11

Cellomics: Characterization of Neural Subtypes by High-Throughput Methods and Transgenic Mouse Models

JOSEPH DOUGHERTY

INTRODUCTION

Since the birth of neuroanatomy, it has been recognized that the nervous system exhibits a most remarkable cellular diversity (Ramón y Cajal et al. 1899; Sotelo, 2003). Although most cells can be broadly categorized as neurons or glia, there is a dizzying array of diverse cellular morphologies within these categories, particularly for neurons. Using the classic Golgi technique for sparse labeling of random neurons, even within a relatively homogenous structure such as the cerebellum, there exists at least a dozen different neuronal cell types distinguishable by cell size, location, length, and complexity of dendritic arbors or axonal projections (Figure 11.1). Many of the fundamental questions of neuroscience are centered on this diversity of form, and presumably function, within the nervous system. What is the purpose of this remarkable diversity? How does it arise, evolutionarily and ontologically? How does this contribute to the computational capacity of the brain? Which cell types are essential for which behaviors? And, importantly, what changes and in which cell types lead to neurological disorder? The purpose of this chapter is to discuss current methods with which to study these cell types within the natural context of the brain in comprehensive and high-throughput manners. The chapter first addresses the definition of cell types and then the approaches for targeting them genetically. Then it touches on the "omics"-level approaches available to investigate cell types, ending with a particular focus on the molecular characterization of cell type by transcriptome profiling.

DEFINITIONS OF CELL TYPE

Historically, neuroscientists have focused on several classical approaches for defining cell types: morphological, physiological, functional, and molecular (Box 11.1). In an ideal taxonomy, each of these four levels of definition would be redundant and thus would be mutually interchangeable. For example, every Purkinje neuron would express exactly the same suite of genes (morphological = molecular), have the same peak firing rate and membrane capacitance (morphological = physiological), and have the same impact on the circuit when deleted (morphological = functional). We know that this is not always the case. For example, there are two classes of unipolar brush cells that are thus far morphologically indistinguishable yet can be distinguished by molecular markers (Nunzi et al. 2002). There are subsets of Purkinje neurons that express zebrin (Leclerc et al. 1992), or tyrosine hydroxylase (Takada et al. 1993). And the relationship between physiological classes of cortical interneurons and available molecular markers is particularly complex (Ascoli et al. 2008; Yuste, 2005). However, this likely reflects our own imperfect characterization of these cells across multiple levels of investigation.

The Centrality of Gene Expression to Definitions of Cell Type

To quote the American architect, Henry Louis Sullivan, "it is the pervasive law of all things organic and inorganic…that form ever follows function." This maxim, usually paraphrased as "Form follows function," states that buildings should be structured to suit the roles for which they are intended. This maxim can be applied

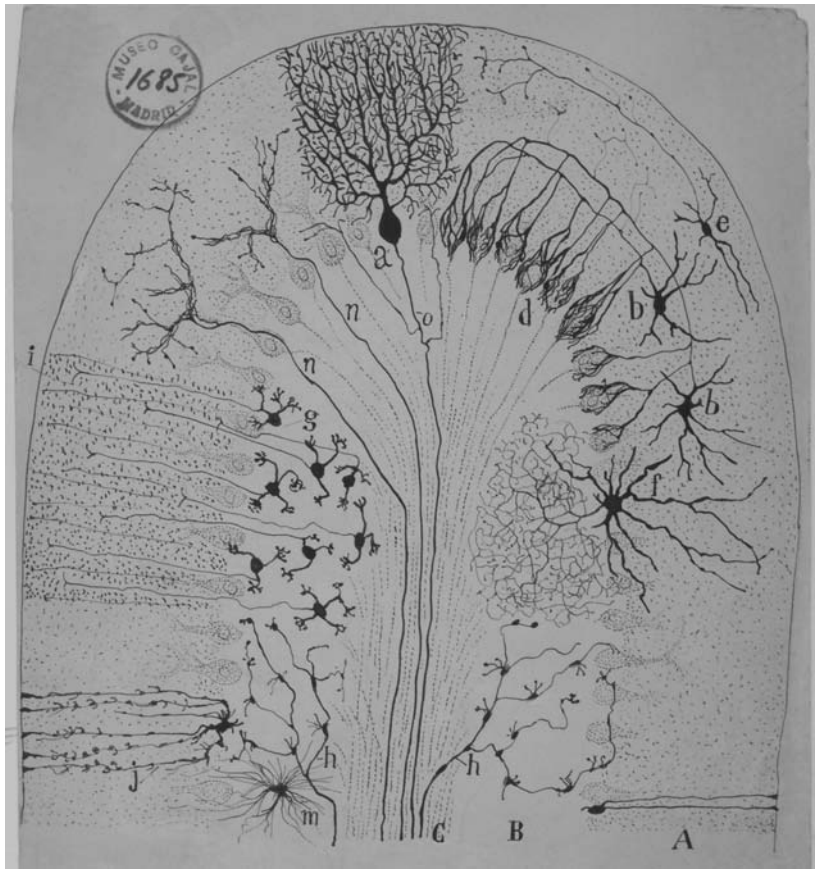


FIGURE 11.1: Illustration of the morphologically diverse cell types present in a single CNS structure: the cerebellar follia. Reprinted with permission. Santiago Ramón y Cajal. Legado Cajal. Instituto Cajal (CSIC). Madrid (Spain).

to cell types as well: The form and physiology of each cell type has been carefully tuned to its particular function in the nervous system. However, in biology, adaptation is mediated not by an architect but by gene expression. Thus, to a biologist, the phrase would be: "Physiology, function, and form follow gene expression."

The brain is a biological structure, tuned by evolution to input sensory information and output behavior. The embodiment of evolution is the cell, but the substrate of evolution is the inherited material—the DNA. To evolve a new cell type, DNA must change to generate new genes, or alter the regulation of existing genes, to provide a set of coherent specifications for the new cell type. While every cell has essentially an identical copy of the genome, each cell type has to be tuned to express the set of genes required for its particular functions at appropriate levels and times. A neuron with a large elaborate dendritic arbor, like a Purkinje neuron, will need to generate more postsynaptic proteins, dendritic microtubules, and ribosomes. A fast-spiking neuron will need to express more proteins for rapid repolarization of the membranes, the buffering intracellular calcium, or a unique set of channels. Neurons that use particular transmitters need to generate enzymes for transmitter synthesis and loading into vesicles. And all cell types will need to express a unique set of receptors to respond to the local and distal cues important for their particular functional roles in the calculus of the brain. Therefore, while cell types may be operationally defined by morphology, physiology, or function, all of those unique features must be mediated by the gene expression of that cell.

Thus the molecular definition of cell type (Box 11.1), when comprehensive enough to be genomic, will allow a final taxonomy of cells, even before the relationship between particular genes and morphology or physiology are understood. However, the classic molecular methods described in Box 11.1 are not comprehensive

BOX 11.1

CLASSICAL APPROACHES TO THE DEFINITION OF CELL TYPES

Morphological/Anatomical

The earliest definitions of cell type were strictly morphological or anatomical (Ramón y Cajal et al. 1899). Cells were defined by location and shape of the cell body, orientation and elaboration of the dendritic arbor, and the projection pattern of the axons. These early descriptive studies were fundamental to understanding the basic texture of the nervous system, and early inferences from morphology regarding development, information flow, and computational processes of the central nervous system were surprisingly insightful. However, rigorous, comprehensive classification schemes and robust quantification were not the focus of these early studies. Many clear cell types are apparent from these early investigations. For example, Purkinje neurons are unambiguously distinct from other cell types. Yet other gradations made by Cajal and compatriots, such as the distinctions between cerebellar Golgi neurons of long axon and Golgi neurons of short axon are less obviously a distinct classification.

Functional

Perhaps the most pragmatic manner to define a cell type is functional. A cell type serves as a particular computational node within a circuit, with the ultimate purpose of all circuits in the brain being to mediate behaviors. Thus, a cell type could be defined by its computational role in a circuit and its ultimate influence on an animal's capacity for a certain behavior. Early investigations involved the physical ablation (Lashley 1930) or stimulation (Penfield & Rasmussen 1950) of entire regions followed by behavioral analysis, permitting connections between anatomical and functional levels of analysis. With time and technical development, the resolution of these approaches gradually progressed. Stereotactic targeting of smaller and smaller structures became practical. Excitotoxic agents permitted the ablation of cell bodies rather than traversing fibers within a region. Pharmacological agents permitted the selective manipulation of cells expressing just a particular receptor, linking functional and molecular levels of analysis to some extent. Finally, in the modern era, genetic approaches permit cell specific targeting of specific transgenic constructs for functional manipulation of cell types (Boyden et al. 2005), allowing for definitions of cells important for behavioral functions in an intact animal, ranging from sleep and memory consolidation (Rolls et al. 2011) and feeding (Domingos et al. 2011) to fear conditioning (Haubensak et al. 2010; Letzkus et al. 2011).

Physiological

With the development of electrophysiological techniques it became possible to define a cell type by a coherent and consistent set of physiological properties, such as firing rates and patterns, membrane capacitance, afterhyperpolarization, and electrical response to pharmacological agents. Data collection in the early studies was typically blind to any morphological or molecular features, and there were sometimes quite surprising divergences between these different levels of definition. In the modern era, improvements in microscopy, particularly the advent of calcium dyes and two-photon imaging, allow some combination of morphological, physiological, and molecular investigation. The development of transgenic mouse lines expressing GFP to molecularly define cell types has also been a boon to prospective physiological analysis of cell types, especially those that are present at low frequency in a tissue.

Molecular

Finally, the development of early fluorescent microscopic techniques and of molecular reagents such as immunohistochemistry and in situ hybridization led to the recognition that

(*continued*)

BOX 11.1
CONTINUED

even morphologically identical neurons may have somewhat distinct molecular composition (Coons AH et al. 1941, 1950; Hyden & McEwen, 1966); thus methods developed permitting neurons to be classified by neurotransmitter phenotype (Hillarp et al. 1966; Saito et al. 1974), receptors for drugs or neurochemicals (Roth & Barlow 1961), or expression of particular calcium binding proteins (Celio & Heizmann 1981; Celio & Norman 1985; Hyden & McEwen 1966). Again, these molecular features sometimes showed direct correspondence with morphological or physiological criteria and sometimes did not, most notoriously in the case of interneurons, where the relationship between particular markers and physiology is quite complicated (Yuste 2005).

It is also clear that there are particular molecular markers that are thought to be more indicative of a particular *state* of a neuron, rather than the presumably more stable *trait* of cell type. The classic example being the expression of an intermediate early gene, such as cFos, thought to correspond to a recent bolus of activity in a neuron.

Finally, although we refer to molecular approaches rather broadly here, with current technology the most scalable and comprehensive will be those based on nucleic acid, such as measurement of transcript levels in particular cell types.

and can be conducted only postmortem. To approach a true molecular taxonomy of cell type requires prospective and comprehensive methods of analysis.

TARGETING CELL TYPES FOR PROSPECTIVE ANALYSIS

While early methods for studying specific molecules in a cell, such as immunohistochemistry, collectively allowed the study of specific cell types post hoc, they did not directly permit in vivo observations of the cell types of interest, nor did they provide access for experimental manipulation. However, all cellular macromolecules are either direct or indirect products of genes. Thus the cell-specific localization of a molecule often indicates the cell-specific expression of certain genes. Therefore a particular gene's genomic regulatory information can be coopted to drive the expression of foreign transgenes in specific cell types. A variety of methods now exist for doing just this, as summarized in Box 11.2. Many of these methods are applicable across a variety of species, although this chapter focuses on applications in mice—the mammalian model organism most amenable to genetic manipulation.

While not as high-throughput as readily scalable molecular methodologies such as sequencing or microarrays, several projects have systematically generated and characterized a large number of mouse lines targeting a variety of transgenes to specific cell types (Gong et al. 2007; Heintz 2004; Madisen et al. 2010; Portales-Casamar et al. 2010; Smedley et al. 2011; Taniguchi et al. 2011). These various transgenes (Box 11.3) facilitate the direct observation and, in some cases, manipulation of the targeted cell types. These techniques have been a boon to the parallel physiological, functional, and molecular characterization of the same cells.

