raw greyscale data (**Figure 12**). The switch from manual tracing of neuronal skeletons to automated segmentation has resulted in >50x speedups in connectome generation essential to the largest connectome to date (the *Drosophila* hemibrain). The end goal is to extract all neuronal morphologies and the location and strength of (chemical) synaptic connections between neurons. In most instances it is not possible to extract information about electrical synapses or other modes of communication.

Automated segmentation performance interacts crucially with both incoming image quality as well as downstream manual proofreading to correct segmentation errors. Ultimately the primary measure of segmentation quality is in the reduction of the amount of proofreading required. Automatic segmentation quality has improved substantially over the last few years. Some of our interviewees feared that it may be plateauing; others remained optimistic and commented that those who develop and deliver segmentation are naturally cautious about promising large gains.

Segmentation is very dependent on both the resolution and quality of the input data. Throughout this report, we have used 10 nm isotropic voxels as a standard size. However, 10 nm voxels may not reveal all the ultrastructural features in the specimen (e.g. small vesicles or lipid bilayers). Raw image quality (which depends on many factors including sample preparation, signal homogeneity, contrast, alignment, ground truth, etc) and other image-processing steps may be crucial, e.g. ensuring accurate alignment from one section to the next so that segmentation algorithms can successfully follow neuronal processes (Popovych et al., 2022). At the same time, the semantic information present only in large fields of view (e.g. whole-neuron morphology or location) might become a valuable source of information to obtain accurate automated segmentations of all cells and synapses in the dataset. Therefore, it is expected that the optimisation of the imaging and segmentation steps may require bidirectional feedback.

Figure 12: Segmentation

Segmentation of the volumetric image data extracts neuron morphology, synaptic connectivity and other ultrastructural features.



Numerous groups were working on 3D image segmentation as applied to EM connectomics ten years ago. There was a relatively sudden transition about five years ago when automated segmentation suddenly improved enough that it was worth only using humans to fix residual errors. One key driver for this was the flood-filling neural network algorithm developed at Google Research (Januszewski et al., 2018).

Underlying most segmentation algorithms is a distinction between the inside and the outside of a cell defined by the plasma membrane. Early machine-learning approaches applied neural networks to identify cell boundaries, but then used simple but computationally highly efficient methods to fill in the space between the membranes. Flood-filling networks improved on this by using an end-to-end neural network approach to simultaneously detect cell boundaries and fill the cell. This effectively increased the field of view over which the algorithm can collect features to make predictions and turned out much more effective than earlier algorithms, albeit at significantly increased compute cost. Other competitive approaches are now appearing (Macrina et al., 2021; Sheridan et al., 2021) but they have not yet been used extensively.

For the last few years, there has been a relatively clear distinction in the field between *automated* segmentation and *manual* proofreading. Experienced human proofreaders still retain an advantage in many situations, quickly spotting residual errors. This is likely because they can use additional domain-specific knowledge, such as the expected branching pattern of neurons (sometimes based on very precise knowledge from the same specimen or another specimen containing the same cell type). However, using this human domain knowledge directly will be too slow and expensive at mouse brain scale; therefore it is essential to try and capture as much as possible in automated strategies. One recent attempt to do this is RoboEM (Schmidt et al., 2022) which effectively flies an AI agent down long neurites to identify potential false splits or false merges. Other strategies include unsupervised learning methods that use morphological and ultrastructural properties of neurites to predict their cell types of origin (e.g. Dorkenwald et al., 2022a). This could be a good way e.g. to identify false merges between neurites which appear to belong to different cell types. These are just two examples of a trend which is starting to see the distinction between segmentation and proofreading blurring, as attempts are made to make proofreading much more automated and to incorporate more domain knowledge.

Gap analysis

- Improvements in segmentation to reduce the amount of human proofreading will likely be the key determinant in making a whole mouse connectome feasible. Specifically, improvements of at least 30x seem a requirement to bring the cost of proofreading into the same realm as the rest of the pipeline. However, even this would represent a huge logistical challenge (thousands of person-years of proofreading).
- The number of groups working on connectome segmentation is actually quite small in proportion to its significance for connectomics.
- Segmentation still does not leverage all the cues of experienced human proofreaders.
- There is limited work on segmentation of glia or other non-neuronal cells in the brain.
- Further work will also need to consider how neuroanatomical features, such as the presence of white matter, will impact segmentation quality. This would be one argument to develop methods on a subvolume of the mouse brain, like MICrONS.
- There is limited ground truth data for mammalian connectomics.
    - For example, about 1% (601 neurons) of the 1 mm^3 MICrONs dataset has been proofread.
    - Similarly just 100/~50,000 neurons in the H01 human temporal lobe dataset have been proofread (Shapson-Coe et al., 2021).
- Development has focused more on performance than efficiency. But efficiency may be critical for scale-up and is important for equitable access and environmental impact.
- Segmentation algorithms may be clearly described in publications or released as open source, but still be challenging to implement in practice e.g. because of dependencies on specific cloud compute environments and/or closed-source infrastructure.

Recommendations

All these are [SML]:

- Research on automated image segmentation for connectomics should be strengthened and diversified. This should be a win for everyone.
- Improvements could come from using:
  - Subcellular embeddings: here the high-resolution surface morphology of neurites, synaptic ultrastructure and intracellular features, such as the cytoskeleton or endoplasmic reticulum or ultrastructure, could help define long-range identity across neurites. There is promising preliminary work (Dorkenwald et al., 2022a).
  - Global cues, including cell typing or unsupervised machine-learning algorithms (for both segmentation and cell typing).
  - The history of edits within previously proofread connectomes, or ongoing edits in new ones.
- We recommend supporting the generation of ground truth data based on proofreading of existing mammalian datasets, and of datasets acquired with promising new imaging techniques.
  - Ground truth data should probably be a mix of locally dense and globally sparse to ensure that numerous full morphologies are available, as well as regions in which all profiles have been reconstructed.
  - Higher-order features, including annotation of cell types, may be an important part of this ground truth.
- Development of methods and testing will need to be flexible and easily deployable.

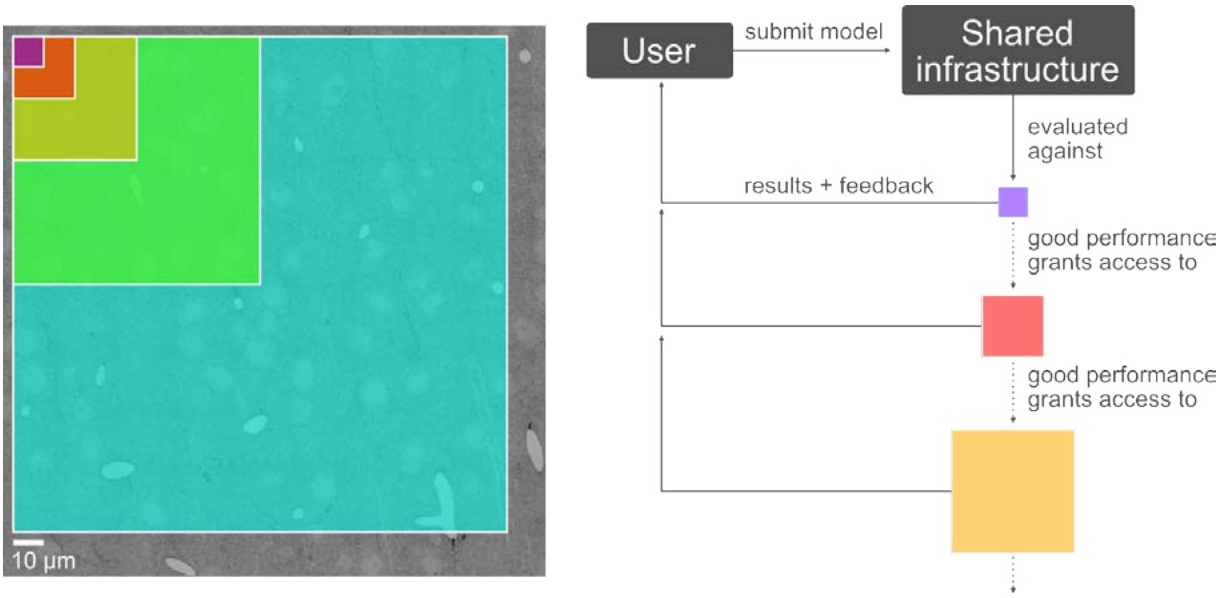
Developing segmentation competitions could be one effective (and efficient) way to generate more ideas and progress in segmentation (**Figure 13**).

- These should be planned and run by experts in the field who could engage with their community.
- The CREMI competition (Stephan Saalfeld, Janelia; https://cremi.org) should provide useful ideas, but used much smaller volumes (order 5 μm^3 cubes) than we would now recommend.
- Investment in storage and compute infrastructure to support these would be necessary to ensure competitions are in fact open to all.
- Since it would be better to make substantial progress before attempting to acquire a whole mouse volume, competitions could use existing fully proofread volumes for training or resegmenting existing datasets.
  - This could include fly datasets which have been extensively proofread and validated and are therefore ready to act as a testbed. This should be of general significance since most segmentation methods to date seem to be surprisingly portable; furthermore, the fly offers a lot of opportunities for functional and comparative connectomics so there will be significant scientific value in improved segmentations.
  - However, leveraging new mouse ground truth data as it becomes available would also be essential.

- Holding back some data might be necessary to run competitions.
- The competition could be organised in a pyramid style with low entry requirements (more computing available in each stage), moving to the next stage only if successful. These funding initiatives/competitions would require some management.
- The same computational platform could potentially be used by larger imaging centres for testing and deploying new algorithms and for smaller centres to process intermediate scale datasets (see also **Section 5**).

Figure 13: Segmentation challenges

A tiered system with increasingly large test data. Users submit their model to a shared computational infrastructure which runs it. Compute costs are funded as part of the competition. Good performance grants access to the next tier with larger test datasets.



4.6 Proofreading

Current state of the art

Despite major advances in automated segmentation, substantial human **proofreading** effort **(Figures 2** and **6**) is still required to produce a connectome of sufficient quality for brain-scale analysis. However, no connectome is perfect, so it is important to consider which errors can be tolerated. For example, correlations in the size of nearby synapses involving the same partner neurons have been used to identify potential sites of learning-related synaptic plasticity (Motta et al., 2019a). If the requirement for analysis is that synapses 10s of µm apart have a <5% chance of being incorrectly separated or joined, then proofreading requirements are much weaker and may already be met by automated approaches (Schmidt et al., 2022). In contrast, if the goal is to analyse complete connectomes at brain scale, then entire neurons should have a large (and largely unbiased) fraction of their synapses correctly attached and very few incorrect synapses misattached. This is a much stronger requirement since large missed branches, and especially false merges, must be eliminated over cms of neuronal cable, whereas typical error rates are an error every few mm of cable (Dorkenwald et al., 2022b). Finally, it is also worth considering that relevant insights might be obtainable from partial or noisy connectome reconstructions, provided that the questions posed to the analysis are framed by provable models/ hypotheses (Klinger et al., 2021).

In considering the acceptable error rate for a connectome, it is not just the frequency of errors but the nature of those errors that is important. For example, the *Drosophila* hemibrain connectome only identifies about 36% of the input connectivity of the reconstructed neurons; the remaining 64% of inputs are associated with very small pieces of dendrite that could not be easily attached to a neuron (Scheffer et al., 2020). The segmentation algorithm effectively terminated prematurely, resulting in a ‘false split’ between each fragment and the main part of the neuron. This difficulty of tracing fine dendrites is typical of *Drosophila* and probably many other insects (Schneider-Mizell et al., 2016). However, this loss appears rather uniform across different partners; furthermore, strong neuronal partners are connected by tens or hundreds of individual synapses in the adult fly, so that the resultant connectome graph appears to be missing only the weakest connections. In contrast, state-of-the-art imaging in mammals leaves fine unmyelinated axons requiring extensive proofreading (MICrONS Consortium et al., 2021). This is a potentially serious issue for the connectome as one uncorrected ‘false split’ error could remove nearly all of the output partners of a neuron.

All connectome projects use software tools that allow many users to work concurrently on an in-progress dataset. These proofreading teams can include different members, including specialised scientists at both postdoctoral or post-baccalaureate level, full-time contract workers without extensive training, postdoctoral fellows or postgraduate students carrying out proofreading as one element of their research in neuroscience, baccalaureate or high school students using it as a research experience, and finally, citizen scientists. Some of these individuals are effectively paid to proofread, others do it to advance their research goals, while some are volunteers without a specific financial or research motivation. Different projects have placed more or less reliance on different categories. The projects with the largest proofreading burden so far have been on *Drosophila*, so they provide some insight into the different strategies adopted. For example, the pioneering larval *Drosophila* project depended from the outset on volunteer scientists. They traced out specific neurons of interest (using the collaborative web application CATMAID (Saalfeld et al., 2009)) in exchange for insight into their circuits of interest. This arrangement worked well, and the first publication appeared in 2015, just three years after the project started (Ohyama et al., 2015), with many more following. However, it was noticeable that finishing the larval brain connectome required a more professionalised effort with a few individuals dedicated to completing all the remaining neurons in the brain that had yet to attract attention (Winding et al., 2022).

The adult *Drosophila* hemibrain connectome took a quite different approach. FlyEM, a large, dedicated team based at Janelia, collectively spent over 50 years proofreading an automated segmentation (Scheffer et al., 2020). This team included a small number of experienced neuro-anatomists and a large number of research assistants who were trained locally. This combination allowed the proofreading aspect of the project to be completed in just two calendar years. The FlyWire project (Dorkenwald et al., 2022b), which is expected to finish proofreading a full adult fly brain connectome in 2023, has adopted a more heterogeneous strategy. Both research scientists and citizen scientists have contributed their time and expertise. Nevertheless, the vast majority of proofreading has been carried out by professional proofreaders. Besides experienced neuroanatomists based in Princeton and Cambridge, this has also included contract proofreaders based in the Philippines and India, either contracted directly or through ariadne.ai, one of a small number of private enterprises in this space. This effort is therefore more similar to the hemibrain strategy, but with a distributed team. Unless and until the requirement for proofreading falls by at least 10,000x, it seems likely that proofreading a mouse brain connectome would rely on similar distributed teams.

Software tools are key to efficient proofreading (see **Appendix B**). NeuTu and Neu3 developed by the FlyEM team at Janelia have probably been used for more connectome edits than any other (Hubbard et al., 2020; Zhao et al., 2018). However, these tools were never designed for use by globally distributed teams as is now typical. Furthermore, Janelia's change in scientific leadership, together with an accompanying shift in research direction away from large-scale connectomics, means that these tools are now essentially in maintenance mode. Neuroglancer is an efficient web application specialised for data visualisation and exploration, developed by Google Research. It has been customised by the Seung lab at Princeton to enable efficient distributed proofreading. This has worked effectively for a wide range of end users working on the FlyWire dataset. However, it has no direct support for coordinated proofreading – for example, for a queue of centralised tasks containing locations of possible errors to review. This is an area where efficiencies seem very possible (webKnossos: Boergens et al., 2017; e.g. NeuVue: Xenes et al., 2022). Nevertheless, further development work will be required to combine the most effective aspects of these different tools into a single modern application. An improved user interface would also need to be combined with analysis methods that can identify locations needing attention.

Although proofreading is currently essential for good-quality connectomes, some of our interviewees questioned whether this would ever be practical for a mouse connectome. In this view, the feasibility of a mouse connectome depends not on reducing the amount of proofreading but on completely eliminating it. While there is promising research into (semi-) automated quality control (e.g. Dorkenwald et al., 2022a; Schmidt et al., 2022), it will require intensive research to reach a ~zero-proofreading state. It may be noteworthy that authors for both the MICrONS and H01 datasets commented that the principal segmentation errors that still needed fixing were associated with image artefacts (Popovych et al., 2022; Shapson-Coe et al., 2021). Reaching zero proofreading will probably depend on very high-quality input data and/or training machine-learning models about the causes and consequences of different image artefacts.

