8

An Integrative Neuroscience Program Linking Mouse Genes to Cognition and Disease

Seth G. N. Grant

The year 2001 marked the arrival of almost complete genome sequence for four organisms with nervous systems, which are also the subject of intense genetic studies: humans, mice, flies, and worms. In the same way that the advent of molecular biology or electrophysiology opened grand new insights into the function of the nervous system, the genome era offers yet another exciting platform for discovery. As with any new set of tools, we want to see them result in improved forms of health care and in the discovery of new biological processes. This chapter does not aim to predict either the future of this "postgenomic revolution" or all of the ways genome information can be applied to understanding behavior. Here I outline a structure for a program of study, referred to as the Genes to Cognition program (G2C), which takes advantage of genomic information and combines it with a diverse set of methodologies. The G2C is driven by studies in basic genetic organisms with the aim of using this information to understand mechanisms of behavior and diseases of the human nervous system.

The potential of the G2C is illustrated using the biology of learning and memory. Learning is a fundamental cognitive process that has been at the center of mechanistic studies of neural function for more than a century. In the past decade, studies in rodents have led to the identification of a large number of genes involved with learning, which far outstrips those known in any other area of cognitive science. This knowledge can now be applied to humans, where it is likely to be relevant to the pathologies of learning impairment in children, dementias, schizophrenia, and brain injury.

By the very nature of the quest to link genes with behavior, it is necessary to construct a broadly integrative program that encompasses many distinct methods apart from genetics and psychology. These other areas include cell biology, electrophysiology, biochemistry, proteomics, microar-

I thank the following colleagues and collaborators for stimulating discussions and input: David Bentley, Douglas Blackwood, Robert Plomin, and Jane Rogers.

rays, brain imaging, and more. This raises a new and fascinating problem —of how to construct information networks that facilitate linking of these areas within a framework that not only provides rapid and simple access to information but also leads to the generation of new hypotheses and insights.

Despite the logic of constructing a framework linking large datasets ranging from molecular biology to psychology, and the inevitability of information accumulating and being organized in this manner, there needs to be a more purposeful drive to this goal. A potentially powerful focus of organization is based on the general recognition that biological functions are performed by sets of proteins (or genes) working together in pathways or as macromolecular machines (Alberts, 1998; Brent, 2000; Tjian, 1995). This logic can be adapted and applied to the study of behavior. Below are some guiding principles that underpin the G2C strategy:

1. Basic aspects of behavior and brain function are evolutionarily conserved.
2. Core biochemical processes (e.g., signaling pathways and protein complexes) conduct these functions.
3. Sets of genes encoding the core processes can be defined.
4. Core processes are disrupted by mutations and produce phenotypes in humans.
5. Multiple mutations or alleles cooperate to disrupt the core processes.
6. Behaviors and some psychiatric diseases are polygenic; this nature may reflect multiple genes encoding core processes.

The following sections outline aspects of a general strategy for the assembly of a knowledge base that encompasses a biological spectrum from gene to behavior.

General Strategy and Outline of the G2C

A key feature of this strategy is the linking of mouse and human genetics. Genes involved with learning can be discovered in mice (Grant et al., 1992; Grant & Silva, 1994; Silva, Paylor, Wehner, & Tonegawa, 1992). Because of the homology between mice and humans at the levels of the gene, the protein, the synapse, the brain region, and the behavior, there is a high probability that a gene involved with learning in a mouse will also be important in humans. This notion has been substantially supported by many studies.

An alternative approach to finding the genes involved with learning, or other behaviors, is to score for variation in the phenotype between individuals and then seek the genetic differences that correlate with these changes (see Plomin, chapter 11, this volume). This approach has been widely used and particularly successfully for finding large-effect genes underpinning some disorders (e.g., Huntington's disease).

Gene-targeting technology and the use of embryonic stem (ES) cells allow the experimentalist to modify the structure of any given gene or chromosome in the mouse in a controlled manner (Bradley, Zheng, & Liu, 1998). The widespread use of this technology has led to many hundreds of mouse genes being disrupted or modified. These mutant mice are routinely examined in a wide variety of assays, many of which are aimed at exploring the dysfunction of the nervous system. In this way, lists of genes are being developed in which the named genes are known to be required for the normal physiological function in question. These lists can be used to design experiments in humans, in which one asks, Does the behavioral abnormality in humans correspond to an altered structure of the corresponding human gene? To illustrate how this might work, let us consider the molecular mechanisms of learning as revealed by studies in the mouse and ask if this is informative for studies in humans. First, I provide a brief overview of the history of the molecular biology of learning.

Connecting the Molecular Mechanisms of Learning Between Mouse and Human

The recognition that sensory information is encoded in patterns of action potentials and transmitted into the brain (Adrian, 1928) led to the prediction that there must be some "metabolic" mechanism in neurons that is capable of detecting specific patterns of activity and converting them into some "structural" changes (Hebb, 1949). This hypothesis was made experimentally tractable when electrophysiologists found that synapses from the hippocampus, a region of the brain involved with learning, could be stimulated with different patterns of action potentials, and these patterns would induce increases or decreases in the efficiency of communication between two neurons (Bliss & Lomo, 1973). This system allowed pharmacological studies (Collingridge, Kehl, & McLennan, 1983) and mouse genetic studies to be used to identify molecules previously unknown in this process of synaptic plasticity (Grant et al., 1992). By using these methods, a large body of data accumulated during the 1990s, which essentially implicated in excess of 100 proteins in this biology without providing any unifying scheme or molecular hypothesis (Sanes & Lichtman, 1999).