There now exist relatively efficient means of generating targeting constructs for transgenesis that are scalable for high-throughput approaches and potentially automation (Gong et al. 2010; Poser et al. 2008). However, the rate-limiting step will remain to be the actual genesis and husbandry of the mouse lines. For some reagents, such as BAC transgenics, there is some potential for multiplexing constructs in vivo into a single locus, stably inherited, with independent regulation (Figure 11.2), although this could improve efficiency to only a limited extent.

The Limitations of Genomic Information as a Method for Targeting Cell Types

Unfortunately for the neuroscientist, nature did not necessarily provide us with a single uniquely expressed gene for every cell type. Even long-established drivers—such as Pcp2

BOX 11.2

GENETIC TOOLS TO TARGET SPECIFIC CELL TYPES

Genomes contain regulatory sequences, such as promoters, enhancers, and repressors, that collectively direct the expression of the surrounding genes—often with exquisite temporal and spatial selectivity, or in response to particular cellular states such as sustained neuronal depolarization. Promoters and enhancers are both thought to be elements that increase the expression of the adjacent gene in appropriate contexts, while repressors are elements that suppress the expression of the gene in inappropriate contexts. Promoters are found at the start site of transcription, while enhancers can be found nearly anywhere, including in introns, or even many kilobases from the closest exons. The current model is that the activity of these regulatory sequences is mediated by the presence of transcription factor binding sites and corresponding epigenetic modifications. While there are clear examples of genes in which these elements have been studied to a high degree of detail (Johansson et al. 2002; Oberdick et al. 1990), for the vast majority of genes these features are inferred from sequence of epigenetic marks (Muers, 2011; Myers et al. 2011) but are functionally uncharacterized. Regardless, these genomic regions can be coopted to drive the expression of particular transgenes.

Transgenesis

Transgenesis is the insertion of a distant or even exogenous gene into a new genomic context. In many species, including mouse, this can be accomplished by the injection of DNA fragments containing the gene into a fertilized oocyte (Gordon et al. 1980; Jaenisch 1976; Palmiter & Brinster 1986). With some low but experimentally tractable frequency, this DNA will be integrated into the oocyte's genomic DNA at one or more random locations. The fragments typically integrate into tandem repeated copies, often hundreds of copies in length (Chandler et al. 2007; Palmiter & Brinster 1986). If the DNA integrates early enough in development to contribute to the germline, mice derived from these oocytes can transmit this transgenic locus to their progeny, establishing a transgenic line. The random locus of integration of the transgene is both a strength and a weakness. If the transgenes, by chance, land in a locus containing a cell specific promoter, these transgenes will demonstrate specific and heritable celltype–specific expression patterns. However, most frequently the injected fragment of DNA will contain a small (0.5- to 5-kb) promoter or enhancer upstream of the transgene. Depending on the locus of integration, this small regulatory element may be sufficient to direct cell-specific expression in some mouse lines, such as in the examples of the PCP2 promoter (Oberdick et al. 1990) and Nestin enhancers (Johansson et al. 2002). In other cases, the locus of integration interacts with the promoter to generate experimentally useful (but irreproducible) patterns of expression. For example, the small and fairly ubiquitous neuronal promoter Thy-1 was utilized to generate a range of transgenic lines, each having a heritable pattern of expression in selective subsets of neurons (Feng et al. 2000).

Pros:

- Faster and cheaper than knockins
- Less sophisticated molecular skills needed to generate construct than BAC transgenesis
- Higher copy number can lead to high transgene expression

Cons:

- Locus of integration effects more common than BACs
- Retargeting the same cell type difficult in some cases

(*continued*)

BOX 11.2
CONTINUED

- Transgenerational silencing
- Unintended disruption of a random gene at the locus of integration

BAC TRANSGENESIS

There are two primary limitations of small promoters in transgenesis. First, for many genes (and therefore many cell types), the appropriate promoters and enhancers regions for directing cell specific expression have not been characterized. Second, those promoters that are characterized are often strongly influenced by locus of integration effects. BAC transgenesis was developed to circumvent these difficulties (Yang et al. 1997). Bacterial Artificial Chromosomes (BACs) are large 100- to 200-kb fragments of mouse or human genomic DNA, maintained in bacteria. As these originally served as a mechanism to fragment genomes into manageable sizes for genome sequencing projects, large libraries of BACs exist, tiling essentially the entire mouse genome. These fragments are thought to be large enough to contain most if not all of the enhancer, repressor, and promoter elements that direct cell specific expression of a gene, even if the individual elements are undefined. These BACs can be readily modified in bacteria, utilizing recombination-based methodologies (Gong et al. 2002; Hollenback et al. 2011; Poser et al. 2008; Yang et al. 1997), to insert transgenes into the translation start site of cell-specific "driver" genes. The modified BAC is then utilized for transgenesis, creating a novel genomic context for these transgenes. Performance is dependent on the BAC utilized, but overall, BACs have been shown to accurately and reproducibly target a wide range of cell types in the CNS (Heintz 2004).

Pros:

- Multiple copies can insert (higher potential expression level than knockins). Greater accuracy (less locus of integration effects than small promoters).
- Speed (more rapid to generate than knockins).
- Little to no transgenerational silencing.
- BACs are large enough to provide sufficient insulation from adjacent sequence, whether that is genomic sequence or even other BACs. Thus it is possible to multiplex BACs in a manner not possible in other techniques (Figure 11.2).

Cons:

- Copy number not as high as small constructs (but higher than a knockin).
- Some locus of integration effects possible (less than small transgenes).
- BAC constructs require more expertise to generate.
- BACs may carry extra copies of unmodified genes adjacent to driver gene and thus result in increased gene dose for these genes. Often these genes are expressed in unrelated cells types (or even tissue) and thus may have little consequence on the phenotypes of interest in the brain, or they may be buffered by biological mechanisms that keep RNA levels constant even when extra copies of the gene are present. But this may need to be checked by qPCR for some experiments.

KNOCKINS

There are genes in the genome that exceed even the size of a BAC as well as known examples of enhancers that are hundreds of kilobases distant from their target genes. Therefore even BACS may not contain all of the regulatory information necessary to recapitulate the endogenous expression pattern of a gene. Thus, in many ways, inserting a transgene directly into a particular locus by targeted homologous recombination provides the ultimate opportunity

BOX 11.2
CONTINUED

to coopt genomic regulatory information to control the expression of an exogenous transgene. While there may be gains in accuracy for some genes, there are several subtleties to the strategy that should be mentioned. First, knockins are slower and more labor-intensive than transgenics—homologous recombination requires many extra steps and screening in embryonic stem cells and additional rounds of breeding to screen chimeras. Second, in many strategies, knocking in the transgene simultaneously knocks out one copy of the endogenous gene, meaning that all mice will be haplo-insufficient for the driver—a gene that may be of particular importance to the cell type of interest. (In some designs this can be overcome by careful use of bicistronic sequences such as IRES or viral 2a sequences)(Taniguchi et al. 2011). Third, compared with transgenics, knockins may tend to have lower expression of the transgene, as they will be present in only a single copy rather than tandem arrays. Finally, there are several reasons that even with the nearly perfect genomic context of a knockin, the transgene may differ in observed expression from the endogenous gene—including differences in transcript stability, protein stability, and translation efficiency between the gene and the transgene.

Pros:

- Accuracy and reproducibility (no unintended locus of integration effects).
- New technologies (TALENs and CRISPR/Cas9) improving efficiency and applicability beyond mouse.

Cons:

- Slower and more expensive than transgenesis.*
- Limited to at most one or two copies of the transgene.
- Can create haploinsufficiency of the driver gene.

COMBINATIONS AND INTEGRATIONS

It is worth noting that variations of these techniques exist, for inserting transgenes with small promoters (Portales-Casamar et al. 2010), or BACs (Heaney et al. 2004) into specific loci (as is the case in knockins), thus balancing some of the pros and cons of knockins. Overall, the selection of the technique depends largely on the influence of the relevant pros and cons on a particular individual project as well as the experience and expertise of the investigator.

Also, it is worth noting that many genes in the genome are heavily regulated at the level of splicing and selection of alternative transcription start sites. Thus, the selection of one transcription start site over another in targeting by BAC or knockin could clearly influence the expression of a transgene regardless of which method is chosen for targeting. This level of complexity is largely ignored by the field currently.

**While this chapter was in press, new CRISPR technologies have alleviated this issue.*

for Purkinje neurons, Nestin for neural stem cells, and emerging drivers such as Aldh1L1 for astrocytes—which each have apparently good specificity in the brain, are often expressed robustly in other tissues as well (Anthony & Heintz 2007; Cahoy et al. 2008; Day et al. 2007; Doyle et al. 2008; Dubois et al. 2006; Foo & Dougherty, 2013; Zhang et al. 2005). It is also a common observation in the characterization of Cre lines that many drivers are expressed in unexpected populations during development, or even the egg, resulting in widespread early recombination. This highlights the need for strategies that permit additional layers of experimental control for the temporal or spatial expression of genes.

Temporal Control of Transgenic Expression

Experimentally, it is often important to control the time of transgene expression to prevent widespread recombination, particularly for studies of development (lineage tracing) or cell

BOX 11.3
TRANSGENES OF NOTE

FLUOROPHORES

The most common transgene providing anatomical access to particular cell types is the green fluorescent protein (GFP), derived from Aequorea victoria, or its variants. Over a decade of engineering and directed evolution have resulted in variants covering a range of fluorescence wavelengths, of which CFP, YFP, and GFP have been shown to have robust fluorescent properties and a lack of overt toxicity in vivo. The other major family of fluorescent proteins in use in laboratories is the variants of the red fluorescent DSRed gene, originally isolated from *Discosoma*. Though frequently incredibly bright and stable, several of these, particularly the monomeric forms, have been reported to have toxicity or to aggregate in vivo (Dougherty et al. 2012b; Strack et al. 2008). Tandem dimerized variations of this protein (tdTomato) apparently do not display these detrimental properties and have been gaining popularity both as transgenes and as the new standard reporter lines for Cre recombinase (Madisen et al. 2010). Finally, this is quite an area of active research and new proteins with novel properties are continuously being discovered, designed, or evolved. In addition to continual modifications to adjust wavelengths and increase maturation, stability, and brightness, there also exist variants with photoswitchable wavelengths (Andresen et al. 2008), new far-red variants (Dieguez-Hurtado et al. 2011; Shcherbo et al. 2009), variants whose wavelengths change over time (Terskikh et al. 2000; Yanushevich et al. 2003), as well as proteins that alter fluorescent properties in response to changes in intracellular calcium (Looger & Griesbeck, 2012; Tian et al. 2012).

RECOMBINASES

Recombinases are proteins that recognize specific sequences in DNA and re-arrange or recombine the DNA in particular predictable manners. By far the most commonly used recombinase is the Cre recombinase, derived from bacteriophage. Cre canonically recognizes specific DNA sequences called LoxP sites. Depending on the relative orientation of the sites, Cre can be used to either excise or invert DNA flanked by Lox P sites ("Floxed") or insert large fragments of DNA into single sites or even mediate engineered chromosomal rearrangements (Mills & Bradley, 2001; Nakatani et al. 2009; van der Weyden & Bradley, 2006). Since the introduction of the technique, the most important adaptations have included modifications to optimize codon usage for mammalian systems (Shimshek et al. 2002), the fusion with the estrogen receptor to permit tamoxifen inducible nuclear translocation (and thus recombination) (Indra et al. 1999), the identification of alternative variations of LoxP sites (Siegel et al. 2001), and the development of split Cre reagents for recombination mediated by the intersection of expression of two separate loci (Hirrlinger et al. 2009).

As a distant second, the next most frequently utilized recombinase is Flpe, an optimized version of the *Saccharomyces cerevisiae* Flp-1 recombinase. This recombinase recognizes Frt sites rather than LoxP. It has also now been optimized for mammalian codon usage and thermostability (Kranz et al. 2010; Rodriguez et al. 2000) and has inducible variations available (Hunter et al. 2005).

