Published in final edited form as:
Cell. 2008 November 14; 135(4): 749–762. doi:10.1016/j.cell.2008.10.029.

Application of BACarray for comparative analysis of CNS cell types

Joseph P. Doyle[1,*], **Joseph D. Dougherty**[1,*], **Myriam Heiman**[2], **Eric F. Schmidt**[1], **Tanya R. Stevens**[1], **Guojun Ma**[1], **Sujata Bupp**[1], **Prerana Shrestha**[1], **Rajiv D. Shah**[1], **Martin L. Doughty**[3], **Shiaoching Gong**[1,3], **Paul Greengard**[2], and **Nathaniel Heintz**[1,3]

[1] Laboratory of Molecular Biology, Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA

[2] Laboratory of Molecular and Cellular Neuroscience, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA

[3] GENSAT Project, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA

Summary

Comparative analysis of BACarray data can provide important insights into complex biological systems. As demonstrated in the accompanying paper, BACarray translational profiling permits comprehensive studies of translated mRNAs in genetically defined cell populations following physiological perturbations. To establish the generality of this approach, we present BACarray translational profiles for twenty four CNS cell populations, and identify known cell-specific and enriched transcripts for each population. We report thousands of cell-specific mRNAs that were not detected in whole tissue microarray studies, and provide examples that demonstrate the benefits deriving from comparative analysis. To provide a foundation for further biological and *in silico* studies, we provide a resource of sixteen transgenic mouse lines, their corresponding anatomic characterization, and BACarray translational profiles for cell types from a variety of CNS structures. This resource will enable a wide spectrum of molecular and mechanistic studies of both well known and previously uncharacterized neural cell populations.

Introduction

The histological, molecular and biochemical complexities of the mammalian brain present a serious challenge for mechanistic studies of brain development, function and dysfunction. To provide a foundation for these studies, we applied several classical principles to the exploration of anatomical and functional diversity in the mouse central nervous system (CNS). First, as exemplified by Ramon y Cajal, detailed comparative analysis of myriad cell types can permit strong inferences about their specific contributions to CNS function (Ramon y Cajal et al., 1899). Second, as demonstrated from invertebrate studies, a deep understanding of the contributions of specific cells to behavior, can best be achieved if one has reproducible, efficient genetic access to these cell populations *in vivo* (Bargmann, 1993; Zipursky and Rubin, 1994). Third, as illustrated by detailed studies of signal transduction in striatal medium spiny neurons

Contact: Dr. Nathaniel Heintz, Laboratory of Molecular Biology, HHMI/The Rockefeller University, 1230 York Avenue, Box 260, New York, NY 10065, Tel: 212-327-7955, Fax: 212-327-7878, heintz@mail.rockefeller.edu.
*These authors contributed equally to this work.

(Greengard, 2001; Svenningsson et al., 2004), the highly specialized properties of even closely related neurons arise from the combined actions of their many protein components.

Previously, we have broadly applied the BAC transgenic strategy (Heintz, 2004; Yang et al., 1997) to provide high resolution anatomical data and BAC vectors for genetic studies of morphologically defined cells in the CNS (Gong et al., 2003). In the accompanying paper (Heiman et al., 2008), we have reported the development of the BACarray translational profiling; a methodology for the discovery of the complement of proteins synthesized in any genetically defined cell population. Here we describe the generation of BACarray transgenic mice and translational profiles for twenty four distinct cell populations, including all of the major cerebellar cell types. We also demonstrate some of the analytical tools that can be employed for comparative analysis of selected cell types, and illustrate as an example of this analysis, the many features of spinal motor neurons that can be discovered using this approach.

As anticipated in the studies of Heiman et al., this resource will allow molecular phenotyping of CNS cell types at specified developmental stages, and in response to a variety of pharmacological, genetic or behavioral alterations. The mice and data we present here confirm the generality of the BACarray approach and provide an important new resource for studies of the molecular bases for cellular diversity in the mouse brain.

Results

Selection of BAC drivers to target specific CNS cell types

As illustrated by Heiman et al, the generation of BACarray translational profiles requires accurate targeting of the EGFP-L10a ribosomal protein fusion to desired CNS cell types, affinity purification of polysomal RNAs from these cell types, and interrogation of the resultant mRNA populations. We selected BACs reported by the GENSAT Project to specifically target a wide range of neurons and glia from different structures throughout the CNS, including BAC drivers expected to target less well-defined populations (www.gensat.org).

Anatomic characterization of BACarray transgenic mouse lines

To ensure that expression of the EGFP-L10a fusion protein is accurate, and to clearly define the cell types to be further analyzed by BACarray, detailed anatomic studies were conducted. For each line, transgene expression was carefully assayed by immunohistochemistry (IHC) using an antibody against EGFP (Figure 1). The regions covered in this survey include cerebellum (panels 1–9), spinal cord (10), basal forebrain and corpus striatum (11–14), brainstem (15), and cerebral cortex (16–25).

For well characterized cell types, confirmation of transgene targeting was straight forward. For example, one can easily identify Purkinje cells (*Pcp2*, panel 1), granule cells (*Neurod1*, panel 2), Golgi neurons (*Grm2*, panel 3), and Bergmann glia (*Sept4*, panel 8), based on their morphology and position in the cerebellum. Two points can be made from this analysis. First, the expression of the EGFP-L10a transgene from each BAC driver is correct, conforming both to the published literature and to the GENSAT atlas. Second, the cytoplasmic distribution of the EGFP-L10a fusion protein, while more limited than soluble EGFP, provides sufficient morphological detail to unambiguously identify well described CNS cell types.

However, in many cases cell identity cannot be assigned by morphology and regional position alone. Therefore, we confirmed the presumed cellular identity using double immunofluorescence (IF) for the EGFP-L10a fusion protein and cell type specific markers (Figure 2). In most cases, these studies established that any further analyses would be restricted to a well-defined cell population, such as Purkinje cells (Figure 2B), Golgi cells (Figure 2E), or glial cell types including astrocytes (*Aldh1L1*), mature oligodendrocytes (*Cmtm5*), and a

mixed oligodendroglial line that included mature oligodendrocytes and oligodendrocyte progenitors (*Olig2*) (Figure S1).

These studies also identified BACarray lines in which the transgene is expressed in two or more cell types. For example, the IF analysis of the *Lypd6* BACarray line (Figure 2D) revealed that EGFP-L10a is found in all Pvalb positive and NeuN negative interneurons of the cerebellar molecular layer, suggesting that this line targets both stellate and basket cells. Also, in certain lines it is apparent that the transgene is expressed in only a subset of a particular cell type. For instance, in the *Grp* BACarray line the EGFP-L10a fusion protein is restricted to the subpopulation of unipolar brush cells (Nunzi et al., 2002) which are immunoreactive for Grm1 but not Calb2 (calretinin) (Figure 2F).

BACarray lines that express as anticipated from GENSAT, but do not conform to readily identified cell types, were also analyzed by IF analysis to provide data concerning the broad classification of cell populations targeted. For example, in the cerebral cortex of the *Cort* BACarray line, Calb1 was detected in nearly 50% of EGFP-L10a positive cells, Pvalb was found in less than 5% of these cells, and Calb2 was not detected (data not shown). In *Pnoc* BACarray mice, the majority of EGFP-L10a positive cells in the superficial layers of the cerebral cortex were multipolar and GABA positive, although some cells in deeper layers of cortex were GABA negative and appeared to have a single apical dendrite. The multipolar cells in this case were often positive for Calb2, but not Calb1 or Pvalb (data not shown). Both IHC and IF studies of the cortex of the *Cck* BACarray line clearly demonstrate that EGFP-L10a is detected in small neurons positive for Calb1 but not Pvalb or Calb2, as well as in pyramidal cells (data not shown), consistent with previous *in situ* hybridization (ISH) data (www.stjudebgem.org; www.brain-map.org) (Lein et al., 2007; Magdaleno et al., 2006).

Unfortunately, good markers are not known for every cell type. For example, the markers for cortical interneurons assayed above have only limited correspondence to physiological properties of the cells (Markram et al., 2004), and markers for various pyramidal cell populations have not been established. Rather, since the initial studies of Ramon y Cajal (Ramon y Cajal et al., 1899) and Lorente de No (Lorente de No, 1934), projection neurons in the cerebrum have been identified by their pyramidal shape, and broadly classified by their laminar specificity, dendritic arbor, and axonal targets. Accordingly, we have produced lines which clearly label large pyramidal cells of layers 6 (*Ntsr1*, panel 16), 5b (*Glt25d2*, panel 17), and 5a (*Etv1*, panel 18). Although axons were not clearly labeled in these EGFP-L10a BACarray mice, morphometric studies provide additional data indicating that the GENSAT EGFP lines and EGFP-L10a BACarray lines target similar cortical pyramidal cell populations (Figure S2). In the corresponding GENSAT lines these cell populations were shown to project to the thalamus, pons/spinal cord, and striatum, respectively (www.gensat.org).

It is important to note that in most of the BACarray lines, the EGFP-L10a fusion protein is detected in multiple CNS structures. A salient example is the cholinergic cell populations targeted in the *Chat* BACarray lines. In this case, we have clearly demonstrated correct expression in spinal cord motor neurons, neurons of the corpus striatum, basal forebrain projection neurons, brainstem motor neurons (Figure 1, panels:10,13,14,15) and neurons of the medial habenula (data not shown). As detailed below, we have collected BACarray translational profiles for the first four of these cholinergic cell populations by separately dissecting these regions prior to affinity purification of the EGFP-L10a tagged polysome populations. Likewise, we assayed the glial cell lines in both cerebellar and cortical tissue. Since specifically expressed genes are often found in distinct cell types from physically separable brain structures, the lines we present here offer opportunities for the study of additional cell types.

BACarray polysome purification, RNA extraction and control microarray experiments

In total, we identified twenty four cell populations in five regions that we chose to assay by BACarray translational profiling (Heiman et al., 2008). As shown in Figure S3, this procedure yielded the purification of EGFP-ribosomal fusion protein along with cell-specific mRNAs. We also harvested RNA from the unbound (UB) fraction of the immunoprecipitation to measure the genes expressed in the dissected region as a whole.

As shown by Heiman et al., and in Figure S4A, replicates for the same cell type gave nearly identical genome wide translational profiles. The average Pearson's correlation between independent replicates was above .98 across all cell types. To determine whether the transgene's integration position would influence the data, we also examined independent BACarray lines prepared with the same engineered BAC. This analysis revealed that the variation between independent founder lines was low, and no more extensive than for replicate samples isolated from the same founder line (Figure S4D). Thus, the location of the transgene insertion into the genome had little global impact on the data. Finally, we tested four different custom monoclonal antibodies and one goat polyclonal against EGFP. Each antibody immunoprecipitated comparable levels of mRNA and yielded similar global gene translational profiles (data not shown). Thus, the monoclonal antibodies, a renewable reagent for future BACarray studies, were used for the remainder of the work.