Within this dataset, it was well established that a receptor-ion channel, known as the *N*-methyl-D-aspartate receptor (NR), was an essential component. The NR is a receptor for the excitatory neurotransmitter glutamate and on activation allows Ca^{2+} influx via its central pore. As is inappropriately illustrated in many textbooks and reviews, it would appear that this receptor simply sits in the membrane at the postsynaptic side of the synapse, where it injects Ca^{2+} into the dendrite, which then diffuses to activate a variety of enzymes that seem to float freely in the cytoplasm. These enzymes then drive various poorly understood signaling pathways that control neuronal properties. The first evidence that the NR and signaling proteins do not function in this way came when transgenic

mice carrying a mutation in the Post Synaptic Density 95 protein (PSD-95), which normally binds the NR, were found to produce striking changes in the properties of synaptic plasticity and learning (Migaud et al., 1998). This work predicted that there are multiprotein signaling complexes comprised of NR and PSD-95 with other proteins, which control learning.

This genetic evidence for a multiprotein complex was used to justify a proteomic analysis: biochemical isolation of the protein complexes from brain and identification of proteins using mass spectrometry and immunoblotting (Husi & Grant, 2001a; Husi, Ward, Choudhary, Blackstock, & Grant, 2000). These methods, which have general applicability to other receptor complexes (Husi & Grant, 2001b), showed that the NR PSD-95 complexes were approximately 2,000–3,000 kDa, which is several-fold more than would be expected if it was simply the NR subunits alone. A picture emerged of 75 or more proteins that could be broadly categorized into five classes: neurotransmitter receptors, cell adhesion molecules, adaptors, signaling enzymes, and cytoskeletal proteins (for more details, see Husi et al., 2000). A major surprise in this study was that at least 27 proteins from the complexes were known to be required for synaptic plasticity and 18 for learning in rodents and were from each of the five classes of complex components. Thus the organization of these proteins into these multiprotein complexes suggests that they work together in a large "machine," not unlike many other multiprotein molecular machines. The importance of this concept is that it removes the focus of interest away from the individual molecules onto the function of the overall machine. My colleagues and I (Migaud et al., 1998) have proposed that these complexes are a "device" for detecting patterns of synaptic activity and for converting this information into intracellular signals that store the information in the cell. In this way, electrical information can be translated into cellular memory.

These properties were at the basis of Hebb's postulate, and these complexes have been described as *Hebbosomes,* multiprotein complexes that convert patterns of neuronal activity into cellular changes underlying learning. It is beyond the scope of this chapter to broadly discuss Hebbosomes, except to indicate that there are families of such complexes, with different molecular composition, which confer specific properties to different synapses.

The characterization of Hebbosomes has significant implications for human genetics. Three genes previously known in humans to be involved with cognitive deficits were also found to encode proteins found in the complexes. These include two signal transduction enzymes: neurofibromin (also known as NF-1 and mutant in the neurofibromatosis syndrome) and RSK-2 (mutant in the Coffin Lowry syndrome) and the adhesion protein L1 (mutant in CRASH syndrome). These observations open the exciting possibility that other human cognitive disorders that have a genetic component may involve genes encoding the proteins in these complexes. In this way, the named genes from the mouse studies can be used as candidate genes in a human association study.

In the simplest setting, knowing that a mouse gene is important for

behavior is a reasonable starting point for a human study. There are a number of potential weaknesses in this setting. For example, the human gene may not be as important to the human as it is to the mouse. A stronger starting point is not to rely on a single gene but to use the information about that gene to build up a set of genes. As described in the example above, one could consider that PSD-95 was a starting point, because the mouse knock out had severe learning deficits (Migaud et al., 1998). By isolating the PSD-95 containing complexes and using proteomic tools, it became clear that at least 75 proteins could now be considered candidates for association studies. Thus a nongenetic strategy, such as proteomics, can be used in conjunction with the genetics to identify molecules involved with learning.

A Multilayer Organization

The G2C can be organized into four layers (see Figure 8.1). These layers are briefly summarized here and discussed in more detail later.

The entry point for this strategy (Layer 1) is molecular information derived from basic science studies. Strong emphasis is placed on the value of genetically modifiable organisms with nervous systems (invertebrates: fruit fly, *Drosophila*; worm, *Caenorhabditis elegans*; vertebrates: mouse, *Mus musculus*; zebra fish, *Danio rerio*). Through the use of genetic screens and mutations, these organisms have generated lists of proteins that are involved with various phenotypes. Compiling the set of genes that are involved in a common phenotype (e.g., learning) or involved in a multiprotein complex, or some other ways of classifying sets, produces useful information for a human genotyping study. A prototype for this set is that derived from the molecular studies of the multiprotein complexes (Hebbosomes) underlying acquisition of learning (Husi et al., 2000).

Layer 2 of the G2C takes forward the candidate genes from Layer 1 into human genotyping. Using genome sequencing technology, human single-nucleotide polymorphisms (SNPs) can be determined for all genes in the set and DNAs from relevant humans genotyped. Given the rapid pace of the SNP identification and characterization, information covering the first phase of this should be available in the public domain in the very near future.

