NEWS AND COMMENTARY

. Genetic differences between populations

RB O'Hara

Heredity (2007) 98, 245–246. doi:10.1038/sj.hdy.6800970; published online 4 April 2007

t has often been observed that people are different. Indeed, some observers have gone further to suggest that this diversity exists between people from different parts of the world, or of different ethnic groups, and it is hereditary. This latter observation has led to a certain amount of contention over time. As a result, many geneticists have been wary about asserting that such differences exist.

Of course, this does not mean that no differences exist. Evidence for variation is presented in a recent article by Spielman et al. (2007). They looked at gene expression in cell lines derived from two populations: one of European origin (actually sampled in UT, USA), and the other an amalgam of two very similar Asian populations, from China and Japan. When they looked at the expression of 4197 genes, they found that 1097 of them showed significant differences between the two populations even when conservative statistical tests were used. Most of the variation between the two populations was small, with less than a twofold difference in mean expression levels between populations in all but 35 genes.

It may be surprising that over a quarter of the genes show differences between populations, but can these be explained away? It seems difficult to do this statistically: the P-values were corrected for the many tests that were carried out, and the authors carried out a couple of other tests, getting similar results.

It could also be that the differences are artefacts of the sampling. However, the authors checked that the results were replicable by examining expression in 24 individuals of Han Chinese descent from Los Angeles. Of the 35 genes that showed different expression with the original samples, only one showed a difference from the Asian population, but 32 showed a difference from the European-derived population. So, the results are not a quirk of the populations sampled, they can be replicated.

Another explanation is that the differences are an environmental effect. Well,

the authors went on to ask whether they could find loci that would explain the variation in expression. They carried out a genome-wide association analysis with about 2 million single-nucleotide polymorphisms, and for the two populations independently, tested whether they could explain the variation in expression levels. From these tests, 104 markers showed a significant association in the European-derived population, and 89 in the Asian population (about 55 significant markers would be expected by chance as type I errors). Of these markers, 11 were common to both populations. All were cis-regulators (i.e. they were close to the expressed gene). Although there was evidence for transregulation, this was not consistent between populations. Because the authors were being conservative throughout, it seems reasonable to expect that some of these are actual genuine effects: this will lead to the inevitable call that more research is required.

So, there are differences between populations of cell lines derived from different human populations, and some (almost certainly more than the 11 the authors found) are under genetic control. Perhaps, we can use the standard complaint about laboratory studies, and argue that the populations came from cell lines that have been cultured in the laboratory, rather than directly from humans. But it is difficult to see why so many genes would react differently, unless they were all (or almost all) regulated in the same way. This seems unlikely, as the genes cover a wide range of activities.

Thus, while the exact numbers of differences could be debated, the data at least suggest that there should be considerable variation in actual humans. But, does this matter? One problem with establishing the importance of microarray results is that they only tell about gene expression, not about the physiological effects of the gene. Earlier studies on the dynamics of metabolic pathways have shown that fluxes through the pathways may be relatively insensitive to changes in the concentration of many of the enzymes in a pathway (e.g. Fell, 1992). This theory provides one explanation for dominance (Kacser and Burns, 1981): alleles that result in a lack of function of an enzyme are often recessive, which implies that a reduction in half of the concentration of the enzyme has no observable effect on the phenotype. Most of the differences in expression in the study were smaller than the twofold difference that would be seen in dominance, so many of the differences in expression may have little or no effect on physiology. Of course, many effects would still remain, either because the gene is an enzyme that is rate limiting, or because it codes for a protein with another function, where the concentration is more important. Overall, it is not clear how the variation in gene expression relates to phenotype, and fitness (Townsend et al., 2003).

In this study, we can see some of both the strengths and weaknesses of microarray studies. One of their strengths is that they can produce a large amount of data, by screening a large number of genes. The weakness is that these data have to be interpreted, which can be difficult when the significance of differences between genes has to be assessed. Finding out what all this means for real humans, for their biochemistry, their physiology and their health, will require more detailed investigation of the genes that differ, and the roles they play in the body. This will require hard work in the laboratory, going beyond expression to looking at how it translates into variation between people. Perhaps, Heredity is not the best place to declare this, but there is more to life than genes.

Dr RB O'Hara is at the Department of Mathematics and Statistics, University of Helsinki, PO Box 68 (Gusfaf Hällströmin katu 2b), FIN-00014, Finland.

e-mail: bob.ohara@helsinki.fi;

http://www.RNI.Helsinki.FI/~boh/

- Fell DA (1992). Metabolic control analysis: a survey of its theoretical and experimental development. Biochem J 286: 313-330.
- Kacser H, Burns J (1981). The molecular basis of dominance. Genetics 97: 639-666.
- Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG (2007). Common genetic variants account for differences in gene expression among ethnic groups. Nat Genet 39: 226-231.
- Townsend JP, Cavalieri D, Hartl DL (2003). Population genetic variation in genome-wide gene expression. *Mol Biol Evol* **20**: 955–963.

Editor's suggested reading

Hoti F, Sillanpaa MJ (2006). Bayesian mapping of genotype \times expression interactions in quantitative and qualitative traits. Heredity 97: 4 - 18

- npg 246
- Li J, Ji L (2005). Adjusting multiple testing in multilocus analyses using the eigen values of a correlation matrix. *Heredity* 95: 221–227.
 Oldham MC, Geschwind DH (2006). Comparative genomics: grasping human transcriptome evo-

lution: what does it all mean? *Heredity* **96**: 339–340.

Siju SH, Borevitz JO (2006). The next generation of microarray research: applications in evolu-tionary and ecological genomics. *Heredity*

advance online publication, 8 November 2006; doi:10.1038/sj.hdy.6800916. Svingen T, Tonissen KF (2006). Hox transcription factors and their elusive mammalian gene targets. *Heredity* **97**: 88–96.