REPORT

BRAIN CONNECTIVITY WORKSHOP SERIES

SUMMARY OF DISCUSSION

FEBRUARY 5 – MARCH 31, 2021

NIH The BRAIN Initiative®

U.S. DEPARTMENT OF ENERGY | Office of Science

Preface

Co-hosted by the National Institutes of Health (NIH) BRAIN Initiative and the U.S. Department of Energy (DOE) Office of Science, this workshop series convened experts to discuss the state-of-the-art, opportunities, and challenges to generating comprehensive atlases of brain connectivity (i.e., wiring diagrams) of mammalian brains. Each workshop explored a facet of this ambitious goal: the biological significance of such an undertaking, methods of sample preparation, experimental modalities, data pipelines, and optimized use of the resulting product.

Comprehensive atlases of brain connectivity will dramatically enhance the capability of researchers to formulate and test models of how activity in brain circuits drives coordinated function and behavior. The knowledge gained will accelerate the development of the next generation of circuit-based therapeutics. However, because data have never been collected at the scale needed to characterize a mammalian brain, the full implications of such an undertaking remain uncertain, much in the way that the transformative effect of sequencing the first human genome could not have been imagined at the start of the Human Genome Project. Further, the field has already learned important lessons from previous brain mapping efforts in other species such as *Drosophila* and roundworms.

Mammalian brains, however, are orders of magnitude larger than the nervous systems of fruit flies and roundworms. The construction of high-resolution anatomical maps of mammalian brains will require new innovations in high-throughput automated imaging, high-performance computing, and data science, including artificial intelligence and machine learning. This effort will also require collaboration among scientists and engineers across many diverse disciplines and sectors to invent, refine, scale up, and democratize the new technologies and resources. Collaboration between the BRAIN Initiative and DOE, as well as other federal agencies and private partners, will be critical to achieving this goal.

NIH and DOE have a long and successful history of collaboration to advance their shared goals. As a recent example directly relevant to this workshop series, DOE hosted a roundtable in November 2020, during which experts in neuroscience, bioimaging, advanced computing, and data science articulated the needs of the neuroscience community, considered the limitations in current technology, and proposed opportunities for DOE to advance the community's capacity to develop a comprehensive atlas of brain connectivity. Moreover, DOE's engagement in this effort will help to advance its own mission in science, energy, security, and environment because the scientific benefits of mapping complex brains will extend far beyond human health.

This meeting summary was prepared by Rose Li and Associates, Inc., under contract to the National Institute of Neurological Disorders and Stroke (NINDS). The views expressed in this document reflect individual opinions of the meeting participants and not necessarily those of NINDS, the NIH BRAIN Initiative, or the U.S. Department of Energy Office of Science (DOE-SC).

DOI: https://doi.org/10.2172/1812309

Table of Contents

Executive Summary

The National Institutes of Health (NIH) BRAIN Initiative aims to accelerate development and application of technologies that show how individual brain cells in complex neural circuits interact at the speed of thought and action. Complete wiring diagrams of the mammalian brain will revolutionize the capabilities of researchers to formulate and test models of how activity within brain circuits drives coordinated function and behavior. A top priority for the BRAIN Initiative, motivated by the strategic guidance of the BRAIN 2.0 Working Group report, is to muster the resources and foster the collaborations needed to generate these wiring diagrams at the level of long-range projections ("projectomes") and synapses ("connectomes") in whole mammalian brains, including those from rodents, humans, and other large-brained mammals. Achieving the scale and resolution necessary to characterize and use these projectomes and connectomes will not only push the boundaries of current imaging and computational capabilities to generate faster, cheaper, and more scalable technologies, but also drive broad innovation in data science, artificial intelligence, and machine learning.

Collaboration among scientists and engineers across diverse disciplines and sectors is needed to invent, refine, scale up, and democratize new technologies and resources to reach these ambitious biological and technical goals. The mission of the Department of Energy (DOE) Office of Science is to deliver scientific discoveries and major scientific tools to transform our understanding of nature and advance the energy, economic, and national security interests of the United States. To that end, the NIH BRAIN Initiative partnered with the DOE Office of Science to convene five workshops in February and March 2021 for scientists and engineers to discuss the state of the art, opportunities, and challenges in mapping complete neural circuits. Each workshop explored a different facet of achieving whole-brain projectomes and connectomes: the biological significance of mapping mammalian brains, sample preparation methods, imaging modalities, data pipelines, and access to resulting products and technologies. A healthy and productive dialogue between NIH, DOE, and the community continues about potential strategies and approaches. "Success" will generate not only complete connectomics datasets with the resulting knowledge they provide, but also a legacy of widely available tools to make projectomic and connectomic studies routine. Thus, a flexible research program that is accessible to and inclusive of the breadth of the research community is needed to adapt existing tools and propel technology innovation. The following considerations emerged from the workshop series.

Significance of Mapping Complete Neural Circuits

A comprehensive map of an entire mammalian brain would dramatically enhance the capability of researchers to formulate and test models of how activity in brain circuits drives coordinated function and behavior. Projectomes and connectomes comprise images, segmentations, and reconstructions in databases that allow scientists to query cell types and their connections throughout the brain. To understand relationships between structure and function of neurons, circuits, and regions, individual axons must be traced from somas to all of their synapses in all their final destinations, and their postsynaptic targets must be identified—including aminergic,

peptidergic, or other types of axons that do not form discrete synapses. Ultimately, all glial cells and their processes will also be fully mapped. New knowledge will emerge regarding the global and local functions of specific circuits and individual neurons or glia based on connection patterns replete with divergence, convergence, and feedback loops. Emerging computational tools will drive the incorporation of structural circuits with functional (e.g., behavioral, physiological, genetic, and molecular) data. As the technologies scale, the hope is that obtaining full projectomes and connectomes across many individual brains within and across species will become feasible. Ultimately, application of the new knowledge and technologies to the human brain will accelerate the development of next-generation circuit-based therapeutics. In addition to sizable implications for human health, such an initiative would also pave the way for major advances in both conventional and novel computing strategies and technologies. Current computers are profligate in energy consumption and have poor performance at human-like tasks. New knowledge about how the brain works will undoubtedly help to inform the next generation of machine learning algorithms and potentially new physical hardware, helping to overcome these limitations.

Sample Preparation in Mammalian Whole-Brain Connectomics

Sample preparation is absolutely crucial for all downstream steps. Current sample preparation protocols produce high-quality fixation and staining of small subsamples of brain tissue. New approaches are emerging that could evenly preserve, stain, and embed samples as large as a whole mouse brain. However, profound challenges exist to develop technologies for lossless subdivision of the brain into manageable sample sizes. “Ultra-smooth” vibratomy may provide opportunities to solve this problem. Another area of considerable discussion centered on how to account for the potential variability in brain structure and connectivity across individual animals. This variance could be assessed by comparing relatively small but complete maps of specific brain regions across many individuals. This strategy is already under way in a subset of brain regions (e.g., neocortex, hippocampus, and retina) and will likely continue with available and emerging technologies. On the other hand, a complete connectome from a single brain will provide a cohesive map of connectivity, obviating confusion and information loss that may arise from linking smaller maps across pieces of brains from different individuals. When technology permits production of a whole mouse brain connectome, it should ultimately be obtained in a wildtype mouse to generate baseline data of the widest utility, against which variation can be compared.

Sample quality will also influence which imaging approaches are adopted. Many factors contribute to the acquisition of uniform ultrathin sections, as would be required for non-destructive imaging approaches involving transmission or scanning electron microscopy (SEM). Similarly, specific sample preparation approaches will influence the depth of reliable imaging within block faces if destructive surface ablation techniques are adopted (e.g., focused ion beam SEM or serial block-face EM).

Workshop discussions emphasized the use of complementary multi-modal and multi-scale approaches to aid in the interpretation of synapse-level circuits. These include but are not restricted to multiphoton, light-sheet, expansion, and X-ray microscopy. In addition, single-cell

transcriptomics, spectral connectomics, nanobody immunostaining, enhanced molecular genetic probes, and chemical labeling approaches could be integrated across spatial and temporal scales. Identical tissue preparations should preferably be imaged by different methods on multiple scales, beginning with volumetric imaging at lower resolution before proceeding to the highest resolution EM. Prior work in retina connectomics has successfully used immunogold antibody labeling directly on ultrathin sections to provide positive identification of various synapse-specific tags, such as for various neurotransmitters and related receptor proteins. This labeling could be done after imaging on select sections and could also help to follow identified axons. A challenge will be to develop sample preparation methods that retain brain ultrastructure while obtaining these complementary signals.

Evaluation of the quality of a sample preparation can only be secured after imaging, montaging, alignment, segmentation, annotation, reconstruction, and 3D visualization. Ultimately, metrics must be established that determine when staining is sufficient, when artifacts are minimized, and when speed of acquisition is optimized.

Imaging Modalities and Post-Imaging Pipelines to Map Whole-Brain Connectivity

An important open question concerns image resolution: What is the appropriate tradeoff of time and other resources for improved resolution? Image resolution should be sufficient to distinguish and trace the finest axon fibers, to identify the nature of connections as excitatory, inhibitory, peptidergic, spinous, dendro-dendritic, axo-axonic, reciprocal, and electrical synapses, and to distinguish them from other forms of adhesion junctions. One possibility is to couple rapidly-obtained, relatively lower-resolution imaging of large fields for cellular and large process tracing with higher-resolution re-imaging of any ambiguous areas from those same fields. Indeed, algorithms can be trained at high resolution (e.g., 0.5-2.0 nm pixels depending on the object) to recognize objects such as synapses in lower-resolution (e.g., 4-8 nm pixels) images. Thus, choosing a non-destructive imaging strategy would allow re-imaging to answer other specific high-resolution questions about synapse function from lower-resolution images—such as, for example, whether ribosomes are present, where microtubules are located, whether presynaptic vesicles are docked and, if so, how tightly, and whether tiny glial processes touch or surround a synapse.

At present, EM is the only approach with sufficient resolution to identify and characterize synaptic connections. However, despite great strides, all EM approaches remain prohibitively slow for large-scale studies. For example, current technologies limit the timeline for accomplishing the ambitious goal of imaging a whole mouse brain connectome to a 5- to 10-year period. Nevertheless, aiming for this goal will undoubtedly result in major advances in high-throughput tissue handling, EM and complementary modalities, and modern machine learning algorithms. In addition, opportunities abound to enhance analysis of those images to create connectomes. A useful compromise and starting point could be to image globally at 4 nm *x-y* and 40 nm *z* resolution to obtain knowledge about cell locations and large-process connectivity, and then layer on the local connectomes at higher resolution as they become available. Of course, a non-destructive approach would be required for re-imaging the same

sample at different resolutions. X-ray imaging can be performed on three-dimensional preparations at sub-micron resolution with large fields of view. The same sample can then be subsequently imaged at nanoscale resolution with EM. Thus, integrating EM-level resolution with light or X-ray microscopy and molecular or activity labeling may help to overcome some challenges associated with connecting tiny unmyelinated axons across long distances. A bolder opportunity discussed is the possibility to use EM data to train machine learning methods on optical or X-ray data to infer structure below the resolution limit. In the medium term (3-5 years), rapid technology development in X-ray imaging with the availability of million-times brighter storage rings, together with the continuing advances of optical imaging, could push down the resolution closer to that of EM. However, these tools are as yet unproven at the needed scale and resolution, and have many hurdles to overcome.

Workshop participants noted that a few nanometer-resolution datasets over a single whole mouse brain will produce multi-exabytes of raw data at a terabit-per-second rate. These are daunting but not impossible numbers, and the growth of computational processing power may be rapid enough to make such data rates routine in 5 to 10 years. Nonetheless, data compression throughout the pipeline will be important, particularly if large volume imaging “as a service” is to be an outcome of the project. The quality of the reconstruction process remains a primary challenge in the analysis of large connectomics datasets. New machine learning and artificial intelligence technology that accurately and automatically segments high-resolution image datasets aims to reduce manual segmentation and proofreading. Further development of automated methods for data collection, reconstruction, and analysis will help to mitigate bottlenecks in the pipeline. Such methods are generic for large-scale image and video processing and can be expected to have broad utility in data science.

Given the current state of the art, production of projectomes of myelinated axons spanning entire non-human primate and human brains are under way, and can be expected in the next couple of years. Based on great advances over the past decade, multiple groups are developing protocols for each aspect of the problem, including tissue clearing, staining, expansion, high-throughput light-sheet microscopy, and automated segmentation. These techniques have been applied specifically to axonal connectomics based on antibody staining, while other approaches, such as genetic barcoding, offer alternatives to achieve whole-brain projectomes.

Ultimately, the imaging modality and resolution required to attain meaningful wiring models will depend on the specific questions being addressed. Current technologies appear capable of obtaining small- and large-scale projectomes on practical time scales. Opportunities exist to develop systematic metrics that assess when the resolution and volume of the sample are sufficient to answer the questions posed and make new discoveries. Although some tradeoff between resolution and data rate exists, there are opportunities to independently increase both through combined research investment in multiple imaging modalities supported by data compression and artificial intelligence.

Optimizing the Use of Connectomic Data to Drive Data Science and Scientific Discovery

One of the most challenging aspects of large connectomics projects is scaling up current technologies and approaches to a whole brain—whether mouse, NHP, or human. The field of connectomics has learned important lessons from previous brain mapping efforts in model systems such as *C. elegans* and *Drosophila* (e.g., the full adult fly brain and the hemibrain), as well as smaller pieces of mammalian brain (e.g., the Intelligence Advanced Research Projects Activity MIcRONS project). These small connectomes have already had a transformative influence on neuroscience. Mammalian brains, however, are orders of magnitude larger than the nervous systems of fruit flies and roundworms. While the adult fly brain has about a 100,000 neurons and 100 meters of neuronal cable, the mouse brain has about 100 million neurons and 500 kilometers of neuronal cable. The resulting three-dimensional dataset from even one mammalian brain will be about 1 exabyte—thousands of times larger than previous endeavors. A human brain at the same resolution will be one thousand times larger (roughly a zettabyte, comparable to the total digital data stored globally in 2020).

The generation of intermediate steppingstone connectomes—of smaller regions of the mouse brain or in other species—between these dramatically different scales would facilitate solutions to the acquisition and analytical challenges for scaling to the whole mammalian brain. Scaling up will require building on the recent advances in high-throughput approaches and the ability to generate, store, and analyze large amounts of data. Experience with building data pipelines for the current petascale datasets has identified paths to scaling up, and projections of the growth in computing and data-processing power indicate that the analysis of a whole mouse brain at synaptic scale will be feasible. Smart data compression early in the pipeline will reduce large data volumes. Automated methods must be developed to minimize human interaction with the data for training sets and for proofreading. The data pipeline—collection, montage, alignment, and segmentation—must run synchronously. A segmented connectome can be much smaller than the raw dataset (which may be itself infeasible to store in totality) depending on the level of detail retained; that is a decision to come. In such a highly automated procedure, perfect accuracy is not attainable, and the level of accuracy (and indeed the measure of accuracy) will depend on the use of the data. Again, combining methods that link across scales will likely be more reliable than a single monolithic approach.

Analyzing a (set of) connectomes will require new mathematical tools and biological understanding. Assumptions from classical statistical methods may not apply; the level of heterogeneity is not understood, and even when complete graphs have been reconstructed, they must be annotated with spatial, functional, and transcriptomic data. There is seen to be enormous value in cross-species comparison, assessment of statistical fluctuations within species, and longitudinal developmental studies. At the end of the pipeline (i.e., following segmentation), data sizes are manageable, and comparisons between many individuals appear to be possible—notwithstanding the current paucity of user-friendly tools.

Parallel Efforts, Democratization, and Data Sharing

The anticipated massive imaging datasets will require immediate and widespread sharing and dissemination as they are collected. This goal presents an opportunity to develop new ways to manage the data so that scientists can assess the quality of the sample and images, improve segmentation and reconstruction algorithms, and rapidly begin to build partial connectomes. Because the volume of image data is very large, tools of "computation as a service" for widely accessible cloud-based data must be developed. Current resources include open-source libraries with a wealth of analytical tools. Two open questions are how to access the data repositories, and at what redundancy would researchers download or work online with small versus large portions of the data. Even with increased use of machine learning and artificial intelligence to segment, there is likely to be a continuing need for human annotation to provide training data and for proofreading. What level of metadata flexibility versus standardization will be needed? Parallel efforts to train the next generation of human experts may occur with training artificial intelligence to guide proofreading and "challenge" or correct human errors. Proofreading has been deployed in both expert-centered environments as well as large-scale citizen proofreading environments and provides a basis from which to design the next generation of these tools.

These efforts will inspire innovation, collaboration to assess metrics, and cooperation to settle on ultimate strategies. It will require support from the BRAIN Initiative and other federal agencies, private partners, academic institutions, and scientific and non-scientific communities. At present, high-throughput technologies capable of producing and handling the whole process of even a tiny part of a single brain region require unique state-of-the-art instruments and technical support. Ultimately, broader access to imaging devices and other connectomics resources would lead to competency in connectomic data collection for neuroscientists across the world. These goals may be reached through wide distribution of tools to many laboratories or via broad access to singular centralized facilities, or indeed likely some combination of strategies. Ultimately, access of researchers to instruments and data should not require physical proximity or control, and opportunities exist to adapt current research practices. Workshop participants noted that other big data scientific disciplines (e.g., cosmology) are transitioning to the cloud, and that their strategies for democratizing access could be a guide.

Future Perspectives

Many good ideas emerged during these workshops that will be useful for developing a strategic roadmap for pursuing a mammalian connectivity project at scale. Despite the complexity of obtaining a whole mouse brain connectome, it is clear that this ambitious goal can ultimately be achieved. Emerging technologies will be driven by scientific developments in which connectomics can both ride a rising tide and play a catalyzing role, such as in computing, information technology, artificial intelligence, and photon imaging. Addressing challenges particular to this project—in sample preparation, staining, handling, and EM (including approaches to data analysis)—will require the development of bespoke technologies that may have spillover advantages for other fields. The technological advances and infrastructure developed in pursuit of accurate projectomes and connectomes must make subsequent

acquisition and measurement of connectivity routine. The legacy of the new technologies would provide a framework that readily incorporates functional data. Furthermore, the aggregate effort of multiple scientific communities will facilitate rapid integration of new findings into the shared datasets in a transparent and publicly accessible environment.

This will be a large research program that requires integration across a broad community, involving new tools, new methods, and new researchers. While the scientific agenda arguably lies firmly within the remit of NIH and the National Science Foundation, the scale of physical technology needed, and the unique management required in conducting projects of this magnitude that engage distributed communities of researchers, make a strong case for DOE to be solidly engaged in the program from the beginning. It was noted that DOE's national laboratories are structured to enable large computing and large user facilities, and can build infrastructure so that any university can contribute to the program. The outcomes can be expected to have a positive impact on the distinct missions of all the agencies (and probably more, including the Department of Defense). Indeed, the history of interagency collaboration on programs of such national and scientific significance is a good one.

The access to a connectome has already been transformative to the understanding of neural function in the worm and fly communities. Now that a mammalian connectome is feasible, our goal might be to make it routine—the value of the illumination it will provide not only for neural architecture but also for longitudinal studies and development, for comparison between species, and for disease is likely to be profound. Let us not be limited by current technologies, but instead allow our imaginations to embrace new developments.

Workshop 1: Significance of Mapping Complete Neural Circuits

Precis

The first Brain Connectivity Workshop focused on the potential for connectomics to advance our understanding of a range of neural circuits. Participants were asked to address a set of questions: What issues would you be able to address if you had detailed connectome data that cannot currently be investigated? Are there conceptual issues that would be approached entirely differently if detailed connectomic data were available? What would be the value added between having one versus many samples—across individuals, pathological conditions, and species—and how many is enough?

A series of talks probed the opportunities provided by connectomics for studies of flies, zebrafish, songbirds, and the mammalian retina and basal ganglia as well as subcortical circuits involved in mating, sleep, feeding, parenting, thermal regulation, aggression, and other innate behaviors. A panel discussion explored the possibility and implications of a whole mouse brain connectome.

Recent advances in *Drosophila* connectomics have transformed fly research. The correspondence between genetically and functionally defined cell types in the fly brain is a critical component of this success. A challenge for the broader use of connectomes is that the catalog of cell types may be less complete and the connection to function more complex and mysterious. If this and other challenges can be met, connectomics is likely to revolutionize the study of neural circuitry.

Introduction

Larry Abbott, PhD, and Eve Marder PhD (Workshop Co-Leads)

One might attribute the dawn of the connectome era to early days when workers sought "simple" systems in which it was possible to set the goals of defining the neurons and their connections responsible for generating easily measurable behaviors. To this end researchers sought biological preparations that were easy to record, had identifiable neurons, and had reliable activity patterns that could be characterized in vitro and in vivo. Many of the fundamental insights into how circuits work came from invertebrate and lower vertebrate preparations, including insects, crustaceans, annelids, molluscs, lampreys, and frogs. In these systems, laborious electrophysiological and anatomical studies were done to "crack" the circuit and attempt to arrive at connectivity diagrams. By the 1980s several simple systems had first pass connectivity diagrams, and their significant differences led Peter Getting to develop his idea of building blocks to try to define common features of circuit design. This early work made it evident that without connectivity diagrams, it would not have been possible to develop many important concepts such as neuromodulation and circuit reconfiguration. Nonetheless, even by 1980 it was understood that the connectivity diagram represents only an initial step toward a full understanding of circuit function and dynamics.

Many of the lessons learned from the study of small circuits are relevant to the study of larger and more complex brain circuits. For example, the field must be able to (1) measure and characterize the behavior at issue, (2) identify the neurons involved, (3) determine characteristic neuronal firing properties, (4) determine the connectivity among neurons in sufficient detail, and (5) reliably manipulate the activity of identifiable neurons while measuring circuit dynamics. A core task for Workshop 1 participants was to determine the type of connectomic data needed and the level of detail needed for those data to inform heuristic understanding of circuit function.

Another group of questions relate to what happens as the scale and complexity of the system increase (i.e., when the entire brain and the body are also considered). Workshop 1 participants were also tasked with considering ways to fill the gaps in knowledge about the outside world and to best leverage the very detailed information that will be obtained about the inside world from connectomics.

To set the stage for Q&A and panel discussions, scientists presented on studies of the fly, zebrafish whole-brain connectomes, songbirds connectomes, the retina, subcortical and brainstem nuclei, and basal ganglia. Each presenter focused on (1) issues that could be addressed if detailed connectome data were available; (2) conceptual issues that would be approached entirely differently if detailed connectomic data were available; and (3) the value added between having one versus many samples—across individuals, pathological conditions, and species—and how many is enough. The presentations can be viewed by accessing the archived NIH VideoCast.

Presentations

Beyond the Lamppost: How the Fly Connectome Is Helping Us to See the Whole Picture

Gwyneth Card, PhD, and Gerry Rubin, PhD

The first image volume for the fly brain was the Full Adult Fly Brain (FAFB) female dataset, which was generated by serial transmission electron microscopy (TEM). Although the data exist to fully reconstruct neurons and their connections, that task has not yet been performed. With this dataset, users must manually trace neurons of interest. Nevertheless, it has been used by more than 50 laboratories, aided by subsequent machine-vision segmentation efforts. The *Drosophila* hemibrain female dataset was obtained with focused ion beam SEM imaging. These data were used to produce a connectome, defined as an organized and accessible map of neuronal connections. With a connectome, all paths between any two neurons that involve fewer than two interneurons can be obtained almost instantaneously. To produce a connectome from an image volume, as was done with the *Drosophila* hemibrain female dataset, the data are segmented using machine-vision to predict neuronal boundaries. The segmentation must then undergo a time-intensive manual proofreading and synapse identification process. Additional datasets are forthcoming, including the full connectome of

the ventral nerve cord of a male fly (i.e., the fly spinal column); the ventral nerve cord of the female fly is now available as an image volume.

Comprehensive knowledge of all circuit components empowers vastly improved experimental design and theory. Even in well-studied regions, new cell types and unanticipated connectivity patterns and classes of inputs have been discovered. Connectivity appears to be much better than morphology or gene expression for defining cell types. More information can be gleaned from EM images than neuronal connectivity and synapse number; neurotransmitters and other cellular features (e.g., mitochondrial position) can be discovered by applying machine-vision to the image dataset.

Knowledge of the connectome is transformative in the same way that knowledge of genome sequences transformed genetics, for several reasons. First, nearly all aspects of neuroscience become more efficient and cost-effective as powerful new approaches to experimental design, modeling, and theory are enabled. Second, although past theory and modeling provided conceptual ideas about how the brain might perform computations, the connectome is providing theorists with the ability to generate detailed mechanistic insights. Moreover, the connectome suggests incisive and tractable experiments to test those ideas by revealing which cells to manipulate, image, and record. Finally, modeling and theory based on the connectome have predicted that the fly brain is endowed with computational capabilities that have not yet been experimentally established.

One connectome (but not an image volume) is enough to realize a dramatic impact. Furthermore, comparison within and between individual flies suggests that variability is similar between the brain's hemispheres. Thus, comparing small brain regions from multiple individuals would be a prudent approach to explore these questions in other animal models. Importantly, cost reductions due to technological advances will likely make this issue of sample size largely irrelevant from a practical perspective.

Dr. Card explained that connectomes provide a critical resource that will allow individual researchers and the neuroscience community as a whole to move beyond analysis of circuits in small, discrete regions using current tools and toward engagement with the whole brain and nervous system. Dr. Card highlighted five takeaways from the fly: (1) Connectomes are democratizing; (2) Connectomes let us think on a brain-wide scale; (3) Connectomes enable a different approach to experiments; (4) Connectomes reveal new functional motifs; and (5) Connectomes bridge theory and experiment, speeding conceptual level insight.

Mapping the Zebrafish Whole-Brain Connectome

Florian Engert, PhD, and Claire Wyart, PhD

In Dr. Engert's view, the purpose of connectomics is to validate or invalidate circuit models derived from activity maps. He believes that connectomics must be a high-throughput process in order to collect connectomes for each animal used in an experiment, because this approach enables better anatomical and functional comparisons. Dr. Engert's team visualized blood vessels and glutamatergic and GABAergic neurons in the brain of a 6-day-old larval zebrafish

and, following euthanasia, subjected it to high-resolution electron microscopy (EM). The team used the EM, vasculature, and activity maps of this single animal overlaid into the same volume as proof of concept for using connectomics for circuit validation and individual neuron identification. In collaboration with Google, the team has begun to scale up this approach by designing algorithms that automatically segment the zebrafish brain's whole volume to reconstruct every individual neuron. Uses for these datasets include classifying neurons by cell type, creating network structures and motifs, and validating circuit models previously created with alternate models.

Connectomes may also enable experimentation on modulation of neuromasts (sensory organs within the lateral line) and their hair cells. EM volumes of neuromasts have already aided in labeling dopaminergic and cholinergic afferent neurons that presumably modulate them, changing existing hypotheses on how these organs function. Connectomes may provide additional fundamental insights into neuromast function.

Dr. Wyart used the vertebrate interoceptive system in zebrafish to illustrate a single-cell optogenetics and electrophysiology approach to linking connectivity to function that is complementary to Dr. Engert's imaging and connectome development approach. Many vertebrates have cerebrospinal fluid-contacting neurons (CSF-cNs) that resemble hair cells and are seen in the spinal cord. The question of whether and how CSF-cNs interact with the Reissner fiber, a long polymer thread present in the spinal cord, has not been easily investigated with earlier methods. Dr. Wyart's approach, however, has revealed that these sensory cells respond to spinal bending on the side of the spine with negative curvature; CSF-cNs extend toward the Reissner fiber in the central spinal canal. CRISPR/Cas9-generated mutants with no Reissner fiber do not exhibit a response to bending in their CSF-cNs. Targeted ablation of CSF-cNs that project to the hindbrain revealed that they control the power and speed of locomotion, changing the amplitude of tail bends during escape. Some of these neurons affect posture as well, and ablation of those cells prevents posture correction after animals are physically rolled. These posture-related cells appear to project specifically onto the occipital motor neurons and to command reticulospinal neurons expressing Chx10, which are conserved across all vertebrates. The transparent nature of larval zebrafish allows for greater experimentation on this system, and the results may be translatable to all vertebrates.

Songbird Connectomes

Michale Fee, PhD, and Sarah M. N. Woolley, PhD

Dr. Woolley's lab focuses on how neural activity in the songbird auditory cortex corresponds to self-produced song acoustics. Her team established a basic wiring diagram of the finch auditory cortex that is analogous to the flow of information in mammalian brains: information from the thalamus arrives in the intermediate layer, is passed to the superficial layer, and then to the deep output layer. The superficial layer projects to the song premotor area, and the deep layer has descending outputs to vocal production circuits. Single neuron recordings throughout the region revealed that activity in the deep output layer maps onto the acoustics of the finch's song. Because different species produce unique songs, this activity map varies dramatically

across species. Moreover, activity maps in finches taught by a finch of another species resemble the activity maps of the tutor bird, and as a result activity maps can vary *within* a species, as well.

Through this work, Dr. Woolley's team has determined the locations of excitatory and inhibitory neurons and has identified firing patterns, their relative proportions, and how tuning affects each type of neuron differently. However, to understand how selectivity for song features develops, how tuning for those learned acoustics emerges in this cortical circuit, and how the deep output layer is tuned to the songs while the input region neurons are not, researchers need to identify the inputs for each of these auditory cortex neurons and how they converge on deep output layer neurons. In short, they need a detailed connectome of the auditory cortex of finches.

Dr. Fee's team investigates the song production and vocal learning circuits in the avian cortical basal ganglia, which follow the deep output layer from the auditory circuit. Specifically, the team focuses on determining mechanisms underlying how a region called Area X influences a variability generator (LMAN) to produce the correct melody. The team hypothesizes that individual medium spiny neurons in Area X receive inputs from both LMAN and HVC, a region that controls song timing: HVC inputs terminate on dendritic spines, and LMAN inputs terminate preferentially on dendritic shafts. When LMAN activation depolarizes the entire dendritic shaft, coincident activation of HVC synapses causes the formation of a biochemical eligibility trace that correlates action and time. Subsequent arrival of dopaminergic input from the ventral tegmental area (VTA) selectively strengthens the activated HVC synapses. Afterward, the bird will experience activation of the medium spiny neuron from the HVC at the relevant point of the song, leading to reactivation of the correct LMAN neuron.

This hypothesis can only be tested using the principles of connectomics, because it fundamentally relies on determining the specific connectivity of individual neurons from distinct regions. Further hypotheses regarding how these circuits developed could be answered by developing connectomes at different stages of vocal learning.

Connectomics and Computation in the Retina

Stephan A. Baccus, PhD, and Rachel Wong, PhD

Dr. Baccus described the tight link between connectomics and computation in the retina. There are more than 100 subtypes of retinal cells, each of which conduct different analyses. In a simplified model of the retina, these cell types tile to form a mosaic of the entire visual field. The retina conducts data compression to unify that mosaic, including adaptation to various statistics such as movement, contrast, and patterns.

Convolutional neural network and machine learning models have accurately characterized visual computation for any arbitrary stimulus in the retina but can do so over only limited areas. These models are successful especially because of the repeating nature of retinal cells, but non-uniformities in the retina can also be programmed. The models correlate with retinal variability and, in addition to accurately predicting response, have units that are highly correlated with

Baccus's lab's interneuron recordings. The models therefore provide information about not only computation, but also structure. A model that uses a retinal connectome to incorporate individual cell type's specific biophysics would enable researchers to test theories of retinal function and determine mechanisms that produce visual computation.

Three specific approaches that might be enabled by a full connectome include determining which existing computational models are informative about anatomical structure; training a model from the outset with the connectome's anatomical data before presenting it with visual stimuli; and using known biophysical and structural data to build a model to explain emergent neural circuit function. Because deriving emergent function out of lower-level structures is very challenging, this third possibility would be a major advance.

Serial block-face SEM has enabled significant reconstruction of the retina, enabling classification of cell types and further differentiation of subtypes. Cell morphology and stratification of axons and dendrites generally suffice to differentiate cell types. However, in some regions it can be difficult or impossible to define and separate observed cells. Dense reconstructions of the retina can aid classification by applying known principles of retinal organization. Standard cell mosaics and arbor tiling also clarify individual cell types. Volume reconstruction methods have been instrumental in identifying less common cell types in the retina and have enabled researchers to identify circuit features underlying specific retinal functions. Connectomics have further revealed unexpected contact and synapse arrangement, including Dr. Wong's favorite example of the M1 melanopsin ganglion cell, which has arbors in the OFF sublayer despite receiving connections from ON bipolar cells.

Volumes that include connections within the inner retina have been reconstructed, but volumes encompassing the vertical connectivity from photoreceptors-to-bipolar cells-to-ganglion cells have yet to be collected. In addition, larger volumes are needed to capture horizontal connections involving long-range amacrine cells, which are not fully encompassed by the existing smaller volumes, and thereby to build a more complete circuit map. Although the retina is composed of repeatable units, those units can vary dramatically across different retinal locations (e.g., dorsal vs. ventral retina, central vs. peripheral) and exhibit different functional properties depending on peripherality or centralized location. More human retinal connectomes are also needed to provide deeper insight into the nature of circuitries that are unique to primates. Connectomes can also aid the design of strategies to repair neuronal damage in injury or disease by showing how retinal circuitry neurons and their connections remodel upon cell loss, or how transplanted cells integrate synaptically into the damaged environment.

The Mouse Brain Connectome Would Transform Understanding of Core Behaviors & Physiology

Catherine Dulac, PhD, and Scott Sternson, PhD

Subcortical structures have been understudied relative to cortex, despite much brain function and dysfunction being controlled subcortically. Dr. Dulac asserts that a brain connectome

applied to subcortical structures would transform understandings of core behavior and physiology. For example, subcortical regions' functions are anatomically intermingled; the preoptic area is subdivided into dozens of anatomically defined subnuclei, but each subnucleus contributes to a variety of functions, including mating, sleep, feeding, parenting, thermal regulation, aggression, and more. A connectome could elucidate the organizational logic within and between each anatomically defined structure as well as the connectivity of these subcortical regions and their arborization. By enabling a granular, cell type–specific understanding of the brain's wiring schema a connectome could improve and revolutionize understanding of subcortical function and control. Dr. Sternson spoke further about circuit mapping and connectivity between different types of neurons in the hypothalamus, with a focus on those involved with appetite and body weight control. Intermingled AGRP neurons (which promote appetite) and POMC neurons (which suppress appetite) in the hypothalamic arcuate nucleus have four theoretically possible connective relationships, both homotypic and heterotypic, although only one is supported by evidence. The evidentiary support was collected with inherently unscalable experiments demanding multiple transgenic lines.

An EM ultrastructure can provide cell biological insight into the wiring of modulatory neurons like AGRP and POMC and their peptide or transmitter connections. However, a focus on these cell-biological insights prevents a comprehensive view of connectivity and restricts the observable volume. To truly transform subcortical neuroscience, a comprehensive large-scale connectivity map showing complete synaptic connectivity interactions is necessary, which may be possible with EM ultrastructures or expansion microscopy.

In the future, machine learning predictors could use ultrastructural features to identify cell types. Cell type–specific connectivity measurements in subcortical regions would facilitate comparison across sexes, physiological states, or disease models. Pairing connectivity among cell types with detailed transcriptomes would enable identification of the molecular logic of cell-to-cell nervous system recognition and, furthermore, perturbation of those functions, which holds both clinical and basic science relevance.

Functional imaging during behavior could be integrated with molecular markers of connectivity to bridge the gaps among neuron dynamics, behavior, gene expression, and circuit organization. Replicates that enable basic statistical comparisons (Dr. Sternson suggests three connectomes per sex) and comparisons across physiological or disease states would be critical to full understanding of neurophysiology and provide an anchor for subsequent investigations into the mouse brain.

Basal Ganglia Connectome: Towards an Accurate Neural Circuit Model of Reinforcement Learning

Bernardo Sabatini, MD, PhD, and Ilana Witten, PhD

Basal ganglia research that traditionally links structure and function has led to the development of go/no-go, complementary, and competitive models of movement control. Connectomic analysis may distinguish among these models and would further aid in differentiating between

two propositions involving either massive convergence or highly selective innervation. The projection from the thalamus to the cortex in particular demands further connectomics research to clarify the organization of this leg of the movement control circuit. Proposals regarding microcircuits in the striatum, which explore roles for the region's few excitatory cholinergic interneurons, could distinguish among models of the basal ganglia, including Dr. Sabatini's preferred hierarchical hybrid model.

Dr. Witten believes that midbrain dopamine neurons in the basal ganglia are essential to reinforcement learning, and that a murine connectome could advance understanding of these neurons' computations. Dopamine neurons encode reward prediction error, responding not to fully predicted reward but to unexpected rewards or reward-predicting stimuli to produce reinforcement learning. A classic hypothesis for the mechanisms underlying reinforcement learning involves cortical encoding of environmental status projecting to the striatum or pallidum, followed by modification of synaptic weights by the dopamine reward prediction error, causing the relevant synapses to become predictive of reward likelihood. Three potential hypotheses relying on interneuron activity in different basal ganglia regions attempt to explain how the temporal derivative of reward prediction is determined. All three are partially consistent with existing coarse anatomical data. Distinguishing among them may be possible only with a full connectome.

Dr. Witten argued that a connectome could also reconcile recent findings that dopamine neurons' contributions to reinforcement learning are heterogenous. Dr. Witten hypothesizes that multiple dopamine neurons conduct the same circuit computation using different inputs from aspects of the larger environment. Anatomically, either all dopamine neurons converge at the target site, or each converges on different regions, leading different brain regions to conduct different predictions. This latter model conceptually reimagines reinforcement learning implementation.

Dr. Witten stressed that a single sample would be valuable in pinning reinforcement learning to higher-resolution anatomical data. She additionally speculated that assessment of connectomes across disease states or sexes could expand research in multiple fields.

Q&A Session

Characterization of Cell Types

Cell types can be defined in many ways. Dr. Rubin remarked that fruit fly researchers agree that clustering by connectivity is the single best method to define cell types in that model organism, because this method provides brain-wide information that enables researchers to categorize cells that are indistinguishable in other domains. These connectome-based definitions can then be refined by adding other features, including transcriptional profiles, electrophysiological activity, and morphology. Dr. Jayaraman underscored the value of connectivity information for identifying functionally distinct neuronal populations in the fruit fly central complex that are morphologically indistinguishable. Dr. Marder agreed, citing specific projection patterns as the most reliable way to identify cells in the stomatogastric ganglion (STG). Dr. Dulac added that

connectomics will enable researchers to investigate the nature of neuronal cell types as discrete or continuous populations.

Dr. Dulac emphasized the importance of considering the interplay between transcriptional profiles and projection patterns when clustering cell types, noting that neuronal populations as defined by *both* transcriptional profile and projection pattern have demonstrated functional similarity for some behaviors (e.g., parenting). Dr. Rubin agreed that more information across multiple domains is needed to characterize cell types, but noted that obtaining brain-wide transcriptional data is laborious and that transcriptional profiles become more similar across neurons as they mature.

In order to align a connectome with transcriptionally-defined cell types, Dr. Rubin posited that anatomical information (e.g., from fluorescent in situ hybridization) is essential. For species that lack anatomical information about the transcriptome at the single-cell level, Dr. Engert suggested an intermediate step that aligns functional imaging data to the connectome and then cross-references these data with EM images. Dr. Sabatini added that sequencing-based connectomics approaches currently in development (e.g., methods that employ synaptically-targeted barcodes) may be used to merge the transcriptome and the connectome via sequencing the connectome.

Opportunities and Limitations for Connectomics Approaches to Discovery

Dr. Rubin noted a tradeoff between speed and resolution in connectomics. In fruit flies, for example, current approaches to connectomics will not shed light on gap junctions, neuromodulators, or receptor localization. Therefore, it is important to recognize the limitations of the methods and technologies that are employed to build a particular connectome. Dr. Marder remarked that constructing the whole connectome may be viewed as an iterative process that proceeds with increasingly higher resolution to address this tradeoff.

Different connectomics approaches will produce different levels of insight into nontraditional synaptic structures, such as dendrodendritic synapses and gap junctions. Drs. Baccus and Dulac commented on the important functional implications that are specific to these structures, particularly for regions such as the retina and the hypothalamus. Dr. Sabatini cautioned against connectome-based approaches with narrow definitions of synaptic structure and function because they may preclude the discovery of fundamental and functionally-relevant building blocks for neural circuits. Dr. Wong added that synapses are complex structures whose definitions include molecular and spatial features that have not been fully characterized.

Integration of neuromodulators with the connectome is another outstanding challenge. Dr. Engert suggested that one approach to this complex task is to focus on individual neuromodulators and their contributions to circuit activity. Drs. Dulac and Marder somewhat disagreed and cautioned that an overly simplistic approach that is too narrowly focused may not adequately assess the host of neuromodulators that any given neuron will receive from a variety of inputs, thereby failing to reveal the bigger picture of modulatory systems within the connectome.

Plastic Connections

When constructing a connectome, researchers must consider the dynamic nature of neural connections. Individual synapses will vary in their relative plasticity or stability. For example, Dr. Jayaraman noted that some neurons in the fruit fly undergo relatively rapid changes to enable the flexible mapping of visual scenes and that the time scale of plasticity may influence its structural manifestation. Dr. Dulac added that some gap junctions in the suprachiasmatic nucleus will fluctuate according to a circadian clock, and therefore it is important to have multiple samples for particular connectome regions that correspond to different states. Similarly, Dr. Woolley commented on the dynamic nature of dendritic spines that are modulated by the estrous cycle of rats in some brain regions (e.g., hippocampus) but not in others (e.g., brain stem), and added that it would be useful to develop a metric for the expected degree of change at a given synapse based on the specific conditions (e.g., neurons, circuits, regions) at play in that synapse in order to set a tolerance level for a connectome based on the relative resistance of its connections to plasticity.

To explore the influence of learning on projection patterns from a connectomics perspective, Dr. Woolley suggested that species differences in learning can be used to expand the capacity to identify cell types, interactions, and physiological response properties. For example, she has leveraged the tonotopic mapping of the finch auditory system to selectively induce learning-related changes in some circuits but not others. This approach allows for the controlled comparison of portions of the connectome that have undergone learning-driven plasticity to those that have not within the same brain region of an individual animal.

Neural Circuit Modeling

Comprehensive connectome data enable the identification of fundamental organizational principles that underlie neural circuits. Dr. Card responded to an observation that mapping of mammalian primary visual cortex has revealed a projectome that bears similarity to organizational properties in the fruit fly. This observation underscores the power of connectomes to elucidate fundamental neural circuit architectures as well as facilitate cross-species comparisons. Theoretical models have been developed that propose the existence of some of these fundamental circuit features that connectomics may be able to confirm. Dr. Fee spoke to the value of leveraging these models to guide the identification of microscale circuit mechanisms within a connectome. For example, emerging computational models suggest mechanisms by which a songbird may process mismatches between the song it produced compared to the song of its tutor. Upon the collection of comprehensive connectomic data, researchers can search for connectivity motifs reminiscent of features in these computational models. Dr. Engert agreed that connectomes are important tools for testing and validating the hypotheses put forth by circuit models. Dr. Baccus concurred and remarked that models are a crucial resource for guiding the meaningful interpretation of connectome data.

The relationship between connectomes and models is bidirectional: as models can guide the interpretation of connections, so too can connectome data push researchers to rethink and reshape circuit models. Dr. Baccus explained that connectomes are used to shape and constrain computational models, such that the outputs of a model are weighted in accordance with

connectome data. It is important to consider how strictly a model should be constrained by connectomic data. For example, if a feature is missing from a connectome, then the model may be inappropriately constrained. As such, Dr. Baccus suggested that weighted—rather than strict—penalties may be applied to models that violate related connectomic data.

Functional Implications of Connectomes

Dr. Fee remarked that circuits are often more complex than current computational models would suggest. Various biological characteristics of a neural circuit will modulate its functional activity (e.g., the intrinsic properties of its component neurons, or the specific location of its synapses on the soma or neural processes). Dr. Marder expanded on this notion by suggesting that connectomics will help researchers to investigate parallel and polysynaptic pathways that models do not require to perform basic functions but that may nonetheless be critical to the biological circuit.

The structure of a connectome is influenced by any thresholds that are applied to it; for example, a synaptic count threshold will directly modulate the number of mapped connections between neurons. As such, Dr. Callaway remarked that a connectome does not portray the functional impact of its connections. Dr. Card reinforced this point, noting that a connectome is a starting point for circuit analysis that must be interpreted according to an appropriate threshold. Dr. Abbott suggested that a threshold of one to two synapses may be useful to prevent the mapping of erroneous connections. Dr. Wong added that the number of synapses may not be the only relevant constraint, noting that the number of synapses decreases as neural circuits mature.

Connectomes may offer varying degrees of insight into the functional properties of its purported circuits. Dr. Baccus suggested that connectomes may offer clues that suggest linear networks (e.g., fewer cell types or sequential stages), but that aspects of nonlinear processing (e.g., those governed by membrane biophysics) may remain difficult to identify from a connectomics approach. Dr. Abbott reasoned that nonlinearity may be suggested by differential targeting of synapses.

Discussion Panel

Panel Discussion Topic: How Important Is a Whole Mouse Brain Connectome?

Leaders: *Adrienne Fairhall, PhD, and Larry Zipursky, PhD*
Discussants: *Timo Bremer, PhD, Anne Churchland, PhD, Vivek Jayaraman, PhD, Jeff Lichtman, MD, PhD, Josh Sanes, PhD, and David Tank, PhD*

During this portion of the agenda, discussants presented their views on the topic, commented on the presentations by the earlier speakers and other discussants, and considered the questions posed in the online Q&A.

Dr. Bremer suggested that the questions and challenges posed by the connectome effort may be aided by a number of traditional mission-critical areas of DOE, where Dr. Bremer is a computer scientist. At a basic level, DOE's work on the complex networks that compose the

national power grid is structurally homologous to neuroscientists' efforts to map complex networks in the brain. DOE has thus built expertise in techniques that are highly relevant to mapping the connectome, including management of very large images and analytic approaches on both the small scale (e.g., segmentation, feature detection) and large scale.

DOE can also help the whole mouse brain connectome endeavor by bringing together scientific communities to enable successful solutions. DOE has expertise in coordinating large integrated multidisciplinary teams that are essential to a project of this magnitude. Its significant computational resources can potentially ease the transition from work by individual research teams and small groups with only local resources to a unified large-scale effort that leverages the latest computing technologies in hardware and software.

While DOE can aid the whole mouse brain connectome in various ways, the effort has potential advantages for DOE, as well. It promises a number of long-term benefits for computation-driven sciences, including insight into how the human brain learns, develops, assesses risks, processes uncertainty, and makes decisions—all of which are of interest to DOE. These insights speak to the potentially vast lessons of the initiative for society more broadly.

Dr. Churchland and her team at the Churchland lab study sensory processing and decision-making, with the objective of optimizing the interpretation of key measurements (e.g., identification of inputs and outputs for every neuron). Central to this work is the acquisition of sufficient information to arrive at a sensible interpretation for the purpose of pooling measurements from related cells. However, the intermingling of different types of neurons creates challenges for determining neuronal identity. Although tools that label cells according to their projection target can help identify neuronal types, these tools are in the early stages of development. Dr. Churchland acknowledged that defining cell types is difficult but suggested that developmental lineage is one relevant factor.