We noticed that a small number of probesets (Table S2) are consistently enriched in every BACarray dataset analyzed. Since these same probesets were also enriched in immunoprecipitates from control mice with no transgene expression, we conclude that they represent background which we systematically eliminated from further analysis.

BACarray data analysis and confirmation

To provide a measure of the enrichment for each mRNA immunoprecipitated from the targeted cell type (IP) versus its expression in the tissue sample dissected for the analysis (UB), we calculated the ratio of IP/UB. Figure S4B shows scatter plots for three representative cell types of the cerebellum. Dramatic differences are evident between the genome-wide translational profiles of IP samples compared to whole tissue, with each cell population displaying a unique profile of thousands of enriched genes (Figure S4C). Venn diagrams of the top 1000 most enriched probesets for each cell type illustrate this point. Thus, approximately 75% of the enriched probesets are not shared between Purkinje cells, granule cells and unipolar brush cells, and only 52 of the probesets enriched in these three cell types are shared between them. To aid in the use of these lines, and allow users to investigate mRNAs in specific CNS cell types, IP/UB data for each cell type is presented in Table S5 online.

To determine if this methodology accurately enriched for cell-specific genes, we examined the BACarray data for known markers (positive controls) for each cell type. We also examined genes expressed exclusively in other cell types (negative controls). Figure 3A shows a scatter plot of IP/UB for spinal cord motor neurons. Probesets for markers of motor neurons with measurable signal (green dots) are clearly enriched in the IP sample, whereas probesets for glial-specific RNAs (red dots, negative controls), are clearly enriched in the UB sample. To establish the generality of this finding, we quantified the enrichment by calculating an average ratio of IP/UB for positive and negative controls for each cell type with at least three known markers. As shown in Figure 3B, all IPs showed a clear enrichment for appropriate known markers, (Figure 3B, plotted in log base 2). Even for cell types with only one known marker (such as *Pnoc* or *Grp* positive cells), probesets for these genes were consistently enriched in the IP. In the IPs with the lowest relative yield of RNA, such as those for mature oligodendrocytes (Figure 3B), and *Cort* expressing interneurons (not shown), background was

proportionally higher, and enrichment was less robust. Nonetheless, BACarray successfully identifies the known markers for these cells as well.

We next attempted to identify novel cell-specific markers for rare cell types. For this, we screened eleven genes predicted by BACarray to be enriched in either the *Pnoc* expressing cells of the cerebral cortex, or *Grm2* expressing cerebellar Golgi cells with confocal microscopy for fluorescent ISH and IF for EGFP-L10a. For the nine genes where ISH gave clear results, all were clearly overlapping with EGFP-L10a (Figure 3 and data not shown).

In Golgi cells, there is a high degree of overlap between EGFP-L10a expression in the BACarray line and expression of the ISH analysis (Figure 3C). This substantial overlap confirms the specificity of the results we have obtained for this and other cell types. Nonetheless, the enrichment of a particular mRNA in the IP sample cannot be used to conclude that it is expressed in the cell type exclusively or that it is expressed in all cells of that type. For example, consistent with the ISH databases (www.stjudebgem.org; www.brain-map.org) our data clearly indicate that *Penk1* is expressed in Golgi cells and in scattered cells in the molecular layer (Figure 3C, panel 1). Finally, some mRNAs were not detected using the ISH technique, perhaps reflecting limited sensitivity of ISH for poorly expressed genes (Figure 3C, panel 6).

In order to validate the quantitative aspects of the BACarray datasets, we measured the enrichment of a variety of mRNAs isolated from the *Chat* (motor neuron) and *Pcp2* (Purkinje cell) BACarray transgenic lines with quantitative real time PCR (qRT-PCR) (Figure 3D). For all of the control genes tested, this methodology confirmed our BACarray results. For genes not previously known to be expressed in a specific cell type, results from qRT-PCR demonstrated that seven out of the eight mRNAs assayed were in fact cell type enriched. Moreover, despite an inconclusive ISH result (Figure 3C, panel 6), qRT-PCR validated the expression of *Ceacam10* in the cerebellum and its enrichment in Golgi cells (Figure 3D). In some cases, therefore, the BACarray methodology appears to be more sensitive than ISH.

Comparative analysis of BACarray data collected from many cell types

Having established that the BACarray data accurately reflect expression of known controls for each cell type, and can be confirmed by independent experimental analysis (Heiman et al., 2008), we were next interested in illustrating the broad properties of these cells that could be inferred from their comparative analysis. We first performed a hierarchical clustering of all twenty four IP and six UB samples using the 20% of probesets with the highest coefficient of variation (Figure 4A). This unsupervised clustering essentially recapitulates the known biology of CNS cell types. Thus, the three populations of cortical projection neurons are more similar to one another than they are to cortical interneurons, Purkinje cells, or motor neurons. Astroglial BACarray data collected from different regions of the brain are, as expected, more similar to one another and to Bergmann glia than they are to oligodendrocytes. Oligodendroglia are more similar to each other than they are to any neuronal population, etc. These findings support the concept that cells with similar gene expression patterns share similar functions, and suggest that analysis of BACarray data will permit the identification of those gene products responsible for the distinguishing characteristics of each cell type.

There are some surprising features to this analysis. Remarkably, the diversity of translational profiles across neuronal types nearly rivals the diversity between neurons and glia. Although related cell subtypes, such as different motor neurons, are clearly tightly clustered, many neuronal types (e.g. Purkinje cells) are not strongly clustered with any other cell type. This suggests that comparative analysis of BACarray translational profiles obtained from highly specialized cell types may yield unanticipated insights into their biochemical properties. Finally, individual cell types did not generally cluster tightly with their tissue of origin. In fact,

profiles from different brain regions were clustered more closely to each other than to their respective IP samples, suggesting that microarray data produced from dissected regions are significantly less informative than BACarray analysis of individual cell types.

To examine this point in more detail, we compared the data from total cerebellum to that of the individual cerebellar cell types analyzed in this study. As can be seen in Figure 4B, any single cell type has fewer probesets detectable than the whole cerebellar sample, since the whole cerebellar sample represents an aggregate of different cell types. However, taking the union of the probesets detectable in each of the six individual cerebellar cell types reveals over 4000 probesets that are undetectable in the microarray from whole cerebellum. Importantly, these undetectable probesets tend to represent cell-type enriched genes (Figure 4C). In fact, for rare cell types, up to 42% of the genes enriched in that cell may not be detectable at all in whole tissue microarray studies. For detection of genes expressed in specific cell types within complex brain regions, therefore, the BACarray methodology is more sensitive than microarray analysis of dissected brain regions.

The increased sensitivity of the BACarray methodology results in identification of more mRNAs in each cell type, yielding a more complete picture of the translational profile for each cell type or, simply put, more information. To assess if this increased sensitivity in fact does give better information, we calculated the Shannon entropy for each probeset across the six whole tissue samples, and across the twenty four individual cell populations (Fuhrman et al., 2000; Shannon and Weaver, 1969). Shannon entropy is a measure of information content that describes the complexity of a signal across samples, with values ranging from 0 (low information) to 2 (high information). The average Shannon entropy in cell type specific experiments (IP's) is over twice as high as that calculated from microarray data of whole tissue samples (Figure 5A) (T-test, $p<.0001$, average across all IPs: 0.88 +/− .002, whole tissue: .41 +/− .003). Examples of probesets with low and high information are shown in Figure 5B. This analysis directly demonstrates that microarray data collected from specific cell types using the BACarray strategy can provide significantly better information than traditional microarray studies of dissected brain tissues.

Using this measure of information, we next classified the ten percent of the probesets with either the highest or the lowest entropy (Figure 5C) with Gene Ontologies and searched for functional categories that were significantly over-represented. According to this analysis, cell type diversity in the nervous system is driven primarily by the expression of cell surface proteins, such as channels and receptors, and also to some extent by the specific expression of transcription factors and calcium binding proteins. Genes with less information content tend to be those that are more ubiquitously expressed, such as ribosomal and mitochondrial proteins, the expression of which certainly does vary across cell types, but much less dramatically than that of receptors and channels.

Comparative analysis of translational profiling across a large number of cell populations may also identify novel co-regulated genes that encode the highly specialized properties of individual cell types. To test this, we selected a probeset for a gene known to be involved in myelination – the myelin basic protein (*Mbp*). We next examined its highest correlates across all samples. In the top 35 genes correlating with *Mbp* expression (min correlation, .86), we identified 6 genes also involved in myelination, including *Plp1*, *Cnp*, *Mog*, *Mal* and *Mobp*, and another three genes previously identified in a proteomic screen of myelin components (Table S6), in addition to many novel genes that could contribute to myelination. Although this type of correlative information will become increasingly useful as more BACarray data become available, this experiment provides an illustrative example that the large amount of information we have provided in this study can already be of use for generation of hypotheses concerning the biological functions of poorly studied CNS expressed genes.

Next, to elucidate those genes potentially involved in specifying cell type, we undertook a comparative analysis to identify the most highly specific genes in each population. We performed an iterative comparison: one-by-one, each sample was compared to each other sample in the dataset, and for each population, probesets were sorted by their average ranking across these comparisons. We then combined and clustered by expression the top one hundred ranked probesets for each population in a heatmap (Figure 6A). This illustrates the extent to which distinct cell types are characterized by specific cohorts of genes. For example, none of the top twenty five most specific probesets observed in the Purkinje cell sample are found in any of the top twenty five most specific probesets for any of the other cell types (Figure 6B). In contrast, *Drd1* and *Drd2* medium spiny neurons, two closely related cell types, co-express many genes that are not found in the other cell populations analyzed, yet they also express distinct subsets of genes that differentiate them (Heiman et al., 2008). Thus, comparative analysis of BACarray data can be used to characterize CNS cell populations with very unique biochemical and physiological properties, and to distinguish between closely related cell types at the molecular and biochemical level.

