SUPPLEMENTAL MATERIAL

Considerations for normalization and analysis across cell type specific data

There is a major distinction between microarray experiments of TRAP RNA compared to whole tissue, unbound or Total RNA, and this distinction has an important impact on the assumptions regarding normalization: any given cell only translates those mRNAs required for its functions, and at the levels required by that particular cell type. Thus any given cell will have a smaller number of detectable RNA species than a whole tissue sample, which consists of an aggregate of RNAs from cells with a variety of roles. Therefore the distribution of measurable RNAs between IP and Total samples should be different. This is shown in the histograms of Supplemental Figure 3a. Total samples show more RNA's with detectable signal, consistent with the measurement of a more complex population of mRNAs from a mixture of cells. This is an important consideration because some normalization and analysis methods assume only minimal differences in the distributions between samples, and may by default filter to remove those probesets with signal in a small number of samples(27), or force all samples to have identical distributions(13). This is clearly inappropriate when comparing widely divergent cell types in which most genes are expected to vary in expression, with many genes being highly enriched in a certain cell type.

The IPvTotal plot was used to examine the impact of different normalization methodologies. Proper normalization should minimize IP/Total ratios for negative control genes, and maximize it for positive controls. Among the most common methods for normalization of Affymetrix data are the robust multi-array normalization (RMA) and GeneChip RMA (GCRMA), both of which apply quantile normalization to all data sets using the assumption that all samples should have the same RNA distribution (13). However, the assumption that any one cell type should express the same number of mRNAs at similar proportions as any other cell type and/or that the distribution of the aggregate of many cell types (Total) should be similar to the distribution of a single cell type is not supported by our data (Supplemental Figure 3a and (1,2)). Consequently, quantile normalization across IP and Total samples resulted in forcing both samples into an artificial distribution that represents neither. Thus, RNAs that are present in the Total sample will have their signals reduced and RNAs that are not present in the TRAP RNA will be artificially inflated.

The impact of these considerations for specific genes is shown in Supplemental Figure 3b, a scatterplot of Purkinje cell IP vs. cerebellar Total for all probesets on the array, with quantile normalization performed either within groups (separately) or across groups (together). On average, normalization across groups results in a decrease in the ability to detect enriched messages (IP/Total for positive controls 8.17 separate vs. 6.92 together), and higher signal in negative controls (0.23 separate vs. 0.37 together). In the case of specific probesets, particularly those for negative controls with low signal, the difference can be quite dramatic. For example, those from the *Cnp1* gene change from a 0.1 IP/Total to a 0.5 IP/Total when the samples are normalized together, and the glial genes *Mog* and *Plp*, which are not detected in the IP when samples are normalized separately, appear as if they are present in the IP in Purkinje cells.

Quantile normalization still functions well in removing non-biological variability from biological replicates (multiple independent TRAP samples from the same line and

tissue). Thus we first GCRMA normalized within replicate samples. Then, to correct for any global biases in hybridization or scanning conditions, we performed global normalization to the biotinylated spike in controls provided by Affymetrix across all cell populations.

Background: sources and removal

Following normalization, for each cell population we plotted IPvTotal and displayed positive and negative controls to make initial judgments regarding the quality of a particular TRAP dataset. In Figure 1a, it is apparent there are some glial RNAs with detectable signal in the IP, though they are enriched eight to ten fold in the Total RNA (*red genes*, 0.12 average IP/Total). There are three possible explanations, which are not mutually exclusive.

First, it is possible that neurons are translating a very low level of glial genes. For example, it is known that Vimentin, expressed highly in glial progenitors(28), can be translated in adult neurons following injury(29). Second, in some cases there may be low levels of eGFP-L10a transgene expression in another cell type. For example, anatomical analysis (Supplemental Figure 9c) demonstrates that low levels of transgene expression in 'non-targeted' cell types can contribute signal to the TRAP microarray from the Lypd6 JP48 line. Though rare, careful driver selection can avoid this complexity. It is also possible to exclude data from these contaminating cell types in many cases posthoc by comparative analysis (1). Finally, as TRAP is an affinity purification method, there may be a small amount of RNA binding to the affinity purification reagents that is not derived from the labeled cells. To test this possibility, we performed TRAP on a wildtype brain and determined that the affinity purification reagents can bind a very small amount of RNA (Supplemental Figure 9a) in a manner proportional to the concentration of the lysates (Supplemental 9b). For most cell populations, this background represents a small fraction of the TRAP yield. However, in TRAP experiments with exceptionally low yield (<10 ngs), non-specific background can become problematic. Consistent with this, Supplemental Figure 2 shows increasing relative levels of negative control probesets as yield decreases for examples of experiments with good (*Pcp2*), low-moderate (*Cmtm5*), and very low yield (*Cort*).

Since the low yield IPs contain a larger proportion of non-specific background RNA that comes from unlabeled cell types in the tissue, it is more difficult in these samples to make the distinction between non-specific background and broadly translated messages. In spite of this difficulty, even low yield samples (eg *Cort)*, which have a substantial contribution from non-specific background, still show remarkable enrichment of the positive control (*Cort*) (Supplemental Figure 2b). Thus, these experiments also provide valid information (see also Figure 4b), although not of the same quality as those with minimal background.

We quantified this level of non-specific background as the average IP/Total ratio of those negative control genes that have measurable signal. Thus, from the examples in Supplemental Figure 2, *Cort* has an average non-specific background of 1.1, while *Cmtm5* has .48, and *Pcp2* has .05. We then tested if the background could be removed with a relatively simple filter using this measure. We excluded those probesets falling below this average non-specific background, plus two standard deviations. Assuming a linear contribution of non specific background to TRAP signal, and a normal distribution

of background signal intensities, theoretically this should remove the vast majority (96%) of those probesets that derive signal uniquely from background RNA. This threshold is shown as the red lines on Supplemental Figure 2. Filtering to remove these probesets prior to further analysis has the added advantage of reducing the number of probesets tested, thus reducing the requisite number of multiple testing corrections for downstream statistical analyses.

To determine if this filter is effective, we examined comparisons of two cell types from different tissues (Supplemental Figure 5b), or with differential levels of non-specific background contamination (Supplemental Figure 5a). Comparing cerebellar Purkinje cell IP data to Drd1+ medium spiny neuron IP data, without accounting for background, results in the apparent expression of the cerebellar granule cell-specific gene *Neurod1* in Purkinje cells (Supplemental Figure 5b, left panel). Simple filtering prior to the IPvIP comparison successfully removed this false positive result (Supplemental Figure 5b, right panel). Thus, regardless of the source of the non-specific background, simple filters based on negative controls can be used as a generic method to remove most probesets deriving from non-specific background.

As previously reported, there are also a group of mRNAs that apparently specifically bind the affinity reagents even in the absence of eGFP-L10a protein(1). These probesets have extremely high IP/Total ratios in every IP, including those from control, non-transgenic mouse brains. These may represent specific interactions between anti-eGFP antibodies or protein G beads and nascent peptides on the ribosomes. They were removed from subsequent analysis.

