Minireview

Mouse maps of gene expression in the brain

Susan E Koester* and Thomas R Insel[†]

Addresses: *Division of Neuroscience and Basic Behavioral Science, National Institute of Mental Health, Executive Blvd, Bethesda, MD 20892-9645, USA. [†]National Institute of Mental Health, Executive Blvd, Bethesda, MD 20892-9645, USA.

Correspondence: Thomas R Insel. Email: tinsel@mail.nih.gov

Abstract

The completion of the Allen Brain Atlas generated a great deal of press interest and enthusiasm from the research community. What does it do, and what other complementary resources increase its functionality?

Since the 19th century, neuroscientists have struggled to categorize the types of cells in the brain using anatomical or physiological markers as an indication of cell function [1]. Subdivisions of the brain are defined by clusters of neurons related by cellular architecture, connectional specificity or physiological properties. In the past 20 years, neuroscientists have added molecular biology to their arsenal of tools for categorizing brain cells. With the complete sequencing of the genomes of many organisms, neuroscientists have for the first time an opportunity to understand the full range of gene expression across the brain. Until recently, individual researchers have been assembling this information in stages appropriate to individual laboratories (for example, an expression map of all known transcription factors in the mouse brain [2]). Now three large-scale projects are working toward a map of all gene expression in the mouse brain, bringing us a giant step closer to understanding the classification and function of different types of neurons in the brain.

A complete neuroanatomical atlas of brain expression

Arguably the most complete of the molecular brain atlas efforts (in terms of coverage of the genome) is the Allen Brain Atlas [3], which was recently described by Ed Lein and colleagues [4]. The Allen Brain Institute (ABI), established in 2003 by Paul Allen (one of the original founders of Microsoft), set out to describe the expression of all known genes in the adult mouse brain. The atlas project used high-throughput, semi-automated *in situ* hybridization

methods developed by collaborators Gregor Eichele and Axel Visel [5] on rigorously controlled coronal sections through the entire postnatal day (P) 56 (adult) male mouse brain. Colorimetric *in situ* hybridization data for 21,000 genes are now posted online with open access [4]. The data are searchable by gene names or symbols, as well as by large anatomical regions (for example, cortex, midbrain, cerebellum) or by a growing number of smaller-scale brain nuclei (such as the substantia nigra and the ventral tegmental area).

A major advantage of the atlas is the near-saturation coverage of the genome in the brain-expression data. This allows researchers to understand brain gene expression at a genomic level: surprisingly, nearly 80% of all genes are expressed in the brain of the adult mouse. In addition, the vast majority of these show regionally or cell-type restricted patterns of expression. ABI staff members have developed a reference anatomical atlas that can be directly compared side by side with the *in situ* expression data. An add-on module also allows three-dimensional reconstruction of gene-expression patterns. Another useful feature is that the atlas links directly from images of particular gene-expression patterns to the corresponding gene entry in a variety of databases of gene structure and function, including the developmental GenePaint database described below. This significantly increases the versatility of the atlas.

While the recent paper by Lein *et al.* [4] is a landmark for molecular neuroanatomy, the Allen Brain Atlas needs to be

understood as a work in progress. As with any effort of this size, there are errors that will need to be corrected, hopefully in response to feedback from users. The limitation to one age and one gender will no doubt be addressed in future efforts. And, at this point, the annotation of the anatomical location of expression is only semi-automated and requires expert human validation. This is currently being done by a small number of expert collaborators, but the addition of new subregions will clearly be incredibly laborious and time-consuming in the absence of automated methods.

Collections of mouse knockouts for brain-specific genes

Meanwhile, the National Institutes of Health (NIH) has been working in parallel on the Gene Expression Nervous System Atlas (GENSAT) to establish a resource of targeted knockouts of brain-specific mouse genes with interesting expression patterns in development as well as in the adult brain [6,7]. There are two arms to the GENSAT project. A group at St Jude Children's Research hospital, Tennessee, led by Thomas Curran (now at the Children's Hospital of Philadelphia, Pennsylvania), has carried out radiometric *in situ* hybridization in mouse brains at four stages of development. A separate group led by Nathaniel Heintz at Rockefeller University, New York, cloned the DNA flanking the genes with restricted expression patterns in brain into bacterial artificial chromosomes (BACs) with the gene-coding sequence replaced by an enhanced green fluorescent protein (EGFP) reporter gene. The BAC is then used to create a transgenic mouse line that ideally expresses EGFP in the cells that normally express the gene of interest. The advantage of using the BAC as a cloning vector is that it can carry as much as 200 kb of genomic DNA - typically enough to include most, if not all, of the appropriate regulatory elements of the gene. The mice can then express the EGFP reporter construct in the same anatomical and cell-type-specific locations as the original gene. Although the subcellular localization of the native protein cannot be inferred from the EGFP location, the reporter fills the entire cell, allowing a clear characterization of the anatomical cell type in which the gene is expressed. Multiple lines are analyzed to ensure that the expression pattern is consistent, indicating that the pattern is due to the promoter elements of the gene of interest rather than the genomic insertion point. The anatomical annotators of the GENSAT data have embraced the variability between their *in situ* data and the expression patterns seen in the BAC transgenic mice, noting where there are differences for each gene catalogued.