Gap analysis

- Proofreading is the bottleneck and the most resource-intensive stage, so it must be reduced as much as possible. This is likely to require significant optimisation of all the steps leading to image segmentation.
- It will be necessary to develop improved error metrics and decide what errors are an acceptable trade off between the amount of human effort and the resultant cost.
- Error metrics should be informed by a quantitative understanding of connectome variability, for which limited data are currently available. This need is particularly critical in mammalian connectomes.
- With today's technology, proofreading is still needed to correct errors in the segmentation. It is not clear when a goal of zero proofreading could be achieved.
- Distributed proofreading teams will increase capacity, as well as provide opportunities for the research community to contribute to the effort.
- As the quality of automated segmentation goes up, it is likely that proofreading tasks will become more complex.
- Contracted (non-scientist) proofreaders will continue to need in-depth training and supervision to be effective.

- Citizen science (CS) strategies could contribute to a long tail of proofreading, but they require very active management. However, CS projects can increase public understanding of science and may fit with Wellcome's goals. Nevertheless, developing CS projects requires specific resources (software, management, professional proofreaders for support/training), and might need different proofreading tools.

Recommendations

All these are [SML] unless noted:

- There is definitely scope to improve the efficiency of current proofreading software by funding further development. Even gains of 2x (which seem easily achievable) would have a major impact, whether for proofreading a whole mouse connectome or developing improved ground truth. [S]
- Ensure that tools can support distributed instead of centralised proofreading, including in low-resource settings. [S]
- Further investment could help with identification of problem areas and developing solutions for a seamless user interface appropriate for correcting different kinds of errors.
- Infrastructure supporting proofreading needs to be able to cope with large-scale, real-time changes and queries, and have robust links to annotation.
- There was consensus that, ideally, any lab should be able to participate in proofreading – i.e. a distributed approach (see also **Sections 3** and **5**) – but this would require discussion and investment in how consistency in quality can be maintained, how to ensure effective communication and how training would be given. A community management strategy and team would need to be resourced for some centralised quality control to ensure standards, consistency and continued progress.

4.7 Dissemination, annotation and analysis

Current state of the art

In **Section 4.3** we already discussed that planning for data sharing must start at the beginning of a project and be integrated with the whole data-processing pipeline.

Connectomics data is heterogeneous, complex and large, which makes dissemination non-trivial. It includes image volumes, neuron morphologies (as skeletons or surface meshes), segmentation of neurites, synapses and other ultrastructural features, connectivity data (adjacency matrices or edge lists). In the past, small, sparsely annotated datasets have often been shared as supplemental files available for download directly through the publisher's website (Bates et al., 2020; Eichler et al., 2017). With larger datasets, however, this quickly becomes impractical. Instead, the data are typically made available on the web. This can be done statically by providing files for download or interactively via browser-based web applications. Static download is most suitable for smaller files such as summary spreadsheets or datasets that must be downloaded in their entirety to be useful, e.g. a dump to be ingested into a database system to be run locally. Suitable locations include standard repositories such as Zenodo, Figshare and Dryad. However, these general purpose data-sharing options do not allow interactive data exploration and may not be able to handle very large data such as the raw images.

Given the heterogeneous nature of connectomics data and the fact that most users are interested in exploring only parts of the dataset, specialised applications that allow selective, interactive, multiresolution data viewing, in the style of Google Maps, have become key. Although some software (e.g. Knossos) is available as a downloaded application, the three most widely used for connectomics data are CATMAID, Neuroglancer and webKnossos (see **Appendix B**), all web browser-based applications. These need only a modest computer, a browser and an internet connection to run. CATMAID was initially developed for tracing and analysis of the *Drosophila* larval connectome (Schneider-Mizell et al., 2016); it is a rich and powerful tool for viewing and analysing connectome data. However, it started life as a skeleton-based tracing tool and it has rather limited support for proofreading based on volumetric autosegmentation which is now the state of the art in connectomics. In contrast, Neuroglancer and its variants are outstanding in displaying mesh, 3D segmentation and chunked image data, with effective level-of-detail optimisations. However, Neuroglancer has limited search ability, almost no built-in annotation options and no analysis functionality. There have been some initial attempts to use Neuroglancer as a viewer, embedded within a web application that allows annotations, but these have not reached maturity. webKnossos (initially developed at the Max Planck Institute for Brain Research and now developed by Scalable Minds) provides dataset storage and organisation that enables rapid control of the dataset library, allows for multiuser dataset exploration and annotation, and contains an embedded task-management system offering a direct management dashboard for distributed tracing tasks.

One example of state-of-the-art dissemination is the datasets produced as part of the MICrONs project (e.g. MICrONS Consortium et al., 2021). Here, a dedicated website allows users to explore and download (parts of) the data (**Table 2**). Importantly, new versions are released as proofreading efforts continue. Expert access is provided via the CAVE (Connectome Annotation Versioning Engine) service and a Python client (see **Appendix B**). The *Drosophila* hemibrain dataset was also released with an interactive web platform, neuPrint (**Table 2**). This allows efficient exploration of the data, querying neurons by name or connectivity. More sophisticated analysis is available via command-line tools.

The Janelia FlyEM team and collaborators use neuPrint for connectome analysis while proofreading is in progress. This provides many opportunities to identify problems by irregularities in connectivity. Similarly, the FlyWire and MICrONS teams use their CAVEclient for this purpose. This means that ongoing analysis and curation use the same system and will support analysis by end users after data release. This ensures the robustness of the system and also allows for more flexibility about when data can be released. There is a virtuous circle between proofreading, annotation and analysis. Ideally this can be exploited both before and after data release.

In addition to dataset- or project-specific sites, there are a number of repositories that either curate lists of available datasets or host the data themselves (**Table 2**).

Table 2: Some of the websites disseminating connectomics data

Name	URL	Description
BossDB	https://bossdb.org/	Open-data repository for various connectomics datasets.
NEURODATA	https://neurodata.io/	Repository collecting various connectomics datasets and tools.
NeuroMorpho	https://neuromorpho.org	Repository for neuron morphologies across species and modalities.
neuPrint	https://neuprint.janelia.org	Data-exploration website for various *Drosophila* connectomes.
mapzebrain	https://mapzebrain.org/home	Zebrafish data, including an imaged and segmented larval brain.
MICrONS	https://www.microns-explorer.org/	Project page for the MICrONS datasets.
Virtual Fly Brain	https://catmaid.virtualflybrain.org/	Hosts interactive CATMAID browser for larval and adult datasets. Virtual Fly Brain also integrates connectomic data and the published literature.
webKnossos	https://webknossos.org	Web-based exploration/annotation/ proofreading tool for connectomics data. Also hosts various published datasets.

Gap analysis

- Interviewees agreed that data should be released to the community, sooner rather than later, but detailed advice varied: some argued it should be from acquisition, others that later stages were sufficient.
- In some contrast with genome sequencing, for connectomics there is a significant delay between initial data acquisition and outputs that are useful for end users. In detail:
  - EM greyscale data, even after alignment, is of relatively little use to most end users. It is only after segmentation that the value increases. Note however that for the development of segmentation methods, release of the aligned EM greyscale data is essential.
  - The next stage might be when neurons (or neuron fragments) can be queried by connectivity to a starting object, and eventually when these queries can either search for or return annotations about cell types.
  - And beyond that, an annotated connectome is the most useful one to release for exploration and use of the data by neuroscientists, computer scientists and theorists.

- The utility of a connectome to the wider community depends strongly on the amount of annotation available when released. For example, it is difficult for humans to explore a dataset in which neurons are identified solely by numeric identifiers. Annotations such as cell type, other anatomical features or broader groupings of neurons are all useful to end users.
- There is no general purpose web tool for viewing and detailed analysis of today's connectome data.
    - CATMAID has strong exploration, annotation and analysis capabilities, but is specialised for skeleton data rather than the volumetric data that represents the current state of the art.
    - Neuroglancer has a very high-performance display of neuronal meshes, but limited data-exploration capabilities and essentially no capacity for annotation or analysis.
    - Developing efficient front ends for connectomics web applications appears to be a limiting factor at the moment, perhaps because the relevant software engineers have more lucrative options in the private sector.
- Proofreading, annotation and analysis feed off each other, and any infrastructure will need to take this into account (local or centralised).

Recommendations

All these are [SML] unless noted:

- Dissemination should use a centralised release infrastructure that is linked to proofreading and annotation (whether local or distributed).
- Sequential versioned releases were strongly recommended by interviewees (compare assemblies of the Human Genome Project).
- Semi-automated annotation of neurons could be used to improve segmentation and proofreading performance (Motta et al., 2019b). Funding development and validation of tools to create such annotations is likely to have a good return on investment. Possible starting points include Dorkenwald et al. (2022a) and Schubert et al. (2019).
- Labs that are specialists in a particular brain area may be best placed to identify and annotate cell types, but this will require a distributed approach to annotation, which is complex to coordinate and for which there is currently no good software solution. Attribution would be particularly important for a distributed strategy, including low-resource labs.
- If a centralised approach to annotation is taken, neuroanatomy specialists will be needed within the connectome teams. They could still work in collaboration with external labs.
- Linking the connectome to other data in the literature and beyond is vital. Possible models include the Ensembl genome portal (https://www.ensembl.org/) and the Virtual Fly Brain project (http://www.virtualflybrain.org/). [ML]
    - Ongoing support for appropriate online data-integration resources targeted at mammalian connectomics would be essential.

5. Specific issues for low- and middle-income countries

Connectomics data acquisition is resource-intensive, but this does not mean that connectomics cannot be inclusive. In particular, a *finished* connectome (revealing which neurons are connected without any further work) can level the playing field by removing the upfront cost of huge amounts of experimental circuit-mapping work. Furthermore, as noted earlier in this report, data in digital form should be accessible from around the world by anyone with a computer and an idea. Later stages in the pipeline, such as proofreading (if necessary), annotation and analysis, should be globally accessible. Earlier computational steps can also be made accessible if computer resources are provided to eligible researchers. We recommend that the needs and opportunities of researchers from low- and middle-income countries (**LMICs**) should be considered explicitly as we scale up connectomics, including in pipeline design, global hubs, collaboration culture and training.

LMICs face a wide range of challenges to equitable participation in research. Using the World Bank definitions (https://data.worldbank.org/country/XO) there are 136 such countries, ranging from countries with prestigious scientific centres of excellence (e.g. China, India, Argentina) to the poorest countries in the world. So one-size-fits-all recommendations are not possible, and our only further firm recommendation in this area is that it is vital to begin by consulting with scientists in LMIC countries before taking action.

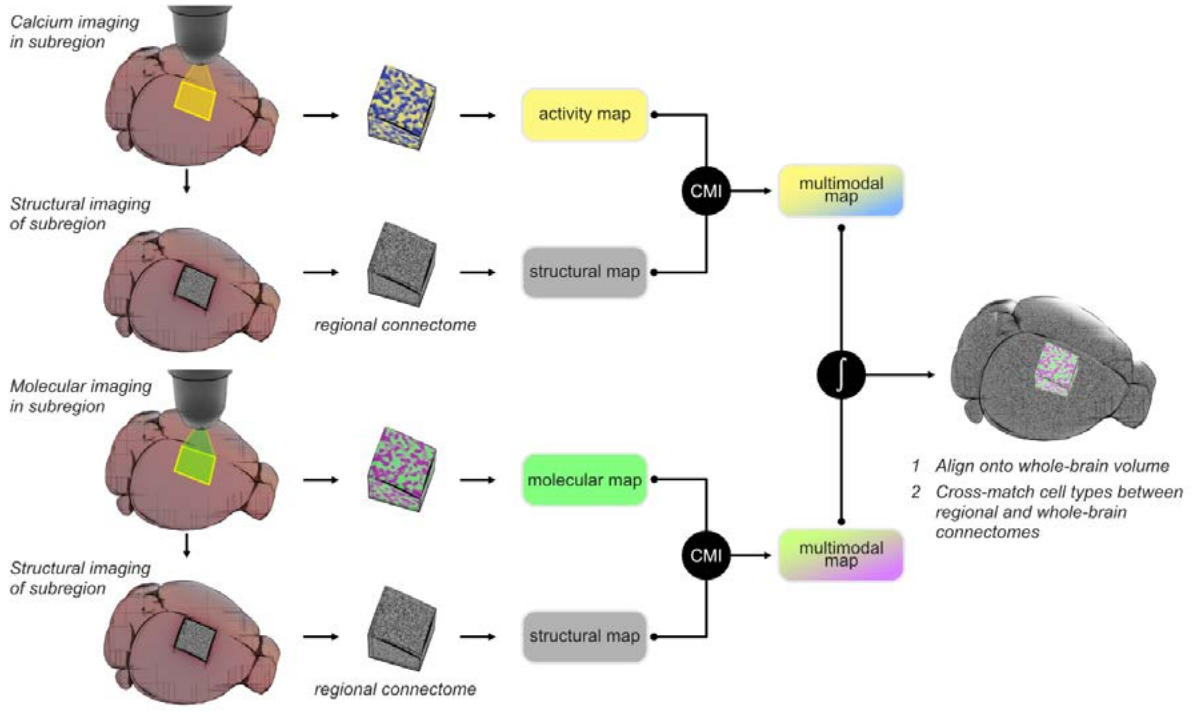
With this proviso clearly stated, we offer the following ideas, based on expert interviews, as a catalyst for discussion:

- Identify globally distributed 'connectomics champions' who can be engaged to advise on what their colleagues want.
- Lower the barrier for global contributions to automatic segmentation and manual proofreading of connectomes (see also **Sections 4.5** and **4.6**).
    - For example, in the area of segmentation, providing shared compute infrastructure for a segmentation competition could allow scientists in lower resource settings to test and demonstrate their ideas.
    - If successful, this could provide a starting point for ongoing support of such a platform e.g. to assist with the generation of connectomes of interest to LMIC countries.
    - Related to this, LMICs could be involved at an early stage in collaborative proofreading to improve the ground truth for mammalian connectome datasets. Initiatives such as the CIRCUIT programme at Johns Hopkins that have been developed to engage underserved communities in the US could be a useful starting point (Johns Hopkins Applied Physics Laboratory, 2022; Kerrigan, 2022).

- Initial discussion during the interview process with scientists based in LMICs indicates that scale-up of proofreading could be of significant interest if developed in tandem with training opportunities, technical infrastructure for connectomics and full project participation.
- Develop fellowship/training programmes hosted jointly between a global 'connectomics hub' and a home institution (e.g. 2 + 2-year travel/return) (see also **Section 3**).
- Run workshops on connectome analysis for finished connectomes, including the use of connectomes in computational neuroscience.
- Develop a local imaging centre suitable for 'integrative connectomics' imaging at the 1–20 mm^3 scale (as opposed to the whole-brain 500 mm^3 scale).
- Also use this to image small brains of economic/medical interest in LMIC, e.g. insect disease vectors such as *Anopheles stephensi* or pollinators like *Apis mellifera* or *Apis dorsata* (see also **Section 3**).
- Develop regional hub relationship (e.g. across India/Singapore/Korea/Japan) if one country were ready to fund a whole-brain connectomics centre.

6. Integrative connectomics: adding function to structure

Figure 14: Example of integrative connectomics
Physiological, molecular or other properties are collected and combined with a structural map of the same subregion using correlative multimodal imaging (CMI). The structural map (based on EM or perhaps XRM imaging) provides a connectome of that subregion. CMI alone will provide structure-function insights that are relevant to the study of how neuronal circuits operate. However, further insights can be reached if multiple maps of different specimens can be cross-integrated. To begin the process of integrating data across specimens, this multimodal map should be physically aligned onto the reference whole-brain connectome by 3D registration. It is the anatomical information contained within the regional connectome (both cell morphology and connectivity) that allows these two datasets from different animals to be fully integrated by cell type matching. Sophisticated CMI approaches have already been demonstrated in the mouse, although there is plenty of room for development. However, the second step, integrating across specimens, is not yet well characterised, although well-established in invertebrates.



