ALLEN INSTITUTE for BRAIN SCIENCE
Fueling Discovery

ALLEN **Human Brain Atlas**

TECHNICAL WHITE PAPER:
MICROARRAY PLATFORM SELECTION FOR THE ALLEN HUMAN BRAIN ATLAS

INTRODUCTION

This document describes the process and rationale by which the Allen Institute for Brain Science selected the microarray platform and service provider for the ALLEN **Human Brain Atlas**. The four most prominent commercially available microarray platforms and five different service providers for processing were assessed. Microarray data quality was compared in terms of reproducibility, ability to correct for batch bias, sensitivity to differentially expressed genes over two structure samples, and expression agreement with *in situ* hybridization (ISH) data. Additional considerations were throughput capacity and experience with high volume projects at each service provider, as well as the overall cost of the arrays and processing. After site visits to the top two service providers, as determined by data analysis and external expert review of this study, the decision was made to use the Agilent microarray platform. Microarray data were generated initially by Beckman Coulter Genomics, (previously Cogenics), then by Covance Genomics Laboratories after completion of the third brain. Covance was selected for array data generation following pilot testing and analysis verifying that data generated by Covance was highly comparable to data from Beckman Coulter Genomics.

MATERIALS AND METHODS

Comparisons were made among four human microarray platforms, which represented the four most prominent commercially available microarray platforms at the time of this study (third quarter of 2008). Five service providers were selected to process these microarrays. The four platforms were:

- **Agilent** 4x44 Whole Human Genome array;
- **NimbleGen** 12x135K bead array;
- **Illumina** Human HT-12 BeadChip; and
- **Affymetrix** GeneChip Human Exon 1.0 ST Array.

Fresh-frozen, post-mortem, non-diseased human brain tissue was used. Tissue samples for microarray analysis were obtained using a tissue punch to sample the structures of interest. Four human brain structures from three different individuals were analyzed. The four brain structures comprised two cortical areas (from the same tissue sample, i.e., single individual) and two sub-cortical areas (from two additional individuals) as follows:

- Brodmann Area 17 (also known as primary visual cortex or V1, and labeled as **Cortex 17** here);
- Brodmann Area 18 (also known as secondary visual cortex or V2, and labeled as **Cortex 18** here);
- **striatum**; and
- **thalamus**.

Three identical RNA isolates (technical replicates) were prepared for each of the four brain structures, resulting in a set of 12 distinct samples for microarray analysis. RNA was isolated using Ambion's MELT chemistry on the MagMaxExpress-96 instrument, quantitated using the Nanodrop 8000, and qualitated using

Agilent's Bioanalyzer 2100 Nano and Pico chips. All RNA for each of the four structures was pooled by structure after original quality assessment, normalized to 10ng/µl, re-quantitated on the Nanodrop 8000 and re-qualified on the Bioanalyzer Pico chip. The quality results are summarized in Table 1, and Appendix C displays the Bioanalyzer Pico electropherograms of the pooled RNA samples at 10ng/µl. The RNA was then aliquoted (13µl (130ng) per tube) into seven sample sets (Table 2, labeled A-G), each comprising all 12 samples. These sample sets were shipped on dry ice to the service providers. Some, but not all, service providers tested more than one microarray platform.

The thalamus RNA samples were of poorer quality than the others, and the service providers were expected to report these samples as outliers based on their quality control processes. The thalamus samples were excluded from the analyses described below and did not factor into the platform selection process.

Table 1. RNA quality for each of the four brain structures sampled.

Tissue	RNA quality	
	18s/28s	RIN
Cortex 17	1.2	6.4
Cortex 18	0.9	5.8
Striatum	0.7	4.7
Thalamus	0.3	3.1

Table 2. Microarray platforms and de-identified service providers.

Platform	Service Provider	Provider/Platform Combinations
Agilent 4x44 Whole Human Genome array	A	A.Agilent
NimbleGen 12x135K bead array	B	B.Nimble
Illumina Human HT-12 BeadChip	C	C.Illumina
	D	D.Illumina
Affymetrix GeneChip Human Exon 1.0 ST Array	E	E.AffyN*
	F	F.AffyG*
	G	G.AffyN*

* Two distinct amplification methods are used with the Affymetrix exon array platforms: NuGEN (AffyN) and Genisphere (AffyG)

To assess data variation potentially introduced by sample processing on different days (batch bias), service providers were asked to process (amplify and label) one replicate sample per day over 3 different days. The amplification and labeling steps are typically seen as the point where the most variance is introduced in comparison with hybridization and scanning. Due to the long duration of the Allen Human Brain Atlas project, evaluation of the service providers in terms of between-batch data reproducibility was critical.