Layer 3 of the G2C is aimed at validating the biological significance of variant alleles found in humans. Here functional assays are required, and mouse ES cell technology again is used to provide several complementary in vivo and in vitro approaches. One could assemble a wide range of molecular and neuroscience methods in a highly integrative research program. These neurobiological studies can be linked to human neurobiological studies, thus providing a broad framework of connections at many levels of analysis.

There will be an important role for informatics at all stages of the G2C, and Layer 4 is the platform for this technology. This will include access to existing databases as well as generating new databases. These

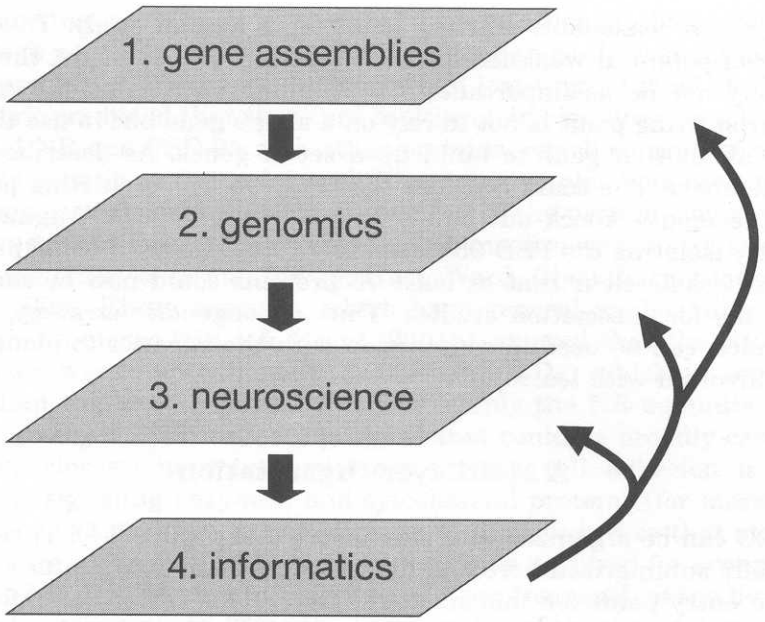

Figure 8.1. Overview of organizational layers. The four main layers of the Genes to Cognition program (G2C) are shown as flat planes. The entry point to the program is primarily via Layer 1 (see Figure 8.2 for detail). In Layer 1, a set of genes is defined according to various criteria and driven by basic neuroscience studies. In Layer 2 (see Figure 8.3 for detail), the genes in Layer 1 are used in human genotyping assays to seek putative functional variants. These variants are examined in biological assays in Layer 3 (see Figure 8.4 for detail). Underpinning and central to the integration of the data in all layers is bioinformatic tools (Layer 4). This information is transferred between layers and is used to modify the experiments as the G2C develops. The arrows indicate the simplest flow of information between layers.

databases and links should generate a novel and valuable resource for the scientific and medical community.

Layer 1: Identification of Genes Encoding Assemblies

Sets of genes are defined using several sources of information (see Figure 8.2). This layer requires bioinformatics and expertise of scientists within the area of basic biology. Types of molecular information that will be used to select genes include the following: (a) mutant phenotypes of mice and other genetic organisms, (b) knowledge of molecular pathways, (c) protein interaction networks obtained from proteomic and yeast 2-hybrid screens, and (d) gene families, chromosomal organization, and syntenic regions between human and mouse. This prioritization of genes will provide the information for Layer 2.

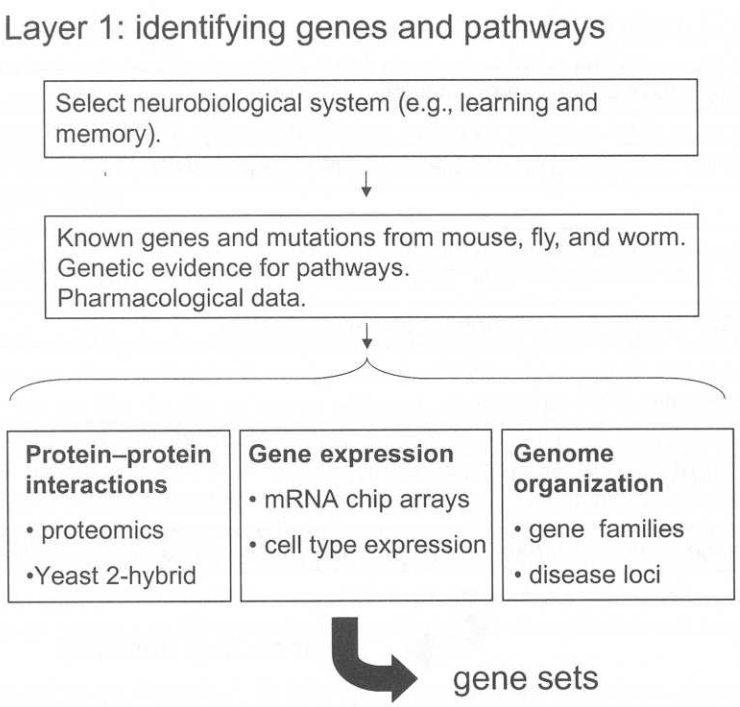

Figure 8.2. Schematic representation of Layer 1. Layer 1 of the Genes to Cognition program (G2C) is aimed at defining sets of genes encoding "assemblies" or functional sets of genes. The first step is choice of neurobiological problem for which there is considerable molecular data from basic neuroscience experiments, such as learning and memory. Using all available molecular data, with an emphasis on the value of genetic organisms, as well as other data (illustrated in boxes), sets of candidate genes for future human studies are chosen. These genes are used in Layer 2.