In addition to providing flexibility of experimental strategies, the development of a second recombination system has opened the door to strategies that respond to particular combinations of gene expression. For example, there are now reporter mice that express GFP only after recombination by both Cre and Flpe (Farago et al. 2006), allowing even more specific molecular definitions of cell type.

BOX 11.3
CONTINUED

REAGENTS FOR ACTIVATING AND SILENCING CELL TYPES

For neuroscientists interested in the physiological properties and functional roles of particular cell types, the development of light-activatable ion channels has been an extraordinary advance. These channels combine the millisecond temporal resolution required for sophisticated and naturalistic manipulation of neurons together with all the potential for targeting of genetically encoded tools to provide extraordinary specificity for cell type. The basic tool for activation, Channelrhodopsin, derived from *Chlamydomonas reinhardtii*, is a channel that passes cations upon stimulation by blue wavelength light. This allows depolarization sufficient to trigger action potentials in neurons. For inhibition, the basic tool is Halorhodopsin, a yellow light–driven chloride pump derived from Halobacterium, that could be utilized to hyperpolarize neurons and prevent them from firing. Both of these tools, as well as related proteins derived from other species, have now been extensively modified to adapt them to mammalian systems, increase stability, maturation, membrane localization, and provide a wider variety of wavelengths. The most current variants of each are recently reviewed (Chow et al. 2012; Lin, 2011). These tools have been shown to be effective when inserted in a cell-specific manner with virus (Haubensak et al. 2010; Letzkus et al. 2011) or under the control of specific BACs (Zhao et al. 2011a), or Cre reporters (Madisen et al. 2010).

Experimentally activatable G protein–coupled receptors are also now available. Examples include chimeras of vertebrate rhodopsins with the intracellular loop from a β2-adrenergic receptor allowing light activation of second messenger signaling (OptoXRs) (Airan et al. 2009), as well as a variety of receptors lacking endogenous mammalian ligands, such as DREADS, RASSLs, and Allostatin receptors (Masseck et al. 2011). These tools will permit the investigation of a variety of neuromodulatory and secondary messenger signaling within genetically targeted cell types.

Finally, it is worth noting a variety of tools for relatively permanently silencing neurons, including toxic transgenes (Garcia et al. 2004; Hara et al. 2001) and tethered toxins (Auer et al. 2010), now exist. While these tools lack the exquisite temporal resolution of their light-activated analogs, they are advantageous for studies requiring long-term silencing to study the behavioral consequences of the functional ablation of particular cell types.

TAGS AND FUSIONS (CHIP, CLIP, TRAP, AND SUBCELLULAR PROTEOMICS)

A variety of strategies have been developed to adapt biochemical purification techniques to cell-specific profiling, most notably affinity purification of RNA, DNA, protein complexes, and specific subcellular organelles. One example of this is the TRAP methodology for translating ribosome affinity purification (Heiman et al. 2008). In this method a transgene that is a fusion protein of eGFP and the ribosomal protein L10a (Rpl10a) is targeted to specific cell types. This is typically done using BACs, although any of the methods in Box 11.2 could be applicable and Cre-responsive lines now exist. Because of the Rpl10a moiety, the protein is incorporated into ribosomes and is thus associated with mRNA that is undergoing translation only in the targeted cell type. The eGFP moiety then serves both as a fluorescent tag for anatomical studies as well as an affinity tag for biochemical purification: mouse brains are rapidly homogenized, and the GFP-tagged ribosomes (and affiliated mRNAs) are captured with anti-GFP antibodies coupled to magnetic beads. The method has been shown to be effective across a range of cell types in the nervous system (Doyle et al. 2008) and, when coupled to microarray or RNAseq, allows a genome-wide snapshot of mRNAs in use in a particular cell. Although the method has not yet been validated with any genomic assays, the very similar Ribotag strategy utilizes an HA tag and a Cre-responsive knockin into the Rpl22 locus to capture ribosomes (Sanz et al. 2009).

(*continued*)

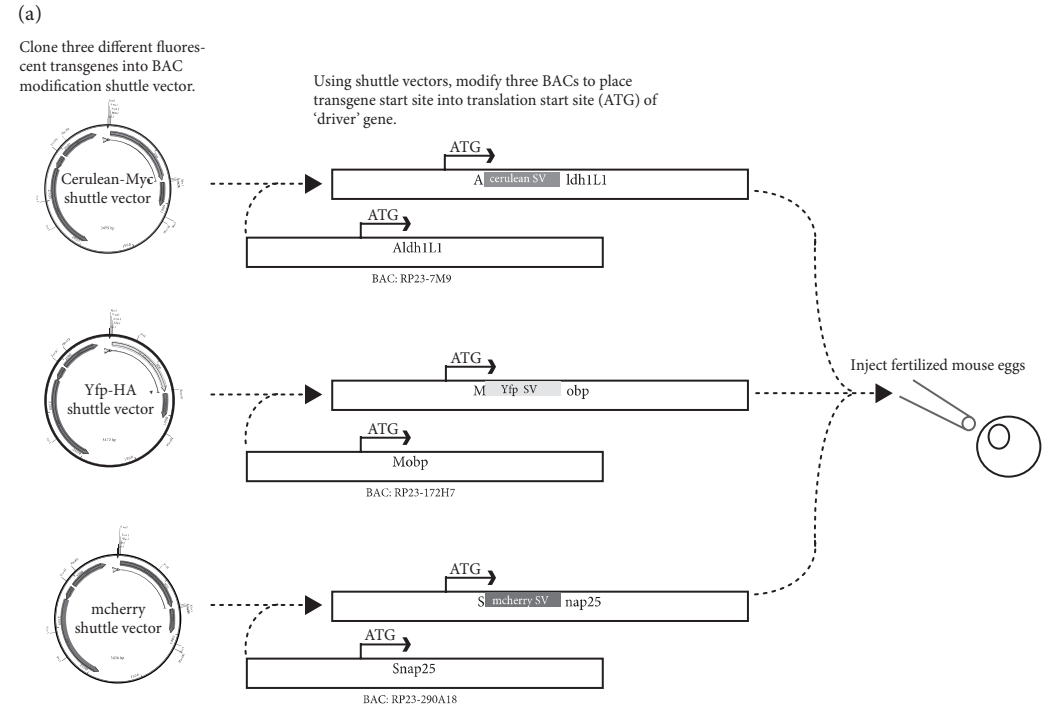
BOX 11.3
CONTINUED

Similar strategies have been applied to study the interaction of mRNAs and microRNAs to particular RNA-binding proteins using the more generic CLIP (cross-linking and immunoprecipitation) approach, utilizing either tagged versions of the protein or antibodies against the endogenous protein (Chi et al. 2009; He et al. 2012; Jensen & Darnell 2008). Likewise any protein of interest can be expressed in a tagged form in specific cell populations, or under its endogenous promoter and be utilized for affinity purification followed by proteomic strategies to study protein-protein interactions and protein modifications in particular cell types (Bateup et al. 2008; Zhong et al. 2009). Tagged DNA-binding proteins can also be expressed in a cell-specific manner to permit cell-specific epigenetic approaches, such as studying the interaction of transcription factors to DNA with CHIP (chromatin immunoprecipitation) (Zhang et al. 2008). This can be helpful both to study the differential binding patterns of the same protein in distinct cell types, or as an alternative strategy when antibodies to the endogenous protein are unavailable or ineffective. Finally, it has been demonstrated that tags of the correct protein can permit the cell-specific purification of entire organelles in a manner sufficient for proteomic analysis (Heller et al. 2012; Selimi et al. 2009). This approach opens completely new avenues of investigation for cell biologists interested in the nervous system.

function at discrete time points, such as during acquisition of a new behavior in a learning assay or after early development. Currently there exist several methods for integrating additional temporal regulation into genetic modifications. All have some utility, although there are particular advantages and disadvantages to each.

First, utilization of inducible recombinases, such as a fusion protein of Cre or Flpe together with an estrogen receptor (Cre-Ert2), can limit recombination events to a particular time window when a drug (tamoxifen) is added, avoiding recombination in early development (Danielian et al. 1998; Hunter et al. 2005). Although Cre-Ert2 can be less efficient than normal Cre under the same driver (S. Gong, personal communication) (Gong et al. 2007), this approach is becoming more widespread. Likewise, there exists a smaller set of lines utilizing tetracycline-responsive promoter and repressor elements coopted from *Escherichia coli* (Passman & Fishman 1994; Saez et al. 1997; Zhou et al. 2009). The advantage over Cre is reversibility, with the ability to add and then remove the drug again later to turn the gene on and back off. The disadvantages are reported leakiness of the system in some lines, toxicity of tTA transgene at high levels, and the lack of lines targeting a wide range of cell types compared with the number of available Cre drivers.

Second, utilizing combinations of recombinases (Farago et al. 2006), or split recombinases (Hirrlinger et al. 2009), one can design logic gates such as "and" or "or" that would require the concurrent or sequential expression of two genes in a cell type for recombination to occur (Dymecki & Kim 2007). This would also permit the manipulation of cell types for which no single driver gene exists but which are unique in their expression of a particular combination of genes, with a particular timing. The drawbacks to this are the "allele problem" of combining three or more alleles into a single animal and the sometimes imperfect efficiency of recombination. The allele problem follows from simple Mendelian rules—in attempting to combine three or more alleles utilizing heterozygous breedings, only a small fraction of the progeny generated will have the correct combination of alleles to be experimentally useful. Thus these approaches may be particularly amenable to multiplexing transgenic approaches (Figure 11.2) (Dougherty et al. 2012b). Still, even multiplexing may be stymied by imperfect efficiency of recombination—it is a common observation that even with accurate expression of recombinase, recombination may not occur in all cells. This is likely to be a function, in part, of the level of recombinase expression and, in part, of the accessibility and structure of the targeted region in the genome. Regardless,



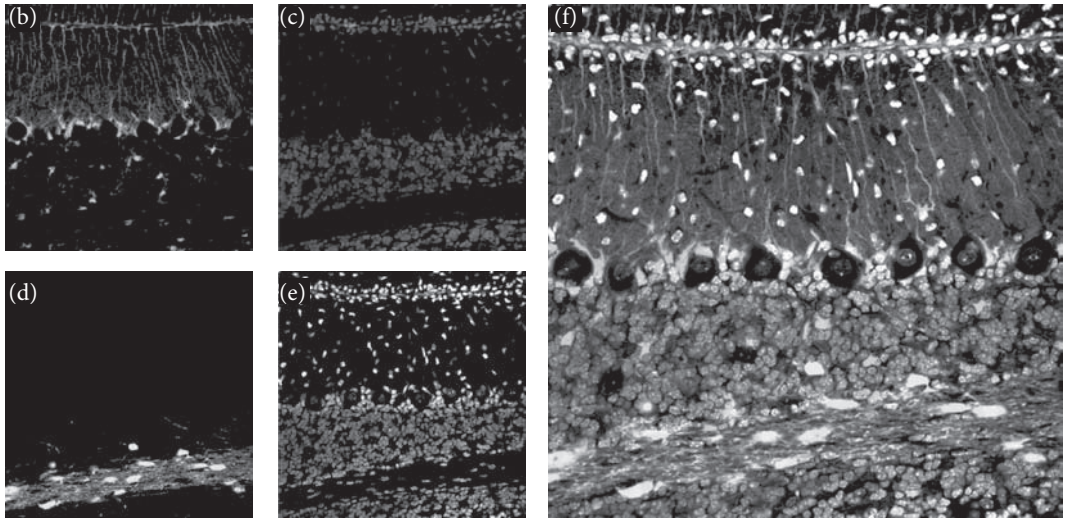


FIGURE 11.2: Multiplexed genetic targeting of cell types in the mouse central nervous system. (a) Three different BAC promoter constructs, with drivers for neurons (Snap25), astrocytes (Aldh1L1), and oligodendrocytes (Mobp), are modified in parallel with three spectrally distinct fluorophores, then co-injected into fertilized mouse eggs. The resulting lines demonstrate distinct labeling of (b) Aldh1L1 BAC drives Cerulean Fluorescent Protein specifically in astroglia of a 13-day-old mouse cerebellum. (c) Snap25 BAC drives mCherry fluorophere specifically in neurons. (d) Mobp BAC drives Yellow Fluorescent Protein specifically in oligodendrocytes. (e) DAPI counterstain shows nuclei. (f) Overlay demonstrates mutually exclusive expression of the three fluorophores in the same mouse.

these inefficiencies would be estimated to be multiplicative when dealing with strategies requiring combinations of recombinases.

