Genome **Biology**

REVIEW

Mapping of disease-associated variants in admixed populations

Daniel Shriner*, Adebowale Adeyemo, Edward Ramos, Guanjie Chen and Charles N Rotimi*

Abstract

A more refined estimation of ancestry would benefit admixture mapping and association mapping, making disease loci identification in admixed populations more powerful.

A major goal of human genetics is to identify genetic variation causally related to either Mendelian or complex diseases. More broadly, a fundamental goal of genetics is to describe the genetic architecture underlying any trait of interest. Most candidate gene studies, linkage studies, or genome-wide association studies to date have focused on European populations, for which large samples of ancestrally homogeneous individuals from relatively homogeneous environments have been established. For example, approximately 90% of genome-wide association studies have been performed using samples of individuals with European ancestry [1]. However, expanding human genetic studies to include diverse worldwide populations is needed to: (i) identify novel loci absent or not readily identifiable in European populations due to low allele frequencies and the resulting low statistical power; (ii) establish the extent to which findings from studies of European populations generalize or transfer to non-European populations; and (iii) study diseases or traits, such as podoconiosis and human African trypano-somiasis (sleeping sickness), present in non-European populations only [2-4].

In this article, we highlight the special value of admixed populations in disease mapping studies. Admixed populations are not ancestrally homogeneous but rather are populations with ancestry from more than one parental population. Admixed populations that have successfully contributed to the mapping of susceptibility
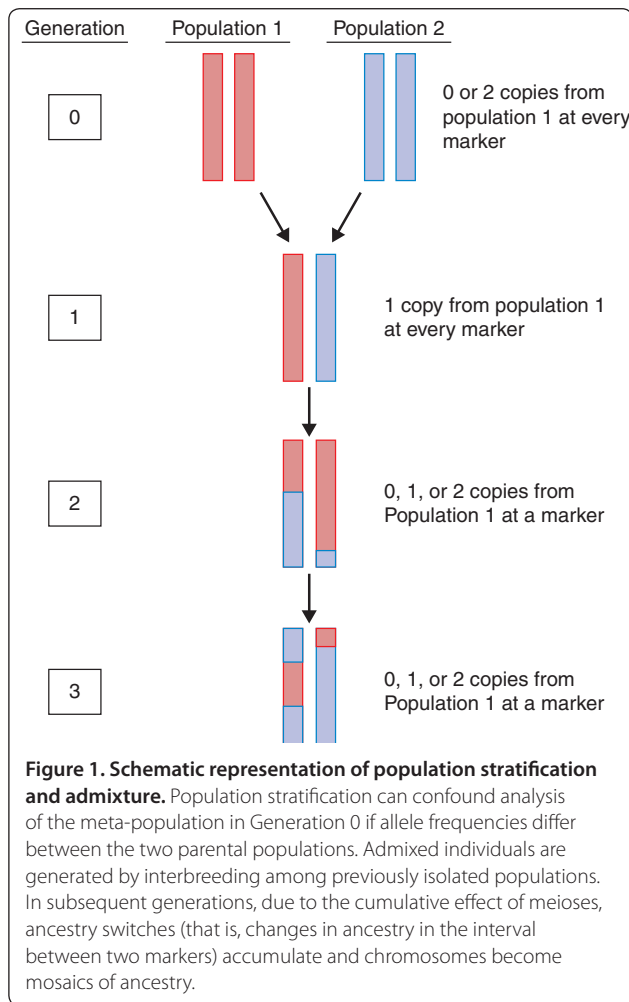
loci include African Americans, who have African and European ancestry, and Latino Americans, who have African, European, and Native American ancestry. These admixed populations afford opportunities for the study of health inequalities or group differences, which can occur when there are differences in traits such as disease susceptibility (for example, a 2.8-fold increase in risk for hypertensive heart disease in African Americans compared with European Americans [5]) or drug response (for example, differential response between populations with African and European ancestry to peginterferon α-2a or peginterferon α-2b combined with ribavirin, which are used to treat chronic infection with the hepatitis C virus [6]). Such differences result from a combination of environmental and genetic differences, the latter of which are our focus.

Admixture mapping and association mapping studies in admixed populations are poised to enter a new era as a result of the availability of economical high-throughput genotyping and sequencing. To date, ancestry has been estimated using panels containing <10,000 highly ascertained markers known as ancestry-informative markers (AIMs) [7]. Estimation of ancestry improves with panels of approximately 1,000,000 random markers compared with sparse panels of AIMs, leading to increased statistical power and resolution for admixture mapping. Improved estimation of ancestry at the marker level results in decreased false-positive error rates and increased power for association mapping due to the elimination of confounding by ancestry. Highly resolved estimation of local ancestry can also facilitate detection of natural selection in admixed samples. Improved estimation of ancestry will therefore contribute to mapping of disease loci as well as contribute to understanding demographic and adaptive history.

Population genetics of admixture

Conceptually, an admixed human population resembles an advanced intercross between outbred populations with the admixed individuals having variable ancestry [8]. To illustrate the salient features of variation in ancestry, consider two isolated populations that have experienced no interbreeding (Figure 1). In Generation 0,

---

*Correspondence: shrinerda@mail.nih.gov, rotimic@mail.nih.gov
Center for Research on Genomics and Global Health, National Human Genome Research Institute, Building 12A, Room 4047, 12 South Drive, MSC 5635, Bethesda, MD 20892, USA.

Figure 1. Schematic representation of population stratification and admixture. Population stratification can confound analysis of the meta-population in Generation 0 if allele frequencies differ between the two parental populations. Admixed individuals are generated by interbreeding among previously isolated populations. In subsequent generations, due to the cumulative effect of meioses, ancestry switches (that is, changes in ancestry in the interval between two markers) accumulate and chromosomes become mosaics of ancestry.

