Table 2
Quality control assessment

Brain	Criteria	Clinical quality		Tissue quality			RNA quality				Microarray chip quality			
		Age	AFS* 0/≥1	pH	PMI (hours)	Freezer time (years)	28S/18S	3′/5′ GAPDH	3′/5′ ACTB	RIN	% PC	SF	ACI	Type 1/2*
ACC	AFS = 0 (n = 75)	50.49 (13.87)	N/A	6.87 (0.20)	25.04 (8.48)	2.85 (3.08)	1.86 (0.36)	1.68 (0.72)	3.02 (1.97)	6.05 (1.08)	43.19 (4.87)	1.06 (0.93)	0.95 (0.04)	61/14
	AFS ≥ 1 (n = 23)	56.83 (10.57)	N/A	6.69 (0.31)	21.22 (8.67)	3.33 (4.15)	1.60 (0.31)	1.66 (0.64)	2.62 (1.06)	5.73 (1.81)	42.45 (6.33)	2.58 (2.48)	0.90 (0.08)	10/13
	p-Value	**0.02**	N/A	**0.01**	0.07	0.61	**0.001**	0.90	0.21	0.44	0.61	**0.01**	**0.002**	**0.0009**
CB	AFS = 0 (n = 75)	50.49 (13.87)	N/A	6.87 (0.20)	24.27 (8.28)	2.85 (3.08)	1.96 (0.45)	1.62 (0.77)	3.62 (3.28)	6.58 (1.66)	45.18 (3.89)	1.37 (2.10)	0.97 (0.04)	61/14
	AFS ≥ 1 (n = 16)	57.06 (7.65)	N/A	6.64 (0.33)	21.16 (8.80)	4.53 (4.48)	1.56 (0.39)	1.97 (0.71)	3.24 (1.35)	4.99 (1.64)	38.09 (7.83)	7.01 (5.95)	0.90 (0.07)	3/13
	p-Value	**0.01**	N/A	**0.02**	0.21	0.17	**0.00125**	0.09	0.45	**0.0001**	**0.0025**	**0.0017**	**0.01**	**<0.000005**
ACC	pH ≥ 6.6 (n = 81)	51.23 (13.67)	67/14	N/A	23.89 (8.22)	2.97 (3.26)	1.75 (0.37)	1.52 (0.53)	2.55 (1.40)	6.09 (1.25)	44.13 (4.26)	1.19 (1.20)	0.96 (0.03)	69/12
	pH < 6.6 (n = 17)	55.53 (11.71)	9/8	N/A	25.35 (10.60)	2.96 (3.82)	1.49 (0.41)	2.46 (0.88)	4.70 (2.40)	5.34 (1.56)	37.71 (6.20)	2.46 (2.54)	0.91 (0.06)	2/15
	p-Value	0.19	**0.02**	N/A	0.60	1.00	**0.00017**	**0.0005**	**0.0022**	0.12	**0.0006**	0.06	**0.01**	**<0.000005**
CB	pH ≥ 6.6 (n = 74)	50.76 (13.44)	66/8	N/A	23.81 (8.42)	3.19 (3.32)	1.92 (0.43)	1.58 (0.72)	3.34 (3.06)	6.73 (1.66)	44.81 (4.65)	2.03 (3.41)	0.97 (0.03)	62/12
	pH < 6.6 (n = 17)	55.53 (11.71)	9/8	N/A	23.37 (8.64)	2.96 (3.82)	1.73 (0.56)	2.11 (0.84)	4.49 (2.81)	6.03 (2.13)	40.14 (7.18)	3.80 (4.92)	0.91 (0.09)	2/15
	p-Value	0.15	**0.002**	N/A	0.85	0.82	0.21	**0.03**	0.15	0.27	**0.02**	0.17	**0.02**	**<0.000005**
ACC	28S/18S ≥ 1.5 (n = 73)	49.84 (13.62)	64/14	6.86 (0.22)	23.99 (7.92)	2.83 (3.31)	N/A	1.55 (0.57)	2.73 (1.78)	6.08 (1.35)	43.57 (5.10)	1.29 (1.53)	0.96 (0.03)	60/13
	28S/18S < 1.5 (n = 25)	58.24 (10.67)	11/9	6.72 (0.27)	24.59 (10.62)	3.35 (3.48)	N/A	2.07 (0.89)	3.48 (1.78)	5.56 (1.45)	41.42 (5.35)	1.76 (1.69)	0.93 (0.05)	11/14
	p-Value	**0.01**	0.02	**0.00002**	0.51	0.93	N/A	**0.0001**	**0.002**	0.11	**0.004**	**0.02**	**0.00003**	**0.0005**
CB	28S/18S ≥ 1.5 (n = 76)	51.33 (13.50)	66/10	6.86 (0.24)	24.17 (8.10)	3.03 (3.28)	N/A	1.50 (0.54)	3.06 (2.53)	6.91 (1.45)	45.00 (4.55)	1.80 (2.89)	0.97 (0.03)	60/16
	28S/18S < 1.5 (n = 15)	53.27 (11.84)	9/6	6.63 (0.19)	21.49 (9.87)	3.72 (4.04)	N/A	2.58 (1.10)	6.07 (4.06)	4.49 (1.44)	38.53 (6.69)	5.20 (6.02)	0.91 (0.08)	4/11
	p-Value	0.58	**0.02**	**0.0002**	0.34	0.55	N/A	**0.002**	**0.01**	**0.0006**	**0.002**	**0.05**	**0.02**	**0.0001**

Each quality control measure is based on the criteria listed below for both anterior cingulate (ACC) and cerebellum (CB). The averages (±standard deviation) are shown for each criterion and significant differences between the groups were determined by t-test. The resulting significant p-values are shown in bold. The abbreviations are the same as used in Table 1. N/A indicates that due to *a priori* selection of quality by AFS, pH, and RNA these produced highly significant differences therefore the p-value was not shown.

* p-Values for AFS and Type 1/2 were calculated using Fisher's exact test.

Table 3
Differential expression of genes in the mitochondrial pathway

Brain region	Pathway search	Pathway	Subjects with AFS = 0 compared to subjects with AFS > 0	Subjects with pH ≥ 6.6 compared to subjects with pH < 6.6	Subjects with 28S/18S ≥ 1.5 compared to subjects with 28S/18S < 1.5
ACC	KEGG	Oxidative phosphorylation	4.85E−02	NS*	1.00E−02
	Gene ontology	Mitochondria	NS	4.30E−02	1.30E−02
CB	KEGG	Oxidative phosphorylation	4.00E−03	2.70E−02	NS
	Gene ontology	Mitochondria	4.30E−02	6.50E−03	NS

Differential gene expression was determined in two brain regions and used for ranking in an over-representation analysis (GSEA gene set enrichment analysis, Mootha et al., 2003a,b). The over-representation analysis used subject permutation to correct p-values for multiple test comparisons.
* NS: not significant.

3.2. Tissue quality for gene expression profiling

Some brain banks have collected brains ~10–30 years older than the current sample (Breese et al., 1997; Hakak et al., 2001; Konradi et al., 2004). Because of the high average age of these cases, investigators routinely utilized microscopic neuropathological examinations to rule out age-related neurodegeneration in cases and controls. In the present study, subjects were relatively young at the time of death (Table 1) and were excluded if they showed clinical signs of dementia prior to death. Therefore a neuroanatomist performed gross neuropathological examinations of all subjects after the brain had been sliced into 1 cm coronal slabs to rule out hemorrhage, gross infarcts, and lesions. Any subjects with gross neuropathology were excluded.

The pH and AFS measures were significantly correlated ($r = -0.44$, $p < 0.01$, d.f. = 89) therefore the residual variation between pH and AFS must be accounted for by other unmeasured variables. We found that subjects with zero agonal factors, which included only short agonal duration of minutes and no prolonged illnesses had a range of pH from 6.3 to 7.26 (Fig. 2). The average pH was 6.83 ($n = 98$) and the median was 6.87 (Fig. 2).

In order to assess the statistical relationships of pH on RNA integrity and microarray quality, the subjects were arbitrarily divided into high pH (≥6.6) and low pH (<6.6). In a statistical sense, pH significantly affected measures of RNA quality and microarray quality for both brain regions by t-test comparisons of the high and low pH groups (Fig. 1; Table 2).

The PMI was not statistically associated with pH or RNA integrity (Fig. 1; Table 2). This was not surprising because the average PMI was 24.1 ± 8.6 h and during the collection process any sample with a PMI over 48 h would not be collected. Furthermore, a majority of the postmortem cases received from the coroner's office were placed in the refrigerator within hours after death pending autopsy. The average ±standard deviation of the postmortem time to refrigeration was 6.5 ± 4.2 h. The average freezer time ±standard deviation for the 98 ACC samples was 2.9 ± 3.3 years and there were no significant relationships between freezer time and RNA quality (Tables 2 and 4). However, an in vitro perturbation of RNA to assess one freeze–thaw cycle that could occur during prolonged freezer storage conditions of small RNA aliquots modestly increased RNA degradation (next section).

