

Neurogenomics: at the intersection of neurobiology and genome sciences

Mark S Boguski & Allan R Jones

Neurogenomics is the study of how the genome as a whole contributes to the evolution, development, structure and function of the nervous system. It includes investigations of how genome products (transcriptomes and proteomes) vary in time and space. Neurogenomics differs markedly from the application of genome sciences to other systems, particularly in the spatial category, because anatomy and connectivity are paramount to our understanding of function in the nervous system. We focus here on some of the influences of genomics and its associated technologies on neuroscience. We discuss comparative genomics, gene expression atlases of the brain, network genetics and applications to behavioral phenotypes, and consider the culture, organization and funding of genome-scale projects.

In the years immediately before and after the Human Genome Project (HGP) began, the biology community at large viewed the project with considerable skepticism. As Princeton University president Shirley Tilghman recalled¹, there were three main criticisms: first, that the sequence would not be interpretable; second, that generating the sequence would be boring production work; and third, that the 15-year project would divert resources away from the most creative and imaginative way to do science, that is, in small groups consisting of a principal investigator and a number of students and postdoctoral fellows. However, by 1996, when large amounts of transcriptome data from humans and large amounts of genomic sequence from 'model organisms' had become available, most biologists had changed their minds. As Tilghman put it¹, "A genome enthusiast was a genome critic that just got a hit in the EST database"². (Expressed sequence tags, or ESTs, are cDNA subsequences that are generated rapidly to inventory the transcribed components of a genome—mostly protein-encoding mRNAs.)

By 1996, the first large-scale transcript map of the human genome, containing about 15,000 genes, was published³ and made available on the still-nascent World Wide Web, thus setting the stage for the accelerated positional cloning of hundreds of disease genes, including many mutant alleles underlying neurodegenerative and neuropsychiatric disorders. By 1998, the genomic sequence of the first metazoan organism, *Caenorhabditis elegans*, was completed, and the enormous ramp-up to produce the first complete draft of the human genome in the next three years was undertaken⁴. But at that time, eight years into

both the HGP and the 'Decade of the Brain', genomics seemed to have had little impact on neuroscience: a perspective in the journal *Neuron*, marking the end of the first ten years of that journal and the sixth decade of modern neuroscience⁵, mentioned neither the HGP nor the genomics of any other organism (although this article did point out that progress had been made in what might be called the neurogenetics of Huntington and Alzheimer disease). It was not until 2002 that a consortium of three institutes of the US National Institutes of Health (NIMH, NIDA, NINDS) held a meeting to discuss and set priorities for molecular neurobiology in the 'postgenomic' era (<http://www.drugabuse.gov/MeetSum/Postgenomic.html>).

Currently there are nearly 1,000 completed or ongoing genome projects representing an impressive array of species. This creates a wealth of new data for comparative genomics studies and for launching new 'model organisms' for biomedical research. Comparative studies of the human and mouse genomes (with experimental follow-up), for example, are expected to contribute enormously to our understanding of physiology, behavior and psychiatric disease^{6,7}. A draft of the chimpanzee genome has recently become available and has already led to some intriguing observations pertaining to the evolution of speech, hearing and language in the human lineage^{8,9}.

Here we consider some of the "lessons learned and promises kept"¹⁰ of genomics¹⁰, and consider how this knowledge might be applied to neurobiological problems. We describe factors that characterize a genomic approach to a problem, some functional genomics applications to studying gene expression in the brain, as well as comparative genomics and 'network genetics'. We conclude with some important considerations of the culture, organization and funding of genome-scale projects.