As shown in Figure 6B, the top twenty five most specific probesets in each cell type include probesets for both well known cell-specific markers and novel, previously uncharacterized genes. For example, Pcp2, the calcium binding protein Calb1, the scaffolding/synaptic protein Homer3, and the transcription factor Ebf2, all of which are known to be specifically expressed in Purkinje cells (Malgaretti et al., 1997; Shiraishi et al., 2004; Wang et al., 1997), are among the most highly ranked probesets in the Pcp2 BACarray list. Mobp, one of the most abundant components of the CNS myelin sheath (Montague et al., 2006), is prominent in the *Cmtm5* myelinating oligodendrocytes' list. The expression of *Tcrb* in deep layer cortical neurons (Nishiyori et al., 2004) is confirmed in the *Ntsr1* BACarray data. The large number of uncharacterized genes with cell specific translation identified here provide an important resource for discovery of novel biochemical pathways operating in these cell types, or for the identification of new proteins operating in well known pathways. Finally, comparative analysis can reveal discrepancies that are not apparent from anatomical studies. For example, the most specific probesets for the *Etv1* line identify several genes well-known to be expressed in lymphoid cells, suggesting that in this line the EGFP-L10a transgene may also be expressed in circulating cells in the CNS vasculature. For this reason, we are currently characterizing additional BACarray lines for corticostriatal neurons. Taken together, the data shown above demonstrate two important strengths of large-scale comparative analyses of BACarray data. First, molecular relationships between cell types can be easily established with hierarchical clustering (Figure 4); second, groups of genes that encode the biochemical functions of specific cell types can be identified using this sort of systematic comparative approach (Figure 6).

Analysis of BACarray data collected from spinal motor neurons

Due to their involvement in a variety of serious neurological disorders and severe, acute injuries, spinal cord motor neurons (MN) have been extensively studied, and a wealth of anatomical, molecular, and physiological data exists for them. This fact has allowed us to compare the BACarray data presented here with the published literature. As shown in Figure 7, in a single BACarray experiment, we can rediscover most of the MN expressed molecules that have been documented in prior studies. In most cases, where microarray probesets were present and informative, BACarray results agree well with the literature. Thus, it has been reported that MNs express glutamate receptors sensitive to AMPA, kainate, and NMDA (Rekling et al., 2000). Our results suggest that the specific receptor subunits mediating these responses include Gria3 and 4, Grik2 and 4, and Grin1, 3a and 3b. Inhibition in MNs should be due the actions of the Glrα2 and Glrβ glycine receptor subunits and both metabotropic (Gabbr1) and ionotropic GABAergic receptors, potentially composed of Gabrα2, α5, and β3 subunits. Our data predict that MNs should respond to all classic neurotransmitters, including

acetylcholine, via Chrnα4/β2 and/or Chrnα7 receptors, and serotonin, via the Htr1d receptor. In disagreement with prior immunohistochemical findings (Rekling et al., 2000), we do not detect the expression of *Drd1* and or *Drd2* in MNs. Moreover, our transgenic mice for *Drd1* and *Drd2* do not show transgene expression in MNs, nor does the Allen Brain Atlas ISH show expression in brain stem MN, supporting the BACarray results.

MNs also express a variety of newly characterized receptors and orphan receptors. For example, our BACarray data has successfully identified *Grin3b* as a MN specific gene encoding an NMDA subunit. This receptor was recently characterized as creating a unique glycine gated channel in MNs (Nishi et al., 2001). We have also identified several other genes enriched in MNs which potentially encode for MN specific receptors that either have not been previously characterized in MNs or are entirely unstudied. Two that are particularly interesting are the vitamin D receptor (Figure S5) and the orphan receptor P2rxl1 (Figure 7). Future studies investigating the role of these receptors in MN behavior may explain cases of reversible muscle weakness in patients with vitamin D deficiency (Ziambaras and Dagogo-Jack, 1997), or suggest new pathways important to MN function. An important caveat to these conclusions, as highlighted by our ISH studies of Grm2 positive neurons of cerebellum, is that these BACarray results reflect the average expression of all the cholinergic cells of the spinal cord - some of the receptors listed in Figure 7 may be expressed in separate pools of cholinergic cells.

Perspectives

In this study, we have extended the findings of Heiman et al, to establish the generality of the BACarray methodology by demonstrating that it allows robust and reproducible isolation of mRNA across a variety of regions and cell types. The method correctly identifies known cell-specific and enriched transcripts for the twenty four lines reported here, and reports the comprehensive translational profiles for each of them. These profiles identify thousands of novel cell-specific mRNAs that are not detectable by whole tissue microarray analysis. We have provided a primary analysis of many previously uncharacterized neurons and glia and shown that much of the diversity of the nervous system is driven by the suite of proteins expressed on the surface of specific cell types. To illustrate the depth of the available information, the profile of the spinal cord motor neurons was examined in detail with regard to the receptors they produce and the ligands they secrete, as these are genes that determine the responsiveness of a cell to its environment and the behavior of the cell within a circuit. These data demonstrate that a single BACarray experiment can confirm the findings of decades of gene-by-gene expression studies, while at the same time identifying a vast number of novel genes that may be essential for motor neuron function.

Further applications of the BACarray transgenic mouse lines

It is important to note that with these sixteen initial mouse lines, there are additional cell populations from which BACarray data could be collected. For example, the EGFP-L10a transgene is strongly expressed in CA1 neurons in the *Cck* line as well as in the substantia nigra of the *Ntsr1* line. Additional studies of these and other uncharacterized populations could provide important information for a variety of critical CNS cell types. Given the presence of the EGFP-L10a fusion protein in dendrites and axons of some of the cell types presented here, it would also be interesting to combine laser-capture microdissection and BACarray affinity purification to identify translated mRNAs that are localized to specific subcellular compartments.

Beyond the initial characterization we have reported here, there are a variety of biological applications for these BACarray lines that are of significant interest. As established by Heiman et al., the BACarray strategy can be used as a sensitive method to detect changes in single cell populations due to whole animal pharmacological manipulations. Related studies could readily

be conducted to assess the cell-specific translational profile across development, in aging, following injury, or in response to behavioral manipulations and genetic perturbations. For example, BACarray mice can be readily crossed with knock out mice modeling human diseases, particularly those that impact clearly defined cell types, such as Purkinje cells in some cerebellar ataxias, or oligodendrocytes in multiple sclerosis. In other diseases or conditions such as stroke, which can broadly impact neurons, astrocytes, and oligodendrocytes, the responses of each cell type can be parsed individually by assessing selected BACarray lines. This technology will allow us to systematically answer fundamental biological questions regarding the magnitude and particulars of changes in mRNA translation consequent to whole animal manipulations.

Further applications of the BACarray data

In addition to the available lines, the data from these twenty four cell types provide a resource for a variety of studies. The most direct result available from the analysis across these cell types is the identification of novel cell-specific markers. Even the data we present from mixed cell populations can be useful. For example, the *Grp* BACarray data clearly identifies mRNAs expressed in both unipolar brush cells and Bergmann glial, as we would predict from the anatomic characterization of this line. Comparative analysis of these data with that obtained from the *Sept4* BACarray line allows us to subtract Bergmann glial cell mRNAs from this data, and identify additional mRNAs specific to unipolar brush cells in order to uniquely target these cell types for second generation BACarray studies. Furthermore, in cases of relatively weakly expressing lines, such as *Cmtm5*, new drivers can be selected to more effectively target the same cell type. Indeed, preliminary studies with a new mature oligodendrocyte line (*Cnp* JD368) have demonstrated improved RNA yield and data quality from a more strongly expressing transgene (Figure S6). In addition, BACarray data can be used to identify suites of candidate epitopes for cross species comparisons to examine the evolution of the distribution of cell-types in species from mouse through primates.

A number of sophisticated analytical methods have been applied with some success to traditional microarray datasets in order to infer complex biological information from gene expression data. For instance, there are now a variety of approaches to identify transcriptional regulatory networks and novel transcription factor binding sites from microarray data in yeast or cell culture (Blais and Dynlacht, 2005). In higher organisms, however, these analyses are complicated by the cellular complexity of the tissue samples involved, thus confounding the elucidation of “co-expressed” genes. This type of analysis should therefore benefit from the collection of numerous cell-type specific BACarray datasets. For the same reason, the use of the BACarray methodology should enhance the search for functional ‘modules’ of genes in the CNS (Oldham et al., 2006). The combination of these analytical methods with our cell-type specific BACarray datasets should significantly enhance our understanding of global gene expression patterns within the CNS. We believe, due to the enhanced information content, BACarray and other cell-specific technologies should become the standard for microarrays in neuroscience. To this end, we have generated this BACarray resource.

The BACarray Resource

The BACarray Resource we provide here includes: sixteen BACarray transgenic lines expressing the EGFP-L10a fusion protein in well characterized cell types; an anatomic database showing serial coronal sections of EGFP-L10a fusion protein expression in the adult mouse brain; IPvUB datasets for twenty four cell types listing all mRNAs enriched in that cell type; detailed protocols for conducting BACarray translational profiling experiments from the brain; and anti-EGFP monoclonal antibodies. This resource provides the materials, data and knowledge to enable a wide variety of studies not previously available to the neuroscience community.

Experimental Procedures

BAC Modification, Transgenesis, and Animal Husbandry

All protocols involving animals were approved by the Rockefeller University Institute Animal Care and Use Committee. BACs from Table S1 were modified as described to insert an EGFP-L10a fusion protein into the translation start site of the driver gene (Gong et al., 2002;Gong et al., 2003). Founders and subsequent generations were bred to either Swiss-Webster or c57bl/6 wildtype mice. Lines were maintained as trans-heterozygotes.

Immunoprecipitation of Polyribosomes

All immunoprecipiations except for the *Drd1* and *Drd2* lines, which used the goat anti-EGFP described in the accompanying paper (Heiman et al., 2008), were done using a mix of two monoclonal antibodies (19C8, 19F7). Three to six mice for each replicated sample were euthanized with CO_2 and distinct brain regions were dissected. Each cell population was assayed in triplicate. RNA quantity and quality were determined with a Nanodrop 1000 spectrophotometer (Wilmington, DE) and Agilent 2100 Bioanalyzer (Foster City, CA). For each sample, 15ng of total RNA were amplified with the Affymetrix two cycle amplification kit and hybridized to Affymetrix 430 2.0 microarrays following the manufacturer's instructions.

Histological Methods

Brains were processed identically using MultiBrain™ Technology (NSA, NeuroScience Associates, Knoxville, TN) for DAB IHC using a 1:75,000 dilution of Goat anti-EGFP serum (Heiman et al., 2008) following the Vectastain elite protocol (Vector Labs, Burlingame, Ca). Serial sections were digitized with Zeiss Axioskop2 microscope at 10x magnification.

For IF, sections were blocked with 5% normal donkey serum, 0.25% triton and incubated with primary antibodies (Table S7), and appropriate Alexa dye conjugated secondary antibodies (Molecular Probes/Invitrogen, Carlsbad, CA).