In this study, the seven provider/platform combinations were assessed and compared for reproducibility, ability to correct batch bias, sensitivity to differentially expressed (DE) genes over two anatomic regions, specificity of DE genes in replicates from the same anatomic region, and correlation with *in situ* hybridization (ISH) data in human brain for the cortical regions. These results are discussed in the Data Analysis section and summarized in Conclusions, below. The pre-processing steps for each dataset are given in Appendix A.

DATA ANALYSIS

Reproducibility

Reproducibility within each microarray platform was analyzed by determining pairwise correlations among three replicates of each brain region. Agilent, NimbleGen and Illumina platform data each had average correlation values ranging from 0.974-0.987, while Affymetrix platforms had correlations in the range 0.864-0.934. Affymetrix data was provided at the probe set level, which is summarized from the exon level, potentially accounting for higher variance and lower pair-wise correlations compared to other platforms.

Figure 1 shows pair-wise scatter plots of all possible two-sample comparisons from the A.Agilent dataset, illustrating high correlation among the replicates. Data for all platforms is summarized in Table 3.

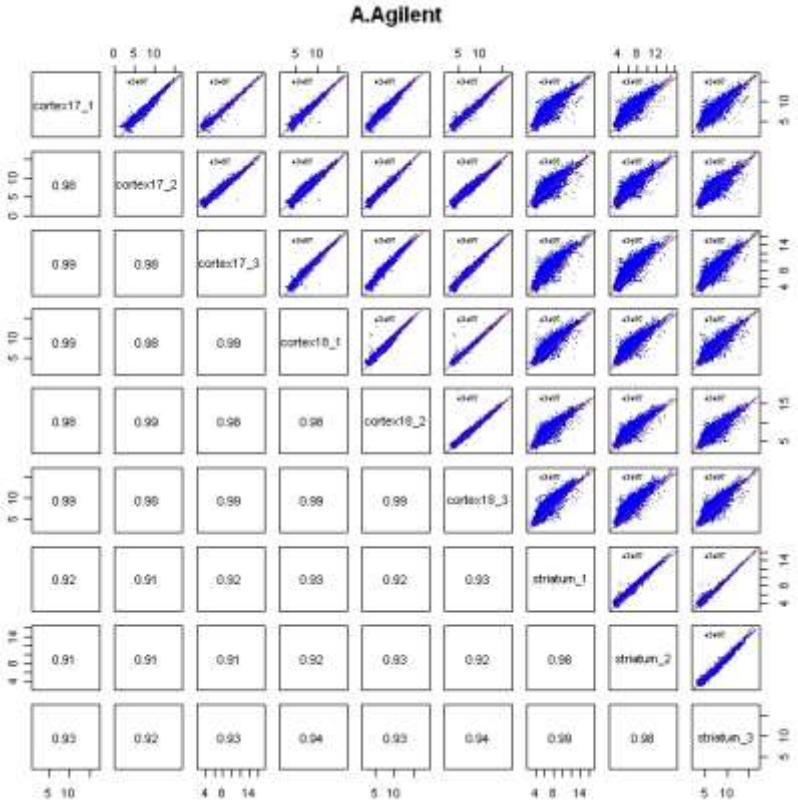


Figure 1. Pairwise correlation among replicates. Scatter plots showing high pairwise correlation between replicates for the A.Agilent dataset. Three replicates each of Cortex 17 (cortex17_1, cortex17_2, cortex17_3), Cortex 18 (cortex18_1, cortex18_2, cortex18_3) and striatum (striatum_1, striatum_2, striatum_3) samples are shown in the top, middle, and bottom three rows, respectively.

Table 3. Summary of average correlations among the three replicates for each brain area.

	A.Agilent	B.Nimble	C.Illumina	D.Illumina	E.AffyN	F.AffyG	G.AffyN
Cortex17	0.984	0.966	0.975	0.979	0.917	0.934	0.882
Cortex18	0.988	0.977	0.976	0.986	0.916	0.918	0.875
Striatum	0.983	0.978	0.971	0.971	0.880	0.950	0.834

Batch Bias Correction

For each brain region sampled, the three replicates were labeled and amplified on three different days by each service provider. We assessed batch bias, defined as variation in gene expression results due to processing of samples in different batches, as opposed to variation due to biological differences such as differences across anatomic regions or between individuals. Two-way ANOVA was applied to measure the effect of batch and tissue variance of gene expression signal. If a significant batch effect was detected, additional normalization algorithms were applied to correct this bias. The normalization methods applied are given in Table 4.

