Review
Risk for rheumatic disease in relation to ethnicity and admixture
Mariam Molokhia and Paul McKeigue

London School of Hygiene and Tropical Medicine, London, UK

© Current Science Ltd

Abstract

Risk of systemic lupus erythematosus (SLE) is high in west Africans compared with Europeans, and risk of rheumatoid arthritis (RA) is high in Native Americans compared with Europeans. These differences are not accounted for by differences in allele or haplotype frequencies in the human leucocyte antigen (HLA) region or any other loci known to influence risk of rheumatic disease. Where there has been admixture between two or more ethnic groups that differ in risk of disease, studies of the relationship of disease risk to proportionate admixture can help to distinguish between genetic and environmental explanations for ethnic differences in disease risk and to map the genes underlying these differences.

Keywords: admixture, genetic, lupus, rheumatoid arthritis

Introduction

The risks of developing SLE and RA vary with ethnic origin. Prevalence of lupus has been found to be up to eight times higher in African-American and Afro-Caribbean populations than in people of European descent. The consistency with which high prevalence of lupus occurs in populations of west African descent who are living in different environments suggests that genetic factors are likely to underlie the high risk in this group. The high risk for lupus in west Africans compared with in Europeans does not appear to be accounted for by differences in allele frequencies at any of the loci at which associations with SLE have been found, including those in the HLA region. Other populations at high risk for lupus include Pacific islanders and Chinese-Americans, with prevalence rates up to three times higher than in people of European

descent living in the same countries. For RA, the highest risks are recorded for Native American populations, in which prevalence is up to four times higher than it is in Europeans. Although the frequency of alleles that code for the high-risk 'shared epitope' at the HLA-DRB1 locus is higher in Native Americans than in Europeans, this only accounts for a population risk ratio of about 1.4, compared with the observed risk ratios of around 4.

Where there has been admixture between ethnic groups that differ in risk for disease, studies of how the risk for disease varies with proportionate admixture can help to distinguish between genetic and environmental explanations for the difference in disease risk and can help to define the genetic model. No adequate studies have yet been undertaken on the risk for lupus in relation to admix-

Table 1

Prevalence of systemic lupus erythematosus in European, Afro-Caribbean and African American women

Location	Ethnicity	Age range	Age-standardized prevalence/100 000
Birmingham, UK [6]	European	≥18 years	36
Birmingham, UK [6]	Afro-Caribbean	≥18 years	197
Nottingham, UK [5]	Afro-Caribbean	≥20 years	251
Curacao, West Indies [12]	Afro-Caribbean (predominantly)	≥15 years	117
San Francisco, USA [9]	African-American	≥15 years	375
San Francisco, USA [9]	European-American	≥15 years	81

ture in populations of mixed European/west African descent, or on the risk for RA in relation to admixture in populations of mixed European/Native American descent. If ethnic differences in risk for lupus and RA have a genetic basis, it is possible in principle to map the genes that underlie ethnic differences in risk by studying affected individuals of mixed descent. This approach is an extension of the methods used for linkage analysis of a cross in experimental genetics. Before this can be applied in practice, it is necessary to assemble sets of marker polymorphisms that can be used to assign ancestry on chromosomes of mixed descent.

Ethnic variation in risk for rheumatic disease
Studying ethnic variation in disease risk can yield clues to environmental or genetic factors that influence disease risk. Where variations in risk between populations have been identified, studies of migrants between low-risk and high-risk areas can help to distinguish between genetic and environmental explanations for these differences in risk, and to determine the age at which risk is set. If an ethnic difference in disease risk has an environmental explanation, this difference is expected to 'wear off' within a few generations after migration [1]. For instance, the low risk for multiple sclerosis in tropical countries compared with the risk in northern latitudes persists in those who migrate from tropical countries to the UK as adults [2], but in second-generation migrants the risk for multiple sclerosis is similar to that in the host population [3]. If, on the other hand, genetic factors underlie an ethnic difference in disease risk, we would expect this ethnic difference to persist in populations where migrants have been settled overseas for many generations, and to be observed consistently in all countries where a given migrant group has settled.

Ethnic variation in risk for systemic lupus erythematosus
Most estimates of SLE prevalence have relied on ascertainment of cases already diagnosed. A study in the UK that used a questionnaire followed by an antinuclear antibody test to screen for undiagnosed cases of SLE [4] suggested that about 40% of cases are undiagnosed.

Estimates of the risk ratio for SLE in one ethnic group compared with another generally assume that the proportion of cases that are undiagnosed and the average survival of diagnosed cases are the same in both groups. If these assumptions hold, the risk ratio can be estimated from the ratio of prevalence rates of diagnosed SLE in the two groups. The assumption that the proportion of cases that are undiagnosed does not vary between ethnic groups is more likely to hold in the UK, where access to health care does not depend on income, than in the USA. Table 1 summarizes prevalence studies in populations of west African descent, with prevalence in populations of European descent in the same countries for comparison. In general, the prevalence of SLE has been found to be far higher in people of west African descent settled in Europe and the Americas than in Europeans [5,6]. In the UK, prevalence in adults is estimated to be around 35 per 100 000 in women of European descent [5,7], compared with 200–250 per 100 000 in women of Afro-Caribbean origin [5,6].

It has been suggested that rates of SLE are low in west African because exposure to malaria and parasitic infections in west Africans in the tropics may result in an altered immunological state, leading to a tolerance of host antigens or absorption of autoantibody [8]. However, there are no population-based data on prevalence of SLE in west Africa itself.

Similar risk ratios have been reported in studies performed in the USA [9,10] that have compared prevalence rates in African-Americans and European-Americans. In a study of patients registered with the Kaiser Foundation Health Plan in San Francisco, age-adjusted prevalence of SLE in African-American women aged over 15 years was estimated to be 375 per 100 000, compared with 81 per 100 000 for European-American women [9]. The relatively high prevalence rates in this study may be attributable to the unrestricted access to health care. In other studies from the USA, coverage of the population by the health care system is likely to vary considerably between ethnic

groups. If poorer access to health care results in a lower proportion of cases ascertained in African-Americans, and shorter survival among diagnosed cases, this would reduce the ratio of prevalence in African-Americans to that in European-Americans. The incidence of SLE was estimated to be four times higher in African-American than in European-American women [10], which is approximately the same as the ratio recorded in studies that compared prevalence rates. From the studies that are available at present, it is not possible to establish whether the difference between the risk ratios of 6–8 reported in the UK and the risk ratios of 4–5 reported in the USA are real or attributable to differences in the extent to which the risk ratios are biased by differential case ascertainment. In most African-American populations the proportion of European admixture varies from 14 to 22%. Most Afro-Caribbean migrants to the UK are from Jamaica, where the average proportion of European admixture is only 7% [11]. If the risk for SLE varies with the proportion of the genome that is of African origin, we would expect the risk ratio for African-Americans compared with European-Americans to be higher than the risk ratio for Afro-Caribbeans compared with Europeans in the UK.