Revolutions and evolutions

One of the technological revolutions that made genomics and the HGP possible was the Nobel prize-winning innovation of 'rapid' DNA sequencing¹¹ in 1975. However, it took roughly the next two decades for the evolution and development of this technology (and others, such as the polymerase chain reaction or PCR) to progress to the point when another revolutionary period, the genome era, could begin. This second revolution was largely a matter of the rate of data generation (throughput) and the almost unimaginable scales that could consequently be achieved. Today entire genomes can be rapidly sequenced and interpreted, whereas in 1977 the cloning and sequencing of a single mRNA (rabbit β -globin) merited the cover and three separate papers in an issue of the journal *Cell* (vol. 10, issue 4): one for the 5' region, one for the coding sequence and one for the 3' region. One of the important 'promises kept' by the HGP was the 'freedom to

Mark S. Boguski and Allan R. Jones are at the Allen Institute for Brain Science, 551 N. 34th Street, Seattle, Washington 98103, USA.
e-mail: markbo@alleninstitute.org

Published online 27 April 2004; doi:10.1038/nn1232

do biology'. This meant that biologists would no longer have to spend their time laboriously producing reagents (such as clones and sequences) in their own laboratories. Instead, genomics has created vast databases of basic information and corresponding physical resources that allow investigators to focus on research in their fields.

Functional genomics refers to the development and application of global (genome-wide or system-wide) experimental approaches to assess gene function by making and using genome-scale information and reagents, and is characterized by high-throughput or large-scale experimental methodologies combined with statistical and computational analysis of the results¹². Functional genomics enables scientists to ask entirely new types of questions that require the analysis of large numbers of a system's components simultaneously, for example to enhance the understanding of complex physiological functions¹³.

Strategies for deploying genomics technologies and analyzing the resultant data also had to evolve. For example, it was not obvious that incomplete and inaccurate, but inexpensive and rapidly generated survey data (such as ESTs) could be very useful and greatly expand and accelerate hypothesis-driven, follow-up studies. The practice of systematically collecting data, even before knowing precisely all of the ways in which it might be used, has been referred to as 'discovery-driven' science and has even led to some reconsiderations of biological epistemology^{14,15}. It took some time for researchers pursuing hypothesis-driven versus discovery-driven approaches to realize that they were allies and not antagonists.

Genomics, particularly high-volume DNA sequencing, has proven the value of discovery-driven science. For example, ESTs¹⁶ were initially viewed as kind of an illegitimate offspring of the HGP¹⁷, but they have provided at least six generations of utility over the past dozen years, including gene discovery¹⁸, gene mapping¹⁹, the design and construction of microarrays²⁰, exon detection and alternative splicing analysis²¹ and discovery of an important class of SNPs—single-nucleotide polymorphisms—in coding sequences^{22,23}. EST sequencing also remains a mainstay method for transcriptome analysis. (See applications to the nervous system below.)

Potential perils in the application of genome-scale technologies

When new technologies become available, many investigators are understandably eager to try them out (and journals are eager to publish the results). Examples include comparative sequence analysis^{14,24} in the early-to-mid-1990s and DNA microarrays^{25,26} in the late 1990s. Such periods can be perilous because the limits and optimal use of the technologies may not be well understood, from experimental design to data analysis and interpretation^{27,28}. Reproducibility may even be a question because the expense of and/or limited access to the technology may lead to some experiments being under-controlled, or insufficiently validated by other research groups. Moreover, new statistical methods and analytic tools are often required to properly interpret the results and are usually unavailable or incomplete when the first data appears. Kohane and colleagues²⁷ provide a number of cautionary tales derived from published studies that used comprehensive genomic measurement technologies; misinterpretation and over-interpretation of results are apparently distressingly common. Mirnics and Pevsner²⁹ (p. 434–439 of this issue) describe many of the difficulties in applying microarray technology to brain tissue. Proteomics technologies are now undergoing their early deployments, with associated growing pains²⁸. Both technologies are limited by non-standardized and non-homogeneous starting materials removed from their anatomical context.

Large-scale, high-throughput biology requires an industrial mindset and a different level of management than does a more traditional

research project environment. Strict standardization of protocols and procedures is essential, along with rigorous definition of and adherence to quality control and quality assurance standards. Although these principles are applicable to large-scale projects in any field, they may be particularly challenging when applied to neurobiology because of the anatomical and physiological complexity of many of the experimental systems.

As a final caution, discovery-driven projects usually generate prodigious amounts of data. However, as Geoffrey Duyk of TPG Ventures puts it, confusing throughput with output often blurs the distinction between data and knowledge²⁵. For example, it is very unlikely that generating large repositories of mass spectrometry, proteomics data will yield the same value as high-volume DNA sequencing²⁸.