IPvTotal may indicate rarity of a cell type

The magnitude of IP/Total for positive controls can be used as a crude measure of the contribution of the targeted cell type's mRNAs to the total mRNA pool in the tissue of interest. Supplemental Figure 4 shows IPvTotal plots for a less frequent (a), and extremely common cell type (b), with similar levels of non-specific background (*red line*). One can see that the ratio of the IP/Total for the driver gene (*blue)* increases with the rarity of the cell. Logically, if a cell contributes 5% of the RNA in the total tissue, then the cell-specific genes should be 20 fold enriched. From this, we can estimate that Purkinje cells, with an average enrichment of 8 fold for their positive control genes (Figure 1a), contribute 12% of the RNA in the cerebellum. While this is a disproportionately high amount relative to their numbers (0.3-0.4% of cerebellum, (30)) this number is not unreasonable given their relatively large cytoplasmic compartments (estimated at 40x the volume of most common cerebellar cells). In addition to magnitude of ratio, there is a broad difference in the number of RNAs with high IP/Total values across the samples in Supplemental Figure 4a and 4b. In general, the number of RNAs that are differentially expressed in this analysis correlates with the distinctive properties of that cell type relative to their tissue. Of course, the exact magnitude of these fold changes can depend on the level of background signal in the IP, thus currently these rules serve as useful heuristics rather than precise measures.

Anatomical Considerations

Careful characterization of transgene expression is essential to the interpretation of the TRAP data. We typically characterize the eGFP levels both with and without

antibody staining. Those mouse lines with more robust expression have better yield and hence lower non-specific background. If there is visible eGFP without antibody in mouse brain sections, yields will generally be sufficient for microarray experiments. However, it is important to detect the presence of trace labeling in additional populations using anti-eGFP antibodies, as some signal from these populations would be detectable in the microarray data (Supplemental Figure 9c). Most mouse lines will express in multiple cell populations. Normally, these populations are present in distinct structures, and can thus be separated by careful dissection. Otherwise, microarray data from mixed populations can also be approached post hoc: for example a Bergman glial IP can be compared to a mixed Bergman glial/Unipolar Brush cell IP to identify Unipolar Brush cell specific genes (1).

Finally, it is important to consider if the experimental manipulation will impact the expression of the transgene itself. If so, this could have a dramatic impact on microarray results, particularly if the manipulation induces a dramatic change in the populations expressing the transgene, or the level of the transgene expression. This will need to be considered in the interpretation of the data.

Recommendations for design of TRAP experiments

Supplemental Figure 10 provides an example for good TRAP study design. For TRAP, standard good practices for microarray experimental design, execution, and analysis should be followed (31). Among these, it is particularly important to include careful checks of RNA quality and quantity before amplification. We recommend fluorometric measures for quantification, such at the Ribogreen assay, when measuring RNA concentrations of less than fifty nanogram per microliter, as well as Agilent Bioanalyzer assays to determine RNA integrity. Also, when amplifying RNA, it is important to start with the same amount of RNA from each sample, and use identical protocols. It is absolutely essential that experiment and control samples should be collected and amplified simultaneously or in balanced pairs, to control for non-specific amplification biases and batch effects, which afflict all microarray experiments (31). This is especially important when investigating more subtle manipulations such as drug treatments or the impact of knockouts on specific cell types. Finally, it is frequently advisable to pool tissue from multiple animals for each condition to increase yield in the case of small structures, as well as to help average out minor variations in dissection or treatment from animal to animal. We conduct at least three replicate affinity purifications, per experimental condition, and typically pool from three to six animals per replicate. However, it is clear the amount of background is dependant of the concentration of tissue homogenized (Supplemental Figure 9b). Maintaining an approximate 100mg/ml (or less) ratio of tissue to homogenization buffer, is recommended when pooling tissue to reduce non-specific background.

Future improvements of the TRAP methodology may eliminate the need to collect a Total measure and subtract non-specific background, and several strategies are actively being pursued to allow this. Currently, low level transgene expression in alternate cell types can be controlled by selecting more specific drivers. Weak drivers can be replaced with stronger ones. Often TRAP data with high background is mined to select stronger drivers yielding lines targeting the same cell type, but with better yield and lower

background, such as replacing the *Cmtm5* line with the *Cnp1* line(1) for mature oligodendrocytes.

Supplemental Figure 1. Illustration of the TRAP method. BAC transgenic mice are generated to target the expression of a fusion of eGFP and a ribosomal protein (L10a) to a specific population of cells in the mouse brain (shown here are motor neurons, targeted using a BAC containing the motor neuron specific Choline Acetyl Transferase gene). To isolate cell specific translational profiles, the entire tissue is homogenized, treated with detergents to solubulize the endoplasmic reticulum, and centrifuged to prepare a crude homogenate containing a mix of eGFP tagged and untagged polysomes. Importantly the tagged polysomes come uniquely from the motor neurons. Tagged polysomes, and associated mRNA, are then purified from homogenate using antibodies against eGFP, which are bound to protein G coated magnetic beads. Both purified and unpurified RNA are them amplified and hybridized to microarrays to profile the mRNA populations.

Supplemental Figure 2. Assessment of IPvTotal plots. **a)** A ratio threshold (*red line*) can be set between non-specific background and broadly translated genes, based on the values of the negative control genes *(red,* glial genes for neurons, and neuronal genes for glia). High yielding lines (Pcp2) generally have low background, while low yielding lines, (Cmtm5), have a correspondingly higher background. **b)** Compared to higher yielding lings, with very low yielding line (*Cort*), broadly translated RNA's can not be easily distinguished from non-specific background, though RNAs representing driver genes (*blue)* are consistently enriched. *Black lines, all plots, 0.5, 1, 2 IP/Total ratio lines.*

Supplemental Figure 3. Improper normalization can create false positive signals. **a)** Average histogram of probeset signal intensities for all IP samples (*dotted lines)* compared to all Total samples (s*olid lines*) reveals differences between distributions. In particular, IP samples have more undetectable probesets (*first bin, red arrow*), consistent with RNA purified from discrete cell populations compared to RNA from a mix of cell types. **b)** As illustrated with IPvTotal plots for Purkinje cells, quantile normalization (forcing identical distributions) of IP and Total samples together (right panel) produces artificial signal in negative controls (red genes, in yellow circle, shifted right). *Black line: 1 fold. Red line, line of best fit through negative controls. Blue line, line of best fit through positive controls.*

Supplemental Figure 4. IPvTotal is dependant on the composition of the Total. **a)** IPvTotal for astroglial sample, under control of the driver *Aldh1L1* (blue), shows the enrichment of many genes, as illustrated by number of probesets falling above the two fold line, suggesting glia contribute relatively a small fraction of the RNA pool of the whole cerebellum. **b)** IPvTotal for an extremely common cell type, the granule cell of the cerebellum, fails to show enrichment of granule cell-specific driver *Neurod1*, in spite of low background (*red line)*, suggesting granule cells contribute a significant fraction of whole cerebellar RNA. **c)** A scatterplot of an IPvIP comparison of the *Neurod1* IP to *Pcp2* IP reveals clear enrichment of the *Neurod1* probeset in the *Neurod1* IP demonstrating the *Neuod1* IP is enriched in granule cell RNA. *Black lines, 0.5, 1, 2 fold. Red genes, cell-specific negative controls, as Figure 1*

Supplemental Figure 5. Simple thresholds improve IPvIP comparisons. **a)** An illustration of how to filter data using simple thresholds for background and low expressed genes. Background threshold was set at mean plus two standard deviations of the detectable negative control genes. Probesets with expression below 50 were also removed. **b)** IPvIP comparison of cerebellar Purkinje cells (y-axis) to cerebellar Granule cells (x-axis), which have slightly different levels of background, have corresponding 'differential' expression of glial mRNAs (red genes, left panel), which can be removed by applying simple thresholds (right panel). **b)** Likewise an IPvIP of cerebellar Purkinje cells to Drd1a medium spiney neurons, which have background from different tissues, have a corresponding 'differential' expression of background mRNAs from other tissue specific cell types *(red, Drd2 from the striatal Drd2+ medium spiny neuron, and Neurod1 from the cerebellar granule cell)* which can be removed with simple thresholds (right panel).