There are several characteristics of admixed populations that are relevant to disease loci mapping. First, not all individuals in an admixed population necessarily have the same proportion of ancestors from each parental population. Second, all loci do not have to share the same genealogical history. These two characteristics of admixed populations are sources of variance that must be accounted for when estimating local ancestry. Third, at any given locus, allele frequencies can vary between the parental populations. The expected allele frequency in the admixed population is the linear combination of the allele frequencies in the parental populations with weights determined by the sample admixture proportion. Fourth, the admixed population can be more genetically diverse than the parental populations if a locus is not polymorphic with the same alleles in all parental populations. For example, suppose a locus is polymorphic in one parental population and monomorphic in a second parental population. In addition, suppose a second locus is monomorphic in the first parental population and polymorphic in the second parental population. The admixed population is expected to be polymorphic at both loci. Fifth, similar to the way in which allele frequencies at a locus may vary, the patterns of covariance between allele frequencies at linked loci, known as linkage disequilibrium, can also differ. As a result, the distribution of haplotype frequencies in the admixed population can be substantially different from the distributions of haplotype frequencies in the parental populations. These latter three characteristics of admixed populations affect how many markers are needed for association mapping and the fine-mapping of functional variants.
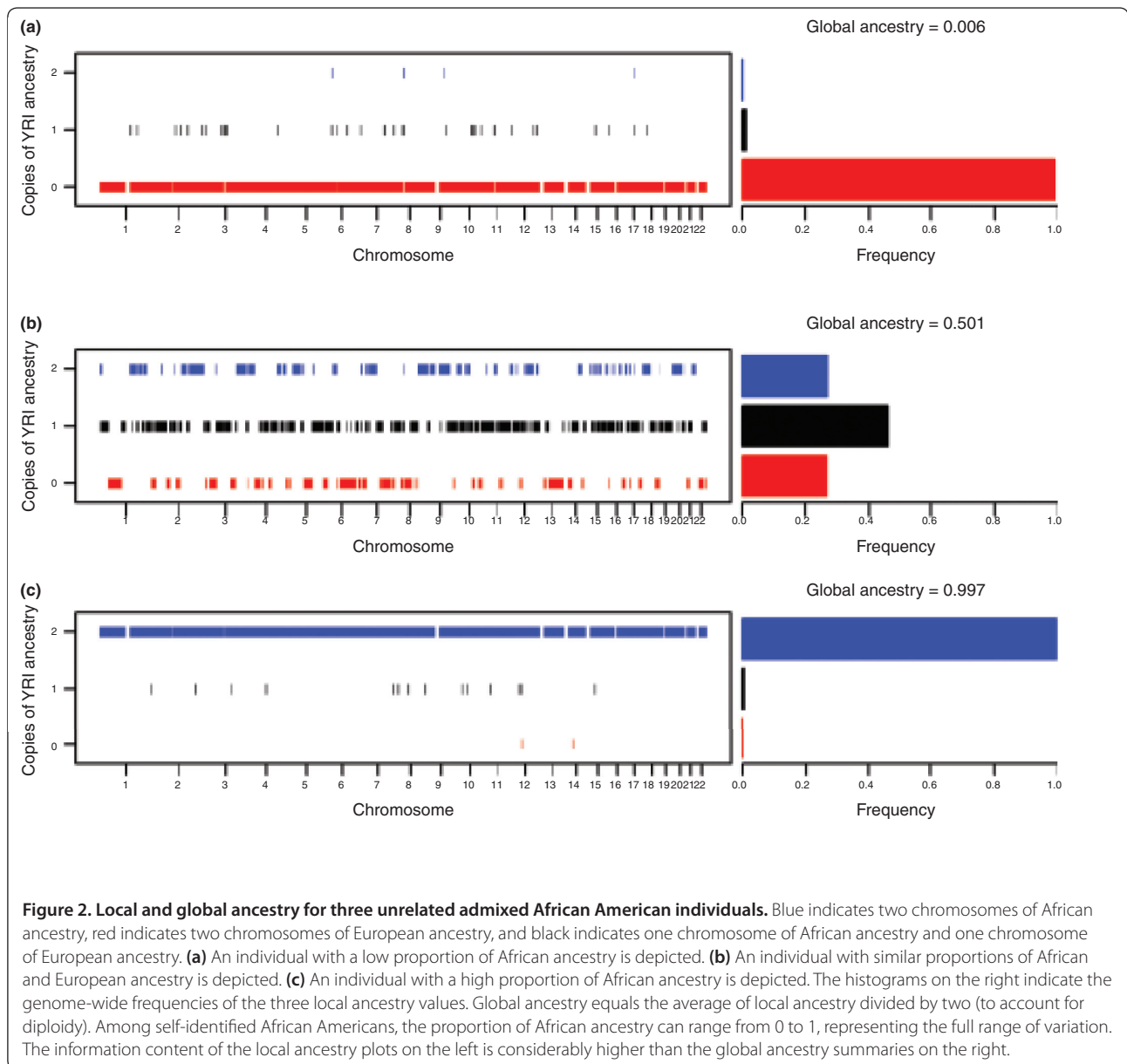
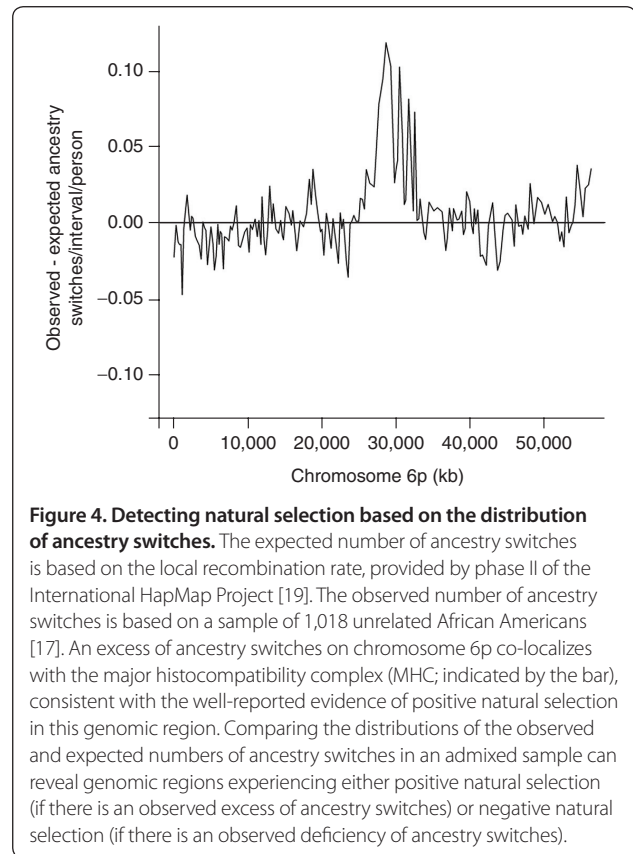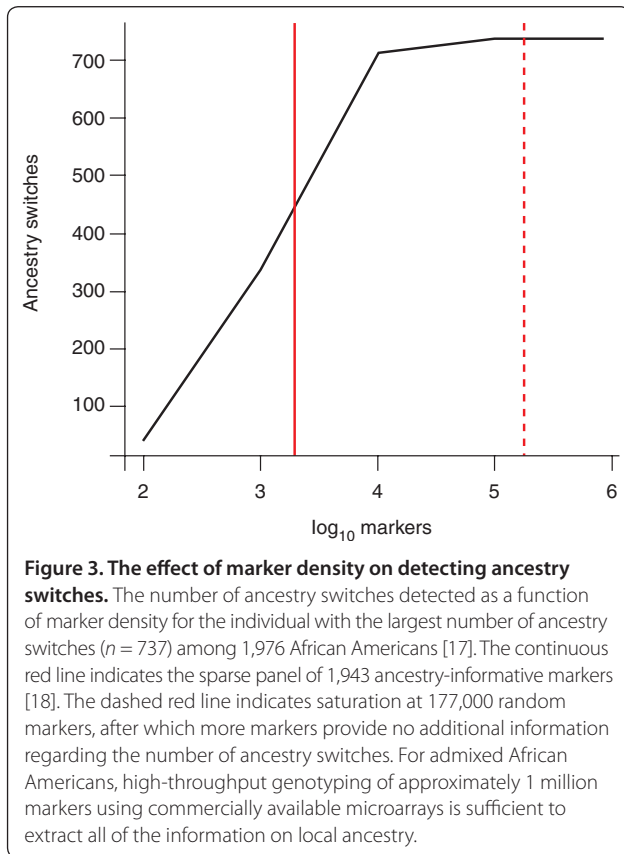**Admixture mapping: from AIMs to random markers**
Linkage disequilibrium caused by admixture that extends further than background linkage disequilibrium in the parental populations is the basis of admixture mapping [9,10]. Admixture mapping became practical in 2004 with the development of statistical methods [11] and panels of AIMs developed from reference databases [12] (for a historical overview of progress, see Winkler *et al.* [7]). Essentially, admixture mapping is designed to evaluate variation in ancestry. Tests of linkage are based on assessing the relationship between phenotype and local ancestry [13]. The standard case-control design involves a comparison of local ancestry between cases and controls, whereas the case-only design involves a comparison of local ancestry to the average local ancestry [11,14,15]. The average local ancestry is synonymous with global ancestry and the individual admixture proportion.

