were able to compare the ASRs based on data from two different platforms. For each sample, we averaged the Affymetrix ASR values from two sites to obtain the six values for six regions, and calculated their correlation to the corresponding six ASR values from the Illumina data. We obtained 67 such correlations in 67 samples for which we had both the Affymetrix and Illumina data. Of these 67, 58 correlations were above 0.33, 43 were above 0.67, and 28 were above 0.8, with most of the smaller correlations explained by very low between-region variations, that is, when the six scores are similar to begin with for a given sample, it is more difficult to observe a highly correlated pattern for this sample in a different platform.

Discussion

Many gene expression analyses have shown that agonal stress at the time of death can significantly alter expression patterns in postmortem tissues [29-33]. Because of this, comparative studies that rely on postmortem material must take every precaution to ensure that cases and controls are properly balanced with regard to this well-established confounder. What remains unclear, however, is to what extent residual imbalance still accounts for the most significant findings. The study presented here is motivated by two considerations. First, a dichotomous classification of good-quality versus bad-quality samples is often inadequate in assuring that cases and controls are well matched among the "good samples". Secondly, clinical variables such as AFS and continuous variables such as tissue pH have several practical limitations, both in terms of accuracy and in their inability to assess each brain region separately. To address these concerns, we developed a rating metric to infer agonal stress based on expression data, and strongly advocated a conservative approach that uses these empirical ratings to select a homogeneous group or to balance cases and controls. The adoption of continuous

ratings instead of the more commonly used categorical assignment is expected to bring several advantages: providing a more reliable measurement of underlying heterogeneity that does not show natural boundaries, retaining important information regarding sub-threshold variability, and allowing flexible integration with other variables.

Between-region differences

Our results revealed striking between-region differences in stress outcome. This complexity presents additional challenges to gene expression analyses of the brain, especially if the goal is to not only study each region separately, but also to study biological regulation across brain regions, as well as their disruption in the diseased state. During our initial analysis, we had hoped that all regions of the same brain would have a similar stress profile, making it possible to use one or more regions as the same-subject control for the other regions. In this scheme, between-region differences in gene expression, instead of gene expression levels in each region separately, are compared between cases and controls, under the assumption that agonal stress is acting more extensively across the entire brain, whereas expression signatures of psychiatric disorders are restricted in some regions. This approach is reminiscent of the genetic method of using family controls to remove the impact of population structure. Our analysis, however, revealed considerable heterogeneity between brain regions in stress-related expression patterns, suggesting that this between-region analysis method is still problematic and requires further investigation.

The need for a conservative approach

One of the concerns in adopting this approach is that gene expression changes in response to near-death conditions may overlap with those affected by chronic psychiatric states. If the Agonal Stress Ratings reflect the blending of both effects, how do we know whether we have

under- or over-corrected the main effect of interest? Unfortunately, at the present time we do not know many genes or pathways that are clearly and reproducibly altered by major depression and bipolar disorder; as a result we currently do not have a definitive set of "true positives" to fine-tune the stringency of the sample matching strategy. To address this type of question would require a sufficiently large collection of samples to directly investigate whether patients of mental disorders are more likely than the healthy controls to have more severe terminal stress, lower tissue pH values, and more pronounced changes in gene expression patterns. When the sample size is below or near 100, as is currently the norm in expression profiling of human brains, a study is not properly powered to formally answer this question, and the gene expression signature due to mental disorders, if it overlaps with that due to agonal stress, will be easily obscured. However, despite this uncertainty, it is widely recognized that the effects due to agonal stress usually have a much more potent influence on gene expression patterns than do mental disorders. In such a situation, the chance of finding false positives due to residual imbalance is quite high. A continuous, quantitative rating system, like the ASR that we developed here, can greatly reduce the number of false positives in such studies by allowing a quantitative assessment of the broadly recognizable outcome of the most prominent confounder. By applying the ASR, it remains possible to detect a true signal for psychiatric disorders as long as it is strong enough or sufficiently independent of the confounding effect, particularly if the disease related changes involve a small number of transcripts whereas the agonal conditions usually affect a large number of genes [6]. In essence, the practice of sample matching by data-derived indices constitutes a higher standard for a "true positive", requiring that its main effect (for psychiatric disorders) to rise above its potential colinearity with the major extraneous covariates (changes due to agonal stress). The situation is analogous to the case of a genetic association study of, for example, cardiovascular disorders, where the goal is typically to detect genetic association with the disease while adjusting for the traditional risk factors, such as blood

