Towards a global view of multiple sclerosis genetics

Benjamin Meir Jacobs[1,2†], Michelle Peter[3], Gavin Giovannoni[1,2,4], Alastair Noyce[1,2,5], Huw R Morris[5,6] and Ruth Dobson[1,2,6]

[1]Preventive Neurology Unit, Wolfson Institute of Population Health, Queen Mary University London, London, UK
[2]Department of Neurology, Royal London Hospital, London, UK
[3]NHS North Thames Genomic Laboratory Hub, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

[4]Blizard Institute, Queen Mary University London, London, UK
[5]Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, University College London, London, UK

[6]These authors contributed equally
[†]email: b.jacobs@qmul.ac.uk

Abstract | Multiple sclerosis (MS) is a neuroimmunological disorder of the CNS with a strong heritable component. The genetic architecture of MS susceptibility is well understood in populations of European ancestry. However, the extent to which this architecture explains MS susceptibility in populations of non-European ancestry remains unclear. In this Perspective article, we outline the scientific arguments for studying MS genetics in ancestrally diverse populations. We argue that this approach is likely to yield insights that could benefit individuals with MS from all ancestral groups. We explore the logistical and theoretical challenges that have held back this field to date and conclude that, despite these challenges, inclusion of participants of non-European ancestry in MS genetics studies will ultimately be of value to all patients with MS worldwide.

[H1] Introduction

Individuals of non-European ancestry are systematically under-represented in complex genetics research and remain so despite efforts in recent years to address this disparity[1,2]. For example, despite people of European ancestry constituting only ~16% of the global population, 78% of individuals included in GWAS are of European descent; by contrast, only 11% are of Asian ancestry, 2.4% are of African ancestry, and the remainder comprises other populations[2,3]. This under-representation is a pressing scientific and societal concern[1,4–6].

In the field of multiple sclerosis (MS), GWAS in populations of European ancestry[7,8] have shown that susceptibility to MS is associated with 32 independent signals at the major histocompatibility complex (MHC), 200 additional autosomal loci and one X-chromosome locus. These signals collectively explain up to ~50% of the estimated genetic heritability of MS[7,8]. However, these findings alone tell us little about the genetic architecture of MS in other ancestral populations[9–15].

There are strong ethical and scientific arguments for broadening participation in the genetic study of complex traits and diseases, including MS[1,3]. To some extent, we believe that the distinction between ‘ethical’ and ‘scientific’ arguments here is artificial. The primary ethical concern is that failure to include individuals of diverse ancestry in genetic studies risks perpetuating health inequalities, as developments in personalized genomic medicine that are based on findings from populations of European ancestry might not translate into benefits for people with non-European ancestral backgrounds. Examples of such developments include improved disease prediction with polygenic risk scores, enrolment in genetically-stratified prevention studies guided by risk scores, pharmacogenetics-informed treatment decisions, and genetics-informed prognostication. If genetics begins to inform routine clinical practice in MS, individuals of non-European ancestry could receive inferior clinical care, potentially leading to worse disease-related outcomes. This possibility is particularly concerning given that existing evidence indicates that diagnosis and initiation of treatment are slower for people of non-European ancestry with MS and that people in these populations experience greater disability and higher overall mortality than people of European ancestry. Advances in our understanding of how genetics shapes risk in diverse populations will hopefully address this concern. Examination of more complex ethical aspects is beyond the scope of this article, but these arguments are explored in greater depth elsewhere[2,16–19].

In this Perspective article, we discuss the major scientific benefits of studying MS genetics in populations from diverse ancestral backgrounds (Box 1). We draw on recent

successes of this approach in other complex traits and diseases, including schizophrenia[20], chronic kidney disease[21], cerebrovascular disease[22], blood cell traits[23], lipid traits[24], rheumatoid arthritis (RA)[25], type 1 diabetes mellitus (T1DM)[26,27] and inflammatory bowel disease (IBD)[28,29]. These studies have led to advances in fine-mapping and polygenic risk score prediction, and to the discovery of novel associations within and beyond known loci. We argue that well-powered GWAS in these populations could help to explain variation in MS epidemiology and clinical course, elucidate the ways in which MHC variation affects risk, identify causal variants at known loci, uncover novel risk loci, and improve prediction of MS risk.

[H1] Genetics to understand epidemiology

The Global Burden of Diseases Study revealed global variation in the prevalence of MS[30]. Age-standardized estimates range from 2.0 per 100,000 people in Oceania and 2.8 per 100,000 in central sub-Saharan Africa to 164.6 per 100,000 in high-income North America[30]. Though the estimates for many of the countries with low prevalence are influenced by ascertainment bias and missing data, the differences are striking[30]. Potential contributors to this variation include differences in documentation of cases owing to, for example, differential access to medical facilities and health data recording practices, and differences in the risk of disease owing to environmental exposures and genetic architecture[31].

Rapid changes in the apparent epidemiology of MS provide evidence that non-genetic factors make important contributions to differences between populations. For example, one study published in 2015 indicated that the incidence of MS among women of East Asian or South East Asian descent in British Columbia, Canada, doubled between 1986 and 2010[32]. Such a rapid change cannot be attributable to genetics. Plausible explanations include changes in case ascertainment as a result of, for example, changes in diagnostic criteria, access to healthcare services, and the availability of MRI scans, or increased exposure to MS risk factors, such as smoking and obesity during adolescence.

A complete understanding of the genetic architecture of MS across ancestral populations would help to clarify the proportion of global variation in MS that is due to genetics and the proportion that is due to alternative explanations. Combining a pan-ancestral map of MS genetics with worldwide allele frequencies could even enable prediction of the ‘hidden burden’ of MS in low-resource settings, where prevalence and incidence estimates are limited by access to specialist neurological care and diagnostic facilities[30].

Some evidence indicates that in high-income countries, the incidence of MS is higher among individuals of non-European ancestry than among those of European ancestry. For example, in a study of the Gulf War cohort of US veterans published in 2012, age-standardized incidence rates were highest among Black American people

(12.1 per 100,000 person–years compared with 9.3 per 100,000 person–years in white American people)[33]. In this study, ethnic groups were based on categories recorded in the US Defense Medical Epidemiological Database. This finding was replicated in smaller cohorts, in which the disparities were more pronounced among women[34]. Similarly, in a primary care record-based study conducted in London, UK, British individuals aged <40 years whose ethnicity was recorded as 'Black British' had an equal, if not higher, risk of MS than individuals whose recorded ethnicity was 'White' (OR 1.15, 95% CI 0.81 – 1.62)[35]. Data published in 2022 from the Kaiser Permanente cohort in Southern California, USA, indicate a similar prevalence of MS among Black American people (225.8 per 100,000 people) and white American people (237.7 per 100,000 people)[36]. In this study, race and ethnicity was based on medical records and birth certificates. Taken together, these data suggest that in countries with a high MS prevalence, the risk of MS is similar or slightly higher among Black people than among white people, although the classification of race and ethnicity in these studies does not necessarily provide accurate information about genetic ancestry in these populations.

Studying migrant populations in high-resource settings could also be helpful in determining the proportion of variation in MS risk that can be attributed to genetics alone. Such studies offer a unique opportunity to distinguish the roles of genetic and environmental factors, and to determine critical windows within which environmental factors influence risk. Contemporary studies from Denmark and Sweden support the view that migration from a low-prevalence to a high-prevalence country is associated with an increase in the risk of MS and that this increase is particularly prominent among people who migrate before adolescence[37,38]. In principle, if a migrant population retains similar allele frequencies and linkage disequilibrium structure to individuals who remain in their country of origin, the discordance in disease risk could be used to estimate the heritability of MS, and substantial differences in the estimated heritability could indicate gene–environment interactions.

For example, a population could harbour several common risk alleles for MS that only increase the risk if the individual smokes cigarettes. If people from this population move to a country where the prevalence of MS is higher and exposure to cigarette smoking is higher than in their country of origin, the estimated heritability of MS among this population will be greater than in the population in the country of origin, as these alleles have a greater impact on MS risk with increased exposure to smoking. This scenario is extreme but illustrates the point that heritability estimates could, in principle, be affected by gene–environment interactions.

Though migration studies are an elegant approach to disentangling the relative contributions of genetic and environmental risk factors for MS, sound interpretation of these data relies on the assumption that cases are ascertained with equivalent accuracy in the two countries. This assumption is unlikely to hold true in the case of migration from a low-income country to a high-income country.

[H1] The MHC locus

The role of the MHC locus (6p21) in determining MS susceptibility is well-established[39,40]. However, studying the role of the MHC in populations of non-European ancestry is valuable because population-specific alleles exist and haplotype structures differ between populations (Figure 1). Such studies can reveal the role of HLA alleles that are not present in European populations and clarify independent and/or shared effects of alleles that highlight core disease pathways. Differences between populations of different ancestry could also reveal a role for specific pathogens or selection pressures.