Supplemental Figure 6. a) Illustration of method for determining p-values for specificity index for one cell type. **b)** Illustration of analytical flow to identify cell specific and enriched genes for all cell types.

Supplemental Figure 7. Updated chip definitions improve accuracy and interpretability of TRAP experiments. **a)** IP/Total (log2, red) and specificity p-value (-log 10, blue), for all cell types (y-axis), for the probeset representing the known oligodendrocyte gene MBP as measured using custom chip definition files (cdf) which remove misaligned probes(19). **b)** Four examples of probesets for MBP using Affymetrix cdf files. *Oligodendrocyte populations marked with *.* **c)** Probeset for Purkinje cell-specific gene, PCP2, using custom cdf. **d)** Four of the probesets for PCP2 using Affymetrix cdf files.

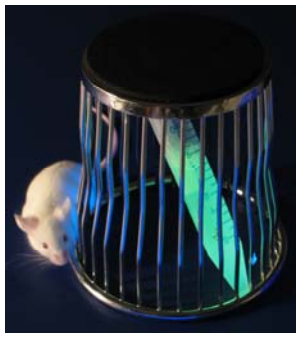
Supplemental Figure 8. Dramatic differential translation of the GalNT gene family, suggest cellular specialization of Golgi apparti. **a-f)** Combined Specificity Index p values (blue bars, -log 10 scale) and IP/Total values (red bars, log 2) across all cell populations for a selection of the *UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase* golgi protein family. **a)** Galnt3 shows specific translation in oligodendrocyte progenitors. **b)** Galnt4 shows translation in astrocytic cell types. **c)** Galnt6 shows specific translation in mature oligodendrocytes. **d)**Galnt14 shows specific translation in Corticospinal/Corticpontine neurons. **e)** GalntL2 shows specific translation in granule cells of cerebellum.

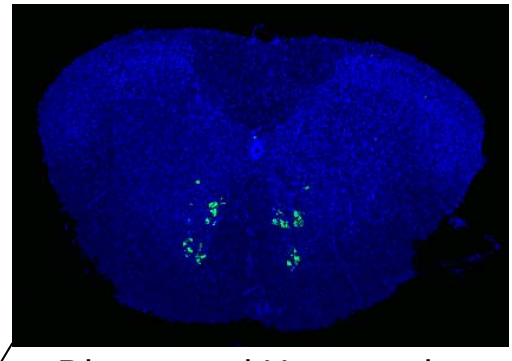
Supplemental Figure 9. Two potential sources of non-specific background **a)** Representative Picochip capillary electrophoretic traces from Agilent Bioanalyzer for RNA from a TRAP experiment on two cerebellums from Bergman glial (*Sept4*) bacTRAP mice (left panel) or wild type mice (center panel), suggest a small amount of RNA may derive from non-specific interactions of unlabeled RNA with affinity purification reagents *Arrows: 18 and 28s Ribosomal RNA peaks.* IPvTotal plot (right panel) shows low level of signal in known negative control genes (*neuronal genes, red).* Driver genes known to be highly expressed in Bergman glia (*Sept4, Aldh1L1, blue)* show strong enrichment, while drivers for other cerebellar cell types (*Neurod1, Pcp2, Lypd6, blue)* show IP/Total ratios similar to negative controls. *Black Lines: 0.5, 1, 2 fold lines. Red Line: Average IP/Total ratio of negative controls. Green Line: background IP/Total ratio level suggested by non specific yield (4.5ng) divided by TRAP yield.* **b)** Amount of non-specific RNA binding to affinity purification reagents depends on amount of tissue. Various amounts of brain tissue from wild type mice were homogenized in a consistent volume of homogenization buffer, and TRAP methodology was carried forward. Increasing amount of tissue increases non-specific background. A 1:10 w/v ratio, or less, is recommended to minimize this. **c)** Confocal immunofluorescence for eGFP in Stellate/Basket neuronal (*Lypd6)* bacTRAP line shows low level transgene expression in additional cell types. Top left, DAPI nuclear counterstain delineates layers of cerebellum (*WM, white matter, GCL, granule cell layer, PCL, Purkinje cell layer, ML, molecular layer).* Stellate and Basket cells of molecular layer clearly contain eGFP-L10a (top center, top right). The same eGFP image shown in range scale (*blue pixels: no signal, red pixels: saturated)* with excessive gain (bottom center), or normal gain (bottom left) shows trace eGFP-L10a in white matter glia (*red arrow).* IPvTotal (bottom right) shows clear enrichment of driver (*Lypd6, blue)*, but moderate levels of signal from negative control genes (*glial genes, red)* or drivers for other cell types showing trace expression *(Sept4, blue).* Note that granule cell driver, neurod1 (*green*), has low signal, consistent with lack of expression in granule cells. *Green line: average IP/Total ratio of neurod1 probesets. Red line: average IP/Total ratio of glial genes.*

Supplemental Figure 10. Recommendations and examples for TRAP experimental design.

Supplemental Table 1. Positive and negative Controls. **a)** List of genes scored as specific to the Purkinje cell layer in the cerebellum by three independent reviewers, based on online ISH atlases(11,12). **b)** List of known markers for glial cell types, which may be used as negative control genes for neuronal samples. **c)** List of markers for neurons (neurofilaments and synaptic proteins), which may be used as negative controls for glial samples.

Supplemental Figure 1

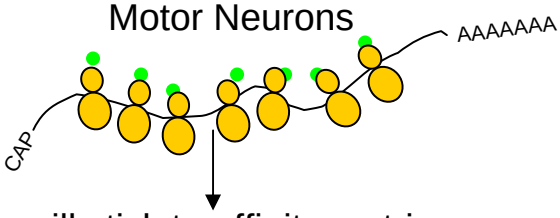




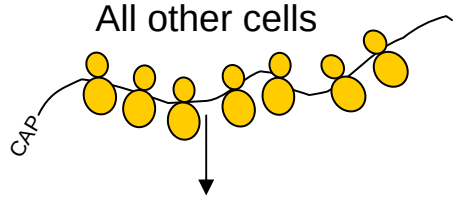
Generate and Characterize
BAC transgenic mice