Finally, there is the powerful combination of Cre reagents with lentiviral and adenoviral constructs. There are two variations of this approach—a virus with a floxed transgene can be injected into a mouse line with cell specific Cre expression or a virus expressing Cre can injected into a mouse line with a floxed gene or transgene in the genome. There are several advantages to these approaches. First, the time of injection permits temporal control of transgene expression, thus circumventing difficulties with Cre lines with early expression. Second, locus of injection allows for a level of anatomical selectivity, or the intersection of region and cell type, in a manner that genomic drivers alone may not permit. Third, the strong promoters available in the virus may permit more robust transgene expression than most endogenous promoters, whether knockin or BAC. This is a particularly important benefit for transgenes such as channelrhodopsins that require high levels of expression for efficacy.

Nongenetic Methods of Targeting Cell Types

Overall, most of these methods generally rely on genomic information for targeting and thus are reliant on a molecular definition of cell types. However, alternatives do exist. For example, injecting retrogradely labeling fluorescent dyes into the spinal cord can permit the sorting and profiling of corticospinal upper motor neurons (Arlotta et al. 2005), thus using an anatomical definition of cell type to harvest molecular information. As long as perfect drivers do not exist for all cell types, clever use of retrogradely integrating reagents, including an interesting set of engineered rabies viruses (Wall et al. 2010), could also permit the expression of transgenes in a cell-specific manner based on projection patterns and anatomical connections.

CELLOMICS: HIGH-THROUGHPUT AND COMPREHENSIVE CHARACTERIZATION OF CELL TYPES

While the act of transgenesis and animal husbandry remain low-throughput, once they are targeted there now exist several related methods for high-throughput characterization of the molecular properties of specific cell types. These -omics and -omics-like approaches can be organized roughly by the tradition of cell-type definition (Box 11.1) that they best inform. Of these, the molecular approaches are the most readily scalable, although several technologies currently enhancing the throughput of other approaches should be mentioned.

Morphological "Anat-omics"

Over the last decade, several complimentary projects have undertaken efforts to characterize the expression of hundreds or thousands of genes in the genome within the mouse brain and embryo (Easterday et al. 2003; Geschwind et al. 2001; Gray et al. 2004; Heintz 2004; Jones et al. 2009; Lein et al. 2007; Magdaleno et al. 2006; Shimogori et al. 2010; Visel et al. 2004). Although, as a general rule, quality is often inversely proportional to throughput, these resources have transformed the manner in which we analyze expression studies and demonstrated the potential of automation and informatics even when applied to techniques with components that are not as inherently scalable as sequencing or microarrays.

These efforts collectively have generated a new encyclopedia of knowledge about gene expression, and we cannot overemphasize the importance of this work. However, there are several important caveats to consider in examining these sources of data (Jones et al. 2009). First, null results for a given gene (lack of any expression in the brain) should be viewed with extreme caution, as these assays are typically executed with little optimization for any given gene and each technique has a different threshold of detection and dynamic range. RNA for genes with no measurable expression by *in situ* hybridization can be detected by other methods (Dougherty et al. 2010; Lee et al. 2008). Second, in most cases only a single isoform per gene has been assayed; thus our knowledge is likely incomplete for the other products of a given gene. Finally, these approaches are not inherently quantitative, and time of development for the enzymatic reactions may be variable. With these techniques, comparison across different experiments regarding the relative strength of gene expression can be misleading. Nonetheless, these resources provide an essential baseline analysis for the anatomy of gene expression in the nervous system. It is also worth noting there are methodological advances that may improve some of these shortcomings.

One important advancement that may serve to increase the throughput of anatomical studies using genetically encoding fluorophores is the direct coupling of tissue sectioning with automation of fluorescent microscopy and data acquisition (Ragan et al. 2012). Scanning the brain in real time as it is sectioned removes many intermediate steps that may result in experimental variability (such as development with enzymatic reagents) and provides data that are more readily quantifiable, with the exquisite morphological detail available from high-resolution two-photon imaging and the capacity for three-dimensional reconstruction. If the eventual cost of these systems permits it, future genome-wide efforts to characterize fluorescent molecule expression may bypass expensive, time-consuming, and nonquantitative enzymatic detection and amplification steps. Likewise, advances in imaging may remove the need for sectioning altogether (Chung & Deisseroth, 2013) .

Another interesting advance has been the combination of multiple fluorophores in single animals both in targeted strategies (Dougherty et al. 2012b; Feng et al. 2000; Shuen et al. 2008) and the "Brainbow" approach for a fluorescent variation of a high-throughput Golgi-like labeling (Weissman et al. 2011). These approaches, particularly in combination with digitization efforts, have the potential to permit some parallelization and automation of the analysis of morphological studies.

Finally, one emerging technology proposes to adapt the throughput of sequencing technology to study connectivity (Oyibo et al. 2011). Retrogradely transported viruses, capable of genomic integration, lift unique DNA "barcodes" from the genome of one cell to into the genomes of synaptically connected cells. Potentially, sequencing of these tags from the genomes of each cell might permit the construction of an all vs. all map of connectivity of a mouse nervous system.

Physiological and Functional Cellomics

The mature high-throughput correlate for physiological studies would be the implantable multielectrode array. Developed over the last two decades, these arrays have permitted the simultaneous physiological assessment of tens to hundreds of individual neurons within a region. This advance has provided the opportunity to understand the behavior of cells in situ and in a relatively unbiased manner. A variety of important conceptual advances have been made possible by this technology. The understanding of the importance of neural synchrony to perceptual processing within and across regions has been strongly advanced by these studies (Miller & Wilson 2008), as have the hypotheses regarding memory consolidation during sleep (Sutherland & McNaughton 2000). However, aside from some recent examples utilizing these tools in clever combination with genetically encoded constructs for silencing or activating neurons (Haubensak et al. 2010; Letzkus et al. 2011), there is little about the technique that lends itself to analysis of prospectively targeted cell types.

This niche is being filled now by the development of genetic indicators of neural activity—especially genetically coded calcium indicators (Looger & Griesbeck 2012; Tian et al. 2012). Although early indicators have existed since the 1980s, only recently have proteins been developed with properties sufficient to merit their widespread adoption. The parallel development of enhanced capacity for in vivo imaging, notably multiphoton microscopy and derivatives, has facilitated the adoption of these technologies. Especially now with the advent of multicolor genetically encoded dyes (Zhao et al. 2011b), the potential begins to exist to study the physiological interactions between genetically targeted cell types in awake and behaving animals. In the future, this will permit descriptive correlational studies between the behavior of neurons and the behavior of animals that will allow some inferences as to the final functional role of particular cell types. However, by their very nature, behavioral analyses are going to remain relatively low-throughput methods as compared with molecular approaches.

Molecular Cellomics

Epigenomics

Between the DNA and the transcription of RNA lies the epigenome. Although, in classical genetics, it had a somewhat different definition (Waddington 2012), epigenetics has currently come to be defined as the suite of modifications occurring to the DNA molecule itself (methylation, hydroxymethylation) or associated histones (methylation, acetylation, etc.), or the presence of other DNA binding proteins that appear to correspond to changes in gene expression (Goldberg et al. 2007). Whether

these changes should be considered causes or corollaries of gene expression is currently unclear. Their accurate assessment and interpretation is an area of very active investigation (Maunakea et al. 2010).

Current technologies for assessing epigenetic state have been moving from hybridization based (CHIP-Chip) assays to sequencing-based approaches (Chip-Seq, MRE-Seq, Methyl-Seq) (Laird 2010). These technologies share the common feature of using an affinity reagent or enzymatic reaction (bi-sulfite sequencing) to capture epigenetically modified DNA, or epigenetic marks on proteins cross-linked to DNA, followed by sequencing to identify the approximate or exact sites of modifications (Harris et al. 2010).

To date, several projects, notably ENCODE/modENCODE, have conducted surveys across tissues to examine tissue-specific patterns of epigenetic modification (Muers 2011; Myers et al. 2011). Despite the fact that these methods are DNA-based and thus amenable to working with small amounts of starting material, surprisingly little of the work has focused at the level of individual cell types with the notable exception of the work that discovered DNA hydroxymethylation in mammalian cells via examination of Purkinje cell nuclei in the brain (Kriaucionis & Heintz 2009). As an intermediary between the DNA and transcription, many of the changes apparent from the transcriptomics ought to be reflected in any epigenetic investigation of cell type, and direct comparisons of transcription of the epigenome may help deconvolute the meaning of many of the epigenetic marks in real in vivo contexts. This will likely be an area of active investigation in the next few years: technology exists for purification of nuclei of specific cell types (Kriaucionis & Heintz 2009), and many epigenetic assays are scalable and applicable to small amounts of material.

Proteomics/Metabolomics

If epigenomics is the study of what comes between the RNA and the DNA, then proteomics is the study of what comes after RNA. Nearly any substance present in a cell is in some manner the consequence of the expression of a particular gene. Most proteins will be coded for directly by a particular RNA, while many small molecules (metabolites) and modifications of larger proteins are the consequence of the expression of particular enzymes.

RNA expression, while necessary for the generation of many proteins, is not sufficient. The RNA must be translated—a process regulated extensively and in a transcript-dependent manner. By the best estimates we have available, RNA levels will correlate with the expression of their corresponding proteins at about 0.6 or better (Hegde et al. 2003; Kislinger et al. 2006). Thus one could argue that knowledge of protein levels directly is more important to the prediction of cell behavior and function than RNA levels alone. Proteomics also gives additional information about the post-translational modifications upon the products of genes. Many, if not all, proteins are regulated by extensive post-translational modifications such as phosphorylation, methylation, myristoylation, ubiquitination, acetylation, and many others that can profoundly alter their activity and function, and these modifications are impossible to predict from RNA sequence alone. There are many technical challenges, however, in performing global proteomic profiling. Given their size, proteins are typically digested and analyzed by mass spectrometry. To achieve profiling coverage of the hundreds of thousands of peptides generated, multidimensional separation strategies are used. Although effective, such strategies usually require 10 to 20 hours of analysis time and therefore impose significant limitations on throughput. Moreover, generally the global profiling approach does not screen for post-translational modifications or other potential protein-protein interactions that may strongly influence protein activity.

To more directly examine protein activity, activity-based proteomic strategies have been developed that use chemical probes engineered to react with enzymes of interest (Kam et al. 1993). Alternatively, as another mechanism to monitor protein activity, the substrates and products of enzymes may be measured. This approach, known as metabolomics, aims to simultaneously measure small molecules involved in cellular pathways such as glycolysis and the Krebs cycle. By using state-of-the-art mass spectrometry technology, tens of thousands of small molecules—metabolites—can be detected within cells (Patti et al. 2012; Scalbert et al. 2009). While theoretically knowledge of the levels of all proteins present in a cell would predict the presence of particular metabolites, too little is known about most of these molecules. No such comprehensive relationships can

be built. Only the high-level integration of data at multiple -omics levels of investigation will permit the generation of these sorts of predictive models.

Here, we have lumped together the assessment of proteins and their products largely because the technologies available for their assessment are similar. As with epigenetic investigations, little effort has been focused on assays of protein or metabolite levels in specific cell types. Here the barriers are technological—unlike with nucleic acid, there are no readily apparent methods for amplifying proteins and metabolites, and as sensitive as these sophisticated devices have become, collecting sufficient material of a particular type will still be limiting for many cell types in the nervous system. Furthermore, unlike nucleic acids, individual proteins will vary widely in the biochemical properties, therefore these assays are much more sensitive to handling and extraction conditions than work with DNA and RNA. Nonetheless, there are at least two interesting advancements to be highlighted in the application of proteomics to the nervous system: organellomics and in situ metabolomics.