Determining the precise mechanisms by which MHC variation affects MS biology has been challenging owing to the density of genes in this region, the complex linkage disequilibrium, and the existence of long-range haplotypes. A GWAS in populations of European ancestry has identified 32 statistically independent signals within the MHC locus, including class II alleles that increase risk (HLA-DRB1*15:01, HLA-DRB1*03:01,HLA-DRB1*13:03, HLA-DRB1*08:01 and HLA-DQB1*03:02), class I alleles that are protective (HLA-A*02:01, HLA-B*44:02, HLA-B*38:01 and HLA-B*55:01), and some risk variants outside of classical HLA genes[8,41]. Gene–gene interactions also occur at this locus in populations of European ancestry — several alleles modulate the effect of HLA-DRB1*15:01[8,41]. There is little evidence to suggest that the MHC influences MS phenotypes (for example, relapse rate, severity or relapsing–remitting versus progressive disease) besides age of onset in European populations[41–43].

The frequency, distribution and haplotypes of HLA alleles differ between ancestral populations (Figure 1)[44,45]. Some alleles are absent in the European population so their influence on MS risk cannot be studied in populations of European ancestry. Conversely, some MS risk alleles, such as HLA-DRB1*15:01, are rare in populations of non-European ancestry. The class II allele HLA-DRB1*04:05, which is essentially absent in populations of European ancestry, has been associated with MS in Japanese[46], Turkish [47], South American [48], African American[10], and Sicilian populations[49]. The combination of HLA-DRB1*04:05 positivity and HLA-DRB1*15:01 negativity seems to be associated with a distinct clinical phenotype of MS characterized by early onset, a relatively benign course, and an unusually low rate of Epstein–Barr virus seropositivity[46,50,51]; this phenotype contrasts with that in European populations, in which HLA-DRB1*15:01 positivity is associated with earlier onset. Similarly, the HLA-DRB1*15:03 allele is essentially absent in populations of European descent[10,13] but is consistently associated with MS susceptibility in cohorts from Iran, Brazil and Martinique[52–54].

The increased risk of MS associated with HLA-DRB1*15:01-containing haplotypes in populations of European ancestry has been replicated in populations of

Ashkenazi Jewish[55,56], Sardinian[57], African[10,58,59], Hispanic[59,60], Japanese[46] and Indian ancestry[12,61]. This allele and the MS-associated HLA-DRB1*15:03 allele (which is most common among populations of African ancestry) are characterized by an alanine residue at position 71, which alters the peptide groove and might enable presentation of epitopes derived from myelin basic protein[45]. Larger studies in diverse cohorts are required to determine the effects of HLA alleles other than DRB1 alleles on MS risk in populations of non-European ancestry. Combining findings of HLA allelic associations with MS across ancestries could help to determine the common features of MS-associated MHC molecules.

[H2] Ancestral variation and MHC fine mapping

Differences in linkage disequilibrium between HLA alleles across ancestries can help to disentangle the effects of individual alleles. Across the genome, blocks of alleles in linkage disequilibrium are shorter in African populations than in European populations. This difference is due to the greater number of ancestral generations (and therefore meiotic recombination events) without genetic bottlenecks or other causes of loss of genetic diversity[62]. This observation has helped to determine that the association of MHC with narcolepsy is attributable to the DQB1*06:02 allele[63]. Similar approaches have been used in populations of non-European ancestry to demonstrate the association of the non-classical HLA-DRB1*04:01 allele with rheumatoid arthritis[64], and of the HLA-DRB1*04:03 allele with type 1 diabetes mellitus[65].

In MS, the most notable demonstration of the power of this approach was the finding that the association of the European DRB1*15:01–DQB1*06:02 haplotype was driven by DRB1*15 alleles (either 15:01 or 15:03)[13]. This insight was possible because the two alleles are not in complete linkage disequilibrium in African American populations, so their associations could be tested independently in African American[13] and Martinican[54] cohorts. Several years later, the availability of larger European sample sizes and sophisticated statistical approaches for fine mapping confirmed that the association of this haplotype with MS is primarily driven by HLA-DRB1*15:01 in populations of European ancestry[41,66]. The increasing availability of multi-ancestry HLA reference panels, increasing sophistication of algorithms for imputing HLA alleles from single nucleotide polymorphism (SNP) data, and the shift towards MHC sequencing over genotyping are likely to facilitate further insights from cross-ancestral MHC fine mapping[65,67].

[H2] HLA effects and ancestry

Even if the same HLA alleles are associated with MS in populations with different ancestries, they could exert different effects. Heterogeneity in effect sizes could reflect differences in the linkage disequilibrium structure at the locus (for example, the same

allele could be associated with other haplotypes or other genetic modifiers of the effect), differences in the noncoding region of the allele itself, or gene–environment interactions.

Whether HLA alleles do have differential effects across ancestries in MS remains unknown. In one admixture mapping (See Discovery of novel loci below) study of a cohort predominantly from California, USA, excess European ancestry was apparent at the MHC among African American people with MS[67]. Among African American people who were heterozygotic for the HLA-DRB1*15:01 allele, carriage of an ancestrally European allele was associated with a threefold greater risk of MS than was carriage of an ancestrally African allele[59]. However, another admixture mapping study of African American people (with a partially overlapping cohort) found no evidence of excess European ancestry at the MHC despite a similar study design and comparable statistical power[14,15]. These findings should be interpreted with some caution because the admixture peak at the MHC did not pass genome-wide significance in either study. Heterogeneity of HLA effects across ancestries would have interesting biological implications, but further studies with greater statistical power are required to prove or refute this hypothesis.

[H1] Beyond the MHC

A common theme emerging from studies of MS genetics in populations of non-European ancestry is that the risk alleles with large effect sizes identified in European populations tend to have broadly similar effects across different ancestral backgrounds: this observation has been made in African American[9,10,68,69], Hispanic[68], Sardinian[70], Greek[71], South Asian[11,72] and Japanese[73] cohorts. Identifying variants with truly heterogeneous effects across ancestries is challenging for two major reasons. First, marginal effects of variants are expected to differ across ancestries purely as a result of differences in linkage disequilibrium and allele frequency, regardless of whether the causal effect differs across ancestries[74]. Proving that an individual variant's independent effect on risk differs between ancestries therefore requires adeqaute controlling for the difference in frequency and linkage disequilibrium.

Second, statistical tests to identify heterogeneity involve comparing estimates of two effects, and statistical power is lower when the uncertainty in these estimates is greater. Given the relatively small samples in MS GWAS in populations of non-European ancestry to date, the power to detect heterogeneity between ancestries is relatively low. For this reason, studies that have been done have focused on replication of European risk variants — often defined operationally as demonstrating an association with one-sided $P < 0.05$ — rather than discovery of novel variants. Nevertheless, some intriguing signals that require follow-up have been detected to date.

In one study published in 2019, use of a custom genotyping array called the MS Chip (Box 2) to genotype 1,398 Hispanic people with MS, 1,305 African American

people with MS and matched control individuals revealed less replication of European risk variants in African American people than expected on the basis of power calculations — 41 of 200 variants were replicated, whereas 69 (95% CI 57–82) were expected[68]. This study also provided nominal evidence (P <0.05) of effect size heterogeneity between all three populations for one variant: rs4545915 in *MALT1*, which encodes a protease involved in B-cell lymphoma/leukaemia 10 (BCL10) signalling. The estimates of effect size for this variant were more pronounced in the Hispanic population (OR 1.29, 95% CI 1.14–1.47) and the African American population (OR 1.18, 95% CI 1.04–1.34) than in populations of European ancestry (OR 1.1, 95% CI 1.07–1.16)[68]. The same study also demonstrated nominal heterogeneity of effect size for the rs11740512 variant at 5p13.1 between the African American population (OR 1.29, 95% CI 1.12–1.49) and populations of European ancestry (OR 1.21, 95% CI 1.16–1.27). This variant maps to the prostaglandin E2 receptor, encoded by *PTGER4*, which is involved in vitamin D signalling. Given that vitamin D status differs between ancestral groups, the different effect sizes for this variant between ancestries has some appealing biological plausibility[68].

Various caveats need to be kept in mind when interpreting these replication results. First, the expected number of replicating variants under the null hypothesis of homogeneity between populations must be specified. Even if allele frequencies and effect sizes of risk variants identified in European populations were identical between populations, the power to detect associations in the replicating cohort is limited by sample size, especially as all such risk variants outside the MHC have relatively small effects (OR <1.3)[8]. Explicit power calculations help to clarify whether the degree of replication is within the expected range for homogeneous effects across ancestries[9,68]. Second, attempts to replicate associations are likely to produce effect sizes and P values of smaller magnitude than the original study simply as a result of the so-called winner's curse — the likelihood that discovery-stage GWAS hits are likely to be inflated — rather than true population differences. This difficulty is mitigated to some extent by the two-step design of the European GWAS and the large sample sizes included, but could still create a false impression of heterogeneity if fewer associations are replicated than expected. Third, as noted above, replication studies are not statistically powered to detect heterogeneity of effect sizes so might not be enable detection of subtle differences in effects between populations owing to sample size. Furthermore, the biological interpretation of heterogeneity is complicated; although variants with opposite effects across ancestries have a clear interpretation, if variants have effects in the same direction but different magnitude, the implications for disease biology are less clear.