The question of whether multiple connectome samples are needed is also relevant to Dr. Churchland's work because in complex animals such as mammals, connectomes will vary across individuals as well as across neurons within an individual. Ideally, the connectome of every animal would be known during an experiment or post hoc reconstruction. Short of this ideal, information about the connectomes in a small number of mice would be useful. Nonetheless, the neuroscience community must bridge the gap between general knowledge of connectomics and specific knowledge for a particular experiment. From Dr. Churchland's perspective, a large gap remains between the current reality and the long-term goal.

Finally, Dr. Churchland advocated for the importance of considering multiple approaches to obtain the connectivity information desired in larger brains. Although a satisfactory technique for animals with small brains or to obtain connectivity within a specific region of a larger brain, EM reconstruction to obtain connectivity in a larger brain is time intensive and not always feasible. Thus, although EM reconstruction is very informative, it must be bolstered by sequencing-based approaches, which are more suitable for investigating long-range connections.

Dr. Jayaraman's perspective on the whole mouse brain connectome is largely informed by his lab's efforts to understand navigational circuits in a highly conserved fly brain region called the central complex. The lab has used physiology and cell type–specific perturbation in behaving animals in conjunction with connectomics to explore the dynamics and function of well-defined recurrent circuits in the middle of the brain. His perceived value of a whole (versus partial) mouse brain connectome is based on two assumptions. First, this question should be evaluated within the framework of understanding circuit function in the context of behavior. Second, a "connectome" is not only an imaged and segmented volume or a piecemeal approach, but also a queryable database that is comprehensive and sufficiently organized such that synaptic connectivity can be extracted across large networks. Based on these assumptions, Dr. Jayaraman suggested that it would be a mistake to fixate exclusively on an ambitious whole-brain connectome because partial connectomes that are complete for specific brain regions can be very useful. These partial connectomes not only serve as technological stepping stones, but also provide functional insight if additional complementary information is available.

A great deal of information can be extracted about function of the sensory periphery based on the connectomic information alone because the inputs are known, which in turn provides a starting point for working through the system. Although additional information about neuronal biophysics is needed to address the temporal transfer functions, substantial information can be extracted without it. However, these questions become more difficult to address for deeper brain regions because the inputs are often unknown and neuronal processes may be severed. The motifs in deeper brain structures can be interpreted more readily if information about inputs can be obtained by labeling the key projection neurons types in EM, matching morphology with light level data, or following a similar approach.

If the goal of the connectomic effort is to build cumulative knowledge about a system, then reproducible, targetable cell types in different species must be defined. Much of the progress of Dr. Jayaraman's lab in densely recurrent circuits is rooted in the known function in some set of neurons. The connectome is then applied to this functional knowledge. The foundation for this approach is a working hypothesis based on findings from previous studies and the ability to reproducibly identify specific cell types in the fly. An understanding of the circuit function is also needed to ground this approach, even if that understanding is a speculative conceptual model or a quantitative model.

Based on these assumptions, a connectome of only one brain region is itself transformative. This regional connectome enables constraining and testing implementation-level models, although the connectome often reveals many more connections than do the models. Indeed, the identification of new connections helps to generate novel and testable hypotheses, which is one powerful outcome of having the connectome. Another benefit of a region-specific connectome is that it can serve as a foundation for the analysis and decoding of population data. There is a current trend to collect data from large numbers of cells, but to analyze those data as if all of the cells are roughly equal. However, different cell populations project to particular post-synaptic partners, and some lines never converge, suggesting functional differences. Having such connectivity information from the connectome thus informs data

analysis and extraction that is biologically relevant for the system. In the same vein, features such as mixed selectivity can be demystified by simply identifying inputs.

Dr. Jayaraman explained that his advocacy for the benefits of a partial connectome does not imply that the advances enabled by a whole-brain connectome would not be tremendous. Moreover, although large-scale connectivity at low resolution is useful when no background information is available, higher resolution at small volumes can be very informative when determining the relative importance of inputs and synaptic weights.

Dr. Lichtman recently co-authored an article in *Cell*, titled "The Mind of a Mouse," that elaborates on the importance of a mouse whole-brain connectome. As that article notes, much of a neuron's function is based on its pre- and post-synaptic connections. However, no complete set of inputs and outputs for any single mammalian neuron have yet been mapped. The mouse whole-brain connectome would create the first such map by cataloging the connections among nearly 100 million neurons within the mammalian brain. In Dr. Lichtman's view, this connectomic profile will be a more functionally relevant descriptor of any given neuron than the information offered by that neuron's transcriptome.

Because the mammalian brain sends and receives information from a wide variety of regions, Dr. Lichtman suggested that connectivity maps of piecemeal parts of the brain would be insufficient to generate a complete mammalian connectome. Obtaining all neural connections and projections in the same animal will close the gap between mesoscale projection maps and nanoscale synaptic circuits and will enable the construction of a complete mammalian brain circuit.

Although completing a 1 exabyte–scale whole-brain connectome will be extremely challenging, the knowledge generated by this effort may create new research trajectories. For example, this initiative could pave the way for comparative neuroscience research based on various animal species' development, sex, aging, and disease, including more complex animals such as NHPs and humans. For the study of disease, one mouse whole-brain connectome enables the identification of connectopathies—diseases in which connections are abnormal—and may ultimately spur the creation of a new branch of neuropathology. A connectopathy approach may be especially illuminating for common but presently incurable and poorly understood chronic brain disorders, including autism spectrum disorder and schizophrenia.

A mouse whole-brain connectome would also illuminate the synapse-level structure of long-term memories in mammals, an area where highly stereotyped non-mammalian nervous systems are less informative. Moreover, the connectome would continue progress toward designing Nonbiological Thinking Systems based on biological insights. For example, the whole-brain connectome could help produce a blueprint for (1) cognitive computing systems with general intelligence (in comparison to most more narrowly trained artificial intelligence); (2) data-efficient learning (e.g., one-shot learning, obviating the need for millions of samples needed to train a machine learning algorithm; and (3) energy-efficient computing, given that brains can perform computations that supercomputers require megawatts to complete.

Finally, beyond these specific advantages, Dr. Lichtman argued that the whole-brain connectome can yield unanticipated positive technological consequences, such as potential advances in the ability to solve problems based on vast amounts of exabyte-scale data.

Dr. Sanes remarked that dense reconstruction of brain tissue provides valuable insights and agreed with Dr. Lichtman's comments regarding the importance of the mouse whole-brain connectome as a target for this endeavor. Dr. Sanes focused on the challenges and potential drawbacks of reconstructing a whole mouse brain in the near future by highlighting a set of conceptual issues and underscoring the highly repetitive and modular structure of the mouse brain.

Dr. Sanes cited three concerning ways in which the connectome and genomes differ, which were first delineated in an essay that he co-authored with Dr. Lichtman more than 10 years ago. First, connectomes are vastly more complicated than genomes and therefore are more difficult to obtain, analyze, and interpret. Second, while genomes are extremely similar among individuals of a species—and especially similar for individual animals from inbred experimental lines—connectomes are quite variable. The mammalian connectome is particularly heterogeneous across individuals because many connections are refined by activity and experience. Third, whereas genomes change only minimally throughout the lifespan, connectomes undergo substantial changes, even in adulthood.

While these reservations are relevant to any connectomic project, Dr. Sanes argued that as the cost and scale of a project increases with limited resources, greater scrutiny of the risks and benefits that accompany those projects is needed. Thus, it is important to consider whether a whole-brain reference connectome is truly feasible relative to its potential utility.

In assessing the benefit of the whole-brain mouse connectome, Dr. Sanes commented that many regions of the brain, such as the barrel cortex and retina, have a highly repetitive structure. In the mouse barrel cortex, for example, Dr. Sanes suggested that the field would benefit from the reconstruction of one, or perhaps four, whole barrels, but that it is less clear whether scaling up to reconstruct the full barrel cortex connectome relative to the cost is beneficial.

Dr. Sanes emphasized that reconstructing a whole mouse brain would certainly be desirable if resources were unlimited. Given limited resources, he instead advocated for funding connectome efforts for a wider variety of species, brain regions, and developmental stages, as well as constructing connectomes from multiple individuals, genotypes, and conditions.

Dr. Tank is an advocate for the construction of the mouse whole-brain connectome. Theorists have long considered the value of large connectomes, and recent work has revealed a tight link between structure and function. The connectedness doctrine—which states that connections between neurons largely define neuronal computation—has been essentially taboo in the field of circuit neuroscience until this point. Yet the connectedness doctrine is manifest in the success of current AI and deep learning technologies. Thus, Dr. Tank contends that

connectomes are a necessary component to achieving a mechanistic understanding of neural circuits, although they may not be sufficient.

Dr. Tank shared an example from his lab to illustrate the potential utility of a mouse connectome. By imaging the activity of grid cells in the entorhinal cortex of awake behaving mice, his team showed that grid cells of the same phase are organized topographically on the surface of the entorhinal cortex in a hexagon-like lattice structure. Furthermore, neurons with different phases demonstrated a geometric correspondence to their relative phase. This elegant finding may be solidified in rigorous theoretical circuit models, which would then allow researchers to explore whether this two-dimensional attractor circuit is analogous to, for example, ring attractor–like circuits in the central complex of the fruit fly.

Similarly, researchers have recently discovered and characterized novel forms of dynamics in the mammalian brain (e.g., choice-specific sequences, rapid replay of sequences, and ripples in the hippocampus). Although conceptual and theoretical models for these types of dynamics exist, they are unconstrained by real connectomic data and an evidence-based understanding of the underlying circuits. Knowledge of the circuits that underlie these dynamics is essential, just as it is for understanding the ring attractor in the fly, learning in the basal ganglia, and the structure of connections in the songbird HVC.

Finally, the ultimate goal is not solely to generate a mouse connectome; it is also to enable connectomics at the exascale, which is a 1,000-fold leap from the field's current technical ability. The mouse connectome is an intermediate step for a field that hopes to advance beyond the mouse to do connectomics of even larger structures such as the human brain. Until then, a connectome for a cubic centimeter–size volume derived from the mouse can be applied to many other species and to regions of the human brain in different pathological conditions.

Follow-up Discussion

Benefits of This Initiative

Dr. Bremer reinforced the point (made also by Drs. Lichtman and Tank) that we do not yet know what technological advances and other impacts may emerge from a "moonshot" approach to the mouse whole-brain connectome. It is difficult to predict the downstream ramifications for AI, computer chip designs, and understanding of structure-function relationships. However, he believes the knowledge gained from this effort will lead to immense technological advances, both anticipated and unanticipated.

Dr. Tank envisions potential benefits unfolding in parallel. As the technological advances needed to complete the whole-brain connectome are developed, those emerging technologies will also benefit other areas of brain research, similarly to the advances in genetics research that emerged from the human genome project. Experiments that currently take 6 months to 1 year to complete could be conducted on vastly shorter timeframes of days to weeks. Advances such as this would have widespread implications for the neuroscience field beyond the specific outcome of a single mouse whole-brain connectome.

Dr. Jayaraman addressed one of Dr. Sanes' comments about the types of changes in the mouse brain connectome that might occur over the lifespan of a mouse or appear between different mice. He emphasized that although changes will occur at the synapse level (i.e., synapse numbers and weights), the types of insights that can be extracted from connectomes lie at the level of motifs and topologies, much of which will be pre-built, genetically prespecified, and highly conserved. He thus suggested that these intra- and inter-individual differences do not diminish the value of generating a single whole-brain connectome.

Computational Capabilities

Dr. Churchland noted that this effort will require major advances in data storage and computing. She asked whether existing technologies such as the Open Connectome Project, which have devoted substantial resources to store and compute data, are sufficient for a project of this scale, or whether new resources would have to be built in a potentially open-source way.

Dr. Lichtman remarked that in addition to national labs, private entities such as Google are working on the computational issues required to conduct exascale computing. Surpassing current computational boundaries is in fact one of the main goals of the connectome project. In particular, scientists seek to improve techniques of storing and handling data so that these factors are no longer a limiting problem.

Dr. Bremer further emphasized that the computational tools developed for the whole-brain connectome can enable solutions to other problems. Such broadly applicable technological advances have resulted from previous long-term DOE efforts to solve computational challenges for analogous large-scale projects. For example, this endeavor will generate not only one whole-brain connectome, but also the toolchain capable of generating it. These advances in tool chains and processes might enable other labs to complete a connectome for a piece of brain tissue that is orders of magnitude larger than previously possible, even if not every lab will be able to generate a whole mouse brain connectome.

Cost/Benefit Analysis

Dr. Lichtman countered the arguments made by Dr. Sanes and others that the mouse whole-brain connectome would be too expensive, and that its cost would not outweigh its benefit. He proposed that the democratization of a connectomic dataset on this scale would pay for itself almost immediately. Moreover, he argued that the expense of generating the connectome for one whole mouse brain is probably less than the amount of money that has already been expended mapping brain connections using projections, which does not even fully construct single neurons.

Dr. Sanes reiterated his view that with unlimited resources—including money, time, and equipment—generating a whole-brain connectome would be worthwhile. He questions, however, whether the best value can be obtained by completing a single mouse dense reconstruction rather than adopting other suggested approaches, such as analyzing more species, regions, and developmental stages, many individuals (to study variability), multiple genotypes and disease models, and a range of other conditions.

Dr. Churchland suggested that the mouse connectome initiative could take many forms, all of which may dictate different allocations of resources. For example, the initiative's scope could range from one reconstruction of one mouse brain to reconstructions from multiple brains under different conditions using a combination of high-resolution and lower-resolution techniques.

Dr. Lichtman responded that the neuroscience community at large has proposed generating one whole mouse brain connectome at the resolution of synapses, so that every projection generates a connection. That means that for the first time in a mammal, the complete wiring input and output of all of the 70 million neurons in one animal will be available. At this point, this wiring diagram has not been obtained for even a single neuron.

Dr. Tank noted that although he shares Dr. Lichtman's optimism about this effort, he recognizes the hurdles left to overcome. For example, steps such as staining or sectioning an entire brain, imaging the brain, and stitching together images are all immense challenges. Moreover, he suggested that even proponents do not expect funding for a mouse whole-brain connectome to be immediately available. Research will need to occupy an intermediate stage during which technology pipelines are developed to the point of being practical and working well at scale.

Workshop 2: Sample Preparation in Mammalian Whole-Brain Connectomics

Precis

This workshop focused on current methods, opportunities, and open issues in sample preparation for (1) whole mouse brain EM volumes; and (2) complementary light microscopy (LM) imaging of the brains of mice and larger organisms, with a focus on NHPs and humans. The potential utility of additional complementary data, from tissues such as retina, olfactory bulb, peripheral nervous system, and whole body from the same or different animals, was also raised.

Sample preparation for whole mouse brain EM remains an unsolved problem. The field is, however, making rapid progress. Open issues include (1) uniform, high-quality staining of the whole mouse brain with EM-dense metals; (2) high-quality embedding of the whole mouse brain in resin; and (3) lossless subdivision of the embedded brain for imaging either by destructive or non-destructive methods (e.g., serial section TEM or gas cluster ion beam (GCIB) multibeam SEM, respectively). High-quality results were presented for staining and embedding of an early postnatal mouse brain (6x smaller by volume than the adult), and preliminary data were shown indicating that a similar process will scale to the adult. Progress has also been seen in near-lossless subdivision via serial thick “hot knife” sections (1 micron or less, for larger samples; 20 microns for fly brain), although this has not yet been demonstrated on samples as large as the whole mouse brain. An approach was proposed to achieve “ultra-smooth” vibratome sections, which would allow fine neurites to be traced across section boundaries. This would enable direct application of existing diverse sample preparation and volume EM imaging methods, at the scale of the whole mouse brain.

Numerous opportunities and diverse methods exist for complementary sample preparation and LM of mouse brains and larger brains, although they do not provide complete connectomic information. In the mouse brain, complementary imaging could be done within the same animal as whole mouse brain EM, and/or with separate animals, to associate cell type or functional information with the connectome. Complementary approaches discussed included multiphoton, light-sheet, and wide-field microscopy, with large fields of view and high scan rates; classical tracer injections, performed at scale; clearing and embedding of samples in elastic hydrogel for better antibody access; viral vectors; X-ray microscopy; EM-dense and correlative light and EM (CLEM) chemical and genetic labels; immuno-EM labeling at the surface of vibratome-cut sections; and axonal projectomes. The diversity of potential options indicates both the need for careful planning and coordination, and the huge opportunities available, to maximize the impact of these efforts.

Briefly summarizing the way forward, the field’s efforts should include multi-scale correlative microscopies, functionalized microscopies with tagging and barcoding to provide point-to-point long-range connections, and kerfless sectioning to improve alignment and segmentation accuracy.

Introduction

Davi Bock, PhD, and Hongkui Zeng, PhD

The goal of Workshop 2 was to identify both established methods and workflow gaps in sample preparation for (1) whole-brain EM connectomes of mouse, (2) complementary LM methods (on the same individual or different animals) in conjunction with the EM connectomes, and (3) whole-brain LM connectomes/projectomes of larger organisms such as humans and NHPs.

Much can be learned from whole mouse brain connectomes and dense reconstructions from those connectomes. First, researchers can learn about the rules of organization at both the cell type level (e.g., cell type diversity) and at the network level (e.g., network properties and brain-wide cell interactions). Second, researchers can learn about structure-function relationships by building connectomes of functionally integrated brains to understand the underlying connectivity patterns and rules, and their relationships to the functional properties of different cells, pathways, and networks. Further, the connection properties obtained from the connectome can be related to cell types for targeting using molecular tools and then further monitoring and functional manipulation in living animals.

Regarding the rules of brain organization, the field should identify the best definition of cell type: Do cells belonging to different types have different connection properties, and do individual cells within a type have the same connection properties? These questions should be addressed for the thousands of cell types in the brain. At the network level, it is important to investigate the divergent and convergent connections within and across brain regions at different levels of hierarchical connections to learn about higher order network rules. These network patterns strongly influence functional activities of individual cells and can be shaped by plasticity and learning. To identify the patterns and rules of these networks and the possible existence of subnetworks, single axons must be traced from starter cells to their terminal synapses, and multiple cells converging onto the same target cell must be traced simultaneously in order to understand their relationship. These requirements argue for whole brain–level dense reconstruction.

In order to describe local and long-range subnetworks and their structures across brain regions, the field should coalesce around a concept of “complete.” Sidney Brenner has stated that large-scale endeavors, such as this, should strive for complete, accurate, and permanent (CAP) datasets. A whole mouse brain dataset would likely lead to a van Essen diagram across all brain regions at synaptic resolution. However, the brain does not exist in isolation, leading the research community to question whether the sample preparation pipeline should be geared toward complementing the whole mouse brain dataset with additional datasets such as retina, spinal cord, peripheral nervous system, and the whole body.

Three possible experimental scenarios could influence the specific approaches used in the connectome endeavor, and conversely, the outcomes: (1) integration of cross-modal datasets within the same sample, (2) integration of multi-resolution datasets from within the same sample, and (3) integration of datasets across samples using different protocols. For all of these

scenarios, data quality is a function of the entire pipeline: perfusion, fixation, staining, embedding, sectioning, imaging, alignment, and segmentation, many of which involve sample preparation. Therefore, the design of the sample preparation pipeline must reflect the overall plan for cross-modal labeling and imaging within and between samples.

Session 1: Continuous Whole Mouse Brain EM Connectomes

For Session 1, the speakers were asked to address four questions. The presentations can be viewed by accessing the archived NIH VideoCast.

1. What current sample preparation methods for whole mouse brain EM connectomics maximally conserve structural continuity across the entire brain? What are the pros and cons for each approach?
2. All currently available imaging methods suitable for whole mouse brain connectomics require subdivision of the sample prior to imaging. What amount of loss is expected with current subdivision methods? Can continuity at single axon (~50 nm) level across the entire brain be maintained?
3. What are the prospects for improved methods to reduce loss during subdivision, either prior to resin embedding (e.g., vibratome sectioning) or after (e.g., hot knife)?
4. Prior to imaging, how to assess the quality of the sample preparation? How to predict and/or validate continuity before, during and after imaging?

EM Staining for Whole Mouse Connectome

Xiaotang Lu, PhD

Volume EM, the gold standard for reconstructing connectomes, will likely be the approach used for the whole mouse brain connectome project. Dr. Xiaotang Lu described three specific requirements for adequate EM staining for the whole mouse brain connectome. First, membranes must attain higher contrast than cytosol. High membrane contrast greatly facilitates the identification of cell boundaries, a crucial first step in the segmentation process during EM reconstruction. Second, staining must be sufficiently intense to enable reasonably fast image acquisition times; this can be achieved by obtaining high metal levels in the tissue. Third, the staining methods must be scalable to large brain volumes. The scale-up needed for EM imaging of the whole mouse brain connectome is the most challenging of these requirements.

The most commonly used tissue staining protocols for volume EM—ROTO, ORTO, and BROPA—use “two-round enhanced osmium staining,” which entails three basic steps: first osmication, ligand binding to improve contrast, and second osmication. Osmium is used to create contrast in both scanning and transmission EM images, and two-round enhanced osmium staining approaches provide high membrane contrast. These protocols differ in the implementation of several key steps required for optimal staining.

Each protocol has pros and cons for achieving staining suitable for the whole mouse brain connectome project. ROTO, the most widely used protocol, offers the best membrane contrast and staining intensity; however, it does not provide usable staining for large brain blocks because staining beyond 200 µm in depth becomes light and muddy. The ORTO protocol was developed to increase the staining depth. In this protocol, osmication and reduction are separated into sequential instead of concurrent steps; this proved to be crucial for increasing staining penetration and contrast. ORTO's main drawback is its use of thiocarbohydrazide (TCH) as the linker molecule, which leads to crack formation in large tissue blocks. The BROPA protocol maintains the improved staining depth of ORTO but decreases crack formation by using pyrogallol as a linker molecule instead of TCH, as well as formamide in the first osmication step to increase osmium penetration. BROPA can achieve brain-wide staining, but its contrast is inferior to that of ROTO or ORTO. Thus, none of these methods simultaneously achieves contrast, intensity, and scalability.

Dr. Lu and colleagues are developing a new two-round enhanced osmium staining protocol—ORPO—to meet all three staining requirements of the whole mouse brain connectome project. X-ray micro-computed tomography (micro-CT) is being leveraged as a key tool in the development of ORPO because it enables monitoring of tissue during the staining process, evaluation of the effects of protocol modification, and rational optimization. Micro-CT thus lifts a major barrier to protocol development, because historically it has been impossible to assess tissue quality for EM until the final block is cut and imaged.

Several key adjustments to reaction conditions have been critical to the success of the ORPO protocol thus far. For example, the first osmication step was separated from the reduction step by an intervening wash to facilitate removal of unbound osmium and increase the efficiency of the first osmication. Moreover, reaction conditions were optimized by including a buffer, instead of water, in the osmication and dehydrations steps to prevent osmotic damage to the tissue. This protocol has been optimized for P0 mouse brain and is currently being scaled to the adult mouse brain. As this process is scaled up, such attention to detail and rigorous optimization will be even more important to achieve the high quality of tissue staining required for the whole mouse brain connectome.

Sample Preparation in the Context of a Large-Scale Connectomics Pipeline

Nuno Maçarico da Costa, PhD

Dr. da Costa discussed optimizing sample preparation in the context of a full pipeline. Although sample preparation is absolutely crucial for all downstream steps, success of the entire project is only measured and evaluated during data analysis—the final step of the pipeline. Thus, histology must be optimized in the context of the end product and what is considered a measure of success for the whole project.

In this context, several keys steps for optimizing sample preparation were identified. (1) Success must be defined. For example, in the Intelligence Advanced Research Projects Activity (IARPA) MICrONS project, success was finding chemical synapses between all neurons in the volume within a specified timeframe necessary for that project. (2) Sample preparation must then be

optimized to achieve that specific definition of success and the end-to-end results of the entire pipeline. For example, although post-staining increases sample contrast, its lack of reliability may pose too much risk. (3) The protocol must yield consistent and reliable results, which increases the predictability of the process and decreases risk. (4) Once these steps have been achieved, the entire pipeline should be iterated as early and often as possible to identify and correct points of failure. (5) Backup plans for pipeline components should also be developed to ensure enough diversity of approach to be robust to unknowns. (6) Finally, samples should be systematically evaluated before selecting a production sample for sectioning.

Dr. da Costa stressed that consistency and reliability are critical for scalability. The sample preparation protocols selected for the MICrONS project ensured that most of the animals were successful and that within animal samples appeared very similar. The staining protocol was based on ORTO, which was the most reliable protocol in the researchers' hands, supplemented by parts of the BROPA protocol. This combination allowed for a reliable and consistent pipeline for the MICrONS project. Moreover, this robust infrastructure was subsequently leveraged to produce a collection of datasets for different mouse brain regions and species.

Two approaches were used to assess sample preparation quality for the MICrONS project. First, there was a systematic ultrastructural evaluation of samples that were processed identically (i.e., same days and same solutions) as the production samples. Second, evaluation of the production sample staining using micro-CT-assisted staining was especially valuable because it enabled visualization of steps where the protocol was failing as well as early recognition of the sample's insufficient quality to proceed. Continued development of more efficient quality assessment protocols would allow more time for optimization of the sample preparation itself.

Blocking protocols required to prepare tissue for EM lead to some inevitable loss of tissue between blocks, which raises the issue of how to bridge tissue gaps between adjacent blocks. Dr. da Costa asserted that, in addition to strategies for reducing the amount of actual tissue that is lost, computational approaches may be developed to bridge tissue gaps and produce continuous reconstructions. These approaches must be guided by a detailed understanding of the morphologies of different neurons. Understanding of these well-defined features could also be used to help bridge the physical loss of tissue between sections. The completed IARPA MICrONS mouse dataset provides a large library of neuronal morphologies and can also be leveraged to digitally remove sections of a specified thickness (e.g., 100 nm or 1 µm) to analyze the amount of tissue loss that still allows for bridging and recovery of neuronal reconstructions.

Specimen Preparation and Screening for Volume EM

Eric Bushong, PhD

The Ellisman group in the National Center for Microscopy and Imaging Research (NCMIR) prepares mouse brain samples to optimize ultrastructural preservation for volume EM imaging. Its staining procedure results in highly preserved ultrastructure details, including intact, smooth, high-contrast membranes and intact organelles (e.g., mitochondrial cristae) and cytoskeleton that are well-captured in block-face imaging—setting the gold standard for well-preserved and well-stained mouse brain.

Transcardial perfusion with aldehydes is the most critical step to preserving ultrastructure. Several important details are essential for optimal primary perfusion fixation. First, the brain should be exposed to formaldehyde-glutaraldehyde fixative as quickly as possible, which can be achieved by a brief but efficient clearing of blood with an oxygenated and heparinized flushing solution. Importantly, all solutions used during perfusion fixation should have the correct osmolarity, and the buffering system should contain divalent cations ($CaCl_2$) to stabilize membranes and avoid membrane fusion artefacts. Because aldehydes do not fix cell membranes, the brain should be chilled immediately after perfusion (prior to extraction from the skull) to stabilize membranes and reduce membrane fluidity, and kept cold through at least the first osmication step, when membranes are fixed.

For optimal post-fixation, the brain should be incubated with glutaraldehyde at 4°C to help prevent tissue damage caused by the harshness of osmium fixation. Because the dehydration process can also strongly influence membrane ultrastructure, it is best performed as cold as possible (e.g., on ice). Although typical freeze substitution protocols for fixation and dehydration do not stain tissue heavily enough to enable imaging at high vacuum or preserve genetics probes (e.g., APEX2), the CryoChem freeze substitution protocol developed by Dr. Bushong and colleagues overcomes these limitations and yields well-preserved tissue compatible with EM imaging, including intact myelin sheaths, smooth membranes, and easily detectable synapses.

The optimal staining protocol is heavily influenced by the EM imaging method; block-face imaging, thin sections on substrate (for serial section TEM or SEM), and hybrid approaches (GCIB-SEM) have different requirements. For en bloc staining with heavy metals prior to serial block-face SEM imaging, Dr. Bushong and colleagues developed the ROTO protocol to provide optimal staining for visualizing ultrastructural details. The primary limitation of ROTO is that specimens of up to only a few hundred microns thick can be stained well, although modifications of some steps, as in ORTO and BROPA, increase tissue penetration to allow staining of specimens up to 1 mm thick.

X-ray microscopy is used by NCMIR to screen specimens before volume EM imaging to ensure that staining is uniform, sufficiently intense, and free of defects or artefacts (including holes and cracks in tissue). With typical laboratory-based X-ray microscopy, imaging with sufficient resolution (< 1 μm) to assess ultrastructural preservation is not yet possible. Recently, however, X-ray holographic nano-tomography was used to obtain pixel resolution down to 20 nm, suggesting the potential to assess tissue ultrastructure before moving to EM. X-ray microscopy can also be used to precisely target regions of interest to limit volume EM to specific areas and to correlate volume EM datasets with other imaging modalities, such as confocal and in vivo 2-photon (2P) imaging.

Ideas for Thick Sectioning Brain Tissue

Kenneth Hayworth, PhD

To put into context the need for high-quality staining and embedding, Dr. Hayworth asserted that GCIB-SEM is one viable route to achieving the connectome of a whole adult mouse brain. GCIB-SEM is a hybrid approach that combines serial thick sectioning and EM block-face imaging. The approach entails cutting plastic-embedded mouse brain into relatively thick sections, which are then individually block-face imaged using the multi-beam SEM and broad ion beam milled. The key advantage of such a hybrid approach is compatibility with high-throughput multi-beam SEM imaging. Another important advantage is that it decouples the z-resolution that can be obtained from the actual section thickness; the imaging resolution can be set to whatever voxel size is needed for reliable automated tracing (typically 10 nm^3) and the section thickness to whatever provides the most reliable sectioning and collection (typically 100 to 1,000 nm). Thus far, a prototype GCIB-SEM system that employs single-beam SEM has successfully been used to image brain tissue. Current efforts are focused on integrating the process with a 61-beam Zeiss MultiSEM, with the goal of demonstrating precisely the type of fast wide-area automated imaging that would be required for a whole mouse brain connectome project.

If GCIB-SEM with multi-beam SEM were used to image an entire mouse brain, crude estimates of the imaging requirements suggest that in order to achieve the whole-brain connectome in 5 years, 100 GCIB-SEM microscope systems would be required for typical levels of tissue staining, while only 10 GCIB-SEM microscope systems would be required with the best possible tissue staining to date. These estimates illustrate that the staining level of the tissue is a critical factor in determining the economic and logistical feasibility of the entire project. Other key factors include obtaining uniformly good resin infiltration and optimizing the resin for reliable semithin section and quality ion milling. Thus far, the necessary stain intensity and infiltration quality have only been achieved in vibratome sections less than 1 mm thick, and such specimen quality is not easily reproducible. There are two possible paths to overcoming these limitations. The first is to stain and embed the whole intact brain prior to sectioning. The second is to use “ultra-smooth” vibratome sectioning after aldehyde fixation but before osmium staining.

The BROPA protocol is currently the only approach that produces sufficiently intense staining of a whole mouse brain for this project. However, while specimen quality is very good in the periphery, the tissue and staining level decreases dramatically in deeper brain tissue. Thus, while the BROPA protocol and more recent work show great promise as potential approaches to the whole mouse brain connectome, the limits of diffusion make staining an entire adult mouse brain extremely difficult.

Dr. Hayworth advocated for a parallel effort that does not rely on the success of whole mouse brain staining. This parallel effort entails vibratome sectioning of fixed brain tissue into 1 mm slices prior to staining and infiltration. The slices would be imaged separately and computational methods leveraged to stitch the images together for the final volume. However, the main challenge of using traditional vibratomy is the tissue damage produced at cut surfaces, which might make it impossible to reliably trace cells across tissue boundaries. Ultra-

smooth vibratomy may be the solution to this problem. Images of vibratome slices of glutaraldehyde-fixed tissue suggest that such well-fixed tissue should be sturdy enough to remain intact at cut boundaries, and tissue ripping and surface tearing may be alleviated by using an extremely thin and sharp blade with precise and well-controlled vibration. Dr. Hayworth suggested that blades made from 20 µm thick single crystal diamond wafers and sharpened by ion milling might overcome current limitations in blade thinness with mechanical polishing. Several additional parameters would also need to be optimized for ultra-smooth vibratome sectioning, including the tissue-holding mechanism, cutting speed, side-to-side oscillation, and the maximum level of tissue fixation compatible with good quality staining.

Lessons from "hot knife" sectioning to obtain ultrathick sections of plastic embedded brain tissue may inform the optimization of ultra-smooth vibratome slicing of the whole mouse brain. The use of a heated and oil-lubricated diamond knife makes it possible to section plastic-embedded tissue to a thickness of 20 µm and to achieve a very smooth cut surface such that surfaces of serial sections can be matched and computationally stitched back together to enable tracing of neuronal processes across gaps between sections.

Session 1 Q&A

Brain Tissue Staining

Several technical considerations related to the staining of brain tissue influence tissue contrast, block brittleness, and staining reliability. Very intense staining to produce high contrast can make sample blocks brittle and difficult to section. Dr. da Costa and colleagues considered using post-staining in their process to mitigate this issue; however, the lack of reliability and risk of potential damage by debris outweighed the potential gain. Nonetheless, Dr. Lu noted that the Lichtman lab post-stains samples to increase contrast and reduce dwell time, although the contrast achieved with one round of osmication followed by post-staining is never comparable to that from two rounds of osmication followed by post-staining. Dr. Bushong remarked that post-staining with lead often risks introducing precipitates and artefacts on the sample. Carbon dioxide–free environments—including carbon dioxide–free solutions and instruments—may improve staining consistency. Moreover, post-staining alone cannot solve contrast issues, and, in Dr. Lu's experience, two rounds of enhanced osmium staining always improves imaging. The contrast achieved with one round of osmication followed by post-staining is never comparable to two rounds of osmication followed by post-staining.

Dr. Lu further advised that enhanced osmium staining does not necessarily increase brittleness of the tissue specimen if osmium is deposited specifically to the membranes. Dr. Bushong noted that the embedding resin also influences brittleness. The field has used the same resins (Epon and Durcupan) for decades and would benefit from identifying novel embedding media that cause fewer problems with brittleness.

Staining techniques may also alter tissue properties and brain morphology. In Dr. Lu's ORPO protocol, X-ray micro-CT revealed that brain volume expanded by 10-20 percent during the reduction step. Although small gaps in the tissue may form during expansion, the brain shrinks

back to its normal size during the second osmication and dehydration; subsequent EM suggests that ultrastructural morphology is not permanently altered.

Although perfusion could be used in some of the tissue staining steps, this method would require a large amount of osmium tetroxide and osmium might block perfusion. Moreover, Dr. Lu does not consider the first osmication to be the limiting step; rather, reduction and, more significantly, second round osmication are limiting steps. Recent protocols that employ a series of solutions to large samples for light-sheet imaging raise the possibility that perfusing with multiple solutions could also be done for EM tissue preparations.

Brain Fixation

Perfusion fixation may cause changes in morphology, ultrastructure, and molecular localization. Exposure of brain tissue to low temperatures during fixation can alter the concentrations and localization of export and reuptake pumps and neurotransmitters used to identify cell types, collapse microtubes, or alter synapse morphology. Dr. Bushong explained that his perfusion fixation approach is designed to avoid such problems. Specifically, the initial perfusion solutions are introduced at 37°C and the temperature gradually decreased over the 10- to 15-minute perfusion period so that the brain is completely chilled when removed. This perfusion protocol is intended to sufficiently fix the brain with aldehydes before it is cooled. The preservation of good ultrastructure details, such as microtubules and dendritic spines, supports the effectiveness of this approach.

Other potential approaches to brain fixation could be pursued. One technique, developed in the late 1970s, entails perfusing the brain at high pressure with sucrose solution, then slowly replacing the sucrose with fixative to preserve the extracellular space. However, this approach may transiently expose the brain to hypoxic conditions, which can alter ultrastructure. Another option for preserving morphology and ultrastructure is to use high pressure freezing to quickly preserve brain tissue before anoxia sets in, which has revealed interesting differences in the glia compared to classically fixed tissue; however, this approach is not currently feasible for the whole brain. These alternative approaches offer many advantages if they can be shown to preserve brain tissue closer to its native state than current techniques.

Tissue Sectioning

Dr. Haywood discussed the conditions that influence the ability to cut tissue samples using the diamond knife, ion milling, and hot knife techniques. It is relatively easy to cut plastic embedded samples to yield very smooth surfaces because the tissue is held together by good resin infiltration. In contrast, sectioning of aldehyde-fixed tissue before introduction of osmium—as outlined in Dr. Haywood's speculative ultra-smooth vibratomy approach—will be difficult because the tissue is much more fragile. However, an ultrasonically vibrating diamond knife could decrease the forces to almost zero as it cuts through the fixed tissue. Dr. da Costa remarked that, given the reliability of sample processing for 1 mm tissue slices, these improvements to vibratomy could solve many of the issues with sample processing and is a worthy pursuit for the mouse whole-brain connectome initiative.

The ion milling approach imposes additional criteria for resin embedding, and some resins perform better than others. If the resin infiltrates effectively but nonetheless fails to ion mill well, then the overall processing pipeline could be disrupted. Dr. Haywood noted that he previously used the hot knife technique to optimize section thicknesses for focused ion beam. Although the hot knife approach could work for the whole mouse brain, it would impose extra requirements on consistency, the type of embedding material, and staining levels to avoid brittleness, which Dr. Haywood does not view worthwhile unless the goal is ultrathin sectioning.

For serial block-face imaging, a diamond knife is used to repeatedly scrape off the surface of the tissue block between sequentially acquired images. One drawback to this method is that the size of the area that can be scraped is limited by the size and rigidity of the diamond knife, which may prevent scaling to larger tissues. Dr. Mark Ellisman proposed the use of an ultrasonic diamond wheel as an alternative to a diamond knife. Dr. Haywood further suggested that use of an abrasive method to grind down the tissue surface, such as a diamond plate, might also obviate difficulties posed by diamond scraping. Dr. Ellisman added the possibility of grinding with a propellor and imaging between the blades. Dr. Hayworth stressed that the grinding tool cannot be moved more than a few µm back and forth unless its surface flatness is perfectly matched (on a nm scale) to the actual surface of tissue.

Is the Mouse Whole-Brain Connectome Feasible from the Sample Prep Point of View?
Finally, the speakers shared their views about whether the mouse whole-brain project—from the sample preparation point of view—is feasible, likely to be achieved with enough effort, or unlikely to be achieved. Drs. Lu, Bushong, Hayworth, and da Costa all believe that the mouse whole-brain project will likely succeed. The significant advances realized by Dr. Lu and the BROPA protocol strongly suggest that mouse whole-brain staining will be achieved. Dr. da Costa cautioned that several engineering and other issues remain to be solved and commented that the community should remain open to new approaches, such as the ultra-smooth vibratomy approach, and that parallel approaches will be crucial for success of the overall mouse whole-brain project.

Session 2: Complementary Whole-Brain Imaging in Mouse and Larger Species

For session 2, the speakers were asked to address four questions. The presentations can be viewed by accessing the archived NIH VideoCast.

1. What types of complementary LM data should be collected on the same brain and/or on different brains to help with processing and/or interpretation of the whole mouse brain EM connectome? E.g., projectomes at population or single cell level; functional imaging in behaving animals before EM; LM/X-ray/EM of central nervous system (CNS), peripheral nervous system (PNS), and/or whole body; cell type–specific labeling/tagging; immunolabeling.

2. Why is such information useful, what questions can it be used to address? What is the priority and order of such data generation in conjunction with the whole mouse brain EM connectome data generation?
3. Which techniques can be extended/scaled to larger brains (e.g., human and NHP brains)? Will these techniques require pre-labeling in live tissues?
4. What considerations are needed in sample selection (e.g., sex, strain or race, age, individual variation, health status, life history) and sample preparation (e.g., technical requirements, desired data types from the living brains)?

Scalable Approaches for Functional and Structural Light Microscopy of the Mammalian Brain

Elizabeth Hillman, PhD

Dr. Hillman addressed the utility of performing functional imaging in behaving animals before building the connectome. She presented a range of new imaging technologies in development in her lab that enable the visualization of dynamic cellular activity in three dimensions (3D) and described how these technologies can be used to investigate a range of functional questions.

One approach for high-speed 3D in vivo microscopy, known as SCAPE, uses an oblique scanning light sheet to enable 3D imaging at much higher speeds than point scanning. SCAPE microscopy enables the capture of dynamics of cellular activity (e.g., via genetically encoded calcium indicators such as GCaMP6f) and connectivity in superficial layers of the cortex in a way not possible in 2D. Dr. Hillman's lab has also developed a 2 photon (2P) version of SCAPE, which enables imaging deeper into the mouse brain (e.g., cortical layers 2 and 3) and visualization of not only the dendrites, but also entire neurons.

Dynamic 2P data are usually analyzed using spatiotemporal unmixing. This approach correlates pixels with a calcium signal of a particular time course to predict signals belonging to the same cell. Data from SCAPE can thereby be used to render entire cells and to extract the different calcium signal time courses derived from the soma, dendrites, and axonal projections. Because segmentation is based on timing, these calcium signals can only be extracted using a real-time 3D imaging technique. SCAPE can also be used to analyze neuronal activity while an animal performs tasks or under different stimulus conditions and to examine structure-function relationships.

Wide-field optical mapping is an approach to multi-scale imaging in which the ensemble activity of neurons across many brain regions can be captured by expressing genetically-encoded calcium indicators across large populations of cells. With this approach, the skull is thinned to provide a view of the entire surface of the cortex, which enables recording of brain-wide activity of the cells expressing the calcium indicator. For example, wide-field optical mapping led to the first visualization of high levels of activity in the somatosensory cortex of a running mouse, including signals that corresponded to the movement of individual paws. These data can be used to create segmentations of the cortex and derive a functional atlas that is analogous to a structural atlas of the brain. Functional connectivity can also be analyzed by

correlating the time courses from each active region, much as functional magnetic resonance imaging (fMRI) measures functional connectivity as synchrony between distant brain regions. These analyses enable the determination of brains regions that remain highly correlated versus those that change over time. Advances in viral transfection methods (such as those described by Dr. Viviana Gradinaru) can be paired with this approach to measure the activity of different types of neurons simultaneously.

Dr. Hillman also addressed the use of LM to map the brain structure and cell identities. For these types of questions, light-sheet approaches will be vastly better than point scanning approaches. Although increasing the speed of point scanning decreases the time available to integrate individual pixels, light-sheet imaging can integrate signals over a large area simultaneously. Moreover, light-sheet imaging entails less physical movement and stitching, lower photodamage, greater efficiency, and a higher signal-to-noise ratio. Dr. Hillman's group has demonstrated the potential of light-sheet imaging for acquiring blocks of 3D data across large areas of the mouse brain, that is, the complete projections of these cells can be visualized and traced over long distances.

With SCAPE imaging, a full mouse brain can be acquired in about 30 minutes—compared to 34 hours with resonance imaging. Importantly for a human brain, SCAPE can be used to image samples that are 5-6 mm thick (because the lateral size does not increase image acquisition time); therefore, an entire human brain can be imaged in about 3 days at 3 µm resolution.

Brain Connectivity in Primates

Helen Barbas, PhD

Dr. Barbas studies connectomes in large brains from NHPs and humans from the systemic to the synaptic levels in specific brain areas of interest. This approach enables the study of pre- and postsynaptic features and pathways and their functional interactions across various types of neurons.

Dr. Barbas employs a range of approaches to study large animal connectomes, including EM, double- and triple-labeling, and tracer injections. Pathways and connectivity are first analyzed by light and confocal microscopy, followed by EM and serial reconstruction of small brain regions of interest. Pathway labeling is carried out in live animals using tracer injections. After perfusion fixation and cryoprotection, the brain is frozen and cut into matched series, which enables the analysis of multiple features in comparable brain regions. Individual series are processed to simultaneously view pathways and postsynaptic sites using double- or triple-labeling. For EM, regions are double- or triple-labeled to visualize pathways and postsynaptic sites at the ultrastructural level. Matched series of sections also enable the comparison of a range of microarchitectural features (e.g., relative myelination, cytoarchitecture, laminar features, receptors) within and between brain areas. The group's overall approach follows a connectivity roadmap based on a theoretical structural model that relates cortical architecture to the patterns and strength of connections.

Matched series also enable comparisons of analogous regions between NHP and human brains, supporting investigation of distinct features in human psychiatric and neurological diseases. For example, parallel analyses of myelination patterns in prefrontal areas in human and rhesus macaque brains showed that intra-species myelination patterns are not equivalent across these cortical areas and that analogous regions between humans and monkeys exhibit parallel myelination patterns. Dr. Barbas and her team have also analyzed axonal pathology in post-mortem brains from humans with autism and have shown that there are fewer thin and medium-sized axons below the anterior cingulate cortex in brains of adults with autism compared to control brains. Moreover, the axons that are present connect with nearby areas of prefrontal cortex, supporting clinical data that suggest overconnectivity within prefrontal cortex and diminished connectivity with distant brain areas in individuals with autism.

Dr. Barbas also addressed considerations and information needed for sample selection for analyses of primate and human brains. Important factors include age, sex, species (for NHPs), health status, and postmortem interval (for humans).

Towards Holistic Phenotyping and Understanding of the Human Brain

Kwanghun Chung, PhD

To understand disease mechanisms and develop new therapeutic strategies for many neurological disorders, researchers must study the brain at multiple scales, from the whole brain to individual molecules. Moreover, they must study the complex interactions among various molecules, cell types, circuits, and pathological factors following holistic rather than reductionist approaches. However, extracting multi-scale molecular and anatomical information from the human brain remains a major challenge. Dr. Chung's team is developing a full technology pipeline for scalable and holistic imaging and phenotyping of the human brain that enables extraction of multi-scale information in a fully integrated manner. The resulting platform of technologies aims to allow the extraction of spatial, molecular, morphological, environmental, and connectivity information from the same brain at single-cell resolution.

A major goal of this endeavor is to develop a technology platform that is compatible with the tens of thousands of banked human brains worldwide. These brains cover different sexes, races, ages, and disease conditions, and many have decades of rich functional and medical data associated with them. These brains are one of the most valuable resources available for studying neurological disorders; yet, they often sit on shelves and are poorly utilized. The development of a cost-effective and scalable technology platform would enable routine mapping of these valuable brains.

To achieve these goals, Dr. Chung's team has developed a set of novel and transformational technologies that engineer the physicochemical properties of tissue. CLARITY, which makes tissue optically transparent and macromolecule permeable, has proven valuable for light microscopic imaging of brain tissue. SHIELD preserves molecular and anatomical information in tissue, such as fluorescent labeling and protein antigenicity, under harsh conditions. MAP linearly expands tissue four-fold, which enables multi-scale imaging down to the level of axonal connectivity and chemical synapses. ELAST turns brain tissue into a rubber-like hydrogel that is

reversibly stretchable and compressible. Together, these technologies preserve critical anatomical, morphological, and cellular information in human brain specimens, which can then be accessed, captured, and analyzed.

By combining the unique properties of ELAST with the other technologies it has developed to engineer molecular interactions and transports, Dr. Chung and his team can significantly increase the scalability of their approach. For example, the depth of tissue staining is limited by the thickness of the tissue. However, an interdependence exists between the penetration time and length scale, such that a 5-fold increase in length leads to a 25-fold increase in time (i.e., d^2). In tissue prepared with ELAST, however, a 5-fold compression of the tissue decreases the time required for staining by 25-fold. Thus, ELAST enables transient compression of the tissue and much more rapid and uniform staining. Once the tissue is stained, compression can be reversed, and the tissue imaged in its native state. Using this approach, Dr. Chung's team stained 5-mm-thick slabs of human tissue within 1 day, which is 100 times faster than it previous demonstrated using CLARITY.