Organellomics

Two limitations of proteomics include the need to isolate large amounts of material from specific cell types, and the difficulty in detecting low abundance proteins. One possible approach to tackling this question is to focus on specific subcellular compartments in specific cell types. The focus on just a particular organelle or subcellular compartment (such as the synapse) should increase the relative abundance of important proteins that may be of relatively low abundance cell wide. The use of transgenic technologies can permit adding tags for affinity purification of specific organelles. For example, Selimi and colleagues tagged just the parallel fiber synapse on Purkinje neurons by expressing a fusion of GFP to the GluRδ2 protein specifically in Purkinje neurons. Affinity purification of these specific synapses from biochemical preparations of the mouse brain permitted proteomics of this specific synapse in a specific cell type, leading to important insights into signaling within this structure in these cells (Selimi et al. 2009).

One could imagine parallel approaches for purification of other organelles from specific cell types. Even existing mouse lines with GFP tagged ribosomes could be used to study differences in protein composition of ribosomes from particular cell types (Doyle et al. 2008). The drawback to these approaches is that they may depend on de novo transgenesis for each cell type and biochemical optimization for each organelle of interest, but they permit more comprehensive information than would be available by any other method.

In Situ Metabolomics

One major challenge of the application of proteomics and metabolomics to the nervous system is the requirement to isolate a sufficient number of cells in a manner that is not too disruptive to the very profile to be studied. Any technique that involves physical dissociation of living cells from the nervous system is likely to have some effect on the profile. Therefore, new technologies that combine classic anatomical preparations with mass spectrometry are of great interest to study the in situ profile of particular cell types. The nanostructure initiator mass spectrometry (NIMS) imaging technique rasters a laser across a slice of tissue on a special substrate and measures by mass spectrometry the ionized particles as they come off the tissue (Patti et al. 2010). Currently the technique is limited in that only the most abundant products can be measured given ion suppression effects. Nonetheless, resolution is on the level of the individual cells, and combination with genetically labeled tissue may permit proteomic and metabolomic investigation of specific cell types in situ.

Transcriptome/Translatomics

Measurement of relative RNA abundances is the most inherently scalable and accessible approach to the molecular characterization of cell types. RNA, once converted to cDNA, can be amplified essentially limitlessly. Compared with genomic DNA (for epigenetic applications), most RNAs are present in more than two copies per gene per cell. And a variety of robust platforms exist for profiling RNA pools, ranging from the hybridization-based microarrays to emergent RNA-sequencing technologies. Because of these features, inputs as low as even single cells or tens of cells have successfully been utilized for measurement of gene expression by qPCR, sequencing and microarrays (Burgemeister et al.

2007; Dixon et al. 2000; Hempel et al. 2007; Islam et al. 2011; Mary et al. 2011).

These studies have shown a remarkable diversity of gene expression across cell types in the nervous system. Comparison of even very closely related cell types, such as Drd1 and Drd2 medium spiny neurons (Heiman et al. 2008; Lobo et al. 2006) or corticospinal neurons to corticostriatal neurons (Arlotta et al. 2005; Schmidt et al. 2012) has identified hundreds of transcripts that are differentially regulated. For more distantly related cell types, such as comparing a Purkinje cell to an astrocyte or to a cholinergic neuron of the medial habenula, there are thousands of differentially regulated genes (Doyle et al. 2008), with differences so robust they violate some of the assumptions normally utilized in microarray analysis (Dougherty et al. 2010). Indeed, one of the most striking findings of these studies was that different types of neurons were as specialized from one another as neurons were from glia. Genes that were particularly prone to high magnitudes of variation across cell types (high entropy) coded for classes of molecules typically found at the cell surface, such as receptors and channels, that mediate the response of particular cell types to the environment, as well as some transcription factors and calcium binding proteins (Doyle et al. 2008).

These -omics-level approaches to gene expression in particular cell types provide a wealth of information beyond the capacity of pursuit of individual laboratories. These data can broadly inform both our understanding of the cells in the normal brain as well as their importance to pathological states. Continuing to capture and distribute this information is going to be an important part of any -omics level study. Unfortunately, the actual measurements are so platform-dependent that comparisons across studies are difficult. Nonetheless, analysis within a study can still provide novel biological insights regarding these cells in health and disease.

CURRENT AND EMERGING USES OF TRANSCRIPTOMIC APPROACHES AND DATA

Finding Novel Molecules Important for the Function of Particular Cell Types

Studies profiling particular cell types have been conducted by laboratories interested in identifying novel features of these cells. The common finale of these transcriptomic studies of a particular cell type is a functional assay with one or a small number of the enriched transcripts discovered in the screen, often transcription factors (Arlotta et al. 2005; Dougherty et al. 2012a; Lai et al. 2008; Lobo et al. 2006; Molyneaux et al. 2005). For example, Arlotta et al. compared gene expression during development of corticospinal neurons and callosal projection neurons and identified a corticospinal neuron-specifically expressed gene, *Ctip2,* that was then found to be necessary for normal development of these cells in functional assays (Arlotta et al. 2005; Dougherty & Geschwind, 2005). While this selection of a single enriched transcript is necessary when pursing labor-intensive functional studies, these same profiles could easily be mined to identify many more transcripts key to cellular function.

Prediction of the Physiological Properties of the Cell Types

A related approach is to utilize the profiles as a means to predict the physiological properties of the cell type of interest. This is of particular interest in the identification of drug targets, such as receptors, that may have cell-specific expression and thus present unique therapeutic opportunities to alter the behavior of particular cell types in vivo. An example of this work was the identification by Heiman et al. of a novel G protein–coupled receptor enriched in a subclass of medium spiny neurons (Heiman et al. 2008). A separate group identified from microarray data particular potassium channels that likely mediate the development of fast spiking behavior in some cortical interneurons (Okaty et al. 2009). Another example was the characterization of a new cell type that appears necessary to mediate much of the response to antidepressants (Schmidt et al. 2012). In the future, profiling a particular, medically relevant cell type may be a powerful method to identify new drug targets for psychiatric or neurological disorders where the relevant cell types are known (Bartfai et al. 2012).

Characterization of Cell Types in Disease Models and Other Manipulations

As the reproducibility of cell-specific profiling has improved, it has become possible to study not just the profile of these cells in the normal

state but also to conduct comparative studies of pathological conditions, such as injuries, or genetic models of human diseases. The ability to study the response of particular cell types to stimuli as varied as drug exposure (Heiman et al. 2008; Schmidt et al. 2012), tumorigenesis (Dougherty et al. 2012a; Fomchenko et al. 2011), or gene knockout and overexpression (Warner-Schmidt et al. 2012) gives us the potential to learn a great deal of new neurobiology about each of these processes at a resolution and throughput previously unavailable. With its high reproducibility (Doyle et al. 2008), TRAP is particularly amenable to this approach, and these studies represent a large fraction of the TRAP studies currently under way.

Interpretation of Human Gene Expression Data from a Cellular Perspective

Finally, even for those investigators not conducting transcriptome profiling in their own labs, these resources have proven very useful datasets to aid in the interpretation of other gene expression data. One recent example of this work is the comparative analysis of different subclasses of glioblastoma tumors with the expression profiles of the major cell types of the brain. This work hinted that different subtypes of glioblastoma likely had emerged from different classes of normal cell types in the brain (Cahoy et al. 2008; Verhaak et al. 2010). Likewise, work profiling whole human brains across regions, time, and species has suggested that many, and perhaps most, gene expression differences seen in these studies are really driven by differences in cellular composition across the samples (Kang et al. 2011; Oldham et al. 2006, 2008). Emerging approaches are attempting to explicitly incorporate cellular information into analytical models for human data (Kuhn et al. 2011; Shen-Orr et al. 2010, Xu & Dougherty, 2013). Future approaches more explicitly utilizing the cellular profiles available from model organisms may lead to further improvements in the analysis and interpretation of human gene expression data. This is particularly important in psychiatric diseases, such as autism, where the relevant cell types are not known. If the human expression data can guide us to consistent cellular alterations, even in the context of distinct genetic or environmental causes across different individuals, then treatments could be tailored to address the common cellular deficits. Recent work in this direction for autism is very promising (Voineagu et al. 2011).

Interpretation of Human Genetic Data from a Cellular Perspective

Philosophically similar is the interpretation of human genetic data from a cellular perspective. Ongoing high-throughput genetic association studies utilizing single nucleotide polymorphisms (SNPs), copy number variations, or resequencing of exomes and genomes are identifying catalogs of common and rare variants that contribute to risk of developing various diseases of the nervous system. Cell-specific transcript profiling can contribute to these studies in at least two major ways.

1. If a particular cell type is already implicated in a disorder (such as dopaminergic neurons in Parkinson's disease or hypocretin neurons in narcolepsy), then cell-specific transcriptional profiling can be utilized to identify candidate genes for genetic analysis in human populations. As an example, it is well documented that a subset of individuals with autism have hyperserotonemia (Lam et al. 2006), suggesting that there may also be fundamental differences in the regulation of serotonin in their brains. We profiled serotonin-producing cells in the brain and identified a set of transcripts enriched in these cells. Polymorphisms in two of these genes are associated with autism in humans, and mutations in one of these had the effect of altering serotonin levels in mouse brains as well as resulting in behaviors reminiscent of autism (Dougherty et al. 2013. This candidate-gene-list approach, while having merit, is going to be eclipsed in human genetics by the rapidly falling cost of genome-wide sequencing studies. However, cell-specific transcript profiling can still make important contributions to the interpretation of these studies.
2. It is clear from the current studies that there are many, if not hundreds, of genetic routes to manifesting a complex psychiatric disorder like schizophrenia (Lee et al. 2012) or autism (Bill & Geschwind 2009). Therefore treatment strategies focused on specific genetic

> defects will have limited applicability. However, if these defects converge at the level of particular cells, then, as noted above, the cell becomes the target for treatment. One manner in which a diverse set of genes may converge on a particular cell type is through gene expression. There are two ways in which to conduct these analyses—first statistically: if the set of genes implicated in a human psychopathology are expressed more often than expected by chance in a particular cell type, the suggestion is that cell type is important to the disease process. Second, biologically: if even a single strongly implicated gene is expressed in only one cell type in the brain, then that provides extremely robust evidence for that cell type in the disorder. For example, an apparent autosomal dominant form of Tourette's syndrome was recently associated with a strongly deleterious mutation in the *HDC* gene in a large family (Ercan-Sencicek et al. 2010). As Hdc is expressed almost uniquely in histaminergic neurons in the brain, this strongly suggests that dysregulation of CNS histamine neurons can cause Tourette's syndrome. Therefore drugs that influence these cells may be of use in the treatment of Tourette's syndrome, at least in this family if not more broadly (Fernandez et al. 2012).

Finally, it is worth noting that the same cell-specific profiles that give the information for these analyses also simultaneously provide a list of potentially druggable molecules (receptors, kinases) that are enriched or uniquely expressed in the candidate cell types. This information could be essential to the design of treatments targeting these specific cell types (Bartfai et al. 2012; Doyle et al. 2008; Nelson et al. 2006).

The Grand Correlation—Assigning Putative Functions to Novel Genes

In the current era, there are roughly twenty thousand protein coding genes in the genome of the mouse and approximately the same number in humans. Of these, only a subset has even been named and even fewer are the focus of at least one publication. Thus the majority of the genome remains essentially unstudied. There is a great as yet untapped potential in these cell-specific profiling data for putative functional categorization for novel genes. Given enough cell types and good systematic phenotypic data regarding them it may be possible from a grand correlation to infer rough functions for these unstudied or understudied genes.