pressure, diabetes, body-mass index, and serum levels of HDL and triglyceride [34]. Because these traditional risk factors are also influenced by genetic variation, making such corrections constitutes a more focused goal: to discover genetic contribution to the disease *independent of the genetic basis of the other risk factors*, even though in the long run it is ideal to be able to study all the contributing factors in conjunction with each other. By the same token, the near-term goal of our study is to discover gene expression signatures of psychiatric disorders independent of gene expression changes due to confounding factors, especially when these factors (stress at death) are not likely to be closely related to the biology of mental illness.

Conventional versus empirical indicators

In this study, we found that the Agonal Stress Ratings are sometimes at odds with tissue pH values or clinically-recorded medical factors. These indices, though useful in most situations, have several shortcomings when serving as surrogates for the underlying biological heterogeneity. First, different brain regions may carry different stress outcomes, while pH values were usually measured in one region. Likewise, the clinical indicators do not inform regional heterogeneity. Secondly, these conventional variables inevitably contain measurement errors or incomplete information. Thirdly, brain pH is both the outcome of prior episodes of stress and the trigger for subsequent physiological responses. As a result, even if pH is measured in all regions, and even if pH is similar across regions [32], the samples that have similar pH are still not likely to have a uniform stress-induced gene expression profile. For these reasons, we recommend using ASR as a primary criterion in sample selection and sample matching, analogous to using inferred ancestry in genetic association studies [15].

It is also important to point out that the tissue pH measurements and clinical records are still of immense value in sample matching. They provide biological meaning to the ASR values, and

allow an integrated strategy that combines these different indicators in sample selection [35]. We believe that future studies need to continue collecting clinical information as extensively as possible, and if feasible, measure pH in multiple brain regions to directly assess regional differences of degree of stress.

Assumption of neutrality

An implicit assumption of using inferred ancestry to control population stratification in genetic studies is that most of the genetic markers tested are phenotypically neutral; that is, not associated with the disease under study. Under this assumption, when data have been collected for a large number of loci (that is, not just for a few candidate genes), the majority of these loci can be used to infer population structure [14, 16, 36, 37]. Alternatively, they can be used as Genomic Control loci to detect over-dispersion of the standard test statistics due to stratification [38-40]. This assumption of neutrality generally holds when we expect to detect large or moderate signals of association for only a small number of causative genes. In gene expression analysis, however, it is possible that a considerable proportion of the transcripts are affected both by the extraneous confounders such as agonal stress and by the disease being studied. As a result, some of the genes used to infer agonal stress may also be associated with the disease, making it difficult to separate the two effects. However, this apparent difference between genetic and gene expression studies depends on the actual context of the study. Population stratification can be a very strong confounder for traits or diseases that have subtle genetic effects but large population differentiation (in incidence rates as well as in allele frequencies), or are studied in recently admixed population where chromosomal segments of defined ancestry can be as long as 5 cM [41]. In these situations, a substantial fraction of genetic markers may show weak association, while most of such signals may disappear after correcting for inferred individual ancestry or locus-specific ancestry. For both gene expression and genetic studies, using the data *per se* to

detect sample heterogeneity is a practically useful yet ultimately exploratory method, especially when true positives are not known beforehand.