Engineered Gene Delivery Vectors for High-Precision Broad Coverage of the Mammalian Brain

Viviana Gradinaru, PhD

The introduction of genetically encoded fluorescent indicators to label neuronal circuits in primates requires non-transgenic methods for gene delivery. Unlike the mouse, which has many transgenic driver lines, significant challenges are related to the delivery of genes and photons to image specific cell types and circuits for large-scale brain connectomics in NHPs and humans. The goal of genetic labeling for cell and circuit mapping is to achieve sparse golgi-like genetic labeling, which is a major challenge for the NHP brain. Approaches to overcome this issue require the systemic delivery of adeno-associated viral vectors (AAVs) that cross the blood–brain barrier.

In the absence of transgenic access to the brain, vectors that encode fluorescent protein labels can be delivered directly into the brain tissue, often by direct injection. However, challenges to this approach include the small area and high density of labeling, which complicates reconstructing morphology and circuit mapping. Moreover, scaling to the larger NHP brain requires a substantial number of invasive injections, each of which produces dense labeling and a lack of homogeneity.

Recent efforts have focused on devising strategies to engineer better methods for labeling circuits across the mammalian brain using either rational design of gene delivery vectors or directed evolution. Both approaches entail the engineering of viral capsids that cross the blood–brain barrier to achieve genetically targeted access to large neuronal circuits. The engineering of only the capsid allows flexible access to and labeling of neural circuitry (e.g., the ability to label few or many cells), and transgenic tools grant access to larger areas as well as regions that are difficult to target by direct injection.

Capsids can also be combined to simultaneously label multiple circuits with different fluorescent proteins by using different gene regulatory elements and transgenics to enable visualization for tracing morphology or connectivity mapping. However, multi-color labeling of different cells types at a high density can cause overlap among cells that are the same color, creating challenges for morphological studies and neuronal reconstructions. To overcome this issue, a two-component vector approach can be used to deliver a transgene that encodes multi-color fluorescent proteins under the control of an inducer, and the density of multi-color labeling controlled by inducer dose. This design enables the fluorescent labels to be turned on in a sparse set of neurons in a way that generates different colors, which is particularly useful for morphological studies and neuronal reconstructions.

To obtain macroscale connectomes of the NHP brain, gene delivery requires cell type–specific methods that do not rely on transgenesis. Directed evolution can be exploited to generate a great deal of diversity that can be mined for patterns, such as families of related sequences or tropisms, which provide mechanistic insights to help guide cell type–specific access across strains and species. Capsids can be engineered to bias tropism toward different cell types. High-throughput methods (e.g., single cell RNA-seq, in situ transcriptomics) can be used to break down these properties to target sub-types of cells. Connectomics in NHPs also carries ethical considerations, including the obligation to maximize the number of traceable, genetically distinct circuits for each NHP brain. Logistical considerations related to seasonal breeding, housing costs, and feasibility associated with scaling up of labeling reagents present additional hurdles.

Projectomes and Connectomes from Mice to Primates

Bobby Kasthuri, MD, PhD

Dr. Kasthuri considered how projectomes can be merged with connectome data to provide complementary connectivity information at multiple scales. For certain insights into the brain, imaging bigger is better; questions that relate to how different regions of the nervous system are connected to each other might require fine-grained resolution. The efficient tracing of the projectome could provide a roadmap for exploration of information processing in the brain, and the projectome of a single control adult animal could ultimately be compared to animals across ages and disease states. Such a mesoscale projectome could provide information about brain connectivity that is complementary to that obtained by a higher-resolution whole-brain connectome.

Dr. Kasthuri and colleagues are developing a universal pipeline for merging projectomes and connectomes across species that leverages in vivo magnetic resonance imaging (MRI), ex vivo MRI, synchrotron X-ray tomography, and automated serial EM. The basic sample preparation workflow for the pipeline is to perfusion fix with aldehydes, stain brains with osmium and embed them in plastic, and then perform synchrotron X-ray tomography and automated serial EM. The strategy of combining X-ray tomography and serial EM enables imaging of the brain over six orders of magnitude. X-ray data provide a mesoscale image of entire brain regions and—at higher magnifications—cell bodies, dendrites, and myelinated axons. Such X-ray data

can be compared and co-registered with structural MRI on the same sample. In addition, the combination of X-ray tomography and serial EM, called X-ray microscopy, allows researchers to address both long-range and short-range connectivity questions.

This approach to combine analysis of projectomes and connectomes can also be applied to primates. Thus far, researchers have learned that for every type of neuron and connection, primate neurons receive many fewer synaptic connections than mouse neurons on both the soma and dendrites. This unexpected result portends future insights into synaptic connectivity throughout the brain. Current efforts focus on addressing challenges to utilizing this approach in larger brains, including subdividing brains to minimize tissue loss, staining sub-volumes, performing X-ray tomography on each volume, and performing EM in selected brain regions.

An important potential application of large volume connectomics is the elucidation of how human brains develop to produce the unique cognitive characteristics of individual adults. Variations in the connectome across individuals are not a potential artefact, but the goal. Ultimately, this work would help to elucidate how lifetime experiences shape the function of the adult brain.

For the analysis of any small volume of brain, there will be about 100,000 neurons and 1 billion synapses. Identification of the parent neuron will only be possible in about 20 percent of these synapses. This lack of comprehensive information about synaptic connection will make it extremely difficult to correlate the activity of neurons in a sub-volume to their overall connectivity. This issue should be addressed when thinking about the differences between collecting small samples and large samples for the mouse whole-brain connectome project.

Chemical Tools for CLEM and Color EM

Stephen Adams, PhD

Dr. Adams spoke about chemical imaging tools for color EM and their visualization at both the fluorescence and electron microscopic levels using CLEM. In early versions of CLEM, diaminobenzidine (DAB) is photooxidized by singlet oxygen produced by excitation of a photosensitizing fluorescent dye or genetically encoded fluorophore, which generates a localized polymer laid down by the tag that stains with osmium. This additional osmium staining is visualized by EM. However, control of this reaction is limited, and few fluorescent proteins are photosensitizers. In addition, this method only generates additional osmium staining, in contrast to color EM. As an alternative method, a peroxidase (e.g., HRP or APEX) can be used to produce the precipitate; here, the peroxidase oxidizes DAB to generate a precipitate that is also stainable with osmium. However, peroxidases are not fluorescent, so they require tagging with a fluorescent tag for CLEM. One advantage of this approach is that light is not needed for these reactions.

A toolbox of genetic labels has been developed for CLEM. One of these labels is HaloTag-JF570, a small molecule photosensitizer/fluor dye that can generate singlet oxygen and selectively label neurons in live mice. Another type of label, the small molecule peroxidase Fe-TMAL, can be similarly used to detect endogenous proteins with immunolabeling.

For color EM, osmium staining is avoided by precipitating another element with DAB. Lanthanides (La, Ce, Pr, Nd) were chosen to precipitate DAB (Ln-DAB) because they have suitable electron energy loss spectroscopy (EELS) signals for energy filtered transmission EM (EF-TEM) imaging. Limiting factors of the Ln-DAB approach are related to tissue penetration and the amount of metal that can be added to DAB to increase contrast and sensitivity. EELS can be used to distinguish between different elements in the specimen; these signals can be detected spectroscopically or imaged through an energy-selecting slit onto a charge-coupled device (CCD) camera or direct detector.

For "multicolor" EM, precipitation of different lanthanides (e.g., Ce-DAB and Pr-DAB) can be distinguished using EF-TEM. The ability to combine multicolor EM with EF-TEM is restricted by limited methods for orthogonal precipitation of Ln-DABs as well as cross-talk between different fluorophores. The need to acquire multiple images per section makes this process time consuming. The sensitivity is also limited but can be greatly improved by direct detectors. Finally, EF-TEM instruments are not widely available, although the more common cyro-EMs can be used for multicolor EM imaging. Single-color EM remains a useful approach to derive specific signals from labels that can be distinguished from background osmium stains, and benefits from fewer technical limitations.

One strategy to increase sensitivity of multicolor EM is element-specific low-loss EF-TEM, which uses a different part of the spectrum than EEL high-loss spectrum. With this approach, the same electron dose yields a 4-fold increase in signal to noise and a 2-fold increase in resolution. Poorer separation of the peaks from adjacent elements can be overcome by using elements that are farther apart on the spectrum.

Dr. Adams and colleagues are developing another method called Color-Scanning TEM (Color-STEM; Z-contrast imaging using elemental tags for 4D STEM) to facilitate imaging of color EM samples. With STEM microscopy, researchers can distinguish between some of the metals used for color EM and the endogenous EM stain used to visualize cellular structures. For example, if ruthenium is used instead of osmium to stain cells, the color EM metals can be distinguished from the cell stain with minimal overlap between signals. Color-STEM is a more accessible, higher throughput, and dose-effective alternative to EELS and other spectroscopic methods.

Nanoscale Map of Whole Cells and Tissue Using Genetic Probes Technologies and 3D Electron Microscopy

Daniela Boassa, PhD

A common challenge in biomedical research is to observe and map biomolecules in their cellular and tissue context. Dr. Boassa and colleagues are developing enhanced molecular genetic probes and chemical labeling approaches to facilitate the application of correlated, multi-scale, multi-modal imaging to the integrated study of biomolecules across spatial and temporal scales.

The NCMIR group has developed a range of genetic EM probes. One key feature of the genetic approach is the incorporation of tags directly onto a target protein of interest, delivering

uniform 3D labeling and adequate preservation of ultrastructure. More recently, probes have been developed that allow detection of both extracellular (Split HRP) and intracellular (Split APEX2 and Split miniSOG) protein–protein interactions. All of these tags induce the oxidation of DAB by either light or chemically, which leads to formation of a polymer that can be stained with osmium for EM visualization.

To achieve 3D imaging and analysis of biomolecules in their cellular and tissue context, NCMIR leverages various imaging platforms. LM is used to observe the dynamic behavior of proteins. By doing correlative microscopy and taking advantage of other platforms such as tomography, serial block-face SEM (SBEM), and X-ray, multiple imaging modalities can be used to follow molecular constituents across scales. SBEM is often used for ultrastructural studies of the brain, because it produces large 3D volumes with nanometer resolution; however, the relatively large fields of view coupled with nanoscale resolution present issues for accessing the vast amounts of information contained in those images. Deep neural networks may mitigate this challenge, including an NCMIR cloud-based tool for image segmentation tasks (CDeep3M). This tool allows for measurement of specific organelle parameters (e.g., volumes, surface area, distribution), which provides metrics that can improve the understanding of the roles of these organelles and their alterations in disease.

With these tools, researchers can obtain nanoscale maps of whole cells in a tissue. However, one challenge is the ability to differentiate cell types based on key molecular constituents. To achieve this goal, Dr. Boassa and colleagues developed specific EM reporters to differentiate cell types by EM or to target proteins within specific pathways to access their location and function. For example, Dr. Boassa leveraged well-characterized neurotransmitter properties to identify and analyze neurotransmitter-defined synapses in a given microcircuit by tagging the vesicular glutamate transporter VGLUT2 with a protein detector (miniSOG) to label synaptic vesicles. Labeled vesicles were then used to visualize projection pathways by LM and SBEM; CDeep3M was used to distinguish individual vesicles in presynaptic terminals. This approach enables a detailed nanoscale analysis of microcircuits where the identity of presynaptic terminals is specified by their neurotransmitter.

The combination of these tools allows for study of the morphology of excitatory synapses in different brain areas, as well as quantitative analysis of synaptic vesicle profiles (i.e., distribution, size, density), while providing information from other organelles. This approach will be powerful in the context of many neurological conditions that lead to synaptic dysfunction. The capability to combine genetic EM probes with correlative imaging approaches, 3D imaging methods, and sophisticated computational tools for image analysis has expanded the possibilities for visualizing biological specimens. Future developments and enhancements in these approaches will continue to push forward a more comprehensive understanding of biological functions.

Session 2 Q&A

Variability Within and Across Animals

The value added of having one versus many samples was discussed in the context of stereotyped and variable patterns of functional activity and structural connectivity in the mouse brain. Dr. Hillman explained that brain-wide activity maps reveal that activity patterns differ from mouse to mouse, although her lab has not yet investigated the extent or nature of these differences. Analysis of these data is complicated by the fact that correlation-based connectivity varies over time as an animal changes its behavior. Dr. Hillman believes it will be particularly informative to focus on understanding which correlations are stable and which change over time. She noted that knowing the extent to which prior atlasing approaches have analyzed stereotypical brain activity between mice would also be instructive. In any event, the inter-animal and intra-animal variability in functional brain activity her lab has observed thus far highlights the importance of gaining a better understanding of the functional heterogeneity within individual animals before making comparisons across animals.

Dr. Zeng noted that structural imaging and connectivity imaging of more than 3,000 C57BL/6 mice by her group revealed some variability in anatomical regions across different animals, although it is relatively small. However, the degree of variability of specific connectivity patterns is not yet understood. Together, these functional and structural findings argue for performing as much analysis as possible on a single brain. The degree of variability is an important question to address, even for inbred mice, and will be a much larger question for human and NHP brains.

Compatibility and Complementarity of Modalities

A range of factors influence the use of multi-modal approaches to investigate brain connectivity on different spatial and temporal scales, including compatibility, comparability, and the ability to correlate results within and across brains. Dr. Boassa emphasized that the capacity to correlate and integrate functional activity information obtained by LM and structural information obtained by EM reconstructions provides a powerful approach to defining cell types, for example. Data obtained across different modalities can be correlated by use of intrinsic features in the images (e.g., the vasculature) or the addition of extrinsic labels as fiducials. Dr. Hillman suggested that the types of modality comparisons under consideration for the brain connectomics project can be performed first on smaller areas of the brain before scaling up to the whole brain. For her research, where connectivity between projections is inferred based on activity patterns, obtaining structural connectomics would be a valuable first step to understanding how to interpret the relationships between functional and structural data.

Dr. Barbas discussed the particular challenges of doing connectomics in the primate brain. Thus far, mapping of neuronal connections in the NHP and human brain has been done on very few regions of the brain because of the time-consuming nature of that work. One approach to using the existing connectivity mapping is to create a database of NHP and human data compiled from many investigators. A key component of Dr. Barbas' strategy entails analyzing broad

patterns of connectivity using LM, where analyzing relatively large numbers of samples is feasible, and then focusing on relevant synaptic details at the EM level. Dr. Barbas' work is also guided by principles derived from studies in the mouse to provide a framework for answering how synaptic connections are disrupted in regions of the human brain that are vulnerable to specific diseases.

Advances in tissue clearing approaches such as CLARITY have expanded the capabilities of LM approaches to image large, thick pieces of brain tissue. Dr. Chung addressed the compatibility of CLARITY with other imaging modalities and factors that might influence its suitability for use in the mouse whole-brain connectome project. Tissue clearing with CLARITY removes all lipids from the tissue; therefore, it would be difficult to obtain enough contrast along cell membranes for EM or X-ray imaging. Although some membranes are still visible with osmium staining, the staining quality is very poor. Current tissue clearing approaches are not compatible with downstream EM imaging, but future advancements may overcome this limitation by enabling preservation of a subset of lipids sufficient for acceptable osmium staining.

It is not yet clear whether it will be feasible to trace axons between blocks in tissue that has been cleared using CLARITY. To facilitate axonal tracing across blocks of human brain, Dr. Chung's team developed an ultra-precision vibratome to minimize tissue damage during sectioning. His team is currently characterizing the surface profile of vibratome-sectioned human brain tissue and investigating whether axons can be traced between adjacent blocks. Dr. Hillman noted that although her lab uses tissue clearing to image large, relatively intact pieces of brain, it can also image thinner pieces of tissue that have not been cleared. For such uncleared samples, Dr. Hillman's LM imaging approaches would be compatible with EM.

Projectomic Approaches

Several pros and cons should influence consideration of whether to use and X-ray tomography or LM imaging approaches for obtaining mesoscale projectomic data. Unlike LM imaging, X-ray tomography, as well as MRI and X-ray microscopy, are species agnostic and do not rely on genetically encoded fluorophores, so they can be used for NHP and human samples. Another advantage of X-ray tomography is that most of the key neuronal features, such as cell bodies, dendrites, and myelinated axons, can be resolved. Moreover, X-ray imaging is compatible with some of the fluorescent tags used to label specific types of neurons (e.g., Apex) and can also be performed after in vivo calcium imaging and before EM.

Additional technical considerations for X-ray tomography include the spatial and temporal resolution. Dr. Kasthuri imaged an entire mouse brain (stained using the BROPA protocol) in 7 hours at a resolution of 600 nm, although the resolution was limited in some areas because of inconsistent osmication. Imaging of large samples requires montaging across the brain to reconstruct the entire sample, although translation of the brain while imaging simplifies this process.

Biological Insights

The development and use of novel chemical, genetic, and imaging tools to investigate brain structure and function are undoubtedly leading to profound advances in our understanding of neuronal function. In one intriguing example, Dr. Boassa presented the unanticipated finding that VGLUT2 only labeled a limited subset of synaptic vesicles in excitatory neurons in the VTA. This result was likely not due to problems with labeling, because all reagents used in their labeling approach were freely accessible to the tissue. Rather, one possible explanation is that, because a subset of glutamatergic VTA neurons also express other neurotransmitters (i.e., dopamine or GABA), some vesicles might contain transporters for those transmitters, but not express VGLUT2. It is also possible that some synaptic vesicles in VTA neurons do not contain any vesicular transporters. These biological questions will be pursued through analysis of other brain regions.

Discussion Panel

Panel Discussion Topic: Charting a roadmap for mouse whole-brain connectomes and larger brain projectomes—sample preparation feasibility as a prerequisite for choosing imaging platforms, sample selection criteria for scaling to multiple brains

Panel Chairs: *Davi Bock, PhD, and Hongkui Zeng, PhD*
Discussants: *JoAnn Buchanan, MS, Adam W. Hantman, PhD, Jeff Lichtman, MD, PhD, Lisa Miller, PhD, Linnaea Ostroff, PhD, and Clay Reid, MD, PhD*

During this portion of the agenda, discussants presented their views on the topic, commented on the presentations by the earlier speakers and other discussants, and considered the questions posed in the online Q&A. Discussants were further charged to summarize the current state of tissue preparation; identify key issues to be resolved in sample preparation for the mouse whole-brain EM connectome in conjunction with different EM imaging platforms; articulate the needs and advantages of complementary data types for mouse whole-brain EM connectomes; and define the limitations, potentials, and greatest opportunities in moving to larger brains.

Ms. Buchanan advocated for the development and adoption of new chemistry for staining samples to image the mouse whole-brain connectome. Techniques that leverage osmium and plastic embedding resins were developed for use with small tissue samples, and therefore new methods are needed to scale up to the larger samples needed for the mouse whole-brain connectome. Ms. Buchanan highlighted two primary limitations related to the use of osmium with large samples. First, osmium fails to penetrate tissue completely in larger samples (e.g., 1-1.2 mm), resulting in an unstained “core” at the center of tissue sections. Myelin is prone to faster osmium uptake, which may contribute to the formation of the core. Second, osmium staining is more likely to yield inconsistent results when performed in larger samples. Given the time-intensive nature of EM and the value of the samples, Ms. Buchanan explained that the ability to evaluate sample quality at early stages is important. Technology such as the micro-CT, which facilitates core visualization without sectioning and enables the tracking of osmium as it permeates tissue, has therefore been a breakthrough for sample preparation. Ms. Buchanan

also suggested that new technologies may be leveraged to improve the penetration of osmium into larger samples (e.g., microwave- or ultrasound-based techniques). Dr. Ostroff agreed with Ms. Buchanan's call for new chemistry and suggested that the qualities that make osmium a valuable reagent—namely its utility as an oxidizing agent, fixative, and heavy metal contrast agent—might be separated among separate chemical entities.

Ms. Buchanan remarked that the neuroscience community must determine how labs will coordinate the approach to reconstructing the mouse whole-brain connectome. Multiple labs could focus on the whole brain, or rather different labs could reconstruct pieces of the brain or other areas of interest (e.g., spinal cord). She suggested that as data are generated and protocols are developed, labs should share these resources to advance the field more generally.

Dr. Lichtman presented a roadmap that represents the minimum essential steps for reconstructing the whole mouse brain. This process begins with the preservation and staining of tissue followed by sectioning and then imaging of the samples. Dr. Lichtman acknowledged that some variation may occur in the order of these first steps. For example, researchers may choose to slice tissue into thick sections prior to staining and then proceed to thinner sectioning, or may combine the sectioning and imaging processes (i.e., using block-face imaging). A series of computational steps follow the imaging of tissue samples, including evaluation, stitching, alignment, segmentation, annotation (including of synapses), and finally analysis and sharing. Dr. Lichtman stressed the importance of optimizing the earlier steps in this process so that the later steps can be meaningful, and expressed optimism that large volume osmium staining (i.e., up to 250 μm^3) will become a reality in the foreseeable future (although plausible alternative staining methods exist).

Dr. Lichtman posed a series of outstanding questions pertaining to sample preparation for the mouse whole-brain connectome. First, he asked whether block-face imaging or preservative approaches to sample preparation are preferable for this endeavor, noting that time savings may be gained from the latter. Preservative approaches enable reanalysis of samples, such that the whole brain can be imaged more quickly at lower resolution and regions that require higher resolution (e.g., neuropil) can be imaged separately at a later stage. Second, Dr. Lichtman questioned whether a whole mouse brain could be reliably sectioned at an ultra-thin thickness. If this sectioning cannot be achieved consistently, a strategy to divide the mouse whole-brain connectome into more manageable segments becomes even more critical.

Finally, Dr. Lichtman considered whether the mouse whole-brain connectome should include molecular or activity labeling. Although numerous ways to perform such labeling were presented at this workshop, Dr. Lichtman suggested that the first mouse whole-brain connectome should not include molecular or activity labeling of any kind. Dr. Zeng agreed with this suggestion, adding that the first connectome should be performed with a wildtype C57BL/6 mouse in the absence of any functional manipulations to avoid overcomplication. She posited that this first connectome may be used to correlate morphological projections of individual cells and that these data may be used as a "Rosetta stone" to link other types of data to the complete EM volume. Dr. Hillman echoed support for this suggestion, adding that—as was the

case for the human genome—a clear vision for the connectome's ultimate application has not emerged to guide decisions about what types of features to add.

Dr. Miller emphasized that imaging the whole brain will require a multi-modal approach. While EM provides the highest available spatial resolution, it is limited by its requirement for heavy metal staining and associated challenges with obtaining sufficient contrast. Because X-rays penetrate tissue more deeply than electrons, imaging methods such as micro-CT are particularly valuable for their ability to assess the quality of whole-brain staining without sectioning. However, these methods also rely on heavy metal contrast and have more limited spatial resolution compared to EM, although synchrotron-based methods can improve this resolution. Dr. Miller added that X-rays are capable of multi-color imaging. Visible LM can leverage the power of genetically encoded tags, but its spatial resolution is insufficient to image the whole-brain connectome. The development of new tags that are sensitive to all three of these imaging modalities—electrons, X-rays, and visible light—would enable a multi-modal approach to imaging the mouse whole-brain connectome. Dr. Miller proposed that one way to achieve this multi-modal approach would be to develop genetically encoded dual-type fusion tags that merge classic fluorescent reporters (e.g., GFP) with lanthanide-binding tags (LBTs). LBTs are short peptide tags that contain a tight binding site for lanthanides. Thus, these dual-type fusion tags can provide sufficient contrast for use with X-rays and electrons, while enabling the imaging of the same tissue by visible LM; when used with a synchrotron source, sufficient resolution can be achieved to image membrane proteins in single cells. Dr. Miller noted that more work is needed to rigorously test these tags for EM, but stated that dual-type fusion tags are a promising avenue for enabling multi-modal imaging of the whole-brain connectome.

Dr. Ostroff highlighted the utility of immuno-EM methods for building large connectomes. Immuno-EM can be used to identify morphological markers of different cell types in an unlabeled brain. In samples that have labeled axons and dendrites, immuno-EM can be used to trace processes across serial sections. Dr. Ostroff acknowledged that researchers sometimes object to the use of immuno-EM techniques for reconstruction because they may impair morphology, but she contends that standard antibody labeling techniques preserve morphological detail. Dr. Ostroff showed a slide in which only the surface of a vibratome-cut section was immuno-EM labeled. Neurites labeled at the surface can then be followed through the EM volume deep into the section, where the labeling was lost, but ultrastructure quality is maintained. Moreover, quality images for reconstruction projects can be achieved using immuno-EM with tissue that has been prepared and stored in a variety of ways (e.g., the tissue can be stored at a range of temperatures). Staining can be performed pre- or post-embedding, and the techniques can be optimized post-staining. However, pre-embedding labeling tends to be more sensitive.

Dr. Reid underscored the two projects at the heart of the discussions during this workshop: the mouse whole-brain EM connectome and the large animal projectome—including the human projectome. Although both projects are valuable, Dr. Reid suggested that the differential between current knowledge and that which would be gained by completion of either project is potentially higher for the human projectome than for the mouse whole-brain connectome. The

human projectome will have long-term clinical importance. For example, deep brain stimulation functions largely by the activation of axons, and therefore a human projectome would facilitate the study of this therapeutic modality by mapping axons within the human brain. Dr. Reid further posited that the human projectome will be more valuable for building large-scale models of cognition. Dr. Barbas also noted that lab mice lack the genetic variation of primates and therefore may not provide sufficient insight into human diseases, suggesting that parallel efforts that are more targeted to understanding regions vulnerable to disease may be necessary.

Dr. Reid acknowledged that the human projectome is an ambitious goal. However, he emphasized that intermediate goals of various sizes can be set and attained that will contribute both important science and valuable new techniques. These intermediate goals may include a dense mouse projectome as well as projectomes from primates of various sizes.

Follow-up Discussion

Dr. Ostroff asked the group whether there is greater value in complete reconstructions (e.g., a full sensory modality from input to output) or in examination of smaller functional areas across different species. Dr. Hillman commented that both goals are important and valuable. Dr. da Costa remarked that the collection of smaller volumes from multiple regions across multiple animals will be a necessary aspect of scaling to reconstruct a full connectome, and that the two goals are not mutually exclusive. Dr. Lichtman expressed an interest in generating a dataset in which sensory modalities are fully traced from input to output, and noted that sample preparation methods may require adaptation to capture relevant CNS tissues outside the brain (e.g., retina). Dr. Lichtman acknowledged that reconstruction of the whole spinal cord would also be valuable, but remarked that the volume of tissue required for such an effort would be considerably time-consuming.

Although reconstruction of the mouse whole-brain connectome may be pursued by distinct groups, Dr. Lichtman cautioned against multiple labs reconstructing distinct connectomes rather than focusing on contributions to a single connectome. He emphasized that the first connectome will be the most expensive because the infrastructure demands will already be established for subsequent connectomes. Thus, connectomics efforts more broadly may be accelerated after completion of the first connectome. Dr. Reid agreed with these points and extended them to apply to the human projectome as well, adding that the barrier to entry for even small labs to contribute to these projects is low given the relative accessibility of useful techniques such as LM and standard antibody labeling techniques. Drs. Bock and Zeng emphasized that the exploration of parallel technical paths, including potentially redundant efforts and contingency plans, are vital to a project of this scale, particularly when the best approach is not clear.

The discussion panel concluded with recognition that even an imperfect first connectome would bring about great scientific advancements. As some refer to the connectome as a “moon shot” project, Dr. Hillman reminded participants that the race to the moon yielded scientific innovations that were wholly unrelated to achieving that goal, and that the connectome will

bring about similar tangential discoveries. Dr. Reid extended this commentary by nothing that the connectome represents a grand challenge in exascale data and, by virtue of this fact, the non-neuroscience computer science field will greatly benefit from this effort.

Workshop 3: Experimental Modalities for Whole-Brain Connectivity Mapping

Precis

Which currently operational serial EM or other imaging methods will scale? What is the time budget for complete pipelines from tissue block to stack of images to provide adequate field of view, lateral (x,y) and z resolution? For each of the competing potential technology pipelines, which have temporal or spatial scaling bottlenecks that may be successfully overcome by directed technology development in 1 to 5 years?

Pioneering strategies with the potential to scale are multibeam SEM and TEM methods, but each has different risks and bottlenecks that need to be addressed. X-ray microscopy was also discussed but currently lacks suitable resolution. More work here on sample preparation and disambiguation of what can be done on epoxy-embedded tissue slabs versus whole mouse brains was not fully explored, raising questions about near-term feasibility of beam line-based strategies to produce connectomes. For the more proven EM methods or hopeful X-ray methods, goal-directed suitability should be explored rapidly and rigorously. Projects exploring each with key subareas of mouse brain might be undertaken to strategically address bottlenecks and risks, such as connecting between thick slabs, addressing distortions due to preparation, and imaging optics, or to address losses of data associated with occasional failures in automated processes.

What level of detail (resolution of detail in x,y and the more challenging z dimension) would be sufficient to drive toward a well-defined goal of charting the synaptic connectivity between ALL neurons in the mouse brain (a "synapto-projectome")?

At a minimum, this would require the ability to track and subdivide synapses belonging to key synaptic classes such as excitatory, inhibitory, peptidergic, spinous, dendro-dendritic, axo-axonic, and reciprocal (e.g., to define sites where synapses include "presynaptic autoreceptors"). As synaptic weights are also part of current considerations in brain structure and function, features of synaptic structure that correlate with synaptic weight, such as synaptic apposition size, vesicle numbers in terminals, and vesicles in docked or ready releasable states (e.g., numbers of vesicles located within a vesicle diameter of an active zone) should be accurately revealed and chronicled brain wide. Non-chemical neurotransmitter-based synaptic interactions, defined by electrical synapses with gap junctions of different connexin types is an important but critical stretch goal, because a synaptic wiring diagram without these may be remarkably incomplete. Lastly, because non-synaptic release of neurotransmitter and receptors away from EM-defined synaptic appositions are common, one must recognize the limitations of a large-scale effort to map morphologically-defined synapses as traditionally defined, because it will not match current conceptions of neuron-to-neuron or neuro-glial interactions, which modulate circuit function.

The discussion of the overall initiative considered an arc of work to progress from establishing the capability to build one complete ultrastructurally-sufficient connectome to applying the methods to deliver perhaps 100 mouse brain connectomes. Accordingly, after one or two campaigns are in process, participants considered which additional parameters should be tracked, to add important details to the initial connectome maps, given practical limitations of what it will take to scale to a complete mouse brain from where we are today, technologically. Important additional details could be provided by "time-resolved EM," with an example offered for simply fixing animals to be mapped in different states or ages and comparing static brain samples across time points. Use of marker mice, possibly exploiting labels for correlated LM and EM, to facilitate identification of synaptic chemistry, non-classical synapses, glial dynamics and modulation, gap junction locations, and so on was also discussed. To propel the adoption and expand the utility of these efforts, it was strongly suggested that some efforts exploit mouse models of human disease. In particular, models designed to incorporate disease risk-linked gene modifications with properly integrated and regulated gene substitution methods should be considered, perhaps with added genetic probes for CLEM, to make the connectomes easier to read out and compare to the naïve first mice done to advance the scalability of methods required.

Democratization of methods and platforms was also discussed. This included discussion of novel and lower-cost platforms in regional facilities and eventually small laboratories. This could be facilitated by the distribution of standardized sample blocks, made in a central facility, allowing multiple groups in suitable smaller laboratories to obtain data and, with the aid of software tools, to dock small subdomain connectomes into most complete versions resulting from a few other much larger-scale centralized efforts. In general, as the community explores and refines methods to produce full brain connectomes with sufficient fidelity, efforts should be taken to lower the cost of such efforts and to distribute the work whenever and where feasible. This dissemination effort should include new instrument development and progressive refinement of new and perhaps novel platforms that are more suitable for laboratory-scale efforts, nationwide and worldwide.

Several ways to distribute work and add value to sets of large-scale complete connectomes were discussed, including ideas such as genetically engineered mice for the *n* beyond the first one or two done with more heroic methods. This would encourage the testing of variations in more specific subareas of the brain, allowing docking of larger but more restricted brain regions into a shared map, as technology throughput becomes more mature. Specifically, there was strong sentiment expressed to encourage multiple groups to pursue work on smaller scale EM-enabled microconnectomes of key brain regions, particularly for those most relevant to fundamental challenges in brain structure and function. Examples of targets include learning and memory, sexual dimorphism, development, and aging, as well as the most obvious health-related underpinnings of neurodegenerative disease or developmental disorders. Additionally, the acquisition of mesoscale projectomes, obtained by fluorescence microscopy techniques, which allow for the linking of the connectome maps with neural activity, or potentially higher-throughput X-ray CT maps, obtained with either laboratory-scale XRMs or more sophisticated and powerful synchrotron-based beam line platforms, were considered.

Introduction

Mark Ellisman, PhD, Galya Orr, PhD

In Dr. Ellisman's view, the overall Brain Connectivity Initiative is aimed at advancing technology to reveal the secrets of the brain at the level of detail of neurons and neuronal connections. This goal gathers inspiration from the selective staining of neurons and glial cells, developed by Camillo Golgi and used to great effect by Ramon y Cajal to characterize the vectoral basis of point-to-point connections made between individual cells to form a brain-wide network. Because he could highlight objects with selective Golgi staining, Cajal was able to tease out many aspects of brain complexity. The immediate question is how to decomplex the brain while viewing it in its entirety using microscopic and computational methods to extract all of the details. This workshop focused on the tremendous technical challenge of analyzing brain connectivity across the wide range of spatial and temporal scales (i.e., "tyranny of scale"), including time-dependent changes in the context of neuronal wiring and the cells that modulate wiring.

Gaining a comprehensive understanding of brain structure requires that the architecture of the cells that support and modulate neuronal function be characterized in addition to the neurons themselves. The various types of glial cells play critical roles in the nervous system by modulating a wide range of neuronal functions, including wiring, connectivity, synaptic transmission, excitability, firing patterns and coherence, and learning and plasticity. The importance of glia in brain function is highlighted by the increase in the proportion of glial cells per neuron during the evolution of the nervous system, particularly in humans. Although the intended goal of the connectomics initiative is to determine the point-to-point wiring maps of neurons, the depiction of glia should be considered an integral part of the initiative.

For Workshop 3, the speakers were asked to address the four questions related to the modalities suitable for whole-brain mapping and to discuss the state of the art, opportunities, and challenges relevant to each of the technologies and approaches they represent.

First, what resolution is required to attain meaningful wiring models? Dr. Ellisman framed this fundamental question in the context of synaptic connectivity. Synapses are among the most dynamic structures in the brain, and the many forms of postsynaptic dendritic spines relate to a readout of synaptic efficacy. Yet, the field should consider the level at which dendritic spines and synapses should be characterized to yield the most useful information about synaptic efficacy and connectivity for the brain connectome project. For example, recent advances in 3D imaging technologies, such as multi-tilt tomography, now enable comprehensive description of dendritic spines at the molecular scale. Thus, while imaging technologies proposed for use in this initiative now enable extraction of these synaptic details, the appropriate level of synaptic detail for the brain connectome project remains to be determined.

Second, which imaging technologies could be driven to scale and most easily disseminated? Concurrent goals for this initiative include not only development of a plan to achieve the connectome in a few years, but also proliferation of new technologies. The many methods

described in this workshop provide a very likely basis for how the mouse whole-brain connectome can be achieved within the next several years. Moreover, parallel efforts, such as the multibeam SEM, will likely produce collateral benefits even if they do not contribute directly to achieving the intended goal. Advances in electron, X-ray, and LM; genetic and other techniques for labeling molecules with heavy metals and fluorescent tags; camera speed and spatial resolution; and computational approaches (e.g., deep learning algorithms) are revolutionizing the ability of neuroscience to tackle this tremendous challenge.

Third, participants were charged to compare and contrast single large-scale efforts to democratize data collection. Dr. Ellisman has long held the view that the whole-brain connectome can only be achieved through democratization of the endeavor so that scientists everywhere can stitch in their individual pieces. His early realization of the democratization of brain mapping led to a prototype strategy called "The Whole Brain Catalogue."

Fourth, how can static brain maps encode dynamics? The simplest idea for introducing dynamics into connectivity maps is to construct connectomes at two different time periods or states (e.g., connectomes during wakefulness and sleep). Capturing faster temporal changes is more difficult, although dynamics can now be encoded genetically.

For Sessions 1 and 2, the speakers were asked to address four questions. The presentations can be viewed by accessing the archived NIH VideoCast.

1. What resolution is required to attain meaningful writing models?
2. Which imaging technologies could be driven to scale and most easily disseminated?
3. Compare and contrast single large-scale efforts to democratize data collection.
4. How do you encode dynamics in statis brain maps (adding the time domain)?

Session 1: Multi-Scale Imaging of the Connectome

Multi-Scale Imaging of the Connectome

Jeff Lichtman, MD, PhD

Several criteria determine the resolution required to attain meaningful wiring models. Foremost, imaging resolution must be sufficient to identify synapses. With slightly higher resolution, researchers could analyze synaptic vesicles to differentiate synapse types and analyze synaptic strength. The resolution must also enable tracing of the finest axonal processes across many sections and neurons so that segmentation can be done without merge errors. In Lichtman's experience, raw data that are 4 x 4 nm in plane and 30 nm in the depth axis (i.e., a 480 nm^3 voxel) is sufficient if there is at least 5 percent extracellular space, which facilitates identification of putative inhibitory or electrical junctions. Other important factors include optimal osmium staining, minimal artefacts, and excellent microscope alignment to enable extended automatic segmentation run lengths.

In Dr. Lichtman's view, four leading EM imaging technologies could be driven to scale and easily disseminated for use in the connectome project, each of which has been used for high-throughput connectomics: (1) TEMCA is a fast, robust, and inexpensive tape-based TEM method, although the current grid tape is too thin. (2) ATUM (Automatic Tape-Collecting Ultramicrotome) is a SEM tape-based technique that uses thick tape that can be used with multibeam SEMs. (3) FIB-SEM is a focused ion beam SEM approach that was used for the *Drosophila* hemi-brain. It has many advantages, but its acquisition speed is slow. (4) GCIB is a new SEM technique that, like ATUM, can be used with multibeam SEMs to enable fast imaging.

Dr. Lichtman believes that the mouse whole-brain connectome project is ideal for democratization because it will require many imaging devices running in parallel on different parts of the same brain sample. Completing a whole mouse brain in 5 years would require the generation of about 1,000 terabytes (1 petabyte) of image data per day and 20 to 30 high-end, fast imaging devices. At present, such high-throughput technologies are largely siloed at elite institutions (e.g., Allen, Janelia, Harvard, Princeton, the Max Plancks), although imaging devices could be distributed more broadly. Wider distribution would challenge efforts to consistently monitor and routinely service the devices. However, such distribution would broaden geographic access to connectomic data-generating machines, which would lead to competency in connectomic data collection for neuroscientists across the country and to local access to the technology for use in other projects. As an alternative democratization strategy, devices could be located in local public sites, such as public National Laboratories, for use by any neuroscientist.

Dr. Lichtman believes that it is possible to infer dynamics from static connectivity maps of the mammalian nervous system. The entire brain likely has a large imprint of experience-based neural activity (e.g., learning, memory, sensory, and motor function) encoded in the wiring diagram, such that the wiring diagram itself is a record of the activity patterns relevant to the animal and becomes instantiated in the connectome. Ocular dominance columns in the visual system are a prototypical example of the importance of neural activity for shaping brain connectivity. Dr. Lichtman suspects the existence of many other as yet unknown connectivity patterns that are similarly experience-based and can be rendered by connectomics. Connectomics offer the great opportunity to describe the organizational principles by which such experience is stored in wiring diagrams.

Raw Speed

Winfried Denk, PhD

Dr. Winfried Denk began his remarks by addressing the four questions posed to the speakers. Although a great deal of connectomics data have already been generated that are meaningful at some level, to obtain a comprehensive connectivity map of the brain that contains all of the wires and that can be largely analyzed computationally without substantial human proofreading effort, the resolution should be around 10 nm. Regarding the technologies that could be driven to the scale needed to complete a mouse whole-brain connectome, Dr. Denk believes that the only promising approach is the combination of the ion cluster gun and the

multibeam SEM, as being developed by Dr. Ken Hayworth. Although the best brain connectivity models are not dynamic, a perfect wiring diagram contains information about how the wiring has evolved over time. However, efforts to also record dynamic activity in the same animal that is used to generate the wiring diagram risks compromising the quality of the structural data and should be avoided. Finally, a large-scale effort such as the connectome project should be democratized from the standpoint of basing decisions about this project on a broad consensus of the field. However, as for other large technology projects, the initial phases of the mouse brain connectome project will require unique state-of-the-art instruments and a great deal of other resources.

Dr. Denk described his recent advances in SBEM imaging technology that are focused on improving acquisition speed and scaling, which he hopes will contribute to the brain connectome enterprise. In the current implementation of SBEM, the time required to scan an entire field of view (FOV) of 150 µm is only a few hundred milliseconds (if some resolution is sacrificed). Following acquisition of each image, the FOV is shifted by the size of the FOV to acquire the next image. If a mechanical stage is used to shift the FOV, this movement causes oscillations—or "ringing"—which requires a pause of 1-2 seconds before the next image can be acquired. This pause represents 80-90 percent of the raw speed and is rate limiting for the rate of image acquisition. However, if a piezo-driven stage is used, movement will cease within tens of milliseconds and the entire scan takes only 3 seconds, which is a substantial improvement.

Dr. Denk is trying to solve the challenge of improving acquisition speed through designing and engineering several technical modifications to the scanning process and mechanism, including the incorporation of piezos and other alterations. Dr. Denk illustrated the key concepts for his design with a series of computer-aided design (CAD) drawings. One change was to acquire a number of frames in quick succession, such that the mechanical stage was not moved and the only moving part was a lightweight wafer that could be controlled very precisely. During imaging, the wafer shuttles between the milling chamber and the imaging chamber. A key feature to achieving the degree of repeatability required for this process is the incorporation of a suspension mechanism made from a sapphire rod that is precisely designed to ensure overlap between FOVs and to avoid focus problems. Piezos are used for the scanning mechanisms, including the slow y-axis scan. This approach has the advantage that the data are written to a single file per beam, which decreases the number of individual frames that must be written.

If sufficient contrast in the sample can be achieved to enable a 50 ns dwell time, data rates of about 1 giga-voxel per second can be achieved. This acquisition rate represents a 40-50 thousand-fold increase in the speed of acquisition compared to Dr. Denk's initial version of the SBEM instrument in 2004.

Multi-Scale Imaging of Connectomes Using TEM

Davi Bock, PhD

Recent trends and lessons learned from the use of TEM for acquisition of large-scale connectomic EM volumes has several implications for scaling up to the mouse whole-brain connectome. With regard to meaningful resolution for TEM, many investigators have

converged on about 4x4x40 nm voxels. However, other considerations influence the meaningfulness of this resolution. To obtain the minimum viable connectome that enables complete reconstruction of neurons and their chemical synapses, the necessary voxel size depends on the image signal-to-noise ratio (S/N), which is a function of the quality of the sample, the electron dose per voxel, and the signal. Quantification of the S/N from images obtained by multiple investigators revealed that much of the field has converged on a similar S/N if normalized for voxel size, suggesting that the information content of acquired data is similar across labs and that researchers often arrive at similar decisions regarding the balance of imaging throughput, image quality, and volume acquired. Whatever the EM imaging modality chosen for the mouse whole-brain connectome, target S/N should be determined quantitatively. The near consensus shown in datasets acquired to date should inform this target.

Error rates and completeness also affect the meaningfulness of the connectome. In general, false negative connections are preferred to false positives, and the more connections that are mapped, the greater the statistical power to test null models. In addition, analyzing the empirical reconstruction quality of existing efforts would be useful in planning the scale-up to the whole mouse brain. Issues affecting error rates include staining, embedding, and lossless subdivision, as discussed in Workshop 2. The “ultra-smooth” vibratome concept introduced by Dr. Hayworth would solve many sample preparation issues and would enable the use of nondestructive imaging methods (e.g., TEM plus GridTape or ATUM-multibeam SEM) at the scale of the whole mouse brain.

Nondestructive imaging methods could advance the field because the ability to reimage sample offers many advantages. Because many important subcellular features cannot be easily resolved at the resolution useful for the mouse brain connectome, tremendous value would be derived from additional information about finer ultrastructure details within the context of the mouse whole-brain connectome by re-imaging samples and annotating higher resolution data. These include the size of post-synaptic densities, which has recently been shown to correlate with excitatory postsynaptic potential (EPSP) size (i.e., the functional strength of the synapse).

Dr. Bock’s assessment is that two imaging modalities could be driven to the scale of the whole mouse brain—multibeam SEM and TEM plus GridTape. Considering cost and throughput, TEM is an order of magnitude less expensive than multibeam SEM, and throughput is currently about three times slower. However, in recent years, engineering efforts have focused on improved sample handling, and tremendous opportunities exist to increase throughput in TEM. Moreover, TEM throughput follows commercial trends in sensor size and readout speed, which continue to improve. Despite its relatively low cost, dissemination of TEM is challenging, primarily because downstream post-processing of data is difficult. However, free or inexpensive services for image storage, stitching, alignment, segmentation, and analysis would facilitate data processing.

Among other efforts to democratize data collection, open data are key and should be available to all, including raw image sets, aligned EM volumes, segmentations, and analysis tools. If more

labs could access connectomics-scale EM volume acquisition methods, data processing software and services, and analysis tools, then the field's creativity could be unleashed.

Finally, encoding dynamics in the connectome can be realized through several approaches. These include in vivo imaging followed by fixation, LM, and EM, as well as performing EM on brain samples obtained at different ages (e.g., development or senescence) to capture structural differences across life stages.

Multi-Scale Imaging of Connectomes with Photons and Electrons

Wei-Chung Allen Lee, PhD

Two key lessons learned can be learned from Dr. Lee's work in multi-scale connectomics. First, big problems should be addressed with a diverse set of approaches—even with some redundancy—to increase the likelihood of success. Second, in the context of data acquisition for a connectome, the goals should be to minimize the number of sections acquired as well as the physical movement of sections and, if possible, to image nondestructively.

For multi-scale connectomes, the resolution to attain meaningful wiring models depends on the questions one wants to address. For questions about information flow, a white matter projectome of < 200 nm is sufficient. For questions about neuronal computations, a chemical synapse connectome with 4-10 nm resolution is required. However, for questions about electrical synapses that entail resolving non-specifically labeled gap junctions, the resolution must be about 2 nm.

Among the approaches that have proven useful for multi-scale imaging of connectomes, more high-resolution connectomic data have been generated from TEM than from SEM or X-ray methods. With regard to imaging technologies that could be driven to scale and most easily disseminated for multi-scale connectomics, Dr. Lee focused on his own experiences with synchrotron-based X-ray and EM. Key advantages of using X-rays as probes include their high penetration, which enables imaging through thick samples, and the ability to image nondestructively, which enables correlative and multi-resolution imaging. Moreover, some approaches allow the flexible trade-off between resolution and field of view. The European synchrotron has some unique aspects that enable high-resolution imaging, which allowed Dr. Lee's team to achieve < 100 nanometer resolution, approaching EM resolutions and contrast that were comparable to EM. The team could then comprehensively reconstruct and follow myelinated axons and large dendrites across long distances in these datasets. Following the upgrade of the European synchrotron facility last year, Dr. Lee expects to gain another 15-fold increase in throughput and to approach synapse resolution.