Comparative (evolutionary) genomics and genome annotation

The field of comparative genomics was born from the comprehensive and systematic cross-species comparisons of gene products³⁰ with the underlying assumption that aspects of function can be inferred from evolutionary conservation of DNA sequence. Such studies led to powerful new connections for investigating the biochemistry of human disease genes, even in model organisms as simple as yeast³¹. However, the recent completion of the mouse genome portends an unprecedented access to highly informative models of human somatic and psychological diseases^{6,7}.

With 'excess' sequencing capacity in the wake of the HGP, the National Human Genome Research Institute (NHGRI) and its advisors devised and implemented a process for selecting and sequencing additional genomes. This served to enhance the interpretation of the human genome by comparative genomic analysis, to fill in major phylogenetic gaps in our sampling of extant genomes in the biosphere, and to create new model organisms. This program, dubbed GRASP (for Genome Resources and Sequencing Priorities), has been operating since early 2002 and reviews white-paper proposals (www.genome.gov/page.cfm?pageID=10002154) from the community. Among the organisms selected to date, the honeybee *Apis mellifera* was chosen, in part, for its ability to learn and engage in complex social behaviors: a complete knowledge of the genome is expected to lead to rapid advances in honeybee genetics and its application to the analysis of learning and behavioral phenotypes (www.genome.gov/11509819).

Publications describing the human genome sequence included extensive bioinformatics-derived annotations of various features, including genes and gene products, often with only inferred or hypothetical functions: it was quickly noted that experimental validation of these computational predictions would be valuable³². Large-scale anatomical and developmental transcriptome mapping, combined with comparative genomics analysis, has the potential to both substantiate hypothetical (predicted) genes and lead to a more focused pursuit of their functions.

Transcriptional profiling of the mammalian brain

The term 'transcriptome' was coined in 1995 by Velculescu and colleagues³³ in the context of applying the technique of serial analysis of gene expression (SAGE)³⁴ to the discovery, description and cataloging of mRNA transcripts in yeast. However, transcriptome analysis has a long and interesting history of applications in neuroscience. The greater transcriptional complexity of the brain compared with other organs was first addressed by hybridization analysis in the 1970s, leading to the claim that one-third of the mammalian genome was exclusively dedicated to brain function^{35,36}. About a decade later, it was possible to summarize the structural properties of 39 'brain-specific'

mRNAs³⁷, although it was still necessary to remind readers that “the brain is made of proteins” and that “proteins are encoded by mRNAs”.

The first ‘genome era’ approach to this question appeared in 1992, when one of the earliest applications of EST sequencing identified 2,375 genes expressed in the human brain¹⁸. This EST survey approach later formed the basis of the NIMH-NINDS ‘Brain Molecular Anatomy Project’ (BMAP, www.nimh.nih.gov/grants/0006-cbd1.cfm), and EST surveys continue to be used to study neural tissue transcriptomes. Both SAGE and microarray technologies have been more recently applied to transcriptional profiling of single cells in retina³⁸ and neuronal progenitors³⁹, respectively. However, gene expression profiling in the central nervous system may culminate in detailed anatomical atlases of gene expression in the developing and adult brain^{40–42}.

According to Cowan⁴³, new methods for anatomical analysis that were first developed in the 1970s led to a renaissance in morphologic studies of the nervous system. Gene expression profiling of cell populations in the brain has the potential to lead to another quantum advance characterized by a new understanding of the number of and relationships among cell types as defined by what portion of the genome is uniquely expressed in them. New insights into function may be gained by analyzing ensembles of transcriptional activities. The full potential of this approach, however, will only be realized by genome-scale, discovery-driven research.

One genome-scale approach to studying the neuroanatomy of gene expression is micro-dissection of particular brain structures followed by mRNA isolation from cells and analysis on microarrays⁴⁴. This method is, in principle, quantitative, but anatomical resolution and fidelity are limited, and confusion may arise from the combinations of different cell types in a dissected specimen.