Model considerations and the need for canonical patterns

In inferring individual ancestry from genetic data, it is helpful to know the genetic profile of the putative ancestral populations, for example, the allele frequencies determined from samples that represent the progenitors of the admixed individuals. Without these "pure" canonical profiles, while it is still theoretically possible to simultaneously determine both the canonical ancestral patterns and the mixing ratios of each sampled individual, the results are less reliable and often improperly scaled [16]. Similarly, our strategy to derive Agonal Stress Ratings relies on having obtained data for both "normal" tissues and severely stressed tissues. These opposing extremes serve as the two reference points to measure the grades of membership in between. Hypothetically, in a different study, if none of the samples (or too few of them) are from the severely stressed low-pH samples, the canonical signature of the confounder (or the principal components) would be much less precisely defined. However, even when the absolute values of ASR only scale from "normal" to mildly stressed, the relative ratings among samples can still be used as the empirical basis for sample matching, or as the basis for a variety of *post hoc* strategies, such as stratified analysis, or incorporating covariates in regression analysis. We have found that the main differences due to agonal stress can be robustly observed, i.e., are highly similar between study sites and across two microarray platforms. This means that the canonical patterns that we describe here, featuring prominently energy metabolism genes and stress response pathways, can be transferred to other studies of postmortem tissues even if these studies characterized a narrower range of the confounding effect. In Additional files 3 and 6 we provide the genes most strongly affected in our dataset; these genes can be used as informative stress markers for pre-evaluating future sample collections, even before microarray experiments. The

parallel example in genetic studies is that one can profile the genetic signature for a few major continent groups (such as in the International HapMap Project) and use this information repeatedly to infer admixture in individuals from a heterogeneous population such as the African Americans, sometime by using a small number of most informative genetic markers [26]. We also wish to emphasize that using only the controls to define the two canonical expression patterns did not change the ASR ratings in any meaningful way, reflecting the fact that the case-control differences are far subtler than the stress-related differences.

The situation would be much simpler if the mixing ratios are known by other, independent methods, or if the canonical patterns were profiled in independent studies. Such an "expression signature bank" can be established by studying the response to controlled stress treatment in animal models, or by analyzing "pure" classes of brain cells, such as laser-captured defined cell populations. Stuart et al. [42] gave an example where the mixing ratio is known for tumor samples when tissue pathologists reported *a priori* the fractional composition of different cell types in each sample. Conversely, Lu et al. [43] used expression data for synchronized yeast cells in defined phases to infer mixing ratios of asynchronized samples—this is a case where the basic patterns were known, with mixing ratios being the target of inference. When neither is known, as in our case, the solution cannot be arrived at analytically but can be optimized iteratively. For example, several methodology studies [44-46] have implemented variants of the Expectation-Maximizing (EM) algorithm to simultaneous search for the unknown canonical patterns and the unknown mixing ratios in gene expression data. Similar methods, including ones involving Markov Chain Monte Carlo methods, have been developed for inferring individual's genetic ancestry [14, 16, 47]. In our dataset, as the principal pattern is relatively strong, we expect the result to converge quickly and decide to adopt a simple definition of ASR. In more complicated situations, for example, when the second or higher Principal Components

can be interpreted as signatures of other confounders such as age, medication, or cell types, it would be important to apply these more sophisticated algorithms.

Conclusions

In this study 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. We found that ASR-based matching provides tighter control of the agonal effect than by using pH. We also found that different brain regions exhibited different stress outcomes and that such regional patterns also varied between individuals. Our results once again highlight one of the main challenges in gene expression studies of psychiatric disorders: transcript levels are under the influence of a large number of confounding factors. Agonal stress undoubtedly plays a major role; as a result the stress-related, pH-sensitive genes or pathways are prone to appear as the top findings in case-control comparisons. We propose to adopt a conservative approach for the genes and pathways that are clearly altered by the confounding factors, even if it is possible that they are also influenced by the disease of interest. Deriving an ASR from gene expression data and using it for sample matching is one example of such an approach. This approach involves using continuous ratings as opposed to categorical assignments, thus represents an early attempt to apply graded classification methods to a sample heterogeneity problem for brain tissues, similar to the application of dimensional models in psychiatric diagnosis. In practice, we believe that this method can and should be expanded to the characterization of other sources of biological variation, such as medication, age, and cell-type composition of the dissected brain tissues. Drug use by psychiatric patients is likely to affect gene expression, regardless whether the medication was effective in treating the disease or not. But medication history is one of the most elusive aspects of clinical information to accurately record and quantify. Likewise, an

individual's chronological age may vary greatly from the actual physiological age of the tissue under study, while the neuronal-glial ratios of the dissected tissues may also vary from brain region to brain region, from subject to subject. Proper monitoring of these additional sources of phenotypic variation is an important prerequisite for the eventual identification of the true gene expression signature of mental disorders, and in this regard, animal models of either drug treatment or stress, independently applied, in the absence of psychiatric history and genetic heterogeneity, represents a powerful alternative strategy for defining canonical expression patterns and assessing their relative contributions.