Dr. Lee's group has made significant advances in automating TEM, with a specific focus on improvements in sample handling. Several developments have made TEM more scalable, including camera arrays, automated sample handling, automated segmentation, and synapse inference. Dr. Lee's group has recently developed a sample collection substrate called GridTape to leverage the advantages of transmitted electron detection. With this approach, tens of thousands of thin or semi-thin sections can be collected automatically onto tape. Together with

their open-source reel-to-reel imaging stage, imaging of thousands of collected sections through a standard TEM column can be achieved without breaking vacuum. The power of this technique was illustrated with a dataset of a fly nerve cord acquired with GridTape, which took only 2 months to acquire. The segmentation of this large dataset highlights how such TEM data can now be routinely segmented by modern deep learning approaches.

Dr. Lee's team is now working toward improving axial resolution and imaging through thicker samples by combining tape-based imaging with TEM tomography. For this, the team can either rotate the tape to tilt the samples or optically section by tilting the beam to generate 3D data. Tilting the beam in this way allows fewer but thicker sections to be cut and imaged and minimizes the physical movement of samples while remaining nondestructive. This approach leverages new advances in sample handling, detectors, and high-speed scope control, although it requires investment in aberration correction, either with engineering of optics or with machine learning approaches.

Regarding democratization of data collection, TEM is highly accessible and relatively inexpensive, and thousands of instruments are disseminated in field. Moreover, institutions world-wide are now adopting Dr. Lee's GridTape-based pipeline, thereby broadly accelerating imaging throughput. By contrast, only about 50 synchrotrons exist around the world. Although dedicated beam lines would be required for a connectome project, beamlines could also be turned into shared resources for the other endeavors in the broader scientific community.

Although there are multiple ways to think about encoding dynamics on static brain maps, Dr. Lee's preferred approach is to combine large-scale structural imaging with functional imaging. His group now routinely correlates cellular resolution in vivo calcium imaging data with EM connectomics datasets. Such a functional connectomics approach enables bridging of multiple levels, from behavior to function to the structure of neural networks. Moreover, with a multi-scale connectomics toolkit, different parts of the brain can be imaged at different resolutions, saving time and data.

Synchotron Hard X-ray Microscopy for the Connectome

Si Chen, PhD

Synchrotrons are large-scale facilities capable of producing highly intense X-ray photon beams. The high penetration power of the X-ray beam provides unique opportunities for the brain connectome project, including the nondestructive study of large brain volumes, which enables multi-modal and multi-scale correlative studies. Several X-ray techniques could be used for the brain connectome, including ptychography, holography, X-ray fluorescence microscopy (XFM), and full-field tomography (i.e., microCT and nanoCT). Recent advances in these technologies combined with ongoing upgrades to the Advanced Photon Source (APS) facility at Argonne National Laboratory (to be completed in 2023) will further expand the range of opportunities for using X-ray techniques for the brain connectome project.

Phototomography, also known as microCT or nanoCT depending on its resolution, is the most common X-ray imaging technique in connectomics. Combined microCT and SEM can be used for

multi-scale analysis, as exemplified by the collaborative project described by Dr. Kasthuri in Workshop 2. Current nanoCT techniques can achieve sub-nanometer spatial resolution, although the FOV is limited. X-ray holographic nano-tomography could combine both high spatial resolution (synaptic scale) and a large FOV (whole mouse brain). This approach has recently been used to reconstruct dense neuronal networks of millimeter-sized tissue at single-cell resolution, with synaptic-scale resolution expected following the APS upgrade.

Ptychography is an emerging technique for studying ultrastructures. With ptychography, a sample is raster scanned across the focus beam, while a far-field diffraction pattern is acquired at each scanning position using an array detector, and the image of the sample is recovered. Key advantages of ptychography for the brain connectome effort include superior spatial resolution beyond the optical limitation (<20 nm), enhanced image contrast, and the potential for high-throughput measurements.

Ptychography has been used extensively to image integrated electronic circuits (i.e., chips). Technological advancements now make it possible to obtain sub-20 nm resolution of a chip sample that is more than 200 μm thick. Current efforts at the APS aim to increase the spatial resolution, expand the FOV, and extend imaging into 3D. Following the APS upgrade, the brightness of the ptychography X-ray beam for ptychography will increase more than 2 to 3 orders of magnitude. For the brain connectome, this upgrade translates to the ability to image a 1 mm^3 piece of brain tissue with a 10 nm isotropic resolution in 3D within half a day, which would make it possible to collect data from a whole mouse brain in about 1 year.

Goals for further advancements in ptychography include adding colors that are physiologically relevant and extending the multi-scale capabilities. A powerful strategy to address these challenges is to perform correlative studies that combine the strengths of multiple techniques to obtain complementary information and extend the imaging scale. As one example, XFM was used to measure the endogenous distribution of elements (i.e., sulfur, phosphorus, potassium, calcium) within a sample, and ptychography was used to show subcellular structures; the combined images facilitated identification of the subcellular organelles. To achieve both elemental sensitivity and higher resolution, correlative XFM and TEM imaging can also be used. With such compositional information, it may be possible to obtain chemical signatures of healthy and diseased brains within the context of the connectome.

Additional objectives for correlative imaging and analysis include the ability to collect multi-modal data simultaneously, which is X-ray dose efficient and makes image feature registration straightforward; the need to transfer samples across multiple platforms for some correlative approaches creates technical challenges. Also needed are further developments to enable a workable pipeline, including improved sample handling methods and universal labeling that allows for identification of features across platforms. Finally, correlative data analysis is also essential to connect data across different scales and modalities.

tomoSEM—Tomography in SEM

John Mendenhall, PhD

Dr. John Mendenhall described work in Dr. Kristen Harris' lab toward developing a tomography approach for a scanning platform based on tSEM (transmission-mode scanning EM) methods, known as "tomoSEM" (tomography within tSEM). The goal for tomoSEM development is to obtain images at high lateral (2 nm) and axial (10-15 nm) resolution across large field sizes. TomoSEM is a nondestructive EM method and is intended to be low cost and available to any researcher with basic field-emission SEM.

tSEM imaging on a field emission SEM platform was developed for efficient and cost-effective imaging of circuit-scale volumes from brain. tSEM yields TEM-like results and maintains a resolution of about 2 nm. However, whereas TEM has small fields, tSEM enables acquisition of large-field images and is capable of automated acquisition and automated generation of field mosaics of sections and grids. The tSEM FOV is 50-100 µm, which is about the size of a small confocal field. One limitation of tSEM is that the axial resolution is limited such that, even within the very fine ultrasections, membranes of obliquely sectioned objects can obscure one another and vesicles can overlap. This axial limitation was the main motivation for the development of tomoSEM.

Tomography is a relatively straightforward method to reveal obscured ultrastructure. Its main limitations are the small field size and the many images that are required. However, tomography can discriminate between oblique membranes and complex structures. TomoSEM is being developed to obtain a similar result within the tSEM methods using tomography and thicker sections. The goal of tomoSEM is to retain the large field of tSEM (50x50 µm to 100x100 µm at 2 nm pixels), but to use fewer, thicker (250 nm) sections for a given volume, which will result in fewer cut-loss intervals, stronger and potentially larger sections, and few or no flaws. The target resolution is 10-15 nm virtually in the axial axis, which would reveal buried structures and improve autosegmentation of finer structures. The total number of images will be automatable as it is for tSEM. Current efforts are focused on optimizing the design, development, and implementation of the tomoSEM technology. This includes assessing beam scatter, calculating virtual sections, determining the minimum number of images required, and increasing the speed of throughput with new engineering.

The realization of the tomoSEM technology will result in a high-resolution nondestructive EM approach that will enable re-imaging of brain samples and correlative imaging approaches. The ability to use thicker sections and consistent performance of the acquisition style should provide for reliable imaging. Based on its use with a basic SEM platform and low cost, this technology should also lead to broad accessibility. Ultimately, tomoSEM will provide many advantages for efforts to map brain connectivity.

Session 1 Q&A

Neurocomputations

The extent to which the temporal dynamics of neurocomputations are encoded in the structure and connectivity of the brain remains an open question, with many alternative views. Drs. Lichtman and Denk argued that temporal dynamics are encoded in brain structure. For computations based on learning that are manifested in changes in synaptic strength, analysis of structural components alone may be sufficient. However, that may not be the case for neurocomputations that occur on a millisecond timescale (e.g., decision-making). Other views hold, for example, that although viewing connectivity and synaptic components at a small scale facilitates the understanding of neurocomputation and may provide a constraint on neurocomputations, these structural features alone are not sufficient to allow inference of actual neurocomputation. Moreover, the extent to which changes in connectivity are a cause or a consequence of the computations that are performed based on a given state of connectivity is also not clear.

Dr. Lichtman concurred that very high-resolution temporal dynamics may not be reflected in a connectome. However, he suggested that because the purpose of mammalian decision-making is to find a solution that can then be repeated automatically in the future, experience probably does become stored in the wiring diagram once sufficient experience is gained to achieve a post-cognitive, automatic state. Lichtman further emphasized that moment-to-moment dynamics may be overemphasized as a critical determinant of behavioral responses and that many behaviors—including human behaviors—are built into the wiring diagram based on experience and are ultimately automatic. Dr. Mao added that one potential mechanism for generating automated responses would be for information to flow through short-cut circuits that are built into the wiring diagram.

Dr. Ellisman raised the issue of how the brain manages the coherence of neuronal firing that emerges during learning, where interneurons and glia play roles in harmonizing the firing of output neurons. He contended that one goal of the connectome project is to identify approaches that could be used to mark neurons that have been harmoniously active. Such labeling approaches are needed to fully reveal how information actually flows through those circuits.

Software Computations and Solutions

Advances in software and computational approaches, including machine learning algorithms, automatic data fusion for multi-scale imaging, and image segmentation, are facilitating solutions to challenges in imaging resolution. Dr. Lichtman commented that with enough data collected at both high and low resolution, a machine learning algorithm can eventually be built to turn the low-resolution data into high-resolution data. Such a machine learning approach could enable the use of faster imaging speeds to generate data more quickly.

Dr. Ellisman cited a study that employs this approach on supercomputer-scale hardware to increase the ability to predict intervening sections, which provides a powerful means of

improving both speed and accuracy. Dr. Chen concurred that this approach is promising and explained that while gaining a two- to four-fold increase in spatial resolution is straightforward, achieving an eight-fold increase would be a challenge and that obtaining ground truth is also critical for conforming to real and not artificial features in image data. Dr. Bock added that some upsampling might already be implicit in current segmentation algorithms, which perform well on data that appear noisy to the human eye.

Synchrotron X-ray Microscopy

Ptychography offers several advantages for imaging brain ultrastructure, and system upgrades at the APS will make this approach feasible for potential use in the brain connectome project. Currently, the thickness of the sample that can be imaged by ptychography is limited by the depth of field; if the sample is thicker than the depth of field, then the beam size is larger. Ongoing efforts at Argonne and other synchrotron facilities are also trying to address imaging samples that extend beyond the depth of field using computational approaches. It is anticipated that 10 nm resolution can ultimately be achieved as long as the X-ray beam can penetrate the sample.

After the APS upgrade, it is expected that imaging of a 1 mm^3 sample at nanometer isotropic resolution could occur within half a day. This estimate relies on improvements in beam coherence and brightness as well as faster detectors. Current detectors can image at 2-3 kHz, whereas 100 kHz detectors will be needed to achieve faster imaging speeds. Current efforts to image integrated electronics can be translated to work on the brain connectome, although high dose irradiation must be avoided to prevent potential damage to brain samples.

Sample Preparation

Several factors influence the suitability of brain samples for particular types of studies. For human brain, the postmortem interval (PMI) is critical. Dr. Lichtman remarked that many fine processes will be destroyed if the PMI exceeds 24 hours. However, fresh human brain samples can be obtained from surgical procedures, such as deep brain stimulation. He further noted that a range of ethical concerns must be addressed before construction of a full human brain connectome is considered.

The connectome of one brain represents only a single timepoint. Dr. Lichtman described an approach known as “time lapse EM,” in which dynamic changes in brain structures are analyzed by comparison across animals that represent different time points. There is a history of doing correlated EM across time points in dynamic processes (e.g., circadian rhythms), although this work is easier when comparing across hours to days instead of milliseconds.

Identifying electrical synapses in the brain remains a significant challenge. Although membranes in close proximity appear to be a good indicator of electrical synapses, other features of EM images should also be considered, and labeling methods could be utilized to visualize gap junctions. Dr. Lichtman added that the inclusion of extracellular space in brain sample preparations boosts confidence in electrical and chemical synapse identification, because cells are generally not closely apposed unless there is something adhesive between

them. He further commented on the advantage of 4 nm resolution for imaging electrical junctions and inhibitory synapses, where postsynaptic specializations are difficult to visualize. This problem raises considerations for whether 10x10x10 nm or 4x4x30 nm EM imaging resolution should be used for the connectome project, because the higher resolution of 4x4x30 nm imaging along two axes would enable resolution of such small structures.

Session 2: Projectome to Connectome Imaging—Synapto-Projectomes: Toward Connectome Relationships in Target Fields

Axonal Connectomics: Projection Maps in Mice, NHPs, and Humans

Clay Reid, MD, PhD

Dr. Reid is an advocate for parallel and complementary efforts to pursue the mouse whole-brain connectome AND a mesoscopic connectome-projectome of the human brain. Microscale EM connectomics requires 10 nm voxels to study local circuits of synaptic connectivity at the levels of axons, dendrites, and synapses. The state of the art is 1 mm^3 volumes of mouse brain, which generates very large datasets in the petascale (10^{15}). Achieving microscale EM connectomes that also capture the network of inputs and outputs to those local networks throughout the whole mouse brain (1 cm^3) will require generation of datasets

in the exascale (10^{18}). For the human brain (10 cm^3), a microscale EM connectome would generate datasets in the zettascale (10^{21}). Although exascale is a plausible technological jump that makes completion of the microscale mouse brain connectome within the next 5 years potentially feasible, achieving this goal for the human will take decades.

An alternate strategy for the human brain would be to generate a mesoscale connectome. A mesoscale connectome would require 100 nm voxels to visualize myelinated axons, instead of 10 nm voxels, which would require 1,000-fold less data than a microscale connectome and turn the whole mouse brain into a petascale problem and the human brain into an exascale problem. This strategy would enable completion of a human whole-brain projectome in the same timeframe it will take to complete the mouse whole-brain connectome.

While previous efforts to obtain the human connectome have used diffusion imaging of white matter, the resolution (0.5-1.0 mm) is not sufficient to obtain a wiring diagram at the axonal scale; the higher 100 nm resolution is required to resolve axons traversing the brain in all directions. With current antibody staining and microscopy technologies, such axons can be traced both by eye and with segmentation algorithms within a section, and efforts to trace axons between sections are in progress.

The methodologies to be used for mesoscale axonal connectomics include fluorescence labels and high-throughput microscopy. For this connectome-projectome project, 50-80 percent of the 100,000 kilometers of myelinated axons in the white matter could be mapped using antibody staining for axonal proteins and ultimately cell type–specific markers. Further, 0.1-1.0 mm sections with modest expansion could be used to achieve 100 nm isotropic voxels and light sheet microscopy used for imaging. Importantly, all necessary methods have been established

and the microscopy platforms are relatively inexpensive. The postprocessing steps, such as data handling, stitching, segmentation, proofreading, databasing, and discovery, are routine in microscale connectomics and can all piggyback on approaches developed for the EM connectomics.

Undertaking the mesoscale human axonal connectome project would lead to several fundamental advancements. Although a single human axon has never been traced in its entirety, billions of axons could be traced in about 5 years. There is the potential to provide the first modern map of cortical and subcortical areas as defined by architecture and patterns of interconnectivity. An axon map would also have long-term clinical importance, with implications for next-generation neuropathology, a greater understanding of deep brain stimulation, and the many neurodegenerative diseases that have a white matter axonal component. This project would also produce a complete cell atlas for the human brain, eventually with molecular labeling for cell types, and would yield the largest biological network to date, leading to widespread impacts for network science, large-scale models of cognition, and artificial intelligence. Although the ultimate goal is to obtain an exascale atlas of the human brain, the intermediate goal of a mesoscale connectome-projectome would thus have tremendous and immediate value.

Mapping Projections of Melanopsin Retinal Ganglion Cells to Brain

Keun-Young Kim, PhD

Dr. Kim described a synapto-projectome collaboration aimed at determining the connectivity of melanopsin-expressing intrinsically photo-sensitive retinal ganglion cell (ipRGC) axons within the brain using genetically engineered version of EM probes and SBEM imaging. ipRGC projections to diverse brain targets mediate non-image-forming responses, such as circadian phase-shifting (via projections to suprachiasmatic nucleus [SCN]) and the pupillary reflex (via projections to olivary prefrontal nucleus [OPN]). However, very little is known about the synapto-projectome patterns because tools for labeling ipRGCs and imaging synapses are limited, and precise 3D EM of target brain structures is a challenge.

The strategy for this study entailed AAV-mediated expression of mini-Singlet Oxygen Generator (miniSOG)—a genetic probe for CLEM—selectively in ipRGCs and a multi-modal multi-scale imaging approach to localize and analyze ipRGC axons in target brain regions. Following miniSOG photooxidation, correlated light and 3D SBEM imaging were used to identify ipRGC axons in multiple regions, including SCN, OPN, and lateral geniculate nucleus (LGN). Because the intense heavy metal staining required for sufficient EM contrast made it difficult to track regions of interest in SBEM imaging, high-magnification X-ray microscopy was used to facilitate the identification of tissue landmarks in target brain areas to subsequently guide data collection in the SBEM.

SBEM reconstruction and analysis of brain target regions showed that ipRGCs form distinct types of axonal boutons in different targets areas. Double-tilt electron tomography of brain sections confirmed that synaptic swellings contained clear synaptic vesicles, mitochondria, and spiney intrusions. High-resolution analysis using SBEM of brain sections revealed that synaptic

boutons in five brain targets exhibit region-specific differences in bouton size, mitochondrial content, and the number of dendritic spine-like processes, which likely contributes to differences in synaptic strength. The largest boutons were in the LGN and contained large numbers of mitochondria and multiple spine intrusions, which likely ensures higher information flow through those synapses than synapses in other brain target regions.

This study demonstrated the in vivo utility of miniSOG as a genetically-encoded CLEM tag to label and characterize the ultrastructure of long distance projections and their plasticity. The methods developed in this project should propel other CLEM synapto-projectome studies.

Connectome at the "Operational" Resolution and in the Context of Protein Dynamics

Tianyi Mao, PhD

The mesoscopic connectome depicts connections across subregions of the brain at an "operational resolution." This resolution is often achieved by leveraging viral infection methods that can be directly translated to optogenetic or chemogenetic manipulation of circuits. A novel application of the mesoscopic connectome is to leverage the comprehensive nature of this approach to define otherwise ambiguous subregions. For example, Dr. Mao's lab employed unsupervised hierarchical clustering to quantitatively define the three classical striatal subregions—dorsolateral, dorsomedial, and ventral—as well as a previously unappreciated fourth posterior subregion that has unique input patterns with important functional implications. These mesoscopic connectomes can be generated in a high-throughput manner to functionally identify and precisely locate subregions across multiple contexts (e.g., developmental stages, sexes, diseases) and to detect stereotypic connectivity across individual brains; a whole-brain mesoscopic connectome can be processed in approximately 5 hours.

Mesoscopic connectomes sensitive enough to capture all possible axons and synaptic targets throughout the brain can be generated by a combination of methods, including viral injection to label neuronal somata, whole-brain clearing, antibody staining, whole-brain light sheet imaging, and algorithms designed to integrate large numbers of individual brains. Dr. Mao presented two sample visualizations of thalamo-cortico-striatal connectivity maps collected by these methods: a quantitative 2D representation of probabilistic connections across a series of coronal sections and a qualitative heat map of projections organized by subregion. Given that axonal projections are a necessary but not sufficient criterion for synaptic connections, these probabilistic connections can be used to guide functional circuit mapping tools to validate possible connections. For example, Dr. Mao's lab calculated the optimal injection site for subcellular channelrhodopsin-2-assisted circuit mapping based on probabilistic thalamocortical connections and demonstrated that 96 percent of the neurons in targeted presynaptic regions formed synaptic connections.

The Mao lab applies computational methods and genetic tools to further characterize synaptic connections within mesoscopic connectomes. For example, machine learning algorithms were used to differentiate axons in a manner that relates their morphology to their function (i.e., traveling vs. forming connections). In order to contextualize these static structural connections with respect to protein dynamics, synaptic proteins may be labeled at endogenous levels for

longitudinal imaging. This protein labeling may be achieved by either mouse ENABLED labeling or a more flexible CRISPR-based labeling strategy that could allow systematic labeling of many synaptic proteins.

The mesoscopic connectome is limited by its resolution, which is on the order of hundreds of microns for viral injections and approximately 1 µm for the imaging of axonal projections. Higher sampling rates (i.e., many overlapping injections within a region) may improve the resolution at the injection site. Because the resolution of mesoscopic connectomes is comparable to that of other scalable physiological and functional circuit mapping tools, these connectomes may be integrated with other datasets by computational methods to further improve resolution. For example, cortical data from the Allen Institute were analyzed with the Mao lab's thalamic data to produce connectome-level network information. Thus, data may be reanalyzed to extract new biological insights.

Dr. Mao enumerated several challenges and opportunities for mesoscopic connectomes, including (1) increasing the efficiency for physiological and functional validation of synaptic connections at the whole-brain level, (2) bridging the gap between structural connectomes and functional experiments, (3) creating a cytoarchitecture-based and user-friendly 3D reference atlas, (4) preparing tissue in a rapid and streamlined fashion suitable for scaling up, (5) integrating physiological properties and machine learning to better track axons and predict synaptic connections, (6) improving the visualization of multidimensional data, and (7) systematically labeling endogenous synaptic proteins for imaging in connectome-defined circuits.

Session 2: Projectome to Connectome Imaging—Strategies to Bridge Spatial and Temporal Gaps

Novel Mouse Genetic Tolls for Scalable Analyses of Brain Cell Morphology

X. William Yang, PhD

There is a pressing need for scalable tools to study mammalian brain cell morphology. Such tools can be used to study neuronal cell–type classification and changes in morphology of single neurons in the context of brain disease. To date, a very small number of neurons in the mammalian brain have been completely reconstructed. Dr. Yang described a suite of scalable mouse genetic tools for sparse, stochastic, and bright labeling of brain cells. These MOnonucleotide Repeat Frameshift Cre reporter mice are known as "MORF" mice.

MORF mice take advantage of long mononucleotide repeats, which are inherently unstable in a stochastic way and can undergo frameshifts at the genomic DNA level during DNA replication or repair. This phenomenon was leveraged to engineer constructs in which long mononucleotide repeats are located between the translation start site and a fluorescent reporter, such that the reporter would normally be out of frame and not expressed. However, in a small subset of cells, the repeat would undergo a frameshift and the fluorescent reporter would be expressed. This MORF construct was then put into a traditional Cre reporter context to achieve cell type–

specific labeling. The average labeling frequency in multiple MORF mouse lines is 2.2 percent, which results in the sparse and stochastic labeling of thousands to tens of thousands of cells per mouse with their complete morphology revealed.

Dr. Yang's group has developed several MORF mouse lines, two of which are particularly useful for the study of brain-wide neuronal morphology. MORF3 mice have a very bright membrane-localized Td-tomato immunoreporter: tandem "spaghetti monster" V5. MORF3 mice require anti-V5 staining and work well with all Cre lines tested thus far. TIGRE-MORF mice use EGFP-F as the reporter. TIGRE-MORF mice work well with inducible Cre lines (although not compatible with all Cre lines) and directly express bright GFP fluorescence.

Use of the MORF approach with different Cre mouse lines enables labeling, analysis, and reconstruction of the complete morphology of many types of neurons (including dendritic spines and axons) and glia throughout the brain. MORF3 is also compatible with EM, which creates the potential for use with multi-scale light and EM approaches to study neuronal ultrastructure.

A key question pertains to the imaging strategies that can be used to image MORF brains to gain insight into neuronal architecture and connectivity. One strategy is to cross MORF3 mice with a light-inducible Cre mouse to yield ultra-sparse labeling. This approach has been used to label only dozens of neurons per brain, which enables imaging of complete morphologies using high-resolution automated fluorescence imaging systems. A second strategy is to use sparse labeling and then utilize different imaging modalities to obtain images at different resolutions. For example, confocal microscopy can be used to resolve dendrites, and light sheet microscopy can be used to image dendrites plus partial axons and their targets. The brain tissue can undergo expansion and be imaged again using light sheet microscopy, to obtain the morphology of entire neurons.

Dr. Yang's goal is to develop a truly scalable and democratized tool that can be transferred to other labs and used to analyze neuronal types across many mouse brains. Scaling is facilitated by the suite of tools developed for tissue clearing, brain expansion, and light sheet imaging. Additional technological challenges for scaling this approach include the need for analytical tools that enable automatic reconstruction of neuronal morphology for large numbers of neurons. Current efforts are focused on using machine learning approaches that automatically generate dendritic reconstructions of up to several thousand neurons and a program that accelerates graph-based reconstruction steps. Finally, in an effort to democratize this approach, Dr. Yang's lab has developed a MORF mouse database, and all MORF mouse lines have been deposited at the Jackson Laboratory.

Towards Whole-Cortical Cellular Resolution Recording of Neuroactivity

Alipasha Vaziri, PhD

Dr. Vaziri discussed the development of optical neurotechnologies that are aimed at pushing the limits of speed and scale of neuronal population recording. He concurs with the consensus in the field that to obtain complete functional circuit models, a wiring diagram at the level of

synaptic resolution is needed. However, the agreed-upon resolution also must account for the resolution at which neural activity can be obtained. Although it is not necessary for these two resolutions to be the same, together they need to confine and inform possible circuit models of the brain.

Dr. Vaziri's work has focused on development of tools for whole-brain and large-scale recording of neural activity. Obtaining large-scale functional brain maps faces unique challenges, such as the volumetric imaging of 1 million neurons across cortical regions at cellular resolution. First, the imaging depth of 2P microscopy in the mouse cortex, for example, is limited to 0.8 mm because of light scattering in the tissue. While 3 photon (3P) microscopic approaches might extend the imaging depth by a factor of 2, 3P approaches would not solve the challenge of achieving large-scale functional recordings at high-speed in deep brain tissue. However, scaling up 2P within its depth reach to laterally cover the majority or entire mouse brain is a tractable engineering problem.

The creation of an imaging technology capable of large-scale volumetric recording requires careful and simultaneous optimization of a coupled multi-parameter space while minimizing exposure of the sample to laser light. By incorporating several key optical concepts into their design, Dr. Vaziri and his team developed a 2P imaging system for high-speed imaging of large brain areas that utilizes 30 axially-focused beams—referred to as "light beads"—to laterally scan the sample to generate a volume. Their imaging system enables recording of the activity of about 1 million neurons in both hemispheres of the mouse brain and across different layers of cortex at 2.2 Hz. This approach can also be applied at synaptic resolution, albeit with some sacrifice of volume size or temporal resolution.

One key limitation for the widespread use of multi-photon technology is the cost (approximately $500,000), which presents a critical barrier to dissemination in the user community. Although laser systems have historically been built around available molecular probes, leveraging biochemical and protein engineering to design probes around lasers could be a more cost-effective approach. For example, probes could be designed for use with fiber-based laser systems at 1,030 nm, which are a fraction of the cost of amplified laser systems. Moreover, the wavelengths of some of the current optimal fluorescent probes can also by imaged using less expensive fiber-based laser technologies. Cost reductions of lasers and other key hardware components could conceivably decrease the overall cost of such large-scale 2P systems by a factor of 10, thereby enabling democratization of this powerful functional imaging approach.

Next-Generation Connectomics (NGC) via Optical In-Situ Multiplexing

Adam Marblestone, PhD

Dr. Marblestone described an optical approach to connectomics that might have very favorable properties going forward as a next-generation technology. The key conceptual difference between this approach and EM is the greater information that can be obtained from each pixel by using an optical approach. In EM, structure is derived from tracing nanoscale membrane morphology such that each pixel contains information about whether or not it is part of a

membrane. In contrast, optical approaches can read out combinatorial color-codes from each pixel, which can provide more information content for each pixel. Such increased information can be achieved through cyclically assaying each pixel by sequentially labeling and imaging the sample using a series of fluorescent tags. This "in situ optical multiplexing" approach enables attainment of a combinatorial optical color code with an exponential number of possibilities for the label that is assigned to each pixel; specifically, with four fluorescence optical channels and 15 cycles of successive staining and imaging, there will be 4^{15} (i.e., billions) of effective colors, or labels, to which a pixel could be assigned.

Another technique known as molecular barcoding assigns a unique molecular barcode to each neuron. These barcodes can be combined with in situ optical multiplexing to assign each pixel to a particular cell, which would allow for computational filling of spatial gaps between parts of a cell. This approach may be leveraged to construct a wiring diagram that realizes a form of connectomics distinct from a full volumetric EM reconstruction. Likewise, in situ optical multiplexing technologies can also be applied to endogenous molecules proteins or RNAs (instead of barcodes) for molecular annotation of the connectome.

There are a number of potential goals for this next-generation optical technology. In the long term, such optical approaches could be used for dense connectomics. However, in the near term the greatest value might be gained by focusing on a synapto-projectome or sparse connectome, in which the synaptic connectivity of individual synapses over long distances can be obtained by bridging spatial gaps. Despite its inability to resolve every membrane, neuron, or glial cell, this approach would provide a great deal of spatial information and could be cell type–resolved, cortical layer–resolved, or dendritic compartment–resolved. Key advantages of this approach include reduced resolution requirements, easier sample handling, simpler computation, and lower cost.

Several use cases would be unique to this technology. One would be the ability to determine neuronal connectivity between distal brain areas without having to trace the intervening wiring to obtain an inter-areal connectome. Another use would be to link the molecular composition of synapses to the molecular composition of parent cells and to use that as a way to investigate the effects of genetic or drug perturbations.

Regarding resolution, a 50 nm resolution in all three axes would be required to identify individual synapses and barcodes. This resolution should be achievable with relatively conventional optical microscope instrumentation combined with 8- to 10-fold expansion microscopy. Data obtained using this optical approach should be able to be co-registered with calcium or voltage imaging data at the cellular level and should also enable the read-out of other time-dependent information, as is done for recording cell lineage trees into RNA with CRISPR. These optical technologies are similar to already disseminated and deployed optical microscopes equipped with fluid cycling; further commercialization of in situ optical technologies will help drive dissemination of optical connectomics. Although both centralized and democratized data collection are required to move this technology forward, a centralized and focused effort is required to develop, robustly integrate, and optimize this toolkit.

Session 2 Q&A

Color Labeling Approaches for Light, EM, and Correlative Microscopy

Advances in multi-color fluorescence labeling were discussed in the context of increasing the number of colors available for EM. To enable multi-color EM, Drs. Mark Ellisman and Roger Tsien made a family of lanthanide (Ln)-containing DABs, which can be detected using the relatively slow EELS technique. To render faster color EM imaging more amenable to scaling up, Dr. Ellisman's team is developing an approach that uses scanning TEM with multiple annular detectors in series. Dr. Marblestone's cathodoluminescence approach could by imaged using this technology, although not with probes that penetrate in 3D very well.

The speakers discussed potential strategies to increase the number of colors that can be used to label proteins at the EM level. Dr. Marblestone explained that the eventual goal of his optical approach is to achieve more than several colors for each sample through sequential staining and expansion. Dr. Yang commented that developing genetically encoded protein labels would allow visualization of the spatial distribution of different types of synapses or synaptic surface proteins by LM and EM, which would reveal a great deal about the potential properties of those neurons.

Dr. Ellisman suggested that another possible strategy to obtain more colors for EM is to combine the fluorescence labeling approaches described by Drs. Mao and Yang with the color EM labels developed by Dr. Ellisman's group (e.g., APEX). Dr. Yang was enthusiastic about this idea. While "spaghetti monster" fluorescent proteins are already compatible with EM, bringing the other color EM reporters into the context of the MORF approach offers great potential for imaging across both light and EM levels. He noted that the MORF model is a general concept that can be readily transferred to other proteins. Current efforts, for example, are focused on determining whether the MORF approach can be used to label endogenous synaptic proteins. Dr. Ellisman suggested that one solution to the problem of identifying electrical synapses would be to label connexins using this approach.

Dr. Mao linked this discussion back to Dr. Ellisman's question about whether CLEM studies should be a high priority for the mouse brain connectome effort, because this provides important context for the work being done on protein labeling. A key motivation for efforts to develop new protein labeling methods is to enable labeling of individual proteins at endogenous levels without using antibodies (i.e., without overexpression), with the goal of achieving high-density labeling of endogenous proteins. The strategy for achieving this goal is to label many protein molecules while also ensuring that the fluorescence labeling does not affect protein function. When such protein labeling is combined with expansion microscopy, a great deal of protein context can be achieved; combining this approach with EM would provide another dimension. A cell atlas, as described by Dr. Reid, with systematic labeling of a large number of endogenous proteins would provide an important framework for this work.

The value of sparse labeling was also discussed in the context of dense reconstruction. Dr. Reid suggested that sparse labeling is important for obtaining ground truth, absent a revolution in

segmentation of densely labeled structures and proteins. For EM, that revolution is already happening. When looking at available EM datasets, single cells can be selected and labeled computationally using what he refers to as "one-click Golgi." While this overall endeavor is called connectomics, a new key principles is that the imaging is dense. The ability to have one-click Golgi in EM and perhaps eventually "one-click HRP" to label axons for projection mapping is a major aim of the entire connectome/projectome endeavor.

Labeling Approaches to Analyze Synaptic Strength

Dr. Kim's multi-modal and multi-scale correlative microscopy approach for investigating the synaptic connectivity of ipRGC projections led to the identification of different classes of ipRGC boutons based on ultrastructural differences in size and synaptic properties and target brain region. Dr. Kim postulated that engulfed spiny intrusions play a role in modulating synaptic strength, such that a larger number of spiny intrusions increases the strength of synaptic transmission. She further hypothesized that the diverse types of bouton and spiny intrusion patterns parallel the diversity in response properties of each brain target to ipRGC inputs. Although a lot of progress has been made on the direct correlation of structural-functional data between synaptic strength, size, and shape in EM, much less work has focused on synaptic strength in light level approaches. Using expansion microscopy, the geometries and orientations of synapses can be visualized with sufficient expansion of the sample, although higher resolution than Dr. Marblestone's proposed approach would be required to count synaptic vesicles. Although the potential for light microscopic approaches to enable counting of vesicles is remote, a combinatorial approach to labeling different populations of vesicles could provide useful information. In spatial transcriptomics, a similar challenge arises regarding the de-mixing of multiple molecules based on their bar code sequence. Dr. Ellisman noted that while EM enables the counting of synaptic vesicles, there is also a substantial literature suggesting that active zones are a key governor of synaptic strength. He suggested that using fluorescent labeling approaches, particularly together with expansion microscopy, to label active zone–defining proteins or complexes might facilitate measuring synaptic strength.

Labeling Approaches to Analyze Neuronal Activity

Although genetic coding is already available for dynamics and colocalization of molecules across length scales, the possibility of using molecular "ticker tapes" that can synchronize a triggering start time from the time of sacrifice was considered. While Dr. Marblestone believes this approach is possible in principle, he cautioned that work in this area is much less mature than next-generation connectomics with sequential staining. Dr. Ellisman added that his group developed TimeSTAMP, a time-specific tag to measure the age of a protein, which can be used as a pulse-labeling approach for both LM and EM. TimeSTAMP was previously used on synapses and could be employed in an activity-dependent manner.

Another question is how to mark time in a population of neurons, such as through the use of ticker tapes to synchronize harmoniously firing events across the brain at the time of sacrifice. Dr. Marblestone believes that this approach might be possible by correlating activity with 2P imaging, along the lines described by Dr. Vaziri, if its speed becomes fast enough. Dr. Vaziri responded that imaging at synaptic resolution will come at the cost of speed or volume size

with current technology. However, Dr. Vaziri has not yet exploited combining spatial multiplexing with temporal multiplexing, which could further accelerate acquisition.

Mesoscale Connectomics in Mice and Humans

One mandate of the Brain Connectivity Workshops is to address the human and NHP connectomes, in addition to the mouse whole-brain connectome. There has been a focus on the different types of information that can be obtained from mesoscale and nanoscale connectomes. Dr. Reid emphasized that although a mouse is similar to a human when examining the finest scales of brain connectivity, humans are profoundly different from mice at the systems scale, particularly regarding the explosion of the forebrain in humans. Studying the human brain directly is essential for understanding human neurobiology and human medicine.

Several technical advances have been discussed throughout these workshops that are better enabling the study of human whole-brain samples. Previous speakers have highlighted several of these approaches, including Dr. Chung's suite of tools for brain clearing and robustifying brain tissue with elastic embedding, expansion, and staining and Dr. Hillman's light sheet microscopy technologies for imaging of large human brain samples.

Discussion Panel

Panel Discussion Topic: The Tyranny of Scale: Challenges presented by spatial and temporal scales

Panel Chairs: *Moritz Helmstaedter, PhD, Hongkui Zeng, PhD*
Discussants: *Moritz Helmstaedter, PhD, Hongkui Zeng, PhD, Anton Maximov, PhD, Kenneth Hayworth, PhD, Lisa Miller, PhD, and Mark Schnitzer, PhD*

During this portion of the agenda, discussants presented their views on the topic, commented on the presentations by the earlier speakers and other discussants, and considered the questions posed in the online Q&A.

Dr. Helmstaedter remarked on the tyranny of scale with regard to spatial scaling. While the reconstruction of cubic millimeter volumes is currently achievable and can reproduce most local connectivity, a dense mouse whole-brain connectome will provide insight into long-range connectivity (e.g., thalamo-cortical loops) that cannot be attained from millimeter-scale cubes. It will be a challenge, however, to maintain complete section interfaces throughout the whole-brain connectome. When sections are stitched together, axons must be reconstructed across sections, and some fraction of axons will be lost. The rate of loss must be commensurate with the number of total sections in the whole-brain reconstruction; for example, the loss of 5 percent of axons at an interface is far greater across 100 section interfaces than across 10. Thus, while the retention of 95 percent of axons may faithfully reconstruct the whole mouse brain across 10 mm^3 blocks cut at 250 nm thickness, this retention must be closer to 99.9 percent to produce a similar reconstruction across 100 sections, or even 99.999 percent across 40,000 sections. Moreover, the neuroscience community must be able to distinguish these error rates to accurately evaluate a connectome relative to its number of sections.

Dr. Helmstaedter acknowledged that it may be possible for computational reconstruction methods to compensate for incomplete interfaces if there is no serious data loss. Dr. Hayworth highlighted the need for a set of established gating criteria to define the degree of completeness that will be required before devoting substantial time and resources to a whole-brain reconstruction.

Dr. Zeng commented on the numerous important considerations for brain connectivity as they relate to challenges of scale. A single mouse brain contains nearly 75 million neurons. To capture the variety of the thousands of cell types and billions of synapses in the whole mouse brain, connectomics must operate at several orders of spatial magnitude—from nanometer to centimeter. Connectivity will vary across individual animals as well as more broadly across sexes and disease states. To probe circuit mechanisms, it will be necessary to obtain functional readouts at multiple spatiotemporal scales, including the individual cell and circuit levels across variable periods of time that reflect plastic changes during learning, development, and aging. It will also be important to consider brain connectivity in the larger context of the CNS and PNS as a whole in order to study how the brain perceives information and generates behavior, which will require analysis from sensory input to motor output. Dr. Zeng acknowledged that multiple connectomes will be required to address all of these considerations.

To focus efforts in connectomics, Dr. Zeng proposed that the first whole-brain connectome should be an EM volume from a naïve, wildtype, C57BL/6, P56 mouse with no environmental or genetic manipulations. This connectome could serve as a reference connectome for the entire neuroscience community. After this landmark connectome is completed, the second connectome should be a whole mouse head (i.e., including sensory organs) from a naïve, wildtype, C57BL/6, P56 mouse of the opposite sex of the mouse used for the first connectome. This second connectome would extend connectomics by deriving the complete inputs and outputs of representative neural systems that may be correlated with behavioral and functional studies already being performed in those systems. These first two connectomes would lay the foundation for many more connectomes to address a wide range of questions pertaining to brain connectivity, circuit mechanisms, and neural computation. Dr. Zeng believes that these reference connectomes are achievable within the coming decades and would revolutionize neuroscience.

Dr. Zeng acknowledged that although C56BL/6 mice are a useful strain for the first connectomes because of their popularity for neuroscientific research, phenotypic challenges (e.g., progressive hearing loss) could influence the integrity of the connectome. An alternative subject for the first connectomes that may mitigate these phenotypic issues could be an F1 hybrid mouse. Dr. Ellisman added that a wild mouse should also be considered, because the neural manifestations of sensory loss are substantial.

Drs. Schnitzer and Yang agreed that Dr. Zeng's proposed connectomes would be excellent resources for a variety of subsequent applications (e.g., human health, computation, artificial intelligence). Dr. Schnitzer added that the most powerful resources would allow researchers to merge their own neural activity datasets with their own connectomic datasets, because they could be more readily adapted to particular questions than a static online database.

Dr. Maximov's approach to connectomics entails imaging limited, relatively small volumes with genetic manipulations. With this approach, his lab demonstrated that the structural diversity of chemical synapses is largely intrinsic by systematically silencing glutamatergic neurotransmission during development and comparing the structural makeup of synapses with and without histories of activity. Dr. Maximov remarked that inactive circuits are a useful frame of reference for the study of ultrastructural changes that correlate with learning. Genetically encoded reporters may be selectively introduced to neurons based on their history of activity (e.g., through a c-Fos-based system) in order to image and reconstruct those cells that are active during specific cognitive tasks. This can be performed at different spatial scales, from imaging of whole cleared brains to nanoscale microconnectomes. The goal of Dr. Maximov's work is to identify long-lasting ultrastructural signatures of activity that arise during memory acquisition or sensory experiences.

Dr. Helmstaedter asked how well a static connectome can capture the highly dynamic nature of brain circuitry. Dr. Maximov acknowledged the lack of a straightforward answer to this question; the relationship between different synaptic and subcellular components as they pertain to dynamic processes (e.g., long-term potentiation [LTP]) is still not fully understood. Moreover, plastic mechanisms are often transient and the extent to which sustained changes are observed at an ultrastructural level is unclear. Nonetheless, Dr. Maximov suggested that LTP is a convenient avenue to pursue these questions, given its relevance to learning and cognition.

Dr. Hayworth explained that the ultimate goal of neuroscience is to understand how neural computations give rise to intelligence. Neuroscientists currently have only a crude understanding of the intricate wiring that enables these computations in the mammalian brain. Dr. Hayworth believes that this lack of knowledge is the key impediment to neuroscientific progress, and that a complete mouse connectome could directly address this obstacle. However, Dr. Hayworth remarked that the immense challenge of a mouse whole-brain connectome cannot be achieved with the current paradigm in which the Z-resolution is inextricably linked to section thickness. Thus, Dr. Hayworth and colleagues are developing a new method that decouples Z-resolution from section thickness: gas cluster ion beam multibeam SEM (GCIB-MultiSEM). In GCIB-SEM, semithin sections (100-1,000 nm) are collected onto wafers. Block-face ion milling is leveraged to obtain higher Z-resolution within each thick section. After the surface of a section is imaged by multibeam SEM, the wafer is automatically transferred to an ion milling station to remove 10 nm from the surface. This imaging and milling cycle is repeated until the entire section has been imaged. The resolution achieved by this method is sufficient to reliably trace axons across long distances. Dr. Hayworth suggested that, with the incorporation of proper automation, the entire mouse brain could be thick sectioned and collected onto wafers for imaging by a fleet of multibeam microscopes with zero losses, resulting in an automatically traceable volume of the whole mouse brain.

Dr. Miller commented on the potential contributions of X-ray imaging to connectomics and reiterated her comments from previous workshops: imaging the connectome will require a multi-modal approach. X-ray sources offer a complementary imaging approach that can bridge

the gap between the unsurpassed resolution of EM and the wide variety of elegant tagging, labeling, and staining techniques offered by LM. X-rays are capable of nondestructive imaging across spatial scales from meters to nanometers due to their long penetration depths and short wavelengths.

Dr. Miller described a sample workflow for multi-scale brain imaging that incorporates X-ray techniques at every step. The workflow begins with hierarchical phase contrast tomography (HiP-CT), a new technique that was recently used to nondestructively image an entire mouse brain in 3D with 1.4 µm spatial resolution in a matter of hours. HiP-CT can be used to assess tissue quality at early stages of processing. At the next step, X-ray imaging can be combined with LM on thicker brain sections to merge the information obtained from visible fluorescent tags with the element information generated by X-ray fluorescence. This multi-modal imaging can also employ new genetically encoded double tags that combine fluorescent reporters with LBTs, which allow visualization of individual proteins with visible light, X-rays, and EM at once.

In preparation for EM imaging, micro-CT may be leveraged to observe the quality of osmium staining. In addition, K-edge subtraction micro-CT can provide added contrast to a tomogram by imaging above and below the X-ray edge of osmium. Tissue can then be imaged by nano-holography. This X-ray technique can produce 20-200 nm resolution on osmium-stained, resin-embedded blocks of tissue, and was recently used to achieve a resolution (87 nm) sufficient to trace individual motor axons from muscles to the CNS of the fruit fly. Importantly, nano-holography can be used as a template for and accelerator of the subsequent collection of high-resolution EM images from the same tissue because the technique is nondestructive and follows the same sample preparation protocols as EM.

Dr. Miller added that next-generation synchrotron light sources have the design characteristics of brightness, coherence, and color to enable the construction of beamlines that can accommodate the techniques described in this workflow. Beamlines can be designed for full-field and scanning microscopy, and images can be collected with pink and monochromatic beam.

Dr. Helmstaedter remarked that X-rays are valuable for their rapid and noninvasive qualities, but wondered whether it was feasible for X-ray imaging to replace the need for dense EM reconstructions. Although X-ray optics continue to improve, Dr. Miller emphasized that the real benefit of X-ray microscopy lies in its nondestructive nature, which allows LM or EM to be performed on the same tissue samples that have been imaged by X-rays. Thus, X-rays are unlikely to replace EM but rather can be leveraged as a complementary approach across a range of spatial scales (i.e., from full brain to nanometer resolution).

Dr. Miller acknowledged that some of the highest resolution X-ray imaging (e.g., 5-10 nm) has been performed on materials with higher contrast than a stained mouse brain and that reasonable resolution limits for X-rays are probably closer to 20 nm. More work in sample preparation is needed to improve the contrast of stained tissue, which will in turn improve spatial resolution. Dr. Lee added that machine learning techniques may be leveraged to improve resolution by upsampling.