The mitochondrial gene pathways were significantly different between the groups with high pH (≥6.6) and low pH (<6.6) using an over-representation analysis of GO terms for cellular, biological, and molecular components (Vawter et al., 2006). The mitochondrial pathway was significantly over-represented between the high pH and low pH groups for the KEGG analysis and the GO mitochondrion cellular component term (Table 3). These results were similar to the comparisons of the AFS groups (see Section 3.1 above), as pH and the AFS (sum of the number of agonal factors) were significantly correlated ($r = -0.44$, $p < 0.01$, Fig. 2).

Table 4
The top 5% of dysregulated genes in ACC comparing different groups based on AFS, pH and 28S/18S were compared to genes listed in Auer et al. (2003) that were found to be dysregulated due to RNA degradation

Gene symbol	Auer et al. up/down	AFS = 0 vs. AFS = +1	pH ≥ 6.6 vs. pH < 6.6	28S/18S ≥ 1.5 vs. 28S/18S < 1.5
GNAS	Down		X	X
CHD1	Down	X		
GDI2	Down		X	
ATP5A1	Down		X	X
YWHAZ	Down		X	X
PCBP1	Down		X	X
TEGT	Down	X		
B2M	Down	X		X
NME1	Up		X	X
ATP5B	Up		X	X

There were 31 genes listed in Auer et al. (2003) changed by RNA degradation. Comparisons of the top 5% of dysregulated genes between our ACC samples revealed 194 genes in common between AFS and pH, 202 genes in common between AFS and 28S/18S and 417 genes in common between pH and 28S/18S. There were 146 genes in common between the three groups.

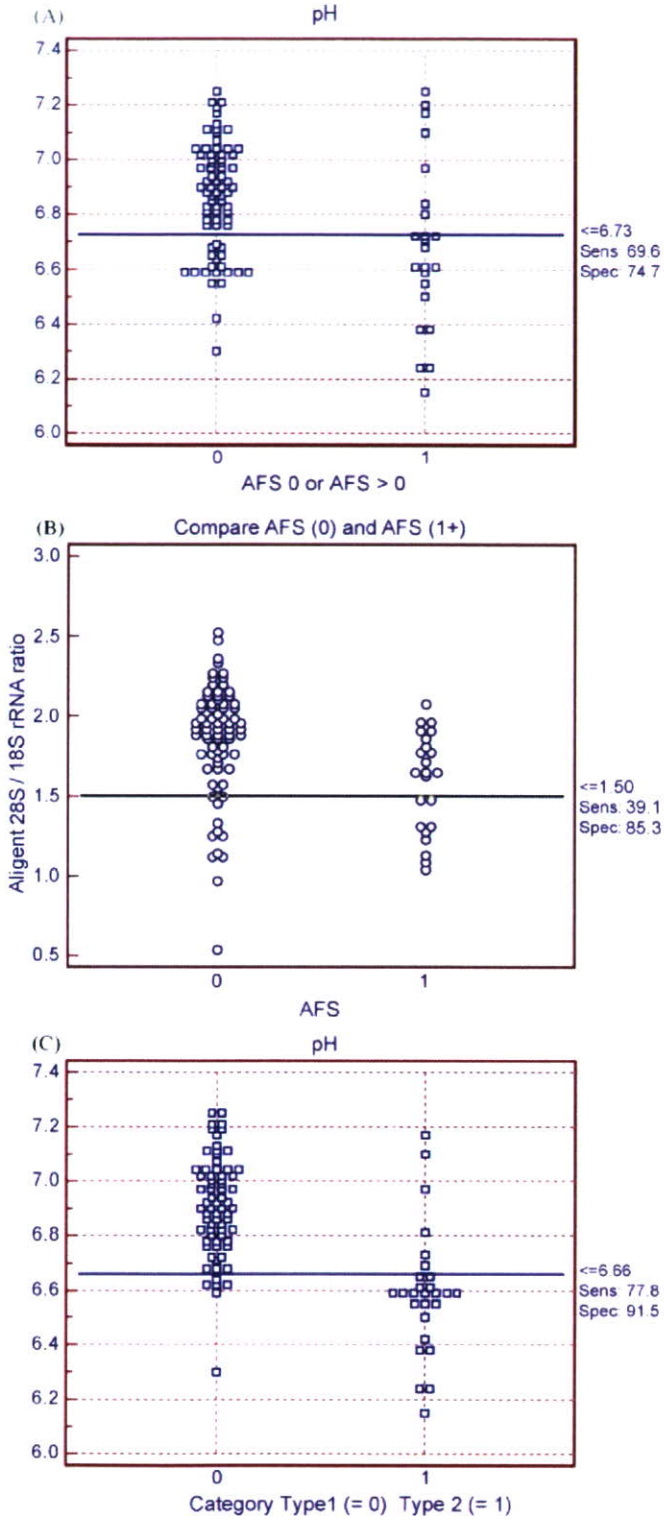

Fig. 2. (A) Sensitivity of pH and (B) RNA to AFS 0 and AFS > 0. The majority of 98 cases show a pH above an arbitrary cut point of 6.73 and RNA quality above a cut-point of 1.4, however there are outliers within AFS = 0 and AFS 1 categories. The combined use of three quality measures of pH, RNA, and AFS results in less differences in gene expression as a result of agonal-pH effects (Vawter et al., 2006). (C) Type '1' and '2' microarray outcome strongly depends on pH (Li et al., 2004). (B) Histogram of RNA quality (28S/18S) by AFS. (C) Histogram of pH by Type '1' and Type '2' samples. The type '1' and '2' refer to groups formed in hierarchical clustering that were associated with pH differences between clusters.

3.3. RNA quality

We induced RNA degradation in vitro by freeze–thawing RNA extracts from brain to simulate what might be observed during prolonged freezer storage, as aliquots in small volumes may freeze–thaw on occasion especially when moving samples in and out of the freezer. RNA samples that were fresh frozen, thawed once at room temperature and then subjected to an additional freeze–thaw cycle showed no differences in the integrity of total RNA determined by quantification of 28S/18S ribosomal RNA ratios on the Agilent 2100 Bioanalyzer. However, the Agilent RNA integrity number (RIN) showed significantly more degradation in the twice thawed samples compared to the thawed once samples ($p = 0.0018$). These results support the ability of the RIN algorithm to robustly detect mild degradation differences between samples compared to the 28S/18S ratio which did not detect any difference.

Total RNA 28S/18S ratios from a set of samples were measured on both the Agilent Bioanalyzer and on a denaturing agarose gel. The paired t-test between the methods was significant ($p = 0.03$) and the Agilent 28S/18S readings were consistently higher than the agarose gel 28S/18S ratio, as previously observed (Boris Sokolov, personal communication). Thus, setting a low threshold for Agilent quality translated to a lower quality measurement by conventional formaldehyde gel electrophoresis.

To measure the effect of RNA degradation on other quality parameters, the samples were divided into high (≥ 1.5) and low (<1.5) 28S/18S total RNA based upon Agilent measures. The median length of cRNA was significantly decreased in the low 28S/18S group ($p = 0.03$) compared to the high 28S/18S group confirming that shorter transcripts, as shown by cRNA length, are indeed made from degraded total RNA. Additionally, the samples with pH < 6.6 displayed a shorter size of the cRNA in vitro transcript compared to the higher pH group >6.6 ($p = 0.006$). A poor quality total RNA was associated with shorter cRNA transcripts.

We found significant correlations between the 28S/18S ratio and 3′/5′ GAPDH ratio ($r = -0.64$, $p < 0.01$, d.f. = 89) and between the 28S/18S and 3′/5′ ACTB ($r = -0.48$, $p < 0.01$, d.f. = 89). As expected, the negative correlations indicated higher 3′/5′ signal ratios, which was associated with poorer starting RNA quality. GAPDH 3′/5′ ratio presents a view of short transcript degradation while the ACTB 3′/5′ ratio gives a view of longer transcript degradation in postmortem brain. Hence, this was the reason that samples often showed a large disparity between the two ratios.

We employed an arbitrary criterion for 28S/18S Agilent measure (≥ 1.5) which produced a significant difference in the four quality measures in ACC and in CB (Fig. 2; Table 2). There was a technical problem in measuring the RIN in the ACC, so not all samples were measured. The technical issue might have biased our measures towards a lower RIN value for ACC compared to the CB samples leading to this regional variability. However, the 28S/18S measure was very robust and the mitochondrial pathway was significantly different between the 28S/18S groups in both the KEGG pathway and GO terms for cellular component

in ACC. The significance was also observed in CB for the KEGG pathway (Table 3).