GENSAT is quite different from the Allen Brain Atlas in that it is designed to yield abundant information about the expression patterns of a subset of genes, not to be a comprehensive atlas. The GENSAT project produces mouse lines for physiological as well as anatomical analysis and it includes developmental patterns of expression. Currently
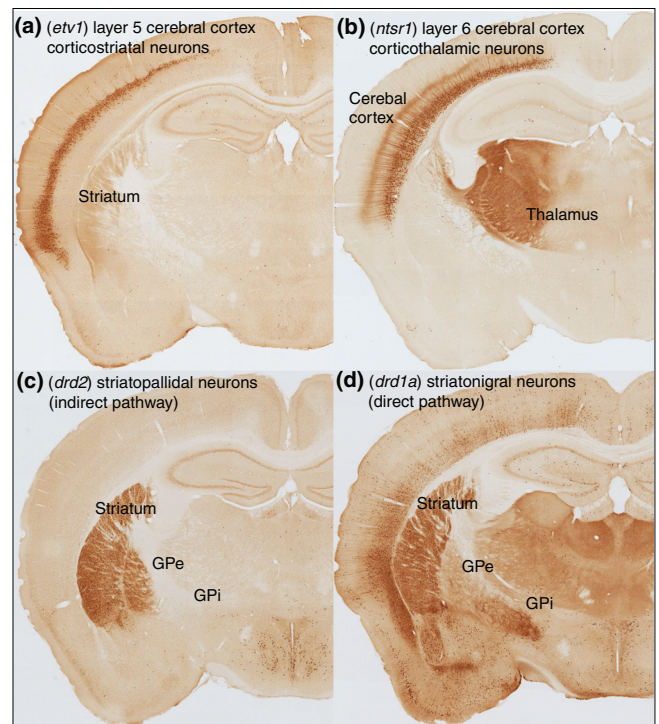

Figure 1
An example of Cre expression in mouse forebrain circuits, with labeling of specific neuronal projection systems in the cerebral cortex and striatum. **(a)** The ETS domain transcription factor (*etv1*) BAC drives Cre recombinase expression in layer 5 corticostriatal nuerons. **(b)** The neurotensin receptor (*ntsr1*) BAC drives Cre expression in layer 6 corticothalamic neurons. The projection axons of these neurons, which terminate in the dorsal thalamic nuclei, are clearly labeled. In the striatum, the majority of neurons are medium spiny projection neurons, which are evenly divided into striatopallidal (indirect pathway) and striatonigral (direct pathway) neurons, which selectively express the dopamine receptors Drd2 and Drd1a, respectively. Cre expression produced in *drd2* BAC-Cre lines **(c)** is directed to striatopallidal neurons. In this line, labeled neurons in the striatum extend axons that terminate in the globus pallidus external segment (GPe). **(d)** Expression produced in *drd1a* BAC-Cre lines, is directed to striatonigral neurons, which have axons that extend through the globus pallidus to terminate in the internal segment of the globus pallidus (GPi) and substantia nigra (not shown). Images courtesy of Charles Gerfen, National Institute of Mental Health.

436 lines [6] are distributed through the Mutant Mouse Research Resource Centers [8,9]. In addition, for some of the most interesting expression patterns, the project is developing lines where the BAC containing the gene promoter elements drive expression of the bacteriophage, Cre recombinase (Figure 1). These lines can then be bred with other lines of mice whose genomes are engineered to contain genes surrounded by *loxP* sites, resulting in a recombinant strain of mouse that eliminates a gene's expression only in the cells that express the Cre. These lines will be a valuable resource for investigators looking to understand the role of individual genes in specific subsets of cells in the nervous system.