There was a clear consensus from our interviews that the aim of the first mouse brain connectome should be to map the structural wiring diagram by tracing all neurons in the volume and identifying connected partners at each synapse. This will have huge intrinsic value. However, there is enormous added value when additional information such as molecular identity and neuronal activity are merged with the structural insights carried by the connectome. We refer to this process as *integrative connectomics* (**Figure 14**), which encompasses two main challenges: the overlay of rich multimodality datasets collected from within a single specimen (**Section 6.1**), and the merging of rich multimodality datasets collected from multiple specimens (**Section 6.2**). The effort to overcome these challenges in connectomics will benefit from advances being made in other research fields that depend on multiscale multimodal imaging of biological soft tissues at different scales, and could extend to integration of insights from preclinical and clinical imaging modalities (**Section 6.3**).

In this section of the report, we explore these different approaches to integrative connectomics, acknowledging that each approach will independently add value to the structural whole mouse brain connectome, and in combination they could deliver transformative insights into the function of neural circuits in health and disease states. Their success depends, however, on a number of significant unsolved challenges, and so these integrative workflows should be developed in parallel with the effort to deliver the first mouse brain connectome.

6.1 Overlaying data within one specimen: correlative multimodal imaging

Correlative multimodal imaging (**CMI**) is an area of intense interest in the biosciences that involves imaging a single specimen with multiple imaging modalities so that the information obtained can be related across imaging modalities. CMI approaches thereby allow interrogation of scales not accessible using a single microscope modality, such as the wiring and physiological properties of a specific neuronal circuit. There are two main applications of CMI approaches:

1. A ‘bridging use’: CMI for locating a region of interest (**ROI**) within a larger sample and imaging that limited volume at high resolution. This enables the same sample to be imaged sequentially in different imaging modalities, to combine the benefits of contextual information at low resolution in large samples (e.g. blood vessels) with mid-resolution features (e.g. cell bodies) and high-resolution structures of interest (e.g. organelles, membranes). CMI is thus a way of minimising the time required to image large tissue volumes at high resolution, by selecting and tracing small ROIs within the volume for high-resolution imaging. Looked at another way, bridging CMI increases the imaging speed by sacrificing high-resolution information outside of the ROIs.

2 A ‘contextual use’: CMI for attribution of orthogonal information to ultrastructural features. Contextual CMI can map information about the localisation of molecules, elements and transcriptomes into the three-dimensional ultrastructure of cells and tissues. In doing so, contextual CMI can confirm cell and organelle state and identity in a way that cannot currently be unequivocally determined by ultrastructure alone: e.g. by revealing which organelles are involved in molecular biogenesis and degradation, signalling, regulation and communication, or which neurons have exhibited specific activity profiles previously during an experiment interrogating the same sample *in vivo*. Contextual CMI depends on conserved landmarks across the imaging modalities used (light, X-ray, electron, ion and spatial elemental analysis imaging systems, to name a few). For neuroscience applications, those probes might target proteins (e.g. connexins, neurotransmitters, pathogenic oligomers such as Tau and α-synuclein), chemicals (e.g. drugs, toxins), signalling indicators (calcium, membrane potential), or metabolites (reactive oxygen species, oxygen consumption, glucose). Therefore, this approach could become a gold standard for extracting structure-function signatures within a sample for systems neuroscience research (e.g. relating the connectome to the tuning properties of the neurons embedded in it). Good probes are essential, and should ideally be multimodal, small, provide high signal-to-noise and low off-target background labelling, easily multiplexed for dense labelling and non-toxic. Overlay of datasets from different imaging modalities with different contrast regimes and different resolutions is particularly challenging and requires dedicated development effort.

Using contextual CMI techniques for the whole mouse brain will require significant advances in all aspects of the correlative pipeline, including:

- Probes for functional brain markers compatible with CMI techniques, including genetically encoded fluorescent protein tags and antibodies, fluorescent dyes, photoconvertible dyes that can be directly detected in electron and X-ray microscopes, RNA probes and stable isotope labels.
- CMI probes that can be (massively) multiplexed to assign functions across multiple features simultaneously.
- Protocols for expressing genetically encoded CMI probes in the brain without disrupting normal molecular localisation and function.
- Protocols for labelling the full volume of the brain with antibodies and dyes without disrupting the ultrastructure of the sample with detergents or solvents (non-permeabilisation techniques).
- A strategy for imaging through the whole intact mouse brain volume (without using clearing protocols that extract lipids from cell membranes and are therefore incompatible with preservation of morphology) using fluorescence microscopy, which is currently limited to an imaging depth of ~100 µm (for confocal) or ~500 µm (2-photon) or ~1,000 µm (3-photon) due to scattering of light in the tissue.

The main features in the brain that might benefit from localisation of molecular, electrical and chemical labels to understand function are cell nuclei (~5–15 µm), synaptic boutons (2–5 µm), neurites (0.1–10 µm), synaptic vesicles (~10 nm) and synapses (~20–40 nm). The required accuracy of a CMI approach is defined by the spatial density of the targeted feature: when the goal is identifying neurons previously imaged *in vivo*, this would often become the distance between any two adjacent cell bodies. This distance will be similar to the diameter of a cell body if their distribution is compact (e.g. in granular layers), but can be coarser where their distribution is sparser (e.g. in molecular or plexiform layers). For identifying specific synapses, a determining factor will be the spatial density of synaptic contacts (~1 synapse/µm in the mouse cortex (Kasthuri et al., 2015; Merchán-Pérez et al., 2014)). Therefore, the CMI technique should be able to deliver a localisation precision of at least 10 µm for localisation of nuclei, 250 nm for localisation of synapses, and <50 nm to localise specific synaptic vesicles.

The former is feasible now using correlative light, X-ray and electron microscopy. The firing pattern of the neurons contained in a ~1 mm^3 volume can be recorded *in vivo* with multiphoton microscopy, alongside their spatial distribution in relation to the fluorescently labelled blood vessel network. The ROI can then be dissected, stained and imaged using hard X-rays to relocate the position of all blood vessels with a contrast regime more similar to vEM. The blood vessel pattern then becomes a conserved landmark that allows correlation of the two datasets and retrieval of the positions of the neurons recorded *in vivo* in the resin-embedded sample, enabling targeted trimming and imaging with vEM. While this is only one approach to track and image biological features of a similar size and density to cell bodies, it serves as an example to showcase its logic and scalability for the whole mouse brain. Limitations might arise in recording neural activity across volumes >1 mm^3 with single-cell spatial resolution and sufficient temporal resolution to resolve action potentials. While advances in probes and *in vivo* multiphoton imaging are expected to expand the accessible brain volume, relevant insights will be obtained by combining regional *in vivo* physiological insight with whole-brain readout of the connectome.

For the latter, CMI approaches are likely to leverage probe detection in the electron microscope. For example, specific synaptic vesicle populations, identified according to *in vivo* activity patterns, could then be relocated in the connectome using genetically encoded horseradish peroxidase tags, which can be photo- or chemically converted into an electron-dense reaction product visible in the electron microscope (Atasoy et al., 2014; Simon et al., 2021), or specific cells or cell populations could be tagged and mapped onto the connectome using genetically encoded tags targeting specific organelles (Martell et al., 2012; Rhee et al., 2013; Zhang et al., 2019). The CMI techniques that have delivered the highest resolution correlations to date are based on ‘in-resin fluorescence’ and ‘on-section labelling’ techniques, which have reached ~50–80 nm lateral localisation accuracy on ~100 nm-thick sections (Simhal et al., 2018). High-resolution correlation through volumes beyond 100 nm could be achieved by labelling serial sections in a process called ‘correlative array tomography’ (CAT). However, CAT imaging strategies to date have either been proof of principle studies on low numbers of sections (<100), or imaging of sparsely labelled sections through a volume. Scale-up of at least four orders of magnitude would be required to deliver any CMI pipeline through the whole mouse brain.

6.2 Integrating data across specimens: comparative connectomics and beyond

The ultimate goal of systems neuroscience is to understand how the brain works. This aim implicitly expresses a will to refer the findings of any experiment to the expected outcomes of a reference system. There is therefore great value not only in acquiring the microcircuitry in one brain connectome (see **Section 2.1**), but also in understanding how connectomes of different individuals compare to each other, what features are consistent and what features stay the same (see also **Sections 7.1** and **7.2**).

There are multiple ways of integrating insights across specimens, each with particular advantages and challenges. A first level involving only connectomics data would involve extraction of meaningful differences between connectomes representing different experimental groups (of distinct e.g. developmental stage, health status or species). This approach is commonly termed *comparative connectomics* and has already delivered insights into developmental changes in the connectome of the somatosensory region of the mouse cortex (Gour et al., 2021), differences between the mouse and human cortices (Loomba et al., 2022), and evolutionary cues through comparison of the connectomes of different Drosophilids (Roberts et al., 2022).

A second level would reside in mapping neurons across individuals using single neurons or, more commonly, cell types – a set of neurons defined by morphology and sometimes connectivity that can be recognised across individuals. The added value of this approach has been clearly demonstrated in the ability to link molecular genetic experimental work in the lab to the *C. elegans* and *D. melanogaster* connectomes. However, this *tour de force* requires approaches to cross-identify neurons across individuals; in the mouse, this will likely involve an extended period of optimising cell type definitions as more connectomics data become available, alongside an array of other cell typing techniques (Zeng and Sanes, 2017).

In the worm, this has been resolved by identifying individual neurons by their shape and the position of their soma. Since the number of neurons is low (302 in the hermaphrodite) and invariant, experts could do this by hand without special software assistance. Over time, this has been used to build a complete mapping of molecular identity to neurons in the connectome. A recent example of a tool that can make this more efficient and accessible is the NeuroPAL multicolour labelling system, which in combination with associated software allows each neuron in a live worm to be recognised by the colour and position of the cell nucleus (Yemini et al., 2021). We think it is fair to say that integrative connectomics is a solved problem in the worm.

In the adult fly, the situation is still more complex although rapid progress is being made. The number of cells and cell types is much higher: about 150,000 neurons across the central nervous system and likely around 10,000 cell types (Galili et al., 2022). Furthermore, the number of cells per cell type varies within and across animals; although whole-brain statistics are not yet available, initial analysis indicates cell number variation occurs in about 20% of cell types per brain hemisphere (Schlegel et al., 2021). There are effective strategies to integrate molecular information with cell types identified in the connectome (Bates et al., 2019). These depend on the use of highly selective genetic driver lines, 3D alignment of different brains with high spatial accuracy (in the order of 3 µm) followed by cross-identification of stereotypical cell morphology aided by computational tools such as NBLAST (Costa et al., 2016). This means that, in theory, the problem of cross-identifying cell types – and therefore by extension the fundamental problem of integrative connectomics – appears largely solved in the fly. However, in practice the great majority of connectomics cell types have not been cross-identified. Furthermore, there are actually fundamental questions remaining about just how variable cell type number and connectivity are across the whole brain or what fraction of the 5,600 cell types identified in the first large, dense connectome in the fly (the hemibrain) will actually be robustly identifiable across individuals. This is an active area of research, and the first complete brain connectomes for the larva (Winding et al., 2022) and for the adult brain (expected 2023) will provide important data that may help to define both analytic tools and an intellectual framework that is helpful for vertebrate connectomics.

Altogether, this second level of comparative connectomics in the worm and the fly (or 'truly' *integrative* connectomics) depends on cross-identifying cell types across datasets through morphology. In the worm it is possible to do this without any special tooling, but in the fly the assistance of precise 3D alignment of EM and light-level datasets is important. Recently, co-registration has been taken to its logical extension in the ragworm *Platynereis*: the highly stereotyped position of individually identifiable cell bodies allowed direct fusion of gene expression data onto an EM volume (Vergara et al., 2021). In some areas of the mouse brain (e.g. the retina) we know that morphological information is sufficient to define functional cell types (Seung and Sümbül, 2014; Zeng and Sanes, 2017). However, in most areas this is very much a work in progress (Peng et al., 2021) that will interact with large-scale molecular cell typing (Yao et al., 2021).

Can integrative connectomics also work in the mouse by cross identifying cell types? What techniques will be necessary to do this?

Although we discussed strategies for integrative connectomics in the mouse with some interviewees, this did not seem to be an area where there was a strong consensus. Rather, this feels like a major knowledge gap at present. Part of the problem lies in the scientific discourse about cell types, which is somewhat contentious. Cell types are shorthand for groups of cells with conserved properties that can be identified across individuals (Zeng and Sanes, 2017). Identifying these conserved groups across individuals will be key to maximising the impact of mammalian connectomics. But they can be defined by many different methods and with many different levels of precision (Zeng, 2022). For connectomics, we will initially only have anatomical definitions of cell type based on morphology or connectivity. The challenge is to see how these are linked to molecular and functional definitions. Patch-seq is one method that can help: neurons are recorded by whole-cell patch clamp in a brain slice, electrophysiologically characterised and then filled with a fluorescent marker so that the dendritic and proximal axonal morphology can be recovered. Patch-seq can link molecular and morphological cell type definitions (Gouwens et al., 2020), but so far has not yielded morphological signatures for the finest subdivisions of molecular typing. This may be because Patch-seq data do not include connectivity information. In the fly there are actually a minority of cases (<10%) where reproducible cell types can only efficiently be inferred based on distinctive patterns of connectivity in addition to morphology (Scheffer et al., 2020; Schlegel et al., 2021). It may well be that connectivity information is much more important, perhaps essential, for fine-scale typing in the mouse.

We can try to draw out the significance of these points for integrative connectomics by considering a specific and much-studied example: orientation tuning in the visual cortex. A cubic millimetre dataset of mouse visual cortex has already been recovered together with calcium-imaging data (MICrONS Consortium et al., 2021). It is certainly possible to identify many discrete anatomical cell types such as excitatory pyramidal cells of different cortical layers and specific inhibitory interneurons such as chandelier cells; this level of cell typing is already hugely informative. But is it possible to identify cells with distinct orientation tuning from the connectome alone? In a smaller volume (0.03 mm^3) that has already been analysed, cells with the same orientation selectivity had statistically distinct connectivity signatures at the population level (Turner et al., 2022). But whether it is possible to go in the opposite direction, i.e. to predict a functional property from the connectome and to do so on a cell-by-cell basis, is not yet clear. Furthermore, we would like to know the absolute nature of this orientation tuning (which way is up): we imagine that it could be inferred from a whole-brain connectome containing all visual pathways from the retina, but we do not know; how small a connectome would suffice is still exceedingly hard to predict.

There is a huge literature on the identification of mammalian cell types by morphology alone, as well as linking these to molecular properties (BRAIN Initiative Cell Census Network (BICCN), 2021; Muñoz-Castañeda et al., 2021; Wheeler et al., 2015). All of these cell types will be distinguishable within connectome data so long as the volume is large enough – cubic-millimetre-scale datasets should suffice in most cases (MICrONS Consortium et al., 2021; Shapson-Coe et al., 2021). So the answer to our earlier question is clear: yes, integrative connectomics can work in the mouse by cross identifying cell types. But the question is, at what resolution? To resolve the finest-scale molecular or functional cell types, we can see three possible scenarios:

1. Functional cell types cannot be identified from connectome data in any meaningful way outside of peripheral areas of the brain.
2. Cell types can be identified so long as both local connectivity and morphology information are available.
3. As in the worm/fly, cell types can be uniquely identified by morphology.

If situation 1 is generally the case, then full success of integrative connectomics would depend on having a very large volume, perhaps even a whole-brain connectome, for every study. Besides the orientation-tuning data already mentioned (Turner et al., 2022) there is already some impressive evidence to suggest this is not the case (Economo et al., 2018; O'Toole et al., 2022), but this is far from exhaustive. At this point scenario 2 appears most likely; but lack of data means there is not yet a clear consensus in the field. Scenario 2 would mean that, for integrative connectomics to be most effective in the mouse, the target should be correlative imaging of connectomes in the scale of 1–20 cubic millimetres; these datasets would contain both extensive connectivity information as well as the desired molecular and functional characterisation. Note also that the extent to which scenario 3 is the case would impact the outcome of studies that use correlative imaging to recover dense neuronal morphology without connectivity. For example, functional imaging could be combined with high-resolution synchrotron X-ray techniques to reveal the morphology of every neuron in a volume.

