

**Manuscript:** Genetic Determinants of Distinct Projection Neuron Circuits in Mammalian Basal Ganglia

## **MIAME Report**

### **Experiment Design**

**The goal of the experiment:** Identification of new genes differentially expressed in direct pathway and indirect pathway striatal projection neurons

**A brief description of the experiment:** Using fluorescent activated cell sorting (FACS) and microarray analysis we have identified reproducible gene expression changes between two different subtypes of medium spiny neurons in P20 and adult mice, striatonigral neurons, which comprise the direct pathway; and striatopallidal neurons, which comprise the indirect pathway.

**Keywords:** Agilent, medium spiny neurons, basal ganglia, fluorescent activated cell sorting, transgenic mouse

**Experimental factors:** gene expression was investigated in two distinct neuronal populations: striatonigral neurons and striatopallidal neurons

**Experimental design:** total RNA from five biological replicates (different animals) of Drd2-EGFP mice (D2-mice) labeling striatopallidal neurons, and two Drd1- and three Chrm4-EGFP mice labeling striatonigral neurons (D1- and M4-mice) on an FVB/N background at P20 were compared in pairs. Two biological replicates of D2-mice and two of D1 mice were compared for 2-month-old animals. These experiments were performed on two different Agilent platforms: Developmental and Mouse I slides. For their annotation see [www.Agilent.com](http://www.Agilent.com) and Methods section of the manuscript.

In addition we performed an independent experiment where different amounts of input RNA (30pg, 300pg and 3 ng) were tested. See folder “Dilution experiment”.

**Quality controls:** quality of total RNA was established using 260/280 ratios and Agilent 2100 Bioanalyzer. Internal controls within the slides were used according to the manufacturers recommendations ([www.Agilent.com](http://www.Agilent.com)).

**Links to the publication, any supplemental websites or database accession**

**numbers:** the data is currently being submitted to GEO, it is also will be available at <http://geschwindlab.medsch.ucla.edu>.

**Samples used, extract preparation and labelling:**

**The origin of each biological sample:** see Methods section of the manuscript.

**Manipulation of biological samples and protocols used:** see Methods section of the manuscript.

**Experimental factor value for each experimental factor, for each sample:** D1 or M4 vs D2 samples. For “Dilution experiment” different amounts of input RNA were tested: 30pg 300pg and 3ng.

**Technical protocols for preparing the hybridization extract:** were performed according to manufacturers recommendations ([www.Agilent.com](http://www.Agilent.com)).

**External controls:** see manufacturer’s website ([www.Agilent.com](http://www.Agilent.com)).

**Hybridization procedures and parameters** were used according to manufacturer’s recommendations (Agilent).

**Measurement data and specifications**

**Data:** The raw data containing raw signal intensities, corresponding local background values, LOWESS normalized values and derived ratios are provided in the MIAME folder “P20 raw and normalized” and “Adult raw and normalized” for P20 and adult

comparisons subsequently. Each file represents data values for one single comparison. The ID columns represent the the following: g – Cy3; r – Cy5; gb – GeneBank accession number; ProbeName – Agilent probe number; LogRatio – log based 10 ratio of Cy5/Cy3 normalized values; ProcessedSignal – local background subtracted and LOWESS normalized signal intensities; MeanSignal – mean of the raw signal intensities, BGMeanSignal – local background of the raw signal intensities. Raw data for “Dilution experiment” is presented in “Table. Raw dilutions data”. Column titles are self-explanatory.

**The normalized and summarized data** is presented in P20 and Adult LOWESS excel file. Normalized data and derived ratios were combined from both Mouse I and Developmental slides. Normalized data for “Dilution experiment” where different amounts of input RNA were tested are presented in Table. LOWESS Dilutions. Annotations are self-explanatory.

**Data extraction and processing protocols:**

**Image scanning hardware and software, and processing procedures and**

**parameters:** Agilent scanner with the default parameters preset by manufacturer ([www.Agilent.com](http://www.Agilent.com)) was used for scanning the slides.