Dissect and Homogenize
Tissue of Interest

Polysomes from
Motor Neurons



Polysomes from
All other cells



will stick to affinity matrix

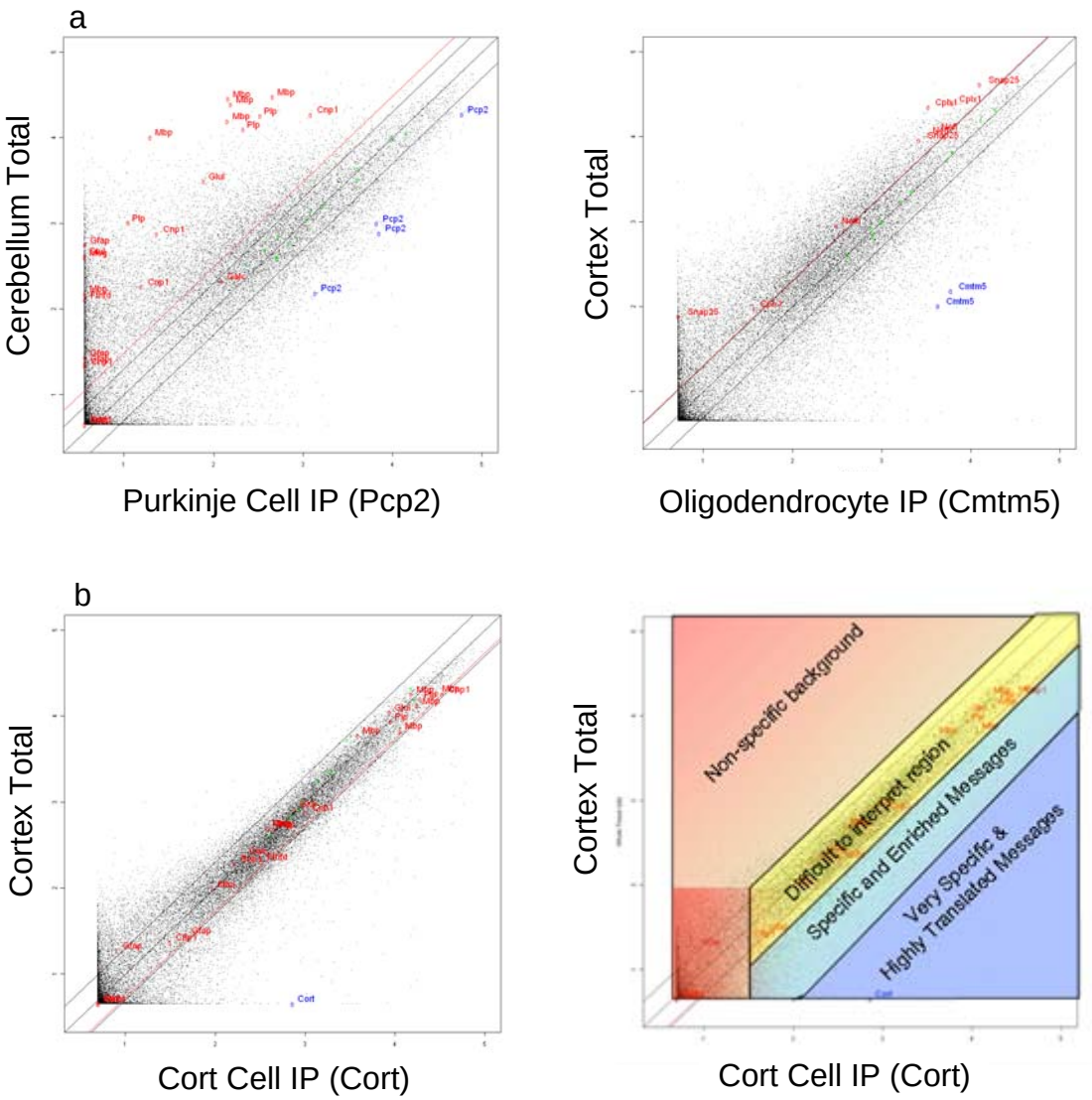
will not stick to affinity matrix

Harvest enriched mRNA (IP)

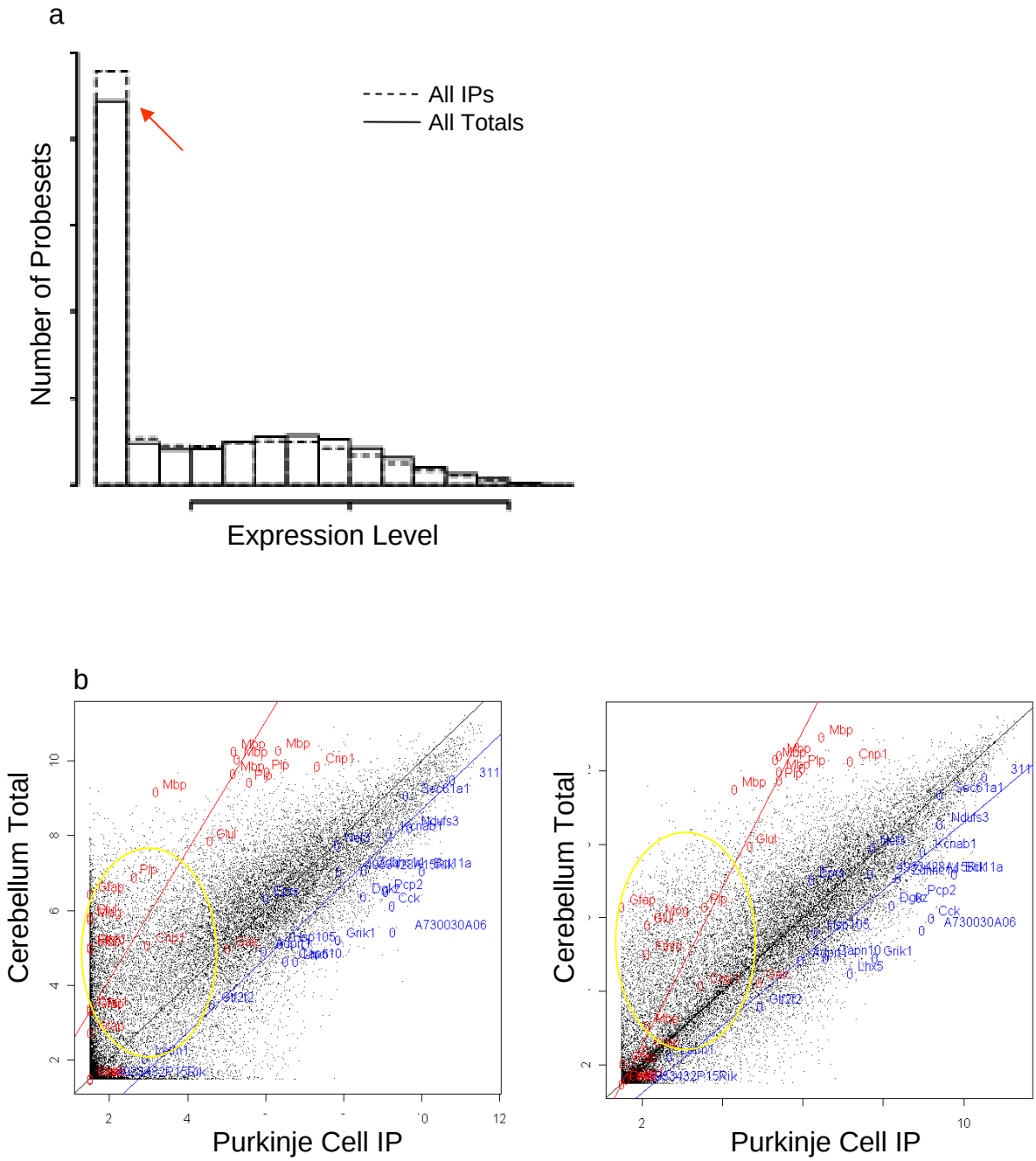
Harvest whole tissue mRNA (Total)