We therefore recommend developing the technological basis for correlative multimodal imaging on the cubic millimetre scale (**Section 6.1**) and developing efficient methods to identify cell types within mammalian connectomics datasets. But crucially, we also recommend research into these fundamental issues of cell and circuit variability using connectomics approaches. This is an area where working on smaller mammals or other vertebrates could be highly complementary to work on subvolumes of the mouse brain (Barsotti et al., 2021).

For the mouse brain, we feel the future of integrative connectomics in the next 10–15 years is through targeted subvolumes. But the pipeline for delivering whole-brain connectomes should be designed from the start to be efficient, reproducible and scalable, so that acquisition of subsequent mouse brain connectomes will become faster and cheaper, much as technology pipeline improvements following the sequencing of the first human genome are now delivering a human genome sequence in just over five hours (Gorzynski et al., 2022). Faster delivery of connectomes will enable comparative connectomics by generating sufficient sample numbers to extract information from the comparison across multiple individuals. Topics may include:

- compare different sexes and genotypes
- compare individuals in a population
- identify patterns of learned and stochastic variation (Motta et al., 2019a)
- identify the statistical rules of the connectome based on variation within and across mammalian species (Loomba et al., 2022)
- compare individuals with different behaviours, including neurological/ neuropsychiatric disease models.

As noted, some initial connectomics investigations of these issues already exist, but whole-brain connectomics will enable new and much more far-reaching conclusions.

6.3 Incorporating preclinical and clinical imaging modalities

Organ-resolving imaging modalities can eventually also be integrated in CMI approaches following a similar logic to the one described above (Walter et al., 2021a, 2021b). These modalities include non-invasive imaging techniques commonly used in clinical frameworks, such as CT (computerised tomography), MRI (magnetic resonance imaging) or PET (positron emission tomography). Doing so would increase the range of techniques available for designing CMI approaches in connectomics frameworks and enable a synergistic development of both clinical and basic research approaches.

6.4 Recommendations

All these are [SML]:

- Fund research to better understand the nature of neuronal cell type variability, as revealed by connectomics across vertebrate and mammalian brains. This will contribute both to a better understanding of the logic of defining cell types as well as informing the practical requirements for integrative connectomics.
    - As this is an important intellectual and practical issue, we believe that it should be an early priority and will likely therefore need to be addressed by imaging large parts or the whole of multiple smaller brains in addition to favourable subvolumes in the mouse (which might include sensory regions such as the retina, subcortical structures or e.g. spinal cord).
    - This work could interact powerfully with efforts such as the BRAIN Initiative Cell Census Network (BICCN, https://biccn.org/) and Human Cell Atlas (HCA, https://www.humancellatlas.org/), and coordination, including at funder level, is recommended.
- Plan for and invest in infrastructure to allow individual labs to image/segment smaller volumes that have been physiologically or molecularly characterised.
    - These might use the infrastructure/technology that was used for stepping-stone volumes, which may well be different (e.g. cheaper/more robust) than whole-brain technology.
- Develop a call for proof of principle CMI pipelines capable of localising molecular, electrical and chemical signatures to ultrastructure in small cubes of mouse brain.
    - The challenge should be set so that it goes beyond current CMI capabilities in terms of overlay accuracy, resolution and volume. The pipelines should have the potential to scale to a volume of 1 cm^3. Two CMI targets could be considered: a localisation accuracy of 15 µm through a volume of 1 mm^3 for labelling neuronal soma; and a localisation accuracy of 50 nm through a volume of 100 $µm^3$ for labelling synaptic vesicles in synaptic boutons. Probe labelling strategies could be dense (i.e. continuous labelling through the volume), or sparse, as long as all features are captured by the sampling strategy.

7. Extended impact

7.1 Neuroscience: from the mouse connectome to *connectopathies*

In essence, the mouse brain connectomics challenge represents one particular application of a broader discipline – that of studying tissue architecture across many cubic millimetres with full ultrastructural detail. This imaging and analysis capacity is likely to result in a revolution in the study of histology and anatomy across scales. There are many applications of this technology beyond the normal mouse connectome and we now give a few examples.

First there are many cell types within the brain besides the neurons themselves that are critical to brain function. These include the classical glia and the brain's resident immune cells, the microglia. There is a huge amount of new biology in this area, with many links to disease processes as well as normal brain function. Within the mouse there will be many additional areas to target including the spinal cord and the peripheral nervous system including the enteric nervous system – so called brain-body interactions are a very active area of current research.

The opportunity to compare brain connectomes from different individuals will also bring many opportunities. Comparative connectomics therefore may cover multiple flavours of knowledge depending on the nature of the other connectomes – raising awareness of inter-individual variability when comparing within cohorts of individuals, or reporting variability across species, developmental stages or health status when configuring the experiments accordingly.

Specifically, while the prospects for whole-brain human connectomics at synaptic resolution still seem remote, there is much that would be possible based on the technology and insight generated by a mouse connectome. The cubic millimetre H01 dataset provides a glimpse of this (Shapson-Coe et al., 2021). That particular sample was removed during surgery to resect an epileptic focus in a patient's temporal lobe. The technological advances required for a mouse connectome would also allow patient biopsies of single or tens of cubic millimetres to be imaged and analysed rapidly. In the near term this should provide a unique opportunity for research in the diseased brain. It is plausible that a number of diseases may have their own blueprint in the multi-mm^3 connectomes of such biopsies, in the form of specific connectivity patterns between neurons. In those situations, such pathologies would classify as 'connectopathies' and their connectomics signature could help in understanding disease aetiology. For example, we could imagine a pathway from human genomics work identifying disease-associated genetic variants, to molecular studies revealing the expression locations of those genes in particular neuronal cell types, to connectomics work revealing circuit differences associated with those same specific cell types. Other possible applications in clinical neuroscience include examining the association of brain pathology with different circuits.

For example the Braak staging of Alzheimer's disease progression (Braak and Braak, 1991) is suggestive of spreading of tau filaments through connected circuits. Large-volume connectomics together with correlative approaches labelling filaments could provide unique insights. Synaptic resolution connectomics provides a fundamentally different way to study the brain and connectopathies may help understand mechanistically brain circuit malfunctions that have a huge impact on human health.

7.2 Research questions that cross scales

Solving the mouse brain connectome challenge will deliver the tools required to solve questions encoded in other biological systems in which the volume of the sample is around 500 mm^3, with features of interest that traverse microns or millimetres, with high-resolution features of interest at the 10 nm scale.

Other biological systems with a volume in the 1 cm^3 range include invertebrate model systems, snails, marine organisms, seedlings, organoids, organs from small mammals and small regions of human organs (e.g. liver, kidney, pancreas, lung, lymph nodes, eyes, ears and gut), embryos, human biopsies, and tumours. Each of these systems contain long-range features such as: polarised structures e.g. pollen tubes, fungal hyphae, root hairs and ciliated epithelia; vascular systems e.g. plant, blood and lymph; neuron-tissue innervations in the peripheral nervous system; damaged and regenerating tissue structures; and cancer cell networks and metastases. Examples of fine features distributed through tissue volumes in these systems include: viruses, plasmodesmata and chloroplasts in seedlings; basement membrane and podocyte foot processes in glomeruli in healthy and pathological kidney; insulin granules in beta cells in the pancreas; intracellular pathogens infecting tissues in model organisms, for example, *Mycobacterium tuberculosis* in lung granulomas; the microbiome attached to the ciliated surface of the gut in relation to the enteric nervous system; and immune cell extravasation during inflammation and tumour cell extravasation during metastasis.

7.3 High-throughput vEM for reproducible research and mapping morphological variation

These are just a few examples to demonstrate the breadth of research areas that can be addressed with large-volume imaging of tissues at EM resolution. Equally important is the impact on reproducibility of high-throughput imaging of smaller volumes with EM resolution. vEM is still often used as a qualitative research tool because of the time and complexity involved in imaging only one sample. High-throughput methods would enable analysis of sample numbers orders of magnitude larger than is currently possible to confirm findings and to understand morphological variation between individuals in a population, much as the speed-up in genomics technology has facilitated our understanding of genetic variation between individuals in the population.

7.4 Improvements in sample preparation protocols for tissue imaging

A new strategy for EM quality preservation, staining and embedding of tissue will be required for delivery of the mouse brain connectome. While advances in this area will inform general approaches, it is unlikely that the same protocol will be transferable to other tissues and model organisms of a similar size. This is because protocol design must consider the sample properties, such as composition, osmolarity, density and architecture of the tissue, whether there is a thick cell wall or cuticle, and how they will impact the diffusion and reaction of chemicals through the volume. Decisions must be taken on what level of preservation is needed for a specific biological question and which artefacts can be tolerated. This may be different even for different regions of the brain (e.g. those rich in myelinated vs non-myelinated axon bundles), let alone for other tissues and organisms. Even for smaller tissue volumes, there is no 'one size fits all' or consensus protocol. In addition, perfusion fixation is not always possible, for example, for insects, small organisms, marine organisms, plants and humans. Nevertheless, improvements made to sample preparation protocols for the mouse brain connectome could have significant benefits for other high-resolution imaging techniques. Currently, fixation of large samples for light sheet microscopy and histopathology often includes harsh chemicals that can induce changes in the ultrastructure of the sample that are effectively invisible to the end user but that may well cause artefacts in fluorescent label distribution and tissue density. As resolutions creep ever higher in light and X-ray imaging modalities, it will become more important to ensure that EM grade fixation is used to maintain near native state molecular distributions and tissue architecture for valid interpretation to be made.

7.5 Impact on big image data handling and analysis

For some time genomics has been the most data-intensive area in the biological sciences. But the development of high-throughput imaging techniques means that image volumes are increasing rapidly. The two- or three-dimensional nature of image content brings new computational challenges. But so too, does the scale of the data and in particular the scale of individual experiments. International genomics archives manage hundreds of petabytes of data, but individual studies are much smaller. A mouse connectome would derive from hundreds of petabytes of raw image data; a mouse genome even with highly redundant sequencing coverage and all raw data present is about five orders of magnitude smaller. Connectomics is now generating petascale datasets that are effectively too large to download in their entirety. This brings new challenges in enabling efficient random access to specific parts of the data. It also raises issues about what data-sharing practices are acceptable upon publication and whether and how it might be possible for other researchers to do analysis on raw connectome data.

There are undoubtedly significant opportunities for mutual support and learning with other research communities including large-scale biological imaging (Hartley et al., 2022) and the physical sciences. The mouse connectome is comparable to some of the largest experiments in physics. CERN, the European Laboratory for Particle Physics, currently has about 600 petabytes of stored data from what are effectively four large experiments each with data preservation, access and re-use policies (opendata.cern.ch). They classify data into four levels, where level 1 data are released as simple supplements to a paper while level 4 includes the raw data about collisions collected during experiments. The CERN Open Data Portal currently shares about 2 petabytes of data, principally at level 2 and 3; there is therefore no general access to raw project data. It would be valuable to engage deeply with these communities before collecting data for a whole mouse connectome.

7.6 Recommendations

- Gather thought leaders from a diverse range of life science fields (e.g. cell biology, developmental biology, neuroscience, infection and immunity, cancer research, ageing, plant biology, clinical research, biomaterials) to build a series of case studies that require imaging of large volumes (in the order of 0.5 cm^3) with EM contrast and resolution (in the order of 10 nm). Develop the detailed requirements for probes, sample preparation, imaging and analysis given the research question, sample type and features of interest for each study. Consider whether these requirements are synergistic with the developments required to deliver the mouse brain connectome, and whether these case studies can be started in parallel with the mouse brain connectome. [SM]

Conclusion

A synaptic resolution connectome of the mouse brain would have a huge impact on neuroscience but we are not yet ready to obtain one. When will we be ready to start? The answer is not clear yet for at least three reasons. First there is no validated sample preparation and imaging strategy suitable for a whole brain. As covered in **Sections 4.1** and **4.2** there are newer sample preparation and imaging strategies that might work in essentially their present form with steady progress in scale-up and reliability. Other approaches which have already been validated at the cubic millimetre scale may also be competitive if specific technical problems, such as lossless subdivision of centimetre-scale brains to millimetre-scale pieces can be solved. Second, as introduced in **Section 2.2** and then detailed in **Sections 4.5** and **4.6**, the cost is large and uncertain (an estimated $7.5–$21bn) because today's automatic segmentation technology still requires a huge amount of manual labour to produce a valid connectome. Third, there is a substantial interdependence between the quality of sample preparation and imaging and the downstream processes of image segmentation and connectome extraction: effectively the pipeline must be optimised end to end.

Despite these challenges, there is a great deal that can be done immediately to scale up connectomics with the long-term goal of a mouse connectome in mind. We recommend a significant investment in technology development, validated using a dual stepping-stone approach. Stepping stones would include large parts of – or complete – smaller brains in addition to subvolumes of the mouse brain. These will provide opportunities not just to solve problems of sample preparation, imaging and segmentation, but also some of the complex analysis challenges that we will face. Further technology development, including in correlative multimodal imaging (**Section 6.1**), will also provide added value, such as extracting structure-function signatures on individual brain samples. Investment into better understanding of how to define cell types in the mouse will maximise the value of a reference mouse connectome through integrative connectomics (**Section 6.2**). In tandem, there are significant scientific questions to address about variability in neuronal circuits across mammalian brains.

How could we know in the future that the time is right to start generating a whole mouse brain connectome? If we wait longer, it will become cheaper. But in the limit, that is a recipe for never starting and a mouse connectome will have great value – indeed a formal attempt to define at least a foreseeable part of that value could be helpful. Inter-individual variability means that a reference mouse connectome should be generated from a single specimen. But a data-generation project for a single mouse connectome that lasts 15 years is probably not viable now, even if it worked for the nematode 40 years ago. Perhaps one way to decide if the time is right is to ask, based on cutting-edge technology: is it possible and how long will it take from mouse to connectome? Significant technology development and demonstration is still required to give confident answers to these questions. The NIH BRAIN CONNECTS programme envisages five years of technology development (2023–28) preparatory to a five-year effort to obtain a first mouse brain connectome. Meeting that timeline will almost certainly require substantial investment beyond that programme, including in areas that we identify as gaps in this report.

We would like to close with one final point made during our interviews. Large-scale technology-intensive projects always face the potential emergence of new and disruptive technology. In connectomics, even though volume electron microscopy looks like the best option for the next decade, there is always the possibility of disruption e.g. from advances in X-ray imaging. Encouraging development of a portfolio of different technologies is vital for the long-term progress of the field. It is an interesting historical comparison that the human genome was completed mostly through evolution of sequencing technology – the revolution of next-generation sequencing occurred at least five years after its completion. Given this changing landscape, projects must be nimble enough to continue to develop and evaluate the best available technologies, and then decisive enough to deliver with them when the time is right.