Methods

Sample acquisition, RNA labeling, and microarray hybridization were carried out as described previously [6, 48]. We used RMA [49] to obtain the probe set summary values for the Affymetrix U133A and U133Plus_2 Genechips. We removed chips that were clearly outliers, or failed the "gender-test", in which we confirm that Y-chromosome transcripts are only detected in male samples. The entire dataset, consisting 1218 Genechips, has been deposited to the Gene Expression Omnibus [50] with the Accession Number GSE6306.

Since the original Affymetrix probe definitions require frequent updates, and are known to contain errors and redundancies, we developed a custom probe definition method that involves re-annotating all Affymetrix probes by sequence alignment to the most recent build of genomic DNA sequences and a variety of transcribed sequence collections, including Unigene, Refseq, ENSEMBL Genes/Transcripts/Exons, and Entrez, etc [51]. For each of these transcript definition systems, the probes that could be uniquely assigned to their transcript targets were assembled in custom Channel Definition Files (CDF files) as individual probe sets, which assumed the names of the matched transcripts, thus replacing the Affymetrix probe set ID's. The analysis presented here used the third generation of our Unigene-based CDF files, which were based on Unigene Build 176. The CDF files can be downloaded for free at [52].

We calculated chip-chip correlation in each region by using RMA summaries values and defined blocks by visual inspection of the correlation heatmaps. Logged expression values for each gene were centered within each block. Median centered values were used for the Principal Component Analysis, whereby the PC1 scores were used to rank samples so that a proportion of samples at one extreme were designated canonical Type 1 samples, while a proportion of

samples at the other extreme were designated canonical Type 2 samples. These two groups of samples were compared by the Student's t test, so that all genes can be ranked by the t scores. The ASR of each sample is then calculated as the "distance" to canonical Type 1 pattern minus its distance to the canonical Type 2 pattern by using a proportion of the genes with the largest absolute t scores. A range of parameter values were tested, including different proportion of samples used to define canonical patterns, different number of genes used for calculating ASRs, and different measured of chip-chip distance. For example, most results in this report were based on using approximately 20% of samples at each extreme of PC1 scores, 25% of the genes of largest t scores, and Pearson's correlation as distance measures. The nearest Shrunken Centroid Classification was carried out by using the Predictive Analysis of Microarrays (PAM) package in R. All R scripts used in the analysis are available from the authors.

The Illumina custom Beadchips were designed to cover ~700 transcripts that represent both biologically candidate genes and preliminary Affymetrix results to be validated. Sample labeling and hybridization were performed according to manufacturer's specifications. In all, about 67 samples were analyzed in each brain region, except for AMY, which we analyzed only 54 samples. All of the nearly 400 RNA samples were randomized with regard to cohort and region. The calculation of ASRs was carried out in a similar fashion as with the Affymetrix data except that there were only 700 genes used in the analysis.

Abbreviations

AnCg, anterior cingulate cortex; DLPFC, dorsal lateral prefrontal cortex; AMY, amygdala; HC, hippocampus; CB, cerebellum; NACC, nucleus accumben

Competing interests

The authors are members of the Pritzker Neuropsychiatric Disorders Research Consortium, supported by the Pritzker Neuropsychiatric Disorders Research Fund L.L.C. There exists a shared intellectual property agreement between the academic and philanthropic entities of the consortium for potential findings.

Authors' contributions

SJE, PVC, MPV supervised the collection of gene expression data by using the Affymetrix Genechips. LT, VS, TC collected gene expression data by using the Illumina custom Beadchips. FM and SJW developed data management and analysis infrastructure and the Affymetrix custom CDF files. DW and WEB collected and compiled sample information that includes diagnosis, agonal stress and other clinical/demographic variables. MPV and WEB collected brain pH data. JL developed the analysis method in this work and performed the statistical analyses. JL and RMM wrote the manuscript. SJE, PVC, HT, MPV, WEB, EGJ, HA, SJW participated in data analysis and drafting of the manuscript. WEB, EGJ, HA, SJW, RMM conceived, coordinated, and supervised the overall study. All authors read and approved the final manuscript.