Another approach is high-throughput *in situ* hybridization analysis⁴⁰. This method can ideally deliver cellular resolution and anatomical fidelity. A major new effort, based on this approach, to create an expression atlas of the adult mouse brain was recently announced⁴⁵ (see www.brainatlas.org).

The NINDS GENSAT project is creating a gene expression atlas of the mouse CNS based upon individual genes that have been inserted into the mouse genome using bacterial artificial chromosomes (BACs)⁴¹. This method is relatively expensive and time-consuming, but yields transgenic mice that may be used for follow-on studies. Other BAC-based technologies are also in use⁴⁶.

All these approaches have their strengths and weaknesses. Sensitivity, specificity, reproducibility, quantifiability, throughput and cost are all important factors. Although substantial challenges will have to be met, the ultimate goals of anatomical fidelity, cellular resolution and connectivity seem increasingly achievable through the application of genomics approaches and technologies.

‘Network genetics’ and neurobehavioral phenotypes

Comparative genomics is ideally capable of elucidating interspecies similarities and differences in genome structure and function, but it is intraspecies variation (polymorphism) among genes that underlies critical differences in phenotypes such as susceptibility to disease. The merging of genomics and statistical genetics^{47,48} may lead to a revolution in our ability to decipher the genetic basis of complex phenotypes, including common, polygenic diseases and complex behavior. Not only can we now examine variation on a genome scale through high-throughput genotyping technologies, but we can also use the outputs of functional genomics technologies (such as microarray-based transcriptional profiles²⁹) as quantitative phenotypes or traits to assess genotype–phenotype correlations. The power of these

approaches, combined with new strategies for high-throughput behavioral phenotyping⁴⁹, could lead to major advances in our understanding of the molecular-genetic basis of complex behavior and neuropsychiatric disease.

Gene expression profiles observed in segregating populations have refined clinical traits into subtypes that are under the control of different genetic loci⁴⁷. Similarly, gene expression patterns in recombinant inbred strains of mice correlate with neurobehavioral phenotypes⁴⁸. Although this approach might be applied to the analysis of any complex trait, it is particularly exciting to imagine that human psychiatric disorders could be accurately diagnosed by biochemical means and broken down into various sub-phenotypes that have different contributing or causative genes. This prospect has enormous implications for the more effective use of existing therapeutics as well as the identification of new drug targets and development of novel therapies.

The culture, organization and funding of genome-scale projects

A recently published study from the Institute of Medicine and the National Research Council explores a new paradigm of biomedical research: hypothesis-generating, discovery-driven science enabled by technology advances that allow for high-throughput data collection and analysis⁵⁰. This new paradigm—of which the HGP is the most obvious but far from only example—has raised many questions about how such projects should be initiated, funded, organized, managed, staffed and evaluated.

In the wake of the HGP, various NIH institutes struggled with how to support research that depends on new high-throughput or large-scale and expensive technologies. A case in point is the National Heart, Lung and Blood Institute (NHLBI), which approached this challenge with the concept of Programs in Genomic Applications, or PGAs. The goals of this program are to develop information, tools and resources to link genes to biological function on a genomic scale and to provide workshops, courses and visiting scientist programs to facilitate the training of researchers in the use of the data and related technologies developed by the PGAs⁵¹. NIH neuroscience institutes have supported some genome-wide initiatives, such as large-scale mutagenesis and phenotypic screening⁵² and national microarray centers (<http://arrayconsortium.cnmcresearch.org/NINDS/home.do>).

In these examples, the NIH provides essentially all of the support. But there is another model in which a consortium of government and industry organizations provides important resources to the community. In the past several years, NHGRI has initiated and led consortia such as the International Haplotype Map (HapMap) Project⁵³ and the SNP (single nucleotide polymorphism) Consortium⁵⁴ to create maps and databases of human genetic variation to accelerate the study of genes associated with complex, polygenic diseases.

Historically, this trend began in 1993 with a project called the Merck Gene Index^{55,56}, which pioneered the concept of government–industry collaboration to provide, for the community, a genome-scale resource of pre-commercial (‘precompetitive’) biological data. In this case, Merck & Co. provided financial support for random, high-throughput EST sequencing of the human, and subsequently, the mouse, transcriptomes at the Washington University Genome Sequencing Center, which then submitted the data to the new dbEST division² of the NIH-sponsored GenBank database for unrestricted public access. The physical cDNA clones corresponding to the EST sequences were also available for unrestricted use⁵⁶.