Another question we addressed was if a consistent set of labile genes in degraded samples existed which could be used to survey the degradation and adjust for degradation effects using the labile gene set. The top 5% of genes dysregulated by the AFS, pH, and 28S/18S comparisons were matched to the Auer et al. (2003) gene list. There were 9 genes in common between the 31 genes found by Auer et al. (2003) and the present study indicating that certain labile transcripts are dysregulated when the postmortem brain sample was degraded (Table 4). Furthermore, in the present study there were 146 altered genes in common between all three group comparisons of AFS, pH, and 28S/18S.

3.4. Microarray chip quality

The results presented thus far, have largely focused on the impacts and inter-relatedness of clinical, tissue, and RNA quality. We have also examined pathway analysis for mitochondrial-related transcripts, and will later address the issue of variability of the quality factors across the entire transcriptome. It can be summarized that using only one quality indicator did not account for the total picture of sample variability and selection because each quality indicator was partially and not perfectly correlated (see Table 5 for correlation matrix of quality indicators; Fig. 2).

After the samples were completely processed the data from the microarray chip provided a set of indicators to measure overall chip quality and derivative measures of RNA degradation. The Affymetrix MAS 5.1 software was used to determine whether each transcript was reliably detected using a percent present call (%PC) and a scaling factor (SF), which adjusted the average signal intensity to a preset constant. Microarray chip quality was evaluated in the present study at cut points for the other three quality measures (AFS = 0, pH ≥ 6.6 and 28S/18S ≥ 1.5). Two microarray quality indicators, %PC and SF, were significantly different in both ACC and CB for RNA quality and for pH (tissue quality) group comparisons (Fig. 2, Table 2). Thus, differences in mRNA quality were related to significant changes in microarray %PC and SF.

ACI was significantly different between all three groups (AFS, pH, and 28S/18S shown in Fig. 2 and Table 2) for both brain regions. Employing these same cut points for AFS, pH and 28S/18S revealed significant differences (p-values ranged from 9.6E−09 to 5.2E−04) between the 'Type 1' and 'Type 2' groups for the current samples (Table 2; Fig. 2). The ROC plots and data table showed similar findings that pH and RNA quality were equally related to 'Type 1' and 'Type 2' membership (illustrated in Figs. 2 and 3; Table 6).

We tested more global gene expression parameters in two further methods. First, the AffyRNAdeg function modeled the extent of RNA degradation and thus could possibly control for this effect across samples at the probe level. The slope generated by AffyRNAdeg, which putatively measured the severity of degradation, was found to be significantly correlated with %PC ($r = -0.36$, $p = 0.0005$), 3′/5′ GAPDH ($r = 0.47$, $p < 0.0001$),

Table 5
Correlations of quality control measures. Each quality control measure is organized by the four categories (clinical, tissue, RNA, microarray) and then correlated for cerebellum data

	Clinical quality		Tissue quality			RNA quality					Microarray quality			
	Age	AFS	pH	PMI	Freezer time	28S/18S	3′/5′ GAPDH	3′/5′ ACTB	RIN	cRNA (~nts)	PC	SF	ACI	Type 1/2
Age	1.00													
AFS	0.19	1.00												
pH	−0.20	**−0.44***	1.00											
PMI	−0.05	−0.13	0.12	1.00										
Freezer time	−0.19	0.17	0.00	**−0.44**	1.00									
28S/18S	−0.04	−0.25	**0.33**	0.11	−0.18	1.00								
3′/5′ GAPDH	0.08	0.21	**−0.35**	0.13	−0.01	**−0.64**	1.00							
3′/5′ ACTB	−0.02	−0.02	−0.20	0.21	−0.18	**−0.48**	**0.86**	1.00						
RIN	−0.02	**−0.52**	**0.31**	0.25	**−0.37**	**0.68**	**−0.63**	**−0.37**	1.00					
cRNA (~nts)	−0.15	−0.21	0.22	−0.19	**0.34**	0.14	**−0.42**	**−0.34**	**0.28**	1.00				
PC	−0.04	**−0.46**	**0.38**	0.17	**−0.40**	**0.53**	**−0.47**	−0.17	**0.63**	0.21	1.00			
SF	0.01	**0.52**	−0.25	**−0.34**	**0.45**	**−0.41**	**0.27**	0.02	**−0.70**	−0.22	**−0.82**	1.00		
ACI	−0.06	**−0.53**	**0.46**	−0.03	−0.10	**0.58**	**−0.70**	**−0.49**	**0.63**	**0.39**	**0.65**	**−0.41**	1.00	
Type 1/2	0.17	**0.49**	**−0.61**	0.16	0.01	**−0.42**	**0.47**	0.24	**−0.34**	**−0.33**	**−0.48**	**0.30**	**−0.59**	1.00

The significant correlations are in bold ($p < 0.01$). Abbreviations are the same as in Table 1.
* Bold correlations were significant ($r > |0.267|$, $p < 0.01$, d.f. = 89).

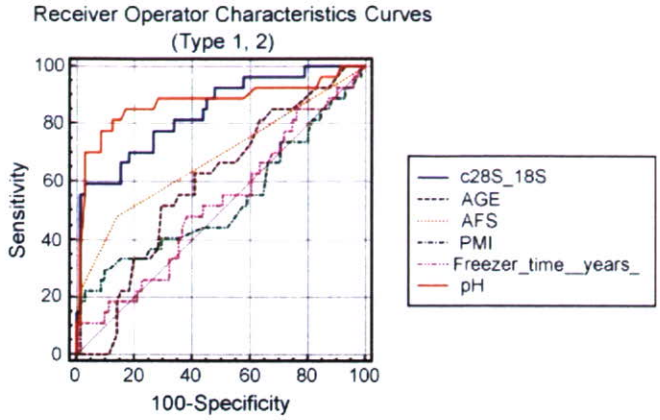

Fig. 3. The diagnostic performance of test variables, or the ability of a variable to discriminate microarray outcome (Type '1' vs. Type '2') is evaluated using receiver operating characteristic (ROC) curve analysis. The ROC curves are shown for six quality variables that relate differentially to microarray outcome. The area under each curve shows a relationship to overall prediction of Type '1' and '2' microarray outcome. The comparison of ROC curves for tissue quality (pH, freezer time, PMI), clinical quality (AFS, age), or RNA quality (28S/18S) shows that tissue quality pH measure is strongly related to Type 1 and Type 2 microarray outcome (Li et al., 2004). Table 6 shows the ROC values for each variable.

3′/5′ ACTB ($r = 0.76$, $p < 0.0001$), and cRNA length ($r = -0.35$, $p = 0.0007$). The AffyRNAdeg slope variable was used as a covariate in PLMfit (Bolstad, 2004) and this reduced the residual variation between duplicate RNA samples run on chips at different laboratories (data not shown).

The second method of testing global gene expression relationships to the four groups of quality covariables was to enter each variable as either a main effect if categorical such as AFS or 'Type 1, 2' or as a covariate if continuous. The F-ratios from the ANCOVA for all transcripts were summarized by average F-ratio for each variable. Of interest, the slope from the AffyRNAdeg showed the highest average F-ratio compared to all other covariates (Fig. 4). The rank of AffyRNAdeg persisted regardless of the type of data normalization used (quantile or grand median centering) or the effect of entering or removing other covariates.

Table 6
The comparison of receiver operator characteristic (ROC) curves for microarray Type 1 or Type 2 outcome (Li et al., 2004)

Quality variable	Area under ROC (maximum area = 1)	Standard error	95% CI
pH	0.874	0.035	0.792–0.933
28S/18S RNA ratio	0.841	0.040	0.754–0.907
AFS*	0.683	0.063	0.581–0.773
Age*	0.597	0.066	0.493–0.694
Freezer time*	0.523	0.066	0.420–0.625
PMI*	0.532	0.066	0.429–0.634

Tissue quality (pH) and RNA quality (28S/18S) strongly associated with accurate detection of Type 1 and Type 2 microarray outcome. The ROC plot for these variables is shown in Fig. 3.

* The area under the ROC column was significantly increased for RNA quality and pH compared to all other variables except the pH.