Table 4. **Pre-normalization and additional bias-correcting normalization.**

	Pre-normalized	Batch bias	Normalization applied
A.Agilent	No	Yes	Bioconductor *normalize.loess()*
B.Nimble	No	Yes	Bioconductor *normalize.loess()*
C.Illumina	BeadStudio	No	N/A*
D.Illumina	BeadStudio	Yes	Bioconductor *normalize.loess()*
E.AffyN	Expression Console RMA	Yes	Bioconductor *normalize.loess()*
F.AffyG	Expression Console RMA	Yes	Bioconductor *normalize.loess()*
G.AffyN	Expression Console RMA	Yes	Partek Batch Remover

***** The C.Illumina dataset was provided in pre-normalized format.

The ANOVA results are shown in Table 5 for A.Agilent and in Appendix B for the remaining provider/platform combinations. Table 5 presents the two-way ANOVA table for the A.Agilent platform before and after bias correcting normalization in the left panel and a boxplot of replicates 1, 2, and 3 of Cortex17, Cortex18, and striatum on the right. Batch biases can be seen by examining the trends of median gene expression values of each sample. As seen in the ANOVA results, significant batch effects in the dataset were eliminated by normalization processes while significant differences between tissue regions were maintained.

Table 5. **Batch bias correction of the A.Agilent dataset using Bioconductor normalization assessed by two-way ANOVA and boxplot comparison.**

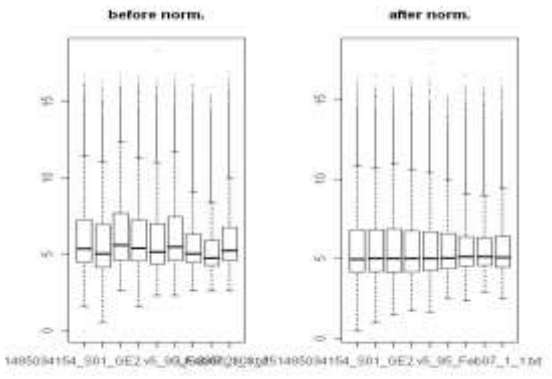
ANOVA2 (before normalization)

source	SS	df	F	Pr(p>F)
batch	9155.9	2	2228.5	0
tissue	21972.3	2	5347.8	0
interaction	8605.6	4	1047.3	0
error	760505.9			
total	800239.8			

ANOVA2 (after normalization)

source	SS	df	F	Pr(p>F)
batch	1.9	2	1.8	0.16
tissue	135.5	2	127.4	0
interaction	25.6	4	12.0	0
error	196889			
total	197052			

Boxplot before/after normalization



Two factors important in assessing effective improvement in data by batch normalization are the variances explained by batch and by tissue as a percentage of total variance. This is graphically displayed in Figure 2 for each of the provider/platform combinations. Each provider/platform is represented by an arrow for which the tail plots the coordinates of the pre-normalized batch and tissue percent variance and the head represents the post-normalized values. A movement toward lower batch variance while maintaining higher tissue variance is desirable. Because C.Illumina was pre-normalized, a single point was plotted for this dataset.

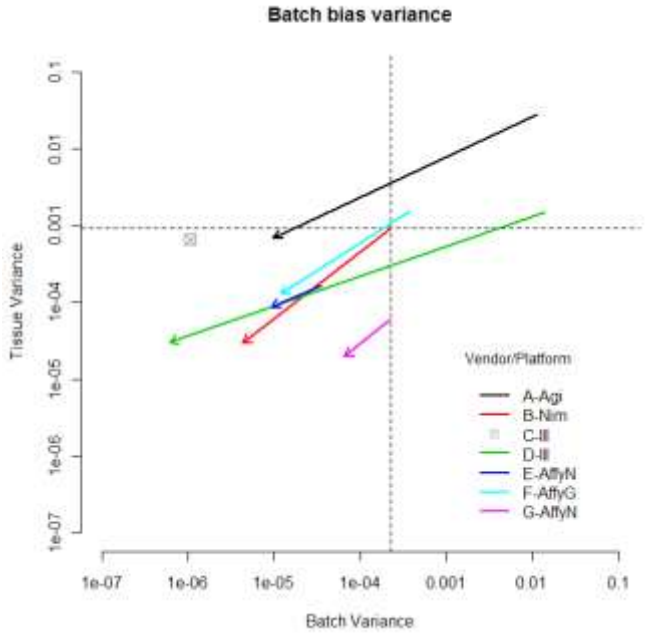


Figure 2. Effect of batch bias correction on percent variance explained by batch and tissue from two-way ANOVA. Each provider/platform is represented as an arrow that moves toward the lower left as the effect of normalization reduces both batch and tissue variance. The tails of the arrows are the pre-normalization values, and the heads represent the post-normalization values. Dotted lines indicate median pre-normalization scores. C.Illumina provides pre-normalized data and is represented by a single point.