Hybridize
Microarrays

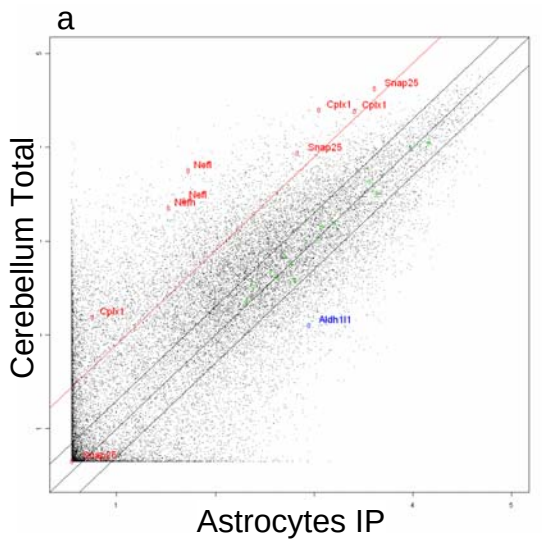
Supplemental Figure 2

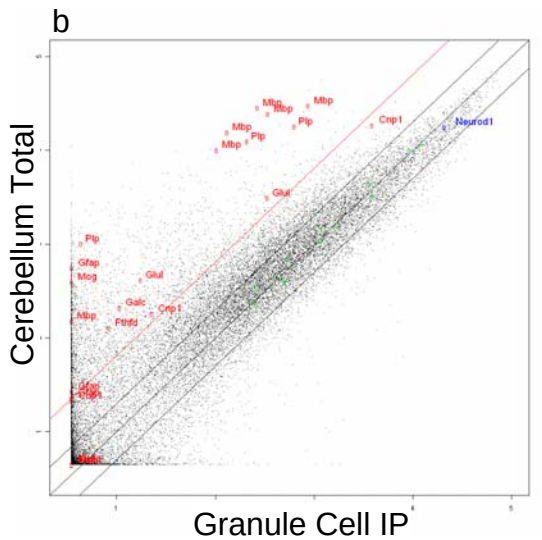


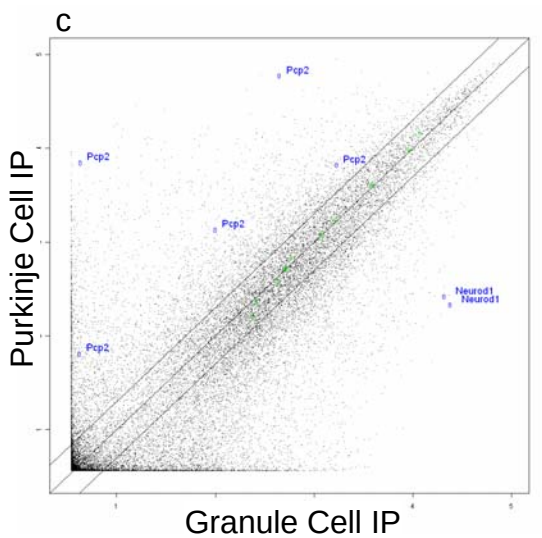
Supplemental Figure 3



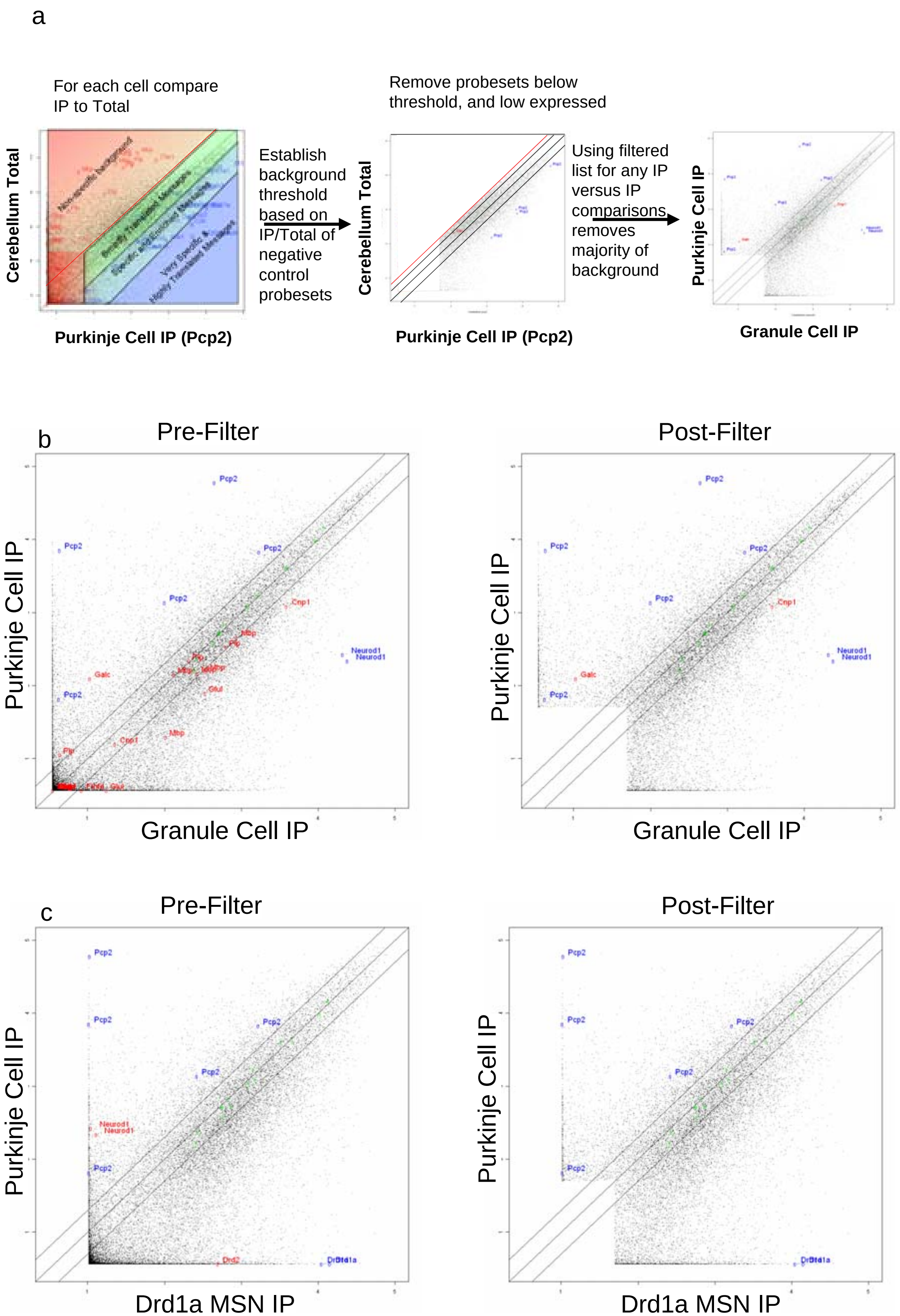
Supplemental Figure 4



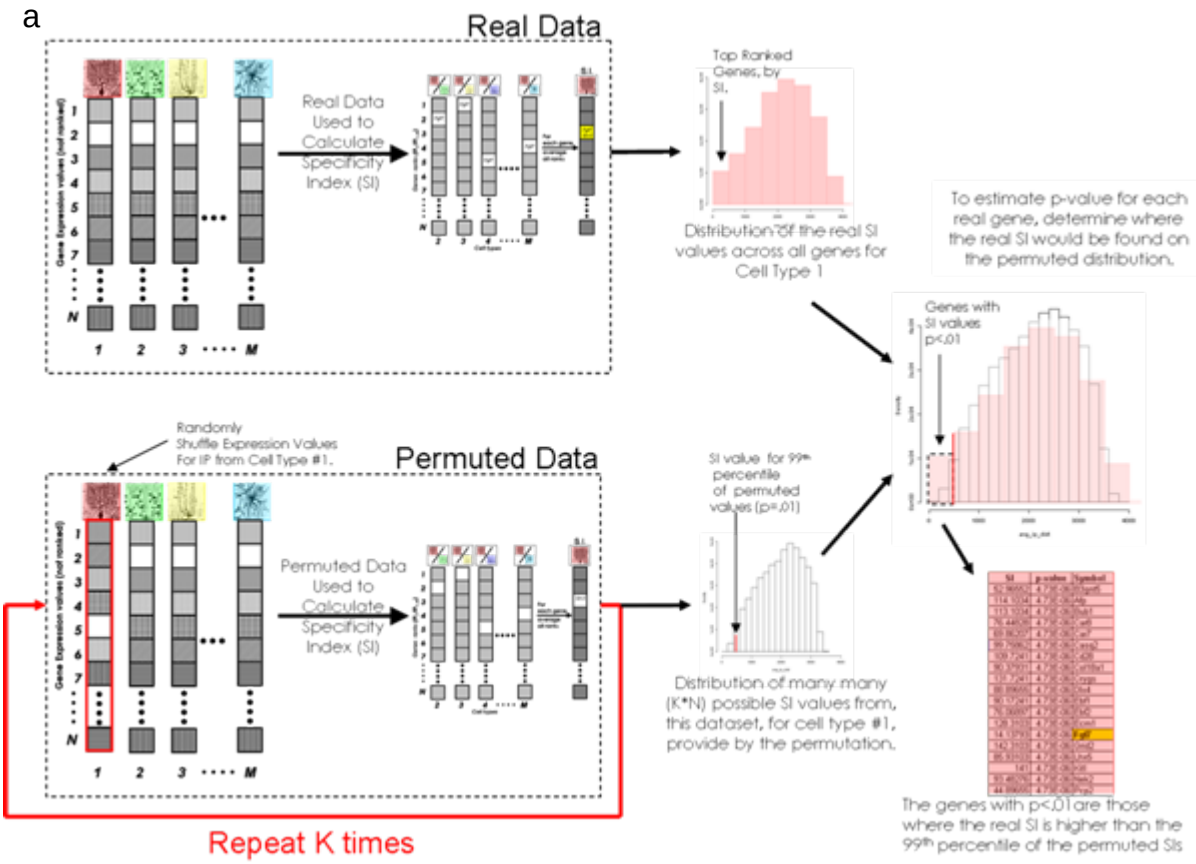


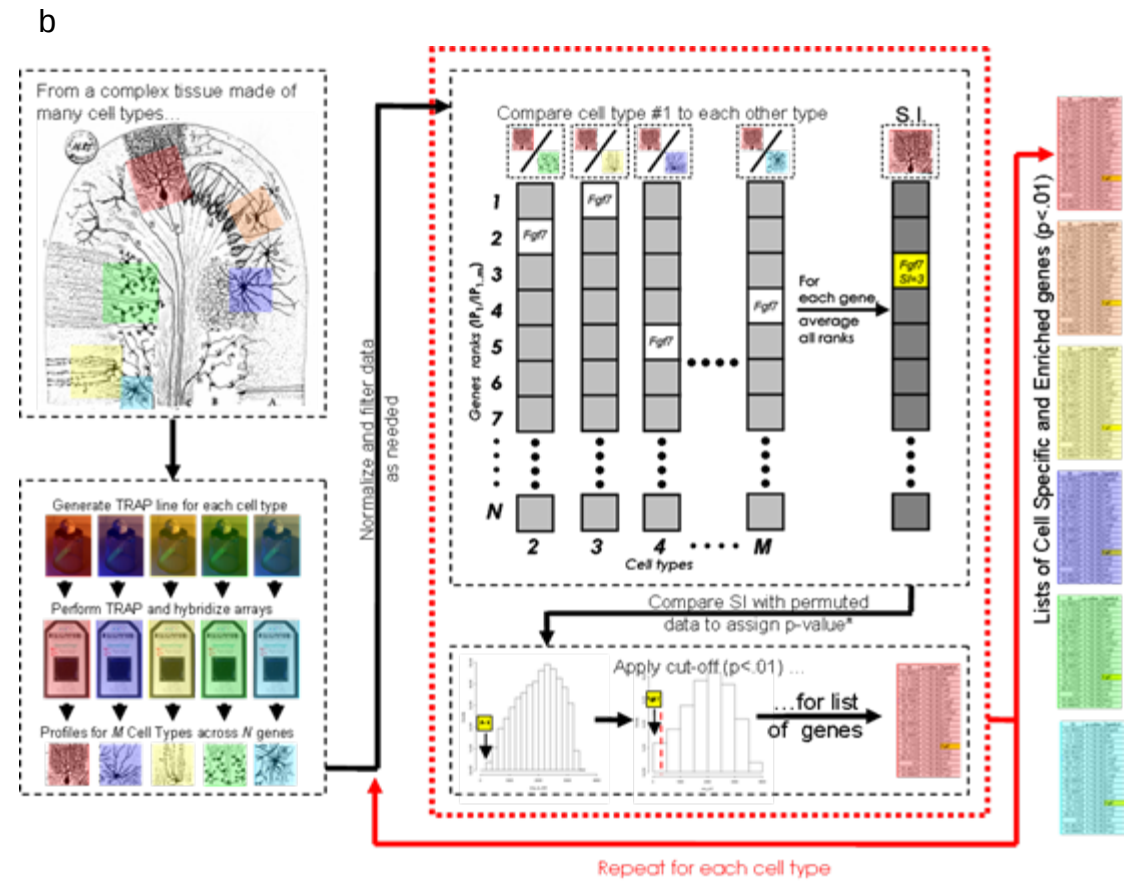


Supplemental Figure 5

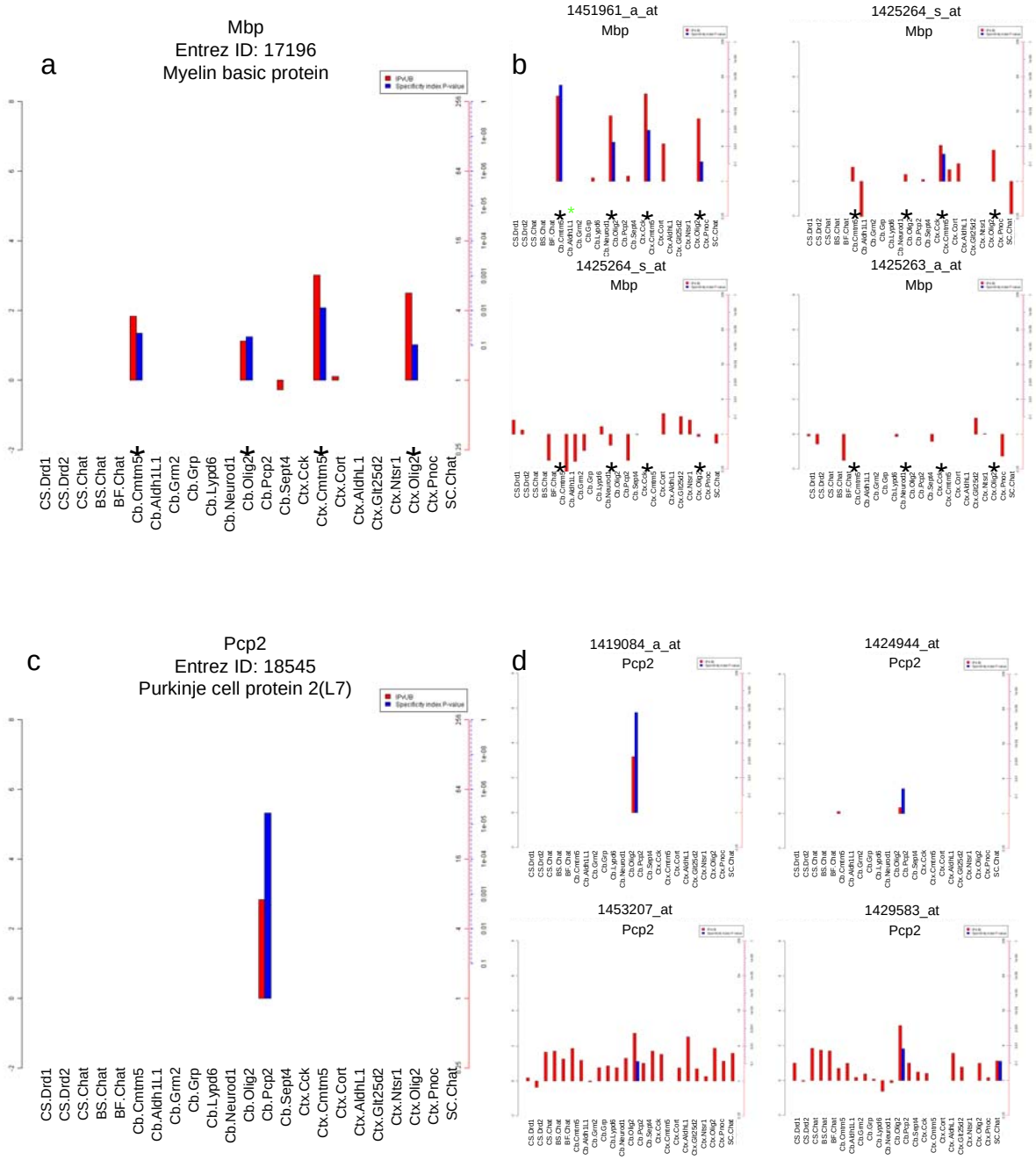


Supplemental Figure 6

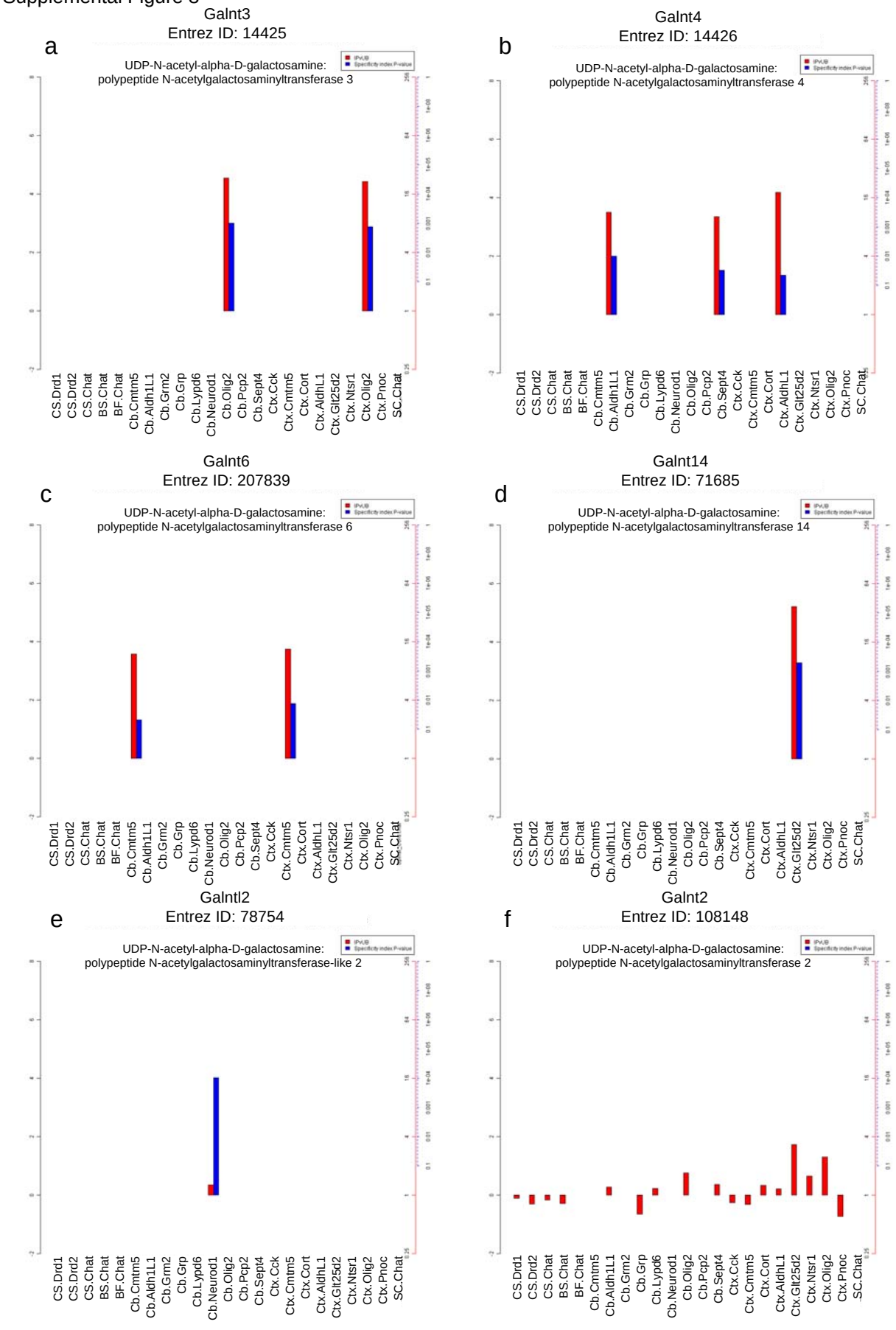




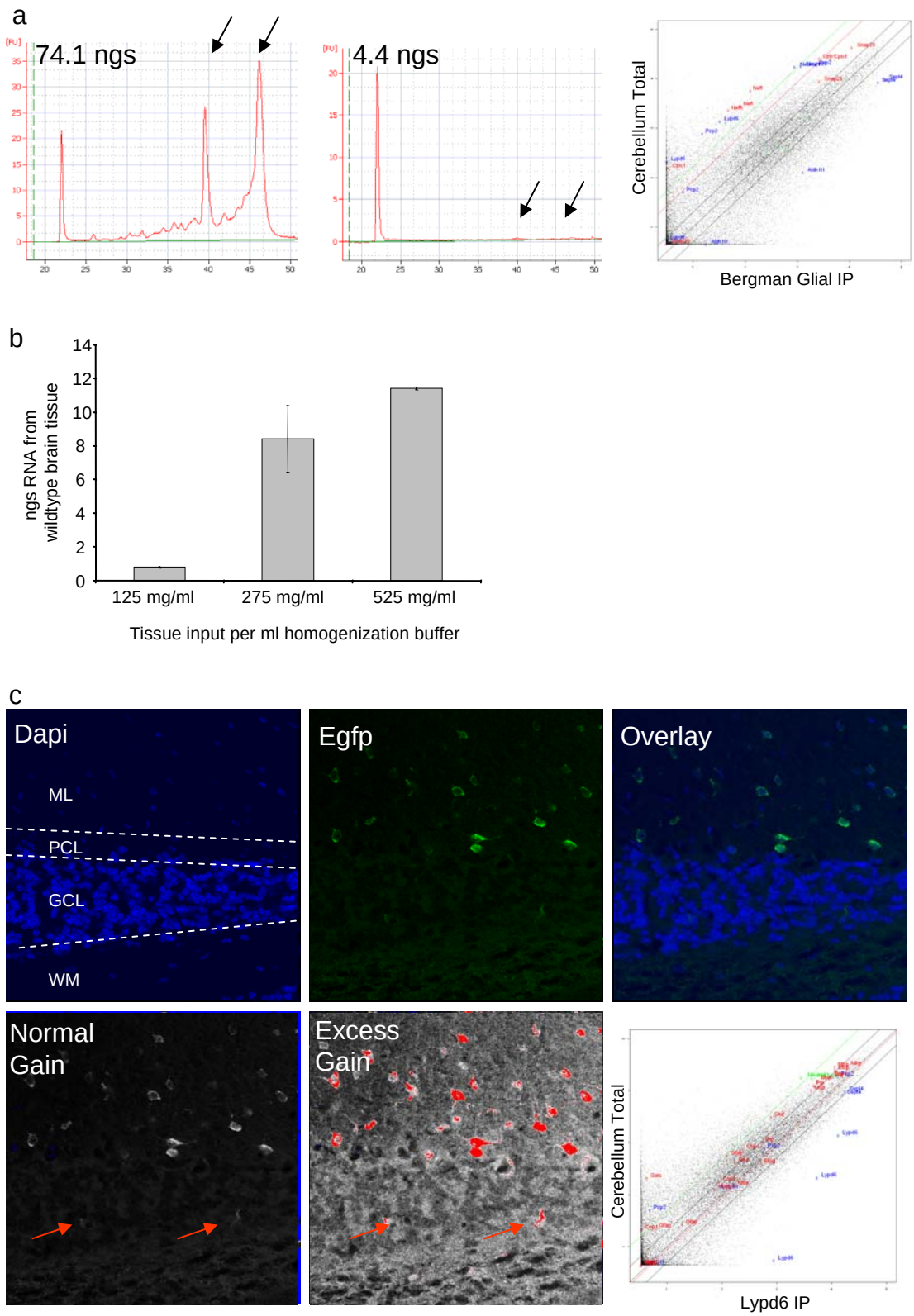
Supplemental Figure 7



Supplemental Figure 8



Supplemental Figure 9



Supplemental Figure 10

Principles	Example
Design **1)** Define question **2)** Select line appropriate for cell type **3)** Plan to balance conditions across batches	**Design: MECP2 in glia** **1)** What is the impact on MECP knockout on cortical astrocytes *in vivo?* **2)** The previously generated Aldh1L1 JD130 line is expressed exclusively in astrocytes [1]. Cross MECP2 KO with Aldh1L1 line to generate breeders. **3)** Plan for three batches. Each batch is three bacTRAP/MECP null mice and three litermate bacTRAP only controls.