Table 7
Example of a strategy that used all three sample cut-points and eliminated the poorest quality samples and chips

Brain	Criteria	Clinical quality		Tissue quality		RNA quality			Microarray Chip quality			
		Age	AFS 0/≥1	pH	PMI (hours)	Freezer time (years)	28S/18S	3′/5′ GAPDH	3′/5′ ACTB	RIN**	%PC	SF	ACI	Type 1/2***
ACC	Include ($n = 60$)	49.43 (13.76)	N/A	6.93 (0.16)	22.96 (7.79)	3.35 (3.36)	1.98 (0.22)	1.46 (0.37)	2.49 (1.42)	5.94 (0.93)	44.00 (4.23)	1.07 (1.03)	0.97 (0.02)	56/4
	Exclude ($n = 38$)	56.26 (12.05)	N/A	6.66 (0.26)	25.52 (9.19)	2.51 (3.41)	1.51 (0.36)	2.03 (0.93)	3.60 (2.13)	5.67 (1.58)	41.47 (6.25)	1.97 (2.08)	0.92 (0.05)	12/24
	p-value	**0.01**	N/A	N/A	0.16	0.24	N/A	**0.00074**	**0.0006**	0.35	**0.03**	**0.02**	**<0.000005**	**<0.000005**
CB	Include ($n = 59$)	50.20 (13.81)	N/A	6.94 (0.16)	23.44 (7.99)	3.11 (3.24)	2.06 (0.34)	1.42 (0.48)	3.02 (2.74)	7.11 (1.31)	45.96 (3.08)	1.66 (3.10)	0.98 (0.02)	57/2
	Exclude ($n = 32$)	54.31 (11.75)	N/A	6.62 (0.25)	24.25 (9.26)	3.22 (3.73)	1.57 (0.50)	2.16 (0.97)	4.54 (3.33)	5.72 (1.81)	40.20 (6.86)	4.24 (5.28)	0.92 (0.07)	7/25
	p-value	0.14	N/A	N/A	0.68	0.89	N/A	**0.0002**	**0.03**	**0.002**	**0.00006**	**0.02**	**0.0005**	**<0.000005**

The criteria applied for inclusion were: AFS = 0 AND pH ≥ 6.6 AND 28S/18S ≥ 1.5. The samples that passed all three measures were placed into the 'included' group and the remaining were placed in the group 'excluded.' The two groups were compared by t-test separately in two brain regions. The significant p-values are in bold. The abbreviations are the same as Table 1. The microarray chip quality was improved in the included compared to the excluded samples. N/A: The *a priori* selection of quality by AFS, pH, and RNA produced highly significant differences in these three measures; therefore the p-value was not shown.

** RIN calculation for ACC only included part of the sample due to technical difficulties in the Agilent run.

*** Type 1/2 p-values were calculated by Fisher's exact test.

Table 8
An example of applying four quality parameters to the results of anterior cingulate for 98 subjects

	Clinical cut-off (AFS = 0)	RNA cut-off (28S/18S ≥ 1.5)	Tissue cut-off (pH ≥ 6.6)	Chip type cut-off (Type '1')
Number of subjects above cut-off	75	64	60	57
Number of subjects below cut-off	23	14	12	9

The methodology shows that 58% of the initial subject pool will pass all four parameters. Other methods to adjust for the effects of these quality and other postmortem variables on gene expression can be tested in post hoc analysis such as ANCOVA.

3.5. *Quality control parameters application in anterior cingulate cortex*

We presented evidence that each of the four quality control indicators (clinical, tissue, RNA and microarray quality) were significantly correlated to one another (Table 5), but were not perfectly correlated. A multiple covariable analysis of all transcripts showed that a post hoc measure of putative RNA degradation (AffyRNAdeg) accounted for a large proportion of variance in transcript expression.

All three quality selection parameters were used simultaneously in the current data set using cut points of AFS = 0, pH ≥ 6.6, and 28S/18S ≥ 1.5 as threshold criteria for sample selection as described. The "included" set (AFS = 0, pH ≥ 6.6 and 28S/18S ≥ 1.5) and the "excluded" (AFS ≥ 1, pH < 6.6 and 28S/18S < 1.5) set were compared. When the poor quality samples were "excluded" each microarray quality control indicator was significantly improved in the "included" group (Table 7). Notably, the microarray quality indicators (all post hoc) were significantly improved in eight measures performed in two brain regions (Table 7). Additionally, the mitochondrial pathway was significantly over-expressed between the groups. The thresholds appeared to work very well for this data set in two brain regions in selecting high quality microarray data. The "included" set shows that the cut points improved the four microarray chip quality indicators (%PC, SF, ACI, and Type 1/2) (Table 7), although it is noted that other methods for determining outlier samples could be equally applied at this point. Other methods might include principal component analysis to determine outliers based upon the entire gene expression profile, or performing an ANCOVA with multiple covariates.

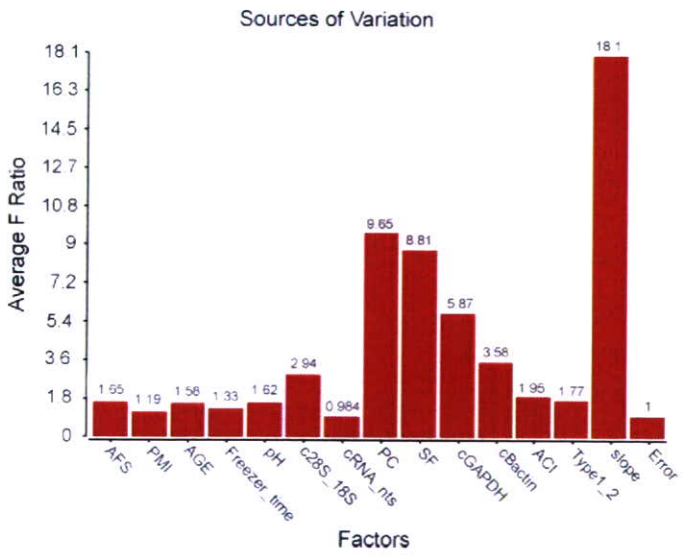

Fig. 4. The sources of variation in an ANCOVA of multiple covariates shows that the RNA degradation accounts for the highest average effect (F-ratio) across the entire transcriptome in ACC measured on an Affymetrix U133A chip. The AffyRNAdeg program (Cope, 2005) provides a slope, which is a microarray chip based indicator of the decline in signal across a transcript, thus a putative index of RNA degradation. The abbreviations for each variable are described in Table 1.

Some regional variability in the outcome of the quality parameters was noted. As an example, the RIN values for the high quality 'included' groups in cerebellum and ACC were not comparable. This was likely due to lower starting RIN values in the ACC the technical result of the revised Agilent software reading of the electropherograms. However correlations were calculated between ACC and cerebellum regions for both RIN ($r = 0.42, p < 0.01$) and SF ($r = 0.85, p < 0.0001$), thus indicating that across regions there was agreement of these values within subjects.

When we employed the four quality cut-offs to the present results the number of samples removed at each stage (Table 8) significantly reduced the study from 98 to 57 samples. This reduction in sample size is only one strategy that we have adopted, and may not be practical when working with degraded samples from fixed tissues, or for certain research questions that require use of tissues obtained under less than ideal conditions.

4. Conclusions

We systematically evaluated four quality indicators in one study with reference to gene expression results in postmortem brain from two regions. When samples in our study were not balanced well or matched for these quality indicators, the results of pathway analysis identified a brain disorders pathophysiology of mitochondria dysfunction. This pathophysiology is correlated to sample groups that have RNA, clinical, and tissue quality differences. Researchers have implicated mitochondrial-related pathways as a cause in schizophrenia (Prabakaran et al., 2004; Altar et al., 2005) and bipolar disorder studies (Konradi et al., 2004). A slight imbalance in group composition based upon an agonal-pH difference will affect this functional pathway and its related pathways such as apoptosis, proteasome, and chaperone functions (Vawter et al., 2006).

The impact of each quality covariable on gene expression may not be linear across wide ranges. As a first approximation researchers have found it useful to examine the readily accessible measures of clinical, tissue and RNA quality prior to microarray analysis and have used matched pairs of subjects or matched groups on these measures. Across a wide

range there was a potentially large non-linear variation observed with matching strategies. Investigators might adopt an approach of using all samples and making adjustments with regression models, however this would presume that adequate models exist.

RNA quality found to have a large statistical impact on the amount of variation across all transcripts. The R-script AffyRNAdeg (Cope, 2005) can be used post hoc to assess RNA quality to measure the slope across all transcripts in the 3′/5′ direction instead of only using the beta-actin or GAPDH transcripts. Other chip measures such as present calls and scaling factor accounted for a large proportion of variance across transcripts. Most investigators agree that severely degraded samples will not provide useful data and the findings are compromised when total RNA is degraded (Schoor et al., 2003; Auer et al., 2003; Buesa et al., 2004; Tomita et al., 2004; Lipska et al., 2006). However, modest degradation of samples such as observed in postmortem brain collections was addressed in the present paper to attempt to describe different strategies to select samples with minimal RNA degradation and with high clinical and tissue quality. It was shown that by removing samples with agonal factors, low pH and degraded RNA higher quality microarray results, based on post hoc measures, will be observed. Consistent recommendations to use the highest quality RNA for gene expression measurements (Bahn et al., 2001; Auer et al., 2003; Schoor et al., 2003; Buesa et al., 2004; Mexal et al., 2006; Lipska et al., 2006) have led one group of investigators to conclude that "The strongest predictor of gene expression was total RNA quality"(Lipska et al., 2006). Our results are consistent with these reports.

