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

between-block differences have been accounted for by a fixed technical effect affecting all samples within the block, for most of the genes studied. In the example of 201 AnCg chips, the median proportion of variance explained by the "block" factor, across all 12,734 genes, is 2.5% of the total variance. Because the 2.5% between-block variance has one degree of freedom, while the 97.5% within-block variance has 199 degrees of freedom, the F test showed that more than half of the genes had significant between-block differences. In this case, 56% of the genes satisfy $P < 0.05$ (one-way ANOVA) for the Block factor.

Biological variation across samples

Previously we showed that most of the brain samples in our study can be classified into two main types of expression patterns [6]: those from individuals who died quickly and had normal tissue pH, and those from individuals who suffered prolonged death, typically with medical complications, and exhibited low tissue pH. The threshold value between low and normal pH is around 6.6 in our samples, but it varied among different studies and different tissue collections [24]. Figure 2a shows the correlation heatmap for the 201 AnCg samples obtained by using expression values already adjusted for the block effect as described above. This is the same plot as in Figure 1b, but with the samples re-ordered so that the previously designated Type 1 samples are grouped to the lower left side of the graph. While the distinction between the two classes can be clearly seen, there are still samples of intermediate patterns that may correspond to varying degrees of agonal stress that do not readily belong to the two opposing prototypes.

Because pH/agonal stress acts as an exceedingly strong confounder in gene expression studies, a dichotomous classification may not be sufficient to ensure that cases and controls are well-stratified, or balanced within each stratus. In addition, different brain regions may carry different stress outcomes; whereas this aspect of data heterogeneity is not well informed by pH values

measured in one brain region, nor by the clinical indicators. One possible approach for dealing with residual heterogeneity is to subdivide the main classes further, such as into Types 1A, 1B, 2A, 2B, etc, effectively classifying samples in multiple tiers to establish nested subsets, within each of which the cases and controls are more homogeneous and better balanced. Figure 2b shows an example of three-type classification; however, even finer subdivisions are clearly possible. In practice, it is difficult to decide the number of clusters or layers of clusters needed, and the samples often show genuinely graduated differences of expression patterns. A natural alternative to a finer-grained classification approach, therefore, is to rate samples on one or several continuous scales. Toward this goal, we first carried out a Principal Component Analysis. We re-ordered the 201 AnCg sample by their first principal component (PC1) scores, and the resulting heatmap (Figure 2c) shows a gradual transition from one end of the spectrum to the other. Similarly smooth progressions are also observed for the other five regions (not shown). Importantly, the PC1 scores are highly correlated with pH ($r = 0.59$ for AnCg, $P < 10^{-13}$), the clinical agonal factors, as well as with the previously determined Type 1-Type 2 designations; whereas PC2 and PC3 scores have almost no correlation. For example, in AnCg, r is 0.04 ($P > 0.35$) and 0.07 ($P > 0.25$) for pH-PC2 and pH-PC3, respectively. Note that with $n = 126$, r needs only to be approximately 0.18 to be significantly non-zero at $p = 0.05$). This result not only confers a biological meaning to PC1, but also suggests that a single continuous variable is likely to be sufficient to capture most of the gradual progression of expression patterns from the low-pH prototype to the normal-pH prototype.

Not all genes contributed equally to the placement of individual samples along the gradient of membership. We selected the 20% strongest Type 1-like samples (on the lower left corner of Figure 2c) and the 20% strongest Type 2-like samples (on the upper right corner), and calculated the Student's t scores that contrasted the group means between these two canonical groups. This

allowed us to rank genes by their Type 1- Type 2 absolute t scores, with the "top genes" being those that are most strongly affected by pH/agonal stress. When we re-plot Figure 2c by using the 25% strongest pH-sensitive genes (instead of all 12,734 genes), a strong gradient is clearly seen (Figure 3, upper left panel). The gradient between samples becomes much dampened when the second 25% of ranked genes (upper right of Figure 3) is used, and fades away almost entirely with the use of the third and last 25% of ranked genes (lower panels of Figure 3). This result indicates that the top 50% of genes are likely to be informative for membership inference, with the top 25% and top 5% of genes carrying increasingly greater discriminating power, as one would expect. In Additional file 2 we showed a heatmap of expression levels of the top 25% of transcripts across 201 AnCg samples. These genes have been used to calculate the sample-sample correlations shown in Figure 3, upper left panel. The genes are ordered from left to right by their coefficients in the first principal component (i.e., the first eigenvector), whereas the samples are ordered from top to bottom by their first principal component scores. The actual expression levels are provided in Additional file 3, which is a .cdt file that can be opened in Java Treeview for flexible browsing. To estimate the number of genes significantly affected by the pH effect, we used the Nearest Shrunken Centroids classifier [25] to calculate the cross-validation errors in a two-class classification analysis, and examined gene panels containing varying numbers of most discriminating genes. Panels having as few as 297 (2.3% of the total of 12,734 probe sets) and as many as 4,720 (37.1%) genes resulted in eight or fewer cross-validation errors out of 201 samples, and formed a plateau of error curve [see Additional file 4], indicating a broad range of the number of informative genes. Similarly, in genetic association studies, some DNA polymorphisms are more informative for distinguishing different populations [26]. In our procedure to construct an Agonal Stress Rating for individual samples (described below), we used 25% or 5% of "top genes", and always found that with the 5% top genes the intermediate ratings are more "stretched out" than with 25% of top gene (not shown), as one