Probes for ISH were prepared from EST clones (Open Biosystems, Table S3) using DIG RNA labeling kit (Roche, Basel, CH) and purified with ProbeQuant G-50 microcolumns (GE Healthcare). 20 micron fixed brain sections were treated with 0.05% Triton X100 and 50 ug/ml Proteinase K, and acetylated (0.1M TEA, 0.25% acetic anhydride). Sections were prehybridized (500 ug/ml salmon sperm DNA, 2.5X Denhardt's, 5X SSC), hybridized overnight with riboprobe, rinsed (5X SSC, 65°C), and washed for 1 hour (0.2X SSC, 68°C). Development with HNPP/Fast Red followed manufacturer's protocols (Roche), with the addition of Goat Anti-EGFP and alexa-488 donkey anti-goat (Invitrogen) for IF.

Images were acquired as Z-stacks (2 micron sections) with a Zeiss Inverted LSM 510 confocal microscope.

Microarray Normalization and Analysis

Briefly, samples were normalized with GCRMA and filtered to remove probesets with low signal and those identified as background (Table S2). Each IP was compared to unbound samples from the same tissue to calculate a ratio of IP/UB as a measure of 'enrichment'. Analyzed data for each cell type are available in Table S5, which contains the IP/UB values for all genes with fold change >2 and $p < .05$ by Welch's t-Test, with Benjamini and Hochberg FDR multiple testing correction. Three cell types were further corrected to remove signal from minor cell types described in Figure 2A (Supplemental Methods). Hierarchical clustering, and Pearson's correlation with MBP were performed in Genespring 7.0 (Agilent technologies). Shannon entropy was calculated in Excel from GCRMA normalized values using published

formulas (Schneider, 2007). Gene Ontologies was performed using the BiNGO plugin for the cytoscape software (Maere et al., 2005). Comparative anlaysis of all cell types, and heatmaps (Figure 6), were generated with the R stastistical software. A full description of analytical methods is in supplemental materials. MIAME compliant raw data are available from GEO.

Quantitative Real-time RT-PCR

cDNA was synthesized from 20 ng of total RNA from the three replicate IP and UB samples using M-MulV reverse transcriptase (New England Biolabs, Ipswich, Ma), using oligo $dT_{23}VN$ as a primer, then purified with the Qiagen Quick PCR cleanup, following manufacturer's instructions (Qiagen, Valencia, Ca).

PCR was performed using Biorad iQ syber green supermix following manufacturer's protocols (Biorad, Hercules, Ca), with 500 nm final concentration of each primer (Table S4). Cycling and quantitation were performed using Biorad iQ5 multiplex real-time detection hardware. PCR was carried out for 45 cycles (94°, 30 seconds, 63°, 30 seconds, 72°, 30 seconds), followed by a melt curve. Each replicate was assayed in triplicate. Conditions yielding significant dimers, as demonstrated by melt curve and/or gel electrophoresis were excluded from further analysis. Primers that did not yield product in at least 2 of 3 replicates prior to 35 cycles were excluded from further analysis. Data were normalized to *ActB* with the ddCT method, via iQ5's optical system software version 2, and averaged across replicates. All qPCR products were subcloned and sequenced to confirm accuracy of PCR. Microarray data were also normalized to *ActB* for comparison purposes (Figure 3).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Cuidong Wang, Huifen Feng, Christine Grevstad, Clint Earnheart, Wenxiang Zhang, Miho Nakajima, George Skabardonis, Ayse Tekinay, and members of the P. Greengard and N. Heintz laboratories for their advice and assistance. We also thank the Rockefeller University Bio-Imaging Resource Center and Genomics Resource Center, the Memorial Sloan-Kettering Cancer Center Monoclonal Antibody Core Facility, and Neuroscience Associates. This work was supported by the Howard Hughes Medical Institute, the Adelson Medical Research Foundation, The F. M. Kirby Foundation, The Picower Foundation, The Jerry and Emily Spiegel Family Foundation, The Peter Jay Sharp Foundation, The Michael Stern Foundation, the Simons Foundation, NIA grant AG09464, NIMH Conte Center grant MH074866, NIDA grant 5F32DA021487 to M.H., and NIDA grant DA10044 to P.G. N.H. is an HHMI Investigator.

References

Bargmann CI. Genetic and cellular analysis of behavior in C. elegans. Annu Rev Neurosci 1993;16:47–71. [PubMed: 8460900]

Berthele A, Boxall SJ, Urban A, Anneser JM, Zieglgansberger W, Urban L, Tolle TR. Distribution and developmental changes in metabotropic glutamate receptor messenger RNA expression in the rat lumbar spinal cord. Brain Res Dev Brain Res 1999;112:39–53.

Blais A, Dynlacht BD. Constructing transcriptional regulatory networks. Genes Dev 2005;19:1499–1511. [PubMed: 15998805]

Fuhrman S, Cunningham MJ, Wen X, Zweiger G, Seilhamer JJ, Somogyi R. The application of shannon entropy in the identification of putative drug targets. Biosystems 2000;55:5–14. [PubMed: 10745103]

Gong S, Yang XW, Li C, Heintz N. Highly efficient modification of bacterial artificial chromosomes (BACs) using novel shuttle vectors containing the R6Kgamma origin of replication. Genome Res 2002;12:1992–1998. [PubMed: 12466304]

Gong S, Zheng C, Doughty ML, Losos K, Didkovsky N, Schambra UB, Nowak NJ, Joyner A, Leblanc G, Hatten ME, Heintz N. A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. Nature 2003;425:917–925. [PubMed: 14586460]

Greengard P. The neurobiology of slow synaptic transmission. Science 2001;294:1024–1030. [PubMed: 11691979]

Heiman M, Schaefer A, Gong S, Peterson J, Day M, Ramsey K, Suárez-Fariñas M, Schwarz C, Stephan DA, Surmeier J, et al. Development of a BACarray translational profiling approach for the molecular characterization of individual CNS cell types. Cell. 2008

Heintz N. Gene expression nervous system atlas (GENSAT). Nat Neurosci 2004;7:483. [PubMed: 15114362]

Kaelin-Lang A, Lauterburg T, Burgunder JM. Expression of adenosine A2a receptors gene in the olfactory bulb and spinal cord of rat and mouse. Neurosci Lett 1999;261:189–191. [PubMed: 10081981]

Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ, et al. Genome-wide atlas of gene expression in the adult mouse brain. Nature 2007;445:168–176. [PubMed: 17151600]

Lorente de No, R. The cerebral cortex: Architecture, intracortical connections and motor projections. In: Fulton, JF., editor. Physiology of the Nervous System. London: Oxford University Press; 1934. p. 291-339.

Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 2005;21:3448–3449. [PubMed: 15972284]

Magdaleno S, Jensen P, Brumwell CL, Seal A, Lehman K, Asbury A, Cheung T, Cornelius T, Batten DM, Eden C, et al. BGEM: an in situ hybridization database of gene expression in the embryonic and adult mouse nervous system. PLoS Biol 2006;4:e86. [PubMed: 16602821]

Malgaretti N, Pozzoli O, Bosetti A, Corradi A, Ciarmatori S, Panigada M, Bianchi ME, Martinez S, Consalez GG. Mmot1, a new helix-loop-helix transcription factor gene displaying a sharp expression boundary in the embryonic mouse brain. J Biol Chem 1997;272:17632–17639. [PubMed: 9211912]

Malosio ML, Marqueze-Pouey B, Kuhse J, Betz H. Widespread expression of glycine receptor subunit mRNAs in the adult and developing rat brain. Embo J 1991;10:2401–2409. [PubMed: 1651228]

Markram H, Toledo-Rodriguez M, Wang Y, Gupta A, Silberberg G, Wu C. Interneurons of the neocortical inhibitory system. Nat Rev Neurosci 2004;5:793–807. [PubMed: 15378039]

Montague P, McCallion AS, Davies RW, Griffiths IR. Myelin-associated oligodendrocytic basic protein: a family of abundant CNS myelin proteins in search of a function. Dev Neurosci 2006;28:479–487. [PubMed: 17028425]

Nishi M, Hinds H, Lu HP, Kawata M, Hayashi Y. Motoneuron-specific expression of NR3B, a novel NMDA-type glutamate receptor subunit that works in a dominant-negative manner. J Neurosci 2001;21:RC185. [PubMed: 11717388]

Nishiyori A, Hanno Y, Saito M, Yoshihara Y. Aberrant transcription of unrearranged T-cell receptor beta gene in mouse brain. J Comp Neurol 2004;469:214–226. [PubMed: 14694535]

Nunzi MG, Shigemoto R, Mugnaini E. Differential expression of calretinin and metabotropic glutamate receptor mGluR1alpha defines subsets of unipolar brush cells in mouse cerebellum. J Comp Neurol 2002;451:189–199. [PubMed: 12209836]

Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. Proc Natl Acad Sci U S A 2006;103:17973–17978. [PubMed: 17101986]

Ramon y Cajal, S.; Pasik, P.; Pasik, T. Texture of the nervous system of man and the vertebrates. Wien; New York: Springer; 1899.

Rekling JC, Funk GD, Bayliss DA, Dong XW, Feldman JL. Synaptic control of motoneuronal excitability. Physiol Rev 2000;80:767–852. [PubMed: 10747207]

Schneider, TD. Information Theory Primer. Frederick, MD: National Cancer Institute; 2007.

Shannon, CE.; Weaver, W. The mathematical theory of communication. Urbana: University of Illinois Press; 1969.