To quantify the improvement in data by batch normalization, we computed the ratio of post-normalization to pre-normalization (percent) batch effect variances using ANOVA. A smaller value will indicate a reduction in the variance due to batch effects through the normalization. For A.Agilent from the ANOVA SS values in Table 5 this value is:

$$\textbf{var.ratio.batch} = 8.458e^{-4} = \frac{1.9/197052}{9155.9/800240}.$$

Maintaining a high final variance explained by tissue after the normalization is also important, and a measure of this variance was also computed. For A.Agilent this value is:

$$\textbf{var.tissue} = 6.876e^{-4} = \frac{135.5}{197052}.$$

To assess which platform responded 'best' to normalization, the ratio **var.tissue/var.ratio.batch** was computed. The higher this ratio, the better the correction is in terms of maximizing variance due to tissue (presumably biological) differences while simultaneously minimizing batch variance presumably due to procedural differences unrelated to biology. In other words, the quantity **var.tissue** should preferably be large while the **var.ratio.batch** should preferably small so that the resulting **var.tissue/var.ratio.batch** ratio is large. These results are presented in Table 6 and show that the A.Agilent and D.Illumina are most improved by batch normalization according to this metric. Other metrics gave similarly comparable results.

Table 6. Effect of batch bias correcting normalization.

	A.Agilent	B.Nimble	C.Illumina*	D.Illumina	E.AffyN	F.AffyG	G.AffyN
var.tissue/ var.ratio.batch	8.15e-01	1.53e-03	6.5e-04	6.93e-01	3.27e-04	4.06e-03	6.49e-05

***** C.Illumina's probe set level data was batch bias free and additional normalization was not required. For this case the denominator **var.ratio.batch** was taken to be 1.

Differential Gene Expression between Anatomic Regions

To compare gene expression results between the provider/platform combinations, we examined gene expression differences between different anatomic structures. The comparisons included cortical and subcortical regions with the expectation that a greater number of differentially expressed (DE) genes would be detected between subcortical and cortical regions as opposed to between two cortical regions. Comparisons between Cortex 17 and Cortex 18 indicated that the number of DE genes in each platform was proportional to the total number of genes in each platform. The total number of DE genes was also lower than the number of genes expected to be DE by chance (p-value*total number of genes).

We measured the sensitivity of expression detection by varying the p-value cutoff for DE gene identification on each platform and measured specificity by the occurrence of DE results within replicate data of the same structure. In particular, if a DE gene is reported between any of the replicates within a structure, this serves as an identifier for a false positive result. The data can then be represented as a receiver operating characteristic (ROC) curve for each platform and tissue combination and shown in Figure 3. In general, curves with larger area under curve (AUC) indicate better performance in maximizing true positives at a given false positive rate. The Agilent and NimbleGen platforms showed better overall performance in their ability to identify DE genes between Cortex 17/18 and Striatum, while the ROC curves for all platforms comparing Cortex 17 and Cortex 18, showed more DE counts occurring as false positives. All values are reported in Table 7.