The only prevalence data from the Caribbean are from the island of Curacao [12], where prevalence in adult women was estimated to be 117 per 100 000 during the period 1980–1989. This is likely to be an underestimate because only a few cases were ascertained from outpatient clinics; even so, it is far higher than the prevalence estimates for women of European descent in the UK. Further evidence for high rates of SLE in the Caribbean comes from the observation that in Birmingham, UK, most Caribbean-born patients reported that their disease had begun before migration [6].

Although no other ethnic group has been found to have such consistently high risk for SLE as people of west African descent, there is some evidence of high risk in people of Chinese descent who are settled outside China. In a study in Malaysia [13], Chinese made up 81% of all hospital admissions (to the only teaching hospital in west Malaysia) with SLE, but only 35% of the local population. A later study [14] estimated prevalence rates (per 100 000) as 46 in Chinese, 26 in Indians and 12 in Malays. High rates in Malaysian Chinese when compared with other ethnic groups might be partly explained by greater access to health care. Excess prevalence of SLE was also reported among Chinese-Americans in Hawaii, however, where prevalence of SLE (per 100 000, both sexes) was estimated to be 24 in Chinese, 20 in Polynesians and six in Europeans [15]. Although these rates were based on small numbers of cases, the differences were statistically significant. These findings were supported by a subsequent study in Hawaii [16] in which age-adjusted prevalence rates for SLE were 33 per 100 000 in Chinese and 10 per 100 000 in Europeans (both sexes),

with cases ascertained from hospital records and death certificates. The high risk in Polynesian islanders is consistent with a study in New Zealand [17], in which prevalence per 100 000 (both sexes) was estimated to be 50 in Polynesians compared with 15 in Europeans.

Role of known genetic associations in ethnic differences in risk for systemic lupus erythematosus
Several loci have been identified at which variation is associated with risk for SLE. In order to determine whether the high risk for SLE in people of west African descent compared with the risk in Europeans can be accounted for by higher frequencies in west Africans of alleles or haplotypes that have been associated with increased risk for SLE within populations, we have calculated the population risk ratio (which is derived from the risk ratio between the exposed groups under study, in this case west Africans and Europeans) generated by each locus. The population risk ratio is calculated as follows:

$$\text{Population risk ratio} = \frac{1 + p_A(R-1)}{1 + p_E(R-1)}$$

where p_A is the frequency of the high risk genotype in Africans, p_E is the frequency of the high risk genotype in Europeans, and R is the risk ratio for the high risk genotype compared with the low risk genotype. This is based on a simple risk assessment where genotypes are classified as either high or low risk.

This calculation provides the risk ratio between west Africans and Europeans that we would predict from the risk ratios between genotypes and the frequency of genotypes in each population. For instance, if the frequency of a high-risk genotype is 50% in west Africans and 0% in Europeans, and this genotype is associated with a twofold risk for disease, this doubled risk for disease in half the west African population will generate a risk ratio of 1.5 between west Africans and Europeans. For each locus we have collated whatever data are available on the risk ratios associated with each genotype within populations and the allele frequencies in west Africans and Europeans. The results are summarized in Table 2.

Associations with variation in the human leucocyte antigen region
Most of the loci at which associations with risk for SLE have been identified are in the HLA region. This region, extending over about 4000 kilobases on chromosome 6, contains the class I genes HLA-A, HLA-B and HLA-C, and the class II gene families HLA-DP, HLA-DQ and HLA-DR. When alleles were typed by their antigenic specificities, the HLA-DR family of genes was considered as a single locus and alleles were designated in the sequence (DR1, DR2 ... DRw6, …) where the 'w' indicated a provisional or 'workshop' designation. Later five distinct genes (DRA,

Table 2

Extent to which known genetic associations can account for west African/European difference in risk for systemic lupus erythematosus

Locus	Alleles	Risk ratio		Allele frequencies controls		Population risk ratio*
		Eur	Afr	Eur	Afr	
Loci within the HLA region						
HLA-DR [20,21,65]	DRB1*0301 (DRw17) DRB1*0302 (DRw18) (HLA-DR2)	2[†]	1.3[†]	0.12	0.14	1.0
	DRB1*1501 (DRw15) DRB1*1503 (DRw15) (HLA-DR3)	2.4[†]	2.7[†]	0.13	0.14	1.0
HLA-B [18]	*0801 (B8)	1.8	1.4	0.12	0.03	0.9
C4A [20]	*Q0 (Null)	3.0	3.6	0.1	0.07	0.9
TNF-α [27]	−308A	3.8	2.7	0.12	0.04	0.9
HSP70-2 [25]	8.5-kilobase Pst1	2.1	2.6	0.36	0.61	1.2
Loci outside the HLA region						
Mannose-binding protein (54) [28,66]	GAC	1.6	1.4	0.19	0.11	1.0
FCγRIIa [29]	R131	1.3	1.8	0.51	0.48	1.0

*The population risk ratio has been calculated using the mean of the genotypic risk ratio in Europeans and the genotypic risk ratio in Africans, calculating genotype frequencies assuming Hardy–Weinberg equilibrium. [†]Risk ratios have been calculated by averaging frequencies of HLA-DR2 and HLA-DR3 across studies [21,65]. HLA, human leucocyte antigen; HSP, heat shock protein; TNF, tumour necrosis factor.

DRB1−4) were identified within the HLA-DR family, and their alleles were numbered (*0101, *0102, …) according to their nucleotide sequences. In addition to the genes for HLA antigens, the HLA region includes the genes for complement proteins C2 and C4, the heat shock protein HSP70-2 and tumour necrosis factor-α.

A difficulty in interpreting the associations with disease that have been reported for genes in the HLA region is that in this region there are strong associations between alleles at nearby loci. This allelic association is termed 'linkage disequilibrium' by geneticists. Association between alleles at different loci is a consequence of their origin on the same ancestral chromosomes. Linkage disequilibrium is commonly observed when two loci are so close together (usually less than 1 million base-pairs) that this association has not had time to decay through recombination.