Replication studies can also be limited by genotyping array design. The genotyping chips used for the successful European MS GWAS — ImmunoChip and MSChip — were designed with European linkage disequilibrium structures in mind[8,75]. These chips do not capture the same genetic variation in different ancestries.

Replication studies that focus on the lead SNP at a locus identified in a European population are, by design, primarily concerned with replication of known signals rather than discovery of heterogeneity in associated loci or alleles across populations. These efforts will be advanced by the development of new genotyping arrays designed specifically for cross-ancestry association testing and fine mapping[76], and by decreasing sequencing costs, which could make sequencing feasible on a consortium scale.

[H1] Phenotypic and genotypic heterogeneity

[H2] Extreme phenotypes

MS is a clinically heterogeneous disease[77]. To date, GWAS findings have explained almost none of this phenotypic variation; notable exceptions are the association of HLA-DRB1*15:01 with age of onset[77,78] and putative associations of relapse rate with variation in *LRP2* and *WNT9B*[43,79]. Possible explanations for the paucity of associations include low heritability of clinical phenotypes, imprecision and inaccuracies in phenotype recording, and a sizeable influence of many confounders, particularly disease-modifying therapy (DMT), which are difficult to measure and adjust for[77]. Despite this complexity, perseverance with genetic studies of MS severity is worthwhile, as insight could help to identify targets for drugs that slow or reverse progression, a major unmet clinical need.

Studying MS genetics in populations with different patterns of MS severity and disease course could provide a window into the genetic architecture of these traits. The concept of extreme phenotype sampling is well-developed in the study of complex quantitative traits. In this approach, rather than performing GWAS with a given quantitative trait (such as a disease) as the outcome, a case–control GWAS is performed in which the extreme values of the trait (for example people with very benign MS versus people with very severe MS) are used as cases and controls[80]. This approach has been used to discover novel loci associated with several complex traits, including blood pressure[81], anthropometric traits[82], time to infection with *Pseudomonas aeruginosa* in people with cystic fibrosis[83], paclitaxel-induced neuropathy[84], and facial structure[85]. Use of this approach to study extreme MS phenotypes — for example, young-onset MS, or highly inflammatory or progressive disease at onset — could shed light on the genetic drivers of susceptibility and the genetic drivers of progression and severity, as the findings would be expected to extend beyond the extreme phenotypes studied.

[H2] Ancestral heterogeneity in MS phenotypes

A substantial body of evidence suggests that MS phenotypes differ between populations with different ethnic and ancestral backgrounds. Importantly, these apparent discrepancies are likely to partially reflect ascertainment biases and discrepancies in healthcare access. Nevertheless, determining whether genetic differences account for heterogeneity could provide insight into the determinants of disease course.

Some evidence suggests that age of MS onset varies according to ancestry. Specifically, African American[13,86–88] and Hispanic American[87,89–91] individuals with MS seem to have a younger age of onset and diagnosis than white American individuals. However, this difference does not seem to be true for British individuals whose ethnicity was recorded in primary care as 'Black' (assumed to indicate African ancestry in this study)[35]. A more universal finding is that individuals of African ancestry — approximated by Black ethnic background in epidemiological studies — with MS have a more aggressive disease course than people of European ancestry, characterized by greater disability at diagnosis, more rapid progression through disability milestones, higher overall morbidity and mortality, a higher likelihood of progressive (rather than relapsing) disease, faster rates of brain and retinal atrophy[92,93], more inflammatory activity[94], and poorer responses to DMTs[88,95–98]. These findings persist despite adjustment for socio-economic status, suggesting that healthcare inequalities are unlikely to explain all of this phenotypic variability[88]. Within the USA, data suggest that Hispanic American individuals also experience a more aggressive disease course than white American individuals[87,99].

Variability in MS severity between ethnicities could plausibly have little genetic basis and be mainly due to systematic differences in healthcare access, cultural differences in healthcare use, socioeconomic confounders and clinicians' biases. However, epidemiological and genetic data suggest that ancestral genetic variation does influence disease severity. For example, correction for measurable confounders associated with socioeconomic status suggests at least some residual effect that is not attributable to social determinants of health, and clinical observations in populations of non-European ancestry living outside the USA support heterogeneity of phenotypes[100]. Furthermore, data from a cohort of ~1,000 Hispanic individuals suggest that higher African or Native American ancestry are associated with an earlier age of onset[91]. Larger studies and studies in other populations are required to examine this question further. Clarification of the extent to which the genetic architecture of MS is similar across ancestries will help to clarify what proportion of variation in phenotypes is attributable to non-genetic factors, such as racism and inequality, and could help to address these factors.

[H1] Discovery of novel loci

[H2] GWAS

GWAS in populations of non-European ancestry can help to uncover novel risk loci for MS. This approach is exemplified by the elucidation of the role of *TNFSF13B*, which encodes B cell activating factor (BAFF), in MS in a Sardinian cohort. Expansion of the Sardinian-specific MS GWAS[101] and integration of the results with those from extensive Sardinian-specific immunophenotyping[102] and whole-genome sequencing[103] determined that variation at the *TNFSF13B* locus is important in determining MS susceptibility[104]. This GWAS analysis, which involved 2,273 people with MS and 2,148 control individuals, identified a genome-wide suggestive association (P <5 x 10^{-6}) of a variant near *TNFSF13B* — rs12874404 — with MS in the Sardinian population. This variant was not associated with MS in case–control data sets from the UK and Sweden. This apparent lack of replication is explained by differences in linkage disequilibrium and allele frequencies between the populations. In the Sardinian population, rs12874404 is in linkage disequilibrium (r^2 = 0.76) with a common (allele frequency 26%) deletion (GCTGT>A; referred to as BAFF-var), which was shown, through conditional analysis, to be responsible for the association. In non-Sardinian European populations, the allele frequency of BAFF-var is low (2%), so the linkage disequilibrium with rs12874404 is substantially weaker (r^2 = 0.44); the unique genetic architecture of the Sardinian population therefore made this discovery possible despite smaller sample sizes than in GWAS conducted in mainland Europe.

Mechanistic studies have demonstrated a plausible mechanism through which this variant could increase MS risk — the variant creates a new polyadenylation signal, which results in a shorter transcript that lacks microRNA binding sites and is therefore expressed at higher levels. These higher levels of soluble BAFF are likely to increase the risk of immune tolerance being broken because they promote B cell proliferation, survival and immunoglobulin production[104]. This association remains one of few identified in GWAS for which the likely underlying mechanism has been established. This example illustrates how discoveries in one population can have value for all. It also emphasizes the critical dependency of genetic studies on statistical power — these investigations were successful because Sardinia is an island with a degree of genetic isolation from the mainland and an extremely high prevalence of MS.

Other than this example, efforts to discover novel associations with MS in GWAS in populations of non-European ancestry have been less successful. The most likely reasons are a lack of statistical power owing to small sample sizes and the use of genotyping arrays that have been designed for European populations, making them less

useful for novel risk allele discovery in other populations. In one study of an African American population in which the ImmunoChip array (Box 2) was used, no novel risk variants were identified but the results did demonstrate highly significant concordance with the variants identified in GWAS of European populations. These results were expected given the statistical power of this study. Seven novel variants (outside of the 110 loci known at the time) were modestly associated with MS (P <1 x 10^{-4})[9] but a study in an independent cohort of 620 African American people with MS and 1,565 control individuals produced nominally significant evidence of association for just one of these variants (rs2702180 in *SMG7,* which encodes a protein involved in mRNA homeostasis). Even in a combined analysis, the level of evidence for an association was modest and well below genome-wide significance.

[H2] Admixture mapping

An alternative approach to identifying novel risk loci is admixture mapping[105], a relatively old technique for mapping loci associated with a disease or trait among individuals with a mixture of genetic ancestries. This technique is based on the hypothesis that if a trait is more common among one of the parent ancestral groups then variants that are associated with that trait will, on average, be inherited in haplotypes derived from the ancestrally higher-risk population. The ancestral origin of each region of the genome is estimated and the extent to which each locus deviates from the genome-average is calculated. Admixture mapping statistics can be calculated on the basis of cases only or in a case–control setting. Both approaches involve estimation of local ancestry (the ancestral origin of any given haplotype in the genome) and global ancestry (the overall proportions of the genome derived from each ancestral group). Case-only analysis involves comparison of local ancestry with global ancestry proportions in a genome-wide scan, with the expectation that disease-associated loci will deviate from the average, global ancestry. Case–control analysis involves comparing local ancestry across the genome between cases and controls, with the expectation that disease-associated loci will be derived from the higher risk ancestral group more often in cases than controls.

