

# nature neuroscience

## Scaling up neuroscience

In the foreword to *Discovering the Brain*, James Watson wrote, “the brain is the most complex thing we have yet discovered in our universe. It contains hundreds of billions of cells interlinked through trillions of connections. The brain boggles the mind.” Obviously to understand brain function, we need to confront its complexity. Various strategies have been proposed to address this issue, but one view is that the way forward is to obtain vast amounts of data characterizing the brain. Gathering the data is only the first step on this path, however; effectively mining this information and interpreting it are as difficult, and as crucial.

Following the success of large-scale efforts in molecular biology, efforts to ‘scale up’ data production and analysis in neuroscience are beginning to escalate. In this issue, we present a focus consisting of nine Perspective articles that highlight the use of high-throughput methodologies in neuroscience, and discuss their current progress and future challenges. This format is a type of review intended for scholarly presentation of a particular viewpoint. For this young field, in which consensus views of the community have not yet emerged in most cases, a collection of Perspectives seemed most appropriate to us.

Mark Boguski and Allan Jones open the focus by discussing the influences of genomics on neuroscience, including comparative genomics, gene expression atlases and the organization of genome-scale projects. Karoly Mirnics and Jonathan Pevsner continue the spotlight on genomics, focusing on the problems of postmortem tissue and the use of gene microarrays to study neurological disorders. Seth Grant and Jyoti Choudhary then discuss some common approaches in proteomics, and the difficulties of applying these high-throughput proteomic approaches to study neurons and their function.

Genomic or proteomic approaches can only take us so far toward understanding neural function, which requires monitoring the activity of individual neurons. Gyuri Buzsaki examines large-scale recordings from neuronal populations, from technological aspects of multiple-electrode recording to the insights into brain function gained from these studies. John Chapin discusses the use of multiple-electrode recording for neural prosthetics.

Data collection remains only the first step across all these disciplines; good data can only be used effectively with good analytical tools for data extraction. Given the quantity of data generated by these high-throughput methods, data mining becomes even more challenging. Emery Brown, Partha Mitra and Robert Kass discuss the use of statistical methods for spike train analysis, and the use of mathematical tools to extract useful information from multiple-electrode recordings.

The final product of the nervous system, of course, is behavior. Larry Tecott and Eric Nestler discuss their views on high-throughput behavioral screening, and the strategies currently being used to rapidly screen mouse lines with automated behavioral tests.

Years of painstaking research have produced huge amounts of data on many different levels—from gene expression in the nervous system to phenotypic differences in mouse behavior. How then can we integrate this information to provide insights into the functioning of the

brain? The complexity of the nervous system presents unique challenges for the creation of databases, particularly those that link different types of data. As Mark Ellisman and colleagues explain, unlike genomic data (mainly various combinations of the same four letters), neuroscience data are much more complex—spanning cellular distributions of proteins, cell connectivity and physiological and behavioral data. In large part, neuroscience, like molecular biology, is being shaped by advances of information technology, and efforts are underway to use neuroinformatics to bring together these various datasets. The authors discuss possible approaches to these issues, and requirements for creating web-accessible databases. Jack Van Horn and colleagues give a personal view of their efforts to archive fMRI data in their publicly accessible database, and the scientific, technical and sociological concerns that hindered immediate acceptance of the fMRI datacenter.

To give our readers a flavor for the database projects currently being developed, we include short descriptions of a few publicly available databases. Nat Heintz and colleagues describe their GENSAT database and their large-scale effort to create an atlas of gene expression in the mammalian brain. Dan Goldowitz and colleagues discuss the consortium of databases available at [neuromice.org](http://neuromice.org) and their efforts to use ENU mutagenesis to create novel mouse mutants. Rob Williams and colleagues describe their collection of databases and analysis software for whole-genome analysis, and Dan Gardner introduces readers to a database for physiology data from mammalian cortex.

Our list of databases is by no means exhaustive. The Society for Neuroscience is leading an effort to link these databases through a common website called the Neuroscience Database Portal (<http://big.sfn.org/NDG/site>). Another such portal is provided by the Human Brain Project (<http://ycmi-hpb.med.yale.edu/hbpd>); we encourage our readers to surf through some of these databases to appreciate the depth and complexity of neuroscience data collection and storage.

We are grateful to a group of institutes from the National Institutes of Health (NIMH, NIDA, NINDS, NIAAA, NEI and NIDCD) for their generous financial support of this focus issue. With their help, we are making all the content of the focus freely available on the web. We share with the NIH a strong commitment to data sharing and promoting dialogue across different disciplines; however, responsibility for the editorial content (with the exception of the sponsors’ foreword) rests entirely with the editors of *Nature Neuroscience*.

Large-scale efforts such as those highlighted here represent only one approach to understanding the complexity of the brain, and high-throughput techniques will not be the whole solution to this puzzle. Still, the approaches for collecting, analyzing and sharing data described in this focus seem likely to contribute substantially to scientific progress. We hope that this collection of articles will stimulate discussion within the community, and highlight some of the challenges and triumphs of large-scale neuroscience.

*Kalyani Narasimhan*  
Associate Editor