Layer 2: Genomics

The overall goal of Layer 2 is to identify variant structures in specific human genes, which are candidates for detailed functional testing (see Figure 8.3). The basic gene structure for those loci that have been selected and prioritized according to Layer 1 of the G2C will be determined for human and mouse using available finished sequence from the various genome sequencing projects (see www.sanger.ac.uk for information on genome projects). The comparative gene structure of mouse and human serves several purposes. First, it provides a basis for comparing gene structure and assigning intron/exon and other regulatory features to the sequence (Wiehe, Guigo, & Miller, 2000). This information is useful in designing genotyping strategies, including those involving SNP detection. A second reason for obtaining mouse sequence is that in Layer 3 of the G2C this information is useful as a guide for construction of gene-targeting vectors for engineering specific mutations into the mouse.

A major collaborative international effort is under way to identify

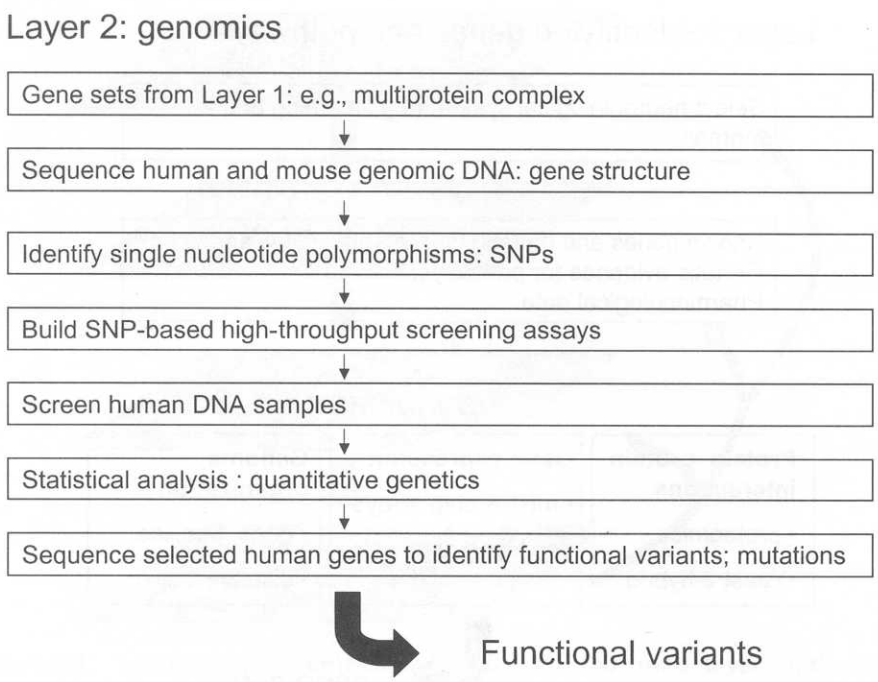

Figure 8.3. Schematic representation of Layer 2. Layer 2 of the Genes to Cognition program (G2C) sets out a scheme for screening DNA from humans for polymorphisms in genes identified in Layer 1. Details relevant to the individual boxes are described in the body of the text. The output from this layer is information on putative functional variants that can be tested in Layer 3 (see Figure 8.4).

SNPs in the human genome. This SNP Consortium (Altshuler et al., 2000; Isaksson et al., 2000; Masood, 1999; Mullikin et al., 2000; Sachidanandam et al., 2001) aims to generate sufficient numbers of SNPs that can then be used in high-throughput genotyping assays (Fors, Lieder, Vavra, & Kwiatkowski, 2000; Kokoris et al., 2000; Kwok, 2000). Statistical analysis (Bader, 2001; Niu, Struk, & Lindpaintner, 2001) of SNP frequency in populations is used to implicate a gene in the phenotype relevant to the human DNAs. The identification of statistical association will motivate resequencing of the alleles in affected individuals to identify potential functional variants. Sequence information may predict the nature of the functional impairment, such as premature termination codons, and these putative functional variants will be tested in Layer 3.

There are potentially interesting features of a genotyping strategy based on genes encoding proteins known to be components of pathways. As has been shown in model genetic organisms, construction of compound mutations allows one to examine the functional relationship between the two genes. *Epistatic* interactions between genes (traditionally defined as the presence of one allele at one locus preventing the expression of an allele at a different locus) is a feature of genes encoding proteins in common pathways. By extension, it may be that some diseases manifest symp-

toms only if a pathway is debilitated, and this may require the presence of two affected genes. Thus, statistical analysis of the set of genes in Layer 1 may show that sets of SNPs identifying particular variant genes will detect these genes. The effects of these variant genes alone may not give statistical significance association with the disease, although the subsets of genes may do so.