For example, imagine that one were interested in identifying genes involved in the maintenance of dendrites. Currently we have measured the RNA profiles of over twenty different types of neural cells (Dalal et al. 2013; Dougherty et al. 2012a; Dougherty et al. 2013; Doyle et al. 2008; Heiman et al. 2008; Schmidt et al. 2012). If one were to carefully measure a phenotype, such as the average dendritic area from each of these cell types, one could then look for genes whose expression is positively correlated with the phenotype as an in silico screen for genes involved in particular processes. The same approach could be taken for axon length, firing rate, nuclear size, *fos* expression in response to agonist, dendritic branching, or density of mitochondrial labeling. Much as webQTL permits in silico genetic investigations by only the phenotypic profiling of a standard set of strains (Wang et al. 2003), these profiling data provide an additional opportunity to leverage existing information to screen for novel contributors to a phenotype of interest. We have taken initial steps in this direction by providing a browsable interface for published bacTRAP data (http://java.bactrap.org/bactrap/index.jsp) (Dougherty et al. 2010).

Likewise, if parallel measurements can be conducted on the metabolic, epigenetic, and transcriptomic levels of a sufficiently large number of cell types, then the possibility exists for a truly grand correlation—a matrix that may permit the prediction of which epigenetic marks correspond to the production of which transcripts, and which transcripts indicate the presence of particular metabolites, and are thus related to the pathways that generate them. This combination of approaches—this overlapping of -omics with -omics—has the alluring potential to unlock many of the puzzles emerging from these intertwined fields.

The End Game of Molecular Analysis

Of the molecular -omics methods, the most readily scalable are the molecular methods based around nucleic acids (RNA and DNA). A final,

comprehensive taxonomy of cell types within the brain may await the moment when every cell type in the brain can be individually and comprehensively assayed for RNA expression—a goal that is not perhaps as unimaginably distant as it might seem. If the rate of decrease in sequencing costs continues at the exponential pace of the last 4 years (NHGRI, 2012), within 13 years it will be feasible to conduct RNA-seq on each of the ~100 billion neurons of a human brain, with 30 million 100-bp reads per neuron, for less than $500,000. By 15 years, it would cost $20,000. This all assumes that every neuron would need to be assayed, rather than just a sufficiently large subset, to identify all extant types. Comprehensive clustering by gene expression would then permit a final molecular categorization of all cell types, down to the level of individual cells.

THE FINAL WORD

It is important to keep in mind, in this era of heady scientific acceleration, that although data generation may be high-throughput, good data interpretation is low-throughput. Careful tool development and even more careful thought will always be needed to cope with this deluge of data.

ACKNOWLEDGMENTS

The author would like to thank Dr. S. Maloney for helpful comments on this manuscript. JDD is supported by the NIH (5R00NS067239-03, 2R21MH099798, 1R21NS083052) and the Malinckrodt Foundation.

REFERENCES

Airan, R. D., Thompson, K. R., Fenno, L. E., Bernstein, H., & Deisseroth, K. (2009). Temporally precise in vivo control of intracellular signalling. *Nature 458*, 1025–1029.

Andresen, M., Stiel, A. C., Folling, J., Wenzel, D., Schonle, A., Egner, A.,… Jakobs, S. (2008). Photoswitchable fluorescent proteins enable monochromatic multilabel imaging and dual color fluorescence nanoscopy. *Nat Biotechnol 26*, 1035–1040.

Anthony, T. E., & Heintz, N. (2007). The folate metabolic enzyme ALDH1L1 is restricted to the midline of the early CNS, suggesting a role in human neural tube defects. *J Comp Neurol 500*, 368–383.

Arlotta, P., Molyneaux, B. J., Chen, J., Inoue, J., Kominami, R., & Macklis, J. D. (2005). Neuronal subtype-specific genes that control corticospinal motor neuron development in vivo. *Neuron 45*, 207–221.

Ascoli, G. A., Alonso-Nanclares, L., Anderson, S. A., Barrionuevo, G., Benavides-Piccione, R., Burkhalter, A.,… et al. (2008). Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nat Rev Neurosci 9*, 557–568.

Auer, S., Sturzebecher, A. S., Juttner, R., Santos-Torres, J., Hanack, C., Frahm, S.,… Ibanez-Tallon, I. (2010). Silencing neurotransmission with membrane-tethered toxins. *Nat Methods 7*, 229–236.

Bartfai, T., Buckley, P. T., & Eberwine, J. (2012). Drug targets: single-cell transcriptomics hastens unbiased discovery. *Trends Pharmacol Sci 33*, 9–16.

Bateup, H. S., Svenningsson, P., Kuroiwa, M., Gong, S., Nishi, A., Heintz, N., & Greengard, P. (2008). Cell type-specific regulation of DARPP-32 phosphorylation by psychostimulant and antipsychotic drugs. *Nat Neurosci 11*, 932–939.

Bill, B. R., & Geschwind, D. H. (2009). Genetic advances in autism: heterogeneity and convergence on shared pathways. *Curr Opin Genet Dev 19*, 271–278.

Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G., & Deisseroth, K. (2005). Millisecond-time scale, genetically targeted optical control of neural activity. *Nat Neurosci 8*, 1263–1268.

Burgemeister, R., Friedemann, G., Schlieben, S., & Hitzler, H. (2007). Laser microdissection: gene expression analysis at the single-cell level. *Nat Methods* An24–An25.

Cahoy, J. D., Emery, B., Kaushal, A., Foo, L. C., Zamanian, J. L., Christopherson, K. S.,… et al. (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J Neurosci* (the official journal of the Society for Neuroscience) *28*, 264–278.

Celio, M. R., & Heizmann, C. W. (1981). Calcium-binding protein parvalbumin as a neuronal marker. *Nature 293*, 300–302.

Celio, M. R., & Norman, A. W. (1985). Nucleus basalis Meynert neurons contain the vitamin D-induced calcium-binding protein (Calbindin-D 28k). *Anat Embryol (Berl) 173*, 143–148.

Chandler, K. J., Chandler, R. L., Broeckelmann, E. M., Hou, Y., Southard-Smith, E. M., & Mortlock, D.P. (2007). Relevance of BAC transgene copy number in mice: transgene copy number variation across multiple transgenic lines and correlations with transgene integrity and expression. *Mamm Genome 18*, 693–708.

Chi, S. W., Zang, J. B., Mele, A., & Darnell, R. B. (2009). Argonaute HITS-CLIP decodes

microRNA-mRNA interaction maps. *Nature 460*, 479–486.

Chow, B. Y., Han, X., & Boyden, E. S. (2012). Genetically encoded molecular tools for light-driven silencing of targeted neurons. *Prog Brain Res 196*, 49–61.

Chung, K. & Deisseroth, K. (2013) CLARITY for mapping the nervous system., *Nature Methods, 10.* 508–513.

Coons A. H., Creech H. J., & RN, J. (1941). Immunological properties of an antibody containing a fluorescent group. *Proc Soc Exp Biol Med 47*, 200–202.

Coons, A. H., Snyder, J. C., et al. (1950). Localization of antigen in tissue cells; antigens of rickettsiae and mumps virus. *J Exp Med 91*, 31–38.

Dalal, J., Roh, J. H., Maloney, S.E., Akuffo, A., Shah, S., Yuan, H., Wamsley, B.,… Dougherty, J.D. (2013). Translational profiling of hypocretin neurons identifies candidate molecules for sleep regulation. *Genes Dev 27,* 565–578.

Danielian, P. S., Muccino, D., Rowitch, D. H., Michael, S. K., & McMahon, A. P. (1998). Modification of gene activity in mouse embryos in utero by a tamoxifen-inducible form of Cre recombinase. *Curr Biol 8*, 1323–1326.

Day, K., Shefer, G., Richardson, J. B., Enikolopov, G., & Yablonka-Reuveni, Z. (2007). Nestin-GFP reporter expression defines the quiescent state of skeletal muscle satellite cells. *Dev Biol 304*, 246–259.

Dieguez-Hurtado, R., Martin, J., Martinez-Corral, I., Martinez, M. D., Megias, D., Olmeda, D., & Ortega, S. (2011). A Cre-reporter transgenic mouse expressing the far-red fluorescent protein Katushka. *Genesis 49*, 36–45.

Dixon, A. K., Richardsen, P. J., Pinnock, R. D., & Lee, K. (2000). Gene-expression analysis at the single-cell level. *Trends Pharmacol Sci 21*, 65–70.

Domingos, A. I., Vaynshteyn, J., Voss, H. U., Ren, X., Gradinaru, V., Zang, F.,… Friedman, J. (2011). Leptin regulates the reward value of nutrient. *Nat Neurosci 14*, 1562–1568.

Dougherty, J. D., Fomchenko, E. I., Afua, A. A., Schmidt, E., Helmy, K.Y., Bazzoli, E.,… Milosevic, A. (2012a). Candidate pathways for promoting differentiation and quiescence in of oligodendrocyte progenitor-like cells in glioblastoma *Cancer Res 72,* 4856–4568.

Dougherty, J. D., & Geschwind, D. H. (2005). Progress in realizing the promise of microarrays in systems neurobiology. *Neuron 45*, 183–185.

Dougherty, J. D., Maloney, S. E., Wozniak, D. F., Rieger, M. A., Sonnenblick, L., Coppola, G.,…et al. (2013). The disruption of Celf6, a gene identified by translational profiling of serotonergic neurons, results in autism-related behaviors. *J Neurosci 33*, 2732–2753.

Dougherty, J. D., Schmidt, E. F., Nakajima, M., & Heintz, N. (2010). Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. *Nucleic Acids Res 38*, 4218–4230.

Dougherty, J. D., Zhang, J., Feng, H., Gong, S., & Heintz, N. (2012b). Mouse transgenesis in a single locus with independent regulation for multiple fluorophores. *Plos One* Accepted.

Doyle, J. P., Dougherty, J. D., Heiman, M., Schmidt, E. F., Stevens, T. R., Ma, G.,…et al. (2008). Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell 135*, 749–762.

Dubois, N. C., Hofmann, D., Kaloulis, K., Bishop, J. M., & Trumpp, A. (2006). Nestin-Cre transgenic mouse line Nes-Cre1 mediates highly efficient Cre/loxP mediated recombination in the nervous system, kidney, and somite-derived tissues. *Genesis 44*, 355–360.

Dymecki, S. M., & Kim, J. C. (2007). Molecular neuroanatomy's "three Gs": A primer. *Neuron 54*, 17–34.

Easterday, M. C., Dougherty, J. D., Jackson, R. L., Ou, J., Nakano, I., Paucar, A. A.,…et al. (2003). Neural progenitor genes. Germinal zone expression and analysis of genetic overlap in stem cell populations. *Dev Biol 264*, 309–322.

Ercan-Sencicek, A. G., Stillman, A. A., Ghosh, A. K., Bilguvar, K., O'Roak, B. J.,…et al. (2010). L-histidine decarboxylase and Tourette's syndrome. *N Engl J Med 362*, 1901–1908.

Farago, A.F., Awatramani, R. B., & Dymecki, S. M. (2006). Assembly of the brainstem cochlear nuclear complex is revealed by intersectional and subtractive genetic fate maps. *Neuron 50*, 205–218.

Feng, G., Mellor, R. H., Bernstein, M., Keller-Peck, C., Nguyen, Q. T., Wallace, M.,… Sanes, J. R. (2000). Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. *Neuron 28*, 41–51.

Fernandez, T. V., Sanders, S. J., Yurkiewicz, I. R., Ercan-Sencicek, A. G., Kim, Y. S., Fishman, D. O.,…et al. (2012). Rare copy number variants in Tourette syndrome disrupt genes in histaminergic pathways and overlap with autism. *Biol Psychiatry 71*, 392–402.

Fomchenko, E. I., Dougherty, J. D., Helmy, K. Y., Katz, A. M., Pietras, A., Brennan, C.,… Holland, E. C. (2011). Recruited cells can become transformed and overtake PDGF-induced murine gliomas in vivo during tumor progression. *PLoS One 6*, e20605.

Foo, LC, and Dougherty, J.D. (2013). Aldh1L1 is expressed in postnatal neural stem cells in vivo. *Glia, 61,* 1533–1541.