References

Abbott LF, Angelaki DE, Carandini M, Churchland AK, Dan Y, Dayan P, Deneve S, Fiete I, Ganguli S, Harris KD, Häusser M, Hofer S, Latham PE, Mainen ZF, Mrsic-Flogel T, Paninski L, Pillow JW, Pouget A, Svoboda K, Witten IB, Zador AM. 2017. An international laboratory for systems and computational neuroscience. Neuron 96:1213–18. doi:10.1016/j.neuron.2017.12.013

Abbott LF, Bock DD, Callaway EM, Denk W, Dulac C, Fairhall AL, Fiete I, Harris KM, Helmstaedter M, Jain V, Kasthuri N, LeCun Y, Lichtman JW, Littlewood PB, Luo L, Maunsell JHR, Reid RC, Rosen BR, Rubin GM, Sejnowski TJ, Seung HS, Svoboda K, Tank DW, Tsao D, Van Essen DC. 2020. The mind of a mouse. Cell 182:1372–6. doi:10.1016/j.cell.2020.08.010

Abdellah M, Hernando J, Eilemann S, Lapere S, Antille N, Markram H, Schürmann F. 2018. NeuroMorphoVis: a collaborative framework for analysis and visualization of neuronal morphology skeletons reconstructed from microscopy stacks. Bioinformatics 34:i574–82. doi:10.1093/bioinformatics/bty231

Allen Institute for Brain Science. 2015. Allen Reference Atlas – Mouse Brain [Version 3 (2015)]. https://atlas.brain-map.org

Arshadi C, Günther U, Eddison M, Harrington KIS, Ferreira TA. 2021. SNT: a unifying toolbox for quantification of neuronal anatomy. Nat Methods 18:374–7. doi:10.1038/s41592-021-01105-7

Atasoy D, Betley JN, Li W-P, Su HH, Sertel SM, Scheffer LK, Simpson JH, Fetter RD, Sternson SM. 2014. A genetically specified connectomics approach applied to long-range feeding regulatory circuits. Nat Neurosci 17:1830–9. doi:10.1038/nn.3854

Badea A, Ali-Sharief AA, Johnson GA. 2007. Morphometric analysis of the C57BL/6J mouse brain. Neuroimage 37:683–93. doi:10.1016/j.neuroimage.2007.05.046

Barsotti E, Correia A, Cardona A. 2021. Neural architectures in the light of comparative connectomics. Curr Opin Neurobiol 71:139–49. doi:10.1016/j.conb.2021.10.006

Bates AS, Janssens J, Jefferis GSXE, Aerts S. 2019. Neuronal cell types in the fly: single-cell anatomy meets single-cell genomics. Curr Opin Neurobiol 56:125–34. doi:10.1016/j.conb.2018.12.012

Bates AS, Schlegel P, Roberts RJV, Drummond N, Tamimi IFM, Turnbull R, Zhao X, Marin EC, Popovici PD, Dhawan S, Jamasb A, Javier A, Serratosa Capdevila L, Li F, Rubin GM, Waddell S, Bock DD, Costa M, Jefferis GSXE. 2020. Complete connectomic reconstruction of olfactory projection neurons in the fly brain. Curr Biol 30:3183–99.e6. doi:10.1016/j.cub.2020.06.042

Berger DR, Seung HS, Lichtman JW. 2018. VAST (Volume Annotation and Segmentation Tool): efficient manual and semi-automatic labeling of large 3D image stacks. Front Neural Circuits 12:88. doi:10.3389/fncir.2018.00088

Berning M, Boergens KM, Helmstaedter M. 2015. Segem: efficient image analysis for high-resolution connectomics. Neuron 87:1193–206. doi:10.1016/j.neuron.2015.09.003

Birney E. 2012. The making of ENCODE: lessons for big-data projects. Nature 489:49–51. doi:10.1038/489049a

Bock DD, Lee W-CA, Kerlin AM, Andermann ML, Hood G, Wetzel AW, Yurgenson S, Soucy ER, Kim HS, Reid RC. 2011. Network anatomy and in vivo physiology of visual cortical neurons. Nature 471:177–82. doi:10.1038/nature09802

Boergens KM, Berning M, Bocklisch T, Bräunlein D, Drawitsch F, Frohnhofen J, Herold T, Otto P, Rzepka N, Werkmeister T, Werner D, Wiese G, Wissler H, Helmstaedter M. 2017. WebKnossos: efficient online 3D data annotation for connectomics. Nat Methods 14:691–4. doi:10.1038/nmeth.4331

Bogovic JA, Hanslovsky P, Wong A, Saalfeld S. 2016. Robust registration of calcium images by learned contrast synthesis. 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). Presented at the International Symposium on Biomedical Imaging (ISBI). IEEE. doi:10.1109/ISBI.2016.7493463

Bosch C, Jefferis GS, Collinson L, Schlegel P, Costa M 2023. In Scaling Up Connectomics: The Road To A Whole Mouse Brain Connectome. wtlandscape_mbc, 10.5281/zenodo.7599974 (Zenodo).

Bosch C, Karreman MA, Pacureanu A, Schaefer AT, Walter A. 2021. Correlative multimodality imaging across scales. In A Walter et al. (eds). Imaging Modalities for Biological and Preclinical Research: A Compendium, Volume 2. Bristol: IOP Publishing. pp. III.3-1–III.3-12. doi:10.1088/978-0-7503-3747-2ch23

Bosch C, Ackels T, Pacureanu A, Zhang Y, Peddie CJ, Berning M, Rzepka N, Zdora M-C, Whiteley I, Storm M, Bonnin A, Rau C, Margrie T, Collinson L, Schaefer AT. 2022. Functional and multiscale 3D structural investigation of brain tissue through correlative in vivo physiology, synchrotron microtomography and volume electron microscopy. Nat Commun 13:2923. doi:10.1038/s41467-022-30199-6

Braak H, Braak E. 1991. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol 82:239–59. doi:10.1007/BF00308809

BRAIN Initiative Cell Census Network (BICCN). 2021. A multimodal cell census and atlas of the mammalian primary motor cortex. Nature 598:86–102. doi:10.1038/s41586-021-03950-0

Briggman KL, Bock DD. 2012. Volume electron microscopy for neuronal circuit reconstruction. Curr Opin Neurobiol 22:154–61. doi: 10.1016/j.conb.2011.10.022

Briggman KL, Helmstaedter M, Denk W. 2011. Wiring specificity in the direction-selectivity circuit of the retina. Nature 471:183–8. doi:10.1038/nature09818

Bushong EA, Johnson DD Jr, Kim K-Y, Terada M, Hatori M, Peltier ST, Panda S, Merkle A, Ellisman MH. 2015. X-ray microscopy as an approach to increasing accuracy and efficiency of serial block-face imaging for correlated light and electron microscopy of biological specimens. Microsc Microanal 21:231–8. doi:10.1017/S1431927614013579

Clements J, Dolafi T, Umayam L, Neubarth NL, Berg S, Scheffer LK, Plaza SM. 2020. neuPrint: analysis tools for EM connectomics. bioRxiv 2020.01.16.909465. doi:10.1101/2020.01.16.909465

Costa M, Manton JD, Ostrovsky AD, Prohaska S, Jefferis GSXE. 2016. NBLAST: rapid, sensitive comparison of neuronal structure and construction of neuron family databases. Neuron 91:293–311. doi:10.1016/j.neuron.2016.06.012

Deerinck TJ, Shone TM, Bushong EA, Ramachandra R, Peltier ST, Ellisman MH. 2018. High-performance serial block-face SEM of nonconductive biological samples enabled by focal gas injection-based charge compensation. J Microsc 270:142–9. doi:10.1111/jmi.12667

Dempster WT. 1960. Rates of penetration of fixing fluids. Am J Anat 107:59–72. doi:10.1002/aja.1001070105

Denk W, Horstmann H. 2004. Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. PLoS Biol 2:e329. doi:10.1371/journal.pbio.0020329

Dorkenwald S, Li PH, Januszewski M, Berger DR, Maitin-Shepard J, Bodor AL, Collman F, Schneider-Mizell CM, da Costa NM, Lichtman JW, Jain V. 2022a. Multi-layered maps of neuropil with segmentation-guided contrastive learning. bioRxiv 2022.03.29.486320. doi:10.1101/2022.03.29.486320

Dorkenwald S, McKellar CE, Macrina T, Kemnitz N, Lee K, Lu R, Wu J, Popovych S, Mitchell E, Nehoran B, Jia Z, Bae JA, Mu S, Ih D, Castro M, Ogedengbe O, Halageri A, Kuehner K, Sterling AR, Ashwood Z, Zung J, Brittain D, Collman F, Schneider-Mizell C, Jordan C, Silversmith W, Baker C, Deutsch D, Encarnacion-Rivera L, Kumar S, Burke A, Bland D, Gager J, Hebditch J, Koolman S, Moore M, Morejohn S, Silverman B, Willie K, Willie R, Yu S-C, Murthy M, Seung HS. 2022b. FlyWire: online community for whole-brain connectomics. Nat Methods 19:119–28. doi:10.1038/s41592-021-01330-0

Dyer EL, Gray Roncal W, Prasad JA, Fernandes HL, Gürsoy D, De Andrade V, Fezzaa K, Xiao X, Vogelstein JT, Jacobsen C, Körding KP, Kasthuri N. 2017. Quantifying mesoscale neuroanatomy using x-ray microtomography. eNeuro 4. doi:10.1523/ENEURO.0195-17.2017

Eberle AL, Mikula S, Schalek R, Lichtman J, Tate MLK, Zeidler D. 2015. High-resolution, high-throughput imaging with a multibeam scanning electron microscope. J Microsc 259:114–20. doi:10.1111/jmi.12224

Economo MN, Viswanathan S, Tasic B, Bas E, Winnubst J, Menon V, Graybuck LT, Nguyen TN, Smith KA, Yao Z, Wang L, Gerfen CR, Chandrashekar J, Zeng H, Looger LL, Svoboda K. 2018. Distinct descending motor cortex pathways and their roles in movement. Nature 563:79–84. doi:10.1038/s41586-018-0642-9

Eichler K, Li F, Litwin-Kumar A, Park Y, Andrade I, Schneider-Mizell CM, Saumweber T, Huser A, Eschbach C, Gerber B, Fetter RD, Truman JW, Priebe CE, Abbott LF, Thum AS, Zlatic M, Cardona A. 2017. The complete connectome of a learning and memory centre in an insect brain. Nature 548:175–82. doi:10.1038/nature23455

Eschbach C, Fushiki A, Winding M, Schneider-Mizell CM, Shao M, Arruda R, Eichler K, Valdes-Aleman J, Ohyama T, Thum AS, Gerber B,

Fetter RD, Truman JW, Litwin-Kumar A, Cardona A, Zlatic M. 2020. Recurrent architecture for adaptive regulation of learning in the insect brain. Nat Neurosci 23:544–55. doi:10.1038/s41593-020-0607-9

Felsenberg J, Jacob PF, Walker T, Barnstedt O, Edmondson-Stait AJ, Pleijzier MW, Otto N, Schlegel P, Sharifi N, Perisse E, Smith CS, Lauritzen JS, Costa M, Jefferis GSXE, Bock DD, Waddell S. 2018. Integration of parallel opposing memories underlies memory extinction. Cell 175:709–22.e15. doi:10.1016/j.cell.2018.08.021

Galili DS, Jefferis GSXE, Costa M. 2022. Connectomics and the neural basis of behaviour. Curr Opin Insect Sci 54:100968. doi:10.1016/j.cois.2022.100968

Genoud C, Titze B, Graff-Meyer A, Friedrich RW. 2018. Fast homogeneous *en bloc* staining of large tissue samples for volume electron microscopy. Front Neuroanat 12:76. doi:10.3389/fnana.2018.00076

Givon LE, Lazar AA. 2016. Neurokernel: an open source platform for emulating the fruit fly brain. PLoS One 11:e0146581. doi:10.1371/journal.pone.0146581

Gorzynski JE, Goenka SD, Shafin K, Jensen TD, Fisk DG, Grove ME, Spiteri E, Pesout T, Monlong J, Baid G, Bernstein JA, Ceresnak S, Chang P-C, Christle JW, Chubb H, Dalton KP, Dunn K, Garalde DR, Guillory J, Knowles JW, Kolesnikov A, Ma M, Moscarello T, Nattestad M, Perez M, Ruzhnikov MRZ, Samadi M, Setia A, Wright C, Wusthoff CJ, Xiong K, Zhu T, Jain M, Sedlazeck FJ, Carroll A, Paten B, Ashley EA. 2022. Ultrarapid nanopore genome sequencing in a critical care setting. N Engl J Med 386:700–2. doi:10.1056/NEJMc2112090

Gour A, Boergens KM, Heike N, Hua Y, Laserstein P, Song K, Helmstaedter M. 2021. Postnatal connectomic development of inhibition in mouse barrel cortex. Science 371:eabb4534. doi:10.1126/science.abb4534

Gouwens NW, Sorensen SA, Baftizadeh F, Budzillo A, Lee BR, Jarsky T, Alfiler L, Baker K, Barkan E, Berry K, Bertagnolli D, Bickley K, Bomben J, Braun T, Brouner K, Casper T, Crichton K, Daigle TL, Dalley R, de Frates RA, Dee N, Desta T, Lee SD, Dotson N, Egdorf T, Ellingwood L, Enstrom R, Esposito L, Farrell C, Feng D, Fong O, Gala R, Gamlin C, Gary A, Glandon A, Goldy J, Gorham M, Graybuck L, Gu H, Hadley K, Hawrylycz MJ, Henry AM, Hill D, Hupp M, Kebede S, Kim TK, Kim L, Kroll M, Lee C, Link KE, Mallory M, Mann R, Maxwell M, McGraw M, McMillen D, Mukora A, Ng L, Ng L, Ngo K, Nicovich PR, Oldre A, Park D, Peng H, Penn O, Pham T, Pom A, Popović Z, Potekhina L, Rajanbabu R, Ransford S, Reid D, Rimorin C, Robertson M, Ronellenfitch K, Ruiz A, Sandman D, Smith K, Sulc J, Sunkin SM, Szafer A, Tieu M, Torkelson A, Trinh J, Tung H, Wakeman W, Ward K, Williams G, Zhou Z, Ting JT, Arkhipov A, Sümbül U, Lein ES, Koch C, Yao Z, Tasic B, Berg J, Murphy GJ, Zeng H. 2020. Integrated morphoelectric and transcriptomic classification of cortical GABAergic cells. Cell 183:935–53.e19. doi:10.1016/j.cell.2020.09.057

Hagstroem L, Bahr GF. 1960. Penetration rates of osmium tetroxide with different fixation vehicles. Z Zellforch Microsk Anat Histochem 2:1–4. doi:10.1007/BF00736486

Hardcastle BJ, Omoto JJ, Kandimalla P, Nguyen B-CM, Keleş MF, Boyd NK, Hartenstein V, Frye MA. 2021. A visual pathway for skylight polarization processing in *Drosophila*. Elife 10:e63225. doi:10.7554/eLife.63225

Hartley M, Kleywegt GJ, Patwardhan A, Sarkans U, Swedlow JR, Brazma A. 2022. The BioImage Archive – building a home for life-sciences microscopy data. J Mol Biol 434:167505. doi:10.1016/j.jmb.2022.167505

Hayworth KJ, Morgan JL, Schalek R, Berger DR, Hildebrand DGC, Lichtman JW. 2014. Imaging ATUM ultrathin section libraries with WaferMapper: a multi-scale approach to EM reconstruction of neural circuits. Front Neural Circuits 8:68. doi:10.3389/fncir.2014.00068

Hayworth KJ, Xu CS, Lu Z, Knott GW, Fetter RD, Tapia JC, Lichtman JW, Hess HF. 2015. Ultrastructurally smooth thick partitioning and volume stitching for large-scale connectomics. Nat Methods 12:319–22. doi:10.1038/nmeth.3292

Hayworth K, Peale D, Januszewski M, Knott GW, Lu Z, Xu CS, Hess HF. 2020. Gas cluster ion beam SEM for imaging of large tissue samples with 10 nm isotropic resolution. Nat Methods 17:68–71. doi:10.1038/s41592-019-0641-2

Helmstaedter M, Briggman KL, Denk, W. 2008. 3D structural imaging of the brain with photons and electrons. Curr Opin Neurobiol 18:633–41

Hider R Jr, Kleissas D, Gion T, Xenes D, Matelsky J, Pryor D, Rodriguez L, Johnson EC, Gray-Roncal W, Wester B. 2022. The Brain Observatory Storage Service and Database (BossDB): a cloud-native approach for petascale neuroscience discovery. Front Neuroinform 16:828787. doi:10.3389/fninf.2022.828787

Hildebrand DGC, Cicconet M, Torres RM, Choi W, Quan TM, Moon J, Wetzel AW, Scott Champion A, Graham BJ, Randlett O, Plummer GS, Portugues R, Bianco IH, Saalfeld S, Baden AD, Lillaney K, Burns R, Vogelstein JT, Schier AF, Lee W-CA, Jeong W-K, Lichtman JW, Engert F. 2017. Whole-brain serial-section electron microscopy in larval zebrafish. Nature 545:345–9. doi:10.1038/nature22356