Alleles at the C2 and C4 loci are in strong linkage disequilibrium with alleles at other loci within the HLA region. SLE in Europeans is associated with allele B8 at the HLA-B locus [18], with allele *0501 at the DQA1 locus [19], with alleles DR2 and DR3 at the HLA-DR locus [20,21], and with the null allele at the C4A locus [20]. Risk ratios associated with DR3 loci are generally similar in African-Americans (2.7) to the risk ratios in Europeans (2.4). The associations with SLE of B8 at the HLA-B locus and DR3 at the HLA-DR locus are no longer detectable when their association with the null allele at the C4A locus is taken into account [20,22,23], but the association of allele *0501 at the DQA1 locus remains statistically significant after controlling for presence of the C4A null allele [19].

The frequency of the C4A null allele is no higher in African-Americans than in Europeans.

Family-based studies of transmission of haplotypes from parents to affected offspring are required to establish definitively which of these loci influence risk for SLE directly, and which are associated with disease only through linkage disequilibrium with other loci that influence risk directly. Table 1 shows that none of the alleles at loci HLA-B, HLA-DR or C4A that are associated with SLE are more common in west Africans than in Europeans.

A 28 base-pair deletion allele at the C2 locus is strongly associated with SLE, but its frequency is so low (1% in Europeans and zero in west Africans) that this mutation does not contribute to the overall population risk ratio for SLE [24]. The 8.5-kilobase-pair allele at the heat shock protein HSP70-2 locus has been shown to be associated with SLE in African Americans (independent of associations with the null allele at the C4 locus or HLA-DR3) [25]. The frequency of this allele is higher in Africans (0.6) than in Europeans (0.4), but this could account only for a risk ratio of 1.2 between west Africans and Europeans. The −308A allele of the tumour necrosis factor-α gene is independently associated with an increased risk for SLE in Europeans, but the frequency of this allele is not higher in African Americans (0.04) than it is in Europeans (0.12) [26,27].

Associations with loci outside the human leucocyte antigen region
Mannose-binding protein is a serum protein that can activate complement. Variant alleles exist in which either

Table 3

Ethnic variation in prevalence of rheumatoid arthritis: studies using radiography for primary screening

Ethnicity	Location	Age range (years)	Criteria	Prevalence/1000	
				Males	Females
Chippewa Native American	Minnesota, USA [34]	>18	ACR	48	59
Pima Native American	Arizona, USA [32]	>30	ARA*	50	60
Pima Native American	Arizona, USA [63]	>20	ARA	32	70
Afro-American	NHES, USA [37]	18–79	ARA*	7	16
European-American	NHES, USA [37]	18–79	ARA*	7	16
European-American	Sudbury, USA [64]	15–75	ARA*	5	15

ARA* refers to diagnosis based on modified ARA criteria. ACR, American College of Rheumatology; ARA, American Rheumatism Association; NHES, National Health Examination Survey.

Table 4

Ethnic variation in prevalence of rheumatoid arthritis: studies using two-stage screening

Ethnic origin	Location	Age (years)	Prevalence/1000		
			Males	Females	Both
Pima	Arizona, USA [31]	>20	10	31	
Yakima	Washington, USA [33]	18–79	N/A	34	
European	Minnesota, USA [35]	35–85	7.4	14	
European and Afro-American	NHANES, USA [30]	18–79	6	15	
European	Manchester, UK [36]	45–65			8
Afro-Caribbean	Manchester, UK [36]	44–66			2.9

NHANES, National Health and Nutrition Examination Survey.

codon 54 or 57 of the mannose-binding protein gene is altered, causing deficient complement activation. These variant alleles are associated with SLE in both African and European populations, with risk ratios of about 1.5. The frequency of the codon 54-variant allele is lower (0.11) in Africans than in Europeans (0.19) [28], however, so this cannot account for higher risk in west Africans.

Allelic variants of FcγRIIa influence the ability to clear circulating immune complexes, and have been associated with risk for lupus nephritis in African-Americans and Europeans. In a multicentre study of 214 SLE patients and 100 non-SLE control-individuals the odds ratio for the risk for SLE in FcγRIIA-H131/R131 heterozygotes and R131/R131 homozygotes compared with H131/H131 homozygotes was 1.8 [29]. R131/R131 homozygotes and R131/H131 heterozygotes are not more common in African populations (73% compared with 76% in Europeans), and thus the higher risk in African-Americans is not accounted for.

In summary, the higher risk for SLE in west Africans compared with the risk in Europeans cannot be accounted for

by higher frequencies of high-risk alleles or haplotypes at any of the loci at which variation is associated with risk for SLE within populations.

Ethnic variation in risk for rheumatoid arthritis

Comparisons of RA prevalence in different populations are difficult to interpret unless standardized methods are used. Prevalence estimates depend on the methods used to ascertain cases, whether point prevalence or period prevalence is recorded, and whether 'probable' or only 'definite' cases are included. The highest prevalence rates are obtained in surveys that include radiographic screening of all individuals sampled. The results are summarized in Table 3. Lower prevalence rates are generally found in surveys that rely on a two-stage screening process – questionnaire screening for RA followed by radiographs of those who report symptoms. The results are summarized in Table 4. Only two surveys of the European-American population included radiographic screening of the entire sample [30].

The highest prevalence rates of RA have been recorded in Native American populations such as the Pima, Chippewa

and Yakima tribes [30–34]. In surveys based on radio-graphic screening of the entire population, prevalence in adults from these populations has been estimated to be in the range 32–48 per 1000 for men and 59–70 per 1000 for women. By comparison, in the National Health Examination Survey conducted from 1960 to 1962, using bilateral hand and foot radiographs and serologic tests for rheumatoid factor, RA prevalence in the general US population aged 18–79 years was estimated at 7 per 1000 in men and 16 per 1000 in women.

Lower prevalence rates have been recorded in surveys that have used a two-stage screening process: 31–34 per 1000 in Native American women [33]. In that study the case definition was more restrictive than in other studies. For comparison, in European-Americans and African-Americans, surveys using a two-stage screening process have yielded prevalence rates of 6–7 per 1000 in men and 14–15 per 1000 in women [30,35].