The next era of admixture mapping will benefit from an increased density of markers. Local ancestry can be

the two parental populations form a meta-population, which is simply a population of populations (Figure 1). At this generation, every marker in the genome of an individual traces its ancestry to only one parental population. Consequently, ancestry for each person at each marker, known as local ancestry, is constant for each individual across all loci.

After one generation of random mating within the meta-population, an individual has inherited one chromosome from each parental population (Figure 1). Local ancestry is still uniform across all loci for a given individual. After a second generation of random mating and beyond, an individual's genome is a mosaic of chromosomal segments with ancestry switching from segment to segment among the parental populations (Figure 1). An ancestry switch refers to a change in ancestry in the interval between two markers and is the result of recombination during meiosis. In this simple example, the proportion of ancestors from each parental population is equal for every admixed individual.

estimated using either AIMs or random markers. Furthermore, given the distribution of local ancestry, we can efficiently estimate global ancestry (Figure 2). However, despite ascertainment for ancestry informativeness, a sparse panel of AIMs does not extract as much information regarding ancestry as does a dense panel of random markers [16]. Dense panels provide two advantages: increased sensitivity to smaller segments and higher resolution of ancestry switches. The average intermarker distance decreases from approximately 1,500 kb for a typical panel of 2,000 AIMs to approximately 3 kb for a dense panel of 1,000,000 markers. As an example, in a sample of 1,976 African Americans [17], we detected a 1,027 kb segment of European ancestry in an African background with the intervals for the left and right switches localized to 35 kb and 1 kb, respectively. This segment is undetectable using a sparse panel of AIMs [18], as it lies entirely between flanking markers 1,176 kb apart. Expanded to the genome-wide scale, how many random markers would it take to detect all ancestry switches? By examining the individual with the most ancestry switches in our sample, we estimate that the number of random markers required is 177,000 (Figure 3), for which high-throughput genotyping will be more than sufficient. Some 'failures' of previous admixture mapping studies might have resulted from peaks of excess ancestry falling between AIMs. Revisiting these studies with denser panels might yield positive findings.

Figure 2. Local and global ancestry for three unrelated admixed African American individuals. Blue indicates two chromosomes of African ancestry, red indicates two chromosomes of European ancestry, and black indicates one chromosome of African ancestry and one chromosome of European ancestry. **(a)** An individual with a low proportion of African ancestry is depicted. **(b)** An individual with similar proportions of African and European ancestry is depicted. **(c)** An individual with a high proportion of African ancestry is depicted. The histograms on the right indicate the genome-wide frequencies of the three local ancestry values. Global ancestry equals the average of local ancestry divided by two (to account for diploidy). Among self-identified African Americans, the proportion of African ancestry can range from 0 to 1, representing the full range of variation. The information content of the local ancestry plots on the left is considerably higher than the global ancestry summaries on the right.