Layer 3: Functional Genomics–Experimental Neuroscience

An output from Layer 2 will be variants in the sequence of a human gene. In addition to the statistical analysis used to make a case that a variant gene may be at the basis of some altered phenotype in humans, there is a need to generate biological data showing this variant has function consequences. The simplest way forward may be to use some kind of specific in vitro assay that is sensitive to the function of the protein involved. The G2C includes this aspect; however, it proposes to use a wider, integrative program of study where the variant is tested in sets of assays relevant to the cells on one hand and the cognitive processes on the other—in other words, many assays at the molecular, cellular, and animal level (see Figure 8.4).

Studying gene function in the nervous system requires general tools applicable to neurons and glia. This is in contrast to some areas of cell biology, such as DNA replication or growth control, which can be studied in generic cells. Moreover, in the context of heritable differences in gene structure and the implications for behavior, it is ultimately necessary to study the gene in the context of the whole animal. Gene targeting in mouse provides an ideal way to bridge the gap between cell biology in cultured neurons and biology of the whole animal. This is because of the pluripotential nature of ES cells and thus the ability to derive cells and animals from the same genetically modified cell. Layer 3 outlines some of the applications of mouse gene targeting and the analysis of the mice.

Gene targeting in mouse ES cells (Box 2) is ideally suited for studies of human gene function in complex organs such as the brain because almost any type of gene or chromosomal engineering is feasible in ES cells. The following are some of the relevant technologies:

1. Gene knock out (complete disruption of expression; Cheah & Behringer, 2000; DeChiara, 2001), including gene traps (Brown & Nolan, 1998; Medico, Gambarotta, Gentile, Comoglio, & Soriano, 2001).
2. Point mutation and other fine mutations (Brown & Nolan, 1998); this may be particularly useful for introducing SNPs into mouse genes.
3. Larger sequence modification, including "humanization" or substitution of human wild-type or mutant genes for mouse genes; this uses techniques of chromosomal engineering (Mills & Bradley, 2001).

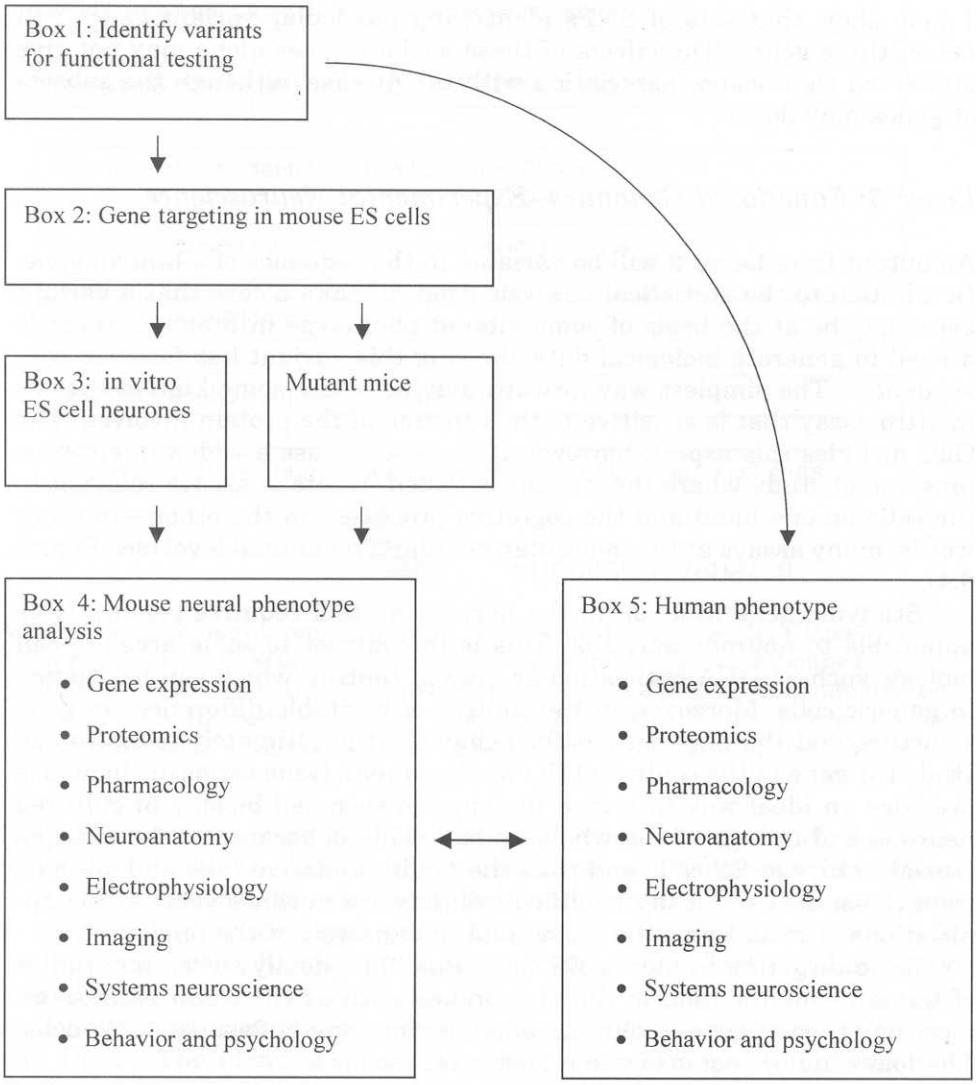

Figure 8.4. Schematic representation of Layer 3. Layer 3 shows a simplified flowchart beginning with the sequence of putative human variant gene and an outline of assays based on mouse embryonic stem (ES) cell technology. These ES cell experiments are based on generating mutant cell lines that can be used to generate mutant mice (Box 4) as well as in vitro differentiation into neurons (Box 3; see Figure 8.5 for more detail). A list of phenotypes that can be scored are shown in Box 4 for mice, and a comparable set of human phenotype data can also be drawn from the profiling of the subjects used for the DNA used in Layer 2. The arrow from Box 1 to Box 5 is meant to indicate that the human phenotypes listed are those relevant to the DNAs that were used in the genotype screen.