Shiraishi Y, Mizutani A, Yuasa S, Mikoshiba K, Furuichi T. Differential expression of Homer family proteins in the developing mouse brain. J Comp Neurol 2004;473:582–599. [PubMed: 15116392]

Svenningsson P, Nishi A, Fisone G, Girault JA, Nairn AC, Greengard P. DARPP-32: an integrator of neurotransmission. Annu Rev Pharmacol Toxicol 2004;44:269–296. [PubMed: 14744247]

Towers S, Princivalle A, Billinton A, Edmunds M, Bettler B, Urban L, Castro-Lopes J, Bowery NG. GABAB receptor protein and mRNA distribution in rat spinal cord and dorsal root ganglia. Eur J Neurosci 2000;12:3201–3210. [PubMed: 10998104]

Wang SS, Tsai RY, Reed RR. The characterization of the Olf-1/EBF-like HLH transcription factor family: implications in olfactory gene regulation and neuronal development. J Neurosci 1997;17:4149–4158. [PubMed: 9151732]

Yang XW, Model P, Heintz N. Homologous recombination based modification in Escherichia coli and germline transmission in transgenic mice of a bacterial artificial chromosome. Nat Biotechnol 1997;15:859–865. [PubMed: 9306400]

Ziambaras K, Dagogo-Jack S. Reversible muscle weakness in patients with vitamin D deficiency. West J Med 1997;167:435–439. [PubMed: 9426489]

Zipursky SL, Rubin GM. Determination of neuronal cell fate: lessons from the R7 neuron of Drosophila. Annu Rev Neurosci 1994;17:373–397. [PubMed: 8210180]

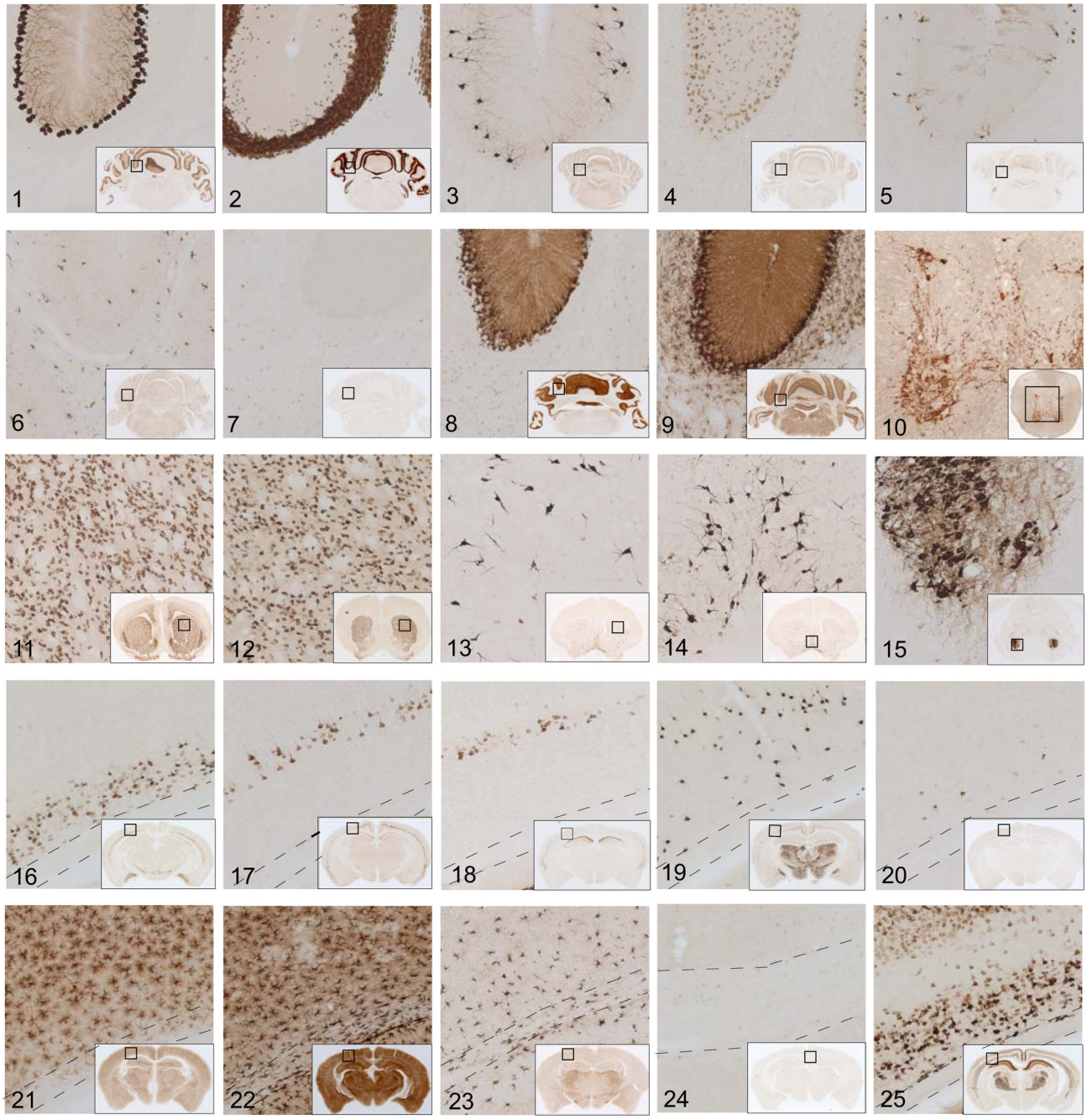


Figure 1. BACarray lines express the EGFP-L10a transgene in specific CNS cell populations

DAB immunohistochemistry with anti-EGFP antibody on each mouse line reveals a unique and specific pattern of expression for the EGFP-L10a transgene. Panels (10x) show the morphology and localization of cell types expressing the transgene, while inset shows location of panel, for cerebellar (1–9), spinal cord (10), striatal/basal forebrain (11–14) brainstem (15), and cortical (16–25) cell types. *A key for all cell types is in Figure 2A. Dashed lines (panels 16–25) indicate corpus callosum.*

A

Summary of Cell Types Studied

Panel	Region	BAC Driver	Line	Primary Cell Type Labeled	Confirmation of Cell Types	Minor Cell Type Labeled
1	Cerebellum	Pcp2	DR166	Purkinje Cells	Morphology, IF: PVALB+ and CALB1+	None
	Cerebellum	Pcp2	DR168	Purkinje Cells	Morphology, IF: PVALB+ and CALB1+	None
2	Cerebellum	NeuroD1	JP241	Granule Cells, Deep Cerebellar Nuclei	Morphology, IF: NEUN+	None
3	Cerebellum	Grm2	JP77	Granule Cell Layer Interneurons (Inner Golgi Cells)	Morphology, IF: GRM2/3+	None
4	Cerebellum	Lypd6	JP48	Stellate and Basket Cells	Morphology, IF: PVALB+, NEUN-, CALB1-	Noe
5	Cerebellum	Grp	JP25	Unipolar Brush Cells (mGluR1 subtype)	Morphology, IF: CALB2-, GRM1+ or S100+	Bergman Glia
6	Cerebellum	Olig2	JD97	Mature Oligodendrocytes and Progenitors	Morphology, IF: OLIG2+, CSPG4+ or CNP+, GFAP-	None
7	Cerebellum	Cmtm5	JD307	Mature Oligodendrocytes	Morphology, IF: CNP+, CSPG4-	None
8	Cerebellum	Sept4	DS152	Bergman Glia	Morphology, IF: S100+, ALDH1L1+, GFAP+	Mature Oligodendrocytes
9	Cerebellum	Aldh1L1	JD130	Astroglia (includes Bergman Glia)	Morphology, IF: ALDH1L1+, GFAP+, CSPG4-, CNP-	None
10	Spinal Cord	Chat	DW167	Motor Neurons, Cholinergic Interneurons	Morphology, IF: CHAT+, SLC18A3+	None
11	Striatum	Drd1	CP73	Drd1 Positive Medium Spiny Neurons	Morphology, IF: PENK-	None
12	Striatum	Drd2	CP101	Drd2 Positive Medium Spiny Neurons	Morphology, IF :PENK+	Cholinergic interneurons
13	Corpus Striatum	Chat	DW167	Cholinergic Neurons	Morphology, IF: CHAT+, SLC18A3+	None
14	Basal Forebrain	Chat	DW167	Cholinergic Projection Neurons	Morphology, IF: CHAT+, SLC18A3+	None
15	Brain Stem	Chat	DW167	Motor Neurons, Midbrain Cholinergic Neurons	Morphology, IF: CHAT+, SLC18A3+	None
16	Cortex	Ntsr1	TS16	Layer 6 Corticothalamic Pyramidal Neurons	Morphology, Morphometric comparison to eGFP lines	None
17	Cortex	Glt25d2	DU9	Layer 5b Corticospinal, Corticopontine Pyramidal Neurons, and Small Pyramidal Cells	Morphology, Morphometric comparison to eGFP lines	None
18	Cortex	Etv1	TS88	Layer 5a Corticostriatal Pyramidal Neurons	Morphology, Morphometric comparison to eGFP lines	Few GFAP+ Astrocytes
19	Cortex	Pnoc	GM64	Neurons	Morph, IF: some GABA+, some CALB2+, CALB1-	None
20	Cortex	Cort	GM130	Interneurons	Morph, IF: some CALB1+, few PVALB+, CALB2-	None
21	Cortex	Aldh1L1	JD133	Astroglia (reactive and non-reactive)	Morph, IF: ALDH1L1+,GLUL+, GFAP+, CSPG4-, CNP-	None
22	Cortex	Aldh1L1	JD130	Astroglia (reactive and non-reactive)	Morph, IF: ALDH1L1+,GLUL+, GFAP+, CSPG4-, CNP-	None
23	Cortex	Olig2	JD97	Mature Oligodendrocytes and Progenitors	Morphology, IF: OLIG2+, CSPG4+ or CNP+, GFAP-	None
24	Cortex	Cmtm5	JD307	Mature Oligodendroctes	Morphology, IF: CNP+, CSPG4-	None
25	Cortex	Cck	GM391	Mixed Neurons	Morphology, IF: some CALB1+, all PVALB-, CALB2-	None

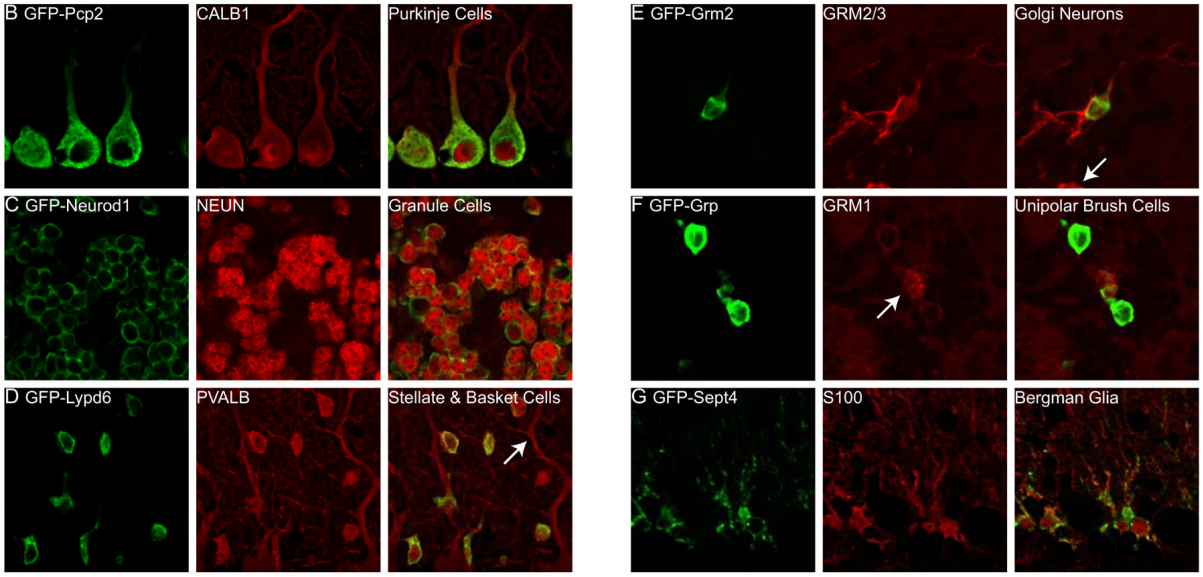


Figure 2. Summary of cell types studied and in-depth characterization of lines

A) All primary cell types expressing EGFP-L10a are listed, as well as the methods used to confirm correct expression. Minor cell types expressing relatively low levels of the EGFP-L10a transgene in the same structure are also listed. Panel # corresponds to Figure 1. B-G) IF on six mouse lines confirms transgene expression in distinct cell types in the cerebellum. First panels show IF for the EGFP-L10a fusion protein (green) in *PCP2* (B), *NeuroD1* (C), *Lypd6* (D), *Grm2* (E), *Grp* (F), and *Sept4* (G) BACarray lines. Second panels (red) show co-staining with appropriate cell-type specific markers: Calbindin positive Purkinje cells (B), NeuN positive granule cells (C), Parvalbumin positive outer stellate and deep stellate (basket) neurons of the molecular layer (D), Grm2/3 positive interneurons (Golgi cells) (E), unipolar brush cells