Some of the technological but also sociological and funding challenges of neuroscience in the ‘information age’ are coming to the forefront^{57–59}. For example, the sharing of primary data, even prior

to publication, is now the norm in genomics and functional genomics studies, but there is apparently a reluctance to adopt this practice in the neuroscience community⁶⁰. Some outmoded aspects of academic culture and reward systems with respect to discovery-driven science are significant obstacles^{57,60}. However, genome-scale projects are too resource-intensive to justify duplication in multiple laboratories. Public and even commercial support of such projects will favor data access policies that promote the broadest possible uses of the data.

Conclusions and future prospects

The historical trends toward large-scale research projects and government-industry collaborations have culminated in a new 'NIH Roadmap' that will result in new models for interdisciplinary team science and novel public-private partnerships⁶¹. It is important to begin thinking about how this trend can advance the neurosciences, both basic and applied. The time seems ripe for a new and vigorous exploration of the frontier between neurobiology and genome sciences.

What might be some of the paths forward? The first, and perhaps the most obvious, will be the application of functional genomics technologies to existing model systems for obtaining more global views of gene expression in the nervous system. Transcriptome surveys and analyses and large-scale neuroanatomical mapping of gene expression are already underway and may define a new functional anatomy of the brain. Major advances in the diagnosis and treatment of psychiatric diseases could result if brain gene expression patterns could be developed into quantitative traits and correlated with genetic variation and neurobehavioral phenotypes.

Other advances will come from the application of comparative genomics studies focused on neurobiology and behavioral genetics. The development of new model organisms (such as the honeybee and zebrafish⁶²) for neuroscience research is being accelerated by extensive genome sequencing and the application of comparative and functional genomics technologies.

Soon perhaps, neurogenomics may be able to begin filling in what has been called 'the gene gap'⁶³—that is, the curious lack of correspondence between genome size and organismal and behavioral complexity. It remains a mystery, for example, why human beings, with only about 20% more genes than a soil nematode (24,000 versus 19,000 genes), can develop a brain with 100 billion neurons and 10¹⁵ connections, whereas *C. elegans* has only 302 neurons with about 7,000 synapses⁶⁴. Neurogenomics could perhaps even serve as a new bridge between reductionist and more holistic approaches to the analysis of the nervous system and lead to a new kind of systems neuroscience akin to the re-conceptualization of 'systems biology' in the genome era.

ACKNOWLEDGMENTS

We thank N. Beauchamp, S. Minoshima and R. Williams for critical readings of the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at <http://www.nature.com/natureneuroscience/>

1. Tilghman, S.M. Lessons learned, promises kept: a biologist's eye view of the Genome Project. *Genome Res.* **6**, 773–780 (1996).
2. Boguski, M.S., Lowe, T.M. & Tolstoshev, C.M. dbEST—database for "expressed sequence tags". *Nat. Genet.* **4**, 332–333 (1993).
3. Schuler, G.D. *et al.* A gene map of the human genome. *Science* **274**, 540–546 (1996).
4. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
5. Jessell, T. & Kandel, E. Introduction: one decade of *Neuron*, six decades of neuroscience. *Neuron* **20**, 367–369 (1998).