Early attempts at admixture mapping in an African American cohort with MS identified an association of a locus on chromosome one in the region of the centromere[14,15]. However, the strength of the association weakened substantially (P = 0.1) in follow-up analysis with a larger number of patients, highlighting the risk of type 1 errors in genome-wide scans with relatively small sample sizes[15]. Another genome-wide admixture mapping study identified no novel loci that were significantly associated with MS in African American or American people of East Asian ancestry but did identify one suggestive locus on chromosome 8 in Hispanic American people[59]. The nearest gene to the identified locus is *ZNF596*, which encodes a zinc finger with unclear relevance to known MS risk pathways. This result has not yet been replicated.

An essential prerequisite for admixture mapping is that the disease risk attributable to genetic factors differs substantially between the two populations. This basic requirement seemed to be supported by data on the epidemiology of MS when admixture mapping was first used in this context, but the latest epidemiological evidence indicates little difference in the risk of MS between ethnic groups, particularly between people of African and European ancestry[33–35,106]. In the absence of this fundamentally important difference, the lack of associations identified is perhaps unsurprising. Theoretically, however, this type of mapping could be undertaken in other admixed groups whose ancestry is of a population in which the risk of MS does seem to be lower, such as South Asian or East Asian populations.

[H1] Pinpointing causal variants

Since the migration of humans from Africa and their spread across the world, populations have emerged with different sets of haplotypes, alleles and linkage disequilibrium structures. Variants that are associated with MS in populations of European ancestry at a given locus are likely to be in linkage disequilibrium with the causal variant (or variants) at the locus rather than being causal themselves; in fact, the causal association for most of the 201 non-MHC risk variants identified in GWAS remains poorly understood[8,107]. Causal variants at a given locus are likely to be shared across populations with different ancestries, whereas the variants that are in linkage disequilibrium with this causal variant are likely to differ between populations of different ancestries[108]. For this reason, combining results from well-powered GWAS in populations with a variety of ancestries would be expected to reduce the size of the credible set of variants that could account for a given GWAS signal[62,109]. Well-powered GWAS in populations of African ancestry are likely to be particularly helpful in this regard because linkage disequilibrium blocks in these populations tend to be far smaller than in other populations[62].

This approach — known as cross-ancestral fine mapping, or trans-ethnic fine mapping (Figure 2) — has been successfully applied to the study of several complex traits and diseases[1,109]. Sample size is a crucial determinant of success in fine mapping studies; achieving a sufficient sample size from populations in which the prevalence of the disease is low is challenging, and this difficulty has limited use of cross-ancestral fine mapping in MS to date[9,15].

Cross-ancestral fine-mapping has been attempted for two MS-associated loci using data from African American patients with MS who were genotyped with the ImmunoChip (Box 2)[9]. The investigators searched for variants that were in linkage disequilibrium with the lead variant identified in European populations at each locus and determined whether the lead variant identified at the same locus in the African American

population could help to narrow the credible set of causal variants at the locus. At the *MMEL1* locus (1p36), the linkage disequilibrium block around the lead variant identified in the African American population (rs111375644) restricted the size of the locus to a single gene — *TNFRFS14* — compared with five genes in populations of European ancestry[9]. Subsequent annotation of this locus using cell-type-specific regulatory elements revealed that the variant identified in European populations affects expression of *TNFRFS14* in immune cell types[110], a finding that increases confidence in the result of the fine mapping. At the *PVT1–MIR1208* locus, the lead variant identified in the African American population (rs1861842) was in linkage disequilibrium with the lead variant identified in the European population (rs759648) only in the European population and not in the African American population[9]. This observation suggests that the rs1861842 variant is more likely than the rs759648 variant to tag the causal variant. However, the African American population sample was small, so the statistical evidence for associations of variants was relatively weak. As a result, the study does not provide definitive evidence for successful fine mapping at either locus.

[H1] Downstream insights

GWAS summary statistics are a starting point for understanding disease biology — a range of statistical and bioinformatics tools can be used to translate these findings into meaningful disease insights. Examples of such applications include polygenic risk score profiling, Mendelian randomization, heritability estimates, including heritability partitioning, genetic correlation analysis with linkage disequilibrium score regression, functional annotation, and fine mapping. Extending these insights to populations of non-European ancestry requires GWAS of these populations.

Polygenic risk scores have several applications, including prediction of future disease, risk stratification for prevention or early treatment trials, and to make inferences about the causal role of genes in disease, where they can be treated as instrumental variables[111]. The reliability of polygenic risk scores depends on the premise that susceptibility alleles have similar effects in the base (GWAS) and target (validation) populations. If this premise does not hold true, polygenic risk scores derived from GWAS of European populations will perform poorly for populations of non-European ancestry[17,19,112]; this is the case in MS[55,68]. New methods are being developed to improve cross-ancestry performance of polygenic risk scores[113,114], but the largest gains in performance will come from conducting GWAS in populations of non-European ancestry that have greater statistical power than those done to date[115].

Mendelian randomization is an instrumental variable approach in which genetic associations with exposures and outcomes of interest are used to determine whether observational associations are causal[116–119]. Two-sample Mendelian randomization

involves use of separate GWAS summary statistics for the exposure and the outcome, so the population structure must be similar in the two GWAS to ensure that population stratification does not bias the result. Mendelian randomization studies in MS in European populations have extended and clarified some important epidemiological observations. For example, this approach has helped to clarify independent causal roles of childhood obesity and low vitamin D levels in MS; more specifically, use of Mendelian randomization demonstrated that the increased risk of MS associated with earlier puberty is likely to be driven by childhood obesity[120]. GWAS of MS in non-European populations are required to determine whether these findings are generalizable across ancestries. This approach has been successful, for instance, in replicating the causal influence of type 2 diabetes mellitus and dyslipidaemia on stroke risk in individuals of African-ancestry[121].

[H1] Challenges

Powerful arguments exist for broadening the scope of MS genetics research to non-European populations, but several challenges need to be overcome to achieve this goal.

The cost of genetic studies is a major challenge. An argument could be made that the genetic architecture of MS seems to overlap substantially across populations on the basis of existing evidence, so additional costly efforts to perform large GWAS in other populations might not yield new insights. However, experience from cross-ancestral genetic studies of other complex traits and related autoimmune disorders suggests that GWAS in non-European populations can lead to important advances in fine mapping and allelic heterogeneity if sample sizes are adequate, and that within-locus heterogeneity can exist between populations even when the genome-wide genetic correlation is high[28].

A related challenge is that large sample sizes are needed for GWAS in order to detect associations of common variants, most of which are weak (OR <1.3). Identifying sufficiently large numbers of patients with disease can be difficult and is complicated by differential access to healthcare facilities, phenotypic heterogeneity and a low incidence in some populations. We anticipate that the increasing number of international efforts to do this work will eventually lead to GWAS in populations of non-European ancestry with comparable sample sizes and statistical power to studies in populations of European-ancestry[76].

A further challenge is the identification of appropriate control participants to avoid false discoveries owing to population stratification. This challenge is particularly relevant for individuals of African ancestry, who have far more genetic diversity, on average, than individuals of European descent. Large-scale biobanks that include genotyping and rich phenotyping of increasing numbers of non-European individuals will make such

GWAS efforts more tractable. Such biobanks include Genes and Health[122], Biobank Japan[51], and the All Of Us Research Programme[123].

Historically, a methodological issue that has limited progress in cross-ancestral genetic studies of complex disease is the relative paucity of tools for cross-ancestry analyses. However, several major advances in this field include new approaches for inferring an individual's genetic ancestry, inferring the ancestral origin of specific haplotypes, conducting GWAS in admixed populations, deriving cross-ancestry polygenic risk scores, calculating cross-ancestry genetic correlations, and performing cross-ancestry power calculations[124–130]. In addition to these advances in statistical software, the increasing representation of a broad range of genetic ancestries in reference datasets, such as TOPMED[131], has been a key driver of recent multi-ancestry GWAS efforts.

Finally, several logistical and ethical issues need to be considered when studying genetics in populations of non-European ancestry. Care must be taken to ensure that results are interpreted and presented in a way that avoids discrimination. External researchers carrying out genetic research in countries with non-European ancestral populations must ensure that this is developed with local researchers, and delivers lasting infrastructure. In addition, historical, racially-motivated human rights abuses have led to understandable skepticism about the motives underlying genetic studies among people of non-European ancestry[132]. One of the most striking examples of such abuses is the Tuskegee Syphilis Study, a natural history study of syphilis among a cohort of largely African American men that ran from 1932 until 1972. Despite treatment becoming available in 1947, these men were denied access to the drug (penicillin) — as a consequence, participants were consigned to entirely preventable deterioration and many died unnecessarily[133]. Addressing such concerns is a crucial part of designing cross-ancestral genetic studies.