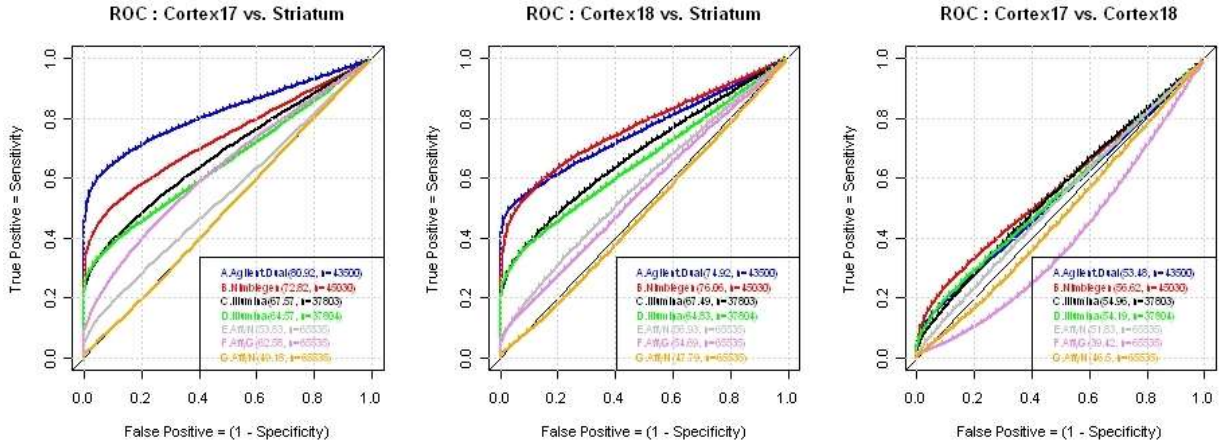


Figure 3. ROC curves for between-structure comparisons. ROC analysis among platforms with DE genes between two anatomical regions as true positive (sensitivity) and DE genes among replicates as false positive (specificity). A.Agilent and B.NimbleGen show better performance with higher true positive and lower false positives. (A.Agilent:blue, B.Nimble:red, C.Illumina:black, D.Illumina:green, E.AffyN:gray, F.AffyG:purple, G.AffyN:gold)

Table 7. ROC area under curve (AUC) by structure.

Provider/Platform	AUC (x100)*			
	Cortex 17	Cortex 18	Striatum	Mean across all sets
A.Agilent	80.91	74.91	53.47	69.76
B.Nimble	72.81	76.05	56.62	68.49
C.Illumina	67.56	67.49	54.96	63.34
D.Illumina	64.56	64.82	54.19	61.19
E.AffyN	53.82	56.92	51.83	54.19
F.AffyG	62.57	54.68	39.42	52.22
G.AffyN	49.18	47.79	46.50	47.82

* A larger value indicates more DE genes at a given false positive level on average.

Comparison of Microarray Results with *In Situ* Hybridization (ISH) Data

As part of the Human Cortex Study—originally a separate, stand-alone dataset and now integrated into the Allen Human Brain Atlas—the Allen Institute has generated ISH data on human visual cortex (Cortex 17 and 18) for 1,000 genes. This dataset provides a valuable framework for comparison with microarray expression results. To compare this ISH dataset with microarray data in Cortex 17 and Cortex 18, ISH gene expression data were quantified as in Lee et al. (2008). This technique essentially involves using an image processing algorithm to segment expressing cells and then calculating a normalized integrated optical density (IOD) across the tissue region. To normalize results, the area of the cortex section is estimated using the closest Nissl section to the ISH section. Following quantification, correlations between ISH data and microarray data from each visual cortical region were calculated (see Figure 4). The summary correlations for each provider/platform combination are given in Table 8. Consistent with previous findings (Lee et al., 2008), the correlations ranged from 0.478 to 0.578.