There is no consistent evidence that any other ethnic group apart from Native Americans is at unusually high or low risk for RA. Although people of Afro-Caribbean origin (mainly first-generation and second-generation migrants) in the UK appear to have a lower prevalence of RA compared with that in the general population [36], in the USA the prevalence of RA is similar in African-Americans and European-Americans [30,37]. In South Africa, prevalence of RA in urban African populations has been estimated to be around 10 per 1000 (both sexes combined), which is similar to the estimates for populations of European descent [38]. However, prevalence of RA in rural South African populations was noted to be significantly lower than in urban populations (0.87% compared with 3.3%, respectively), suggesting that environmental factors play an important role that would explain such marked differences in genetically closely related communities [39].

Prevalence surveys of Chinese populations in native China [40] and Hong Kong [41], using a two-stage screening protocol with questionnaires and radiographs of all those with symptoms, have found prevalence of RA to be lower (around 5 per 1000 in women) than that in comparable surveys of populations of European descent. It is possible that some patients may have been overlooked who did not qualify for radiographic examination. There are no surveys of prevalence of RA in Chinese Americans that can be compared with rates in Europeans.

Other rheumatic conditions such as ankylosing spondylitis appear also to be commoner in Native Americans than in Europeans. Using four different subsamples of Native American adult males aged over 25 years, the prevalence of sacro-iliitis grade 2 or more on radiography (based on Rome 1962 criteria) was estimated at 100 per 1000 [42].

This compares with a prevalence of about 10 per 1000 in men from the general US population.

Studies that compared the sequences of HLA antigens have pointed to a direct effect of variation at the HLA-DRB1 locus on susceptibility to RA [43]. At this locus all of the alleles that are associated with high risk for RA share an identical amino acid sequence in positions 70–74: the 'shared epitope' [44]. These alleles include *0101 (included in DR1), *0401/*0404/*0405/*0408 (included in DR4), *1001 (DRw10), *1402 (included in DRw14) and *1601/*1602 (DRw16) [45–48]. Although others have concluded that the high risk for RA in Native Americans compared with Europeans can be largely explained by the higher frequency of shared epitope alleles in Native Americans, this interpretation has not been supported by calculation of the population risk ratio. In Europeans the total frequency of 'shared epitope' alleles is about 25%, mainly alleles *0101, *0401 and *0404. Thus, approximately 44% of Europeans have at least one copy of a 'shared epitope' allele. In Native Americans about 82% have at least one copy of allele *1402 [48], from which the allele frequency can be estimated as 0.58. The frequency of other 'shared epitope' alleles in Native Americans is about 0.04 [49], and thus approximately 86% of Native Americans have at least one copy of a 'shared epitope' allele. In both Europeans and Native Americans, the risk ratio for RA associated with the presence of at least one copy of a 'shared epitope' allele is about 2.7 [45,48]. On this basis, we can calculate the population risk ratio generated by the HLA-B1 locus to be about 1.4 in Native Americans compared with that in Europeans. This accounts for only a small fraction of the observed risk ratios of approximately 4 between these two groups.

Relation of European admixture to rheumatic disease
As a result of maritime expansion during the past 500 years, admixture between populations that were previously isolated from one another has occurred on a large scale in many parts of the world. Where ethnic differences in disease risk have been observed and admixture between high-risk and low-risk ethnic groups has occurred, studies of these admixed populations can help to advance understanding of the aetiology of disease in several ways. First, studying the relationship of disease risk to the proportionate admixture of individuals can help to distinguish between genetic and environmental explanations for the ethnic difference in disease risk. Second, studying the form of the relationship can yield information about the underlying genetic model. Finally, where a relationship of disease risk to admixture has been identified, it is possible in principle to exploit recent admixture to map the genes that underlie this relationship in a manner that is analogous to linkage analysis of an experimental cross [50,51]. In order to estimate the proportionate admixture of individu-

als, it is necessary to type large numbers of markers that are highly informative for ancestry. Until recently such markers have not been available.

Admixture as a confounder of genetic associations

In most genetic study designs, admixture is considered as a nuisance variable because variation in individual admixture ('hidden population stratification') can generate associations of alleles at marker loci with disease even though the marker locus is not linked to any gene that influences disease risk. In an admixed population, any trait that is present at higher frequency in an ethnic group will show positive association with any allele that happens to be common in that group [52]. For example, risk for type 2 (noninsulin-dependent) diabetes is far higher in Pima Native Americans than in Europeans. In the Pima population, the presence of haplotype $Gm^{3;5,13,14}$ is inversely associated with diabetes [53]. This association arises not because the Gm locus is linked to diabetes, but because haplotype $Gm^{3;5,13,14}$ is more common in Europeans (frequency 0.67) than in unadmixed Native Americans (frequency 0.01), and the proportion of European ancestry varies within the Pima population. The inverse relation of risk for diabetes to the proportion of European admixture within this population leads to inverse associations of diabetes with markers of European admixture, such as haplotype $Gm^{3;5,13,14}$.

In an epidemiological study, a confounder is a factor that is associated with the exposure under study and is independently associated with disease risk. In this sense, admixture is a confounder because it is associated with the 'exposure' under study (an allele at the marker locus) and independently associated with disease risk. Thus, for instance, an association between allele Dw16 at the HLA-DRB locus and RA in Mexican-Americans could arise simply because allele Dw16 is more common in those with a higher proportion of Native American ancestry.

To eliminate this confounding by admixture, one possible strategy is to control for admixture in the design by typing parents of cases, and using the two untransmitted parental alleles as controls for the two alleles transmitted to each affected individual. This is a 'family-based' association study design [54]. An alternative, when parents of patients are not available, is to control for admixture in the analysis. For this it is necessary to type cases and controls for a set of markers that have different allele freqencies in the two (or more) ancestral populations, and are unlinked to the locus under study. The association can then be adjusted either by estimating admixture directly from the observed marker data, or indirectly by using the marker data to derive a measure of genetic distance between cases and controls [55]. An extension of this approach would be to establish whether, for instance, the difference in risk for RA between Native Americans and Europeans can be accounted for by the

HLA region. For this one could study cases and controls from the Mexican-American population, within which the proportion of Native American admixture varies widely. By typing multiple markers within the HLA region, the ancestry of the HLA-DRB locus could be assigned as 0, 1 or 2 alleles of Native American ancestry. Estimating the strength of association between RA and the proportion of Native American ancestry after adjusting for ancestry at the HLA-DRB locus would yield an estimate of the risk ratio between the two populations that is not explained by this locus.