Figure 3. The effect of marker density on detecting ancestry switches. The number of ancestry switches detected as a function of marker density for the individual with the largest number of ancestry switches ($n = 737$) among 1,976 African Americans [17]. The continuous red line indicates the sparse panel of 1,943 ancestry-informative markers [18]. The dashed red line indicates saturation at 177,000 random markers, after which more markers provide no additional information regarding the number of ancestry switches. For admixed African Americans, high-throughput genotyping of approximately 1 million markers using commercially available microarrays is sufficient to extract all of the information on local ancestry.

Figure 4. Detecting natural selection based on the distribution of ancestry switches. The expected number of ancestry switches is based on the local recombination rate, provided by phase II of the International HapMap Project [19]. The observed number of ancestry switches is based on a sample of 1,018 unrelated African Americans [17]. An excess of ancestry switches on chromosome 6p co-localizes with the major histocompatibility complex (MHC; indicated by the bar), consistent with the well-reported evidence of positive natural selection in this genomic region. Comparing the distributions of the observed and expected numbers of ancestry switches in an admixed sample can reveal genomic regions experiencing either positive natural selection (if there is an observed excess of ancestry switches) or negative natural selection (if there is an observed deficiency of ancestry switches).

Fine-scale mapping of ancestry enables additional characterization of the distribution of ancestry switches. By definition, ancestry switches are a subset of meiotic recombination events. The expected number of ancestry switches can be calculated using a fine-scale recombination map such as the one provided by the International HapMap Project [19]. Deviation between the observed and expected numbers of ancestry switches indicates an inconsistency between the recombination map and the sample. A trivial explanation for such inconsistencies is error in the recombination map and/or the estimation of ancestry switches. Alternatively, fewer ancestry switches than expected given the local recombination rate might reflect negative natural selection, whereas more ancestry switches than expected might reflect positive natural selection [20]. To illustrate, on chromosome 6p in our sample of African Americans, the region from 28 Mb to 33 Mb shows an excess of ancestry switches (Figure 4). This region includes the major histocompatibility complex, which includes multiple immune response genes and is well known to be under positive natural selection. However, formal tests to evaluate natural selection in this manner await development. The distribution of ancestry switches also provides information with respect to the number of generations since admixture began [21].

Two major challenges remain in admixture mapping. First, inferring ancestry conditional on two parental populations is generally considered to be solved, but inferring ancestry for admixed populations with three or more parental populations remains challenging, particularly when the number of parental populations is unknown. As an example of three-way admixture, Puerto Rican individuals can have varying proportions of African, European, and Native American ancestry [22,23]. Compared with a prevalence of 7.8% of asthma among European Americans, the prevalence of asthma among Puerto Ricans is 16.6% [24]. Second, admixture mapping involves testing multiple markers across the genome. To maintain control of the false-positive error rate and provide maximum power, the genome-wide significance level must account for the number of markers tested while accounting for correlation of ancestry between markers. Appropriate genome-wide significance levels for admixture mapping are unclear [25-27]: for African Americans, estimates of the number of tests range from 400 [26] to 31,000 [27].

Refining association mapping in admixed populations

Association mapping is designed to evaluate differences in genotype frequencies. The major challenge for this

approach in admixed populations is the risk of false-positive genotype-phenotype associations due to variation in ancestry [28]. There are several techniques to control for this form of confounding for samples of ancestrally homogeneous individuals, including genomic control, structured association testing, principal components, variance components, and propensity scores [16,29,30]. These techniques control for confounding at the level of the individual but none controls for confounding at the level of specific markers such as SNPs [31].

Global ancestry as a covariate will control for confounding due to variation in individual ancestry if there are no marker-specific ancestry effects and genotypic effects are additive [32]. If there are marker-specific ancestry effects, or if genotypic effects are not additive, it is important to measure local ancestry to control for confounding due to admixture (Table 1). Accounting for local ancestry can also improve power of association testing if there are both local ancestry and genotypic effects at the same marker (Table 1). Controlling for local ancestry will not necessarily control for confounding due to global ancestry because local ancestry and global ancestry are weakly correlated [33]. Therefore, control of confounding due to admixture requires conditioning on both local and global ancestry.

One way to control for confounding due to local ancestry in association mapping is to simply include local

Table 1. Controlling for confounding due to admixture in association testing

Generative model[a]		Regression model[d]		
$\beta_{genotype}$[b]	$\beta_{local\ ancestry}$[c]	Genotype	Genotype and global ancestry	Genotype stratified by local ancestry
0	0	0.048	0.044	0.053
0	1	0.570	0.538	0.052
1	0	0.892	0.892	0.893
1	1	0.599	0.626	0.899

The first two rows demonstrate inflation of the false-positive error rate resulting from confounding due to admixture. The second two rows demonstrate the loss of power resulting from confounding due to admixture. In both cases, confounding is controlled by local ancestry but not by global ancestry. [a]Two isolated parental populations were generated with $F_{ST} = 0.115$ (F_{ST} is the ratio of the observed variance in allele frequencies among populations to the variance expected if the populations were randomly mating), mimicking the amount of population differentiation between the African and European ancestors of African Americans. A sample of admixed individuals was generated with 80% of the genome inherited from the first parental population, mimicking the amount of African ancestry in African Americans. A dataset consisted of 1,000 unrelated individuals and 1,000 unlinked markers. The generative model for the phenotype was a linear model with the listed fixed effects for the queried marker, no effects for all other markers, and noise equal to a random deviate from the standard normal distribution. [b]For each marker and individual, the genotype was coded as 0, 1, or 2 copies of the derived allele. [c]For each marker and individual, local ancestry was coded as 0, 1, or 2 copies inherited from the first parental population. [d]The rejection rates (false-positive error rates if $\beta_{genotype}$ = 0, or power if $\beta_{genotype}$ = 1) for testing genotype association at one marker are shown. The significance level was 0.05.

ancestry as an additive covariate. This parametric approach assumes that the effect of local ancestry is additive, analogous to the additive genetic model for genotypes. Alternatively, a non-parametric approach to control for confounding due to local ancestry is stratified regression. Specifically, when testing for association at a locus, one actually performs separate regressions with the subgroups defined by local ancestry. The separate regressions can then be pooled to generate inverse variance-weighted regression coefficients and standard errors. One can also perform subgroup analyses. For example, differences in effect sizes at a locus can be tested by Welch's *t* test, and differences in reference allele frequencies can be tested using a test of proportions between subgroups. These tests help to address the issue of heterogeneity of genetic associations across populations. Furthermore, we can jointly test for genotype and ancestry effects at each marker [34].