Anatomy **1)** Confirm transgene expression **2)** Does manipulation alter expression?	**Anatomy: MECP2 in glia** **1)** Aldh1L1 bacTRAP line was previously and thoroughly characterized [1]. Skip this step. **2)** In first litter of bacTRAP/MECP null mice, confirm transgene is still expressed uniquely in astrocytes, and at comparable levels to littermate controls.
Immunoprecipitation and Microarrays **1)** Harvest paired conditions together in one batch and collect total polysome sample **2)** Quantify RNA carefully, and start with identical amounts of all samples for amplification **3)** Amplify and hybridize all batches together, if feasible	**IP and Microarrays: MECP2 in glia** **1)** Day 1(3-4 hours): Harvest cortices from three MECP2 null/bacTRAP mice and 3 bacTRAP littermate controls. Pool MECP2 null and control tissue separately. Homogenize and prepare polysomes. Prior to immunoaffinity step, set aside 20 uls for Total polysome sample. Complete immunoaffinity purification, and RNA extraction of Total and IP'd RNA until isopropanol precipitation. Store RNA at -80. Day 2:Repeat day 1 with second batch of 3 MECP2 null and 3 controls. Day 3:Repeat day 1 with third batch of 3 MECP2 null and 3 controls. **2)** Day 4 (2 hours): Complete purification all three batches of frozen RNA. Quantify with Ribogreen assay. Assure that RNA integrity is above 8 for all samples with Bioanalyzer assay. **3)** Day 4-6: From each sample, take 20 ngs of RNA, and begin Affymetrix two cycle amplification for all twelve samples. Carry through amplification and hybridization of all samples together.
Analysis **1)** GCRMA normalize together only those samples that should have the same distribution. Global normalize subsequently to biotinylated spike ins. **2)** Compare IPs to Total. Calculate a background threshold using the IP/Total ratio for negative controls. Remove from further analysis those probesets with IP/Total below this threshold. **3)** Conduct statistical analysis on remaining probesets.	