Relationship of disease risk to admixture

Within an admixed population, ancestry of the genome will vary between individuals. This variation in individual admixture can be maintained by continuing gene flow from the ancestral populations or by social stratification, as in Mexican-Americans in whom socioeconomic status is related to the proportion of European ancestry [56]. If the ancestral populations differ in risk for disease, and this difference in risk has a genetic basis, we can predict that within the admixed population there will be a relationship between admixture and disease risk. If, for instance, the higher risk for SLE in west Africans compared with the risk in Europeans results from higher frequencies in west Africans of alleles (at one or more loci) that increase the risk for SLE, then within a population of mixed west African/European descent risk will vary with the proportion of the genome that is of African ancestry. There will therefore be a linear relationship between risk for SLE and the number of alleles (0, 1 or 2 alleles) that have African ancestry at one of the disease loci. In the language of statistical genetics this is an 'additive' model, in the sense that additive effects account for most of the genetic variance.

In order to understand how genetic effects on a trait are partitioned into 'additive' and 'dominant' effects, we can imagine that the risk for disease is plotted against the number of copies of the high risk allele (0, 1 or 2) that make up the genotype. This will give a graph with three points, weighted by the frequencies with which the three genotypes occur in the population. The additive variance attributable to the locus is the variance accounted for by a regression line fitted to these three points. The dominance variance attributable to the locus is the residual variance (variance about the regression line) in this regression model. If the regression line fits the three data points perfectly (linear relationship of disease risk to number of copies of the high risk allele), there is no dominance variance and all of the genetic variance attributable to the locus is accounted for by additive effects [57].

An alternative genetic model for the high risk for SLE in African-American and Afro-Caribbean populations could be that admixture itself leads to increased risk. Admixture between two populations that have different allele frequencies at a locus leads to an increase in heterozygosity

(the proportion of individuals who are heterozygous) at this locus, in comparison with the average heterozygosity of the two populations between which this admixture occurred. To take an extreme example, at the FY (Duffy blood group) locus, the frequency of the null allele is 0 in Europeans and 1 in west Africans [11]. The proportion of individuals who are heterozygous is 0 in unadmixed Europeans and 0 in west Africans, and can be calculated from the Hardy–Weinberg theorem as 32% ($2 \times 0.20 \times 0.80$) in an homogenous African-American population in which 20% of the genome is of European ancestry.

If the risk for SLE is higher in individuals who are heterozygous at one or more loci that controls immune function than in individuals who are homozygous at these loci, the risk will be higher in people of mixed west African/European descent than in unadmixed west Africans and unadmixed Europeans. The highest risk would be in the first generation of mixed descent (equivalent to the F_1 generation in an experimental cross), who have one African and one European allele at each locus. In statistical terms this would be an 'overdominant' model in that dominance effects account for most of the genetic variance [57]. An experimental model for this is the cross between New Zealand Black and New Zealand White strains of mice. In offspring that have one parent from each of these two strains, the risk for lupus is higher than in either of the two original inbred strains [58]. Because human ethnic groups are not inbred strains, the effect of admixture on heterozygosity is likely to be small except at a few loci where different alleles have been fixed in Europeans and west Africans.

Recent studies have shown that the average proportion of European admixture in African-Americans is lower than previously estimated (14–22% in most of the populations studied) and varies only over a narrow range [11]. This narrow range of variation in individual admixture means that studies of the relationship of disease risk to individual admixture in African-American populations have low statistical power. Other populations may exist, for instance in the Caribbean, where admixture varies over a wider range. As yet there has been only one study on the relation of SLE to admixture in humans [59]. Seventy-two cases and 79 controls were sampled from the African-American population of Baltimore. The average proportion of European admixture, estimated by typing classical protein polymorphisms (blood groups, erythrocyte enzymes and serum proteins), was estimated to be the same (29%) in cases and controls. The markers used were not specified. To estimate individual admixture accurately would require at least 40 markers, and to estimate the slope of the relationship of disease risk to admixture accurately would require larger numbers of cases and controls.

The only study to examine the relation of RA to admixture was a case–control study in African-Americans [60]. Eigh-

teen African-American probands with RA and 15 non-RA families were studied. Extended haplotypes of European ancestry were more common in the cases than in the controls, even though haplotypes of European ancestry that were known to be associated with RA had been excluded from the analysis. The interpretation of this result depends on whether this association with alleles of European ancestry is specific to the HLA region (implying that haplotypes in this region that are more common in Europeans predispose to RA) or whether it is present over the entire genome (implying that other unidentified loci underlie the association of RA with European ancestry in African-Americans).

There are no studies of the relationship of RA risk to European admixture in people of mixed European/Native American descent. The Mexican-American population of the USA is of mixed European/Native American descent, and the proportion of Native American ancestry is related to socioeconomic status. In a case–control study in Mexican Americans [61], the associations between RA and HLA-DRB1 alleles containing the shared epitope sequence were of similar magnitude to those in Europeans. No attempt was made to control for the proportion of Native American admixture, however, so that it is not possible to establish whether the association was specifically with the shared epitope, or more generally with the proportion of Native American ancestry.

Mapping genes underlying ethnic differences in disease risk
Allele-sharing and association studies
Current attempts to map genes that influence rheumatic diseases are based on two main approaches: analysis of identity-by-descent sharing ('allele-sharing studies') in families with multiple affected members, or association studies of polymorphisms in candidate genes [62]. Both approaches have limitations. Allele-sharing studies lack adequate power to detect genes of modest effect. Unless a single locus accounts for a sibling recurrence risk ratio of 2 or more [62], very large numbers of families are required for adequate power to detect this locus and map it to a small region of the genome. Association studies have far greater statistical power than allele-sharing studies, but depend on knowing which gene to look at. Genome-wide association studies may eventually become possible, but will require the ability to type more than 100 000 marker polymorphisms in each individual studied.