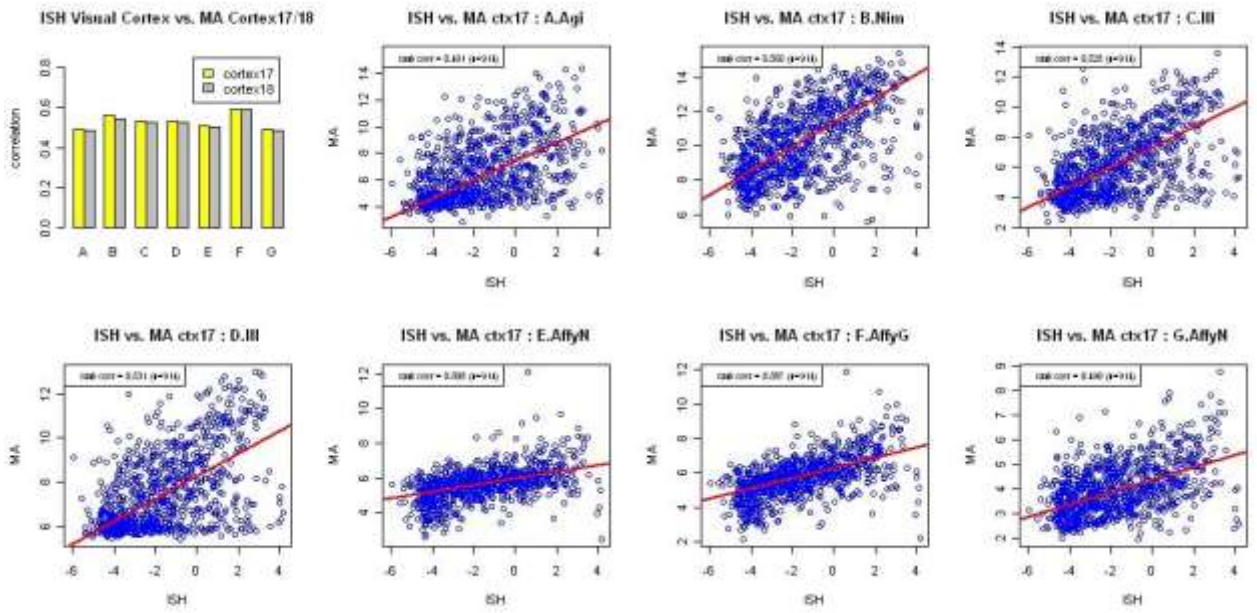


Figure 4. Correlation between microarray (MA) and *in situ* hybridization (ISH) data. The upper left panel is the summary of correlation between visual cortex ISH data and Cortex 17/Cortex 18 (yellow/grey) MA data. Scatter plots of the comparisons with Cortex 17 follow. The scatter plot comparisons between Cortex 18 MA and visual cortex ISH are similar to that between Cortex 17 MA and visual cortex ISH (data not shown).

Table 8. Correlation of ISH data with microarray data from genes differentially expressed in Cortex 17 and Cortex 18.

	A.Agilent	B.Nimble	C.Illumina	D.Illumina	E.AffyN	F.AffyG	G.AffyN
Correlation Cortex 17 /Cortex 18	0.491/0.485	0.56/0.545	0.528/0.524	0.531/0.526	0.506/0.505	0.587/0.588	0.49/0.485

CONCLUSIONS

We performed a series of analyses to measure reproducibility, elimination of batch bias, sensitivity and specificity to differentially expressed genes, correlation with ISH data for each provider/platform combination. The results are summarized in Table 9 with an average ranking score given. The top three provider/platform combinations were A.Agilent, B.Nimble and D.Illumina. Based on other considerations, including cost and customer relationship, site visits were made to sites A and B to examine the capabilities of processing throughput, experience with large scale studies, cost of arrays and service. In addition, these datasets were vetted and analyzed by independent microarray experts who provided recommendations.

As a result, the Agilent platform and provider A (i.e., A.Agilent) were selected for the microarray component of the Allen Human Brain Atlas. We acknowledge that although care was taken in choosing logical metrics, other combinations could be defined that could potentially lead to variation in the results.

Table 9. Results ranking for the comparative evaluation metrics for the microarray platform comparison.

Criteria	A.Agilent	B.Nimble	C.Illumina	D.Illumina	E.AffyN	F.AffyG	G.AffyN
Pairwise correlation	0.985 1	0.980 3	0.974 4	0.984 2	0.904 6	0.934 5	0.864 7
Batch bias correction	8.15e-01 1	1.53e-03 4	6.5e-04 5	6.93e-01 2	3.27e-04 6	4.06e-03 3	6.49e-05 7
Differential expression	69.76 1	68.49 2	63.34 3	61.19 4	54.19 5	52.22 6	47.82 7
Correlation with ISH	0.488 6	0.553 2	0.526 4	0.528 3	0.505 5	0.587 1	0.487 7
Average Ranking	**2.25**	**2.75**	**4**	**2.75**	**5.5**	**3.75**	**7**

ACKNOWLEDGEMENTS

We would like to thank all the service providers for processing these datasets and providing the data in a timely manner. We would also like to thank the array manufacturers for providing the arrays used in this pilot study. We would like to acknowledge and thank the extensive effort and support received from Affymetrix, Illumina, Roche NimbleGen, Expression Analysis, GeneLogic, Asuragen, Cogenics, and the Rosetta technical support group during the data analysis portion of this study. We are also grateful to Dr. Mark Reimers (Virginia Commonwealth University) and Yudong He (Rosetta Inpharmatics/Merck & Co.) for their expertise and support in analyzing these data and developing this white paper. Tissue for this study was generously provided by Drs. Joel Kleinman and Tom Hyde (NIMH); NICHD Brain and Tissue Bank (University of Maryland), and Drs. Marquis Vawter and Preston Cartagena (University of California, Irvine).

RELATED REFERENCES

Affymetrix (2005) Gene signal estimates from exon arrays v1.0. In: *Affymetrix White Papers*. http://www.affymetrix.com/support/technical/whitepapers.affx

Affymetrix (2007) Quality assessment of exon arrays v1.1. In: *Affymetrix White Papers*. http://www.affymetrix.com/support/technical/whitepapers.affx

Fujita A, Sato JR, de Oliveira Rodrigues L, Ferreira CE, Sogayar MC (2006) Evaluating different methods of microarray data normalization. *BMC Bioinformatics* 7:469.

Gentleman RC *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* 5:R80.

Kapur K, Xing Y., Ouyang Z, Wong WH (2007) Exon arrays provide accurate assessments of gene expression. *Genome Biology* 8(5):R82

Lee CK *et al.* (2008) Quantitative methods for genome-scale analysis of in situ hybridization and correlation with microarray data. *Genome Biology* 9(1):R23.