**Analysis: MECP2 in glia** **1)** GCRMA all Total samples together. GCRMA together all IP'd samples (from MECP2 null and controls). Normalize all samples(total and IP) together to spike in controls **2)** Calculate IP/Total for a list (Supplemental Table 2) of non-astrocyte probesets (ie neuron specific genes). Remove all probesets below the Mean + 2 S.D. of the ratios on this list from further analysis for astrocytes. **3)** Use the Limma module of Bioconductor to detect those genes that change significantly between MECP2 null IP and control IP. These represent the astrocyte's response to the knock out. Genes that change significantly between the MECP2 null Total and control Total samples will represent the response of the other cells in the tissue. These can be compared listwise or statistically to determine the astrocyte specific response.

Supplemental Table 1

A **Postive controls for Purkinje cells**
A930006D11
3110001A13Rik
4933428A15Rik
4933432P15Rik
A730030A06
Adprt1
Bcl11a
Capn10
Cck
Dgkz
Eprs
Grik1
Gtf2f2
Hsp105
Kcnab1
Letm1
Lhx5
Ndufs3
Nef3
Pcp2
Sec61a1
Zdhhc14

B **Negative controls for Neurons**
Mbp
Aldh1l1
Cspg4
Galc
Glul
Mag
Mobp
Mog
Olig2
Plp1

C **Negative controls for Glia**
Snap25
Cplx1
Nefh
Nefl
Nefm