genes are differentially expressed between individuals of European and African ancestry[2]. To rule out the possibility that this difference is related to the larger sample size of Spielman *et al.*[1] and is thus not simply a power issue, we randomly sampled eight CEU and eight ASN individuals (corresponding to the sample size of Storey *et al.*[2]) 1,000 times, and for each sample we estimated the proportion of differentially expressed genes as described above. The average proportion was 43% (standard deviation (s.d.) = 8%), demonstrating that differences in power between our study and Spielman *et al.*[1] do not fully account for differences in the estimated proportion of differentially expressed genes among human populations.

A possible explanation for the pervasive signature of differential expression observed in Spielman *et al.* is a systematic bias introduced during sample preparation or microarray expression measurements. The authors clearly recognize the importance of controlling for systematic confounding variables, as they state that "the growing and processing of the HapMap cell lines was randomized by population group to eliminate batch effects that may contribute to apparent population differences in gene expression." In addition to sample processing, it is widely known that technical variation can also be introduced through batch-to-batch variation in microarray manufacturing and through day-to-day laboratory conditions under which hybridization is performed[5]. To explore these issues in more detail, we downloaded the raw CEL files from Gene Expression Omnibus (GSE5859) and extracted from the header line the date on which the file was created. Interestingly, the arrays used to measure expression for the CEU individuals were primarily processed from 2003 to 2004, whereas the arrays used to measure expression for the ASN individuals were all processed in 2005–2006.

We tested for differential expression with respect to the year in which the microarrays were processed and found that at least 94% of genes are estimated to be differentially expressed (**Fig. 1b**). Typically, one would take these batch effects into account before performing any differential expression analyses. When we used a standard method to do so[6], we find no evidence for differential expression between populations (**Fig. 1c**), which is not surprising, given that microarray batch effects seem to be completely confounded with population effects. Obviously, these findings do not mean that all differentially expressed genes in Spielman *et al.*[1] are due to batch effects; rather, the source of population differences in expression cannot be uniquely attributed to biological causes. To gain insight into the magnitude of the batch effects in Spielman *et al.*[1], we tested for differential expression within the CEU sample with respect to the year of processing. Strikingly, 79% of genes among CEU individuals are estimated to be differentially expressed between processing years (**Fig. 1d**). Collectively, these results suggest that the expression data analyzed in Spielman *et al.*[1] possess systematic and uncorrectable bias, raising concerns about the accuracy of their reported results.

The genotype-phenotype correlations made with the data set of Spielman *et al.* should also be viewed with caution. Specifically, because batch effect appears to be the major source of differential expression, any marker with allele frequency differences among batches is therefore also vulnerable to confounding when testing for genotype-phenotype correlations. Even though our primary purpose here was to explore potential explanations for the large number of genes estimated to be differentially expressed between the CEU and ASN samples, the consequences of microarray batch effects on the gene expression association results of Spielman *et al.*[1] also warrant further investigation.

In summary, characterizing patterns of gene expression variation within and among human populations is an important and interesting problem. However, it is critical that experimental design and statistical analyses be carefully thought out and implemented in order for accurate conclusions to be drawn. In particular, components of variation from both measured and unmeasured variables must be taken into account, for example, by balancing or randomizing the study design with respect to sex, time of sample preparation and processing and microarray batch.

Joshua M Akey[1], Shameek Biswas[1], Jeffrey T Leek[2] & John D Storey[1,2]

[1]Department of Genome Sciences and [2]Department of Biostatistics, University of Washington, 1705 NE Pacific St., Seattle, Washington 98195, USA.
e-mail: akeyj@u.washington.edu or jstorey@u.washington.edu

COMPETING INTERESTS STATEMENT
The authors declare no competing financial interests.

1. Spielman, R.S. *et al. Nat. Genet.* **39**, 226–231 (2007).
2. Storey, J.D. *et al. Am. J. Hum. Genet.* **80**, 502–509 (2007).
3. Storey, J.D. *J. R. Stat. Soc. Ser. B* **64**, 479–498 (2002).
4. Storey, J.D. *et al. Proc. Natl. Acad. Sci. USA* **100**, 9440–9445 (2003).
5. Churchill, G.A. *Nat. Genet.* **32** (Suppl.), 490–495 (2002).
6. Jin, W. *et al. Nat. Genet.* **29**, 389–395 (2001).

Spielman and Cheung reply:

In our paper[1], we found that mean gene expression differed significantly between European-derived and Asian-derived populations for approximately 25% of 4,197 genes tested. For the expression phenotypes with the strongest evidence of polymorphic *cis* determinants, the differences in mean expression were largely explained by differences in the frequency of specific polymorphic variants.

Akey *et al.*[2] reanalyzed our data and asserted that their findings "suggest that the expression data of Spielman *et al.* possess systematic and uncorrectable bias, raising concerns about the accuracy of [the reported results of Spielman *et al.*]." To reach this conclusion, Akey *et al.*[2] restricted their analysis to just one part of our results[1]. Here, we first comment on their analysis and conclusions. We then describe results from a published study by others[3] that confirms our findings. Finally, we show how other aspects of our paper contradict the conclusions of Akey *et al.*[2] and support the "accuracy of [our] reported results."

Akey *et al.*[2] obtained our earlier data[1] from the Gene Expression Omnibus (GEO). They found that the expression arrays used for the HapMap CEU sample (of European ancestry) were processed in 2003–2004 (except for four individuals). In contrast, the arrays used for the HapMap CHB+JPT samples (their 'ASN' group, of Chinese and Japanese ancestry) were processed in 2005–2006. Akey *et al.* point out, correctly, that "microarray batch effects appear to be completely confounded with population effects."