Future microarray knowledge will include how different types of transcripts are affected by postmortem and premortem variables as well as the set of transcripts not affected by these variables. Postmortem brain studies generally utilize one or more of the criteria reviewed for study design and statistical analysis. By adopting only one indicator to accept or reject samples at a certain threshold, an investigator may accept marginal samples. However, in post hoc analysis the impact of these parameters can be determined. We have used multiple criteria to form cohorts of postmortem samples based upon these *a priori* quality parameters of clinical, tissue, and RNA quality indicators and post hoc microarray indicators. We have also used these same parameters as covariables. For variables with a simple linear relationship these approaches are satisfactory, however, we have not tested non-linear models which may address a larger proportion of variance than the simple linear models. After forming a cohort with cut-off criteria this essentially narrows the range that strong effects can operate, so that the potential for case–control effects can emerge. As an example, we have used a set of criteria to reduce the range of strong effects with the following criteria: AFS = 0; the samples must have both medical records and next of kin interview information; the sample pH is minimally in the range of 6.4–6.6; and the RNA integrity measured by Agilent must have a 28S/18S ratio greater than 1.4–1.5 while at the same time taking the RIN value into consideration. We finally use post hoc indicators such as the AffyRNAdeg slope, percent present, scaling factor, ACI, and hierarchical clustering approaches to find outlier chips. These steps taken together will minimize the strong effects of these covariates, but other methods can be utilized to study the same effects and to assess the final quality of the chip.

The evaluation of these four quality categories aids in the study design, characterization and assessment of our samples, analysis, and interpretation of results. Not all transcripts are affected to the same degree by these variables as shown in our ANCOVA results. Thus, meaningful data can be derived if post hoc analysis has ruled out confounding effects on specific transcripts. Post hoc use of these quality covariables will help to either strengthen or weaken a candidate gene depending on the impact of the key variables we describe.

The four quality indicators are broadly related and therefore using a certain combination of these factors improves the quality of the data set, but might never truly separate the low, moderate, and high quality samples completely. By having access to alternative parameters to assess the quality of both the sample and the microarray data, we present an investigator with covariates for study design, selection criteria, or parameters for matching strategies depending on the nature of the study. These quality parameters may assist future investigators for meta-analysis of postmortem brain gene expression studies, such as those that can be conducted with the gene expression arrays deposited at the Gene Expression Omnibus.

Acknowledgements

We acknowledge Jacque Berndt and the investigators and medical examiners at the Orange County Coroners Office for procurement of brain tissue. We also appreciate the technical contributions of Kathleen Burke, Sharon Burke, Xiaohong Fan, and Phong Nguyen. F. Warren Lovell, M.D, performed a neuropathological evaluation of the postmortem brains. Tissue specimens were processed and stored at the Human Brain and Spinal Fluid Resource Center, Veteran's Medical Center, Los Angeles under the direction of Wallace W. Tourtellotte, M.D., Ph.D. This project is supported by the NIMH Conte Center Grant P50 MH60398, Pritzker Family Philanthropic Fund, William Lion Penzner Foundation (UCI), NIMH Grant #MH54844 (EGJ), W.M. Keck Foundation (EGJ), and the NIMH Program Project MH42251 (SJW and HA). "The authors are members of a research consortium supported by the Pritzker Neuropsychiatric Disorders Research Fund L.L.C. An agreement exists between the fund and the University of Michigan, Stanford University, the Weill Medical College of Cornell University, the Universities of California at Davis, and at Irvine, to encourage the development of appropriate findings for research and clinical applications".

References

Altar CA, Jurata LW, Charles V, Lemire A, Liu P, Bukhman Y, et al. Deficient hippocampal neuron expression of proteasome, ubiquitin, and mitochondrial genes in multiple schizophrenia cohorts. Biol Psychiatry 2005;58(2):85–96.

Auer H, Lyianarachchi S, Newsom D, Klisovic MI, Marcucci G, Kornacker K, et al. Chipping away at the chip bias: RNA degradation in microarray analysis. Nat Genet 2003;35(1):106.

Bahn S, Augood SJ, Ryan M, Standaert DG, Starkey M, Emson PC. Gene expression profiling in the post-mortem human brain—no cause for dismay. J Chem Neuroanat 2001;22(1/2):79–94.

Barrett T, Cheadle C, Wood WB, Teichberg D, Donovan DM, Freed WJ, et al. Assembly and use of a broadly applicable neural cDNA microarray. Restor Neurol Neurosci 2001;18(2/3):127–35.

Barton AJ, Najlerahim PRA, Harrison PJ. Pre- and postmortem influences on brain RNA. J Neurochem 1993;61(1):1–11.

Berger A, Stierkorb E, Nickenig G. The role of the AUUUUA hexamer for the posttranscriptional regulation of the AT1 receptor mRNA stability. Biochem Biophys Res Commun 2005;330(3):805–12.

Bernstein PPS, Ross J. The poly(A)-poly(A)-binding protein complex is a major determinant of mRNA stability in vitro. Mol Cell Biol 1989;9(2):659–70.

Bird ED, Barnes SE, MacKay J, Iversen AV, Shepherd LLM. Increased brain dopamine and reduced glutamic acid decarboxylase and choline acetyl transferase activity in schizophrenia and related psychoses. Lancet 1977;2(8049):1157–8.

Bolstad Benjamin Milo. Low-level analysis of high-density oligonucleotide array data: background, normalization and summarization. PhD dissertation, University of California, Berkeley, United States–California, 2004.

Breese CR, Marks MJ, Logel J, Adams CE, Sullivan B, Collins AC, et al. Effect of smoking history on [3H]nicotine binding in human postmortem brain. J Pharmacol Exp Ther 1997;282(1):7–13.

Brent DA, Moritz PJ, Allman G, Roth CJ, Schweers C, Balach JL. The validity of diagnoses obtained through the psychological autopsy procedure in adolescent suicide victims: use of family history. Acta Psychiatr Scand 1993;87(2):118–22.

Buesa CMT, Subirada F, Barrachina M, Ferrer I. DNA chip technology in brain banks: confronting a degrading world. J Neuropathol Exp Neurol 2004;63(10):1003–14.

Bunney WE, Vawter BB, Tomita MP, Li H, Evans J, Choudary SJ, et al. Microarray technology: a review of new strategies to discover candidate vulnerability genes in psychiatric disorders. Am J Psychiatry 2003;160(4):657–66.

Burke WJ, Chung OMK, Harmon HD, Miller SK, Berg JPL. Effect of pre- and postmortem variables on specific mRNA levels in human brain. Brain Res Mol Brain Res 1991;11(1):37–41.

Carter DE, Robinson JF, Allister EM, Huff MW, Hegele RA. Quality assessment of microarray experiments. Clin Biochem 2005;38(7):639–42.

Catts VS, Catts SV, Fernandez HR, Taylor JM, Coulson EJ, Lutze-Mann LH. A microarray study of post-mortem mRNA degradation in mouse brain tissue. Brain Res 2005;138(2):164–77.

Cope L. Function to assess RNA degradation in Affymetrix GeneChip data: AffyRNAdeg, 2005, http://www.bioconductor.org.

Corbett RJLA, Sterett R, Tollefsbol G, Garcia D. Effect of hypoxia on glucose-modulated cerebral lactic acidosis, agonal glycolytic rates, and energy utilization. Pediatr Res 1996;39(3):477–86.

Craddock N, Jones I. Genetics of bipolar disorder. J Med Genet 1999;36(8):585–94.

Cummings TJSJ, Yoon LW, Szymanski MH, Hulette CM. Recovery and expression of messenger RNA from postmortem human brain tissue. Mod Pathol 2001;14(11):1157–61.

Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, et al. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res 2005;33(20):e175.

Deep-Soboslay A, Martin AM, Bigelow CE, Herman LB, Hyde MM, Kleinman TMJE. Reliability of psychiatric diagnosis in postmortem research. Biol Psychiatry 2005;57(1):96–101.

Dumur CI, Best NS, Archer AM, Ladd KJ, Mas AC, Wilkinson VR, et al. Evaluation of quality-control criteria for microarray gene expression analysis. Clin Chem 2004;50(11):1994–2002.

Erraji-Benchekroun L, Underwood MD, Arango V, Galfalvy H, Pavlidis P, Smyrniotopoulos P, et al. Molecular aging in human prefrontal cortex is selective and continuous throughout adult life. Biol Psychiatry 2005;57(5):549–58.

Evans SJ, Datson NA, Kabbaj M, Thompson RC, Vreugdenhil E, De Kloet ER, et al. Evaluation of Affymetrix Gene Chip sensitivity in rat hippocampal tissue using SAGE analysis. Serial analysis of gene expression. Eur J Neurosci 2002;16(3):409–13.