The third project mapping gene expression in the mouse brain is the GenePaint atlas [10], a large-scale European effort led by Gregor Eichele at the Max-Planck-Institute of Biophysical Chemistry in Göttingen, Germany. In approach, this is a hybrid of the Allen Brain Atlas and GENSAT. The GenePaint atlas [10] catalogs *in situ* hybridization expression data in selected sections of whole mouse embryos for several thousand genes at embryonic day (E) 14.5 and augments these data with additional data for some genes in E10.5 embryos, E15.5 head, P7 and adult (day 56) brain, often in both coronal and sagittal sections. New data are posted regularly, with a list of new genes updated weekly on the home page. Researchers can search gene expression by gross anatomical area using the E14.5 dataset and can then link to views of the same gene expressed at different ages. GenePaint is a useful addition to the Allen Brain Atlas, giving researchers a flavor of dynamic gene-expression patterns during development. Although it does not have the same coverage of the genome as the Allen Brain Atlas, investigators can submit requests for additional genes to be tested and expression pattern data will be made available within a few weeks.

Mouse genetic resources

Once an investigator identifies a gene with an interesting expression pattern or some information on its function, many would welcome ready access to a targeted knockout mouse. The NIH Knockout Mouse Project (KOMP) [11] is an NIH effort to generate null mutations of all genes in C57BL/6 mice and distribute them to the research community. This effort is proceeding in several parallel directions. First, NIH have licensed 250 knockout lines from the companies Deltagen and Lexicon to make them widely available to the academic community; second, NIH issued contracts to create new knockouts of a list of priority gene candidates; and third, NIH is also 'repatriating' knockout mouse lines for distribution that have been generated by academic researchers, many using NIH funds. The KOMP project will manage the necessary husbandry and record-keeping for the distribution, which is an important advantage to investigators submitting popular mouse strains. The priorities for gene targeting or knockouts for repatriation are derived from public input from the research community, which is actively solicited [12]. The list of genes scheduled for targeting and the current list of knockout lines available are online [13]. In addition, genetically modified mice that were generated with other funding can be nominated to the KOMP for redistribution by the Mutant Mouse Regional Resource Centers, supported by NIH [8].

The European Union has launched a complementary effort to generate 20,000 lines of mice with conditional mutations with the ultimate goal of functionally characterizing all the genes in the mouse genome [14]. The European Conditional Mouse Mutagenesis Program (EUCOMM) brings together several different European projects to generate genetically modified mouse resources under one coordinating group. While the group is collaborating with the NIH KOMP as well as with a related Canadian effort, their emphasis is complementary. EUCOMM will use conditional gene-trap strategies to generate embryonic stem cell lines for eventual worldwide distribution. The project is in the early stages, but its progress can be followed at the EUCOMM website [15].

Future challenges for gene-expression databases

The philosophy of open access has guided the development of all of these resources, which is what makes them such a remarkable boon to neuroscientists and other biomedical researchers. Researchers are already finding the atlases useful in augmenting their data from unbiased screens for gene function [16]. The atlases provide critical reference data that allow investigators to include or exclude candidate genes on the basis of their expression patterns, and provide initial insights leading to a more complete investigation of expression patterns [17,18]. As the developers consider the evolution of the databases, it would be useful to expand that spirit to include data sources from the community. Leveraging the huge body of work by experts and incorporating new discoveries will be key to keeping these resources on the cutting edge.

A problem faced by the gene-expression databases is the lack of complete neuroanatomical annotation. This is not unexpected, as this type of annotation is time consuming, requires great expertise and is not easily automated. The task is made more difficult by the lack of a common neuroanatomical nomenclature. The Allen Brain Atlas devised a broad-scale nomenclature for its own reference atlas, rather than choose sides in the ongoing debate about nomenclature in individual brain areas. GenePaint points the user to several standard published adult and developmental brain atlases for both mice and rats. The GENSAT project goes farthest in describing the anatomy of gene expression and compares data from numbers of individual animals using different methods to assay gene expression. However, the number of genes covered remains small by comparison to the Allen Brain atlas. Furthermore, searching any of the databases for genes based on their expression in a particular cell type remains an elusive goal.

To be sustainable and useful in the long term, all these atlases will need to grow and evolve to incorporate additional data about subregional and cell-type-specific expression. The NIH databases that have become a mainstay for biomedical research, such as GenBank, rely on continuous update by users following a specified submission policy and curated by an in-house staff. In contrast, brain atlases have followed an older, proprietary model, alluding to their tradition of publication as books. A user-annotation model that allows addition of references to peer-reviewed

publications would ensure the atlases continue to support user needs and the current state of the science. This approach has been successfully followed in other model organisms such as the zebrafish, with their Zebrafish Information Network [19].