4. Conditional gene modification; these methods allow the desired genetic modification to be "active" or "inactive" in a desired cell (neurone or specific neuronal population) at a specific time during the lifespan of the animal (Le & Sauer, 2000; Mansuy & Bujard, 2000). For example, a gene that regulates synaptic plasticity, which may be encoded in almost all neurons, can be inactivated in a set of neurons in a selected brain region (e.g., hippocampus) and the effects on cognitive functions assessed.
5. Rescue of knock-out allele with mutant or variant genes (Kojima et al., 1997).
6. Insertion of reporter constructs to monitor gene expression (Migaud et al., 1998) and subcellular localization of proteins (e.g., Green Fluorescent Protein technology).

Although mutant mice are useful, there are some neuronal phenotypes that can be studied in neurons grown in culture (see Figure 8.4, Box 4). A major limitation to the study of synapse function has been the lack of clonal cell lines that form synapses with the properties of central nervous system synapses. Very recently, it was found that totipotent murine ES cells can be induced to differentiate in culture into neurones (*e*mbryonic *s*tem cell *n*eurons; ESNs) comparable with those prepared from neonatal cortex (Bain & Gottlieb, 1998). Importantly, the ESNs display the ability to form functional synapses. Combining gene targeting with ESN technology allows the creation of mutant neurons in vitro. This opens the possibility toward various in vitro screens in mutant neurons (see Figure 8.5).

The phenotype of the cells, animal tissues, and whole animal can be systematically studied in a variety of studies ranging from the molecular to the psychological (see example list in Figure 8.4, Box 5). It is unnecessary here to break this list into further detail but rather to draw attention to the value of multiple lines of experimental analysis. The first advantage of testing a variant allele in multiple assays is that it makes it more likely that a phenotype can be identified. A greater challenge is to understand why a variant allele may be involved with the human phenotype. Here it is necessary to have some information on the brain at many levels. For example, if one were to only examine synapse function, one may overlook some other critical role in, say, glial function. The advantage of the mouse is that it is possible to explore many levels using ethically acceptable approaches, unlike humans, for which it is not possible to perform similarly invasive procedures. Thus, it is necessary to compare and contrast at those levels where it is possible—the phenotype of mouse and human (see Figure 8.4).

Comparison of mouse and human phenotypes can be pursued on two levels: (a) comparing the mouse and human where each carries a mutation in the same gene and (b) comparing similar phenotypes where the genetic basis in humans is unknown. As illustrated in Boxes 4 and 5 of Figure 8.4, it would be important to have detailed annotation of phenotype information, assembled in appropriate databases, so that genotype information could be used to ascribe gene function to a phenotype. Developing "neu-

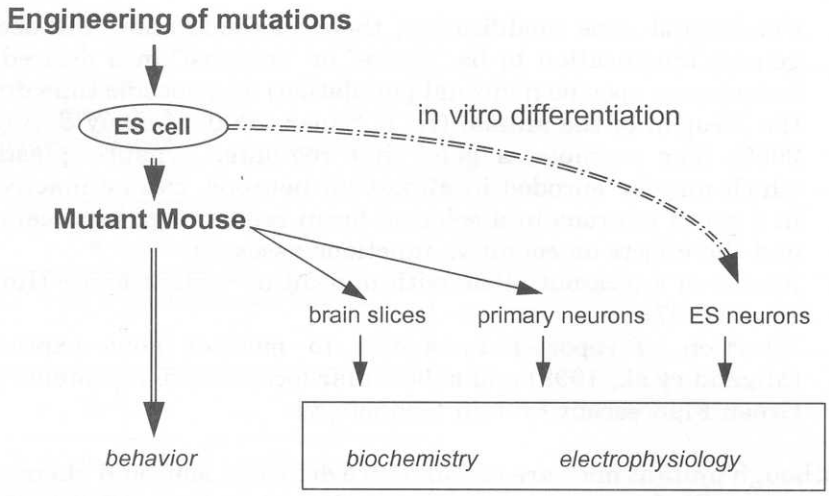

Figure 8.5. Using mutant mouse embryonic stem (ES) cells to generate both mice and neurons in vitro. Mutations or chromosomal alterations are engineered into mouse ES cells. These cells can be used to create mutant mice via blastocyst injection or differentiated in vitro to produce neurons (ES neurons). The mutant mice can be the source of a variety of neural tissue preparations including brain slice and primary neuronal cultures. Thus neurons carrying mutations can be derived from a variety of methods and used in many types of assays, ranging from behavior of the intact animal to electrophysiology and biochemistry.

roscience phenotyping" assays for comparison of humans and mouse is an area that needs further development. Many tests have been developed for rats and are readily transferable to mice. This program of research could promote further efforts to improve and find new ways to examine neurological phenotypes in mice.