Early and sustained involvement of researchers and participants from the study population is essential to build trust in the medical establishment and to ensure that researchers can carry out this scientifically-important work while respecting the ethical issues involved. The challenges and current best practices in this field have been reviewed in detail elsewhere[2]. In the ongoing ADAMS study (https://app.mantal.co.uk/adams) of MS genetics among British individuals of non-European ancestry, we have been working with a group of individuals from diverse ethnic backgrounds who have MS to ensure that the study methods, goals and communications are inclusive and respectful. We believe that inclusion of diverse patient and researcher perspectives at all stages in the research process is a key part of developing equitable research studies that promote the benefits for all.

[H1] Conclusions

In this Perspective article, we have discussed the various ways in which studying the genetics of MS in populations of non-European ancestry is likely to deepen our understanding of the genetic basis of this complex disease. Despite the logistical and theoretical barriers, there is a strong scientific imperative — in addition to the purely ethical need — to doing this work. We anticipate that consortium-scale efforts to tackle this question, if logistically feasible, will ultimately benefit people with MS from all ancestral backgrounds.

References

1. Bentley, A. R., Callier, S. L. & Rotimi, C. N. Evaluating the promise of inclusion of African ancestry populations in genomics. *NPJ Genom Med* **5**, 5 (2020).
2. Fatumo, S. *et al.* A roadmap to increase diversity in genomic studies. *Nat. Med.* 1–8 (2022).
3. Morales, J. *et al.* A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. *Genome Biol.* **19**, 21 (2018).
4. Landry, L. G., Ali, N., Williams, D. R., Rehm, H. L. & Bonham, V. L. Lack Of Diversity In Genomic Databases Is A Barrier To Translating Precision Medicine Research Into Practice. *Health Aff.* **37**, 780–785 (2018).
5. Hindorff, L. A. *et al.* Prioritizing diversity in human genomics research. *Nat. Rev. Genet.* **19**, 175–185 (2018).
6. Ben-Eghan, C. *et al.* Don't ignore genetic data from minority populations. *Nature* **585**, 184–186 (2020).
7. International Multiple Sclerosis Genetics Consortium (IMSGC) *et al.* Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* **45**, 1353–1360 (2013).
8. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* **365**, (2019).
9. Isobe, N. *et al.* An ImmunoChip study of multiple sclerosis risk in African Americans. *Brain* **138**, 1518–1530 (2015).
10. Isobe, N. *et al.* Genetic risk variants in African Americans with multiple sclerosis.

Neurology **81**, 219–227 (2013).

11. Pandit, L. *et al.* Evaluation of the established non-MHC multiple sclerosis loci in an Indian population. *Mult. Scler.* **17**, 139–143 (2011).
12. Pandit, L. *et al.* HLA associations in South Asian multiple sclerosis. *Mult. Scler.* **22**, 19–24 (2016).
13. Oksenberg, J. R. *et al.* Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. *Am. J. Hum. Genet.* **74**, 160–167 (2004).
14. Reich, D. *et al.* A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. *Nat. Genet.* **37**, 1113–1118 (2005).
15. Nakatsuka, N. *et al.* Two genetic variants explain the association of European ancestry with multiple sclerosis risk in African-Americans. *Sci. Rep.* **10**, 16902 (2020).
16. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The Missing Diversity in Human Genetic Studies. *Cell* **177**, 1080 (2019).
17. Duncan, L. *et al.* Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* **10**, 3328 (2019).
18. Peterson, R. E. *et al.* Genome-wide association studies in ancestrally diverse populations: Opportunities, methods, pitfalls, and recommendations. *Cell* **179**, 589–603 (2019).
19. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
20. Lam, M. *et al.* Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat. Genet.* **51**, 1670–1678 (2019).

21. Mahajan, A. *et al.* Trans-ethnic Fine Mapping Highlights Kidney-Function Genes Linked to Salt Sensitivity. *Am. J. Hum. Genet.* **99**, 636–646 (2016).
22. Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
23. Chen, M.-H. *et al.* Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell* **182**, 1198-1213.e14 (2020).
24. Graham, S. E. *et al.* The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675–679 (2021).
25. Laufer, V. A. *et al.* Genetic influences on susceptibility to rheumatoid arthritis in African-Americans. *Hum. Mol. Genet.* **28**, 858–874 (2019).
26. Robertson, C. C. *et al.* Fine-mapping, trans-ancestral and genomic analyses identify causal variants, cells, genes and drug targets for type 1 diabetes. *Nat. Genet.* **53**, 962–971 (2021).
27. Onengut-Gumuscu, S. *et al.* Type 1 Diabetes Risk in African-Ancestry Participants and Utility of an Ancestry-Specific Genetic Risk Score. *Diabetes Care* **42**, 406–415 (2019).
28. Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).
29. Somineni, H. K. *et al.* Whole-genome sequencing of African Americans implicates differential genetic architecture in inflammatory bowel disease. *Am. J. Hum. Genet.* **108**, 431–445 (2021).

30. GBD 2016 Multiple Sclerosis Collaborators. Global, regional, and national burden of multiple sclerosis 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* **18**, 269–285 (2019).
31. Koch-Henriksen, N. & Sørensen, P. S. The changing demographic pattern of multiple sclerosis epidemiology. *Lancet Neurol.* **9**, 520–532 (2010).
32. Lee, J. D. *et al.* Incidence of Multiple Sclerosis and Related Disorders in Asian Populations of British Columbia. *Can. J. Neurol. Sci.* **42**, 235–241 (2015).
33. Wallin, M. T. *et al.* The Gulf War era multiple sclerosis cohort: age and incidence rates by race, sex and service. *Brain* **135**, 1778–1785 (2012).
34. Langer-Gould, A., Brara, S. M., Beaber, B. E. & Zhang, J. L. Incidence of multiple sclerosis in multiple racial and ethnic groups. *Neurology* **80**, 1734–1739 (2013).
35. Dobson, R. *et al.* Ethnic and Socioeconomic Associations with Multiple Sclerosis Risk. *Ann. Neurol.* **87**, 599–608 (2020).
36. Langer-Gould, A. M., Gonzales, E. G., Smith, J. B., Li, B. H. & Nelson, L. M. Racial and ethnic disparities in multiple sclerosis prevalence. *Neurology* **98**, e1818–e1827 (2022).
37. Munk Nielsen, N. *et al.* Multiple sclerosis among first- and second-generation immigrants in Denmark: a population-based cohort study. *Brain* **142**, 1587–1597 (2019).
38. Ahlgren, C., Odén, A. & Lycke, J. A nationwide survey of the prevalence of multiple sclerosis in immigrant populations of Sweden. *Mult. Scler.* **18**, 1099–1107 (2012).
39. Sawcer, S. *et al.* A high-density screen for linkage in multiple sclerosis. *Am. J. Hum. Genet.* **77**, 454–467 (2005).

40. Jersild, C., Svejgaard, A. & Fog, T. HL-A antigens and multiple sclerosis. *Lancet* **1**, 1240–1241 (1972).
41. Moutsianas, L. *et al.* Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nat. Genet.* **47**, 1107–1113 (2015).
42. Jokubaitis, V. G. *et al.* Not all roads lead to the immune system: The Genetic Basis of Multiple Sclerosis Severity Implicates Central Nervous System and Mitochondrial Involvement. *bioRxiv* (2022) doi:10.1101/2022.02.04.22270362.
43. Vandebergh, M. *et al.* Genetic Variation in WNT9B Increases Relapse Hazard in Multiple Sclerosis. *Ann. Neurol.* **89**, 884–894 (2021).
44. Dendrou, C. A., Petersen, J., Rossjohn, J. & Fugger, L. HLA variation and disease. *Nat. Rev. Immunol.* **18**, 325–339 (2018).
45. Hollenbach, J. A. & Oksenberg, J. R. The immunogenetics of multiple sclerosis: A comprehensive review. *J. Autoimmun.* **64**, 13–25 (2015).
46. Yoshimura, S. *et al.* Genetic and infectious profiles of Japanese multiple sclerosis patients. *PLoS One* **7**, e48592 (2012).
47. Saruhan-Direskeneli, G. *et al.* HLA-DR and -DQ associations with multiple sclerosis in Turkey. *Hum. Immunol.* **55**, 59–65 (1997).
48. Alvarado-de la Barrera, C. *et al.* HLA class II genotypes in Mexican Mestizos with familial and nonfamilial multiple sclerosis. *Neurology* **55**, 1897–1900 (2000).
49. Brassat, D. *et al.* The HLA locus and multiple sclerosis in Sicily. *Neurology* **64**, 361–363 (2005).
50. Nakamura, Y. *et al.* Latitude and HLA-DRB1*04:05 independently influence disease severity in Japanese multiple sclerosis: a cross-sectional study. *J.*

Neuroinflammation **13**, 239 (2016).

