Common genetic variants account for differences in gene expression among ethnic groups

Richard S Spielman¹, Laurel A Bastone², Joshua T Burdick³, Michael Morley³, Warren J Ewens⁴ & Vivian G Cheung¹,³,⁵

Variation in DNA sequence contributes to individual differences in quantitative traits, but in humans the specific sequence variants are known for very few traits. We characterized variation in gene expression in cells from individuals belonging to three major population groups. This quantitative phenotype differs significantly between European-derived and Asian-derived populations for 1,097 of 4,197 genes tested. For the phenotypes with the strongest evidence of cis determinants, most of the variation is due to allele frequency differences at cis-linked regulators. The results show that specific genetic variation among populations contributes appreciably to differences in gene expression phenotypes.

Populations differ in prevalence of many complex genetic diseases, such as diabetes and cardiovascular disease. As some of these are probably influenced by the level of gene expression, our results suggest that allele frequency differences at regulatory polymorphisms also account for some population differences in prevalence of complex diseases.

genetic diseases. The marked population differences in prevalence of these qualitative phenotypes (such as cystic fibrosis⁶ and Tay-Sachs disease⁷) are entirely due to differences in frequencies of the mutant alleles. However, genetic differences among populations in quantitative phenotypes are potentially just as important functionally.

Here we extend the comparative genetic analysis of population differences from qualitative phenotypes to a particular quantitative phenotype, the expression level of genes. The choice of gene expression as a phenotype provides a large set of comparable traits, all measured at the same time in each individual. Our goals are to determine what proportion of gene expression phenotypes differs significantly between populations and to what extent the phenotypic differences are attributable to specific genetic polymorphisms. We find that at least 25% of the gene expression phenotypes differ significantly between the major populations studied, and specific genetic variation (in allele frequency) accounts for the difference in the most significant instances among the phenotypes that are cis regulated.

We measured the expression of genes in Epstein-Barr virus (EBV)-

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics
Confounding of population and processing time
On the design and analysis of gene expression studies in human populations

To the Editor:
In a recent Nature Genetics Letter entitled “Common genetic variants account for differences in gene expression among ethnic groups,” Spielman et al. estimate the number of genes differentially expressed between individuals of European (CEU) and Asian (ASN) ancestry and suggest that these differences can be accounted for by measured genetic variants. We recently performed a similar study comparing differences in gene expression among individuals of European and Yoruban ancestry. Given the scientific, medical and societal implications of this research area, it is important for the scientific community to carefully revisit and critically evaluate the conclusions of such studies. To this end, we have reanalyzed the data in Spielman et al. to provide a common basis for comparison with our study. In doing so, we found that important issues arise about the accuracy of their results.

The authors categorized genes as differentially expressed if they had P values < 10^{-3}, corresponding to a Sidak corrected P value of < 0.05 for multiple hypothesis tests. At this significance threshold, they report that approximately 26% of genes are differentially expressed between the CEU and ASN samples (ASN denotes the combined HapMap Beijing Chinese (CHB) and Japanese (JPT) HapMap individuals). As a Sidak correction is similar to a Bonferroni correction, the proportion of genes found to be significant is a conservative estimate of the true overall proportion of differentially expressed genes. A more widely used and less conservatively biased approach is to analyze the complete distribution of P values, which provides a lower bound estimate of the proportion of truly differentially expressed genes. Applying this methodology to the distribution of P values obtained by t tests on genes expressed in lymphoblastoid cell lines as defined in Spielman et al., we estimate that at least 78% of these genes are differentially expressed between the CEU and ASN samples.

(Fig. 1a). Estimates of this proportion were nearly identical regardless of whether P values were obtained from standard t tests, permutation t tests, bootstrap t tests or nonparametric Wilcoxon rank-sum tests (data not shown).

It seems implausible that as many as 78% of genes are differentially expressed between the CEU and ASN samples. For example, based on the complete distribution of P values, we have recently estimated that approximately 17% of
A decent SNP

Plates ordered by date of first run

A not so decent SNP