Holler M, Guizar-Sicairos M, Tsai EHR, Dinapoli R, Müller E, Bunk O, Raabe J, Aeppli G. 2017. High-resolution non-destructive three-dimensional imaging of integrated circuits. Nature 543:402–6. doi:10.1038/nature21698

Hörl D, Rojas Rusak F, Preusser F, Tillberg P, Randel N, Chhetri RK, Cardona A, Keller PJ, Harz H, Leonhardt H, Treier M, Preibisch S. 2019. BigStitcher: reconstructing high-resolution image datasets of cleared and expanded samples. Nat Methods 16:870–4. doi:10.1038/s41592-019-0501-0

Hu F, Shi L, Min W. 2019. Biological imaging of chemical bonds by stimulated Raman scattering microscopy. Nat Methods 16:830–42. doi:10.1038/s41592-019-0538-0

Hua Y, Laserstein P, Helmstaedter M. 2015. Large-volume en-bloc staining for electron microscopy-based connectomics. Nat Commun 6:7923. doi:10.1038/ncomms8923

Hua Y, Ding X, Wang H, Wang F, Lu Y, Neef J, Gao Y, Moser T, Wu H. 2021. Electron microscopic reconstruction of neural circuitry in the cochlea. Cell Rep 34:108551. doi:10.1016/j.celrep.2020.108551

Huang L, Kebschull JM, Fürth D, Musall S, Kaufman MT, Churchland AK, Zador AM. 2020. BRICseq bridges brain-wide interregional connectivity to neural activity and gene expression in single animals. Cell 182:177–88.e27. doi:10.1016/j.cell.2020.05.029

Huang Y-C, Wang C-T, Wang G-T, Su T-S, Hsiao P-Y, Lin C-Y, Hsieh C-H, Chang H-M, Lo C-C. 2014. The Flysim project – persistent simulation and real-time visualization of fruit fly whole-brain spiking neural network model. Front Neuroinform. 8:00043. Conference Abstract: Neuroinformatics 2014. doi:10.3389/conf.fninf.2014.18.00043

Hubbard PM, Berg S, Zhao T, Olbris DJ, Umayam L, Maitin-Shepard J, Januszewski M, Katz WT, Neace ER, Plaza SM. 2020. Accelerated EM connectome reconstruction using 3D visualization and segmentation graphs. bioRxiv. doi:10.1101/2020.01.17.909572

Hulse BK, Haberkern H, Franconville R, Turner-Evans D, Takemura S-Y, Wolff T, Noorman M, Dreher M, Dan C, Parekh R, Hermundstad AM, Rubin GM, Jayaraman V. 2021. A connectome of the central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. Elife 10:e66039. doi:10.7554/eLife.66039

Januszewski M, Kornfeld J, Li PH, Pope A, Blakely T, Lindsey L, Maitin-Shepard J, Tyka M, Denk W, Jain V. 2018. High-precision automated reconstruction of neurons with flood-filling networks. Nat Methods 15:605–10. doi:10.1038/s41592-018-0049-4

Johns Hopkins Applied Physics Laboratory. 2022. CIRCUIT – Cohort-based Integrated Research Community for Undergraduate Innovation and Trailblazing. CIRCUIT – Cohort-based Integrated Research Community for Undergraduate Innovation and Trailblazing. https://www.jhuapl.edu/circuit/

Jovanic T, Schneider-Mizell CM, Shao M, Masson J-B, Denisov G, Fetter RD, Mensh BD, Truman JW, Cardona A, Zlatic M. 2016. Competitive disinhibition mediates behavioral choice and sequences in *Drosophila*. Cell 167:858–70.e19. doi:10.1016/j.cell.2016.09.009

Karagiannis ED, Kang JS, Shin TW, Emenari A, Asano S, Lin L, Costa EK, Marblestone AH, Kasthuri N, Boyden ES. 2019. Expansion microscopy of lipid membranes. bioRxiv 829903. doi:10.1101/829903

Kasthuri N, Hayworth KJ, Berger DR, Schalek RL, Conchello JA, Knowles-Barley S, Lee D, Vázquez-Reina A, Kaynig V, Jones TR, Roberts M, Morgan JL, Tapia JC, Seung HS, Roncal WG, Vogelstein JT, Burns R, Sussman DL, Priebe CE, Pfister H, Lichtman JW. 2015. Saturated reconstruction of a volume of neocortex. Cell 162:648–61. doi:10.1016/j.cell.2015.06.054

Katz WT, Plaza SM. 2019. DVID: Distributed Versioned Image-Oriented Dataservice. Front Neural Circuits 13:5. doi:10.3389/fncir.2019.00005

Kerrigan K. 2022. CIRCUIT sparks investment in the next generation of STEM professionals. Johns Hopkins Applied Physics Laboratory. https://www.jhuapl.edu/NewsStory/221006-circuit-sparks-investment-in-next-generation-of-stem-professionals/

Khairy K, Denisov G, Saalfeld S. 2018. Joint deformable registration of large EM image volumes: a matrix solver approach. arXiv 1804.10019. doi:10.48550/ARXIV.1804.10019

Klinger E, Motta A, Marr C, Theis FJ, Helmstaedter M. 2021. Cellular connectomes as arbiters of local circuit models in the cerebral cortex. Nat Commun 12:2785. doi:10.1038/s41467-021-22856-z

Knott G, Marchman H, Wall D, Lich B. 2008. Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling. J Neurosci 28:2959–64. doi:10.1523/JNEUROSCI.3189-07.2008

Knott G, Rosset S, Cantoni M. 2011. Focussed ion beam milling and scanning electron microscopy of brain tissue. J Vis Exp 53:e2588. doi:10.3791/2588

Koch C, Svoboda K, Bernard A, Basso MA, Churchland AK, Fairhall AL, Groblewski PA, Lecoq JA, Mainen ZF, Mathis MW, Olsen SR, Phillips JW, Pouget A, Saxena S, Siegle JH, Zador AM. 2022. Next-generation brain observatories. Neuron 110:3661–6. doi:10.1016/j.neuron.2022.09.033

Kornfeld J, Denk W. 2018. Progress and remaining challenges in high-throughput volume electron microscopy. Curr Opin Neurobiol 50:261–7. doi: 10.1016/j.conb.2018.04.030

Kuan AT, Phelps JS, Thomas LA, Nguyen TM, Han J, Chen C-L, Azevedo AW, Tuthill JC, Funke J, Cloetens P, Pacureanu A, Lee W-CA. 2020. Dense neuronal reconstruction through X-ray holographic nano-tomography. Nat Neurosci 23:1637–43. doi:10.1038/s41593-020-0704-9

Lee W-CA, Bonin V, Reed M, Graham BJ, Hood G, Glattfelder K, Reid RC. 2016. Anatomy and function of an excitatory network in the visual cortex. Nature 532:370–4. doi:10.1038/nature17192

Leighton SB. 1981. SEM images of block faces, cut by a miniature microtome within the SEM – a technical note. Scan Electron Microsc 73–6.

Li F, Lindsey JW, Marin EC, Otto N, Dreher M, Dempsey G, Stark I, Bates AS, Pleijzier MW, Schlegel P, Nern A, Takemura S-Y, Eckstein N, Yang T, Francis A, Braun A, Parekh R, Costa M, Scheffer LK, Aso Y, Jefferis GS, Abbott LF, Litwin-Kumar A, Waddell S, Rubin GM. 2020. The connectome of the adult *Drosophila* mushroom body provides insights into function. Elife 9:e62576. doi:10.7554/eLife.62576

Li PH, Lindsey LF, Januszewski M, Zheng Z, Bates AS, Taisz I, Tyka M, Nichols M, Li F, Perlman E, Maitin-Shepard J, Blakely T, Leavitt L, Jefferis GSX, Bock D, Jain V. 2020. Automated reconstruction of a serial-section EM *Drosophila* brain with flood-filling networks and local realignment. bioRxiv 605634. doi:10.1101/605634

Lichtman JW, Sanes JR. 2008. Ome sweet ome: what can the genome tell us about the connectome? Curr Opin Neurobiol 18:346–53. doi:10.1016/j.conb.2008.08.010

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

Lu J, Tapia JC, White OL, Lichtman JW. 2009. The interscutularis muscle connectome. PLoS Biol 7:e32. doi:10.1371/journal.pbio.1000032

Luo L, Callaway EM, Svoboda K. 2018. Genetic dissection of neural circuits: a decade of progress. Neuron 98:256–81. doi:10.1016/j.neuron.2018.03.040

Macrina T, Lee K, Lu R, Turner NL, Wu J, Popovych S, Silversmith W, Kemnitz N, Alexander Bae J, Castro MA, Dorkenwald S, Halageri A, Jia Z, Jordan C, Li K, Mitchell E, Mondal SS, Mu S, Nehoran B, Wong W, Yu S-C, Bodor AL, Brittain D, Buchanan J, Bumbarger DJ, Cobos E, Collman F, Elabbady L, Fahey PG, Froudarakis E, Kapner D, Kinn S, Mahalingam G, Papadopoulos S, Patel S, Schneider-Mizell CM, Sinz FH, Takeno M, Torres R, Yin W, Pitkow X, Reimer J, Tolias AS, Clay Reid R, da Costa NM, Sebastian Seung H. 2021. Petascale neural circuit reconstruction: automated methods. bioRxiv 2021.08.04.455162. doi:10.1101/2021.08.04.455162

Mahalingam G, Torres R, Kapner D, Trautman ET, Fliss T, Seshamani S, Perlman E, Young R, Kinn S, Buchanan J, Takeno MM, Yin W, Bumbarger DJ, Gwinn RP, Nyhus J, Lein E, Smith SJ, Reid RC, Khairy KA, Saalfeld S, Collman F, Macarico da Costa N. 2022. A scalable and modular automated pipeline for stitching of large electron microscopy datasets. Elife 11:e76534. doi:10.7554/eLife.76534

Martell JD, Deerinck TJ, Sancak Y, Poulos TL, Mootha VK, Sosinsky GE, Ellisman MH, Ting AY. 2012. Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. Nat Biotechnol 30:1143–8. doi:10.1038/nbt.2375

Meechan K, Guan W, Riedinger A, Stankova V, Yoshimura A, Pipitone R, Milberger A, Schaar H, Romero-Brey I, Templin R, Peddie CJ, Schieber NL, Jones ML, Collinson L, Schwab Y. 2022. Crosshair, semi-automated targeting for electron microscopy with a motorised ultramicrotome. bioRxiv. doi:10.1101/2022.06.02.491151

Merchán-Pérez A, Rodríguez J-R, González S, Robles V, Defelipe J, Larrañaga P, Bielza C. 2014. Three-dimensional spatial distribution of synapses in the neocortex: a dual-beam electron microscopy study. Cereb Cortex 24:1579–88. doi:10.1093/cercor/bht018

MICrONS Consortium, Alexander Bae J, Baptiste M, Bodor AL, Brittain D, Buchanan J, Bumbarger DJ, Castro MA, Celii B, Cobos E, Collman F, da Costa NM, Dorkenwald S, Elabbady L, Fahey PG, Fliss T, Froudarakis E, Gager J, Gamlin C, Halageri A, Hebditch J, Jia Z, Jordan C, Kapner D, Kemnitz N, Kinn S, Koolman S, Kuehner K, Lee K, Li K, Lu R, Macrina T, Mahalingam G, McReynolds S, Miranda E, Mitchell E, Mondal SS, Moore M, Mu S, Muhammad T, Nehoran B, Ogedengbe O, Papadopoulos C, Papadopoulos S, Patel S, Pitkow X, Popovych S, Ramos A, Reid RC, Reimer J, Schneider-Mizell CM, Sebastian Seung H, Silverman B, Silversmith W, Sterling A, Sinz FH, Smith CL, Suckow S, Takeno M, Tan ZH, Tolias AS, Torres R, Turner NL, Walker EY, Wang T, Williams G, Williams S, Willie K, Willie R, Wong W, Wu J, Xu C, Yang R, Yatsenko D, Ye F, Yin W, Yu S-C. 2021. Functional connectomics spanning multiple areas of mouse visual cortex. bioRxiv 2021.07.28.454025. doi:10.1101/2021.07.28.454025

Mikula S, Binding J, Denk W. 2012. Staining and embedding the whole mouse brain for electron microscopy. Nat Methods 9:1198–201. doi:10.1038/nmeth.2213

Mikula S, Denk W. 2015. High-resolution whole-brain staining for electron microscopic circuit reconstruction. Nat Methods 12:541–6. doi:10.1038/nmeth.3361

Minnen D, Januszewski M, Blakely T, Shapson-Coe A, Schalek RL, Ballé J, Lichtman JW, Jain V. 2021. Denoising-based image compression for connectomics. bioRxiv. doi:10.1101/2021.05.29.445828

Montero-Crespo M, Dominguez-Alvaro M, Rondon-Carrillo P, Alonso-Nanclares L, DeFelipe J, Blazquez-Llorca L. 2020. Three-dimensional synaptic organization of the human hippocampal CA1 field. Elife 9:e57013. doi:10.7554/eLife.57013

Motta A, Berning M, Boergens KM, Staffler B, Beining M, Loomba S, Hennig P, Wissler H, Helmstaedter M. 2019a. Dense connectomic reconstruction in layer 4 of the somatosensory cortex. Science 366:eaay3134. doi:10.1126/science.aay3134

Motta A, Schurr M, Staffler B, Helmstaedter M. 2019b. Big data in nanoscale connectomics, and the greed for training labels. Curr Opin Neurobiol 55:180–7. doi:10.1016/j.conb.2019.03.012

M'Saad O, Bewersdorf J. 2020. Light microscopy of proteins in their ultrastructural context. Nat Commun 11:3850. doi:10.1038/s41467-020-17523-8

M'Saad O, Kasula R, Kondratiuk I, Kidd P, Falahati H, Gentile JE, Niescier RF, Watters K, Sterner RC, Lee S, Liu X, De Camilli P, Rothman JE, Koleske AJ, Biederer T, Bewersdorf J. 2022. All-optical visualization of specific molecules in the ultrastructural context of brain tissue. bioRxiv 2022.04.04.486901. doi:10.1101/2022.04.04.486901

Muñoz-Castañeda R, Zingg B, Matho KS, Chen X, Wang Q, Foster NN, Li A, Narasimhan A, Hirokawa KE, Huo B, Bannerjee S, Korobkova L, Park CS, Park Y-G, Bienkowski MS, Chon U, Wheeler DW, Li X, Wang Y, Naeemi M, Xie P, Liu L, Kelly K, An X, Attili SM, Bowman I, Bludova A, Cetin A, Ding L, Drewes R, D'Orazi F, Elowsky C, Fischer S, Galbavy W, Gao L, Gillis J, Groblewski PA, Gou L, Hahn JD, Hatfield JT, Hintiryan H, Huang JJ, Kondo H, Kuang X,

Lesnar P, Li X, Li Y, Lin M, Lo D, Mizrachi J, Mok S, Nicovich PR, Palaniswamy R, Palmer J, Qi X, Shen E, Sun Y-C, Tao HW, Wakemen W, Wang Y, Yao S, Yuan J, Zhan H, Zhu M, Ng L, Zhang LI, Lim BK, Hawrylycz M, Gong H, Gee JC, Kim Y, Chung K, Yang XW, Peng H, Luo Q, Mitra PP, Zador AM, Zeng H, Ascoli GA, Josh Huang Z, Osten P, Harris JA, Dong H-W. 2021. Cellular anatomy of the mouse primary motor cortex. Nature 598:159–66. doi:10.1038/s41586-021-03970-w

Ohyama T, Schneider-Mizell CM, Fetter RD, Aleman JV, Franconville R, Rivera-Alba M, Mensh BD, Branson KM, Simpson JH, Truman JW, Cardona A, Zlatic M. 2015. A multilevel multimodal circuit enhances action selection in *Drosophila*. Nature 520:633–9. doi:10.1038/nature14297

Okubo TS, Patella P, D'Alessandro I, Wilson RI. 2020. A neural network for wind-guided compass navigation. Neuron 107:924–40.e18. doi:10.1016/j.neuron.2020.06.022