51. Watanabe, M. *et al.* HLA genotype-clinical phenotype correlations in multiple sclerosis and neuromyelitis optica spectrum disorders based on Japan MS/NMOSD Biobank data. *Sci. Rep.* **11**, 607 (2021).
52. Amirzargar, A. *et al.* HLA class II (DRB1, DQA1 and DQB1) associated genetic susceptibility in Iranian multiple sclerosis (MS) patients. *Eur. J. Immunogenet.* **25**, 297–301 (1998).
53. Brum, D. G., Barreira, A. A., Louzada-Junior, P., Mendes-Junior, C. T. & Donadi, E. A. Association of the HLA-DRB1*15 allele group and the DRB1*1501 and DRB1*1503 alleles with multiple sclerosis in White and Mulatto samples from Brazil. *J. Neuroimmunol.* **189**, 118–124 (2007).
54. Quelvennec, E. *et al.* Genetic and functional studies in multiple sclerosis patients from Martinique attest for a specific and direct role of the HLA-DR locus in the syndrome. *Tissue Antigens* vol. 61 166–171 (2003).
55. Khankhanian, P. *et al.* Genetic contribution to multiple sclerosis risk among Ashkenazi Jews. *BMC Med. Genet.* **16**, 55 (2015).
56. Kwon, O. J. *et al.* HLA class II susceptibility to multiple sclerosis among Ashkenazi and non-Ashkenazi Jews. *Arch. Neurol.* **56**, 555–560 (1999).
57. Marrosu, M. G. *et al.* Dissection of the HLA association with multiple sclerosis in the founder isolated population of Sardinia. *Hum. Mol. Genet.* **10**, 2907–2916 (2001).
58. Goodin, D. S., Oksenberg, J. R., Douillard, V., Gourraud, P.-A. & Vince, N. Genetic susceptibility to multiple sclerosis in African Americans. *PLoS One* **16**, e0254945 (2021).

59. Chi, C. *et al.* Admixture mapping reveals evidence of differential multiple sclerosis risk by genetic ancestry. *PLoS Genet.* **15**, e1007808 (2019).
60. Rivera, V. M. Multiple sclerosis in Latin Americans: Genetic aspects. *Curr. Neurol. Neurosci. Rep.* **17**, (2017).
61. Vinoy, N., Sheeja, N., Kumar, S. & Biswas, L. Class II HLA (DRB1, & DQB1) alleles and IL7R (rs6897932) variants and the risk for Multiple Sclerosis in Kerala, India. *Mult. Scler. Relat. Disord.* **50**, 102848 (2021).
62. International HapMap Consortium *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
63. Matsuki, K., Carl Grumet, F., Lin, X., Gelb, M. & Gueilleminault, C. DQ (rather than DR) gene marks susceptibility to narcolepsy. *Lancet* **339**, (1992).
64. Okada, Y. *et al.* Contribution of a Non-classical HLA Gene, HLA-DOA, to the Risk of Rheumatoid Arthritis. *Am. J. Hum. Genet.* **99**, 366–374 (2016).
65. Naito, T. *et al.* A deep learning method for HLA imputation and trans-ethnic MHC fine-mapping of type 1 diabetes. *Nat. Commun.* **12**, 1639 (2021).
66. Patsopoulos, N. A. *et al.* Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet.* **9**, e1003926 (2013).
67. Luo, Y. *et al.* A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-mapping in HIV host response. *Nat. Genet.* **53**, 1504–1516 (2021).
68. Beecham, A. H. *et al.* The genetic diversity of multiple sclerosis risk among Hispanic and African American populations living in the United States. *Mult. Scler.*

1352458519863764 (2019).

69. Johnson, B. A. *et al.* Multiple sclerosis susceptibility alleles in African Americans. *Genes Immun.* **11**, 343–350 (2010).
70. Hilven, K. & Goris, A. Genetic burden mirrors epidemiology of multiple sclerosis. *Mult. Scler.* **21**, 1353–1354 (2015).
71. Hadjigeorgiou, G. M. *et al.* Replication study of GWAS risk loci in Greek multiple sclerosis patients. *Neurol. Sci.* **40**, 253–260 (2019).
72. Pandit, L. *et al.* European multiple sclerosis risk variants in the south Asian population. *Mult. Scler.* **22**, 1536–1540 (2016).
73. Kira, J., Matsushita, T., Sato, S. & Yamamoto, K. A genome-wide association study (GWAS) in the Japanese population reveals novel genetic risk factors for multiple sclerosis and neuromyelitis optica. *J. Neurol. Sci.* **357**, e308 (2015).
74. Weissbrod, O. *et al.* Leveraging fine-mapping and non-European training data to improve cross-population polygenic risk scores. *bioRxiv* (2021) doi:10.1101/2021.01.19.21249483.
75. Cortes, A. & Brown, M. A. Promise and pitfalls of the Immunochip. *Arthritis Res. Ther.* **13**, 101 (2011).
76. Beecham, A. H. & McCauley, J. L. Fine-Mapping Array Design for Multi-Ethnic Studies of Multiple Sclerosis. *Genes* **10**, (2019).
77. Jokubaitis, V. G., Zhou, Y., Butzkueven, H. & Taylor, B. V. Genotype and Phenotype in Multiple Sclerosis-Potential for Disease Course Prediction? *Curr. Treat. Options Neurol.* **20**, 18 (2018).
78. International Multiple Sclerosis Genetics Consortium *et al.* Genetic risk and a

primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219 (2011).

79. Zhou, Y. *et al.* Genetic variation in the gene LRP2 increases relapse risk in multiple sclerosis. *J. Neurol. Neurosurg. Psychiatry* **88**, 864–868 (2017).
80. Barnett, I. J., Lee, S. & Lin, X. Detecting rare variant effects using extreme phenotype sampling in sequencing association studies. *Genet. Epidemiol.* **37**, 142–151 (2013).
81. Padmanabhan, S. *et al.* Genome-wide association study of blood pressure extremes identifies variant near UMOD associated with hypertension. *PLoS Genet.* **6**, e1001177 (2010).
82. Berndt, S. I. *et al.* Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat. Genet.* **45**, 501–512 (2013).
83. Emond, M. J. *et al.* Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. *Nat. Genet.* **44**, 886–889 (2012).
84. Boora, G. K. *et al.* Testing of candidate single nucleotide variants associated with paclitaxel neuropathy in the trial NCCTG N08C1 (Alliance). *Cancer Med.* **5**, 631–639 (2016).
85. Crouch, D. J. M. *et al.* Genetics of the human face: Identification of large-effect single gene variants. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E676–E685 (2018).
86. Weinstock-Guttman, B. *et al.* Multiple sclerosis characteristics in African American patients in the New York State Multiple Sclerosis Consortium. *Mult. Scler.* **9**, 293–

298 (2003).

87. Ventura, R. E., Antezana, A. O., Bacon, T. & Kister, I. Hispanic Americans and African Americans with multiple sclerosis have more severe disease course than Caucasian Americans. *Mult. Scler.* **23**, 1554–1557 (2017).
88. Gray-Roncal, K. *et al.* Association of Disease Severity and Socioeconomic Status in Black and White Americans With Multiple Sclerosis. *Neurology* (2021) doi:10.1212/WNL.0000000000012362.
89. Hadjixenofontos, A. *et al.* Clinical expression of multiple sclerosis in Hispanic whites of primarily Caribbean ancestry. *Neuroepidemiology* **44**, 262–268 (2015).
90. Amezcua, L., Lund, B. T., Weiner, L. P. & Islam, T. Multiple sclerosis in Hispanics: a study of clinical disease expression. *Mult. Scler.* **17**, 1010–1016 (2011).
91. Amezcua, L. *et al.* Native ancestry is associated with optic neuritis and age of onset in hispanics with multiple sclerosis. *Ann. Clin. Transl. Neurol.* **5**, 1362–1371 (2018).
92. Caldito, N. G. *et al.* Brain and retinal atrophy in African-Americans versus Caucasian-Americans with multiple sclerosis: a longitudinal study. *Brain* **141**, 3115–3129 (2018).
93. Kimbrough, D. J. *et al.* Retinal damage and vision loss in African American multiple sclerosis patients. *Ann. Neurol.* **77**, 228–236 (2015).
94. Howard, J. *et al.* MRI Correlates of Disability in African-Americans with Multiple Sclerosis. *PLoS ONE* vol. 7 e43061 (2012).
95. Kister, I. *et al.* Rapid disease course in African Americans with multiple sclerosis. *Neurology* **75**, 217–223 (2010).
96. Khan, O. *et al.* Multiple sclerosis in US minority populations: Clinical practice

insights. *Neurol. Clin. Pract.* **5**, 132–142 (2015).