Garcia, A. D., Doan, N. B., Imura, T., Bush, T. G., & Sofroniew, M. V. (2004). GFAP-expressing progenitors are the principal source of constitutive neurogenesis in adult mouse forebrain. *Nat Neurosci 7*, 1233–1241.

Geschwind, D. H., Ou, J., Easterday, M. C., Dougherty, J. D., Jackson, R. L., Chen, Z. G., … et al. (2001). A genetic analysis of neural progenitor differentiation. *Neuron 29*, 325–339.

Goldberg, A. D., Allis, C. D., & Bernstein, E. (2007). Epigenetics: A landscape takes shape. *Cell 128*, 635–638.

Gong, S., Doughty, M., Harbaugh, C. R., Cummins, A., Hatten, M. E., Heintz, N., & Gerfen, C.R. (2007). Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *J Neurosci* (the official journal of the Society for Neuroscience) *27*, 9817–9823.

Gong, S., Kus, L., & Heintz, N. (2010). Rapid bacterial artificial chromosome modification for large-scale mouse transgenesis. *Nat Protoc 5*, 1678–1696.

Gong, S., Yang, X. W., Li, C., & Heintz, N. (2002). Highly efficient modification of bacterial artificial chromosomes (BACs) using novel shuttle vectors containing the R6Kgamma origin of replication. *Genome Res 12*, 1992–1998.

Gordon, J. W., Scangos, G. A., Plotkin, D. J., Barbosa, J. A., & Ruddle, F. H. (1980). Genetic transformation of mouse embryos by microinjection of purified DNA. *Proc Natl Acad Sci U S A 77*, 7380–7384.

Gray, P. A., Fu, H., Luo, P., Zhao, Q., Yu, J., Ferrari, A., … et al. (2004). Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science 306*, 2255–2257.

Hara, J., Beuckmann, C. T., Nambu, T., Willie, J. T., Chemelli, R. M., Sinton, C. M., … et al. (2001). Genetic ablation of orexin neurons in mice results in narcolepsy, hypophagia, and obesity. *Neuron 30*, 345–354.

Harris, R. A., Wang, T., Coarfa, C., Nagarajan, R. P., Hong, C., Downey, S. L., … et al. (2010). Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotechnol 28*, 1097–1105.

Haubensak, W., Kunwar, P. S., Cai, H. J., Ciocchi, S., Wall, N. R., Ponnusamy, R., … et al. (2010). Genetic dissection of an amygdala microcircuit that gates conditioned fear. *Nature 468*, 270–U230.

He, M., Liu, Y., Wang, X., Zhang, M. Q., Hannon, G. J., & Huang, Z. J. (2012). Cell-type-based analysis of microRNA profiles in the mouse brain. *Neuron 73*, 35–48.

Heaney, J. D., Rettew, A. N., & Bronson, S. K. (2004). Tissue-specific expression of a BAC transgene targeted to the Hprt locus in mouse embryonic stem cells. *Genomics 83*, 1072–1082.

Hegde, P. S., White, I. R., & Debouck, C. (2003). Interplay of transcriptomics and proteomics. *Curr Opin Biotechnol 14*, 647–651.

Heiman, M., Schaefer, A., Gong, S., Peterson, J. D., Day, M., Ramsey, K. E., … et al. (2008). A translational profiling approach for the molecular characterization of CNS cell types. *Cell 135*, 738–748.

Heintz, N. (2004). Gene expression nervous system atlas (GENSAT). *Nature 7*, 483–483.

Heller, E. A., Zhang, W., Selimi, F., Earnheart, J. C., Ślimak, M. A., Santos-Torres, J., ... Heintz, N. (2012). The biochemical anatomy of cortical inhibitory synapses. *PLoS One 7*, e39572.

Hempel, C. M., Sugino, K., & Nelson, S. B. (2007). A manual method for the purification of fluorescently labeled neurons from the mammalian brain. *Nat Protoc 2*, 2924–2929.

Hillarp, N. A., Fuxe, K., & Dahlstrom, A. (1966). Demonstration and mapping of central neurons containing dopamine, noradrenaline, and 5-hydroxytryptamine and their reactions to psychopharmaca. *Pharmacol Rev 18*, 727–741.

Hirrlinger, J., Scheller, A., Hirrlinger, P. G., Kellert, B., Tang, W., Wehr, M. C., … et al. (2009). Split-cre complementation indicates coincident activity of different genes in vivo. *PLoS One 4*, e4286.

Hollenback, S.M., Lyman, S., & Cheng, J. (2011). Recombineering-based procedure for creating BAC transgene constructs for animals and cell lines. *Curr Protoc Mol Biol* Chapter 23, Unit 23 14.

Hunter, N. L., Awatramani, R. B., Farley, F. W., & Dymecki, S.M. (2005). Ligand-activated flpe for temporally regulated gene modifications. *Genesis 41*, 99–109.

Hyden, H., & McEwen, B. (1966). A glial protein specific for the nervous system. *Proc Natl Acad Sci U S A 55*, 354–358.

Indra, A. K., Warot, X., Brocard, J., Bornert, J. M., Xiao, J. H., Chambon, P., & Metzger, D. (1999). Temporally-controlled site-specific mutagenesis in the basal layer of the epidermis: comparison of the recombinase activity of the tamoxifen-inducible Cre-ERT and Cre-ERT2 recombinases. *Nucleic Acids Res 27*, 4324–4327.

Islam, S., Kjallquist, U., Moliner, A., Zajac, P., Fan, J. B., Lonnerberg, P., & Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res 21*, 1160–1167.

Jaenisch, R. (1976). Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. *Proc Natl Acad Sci U S A 73*, 1260–1264.

Jensen, K. B., & Darnell, R. B. (2008). CLIP: crosslinking and immunoprecipitation of in vivo RNA targets of RNA-binding proteins. *Methods Mol Biol 488*, 85–98.

Johansson, C. B., Lothian, C., Molin, M., Okano, H., & Lendahl, U. (2002). Nestin enhancer requirements for expression in normal and injured adult CNS. *J Res 69*, 784–794.

Jones, A. R., Overly, C. C., & Sunkin, S.M. (2009). The Allen Brain Atlas: 5 years and beyond. *Nat Rev Neurosci 10*, 821–828.

Kam, C. M., Abuelyaman, A. S., Li, Z. Z., Hudig, D., & Powers, J. C. (1993). Biotinylated isocoumarins, new inhibitors and reagents for detection, localization, and isolation of serine proteases. *Bioconjugate Chem 4*, 560–567.

Kang, H. J., Kawasawa, Y. I., Cheng, F., Zhu, Y., Xu, X., Li, M.,…et al. (2011). Spatio-temporal transcriptome of the human brain. *Nature 478*, 483–489.

Kislinger, T., Cox, B., Kannan, A., Chung, C., Hu, P., Ignatchenko, A.,…et al. (2006). Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell 125*, 173–186.

Kranz, A., Fu, J., Duerschke, K., Weidlich, S., Naumann, R., Stewart, A.F., & Anastassiadis, K. (2010). An improved Flp deleter mouse in C57Bl/6 based on Flpo recombinase. *Genesis 48*, 512–520.

Kriaucionis, S., & Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science 324*, 929–930.

Kuhn, A., Thu, D., Waldvogel, H. J., Faull, R. L., & Luthi-Carter, R. (2011). Population-specific expression analysis (PSEA) reveals molecular changes in diseased brain. *Nat Methods 8*, 945–947.

Lai, T., Jabaudon, D., Molyneaux, B. J., Azim, E., Arlotta, P., Menezes, J.R.L., & Macklis, J. D. (2008). SOX5 controls the sequential generation of distinct corticofugal neuron subtypes. *Neuron 57*, 232–247.

Laird, P.W. (2010). Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet 11*, 191–203.

Lam, K. S., Aman, M. G., & Arnold, L. E. (2006). Neurochemical correlates of autistic disorder: a review of the literature. *Res Dev Disabil 27*, 254–289.

Lashley, K. S. (1930). Basic neural mechanisms in behavior. *Psychol Rev 37*, 1–24.

Leclerc, N., Schwarting, G. A., Herrup, K., Hawkes, R., & Yamamoto, M. (1992). Compartmentation in mammalian cerebellum: Zebrin II and P-path antibodies define three classes of sagittally organized bands of Purkinje cells. *Proc Natl Acad Sci U S A 89*, 5006–5010.

Lee, C. K., Sunkin, S. M., Kuan, C., Thompson, C. L., Pathak, S., Ng, L.,…et al. (2008). Quantitative methods for genome-scale analysis of in situ hybridization and correlation with microarray data. *Genome Biol 9*, R23.

Lee, S. H., DeCandia, T. R., Ripke, S., Yang, J., Sullivan, P. F., Goddard, M. E.,…et al. (2012). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nature Genet 44*, 247–U235.

Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A. F.,…et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature 445*, 168–176.

Letzkus, J. J., Wolff, S.B.E., Meyer, E.M.M., Tovote, P., Courtin, J., Herry, C., & Luthi, A. (2011). A disinhibitory microcircuit for associative fear learning in the auditory cortex. *Nature 480*, 331–U376.

Lin, J. Y. (2011). A user's guide to channelrhodopsin variants: features, limitations and future developments. *Exp Physiol 96*, 19–25.

Lobo, M. K., Karsten, S. L., Gray, M., Geschwind, D. H., & Yang, X. W. (2006). FACS-array profiling of striatal projection neuron subtypes in juvenile and adult mouse brains. *Nat Neurosci 9*, 443–452.

Looger, L. L., & Griesbeck, O. (2012). Genetically encoded neural activity indicators. *Curr Opin Neurobiol 22*, 18–23.

Madisen, L., Zwingman, T. A., Sunkin, S. M., Oh, S. W., Zariwala, H. A., Gu, H.,…et al. (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat Neurosci 13*, 133–U311.

Magdaleno, S., Jensen, P., Brumwell, C. L., Seal, A., Lehman, K., Asbury, A.,…et al. (2006). BGEM: an in situ hybridization database of gene expression in the embryonic and adult mouse nervous system. *PLoS Biol 4*, e86.

Mary, P., Dauphinot, L., Bois, N., Potier, M. C., Studer, V., & Tabeling, P. (2011). Analysis of gene expression at the single-cell level using microdroplet-based microfluidic technology. *Biomicrofluidics 5*.

Masseck, O. A., Rubelowski, J. M., Spoida, K., & Herlitze, S. (2011). Light- and drug-activated G-protein-coupled receptors to control intracellular signalling. *Exp Physiol 96*, 51–56.

Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., D'Souza, C., Fouse, S.D.,…et al. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature 466*, 253–U131.

Miller, E. K., & Wilson, M. A. (2008). All my circuits: using multiple electrodes to understand functioning neural networks. *Neuron 60*, 483–488.

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

Molyneaux, B. J., Arlotta, P., Hirata, T., Hibi, M., & Macklis, J. D. (2005). Fezl is required for the birth and specification of corticospinal motor neurons. *Neuron 47*, 817–831.

Muers, M. (2011). Functional genomics: The modENCODE guide to the genome. *Nat Rev Genet 12.* Nrg2942.

Myers, R. M., Stamatoyannopoulos, J., Snyder, M., Dunham, I., Hardison, R. C., Bernstein, B.E., … et al. (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol 9*, e1001046.

Nakatani, J., Tamada, K., Hatanaka, F., Ise, S., Ohta, H., Inoue, K., … et al. (2009). Abnormal behavior in a chromosome-engineered mouse model for human 15q11-13 duplication seen in autism. *Cell 137*, 1235–1246.

Nelson, S. B., Sugino, K., & Hempel, C. M. (2006). The problem of neuronal cell types: a physiological genomics approach. *Trends Neurosci 29*, 339–345.

NHGRI (2012). http://www.genome.gov/sequencingcosts/.

Nunzi, M. G., Shigemoto, R., & Mugnaini, E. (2002). Differential expression of calretinin and metabotropic glutamate receptor mGluR1alpha defines subsets of unipolar brush cells in mouse cerebellum. *J Comp Neurol 451*, 189–199.