Imputation is commonly used in association studies to 'fill in' genotypes for untyped markers by leveraging external data on patterns of linkage disequilibrium [35]. For example, the HapMap [19] or 1000 Genomes CEU [36] data provide reference data regarding linkage disequilibrium patterns relevant for association studies comprising samples of similar European ancestry. For admixed samples, each parental population may be represented by a separate reference data set. However, it is not yet clear how best to utilize multiple reference data sets to maximize imputation accuracy [37]. Ancestry-aware imputation can be more accurate than not accounting for local ancestry [38].

In direct contrast to the situation with admixture mapping, approximately 1 million markers are barely sufficient to saturate background linkage disequilibrium in association mapping in populations of European or East Asian ancestry and insufficient for populations of African ancestry, as they are more genetically diverse [19]. In addition to weaker background linkage disequilibrium in populations of African ancestry compared with those of European ancestry, there are more low-frequency and rare variants in populations of African ancestry [36]. Whole-genome sequencing can contribute substantially to association mapping by eliminating the use of tagging variants that achieve poor coverage in genomic regions of weak linkage disequilibrium and by discovering all genetic variants regardless of frequency. Compared with high-throughput genotyping, sequencing will also expedite finding causal variants (that is, genetic variants directly associated with the phenotype of interest).

Success stories

Several success stories of disease loci mapping in admixed populations have been reported [7,39],

including studies in African Americans for asthma, cancer, and kidney disease. The two diseases for which fine-mapping of the original admixture signal has proceeded the furthest are prostate cancer and end-stage kidney disease.

An admixture signal on chromosome 8q24 was found for prostate cancer in African Americans [40]. The same locus was detected by linkage analysis in Icelanders [41]. At least three blocks of linkage disequilibrium containing several independently associated variants have been identified within this locus [41-44]. Molecular studies identified enhancer elements for the oncogenic transcription factor *MYC* that regulate tissue-specific expression patterns, which potentially explain why the locus affects risk for breast and colorectal cancer in addition to prostate cancer [45-51]. One of these enhancers interacts with the *MYC* promoter through binding of the transcription factor complex β-catenin/TCF7L2 [46,50]. *TCF7L2* is the most strongly associated gene for type 2 diabetes [52], providing in part a genetic basis for the epidemiological association between these two diseases. Thus, although the original admixture signal was detected for prostate cancer, ongoing follow-up studies indicate that the locus also influences breast cancer, colorectal cancer, and type 2 diabetes.

An admixture signal on chromosome 22q13 was found for focal segmental glomerulosclerosis, a cause of end-stage kidney disease, in African Americans [53,54]. Originally, the candidate gene underlying this signal was thought to be *MYH9*, which encodes non-muscle myosin heavy chain 9 and is highly expressed in kidney podocytes [53,54]. Sequence data from the 1000 Genomes Project [36] included newly discovered variants in a part of chromosome 22q13 that had poor coverage in the International HapMap Project [55]. Using the more comprehensive sequence data, it now appears that the major effect gene is that encoding apolipoprotein L1, which is in linkage disequilibrium with *MYH9* [56]. The protein product apolipoprotein L1 has trypanolytic activity; the gene locus appears to be under balancing selection for protection against sleeping sickness at the cost of increased risk for end-stage kidney disease [56]. Starting from the original admixture signal, sequence data have permitted fine-mapping, with the majority of the signal resolved to two alleles located in the last exon of the gene encoding apolipoprotein L1 [56].

Admixed populations: how many markers should one use?

In the USA, the two most commonly studied admixed populations are African Americans and Latino Americans. Both of these populations are characterized by less than 25 generations since admixture began as a result of maritime European expansion. For both of these admixed populations, the parental populations genetically differ at the intercontinental level. Population differentiation is often measured by F_{ST}, which is defined as the ratio of the observed variance in allele frequencies among populations to the variance expected if the populations were randomly mating. F_{ST} at the intercontinental level is generally greater than 0.10 [55]. Other examples of admixed populations with this level of population differentiation include Ashkenazi Jews (who have Eastern European and Middle Eastern ancestry), South African Coloureds (who have Bantu-speaking African, European, Indian, Khoisan, and Southeast Asian ancestry), Australian Aboriginals (who have Aboriginal and European ancestry), and Pacific Islanders (who have European and Polynesian ancestry) [7].

We estimate that <200,000 random markers (that is, markers not ascertained for ancestry informativeness) are sufficient to saturate the signal of linkage disequilibrium due to admixture in the context of admixture mapping for African Americans. A higher density of markers may enable detection of older admixture, because the number of ancestry switches (and the number of markers required to detect all of those ancestry switches) increases as the number of generations since admixture began increases. As a first-pass approximation, a sample is well powered to detect population structure if $1/F_{ST}$ exceeds the geometric mean of the number of unrelated individuals and the number of independent markers [57]. With a sufficiently large sample and dense markers, it might be possible to detect admixture among parental populations that differ at the intracontinental level, or F_{ST} of the order of 0.01 [57], which is relevant for analysis of Northern Europe versus Southern Europe [57] and East Asian populations [58]. It also may be possible to detect admixture that occurred more distantly in the past (for example, approximately 100 generations ago), for example, in the Uyghurs in western China (who have European and East Asian ancestry) [59]. At a more ancient level, Tishkoff *et al.* reported evidence of admixture between Bantu-speaking Africans and Khoisans [60]. At an even more ancient level, the proportion of Neandertal ancestry of Eurasians has been estimated to be between 1% and 4% [61]. Characterization of populations with complex patterns of admixture can contribute substantially to our understanding of population history and can also contribute to understanding complex disease.