Acknowledgements

We thank Hyungwon Choi, Terry Speed, Margit Burmeister, Laura Scott for their helpful advice. The authors are members of the NIMH Silvio O. Conte Center for Genomic Studies of Bipolar Disorder and Major Depression and the Pritzker Neuropsychiatric Disorders Research Consortium, supported by the Pritzker Neuropsychiatric Disorders Research Fund L.L.C. This work also received support from the William Lion Penzner Foundation.

References

1. Elston R, Olson J, Palmer L (eds.): **Biostatistical Genetics and Genetic Epidemiology**. Chichester, England: John Wiley &Sons, LTD; 2002.
2. Knowler WC, Williams RC, Pettitt DJ, Steinberg AG: **Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture**. *Am J Hum Genet* 1988, **43**(4):520-526.
3. Campbell CD, Ogburn EL, Lunetta KL, Lyon HN, Freedman ML, Groop LC, Altshuler D, Ardlie KG, Hirschhorn JN: **Demonstrating stratification in a European American population**. *Nat Genet* 2005, **37**(8):868-872.
4. Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN, Pato MT, Petryshen TL, Kolonel LN, Lander ES, Sklar P, Henderson B, Hirschhorn JN, Altshuler D: **Assessing the impact of population stratification on genetic association studies**. *Nat Genet* 2004, **36**(4):388-393.
5. Marchini J, Cardon LR, Phillips MS, Donnelly P: **The effects of human population structure on large genetic association studies**. *Nat Genet* 2004, **36**(5):512-517.
6. Li JZ, Vawter MP, Walsh DM, Tomita H, Evans SJ, Choudary PV, Lopez JF, Avelar A, Shokoohi V, Chung T, Mesarwi O, Jones EG, Watson SJ, Akil H, Bunney WE, Jr., Myers RM: **Systematic changes in gene expression in postmortem human brains associated with tissue pH and terminal medical conditions**. *Hum Mol Genet* 2004, **13**(6):609-616.
7. Tomita H, Vawter MP, Walsh DM, Evans SJ, Choudary PV, Li J, Overman KM, Atz ME, Myers RM, Jones EG, Watson SJ, Akil H, Bunney WE, Jr.: **Effect of agonal and postmortem factors on gene expression profile: quality control in microarray analyses of postmortem human brain**. *Biol Psychiatry* 2004, **55**(4):346-352.
8. Pritchard CC, Hsu L, Delrow J, Nelson PS: **Project normal: defining normal variance in mouse gene expression**. *Proc Natl Acad Sci U S A* 2001, **98**(23):13266-13271.
9. Prabakaran S, Swatton JE, Ryan MM, Huffaker SJ, Huang JT, Griffin JL, Wayland M, Freeman T, Dudbridge F, Lilley KS, Karp NA, Hester S, Tkachev D, Mimmack ML, Yolken RH, Webster MJ, Torrey EF, Bahn S: **Mitochondrial dysfunction in schizophrenia: evidence for compromised brain metabolism and oxidative stress**. *Mol Psychiatry* 2004, **9**(7):684-697, 643.
10. Sun X, Wang J, Tseng M, Young T: **Downregulation in components of the mitochondrial electron transport chain in the postmortem frontalcortex of subjects with bipolar disorder**. *J Psychiatry Neurosci* 2006, **31**(3):189-196.
11. Glatt SJ, Everall IP, Kremen WS, Corbeil J, Sasik R, Khanlou N, Han M, Liew CC, Tsuang MT: **Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 up-regulation in schizophrenia**. *Proc Natl Acad Sci U S A* 2005, **102**(43):15533-15538.
12. Vawter MP, Tomita H, Meng F, Bolstad B, Li J, Evans S, Choudary P, Atz M, Shao L, Neal C, Walsh DM, Burmeister M, Speed T, Myers R, Jones EG, Watson SJ, Akil H, Bunney WE: **Mitochondrial related gene expression changes are sensitive to agonal-pH state**. *Mol Psychiatry* 2006, **11**(7):615.
13. Website title [http://www.stanleyresearch.org/programs/brain_collection.asp]
14. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data**. *Genetics* 2000, **155**(2):945-959.

15. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P: **Association mapping in structured populations.** *Am J Hum Genet* 2000, **67**(1):170-181.
16. Tang H, Peng J, Wang P, Risch NJ: **Estimation of individual admixture: analytical and study design considerations.** *Genet Epidemiol* 2005, **28**(4):289-301.
17. Hinds DA, Stokowski RP, Patil N, Konvicka K, Kershenobich D, Cox DR, Ballinger DG: **Matching strategies for genetic association studies in structured populations.** *Am J Hum Genet* 2004, **74**(2):317-325.
18. Bamshad MJ, Wooding S, Watkins WS, Ostler CT, Batzer MA, Jorde LB: **Human population genetic structure and inference of group membership.** *Am J Hum Genet* 2003, **72**(3):578-589.
19. Erosheva E: **Grade of membership and latent structure models with application to disability survey data.** *PhD Thesis.* Pittsburgh: Carnegie Mellon University; 2002.
20. Hauser ER, Watanabe RM, Duren WL, Bass MP, Langefeld CD, Boehnke M: **Ordered subset analysis in genetic linkage mapping of complex traits.** *Genet Epidemiol* 2004, **27**(1):53-63.
21. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: **Principal components analysis corrects for stratification in genome-wide association studies.** *Nat Genet* 2006, **38**(8):904-909.
22. Setakis E, Stirnadel H, Balding DJ: **Logistic regression protects against population structure in genetic association studies.** *Genome Res* 2006, **16**(2):290-296.
23. 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):241-253.
24. Torrey EF, Webster M, Knable M, Johnston N, Yolken RH: **The stanley foundation brain collection and neuropathology consortium.** *Schizophr Res* 2000, **44**(2):151-155.
25. Tibshirani R, Hastie T, Narasimhan B, Chu G: **Diagnosis of multiple cancer types by shrunken centroids of gene expression.** *Proc Natl Acad Sci U S A* 2002, **99**(10):6567-6572.
26. Rosenberg NA, Li LM, Ward R, Pritchard JK: **Informativeness of genetic markers for inference of ancestry.** *Am J Hum Genet* 2003, **73**(6):1402-1422.
27. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: **Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.** *Science* 1999, **286**(5439):531-537.
28. Hosack DA, Dennis G, Jr., Sherman BT, Lane HC, Lempicki RA: **Identifying biological themes within lists of genes with EASE.** *Genome Biol* 2003, **4**(10):R70.
29. Burke WJ, O'Malley KL, Chung HD, Harmon SK, Miller JP, Berg L: **Effect of pre- and postmortem variables on specific mRNA levels in human brain.** *Brain Res Mol Brain Res* 1991, **11**(1):37-41.
30. Hardy JA, Wester P, Winblad B, Gezelius C, Bring G, Eriksson A: **The patients dying after long terminal phase have acidotic brains; implications for biochemical measurements on autopsy tissue.** *J Neural Transm* 1985, **61**(3-4):253-264.
31. Hynd MR, Lewohl JM, Scott HL, Dodd PR: **Biochemical and molecular studies using human autopsy brain tissue.** *J Neurochem* 2003, **85**(3):543-562.
32. 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.