### Study designs that exploit admixture to map genes
A far more powerful method for mapping genes for complex traits is available to experimental geneticists, who detect linkage by studying hybrid populations formed by crossing two inbred strains that differ in the trait under study. The genomic regions that contain loci that underlie this difference in the trait can be detected by testing for association of the trait with inheritance of marker alleles

from the strain that has higher trait values. The same principle can be applied in humans to detecting genes that underlie ethnic differences in disease risk, by studying populations formed by admixture between a high-risk and a low-risk population. Figure 1 shows, for a pair of chromosomes in an F_2 sibship of mixed European/African descent, how crossing two ancestral strains generates variation in states of ancestry in the F_2 and subsequent generations. In the second (F_2) and subsequent generations of mixed descent, ancestry on chromosomes varies between two states, European and African, so that at any given locus an individual of mixed descent may have 0, 1 or 2 alleles of African ancestry. On chromosomes of mixed descent, the states of ancestry at two linked loci will be correlated with each other, because markers are likely to be inherited together if they are near each other on the chromosome. The more tightly linked the two loci are, the higher will be the correlation. This autocorrelation of states of ancestry on chromosomes inherited from parents of mixed descent conveys information about linkage, and can be used to map disease genes. Suppose, for instance that a disease locus accounts for a risk ratio of 3 for SLE in west Africans compared with Europeans. Within a population of mixed west African/European descent, the risk for SLE will be three times higher in those with 2 alleles of African ancestry at the disease locus than in those with 0 alleles of African ancestry at this locus, with other factors held constant. In practice the most efficient design is to study only affected individuals, and to test for regions of the genome where the proportion of alleles that have African ancestry is higher than expected. By comparing the observed proportion of alleles at the marker locus that have African ancestry with the expected proportion given the admixture of each individual's parents, association with the ancestry of alleles at loci unlinked to the disease is eliminated [51]. The analysis does not require parents to be available for typing, because the admixture of an individual's parents can be estimated by typing markers in that individual.

Experimental crosses are generally conducted with inbred strains, so that, in order to assign ancestry at all points on the genome in a hybrid individual, it is sufficient to type markers at loci where different alleles have become fixed in each of the two parental strains. Because human ethnic groups are not inbred strains, marker loci at which different alleles have become fixed in each of the two founding populations are rare. Thus, the information conveyed by typing a single marker will not usually be sufficient to assign the ancestry of each allele at the marker locus to one of the two founding populations. This problem can be overcome by using a multipoint statistical method to combine information from all markers to estimate ancestry at each locus [51]. Although advanced statistical methods are required to apply this approach in practice, the underlying principles on which it relies to detect linkage are simple. For example, to estimate ancestry at each locus in

Figure 1

European   African   European   African

F_1 generation

F_2 generation
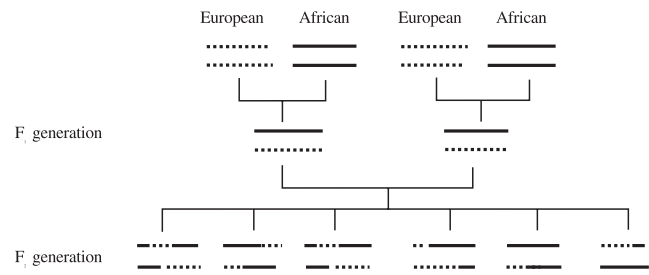
How hybridization gives rise to variation in ancestry on chromosomes of mixed descent. Chromosomal segments of European and African descent are represented by broken lines and continuous lines, respectively.

a population of mixed European/west African descent, one would first choose a set of marker polymorphisms that have large differences in allele frequencies between European and Africans, spacing these markers at much higher density than the transitions of ancestry that occur on the chromosomes of individuals of mixed descent. If we type these markers in an individual together with this individual's parents, siblings or offspring, we can assign haplotypes and reconstruct the sequence of marker alleles on each chromosome. Over any short interval, a haplotype in an individual of mixed descent will be effectively unique to one of the two founding populations. By combining information from these marker alleles, we can reduce the uncertainty with which ancestry at each locus is assigned as 0, 1 or 2 alleles African by descent. This approach, in which information about ancestry on chromosomes of mixed descent is extracted from marker genotypes, differs fundamentally from conventional linkage disequilibrium mapping, which relies on detecting allelic association. By combining information from all markers simultaneously in a multipoint analysis, it is possible to extract all of the information about linkage that is generated by admixture, which is not possible in a conventional allelic association study unless the markers are perfectly informative for ancestry.

The advantage of this approach is that, in comparison with allele-sharing designs, study designs that exploit admixture have far greater statistical power to detect genes of modest effect if these genes contribute to the ethnic difference in disease risk. There are fundamental statistical reasons for this [52]; allele-sharing designs rely on an indirect comparison (allele-sharing in concordant pairs higher than expected), whereas linkage analysis of a cross relies on a direct comparison (risk for disease in those with 0, 1 or 2 alleles with ancestry from the high-risk strain).

Application of this approach in practice will require special sets of markers for genome-wide assignment of ancestry:

African versus European ancestry for studies of SLE, and Native American versus European ancestry for studies of RA. Simulations indicate that for each pair of ancestral populations, a set of about 1000 markers that have large differences in allele frequency (>0.6) between the two populations will be sufficient for a genome search [51]. Several possible strategies for assembling such markers are available: screening libraries of microsatellite of single nucleotide polymorphism (SNP) marker loci to identify those that have large differences in allele frequencies; using representational difference analysis (a subtractive hybridization technique) to generate base sequences that are common in one population but absent in another; and screening multiple SNPs in candidate genes to define intragenic haplotypes that are informative for ancestry. With the availability of large public repositories of SNP markers, identifying a subset of markers that are informative for ancestry will become easier.

Conclusion

The possibilities for exploiting admixture to investigate the aetiology of rheumatic disease have not been fully explored. Studies of how the risk for disease varies with individual admixture could help to distinguish between possible genetic models for the high risk for SLE in people of west African descent and the high risk for RA in Native Americans. Further studies in which markers are used to assign ancestry in regions of interest, or across the entire genome, could help to map, and eventually identify genes that underlie these ethnic differences in risk for rheumatic disease.