Ford LP, Bagga PS, Wilusz J. The poly(A) tail inhibits the assembly of a 3′-to-5′ exonuclease in an in vitro RNA stability system. Mol Cell Biol 1997;17(1):398–406.

Fritz DT, Bergman N, Kilpatrick WJ, Wilusz CJ, Wilusz J. Messenger RNA decay in Mammalian cells: the exonuclease perspective. Cell Biochem Biophys 2004;41(2):265–78.

Gautier L, Cope L, Bolstad BM, Irizarry RA. affy–analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 2004;20(3):307–15.

Greenbaum D, Colangelo C, Williams K, Gerstein M. Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biol 2003;4(9):117.

Grissom SF, Lobenhofer EK, Tucker CJ. A qualitative assessment of direct-labeled cDNA products prior to microarray analysis. BMC Genom 2005;6(1):36.

Hakak Y, Walker JR, Li C, Wong WH, Davis KL, Buxbaum JD, et al. Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia. Proc Natl Acad Sci USA 2001;98(8):4746–51.

Hardy JA, Winblad WP, Gezelius B, Bring C, Eriksson GA. The patients dying after long terminal phase have acidotic brains; implications for biochemical measurements on autopsy tissue. J Neural Transm 1985;61(3/4):253–64.

Harrison PJ, Eastwood HP, Burnet SL, McDonald PW, Pearson BRC. The relative importance of premortem acidosis and postmortem interval for human brain gene expression studies: selective mRNA vulnerability and comparison with their encoded proteins. Neurosci Lett 1995;200(3):151–4.

Harrison PJ, Barton PA, Lowe AJ, Najlerahim SL, Bertolucci A, Bowen PH, et al. Terminal coma affects messenger RNA detection in post mortem human temporal cortex. Brain Res Mol Brain Res 1991;9(1/2):161–4.

Hollams EM, Giles KM, Thomson AM, Leedman PJ. MRNA stability and the control of gene expression: implications for human disease. Neurochem Res 2002;27(10):957–80.

Hosack DA, Dennis Jr G, Sherman BT, Lane HC, Lempicki RA. Identifying biological themes within lists of genes with EASE. Genome Biol 2003;4(10):R70.

Hynd MR, Lewohl JM, Scott HL, Dodd PR. Biochemical and molecular studies using human autopsy brain tissue. J Neurochem 2003;85(3):543–62.

Imbeaud S, Boulanger GE, Barlet V, Zaborski X, Eveno P, Mueller E, et al. Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. Nucleic Acids Res 2005;33(6):e56.

Isometsa E. Psychological autopsy studies—a review. Eur Psychiatry 2001;16(7):379–85.

Iwamoto K, Bundo M, Kato T. Altered expression of mitochondria-related genes in postmortem brains of patients with bipolar disorder or schizophrenia, as revealed by large-scale DNA microarray analysis. Hum Mol Genet 2005;14(2):241–53.

Johnson SA, Morgan DG, Finch CE. Extensive postmortem stability of RNA from rat and human brain. J Neurosci Res 1986;16(1):267–80.

Johnston NL, Shore CJ, Torrey AD, Yolken EFRH. Multivariate analysis of RNA levels from postmortem human brains as measured by three different methods of RT-PCR. Stanley Neuropathology Consortium. J Neurosci Methods 1998;77(1):83–92.

Jurata LW, Bukhman YV, Charles V, Capriglione F, Bullard J, Lemire AL, et al. Comparison of microarray-based mRNA profiling technologies for identification of psychiatric disease and drug signatures. J Neurosci Methods 2004;138(1/2):173–88.

Kelly TMMJ. Validity of DSM-III-R diagnosis by psychological autopsy: a comparison with clinician ante-mortem diagnosis. Acta Psychiatr Scand 1996;94(5):337–43.

Kingsbury AE, Nisbet FO, Cairns AP, Bray N, Eve L, Lees DJ, et al. Tissue pH as an indicator of mRNA preservation in human post-mortem brain. Brain Res Mol Brain Res 1995;28(2):311–8.

Konradi C, Eaton M, MacDonald ML, Walsh J, Benes FM, Heckers S. Molecular evidence for mitochondrial dysfunction in bipolar disorder. Arch Gen Psychiatry 2004;61(3):300–8.

Lee J, Hever A, Willhite D, Zlotnik A, Hevezi P. Effects of RNA degradation on gene expression analysis of human postmortem tissues. Faseb J 2005;19(10):1356–8.

Li J, Meng F, Evans SJ, Choudary PV, Tomita H, Vawter MP, et al. Sample matching strategies in gene expression studies of brain tissues. Program No. 923.18. Abstract Viewer/Itinerary Planner. Washington, DC; Society for Neuroscience, online; 2005.

Li JZ, Walsh VM, Tomita DM, Evans H, Choudary SJ, Lopez PV, et al. Systematic changes in gene expression in postmortem human brains associated with tissue pH and terminal medical conditions. Hum Mol Genet 2004;13(6):609–16.

Lipska BK, Deep-Soboslay A, Weickert CS, Hyde TM, Martin CE, Herman MM, et al. Critical factors in gene expression in postmortem human brain: focus on studies in schizophrenia. Biol Psychiatry 2006;60(6):650–8.

Luo Z, Geschwind DH. Microarray applications in neuroscience. Neurobiol Dis 2001;8(2):183–93.

Mexal S, Berger R, Adams CE, Ross RG, Freedman R, Leonard S. Brain pH has a significant impact on human postmortem hippocampal gene expression profiles. Brain Res 2006;1106(1):1–11.

Middleton FA, Lewis PL, Levitt DA, Mirnics PK. Altered expression of 14-3-3 genes in the prefrontal cortex of subjects with schizophrenia. Neuropsychopharmacology 2005;30(5):974–83.

Miller CL, Diglisic S, Leister F, Webster M, Yolken RH. Evaluating RNA status for RT-PCR in extracts of postmortem human brain tissue. Biotechniques 2004;36(4):628–33.

Mimmack ML, Brooking J, Bahn S. Quantitative polymerase chain reaction: validation of microarray results from postmortem brain studies. Biol Psychiatry 2004;55(4):337–45.

Mirnics K, Levitt P, Lewis DA. DNA microarray analysis of postmortem brain tissue. Int Rev Neurobiol 2004;60:153–81.

Mirnics K, Lewis MF, Levitt DAP. Analysis of complex brain disorders with gene expression microarrays: schizophrenia as a disease of the synapse. Trends Neurosci 2001;24(8):479–86.

Mirnics K, Pevsner J. Progress in the use of microarray technology to study the neurobiology of disease. Nat Neurosci 2004;7(5):434–9.

Mirnics K, Levitt P, Lewis DA. Critical appraisal of DNA microarrays in psychiatric genomics. Biol Psychiatry 2006;60(2):163–76.

Mitchell P, Mackinnon A, Waters B. The genetics of bipolar disorder. Aust N Z J Psychiatry 1993;27(4):560–80.

Mootha VK, Lepage P, Miller K, Bunkenborg J, Reich M, Hjerrild M, et al. Identification of a gene causing human cytochrome *c* oxidase deficiency by integrative genomics. Proc Natl Acad Sci USA 2003a;100(2):605–10.

Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 2003b;34(3):267–73.

Morrison-Bogorad M, Pardue ZAS. Heat-shock 70 messenger RNA levels in human brain: correlation with agonal fever. J Neurochem 1995;64(1):235–46.

Nestler EJ, DiLeone BM, Eisch RJ, Gold AJ, Monteggia SJLM. Neurobiology of depression. Neuron 2002;34(1):13–25.

Newton SS, Bennett A, Duman RS. Production of custom microarrays for neuroscience research. Methods 2005;37(3):238–46.

Prabakaran S, Swatton JE, Ryan MM, Huffaker SJ, Huang JT, Griffin JL, et al. Mitochondrial dysfunction in schizophrenia: evidence for compromised brain metabolism and oxidative stress. Mol Psychiatry 2004;9(7):684–97, 643.

Preece P, Cairns NJ. Quantifying mRNA in postmortem human brain: influence of gender, age at death, postmortem interval, brain pH, agonal state and inter-lobe mRNA variance. Brain Res Mol Brain Res 2003;118(1/2):60–71.

Ryan MM, Huffaker SJ, Webster MJ, Wayland M, Freeman T, Bahn S. Application and optimization of microarray technologies for human postmortem brain studies. Biol Psychiatry 2004;55(4):329–36.

Schoor O, Hennenlotter WT, Corvin J, Stenzl S, Rammensee A, Stevanovic HGS. Moderate degradation does not preclude microarray analysis of small amounts of RNA. Biotechniques 2003;35(6):1192–6, 1198–201.

Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol 2006;7:3.

Shastry BS. Bipolar disorder: an update. Neurochem Int 2005;46(4):273–9.