Bright prospects for the future

The convergence of high-throughput genotyping or sequencing and new methods to infer local ancestry allows for joint admixture and association analysis. Furthermore, sensitivity to detect admixture afforded by the combination of larger samples, denser markers, and

improved inferential methods has increased for ancient admixture events and admixture between more closely related populations. Improved estimation of ancestry will benefit association mapping in admixed populations by eliminating the effects of confounding due to variation in ancestry. High-throughput genotyping and sequencing will enable refined estimation of ancestry, making disease loci identification in admixed populations more powerful.

Abbreviations

AIM, ancestry-informative marker; F_{ST}, the ratio of the observed variance in allele frequencies among populations to the variance expected if the populations were randomly mating; SNP, single nucleotide polymorphism.

Competing interests

The authors declare that they have no competing interests.

References

1. Need AC, Goldstein DB: **Next generation disparities in human genomics: concerns and remedies**. *Trends Genet* 2009, **25**:489-494.
2. McCarthy MI: **Casting a wider net for diabetes susceptibility genes**. *Nat Genet* 2008, **40**:1039-1040.
3. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M: **Genome-wide association studies in diverse populations**. *Nat Rev Genet* 2010, **11**:356-366.
4. Rotimi CN, Jorde LB: **Ancestry and disease in the age of genomic medicine**. *N Engl J Med* 2010, **363**:1551-1558.
5. Davey Smith G, Neaton JD, Wentworth D, Stamler R, Stamler J: **Mortality differences between black and white men in the USA: contribution of income and other risk factors among men screened for the MRFIT**. *Lancet* 1998, **351**:934-939.
6. Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, Heinzen EL, Qiu P, Bertelsen AH, Muir AJ, Sulkowski M, McHutchison JG, Goldstein DB: **Genetic variation in *IL28B* predicts hepatitis C treatment-induced viral clearance**. *Nature* 2009, **461**:399-401.
7. Winkler CA, Nelson GW, Smith MW: **Admixture mapping comes of age**. *Annu Rev Genomics Hum Genet* 2010, **11**:65-89.
8. McKeigue PM: **Prospects for admixture mapping of complex traits**. *Am J Hum Genet* 2005, **76**:1-7.
9. Rife DC: **Populations of hybrid origin as source material for the detection of linkage**. *Am J Hum Genet* 1954, **6**:26-33.
10. Chakraborty R, Weiss KM: **Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci**. *Proc Natl Acad Sci USA* 1988, **85**:9119-9123.
11. Patterson N, Hattangadi N, Lane B, Lohmueller KE, Hafler DA, Oksenberg JR, Hauser SL, Smith MW, O'Brien SJ, Altshuler D, Daly MJ, Reich D: **Methods for high-density admixture mapping of disease genes**. *Am J Hum Genet* 2004, **74**:979-1000.
12. Smith MW, Patterson N, Lautenberger JA, Truelove AL, McDonald GJ, Waliszewska A, Kessing BD, Malasky MJ, Scafe C, Le E, De Jager PL,