6. Tecott, L.H. The genes and brains of mice and men. *Am. J. Psychiatry* **160**, 646–656 (2003).
7. Boguski, M.S. Comparative genomics: the mouse that roared. *Nature* **420**, 515–516 (2002).
8. Enard, W. *et al.* Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature* **418**, 869–872 (2002).
9. Clark, A.G. *et al.* Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* **302**, 1960–1963 (2003).
10. Boguski, M.S. The only thing permanent is change. *Genomics* **82**, 253 (2003).
11. Sanger, F. & Coulson, A.R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **94**, 441–448 (1975).
12. Hieter, P. & Boguski, M. Functional genomics: it's all how you read it. *Science* **278**, 601–602 (1997).
13. Liang, M., Cowley, A.W. & Greene, A.S. High throughput gene expression profiling: a molecular approach to integrative physiology. *J. Physiol. (Lond.)* **554**, 22–30 (2004).
14. Boguski, M.S. Biosequence exegesis. *Science* **286**, 453–455 (1999).
15. Kell, D.B. & Oliver, S.G. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *Bioessays* **26**, 99–105 (2004).
16. Adams, M.D. *et al.* Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**, 1651–1656 (1991).
17. Zweiger, G. *Transducing the Genome: Information, Anarchy, and Revolution in the Biomedical Sciences* (McGraw-Hill, New York, 2001).
18. Adams, M.D. *et al.* Sequence identification of 2,375 human brain genes. *Nature* **355**, 632–634 (1992).
19. Boguski, M.S. & Schuler, G.D. ESTablishing a human transcript map. *Nat. Genet.* **10**, 369–371 (1995).
20. Ermolaeva, O. *et al.* Data management and analysis for gene expression arrays. *Nat. Genet.* **20**, 19–23 (1998).
21. Mathe, C., Sagot, M.F., Schiex, T. & Rouze, P. Current methods of gene prediction, their strengths and weaknesses. *Nucleic Acids Res.* **30**, 4103–4117 (2002).
22. Picoult-Newberg, L. *et al.* Mining SNPs from EST databases. *Genome Res.* **9**, 167–174 (1999).
23. Gu, Z., Hillier, L. & Kwok, P.Y. Single nucleotide polymorphism hunting in cyberspace. *Hum. Mutat.* **12**, 221–225 (1998).
24. Boguski, M.S. Hunting for genes in computer data bases. *N. Engl. J. Med.* **333**, 645–647 (1995).
25. Duyk, G.M. Sharper tools and simpler methods. *Nat. Genet.* **32** (Suppl.), 465–468 (2002).
26. Anonymous. Taking stock. *Nat. Genet.* **32**, 461 (2002).
27. Kohane, I.S., Kho, A.T. & Butte, A.J. *Microarrays for an Integrative Genomics* (MIT Press, Cambridge, Massachusetts, USA, 2003).
28. Boguski, M.S. & McIntosh, M.W. Biomedical informatics for proteomics. *Nature* **422**, 233–237 (2003).
29. Mirnics, K. & Pevsner, J. Progress in the use of microarray technology to study the neurobiology of disease. *Nat. Neurosci.* **434–439** (2004).
30. Bassett, D.E., Jr *et al.* Comparative genomics, genome cross-referencing and XREFdb. *Trends Genet.* **11**, 372–373 (1995).
31. Tugendreich, S., Bassett, D.E., Jr., McKusick, V.A., Boguski, M.S. & Hieter, P. Genes conserved in yeast and humans. *Hum. Mol. Genet.* **3** (Spec. No.), 1509–1517 (1994).
32. Shoemaker, D.D. *et al.* Experimental annotation of the human genome using microarray technology. *Nature* **409**, 922–927 (2001).
33. Velculescu, V.E. *et al.* Characterization of the yeast transcriptome. *Cell* **88**, 243–251 (1997).
34. Velculescu, V.E., Zhang, L., Vogelstein, B. & Kinzler, K.W. Serial analysis of gene expression. *Science* **270**, 484–487 (1995).
35. Bantle, J.A. & Hahn, W.E. Complexity and characterization of polyadenylated RNA in the mouse brain. *Cell* **8**, 139–150 (1976).
36. Chikaraishi, D.M. Complexity of cytoplasmic polyadenylated and nonpolyadenylated rat brain ribonucleic acids. *Biochemistry* **18**, 3249–3256 (1979).
37. Sutcliffe, J.G. mRNA in the mammalian central nervous system. *Annu. Rev. Neurosci.* **11**, 157–198 (1988).
38. Blackshaw, S., Fraioli, R.E., Furukawa, T. & Cepko, C.L. Comprehensive analysis of photoreceptor gene expression and the identification of candidate retinal disease genes. *Cell* **107**, 579–589 (2001).
39. Tietjen, I. *et al.* Single-cell transcriptional analysis of neuronal progenitors. *Neuron* **38**, 161–175 (2003).
40. Carson, J.P., Thaller, C. & Eichele, G. A transcriptome atlas of the mouse brain at cellular resolution. *Curr. Opin. Neurobiol.* **12**, 562–565 (2002).
41. Gong, S. *et al.* A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* **425**, 917–925 (2003).
42. Singh, R.P. *et al.* High-resolution voxelation mapping of human and rodent brain gene expression. *J. Neurosci. Methods* **125**, 93–101 (2003).
43. Cowan, W.M. The emergence of modern neuroanatomy and developmental neurobiology. *Neuron* **20**, 413–426 (1998).
44. Lockhart, D.J. & Barlow, C. Expressing what's on your mind: DNA arrays and the brain. *Nat. Rev. Neurosci.* **2**, 63–68 (2001).
45. Anonymous. Mapping the mouse brain. *Nat. Neurosci.* **6**, 1113 (2003).
46. Valenzuela, D.M. *et al.* High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nat. Biotechnol.* **21**, 652–659 (2003).
47. Schadt, E.E., Monks, S.A. & Friend, S.H. A new paradigm for drug discovery: inte-