97. Cree, B. A. C. *et al.* Clinical characteristics of African Americans vs Caucasian Americans with multiple sclerosis. *Neurology* **63**, 2039–2045 (2004).
98. Naismith, R. T., Trinkaus, K. & Cross, A. H. Phenotype and prognosis in African-Americans with multiple sclerosis: a retrospective chart review. *Mult. Scler.* **12**, 775–781 (2006).
99. Kister, I., Bacon, T. & Cutter, G. R. How Multiple Sclerosis Symptoms Vary by Age, Sex, and Race/Ethnicity. *Neurol. Clin. Pract.* **11**, 335–341 (2021).
100. Jamal, I. *et al.* Multiple sclerosis in Kenya: Demographic and clinical characteristics of a registry cohort. *Mult Scler J Exp Transl Clin* **7**, 20552173211022784 (2021).
101. Sanna, S. *et al.* Variants within the immunoregulatory CBLB gene are associated with multiple sclerosis. *Nat. Genet.* **42**, 495–497 (2010).
102. Orrù, V. *et al.* Genetic variants regulating immune cell levels in health and disease. *Cell* **155**, 242–256 (2013).
103. Sidore, C. *et al.* Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. *Nat. Genet.* **47**, 1272–1281 (2015).
104. Steri, M. *et al.* Overexpression of the Cytokine BAFF and Autoimmunity Risk. *N. Engl. J. Med.* **376**, 1615–1626 (2017).
105. Shriner, D. Overview of admixture mapping. *Curr. Protoc. Hum. Genet.* **Chapter 1**, Unit 1.23 (2013).
106. Romanelli, R. J. *et al.* Multiple sclerosis in a multi-ethnic population from

Northern California: a retrospective analysis, 2010-2016. *BMC Neurol.* **20**, 163 (2020).

107. Caliskan, M., Brown, C. D. & Maranville, J. C. A catalog of GWAS fine-mapping efforts in autoimmune disease. *Am. J. Hum. Genet.* **108**, 549–563 (2021).

108. Wang, Y. *et al.* Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. *Nat. Commun.* **11**, 3865 (2020).

109. Li, Y. R. & Keating, B. J. Trans-ethnic genome-wide association studies: advantages and challenges of mapping in diverse populations. *Genome Med.* **6**, 91 (2014).

110. International Multiple Sclerosis Genetics Consortium. A systems biology approach uncovers cell-specific gene regulatory effects of genetic associations in multiple sclerosis. *Nat. Commun.* **10**, 2236 (2019).

111. Jacobs, B. M. *et al.* Gene-Environment Interactions in Multiple Sclerosis: A UK Biobank Study. *Neurol Neuroimmunol Neuroinflamm* **8**, (2021).

112. Privé, F. *et al.* Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. *Am. J. Hum. Genet.* **109**, 373 (2022).

113. Márquez-Luna, C., Loh, P.-R., South Asian Type 2 Diabetes (SAT2D) Consortium, SIGMA Type 2 Diabetes Consortium & Price, A. L. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet. Epidemiol.* **41**, 811–823 (2017).

114. Amariuta, T. *et al.* Improving the trans-ancestry portability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements.

Nat. Genet. **52**, 1346–1354 (2020).

115. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).

116. Harroud, A. *et al.* Childhood obesity and multiple sclerosis: A Mendelian randomization study. *Mult. Scler.* 13524585211001780 (2021).

117. Jacobs, B. M., Noyce, A. J., Giovannoni, G. & Dobson, R. BMI and low vitamin D are causal factors for multiple sclerosis: A Mendelian Randomization study. *Neurol Neuroimmunol Neuroinflamm* **7**, (2020).

118. Vandebergh, M. & Goris, A. Smoking and multiple sclerosis risk: a Mendelian randomization study. *J. Neurol.* **267**, 3083–3091 (2020).

119. Mitchell, R. E. *et al.* The effect of smoking on multiple sclerosis: a mendelian randomization study. doi:10.1101/2020.06.24.20138834.

120. Harroud, A. *et al.* Effect of age at puberty on risk of multiple sclerosis: A mendelian randomization study. *Neurology* **92**, e1803–e1810 (2019).

121. Fatumo, S. *et al.* Metabolic Traits and Stroke Risk in Individuals of African Ancestry: Mendelian Randomization Analysis. *Stroke* **52**, 2680–2684 (2021).

122. Finer, S. *et al.* Cohort Profile: East London Genes & Health (ELGH), a community-based population genomics and health study in British Bangladeshi and British Pakistani people. *Int. J. Epidemiol.* **49**, 20–21i (2020).

123. All of Us Research Program Investigators *et al.* The “All of Us” Research Program. *N. Engl. J. Med.* **381**, 668–676 (2019).

124. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**, 246 (2011).

125. Maples, B. K., Gravel, S., Kenny, E. E. & Bustamante, C. D. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* **93**, 278–288 (2013).

126. Atkinson, E. G. *et al.* Tractor uses local ancestry to enable the inclusion of admixed individuals in GWAS and to boost power. *Nat. Genet.* **53**, 195–204 (2021).

127. Brown, B. C., Asian Genetic Epidemiology Network Type 2 Diabetes Consortium, Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic genetic-correlation estimates from summary statistics. *Am. J. Hum. Genet.* **99**, 76–88 (2016).

128. Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat. Genet.* **54**, 573–580 (2022).

129. Weissbrod, O. *et al.* Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores. *Nat. Genet.* **54**, 450–458 (2022).

130. Huang, Q. Q. *et al.* Transferability of genetic loci and polygenic scores for cardiometabolic traits in British Pakistanis and Bangladeshis. *bioRxiv* (2021) doi:10.1101/2021.06.22.21259323.

131. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Cold Spring Harbor Laboratory* 563866 (2019) doi:10.1101/563866.

132. Kraft, S. A. *et al.* Beyond consent: Building trusting relationships with diverse populations in precision medicine research. *Am. J. Bioeth.* **18**, 3–20 (2018).

133. Nuriddin, A., Mooney, G. & White, A. I. R. Reckoning with histories of medical racism and violence in the USA. *Lancet* **396**, 949–951 (2020).

134. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to

candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **19**, 491–504 (2018).

135. International Multiple Sclerosis Genetics Consortium*†. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* **365**, eaav7188 (2019).

Acknowledgements

The authors thank Professor Stephen Sawcer, University of Cambridge, UK, for his helpful comments on an early draft of the manuscript.

Competing interests

The authors declare no competing interests.

Figure 1 | Global variation in frequency of the HLA-DRB1*15:01 allele. The HLA-DRB1*15:01 allele has the largest effect on MS risk in European populations. Note that numbers are allele frequency, not carrier frequency; for an allele frequency of *n* the carrier frequency is *2n*. For example, in the UK, the reported allele frequency is 0.15 and the carrier frequency is 0.30, meaning that 30% of individuals carry at least one copy of the DRB1*15:01 allele. Overall, there is a latitudinal gradient of DRB1*15:01 frequency – the allele is rare in Africa and South Asia and relatively common in North American, Scandinavia and Central Europe. Data from http://www.allelefrequencies.net/, accessed 11th July 2022. Data included are from studies that involved ≥500 individuals and for which the evidence was classed as gold-standard. Note that these data are not specifically from individuals with MS.

Figure 2 | Illustration of cross-ancestral fine mapping. Cross-ancestral fine-mapping (also called trans-ethnic fine-mapping) is a statistical method for identifying the probable specific causal variant (or variants) within a locus associated with disease in a genome-wide association study (GWAS). This approach involves triangulation of associations between variants (single-nucleotide polymorphisms) and disease across ancestries. Even in situations where the same causal variant explains the association of the locus with disease in all populations, the variant with the strongest association (the lead variant) is likely to differ between ancestries owing to differences in linkage disequilibrium and allele frequencies. a | The graphs depict the strength of association with disease for variants at different genomic positions within a single locus associated with disease in a GWAS in populations of African ancestry (top) and European ancestry (bottom). The correlation between each variant and the lead variant is indicated by the colours. The lead variant (red triangles) differ between the populations. b | This graph depicts the probability that each variant is the causal variant, calculated from cross-ancestry fine-mapping based on the GWAS data above. Combining the data from the two populations narrows the window within which the causal variant is expected to lie (the width of the peak).

Box 1 | Insights from cross-ancestral genetic studies of complex diseases

[b1] Identification of causal variants

Most variants that are identified as being associated with a disease in GWAS are not causal. Instead, the true disease-causing variant is likely to be in linkage disequilibrium with the identified GWAS variant. The process of determining which variant is the causal variant on the basis of GWAS signals is called fine mapping. Linkage disequilibrium between variants differs across ancestries, so combining associations identified in GWAS from different ancestral populations can help to narrow the list of variants that could plausibly be causal[134]. For example, cross-ancestral fine mapping on the basis of findings from cross-ancestral GWAS of blood lipids reduced the 99% credible set — the variants among which the true causal variant lies with 99% confidence — from a median of 13 variants to 8 variants[24].