Dr. Schnitzer commented on how large-scale imaging of neural activity can be democratized. By way of example, he described two major classes of next-generation multiphoton microscopes that can be used to image neural activity in multiple brain regions, both of which emerged in the past 5 years: 2P mesoscopes and robotic 2P microscopes. 2P mesoscopes have a large FOV for imaging contiguous brain regions. Robotic 2P microscopes have a narrower FOV but can image two or more distal brain regions. Despite the promise of these technologies, very few biological studies have been published beyond initial demonstrations of these approaches. Dr. Schnitzer suggested that multiple obstacles, including a lack of streamlined and standardized workflows as well as analytical challenges, result in the underutilization of these sophisticated imaging platforms. To mitigate these obstacles, Dr. Schnitzer proposed the establishment of national facilities to provide in vivo brain imaging datasets aligned with connectomic data. Such facilities represent a cost-effective way to operate these imaging platforms on a consistent basis that would be difficult for individual labs to sustain. Moreover, these facilities would provide a standardized means of aligning activity to connectomes. Dr. Schnitzer believes that these facilities are an inevitable component of the journey toward democratized participation in large-scale neuroscience and would produce more biological insights than the current paradigm.

Dr. Schnitzer recommended adoption of a portfolio of approaches to connectomics, so that multi-modal datasets will be available to researchers to compare their own data and models to different portions of the connectome. These resources may include a few mouse whole-brain connectomes accompanied by partial complementary connectomes. It will not be essential for each of these resources to be collected at nanoscale resolution, and therefore contributions could be made from more distributed entities that may not have the same resources as a centralized facility.

Follow-up Discussion

Dr. Reid emphasized that a major motivation for building the mouse whole-brain connectome is to gain further insight into neural cell types. The cubic millimeter volumes that have been collected to date already demonstrate extraordinary cell type specificity in neural connectivity that will be interesting to analyze in the context of their morphological and transcriptomic identities. Dr. Zeng also commented that it will become essential to connect static connectomic maps to the molecular identity of cells via multi-modal techniques to label, image, monitor, and manipulate cells in vivo. Dr. Ellisman added that methodological barriers have limited current understanding of the full range of cell types and molecular characteristics that can be linked to the connectome; for example, carbohydrates remain poorly mapped because they are difficult to label. Dr. Reid agreed, adding that it would be most beneficial for the initial connectomes to not include any sort of overlay with molecular data.

Workshop 4: Connectome Generation and Data Pipelines

Precis

This workshop focused on computational techniques and human-computer interactions that produce useful data structures (graphs, identified cell types, etc.) from raw EM data. State-of-the-art methods for stitching, aligning, and reconstructing neural morphology and connectivity were presented, along with challenges for scaling these computational techniques to the exascale (needed for the whole mouse brain).

Volume assembly (i.e., automated transformation of raw EM sections into coherent 3D volumes) has seen enormous progress in recent years, but, in order to address the whole mouse brain, key improvements must be realized. In particular, methods for decomposing montage stitching and serial section alignment into a hierarchy of independent problems must be further developed, which would enable scaling up to much larger datasets. Automated and semi-automated tools for detecting volume assembly errors at various stages of the pipeline should also be further developed. Integration of lower-resolution imaging modalities, such as X-ray microtomography, could also be used as a rough guide for EM alignment, which will be particularly valuable at the scale of the whole mouse brain.

Neuron reconstruction represents one of the most challenging analysis tasks to fully automate; progress has been exponential and state-of-the-art methods can reach 10 mm “expected run length” (a 100-fold improvement from 0.1 mm in 2015). For the whole mouse brain, another 10-fold improvement is likely needed, which is plausible within the next 5 years given general advances in machine learning and intensified efforts in connectomics specifically. Efforts should be made to evaluate the type of imaging errors/defects that have the greatest and least impact on automated reconstruction quality. Improvements in neural network inference efficiency are currently meeting or exceeding a Moore’s-law scaling curve, meaning that computational requirements for this step of analysis are not considered especially concerning.

Methods for determining cell types and synapses and for querying the resulting connectome data are in development. However, given the relative novelty of this type of analysis and related workflows, standards are not yet in place for exchanging data, nor is it clear which types of analyses are most fundamental. Therefore, as connectomics evolves over the next years, the community should focus these efforts and ensure that as conclusions emerge on these issues they are transformed into specific requirements for whole mouse brain–scale analysis tools.

The use of human proofreading is a fundamental part of the connectomics workflow and has been deployed in both expert-centered environments as well as large-scale citizen proofreading environments, both of which will prove useful going forward and will undoubtedly need refinement for a 1,000-fold bigger scale of the whole mouse brain.

In summary, many of the existing tools can potentially be scaled but will require effort (scaling, testing, validating). In particular, the automated accuracy and reliability of each step will need to dramatically improve in order to keep the overall human involvement manageable. The

fundamental requirements related to computational resources are not especially concerning, given overall progress in exascale compute and storage capabilities, but plans will need to be made to include access to such facilities specifically for the whole mouse brain project.

Introduction

Nicola Ferrier, PhD, and Viren Jain, PhD

The goal of this workshop was to address the technological approaches, innovations, and advances in computational techniques and human-computer interaction that can turn petabytes of raw EM data into data structures that biologists can use and understand. This area is mission critical for the overall goal of brain mapping because a whole-brain connectome cannot be achieved without the requisite computational techniques that can process the data. This field involves a diverse set of areas within contemporary computer science and engineering, including machine learning, computational photography, vision, compression, scalable compute frameworks, software, hardware accelerators, and visualization.

A pioneering history of computational approaches is being developed within the context of brain connectomes. Early efforts by Sidney Brenner and colleagues to map the complete nervous system of *C. elegans* in the 1970s—although inspiring—made clear that the amount of human effort required to trace the connectivity data was unsustainable and not scalable. By the late 1970s, there was emerging interest in using computers to assist in the reconstruction of EM data, although at that time it was not even possible to store one EM image on a computer, much less to process EM images in any compelling way. Thus, the goal of using computers for analysis of EM data in a comprehensive way fell by the wayside for many years.

The mid-2000s saw significant progress in the automation of EM data acquisition techniques for volume EM. The pioneering approach to applying computation to this problem resumed at that point. There were substantial advances in many areas, including registration of sections, application of convolutional networks, combination of human and computer effort, automated identification of neurotransmitter types, correction of computer-automated reconstructions use of advanced visualization techniques, and new ways to implement workflows for data processing using modern computational frameworks. This pioneering spirit must continue as the field looks toward scaling up at least 500-fold to the mouse and, in the future, beyond that to the human.

Workshop 4 was divided into four sessions that addressed the following topics: volume assembly from raw data, automated reconstruction and annotation, proofreading and verification, and high-performance computing and infrastructure. For each session, the speakers addressed three questions:

1. What are the prospects and challenges you see in scaling your approach (and/or analysis goals in general) to volumes comprising hundreds of petavoxels of imaging data?
2. What kinds of general improvements in technical infrastructure (hardware and/or software) would mostly dramatically aid in scaling your analysis?

3. What new frontiers of analysis (especially as may be enabled by, or complementary to, a whole mammalian connectome) are you most excited by? What challenges do you see in pursuing those frontiers?

The Discussion Session focused on exascale computing and its relevance to the brain connectome project. The presentations can be viewed by accessing the archived NIH VideoCast.

Session 1: Volume Assembly (Stitching, Alignment, Registration)

Volume Assembly in the Context of a Large-Scale Connectomics Pipeline

Gayathri Mahalingam, PhD

Dr. Mahalingam discussed the stitching and alignment component of EM volume assembly that is used within the Allen Brain Institute's petascale EM reconstruction pipeline. This approach, adapted from the stitching and alignment pipeline developed at Janelia and applied to the whole fly brain, involves steps to stitch raw tiles into montages and to roughly align montages in 3D and has features that are promising for scalability. The roughly aligned volume is then finely aligned by Dr. Macrina and colleagues, as discussed in the third presentation of this session.

Several software improvements facilitated scaling to efficiently process petascale datasets and could be scaled further. These scalable software tools include (1) adaptation of existing tools to perform efficient data transfer for a multi-scope infrastructure; (2) modularized tools for parallelized processing of sections; (3) a set of tools for semi-automated quality control (QC) assessment of stitching quality; and (4) a scalable matrix-based registration system to allow processing in a distributed computing environment. The Allen Brain Institute team also developed a workflow manager to enable automated data processing and reduce manual human intervention. Further, the team is implementing a graphics processing unit (GPU)-based system on the EMs to enable computation of lens distortion corrections and image montaging during the image acquisition process, which will significantly accelerate the stitching and alignment steps of the volume assembly pipeline.

Additional challenges remain to be addressed for scaling up the stitching and alignment process to exascale datasets. First, because high-quality stitching is critical for the success of subsequent pipeline steps, metrics to evaluate quality must be determined based on downstream processing requirements. Second, because QC is essential throughout the pipeline and manual QC is by far the most time-consuming step, automation of QC should be considered throughout the entire pipeline. However, automating QC would require constant adaptation to new issues that evolve with different datasets, such as changes to the acquisition system or tissue sectioning. The potential for data misalignment caused by poor data quality (e.g., minimal or no overlap between tiles) highlights the importance of evaluating data quality early in the pipeline. To evaluate data quality early and to avoid potential downstream problems, the Allen Brain Institute team uses a comprehensive set of QC tools to assess and provide feedback about image quality during imaging. This set of tools has been optimized to run on GPUs and could easily be scaled up to larger datasets.

For scaling up to whole-brain volume assembly, a brain sample will likely be subdivided prior to imaging. Within a subdivision, the stitching and alignment process described by Dr. Mahalingam will likely scale well with sufficient computational resources. A potential strategy for aligning subdivided volumes is to use morphological cues (e.g., blood vessels, nuclei) obtained using other modalities (e.g., X-ray microtomography) as a template for sub-volume alignment. In addition, the newly implemented GPU-based on-scope stitching system will speed up processing 10- to 100-fold, further reducing computational time and cost.

Scalable Stitching and Alignment for 3D Electron Microscopy

Stephan Saalfeld, PhD

Dr. Saalfeld, at Janelia, discussed the current state of some stitching and alignment pipelines for connectomics and what will be needed to scale these approaches to the whole mouse brain. The typical workflow for stitching and aligning 3D EM entails applying the following processing steps to image tiles: (1) flat-field correction and local contrast normalization (to reveal meaningful content), (2) field lens distortion correction, (3) image stitching into montages, (4) automatic sorting of stitched montages, (5) serial section alignment, (6) axial distortion correction, and (7) global contrast alignment across sections.

The scaling issues for a mouse whole-brain volume compared with current large-scale datasets depend on the imaging resolution that will be used. A resolution of 5 nm would generate 4 exavoxels of data, 1.6 million sections, and 200,000 tiles per section, whereas a 40 nm resolution would generate 8 petavoxels of data, 200,000 sections, and 3,000 tiles per section—which is on par with large-scale datasets that are currently being processed, such as the FAFB.

The stitching and alignment steps that would benefit from efforts to improve scalability are montage stitching and serial section alignment, because both of these steps consider all data at once and minimize global objectives. In order to truly scale these steps, approaches must be developed to simplify and decompose these methods into a hierarchy of independent problems. The frequency of errors and ability to detect and correct them must also be improved.

The stitching process starts with identifying corresponding landmarks in neighboring images. In recent work on a large FIB-SEM dataset, the matching density in empty resin was improved by matching constellations of small particles (e.g., fluorescent beads), which can be used as landmarks to facilitate image alignment. If automatic landmark correspondences are not dense enough, dense optic flow can be calculated using block-matching or deep neural networks that can be trained to focus on relevant parts of the signal and ignore the noise. An important aspect of this process is the minimization of false estimates and automation of rare error detection, to avoid the need for human intervention.

A similar strategy is used for serial section alignment. In order to register a section series correctly, content is estimated across images and then a deformation is found that minimizes local distortion to prevent accumulation of distortions across the section series. Although solving this problem globally for an entire stack does not scale, overlapping sub-series can be

independently aligned, and the aligned sub-stacks rigidly aligned to each other and interpolated to generate a perfect alignment of the entire series. This approach can be parallelized on a large compute cluster. A similar strategy could also be used for stitching montages.

Dr. Saalfeld outlined several key considerations for scaling the stitching and alignment pipeline:

- Scaling up current approaches is possible, but it will require some engineering.
- Error minimization and detection are crucial and will require more attention.
- Implementing ensembles of complementary methods make stitching and alignment more robust.
- The parallelized approach for aligning serial sections could also be applied to stitching mosaics.
- Embedding brain samples in textured material to highlight boundaries and the use of particle constellations would help overcome the obstacle of stitching and aligning disconnected islands of brain tissue.
- Sample preparation should aim to facilitate image processing and avoid unnecessary ambiguities through high-quality staining and imaging with sufficient overlap.
- Interactive tools are crucial. Multi-scale cloud store, memory caching, and interaction transformation with BigDataViewer is the best interactive tool-chain at present.
- Healing tissue artifacts in images needs additional work.

Alignment of Serial Section EM for Automated Reconstruction

Thomas Macrina, BSE

The many advantages of serial section EM (ssEM) make it a highly desirable technique for automated reconstruction. However, automated reconstruction of ssEM datasets faces some challenges because the many tissue defects, specifically cracks and folds, caused by tissue handling lead to reconstruction errors. Because of extensive neuronal branching, each neuron might encounter thousands of folds, and traditional approaches that entail manually review of those errors are not feasible in petascale datasets. To overcome these challenges, Mr. Macrina, CEO of Zetta AI, and his colleagues developed an alignment pipeline that vastly improves automated reconstruction of ssEM datasets.

This novel pipeline for scaling automated reconstruction of ssEM datasets entailed complete rebuilding of the alignment process by utilizing convolutional networks and deep learning to directly compute a displacement for every pixel to obtain a dense displacement field. The image is then warped by the predicted displacement field, which results in well-aligned neighboring images with overlapping objects. Extending this approach to align every image in the dataset requires obtaining a displacement field for every image. The serial dependency of this approach was overcome by parallelizing serial alignment into subsets, called blocks, and using the convolutional network to align each block to its predecessor, thereby enabling alignment of petascale datasets with dense fields in a reasonable timeframe. Importantly, the influence of stitching fields was decayed over distance, so that a given section is only influenced by a limited number of previous blocks. This decay eliminates the global dependency of previous techniques, which makes this method highly scalable. To date, Mr. Macrina and

colleagues have used this pipeline to align four petascale datasets, including the MICrONS mouse cortex and the female adult fly brain TEM dataset (which is the basis of Flywire), and have set up this pipeline within Zetta AI.

For the exascale mouse whole-brain connectome, the brain sample will likely be subdivided for imaging. Within each subdivision, the pipeline described by Mr. Macrina will be highly scalable. With enough compute, alignment time is constant and costs are comparable to other segmentation approaches, which scale linearly with the volume. Challenges for scaling the pipeline to exascale include stitching subdivided brain samples, tighter feedback on quality control, and the dependency on rough alignment. Specifically, without global dependency this reconstruction process cannot remove long-range distortions such as shear. One improvement could be to first image the volume at micron resolution with X-ray microtomography and then use those images for rough EM alignment. This multi-modal approach leads to the near-term frontier of inter-areal connectomics, where axons in the X-ray are traced between two EM volumes.

Session 2: Automated Reconstruction and Annotation (Segmentation, Synapses, Types, Compartments)

Automated Dense Reconstruction of Neural Tissue

Michal Januszewski, PhD

Segmentation is one of the first problems that must be solved in the process of extracting information from volume EM datasets. To segment EM data, voxels must be classed into objects within single sections and also extended to 3D to produce volumetric reconstructions of cells and their neurites. Challenges for the reconstruction process are to minimize unintentionally tracing neurites into other nearby neurites, which would cause a merge error, and to extend neurite branches as far as possible without creating false splits. Dr. Januszewski and the Connectomics group at Google developed a new neural network approach—"flood-filling networks" (FFNs)—that automatically performs the segmentation process by tracing neurites one object at a time.

In the FFN segmentation approach, an object is defined in a small region of interest in the EM data, and the FFN creates predictions about whether neighboring voxels belong to the object. This process is continued in 3D until the FFN finds that no additional voxels should be added to the current object and no more unsegmented data exist with which to create new objects. Unlike other approaches, the FFN segments the image directly without creating an intermediate representation that would require further processing. In practice, the reconstruction process also takes advantage of multi-scale processing to first segment down-sampled versions of the images followed by processing higher resolution versions of the images to achieve more complete segmentation, which can increase the processing speed by more than an order of magnitude. This strategy improves both the accuracy and performance of reconstruction.

The scaling of the FFN segmentation approach has advanced substantially since it was first introduced, as evidenced by both the quality and volumes of the reconstructions. Reconstruction quality is measured using a metric called "expected run length" (ERL), which quantifies how far the machine learning models can trace neurites before encountering either merge or split errors, as determined by comparing automated to manual reconstructions. When FFNs were first used in 2016, the best ERLs reached 1 mm; by 2020, the ERLs extended to about 10 mm, which is sufficient to reconstruct highly complex dendritic and axonal arbors. Between 2016 and 2020, the volumes that can be reconstructed have also increased 1,000-fold—a similar scale to what is currently possible for large-scale connectomics.

Dr. Januszewski highlighted several trends in computation that are highly relevant to the planned mouse brain connectome project. First, the costs of both computational power and storage are declining exponentially over time, with about a 10-fold reduction in computational costs every 5 years and a steady but slower decrease in storage costs. This differential in cost reduction creates not only an exponentially increasing amount of computer power per unit of data that can be stored, but also an ever-widening gap between what can be computed and what can be stored. Ongoing work to apply data compression techniques will help to address this potential problem. Second, continued progress in machine learning to develop better models and training techniques (e.g., self- and weakly-supervised learning) will facilitate optimization of segmentation approaches.

Despite recent progress, several major challenges persist. First, the quality of the reconstruction process remains a primary challenge in the analysis of large connectomics datasets. With current methods, the time and cost associated with human proofreading would be immense. Further development of automated methods will help mitigate this issue, but will require testing of new methods on ever larger datasets and the availability of sufficient high-quality, human-verified ground truth data. Second, because reconstruction relies on the quality of the inputs, the robustness and reliability of sample preparation and acquisition methods must be improved in order to minimize errors in reconstruction. The planning of large-scale projects should consider the potential for unexpected sample acquisition errors (e.g., section loss, tissue deformations) and build in a margin of safety to mitigate the risk of reconstruction errors. Looking ahead, with some engineering and a non-trivial cost, Dr. Januszewski believes that a whole mouse brain, which is 500-fold larger than reconstructions to date, could be reconstructed with existing methods. Moreover, by the time the brain volume for this project is imaged, the computational resources for segmentation will become less expensive and faster.

Synapse Detection in a Fly Brain

Julia Buhmann, PhD

Identification of both neurons and synapses is required to extract the underlying neuronal circuit from EM images. Dr. Buhmann presented an approach for the automatic detection of synaptic partners in *Drosophila*, which was developed and utilized for synapse identification in the FAFB dataset.

The "Synful Method" is a direction-based method based on a U-net, which is a specific type of convolution neural network (CNN). The U-net predicts two channels—the presence of postsynaptic sites and a direction vector field with vectors pointing to presynaptic sites—which when merged makes extraction of synaptic partners straightforward. To enable analysis of the entire FAFB dataset, initial methodological developments by Dr. Buhmann and colleagues for automatic synaptic partner identification were adapted to run on large-scale datasets by building up the infrastructure to process datasets blockwise, increasing algorithm efficiency and accuracy and using more computational power (80 GPUs). Overall, these adaptations to bring this method to scale decreased the processing time for the FAFB dataset from 27 years using their initial method to 3 days.

To address the accuracy and performance of their analytical approach, which is crucial for scaling to large datasets such as a whole brain, Dr. Buhmann and colleagues evaluated the Synful Method on several *Drosophila* brain regions by mapping the predicted synaptic partners to ground truth neurons in which synapses had been previously segmented. Comparing the predicted and ground truth synapse counts confirmed that the Synful Method was highly accurate across brain regions (92-96 percent). Evaluation of this approach's performance with different types of neurons showed that there were differences in detection precision for synapses from different cell types and that synapses from the same brain region as the training dataset were predicted with higher accuracy. Moreover, this analytical method was also capable of classifying synapses by neurotransmitter type.

Additional improvements to the Synful Method will be required to scale this approach to the whole mouse brain, which is 10,000-fold larger than the fly brain (and would take 82 years to analyze using the current approach). One straightforward solution for further scaling is use of more GPUs. In addition, using more diverse training datasets will lead to more robust algorithms and improving the efficiency of the algorithms will save energy, money, and time.

Towards a Parts List for a Whole Mouse Brain

Nicholas Turner, MA

Mr. Turner looked just beyond the initial reconstruction effort to consider how to develop a parts list for the mouse whole-brain connectome. Understanding cell types is crucial for a clear understanding of brain function. Reconstruction of a whole mouse brain would generate an incredibly vast sample of neuronal morphology and connectivity. Parsing useful neuronal types from this wealth of data will be a challenging although crucial step toward extracting real insight. Most cell type classifications rely on judgements of human experts, which, while generally useful, cannot be scaled to the whole mouse brain. However, most of this work to date has not had access to the connectivity information that can be extracted from a connectome, which could be leveraged to help inform these cell type classifications.

"Quantitative morphological clustering" is an approach that quantifies cellular morphology using a set of predefined features over hundreds of cells and standard clustering algorithms to parse putative cell types. The features for clustering are often computed from derived representations of cell morphology, such as a computational neuronal skeleton (which

represents the topology of a cell's neurites using a graph of vertices and edges) or a computational mesh (which sparsely represents a cell's 3D morphology). A main challenge to computing features is how to actually derive the representations. This task would be difficult to scale up to the full brain, although recent work of the Seung lab has drastically increased skeleton computation efficiency as part of the IARPA MICrONS program. Scaling up a candidate approach to generate the skeletons would require a large enough compute cluster to store a few petabytes of data in readily accessible memory or fast disk storage, as well as ready access to these derived representations to tune different approaches and read results.

Another notable approach to clustering neuronal morphology is NBLAST, which has been used for preliminary clustering of tens of thousands of cells in the fly hemibrain dataset. This method uses a derived representation similar to a skeleton, but compares neurons on a point-by-point basis rather than with a suite of summary features, computes a similarity score between new cells and a database of pre-defined cell classes, then uses iterative clustering to group similarity scores. The scaling properties of this approach are promising, although deriving a suitable classification database may be difficult across large nervous systems.

To overcome the limits of cell type determinations based solely on morphological clustering, a complementary strategy is to leverage the connectivity information within the dataset. For the fly hemibrain, this strategy was combined with NBLAST to enable further subdivision of initial clusters based on connectivity using human feedback. These types of analyses have raised important questions for petascale and exascale analyses, including how much human input will be needed to extract insightful structure and how much insight can be gained with basic cell type distinctions. In addition to providing connectivity information, these neuronal reconstructions are a rich source of new cellular and morphological data, and current representations and approaches might not highlight the best features to reliably distinguish cell types. Leveraging machine learning techniques may also help advance the capabilities of cell type classifications.

In direct response to the three overarching workshop questions, Mr. Turner believes that the prospects of a mouse whole-brain connectome are immense, although extracting enough faithful reconstructions, computing useful derived representations, and extracting the "right" features will be major challenges. Technical infrastructure improvements to aid in scaling include having enough compute resources to create useful representations, flexible access to derived representation for each reconstructed cell, and systems for human interaction with automated clustering results. Reliably automating these steps will require significant innovation. Having a vast sample of cells will likely be crucial for discovering useful features, and reliably parse cell types and adding connectivity information to large-scale cell typing may be transformative.

Session 3: Connectome Proofreading, Verification, Crowd-Sourcing

Proofreading the *Drosophila* Hemibrain Connectome

Patricia Rivlin, PhD

Dr. Rivlin, Manager of the flyEM Project proofreading team at Janelia, described the proofreading process used for the *Drosophila* hemibrain connectome. The hemibrain connectome is the largest and most detailed connectome of any brain to date, consisting of 25,000 neurons and 25 million synaptic connections, and it took 50 proofreaders from three teams more than 2 calendar years to complete. The large size of the hemibrain connectome relative to previous "densely proofread" connectomes from FIB-SEM volume datasets required a more than 50-fold speed-up and advancements in all aspects of the reconstruction pipeline. For proofreading, this work entailed developing proofreading software and protocols as well as increasing the proofreading workforce.

Proofreading is essential because every segmentation will contain errors, even with the high quality of ultrastructural preservation and state-of-the-art machine segmentation of the fly hemibrain. The roles of proofreading in the dense connectome are to detect and correct segmentation errors (i.e., falsely merged or split segments), validate automated synapse predictions, validate cell type groupings, and provide feedback to software developers to optimize tools.

To generate a dense connectome, the flyEM proofreading team designed an iterative proofreading strategy to refine the reconstruction, starting with the largest segments and progressing from coarse to fine processes. The high quality of the starting segmentation and availability of synapse predictions allow for ranking of segments by numbers of synapses. A connectome can then be extracted at different stages of completeness (the percentage of pre- and postsynaptic connections associated with identifiable neurons). To facilitate this iterative strategy, four proofreading workflows are used: cleaving, to correct false merges; false split review; focused merging; and orphan linking. Workflows 2-4 correct false splits. These workflows are implemented with software developed specifically for proofreading of visualization (i.e., NeuTu, Neu3, and Neuroglancer) and other specialized tools (e.g., Focused Merging). In addition to correcting segmentation errors, expert proofreaders also validate cell typing as neurons are reconstructed. Finally, to coordinate proofreader effort, personnel and analytics are required for data management, monitoring, and streamlining assignments and training.

To address the three overarching questions for this workshop, Dr. Rivlin noted that scaling any process is difficult and requires good engineering. She recommended optimizing proofreading approaches on connectomes of intermediate scales, as in flyEM, before attempting the whole mouse brain. The mouse whole-brain connectome may require multiple proofreading teams at multiple institutions, which raises the issue of uniformity across teams. Any improvements to allow fast fetching of high-resolution 3D meshes and to support greater numbers of proofreaders working simultaneously on the same dataset would be helpful. Regarding new

frontiers of analysis, Dr. Rivlin is particularly excited about the possibilities of comparative connectomics between species and animal models of human disease and of adding cell biological data (e.g., mitochondria) to a connectome. The potential to add cell biological data obtained at high resolution raises the question of whether obtaining cell biological information from large-scale volumes can be achieved using destructive methods such as FIB-SEM, or will require nondestructive methods such as TEM so that sections can be re-imaged and re-analyzed as needed.

Connectome Querying and Verification

Joergen Kornfeld, PhD

Dr. Kornfeld discussed lessons learned about querying EM connectome datasets, query result verification, and scalability of queryable connectomes to a whole mouse brain. To make an EM dataset queryable the task is to create a catalogue of indexed objects. This entails creating an ID for each neuron in the dataset as well as links to all synapses (and their pre- and postsynaptic partners and other properties) and intracellular organelles (e.g., synaptic vesicle clouds, mitochondria). In short, everything that can be thought of that is contained in such a dataset should be catalogued and indexed. Once this catalogue is created, it becomes easy to ask these data neuroscience questions by running queries against the catalogue and index. At present, there are no common standards for querying EM datasets, and therefore many different solutions have been developed and are used across labs and consortia. Currently, manual EM queries (i.e., synapses and wires are manually annotated) remains the state of the art, although automated approaches to reconstructions with proofreading are on the rise.

Dr. Kornfeld and colleagues developed a connectome analysis toolkit called “SyConn v2” to query songbird basal ganglia connectomes. In collaboration with the Google connectomics team, they used FFNs to turn a large basal ganglia connectome into a complete connectivity matrix (8,000 neurons with the soma, axons, and dendrites, 200 interneurons, and 4.5 million synapses) with all objects indexed and queryable. They created a simple data structure for the SyConn v2 platform by using a Python codebase with optimized code for bottlenecks, placing Python objects in a scalable filesystem, and used parallel processing on the Google cloud or in a traditional high-performance computing environment. The resulting SyConn v2 toolkit enables straightforward querying of specific neuronal features. For example, a query for “soma synapses” filters all synapses to retrieve those on neuronal somata and allows synapse visualization and properties (e.g., size) to be obtained.

A critical step in the construction of a queryable catalogue pipeline is the verification of query results. During platform development, sanity checks should be performed on subsamples of the raw data and fixes made to the pipeline as problems are discovered. It is also crucial to make measurement uncertainty explicit, especially for corner cases such as small synapses. However, a common problem in connectomics is the lack of sufficient ground truth that is available, especially for corner cases. One way to mitigate this issue to some degree is to compare unsupervised results with subjective expert classifications.

Regarding the scalability of a queryable connectome to a whole mouse brain, one consideration is that a platform such an SyConn v2 currently doubles the amount of data relative to the raw dataset, which, for a whole mouse brain, would add approximately 200 petabytes of data (for 8 bit raw EM image data at 10x10x25 nm resolution). Although this amount of added data will not be an issue, the potential computational complexity of some queries and algorithms might become problematic. Directly accessing a few cells will be fast, but a single pass query to find all synapses with specific properties could take much more time and will present an even greater challenge for queries with higher order dependencies.

Several key takeaways from this work are relevant to developing a queryable mouse whole-brain connectome. First, standard connectome data formats must be established by the field. In the same way, quality standards must be established and the minimum resolution and annotation required determined. Finally, running complex graph queries against a mouse whole-brain connectome might become challenging from an algorithmic point of view.

Crowdsourcing Connectomics at Scale

Amy Sterling, BS

Ms. Sterling, Director of the citizen neuroscience crowdsourcing project Eyewire, discussed the prospects for crowdsourcing connectomics at scale. In the field of citizen science, members of the general public who largely have no formal background in science contribute in many meaningful ways to research. Thousands of citizen science projects have gathered tens of millions of citizen scientists and produced thousands of publications.

Eyewire was one of the first citizen neuroscience projects. The project invited the public to help reconstruct retinal neurons from a 300x350x60 µm dataset generated by the Max Planck Institute. Eyewire game play is quantized into cubes, in which players fit together super-voxels that have been agglomerated by artificial intelligence. In brief, Eyewire game play entails solving 3D puzzles to, for example, extend dendrites from one side of the cube to the other. The game then compares the reconstructions of multiple accurate players who have reconstructed the same volume in order to generate a consensus and entire neurons are reconstructed from individual cubes. Thus far, tens of millions of cubes have been submitted by hundreds of thousands of Eyewire players representing 140 countries.

Scaling Eyewire to a 1 mm^3 piece of cortex in a new project called “Peer,” which is the crowdsourcing component of the IARPA MICrONS program, will require the implementation of several new elements. One of the greatest challenges will be that, as automation improves, it becomes more difficult for humans to outperform AI; therefore, future crowdsourcing endeavors will become games for experts with a higher benchmark for participation. Eyewire will solve this issue by making the onboarding and training experience an exciting and engaging part of the game and by providing players with real-time personalized feedback that helps them learn in a fun and efficient way.

Another strategy for scaling is to take advantage of automated reconstruction improvements to transition from the cube approach toward the analysis of individual cells. For example, the

FlyWire project invites researchers and members of the general public to proofread neurons to map the *Drosophila* connectome using a Wikipedia-style platform. In order for the proofreading systems to work for the FlyWire platform, the speed of the system itself must be increased, to enable faster merges and splits and to improve image and 3D mesh loading. In addition, future citizen science projects should provide tools that enable collaboration among users and team-based approaches to tackle individual cells or groups of cells. Another challenge for scaling is the accuracy of the game platform, which is required to achieve professional-level proofreading.

New modes of interaction, such as shifting from the cube to the individual cell approach, require creative thinking. In addition, improved automated systems could improve game play, including ways to level players up and to implement systems for reward. Citizen science games should focus on maintaining their user base over longer periods of time by incorporating player feedback. The success of any citizen science project rests on the sense of community among the players and researchers that compels them to continue to contribute for long periods of time.

Session 4: HPC, Data Infrastructure, Visualization

Research Data Management at Scale

Rachana Anathakrishnan, MS

Ms. Anathakrishnan discussed data management in the context of a large initiative such as whole-brain connectome and some of the approaches and solutions that can be applied to manage data at such a scale. As one fundamental data management requirement, researchers must be able to reliably and securely move or replicate data across a variety of systems (e.g., cloud, on-premises, personal laptops, instruments, national resources, research computing HPC). The heterogeneous nature of the storage and data movement within and across institutions and sites requires high-performance data transfer. A second requirement is that in a collaborative enterprise, researchers need tools to enable data access for authorized collaborators without unnecessary data replication.

Scaling data management across operations, resources, people, and all dimensions needed for a large initiative requires intrinsic and efficient automation. Scaling also requires platform capabilities that facilitate development of a wide variety of applications (e.g., analysis, visualization) on top of platform capabilities and integration with tools such as Jupyter notebooks. Finally, to scale nationally and internationally, the research ecosystem should leverage institutional or federated identities that are vetted and have identified life cycle management.

Traditional network architecture, in which the data and the logic flow through the enterprise firewall, significantly reduces the rate of data flow and is not suitable for data-intensive applications. Optimized network architecture uses the well-recognized Science DMZ model in which there are two paths for data flow: one path uses the enterprise firewall, meets security requirement, and is used to browse and query data, and a second transfer data path is optimized for data movement and provides high-performance data transfer.

Several services and tools leverage such networks and support data management. The University of Chicago's Globus transfer service has a managed, reliable, high-performance transfer that comprehensively manages all aspects of data transfer. The service also provides data sharing from any storage, which eliminates the need to replicate the data or to create local accounts for collaborators, which is a real practical concern in terms of cost, time, and security. The platforms needed to enable a rich set of custom domain applications are also provided within Globus. The Globus Automate platform also provides hosted services to orchestrate task automation by enabling researchers to create a workflow that can span heterogeneous resources.

One large-scale data management service that is relevant to brain connectomics is the Petrel Data Service, which is a collaboration between the Argonne leadership computing facility and the Globus team. Using the Globus sharing feature, Argonne researchers are provided a self-managed data sharing service that uses optimized network architecture and allows management of project data permissions. As a relevant example of the use of these services, the Kasthuri lab used the APS at Argonne to image mouse brains and generated segmented datasets to share with the community. This work utilized both the Petrel service and Global Automate to manage an involved set of tasks that required task automation and manual user intervention, and crossed a number of security boundaries.

Connectomics at ALCF

Tom Uram, BS

Mr. Uram discussed connectomics efforts at Argonne Leadership Computing Facility (ALCF). The Data Science Program at ALCF emphasizes the use of Argonne supercomputers for large-scale machine learning and deep learning. ALCF has a tremendous history of infrastructural support for supercomputing and currently supports several systems with petaFLOP processing speeds, one of which has enough global storage space to hold 20 percent of the data for a whole mouse brain. These supercomputing capabilities will culminate in the new Aurora exaFLOP system that is currently under way.

When Argonne receives a new supercomputer, a collection of Early Science Projects are awarded to teams that can prepare their codes for the machine. For the Aurora exascale system, the Early Science Project dedicated to connectomics aims to establish a pipeline for running the code for connectome reconstruction using large-scale systems and to adapt related individual applications for next-generation supercomputing systems through a multi-institution collaboration. In addition to the reconstruction component, the project will also address online processing to optimize throughput from image acquisition to reconstruction to maximize concurrency as well as enable remote access to data for visualization, analysis, and download.

Because the time required for imaging a whole mouse brain and the downstream computing will be substantial, imaging and computing should overlap as much as possible. To this end, the online processing from the SEM to ALCF has been simulated to begin the first stages of the reconstruction pipeline, which is preparatory work for online processing. To prepare for future imaging systems, ALCF can also simulate multiple SEMs, offsite EMs, and X-ray imaging from

Argonne and other facilities. Analyses of data transfer performance between HPC facilities also indicate that both the ALCF network and the machine can transfer large-scale data effectively and simultaneously, and return the data to the originating institution within 48 hours. Current efforts on the large-scale reconstruction component of the project suggest that adapting the current codes for processing connectomics data (e.g., stitching, alignment, scaling FFN training and inference) to meet desired level of concurrency might be a challenge, and current efforts are focused on improving the codes.

The mouse whole-brain connectome effort faces significant challenges in terms of data, imaging, and supercomputing. Projecting an existing connectomics dataset to future reconstruction time, on the current petabyte system, would require tens of years of reconstruction time. The future Aurora exascale system will still require years of reconstruction time. Moreover, connectomics will not have exclusive access to the ALCF supercomputers, so continued efforts to improve and accelerate all stages of the reconstruction pipeline are critical.

Scaling Pipelines for EM Connectomics

Eric Perlman, PhD

Dr. Perlman discussed approaches to scaling pipelines for EM connectomics based on lessons learned during reconstruction of the FAFB in the lab of Dr. Bock at Janelia. The FAFB dataset consists of 21 million micrographs totaling 106 terabytes of raw data and 50 teravoxels when fully stitched. The team developed solutions for dealing with this massive dataset when existing tools did not scale or tools did not yet exist. The FAFB project overcome three technical challenges that are particularly relevant to the mouse brain connectome: (1) data storage and networking, (2) infrastructure for early and ongoing data analysis, and (3) pipeline software for data visualization.

The first challenge is to ensure the availability of ample bandwidth between microscopes and storage and storage and compute. Although the FAFB project was able to circumvent saturation of network and storage systems through a combination of local data storage and processing capabilities, this approach would not be feasible for a whole mouse brain. By uploading to the cloud, data can live in an environment with approximately infinite aggregated bandwidth for compute, which is a key for alignment, segmentation, or analysis. Potential network and infrastructure failures make direct upload from microscopes to the cloud unattractive, although this issue could be overcome through use of local storage as buffer.

The second challenge is to ensure that architecture to do science is in place by building the infrastructure to start data analysis in parallel with ongoing data acquisition. During the FAFB project, an analysis performed early in the acquisition process showed viability of the data by beginning manual annotations with only 7 percent of the data imaged. Maintaining this manual effort despite stack realignments with the addition of new data was made possible through the use of Render, a service for interacting with stacks of aligned images. Similarly, FlyWire provided a deformation field that is now used to compare skeletons and meshes across alignments, which has enabled smooth collaborations across the fly connectomics community.

The third challenge is to ensure that visualization is available from project onset, not just for analysis or publication. Learnings from the FAFB project highlight the importance of an architecture for how to scale over time to move the science forward smoothly. One of the most useful software tools was a quick tool for viewing data after acquisition that enabled identification of a variety of problems (e.g., debris, focus, contrast, and mechanical stage errors). Another valuable strategy was generation of downsampled versions of the data that were collected and still residing in memory, because such data are much faster, easier, and less expensive to access and are versatile for many early processing steps (e.g., data inspection and coarse alignment).

Dr. Perlman suggested that any effort on a full mouse brain should consider these key take-aways: (1) Consider data storage and networking before the data flows. Local bandwidth and storage are not sufficient for scaling, and cloud solutions should be incorporated. (2) Extract early value by supporting analysis on data. Confirm that data quality will be sufficient for scientific discoveries and that early analysis will be preserved. (3) Treat visualization as essential, and generate and provide access to downsampled data in parallel with acquisition.

Q & A Session

Compression of 3D EM Data

Dr. Januszewski discussed the key objectives for compression of 3D EM image data. He asserted that fewer bytes of data at a constant quality of the downstream reconstruction is always better and suggested that targeting less than 100 petabytes of data would be reasonable for the mouse. For determining needed compression ratios, it is useful to consider the ratio based on the final dataset size that will enable effective work with the data, the size of the infrastructure that will be used, and the size of the raw data before processing. Dr. Januszewski also emphasized the importance of addressing the effects of compression on the downstream segmentation and pipeline as well as the other deep learning models that are used on those datasets. Moreover, because the correlation of image quality metrics with the downstream metrics of interest to the connectomics community remains unclear, he recommended that the consequences of data compression on the downstream pipeline should first be measured on small samples, before proceeding to large datasets.

Learning from Interactive Tools

The speakers were asked to discuss potential approaches to gaining insight about data processing errors and potential corrections from their interactive tools. Dr. Saalfeld remarked that the tools he uses work directly on the transformations and maintain the original data; it is important that image data not be modified. He noted that, when performing several iterative alignments, the annotation can always be mapped back to the original image space. The modified transformations will inform understanding of what has been manually edited and will be useful in training deep neural networks to identify places of interest.

Dr. Rivlin responded that as the proofreading process corrects segmentation errors, those corrections can be utilized to train better automated segmentation. The study of different

versions of the proofread segmentation also informs understanding of areas within a dataset that are particularly problematic, as well as preparation of future volumes. Ms. Sterling also noted that previous analyses can aid in training new data users. For example, FlyWire is developing a training system that includes thousands of proofread cells, which will be used to train new users until they develop a sufficient level of accuracy to contribute to the database. Dr. Turner described an active line of research into using various kinds of proofreading edits to determine machine learning classifiers that enable comprehensive error detection in morphological segmentations.

Multi-Modal Data

Several speakers discussed potential applications of multiple modalities in the analysis of EM data. Dr. Mahalingam recalled comments from her presentation about co-registration of complementary imaging modalities such as microCT and EM, and noted that, with feature-based approaches, the identification of features (e.g., nuclei or blood vessels) shared between the datasets can be challenging. Identifying such features in the EM might entail, for example, segmenting the aligned volume and then relating that back to the microCT to find those point correspondences.

Dr. Buhmann and colleagues have used multi-modal approaches to obtain annotations for labeling synapses by neurotransmitter type using information obtained from LM. In this approach, LM determinations of neurotransmitter expression in individual neurons were correlated between light level and EM images. The neurons identified in EM data were then used to create a training dataset that enabled attainment of neurotransmitter types from EM images alone.

Dr. Kornfeld further suggested that such multi-modal approaches could help address the lack of available ground truth. For example, the connectome could be calibrated with different modalities (e.g., LM or electrophysiology) to determine synapse sizes, and those data could inform analysis of datasets in which ground truth has not been established. Mr. Macrina and Mr. Turner concurred that ground truth one step down in resolution (e.g., microCT) would be incredibly useful for alignment and cell typing as well. The problem with cell type clustering persists because consensus on the precise level of splitting of cell types needed has not been reached. Multi-modal approaches are a valuable resource that should be leveraged to address these types of issues.

Multi-Scale Data

Dr. Januszewski addressed the utility of having data available at multiple scales. In production workflows, he and his colleagues tend to use three resolutions for segmentation of dense neuropil. They find that a resolution higher than 8-10 nm per voxel provides no additional benefit for segmentation. While down-sampling to 30-40 nm per voxel is useful, neurites become difficult to resolve with further down-sampling. He added that segmentation can be done at lower resolution for some regions of the brain (e.g., white matter). Dr. Buhmann noted that many of the models in use, such as the U-net, explicitly operate at three different scales.

She believes that it is also important to operate at different resolutions for neuron segmentation.

Dr. Ferrier remarked that the scale of multi-modal approaches can differ by many orders of magnitude and asked the speakers to comment about tools and techniques that can handle this large difference. Dr. Saalfeld responded that, from the alignment perspective, X-ray has not proven useful for providing ground truth or for registering section series. For example, a FAFB brain might appear registered at the level of an X-ray (1 µm per pixel), but it could be very poorly aligned at the EM level. He also expressed skepticism that this type of multi-modality would be very helpful for reconstructing a connectome. Mr. Macrina explained that, in his experience, the 1 µm resolution of an X-ray is sufficient for an initial assessment of rough alignment before proceeding with a nonglobal alignment pipeline. Dr. Saalfeld agreed that X-ray can be useful as a global reference, but added that a 1 µm pixel is not sufficient to provide any structure. For example, the diameter of a single neurite is not visible in an X-ray image, and the precision of the registration could be off by several microns.

Pipeline Feedback

A question was raised about the possibility of using feedback from segmentation algorithms to correct errors in alignment and prior steps. Dr. Saalfeld remarked that, although an interesting concept, optimizing for all neuronal shapes simultaneously is problematic. In his view, the mouse brain connectome should move away from complicated global problems toward minimum feasible solutions. Dr. Mahalingam remarked that feedback from segmentation could serve as a useful tool to correct the alignment. She added that the costs and computational resources required to provide feedback from segmentation to alignment should be considered and that better, simpler solutions to handle local or even global misalignment likely exist. Mr. Macrina also believes that segmentation is a good indicator of misalignments, and Ms. Sterling noted that her team is considering implementing this type of feedback in FlyWire.

Dr. Januszewski offered a segmentation perspective on the feedback issue. In his work on segmentation, he pre-detects any alignment problems before running the segmentation algorithm. He also expressed skepticism about the feedback cycle and suggested that more a classical method could likely improve alignment. Usually, those problems can be detected with relatively simple measures and, once detected, can be addressed before proceeding downstream in the pipeline.

Ground Truth and Machine Learning

Several speakers discussed the importance of ground truth. Dr. Januszewski made several points, suggesting that quality and relevance are more important than sheer quantity. The data should be checked after annotation is complete to ensure that there are no obvious errors, which would be propagated by the algorithm. To improve the efficiency and scale of the annotation process, researchers should (1) train models iteratively to ensure that the ground truth generated is relevant to errors that the models tend to make and (2) move away from de novo painting toward existing segmentation models to create a candidate segmentation that can then be strategically corrected through de novo painting or supervoxel manipulation. Dr.

Januszewski also addressed the state of "transfer learning" to enable bypassing of ground truth generation altogether. In his experience, existing models can be applied to new datasets without additional ground truth and with minimal loss in accuracy, provided the acquisition settings are similar.

With regard to obtaining ground truth data, Ms. Sterling stressed the importance of the user interface and user experience. For example, adding subtle game mechanics to the reconstruction process was highly motivating to EyeWire tracers. Dr. Rivina stated that finding faster methods to generate ground truth would be very helpful. She agreed that the proofreading experience is very important to generating high-quality ground truth. Although the Allen Brain Institute proofreading team routinely creates ground truth from initial segmentations of small volumes (i.e., cubes), its goal is to obtain as much densely proofread volumes of segmented data as possible. However, this labor-intensive process can only be done by the best proofreaders. Proofreader experience, a diversity of datasets, and well-designed proofreading tools facilitate the generation of high-quality ground truth data. Dr. Kornfeld agreed that obtaining high-quality ground truth is difficult and requires experienced annotators. The neuronal segmentation problem is generally not ambiguous as long as these data are of sufficient quality, although other problems can be more difficult to solve. For example, dendritic spine morphologies do not always fit into straightforward categories, and those corner cases require more careful consideration for annotation and training of models.

Cell Type Classification

Dr. Kornfeld suggested that the inclusion of connectivity in cell type analysis can be problematic. This approach might be useful if the goal is to simply determine the cell types in the connectome. However, if the goal is to answer neuroscientific questions about the connectivity of a particular cell type and the connectivity is used to determine the cell type classification, then the approach is fully circular. Instead of combining connectivity with morphology in EM datasets, another modality (e.g., LM) is needed to provide ground truth. Dr. Turner concurred, noting the importance of finding a signal that crosses modalities in a meaningful way.

Dr. Turner asked about the types of information that the field should try to derive from a first pass at cell type clustering in the connectome. He observed that much insight can be obtained from the morphology work done to date (e.g., classification of retinal cell types). Although new cell type information will be derived from the cell type classification process, it will not be sufficient to address every cell type question in neuroscience.

Changes in EM Dataset Size During Its Lifetime

Several speakers from Session 4 discussed the changes in size that EM datasets undergo during the various stages of processing and analysis. Dr. Perlman observed that the dataset shrinks as raw data are processed to aligned data and shrinks further during segmentations. Dataset growth is a function of how the segmentations are stored and represented (e.g., meshes, graph sizes). Dr. Perlman further noted, however, that storage space itself will not likely be a concern

for data beyond raw and aligned data, although the size of the analysis graph will be extremely large.