with Grm1 positive brush (arrow) (F), and S100 positive Bergman glia (G). The third panels show merged images combining EGFP-L10a and cell-type markers. Note that EGFP-L10a is not detected in the parvalbumin positive Purkinje cells of the *Lypd6* line (D, arrow), nor in the glomeruli of the *Grm2* line (E, arrow).

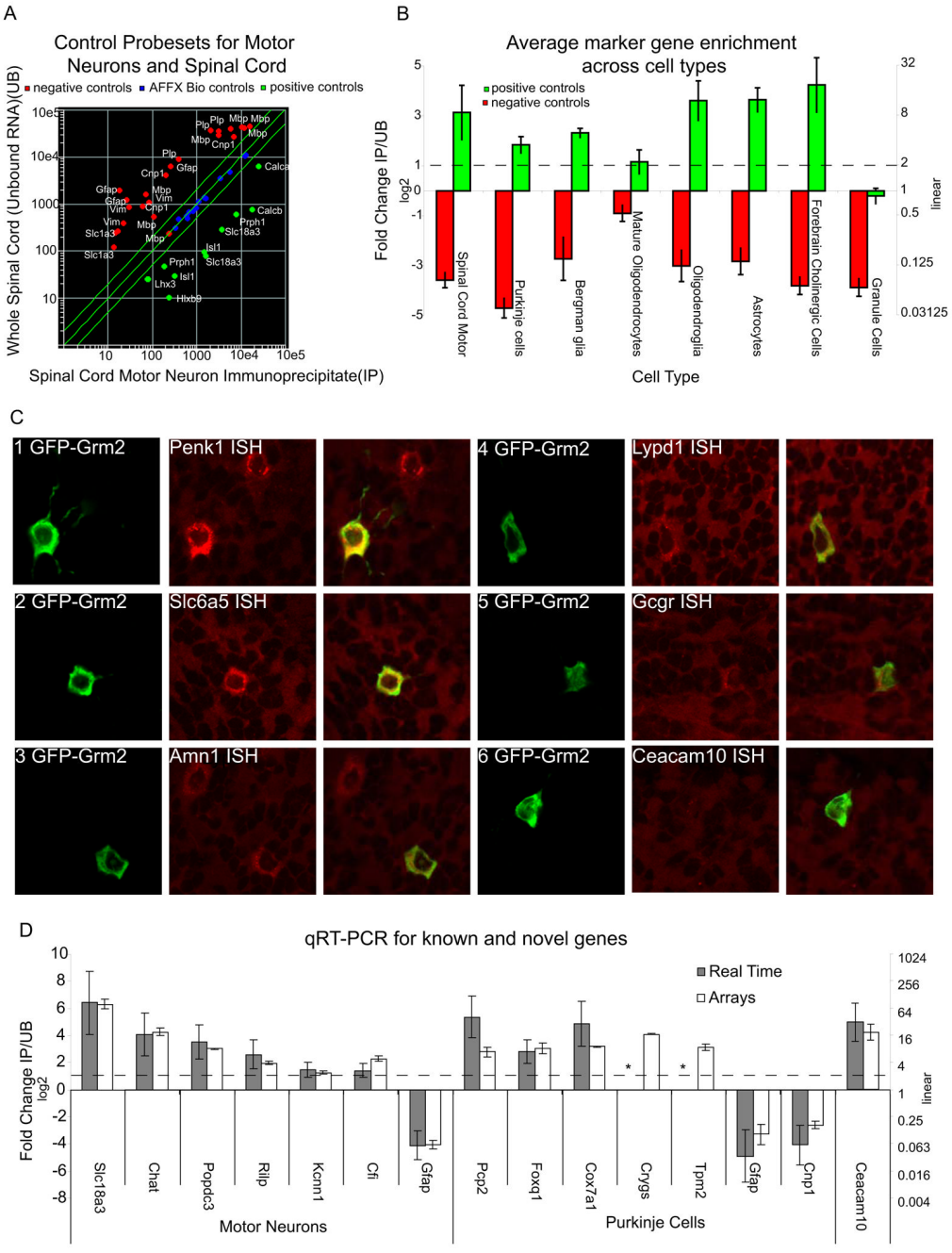


Figure 3. BACarray identifies both known and new markers

A) Known markers of spinal cord motor neurons (green dots) are highly enriched in the BACarray RNA (IP) (X-axis) while non-motor neuron genes (glial genes, red dots) are enriched in the whole tissue (UB) RNA. B) Averages of known cell-specific markers (green bars) are consistently enriched in the IP RNA, while negative controls (red bars) are not. Exceptions are the mature oligodendrocytes (*Cmtm5*) with low transgene expression, and granule cells (*Neurod1*) which contribute the majority of the cerebellar UB RNA thus precluding enrichment. C) ISH (red) and IF (green) images for genes predicted to be expressed in cerebellar Golgi cells show five of six genes with clear double labeling. D) qRT-PCR confirms that the sixth gene, *Ceacam10*, is expressed in the cerebellum and enriched in Golgi cells. qRT-PCR

also confirms BACarray data for genes in Motor Neurons and Purkinje Cells. *Slc18a3*, *Chat*, *Gfap*, *Pcp2*, and *Cnp* are positive and negative controls for these populations. * *Crygs* and *Tpm2* failed to amplify by RT-PCR for either IP (*Tpm2*) or UB (*Crygs*), and thus no ratio could be calculated. *All plots Mean +/− SEM.*

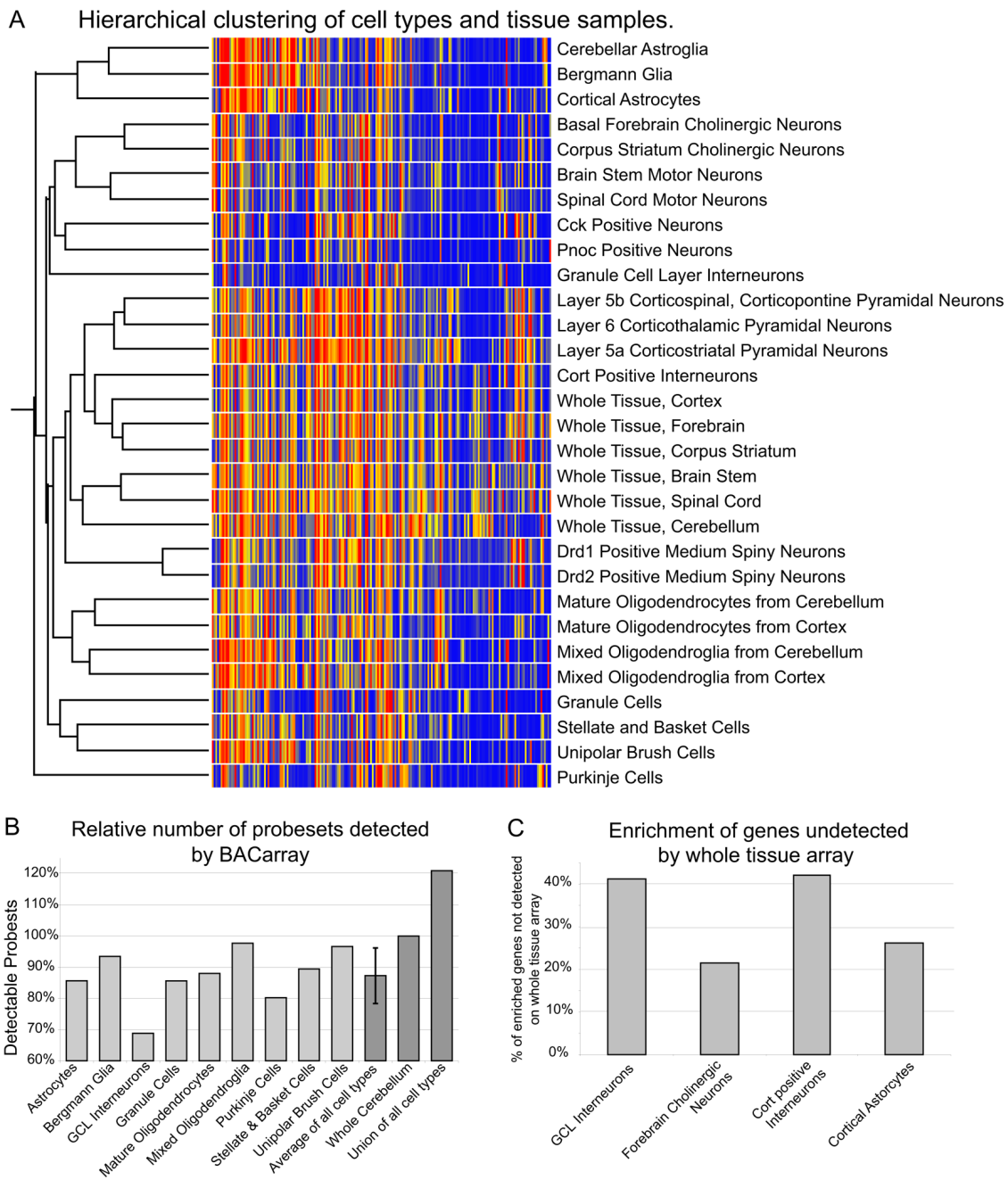


Figure 4. BACarray clusters cells by type, and provides greater sensitivity than whole tissue arrays

A) Hierarchical clustering on high coefficient of variation genes from all samples describes the relationships between cell types. B) Counting detectable (signal >50) probesets in cerebellar samples reveals that while fewer probesets will be detected in any given cell type than are detectable in whole tissue, across all cell types in total, more probesets have measurable signal. *Data normalized to number of probesets in whole cerebellum.* C) For four representative cell types, up to 42% of cell-specific or enriched probesets (IP/UB >2) are undetectable on whole tissue microarrays.

