Principal component analysis of genetic data

David Reich, Alkes L Price & Nick Patterson

Principal component analysis (PCA) has been a useful tool for analysis of genetic data, particularly in studies of human migration. A new study finds evidence that the observed geographic gradients, traditionally thought to represent major historical migrations, may in fact have other interpretations.

Principal component analysis (PCA) has been used for several decades to study human population migrations, resulting in remarkable inferences about history. On page 646 of this issue, John Novembre and Matthew Stephens[1] show that the geographic gradients that emerge when PCA is applied to genetic data—and that are sometimes interpreted as highly suggestive of major historical migrations—can also have other explanations. We suggest guidelines for scientists interested in using PCA in genetic