Shergill IS, Shergill NK, Arya M, Patel HR. Tissue microarrays: a current medical research tool. Curr Med Res Opin 2004;20(5):707–12.

Soverchia L, Ubaldi M, Leonardi-Essmann F, Ciccocioppo R, Hardiman G. Microarrays—the challenge of preparing brain tissue samples. Addict Biol 2005;10(1):5–13.

Spokes EG. GABA in Huntington's chorea. Parkinsonism and schizophrenia. Adv Exp Med Biol 1979;123:461–73.

Spokes EG, Iversen GNLL. Differential effects of agonal status on measurements of GABA and glutamate decarboxylase in human post-mortem brain tissue from control and Huntington's chorea subjects. J Neurochem 1979;33(3):773–8.

Spokes EG, Rossor GN, Iversen MNLL. Distribution of GABA in post-mortem brain tissue from control, psychotic and Huntington's chorea subjects. J Neurol Sci 1980;48(3):303–13.

Taylor GR, Crow CG, Johnson TJ, Fairbairn JA, Perry AF, Perry EKRH. Recovery and measurement of specific RNA species from postmortem brain tissue: a general reduction in Alzheimer's disease detected by molecular hybridization. Exp Mol Pathol 1986;44(1):111–6.

Tomita H, Vawter MP, Walsh DM, Evans SJ, Choudary PV, Li J, et al. Effect of agonal and postmortem factors on gene expression profile: quality control in microarray analyses of postmortem human brain. Biol Psychiatry 2004;55(4):346–52.

Torrey EF, Webster BB, Bartko MJ, Meador-Woodruff JJ, Knable JHMB. Neurochemical markers for schizophrenia, bipolar disorder, and major depression in postmortem brains. Biol Psychiatry 2005;57(3):252–60.

Vawter MP, Tomita H, Meng F, Bolstad B, Li J, Evans S, et al. Mitochondrial-related gene expression changes are sensitive to agonal-pH state: implications for brain disorders. Mol Psychiatry 2006;11(7):663–79.

Wilusz CJ, Wilusz J. Bringing the role of mRNA decay in the control of gene expression into focus. Trends Genet 2004;20(10):491–7.

Wu Z, Irizarry R, Gentleman R, Murillo F, Spencer FA. A model based background adjustment for oligonucleotide expression arrays. Department of Biostatistics Working Papers. Baltimore, MD; Johns Hopkins University; 2003.

Yang E, van Nimwegen E, Zavolan M, Rajewsky N, Schroeder M, Magnasco M, et al. Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. Genome Res 2003;13(8):1863–72.

Yates CM, Tennant BJ, Gordon MCA. Enzyme activities in relation to pH and lactate in postmortem brain in Alzheimer-type and other dementias. J Neurochem 1990;55(5):1624–30.

BMC Genomics

BioMed Central

Sample matching by inferred agonal stress in gene expression analyses of the brain

BMC Genomics 2007, **8**:336 doi:10.1186/1471-2164-8-336

Jun Z Li (junli@shgc.stanford.edu)
Fan Meng (mengf@umich.edu)
Larisa Tsavaler (tsavaler@shgc.stanford.edu)
Simon J Evans (evanssi@umich.edu)
Prabhakara V Choudary (pvchoudary@ucdavis.edu)
Hiroaki Tomita (htomita@uci.edu)
Marquis P Vawter (mvawter@uci.edu)
David Walsh (dwalsh@uci.edu)
Vida Shokoohi (vida@shgc.stanford.edu)
Tisha Chung (tisha@shgc.stanford.edu)
William E Bunney Jr. (webunney@uci.edu)
Edward G Jones (ejones@ucdavis.edu)
Huda Akil (akil@umich.edu)
Stanley J Watson (watsons@umich.edu)
Richard M Myers (myers@shgc.stanford.edu)

ISSN 1471-2164

Article type Research article

Submission date 5 March 2007

Acceptance date 24 September 2007

Publication date 24 September 2007

Article URL http://www.biomedcentral.com/1471-2164/8/336

http://www.biomedcentral.com/info/authors/

Sample matching by inferred agonal stress in gene expression analyses of the brain

Jun Z Li[1,2], Fan Meng[3], Larisa Tsavaler[1], Simon J Evans[3], Prabhakara V Choudary[4], Hiroaki Tomita[5], Marquis P Vawter[5], David Walsh[5], Vida Shokoohi[1], Tisha Chung[1], William E Bunney Jr[5], Edward G Jones[4], Huda Akil[3], Stanley J Watson[3], Richard M Myers[1,2]

1. Stanford Human Genome Center, Stanford University, Palo Alto, CA, USA
2. Department of Genetics, Stanford University, Palo Alto, CA, USA
3. Molecular & Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI, USA
4. Center for Neuroscience, University of California, Davis, CA, USA
5. Department of Psychiatry & Human Behavior, University of California, Irvine, CA, USA

Corresponding author:

Jun Li
Stanford Human Genome Center
975 California Ave.
Palo Alto, CA 94304

Telephone: 650-320-5841
email: junli@shgc.stanford.edu
FAX: 650-320-5801

E-mail addresses:

JZL: junli@shgc.stanford.edu,
FM: mengf@umich.edu,
LT: tsavaler@shgc.stanford.edu,
SJE: evanssi@umich.edu,
PVC: pvchoudary@ucdavis.edu,
HT: htomita@uci.edu,
MPV: mvawter@uci.edu,
DW: dwalsh@uci.edu,
VS: vida@shgc.stanford.edu,
TC: tisha@shgc.stanford.edu,
WEB: webunney@uci.edu,
EGJ: ejones@ucdavis.edu,
HA: akil@umich.edu,
SJW: watsons@umich.edu,
RMM: myers@shgc.stanford.edu

Abstract

Background: Gene expression patterns in the brain are strongly influenced by the severity and duration of physiological stress at the time of death. This agonal effect, if not well controlled, can lead to spurious findings and diminished statistical power in case-control comparisons. While some recent studies match samples by tissue pH and clinically recorded agonal conditions, we found that these indicators were sometimes at odds with observed stress-related gene expression patterns, and that matching by these criteria still sometimes results in identifying case-control differences that are primarily driven by residual agonal effects. This problem is analogous to the one encountered in genetic association studies, where self-reported race and ethnicity are often imprecise proxies for an individual's actual genetic ancestry.

Results: We developed an Agonal Stress Rating (ASR) system that evaluates each sample's degree of stress based on gene expression data, and used ASRs in *post hoc* sample matching or covariate analysis. While gene expression patterns are generally correlated across different brain regions, we found strong region-region differences in empirical ASRs in many subjects that likely reflect inter-individual variabilities in local structure or function, resulting in region-specific vulnerability to agonal stress.

Conclusion: Variation of agonal stress across different brain regions differs between individuals, revealing a new level of complexity for gene expression studies of brain tissues. The Agonal Stress Ratings quantitatively assess each sample's extent of regulatory response to agonal stress, and allow a strong control of this important confounder.

Background

Comparing cases and controls is one of the most widely used methods in genetic and epidemiological research to identify disease risk factors at the population level. From the study design standpoint, to maximize the power of detecting a true effect it is important to understand the major sources of phenotypic variation, and to minimize sample heterogeneity accordingly. Furthermore, to reduce the number of spurious positive findings due to confounding factors, it is important to match cases and controls on "well-established determinants" [1] that are not themselves the variables of direct interest. In practice, however, it is often difficult to declare *a priori* which variables, out of many that are examined, are the *established* risk factors. Occasionally, the major factors affecting the phenotypic outcome may be truly strong and well known, such as cigarette smoking as a risk factor for lung cancer, or older age for Alzheimer's disease. In most other situations, however, particularly those concerning multifactorial diseases such as cancer and psychiatric disorders, there are usually numerous contributing factors for the observed phenotype, but their relative importance is not always known beforehand. While many case-control studies automatically include age and gender in sample matching, additional variables that are important for the phenotype need to be chosen on a case-by-case basis, and sometimes only after the data have been collected and analyzed. In genetic association studies, for example, the ancestral background of human subjects can have a strong confounding effect [for example, 2, 3-5]. A parallel situation exists for gene expression analyses involving the use of postmortem samples, where tissue pH and near-death physiological stress can exert a major influence on the inter-individual variation of expression patterns [6, 7]. The impact of pH/agonal stress is so strong that it often far outweighs the influence of all other factors, including age and gender, and can obviate the detection of the impact of the illness. Because of this, more and

more gene expression studies in recent years take special precaution to match samples by pH and agonal factors, just as a well designed genetic study seeks to balance cases and controls by self-reported racial identity or continental ancestry.