Dr. Kornfeld stated that getting the data into a queryable state doubles the data size compared to the raw data, although the required storage space could be reduced by about 10-fold with careful optimization. If the goal is to store pre-computed meshes and all of the synapses, which is desirable for database querying, the data must be stored in an indexed way. Regarding the scalability from an algorithmic point of view, if the connectome graph is large, complex queries will be expensive to perform (compared to selecting a single cell), which could be a fundamental problem for some analysis. Dr. Perlman added that access to data in the more processed states must be faster and lower latency, particularly with graph algorithms.

Cloud Versus National Laboratory Supercomputers

The speakers discussed the pros and cons of data storage and processing using commercially available cloud services versus national laboratory supercomputing centers, with a focus on alignment and segmentation of a mouse whole-brain EM dataset. Mr. Uram commented that a major advantage of the DOE supercomputers is the significantly lower cost than cloud-based services. DOE supercomputers also offer high-speed internet connections that can be leveraged for fast data queries and analysis. Dr. Perlman noted that, independent of cost, both DOE supercomputers and cloud-based systems would work for the mouse whole-brain project because both environments are designed to store large amounts of data and to enable fast data access in parallel for processing. Ms. Anathakrishnan added that cloud-based storage would be useful for certain parts of the data lifecycle and for democratizing access to the data.

Discussion Panel

Panel Discussion Topic: Reflections on how experience as a computer scientist might inform an exascale project aimed at generating comprehensive maps of brain connectivity

Panel Chairs: *Nicola Ferrier, PhD. and Viren Jain, PhD*
Discussants: *Frances Chance, PhD, Adrienne Fairhall, PhD, Ian Foster, PhD, Eugene Myers, PhD, Olga Ovchinnikova, PhD, and Andreas Tolias, PhD*

During this portion of the agenda, discussants presented their views on the topic, commented on the presentations by the earlier speakers and other discussants, and considered the questions posed in the online Q&A.

Dr. Chance commented on the remarkable breadth of potential impacts that a comprehensive connectome could have both within and beyond the field of neuroscience, noting that she had attended two recent workshops in distinct scientific fields in which papers had been published describing neuroscience as an inspiration for advances in next-generation computer architecture or novel neuro-inspired machine learning algorithms. Although a mouse whole-brain connectome appears to be within reach, it will be challenging not only to design a robust data analysis pipeline, but also to ensure that the tools and representations of data within that

pipeline are accessible to a wide range of researchers (i.e., not only neuroscientists) to maximize the impact and generalizability of the data.

Dr. Fairhall challenged the panel to consider what approaches to modeling the brain will be needed to make the best use of connectomic data and whether those approaches will necessarily amplify current computational needs. Among the considerations for those approaches is the appropriate level of modeling, because a wealth of information about cellular identities, morphologies, and neural dynamics can have implications for models of network function and activity even when constrained by the connectome. Dr. Fairhall remarked that the IARPA MICrONS Project could serve as a valuable testbed for the extent to which connectomic knowledge can constrain models well enough to predict activity in a given tissue.

Dr. Jain remarked that much of the current analysis of connectomic datasets is focused on answering specific queries about forms of connectivity rather than simulation of activity, and asked whether the infrastructure for a whole-brain connectome project should include the capacity to pursue forward-modeling approaches. Dr. Fairhall responded that, if the field is interested in these sorts of analyses, then it should consider these infrastructure needs in advance, noting that the Human Brain Project was focused on building infrastructure for analyses of this kind. To support this kind of modeling, the field should determine the appropriate level of data simplification at the single neuron level to obtain sufficient functional insights. Dr. Tolias added that the ability to understand how a neuron computes from connectivity data also needs functional constraints, insofar as one needs to understand the function of the inputs to understand the activity of the target neuron.

Dr. Foster emphasized that it is easy to underestimate the tremendous scale and scope of the challenge to enable a community to work effectively with petabytes and exabytes of data. For example, Dr. Foster explained that to read a petabyte at 100 MBps would take approximately 4 months, and to read an exabyte-scale mouse brain connectome at that speed would require hundreds of years. These challenges become larger when accounting for the vast analytics and processing that will be required to build the connectome. Dr. Foster therefore suggested that a new cyberinfrastructure, with new tools and methods that leverage parallelism and automation at every stage, will be critical to organize, manage, compute on, analyze, compare, and search exabytes of data. This co-design challenge should be addressed as soon as possible rather than when data become available.

Dr. Myers remarked that the achievement of comprehensive connectomic maps is not a question of *if* but rather a question of *when* and *how*. To reconstruct a whole mouse brain at EM resolution, Dr. Myers suggested that the field should sponsor multiple independent efforts but emphasized the need for coordination across those efforts. This coordinated effort could define the critical sub-tasks that must be performed to complete the whole-brain connectome, paying attention to the specific input-output relationships entailed for each task. These distinct sub-tasks could be pursued by independent groups toward a common end goal. The tools required to perform these tasks should produce comparable outputs in order to promote comparison across tools, aid competition, and allow for parallel efforts. The generation of

benchmark datasets, including intermediates and simulated data, would support the generation of the whole connectome.

Dr. Ovchinnikova commented on the infrastructure that is needed to connect neuroscience and computer science in pursuit of the whole-brain connectome. At the level of instrumentation, data standardization must be democratized to foster compatibility of data formats. Furthermore, the instruments used to build the connectome can incorporate different modalities that enable linkages between chemical identities and morphology in service of better segmentation and ground truth identification. The data pipeline should leverage edge computing to allow for data compression, extraction of connectomic data, and adaptive filtering for selective data retention and active learning. Segmentation algorithms that were optimized by HPC should be implemented closer to the point of data acquisition (i.e., at the instruments) to mitigate issues associated with storage of these very large datasets. Only extracted data would then progress to HPC networks to enable parallel processing. This kind of pipeline would promote an integrated, interconnected ecosystem of different tools at different locations to produce globally accessible data.

With regard to reaching agreement on data standardization, Dr. Myers suggested the formation of two working groups composed of major current stakeholders: one to focus on alignment and volume creation and one to focus on extraction of digital models. Dr. Ovchinnikova noted that Oak Ridge National Laboratory has developed an ecosystem infrastructure that enables streaming of data from microscopes directly to a user facility that standardizes combined data into a format suitable for HPC-type processing.

Dr. Tolias advocated for the collection of functional data to complement the connectome. In the mammalian brain, neurons are organized by canonical rules and principles that repeat within and across brain areas. Thus, Dr. Tolias believes that the function of a neuron cannot be inferred from connectivity of morphologically defined cell type data alone. Moreover, the inclusion of functional data could increase the value of connectomic datasets for understanding computations from a neuroscience as well as a computer science perspective. The MICrONS project includes both connectivity and functional imaging data, which enables inference of neuronal function based on the activity of a neuron as well as of the cells to which it is connected. To collect functional data for the connectome, the infrastructure must include a capacity to perform functional recordings that can be combined with connectivity data at a large scale.

Dr. Tolias cautioned against deciding what the nature of the first whole-brain connectome (e.g., EM volume with or without manipulation) should be at the outset and instead advocated for parallel efforts to build a connectome as well as large-scale functional data. When technology becomes available to image the whole mouse brain, these parallel and complementary datasets can inform the decision of what the first whole-brain connectome should look like, and infrastructure to collect the right kinds of data would already be in place.

Follow-Up Discussion

Dr. Fairhall questioned whether connectivity data are more useful with or without biophysical data. Although interesting findings have been generated from reduced models that contain only connection matrices, Dr. Fairhall speculated that, similar to Dr. Tolias' views on functional data, biophysics will be an important aspect of recapitulating true brain activity. Dr. Chance added that a co-registered functional dataset could capture subthreshold computations that may underlie certain behaviors, although a high-resolution connectomic map alone could identify structures that suggested these computations as well.

Dr. Myers asked what information can be gleaned from an EM reconstruction that could not be gleaned from light-based reconstruction. Dr. Rivlin explained that comparisons of EM reconstructions (including the *Drosophila* hemibrain) to LM libraries reveal that some features will be missed by LM, including identification of some cellular subtypes. Dr. Jain shared that approximately 50 percent of the cells in the *Drosophila* hemibrain dataset were not previously found by LM, and that synaptic connectivity is almost completely missed at the light level. Nonetheless, an EM dataset could also miss some features, particularly if the reconstruction is in a less stereotypic brain region. Dr. Rivlin noted that pairing LM data with EM data to collect complementary datasets is a valuable effort. Dr. Chance added that the field is not yet confident of the true value of high-resolution datasets and cautioned that efforts to improve data compression must be approached with care to avoid unintentional loss of important information in the process.

Dr. Ngai asked the group to identify the greatest computational challenges to integrating data across scales and modalities. Dr. Ovchinnikova responded that one challenge is the need to standardize data streams and sources so that data can be cross-correlated and that a second challenge is the exponential growth in data volumes as more modalities are added. Dr. Ovchinnikova highlighted the value of developing an ecosystem for cross-correlation of data for all instrumentation across modalities, which would support accurate modeling of how these data can be integrated across scales. Dr. Tolias added that reconstructing neural contacts across experiments performed at different times represents additional computational challenges, because structural changes may have occurred between experiments.

Dr. Ellisman suggested a use case for the mouse whole-brain connectome: provide a template onto which other mouse data can be fitted. Dr. Jain noted that mouse brains are less stereotyped than other model organisms (e.g., fly, worm), but that multiple connectomes could be used to derive a quantitative estimation of variability across mouse brains at the partial or whole brain level. Dr. Fairhall remarked that this template would rely on the presence of multiple datasets, and that this goal would support Dr. Tolias' view that parallel efforts on partial brains should be prioritized to better understand the parameters needed to maximize the impact of the mouse whole-brain connectome. Artificial neural networks could provide some insight into the methods needed to assess variability, although translation of these insights to a real brain context may be limited. Dr. Myers commented that sampling only 1 percent of the neurons in the mouse brain at the light level could provide insight into variability in the flow of information across individual brains, provided these neurons can be reproducibly

accessed. Dr. Rivlin liked the idea of comparing the variation between left and right hemispheres to estimate variability across individuals, as well as to assess reconstruction quality, but noted that developmental anomalies could lead to differences between the hemispheres.

Workshop 5: Optimizing the Use of Connectomic Data to Drive Data Science and Scientific Discovery

Precis

The goal of this workshop was to define the current status and limits of analysis, interpretation, and dissemination of brain connectomic data. The workshop focused on recent analysis methods for partial, full, and multiple connectomes emerging from a number of model systems, in addition to the problem of making connectomic data accessible to the community. In particular, the following questions were addressed:

1. What scientific breakthroughs have analyses of connectomes already enabled?
2. What current methods for analyses of connectomes are most promising?
3. What new types of analyses (not developed yet) are needed to make progress and enable scientific breakthroughs?
4. How do we integrate analyses across scales (from nano to micro to meso)?
5. How do we integrate connectomic data with functional data?
6. What types of data and analyses are needed to make interesting comparisons between connectomes?
7. What kinds of infrastructure are needed to make connectomic data and analyses widely available to the neuroscience community?

To address these questions, the workshop covered a series of talks and discussions in three complementary sessions: (1) Analysis of Partial and Full Connectomes, (2) Analysis of Multiple Connectomes, and (3) Making Connectomic Data Accessible to the Community, and finished with a unifying discussion.

In the analysis of partial and full connectomes, talks related to discoveries from analyzing connectomes in *C. elegans*, *Ciona intestinalis, Drosophila*, zebrafish, and at the mesoscale, in mice. The community is making progress in deciphering these partial and full connectomes, and is beginning to assemble tools to facilitate their analysis, including leveraging connectomes to build models of brain function. Because there are multiple different *Drosophila* connectomes, different researchers are beginning to compare across them. Others are using these datasets to build connectome-constrained models, or explain connectivity as a function of cell type. However, some words of caution were expressed as well. First, the assumptions we make in classical statistics do not typically apply to connectomic data, and therefore inferential tools from classical statistics are typically not statistically valid, rendering p-values uninterpretable. Second, even if one uses proper statistical network theory, our sample sizes are approximately 1, meaning we are incapable of assessing heterogeneity, or making any valid inferences about any future connectomes we will collect. Third, existing work in graph theory tells us that subgraphs and the complete graphs from which they were sampled may not, in general, have the same properties at all. For example, if ipsilateral connections in the larva and adult *Drosophila* are similar, it does not mean that contralateral connections will also be similar. Therefore, any analyses of partial connectomes must be interpreted cautiously. Fourth, existing

models fail to adequately account for spatial location, transcriptomics, and other properties of the neurons that could potentially explain connectivity to a large degree.

In the session on analysis of multiple connectomes, presenters spoke about both invertebrate and vertebrate connectomes, including mouse, NHP, and human connectomes, because those are the only species for which we currently have multiple estimated connectomes from different individuals. Of note, the vertebrate connectomes discussed define nodes to be regions of the brain typically, in contrast to the invertebrate connectomes, which defined nodes to be individual cells. Tools for analysis of multiple populations of connectomes are reasonably well established in the network machine learning community, although they have not yet percolated into practice in most labs. Moreover, tools for time-varying connectomes are really just beginning. Nonetheless, the idea of longitudinal studies of connectomes across development came up. While existing time-series of single-cell connectomes are limited to cross-sectional *C. elegans data*, in vivo imaging techniques have enabled longitudinal data collection in humans over months or years. There are major open opportunities for developing theory, methods, and user-friendly tools to analyze populations of connectomes with various different assumptions on the relationship between the connectomes, including nodes are “matched” across the connectomes (as in longitudinal data) and not matched (as in cross-species data).

Third, workshop presenters spoke about making connectomic data accessible to the community. Several groups have collaboratively developed open source tools that enable sharing these data at different levels and scales, including sharing the stitched and aligned 3D volumes from EM and/or LM data. But the image data are inadequate for quantifying neurobiological insights. To do that, the segmentations—be they generated manually, by machines, or a hybrid approach—are required. Better yet, derivatives from those morphological segmentations, to include actual brain networks, the spatial distributions of various properties, and more, are the objects that neuroscientists can use to make inferences.

Finally, the workshop concluded with insightful discussion with panelists. Panelists revealed that other big data scientific disciplines, such as cosmology, have essentially transitioned entirely to the cloud, and so the field of neuroscience should expect a similar transition soon. Panelists also discussed the value of various network machine learning tools, including both spectral methods and so-called “graph neural network” methods. Further, panelists engaged in an in-depth conversation about causality—connectomes reveal densely recurrent networks and highlight the difficulties in interpreting activation and silencing experiments of small subsets of neurons. Instead, brains should be thought of and analyzed as control systems—with connectome data from a number of species and individuals of a species, we can start making connections between network motifs and mental properties (including normal and impaired cognitive function).

Introduction

Mala Murthy, PhD, Joshua T. Vogelstein, PhD

There is a growing number of large EM datasets driven by innovations in the EM data collection methods covered in prior workshops. These datasets present new challenges for comprehensively, efficiently, and accurately reconstructing neurons, glia, and organelles; building and interpreting connectomes; and disseminating these data to drive new science. The goal of this workshop was to identify the current status and limits of analysis, interpretation, and dissemination of brain connectomic data, as they relate to the goal of generating whole-brain connectivity diagrams in mammalian species.

Scaling up from the brain of a fly to the brain of a mouse will require building on the recent advances in the ability to store and analyze large amounts of data. Currently, storing and analyzing petabytes of data is feasible, but the whole mouse brain will be on the order of 1,000 petabytes. Another advance in the field is the development of new machine vision and artificial intelligence technology for cleaning, aligning, and segmenting large EM volumes. For example, for the neural network–based realignment of the *Drosophila* FAFB dataset, the network was trained to correct for cracks and folds, which resulted in a volume that could be automatically segmented into large pieces of neuron. The accuracy of the automated segmentation means that less time is needed for subsequent manual proofreading. These types of approaches hold tremendous promise for scaling up to larger datasets. A third exciting advance is the development of platforms for visualizing and analyzing connectomics data, which are critical to progress on extracting information from and making comparisons between datasets. Finally, the dissemination of EM connectomic data has seen important advances, such as the Open Connectome Project, which provides a central resource for sharing and working with large connectivity datasets, and crowdsourcing platforms such as EyeWire and FlyWire that enable communities of scientists or citizen scientists to reconstruct and annotate neurons from EM data directly.

Despite these tremendous advances, several challenges lie ahead for connectomics. One challenge is that the data being obtained are much more complex than simple binary inputs and now include multi-connectomes with complex vertex and edge attributes. These connectomic data will provide a powerful resource for addressing many scientific questions once the field has developed approaches to properly analyze these highly complex datasets. During the past 10 years, a number of tools have been developed, including open source free libraries, each of which provides a wealth of tools for the analyses of connectome networks. However, many methods lack statistical rigor, theories for dealing with network-, vortex-, and edge-valued attributes are limited, and existing network science toolboxes do not facilitate analyses on these attributed connectomes.

Thus far the connectomics field has primarily focused efforts and plans to obtain one connectome per species. However, the difficulties of comparing within and across species raises the fundamental challenge of balancing competing scientific agendas in connectomics. Moreover, current connectomes highlight the missing information such as synaptic weights,

gene expression profiles, and behavior. Connectomes represent a single snapshot of brain at a given moment in time, and current technologies cannot answer how the brain changes over time. It is not yet clear what level of detail will be required to address many fundamental questions about how the brain works based on connectomic data.

Workshop 5 was divided into three sessions that address the following topics: analysis of partial and full connectomes, analysis of multiple connectomes, and making connectomic data accessible to the community. Speakers addressed seven questions:

1. What scientific breakthroughs have analyses of connectomes already enabled?
2. What current methods for analyses of connectomes are most promising?
3. What new types of analyses (not developed yet) are needed to make progress and enable scientific breakthroughs?
4. How do we integrate analyses across scales (from nano to micro to meso)?
5. How do we integrate connectomic data with functional data?
6. What types of data and analyses are needed to make interesting comparisons between connectomes?
7. What kinds of infrastructure are needed to make connectomic data and analyses widely available to the neuroscience community?

The Discussion Session focused on analysis of connectomes. The presentations can be viewed by accessing the archived NIH VideoCast.

Session 1: Analysis of Partial and Full Connectomes

From the Fly Connectome to the Mouse Connectome

Greg Jefferis, PhD

Dr. Jefferis discussed strategies for using connectomic data to address biological questions and the tool space under development. Although the adult fly brain has about 100,000 neurons and 100 meters of neuronal cable, the mouse brain has about 100 million neurons and 500 kilometers of neuronal cable. The generation of intermediate stepping stone connectomes between these dramatically different scales would facilitate solutions to many analytical challenges for scaling to the mouse.

Two analytical strategies are crucial for extracting biological insight from the vast amount of data contained within connectomes: “comparative connectomics,” which entails making comparisons across multiple connectomes, and “integrative connectomics,” which entails integrating connectomics data with data obtained from other modalities (e.g., anatomy, physiology, molecular genetics, behavior, and functional imaging). Dr. Jefferis illustrated these two analytical approaches with his work on two *Drosophila* connectomics datasets: (1) FAFB, a sparse whole-brain dataset that is densely segmented and sparsely proofread and (2) the hemibrain, a dense dataset of one-third of a brain that is densely proofread. Additional *Drosophila* connectomes being generated by Dr. Jefferis and colleagues are a full male CNS, a

full male ventral nerve cord, and a full female CNS. Comparative analysis of the FAFB and hemibrain datasets, both within brain and across brain, has revealed that within brain variability is similar to across brain variability, which suggests that left-right variability is a useful estimate of the variability across individuals.

Dr. Jefferis and colleagues have developed a series of tools and resources that enable comparative and integrative analysis of fly connectivity data. Dr. Jefferis described how they employed tools and resources to identify and define the circuit position of a pair of visually responsive neurons involved in fly turning responses. The first step was to use a tool that bridges the registration of images across light, confocal, and EM imaging modalities and a tool called NBLAST that uses a confocal image stack as a template to morphologically query neurons in the EM connectome to identify all closely related neurons. Potential matches were manually compared to neurons in confocal stacks to identity the precise neuronal pair that had been functionally characterized; this step was crucial because the connectivity profiles for morphological similar neurons were often substantially different. Detailed morphological analysis and comparison of left- and right-side partner neurons confirmed that the dendritic fields were located in a brain area that responds to visual input. The next step was to define the synaptic inputs onto the identified neuron pair and determine whether the inputs were excitatory or inhibitory—using a tool developed by Dr. Jefferis and colleagues that enables prediction of neurotransmitters for fly brain synapses. The synaptic inputs onto the neuron pair of interest were determined to be inhibitory, suggesting that they suppress activity on the opposite side of the brain during fly turning responses. With this tool, synaptic neurotransmitters can now be predicted with greater than 90 percent accuracy for all 250 million synaptic connections in the fly brain and applied to any neuron identified from the segmentation. Using the strategy of bridging registrations plus NBLAST, the identified neuron pair was connected to homologous neurons in the densely proofread hemibrain to obtain the complete connectivity information for these neurons. Finally, web resources called neuPrint and neurobridge were leveraged to identify candidate driver lines, which enables linking of molecular and connectomic cell types, a key strategy in integrative connectomics.

Dr. Jefferis shared several key lessons learned from comparative and integrative connectomics in the fly that can inform strategies for scaling up these approaches to the mouse connectome. First, high-quality light level atlases are crucial for integrative connectomics. Second, the main benefits of EM connectomics volumes include not only maps of connectivity, but also a comprehensive catalogue of cell type information; this is one reason it is much more powerful to identify all cells from one brain than a fraction of the cells from multiple brains. Researchers have defined 5,000 neuron types in the fly connectome, which suggests that the mouse might have more than 50,000 types of neurons. Key challenges include integrating molecular and functional data with EM connectomes and developing strategies to design, rather than screen, genetic reagents for cells in the mouse connectome.

How to Simulate a Connectome to Gain a Mechanistic Understanding of Neural Computation

Srini Turaga, PhD

Dr. Turaga and his lab aim to build detailed connectome-constrained simulations of neural circuits of *C. elegans* and *Drosophila* to gain mechanistic understanding of neural computations made by those circuits. While the EM connectome provides tremendous detail about precise cellular morphology and synapse properties (e.g., synapse numbers, shapes, and sizes), other types of information required for developing detailed computational models are not contained within connectomic data. These circuit constraints include biophysical properties of neurons and synapses, a range of physiological measurements (e.g., calcium imaging, optogenetic perturbations, electrical recordings), high-level circuit tasks (e.g., the visual system performs vision), and behavior (e.g., fine motor behavior, decision-making).

Dr. Turaga and colleagues used connectome-constrained simulations to model the first few stages of the *Drosophila* visual system, for which connectivity information was available for a subset of neurons. They constrained the model with biophysical properties and the vision task. A model was built that included 66 cell types based on the known connectivity and anatomical details from the connectome. A simplified deterministic network model was then developed that incorporated threshold linear dynamics, voltage dynamics, and instantaneous synapses. The circuit was trained to determine unknown connectivity parameters by performing a computer vision task, with ground truth given for the optic flow, using linear combinations of the output neurons. This model was validated by showing its ability to identify direction selective neurons; the model also made predictions about the function of neurons that perform different visual processing tasks, which is an important tool for discovery.

Using calcium imaging as a constraint, Dr. Turaga and colleagues also built a model of the entire *C. elegans* nervous system focused on chemosensation. Dr. Albert Lin simultaneously imaged calcium signals in 170 neurons in response to presentation of three odors. In contrast to the fruit fly vision system, the odorant responses of many *C. elegans* neurons were strongly state-dependent and exhibited significant trial-to-trial variability, which cannot be replicated in a deterministic model. A probabilistic latent variable model was therefore built using connectome and biophysical constraints, which could infer the internal state of the network and make predictions about future neural dynamics.

This work by Dr. Turaga and colleagues revealed several strategies for combining modeling constraints with connectomic data to develop detailed models and make predictions about properties of individual neurons in brain circuits. Multiple models were found be consistent with these constraints if insufficient constraining information was used. The next step is to build tools to explore model variability and suggest experiments. The long-term goal is to extract understanding and insights about high-level neural computations through application of these types of connectome-constrained models.

Learning Representations of Neural Architecture Across Many Spatial Scales

Eva Dyer, PhD

Dr. Dyer described her work in representation learning to build models of neural architecture across many spatial scales. To build comparative models that enable the comparison of connectomic datasets across individuals or across brain areas within the same individual, representations must be extracted from high-dimensional data (e.g., EM images) and an abstraction of those representations formed to produce low-dimensional models of brain structure (e.g., graphs) that allow for interpreting and comparing the underlying structure.

Dr. Dyer's team is working on two projects to investigate the formation of low-dimensional representations of underlying structure. The first project uses micron-scale features to understand how the structure of brain areas differ and what information within images enables researchers to divide the brain into different areas. With colleagues at the Argonne National Laboratory, the team acquired synchrotron X-ray microtomographic images over large sections of tissue, which revealed the distinct appearance of brain regions due to differences in the patterning of white matter, cell densities, and other microarchitectural features. To build low-dimensional abstractions from images of heterogeneous brain areas, Dr. Dyer's team developed a tool called Deep Brain Disco(very). This tool trains a convolutional neural network to discriminate tissue features across high-level regions of interest in images from different brain areas and then extracts representations from the trained architecture. Those representations are then studied to gain insight into the underlying structure within different brain regions. Although trained on higher-level brain regions (e.g., cortex, thalamus), the network was able to identify differences within brain regions and similarities across brain regions.

A second project used a neural network approach to address the challenge of extracting individual differences in brain structure across many brains. A generative modeling approach, which creates new images resembling those it was trained on, enabled Dr. Dyer and colleagues to build a model of the underlying distribution of the images that can be used for image classification. To map the underlying patterns of variance captured in a deep generative model, approaches were developed to map a model's receptive and projective fields and were applied to images from 1,700 individuals in the Allen Mouse Connectivity Atlas.

In summary, Dr. Dyer described two different ways that this representation learning can be leveraged to build models of brain structure variability at the micro and macro scales to train a neural network to distinguish brain areas and extract and factorize network representations. This approach can be coupled with a microstructure segmentation network to build multi-scale brain maps. At the macro scale, this neural network approach can be used to extract signatures of individual difference between brain images.

Lessons from Analyzing Navigational Circuits in the *Drosophila* Hemibrain Connectome

Hannah Haberkern, PhD

Dr. Haberkern shared lessons learned from analyzing navigational circuits in the *Drosophila* hemibrain connectome. Insects navigate using a variety of external sensory cues to update an

internal population representation of the head direction, analogous to a "neural compass." Analyzing the fly connectome enabled extraction of the circuit logic of the fly compass system to be extracted.

Analysis of connectivity often involves quantifying the strength of a connection between two neurons. For many neurons, it is vital to not only count synapses, but also consider where these synapses are made on the target neuron. The connection measure should therefore consider neuronal morphology and brain anatomy. Although neuron-to-neuron connection measures can be used to study larger networks, Dr. Haberkern and colleagues found that connectivity in the fly navigational circuit was often highly structured in a way that was well-aligned with categorization by neuron type, as previously defined by using genetics and light level anatomy. This consistency in structure allowed the analysis to focus on type-to-type connectivity instead of neuron-to-neuron connectivity, which facilitated the discovery of network motifs. The integration of prior knowledge about the function of at least some neuronal types in the fly compass was key to gaining insight into the meaning of these network motifs. This was the case in Dr. Haberkern's analysis of input circuits to the fly's head direction system. Prior work had established input pathways that convey sensory information to the neural compass, and analysis of the connectome revealed motifs that likely determine the prioritization of sensory cues that update the neural compass. Matching functional data to the connectome also facilitates hypothesis testing about circuit function.

In another example of scientific insight extracted from the connectome, circuit motifs were identified that enable coordinate transforms, so that the fly can compute its direction of travel when it moves in a direction that is not aligned with its head direction. This transformation from head to body coordinates is a prerequisite for path integration. The connectome analysis made predictions about information that should be carried by specific neuron types, which were well-aligned with previously collected physiology data.

Several infrastructural supports were key for enabling this work. The hemibrain connectome data were made available through a service called neuPrint that provides a web-based database interface. This hemibrain database contains the complete dataset and is queryable and expandable. Moreover, a well-documented programmatic interface allows for rapid development of community-driven analysis tools, such as neuprintr, which was used for this project.

Neural Network Organization for Courtship Song Feature Detection in *Drosophila*

Christa Baker, PhD

In animals that communicate acoustically, timing patterns within sounds can convey information about the signaler such as species and fitness. Relevant temporal patterns within acoustic signals can occur on a range of timescales from seconds and longer down to milliseconds, which means that at some stage the nervous system must detect or compute multiple temporal patterns. Male flies use wing vibration to produce courtship song consisting of a sinusoidal mode and a pulsatile mode, and rapidly and dynamically switch between these modes such that no two fly songs, even from the same individual, are identical. Because of

these richly varying acoustic patterns, as well as available connectomic data and genetic tool, flies are a good model for understanding the circuit computations involved in acoustic pattern analysis. Dr. Baker discussed a collaborative effort to integrate connectomics with neural recordings to understand *Drosophila* auditory circuits.

To overcome previous limitations in knowledge of the neurons involved and the circuit organization of the fly auditory pathway, Dr. Baker and colleagues first used genetic lines to identify 28 new auditory cells types and imaged calcium responses of these neurons to pulse and sine courtship song. Quantification of the song mode preference for each neuron revealed a continuum of preferences for sine or pulse song across auditory cell types, which suggested several potential models for how these cell types connect to one another. In one possible connectivity model, this continuum arises from completely separate pathways for pulse and sine song, with cell types connected hierarchically such that pulse and sine song preference along the two pathways is sharpened. In a second model, interconnections occur among neurons with different tuning without clear hierarchy.

Dr. Baker and colleagues used connectomics to distinguish between these two models. FlyWire was used to identify and map the synaptic connections for almost all known auditory neurons in both hemispheres of an EM volume of the entire female fly brain. This effort generated the first auditory connectome in flies and included the identification of auditory cell types, song mode selectivity, synaptic connection directionality and laterality, and excitatory versus inhibitory neurons. This connectome revealed that auditory neurons are richly interconnected, with many connections between differently tuned neurons. Metrics from network theory revealed that the connectivity of auditory neurons lacks hierarchical structure and is more consistent with a randomly connected network. This type of organization is different from predicted based on other auditory systems and suggests that dense interconnections among the more than 30 auditory cell types are important for shaping neural tuning to acoustic stimuli.

The next step is to ask how auditory neurons extract the many temporal patterns within fly songs. Addressing this question will require identifying neurons that may be missing from the auditory connectome using tools that are part of the natverse toolbox. Going forward, it will be helpful to use connectomes to provide testable hypotheses about the function of microcircuits or individual neurons based on connection patterns, which will likely require developing computational tools that account for both connectomic and functional data. The ability to generate connectomes more quickly would enable further understanding of circuit components that are stereotyped or variable across individuals, the development of connectivity patterns, and the effects of experimental manipulations on connectivity and circuit function.

Ciona Connectome Analyses and Considerations for the Future

Kerrianne Ryan, PhD

The larvae of the invertebrate tunicate Ciona intestinalis (sea squirt) are a valuable model system because of their small size, abundance, transparency, and larval swimming behavior. Dr. Ryan discussed the contributions of connectomic analyses to understanding the Ciona nervous system, including elucidation of sensory-motor pathways, identification of novel neurons,

behavioral predictions, and descriptions of network characteristics. The primary value of connectome analyses is that it generates hypotheses and reveals a substrate for investigating a wide variety of computations, pathways, interactions, and behaviors. Comparative connectomics also highlights plasticity within the nervous system, underlying organizational modules, and inter-individual differences based on genetic, experiential, and environmental substrates.

Integrating across scales has proven particularly useful in Ciona, because each scale offers different advantages and disadvantages. For Ciona, the use of landmarks to match coordinates as well as live imaging and genetic tools is particularly helpful for integration across scales. A future goal for integrating connectomic data and functional data in Ciona is the ability to interrogate single cells or single synapses with high-resolution imaging and to obtain the connectome in a single animal.

Of the many promising methods for enabling analyses of connectomes, those that link structure and form to functional analysis are the most powerful. Recent developments in high-throughput methods enable comparative connectomics at a range of physical and temporal scales, which is critical for the analysis of dynamic systems. Other powerful approaches include linkage of high-resolution connectomics with single-cell transcriptomics, spectral connectomics with expansion microscopy, nanobody immunostaining, and methods for understanding gap junctions. Increasing accessibility of these tools and resources will support more effective contributions to connectomic analyses by a broader community of researchers.

Whole-Brain Functional Studies in *C. Elegans*: A Direct Comparison of Structure and Function

Albert Lin, PhD

Access to the connectomic wiring diagram for the 300 neurons of *C. elegans* has informed biological understanding of circuit maps from sensory inputs to motor outputs. To identify functional circuits, functional recordings from a small number of neurons are often combined with identification of synaptic connections. However, the neural circuits that mediate complex behavioral tasks such as chemosensation, which require large numbers of neurons and sometimes lead to context- or state-dependent responses, cannot be described using this straightforward approach. Recent advances in functional imaging have expanded the ability to capture brain-wide context, by enabling activity recordings of the entire *C. elegans* brain with single-neuron resolution. However, these types of functional data cannot be mapped onto the connectome unless the individual neurons can be identified.

Dr. Lin and colleagues have addressed the problem of comprehensive neuronal identification with deterministic multicolor labeling, using a tool called NeuroPAL that enables simultaneous recordings from many neurons while animals respond to complex inputs. Such functional imaging experiments revealed that chemosensory stimuli engage large portions of the *C. elegans* nervous system. Pairwise correlations for all neurons in the dataset showed that the average patterns of stimulus-evoked activity were unique to each stimulus. Because all of these

neurons were identified using the connectome, the functional activity correlations and structural weights in the anatomical connectome could be directly compared. This linear regression showed that neither the chemical nor electric synaptic connections in the wiring diagram were significantly correlated with any of the functional correlations in the whole-brain data, suggesting that anatomical connectivity is not directly deterministic of neural activity.

To better understand brain-wide context, the researchers analyzed average activity patterns evoked by multiple presentations of a single odorant. Although the responses of sensory neurons were highly consistent, responses across trials and across individuals in the rest of the brain exhibited substantial variability. For example, over multiple trials of the same stimulus, the neuron that commands forward and backward movement (AVA) was sometimes active and sometime inactive during an odor presentation. Access to neuronal identifications enabled the identification of populations of neurons whose response patterns correlated with AVA activity. The goal for this work is to eventually use data such as these to uncover how complex behavioral decisions are made as a function of the activity of all upstream neurons.

Combining labeled whole-brain functional data with the *C. elegans* connectome also makes it feasible to construct whole-brain models constrained by the topology of the wiring diagram. In collaboration with Dr. Turaga's group, a neural network model is being built that is constrained by the anatomical connectivity and trained on the pan-neuronal functional data obtained by Dr. Lin and colleagues. The results of this model show that a connectome-constrained model can more accurately recapitulate ground truth neural activity than an unconstrained model. These models are expected to become more refined as more labeled data are generated. Looking forward, combining these experimental and theoretical approaches should inform understanding of how a whole brain combines deterministic sensory readouts from a large number of neurons with state-dependent context to drive behavioral decisions.

Modularity and Neural Coding from a Synaptic Wiring Diagram

Ashwin Vishwanathan, PhD

Dr. Vishwanathan discussed his work on modularity and neural coding from a synaptic wiring diagram of the larval zebrafish circuit that controls eye movement. Larval zebrafish make several types of eye movements, including sharp deflections (i.e., saccades) in between which the animal fixates on a target. The neural circuit that controls eye movements must be robust to be able to compensate for body movements during swimming so that the eyes remain fixed on what the animal is looking at.

To delineate some of the mechanisms by which the eye movement circuit works, Dr. Vishwanathan and colleagues looked for clues in the brainstem by combining SEM with functional imaging. A partial larval zebrafish connectome was generated that contained 3,000 reconstructed neurons and all synaptic locations. The reconstructions were then aligned to a LM reference atlas, which is useful for providing morphological and molecular information. Next, network science tools and a measure called "centrality" were used to divide the reconstructed neurons into a recurrently connected center, which forms the core of the network, and a periphery. These neurons were represented by a matrix of rows of presynaptic

and columns of postsynaptic neurons, with each point color-coded to indicate the number of synapses per neuron. To determine whether the matrix had an underlying structure, it was organized to maximize a measure called modularity (i.e., the tendency of a neuron to connect with other neurons within its own cluster), which led to division of the matrix into two clusters such that there was more crosstalk within modules than across modules.

The functional implications of these modules were investigated by analyzing the projections of neurons within modules. Neurons in one module tended to project onto spinal projecting neurons (the "body axis module"), whereas neurons in the other module tended to project onto abducens neurons, which control eye movements (the "oculomotor module"). Crosstalk between these two modules could enable compensation of eye movements during body movement. Within the oculomotor module, Dr. Vishwanathan asked whether a linear matrix model could be used to compute a pattern matrix specific for eye movements. By using synapse numbers obtained from the connectome, the model was able to derive measurements for eye position sensitivity, which measures the firing rate as a function of eye movement, to reveal that the predictions from the model closely match the predictions from the functional imaging.

Using this combination of approaches, Dr. Vishwanathan and colleagues identified key circuit elements of both functional and anatomical modules and built a theoretical framework to gain a larger mechanistic understanding, with synapses numbers from the connectome enabling the prediction of tuning properties. With regard to the larger question of how to integrate connectomic data with functional data, reference atlases can help bridge the divide between LM and EM and provide a crucial link to molecular information.

Session 1 Q & A

Steppingstone Connectomes

Dr. Jefferis discussed his recommendation for the use of stepping stone connectomes between the fly and the mouse as well as analysis tools needed to scale up. Stepping stone connectomics in other taxa and smaller vertebrates between flies and mice might solve the many technology problems that hinder the leveraging of these technologies on the scale of a mouse connectome. Moreover, analyzing circuit architecture and molecular cell types in brains smaller than the mouse will provide useful insight into cross-species conservation, particularly in subcortical regions of the brain. Dr. Jefferis also emphasized that much work remains to be done in the fly, especially in the crucial area of connecting connectomic information to molecularly defined cell types. In terms of technology, developing forward genetic labeling strategies instead of screening strategies will be essential for other species and can perhaps be achieved by combining classical bioinformatics and connectomics.

The Role of Functional Modules in Connectome Analysis

A question was raised about the importance of functional modules and whether all connectomic analyses should search for and characterize functional modules because those are likely to be evolutionarily conserved and fundamental to all brains. Dr. Vishwanathan confirmed the importance of looking for such functional modules in all species and noted that they are more

likely to be found in evolutionally conserved areas such as the brainstem. A substantial debate persists about whether connectivity alone can be used to classify functional modules in higher cortical areas, although other factors may be useful for classifying a module as functionally relevant.

Dr. Lin added that because *C. elegans* is so small, it is difficult to cluster neurons based solely on the wiring diagram. The approach that many *C. elegans* researchers follow is to group neurons by role, by activity in response to a certain stimulus, or by correlation of activity with a particular behavior. Moreover, in a system with only 300 neurons, most neurons likely play multiple roles and are only two steps away from any other neuron in the wiring diagram.

Dr. Haberkern suggested that knowing a connectivity motif without knowing the tuning of the neurons and the nature of connection has only limited meaning. She also noted that fly researchers are used to thinking about specific cell types, which works well with the analysis of motifs in the connectome. This framework might not work as well in other models or brain areas, such as mouse cortex. Dr. Baker added that because connectomes in flies are relatively new, inferring function from connectivity alone remains difficult. For example, in the auditory connectomes, she does not expect that every neuron with a strong connection to an auditory neuron will also respond to sound and that they will be connected with many other neurons. Much more data are needed to determine how well connectivity can be used to predict function in the fly brain.

Variability

Dr. Murthy raised the issue of across-animal and within-animal variability that emerged from several presentations. Dr. Dyer noted that her work has revealed a great deal of variability in the connectivity within different areas of the same brain such as cortex, as well as the likelihood of more variance across brains. Determining how to address variability will be challenging, and devising strategies to match different partial connectomes across connectomes will be critical. She is considering an approach that uses a different way to represent the overall connectome with lower-level factors, which may enable comparison of partial snapshots of connectomes across brains. The challenge of sampling, variability, and the scale of data will all come into play. However, combining modalities that can be used to visualize the data across scales may enable across-animal and within-animal comparisons.

Dr. Lin commented that *C. elegans* exhibits a great deal of functional variability within and across animals that is likely not a direct consequence of variability in the connectome. Interneurons exhibit more response variability than sensory neurons, which is consistent with the idea that sensory responses are stereotyped and that downstream circuit processing will lead to variable neuronal activity and different behavioral decisions in response to the same stimulus.

Dr. Haberkern noted that having only one connectome for a particular animal impedes understanding of variability. However, a strategy that maps genetically defined cell types enables analysis of a much larger set of light level images, which can be used to understand the variability in structures (e.g., dendritic arbors) within cell type classes.

Dr. Jefferis reemphasized that analyses of fly brains thus far have found that left-right variability is very similar to across individual variability in inbred animals. He believes that this finding is important because comparing left-right connections is the best available approach to thinking about which connections matter in the connectome and can be studied across animals. There can also be compensation in the face of variability; for example, differences in the number of neurons of a given cell type on the two sides of the brain can be compensated for by differences in the numbers of synaptic connections. Thus, variability in cell numbers does not always imply a difference in the ultimate connection properties.

Connectome Constrained Models

Dr. Turaga further described his work on the interactions between the density of connectivity in a connectome and other types of model constraints. In addition to identifying which neurons are connected, connectomes provide graded information about strength of those synaptic connections by using synaptic counts as a proxy. With synaptic count information, the network does not need to be sparse for successful modeling. However, if connections are known to exist between two neurons but connection strength is unknown, sparse connectivity is required. Dr. Turaga suggested that 10 percent connectivity was required, but explained that the key takeaway from this example is that the required level of connectivity also depends on how strongly the task and other factors constrain the model.

Building Models from the Data

Missing data from the connectome (e.g., synaptic signs or weights) presents problems for scaling up to larger brains and building models based on connectomic data. For example, a connectome may not contain synaptic signs or weights. Connections may also extend beyond the brain area represented by the connectome, as described by Drs. Baker and Haberkern for the fly Hemibrain. When functional data are mapped on the connectome, as described by Dr. Baker, nodes may be missing where neurons have not yet been functionally identified. Several strategies were discussed to compensate for missing pieces of information. Dr. Turaga agreed that not knowing the sign of a synapse complicates his modeling approach. While synapse counts would be useful for estimating the weight pattern and distribution across the network, not knowing whether the sign of synapses is positive or negative, particularly for strong synapses, can present a significant challenge for models that are trained, and knowing the sign of synapses substantially limits the combinatorial search space of sign possibilities. These are technical problems of machine learning rather than fundamental issues, which may be overcome in the future.

Dr. Jefferis reemphasized the recent advances made in neurotransmitter identity prediction and their utility for determining the sign of synapses. It will be important for other species to consider how to obtain that type of neurotransmitter information. One possibility when using a sectioning EM approach for mouse brain is that an occasional section could be used for molecular probing to reveal neurotransmitter identity (e.g., antibody staining) without much loss of connectivity information. A similar strategy might be useful for gaining information about other functional properties. Dr. Ryan added that work in Ciona is applying a similar

strategy of using information about neurotransmitter identities to classify synapses and make predictions.

For systems in which models are based on a great deal of functional understanding, the speakers discussed whether additional insight can be gained from models built from the connectome. For the larval zebrafish eye movement system, Dr. Vishwanathan and colleagues were fortunate that the connectome looks exactly as the models predicted. Models that deviate from connectome data likely need to be altered in order to provide useful information about model parameters. Dr. Haberkern remarked that morphological data from the connectome can identify neuronal sub-compartments (as observed in calcium imaging studies) to inform models that provide insight into local computations within neurons.

Session 2: Analysis of Multiple Connectomes

From a Developmental Connectomics Study: Analyses, Interpretations and Limitations

Mai Zhen, PhD

Dr. Zhen studies developmental connectomics and attempts to determine how maturation proceeds. Her team chose *C. elegans* as a model system due to its compact nervous system, with 300 neurons and fewer than 3,000 connections among those neurons. Using high-resolution EM datasets, Dr. Zhen's team reconstructed seven full contactomes and eight full connectomes from eight isogenic individuals at different developmental timepoints.

Using these datasets, Dr. Zhen studied how the brain wiring matures after birth. Her team found that neurite growth is uniform and the neural scaffold is stable across development. Early in development, robust and continuous synaptic growth established a symmetric framework and a baseline synaptic density, and later developmental stages showed continuous increases in neuron-neuron connectivity in both number and synaptic weight.

These findings show that one connectome per stage, like one connectome per animal, does not apply. Dr. Zhen and her team devised the approach to classify connections by the trend of synaptic weight changes across development. They found that each connectome consisted of two portions: (1) an invariant portion, which consists of developmentally stable and developmentally dynamic subgroups of connections, and (2) a variant portion, with changes in synaptic weights of connections that show no specific trend and appear to be stochastic events. They also found that no threshold of synaptic number or size could cleanly separate these groups of synaptic connections. These findings reveal substantial wiring plasticity even among isogenic individuals.

Separating invariable and variable connections allowed Dr. Zhen's team to examine how they affect brain maturation. An examination across the connectome revealed that (1) connections among interneurons consist of the least developmentally dynamic connection, hence the core-decision circuit exhibits least developmental plasticity and (2) across development, increased connections define an increased modulatory.

Dr. Zhen then asked what are the specific rules that govern maturation and developed a testable question: What guides targeted synapse growth that leads to these network-level changes? Her team studied *C. elegans* connectome and contact across developmental stages, and correlated adult connectomes with the first-stage larval connectome and contactome to determine which features in the larval stage hold predictive values for the adult connectome. Pursuing this specific question established broader rules of connectome maturation: that growth is uniform to maintain a stable scaffold, that differential contact size guides formation of new connections, and that synapse addition is both neuron-type and connection-type preferential.

Combining Connectomics with Functional Studies to Investigate Circuit Mechanisms of Learning and Action-Selection

Marta Zlatic, PhD

To explore the neural implementation and circuit mechanisms of learning algorithms and learning-based action-selection, Dr. Zlatic's team built brain-wide maps of synaptic connectivity, cellular resolution activity, neuron-behavior relationships, and RNAseq gene expression in larval drosophilae. Larval *drosophilae* are appropriate model systems because of their compact nervous systems, each consisting of approximately 15,000 neurons; their transparent cuticles, which enable easy measurement of neural activity; their rich behavioral repertoire, particularly as related to learning; their uniquely identifiable neurons, which enable easier network establishment; and the many genetic tools available to work with them, which expand possible approaches.

Drosophila can be trained to approach or avoid specific odors, a behavior that can be further manipulated with optogenetics. The *Drosophila* learning circuits have three types of cells: Kenyon cells, dopamine neurons, and mushroom body output neurons (MBONs). While the interaction among these cells is well-established, the circuits regulating dopamine neurons and the populations downstream from MBONs that provide integration of positive and negative outputs were unknown until Dr. Zlatic's team used an EM-generated synaptic-resolution reconstruction of the larval *Drosophila* brain to discover novel populations of neurons and novel circuit motifs. Because the team intended to study function as well as structure, it combined structural maps with 2P imaging of neural activity and modeling to discover potential roles of newly identified circuit motifs.

In the future the team intends to image the activity of all neurons in its *Drosophila* subjects using light sheet fluorescence microscopy before, after, and during learning. Combining these maps with connectivity data from EM-generated reconstructions will enable the team to develop testable hypotheses about these circuits and to compare across individuals. To determine structural and molecular correlates of enhanced learning performance, the team plans to study subjects across life stages and species as well as *Drosophila* that are consistently good or bad learners.

