MUSIC

Let the people sing

Is Joe Public really a flop when it comes to singing? Simone Dalla Bella *et al.* (*J. Acoust. Soc. Am.* **121,** 1182–1189; 2006) went looking for the answer. They persuaded 42 passers-by in a Montreal park to sing *Gens du pays* — an anthem of the Quebec sovereignty movement, commonly sung at festive occasions in the province — on the pretext that it was one of the experimenters' birthday. They then subjected the recordings, together with those of a further 20 non-musicians and 5 professional singers recorded in the laboratory, to an acoustic analysis.

This quantified errors in timing and in pitch interval (misjudging the jumps between notes) and contour (going the wrong way on the scale). Pitch and timing consistency were assessed by comparing variations in pitch and deviations from the prescribed tempo (*rubato*) in a repeated phrase in the song's chorus.

Unsurprisingly, the professional participants — among them Gilles Vigneault, the composer and original vocalist of *Gens du pays* — scored better. But they also tended to sing less hurriedly. When, in a second test, the non-musicians were told to slow down, most sang just as accurately as the paid singers.

You could conclude that the ability to sing in tune is a universal human trait. Well, almost — the study uncovered two subjects who, even when singing at the slow tempo, went wildly out of tune.

N. CLEMENTS/TAXI/GETTY

They should perhaps confine their singing to the bathroom.
Richard Webb

SNP sets consisting only of variants within genes[3,4], GWA studies aspire to survey all variation in the genome, including non-coding regions. Exhaustive surveys are not yet feasible, but genotyping a subset of the known variable sites may suffice.

Locations in the genome that are separated by a small number of base pairs are often in 'linkage disequilibrium': that is, there is a substantial association between the variants at the two loci. This phenomenon allows us to survey variation across the genome fairly precisely, simply by genotyping a subset of polymorphic loci. GWA studies rely on the assumption that linkage disequilibrium enables one SNP to act as a marker for association to other sequence variants in that region. The GWA studies carried out thus far differ in terms of the number and criteria for selection of the genotyped SNPs: some use SNPs chosen to be evenly physically spaced[5–7], whereas others[8] choose SNPs to maximize the detection of linkage disequilibrium, based on data from the International HapMap Project[9]. Sladek *et al.* used a marker set based on HapMap linkage-disequilibrium data, supplemented by a gene-centric SNP set; their combined marker set (about 400,000 SNPs) provides the highest-resolution survey of genomic variation of any GWA study so far.

In deciding how many individuals to genotype, recent studies follow either a one-stage[8] or a two-stage[5–7] design: in the first case, statistical significance is sought within the genotyped sample (and other samples may be used for replication); in the second, SNPs passing a loose significance threshold in the GWA study are genotyped in a follow-up sample, to seek significance. The choice of study design usually rests on estimates of the power to detect effects of a given magnitude that is considered realistic for the trait in question.

Sladek *et al.*[1] adopted a two-stage design: in the first stage, they conducted the GWA study in a total sample of about 1,400 cases and controls (the largest sample in a GWA study so far). In the second stage, they genotyped the SNPs showing evidence of association in the GWA study in a new sample (about 5,500 total cases and controls) to test for significant associations. The results now published refer only to the follow-up of the most promising SNPs from the first stage: a more comprehensive second stage is still under way.

Sladek *et al.* identify strong associations for three novel loci (one detected in the linkage-disequilibrium-based marker set and two detected in both marker sets). Although more loci may emerge from the complete two-stage analysis, publication of these initial results provides the opportunity for swift replication (or not) by other research groups, using independent samples, as exemplified by the case of *TCF7L2*. Its association with type 2 diabetes was reported last year[2], and has already been replicated in at least 20 independent studies. In one- or two-stage studies, care must be taken that the specific definition of disease adopted in the follow-up (or replication) samples is comparable to the definition used in the original GWA samples. In this respect it is noteworthy, but of unclear significance, that Sladek *et al.* used more stringent inclusion criteria in the GWA sample than in the follow-up sample.

The identification of a few significant disease associations represents only one outcome of Sladek and colleagues' study. GWA studies should be evaluated primarily from an epidemiological standpoint, focused not just on what new disease-susceptibility genes they propose, but on how they advance our understanding of the composition of genetic risk in the population. Sladek *et al.* take a first step towards such understanding, presenting an evaluation of what proportion of the disease cases can be attributed to variation in the loci they identify as significant in their second-stage analysis. As several additional GWA studies of type 2 diabetes will shortly report their results, we may soon be able to estimate the number — and location in the genome — of the genetic variants that are the main contributors to diabetes susceptibility, at least in some populations.

The results of the first GWA studies may also reveal the degree of genome coverage provided by the chosen SNP panels. Reassuringly, both of the linkage-disequilibrium-based studies reported so far[1,8] were able to replicate one known locus. On the other hand, the most significant new locus in both cases was identified by only a single SNP, suggesting that even the dense marker sets employed in these studies provide insufficient coverage for detection of all important loci[10]. In certain populations that are of recent origin and that have remained isolated, linkage disequilibrium is more extensive than in the populations used in the GWA studies so far[11]; the first GWA surveys in such groups will be watched closely for evidence that they permit more complete coverage using comparable marker sets. Similarly, the results of GWA studies, now under way, that are using even larger numbers of SNPs, are much anticipated.

A final observation is that for both of the linkage-disequilibrium-based studies reported so far (for type 2 diabetes[1] and inflammatory bowel disease[8]), the most significant novel associations were to a variant predicted to alter the protein product encoded by a gene (termed a non-synonymous coding SNP), and thus possibly to have a strong functional effect. Furthermore, in both cases the disease is associated with the more common variant at these loci, suggesting that the less common variant may offer protection against developing the disease. That diseases are associated with such common non-synonymous SNPs suggests that these variants may have offered an evolutionary advantage in previous environments. Clearly, one should not generalize from a sample size of two. Nevertheless, these findings underscore how GWA studies may not only deliver 'new' genes[10], but permit advances in our understanding of how human evolution has 'made' the diseases that are common today. ■

Nelson B. Freimer is in the UCLA Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, 695 Charles Young Drive South, Los Angeles, California 90095-1761, USA.
e-mail: nfreimer@mednet.ucla.edu
Chiara Sabatti is in the UCLA Departments of Human Genetics and Statistics, 695 Charles