Acknowledgements

References
1. Reid DD: **The future of migrant studies.** *Is J Med Sci* 1971, **12**: 1592–1596.
2. Dean G, McLoughlin H, Brady R, *et al*: **Multiple sclerosis among immigrants in Greater London.** *Br Med J* 1976, 1:861–864.
3. Elian M, Nightingale S, Dean G: **Multiple sclerosis among United-Kingdom born children of immigrants from the subcontinent, Africa and the West Indies.** *J Neurol Neurosurg Psychiatry* 1990, **53**:906–911.
4. Johnson AE, Gordon C, Hobbs FD, Bacon PA: **Undiagnosed systemic lupus erythematosus in the community**. *Lancet* 1996, **347**: 367–369.
5. Hopkinson ND, Doherty M, Powell RJ: **Clinical features and race specific incidence/prevalence rates of SLE in a geographically complete cohort of patients**. *Ann Rheum Dis* 1994, **53**:675–680.
6. Johnson AE, Gordon C, Palmer RG, Bacon PA: **The prevalence and incidence of systemic lupus erythematosus in Birmingham, England.** *Arthritis Rheum* 1995, **38**:551–558.
7. Samanta A, Roy S, Feehally J, Symmons DPM: **The prevalence of diagnosed systemic lupus erythematosus in Whites and Indian Asian immigrants in Leicester city, UK.** *Br J Rheumatol* 1992, **31**: 679–682.
8. Greenwood BM: **Autoimmune disease and parasitic infection in Nigerians.** *Lancet* 1968, **2**:380–382.
9. Fessel J: **Systemic lupus erythematosus in the community: incidence, prevalence, outcome and first symptoms; the high prevalence in Black women.** *Arch Intern Med* 1974, **134**:1027–1035.
10. Bae S, Fraser P, Liang MH: **The epidemiology of systemic lupus erythematosus in populations of African ancestry.** *Arthritis Rheum* 1998, **41**:2091–2099.
11. Shriver MD, Smith MW, Jin L, *et al*: **Ethnic-affiliation estimation by use of population-specific DNA markers.** *Am J Hum Genet* 1997, **60**:957–964.
12. Nossent JC: **Systemic lupus erythematosus on the Caribbean island of Curacao: an epidemiological investigation.** *Ann Rheum Dis* 1992, **51**:1197–1201.
13. Frank AO: **Apparent predisposition to systemic lupus erythematosus in Chinese patients in west Malaysia.** *Ann Rheum Dis* 1979, **39**:266–269.
14. Wang F, Wang CL, Tan CT, Manivasagar M: **Systemic lupus erythematosus in Malaysia: a study of 539 patients and comparison of prevalence and disease expression in different racial and gender groups.** *Lupus* 1997, **6**:248–253.
15. Serdula MK, Rhoads GG: **Frequency of sytemic lupus erythematosus in different ethnic groups in Hawaii.** *Arthritis Rheum* 1979, **22**:328–333.
16. Catalano MA, Hoffmeier M: **Frequency of SLE among the ethnic groups of Hawaii [abstract].** *Arthritis Rheum* 1989, **32 (suppl 4)**:s30.
17. Hart HH, Grigor RR, Caughey DE: **Ethnic difference in the prevalence of systemic lupus erythematosus.** *Ann Rheum Dis* 1983, **42**: 529–532.
18. Freedman BI, Spray BJ, Heise ER, *et al*: **A race controlled human leucocyte antigen frequency analysis in lupus nephritis.** *Am J Kidney Dis* 1993, **21**:378–382.
19. Davies EJ, Steers G, Ollier WE, *et al*: **Relative contributions of HLA-DQA and complement C4 loci in determining susceptibility to systemic lupus erythematosus.** *Br J Rheumatol* 1995, **34**:221–225.
20. Howard PF, Hochberg MC, Bias WB, Arnett FCJ, Mclean RH: **Relationship between C4 null genes, HLA-D antigens, and genetic susceptibility to lupus in Caucasians and Black Americans.** *Am J Med* 1986, **81**:187–193.
21. Woodrow JC: **Immunogenetics of systemic lupus erythematosus.** *J Rheumatol* 1988, **15**:197–199.
22. Goldstein R, Arnett FC, Mclean RH, Bias WB, Duvic M: **Molecular heterogeneity of complement component C4-null and 21-hydroxylase genes in systemic lupus erythematosus.** *Arthritis Rheum* 1988, **31**:736–744.
23. Green JR, Montasser M, Woodrow JC: **The association of HLA-linked genes with systemic lupus erythematosus.** *Am J Hum Genet* 1986, **50**:93–96.
24. Sullivan KE, Petri MA, Schmeckpeper BJ, Mclean RH, Winkelstein JA: **Prevalence of a mutation causing C2 deficiency in systemic lupus erythematosus.** *J Rheumatol* 1994, **21**:1128–1133.
25. Jarjour W, Reed AM, Gauthier J, Hunt SI, Winfield JB: **The 8.5-kb Pstl allele of the stress protein gene, Hsp 70-2.** *Hum Immunol* 1996, **45**:59–63.
26. Danis VA, Millington M, Hyland V, *et al*: **Increased frequency of a tumour necrosis factor alpha uncommon gene polymorphism in rheumatoid arthritis and systemic lupus erythematosus.** *Dis Markers* 1995, **12**:127–133.
27. Sullivan KE, Wooten C, Schmeckpeper BJ, Goldman D, Petri MA: **A promoter polymorphism of tumour necrosis factor alpha associated with systemic lupus erythematosus in African-Americans.** *Arthritis Rheum* 1997, **40**:2207–2211.
28. Davies EJ, Snowden N, Chantal Hillarby M, *et al*: **Mannose-binding protein gene polymorphism in systemic lupus erythematosus.** *Arthritis Rheum* 1995, **38**:110–114.
29. Salmon JE, Millard S, Schachter LA, *et al*: **Fc gamma RIIA alleles are heritable risk factors for lupus nephritis in African Americans.** *J Clin Invest* 1996, **97**:1348–1354.
30. Lawrence RC, Helmick CG, Arnett FC, *et al*: **Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the US.** *Arthritis Rheum* 1998, **41**:778–799.
31. Hirsch R, Lin J, Scott WWJ, *et al*: **Rheumatoid arthritis in the Pima Indians: the intersection of epidemiologic, demographic, and genealogic Data.** *Arthritis Rheum* 1998, **41**:1464–1469.
32. O'Brien WM, Bennett PH, Burch TA, Bunim JJ: **A genetic study of rheumatois arthritis and rheumatoid factor in Blackfeet and Pima Indians.** *Arthritis Rheum* 1967, **10**:163–178.
33. Palmer Beasley R, Willkens RF, Bennett PH: **High prevalence of rheumatoid arthritis in Yakima Indians.** *Arthritis Rheum* 1973, **16**: 743–748.
34. Harvey J, Lotze M, Stevens MB, Lambert G, Jacobson D: **Rheumatoid arthritis in a Chippewa Band.** *Arthritis Rheum* 1981, **24**:717–721.
35. Gabriel SE, Crowson CS, O'Fallon WM: **The epidemiology of rheumatoid arthritis in Rochester, Minnesota, 1955–1985.** *Arthritis Rheum* 1999, **42**:415–420.