33. 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.
34. Reiner AP, Ziv E, Lind DL, Nievergelt CM, Schork NJ, Cummings SR, Phong A, Burchard EG, Harris TB, Psaty BM, Kwok PY: **Population structure, admixture, and aging-related phenotypes in African American adults: the Cardiovascular Health Study**. *Am J Hum Genet* 2005, **76**(3):463-477.
35. Atz M, Walsh D, Cartagena P, Li J, Evans S, Choudary P, Overman K, Stein R, Tomita H, Potkin S, Myers R, Watson SJ, Jones EG, Akil H, Bunney WE, Jr., Vawter MP: **Methodological considerations for gene expression profiling of human brain**. *J Neurosci Methods* 2007.
36. Purcell S, Sham P: **Properties of structured association approaches to detecting population stratification**. *Hum Hered* 2004, **58**(2):93-107.
37. Satten GA, Flanders WD, Yang Q: **Accounting for unmeasured population substructure in case-control studies of genetic association using a novel latent-class model**. *Am J Hum Genet* 2001, **68**(2):466-477.
38. Reich DE, Goldstein DB: **Detecting association in a case-control study while correcting for population stratification**. *Genet Epidemiol* 2001, **20**(1):4-16.
39. Devlin B, Roeder K: **Genomic control for association studies**. *Biometrics* 1999, **55**(4):997-1004.
40. Bacanu SA, Devlin B, Roeder K: **The power of genomic control**. *Am J Hum Genet* 2000, **66**(6):1933-1944.
41. Collins-Schramm HE, Chima B, Operario DJ, Criswell LA, Seldin MF: **Markers informative for ancestry demonstrate consistent megabase-length linkage disequilibrium in the African American population**. *Hum Genet* 2003, **113**(3):211-219.
42. Stuart RO, Wachsman W, Berry CC, Wang-Rodriguez J, Wasserman L, Klacansky I, Masys D, Arden K, Goodison S, McClelland M, Wang Y, Sawyers A, Kalcheva I, Tarin D, Mercola D: **In silico dissection of cell-type-associated patterns of gene expression in prostate cancer**. *Proc Natl Acad Sci U S A* 2004, **101**(2):615-620.
43. Lu P, Nakorchevskiy A, Marcotte EM: **Expression deconvolution: a reinterpretation of DNA microarray data reveals dynamic changes in cell populations**. *Proc Natl Acad Sci U S A* 2003, **100**(18):10370-10375.
44. Venet D, Pecasse F, Maenhaut C, Bersini H: **Separation of samples into their constituents using gene expression data**. *Bioinformatics* 2001, **17 Suppl 1**:S279-287.
45. Ghosh D: **Mixture models for assessing differential expression in complex tissues using microarray data**. *Bioinformatics* 2004, **20**(11):1663-1669.
46. Lahdesmaki H, Shmulevich L, Dunmire V, Yli-Harja O, Zhang W: **In silico microdissection of microarray data from heterogeneous cell populations**. *BMC Bioinformatics* 2005, **6**:54.
47. Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG, McKeigue PM: **Control of confounding of genetic associations in stratified populations**. *Am J Hum Genet* 2003, **72**(6):1492-1504.
48. Evans SJ, Choudary PV, Neal CR, Li JZ, Vawter MP, Tomita H, Lopez JF, Thompson RC, Meng F, Stead JD, Walsh DM, Myers RM, Bunney WE, Watson SJ, Jones EG, Akil H: **Dysregulation of the fibroblast growth factor system in major depression**. *Proc Natl Acad Sci U S A* 2004, **101**(43):15506-15511.
49. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP: **Summaries of Affymetrix GeneChip probe level data**. *Nucleic Acids Res* 2003, **31**(4):e15.
50. Website title [http://www.ncbi.nlm.nih.gov/geo/]

51. Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, Bunney WE, Myers RM, Speed TP, Akil H, Watson SJ, Meng F: **Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data**. *Nucleic Acids Res* 2005, **33**(20):e175.
52. Website title [http://brainarray.mhri.med.umich.edu/Brainarray/Database/CustomCDF/CDF_download_v3.asp]

Figure legends

Figure 1. Technical "block" effects due to cohort, chip-type, and laboratory (site)

a. A color-coded correlation matrix among 201 AnCg samples. Red indicates high correlation between pairs of samples; blue indicates low correlation. Samples were ordered by technical batches, which in turn were defined by Cohort, Chip Type, and Site, as indicated below the heatmap. Throughout this paper the sample order in correlation heatmaps is the same from left to right as from bottom to top. Pearson's correlations were calculated by using all U133A transcripts. Rectangular "blocks" of high correlation, indicated in red, correspond to samples that are highly similar to each other in gene expression patterns. These block-block partitions coincide with the natural boundaries of experimental batches as indicated below the heatmap. D1 through D4 indicate Depression Cohorts 1 to 4, while S1 and S2 indicate Schizophrenia Cohort 1 and 2.

b. The same correlation matrix after removing the block effect by median centering each block. The diagonal line of high values are self-self correlations, but the two off-diagonal lines of relatively high correlation values are between replicate chips for the same samples ran at two sites. The sample order and block partitions are the same as in 1a.