O'Toole SM, Oyibo HK, Keller GB. 2022. Prediction error neurons in mouse cortex are molecularly targetable cell types. bioRxiv. doi:10.1101/2022.07.20.500837

Pallotto M, Watkins PV, Fubara B, Singer JH, Briggman KL. 2015. Extracellular space preservation aids the connectomic analysis of neural circuits. Elife 4:e08206. doi:10.7554/eLife.08206

Palyanov A, Khayrulin S, Larson SD, Dibert A. 2011. Towards a virtual *C. elegans*: a framework for simulation and visualization of the neuromuscular system in a 3D physical environment. Silico Biol 11:137–47. doi:10.3233/ISB-2012-0445

Peddie CJ, Genoud C, Kreshuk A, Meechan K, Micheva KD, Narayan K, Pape C, Parton RG, Schieber NL, Schwab Y, Titze B, Verkade P, Weigel A, Collinson LM. 2022. Volume electron microscopy. Nature Reviews Methods Primers 2:51. doi:10.1038/s43586-022-00131-9

Peng H, Xie P, Liu L, Kuang X, Wang Y, Qu L, Gong H, Jiang S, Li A, Ruan Z, Ding L, Yao Z, Chen C, Chen M, Daigle TL, Dalley R, Ding Z, Duan Y, Feiner A, He P, Hill C, Hirokawa KE, Hong G, Huang L, Kebede S, Kuo H-C, Larsen R, Lesnar P, Li L, Li Q, Li X, Li Y, Li Y, Liu A, Lu D, Mok S, Ng L, Nguyen TN, Ouyang Q, Pan J, Shen E, Song Y, Sunkin SM, Tasic B, Veldman MB, Wakeman W, Wan W, Wang P, Wang Q, Wang T, Wang Y, Xiong F, Xiong W, Xu W, Ye M, Yin L, Yu Y, Yuan J, Yuan J, Yun Z, Zeng S, Zhang S, Zhao S, Zhao Z, Zhou Z, Huang ZJ, Esposito L, Hawrylycz MJ, Sorensen SA, Yang XW, Zheng Y, Gu Z, Xie W, Koch C, Luo Q, Harris JA, Wang Y, Zeng H. 2021. Morphological diversity of single neurons in molecularly defined cell types. Nature 598:174–81. doi:10.1038/s41586-021-03941-1

Phelps JS, Hildebrand DGC, Graham BJ, Kuan AT, Thomas LA, Nguyen TM, Buhmann J, Azevedo AW, Sustar A, Agrawal S, Liu M, Shanny BL, Funke J, Tuthill JC, Lee W-CA. 2021. Reconstruction of motor control circuits in adult *Drosophila* using automated transmission electron microscopy. Cell 184:759–74.e18. doi:10.1016/j.cell.2020.12.013

Popovych S, Macrina T, Kemnitz N, Castro M, Nehoran B, Jia Z, Bae JA, Mitchell E, Mu S, Trautman ET, Saalfeld S, Li K, Seung S. 2022. Petascale pipeline for precise alignment of images from serial section electron microscopy. bioRxiv 2022.03.25.485816. doi:10.1101/2022.03.25.485816

Rhee H-W, Zou P, Udeshi ND, Martell JD, Mootha VK, Carr SA, Ting AY. 2013. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. Science 339:1328–31. doi:10.1126/science.1230593

Roberts RJV, Pop S, Prieto-Godino LL. 2022. Evolution of central neural circuits: state of the art and perspectives. Nat Rev Neurosci 23:725–43. doi:10.1038/s41583-022-00644-y

Robinson NB, Krieger K, Khan FM, Huffman W, Chang M, Naik A, Yongle R, Hameed I, Krieger K, Girardi LN, Gaudino M. 2019. The current state of animal models in research: a review. Int J Surg 72:9–13. doi:10.1016/j.ijsu.2019.10.015

Saalfeld S, Cardona A, Hartenstein V, Tomancak P. 2009. CATMAID: collaborative annotation toolkit for massive amounts of image data. Bioinformatics 25:1984–6. doi:10.1093/bioinformatics/btp266

Saalfeld S, Fetter R, Cardona A, Tomancak P. 2012. Elastic volume reconstruction from series of ultra-thin microscopy sections. Nat Methods 9:717–20. doi:10.1038/nmeth.2072

Santuy A, Rodríguez J-R, DeFelipe J, Merchán-Pérez A. 2018. Study of the size and shape of synapses in the juvenile rat somatosensory cortex with 3D electron microscopy. eNeuro 5. doi:10.1523/ENEURO.0377-17.2017

Schalek R, Kasthuri N, Hayworth K, Berger D, Tapia J, Morgan J, Turaga S, Fagerholm E, Seung H, Lichtman J. 2011. Development of high-throughput, high-resolution 3D reconstruction of large-volume biological tissue using automated tape collection ultramicrotomy and scanning electron microscopy. Microsc Microanal 17:966–7. doi:10.1017/S1431927611005708

Scheffer LK, Xu CS, Januszewski M, Lu Z, Takemura S-Y, Hayworth KJ, Huang GB, Shinomiya K, Maitlin-Shepard J, Berg S, Clements J, Hubbard PM, Katz WT, Umayam L, Zhao T, Ackerman D, Blakely T, Bogovic J, Dolafi T, Kainmueller D, Kawase T, Khairy KA, Leavitt L, Li PH, Lindsey L, Neubarth N, Olbris DJ, Otsuna H, Trautman ET, Ito M, Bates AS, Goldammer J, Wolff T, Svirskas R, Schlegel P, Neace E, Knecht CJ, Alvarado CX, Bailey DA, Ballinger S, Borycz JA, Canino BS, Cheatham N, Cook M, Dreher M, Duclos O, Eubanks B, Fairbanks K, Finley S, Forknall N, Francis A, Hopkins GP, Joyce EM, Kim S, Kirk NA, Kovalyak J, Lauchie SA, Lohff A, Maldonado C, Manley EA, McLin S, Mooney C, Ndama M, Ogundeyi O, Okeoma N, Ordish C, Padilla N, Patrick CM, Paterson T, Phillips EE, Phillips EM, Rampally N, Ribeiro C, Robertson MK, Rymer JT, Ryan SM, Sammons M, Scott AK, Scott AL, Shinomiya A, Smith C, Smith K, Smith NL, Sobeski MA, Suleiman A, Swift J, Takemura S, Talebi I, Tarnogorska D, Tenshaw E, Tokhi T, Walsh JJ, Yang T, Horne JA, Li F, Parekh R, Rivlin PK, Jayaraman V, Costa M, Jefferis GS, Ito K, Saalfeld S, George R, Meinertzhagen IA, Rubin GM, Hess HF, Jain V, Plaza SM. 2020. A connectome and analysis of the adult *Drosophila* central brain. Elife 9:57443. doi:10.7554/eLife.57443

Schlegel P, Bates AS, Stürner T, Jagannathan SR, Drummond N, Hsu J, Serratosa Capdevila L, Javier A, Marin EC, Barth-Maron A, Tamimi IF, Li F, Rubin GM, Plaza SM, Costa M, Jefferis GSXE. 2021. Information flow, cell types and stereotypy in a full olfactory connectome. Elife 10. doi:10.7554/eLife.66018

Schmidt M, Motta A, Sievers M, Helmstaedter M. 2022. RoboEM: automated 3D flight tracing for synaptic-resolution connectomics. bioRxiv 2022.09.08.507122. doi:10.1101/2022.09.08.507122

Schneider-Mizell CM, Gerhard S, Longair M, Kazimiers T, Li F, Zwart MF, Champion A, Midgley FM, Fetter RD, Saalfeld S, Cardona A. 2016. Quantitative neuroanatomy for connectomics in *Drosophila*. Elife 5:e12059. doi:10.7554/eLife.12059

Schubert PJ, Dorkenwald S, Januszewski M, Jain V, Kornfeld J. 2019. Learning cellular morphology with neural networks. Nat Commun 10:2736. doi:10.1038/s41467-019-10836-3

Seung HS, Sümbül U. 2014. Neuronal cell types and connectivity: lessons from the retina. Neuron 83:1262–72. doi:10.1016/j.neuron.2014.08.054

Shahmoradian SH, Tsai EHR, Diaz A, Guizar-Sicairos M, Raabe J, Spycher L, Britschgi M, Ruf A, Stahlberg H, Holler M. 2017. Three-dimensional imaging of biological tissue by cryo X-ray ptychography. Sci Rep 7:6291. doi:10.1038/s41598-017-05587-4

Shapson-Coe A, Januszewski M, Berger DR, Pope A, Wu Y, Blakely T, Schalek RL, Li PH, Wang S, Maitin-Shepard J, Karlupia N, Dorkenwald S, Sjostedt E, Leavitt L, Lee D, Bailey L, Fitzmaurice A, Kar R, Field B, Wu H, Wagner-Carena J, Aley D, Lau J, Lin Z, Wei D, Pfister H, Peleg A, Jain V, Lichtman JW. 2021. A connectomic study of a petascale fragment of human cerebral cortex. bioRxiv 2021.05.29.446289. doi:10.1101/2021.05.29.446289

Sheridan A, Nguyen T, Deb D, Lee W-CA, Saalfeld S, Turaga S, Manor U, Funke J. 2021. Local shape descriptors for neuron segmentation. bioRxiv. doi:10.1101/2021.01.18.427039

Simhal AK, Gong B, Trimmer JS, Weinberg RJ, Smith SJ, Sapiro G, Micheva KD. 2018. A computational synaptic antibody characterization tool for array tomography. Front Neuroanat 12:51. doi:10.3389/fnana.2018.00051

Simon A, Roth A, Sheridan A, Fişek M, Marra V, Racca C, Funke J, Staras K, Häusser M. 2021. Ultrastructural readout of in vivo synaptic activity for functional connectomics. bioRxiv. doi:10.1101/2021.07.07.451278

Song K, Feng Z, Helmstaedter M. 2022. High-contrast en-bloc staining of mouse whole-brain samples for EM-based connectomics. bioRxiv. doi:10.1101/2022.03.30.486341

Ströh S, Hammerschmith EW, Tank DW, Seung HS, Wanner AA. 2022. X-ray assisted electron microscopy staining for large biological samples. Elife 11:e72147. doi:10.7554/eLife.72147

Svara F, Förster D, Kubo F, Januszewski M, dal Maschio M, Schubert PJ, Kornfeld J, Wanner AA, Laurell E, Denk W, Baier H. 2022. Automated synapse-level reconstruction of neural circuits in the larval zebrafish brain. Nat Methods 19:1357–66. doi:10.1038/s41592-022-01621-0

Szigeti B, Gleeson P, Vella M, Khayrulin S, Palyanov A, Hokanson J, Currie M, Cantarelli M, Idili G, Larson S. 2014. OpenWorm: an open-science approach to modeling *Caenorhabditis elegans*. Front Comput Neurosci 8:137. doi:10.3389/fncom.2014.00137

Tapia JC, Kasthuri N, Hayworth KJ, Schalek R, Lichtman JW, Smith SJ, Buchanan J. 2012. High-contrast en bloc staining of neuronal tissue for field emission scanning electron microscopy. Nat Protoc 7:193–206. doi:10.1038/nprot.2011.439

Templier T. 2019. MagC, magnetic collection of ultrathin sections for volumetric correlative light and electron microscopy. Elife 8:e45696. doi:10.7554/eLife.45696

The NIH BRAIN Initiative and the Department of Energy Office of Science. n.d. The Brain Connectivity Workshop Series. NIH website. https://www.ninds.nih.gov/news-events/events/brain-connectivity-workshop-series

Titze B, Denk W. 2013. Automated in-chamber specimen coating for serial block-face electron microscopy. J Microsc 250:101–10. doi:10.1111/jmi.12023

Titze B, Genoud C, Friedrich RW. 2018. *SBEMimage*: versatile acquisition control software for serial block-face electron microscopy. Front Neural Circuits 12:54. doi:10.3389/fncir.2018.00054

Turner NL, Macrina T, Bae JA, Yang R, Wilson AM, Schneider-Mizell C, Lee K, Lu R, Wu J, Bodor AL, Bleckert AA, Brittain D, Froudarakis E, Dorkenwald S, Collman F, Kemnitz N, Ih D, Silversmith WM, Zung J, Zlateski A, Tartavull I, Yu S-C, Popovych S, Mu S, Wong W, Jordan CS, Castro M, Buchanan J, Bumbarger DJ, Takeno M, Torres R, Mahalingam G, Elabbady L, Li Y, Cobos E, Zhou P, Suckow S, Becker L, Paninski L, Polleux F, Reimer J, Tolias AS, Reid RC, da Costa NM, Seung HS. 2022. Reconstruction of neocortex: organelles, compartments, cells, circuits, and activity. Cell 185:1082–100.e24. doi:10.1016/j.cell.2022.01.023

Velicky P, Miguel E, Michalska JM, Wei D, Lin Z, Watson JF, Troidl J, Beyer J, Ben-Simon Y, Sommer C, Jahr W, Cenameri A, Broichhagen J, Grant SGN, Jonas P, Novarino G, Pfister H, Bickel B, Danzl JG. 2022. Saturated reconstruction of living brain tissue. bioRxiv. doi:10.1101/2022.03.16.484431

Vergara HM, Pape C, Meechan KI, Zinchenko V, Genoud C, Wanner AA, Mutemi KN, Titze B, Templin RM, Bertucci PY, Simakov O, Dürichen W, Machado P, Savage EL, Schermelleh L, Schwab Y, Friedrich RW, Kreshuk A, Tischer C, Arendt D. 2021. Whole-body integration of gene expression and single-cell morphology. Cell 184:4819–37.e22. doi:10.1016/j.cell.2021.07.017

Walsh CL, Tafforeau P, Wagner WL, Jafree DJ, Bellier A, Werlein C, Kühnel MP, Boller E, Walker-Samuel S, Robertus JL, Long DA, Jacob J, Marussi S, Brown E, Holroyd N, Jonigk DD, Ackermann M, Lee PD. 2021. Imaging intact human organs with local resolution of cellular structures using hierarchical phase-contrast tomography. Nat Methods 18:1532–41. doi:10.1038/s41592-021-01317-x

Walter, A, Mannheim JG, Caruana CJ (eds). 2021a. Imaging Modalities for Biological and Preclinical Research: A Compendium, vol. 1, part I: Ex Vivo Biological Imaging. Bristol: IOP Publishing. doi:10.1088/978-0-7503-3059-6

Walter, A, Mannheim JG, Caruana CJ (eds). 2021b. Imaging Modalities for Biological and Preclinical Research: A Compendium, vol. 2, parts II–IV: In Vivo Preclinical Imaging, Multimodality Imaging and Outlook. Bristol: IOP Publishing. doi:10.1088/978-0-7503-3747-2

Wanner AA, Genoud C, Masudi T, Siksou L, Friedrich RW. 2016. Dense EM-based reconstruction of the interglomerular projectome in the zebrafish olfactory bulb. Nat Neurosci 19:816–25. doi:10.1038/nn.4290

Wheeler DW, White CM, Rees CL, Komendantov AO, Hamilton DJ, Ascoli GA. 2015. Hippocampome.org: a knowledge base of neuron types in the rodent hippocampus. Elife 4:e09960. doi:10.7554/eLife.09960

White JG, Southgate E, Thomson JN, Brenner S. 1986. The structure of the nervous system of the nematode *Caenorhabditis elegans*. Philos Trans R Soc Lond B Biol Sci 314:1–340. doi:10.1098/rstb.1986.0056

Wilke SA, Antonios JK, Bushong EA, Badkoobehi A, Malek E, Hwang M, Terada M, Ellisman MH, Ghosh A. 2013. Deconstructing complexity: serial block-face electron microscopic analysis of the hippocampal mossy fiber synapse. J Neurosci 33:507–22. doi:10.1523/JNEUROSCI.1600-12.2013

Winding M, Pedigo BD, Barnes CL, Patsolic HG, Park Y, Kazimiers T, Fushiki A, Andrade IV, Khandelwal A, Valdes-Aleman J, Li F, Randel N, Barsotti E, Correia A, Fetter RD, Hartenstein V, Priebe CE, Vogelstein JT, Cardona A, Zlatic M. 2022. The connectome of an insect brain. *Accepted for publication*.