would expect, because the strongest pH-sensitive genes are more powerful in distinguishing subtle differences in intermediate grades of membership. We ran the Principal Component Analysis for six regions separately and defined the strongest Type 1/2-like samples for each region. When we subsequently compared the strongest Type-1 against the strongest Type-2 samples and ranked genes by their t scores, the top 5% or 25% genes are similar across the six regions as their t scores are highly correlated across regions [see Additional file 5]. On average, a top (or bottom) 10% gene in one region has a 57% chance to be among the top (or bottom) 10% in another region. We ranked genes by their average t score ranks in five regions (all except CB, as cerebellum is an outlier region for gene expression due to its unique anatomical and physiological properties), and listed the 1000 most strongly up-regulated and 1000 most strongly down-regulated genes in Additional file 6. Genes in these lists can be used as most informative genes in future, independent studies.

Agonal Stress Ratings

Because the PC1 scores successfully arrange the samples into a smooth gradient, these scores by themselves could serve as a measure of agonal stress in individual samples. However, we found that PC1 may be strongly influenced by the scale of variance in individual chips and sometimes by a small number of "outlier" samples, whereas our method of using PC1 as a crude criteria to pre-select the strongest ~20% Type 1- and Type 2-like samples is more robust to these outliers. In addition, the interpretation of the Principal Component scores requires the notion of decomposition of the observed gene expression patterns into the linear combination of multiple components. This interpretation is most natural in cases where each sample represents the actual mixing of multiple cell types, each having its own canonical expression patterns. The meaning of PC scores is also complicated by the process of analyzing log-transformed signal values, while actual transcript levels are "mixed" on the linear scale. Alternative indices, such as the

probability of classification [25], or the prediction strength such as the "margin-of-victory" measures adopted in Golub et al. [27], are most appropriate when there are multiple genuinely discrete outcomes, for which what is uncertain is the strength of evidence for class assignment. In our case, gene expression in the brain is most likely affected by agonal stress in a graded (albeit non-linear) fashion. Although many of the samples belong to the two extreme states--one is minimally affected by stress, the other for samples that have converged to the quasi-steady state of "thoroughly affected"--some tissues in our collection are apparently sampled at an intermediate physiological state, and can best be characterized by the sample's relative distance to the two ends of the spectrum, that is, the distance to the prototypical normal-pH pattern minus the distance to the prototypical low-pH pattern, where "distance" can be the Pearson's correlation, Euclidean Distance, Spearman's rank correlation, or a number of other metrics. The Agonal Stress Rating is thus defined as the difference between each sample's distances to the two prototypical patterns, and can be calculated by using different sets of most informative genes, which in turn can be defined by comparing the most extreme samples (see Methods). Although ASR is formally neither a probability of classification nor a mixing ratio, it is actually quite similar to these other measures as most of them are variants of the linear discriminant function.

A comparison of the median ASR values (for each subject, across regions and sites) with tissue pH (Figure 4) reveals two features. First, there is a general correlation, i.e., low pH samples tend to have low ASRs. Secondly, there is considerable local discrepancy, i.e., among the normal pH samples, the ASR-pH correlation is weak. These discrepancies imply that the ASR values, which are derived directly from the gene expression data, are sometimes at odds with the measured brain pH. In the face of such discrepancies, we need to determine which index is a more accurate surrogate for the actual degrees of stress experienced by the individual subjects in specific brain regions, and therefore provides a safer control of the agonal stress confounder. We