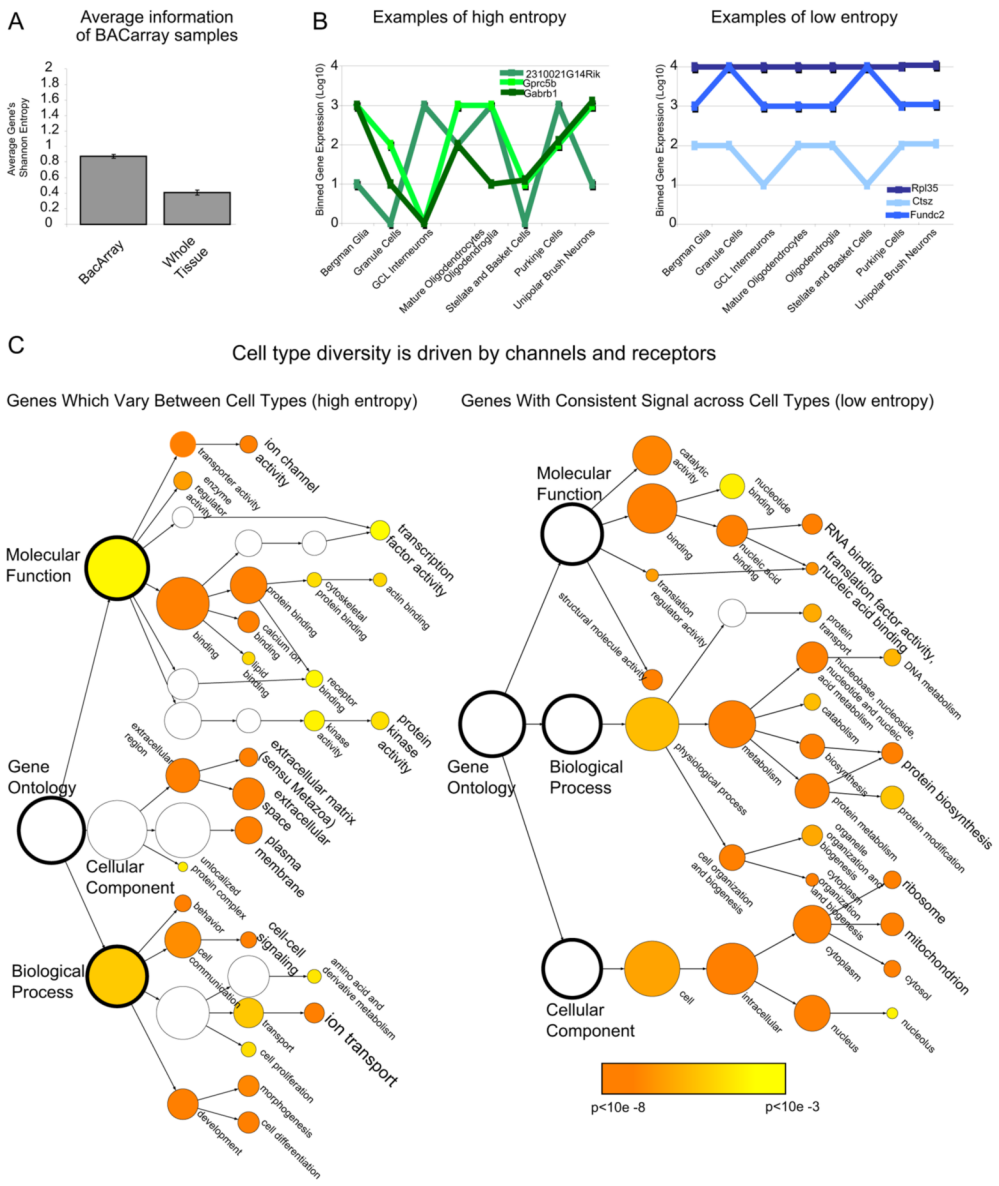
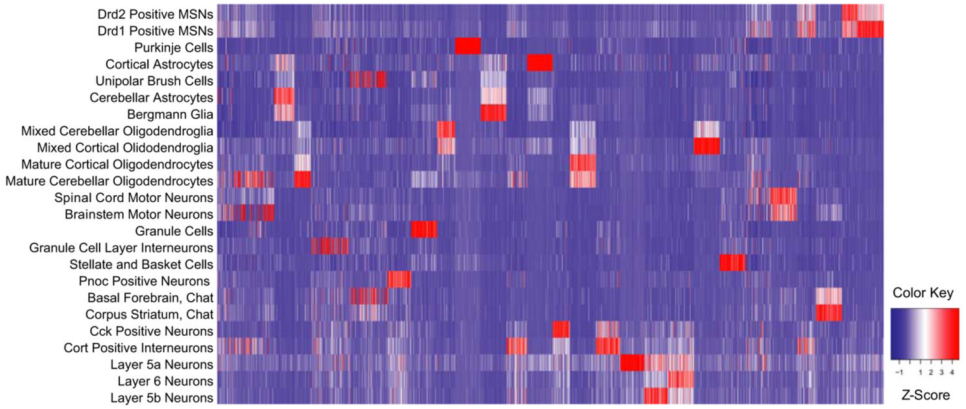


Figure 5. Cell type diversity is driven by cell surface proteins

A) Shannon entropy analysis reveals that BACarray samples have twice as much information as whole tissue *(Mean +/− SEM)*. B) Binned expression of probesets for three genes with high Shannon entropy (average entropy 1.68) or low entropy (average entropy .31) shown for eight representative cell types. C) Gene Ontologies analysis identifies significantly over-represented (p<.001) gene classifications for the 10% of probesets with the highest (left panel) or the 10% with lowest (right panel), information content. *Color bar: significance level for categories by Hypergeometic test with Benjamini Hochberg FDR correction.*

A



B

Str_drd2	Str_drd1	Cb_pcp2	Ctx_aldh1l1	Cb_grp	Cb_aldh1l1	Cb_sept4	Cb_olig2	Ctx_olig2	Ctx_cmtm5	Cb_cmtm5	SC_chat
1457132_at	1457132_at	1419880_x_at	1442917_at	Accn5	1110017I16Rik	8030451F13Rik	1439734_at	5730559C18Rik	1435119_at	1441721_at	0610040B09Rik
Adora2a	4930448N21Rik	1440944_at	1457215_at	Adcyap1	1447307_at	A2m	2310031A18Rik	Adm	1445055_at	1459662_at	Ahnak2
Ahnak	Adra2c	1455970_at	2010002N04Rik	AF529169	1459747_at	Alx3	5730559C18Rik	Brca1	2010001M09Rik	1700063D05Rik	Amhr2
Arg2	Ahnak	1700023E05Rik	2610034M16Rik	Ank1	1500015O10Rik	Aps	9630013A20Rik	Calcrl	4933400E14Rik	2210020M01Rik	Amhr2
Bcl11b	Atp6v1c2	2410124H12Rik	Adora2b	Arg1	1810010K12Rik	BC055107	9630013A20Rik	Ccna2	Abca8a	2700046A07Rik	Arhgap9
Cd4	Bscl2	A730030A06	Adora2b	Cdc25a	Bbox1	C2	9630013A20Rik	Ccnb1	Adamtsl1	2810003C17Rik	Arhgap9
D430042O09Rik	C230078M08Rik	A730090H04Rik	Adora2b	Cxcl11	BC029169	Cables1	A030001D16Rik	Chek1	Apod	4833411O04Rik	Calca
Drd2	Dmkn	Atp2a3	Ai449310	Doc2g	Calml4	Casp12	A930009A15Rik	Cspg4	Bfsp2	A230001M10Rik	Calcb
Foxp1	Drd1a	Calb1	AU067665	Eomes	Cd38	Cox7b2	Adam12	Dmrta1	C030030A07Rik	Apod	Cd24a
Fst	Drd1a	Car7	AU067665	Eomes	Chd9	Gdf10	Ascl1	E2f8	Ctdspl	Gjb1	Cd24a
Fst	Erc2	Cep76	Cyp4f15	Grp	Cpsf4l	Gli	Bmp4	Ect2	Efnb3	Gjb1	Chodl
Gng7	Eya1	Ebf2	Dmrta2	Lek1	Cxxc5	Pax3	C1ql1	EG232599	Elovl1	Gje1	Dnahc2
Gpr6	Frcl1	Eps8l2	Fzd2	Mab21l1	Cyp4f14	Plekho2	Eya2	Fignl1	Galnt6	Gng8	Enpep
Gpr88	Kcnip2	Fgf7	Idb3	Nmb	Dao1	Slc14a1	Limd1	Gpr17	Il33	Hapln2	Gna14
Gprin3	Myo5b	Fgf7	Nr2e1	Nnat	Dmp1	Tbc1d20	Limd1	Lad1	Mass1	Insc	Gna14
Indo	Pde10a	Gchfr	Oaf	Ntf3	Dmp1	Tmem30b	Lims2	Pbk	Mobp	Mobp	Gna14
Iqcb1	Pdyn	Homer3	Ppp1r3g	Otop2	Dnahc8	Tmprss3	Matn4	Pdgfra	Ndrg1	Ninj2	Hlxb9
LOC665252	Peg3	Nrk	Rpe65	Otx2	LOC433016	Ucp2	Meox1	Pnlip	Ninj2	Nkx2-9	Hoxc5
Mn1	Prkch	Nrk	Scara3	Pitpnc1	Mdk	Ucp2	Mmp2	Prc1	Nmral1	Prss18	Hoxc8
Pde10a	Prkch	Pcp2	Slc15a2	Serping1	Msx2	Ucp2	Mmp2	Prkcq	Pcolce2	Serpinb1a	Hspb1
Rasgrp2	Pum1	Slc5a1	Slc15a2	Sgcg	Ramp1	Vim	Pcdh15	Rassf10	Plekhh1	Slc34a3	Hspb1
Rxrg	Rab40b	Smpx	Slc1a2	Sln	Slco4a1	Vim	Ppib	Rlbp1	Rasgrf1	Slc45a3	Lcn2
Sec16a	Spata13	Sos1	Spp1	Sln	Syn2	Vim	Ppil5	Rnf122	Serpinb1a	Slc45a3	Mrvi1
Sec16a	Wnt2	Sycp1	T2bp	Tdo2	Tnfsf7	Wif1	Susd5	Scel	Tmod2	Sypl	Prph1
Ttc12	Zfp261	Tspan11	Upp1	Tmem182	Ttpa	Wnt6	Traf4	Top2a	Tnni1	Tmem125	Rtbdn