**Normalization, transformation and data selection procedures and parameters:** see Methods section of the manuscript.

**Array Design:**

**Title:** Agilent-011472 Mouse Development Oligo Microarray and Mouse I Oligo Microarray.

**Description:** [www.Agilent.com](http://www.Agilent.com)

**Technology type:** *in situ* oligonucleotide

**Distribution:** commercial

**Manufacturer:** Agilent Technologies

**Manufacture protocol:** see manufacturer's web site at <http://www.agilent.com/>

**Catalog number:** G4120A

**The sequence for oligonucleotide based reporters:** see <http://www.agilent.com/>

**The source, preparation and database accession number for long (e.g., cDNA or PCR product based) reporters:** see <http://www.agilent.com/>

**Principal array organism:** *Mus musculus*

## Data Analysis

The intensities were taken by subtracting the background value from the signal value in each spot for each array. These expression intensities were transformed as log base 2, which allows a natural interpretation of differential expression as fold changes and makes the intensity distribution more symmetric and the error variances more homogeneous. Negative control spots in each array were used to assure the data quality. Log-transformed signal intensities below the mean negative control + 2.64sd in each array were excluded from further data analysis. The highest variability between the arrays was observed for brainstem and cerebellum samples. These resulted in higher values of mean + 2.64·sd of negative controls, corresponding to approximately 8 and 8.5, respectively. These values were used as the lower bound to remove noise uniformly over the arrays for each region. One of the control cerebellum arrays (wcb3) had highest variability and was skipped from further analysis. Arrays were lowess normalized in the groups corresponding to a specific brain region (Astrand, 2001; Bolstad et al. 2003). With normalized data, the local-pooled-error (LPE) method for identifying differentially expressed genes was used (Jain et al. 2003). LPE method is known to be appropriate for small number of replicated microarrays (Jain et al, 2003). The LPE was derived by first evaluating the baseline error distribution for each condition (mutant or wild type) for each sample (brain region). For example, with 3 replications of mutant type for a given brain region, all pairwise comparisons of (M vs. A) were pooled together for estimation of variances of M within subinterval derived by quantiles of A. The number of subintervals is fixed as 100 at the first step. At the second step, the number of subintervals is chosen adaptively according to the median

and standard deviation of each subinterval for more precise estimation of baseline error estimate. Then a smooth local regression curve was fit to the error estimates. The baseline error distribution for wild type for a given brain region is similarly derived.

With these estimates of variances in each condition for a given brain region, the LPE test was performed. First, each gene's medians under the two compared conditions are calculated. The LPE statistics for the median difference was then calculated as

$$z = \frac{Med_1 - Med_2}{\sigma_{Pooled}},$$

where  $Med_i$ ,  $i = 1, 2$ , is the median intensity of the mutant and wild type, respectively;

$\sigma_{Pooled}^2 = \frac{\pi}{2} [\sigma_1^2(Med_1)/n_1 + \sigma_2^2(Med_2)/n_2]$  where  $n_1$  and  $n_2$  are number of replicates

in the two conditions;  $\sigma_i^2(Med_i)$ ,  $i = 1, 2$ , is the estimate of the variance from each LPE baseline-error distribution at each log-intensity  $Med_i$ . LPE statistics,  $Z$  follows a Normal distribution under null hypothesis that the gene is not significantly different between two compared conditions. Thus, the raw p-values were obtained for each gene. The FDR (false discovery rate) described in Benjamini and Hochberg(1995) was evaluated under 0.05 level for a p-value adjustment for multiple comparisons.

In addition, total brain RNA was used to evaluate the data variability between independent arrays. Six independent hybridizations have been performed and the ratios have been derived between the targets hybridized onto the different arrays (Table).

**References:**

Astrand M. Normalizing oligonucleotide arrays 2001.

(<http://www.math.chalmers.se/~magnusaa/maffy.pdf>)

Benjamini Y. and Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of Royal Statistical Society Series B: Methodological. 57:289-300.

Bolstad M, R.A Irizarry RA, Astrand M and Speed TP. 2003. A comparison of normalization methods for high-density oligonucleotide array data based on variance and bias. Bioinformatics. 19:185-193.

Bolstad B. Probe level quantile normalization of high density of oligonucleotide array data. (<http://www.stat.berkeley.edu/~bolstad/>)

Jain N, Thatte J, Braciale T, Ley K, O'Connell M and Lee JK. 2003. Local pooled error test for identifying differentially expressed genes with a small number of replicated microarrays. Bioinformatics. 19:145-151.