13. Mignault AA, Yi Z, De Thé G, Essex M, Sankalé J-L, Moore JH, Poku K, Phair JP, Goedert JJ, Vlahov D, Williams SM, Tishkoff SA, Winkler CA, De La Vega FM, Woodage T, Sninsky JJ, Hafler DA, Altshuler D, Gilbert DA, *et al.*: **A high-density admixture map for disease gene discovery in African Americans**. *Am J Hum Genet* 2004, **74**:1001-1013.
13. McKeigue PM: **Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning on parental admixture**. *Am J Hum Genet* 1998, **63**:241-251.
14. Hoggart CJ, Shriver MD, Kittles RA, Clayton DG, McKeigue PM: **Design and analysis of admixture mapping studies**. *Am J Hum Genet* 2004, **74**:965-978.
15. Montana G, Pritchard JK: **Statistical tests for admixture mapping with case-control and cases-only data**. *Am J Hum Genet* 2004, **75**:771-789.
16. Price AL, Zaitlen NA, Reich D, Patterson N: **New approaches to population stratification in genome-wide association studies**. *Nat Rev Genet* 2010, **11**:459-463.
17. Adeyemo A, Gerry N, Chen G, Herbert A, Doumatey A, Huang H, Zhou J, Lashley K, Chen Y, Christman M, Rotimi C: **A genome-wide association study of hypertension and blood pressure in African Americans**. *PLoS Genet* 2009, **5**:e1000564.
18. Chen G, Shriner D, Zhou J, Doumatey A, Huang H, Gerry NP, Herbert A, Christman MF, Chen Y, Dunston GM, Faruque MU, Rotimi CN, Adeyemo A: **Development of admixture mapping panels for African Americans from commercial high-density SNP arrays**. *BMC Genomics* 2010, **11**:417.
19. The International HapMap Consortium: **A second generation human haplotype map of over 3.1 million SNPs**. *Nature* 2007, **449**:851-861.
20. Rieseberg LH, Whitton J, Gardner K: **Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species**. *Genetics* 1999, **152**:713-727.
21. Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, Ruczinski I, Beaty TH, Mathias R, Reich D, Myers S: **Sensitive detection of chromosomal segments of distinct ancestry in admixed populations**. *PLoS Genet* 2009, **5**:e1000519.
22. Bryc K, Velez C, Karafet T, Moreno-Estrada A, Reynolds A, Auton A, Hammer M, Bustamante CD, Ostrer H: **Genome-wide patterns of population structure and admixture among Hispanic/Latino populations**. *Proc Natl Acad Sci USA* 2010, **107 Suppl 2**:8954-8961.
23. Via M, Gignoux CR, Roth LA, Fejerman L, Galanter J, Choudhry S, Toro-Labrador G, Viera-Vera J, Oleksyk TK, Beckman K, Ziv E, Risch N, Burchard EG, Martínez-Cruzado JC: **History shaped the geographic distribution of genomic admixture on the island of Puerto Rico**. *PLoS ONE* 2011, **6**:e16513.
24. Akinbami LJ, Moorman JE, Liu X: **Asthma prevalence, health care use, and mortality: United States, 2005-2009**. *Natl Health Stat Report* 2011, (**32**):1-14.
25. Sha Q, Zhang X, Zhu X, Zhang S: **Analytical correction for multiple testing in admixture mapping**. *Hum Hered* 2006, **62**:55-63.
26. Zhu X, Zhang S, Tang H, Cooper R: **A classical likelihood based approach for admixture mapping using EM algorithm**. *Hum Genet* 2006, **120**:431-445.
27. Nalls MA, Wilson JG, Patterson NJ, Tandon A, Zmuda JM, Huntsman S, Garcia M, Hu D, Li R, Beamer BA, Patel KV, Akylbekova EL, Files JC, Hardy CL, Buxbaum SG, Taylor HA, Reich D, Harris TB, Ziv E: **Admixture mapping of white cell count: genetic locus responsible for lower white blood cell count in the Health ABC and Jackson Heart Studies**. *Am J Hum Genet* 2008, **82**:81-87.
28. Rosenberg NA, Nordborg M: **A general population-genetic model for the production by population structure of spurious genotype-phenotype associations in discrete, admixed or spatially distributed populations**. *Genetics* 2006, **173**:1665-1678.
29. Epstein MP, Allen AS, Satten GA: **A simple and improved correction for population stratification in case-control studies**. *Am J Hum Genet* 2007, **80**:921-930.
30. Tiwari HK, Barnholtz-Sloan J, Wineinger N, Padilla MA, Vaughan LK, Allison DB: **Review and evaluation of methods correcting for population stratification with a focus on underlying statistical principles**. *Hum Hered* 2008, **66**:67-86.
31. McClellan J, King MC: **Why it is time to sequence**. *Cell* 2010, **142**:353-355.
32. Redden DT, Divers J, Vaughan LK, Tiwari HK, Beasley TM, Fernández JR, Kimberly RP, Feng R, Padilla MA, Liu N, Miller MB, Allison DB: **Regional admixture mapping and structured association testing: conceptual unification and an extensible general linear model**. *PLoS Genet* 2006, **2**:e137.
33. Qin H, Morris N, Kang SJ, Li M, Tayo B, Lyon H, Hirschhorn J, Cooper RS, Zhu X: **Interrogating local population structure for fine mapping in genome-wide association studies**. *Bioinformatics* 2010, **26**:2961-2968.
34. Tang H, Siegmund DO, Johnson NA, Romieu I, London SJ: **Joint testing of genotype and ancestry association in admixed families**. *Genetic Epidemiol* 2010, **34**:783-791.