Patterson TA *et al.* (2006) Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. Nature Biotechnology 24(9):1140-50.

Robinson MD, Speed TP (2007) A comparison of Affymetrix gene expression arrays. *BMC Bioinformatics* 8:449.

APPENDIX A: Dataset Preparation

Microarray data were provided in various formats depending on the platform and provider. Each dataset was integrated and analyzed in R using Bioconductor, an open source software package produced by the Bioconductor project (www.bioconductor.org). The Rosetta Resolver system was considered for data integration and analysis but was not available at the time of these studies for the latest Affymetrix Exon ST1.0 arrays. Other specific noteworthy points in the dataset preparation are:

- Resolver's intensity text loader was used to upload NimbleGen (3x135k) data, which was then written out as a text file and integrated in R;
- For the Agilent 4x44 platform, the tab-delimited text data is read into R by the Bioconductor function 'read.Agilent()';
- The Agilent platform is dual channel;
- The Ambion FirstChoice Human Brain Reference RNA was used for the reference channel (Cy5) for the Agilent dataset;
- Agilent data was analyzed in single channel as well as in dual channel since all other microarray data is single channel;
- Affymetrix Exon Array data was provided at probe set level by Affymetrix Expression Console program;
- Illumina BeadStudio program gives the probe set level summary data for Illumina Human HT-12 BeadChip.

Table A1. The number of probe sets and the number of genes available in each data set.

Key	A.Agilent	B.Nimble	C.Illumina	D.Illumina	E.AffyN	F.AffyG	G.AffyN
Probe Set	43349	45030	48803	48803	65535	65535	65535
Gene Count	29907	23611	37804	37804	21980	21980	21980

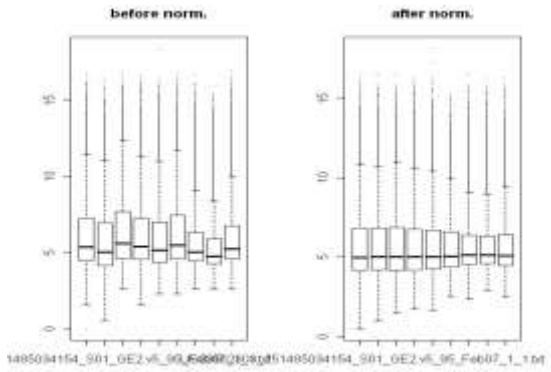
APPENDIX B: Batch Bias and Correction by Normalization

Two-way ANOVA before and after the normalization is shown in the left panel, and boxplots of data before and after the normalization in the right panel. C.Illumina probe set level data did not have batch bias and so no normalization is applied. Data for A.Agilent is shown in Table 5 in the main text, above.

B.Nimble

ANOVA2 before normalization				
source	**SS**	**df**	**F**	**Pr(p>F)**
batch	312.7	2	46.3	0
tissue	1263.7	2	137.2	0
interaction	3330.6	4	246.8	0
error	1367539.6			
total	1372446.6			
ANOVA2 after normalization				
source	**SS**	**df**	**F**	**Pr(p>F)**
batch	5.9	2	0.88	0.41
tissue	39.7	2	5.89	0.0028
interaction	24.5	4	1.82	0.12
error	1366857.9			
total	1366928.0			

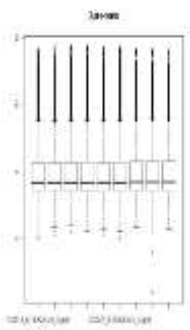
Boxplot before/after normalization



C.Illumina

ANOVA2 before normalization				
source	**SS**	**df**	**F**	**Pr(p>F)**
batch	0.7	2	0.06	0.94
tissue	426.6	2	37.1	0
interaction	16.3	4	0.7	0.58
error	655340.7			
total	655783.9			
ANOVA2 after normalization				
N/A				

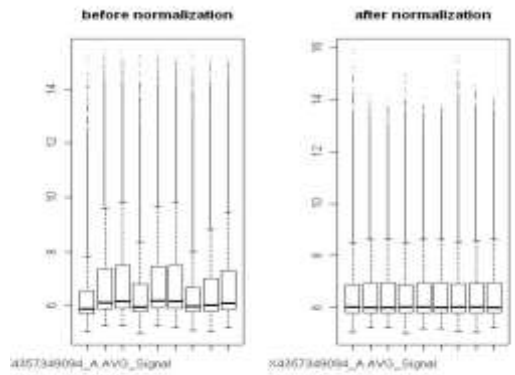
Boxplot before/after normalization



D.Illumina

ANOVA2 before normalization				
source	**SS**	**df**	**F**	**Pr(p>F)**
batch	11802	2	2448.0	0
tissue	1223.4	2	253.8	0
interaction	1331.7	4	138.1	0
error	820117.4			
total	834474.5			
ANOVA2 after normalization				
source	**SS**	**df**	**F**	**Pr(p>F)**
batch	0.4	2	0.11	0.89
tissue	19.6	2	5.13	0.01
interaction	2.5	4	0.33	0.86
error	651214.7			
total	631237.3			