Xenes D, Kitchell LM, Rivlin PK, Brodsky R, Gooden H, Joyce J, Luna D, Norman-Tenazas R, Ramsden D, Romero K, Rose V, Villafane-Delgado M, Gray-Roncal W, Wester B. 2022. NeuVue: a framework and workflows for high-throughput electron microscopy connectomics proofreading. bioRxiv 2022.07.18.500521. doi:10.1101/2022.07.18.500521

Xu CS, Hayworth KJ, Lu Z, Grob P, Hassan AM, García-Cerdán JG, Niyogi KK, Nogales E, Weinberg RJ, Hess HF. 2017. Enhanced FIB-SEM systems for large-volume 3D imaging. Elife 6:e25916. doi:10.7554/eLife.25916

Yao Z, van Velthoven CTJ, Nguyen TN, Goldy J, Sedeno-Cortes AE, Baftizadeh F, Bertagnolli D, Casper T, Chiang M, Crichton K, Ding S-L, Fong O, Garren E, Glandon A, Gouwens NW, Gray J, Graybuck LT, Hawrylycz MJ, Hirschstein D, Kroll M, Lathia K, Lee C, Levi B, McMillen D, Mok S, Pham T, Ren Q, Rimorin C, Shapovalova N, Sulc J, Sunkin SM, Tieu M, Torkelson A, Tung H, Ward K, Dee N, Smith KA, Tasic B, Zeng H. 2021. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. Cell 184:3222–41.e26. doi:10.1016/j.cell.2021.04.021

Yemini E, Lin A, Nejatbakhsh A, Varol E, Sun R, Mena GE, Samuel ADT, Paninski L, Venkatachalam V, Hobert O. 2021. NeuroPAL: a multicolor atlas for whole-brain neuronal identification in *C. elegans*. Cell 184:272–88. e11. doi:10.1016/j.cell.2020.12.012

Yin W, Brittain D, Borseth J, Scott ME, Williams D, Perkins J, Own CS, Murfitt M, Torres RM, Kapner D, Mahalingam G, Bleckert A, Castelli D, Reid D, Lee W-CA, Graham BJ, Takeno M, Bumbarger DJ, Farrell C, Reid RC, da Costa NM. 2020. A petascale automated imaging pipeline for mapping neuronal circuits with high-throughput transmission electron microscopy. Nat Commun 11:4949. doi:10.1038/s41467-020-18659-3

Zeng H. 2022. What is a cell type and how to define it? Cell 185:2739–55. doi:10.1016/j.cell.2022.06.031

Zeng H, Sanes JR. 2017. Neuronal cell-type classification: challenges, opportunities and the path forward. Nat Rev Neurosci 18:530–46. doi:10.1038/nrn.2017.85

Zhang Q, Lee W-CA, Paul DL, Ginty DD. 2019. Multiplexed peroxidase-based electron microscopy labeling enables simultaneous visualization of multiple cell types. Nat Neurosci 22:828–39. doi:10.1038/s41593-019-0358-7

Zhang Y, Ackels T, Pacureanu A, Zdora M-C, Bonnin A, Schaefer AT, Bosch C. 2022. Sample preparation and warping accuracy for correlative multimodal imaging in the mouse olfactory bulb using 2-photon, synchrotron X-ray and volume electron microscopy. Front Cell Dev Biol 10:880696. doi:10.3389/fcell.2022.880696

Zhao T, Olbris DJ, Yu Y, Plaza SM. 2018. NeuTu: software for collaborative, large-scale, segmentation-based connectome reconstruction. Front Neural Circuits 12:101. doi:10.3389/fncir.2018.00101

Zheng Z, Lauritzen JS, Perlman E, Robinson CG, Nichols M, Milkie D, Torrens O, Price J, Fisher CB, Sharifi N, Calle-Schuler SA, Kmecova L, Ali IJ, Karsh B, Trautman ET, Bogovic JA, Hanslovsky P, Jefferis GSXE, Kazhdan M, Khairy K, Saalfeld S, Fetter RD, Bock DD. 2018. A complete electron microscopy volume of the brain of adult *Drosophila melanogaster*. Cell 174:730–43.e22. doi:0.1016/j.cell.2018.06.019

Appendices

Appendix A

Interviewed experts

Wei-Chung Allen Lee (Harvard University)
Assistant Professor of Neurology

Ewan Birney (EMBL-EBI)
Deputy Director General and Joint Director

JoAnn Buchanan (Allen Institute) Scientist

Albert Cardona (University of Cambridge and MRC Laboratory of Molecular Biology)
Group Leader at MRC LMB and Professor at the University of Cambridge

Felipe Court (Universidad Mayor, Chile)
Professor and Director of the Centre for Integrative Biology

Kirk Czymmek (Donald Danforth Plant Science Centre) Director of the Advanced Bioimaging Laboratory

Sven Dorkenwald (FlyWire/Princeton University and Google Research)
PhD student and lead developer at FlyWire, student researcher at Google

Jan Funke (HHMI Janelia Research Campus)
Group Leader

Christel Genoud (University of Lausanne)
Head of the Electron Microscopy Facility

Chris Guerin (VIB Gent)
Emeritus microscopist and training coordinator

Michael Hansen (Microsoft Research)
Director and Principal Researcher in Health Futures

Matthew Hartley (EMBL-EBI)
BioImage Archive Team Leader

Moritz Helmstaedter (Max Planck Institute for Brain Research) Professor and Director

Tony Hey (UK Science and Technology Facilities Council) Chief Data Scientist

Bruno Humbel (Okinawa Institute of Science and Technology) Leader of the Scientific Imaging Section

Viren Jain (Google Research)
Lead of the Connectomics Team

Haibo Jiang (University of Hong Kong)
Associate Professor and Director of JC STEM Lab of Molecular Imaging

Gerard Kleywegt (EMBL-EBI) Senior Team Leader, Lead of the Protein Data Bank in Europe

Anna Kreshuk (EMBL Heidelberg)
Group Leader

Pieter Kruit (TU Delft)
Emeritus Professor of Physics

Jeff Lichtman (Harvard University) Professor of Molecular and Cellular Biology and Santiago Ramón y Cajal Professor of Arts and Sciences

Jennifer Lippincott Schwartz (HHMI Janelia Research Campus) Senior Group Leader and Head of Janelia's 4D Cellular Physiology

Ben Loos (University of Stellenbosch, South Africa) Professor

Xiaotang Lu (Harvard University)
Research Associate, K99 Postdoctoral Fellow

Tommy Macrina (Zeta.ai LLC) CEO

Claire McKellar (FlyWire/Princeton University)
Director of Operations and Community

Kristina Micheva (Stanford University)
Senior Research Scientist

Alexandra Pacureanu (The European Synchrotron Radiation Facility)
Research Scientist

Song Pang (Yale University)
Director, FIB-SEM Collaboration Core

Rob Parton (University of Brisbane) Professor, Role of the Cell Surface in Health and Disease

Lucia Prieto (Crick Institute) Group Leader, founder of TReND (Teaching and Research in Natural Sciences for Development) in Africa

Gerry Rubin (HHMI Janelia Research Campus) Senior Group Leader, founding director of HHMI Janelia Research Campus

Norman Rzepka (Scalable Minds) Co-founder

Andreas Schaefer (Crick Institute, University College London) Principal Group Leader, Assistant Research Director at the Crick and Professor of Neuroscience at UCL

Casey Schneider-Mizell (Allen Institute) Scientist

Yannick Schwab (EMBL Heidelberg) Team Leader and Head of Electron Microscopy Core Facility

Sebastian Seung (Princeton University) Professor of Computer Science and Neuroscience, PI EyeWire and FlyWire

Reza Shahidi (University of Exeter) Postdoctoral researcher

Kun Song (Max Planck Institute for Brain Research) Postdoctoral researcher

Amy Sterling (Eyewire/Princeton) Executive Director of Eyewire

Jason Swedlow (University of Dundee) Professor of Quantitative Cell Biology

David Tank (Princeton University) Henry L. Hillman Professor in Molecular Biology and Co-director of the Princeton Neuroscience Institute

Virginie Uhlmann (EMBL-EBI) Research Group Leader

K. VijayRaghavan (National Centre for Biological Sciences, Bangalore) Emeritus Professor, DAE Homi Bhabha Chair

Miah Wander (Microsoft Research) Principal Researcher with the Biomedical Computing Team

Adrian Andreas Wanner (Paul Scherrer Institute) Group Leader

Rick Webb (University of Queensland) Adjunct Senior Lecturer, Senior Laboratory Manager for the Centre for Microscopy and Analysis and Senior Research Officer

Shan Xu (Yale University) Harvey and Kate Cushing Professor of Cellular and Molecular Physiology

Appendix B

Connectomics software

Image Registration

Name	Description	Link/reference
Big Stitcher	Alignment of multimodal image tiles.	Hörl et al. (2019)
Big Warp	Registration of multimodal image stacks.	JA Bogovic et al. (2016)
Asap	Stitching and alignment pipeline used for the Allen Institute's cortical mm^3 dataset.	https://github.com/AllenInstitute/asap-modules
SEAMLeSS	Fine-scale alignment using neural networks.	Macrina et al. (2021); https://github.com/seung-lab/SEAMLeSS

Proofreading, Exploration & Analysis

Name	Description	Link/reference	Proof-reading	Explo-ration	Analy-sis
VAST	VAST (Volume Annotation and Segmentation Tool) is a utility application for manual annotation of large EM stacks.	Berger, Seung and Lichtman et al. (2018)	yes	some	some
RoboEM	Method for automated proofreading.	Schmidt et al. (2022)	yes		
CATMAID	Open-source, web-based tool for manual tracing and rich interactive analysis of large EM datasets.	Saalfeld et al. (2009)	yes	yes	yes
webKnossos	Open-source, web-based tool for annotating and exploring large 3D image datasets.	Boergens et al. (2017)	yes	yes	some
Knossos	Tool for visualisation and proofreading of large EM datasets. Desktop-based, but supports distributed proofreading.	Svara et al. (2022): https://github.com/knossos-project/knosso	yes	yes	some

Name	Description	Link/reference	Proof-reading	Exploration	Analysis
NeuTu EM	Open-source tool for large-scale collaborative proofreading.	Zhao et al. (2018)	yes	some	
Neuroglancer	Open-source, web-based tool for exploration and proofreading of large EM datasets. Used e.g. for FlyWire.	https://github.com/google/neuroglancer	yes	some	
NeuVue	Neuroglancer-based proofreading environment.	Xenes et al. (2022)	yes	some	
neuPrint	Exploration of large connectomics datasets.	Clements et al. (2020)		yes	some
natverse	R software suite for analyses of neuronal morphology and connectivity, including connectome data.	Bates et al. (2020); https://natverse.org			yes
navis	Python library for analysis and visualisation of connectomic data.	https://github.com/navis-org/navis			yes
Simple Neurite Tracer	ImageJ (Java-based) plugin for morphometric analyses.	Arshadi et al. (2021); https://github.com/morphonets/SNT			yes
NeuroMorphoVis	Blender 3D plugin for morphometric analyses.	Abdellah et al. (2018); https://github.com/BlueBrain/NeuroMorphoVis			yes

Backends

Name	Description	Link/reference
DVID	Open-source, web-based API for petascale EM-based datasets.	Katz and Plaza (2019)
PyChunkedGraph	Proofreading and data management backend built on Google Cloud Bigtable.	Dorkenwald, McKellar, et al. (2022b); https://github.com/seung-lab/PyChunkedGraph
BossDB	Open-source, cloud-based backend for petascale image datasets.	Hider et al. (2022)

Interfaces

Name	Description	Link/reference
cloud-volume	Python library for interacting with connectomics datasets.	https://github.com/seung-lab/cloud-volume
CAVEclient	Python library for interacting with PyChunkedGraph and the Connectome Annotation Versioning Engine.	https://github.com/seung-lab/CAVEclient

Simulation projects

Name	Description	Link/reference
OpenWorm	An international open science project with the aim to simulate the whole *C. elegans*.	Palyanov et al. (2011); Szigeti et al. (2014)
Flysim	Develop a brain-wide computational model for *Drosophila*, based on data from the FlyCircuit database.	Huang et al. (2014)
NeuroKernel / NeuroArch)	GPU-based simulation for the *Drosophila* brain.	Givon and Lazar (2016)

Datasets/repositories

Name	Description	Link/reference
NeuroMorpho	Collection of 186k single neuron morphologies from various different species (mostly from light-level data).	https://neuromorpho.org
MICrONS	Two large EM datasets from the mouse visual cortex.	https://www.microns-explorer.org/
Open Connectome	Collection of various electron microscopy datasets.	https://neurodata.io/project/ocp/
map**zebra**in	Repository for EM and LM (light microscopy) data for zebrafish.	https://mapzebrain.org/home
Virtual Fly Brain	Hub for integrated EM and LM data for *Drosophila*.	https://virtualflybrain.org

Appendix C

Glossary

connectome: A complete inventory of the neurons and synaptic connectivity making up a brain or another region of the nervous system.

connectopathies: Pathologies displaying characteristic connectome patterns.

neurite: Branch of a neuron.

axon: Parts of a neuron that represent the output domain.

dendrite: Parts of a neuron that represent the input domain.

soma: Cell body of a neuron.

synapse: Location of a specialised individual connection between two neurons. Each neuron will normally have thousands of synapses. Distinguishing synapses from other locations where neurons come close together usually requires nanometre-resolution imaging.

electron microscopy (EM): Imaging technique with nanometre resolution employing an electron beam as source of illumination.

transmission electron microscopy/ microscope (TEM): Electron microscopy technique where the image is generated by the incident beam not absorbed by the specimen.

scanning electron microscopy (SEM): Electron microscopy technique where the image is generated by the electrons scattered when every spot in the sample is illuminated by the incident beam.

volume electron microscopy (vEM): Electron microscopy techniques that generate serial images from resin-embedded cells and tissues with a continuous depth greater than 1 µm.

serial section transmission electron microscopy (ssTEM): vEM technique where the third dimension is obtained through sequential imaging of serial ultrathin sections in a TEM.

focused ion beam scanning electron microscopy (FIB-SEM): vEM technique where the third dimension is obtained through sequential imaging of serial block faces after a thin layer of material is vaporised with a focused gallium ion beam.

plasma FIB-SEM (pFIB-SEM): vEM technique where the third dimension is obtained through sequential imaging of serial block faces after a thin layer of material is vaporised with a focused ion beam of a different species to gallium (currently usually xenon or oxygen in biological applications).

X-ray microscopy (XRM): Imaging techniques employing an X-ray beam as source of illumination.

voxel: Three-dimensional unit of spatial data.

resolution: Closest distance between two adjacent features that can be distinguished in an image.

field of view: Size of the landscape mapped by an image (width, height, depth).

automated segmentation: Process by which particular features from an image, such as cellular membranes or synapses, can be identified automatically by machine-learning algorithms. For neurite segmentation, after an initial identification stage, small fragments are agglomerated, according to particular confidence thresholds, with the aim of minimising errors. All segmentation machine-learning algorithms require ground truth data for their initial training phase – that is, annotated data distinguishing between membrane and non-membrane, or synapse and non-synapse.

proofreading: Process by which humans correct errors on the output of the automated segmentation. Errors comprise the most common false split errors (fragments belonging to the same neuron have been split when they should not have) and the rarer false merge errors (fragments from different neurons were merged when they should not have been).

annotation: Process by which crucial metadata for connectomics analysis, such as cell type, is added to neurons during or after proofreading. It requires expert knowledge.

Wellcome supports science to solve the urgent health challenges facing everyone. We support discovery research into life, health and wellbeing, and we're taking on three worldwide health challenges: mental health, global heating and infectious diseases.

Wellcome Trust, 215 Euston Road, London NW1 2BE, United Kingdom
T +44 (0)20 7611 8888, E contact@wellcome.org, wellcome.org