- grating clinical, genetic, genomic and molecular phenotype data to identify drug targets. *Biochem. Soc. Trans.* **31**, 437–443 (2003).
48. Chesler, E.J. *et al.* Genetic correlates of gene expression in recombinant inbred strains: a relational model system to explore neurobehavioral phenotypes. *Neuroinformatics* **1**, 343–357 (2003).
 49. Tecott, L.H. & Nestler, E.J. Neurobehavioral assessment in the information age. *Nat. Neurosci.* 462–466 (2004).
 50. Nass, S.J. & Stillman, B. National Cancer Policy Board (U.S.). Committee on Large-scale Science and Cancer Research. & National Research Council (U.S.). Division on Earth and Life Studies. *Large-Scale Biomedical Science: Exploring Strategies for Future Research* (National Academies Press, Washington, DC, USA, 2003).
 51. Winslow, R.L. & Boguski, M.S. Genome informatics: current status and future prospects. *Circ. Res.* **92**, 953–961 (2003).
 52. Moldin, S.O., Farmer, M.E., Chin, H.R. & Battey, J.F., Jr. Trans-NIH neuroscience initiatives on mouse phenotyping and mutagenesis. *Mamm. Genome* **12**, 575–581 (2001).
 53. International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
 54. Holden, A.L. The SNP consortium: summary of a private consortium effort to develop an applied map of the human genome. *Biotechniques* **26** (Suppl.), 22–24 (2002).
 55. Anonymous. Merck releases first 'gene index' sequences. *Nature* **373**, 549 (1995).
 56. Williamson, A.R. The Merck Gene Index project. *Drug Discov. Today* **4**, 115–122 (1999).
 57. Insel, T.R., Volkow, N.D., Li, T.K., Battey, J.F. & Landis, S.C. Neuroscience networks: data-sharing in an information age. *PLoS Biol* **1**, E17 (2003).
 58. Martone, M.E., Gupta, A. & Ellisman, M.H. e-Neuroscience: challenges and triumphs in integrating distributed data from molecules to brains. *Nat. Neurosci.* 467–472 (2004).
 59. Van Horn, J.D., Grafton, S.T., Rockmore, D. & Gazzaniga, M.S. Sharing neuroimaging studies of human cognition. *Nat. Neurosci.* 473–481 (2004).
 60. Koslow, S.H. Sharing primary data: a threat or asset to discovery? *Nat. Rev. Neurosci.* **3**, 311–313 (2002).
 61. Zerhouni, E. The NIH Roadmap. *Science* **302**, 63–72 (2003).
 62. Neuhauss, S.C. Behavioral genetic approaches to visual system development and function in zebrafish. *J. Neurobiol.* **54**, 148–160 (2003).
 63. Marcus, G.F. *The Birth of the Mind: How a Tiny Number of Genes Creates the Complexities of Human Thought* (Basic Books, New York, 2004).
 64. White, J., Southgate, E., Thomson, J. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **314**, 1–340 (1986).