Layer 4: Informatics

This broadly integrated program places emphasis on the need for user mobility between datasets as well as storage and recall of information. Linking the datasets together is perhaps one of the most difficult challenges, and a fluid interface would be extremely important. Here are some examples of questions that a well-designed informatic interface may be able to handle:

- List all genes that are encoded in a particular region of chromosome 6 and expressed in the hippocampus. Further sort those genes into those that are known to be important for development or synaptic plasticity of the hippocampus.
- Identify the regions of the human brain that are altered in functional magnetic resonance imaging studies in various genetic diseases, and contrast the regions with the known gene expression profiles and the biochemical function of these genes.

- Catalog the multiprotein complexes involved with synaptic signaling, and list the corresponding human syndromes involving those genes.
- List the genotyping assays that could be used to differentially diagnose chosen psychiatric disorders.
- List the human polymorphisms that result in altered expression of neuronal membrane proteins, and link this to drugs known to modulate those proteins.

Although some aspects of these questions could be answered today, the amount of labor involved with the current data mining tools is enormous. In principle, it should be possible to have answers to these questions in just hours with appropriately designed databases and search engines.

Structural Issues for a Large Multidisciplinary Program

The G2C approach is well-suited to be organized as a national or international consortium. The structure can be broken down into component areas, such as the four layers, where these areas are managed by experts. A relatively fixed framework with considerable flexibility would be desirable during the phase of construction. The initial phase of construction could proceed around the existing framework that has developed from the study of learning. This framework, which could be described as the vertical component because it links genes, proteins, cells, circuits, and behavior, presents the major challenges in terms of constructing the new databases, the general management, and coordination. Once this was established, the G2C could be expanded using several strategies. First, using the entry point at Layer 1, new systems or assemblies could be identified and processed. For example, there are areas of developmental neuroscience in which sets of genes and pathways have been identified. Here, experts in these areas could meet and discuss Layer 1, whereupon Layer 2 could be performed by other genotyping experts. In the broadest sense, the consortium could be seen as a platform for scientists in specialized areas to join a program that connects their work with those at other points in the vertical organization.

A second driving force is the collection of human DNA samples. In the area of clinical neuroscience (psychiatry, neurology, and neurosurgery), there has been substantial effort placed on the collection of DNAs from individuals and families with heritable conditions. It is outside the scope of this chapter to review this in detail; however, it is worth noting that there is a considerable need for further collection, including for conditions that have not gained as much attention as some of the major disorders. The availability of these DNAs for Layer 2 of the G2C would influence the choice of gene sets from Layer 1.

A third way to expand the G2C is around the cell biology of neurons and glia. For example, my colleagues and I have initiated the G2C by focusing on a particular synaptic multiprotein complex. The next step

could be to include other synaptic protein complexes and move toward genotyping for all synaptic proteins. The number of synaptic proteins is probably around 2,000, and many of these are now known. These could also be used to constitute a set for Layer 1. Thus, one could expand this approach to include dendritic, axonal, and other sets of neuronal proteins. Similarly, this type of categorization could be applied to glial cells. The use of microarray technology for monitoring messenger RNA expression, will assist in generating these data, especially as arrays encoding all known transcripts are available. In the broadest sense, it would be possible to extend the G2C to examine all genes expressed in the brain.

Conclusion

In the same way that one now understands biochemical pathways underlying various basic biological function including metabolism, cell proliferation, and differentiation, the core molecular pathways utilized in neurons for the generation of behaviors will be found. As is already clear, aberrations in these pathways arising from genetic variation will provide a molecular basis for altered behavior. These molecular insights will provide not only useful diagnostic and therapeutic avenues but also potentially profound insights into human cognition. The development of tools used to link areas of neuroscience through vertical layers ranging from psychology to gene structure will be a valuable resource for new areas of curiosity-led research as well as industrial application.

Perhaps the greatest biological challenge of the 21st century is to understand the mechanisms of human behavior. This intellectual challenge is of enormous practical significance given the burden of neuropsychiatric disease. Here a general discovery framework for understanding behavior at the molecular, cellular, and systems neuroscience levels in humans is outlined. The key hypothesis driving this program is that molecular assemblies are core components of neurobiology and behavior and are the fundamental defective unit in diseases of the brain. A new approach to human studies of cognition and brain disease can be driven by basic science in model genetic organisms and combined with an integrative program of genome sequencing and neuroscience.

References

Adrian. E. (1928). *The basis of sensation, the action of the sense organs.* New York: Norton.

Alberts, B. (1998). The cell as a collection of protein machines: Preparing the next generation of molecular biologists. *Cell, 92,* 291–294.

Altshuler, D., Pollara, V. J., Cowles, C. R., Van Etten, W. J., Baldwin, J., Linton, L., et al. (2000). An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature, 407,* 513–516.

Bader, J. S. (2001). The relative power of SNPs and haplotype as genetic markers for association tests. *Pharmacogenomics, 2,* 11–24.

Bain, G., & Gottlieb, D. (1998). I. Neural cells derived by in vitro differentiation of P19 and embryonic stem cells. *Perspective Developmental Neurobiology, 5,* 175–178.

Bliss, T. V., & Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology, 232,* 331–356.