Despite the widespread use and general success of these sample matching strategies, the risk of residual confounding remains. The key pH-sensitive genes or pathways, such as components of mitochondrial electron transport chain and proteasome genes, are highly variable between samples [8], and often appear as the top findings in microarray-based case-control comparisons of brain samples. For example, while several studies have reported down-regulation of mitochondrial transcripts in schizophrenia [9] and bipolar disorder [10], others have reported up-regulation [11, 12]. The samples used by Prabakaran et al. [9] had a slight case-control difference in pH, but many more controls than cases died of cardiac events [13]. Further analyses of the same RNA samples suggested that most of the findings that implicate mitochondria genes could be explained by effects of medication. Interestingly, the samples used in Sun et al. [10] were not balanced in pH, although the clinical condition appeared to be balanced. These conflicting results suggest that the role of pH-sensitive, stress-related genes in psychiatric disorders is still unresolved. Similarly, in the parallel example of genetic association studies, several recent analyses have highlighted the need for more stringent controls for the very strong genetic confounders [3, 4]. In studies involving highly diverse populations, many human subjects are admixed at the individual level, that is, they carry genetic material derived from several founding populations. For such individuals, a single self-reported racial or ethnic descriptor such as "African American" or "Hispanic" is no longer adequate for representing the proportional contribution from multiple ancestral origins. It has become more desirable, and in fact feasible, to infer the individual admixture ratios from the observed genetic data [14-16], and to apply these empirically derived ratios in *post hoc* sample matching [17, 18]. In effect, what

was initially a sample classification problem, in which subjects are categorized into discrete ethnic groups, has been turned into a "grades of membership" problem [19], in which individual samples are scored on one or several continuous variables. These continuous variables can be used for sample selection, case-control matching, or as a new covariate in regression analyses, stratified analyses, or ordered subset analyses [20-22]. The empirically derived ancestral ratios may be more effective for mitigating the impact of confounding, because they can be more accurate than self-reported ancestry, as the former are derived from the genetic data *per se*, and are less susceptible to survey errors or recall bias.

In this study, we applied a similar strategy to an ongoing gene expression study in which we compare postmortem brain tissues between normal controls and subjects who suffered from major depression, bipolar disorder, or schizophrenia. In a previous report [6], we described a classification-based analysis in which gene expression patterns in most subjects can be assigned to one of two main types: one from a low-pH, highly stressed group of samples, named "Type 2", and the other from a normal-pH, low-stress group of samples, named "Type 1". These two prototypes of expression patterns can be distinguished by strong and systematic changes in several biological pathways, including genes involved in energy metabolism and stress response. Since that report, we have increased the scope of our investigation from three brain regions in 40 subjects to six regions in up to 126 subjects (some regions were studied in fewer than 126 subjects). In carrying out case-control analyses, we found that even among the supposedly "purified" subset consisting of only the Type 1 samples, some residual heterogeneity in pH/agonal stress may still be driving the case-control comparison results, largely because of the overwhelming impact of agonal stress. Meanwhile, the pH- and stress-related genes that we and others have characterized continue to appear in the literature as among the top gene expression findings in comparative studies for a variety of diseases and conditions [9, 23]. This experience

motivated us to seek finer control of this obscuring variable by characterizing sample heterogeneity in greater detail. Specifically, we refined our previous dichotomous classification scheme to one that evaluates group-membership by quantitative ratings. A second rationale for pursuing this study came from the recognition that pH values are typically measured in one or two brain regions (in our case, cerebellum), whereas disease-related changes in gene expression are expected to occur in numerous brain regions. There is no *a priori* reason to assume that altered pH and agonal factors would impact all these brain regions in a uniform manner. Consequently, sample matching based on a parameter derived from a single brain area or the entire brain may not be reliable for all regions examined, whereas gene expression data for individual regions can be used to assess specific regional patterns of agonal stress.

To this end, we developed Agonal Stress Rating (ASR), a quantitative system that measures the degree of stress of each RNA sample on a continuous scale based on gene expression data. We examined the relationship between ASRs and conventional *pre hoc* indicators such as pH and clinically derived Agonal Factor Scores (AFS), compared the stress ratings across six brain regions, and assessed the performance of different sample matching strategies. We also developed rigorous data pre-processing methods, compared different options of defining the ASRs, evaluated the robustness of ASRs in terms of the between-lab and between-platform reproducibility, and explored several related analysis algorithms.

Results

Systematic technical variation and data processing strategies

Before we begin to characterize biological confounding factors, non-biological sources of variation must be identified. The microarray data used in this study were collected in multiple experimental batches, representing the mixed use of two types of Affymetrix Genechips (U133A/B and U133Plus_2), experiments run at three laboratories (at UC Irvine, UC Davis, and University of Michigan), RNA samples from six brain regions (AnCg, DLPFC, AMY, HC, CB, and NACC), and six cohorts of approximately 20 subjects each (four Mood Disorder Cohorts and two Schizophrenia Cohorts), for up to 126 subjects, about half of which were normal controls, the other half were cases of major depression, bipolar disorder, or schizophrenia. Cohort assembly, tissue dissection and RNA extraction took place in multiple stages, typically several months apart. The RNA samples were labeled and hybridized one cohort at a time, one region at a time, in two or three laboratories (called "Sites", not to be confused with the six brain regions) separately. As a result of these technical variabilities, the entire dataset contained systematic differences between sites, chip types, and sometimes, cohorts, although the cohort-cohort technical differences are blended with genuine sample-sample differences across cohorts. This type of technical variation warrants careful scrutiny, and must be adequately controlled to ensure the accuracy of analyzing biological differences.

After array scanning and Affymetrix Genechip data summary (a computational process that combines data from multiple oligonucleotides probes designed to interrogate a given transcript to obtain a single expression value for that transcript, see Methods for more details), we examined chip-to-chip similarity in each region by plotting the pairwise correlation matrices as color-coded

heatmaps, where red indicates high similarities between pairs of chips, blue indicates low similarity, and the samples are ordered by site and by cohort. Figure 1a showed one such correlation map for 201 AnCg chips produced by using logged intensities of all 12,734 Refseq gene-based probe sets on the U133A chip. In this example, as is the case for other brain regions, we analyzed data from two sites in our Consortium for six cohorts, representing 12 naturally occurring experimental batches. Figure 1a shows that the observed patterns aggregate in rectangular "blocks" of high correlation, indicated in red, corresponding to samples that are highly similar to each other in gene expression patterns. Importantly, the block-block partition coincides with the natural boundaries of experimental batches. Not all experimental batches can be definitively separated from each other; typically the 12 batches can be adequately described by 5-9 blocks, as sometimes two adjacent cohorts from one site form a single indistinguishable group, mostly due to relatively homogeneous technical conditions shared across these cohorts. In all, block-like structures are seen in every brain region, and almost always correspond to experimental batches, suggesting that they arose from changes in reagents, hybridization protocols, chip types, or scanning conditions.

At least two other lines of evidence suggest that the "blocks" are derived from technical variation between experimental batches rather than due to genuine biological differences between samples in different cohorts. First, when we set aside data for all human transcripts, and plot chip-chip correlations by using only the 68 Affymetrix control probe sets, which target spiked-in *E. coli* transcripts, the data still exhibit the same block structure as seen with the use of all genes [see Additional file 1, figure 1a], indicating that technical factors play a major role in delineating the blocks. Secondly, when we re-ran all samples on a custom 711-gene Illumina Beadarrays in a validation experiment that was done at one site and randomized samples across cohorts and regions, we did not observe the block-like separation between cohorts [see Additional file 1,

figure 1b], suggesting that biological differences between cohorts made a minor contribution, if any at all, to the observed "blocks".

The correlation matrices not only provide a means to visualize sample heterogeneity, but also allowed us to define a most parsimonious set of blocks for each region for the purpose of data normalization. To adjust for the block effect, we subtracted from each sample's logged expression value the median value of the block, and did so for each block and for every transcript. For example, for each gene, expression values for all Block 1 samples would subtract the median value of Block 1, and likewise for all other blocks, such that after the centering, the median of each block is at zero, effectively transforming the original data into the deviations from the block medians, in a procedure that is similar to adding a Block factor as a categorical variable in robust linear modeling (particularly the median polish method). The goal of this procedure is to remove a block-wide fixed factor, most of which, as we argued above, came from technical sources. The benefit of this adjustment, particularly the assumption of a fixed block-specific effect, can be evaluated by an objective criterion: how well the adjusted data increase the technical reproducibility of the same samples that were run at two or more sites. We found that after removing the block effect by median centering, we improved the between-site similarity for replicate chips run at multiple sites (Figure 1b). For samples that were run on both the U133A and the U133Plus_2 chips, removing a fixed between-chip-type effect produced satisfactory agreement between the two chip types [see Additional file 1, Figure 1c].

Although the systematic differences between blocks can be adjusted in this way, the assumption of a fixed effect is not expected to hold for all genes in all samples equally well. Other types of variation, including within-block heterogeneity, however, are not readily discernible in the data, and are probably impossible to control. Our analysis showed that a major portion of the