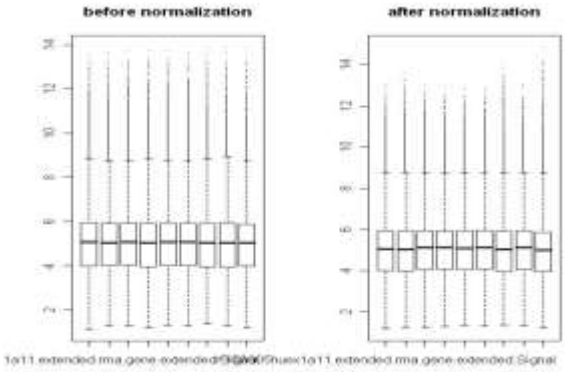
Boxplot before/after normalization



E.AffyN

ANOVA2 before normalization				
source	**SS**	**df**	**F**	**Pr(p>F)**
batch	47.5	2	10.3	0
tissue	222.4	2	48.3	0
interaction	316.8	4	34.4	0
error	1356577.2			
total	1357164.0			
ANOVA2 after normalization				
source	**SS**	**df**	**F**	**Pr(p>F)**
batch	12.2	2	2.7	0.06
tissue	114.0	2	25.6	0
interaction	356.8	4	40.1	0
error	1313125.1			
total	1313608.1			

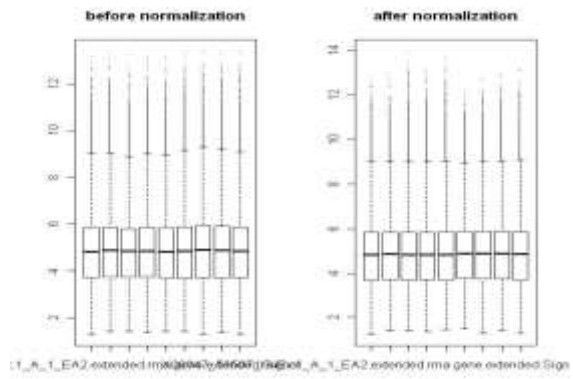
Boxplot before/after normalization



F.AffyG

ANOVA2 before normalization				
source	**SS**	**df**	**F**	**Pr(p>F)**
batch	54.5	2	27.2	0
tissue	219.7	2	109.9	0
interaction	284.4	4	71.1	0
error	1431941.0			
total	143450.0			
ANOVA2 after normalization				
source	**SS**	**df**	**F**	**Pr(p>F)**
batch	16.6	2	3.5	0.03
tissue	177.1	2	37.6	0
interaction	332.5	4	35.3	0
error	1388852.3			
total	1389378.6			

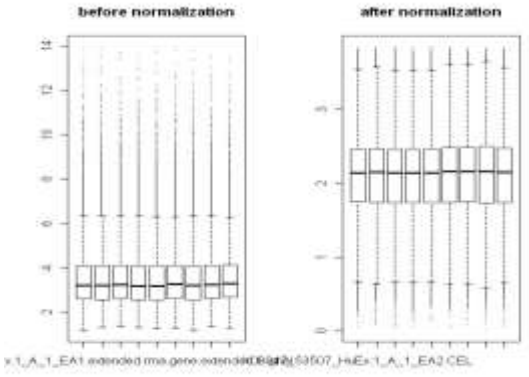
Boxplot before/after normalization



G.AffyN

ANOVA2 before normalization				
source	**SS**	**df**	**F**	**Pr(p>F)**
batch	171.1	2	64.6	0
tissue	46.3	2	17.5	0
interaction	38.0	4	7.2	0
error	780629.4			
total	780884.7			
ANOVA2 after normalization				
source	**SS**	**df**	**F**	**Pr(p>F)**
batch	52.3	2	19.5	0
tissue	15.5	2	5.8	0.003
interaction	39.2	4	7.3	0
error	791103.3			
total	791210.3			

Boxplot before/after normalization



APPENDIX C: Electropherograms

Agilent 2100 Bioanalyzer Pico chip electropherograms of the pooled RNA samples at 10ng/µl. The Thalamus samples are of poor quality and were excluded from the platform selection process.