35. Marchini J, Howie B: **Genotype imputation for genome-wide association studies**. *Nat Rev Genet* 2010, **11**:499-511.
36. The 1000 Genomes Project Consortium: **A map of human genome variation from population-scale sequencing**. *Nature* 2010, **467**:1061-1073.
37. Paşaniuc B, Avinery R, Gur T, Skibola CF, Bracci PM, Halperin E: **A generic coalescent-based framework for the selection of a reference panel for imputation**. *Genet Epidemiol* 2010, **34**:773-782.
38. Paşaniuc B, Sankararaman S, Kimmel G, Halperin E: **Inference of locus-specific ancestry in closely related populations**. *Bioinformatics* 2009, **25**:i213-i221.
39. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA: **Potential etiologic and functional implications of genome-wide association loci for human diseases and traits**. *Proc Natl Acad Sci USA* 2009, **106**:9362-9367.
40. Freedman ML, Haiman CA, Patterson N, McDonald GJ, Tandon A, Waliszewska A, Penney K, Steen RG, Ardlie K, John EM, Oakley-Girvan I, Whittemore AS, Cooney KA, Ingles SA, Altshuler D, Henderson BE, Reich D: **Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men**. *Proc Natl Acad Sci USA* 2006, **103**:14068-14073.
41. Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, Sigurdsson A, Benediktsdottir KR, Cazier J-B, Sainz J, Jakobsdottir M, Kostic J, Magnusdottir DN, Ghosh S, Agnarsson K, Birgisdottir B, Le Roux L, Olafsdottir A, Blondal T, Andresdottir M, Gretarsdottir OS, Bergthorsson JT, Gudbjartsson D, Gylfason A, Thorleifsson G, Manolescu A, Kristjansson K, Geirsson G, Isaksson H, Douglas J, *et al.*: **A common variant associated with prostate cancer in European and African populations**. *Nat Genet* 2006, **38**:652-658.
42. Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, Rafnar T, Bergthorsson JT, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, Xu J, Blondal T, Kostic J, Sun J, Ghosh S, Stacey SN, Mouy M, Saemundsdottir J, Backman VM, Kristjansson K, Tres A, Partin AW, Albers-Akkers MT, Godino-Ivan Marcos J, Walsh PC, Swinkels DW, Navarrete S, *et al.*: **Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24.** *Nat Genet* 2007, **39**:631-637.
43. Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewska A, Neubauer J, Tandon A, Schirmer C, McDonald GJ, Greenway SC, Stram DO, Le Marchand L, Kolonel LN, Frasco M, Wong D, Pooler LC, Ardlie K, Oakley-Girvan I, Whittemore AS, Cooney KA, John EM, Ingles SA, Altshuler D, Henderson BE, Reich D: **Multiple regions within 8q24 independently affect risk for prostate cancer**. *Nat Genet* 2007, **39**:638-644.
44. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello MJ, Fearnhead P, Yu K, Chatterjee N, Wang Z, Welch R, Staats BJ, Calle EE, Feigelson HS, Thun MJ, Rodriguez C, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Giovannucci E, Willett WC, Cancel-Tassin G, Cussenot O, Valeri A, Andriole GL, Gelmann EP, Tucker M, Gerhard DS, *et al.*: **Genome-wide association study of prostate cancer identifies a second risk locus at 8q24**. *Nat Genet* 2007, **39**:645-649.
45. Haiman CA, Le Marchand L, Yamamato J, Stram DO, Sheng X, Kolonel LN, Wu AH, Reich D, Henderson BE: **A common genetic risk factor for colorectal and prostate cancer**. *Nat Genet* 2007, **39**:954-956.
46. Pomerantz MM, Ahmadiyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H, Beckwith CA, Chan JA, Hills A, Davis M, Yao K, Kehoe SM, Lenz H-J, Haiman CA, Yan C, Henderson BE, Frenkel B, Barretina J, Bass A, Tabernero J, Baselga J, Regan MM, Manak JR, Shivdasani R, Coetzee GA, Freedman ML: **The 8q24 cancer risk variant rs6983267 shows long-range interaction with *MYC* in colorectal cancer**. *Nat Genet* 2009, **41**:882-884.
47. Jia L, Landan G, Pomerantz M, Jaschek R, Herman P, Reich D, Yan C, Khalid O, Kantoff P, Oh W, Manak JR, Berman BP, Henderson BE, Frenkel B, Haiman CA, Freedman M, Tanay A, Coetzee GA: **Functional enhancers at the gene-poor 8q24 cancer-linked locus**. *PLoS Genet* 2009, **5**:e1000597.
48. Wasserman NF, Aneas I, Nobrega MA: **An 8q24 gene desert variant associated with prostate cancer risk confers differential *in vivo* activity to a *MYC* enhancer**. *Genome Res* 2010, **20**:1191-1197.
49. Ahmadiyeh N, Pomerantz MM, Grisanzio C, Herman P, Jia L, Almendro V, He HH, Brown M, Liu XS, Davis M, Caswell JL, Beckwith CA, Hills A, MacConaill L, Coetzee GA, Regan MM, Freedman ML: **8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with *MYC***. *Proc Natl Acad Sci USA* 2010, **107**:9742-9746.
50. Sotelo J, Esposito D, Duhagon MA, Banfield K, Mehalko J, Liao H, Stephens RM, Harris TJ, Munroe DJ, Wu X: **Long-range enhancers on 8q24 regulate c-Myc**. *Proc Natl Acad Sci USA* 2010, **107**:3001-3005.
51. Wright JB, Brown SJ, Cole MD: **Upregulation of c-*MYC* in *cis* through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells**. *Mol Cell Biol* 2010, **30**:1411-1420.
52. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, McCulloch LJ, Ferreira T, Grallert H, Amin N, Wu G, Willer CJ, Raychaudhuri S, McCarroll SA, Langenberg C, Hofmann OM, Dupuis J, Qi L, Segrè AV, van Hoek M, Navarro P, Ardlie K, Balkau B, Benediktsson R, Bennett AJ, Blagieva R, *et al.*: **Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis**. *Nat Genet* 2010, **42**:579-589.
53. Kopp JB, Smith MW, Nelson GW, Johnson RC, Freedman BI, Bowden DW, Oleksyk T, McKenzie LM, Kajiyama H, Ahuja TS, Berns JS, Briggs W, Cho ME, Dart RA, Kimmel PL, Korbet SM, Michel DM, Mokrzycki MH, Schelling JR, Simon E, Trachtman H, Vlahov D, Winkler CA: ***MYH9* is a major-effect risk gene for focal segmental glomerulosclerosis**. *Nat Genet* 2008, **40**:1175-1184.
54. Kao WHL, Klag MJ, Meoni LA, Reich D, Berthier-Schaad Y, Li M, Coresh J, Patterson N, Tandon A, Powe NR, Fink NE, Sadler JH, Weir MR, Abboud HE, Adler SG, Divers J, Iyengar SK, Freedman BI, Kimmel PL, Knowler WC, Kohn OF, Kramp K, Leehey DJ, Nicholas SB, Pahl MV, Schelling JR, Sedor JR, Thornley-Brown D, Winkler CA, Smith MW, *et al.*: ***MYH9* is associated with nondiabetic end-stage renal disease in African Americans**. *Nat Genet* 2008, **40**:1185-1192.
55. The International HapMap 3 Consortium: **Integrating common and rare genetic variation in diverse human populations**. *Nature* 2010, **467**:52-58.
56. Genovese G, Friedman DJ, Ross MD, Lecordier L, Uzureau P, Freedman BI, Bowden DW, Langefeld CD, Oleksyk TK, Uscinski Knob AL, Bernhardy AJ, Hicks, PJ, Nelson GW, Vanhollebeke B, Winkler CA, Kopp JB, Pays E, Pollak MR: **Association of trypanolytic ApoL1 variants with kidney disease in African Americans**. *Science* 2010, **329**:841-845.
57. Patterson N, Price AL, Reich D: **Population structure and eigenanalysis**. *PLoS Genet* 2006, **2**:e190.
58. Tian C, Kosoy R, Lee A, Ransom M, Belmont JW, Gregersen PK, Seldin MF: **Analysis of East Asia genetic substructure using genome-wide SNP arrays**. *PLoS ONE* 2008, **3**:e3862.
59. Xu S, Jin L: **A genome-wide analysis of admixture in Uyghurs and a high-density admixture map for disease-gene discovery**. *Am J Hum Genet* 2008, **83**:322-336.
60. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo J-M, Doumbo O, Ibrahim M, Juma AT, Kotze MJ, Lema G, Moore JH, Mortensen H, Nyambo TB, Omar SA, Powell K, Pretorius GS, Smith MW, Thera MA, Wambebe C, Weber JL, Williams SM: **The genetic structure and history of Africans and African Americans**. *Science* 2009, **324**:1035-1044.
61. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y, Hansen NF, Durand EY, Malaspinas A-S, Jensen JD, Marques-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, *et al.*: **A draft sequence of the Neandertal genome**. *Science* 2010, **328**:710-722.