In our paper, we wrote (in the Methods section), "The growing and processing of the HapMap cell lines was randomized by population group to eliminate batch effects that may contribute to apparent population differences in gene expression." Because of the different dates of processing described in the previous paragraph, this was not actually done. (Of course, we did not intend to mislead. Other CEPH cell lines—not HapMap CEU—were grown in the same batch as CHB+JPT, which gave rise to our

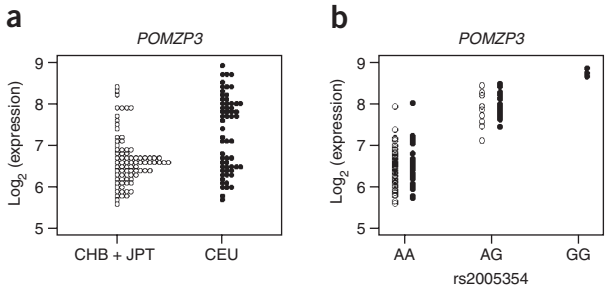

Figure 1 The population difference in expression of *POMZP3* is accounted for by the allele frequency difference at the very closely linked SNP rs2005354. Each circle represents an individual. (**a**) Population distribution of expression level in the two groups. (**b**) Expression level in individuals who have the indicated genotype at the associated SNP. This figure is a reproduction of Figure 3 of Spielman *et al.*[1].

confusion.) We regret our incorrect statement that randomization was carried out, and we appreciate this chance to correct the error.

Nevertheless, data obtained in an independent study[3] strongly support our conclusions for the most extreme population differences. Stranger *et al.*[3] carried out gene expression analysis on cells independently prepared from the same HapMap samples we studied, and with different microarray technology. In Table 1 of our paper, we listed 35 genes whose mean expression level differed between CEU and CHB+JPT samples by a factor of 2 or more (and with $P < 0.05$ after multiple testing correction). We looked for those 35 genes in the data of Stranger *et al.*[3] Of the 35 genes, 32 were on their arrays. Among these 32, 30 were also significantly different between CEU and CHB+JPT in their analysis, and 29 of the 30 differed in the same direction as in our results. Thus, a large proportion (29 of 32) of the genes in Table 1 were also found to differ by Stranger *et al.*[3] In addition, we tested for similarity of the actual expression differences between populations. The rank correlation of the 32 differences was 0.71 ($P < 10^{-5}$). Thus, the results from two independent studies are very similar, despite independent cell culture and expression analysis. Therefore, our results for these genes are unlikely to be an artifact of batch differences.

Akey *et al.*[2] focus mainly on the confounding and the resulting possibility that the many marked population differences are due to effects associated with year of processing (the batch effects). However, by itself the confounding does not imply, nor even suggest, that there is "systematic and uncorrectable bias," the central criticism of Akey *et al.*[2]. Akey *et al.*[2] themselves say, "Obviously, these findings do not mean that all differentially expressed genes in Spielman *et al.* are due to batch effects…" So what do they mean? The confounding means only that the possible batch and population effects cannot be separated in these data simply by comparing population means. And even if the confounding were eliminated by repeating the experiment in a completely randomized fashion, some other kind of evidence would be required for a conclusion about the genetic basis of the expression differences between populations.

That other evidence is provided in our paper itself. Instead of looking only at differences in population means, the ideal approach for genetic analysis would be to find genotypes that determine the level of gene expression and then compare the effects of these genotypes in the two populations. Although we often don't know the actual genetic (allelic) determinant of expression differences, SNPs that are closely linked to an expressed gene are sometimes very good proxies for the unknown determinants[3–5]. Using these *cis*-linked SNPs, we can then dissect the overall difference in population expression into its individual genotypic components.

This is what we did in Figure 3 of our paper[1] (reproduced here as **Fig. 1**). For the example shown (*POMZP3*), the mean expression is significantly higher in CEU ($\log_2 = 7.3$) than in CHB+JPT ($\log_2 = 6.6$) (**Fig. 1a**). **Figure 1b** shows the distribution of expression separately for each genotype of the associated SNP (rs2005354). If there were a batch effect of the kind described by Akey *et al.*, it would apply to all the individuals in each population. Then even within the same genotype, CEU individuals would have higher expression than CHB+JPT. But this is not the case. The results in **Figure 1** make it clear that the population difference in expression is not due to a batch effect; instead, it is due to the population difference in genotype frequencies for the associated determinant. The G allele of rs2005354 is associated with higher expression of *POMZP3* in both populations, but the G allele frequency is 0.28 in CEU and 0.06 in CHB+JPT. Therefore, genotypes with the G allele are more frequent in the CEU population, accounting for the higher mean expression level of *POMZP3* in the CEU individuals. The data for *POMZP3* are representative of many others. Table 2 in our paper[1] summarizes the results for ten more genes. Nine of these show the same effect, with genotype frequencies accounting for most of the genetic differences in mean expression between the populations.

We have shown that the confounding described by Akey *et al.* does not imply "systematic and uncorrectable bias." Furthermore, our findings are strongly supported by other investigators. Finally, we have described the genetic analysis[1] that shows how allele frequency differences account for the population differences in expression.

Richard S Spielman[1] and Vivian G Cheung[1–3]

[1]Department of Genetics and [2]Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. [3]The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA.
e-mail: vcheung@mail.med.upenn.edu or spielman@pobox.upenn.edu

1. Spielman, R.S. *et al. Nat. Genet.* **39**, 226–231 (2007).
2. Akey, J.M., Biswas, S., Leek, J.T., & Storey, J.D. *Nature Genet.* **39**, 807–808 (2007).
3. Stranger, B.E. *et al. Science* **315**, 848–853 (2007).
4. Morley, M. *et al. Nature* **430**, 743–747 (2004).
5. Cheung, V.G. *et al. Nature* **437**, 1365–1369 (2005).