Multi-Scale Statistical Learning for Connectome Data

Jesús Arroyo, PhD

Dr. Arroyo designs statistical models to interpret connectome data. He explained that each individual organism has specific connectivity that can be used as diagnostic biomarkers, for growth charting, or for connectome fingerprinting. A variety of approaches can be used to predict these individual attributes from brain connectivity. The most popular are the bag of features, in which connectomes are represented by functions of the data and which cannot effectively characterize variability; bag of edges, in which a connectome is vectorized before statistical tools are applied; and network-aware statistical tools, which, unlike both bag approaches, display both local and network structure when assessing connectomes.

By exploiting low-dimensional structure in the data with network-aware statistical tools, researchers can impose regularizations or penalties to improve sparsity, smoothness, and low-rankness in their representations and thereby improve their interpretations and accuracy of the connectomes. Analysis can be conducted at a microscale, or edge-and-vertex level; macroscale, or network-level; and mesoscale, which represents communities, or meaningful groups of vertices.

Dr. Arroyo's team developed methods across these scales to predict attributes across connectomes, especially for schizophrenia. At the microscale level, the team developed convex regularization to conduct variable selection and identify predictive edges and vertices. This method has high accuracy, but its granularity narrows the range of useful interpretations it can inform. At the mesoscale level, the team developed simultaneous prediction and community detection to conduct clustering and prediction and identify communities. With this method, it achieved smoothness and regularization and reduced data dimensions more intuitively to enable more efficient analysis and modeling. At the macroscale level, the team developed joint embeddings of multiple networks to conduct graph embeddings and develop the latent space representations of vertices and networks. This method uses diffusion MRI for raw data and both characterizes the variability in connectomes and discriminates between individual connectomes.

These methods perform analysis at different scales and provide insight into how connectome data interact with other attributes. However, many questions remain. Future work may include adding information such as edge, vertex, or network covariates; adding a temporal dimension to connectomes; or integrating data from different sources. Nonparametric methods used on large-scale connectomes may enable more flexible and robust methodologies that do not rely on structural assumptions and capture connectome properties that are currently neglected.

Optimization of the Reliability of Functional Brain Connectome

Ting Xu, PhD

Comparing connectomes across species provides insight into evolution of the human brain. Dr. Xu's team specifically studies the comparison between macaques and humans. The brain surface expansion observed between macaques and humans is comparable to cortical

expansion over human childhood, and both species show similar inter-individual variation and long-distance functional connectivity. Most alignment processes compare brains solely based on anatomical landmarks.

To compare beyond anatomy, Dr. Xu's team used fMRI to compare functional gradients across species and find similar connectivity between areas and among cortical hierarchies. A joint embedding approach simultaneously extracted and matched fMRI data between species, producing a joint-similarity matrix of functional connectivity. A diffusion embedding algorithm extracted components from the joint-similarity matrix.

Each component in each gradient is one dimensional and can be used to build high-dimensional spaces. The gradient score for each node in those high-dimensional spaces provides vertices in both human and macaque cortices and aligns their brains to these high-dimensional spaces. The team added a second component highlighting sensorimotor regions in both species and a third highlighting middle temporal visual areas of the brain. Using these spaces, the team mapped the human and macaque surfaces together. After establishing cross-species alignment, the myelin maps were aligned, which showed that human and macaque myelin maps are comparable.

The researchers used this alignment process to assess cross-species similarity and homology scores in primary areas. Homology indices follow cortical hierarchy from primary to high-order networks. The two species show consistency in a variety of pathways, including in the dorsal-to-ventral topology in sensorimotor regions. Of interest, although homology within regions that control limbs and the trunk is high, face and eye control regions show less functional homology. Dr. Xu suggested this may be due to the complexities of speech and emotional expressions seen in humans. High-order networks associated with self-referencing showed the lowest homology between humans and macaques. The process Dr. Xu described provides the opportunity to study evolutionary processes concerning higher-order cognitive functions as well as general brain deformation.

Network Data Science for Bilateral Brains: Applications in the Larval *Drosophila* Connectome

Benjamin Pedigo, BS

Mr. Pedigo used a larval *Drosophila* brain connectome to study bilaterality and correspondence between the hemispheres. A full reconstruction of the *Drosophila* brain, constructed in part by Dr. Zlatic and Dr. Albert Cardona's groups, was instrumental to this research. Correspondence between left and right hemispheres can be defined by connectivity, represented within each hemisphere by an adjacency matrix. Using network matching, which determines the permutation of one network with regard to another such that their adjacency matrices become as similar as possible, the researchers found approximately 80 percent agreement between their unsupervised method and the manual matching between hemispheres conducted by Dr. Michael Winding, suggesting the utility of these methods for automated matching for future connectomes.

The random dot product graph model, a statistical network model that assumes the connectivity of each neuron is determined by a latent vector, provides parameters with which researchers can align and use to meaningfully compare the hemispheres. Using this strategy, Mr. Pedigo found that the left- and right-hemisphere parameters arise from the same base distribution.

That knowledge informed deeper hypotheses about the connectivity within and similarity between hemispheres. To identify cell type based on connectivity, the group's approach was to learn a joint embedding of the two sides and to use a hierarchical clustering process, finding similar cell type distribution in both hemispheres. This clustering implicitly assumes that the connectivity of neurons is a function of their community. Using a degree-corrected stochastic block model (DCSBM), which explicitly parameterizes probabilities of different connections between nodes as a function of community, the group sampled a synthetic network with a high degree of similarity to the data it obtained from the real *Drosophila* brain.

To formalize similarity, Mr. Pedigo charted the likelihood of the DCSBM correctly predicting connectivity as a function of clustering level. As predicted, when the group fit the model to one hemisphere and compared it to that hemisphere's actual connectivity, the similarity increased with each new level as the model became more complex. However, when it fit the model to one hemisphere and compared it to the opposite hemisphere's actual connectivity, the similarity peaked at approximately level six. Because the researchers expect that a model that most effectively predicts the opposite hemisphere would most effectively predict other brains' connectivity, they ceased adding levels once it became clear that further complexity was not helpful for prediction here and may reflect overfitting.

Mr. Pedigo emphasized the generalizability of these methods, extending to clustering neurons jointly between connectomes or testing for differences between connectomes, and offered for use his open-source Python package graspologic. He mentioned that new types of analyses needed to further the field include means of generalizing across scale (e.g., learning a mapping between a model in a larva and an adult) and relating connectivity models to other properties, such as cell morphology. For interesting comparisons between connectomes, multiple samples across species and conditions are needed.

Session 2 Q & A

Matching Technique Accuracy

Mr. Pedigo described using network matching to determine connected pairs of neurons. He clarified that when using connectivity alone, groups that assessed the accuracy of network matching in comparison to manual matching found approximately 80 percent agreement. When they incorporated morphology as well, the agreement rose to approximately 85 percent.

Variations in Connectivity

The source of the hemispheric asymmetry described in the two talks using larval *Drosophila* brains is not known. Mr. Pedigo stated that on any given EM dataset, some individual neurons appear developmentally aberrant, potentially contributing to this phenomenon. He also emphasized that different models have different degrees of granularity, and the models described in his presentation were high level and smoothed, potentially reducing actual variation. Dr. Zlatic added that while functional activity is approximately the same between the hemispheres in *Drosophila*, there are relatively few strong and reproducible connections among many weak and nonreproducible connections. However, most connections for each individual neuron are strong. Variation may arise in part from these weak connections.

Dr. Zhen explained that asymmetries across hemispheres in *C. elegans* fall into three types of differences. Of 188 neurons in the brain, only 12 are completely without a match on the opposite hemisphere. A second type only appears over the course of maturation. In a final type of asymmetry, pairs of neurons on each side of the brain show similar wiring and positioning within their hemispheres but are molecularly and functionally different. These findings may translate to *Drosophila* hemispheric asymmetry.

Similarly, whether variations across brains in given datasets are wiring noise or meaningfully different is unknown. According to Dr. Arroyo, this question can be answered with supervised graph embedding procedures that incorporate regions or features of interest. With this method, researchers could distinguish differences and identify neural relationships, using slightly modified models dependent on the specific topic of interest.

Functional Connectome Homology

In her talk, Dr. Xu explained that homology across the small species gap between humans and macaques appears primarily in lower cortical hierarchy regions. She does not currently work on cross-species subcortical homologies, because the proportion of total connections found subcortically varies dramatically across species, complicating the comparisons. Her team is currently pursuing cortical alignment using small animals such as mice and marmosets; the mice in particular have proportionally far more subcortical connections than larger primates. To seek homologies across even larger species gaps—for example, from humans to *C. elegans*—researchers would need to develop new methods.

RNA-Seq Analysis Overlay on Connectome Diagrams

Multiple speakers described an intent to develop molecular connectivity maps that can be directly integrated with connectome diagrams. Drs. Cardona and Zlatic collected the RNA-seq data that Dr. Zlatic presented by dissociating the neurons of the entire larval *Drosophila* CNS and performing drop seq on each individual cell. To regain positional data, regions would need to be partitioned, potentially with added genetic markers. Dr. Zlatic added that researchers will likely need to do bulk RNA-seq with labeled neurons to accurately overlay the molecular maps onto EM sets. According to Dr. Zhen, a recently published study using *C. elegans* used combinations of neuropeptide genes to classify neurons and preserve positional information.

Session 3: Making Connectomic Data Accessible to the Community

Engineering the Future of Connectomics

Will Gray-Roncal, PhD

Dr. Gray-Roncal presented the Brain Informatics Program (BIP) and its contributions to reverse engineering the brain. Although the process of reverse engineering consists of eight steps reminiscent of the scientific method, Dr. Gray-Roncal specifically focused on data storage and processing.

BIP provides a cloud-native, scalable repository called bossDB intended for storing, sharing, and analyzing EM and X-ray microtomography data. bossDB stores raw image data, meshes, skeletons, graphs, functional data, and annotated data. BIP also provides a tool called Scalable Analytics for Brain Exploration Researcher (SABER) for data processing, which includes unified and standardized interfaces for many tools used in the community. Using Apache Airflow and common workflow standards, SABER supports processes as diverse as registration and segmentation of the brain. Standardization of the data and interfaces aids secondary analyses and effective data sharing.

Standardized data products and formats necessitate standardized query formats. BIP represents connectomes with graph theory and enables the use of simple syntax to query its database with specific hypotheses (e.g., find all mutual inhibition motifs in this connectome) to search for subgraph isomorphisms. Dr. Gray-Roncal hopes that this database can be used to explore connections between motifs and functions, discover motifs, and develop and implement models.

Neuroscience Research on the Basis of Whole-Brain Connectomics

Albert Cardona, PhD

Dr. Cardona studies the relationship between circuit structure and function using larval *Drosophila* brains and builds computational models based on circuit dynamics and synaptic connectivity. His team collected images of the whole CNS of *Drosophila*, using first serial section transmission EM (ssTEM) and later FIB-SEM, which produces data more suitable for computational processing.

The team invited extensive collaboration to build a full cellular connectome of the larval *Drosophila* using a tool called CATMAID (Collaborate Annotation Toolkit for Massive Amounts of Image Data). Any researcher can access this tool and its related annotated data, and many researchers have used it to study and publish on a variety of topics. For example, the database has enabled extensive research into the regulation of insulin release, uncovering participation from unexpected body and neural systems.

Within the *Drosophila* learning and memory system, Dr. Cardona determined that building new memories depends on older memories, based on how MBONs and dopaminergic neurons are each regulated by their own output neurons. With this computational model, his team

successfully compared connectomes across species and discovered a similar structure in *C. elegans* using data provided by Dr. Zhen. This finding highlights Dr. Cardona's primary takeaway for advancing connectomics research: all data should be shared at all times.

Continuous Proofreading and Analysis of Large EM Reconstructions

Sven Dorkenwald, MSc

Mr. Dorkenwald asserted that proofreading and analysis of connectomes should be conducted simultaneously and discussed the infrastructure needed to make connectome data and analysis widely available to the neuroscience community. Proofreading large datasets is a long process; for example, 50 person-years were required to proofread the *Drosophila* hemibrain connectome. Using a newly launched platform called FlyWire that uses the FAFB dataset, proofreading the *Drosophila* brain requires 1 hour per neuron or approximately 100 person-years for the full brain. With participation of the full neuroscience community, it should be possible to produce a fully proofread fly connectome within 1 year using FlyWire. A completely proofread connectome is possible on a fly scale. However, proofreading a single mouse brain pyramidal cell as part of his team's IARPA MICrONS project requires more than 50 hours because of the complex axonal and dendritic branching. The time required to proofread a mouse brain connectome would far outlast the time required to collect it.

Although proofreading larger datasets is a never-ending process, partial or incomplete connectomes are still valuable datasets that researchers can use to test hypotheses. To enable accessible and dynamic proofreading, Mr. Dorkenwald's team developed a centralized and collaborative proofreading platform called ChunkedGraph. The connectomics pipeline begins with reconstruction of the brain, followed first by adding metadata such as cell types and synapses, and then by analysis. Adding proofreading to this process means that the data being created are always changing. Collaborative annotation infrastructure called CAVE (Connectome Annotation Versioning Engine) enables researchers to proofread reconstruction as it progresses and to share that information while maintaining separation between metadata and the reconstructions. With continuous or frequent data releases, CAVE provides a centralized proofreading platform with timestamped edits and separate annotation storage, providing the infrastructure needed to enable future research.

Exploring Connectomes at Varying Levels of Detail

Stephen Plaza, PhD

The hemibrain initiative reconstructed a large portion of the *Drosophila* brain that became the largest dense connectome ever produced. Dr. Plaza's team made all data in the hemibrain dataset easily accessible to accelerate research and used the data themselves to study associative learning. There are now 20 terabytes of EM image data in FlyEM, consisting of 25,000 neurons and 20 million proofread synaptic connections.

Researchers can use FlyEM to study specific brain regions, the morphology of any of these neurons, the exact locations of synapses, or the number of synapses between two neurons. A connectome analysis tool that facilitates graph-centric queries called neuPrint

(neuPrint.janelia.org) stores data at multiple "levels," including the region, neuron, and synapse levels, facilitating a variety of modeling needs, and only needs 1 second to find all paths between two neurons. Hundreds of unique users access neuPrint daily.

Although neuPrint currently only holds data on *Drosophila* brains, it may be useful for mouse or zebrafish brains. With neuPrint, researchers could add more information beyond the connectome data; for example, mitochondrial research could use this repository. Using an image search tool created with deep neural networks that is featured on clio.jaenlia.org, researchers can search terabytes of imaging data to find patterns within neural ultrastructure.

The Brain Image Library: Designated Repository for NIH BRAIN Initiative Microscopy Data

Alexander Ropelewski, PhD

As part of NIH's Brain Initiative, Dr. Ropelewski serves as the Operations Director for the Brain Image Library (BIL), a designated public repository for brain image data. Scientists can use the BIL to deposit, analyze, mine, share, and otherwise interact with microscopy datasets of the brain. BIL contains partial- and whole-brain images of brains of mice, rats, and other mammals and model organisms and provides computing capability local to the data for pre-submission processing and post-submission exploration. Users can perform targeted experiments including studying the connectivity between cells, spatial transcriptomics, and historical collections. BIL provides a means to standardize, integrate, harmonize, and curate connectome data.

Dr. Ropelewski argued that the infrastructure needed for furthering connectomics research includes networking, storage, and computing. BIL contains three petabytes of public and private data, with the current highest-resolution datasets of mouse connectomes developed with fMOST technology, each consisting of 10 terabytes. BIL expects several connectome datasets each of 1 petabyte, and hopes for a human connectome expected to be 1 exabyte. That quantity of data requires significant technological infrastructure that NIH hopes to develop. Improved networking capabilities, including bandwidth, routing, and last-mile issues, are needed at many institutions. As may be expected with the quantity of data described, storage presents a challenge, whether using a centralized or distributed model and regardless of physical storage systems, although a tiered system may improve functionality. Dr. Ropelewski also mentioned a data seasons model, with data being posted and retracted on a rolling basis. Computing issues that require improvement include memory, GPU, batch and interactive access, virtual machines, and custom web interfaces. The scalability and portability of existing tools need to be further assessed.

Rubin Observatory—Astronomy Big Data

William O'Mullane, PhD

Dr. O'Mullane, an astrophysicist, presented on modeling using big data and collaboration across research groups. The Rubin Observatory collects massive amounts of data, especially with regard to its project surveying the southern sky.

The Observatory provides a Science Platform with three specific objects: the portal, which provides a visualization with which to observe the sky; the notebook, in which most analyses occur; and the application programming interface (API), which provides a set of common tools and catalogued data. Many researchers work directly within the Jupyter notebooks to maintain access to incredible amounts of data (10 petabytes) without needing to store the data on their own machines. The data are kept in Qserv, a horizontally scalable, shared-nothing parallel processing SQL database system. The notebooks also include Science Pipelines established by the Observatory, which contain all algorithms that the observatory staff predict to be useful for processing and consist of more than 1 million lines of code.

The Observatory also generates data about itself and its environment, such as temperatures surrounding the camera, which are stored in an Engineering Facility Database and can be cross-referenced with data in Qserv. Cross-referencing the environmental data with astronomical data is comparable to cross-referencing functional data with connectomic anatomical information.

Dr. O'Mullane envisions all big data projects involving central repositories with researchers downloading only small portions of data as needed. In addition, he envisions full transparency and open-source documentation to be easily accessible to any researcher needing deeper understanding of the methodology and programming involved. In support of that vision, all programs described are available on github.

Session 3 Q & A

Connectomics Analysis Tools

Dr. Gray-Roncal further discussed connectomics analysis tools provided by the Applied Physics Laboratory, particularly the dotMotif tool developed to query connectome graphs for motifs. dotMotif looks for sub-graph isomorphisms in larger graphs; the base version of the tool looks for deterministic answers, such as the features of particular graph pattern. Dr. Gray-Roncal and colleagues are also working on approaches to find signals in graphs with noise. Looking for sub-graphs in big graphs is challenging and requires significant computational power. Although the Laboratory has enough computational power to comprehensively search for patterns for up to seven neurons and their interconnections, analysis of larger datasets may be simplified if the graph structure, cell types, and directivity are considered.

Collaboration and Data Sharing

One of the most challenging aspects of a large connectomics project is that the massive datasets require widespread collaboration and data sharing to process and analyze the data. Moreover, because EM data have a limited lifespan and will eventually be superseded by better and higher-quality data, use of the data must be maximized as quickly as possible. Dr. Cardona discussed strategies and social structures that he has developed to engage the research community and facilitate productive data sharing and collaboration. Collaborative neuroscience should be framed as a win-win strategy in which every contribution is respected and

augmented by the work of others. This strategy builds a community of researchers around a dataset that generates the momentum needed to tackle large analytical problems.

Standardization of Data Sharing and Application Program Interfaces

Several speakers addressed the need for consistent and standardized approaches to sharing large connectomic datasets. Dr. Gray-Roncal is a proponent of using APIs to access the information and suggested beginning with a simple and straightforward API that enables access to imaging data, annotation data, and some graphs. Access to resources that enable researchers to reproduce published data (e.g., pre-written code or Jupyter notebooks) facilitates data sharing.

Dr. Cardona commented that the team that develops the software to host and share the data should write the de facto API. Formalizing the process of accessing the data and versioning APIs would provide stability going forward, but it requires commitment from funding agencies and from the data hosting team, with dedicated personnel to maintain the data repository and API. Dr. Vogelstein agreed that hosting the datasets requires a tremendous commitment of resources and time, which is difficult for a single laboratory to maintain over an extended period of time. Dr. Ropelewski noted that many different standards are periodically updated, which presents challenges for maintaining a metadata or API standards. That said, standardization across datasets, archives, and modalities makes it easier for the end user to access and use the data in new and novel ways.

Dr. O'Mullane explained that the International Observatory Alliance agreed on a set of API standards, which allows a set of common tools to access hundreds of different databases. There is also a series of standards for images, accessing images, and performing simple analyses, which took decades to implement and remains a challenge for the field of astronomy. He encouraged the connectomics field to develop some APIs and standard formats before huge datasets come available.

Discussion Panel

Panel Discussion Topic: Analysis of Connectomes

Panel Chairs: *Mala Murthy, PhD and Joshua T. Vogelstein, PhD*
Discussants: *Sebastian Seung, PhD, Carey Priebe, PhD, Kim Stachenfeld, PhD, Alex Szalay, PhD, Rachel Wilson, PhD, and Terry Sejnowski, PhD*

During this portion of the agenda, discussants presented their views on the topic, commented on the presentations by the earlier speakers and other discussants, and considered the questions posed in the online Q&A.

Dr. Seung remarked on two challenges that he proposed in his book *Connectome* 10 years ago: the mouse connectome and the human projectome. The progress of the past decade has rendered the challenge of the mouse connectome a seemingly imminent reality. Advances in computational abilities have made the connectome a more feasible goal. In addition, Dr. Seung commented on the now more widespread acceptance of the connectome as a significant

contribution to neuroscience. Dr. Seung posits that the function of a neuron is defined chiefly by its connections with other neurons. Although the connectome is often compared to the genome, there is increasing evidence that the relationship between structure and function may be more powerful for the connectome than for the genome; studies of interaural time difference in the barn owl, motion detection in insects, and the organization of the mammalian retina have all lent support to the notion that the function of a neuron is very closely related to its connectivity. Dr. Seung therefore believes that the coming years will see an established acceptance that the connectome is crucial for understanding the function of the nervous system.

Dr. Priebe described why the application of inferential statistics is exceptionally challenging for structural connectomes. By way of example, Dr. Priebe presented a graphical representation of the right hemisphere of the *Drosophila* larva mushroom body. This multigraph connectome is directed, weighted by the number of synapses between any pair of neurons, and has multiple edge colors (for axons, dendrites, etc.). The complexity of this connectome taxes the ability to derive valid, non-trivial, p-values to test properly defined hypotheses. Moreover, this one example does not account for all of the inherent properties of a structural connectome (e.g., synapse location). To account for all of these properties with sound statistical inference is an extraordinary task.

Dr. Stachenfeld presented examples of how deep learning over graphs and building graph neural networks may be leveraged for connectomics. Advances in deep learning models for classifying graphs, nodes, and edges as well as for graph representation learning may be useful tools for analyzing connectomic data. Open source tools such as PyTorch Geometrics, Deep Graph Library, Jraph, and the Graph Nets Library can be leveraged by neuroscientists interested in using neural networks to specify links between graph structures and the computations that perform over them. Datasets are also available to serve as benchmarks for training graph neural networks, including Open Graph Benchmark and Network Repository. Dr. Stachenfeld remarked that the machine learning community is enthusiastic to build graph network models and that neuroscientists would likely find eager collaborators in this space to approach questions about connectomic data; much of the data already on these networks are biological data. Neural networks could be useful for predicting dynamics conditional on a connectome or for classifying connectomes.

Dr. Szalay commented on the scale of research projects as it relates to generating large datasets and computational needs. Historically, scientific research was bimodal: at one end of the spectrum were small projects led by one principal investigator, and at the other end were large, billion-dollar projects. More recently, infrastructure for mid-scale research projects has emerged. These mid-scale projects are often characterized by young researchers developing unique instruments, leveraging cutting-edge technology, and taking bold risks. Dr. Szalay remarked that this scale is the “sweet spot” for science and that breakthroughs will be achieved by creative ideas and unique datasets (e.g., ImageNet). However, mid-scale projects are particularly susceptible to budget overruns, which often affect the computational infrastructure of the project. As a result, a new opportunistic computing model is emerging in which data are

captured at the instrument and analyzed wherever free CPU time is available (e.g., campus HPC clusters, Google cloud, Amazon Web Services). Thus, the location where computations are performed is not important for these projects as long as the data can be moved fast; the Open Storage Network is a distributed data storage service that facilitates the storage and rapid transfer of large datasets. Dr. Szalay noted that this kind of cyberinfrastructure will also be needed to support the longevity of datasets that require many years to generate because the datasets will not be superseded quickly.

Dr. Szalay also commented on how data formats will change as computing models continue to evolve for large datasets. Currently, researchers download whole or partial datasets to perform computations. Soon, these data will likely be streamed from the cloud, which will change the landscape of computational analysis.

Dr. Seung suggested that a possible model for the mouse connectome effort would be for academic centers to build on-premise cloud storage, such that data are distributed at various sites. In this model, the computing is hybrid such that university computing resources, public clouds, and national lab supercomputers would come together to reconstruct the mouse connectome.

Dr. Wilson remarked that connectomics has challenged the concept of causality in neural systems. Moreover, connectomes challenge neuroscientists to consider the brain not as a machine for processing information but rather as a mechanism to control the output of the body in real time. Although scientific projects are often designed to determine the basis for causality (e.g., the cell type or brain regions that causes a behavior), connectomes are replete with divergence, convergence, and feedback loops that blur the line between cause and effect. Dr. Wilson suggested that instead of striving to identify components of neural systems as necessary and sufficient for a behavior or cognitive function, neuroscientists should study the functionality of engineered systems in which divergence, convergence, and feedback loops are ubiquitous. For example, engineered systems for aircraft control have a similar number of wires as flight control circuits in *Drosophila*, and the systems may be compared to understand how essential connection features relate to function. However, it is important to recognize that different constraints exist for engineered systems and living organisms, and researchers must consider the fundamental similarities and differences between the control of machines and the control of living organisms that contribute to similarities and differences in the wiring diagrams that implement these control algorithms. In addition, a correspondence between neural network models and actually implemented algorithms must be established, because the same control algorithm can be implemented by a variety of networks.

Dr. Seung added that another loop in connectomics is the relationship between activity and connectomes: the connectome may cause certain activity patterns, and activity shapes the connectome through learning. Being able to measure the connectome may help elucidate these types of interactions. On the subject of engineered versus biological control systems, Dr. Sejnowski stated that a key difference between these types of systems is that engineered

systems have a central point of control whereas control is very distributed within biological systems.

Dr. Sejnowski suggested that new insights will be gleaned even from simple biological observations. Citing an example from his lab, the observation of paired synapses between the same axon and same dendrite in CA1 revealed a tight correlation between spine head size within a pair as well as a large range of spine head sizes across these synaptic pairs. These observations have implications for the precision of synaptic plasticity algorithms. In another example, the physical proximity of mitochondria to synapses could provide insight into the activity of those synapses, similar to how the observation of high mitochondrial density in cells with high levels of activity elucidated the organizational structure of primary visual cortex. Dr. Sejnowski also commented on how a fully reconstructed connectome will enable new lines of scientific inquiry. Just as the genome has enabled researchers to answer questions about the entire genome rather than one gene, the connectome will similarly expand the potential scope of neuroscience research. Dr. Sejnowski was also optimistic that as reconstruction costs decrease and technologies advance, institutions will be able to establish connectomic cores similar to currently established genomic cores for high-throughput reconstructions.

Dr. Seung agreed that connectomic cores will likely be realized in the future and added that an important aspect of making connectomics widely available is to focus on commercialization and cost reduction.

Follow-Up Discussion

Dr. Priebe noted that the suggestion to study structural connectomes can often be contentious. Dr. Stachenfeld acknowledged that the task of moving from a static graph representation of a connectome to a computation that is performed over that graph is a compelling challenge, and that analogies to controlled systems optimized for a specific goal may be a good approach to conceptualizing this challenge. Dr. Wilson added that the study of a network with a relatively small number of neurons that accomplish a very clear task could inform technical pipelines as well as the novel concepts needed to analyze more complicated brains.

Dr. Wilson remarked that if causality is a less clearly defined concept than previously appreciated by the field, then it may be time to rethink the perturbation experiments that are so common in neuroscience. It is very difficult to observe a meaningful, interpretable phenotype in a perturbation experiment that silences one cell type, for example, which is not surprising given the extensive divergence, convergence, and feedback loops that are fundamental components of the connectome. Instead, Dr. Wilson suggested that more integrative approaches to studying whole-brain connectomes and the dynamics of feedback loops should be incentivized. Dr. Marder agreed that although perturbation experiments are still feasible, they require researchers to move away from linear frameworks and recognize that neural networks are highly interconnected and that multiple parallel pathways exist that can cause a manipulation to fail. Dr. Marder further posited that parallel pathways are the greatest challenge to understanding how circuits work and that the connectome will help to reveal which pathways must be accounted for when conducting experiments.

Dr. Sejnowski agreed that this concept of causality is a challenge for neuroscience and suggested that it may be worth starting with studies of model dynamics before moving on to something as complex as the brain. Dr. Zlatic added that a way to modify traditional perturbation experiments could be to use models to guide manipulation of particular circuit motifs (e.g., feedback loops) to observe how loss of that motif impacts computational power, as opposed to manipulating individual neurons and observing behaviors. Dr. Stachenfeld commented that although it would be difficult to extract causal motifs using neural networks, these tools could be used to draw a link between an underlying graph structure and an objective function optimized by that structure, which in turn could be used to generate a hypothesis for these kinds of experiments. In addition, Dr. Priebe acknowledged that the elucidation of causality in neural networks would be of great value, but noted that causality is a formal inferential result that demands proper statistical analysis.

Dr. Seung suggested that the issue of causality raises a question about the extent to which neuroscience should focus more on observation than manipulation of neural networks, similar to how astronomy is a primarily observational science. Dr. Wilson added that the theoretical underpinnings of neuroscience are relatively weak compared to other sciences, including astrophysics. Dr. Szalay noted that astrophysical and astronomical theories were developed over thousands of years and informed by observational data out of necessity, adding that connectomics may be fundamentally more difficult than astronomy, for example, because it is difficult to study any one part of a connectome in isolation the way an astronomer might study one part of the universe.

Dr. Vogelstein asked the panelists which topics are the most important to teach the next generation of neuroscientists to advance the understanding of how the brain works. Drs. Priebe, Szalay, and Cardona all asserted that students need more training in statistics. Drs. Seung and Stachenfeld both suggested that more training in theoretical concepts (e.g., neural network theory, classic graph theory) would be valuable to the next generation of neuroscientists. Finally, Dr. Marder emphasized the importance of students understanding that scientists must be willing to learn continuously throughout their careers.

Town Hall

Introduction

The Town Hall marked the conclusion of the Brain Connectivity Workshop Series hosted jointly by NIH and DOE. During the series' previous five workshops, speakers addressed the state of the art of connectomics as well as existing opportunities, techniques, and challenges in connectome development. Each workshop focused on a different aspect of connectomes: biological significance, sample preparation techniques, imaging modalities, data pipelines, and access to resulting products. The Town Hall summarized and addressed these topics further.

Keynote Address

Caroline Montojo, PhD

Dr. Montojo, President of the Dana Foundation, discussed the findings of the workshop series in the larger context of the potential impact of mapping a connectome at scale on the field of neuroscience and beyond. She focused on four key areas of significance drawn from the workshop series at the intersection of neuroscience and society. First, developing a whole-brain connectome would provide scientists with a deeper understanding of the brain. Because animal and human brains are organized similarly, studies in mouse brains can provide templates for human brain research. Such a resource would transform the understanding of core behaviors and physiology and would provide a detailed map for information routing in the brain. Second, a whole-brain connectome of a mammalian brain would provide a means to better understand and develop treatments for abnormal circuit functions that underlie many neurological conditions, such as Alzheimer's and Parkinson's diseases and autism spectrum disorder. The connectome could accelerate development of the next generation of circuit-based therapeutics. Third, beyond the value for neuroscience alone, the technical challenges of mapping the connectome will catalyze new innovations in data science, including in artificial intelligence and machine learning. This project will require close and ongoing collaboration among scientists from many fields, including life and physical sciences, mathematics, engineering, and data sciences. A discussion panel in the fifth workshop focused on how our understanding of connectomes might explicate how spontaneous activity arises and invalidate the more simplified "stimulus-response" models of activity. Given the complexity of the brain, gaining new knowledge from the connectome will likely require derivation of new theoretical frameworks. Conversely, the new knowledge about how the brain works will likely drive development of new mathematical concepts. Historically, this type of relationship has been observed between mathematics and theoretical physics. Similarly, a strong partnership between neuroscientists and mathematicians will be important to realizing the goals articulated for connectomics throughout these workshops. Fourth, to process and analyze massive datasets, many connectomics scientists will need to collaborate while practicing widespread data sharing and dissemination. Thus, the effort to achieve a whole mammalian brain connectome will both require and advance the democratization of research and improve scientific culture.

Existing connectomes have transformed research of *Drosophila* especially and neuroscience in general. Developing these connectomes has revealed new functional motifs, provided a bridge between theory and experiment to improve conceptual-level insight, emphasized the need for democratization, and enabled researchers not only to think on a brain-wide scale but also to approach experiments differently. However, connectome projects have uncovered several challenges that the neuroscientific community at large must address. The definition of a complete connectome remains undetermined; the appropriate balance between scale and spatial resolution is unclear; techniques and methods to best manage and analyze the quantity of data involved need to be developed; the difference between the value gained and cost expended is still debated; and the time required to develop these connectomes is enormous.

Even with this complexity and challenge, major value can be derived from large scientific projects. For example, large projects in genomics and astronomy have been influential and significant for science, not because they answer a single experimental question, but because they brought together researchers from different disciplines and locations to work toward shared goals. As such these large projects have enabled investigation of new questions from shared data sources. Dr. Montojo believes that the BRAIN Initiative's model of bringing together the public and private sectors to pursue this common goal with determination and creativity makes this the ideal time to embark on a connectome project.

Panel 1: The State of the Art and the Challenges and Opportunities Ahead

Introduction: *Kristen Harris, PhD, and Peter Littlewood, PhD, Co-Chairs*
Panelists: *Larry Abbott, PhD, Eve Marder, PhD, Davi Bock, PhD, Hongkui Zeng, PhD, Mark Ellisman, PhD, Galya Orr, PhD, Nicola Ferrier, PhD, Mala Murthy, PhD, Viren Jain, PhD, Joshua Vogelstein, PhD*

During Panel 1, the Co-Leads of each workshop presented the key takeaways from their workshop and the challenges and opportunities they foresee in future connectomics work. The series Co-Chairs served as moderators for this panel and introduced the discussion.

Dr. Harris revisited the concept of CAP—Complete, Accurate, and Permanent—in light of the state of the art, challenges, and opportunities that emerged during the workshop series. She suggested adjusting the acronym to emphasize the process of connectome development—Competitive-Collaborative-Coordinated, Accurate, and Public. Dr. Harris proposed that accuracy should be more precisely defined to a given scale, and that connectomics researchers could determine whether gap junctions and neuromodulators must be included to consider a connectome accurate. She replaced "Permanent" with "Public" to emphasize the importance of immediate data sharing and to further underscore the inclusion of Collaborative and Coordinated under "C" in the acronym. Dr. Harris asserted that developing accurate and useful maps of neural circuitry will produce an engineering legacy with new tools that will democratize the ability to obtain accurate and specific maps. Priorities for future work include improving sample preparation methods, developing metrics for determining when goals are achieved,

deciding upon and integrating strategies while retaining the flexibility needed to incorporate new tools as they become available, and ultimately expanding the audience and user groups to spur industry, especially for high-resolution EM.

Dr. Littlewood asserted that connectomics is a large-scale engineering project. He noted that 1 cm^3 of brain tissue imaged at 10 nm resolution will generate approximately 1 exabyte of data, making a whole mouse brain approximately 0.1 percent of global storage in 2021. He emphasized that technologies for computing and storage have exponentially improved in recent years, making achievements that appeared impossible in the recent past now feasible. The general future of computing may be driven by coming connectomics work. The ambition for the connectome project should not be limited by a perception of what can be done now. The field of connectomics can exploit and adapt technologies driven and paid for by other industries (e.g., computing, X-ray and optical imaging, labeling). A key part of the project is determining which technological advancements must be solved by the connectomics field; Dr. Littlewood claimed many of those would be in EM. Dr. Littlewood concluded by asking the workshop co-leads to describe what they believe can be achieved with normal effort at the rate the field is currently progressing, what will require special development to drive progress, and what are the areas of particular concern.

Workshop 1: Significance of Mapping Complete Neural Circuits

Larry Abbott, PhD, and Eve Marder, PhD, Co-Leads

Dr. Abbott stated that the connectomics revolution is already under way. Workshop 1 presentations described several systems that are already benefiting from connectome-inspired insights. Perhaps most spectacular is fly connectomics, which provides many lessons for the field, three of which he described. First, with a large quantity of data researchers can solve problems that appeared to be insurmountable. Researchers assumed that a connectome would establish neuronal connectivity without providing information about transmitter type and, thus, the sign or impact of those connections. However, connectomic data led to machine learning algorithms that can incorporate neurotransmitter phenotype. Second, many scientists believed that models were not falsifiable—a belief that connectomes have reversed by providing the anatomical structures modelers had attempted to derive. Third, the connectome has become a basic and essential tool in the design and interpretation of experiments in species with complete datasets (e.g., *C. elegans*). Development and experimental use of connectomes present conceptual—in addition to technological scale— challenges; the field must learn not only how to obtain new kinds of circuit maps but also how to optimally use them.

Dr. Marder commented that modern progress in connectomics and circuit mapping should drive the field to revisit its choice of organisms. Attempts to “crack circuits” and develop wiring diagrams during 1960-1980 were painstaking yet useful because the focus was on organisms for which much was known about the circuits underlying specific behaviors. At present, relatively few scientists still work in non-genetic model organisms. Dr. Marder suggested that neuroscientists should reconsider the value of studying so-called expert organisms—that is, those particularly good at completing certain tasks—to learn more about those functions. Dr.

Marder also suggested that because the health of the planet depends on human interactions with nature and other species, obtaining connectomes from diverse animals could improve humanity's ability to combat ongoing ecological changes. She advocated for framing connectomics as a new kind of comparative neuroscience that will improve understanding of the human species.

Workshop 2: Sample Preparation in Mammalian Whole-Brain Connectomics

Davi Bock, PhD, and Hongkui Zeng, PhD, Co-Leads

Dr. Bock discussed the challenges to sample preparation for EM images of the whole mouse brain. EM sample preparation has not yet been effectively employed on the scale of the whole mouse brain. He acknowledged the debate in Workshop 2 regarding how to metal stain, embed, comprehensively subdivide, and handle tissue samples. Workshop participants were optimistic about current methodological developments and ideas, including successful staining and embedding of an early postnatal brain; ultrasmooth vibratomy for lossless subdivision prior to embedding; hot-knife sections followed by gas-cluster ion beam milling and multi-beam SEM; and grid tape alongside TEM tomography.

Dr. Zeng described complementary imaging technologies for whole mouse and larger brains, which can provide greater functional information about the brain than EM datasets and, when used alongside whole-brain connectomes, can further understanding of sex differences, individual variability, and more fundamental neuroscientific questions. These technologies include alternative imaging technologies such as X-ray tomography; fluorescence LM coupled with tissue clearing; EM-dense and CLEM chemical and genetic labels; immuno-EM; and axonal projectomes derived from optical, X-ray, or lower-resolution EM. However, techniques for integrating these technologies with EM data have yet to be developed. Using these technologies to answer fundamental neuroscientific questions demands careful experimental planning, collaboration, and coordination across fields. In summary, the interrelated goals of Workshop 2 and Workshop 3 were to address the technological feasibility of a mammalian connectome.

Workshop 3: Experimental Modalities for Whole-Brain Connectivity Mapping

Mark Ellisman, PhD, and Galya Orr, PhD, Co-Leads

Dr. Ellisman addressed imaging and the infrastructure and team science needed to achieve the stretch goals articulated by community leaders. Workshop 3 speakers discussed the resolution that is required to attain meaningful wiring models that represent and differentiate classes of synapses, and they reached the consensus that a resolution great enough to track and subdivide synapses is required. They also asked which imaging technologies could be driven to scale and most easily disseminated; although not reaching a consensus, they agreed that multi-beam SEM and TEM pose different risks and bottlenecks and that X-ray microscopy may be scalable but lacks requisite resolution. Workshop 3 speakers also discussed the value of single large-scale efforts compared to efforts that would democratize data collection.

Dr. Orr commented on the fourth and final issue addressed in Workshop 3—that is, how to incorporate the time domain into static brain maps. One option is time-lapse EM, which captures dynamic changes in brain structure but cannot be easily employed to study changes over short timescales. Correlating cellular resolution in vivo calcium imaging with EM connectomics could enable imaging and reconstruction of cells that are active during specific cognitive tasks, thereby associating activity with structure. Fluorescence microscopy, spatially resolved RNA-Seq, and synaptic protein labelling are additional techniques to measure activity across time, but have yet to be applied at large scales.

Workshop 4: Connectome Generation and Data Pipelines

Nicola Ferrier, PhD, and Viren Jain, PhD, Co-Leads

Dr. Ferrier discussed computational techniques and the human-computer interactions that derive useful information from raw EM data. To develop a coherent 3D volume to address a mouse whole-brain connectome, scientists must develop approaches to scale up montage stitching and serial section alignment for larger datasets, automated and semi-automated tools to detect volume assembly errors, and multi-modal approaches to integrate and further develop sets. Neuron reconstruction and segmentation has improved 100-fold since 2015, but another 10-fold improvement is required for a mouse whole-brain connectome. Researchers must establish standards for workflow and the exchange of data, as well as identify which types of imaging errors have the greatest and least impacts. Improvements in neural network inference efficiency currently meet or exceed a Moore's Law scaling curve. Human proofreading, whether by experts or large-scale citizen science, remains critical. In summary, the workshop speakers agreed that many existing tools can be scaled, but the accuracy of each step must improve to make the overall human effort feasible.

Workshop 5: Optimizing the Use of Connectomic Data to Drive Data Science and Scientific Discovery

Mala Murthy, PhD, and Joshua Vogelstein, PhD, Co-Leads

Dr. Murthy focused her comments on optimizing the use of connectomic data to drive scientific discovery. Several speakers highlighted discoveries that were made using existing connectomes (e.g., *Drosophila* auditory system or mushroom bodies) and how connectomics has spurred development of artificial intelligence and machine learning tools for making meaningful comparisons within and across datasets. In addition, Dr. Murthy noted that connectomics has already enabled the association of neuronal function with wiring diagrams in some species. Speakers also described tools that can be deployed for larger datasets but noted that those tools were developed to handle highly stereotyped systems, which mammal brains are not. Connectomics research cannot rely on single labs, but rather on democratization and collaboration. Lastly, Dr. Murthy highlighted Dr. Cardona's comment that connectomic data must be shared early and openly.

Dr. Vogelstein discussed the challenges identified in Workshop 5. Importantly, developing connectomes is not itself a goal, but a tool with which scientists can study the brain. Exactly

how researchers will use the connectome is yet to be determined, especially because these datasets will vary across and within species. High-resolution datasets will be important but must remain manageable, and for many models and questions, the connectome will have to be augmented with additional data (e.g., morphology) for meaningful interpretation. Finally, Dr. Vogelstein noted that complete connectomes will reveal questions no scientist has yet asked, and that identifying and then pursuing these questions will be a challenge unto itself.

Discussion

The Future of the Field

Dr. Bock stated that he would not predict when a brain tissue preparation suitable for development of a whole mammalian brain connectome will become available; optimistically, sample preparation for such a connectome could be feasible within 6 months to 1 year, and pessimistically, never. Some advancements could propel this effort forward (e.g., ultra-smooth vibratomy). Dr. Bock suggested that inclusion of experts outside of neuroscience (e.g., mechanical and hardware engineers, chemists) could accelerate the development of sample preparation and technical strategies. Dr. Zeng added that serious investment and major effort is needed to allow not just a few but many people to try different ideas to determine the most effective sample preparation strategies. Dr. Littlewood rhetorically asked how scientists should manage the risks of such an operation and endorsed Dr. Bock's suggestion to involve experts from outside of the neuroscience field.

Dr. Littlewood asked about the lessons learned from obtaining the fly brain connectome and the field's next steps. Dr. Abbott replied that, although the existing hemibrain dataset is very useful, many advances remain to be made, including identifying transmitter types across the entire dataset and collecting more connectomes to study plasticity. Dr. Murthy added that researchers can already compare the hemibrain and the FAFB datasets, which are revealing variability even within the stereotypy of *Drosophila* brains.

Involving Undergraduate Students and Faculty

Dr. Vogelstein commented that his team has been passionate about collaborating with undergraduate neuroscience students and faculty in the democratization process since the first big connectomics dataset was developed. While providing the data for inspiration is easy, building tools that enable students to interact with the data and contribute to research is more challenging. Dr. Bock added that he is piloting dissemination of the annotation software CATMAID to undergraduate courses, but technology is not the only limiting factor: a full curriculum is needed to train the next generation in the use of these technologies. Dr. Harris agreed that a full curriculum is needed and noted that she teaches undergraduate students to identify and reconstruct imaging data and provides students the opportunity to analyze actual, as-yet undescribed EM data. Workshops and summer courses are beginning to incorporate similar training. These examples highlight the importance of both the community and curriculum aspects of involving undergraduate students and faculty in the connectomics effort.

Democratization

Dr. Ferrier mentioned that the DOE/NIH collaboration is key to democratization because DOE's National Laboratories are structured to enable large computing projects and can build infrastructure so that any university can contribute to the project. Dr. Marder highlighted that another primary hurdle for democratization of connectome development is that many scientists lack the resources to generate or analyze connectomic data. Dr. Harris added that another requirement will be sharing data early, rather than waiting for publication. Dr. Ellisman noted that each speaker will likely define democratization differently and that he defines it as broadening the engagement and training of anyone who can be involved. Because funding is generally provided for the next incremental step, many small experiments will be conducted along the way that can be used to educate a new generation of scientists in connectomics experimentation, perhaps through specialized courses.

Valuable Datasets

Dr. Littlewood asked the speakers to identify specific comparative connectome projects or species that they would consider most valuable. Dr. Marder expressed particular interest in interactions among species, especially those that change the environment and agriculture, such as bees and ants, and the impact of ants' social structure on interindividual variability. Dr. Abbott also sees value in determining the differences in navigational circuitry between bees and ants, which are navigational geniuses, and *Drosophila*, which are not. Dr. Vogelstein added that understanding how microscopic bee and ant brains can nonetheless achieve cognitive complexity could improve neural network and machine learning algorithms. Dr. Marder stated that working with lab-bred genetic model organisms is problematic because they lack the plastic changes that occur from lived behavior, and that scientists should compare connectomes between laboratory mice and wild mice.

Dr. Bock commented that there is embodied intelligence beyond the CNS in the periphery and that he would be most interested in sample preparations and imaging efforts that connect the CNS with the sensors and effectors in the periphery. The nature of this embodiment might further explain the differences in navigational capability among insects.

Rate-Limiting Steps

Dr. Harris asked about the rate-limiting steps for connectome development and whether the field can rely on Moore's Law. Dr. Littlewood stated that as technology advances and circuits become smaller than a synapse, some problems will arise in computational modeling. However, he compared this issue to engineering speed becoming faster than clock speed, which was overcome by increasing parallelism in circuits. The rate of data development is overtaking the rate of analysis, although newly developed algorithms may close that gap.

Dr. Abbott commented that a rate-limiting step is determining and associating function with a circuit. Dr. Marder agreed, stating that a focus on well-described regions of the mouse brain will therefore be an important starting point. She suggested that another rate-limiting step exists in the current understanding of electrical synapses and their parallel pathways. Dr. Zeng agreed that function is needed to understand circuits but argued that production of a mouse

whole-brain connectome alone would still drive incredible progress in circuit-based neuroscience. Dr. Bock added that defining cell types by connectivity will likely lead to the identification of new cell types and stressed both the importance of lossless subdivision and the limitations of knowledge derived from a single connectome. Dr. Orr stated that a rate-limiting step is correlating a wiring map with activity. Dr. Murthy commented that while neuroscientists have largely focused on obtaining connectomic data, interpretation remains a rate-limiting step.