36. MacGregor AJ, Riste LK, Hazes JMW, Silman AJ: **Low prevalence of rheumatoid arthritis in Black-Caribbeans compared with Whites in inner city Manchester.** *Ann Rheum Dis* 1994, **53**:293–297.
37. Engel A, Roberts J, Burch TA: **Rheumatoid arthritis in adults: US 1960–62.** *Vital Health Stat* 1966, **11**:1–43.
38. Solomon L, Robin G, Valkenburg HA: **Rheumatoid arthritis in an urban South African Negro population.** *Ann Rheum Dis* 1975, **134**:128–135.
39. Solomon L, Beighton P, Valkenburg HA, Robin G, Soskolne CL: **Rheumatic disorders in the South African Negro. Part I. Rheumatoid arthritis and ankylosing spondylitis.** *S Afr Med J* 1975, **49**:1292–1296.
40. Palmer Beasley R, Bennett PH, Chun Lin C: **Low prevalence of rheumatoid arthritis in Chinese.** *J Rheumatol* 1983, **10 (suppl 10)**:11–15.
41. Lau E, Symmons D, Bankhead C, et al: **Low prevalence of rheumatoid arthritis in the urbanized Chinese of Hong Kong.** *J Rheumatol* 1992, **20**:1133–1137.
42. Gofton JP, Bennett PH, Smythe HA, Decker JL: **Sacroiliitis and ankylosing spondylitis in North American Indians.** *Ann Rheum Dis* 1972, **31**:474–481.
43. Wordsworth BP, Lanchbury JS, Sakkas LI, et al: **HLA-DR4 subtype frequencies in rheumatoid arthritis indicate that DRB1 is the major susceptibility locus within the HLA class II region.** *Proc Natl Acad Sci USA* 1989, **86**:10049–10053.
44. Gregersen PK, Silver J, Winchester RJ: **The shared epitope hypothesis.** *Arthritis Rheum* 1987, **30**:1205–1213.
45. Evans TI, Han J, Singh R, Moxley G: **The genotypic distribution of shared-epitope DRB1 alleles suggests a recessive mode of inheritance of the rheumatoid arthritis disease-susceptibility gene.** *Arthritis Rheum* 1995, **38**:1754–1761.
46. Nepom GT, Seyfried CE, Holbeck SL, Wilkse KR, Nepom BS: **Identification of HLA-Dw14 genes in DR4+ rheumatoid arthritis.** *Lancet* 1986, **ii**:1002–1006.
47. Willkens RF, Nepom GT, Marks CR, Nettles JW, Nepom BS: **Association of HLA-Dw16 with rheumatoid arthritis in Yakima Indians.** *Arthritis Rheum* 1991, **34**:43–47.
48. Williams RC, Jacobsson LTH, Knowler WC, et al: **Meta-analysis reveals association between most common class II haplotype in full heritage Native Americans and rheumatoid arthritis.** *Hum Immunol* 1995, **42**:90–94.
49. Nelson JL, Boyer GS, Templin DW, et al: **HLA antigens in Tlingit Indians with rheumatoid arthritis.** *Tissue Antigens* 1992, **40**:57–63.
50. Chakraborty R, Weiss KM: **Frequencies of complex diseases in hybrid populations.** *Am J Physiol Anthropol* 1986, **70**:489–503.
51. 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.
52. Lander ES, Schork NJ: **Genetic dissection of complex traits.** *Science* 1994, **265**:2037–2048.
53. Knowler WC, Williams RC, Pettitt DJ, Steinberg AG: **Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture.** *Am J Hum Genet* 1988, **43**:520–526.
54. Thomson G: **Mapping disease genes: family-based association studies.** *Am J Hum Genet* 1995, **57**:487–498.
55. Pritchard JK, Rosenberg NA: **Use of unlinked genetic markers to detect population stratification in association studies.** *Am J Hum Genet* 1999, **65**:220–228.
56. Relethford JH, Stern MP, Gaskill SP, Hazuda HP: **Social class, admixture, and skin color variation in Mexican-Americans and Anglo-Americans living in San Antonio, Texas.** *Am J Physiol Anthropol* 1983, **61**:97–102.
57. Kempthorne O: *An Introduction to Genetic Statistics.* New York: Wiley, 1997.
58. Drake CG, Rozzo SJ, Hirschfield HF, et al: **Analysis of New Zealand Black contribution to a lupus like renal disease.** *J Immunol* 1995, **154**:2441–2447.
59. Bias WB, Hochberg MC, Mclean RH, Machan C: **Systemic lupus erythematosus joint report.** In: *HLA 1991: Proceedings of the Eleventh International Histocompatability Workshop and Conference.* Edited by Tsuji K, Aizawa M, Sasazuki T. Oxford: Oxford University Press, 1992:740–745.
60. Fraser PA, Yunis MD, Alper CA: **Excess admixture proportion of extended MHC haplotypes of Caucasian origin among rheumatoid arthritis associated haplotypes in African-Americans and Afro-Caribbeans.** *Ethn Health* 1996, **1**:153–159.
61. Del Rincon I, Escalante A: **HLA-DRB1 alleles associated with susceptibility or resistance to rheumatoid arthritis, articular deformities and disability in Mexican Americans.** *Arthritis Rheum* 1999, **42**:1329–1338.
62. Risch N, Merikangas K: **The future of genetic studies of complex human diseases.** *Science* 1996, **273**:1516–1517.
63. Del Puente A, Knowler WC, Pettitt DJ, Bennett PH: **High incidence and prevalence of rheumatoid arthritis in Pima Indians.** *Am J Epidemiol* 1989, **129**:1170–1178.
64. Cathcart ES, O'Sullivan JB: **Rheumatoid arthritis in a New England town: a prevalence study in Sudbury, Massachusetts.** *N Engl J Med* 1970, **282**:421–424.
65. Matsumoto AC, Hochberg MC: **Association between HLA class II alleles and systemic lupus erythematosus in Black Americans: results of a meta analysis [abstract].** *Clin Res* 1989, **37**:857A.
66. Davies EJ, Tikly M, Wordsworth BP, Ollier W: **Mannose binding gene polymorphism in South African SLE [letter].** *Br J Rheumatol* 1998, **37**:465.

Authors' affiliation: Epidemiology Unit, London School of Hygiene and Tropical Medicine, London, UK

Correspondence: Mariam Molokhia, Clinical Research Fellow, London School of Hygiene and Tropical Medicine, Keppel St, London, UK. Tel: +44 (0)171 927 2633; fax: +44 (0)171 580 6897; e-mail: mariam.molokhia@lshtm.ac.uk