

Population Structure

It is very easy to be misled by GWAS if there is hidden population structure in the sample (Figure 6.5). As a trivial and ridiculous example, suppose you were looking for genes that influence religious affiliation in a mixed sample of Americans of Irish and Iranian descent. Any alleles that are at higher frequencies in the Irish-Americans will naturally tend to be more common in Catholics than in Muslims, and hence will give a positive association test result. However, this does not mean that they cause the difference in religious affiliation. They are simply correlated with it because of population structure and environmental and cultural effects. Wherever there is a difference in allele frequency as well as in the prevalence of a disease or

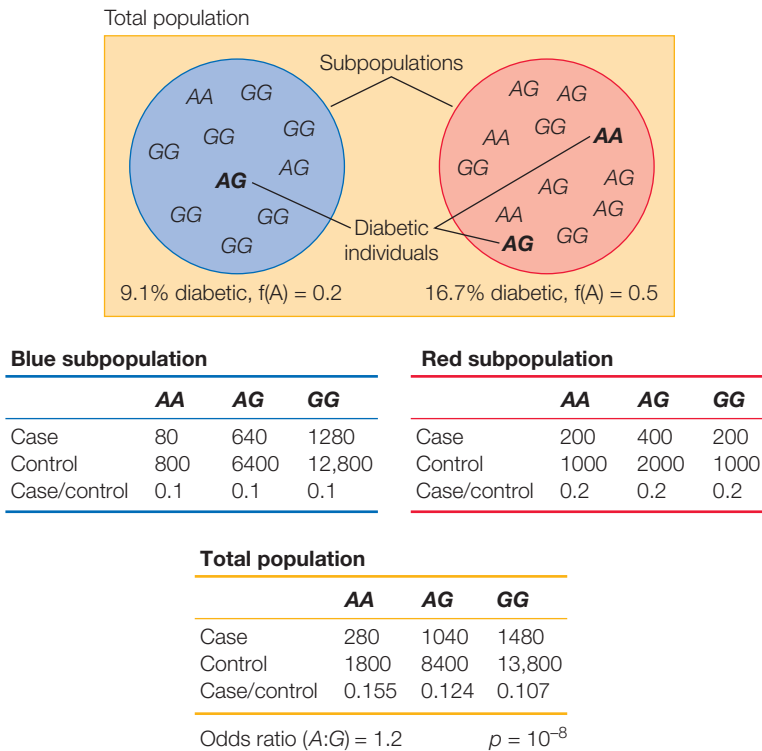


Figure 6.5 Population Structure. If a population consists of two subpopulations, which may be unrecognized, then despite the absence of any association in either subpopulation, there may appear to be a spurious correlation between genotype and disease risk. In this hypothetical example, 9.1% of the individuals in the blue population have diabetes irrespective of genotype, and there is a minor (A) allele frequency of 0.2; in the red population, 16.7% of the individuals have diabetes, and the A allele has a frequency of 0.5. In the total population (yellow), the A alleles are more likely to be observed in people with diabetes, which gives an odds ratio of 1.2 and genome-wide significance, yet the odds ratios in the two sub-populations are 1.0. Note that calculations refer to the tables; the image at the top is illustrative only.