Dr. Vogelstein mentioned that other biological sciences have pursued work with exponentially more data than previously available—for example, the Human Genome Project—and expressed concern that neuroscientists will ask the wrong questions for the data collected. Dr. Ellisman commented that determining the best technologies and instruments will enable researchers to define the bottlenecks and cautioned that connectomics strive to expand knowledge about the complexity of neuroscience, at least initially, until enough is known to simplify the models. Dr. Harris suggested that balancing resolution against the time necessary to collect data is a major rate-limiting step, and that developing approaches to image a brain in multiple modalities will be critical. Dr. Harris also suggested that to determine the minimum resolution needed to achieve the project's scientific goals, the field may need to test the resolution needed to capture gap junctions, interdigitating tiny glial processes, or other objects that traditionally have been difficult to identify at lower resolutions.

Dr. Viren Jain agreed that several different answers exist, but competition for the efforts of computational scientists will be a rate-limiting step. Similarly, Dr. Ferrier stated that human effort is the greatest challenge, and that tools to automate these processes or increase the efficiency of human analysis must be developed.

Developmental Connectomics

Dr. Littlewood commented that understanding of the developmental changes in the connectome and how they vary across species would accelerate developmental neuroscience. Dr. Harris noted that the question of developmental connectomics relates to a broader debate of whether to perfect technologies before pursuing multiple connectomes, or to start with multiple connectomes and improve technologies in the process, because studying connectivity across development necessitates obtaining multiple connectomes.

Additional Information from EM Datasets

Dr. Harris concluded the panel discussion by asking the workshop Co-Leads about their plans to extract and analyze the intracellular structures of individual neurons from EM datasets to determine cell types and function, and the aspects they consider most important. Dr. Bock stated that, orthogonal to mitochondria and cytoskeletons, postsynaptic density area serves as the best proxy for chemical synaptic weight. Dr. Ellisman stated that mitochondria are the easiest subcellular features to segment in modest resolution EM datasets and provide appropriate proxies for energy expenditure.

Panel 2: The Significance of Mapping Mammalian Brain Circuitry at Scale

Introduction: *Cori Bargmann, PhD, and Terry Sejnowski, PhD, Moderators*
Panelists: *Emery Brown, MD, PhD, Elizabeth Buffalo, PhD, Gwyneth Card, PhD, Ed Lein, PhD, and Stephen Streiffer, PhD*

Dr. Bargmann opened Panel 2 by stating that the concept of science-at-scale—the convening of many researchers from across the disciplinary and institutional spectrum to tackle the most complex scientific challenges in a highly coordinated manner—provides a foundation for further work, a concept exemplified by her experience working with *C. elegans*. She suggested that the next step in connectomics should be to learn from other large biological projects, such as the Human Genome Project. The technology used for the first complete human genome was immature, but that work nonetheless accelerated and drove technological development. Importantly, the Human Genome Project did not itself lead to important discoveries but provided a foundation for other discoveries. Similarly, Dr. Bargmann noted that the Cell Census, a science-at-scale project supported by the Brain Initiative since 2017, has reflected these lessons. Improvements in data generation and analytics will arise from connectomic research, and even if every synapse in the human brain cannot be mapped, the attempt will lead to incredible innovation. Multi-modal integration of morphological, physiological, temporal, and spatial data will be a critical future step in connectomics.

Dr. Sejnowski agreed with Dr. Bargmann and commented on the impact of connectomics on psychiatry, which began with talk therapy and then pursued biological approaches. These extremes are both useful; however, computational psychiatry and determining circuit disruption may provide a critical middle path. Using simple circuits with as few as three nodes, some researchers have already pursued this question; by modifying weights or eliminating nodes, modelers could replicate many oscillations observed in different regions of the brain. Future work with larger connectomic datasets will provide exponentially more information.

Dr. Brown stated that connectomes provide an important opportunity to link anatomy with physiology. As anatomical structures are defined at different length scales, the separate datasets must be integrated. For example, ketamine given to a person suffering major depressive disorder produces state changes at multiple levels, and understanding the specific mechanisms involved requires team science to address molecular and systems neuroscience, clinical perspectives, modeling, and data analysis. Using connectomic data alongside integrative modeling work may provide deeper answers about the brain's functions overall.

Dr. Buffalo provided an example from the hippocampus to complement Dr. Brown's comments on merging physiology and anatomy by discussing traveling sharp-wave ripples. Dr. Buffalo's team studies neural mechanisms that support learning and memory using neurophysiological studies in monkeys. Simultaneous chronic recordings from hundreds of individual neurons provide information about not only single unit activity, but also hippocampal ripples, a type of brain wave. When comparing these structures between rodents and primates, Dr. Buffalo's

team discovered significant differences in connectivity between the two species, despite the similar morphology of individual neurons. She hopes that this sort of comparative neuroanatomy could inform comparative neurophysiology. Current network models that describe and map mammalian brain circuitry at scale could enable better prediction and understanding of connectivity patterns.

Dr. Card has used fly connectome data extensively in the previous 5 years and focused her comments on findings from fly connectome research. The ring attractor model, which had been theorized to represent head direction, has now been validated in *Drosophila* with connectomic data. She also discussed parallel circuitry in the *Drosophila* structure for organizing associative memories. This organization resembles Purkinje cells in the cerebellum. She argued that brains have organization, which can be determined topographically or from visual features. Understanding how those topographic maps are interpreted by downstream structures in the brain is the future of connectomics and neuroscience.

Dr. Lein expanded on Dr. Sejnowski's comments about human projectomes and connectomes, stating that most if not all brain-related diseases involve circuit dysfunction. Current scaffolds are based on diffusion tensor imaging that lacks the resolution necessary to understand circuit-level disorders. This lack of knowledge has severe clinical consequences, as effective treatment demands greater understanding of mechanisms and neural organization. Dr. Lein argued that neuroscientists should not wait for development of a complete human connectome but should instead study new treatment strategies while mapping the human brain and other species in parallel. Some methods used for axonal connectomics can be combined with molecularly defined cell types, drawing on the Cell Census Network to better describe these disorders.

Dr. Streiffer agreed with many of the comments from other speakers but stressed that while learning from other scientific communities, neuroscience and connectomics research can also teach or accelerate development in other fields. This will be especially true with artificial intelligence and machine learning; for example, by aiding the development of plasticity in neuro-inspired computing and neural networks synthetic plasticity. Dr. Streiffer argued that researchers should not approach this project as big science, but rather as bench science at a large scale.

Discussion

Unknown Unknowns

Dr. Sejnowski asked the panelists to identify new questions that a mouse whole-brain connectome could help to answer. Dr. Brown stated that integrating information across scales may decipher behaviors, drug effects, and normal physiology. Dr. Sejnowski agreed and related Dr. Brown's answer to Dr. Murthy's description of techniques that enable simultaneous recordings from millions of neurons. Dr. Buffalo mentioned that one of the main oscillation patterns in the brain (theta) is prominent in rodents but has a very different pattern in monkeys. The variance in the patterns of connections that a connectome could provide might inform understanding of whether this difference is due to an intrinsic property of the cells or connectivity itself. Dr. Card stated that the fly connectome has enabled researchers to

determine the "unknown unknowns"—that is, to ask questions they did not know to ask and would not have included in models. Dr. Lein agreed and suggested that connectomes could be used to determine how highly localized or distributed specific functions are. Dr. Streiffer added that before becoming reductionist in modeling, neuroscientists must determine what is common and what differs across species, ages, and developmental states.

Specificity and Sensitivity

Dr. Bargmann asked the panelists about the degree of specificity and sensitivity required for a meaningful connectome and the acceptable level of error. Dr. Card stated that her takeaway from the fly connectome was that the structure of community effort is critical, especially for proofreading and manual tracing. Dr. Sejnowski compared efforts to build connectomes with the first sequencing of the human genome, which was a draft but a useful starting point. Dr. Brown commented that the answer to these metrics varies according to the specific question posed. Dr. Streiffer expanded on that comment, stating that such boundaries cannot be determined ahead of data collection. Drs. Lein and Buffalo agreed and added that the interindividual variability would determine these metrics, and that the first step will be to begin data collection. Both researchers also emphasized the necessity of choosing model animals based on the behaviors they exhibit rather than the techniques that exist for those animals.

Final Comments

Dr. Bargmann asked the panelists to discuss priorities for selecting strains and species, as well as the differences expected between wild and inbred lab animals. Dr. Card stated that mapped connectomes are a critical starting point for circuit investigation across species, but that the first animals selected for connectome mapping should prioritize those where a broad base of knowledge and genetic/physiological tools are already available. As connectomes become more affordable and easier to generate, it will be important to expand to other species that either have specific behaviors that make them compelling for study (i.e., the classic neuroethology perspective) or have a compelling evolutionary relationship to species with a mapped connectome. Regarding the new knowledge that will be generated from connectomes, Dr. Sejnowski concluded that a model that fits the existing data is not as informative as one that fails, because a failed model reveals something missing or wrong that can lead to learning something new—that is, a failed model will be more useful for scientific discovery.

Q & A Session

Resolution and Completeness

Dr. Bock commented on the earlier question regarding acceptable error rates, which depend on the biases in the detection system. Even high error rates can bring about new discoveries. Dr. Vogelstein added that the answer depends on the hypothesis being tested; researchers may want to develop connectomes for fewer brains at higher resolution rather than more brains at coarser resolutions. Dr. Sejnowski explained that high-resolution datasets can be subsampled at lower resolution, and that greater investment at the beginning would be more useful than obtaining more brains at lower resolution. Dr. Marder stated that the question itself is meaningless without knowledge about animal-to-animal variability and suggested that researchers should begin by mapping one region in 10 separate animals and assessing the

variability. If differences between subjects are vast, the greater accuracy will be necessary for meaningful results. Dr. Ellisman echoed these assertions and specifically noted the need to determine areas of variance and invariance between individual, because invariance may indicate regions where consistent structural organization is required for normal function. Dr. Card added that the likely high degree of variability indicates that obtaining the entire whole-brain connectome from a single animal will be more valuable than partial connectomes from multiple animals, because researchers may not be able to construct models from multiple animals.

Dr. Abbott stated that completeness is more important than accuracy and referenced a comment by Dr. Cardona that patterns emerged from data when the connectome neared completion, despite having been obscured to that point. Dr. Harris stated that acceptable error will depend in part on the degree of axonal branching for a given neuron; error early in the branching pattern is more detrimental than error later. She asserted that LM is valuable for long-range connections and that high resolution should be used to study local circuitry. Dr. Zeng expanded on this idea by explaining that EM should be used specifically for questions that cannot be addressed with LM, which is dense reconstruction. She stated that given the multiple parallel neural systems, each with multi-step brain-wide connections, completeness at the whole-brain scale to allow tracing of single axon fibers across long distances would be critical; whereas due to the large numbers of neurons a mouse brain has, absolute 100 percent accuracy is not needed as long as a large proportion of neurons can be traced. Dr. Bock mentioned that sample preservation and the ability to revisit original samples will be critical to managing concerns around accuracy and completeness.

Integration of Data

Dr. Bargmann conveyed a question from the audience about the types of data that the connectome does not provide (e.g., intrinsic properties of cells, kinetics and plasticity of synapses, or effects of neuromodulators on cells and synapses) that will be most important to integrate into connectome-based models. Dr. Bargmann's own answer was neuromodulators, with which Dr. Sejnowski agreed. Dr. Lein answered that functional domain modeling, as determined with fMRI, would be most important. Dr. Vogelstein similarly argued that associating behavior with localized neural activity will be particularly important.

Wrap-Up

Drs. Ngai and Kung provided concluding comments for the series of workshops. Dr. Ngai claimed that at the beginning of the series, developing a whole-brain connectome seemed like an enormous problem. However, speakers clearly described the specific challenges to connectomics and potential solutions. Neuroscientists will develop new theories, methods, and computational tools as they seek to achieve the goal of constructing a whole-brain connectome in mammals. This project poses a valuable opportunity to build new coalitions and collaborations and to reimagine diverse workforces for coming generations of scientists and engineers. It will not only revolutionize brain research but also develop a community of researchers across multiple fields and disciplines.

Dr. Kung added that the Town Hall prompted her to recall an initial meeting with NIH about the Brain Initiative in 2019. She stated that although significant advances remain necessary to achieve the Initiative's goals, DOE's unique ability to integrate scientific discoveries and technology can accelerate the field toward its goals. She stated that a continuous and transparent dialogue with the community across many scientific and engineering disciplines will be critical in the pursuit of a whole human connectome.

Appendix A: Participants

Series Co-Chairs

Kristen Harris, *University of Texas, Austin*
Peter Littlewood, *University of Chicago*

Workshop Co-Leads

Larry Abbott, *Columbia University*
Davi Bock, *University of Vermont*
Mark Ellisman, *University of California, San Diego*
Nicola Ferrier, *Argonne National Laboratory*
Viren Jain, *Google*
Eve Marder, *Brandeis University*
Mala Murthy, *Princeton University*
Galya Orr, *Pacific Northwest National Laboratory*
Joshua Vogelstein, *Johns Hopkins University*
Hongkui Zeng, *Allen Institute for Brain Science*

NIH-DOE Steering Committee

Ruben Alvarez, *National Institute of Mental Health*
Andrea Beckel-Mitchener, *National Institute of Mental Health*
Greg Farber, *National Institute of Mental Health*
Adam Kinney, *Office of Science, Department of Energy*
Harriet Kung, *Office of Science, Department of Energy*
Steven Lee, *Office of Science, Department of Energy*
John Ngai, *National Institute of Neurological Disorders and Stroke*
Michele Pearson, *National Institute of Neurological Disorders and Stroke*
Amanda Price, Eunice Kennedy Shriver *National Institute of Child Health and Human Development*
Ryan Richardson, *National Institute of Neurological Disorders and Stroke*
Amy Swain, *Office of Science, Department of Energy*
Edmund Talley, *National Institute of Neurological Disorders and Stroke*

Speakers

Stephen Adams, *University of California, San Diego*
Wei-Chung Allen Lee, *Harvard University*
Rachana Anathakrishnan, *Globus*
Jesús Arroyo, *University of Maryland*
Stephen Baccus, *Stanford University*
Christa Baker, *Princeton University*
Helen Barbas, *Boston University*
Daniela Boassa, *University of California, San Diego*
Julia Buhmann, *ETH/UZH Zurich*
Eric Bushong, *University of California, San Diego*
Gwyneth Card, *Howard Hughes Medical Institute, Janelia Campus*
Albert Cardona, *Medical Research Council Laboratory of Molecular Biology*
Si Chen, *Argonne National Laboratory*
Kwanghun Chung, *Massachusetts Institute of Technology*
Winfried Denk, *Max Planck Institute of Neurobiology*
Sven Dorkenwald, *Princeton University*
Catherine Dulac, *Harvard University*
Eva Dyer, *Georgia Institute of Technology*
Florian Engert, *Harvard University*

Michale Fee, *Massachusetts Institute of Technology*
Viviana Gradinaru, *California Institute of Technology*
Will Gray-Roncal, *Johns Hopkins University*
Hannah Haberkern, *Howard Hughes Medical Institute, Janelia Campus*
Kenneth Hayworth, *Howard Hughes Medical Institute, Janelia Campus*
Elizabeth Hillman, *Columbia University*
Michal Januszewski, *Google*
Greg Jefferis, *University of Cambridge*
Bobby Kasthuri, *Argonne National Laboratory, University of Chicago*
Keun-Young Kim, *University of California, San Diego*
Joergen Kornfeld, *Massachusetts Institute of Technology*
Jeff Lichtman, *Harvard University*
Albert Lin, *Harvard University*
Xiaotang Lu, *Harvard University*
Nuno Maçarico da Costa, *Allen Institute*
Thomas Macrina, *Princeton University*
Gayathri Mahalingam, *Allen Institute*
Tianyi Mao, *Oregon Health & Science University*
Adam Marblestone, *Massachusetts Institute of Technology*
John Mendenhall, *University of Texas, Austin*
William O'Mullane, *Large Synoptic Survey Telescope*
Benjamin Pedigo, *Johns Hopkins University*
Eric Perlman, *The Jackson Laboratory*
Stephen Plaza, *Howard Hughes Medical Institute, Janelia Campus*
Clay Reid, *Allen Institute*
Patricia Rivlin, *Howard Hughes Medical Institute, Janelia Campus*
Alexander Ropelewski, *Pittsburgh Supercomputing Center*
Gerry Rubin, *Howard Hughes Medical Institute, Janelia Campus*
Kerrianne Ryan, *Dalhousie University*
Stephan Saalfeld, *Howard Hughes Medical Institute, Janelia Campus*
Bernardo Sabatini, *Harvard University*
Amy Sterling, *EyeWire*
Scott Sternson, *University of California, San Diego*
Srini Turaga, *Howard Hughes Medical Institute, Janelia Campus*
Nicholas Turner, *Princeton University*
Tom Uram, *Argonne National Laboratory*
Alipasha Vaziri, *Rockefeller University*
Ashwin Vishwanathan, *Princeton University*
Ilana Witten, *Princeton University*
Rachel Wong, *University of Washington*
Sarah Woolley, *Columbia University*
Claire Wyart, *Brain & Spine Institute*
Ting Xu, *Child Mind Institute*
X. William Yang, *University of California, Los Angeles*
Mei Zhen, *University of Toronto*
Marta Zlatic, *Medical Research Council Laboratory of Molecular Biology*

Discussants

Cori Bargmann, *Rockefeller University, Chan Zuckerberg Initiative*
Timo Bremer, *Lawrence Livermore National Laboratory*
Emery Brown, *Massachusetts Institute of Technology, Harvard University*
JoAnn Buchanan, *Allen Institute*
Elizabeth Buffalo, *University of Washington*
Frances Chance, *Sandia National Labs*
Anne Churchland, *University of California, Los Angeles*
Adrienne Fairhall, *University of Washington*
Ian Foster, *University of Chicago*
Moritz Helmstaedter, *Max Planck Institute for Brain Research*
Vivek Jayaraman, *Howard Hughes Medical Institute, Janelia Campus*
Ed Lein, *Allen Institute*

Anton Maximov, *Scripps Research Institute*
Lisa Miller, *Brookhaven National Laboratory*
Eugene Myers, *Max-Planck-Gesellschaft*
Linnaea Ostroff, *University of Connecticut*
Olga Ovchinnikova, *Oak Ridge National Laboratory*
Carey Priebe, *Johns Hopkins University*
Josh Sanes, *Harvard University*
Mark Schnitzer, *Stanford University*
Terry Sejnowski*, The Salk Institute*
Sebastian Seung, *Princeton University*
Kim Stachenfeld, *DeepMind*
Stephen Streiffer, *Argonne National Laboratory*
Alex Szalay, *Johns Hopkins University*
David Tank, *Princeton University*
Andreas Tolias, *Baylor College of Medicine*
Rachel Wilson, *Harvard University*
Larry Zipursky, *University of California, Los Angeles*

Invited Attendees

Frank Alexander
Nicola Allen
Ilkay Altintas
Paola Arlotta
Alison Barth
Alex Bates
Flavie Bidel
Tammie Borders
Kristofer Bouchard
Edward Boyden
Kevin Briggman
Edward Callaway
Wah Chiu
Dmitri Chklovskii
Chiara Cirelli
Thomas Clandinin
Marlene Cohen
Nuno da Costa
Vincent De Andreade
Alice Dohnalkova
Hongwei Dong
Ali Erturk
Ila Fiete
Jan Funke
Kelly Gaither
Daniel Gardner
Mark Goldman
Julie Harris
Kenneth Harris
Kristen Harris
Harald Hess
Josh Huang
Richard Huganir
Jean Anne Incorvia
Gwen Jacobs
Viren Jain
Vivek Jayaraman
Bryan Jones
Tom Kalil
Garret Kenyon
Kerstin Kleese van Dam
Anna Kreshuk
Carolyn Larabell
Peng Li
Peter Littlewood
Ashok Litwin-Kumar
Tingshan Liu
Jitendra Malik
Kristina Micheva
Partha Mitra
Atsushi Miyawaki
Ryuta Mizutani
Caroline Montojo
Lydia Ng
Tatjana Paunesku
Hanspeter Pfister
Scott Retterer
Kathy Rockland
Aravi Samuel
Nate Sawtell
Philip Schlegel
Katie Schuman
James Sethian
Nathan Shaner
Nir Shavit
Gordon MG Shepherd
Stephen Smith
Rick Stevens
Karel Svoboda
Larry Swanson
Alice Ting
Jessica Tollkuhn
David Van Essen
Axel Visel
Stefan Vogt
Xiao-Jing Wang
Bobbie-Jo Webb-Roberson
Torre Wenaus
Alyssa Wilson
Xianghui Xiao

Appendix B: Workshop Agendas

Workshop 1: Significance of Mapping Complete Neural Circuits

February 5, 2021

Overall Goal: To explore the potential impact of generating comprehensive maps of brain connectivity—"wiring diagrams" spanning the entire mammalian brain.

Agenda:

11:00 am ET **Welcome and Series Overview**
John Ngai, Ph.D.
Harriet Kung, Ph.D.

11:08 am ET **Welcome and Introduction**
Larry Abbott, Ph.D. (Workshop Co-Lead)
Eve Marder, Ph.D. (Workshop Co-Lead)

11:25 am ET **Workshop Logistics**

11:30 am ET **Speakers**

Each speaker will address three questions:

a) What issues would you be able to address if you had detailed connectome data that cannot currently be investigated?
b) Are there conceptual issues that would be approached entirely differently if detailed connectomic data were available?
c) What would be the value added between having one versus many samples—across individuals, pathological conditions, and species—and how many is enough?

11:30 am ET Beyond the lamppost: How the fly connectome is helping us to see the whole picture
Gwyneth Card, Ph.D.
Gerry Rubin, Ph.D.

11:50 am ET Mapping the zebrafish whole-brain connectome
Florian Engert, Ph.D.
Claire Wyart, Ph.D.

12:10 pm ET Songbird connectomes
Michale Fee, Ph.D.
Sarah M. N. Woolley, Ph.D.

12:30 pm ET **Break**

12:40 pm ET Connectomics and computation in the retina
Stephen A. Baccus, Ph.D.
Rachel Wong, Ph.D.

1:00 pm ET The mouse brain connectome would transform understanding of core behaviors & physiology
Catherine Dulac, Ph.D.
Scott Sternson, Ph.D.

1:20 pm ET Basal ganglia connectome: Towards an accurate neural circuit model of reinforcement learning
Bernardo Sabatini, M.D., Ph.D.
Ilana Witten, Ph.D.

1:40 pm ET Q & A Session

2:25 pm ET **Break**

2:55 pm ET **Discussion Panel**

Panel Discussion topic: How important is a whole mouse brain connectome?

Leaders: *Adrienne Fairhall, Ph.D.; Larry Zipursky, Ph.D.*
Discussants: *Timo Bremer, Ph.D.; Anne Churchland, Ph.D.; Vivek Jayaraman, Ph.D.; Jeff Lichtman, M.D., Ph.D.; Josh Sanes, Ph.D.; David Tank, Ph.D.*

3:55 pm ET **Closing Remarks**
Larry Abbott, Ph.D. (Workshop Co-Lead)
Eve Marder, Ph.D. (Workshop Co-Lead)

3:55 pm ET **Workshop Ends**

Workshop 2: Sample Preparation in Mammalian Whole-Brain Connectomics

February 17, 2021

Overall Goal: To identify current capabilities and open issues in sample preparation for (1) whole mouse brain (WMB) electron microscopy (EM) connectomics at synaptic resolution; and (2) complementary imaging at lower resolution, in mouse brains and larger (including human) brains.

Agenda:

11:00 am ET **Welcome and Series Overview**
John Ngai, Ph.D.
Harriet Kung, Ph.D.

11:09 am ET **Welcome and Introduction**
Davi Bock, Ph.D. (Workshop Co-Lead)
Hongkui Zeng, Ph.D. (Workshop Co-Lead)

11:25 am ET **Workshop Logistics**
Ruben Alvarez, Ed.D.

11:30 am ET **Session 1 - Continuous whole mouse brain EM connectomes**

Each speaker will address four questions:

a) What current sample preparation methods for WMB EM connectomics maximally conserve structural continuity across the entire brain? What are the pros and cons for each approach?
b) All currently available imaging methods suitable for WMB connectomics require subdivision of the sample prior to imaging. What amount of loss is expected with current subdivision methods? Can continuity at single axon (~50 nm) level across the entire brain be maintained?
c) What are the prospects for improved methods to reduce loss during subdivision, either prior to resin embedding (e.g. vibratome sectioning) or after (e.g. hot knife)?
d) Prior to imaging, how to assess the quality of the sample preparation? How to predict and/or validate continuity before, during and after imaging?

11:30 am ET EM Staining for Whole Mouse Connectome
Xiaotang Lu, Ph.D.

11:45 am ET Sample preparation in the context of a large scale connectomics pipeline
Nuno Maçarico da Costa, Ph.D.

12:00 pm ET Specimen Preparation and Screening for Volume EM
Eric Bushong, Ph.D.

12:15 pm ET Ideas for thick sectioning brain tissue
Kenneth Hayworth, Ph.D.

12:30 pm ET Q & A Session

1:00 pm ET **Break**

1:10 pm ET **Session 2 - Complementary whole-brain imaging in mouse and larger species**

Each speaker will address four questions:

a) What types of complementary light microscopy (LM) data should be collected on the same brain and/or on different brains to help with processing and/or interpretation of the WMB EM connectome? E.g. projectomes at population or single cell level; functional imaging in behaving animals before EM; LM/X-ray/EM of CNS, PNS, and/or whole body; cell type specific labeling/tagging; immunolabeling.
b) Why is such information useful, what questions can it be used to address? What is the priority and order of such data generation in conjunction with the WMB EM connectome data generation?
c) Which techniques can be extended/scaled to larger brains (e.g. human and non-human primate brains)? Will these techniques require pre-labeling in live tissues?
d) What considerations are needed in sample selection (e.g. sex, strain or race, age, individual variation, health status, life history) and sample preparation (e.g. technical requirements, desired data types from the living brains)?

1:10 pm ET Scalable approaches for functional and structural light microscopy of the mammalian brain
Elizabeth Hillman, Ph.D.

1:20 pm ET Brain connectivity in primates
Helen Barbas, Ph.D.

1:30 pm ET Towards holistic phenotyping and understanding of the human brain
Kwanghun Chung, Ph.D.

1:40 pm ET Engineered gene delivery vectors for high-precision broad coverage of the mammalian brain
Viviana Gradinaru, Ph.D.

1:50 pm ET Projectomes and connectomes from mice to primates
Bobby Kasthuri, M.D., Ph.D.

2:00 pm ET Chemical Tools for CLEM and Color EM
Stephen Adams, Ph.D.

2:10 pm ET Nanoscale map of whole-cells and tissue using genetic probes technologies and 3D electron microscopy
Daniela Boassa, Ph.D.

2:20 pm ET Q & A Session

2:50 pm ET **Break**

3:00 pm ET **Discussion Panel**

Panel Discussion topic: *Charting a roadmap for whole mouse brain connectomes and larger brain projectomes—sample preparation feasibility as a prerequisite for choosing imaging platforms, sample selection criteria for scaling to multiple brains.*

Further charges:

- Summarize current state of tissue preparation
- Identify key issues to be resolved in sample prep for WMB EM connectome in conjunction with different EM imaging platforms
- Articulate needs and advantages of complementary data types for WMB EM connectomes
- Define the limitations, potentials and greatest opportunities in moving to larger brains

Panel chairs: *Davi Bock, Ph.D.; Hongkui Zeng, Ph.D.*
Discussants: *JoAnn Buchanan, M.S.; Jeff Lichtman, M.D., Ph.D.; Lisa Miller, Ph.D.; Linnaea Ostroff, Ph.D.; Clay Reid, M.D., Ph.D.*

4:00 pm ET **Closing Remarks/End of Workshop**

Workshop 3: Experimental Modalities for Whole-Brain Connectivity Mapping

March 5, 2021

Overall Goal: What are state of the art, opportunities and challenges relevant to each of the technologies and approaches you represent?

Agenda:

11:00 am ET **Welcome and Series Overview**
John Ngai, Ph.D.
Harriet Kung, Ph.D.

11:10 am ET **Welcome and Introduction**
Mark Ellisman, Ph.D. (Workshop Co-Lead)
Galya Orr, Ph.D. (Workshop Co-Lead)

11:30 am ET **Workshop Logistics**

11:35 am ET **Speaker Sessions**

Each speaker will address four questions:

a) What resolution is required to attain meaningful wiring models?
b) Which imaging technologies could be driven to scale and most easily disseminated?
c) Compare and contrast single large-scale efforts to democratize data collection.
d) How do you encode dynamics in static brain maps (adding the time domain)?

11:35 am ET **Session 1—Multi-Scale Imaging of the Connectome**

11:35 am ET Multi-scale Imaging of the Connectome
Jeff Lichtman, M.D., Ph.D.

11:45 am ET Raw speed
Winfried Denk, Ph.D.

11:55 am ET Multi-scale imaging of connectomes using TEM
Davi Bock, Ph.D.

12:05 pm ET Multi-scale Imaging of Connectomes With Photons and Electrons
Wei-Chung Allen Lee, Ph.D.

12:15 pm ET Synchrotron Hard X-ray Microscopy for the Connectome
Si Chen, Ph.D.

12:25 pm ET tomoSEM - tomography in tSEM
John Mendenhall, Ph.D.

12:35 pm ET Q & A Session

1:05 pm ET **Break**

1:15 pm ET **Session 2 – Projectome to Connectome Imaging**

Synapto-Projectomes: Toward connectome relationships in target fields

1:15 pm ET Axonal connectomics: projection maps in mice, NPHs, and humans
Clay Reid, M.D., Ph.D.

1:25 pm ET Mapping Projections of Melanopsin Retinal Ganglion Cells to Brain
Keun-Young Kim, Ph.D.

1:35 pm ET Connectome at the 'operational' resolution and in the context of protein dynamics
Tianyi Mao, Ph.D.

Strategies to bridge spatial and temporal gaps

1:45 pm ET Novel mouse genetic tools for scalable analyses of brain cell morphology
X. William Yang, Ph.D.

1:55 pm ET Towards whole-cortical cellular resolution recording of neuroactivity
Alipasha Vaziri, Ph.D.

2:05 pm ET Next-Generation Connectomics (NGC) via optical in-situ multiplexing
Adam Marblestone, Ph.D.

2:15 pm ET Q & A Session

2:45 pm ET **Break**

2:55 pm ET **Discussion Panel**

Panel Discussion topic: The tyranny of scale: Challenges presented by spatial and temporal scales

Panel chairs: *Moritz Helmstaedter, Ph.D.; Hongkui Zeng, Ph.D.*

Discussants: *Moritz Helmstaedter, Ph.D.; Hongkui Zeng, Ph.D.; Anton Maximov, Ph.D.; Kenneth Hayworth, Ph.D.; Lisa Miller, Ph.D.; Mark Schnitzer, Ph.D.*

3:55 pm ET **Closing Remarks**
Mark Ellisman, Ph.D. (Workshop Co-Lead)
Galya Orr, Ph.D. (Workshop Co-Lead)

4:00 pm ET **Workshop Ends**

Workshop 4: Connectome Generation and Data Pipelines

March 17, 2021

Overall Goal: To identify the current state of the art, challenges and opportunities associated with connectome generation and data pipelines.

Agenda:

11:00 am ET **Welcome and Series Overview**
John Ngai, Ph.D.
Harriet Kung, Ph.D.

11:10 am ET **Welcome and Introduction**
Nicola Ferrier, Ph.D. (Workshop Co-Lead)
Viren Jain, Ph.D. (Workshop Co-Lead)

11:20 am ET **Workshop Logistics**

11:25 am ET **Speaker Sessions**

Each speaker will address three questions:

a) What are the prospects and challenges you see in scaling your approach (and/or analysis goals in general) to volumes comprising hundreds of petavoxels of imaging data?
b) What kinds of general improvements in technical infrastructure (hardware and/or software) would mostly dramatically aid in scaling your analysis?
c) What new frontiers of analysis (especially as may be enabled by, or complementary to, a whole mammalian connectome) are you most excited by? What challenges do you see in pursuing those frontiers?

11:25 am ET **Session 1 – Volume assembly (stitching, alignment, registration)**

11:25 am ET Volume Assembly in the context of a large scale Connectomics pipeline
Gayathri Mahalingam, Ph.D.

11:35 am ET Scalable stitching and alignment for 3D electron microscopy
Stephan Saalfeld, Ph.D.

11:45 am ET Alignment of serial section EM for automated reconstruction
Thomas Macrina, B.S.E.

11:55 am ET **Break**

12:05 pm ET **Session 2 – Automated reconstruction and annotation (segmentation, synapses, types, compartments)**

12:05 pm ET Automated dense reconstruction of neural tissue
Michal Januszewski, Ph.D.

12:15 pm ET Synapse Detection in a Fly Brain
Julia Buhmann, Ph.D.

12:25 pm ET Towards a parts list for a whole mouse brain
Nicholas Turner, M.A.

12:35 pm ET **Break**

12:45 pm ET **Session 3 – Connectome proofreading, verification, crowd-sourcing**

12:45 pm ET Proofreading the Drosophila Hemibrain Connectome
Patricia Rivlin, Ph.D.

12:55 pm ET Connectome querying and verification
Joergen Kornfeld, Ph.D.

1:05 pm ET Crowdsourcing Connectomics at Scale
Amy Sterling, B.S.

1:15 pm ET **Break**

1:25 pm ET **Session 4 – HPC, data infrastructure, visualization**

1:25 pm ET Research Data Management at Scale
Rachana Anathakrishnan, M.S.

1:35 pm ET Connectomics at ALCF
Tom Uram, B.S.

1:45 pm ET Scaling Pipelines for EM Connectomics
Eric Perlman, Ph.D.

1:55 pm ET **Q & A Session**

2:40 pm ET **Break**

3:00 pm ET **Discussion Panel**

Panel Discussion topic: Reflections on how experience as a computer scientist might inform an *'exascale' project aimed at generating comprehensive maps of brain connectivity*

Panel Chairs: *Nicola Ferrier, Ph.D.; Viren Jain, Ph.D.*
Discussants: *Frances Chance, Ph.D.; Adrienne Fairhall, Ph.D.; Ian Foster, Ph.D.; Eugene Myers, Ph.D.; Olga Ovchinnikova, Ph.D.; Andreas Tolias, Ph.D.*

4:00 pm ET **Closing Remarks/End of Workshop**

Workshop 5: Optimizing the Use of Connectomic Data to Drive Data Science and Scientific Discovery

March 31, 2021

Overall Goal: To identify the current state of the art, challenges and opportunities in the analysis, interpretation, and dissemination of brain connectomic data. This workshop will focus on new analysis methods for partial, full and multiple connectomes emerging from a number of model systems, in addition to the problem of making connectomic data accessible to the community.

Agenda:

11:00 am ET **Welcome and Series Overview**

John Ngai, Ph.D.
Harriet Kung, Ph.D.

11:10 am ET **Welcome and Introduction**
Mala Murthy, Ph.D. (Workshop Co-Lead)
Joshua T. Vogelstein, Ph.D. (Workshop Co-Lead)

11:25 am ET **Workshop Logistics**

11:30 am ET **Speaker Sessions**

Each speaker will address three questions:

a) What scientific breakthroughs have analyses of connectomes already enabled?
b) What current methods for analyses of connectomes are most promising?
c) What new types of analyses (not developed yet) are needed to make progress and enable scientific breakthroughs?
d) How do we integrate analyses across scales (from nano to micro to meso)?
e) How do we integrate connectomic data with functional data?
f) What types of data and analyses are needed to make interesting comparisons between connectomes?
g) What kinds of infrastructure are needed to make connectomic data and analyses widely available to the neuroscience community?

11:30 am ET **Session 1—Analysis of Partial and Full Connectomes**

11:30 am ET From the fly connectome to the mouse connectome
Greg Jefferis, Ph.D.

11:40 am ET How to simulate a connectome to gain a mechanistic understanding of neural computation
Srini Turaga, Ph.D.

11:50 am ET Learning representations of neural architecture across many spatial scales
Eva Dyer, Ph.D.

12:00 pm ET Lessons from analyzing navigational circuits in the *Drosophila* hemibrain connectome
Hannah Haberkern, Ph.D.

12:05 pm ET Neural network organization for courtship song feature detection in *Drosophila*
Christa Baker, Ph.D.

12:10 pm ET *Ciona* connectome analyses and considerations for the future
Kerrianne Ryan, Ph.D.

12:15 pm ET Whole-brain functional studies in *C. elegans*: A direct comparison of structure and function
Albert Lin, Ph.D.

12:20 pm ET Modularity and neural coding from a synaptic wiring diagram
Ashwin Viswanathan, M.D.

12:25 pm ET Q & A Session

12:45 pm ET **Break**

12:55 pm ET **Session 2—Analysis of Multiple Connectomes**

12:55 pm ET From a developmental connectomics study: Analyses interpretations limitations
Mei Zhen, Ph.D.

1:05 pm ET Combining connectomics with functional studies to investigate circuit mechanisms of learning and action-selection
Marta Zlatic, Ph.D.

1:15 pm ET Multiscale statistical learning for connectome data
Jesús Arroyo, Ph.D.

1:25 pm ET Optimization of the reliability of functional brain connectome
Ting Xu, Ph.D.

1:35 pm ET Network data science for bilateral brains: Applications in the larval *Drosophila* connectome
Benjamin Pedigo

1:45 pm ET Q & A Session

2:05 pm ET **Break**

2:15 pm ET **Session 3—Making Connectomic Data Accessible to the Community**

2:15 pm ET Engineering the Future of Connectomics
Will Gray-Roncal, Ph.D.

2:20 pm ET Neuroscience research on the basis of whole-brain connectomics
Albert Cardona, Ph.D.

2:25 pm ET Continuous proofreading and analysis of large EM reconstructions
Sven Dorkenwald, Ph.D.

2:30 pm ET Exploring connectomes at varying levels of detail
Stephen Plaza, Ph.D.

2:35 pm ET The Brain Image Library: Designated Repository for NIH BRAIN Initiative Microscopy Data
Alexander Ropelewski

2:40 pm ET Rubin Observatory – Astronomy Big Data
William O'Mullane, Ph.D.

2:50 pm ET Q & A Session

3:10 pm ET **Break**

3:20 pm ET **Discussion Panel**

Session Chairs: *Mala Murthy, Ph.D.; Joshua T. Vogelstein, Ph.D.*
Discussants: *Carey E. Priebe, Ph.D.; Sebastian Seung, Ph.D.; Terry Sejnowski, Ph.D.; Kim Stachenfeld, Ph.D.; Alex Szalay, Ph.D.; Rachel Wilson, Ph.D.*

4:20 pm ET **Workshop Ends**

Town Hall

April 28, 2021

12:00 pm ET **Opening Remarks**

Ruben Alvarez, Ed.D.
Deputy Director, Office of Technology Development and Coordination, NIMH

12:05 pm ET **Keynote Address**

Caroline Montojo, Ph.D.
President, The Dana Foundation

12:15 pm ET **Panel 1: The State of the Art and the Challenges and Opportunities Ahead**

Moderators:

Kristen Harris, Ph.D.
Workshop Series Co-Chair; Professor of Neuroscience, University of Texas – Austin

Peter Littlewood, Ph.D.
Workshop Series Co-Chair; Professor of Physics, The University of Chicago; Emeritus Director, Argonne National Lab

Panelists:

Larry Abbott, Ph.D.
Workshop 1 Co-Lead; William Bloor Professor of Theoretical Neuroscience, Professor of Physiology andCellular Biophysics, Columbia University

Eve Marder, Ph.D.
Workshop 1 Co-Lead; Victor and Gwendolyn Beinfield Professor of Biology, Brandeis University

Davi Bock, Ph.D.
Workshop 2 Co-Lead; Research Associate Professor, University of Vermont College of Medicine

Hongkui Zeng, Ph.D.
Workshop 2 Co-Lead; Executive Vice President and Director, Allen Institute for Brain Science

Mark Ellisman, Ph.D.
Workshop 3 Co-Lead; Director, National Center for Microscopy and Imaging Research; Professor, Neurosciences, Bioengineering; Director, Center for Research in Biological Systems

Galya Orr, Ph.D.

Workshop 3 Co-Lead; Biomedical Scientist, Pacific Northwest National Laboratory

Nicola Ferrier, Ph.D.
Workshop 4 Co-Lead; Senior Computer Scientist, Argonne National Laboratory

Viren Jain, Ph.D.
Workshop 4 Co-Lead; Senior Staff Research Scientist, Google

Mala Murthy, Ph.D.
Workshop 5 Co-Lead; Professor of Neuroscience, Princeton University

Joshua Vogelstein, Ph.D.
Workshop 5 Co-Lead; Assistant Professor of Biomedical Engineering, Biostatistics, AppliedMathematics & Statistics, Neuroscience, and Computer Science, Johns Hopkins University

1:20 pm ET **Q&A with Audience**

1:40 pm ET **Break**

1:50 pm ET **Panel 2: The Significance of Mapping Mammalian Brain Circuitry at Scale**

Moderators:

Cori Bargmann, Ph.D.
Torsten N. Wiesel Professor of Genetics and Genomics, Neurosciences and Behavior, The RockefellerUniversity; Head of Science, Chan Zuckerberg Initiative

Terrence Sejnowski, Ph.D.
Professor, Head, Computational Neurobiology Laboratory, Francis Crick Chair, The Salk Institute

Panelists:

Emery Brown, M.D., Ph.D.
Edward Hood Taplin Professor of Medical Engineering and of Computational Neuroscience, Professor ofHealth Sciences and Technology, Massachusetts Institute of Technology; Warren M. Zapol Professor of Anaesthesia, Harvard Medical School, Massachusetts General Hospital; Director, Harvard-MIT Health Sciences and Technology Program, MIT; Associate Director, Institute for Medical Engineering and Science, MIT; Investigator, Picower Center for Learning and Memory, MIT

Elizabeth Buffalo, Ph.D.
Professor and Chair, Department of Physiology and Biophysics, University of Washington School ofMedicine; Interim Associate Director for Research, Washington National Primate Research Center

Gwyneth Card, Ph.D.

Group Leader, Howard Hughes Medical Institute Janelia Research Campus

Ed Lein, Ph.D.
Senior Investigator, Allen Institute for Brain Science; Affiliate Professor of Neurological Surgery,University of Washington

Stephen Streiffer, Ph.D.
Deputy Laboratory Director for Science and Technology, Argonne National Laboratory

2:35 pm ET **Q&A with Audience**

2:55 pm ET **Closing Remarks**

John Ngai, Ph.D.
Director, NIH Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative

Harriet Kung, Ph.D.
Deputy Director for Science Programs, Office of Science, U.S. Department of Energy

3:00 pm ET **Town Hall Ends**

Appendix C: Abbreviations List

2D	2 dimensional
2P	2 photon
3D	3 dimensional
3P	3 photon
AAV	adeno-associated viral vectors
ALCF	Argonne Leadership Computing Facility
API	application programming interface
APS	Advanced Photon Source
BIL	Brain Image Library
BIP	Brain Informatics Program
CAP	complete, accurate, and permanent
CAVE	Connectome Annotation Versioning Engine
CCD	charged-coupled device
CLEM	correlative light and electron microscopy
CNN	convolution neural network
CNS	central nervous system
CSF-cN	cerebrospinal fluid-contacting neuron
CT	computed tomography
DAB	diaminobenzidine
DCSBM	degree-corrected stochastic block model
DOE	U.S. Department of Energy
EELS	electron energy loss spectroscopy
EF-TEM	energy filtered transmission electron microscopy
EM	electron microscopy
EPSP	excitatory postsynaptic potential
ERL	expected run length
FAFB	full adult fly brain
FFN	flood-filling network
FIB-SEM	focused ion beam
fMRI	functional magnetic resonance imaging
FOV	field of view
GCIB	gas cluster ion beam
GCIB-MultiSEM	gas cluster ion beam multibeam SEM
GPU	graphics processing unit
HiP-CT	hierarchical phase contrast tomography
HPC	high performance computing
IARPA	Intelligence Advanced Research Projects Activity
ipRGC	intrinsically photo-sensitive retinal ganglion cell
LBT	lanthanide-binding tag
LGN	lateral geniculate nucleus

LM	light microscopy
LTP	long-term potentiation
MRI	magnetic resonance imaging
NCMIR	National Center for Microscopy and Imaging Re
MBON	mushroom body output neuron
miniSOG	mini-Singlet Oxygen Generator
MORF	MOnonucleotide Repeat Frameshift Cre reporter mice
NHP	non-human primate
NIH	National Institutes of Health
OPN	olivary prefrontal nucleus
PMI	postmortem interval
PNS	peripheral nervous system
QC	quality control
SABER	Scalable Analytics for Brain Exploration Researcher
SBEM	serial block-face SEM
SCN	suprachiasmatic nucleus
SEM	serial electron microscopy
S/N	signal-to-noise ratio
ssEM	serial section electron microscopy
ssTEM	serial section transmission electron microscopy
STEM	scanning transmission electron microscopy
STG	stomatogastric ganglion
TEM	transmission electron microscopy
tSEM	transmission-mode scanning EM
VTA	ventral tegmental area
XFM	X-ray fluorescence microscopy

Appendix D: Continued Participation

This workshop series is the product of collective wisdom and insight from a large swath of invested participants. Members of the scientific community—including from neuroscience, genetics, computer science, engineering, astrophysics, and more—as well as engaged members of the public, from both the private and public sectors, all contributed their knowledge and perspectives to achieve the goal of understanding the brain through mapping its connections. In addition to the 195 people listed in Appendix A who had assigned roles in the series, over 900 people registered to attend the series. The live online broadcast of each workshop in the series garnered 404-795 unique live viewers who had the opportunity to contribute questions and comments to each event. The first workshop alone garnered live viewers from 33 countries, demonstrating the wide interest and engagement in this topic. As of the publication of this report, the online broadcasts of the workshops and Town Hall have collectively amassed 4,407 views across 58 countries. These broadcasts continue to be available through the NIH at the links below, and will remain accessible for later viewing:

- Significance of mapping complete neural circuits (Workshop 1)
- Sample preparation in mammalian whole-brain connectomics (Workshop 2)
- Experimental modalities for whole-brain connectivity mapping (Workshop 3)
- Connectome generation and data pipelines (Workshop 4)
- Optimizing the use of connectomic data to drive data science and scientific discovery (Workshop 5)
- Brain Connectivity Town Hall

In parallel to the series, the NIH and DOE also hosted a crowdsourcing campaign on the IdeaScale platform to further extend the opportunity to provide feedback on these topics.

The campaign sought input from experts in brain connectomics and other key stakeholders, including researchers in academia and industry, clinicians, scientific societies and advocacy organizations, and interested members of the public, particularly soliciting ideas on the **significance** of mapping mammalian brain circuitry in detail, the current **state of the art** in mapping complete neural circuits, and the current **opportunities** for advancing connectomics technologies and **challenges** that need to be addressed to generate comprehensive maps of brain connectivity.

This campaign will remain open an additional 60 days following the publication of this report. The campaign can be accessed from the series website or from the IdeaScale platform directly.