Ultimately, the subdivision of specific regions will need to be based on functional differences among nuclei or brain areas. This work has already begun in a number of species and cell types, correlating gene expression, anatomical and electrophysiological characteristics (reviewed in [20]). For example, Arlotta *et al*. [21] used a combination of axon tracing and gene expression data to categorize pyramidal neurons in layer 5 of the neocortex in mice, demonstrating that the anatomical distinctions also have gene-expression correlates. Sugino *et al*. [22] combined microarray analysis with electrophysiology and neurotransmitter immunocytochemistry to describe the variability within and between 12 known classes of forebrain neurons in mice. Data such as these will be key to understanding the function and development of circuits in the brain and for genetic manipulation of circuit elements. This level of analysis will be critical to the next stages of understanding how changes in gene function affect neuronal circuits and eventually complex behaviors.

References
1. Ramon y Cajal S: *Les Nouvelles Idées sur la Structure du Système Nerveux chez l'Homme et chez Vertébrés*. Paris: Reinwald; 1894.
2. Gray PA, Fu H, Luo P, Zhao Q, Yu J, Ferrari A, Tenzen T, Yuk DI, Tsung EF, Cai Z, *et al*.: **Mouse brain organization revealed through direct genome-scale TF expression analysis.** *Science* 2004, **306:**2255-2257.
3. **Allen Brain Atlas** [http://www.brain-map.org]
4. Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ, *et al*.: **Genome-wide atlas of gene expression in the adult mouse brain.** *Nature* 2007, **445:**168-176.
5. Visel A, Thaller C, Eichele G: **GenePaint.org: an atlas of gene expression patterns in the mouse embryo.** *Nucleic Acids Res* 2004, **32**(Database issue):D552-D556.
6. **NINDS GENSAT BAC Transgenic Project** [http://www.gensat.org]
7. Gong S, Zheng C, Doughty ML, Losos K, Didkovsky N, Schambra UB, Nowak NJ, Joyner A, Leblanc G, Hatten ME, *et al*.: **A gene expression atlas of the central nervous system based on bacterial artificial chromosomes.** *Nature* 2003, **425:**917-925.
8. **Mutant Mouse Regional Resource Centers** [http://www.mmrrc.org]
9. **NINDS GENSAT BAC Transgenics Project: MMRC mouse catalog** [http://www.gensat.org/MMRRC_report.jsp]
10. **Gene Paint** [http://www.genepaint.org]
11. **NIH Knockout Mouse Project** [http://www.nih.gov/science/models/mouse/knockout]
12. **Request for Community Input for the Nomination and Prioritization of Genes to be Targeted in the Knockout Mouse Project (KOMP): NOT-HG-07-004** [http://grants.nih.gov/grants/guide/notice-files/NOT-HG-07-004.html]
13. **Knockout Mouse Project Data Coordination Center** [http://www.knockoutmouse.org]
14. Auwerx J, Avner P, Baldock R, Ballabio A, Balling R, Barbacid M, Berns A, Bradley A, Brown S, Carmeliet P, *et al*.: **The European dimension for the mouse genome mutagenesis program.** *Nat Genet* 2004, **36:**925-927.
15. **European Conditional Mouse Mutagenesis Program** [http://www.eucomm.org]
16. Dugas JC, Tai YC, Speed TP, Ngai J, Barres BA: **Functional genomic analysis of oligodendrocyte differentiation.** *J Neurosci* 2006, **26:**10967-10983.
17. Papassotiropoulos A, Stephan DA, Huentelman MJ, Hoerndli FJ, Craig DW, Pearson JV, Huynh KD, Brunner F, Corneveaux J, Osborne D, *et al*.: **Common Kibra alleles are associated with human memory performance.** *Science* 2006, **314:**475-478.
18. Ponomarev I, Maiya R, Harnett MT, Schafer GL, Ryabinin AE, Blednov YA, Morikawa H, Boehm SL 2nd, Homanics GE, Berman AE, *et al*.: **Transcriptional signatures of cellular plasticity in mice lacking the alpha1 subunit of GABAA receptors.** *J Neurosci* 2006, **26:**5673-5683.
19. **ZFIN: the Zebrafish Model Organism Database** [http://zfin.org]
20. Nelson SB, Hempel C, Sugino K: **Probing the transcriptome of neuronal cell types.** *Curr Opin Neurobiol* 2006, **16:**571-576.
21. Arlotta P, Molyneaux BJ, Chen J, Inoue J, Kominami R, Macklis JD: **Neuronal subtype-specific genes that control corticospinal motor neuron development** *in vivo*. *Neuron* 2005, **45:**207-221.
22. Sugino K, Hempel CM, Miller MN, Hattox AM, Shapiro P, Wu C, Huang ZJ, Nelson SB: **Molecular taxonomy of major neuronal classes in the adult mouse forebrain.** *Nat Neurosci* 2006, **9:**99-107.