BrSt_chat	Cb_neurod1	Cb_grm2	Cb_lypd6	Ctx_pnoc	BFbr_chat	CorpStr_chat	Ctx_cck	Ctx_cort	Ctx_etv1	Ctx_ntsr1	Ctx_glt25d2
1100001E04Rik	1700021K02Rik	1110002E22Rik	1446785_at	Acaa1b	2900001G08Rik	2010007E15Rik	1700101I19Rik	1442650_at	C1qa	1441898_at	1459626_at
Calcb	1700021K02Rik	1700049E17Rik	1458324_x_at	Adarb2	AI606473	6430704N06	9130024F11Rik	1447819_x_at	C1qb	1458271_at	Abcb1a
Cckar	1700021K02Rik	1810033B17Rik	1459359_at	Clnk	Ccdc109b	Ache	A430108G06Rik	2900040J22Rik	C1qb	8430419L09Rik	Acsl4
Cda	2310015B20Rik	4930511J11Rik	6820408C15Rik	Cnr1	Cdkn3	Agtr1	Blnk	4930529M08Rik	C1qg	Arhgap25	Cacna2d1
Gal	5430421F17Rik	Acrbp	9030425E11Rik	Crh	Cox6a2	AI606473	Boll	4933401H06Rik	Cbr2	Cbln4	Cckbr
Glipr1	AI118078	Acyp1	Adamts15	Crispld2	Cox6a2	BB086117	Ccl21b	A230058F20Rik	Cd14	Epha4	Cldn5
Hoxc4	Ankrd25	Cd8a	Akr1b7	Dlx1	Crabp1	Bves	Cct4	Akr1c18	Clecsf10	Epha4	Clstn1
Hspb1	Cabp2	Cited1	Arg1	Dlx1	Dapl1	Cd44	Figf	Atf4	Csf1r	Gpc4	Flt1
Hspb1	Calb2	Fgf5	Asgr1	Htr3a	Edem3	Cpne4	Figf	Casp14	Cxcl4	Il1rap	Glt25d2
Hspb8	Car4	Gcgr	Cdh1	Igf1	Edem3	Crabp2	Fmo2	Cct4	Egr4	Lamc2	Gpr103
Kcnj14	Cbln3	Gdf3	Cxcl12	Igf1	Gbx2	Ecel1	Golt1a	Cort	F13a	Met	Lamc2
Layn	Cbln3	H2-T3	Grid2	Lman1l	Gpr14	Gdnf	Herc5	Crhbp	Fcgr2b	Myh3	Ly6a
Mbl2	Comp	Igf2bp1	Hrc	Myl4	Gpx2	Gm691	Igfbp6	Ddx5	Fcgr2b	Nov	Ly6c
Nppb	Crtam	Klra3	Kit	Necab1	Grem1	Mxra7	Krt1-12	Mid1	Fcgr2b	Nov	Mef2c
Nts	D16Ertd472e	Lbxcor1	Lbx1h	Npy	Hcrtr1	Ngfr	Krt1-12	Moxd1	Gmfg	Ntsr	Mef2c
Phox2b	Gprc5c	Pip5k1b	Lrrc38	Pnoc	Isl1	Ntrk1	Neurod6	Ntng2	Higd1b	Pcdh21	Pcdh19
Pip5kl1	Il16	Serpinb1b	Lypd6	Prox1	Lhx8	Ooep	Ppp1r11	R3hdm	Lyzs	Pde4dip	Pfkp
Ptgis	Il16	St14	Lypd6	Pthlh	LOC545261	Plcz1	Rtn4rl2	Rbp4	Lyzs	Pthr2	Slc39a10
Rgs11	Il17re	Stac3	Plch1	Pthlh	Ngb	Popdc3	Slc22a16	Scnn1a	Lzp-s	Scube1	Slc7a4
Shox2	Negr1	Tcfap2b	Prkcd	Rab3c	Ptger2	Slc7a14	Slc5a9	Shmt2	Mrc1	Tcrb-V13	Slc7a4
Slc18a3	Neurod1	Tpx2	Scnn1g	Sla2	Scn9a	Spink12	Stap2	Slc5a9	Ms4a7	Tcrb-V13	Slco1a4
Slc5a5	Np220	Tssc1	Sla	Syt1	Tacr3	Tmsb10	Tacstd1	Ss18l1	Msr2	Tcrb-V13	Smoc2
Tbx20	Rad51l1	Ush3a	Sla	Tac2	Tacr3	Trap1a	Tcrb-V13	Sst	Vtn	Tcrb-V13	Sox17
Ucn	Rps6ka1	Vwc2	Sla	Tnnt1	Th	Ufsp1	Tcrb-V13	Tacstd1	Vtn	Tcrb-V13	Sox18
Ush1c	Uncx4.1	Yif1	Tcfap2b	Vip	Zar1	Zar1	Tncc	Tnrc6	Xlkd1	Vps11	Wdr35

Figure 6. Comparative BACarray analysis reveals cell type specific translational profiles

A) Heatmap showing the normalized expression of the 100 top ranked probesets from each sample, across all samples. Note blocks of genes detected as specific to each cell type (such as *Pcp2*). Related cell types are evidenced by co-expression of some of these genes (such as Bergman glia, and cerebellar astrocytes). B) Lists of the top 25 probesets of the 100 for each cell population from A include many known cell specific genes (for example *Pcp2* and *Calb1* in Purkinje cells), as well as a variety of novel genes and probesets (such as *2410124H12Rik*). *Columns are headed with the tissue source (Striatum: Str; Cerebellum: Cb; Cortex: Ctx; Spinal Cord: SC; Brain Stem: BrSt; Basal Forebrain: BF; and Corpus Striatum:*

CorpStr) as well as the appropriate BACarray driver. Column order corresponds to cell type order in A.

Not Expressed in M.N.

Expressed in M.N.

Enriched in M.N. (IP/UB >2)

Gene	ip/ub	rf	Title
Glutamate/Aspartate Receptors			
Gria3	1.03	1	ionotropic, AMPA3
Gria4	0.57	1	ionotropic, AMPA4
Grid1	0.73		ionotropic, delta 1
Grik1		1	ionotropic, kainate 1
Grik2	1.35		ionotropic, kainate 2
Grik4	1.31	1	ionotropic, kainate 4
Grik5			ionotropic, kainate 5
Grin1	0.45	1	ionotropic, NMDA1
Grin3a	0.48		ionotropic, NMDA3A
Grin3b	5.66	2	ionotropic, NMDA3B
Grm4	1.85	3	metabotropic 4
GABA Receptors			
Gabbr1	0.92	4	GABA-B receptor 1
Gabra2	1.2	1	GABA-A subunit alpha 2
Gabra5	0.6	1	GABA-A subunit alpha 5
Gabra6		1	GABA-A subunit alpha 6
Gabrb2		1	GABA-A subunit beta 2
Gabrb3	0.44	1	GABA-A subunit beta 3
Gabrd		1	GABA-A subunit delta
Gabrg1		1	GABA-A subunit gamma 1
Gabrg1	0.69		GABA-A subunit gamma 2
Gabrg2	0.79	1	GABA-A subunit gamma 2
Glycine Receptors			
Glra2	0.48	5	alpha 2 subunit
Glrb	1.14	5	beta subunit
Acetylcholine Receptors			
Chrm1			muscarinic 1, CNS
Chrna3			nicotinic, a polypeptide 3
Chrna4	0.74		nicotinic, a polypeptide 4
Chrna5			nicotinic, a polypeptide 5
Chrna6			nicotinic, a polypeptide 6
Chrna7	0.4		nicotinic, a polypeptide 7
Chrna9			nicotinic, a polypeptide 9
Chrnb1			nicotinic, B polypeptide 1
Chrnb2	0.74		nicotinic, B polypeptide 2
Chrnb3			nicotinic, B polypeptide 3
Chrnb4			nicotinic, B polypeptide 4
Serotonin Receptors			
Htr1d	2.56		serotonin receptor 1D
Htr3a		1	serotonin receptor 3A
Htr7	1.23	1	serotonin receptor 7
Adenosine Receptors			
Adora1	0.88	1	A1 receptor
Adora2a		6	A2a receptor
Adora2b			A2b receptor
Nucleoside (ATP) Receptors, P2X			
P2rx4	0.88	1	ligand-gated ion channel, 4
P2rx5	1.32	1	ligand-gated ion channel, 5
P2rxl1	5.36		P2X-like 1, orphan receptor
Nucleoside (ATP) Receptors, P2Y			
P2ry12			P2Y, G-protein coupled 12
P2ry14	0.38		P2Y, G-protein coupled, 14
P2ry5			P2Y, G-protein coupled, 5
P2ry6			P2Y, G-protein coupled, 6
Norepinephrine/Epinephrine Receptors			
Adra1a	0.64	1	Adrenergic alpha 1a
Adra2a	0.43	1	Adrenergic alpha 2a
Adrb1		1	Adrenergic beta 1
Dopamine Receptors			
Drd1		1	Dopamine receptor 1
Drd2		1	Dopamine receptor 2
Histamine Receptors			
Hrh1	1.25		Histamine receptor H 1
Hrh3			Histamine receptor H 3
Thyrotropin releasing hormone			
Trhr	2.05	1	TRH receptor
Trhr2			TRH receptor 2

Figure 7. BACarray data recapitulates known Motor Neuron physiology

Data from MN BACarrays was directly compared to available data for classical neurotransmitters. To perform this analysis, BACarray results were color coded as ‘expressed,’ ‘enriched,’ or ‘not expressed’. This classification was then compared to results reported in the adult rodent literature, color coded simply as either ‘expressed’ or ‘not expressed’ or left uncolored in cases where there were no studies or conflicting data.

IP/UB: *Fold change versus whole spinal cord for expressed genes.* ***RF****: Expression data from published rodent literature.* ***1*** *(Rekling et al., 2000),* ***2*** *(Nishi et al., 2001),* ***3*** *(Berthele et al., 1999),* ***4*** *(Towers et al., 2000),* ***5*** *(Malosio et al., 1991),* ***6*** *(Kaelin-Lang et al., 1999).*