examined a test case: from a set of 55 AnCg samples previously "included" in an intermediate-stage analysis, we selected all of the thirteen Control samples with ASR > 0.8, and divided all of the fourteen Major Depression samples into two groups (Figure 5a): one that was matched with the 13 Controls in pH (6.91 +/- 0.13 in controls, 6.87 +/- 0.11 in cases, p = 0.45, t test) but not in ASR (1.25 +/- 0.27 in controls, 0.18 +/- 0.44 in cases, p = 0.0003), another matched to the Controls in ASR (1.25 +/- 0.27 in controls, 1.14 +/- 0.18 in cases, p = 0.28) but not in pH (6.91 +/- 0.13 in controls, 7.13 +/- 0.08 in cases, p = 0.00015). We carried out MDD-Control comparisons for the two MDD groups separately, and analyzed the top and bottom 4,000 genes in EASE (the Expression Analysis Systematic Explorer) [28] for significantly enriched Gene Ontology terms. Figure 5b shows that the pH-matched comparison yields the gene families and pathways associated with agonal stress that we and others previously discovered, whereas the ASR-matched group significantly reduces the effect. This result is not surprising, as by the definition of ASRs, we effectively have balanced the key stress-related pathways when we match samples by ASR. This test case, however, shows that matching by pH is not always safe for guarding against the agonal stress confounder, whereas the empirically derived ASR values provide a more accurate assessment of the regulatory responses to near-death stress in individual samples, and allow a more stringent control. In a section below we will describe the robustness ASRs, particularly the finding that deriving ASRs by using only the control samples did not substantially change the result.

Between-region differences

After calculating ASR for all six regions and for two sites in each region, we obtained twelve series of ASR values. As AMY and CB had larger numbers of missing chips (samples not analyzed on microarrays), we plotted the eight ASR series for the four remaining brain regions in color codes along with pH values and AFS (in two levels) (Figure 6, see figure legend for color

codes), with the samples sorted by pH (high pH on the top of the figure), and missing data in white. This figure has several broadly recognizable features. First, ASR results from the two sites tend to agree, although not always so. Secondly, ASRs in most samples tend to be correlated across the four regions, reflecting brain-wide patterns. Thirdly, the ASR scores show a coarse correlation with pH and the clinical AFS scores, as samples at the bottom of the figure are generally those with both low pH and low ASR (this can also be seen in Figure 4, where pH is plotted against the median ASR across the six brain regions). The correlations between pH and ASR, across all 126 samples, range from 0.3 to 0.6 among the twelve ASR series (from 0.57 to 0.63 in AnCg, DLPFC, CB, and NACC, 0.47 in HC, and 0.3 in AMY, which has the smallest sample size: 66 out of 126 total samples. Note that pH was measured in CB). But these correlations are much smaller when using only the 90 pH>6.61 samples, indicating that much of the correlation was driven by the large differences between the high-pH and the low-pH samples. Finally, some of the subjects who had normal pH yet low ASR can be explained by having clinically recorded agonal stress (AFS=1 or higher).

However, beyond these large trends, there remain striking between-region differences in many individual subjects. For example, the first subject from the top, indicated by the first arrow on the right, has low ASRs (in green) in AnCg, and normal to high ASRs in the other regions. The third and sixth subjects, indicated by the next two arrows, have unusually low ASRs in DLPFC and HC, respectively. These strong region-region differences in ASRs are robustly observed (see the next section) and suggest genuine differences in stress outcome across different parts of the brain. Such differences may arise due to inter-individual variabilities in local structure or function, resulting in region-specific vulnerability to agonal stress. An additional possibility is that the nature of the illness that caused the death, such as hepatic, renal, or cardiac failures, may have triggered a region-specific brain response. These results underscore the value of using

ASRs to carry out per-region sample matching, as pH and clinical indicators are measures for the entire brain.

Robustness of ASRs

It is difficult to know how much between-region differences arise from variabilities in tissue dissection, but this factor is unlikely to explain most of the differences we observed here, as the effect is seen even for relatively large, easy-to-dissect regions such as DLPFC and CB. Although the nature of the observed regional differences is still not clear, most of such differences are likely due to actual biological differences rather than to technical variability in performing RNA labeling and hybridization or to data analysis methods. There are several reasons for making this conclusion. First, while regional differences of ASRs are often large (+/- 0.5), ASRs varied to a much lesser degree (1) between sites, (2) between different numbers of informative genes used, (3) between different distance measures (correlation, rank correlation, or Euclidean distance), (4) with another round of normalization after median centering, (5) following the selection of only the high ASR samples as the basis for re-calculating block medians, or (6) importantly, following the use of only normal controls rather than both cases and controls for ranking genes by t scores. Secondly, we examined the possibility that some regions may actually have a greater Type 1-Type 2 difference than the other regions, and different samples may have different regional differences. To this end, we calculated additional versions of ASR by using (1) the same ten Type 1 samples and the same 14 Type 2 samples to determine the informativeness of genes in all six regions; (2) combined all six regions (more than 1200 chips) in designating the canonical samples on the two extremes, defining most informative genes, and calculating ASR across all 1200 samples (all regions together). These alternative versions did not significantly alter the ASR patterns across samples and regions, and did not explain most of the observed regional differences. Thirdly, with our 700-gene Illumina data, we