Bradley, A., Zheng, B., & Liu, P. (1998). Thirteen years of manipulating the mouse genome: A personal history. *International Journal of Developmental Biology, 42,* 943–950.

Brent, R. (2000). Genomic biology. *Cell, 100,* 169–183.

Brown, S. D., & Nolan, P. M. (1998). Mouse mutagenesis-systematic studies of mammalian gene function. *Human Molecular Genetics, 7,* 1627–1633.

Cheah, S. S., & Behringer, R. R. (2000). Gene-targeting strategies. *Methods in Molecular Biology, 136,* 455–463.

Collingridge, G. L., Kehl, S. J., & McLennan, H. (1983). Excitatory amino acids in synaptic transmission in the Schaffer collateral–commissural pathway of the rat hippocampus. *Journal of Physiology (London), 334,* 33–46.

DeChiara, T. M. (2001). Gene targeting in ES cells. *Methods in Molecular Biology, 158,* 19–45.

Fors, L., Lieder, K. W., Vavra, S. H., & Kwiatkowski, R. W. (2000). Large-scale SNP scoring from unamplified genomic DNA. *Pharmacogenomics, 1,* 219–229.

Grant, S. G., O'Dell, T. J., Karl, K. A., Stein, P. L., Soriano, P., & Kandel, E. R. (1992). Impaired long-term potentiation, spatial learning, and hippocampal development in fyn mutant mice. *Science, 258,* 1903–1910.

Grant, S. G., & Silva, A. J. (1994). Targeting learning. *Trends in Neuroscience, 17,* 71–75.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory.* New York: Wiley.

Husi, H., & Grant, S. G. (2001a). Isolation of 2000-kDa complexes of *N*-methyl-D-aspartate receptor and postsynaptic density 95 from mouse brain. *Journal of Neurochemistry, 77,* 281–291.

Husi, H., & Grant, S. G. (2001b). Proteomics of the nervous system. *Trends in Neuroscience, 24,* 259–266.

Husi, H., Ward, M. A., Choudhary, J. S., Blackstock, W. P., & Grant, S. G. (2000). Proteomic analysis of NMDA receptor-adhesion protein signaling complexes. *Nature Neuroscience, 3,* 661–669.

Isaksson, A., Landegren, U., Syvanen, A. C., Bork, P., Stein, C., Ortiago, F., et al. (2000). Discovery, scoring and utilization of human single nucleotide polymorphisms: A multidisciplinary problem. *European Journal of Human Genetics, 8,* 154–156.

Kojima, N., Wang, J., Mansuy, I. M., Grant, S. G., Mayford, M., & Kandel, E. R. (1997). Rescuing impairment of long-term potentiation in fyn-deficient mice by introducing Fyn transgene. *Proceedings of the National Academy of Sciences USA, 94,* 4761–4765.

Kokoris, M., Dix, K., Moynihan, K., Mathis, J., Erwin, B., Grass, P., et al. (2000). High-throughput SNP genotyping with the Masscode system. *Molecular Diagnosis, 5,* 329–340.

Kwok, P. Y. (2000). High-throughput genotyping assay approaches. *Pharmacogenomics, 1,* 95–100.

Le, Y., & Sauer, B. (2000). Conditional gene knockout using cre recombinase. *Methods Molecular Biology, 136,* 477–485.

Mansuy, I. M., & Bujard, H. (2000). Tetracycline-regulated gene expression in the brain. *Current Opinion on Neurobiology, 10,* 593–596.

Masood, E. (1999). As consortium plans free SNP map of human genome. *Nature, 398,* 545–546.

Medico, E., Gambarotta, G., Gentile, A., Comoglio, P. M., & Soriano, P. (2001). A gene trap vector system for identifying transcriptionally responsive genes. *Natural Biotechnology, 19,* 579–582.

Migaud, M., Charlesworth, P., Dempster, M., Webster, L. C., Watabe, A. M., Makhinson, M., et al. (1998). Enhanced long-term potentiation and impaired learning in mice with mutant postsynaptic density-95 protein. *Nature, 396,* 433–439.

Mills, A. A., & Bradley, A. (2001). From mouse to man: Generating megabase chromosome rearrangements. *Trends in Genetics, 17,* 331–339.

Mullikin, J. C., Hunt, S. E., Cole, C. G., Mortimore, B. J., Rice, C. M., Burton, J., et al. (2000). An SNP map of human chromosome 22. *Nature, 407,* 516–520.

Niu, T., Struk, B., & Lindpaintner, K. (2001). Statistical considerations for genome-wide scans: Design and application of a novel software package polymorphism. *Human Heredity, 52*, 102–109.

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., et al. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature, 409*, 928–933.

Sanes, J. R., & Lichtman, J. W. (1999). Can molecules explain long-term potentiation? *Nature Neuroscience, 2*, 597–604.

Silva, A. J., Paylor, R., Wehner, J. M., & Tonegawa, S. (1992). Impaired spatial learning in alpha-calcium-calmodulin kinase II mutant mice. *Science, 257*, 206–211.

Tjian, R. (1995). Molecular machines that control genes. *Scientific American, 272*, 54–61.

Wiehe, T., Guigo, R., & Miller, W. (2000). Genome sequence comparisons: Hurdles in the fast lane to functional genomics. *Brief Bioinform, 1*, 381–388.