Figure 2. From discrete classification to grades of membership

a. The same correlation matrix as in Figure 1b but with samples re-grouped so that the Type 1 samples are on the lower left side, showing that samples can be classified into two main types of expression patterns.

b. The same samples can be subdivided into three classes, and re-ordered accordingly in the correlation heatmap, still showing reasonably strong distinction between the three classes.

c. The same heatmap re-ordered by the first Principal Component scores, resulting in a gradual progression from one prototype to the other.

Figure 3. Informativeness of genes for agonal stress

The same heatmap as shown in 2c but with the correlations calculated by using the 25% strongest Type 1-Type 2 genes (upper left), the next 25% (upper right), the third 25% (lower left), and the last 25% of genes (lower right). The genes were ranked by comparing 40 samples (~20% of the total of 201 samples) on the lower left side in 2c against 40 samples on the upper right side in a Student t test.

Figure 4. Correlation of ASR and pH values

Median ASR values across 2-3 sites and six brain regions for 126 subjects versus their brain pH values. Samples indicated in red and black were in general agreement between pH and ASR. The eight samples in blue were previously removed from analysis due to either low pH or some clinical complications (AFS is non-zero), but had normal ASR values and likely had not experienced significant agonal stress. The eight samples shown in green were previously

included but had low ASR values, and should better be removed in order to reduce sample heterogeneity and possible case-control imbalance.

Figure 5. Comparing pH-matching and ASR-matching with an example

a. ASR and pH values for samples used in a test case, where a group of pH-matched major depression samples (in blue) were compared to controls (in red), while a second group of ASR-matched major depression samples (in green) were compared to the same controls.

b. In an analysis of enrichment of Gene Ontology and KEGG pathway terms for the top genes, the pH-matched comparison still revealed significant effects for gene families and pathways known to be affected by the pH effect, while the same effect was much attenuated in the ASR-matched comparison.

Figure 6. Region-region difference of ASRs

A color-coded table of (from left to right) pH values (red is higher pH), AFS scores (0 complication in red, AFS=1 in light blue), and eight series of ASR values (red-higher ASR) for 68 subjects satisfying pH>6.6 and at most two missing values out of eight ASR series. The eight ASR columns represented four brain regions, each measured in two sites. Tan-colored elements indicate missing data because the samples were not run.

Additional files

Additional file 1
File format: PDF
Title: Technical batch effects
Description:
a. The color-coded correlation matrix among the same 201 AnCg samples as in Figure 1a, but calculated by using only the 68 control probe sets targeting spiked-in *E. coli* transcripts. Sample order was the same as in Figure 1a.

b. Similar correlation heatmap based on 700-gene Illumina data. Samples were ordered by Cohort.

c. Correlation matrix among 24 chips that represented eight samples ran three times each. The first time (samples 1-8 counting from lower left) was on U133A chips; the next two times were on U133_Plus2 chips. Shown are results after median centering of the chip-type blocks. The off-diagonal lines of high similarity are for the three replicate chips of the same samples, indicating that sample-sample differences were reproducibly measured across two chip types after removing the block effect.

Additional file 2
File format: PDF
Title: Heatmap of normalized expression levels for "top 25%" genes in 201 AnCg samples
Description:
Shown are log-transformed, normalized expression levels of 3184 transcripts across 201 AnCg samples. These genes have the highest 25% of Type 1- Type 2 absolute t scores, and 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., each gene's "loading" in the first eigenvector), whereas the samples are ordered from top to bottom by their first principal component scores. The color scale is for log2 values of -3 (8-fold lower expression) to 3 (8-fold higher), with some values in the upper left corner being greater than 3. These saturated values were shown in brown.

Additional file 3
File format: CDT
Title: Expression levels of 3841 most changed genes
Description:
Normalized expression levels for 3841 genes in 201 samples as shown in Additional file 2. Samples and genes are ordered by PC results, and formatted in Treeview format.

Additional file 4
File format: PDF
Title: Cross-validation errors in classifying samples
Description:
Number of cross-validation errors as a function of number of genes used in the nearest Shrunken Centroid classification [25] where the 201 AnCg samples were analyzed, and the Type 1-Type 2 designations were taken as known. Lower panel showed the errors for Type 1 and Type 2 samples separately.