Oberdick, J., Smeyne, R. J., Mann, J. R., Zackson, S., & Morgan, J. I. (1990). A promoter that drives transgene expression in cerebellar Purkinje and retinal bipolar neurons. *Science 248*, 223–226.

Okaty, B. W., Miller, M. N., Sugino, K., Hempel, C. M., & Nelson, S. B. (2009). Transcriptional and electrophysiological maturation of neocortical fast-spiking GABAergic interneurons. *J Neurosci* (the official journal of the Society for Neuroscience) *29*, 7040–7052.

Oldham, M. C., Horvath, S., & Geschwind, D. H. (2006). Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc Natl Acad Sci U S A 103*, 17973–17978.

Oldham, M.C., Konopka, G., Iwamoto, K., Langfelder, P., Kato, T., Horvath, S., and Geschwind, D.H. (2008). Functional organization of the transcriptome in human brain. *Nat Neurosci 11*, 1271–1282.

Oyibo, H., Cao, G., Zhan, H., Koulakov, A., Enquist, L., Dubnau, J., & Zador, A. (2011). Probing the connectivity of neural circuits at single-neuron resolution using high-throughput DNA sequencing. Paper presented at the Computational and Systems Neuroscience Meeting (Cosyne) (Salt Lake City, Utah, Nature Proceedings).

Palmiter, R. D., & Brinster, R. L. (1986). Germ-line transformation of mice. *Annu Rev Genet 20*, 465–499.

Passman, R. S., & Fishman, G. I. (1994). Regulated expression of foreign genes in vivo after germline transfer. *J Clin Invest 94*, 2421–2425.

Patti, G. J., Woo, H. K., Yanes, O., Shriver, L., Thomas, D., Uritboonthai, W., … Siuzdak, G. (2010). Detection of carbohydrates and steroids by cation-enhanced nanostructure-initiator mass spectrometry (NIMS) for biofluid analysis and tissue imaging. *Anal Chem 82*, 121–128.

Patti, G. J., Yanes, O., & Siuzdak, G. (2012). Innovation: Metabolomics: the apogee of the omics trilogy. Nat *Rev Mol Cell Biol 13*, 263–269.

Penfield, W., & Rasmussen, T. (1950). *The cerebral cortex of man; a clinical study of localization of function*. Oxford, UK: Macmillan.

Portales-Casamar, E., Swanson, D. J., Liu, L., de Leeuw, C. N., Banks, K. G., Ho Sui, S. J., … et al. (2010). A regulatory toolbox of MiniPromoters to drive selective expression in the brain. *Proc Natl Acad Sci U S A 107*, 16589–16594.

Poser, I., Sarov, M., Hutchins, J. R., Heriche, J. K., Toyoda, Y., Pozniakovsky, A., … et al. (2008). BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat Methods 5*, 409–415.

Ragan, T., Kadiri, L. R., Venkataraju, K. U., Bahlmann, K., Sutin, J., Taranda, J., … Osten, P. (2012). Serial two-photon tomography for automated ex vivo mouse brain imaging. *Nat Methods 9*, 255–258.

Ramón y Cajal, S., Pasik, P., & Pasik, T. (1899). *Texture of the nervous system of man and the vertebrates* (pp. v, 1–3).Vienna and New York: Springer.

Rodriguez, C. I., Buchholz, F., Galloway, J., Sequerra, R., Kasper, J., Ayala, R., … Dymecki, S.M. (2000). High-efficiency deleter mice show that FLPe is an alternative to Cre-loxP. *Nat Genet 25*, 139–140.

Rolls, A., Colas, D., Adamantidis, A., Carter, M., Lanre-Amos, T., Heller, H. C., & de Lecea, L. (2011). Optogenetic disruption of sleep continuity impairs memory consolidation. *Proc Natl Acad Sci U S A 108*, 13305–13310.

Roth, L. J., & Barlow, C. F. (1961). Drugs in the brain. *Science 134*, 22–31.

Saez, E., No, D., West, A., & Evans, R.M. (1997). Inducible gene expression in mammalian cells and transgenic mice. *Curr Opin Biotechnol 8*, 608–616.

Saito, K., Barber, R., Wu, J. Y., Matsuda, T., Roberts, E., & Vaughn, J. E. (1974). Immunohistochemical Localization of Glutamate Decarboxylase in Rat Cerebellum. *Proc Natl Acad Sci U S A 71*, 269–273.

Sanz, E., Yang, L., Su, T., Morris, D. R., McKnight, G. S., & Amieux, P. S. (2009). Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. *Proc Natl Acad Sci U S A 106*, 13939–13944.

Scalbert, A., Brennan, L., Fiehn, O., Hankemeier, T., Kristal, B. S., van Ommen, B.,… Wopereis, S. (2009). Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics 5*, 435–458.

Schmidt, E. F., Warner-Schmidt, J. L., Otopalik, B. G., Pickett, S. B., Greengard, P., & Heintz, N. (2012). Identification of the Cortical Neurons that Mediate Antidepressant Responses. *Cell 149*, 1152–1163.

Selimi, F., Cristea, I. M., Heller, E., Chait, B. T., & Heintz, N. (2009). Proteomic studies of a single CNS synapse type: the parallel fiber/purkinje cell synapse. *PLoS Biol 7*, e83.

Shcherbo, D., Murphy, C. S., Ermakova, G. V., Solovieva, E. A., Chepurnykh, T. V., Shcheglov, A. S.,… et al. (2009). Far-red fluorescent tags for protein imaging in living tissues. *Biochem J 418*, 567–574.

Shen-Orr, S. S., Tibshirani, R., Khatri, P., Bodian, D. L., Staedtler, F., Perry, N. M.,… Butte, A.J. (2010). Cell type-specific gene expression differences in complex tissues. *Nat Methods 7*, 287–289.

Shimogori, T., Lee, D. A., Miranda-Angulo, A., Yang, Y., Wang, H., Jiang, L.,… et al. (2010). A genomic atlas of mouse hypothalamic development. *Nat Neurosci 13*, 767–775.

Shimshek, D. R., Kim, J., Hubner, M. R., Spergel, D. J., Buchholz, F., Casanova, E.,… Sprengel, R. (2002). Codon-improved Cre recombinase (iCre) expression in the mouse. *Genesis 32*, 19–26.

Shuen, J. A., Chen, M., Gloss, B., & Calakos, N. (2008). Drd1a-tdTomato BAC transgenic mice for simultaneous visualization of medium spiny neurons in the direct and indirect pathways of the basal ganglia. *J Neurosci 28*, 2681–2685.

Siegel, R. W., Jain, R., & Bradbury, A. (2001). Using an in vivo phagemid system to identify non-compatible loxP sequences (vol. 499, p. 147). *FEBS Lett 505*, 466–473.

Smedley, D., Salimova, E., & Rosenthal, N. (2011). Cre recombinase resources for conditional mouse mutagenesis. *Methods 53*, 411–416.

Sotelo, C. (2003). Viewing the brain through the master hand of Ramon y Cajal. *Nat Rev Neurosci 4*, 71–77.

Strack, R. L., Strongin, D. E., Bhattacharyya, D., Tao, W., Berman, A., Broxmeyer, H. E.,… Glick, B.S. (2008). A noncytotoxic DsRed variant for whole-cell labeling. *Nat Methods 5*, 955–957.

Sutherland, G. R., & McNaughton, B. (2000). Memory trace reactivation in hippocampal and neocortical neuronal ensembles. *Curr Opin Neurobiol 10*, 180–186.

Takada, M., Sugimoto, T., & Hattori, T. (1993). Tyrosine hydroxylase immunoreactivity in cerebellar Purkinje cells of the rat. *Neurosci Lett 150*, 61–64.

Taniguchi, H., He, M., Wu, P., Kim, S., Paik, R., Sugino, K.,… et al. (2011). A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron 71*, 995–1013.

Terskikh, A., Fradkov, A., Ermakova, G., Zaraisky, A., Tan, P., Kajava, A. V.,… et al. (2000). "Fluorescent timer": protein that changes color with time. *Science 290*, 1585–1588.

Tian, L., Akerboom, J., Schreiter, E.R., & Looger, L. L. (2012). Neural activity imaging with genetically encoded calcium indicators. *Prog Brain Res 196*, 79–94.

van der Weyden, L., & Bradley, A. (2006). Mouse chromosome engineering for modeling human disease. *Annu Rev Genom Hum Genet 7*, 247–276.

Verhaak, R. G., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D.,… et al. (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell 17*, 98–110.

Visel, A., Thaller, C., & Eichele, G. (2004). GenePaint.org: an atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res 32*, D552–556.

Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S.,… Geschwind, D.H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature 474*, 380–384.

Waddington, C. H. (2012). The epigenotype. 1942. *Int J Epidemiol 41*, 10–13.

Wall, N. R., Wickersham, I. R., Cetin, A., De La Parra, M., & Callaway, E. M. (2010). Monosynaptic circuit tracing in vivo through Cre-dependent targeting and complementation of modified rabies virus. *Proc Natl Acad Sci USA 107*, 21848–21853.

Wang, J., Williams, R. W., & Manly, K. F. (2003). WebQTL: web-based complex trait analysis. *Neuroinformatics 1*, 299–308.

Warner-Schmidt, J. L., Schmidt, E. F., Marshall, J. J., Rubin, A. J., Arango-Lievano, M., Kaplitt, M. G.,… Greengard, P. (2012). Cholinergic interneurons in the nucleus accumbens regulate depression-like behavior. *Proc Natl Acad Sci U S A 109*, 11360–11365.

Weissman, T. A., Sanes, J. R., Lichtman, J. W., & Livet, J. (2011). Generating and imaging multicolor Brainbow mice. *Cold Spring Harb Protoc* 2011, 763–769.

Xu, X. & Dougherty, J.D. (2013), Cell Type Specific Analysis of Human Brain Transcriptome Data to Predict Alterations in Cellular Composition. *Systems Biomedicine* (accepted).

Yang, X. W., Model, P., & Heintz, N. (1997). Homologous recombination based modification in Escherichia coli and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nat Biotechnol 15*, 859–865.

Yanushevich, Y. G., Gurskaya, N. G., Staroverov, D. B., Lukyanov, S. A., & Lukyanov, K. A. (2003). A natural fluorescent protein that changes its fluorescence color during maturation. *Russ J Bioorg Chem 29*, 325–329.

Yuste, R. (2005). Origin and classification of neocortical interneurons. *Neuron 48*, 524–527.

Zhang, X., Guo, C., Chen, Y., Shulha, H. P., Schnetz, M. P., LaFramboise, T.,…et al. (2008). Epitope tagging of endogenous proteins for genome-wide ChIP-chip studies. *Nat Methods 5*, 163–165.

Zhang, X. M., Chen, B. Y., Ng, A. H., Tanner, J. A., Tay, D., So, K. F.,… Huang, J.D. (2005). Transgenic mice expressing Cre-recombinase specifically in retinal rod bipolar neurons. *Invest Ophthalmol Vis Sci 46*, 3515–3520.

Zhao, S., Ting, J. T., Atallah, H. E., Qiu, L., Tan, J., Gloss, B.,…et al. (2011a). Cell type-specific channelrhodopsin-2 transgenic mice for optogenetic dissection of neural circuitry function. *Nat Methods 8*, 745–752.

Zhao, Y., Araki, S., Wu, J., Teramoto, T., Chang, Y. F., Nakano, M.,…et al. (2011b). An expanded palette of genetically encoded Ca(2)(+) indicators. *Science 333*, 1888–1891.

Zhong, Y., Wang, Q. J., Li, X., Yan, Y., Backer, J. M., Chait, B. T.,… Yue, Z. (2009). Distinct regulation of autophagic activity by Atg14L and Rubicon associated with Beclin 1-phosphatidylinositol-3-kinase complex. *Nat Cell Biol 11*, 468–476.

Zhou, H., Huang, C., Yang, M., Landel, C. P., Xia, P. Y., Liu, Y.J., & Xia, X. G. (2009). Developing tTA transgenic rats for inducible and reversible gene expression. *Int J Biol Sci 5*, 171–181.