[b1] Improved disease prediction

GWAS results enable calculation of an individual's risk score for a given disease by summing their burden of risk alleles. This approach is known as polygenic risk scoring. Owing to differences in linkage disequilibrium and allele frequency between ancestries, the marginal effects of genetic variants on disease risk are likely to differ between ancestries even if the effects of the truly causal variants are identical. These differences reduce the performance of polygenic risk scores across ancestries — greater ancestral distance leads to poorer predictive performance[112]. Use of ancestry-specific marginal effects is expected to improve the performance of polygenic risk scores among populations of non-European ancestry[74].

[b1] Improved causal inference

Mendelian randomization is an instrumental variable approach that can be used to estimate the extent to which risk factors for a disease are causal. The approach involves use of genetic variants as a proxy for the expected lifetime average exposure to a given risk factor. For instance, if the effects of single nucleotide polymorphism *j* on vitamin D levels and MS risk are known from GWAS, the ratio of these two effects gives an estimate of the extent to which low vitamin D levels lead to an increased risk of MS. Where multiple SNPs are associated with a given exposure, these estimates can be combined to provide a causal estimate. Estimated effects of a given SNP on the exposure and the outcome in one ancestral group are unlikely to be accurate for another ancestral group. Well-powered GWAS from populations of non-European ancestry are required for accurate ancestry-specific Mendelian randomization, which will be helpful to clarify the effects of environmental risk factors across ancestries. This approach has been used to replicate the effects of type 2 diabetes mellitus and dyslipidaemia on ischaemic stroke risk in individuals of African ancestry[121].

[b1] Identification of novel variants

Owing to differences in allele frequencies between populations of different ancestry, the statistical power to detect an effect of a variant on disease risk for a given sample size depends on the genetic ancestry of the population. GWAS in populations of non-European ancestry could identify novel risk alleles, either within known risk loci or in novel loci. For example, a cross-

ancestral GWAS of blood cell traits identified a missense variant in *IL7*, which encodes IL-7, that was associated with low lymphocyte count in South Asian individuals[3]. This variant is more common in South Asian populations that in populations of other ancestry (minor allele frequency 2.6% compared with <0.4%). In vitro studies have shown that this *IL7* variant increases IL-7 protein production without affecting IL-7 mRNA levels, indicating an effect on residues involved in intracellular trafficking[23]. Furthermore, GWAS of lipid traits in multiple ancestries (n ≈99,000) identified 15 novel loci associated with changes in blood lipids in African American individuals – these loci were not associated with blood lipids in individuals of European ancestry, highlighting the power of multi-ancestry cohorts for discovering novel loci when large sample sizes are available[24].

Box 2 | Genotyping array design

Genotyping array is a high-throughput method for determining an individual's genotype at a large number of genetic loci. Each array consists of beads or wells coated with allele-specific oligonucleotides. DNA fragments from the sample bind to complementary oligonucleotides. The double-stranded fragment is hybridized with an allele-specific, fluorescently-labelled base. The fluorescence signal for each allele gives a readout of the genotype distribution in the sample. Contemporary arrays allow multiplexing of many samples across hundreds of thousands to millions of single nucleotide polymorphisms (SNPs).

MS Chip is a custom genotyping array designed by the International MS Genetics Consortium (IMSGC). This chip was designed specifically to fine map known MS-associated loci and to aid discovery and/or replication of novel loci. Other custom content was added to facilitate estimation of global ancestry, to provide coverage of other autoimmune disease-associated loci, and to provide denser coverage of the major histocompatibility complex locus. Approximately 90,000 custom markers were added to the Illumina Exome Core chip; full details are available in the supplementary materials of the 2019 IMSGC meta-analysis[135]. The MS Chip assays ~330,000 markers.

Another example of a custom genotyping array is ImmunoChip, which was designed by the Wellcome Trust Case–Control Consortium to provide dense coverage of loci that are implicated in autoimmune disease. This array assays ~200,000 markers[75].

the MS Chip and the ImmunoChip were designed to maximize coverage of genetic markers in populations of European ancestry. Owing to differences in linkage disequilibrium and allele frequency between ancestries, these arrays are imperfect tools for understanding genetic variation in populations of non-European ancestry. Newer arrays, such as the Illumina Global Screening Array and Global Diversity Array, have been designed to provide better coverage and higher imputation accuracy for populations of non-European ancestry.

Glossary

Haplotype – a group of alleles which tend to be inherited together on the same chromosome. Haplotypes usually span over a short physical distance in the genome.
Long-range haplotypes – haplotypes in which the LD between alleles stretches over long distances in the genome, indicative of limited ancestral recombination events.
Linkage disequilibrium – the co-occurrence of two alleles on the same chromosome relative to the chance that the two alleles would co-occur if they were inherited independently. Usually expressed as the squared correlation coefficient (r^2) or D'. R^2 ranges from 0 to 1, with 0 indicating complete independence of the two alleles, and 1 indicating that every chromosome with allele A also possesses allele B.
Genetic bottlenecks – historical events which reduce the diversity of haplotypes/alleles within a population. Examples include migration events, whereby a small subset of one population migrates to a new location, and extinction events, in which a subset of the population dies.
Admixture – the mixing of haplotypes from more than one ancestral group due to mating between parents with different ancestral origins.
Admixture mapping – a statistical approach to finding disease-associated genetic loci in admixed populations. Admixture mapping is most useful in situations where the disease is more common in one of the ancestral parental populations that the other. This approach is based on the assumption that in admixed individuals, regions of the genome associated with the disease are more likely to be inherited on the higher-risk ancestral haplotype. Genome-wide scans can be used to identify such regions in case-only or case-control designs.
Local ancestry – the ancestral origin of a particular genetic variant or haplotype.
Global ancestry – the genome-wide average proportions of an individual's alleles which are inherited from each ancestral population.
GWAS – genome-wide association studies (GWAS) of diseases compare the allele frequencies of many genetic variants across the genome between cases and controls to determine the genetic determinants of disease risk.
Marginal effect – in GWAS, the estimate of the association between the number of copies of a particular allele and the outcome trait/disease. Due to LD between genetic variants, the marginal effect of a particular allele in GWAS incorporates the effects of many other alleles which are correlated with the tested allele.
Causal effect – in GWAS, the causal effect (or 'independent' effect) of a particular allele refers to the effect of that allele on the outcome if all other alleles are adjusted for.
Conditional analysis – an approach for determining statistically-independent genetic associations with a trait/disease. This is usually implemented in a stepwise regression framework: if SNP x is found to be the most strongly associated SNP with the trait at a given locus, association tests for all other variants at the locus will be re-run adjusting for SNP x. This procedure is often repeated until no further statistically significant results remain.
Winner's curse – the phenomenon of GWAS discoveries regressing to the null (i.e. the initially-observed effect appearing less pronounced) when larger or external studies are performed. This

arises because estimates of SNP effects from GWAS follow a sampling distribution. The effect sizes of variants that surpass an arbitrary statistical significance threshold – traditionally $P<5x10^{-8}$ – are likely to be at the upper end of the sampling distribution for those variants, i.e. they are likely to be overestimates of the true effect. When larger sample sizes become available, on average the estimates for these SNP effects will therefore decrease.

Two-step design (of GWAS) – to mitigate the risk of Winner's curse, GWAS are often conducted within an initial cohort, often termed the 'discovery' population, and then variants with strong statistical support from the discovery stage are taken forward for replication analysis in a separate cohort. Examination of a limited number of genetic variants in the replication stage, and the strong prior probability for those variants being associated with the disease, permit the rational use of a more lax P value threshold for statistical hypothesis testing in the replication phase.

Cross-ancestry fine mapping – a statistical approach which uses genetic association statistics from different ancestries to identify the likely causal SNP/s which account for the association of a locus with the disease/trait.

Heritability partitioning – a statistical approach which aims to identify the relative contribution of different types of genetic variation – broadly-defined – to the overall heritability of a trait/disease. For instance, this approach can be used to distinguish the contribution of genetic variants in a particular region (such as the MHC).

LD score regression – a statistical approach for estimating the heritability of a trait, the shared genetic architecture of >1 trait, or diagnosing population stratification as a cause of inflated P values in GWAS. LDSC considers the relationship between LD scores – a measure of how much genetic variation is tagged by an individual SNP – and the marginal effect of the variant from GWAS. In general, variants which have higher LD scores should have larger marginal effects on the trait as they are correlated with a greater number of variants. The slope of this regression reflects the heritability of the trait. The intercept of this regression reflects the degree of population stratification.

Mendelian randomisation – a statistical approach for inferring the causal impact of a risk factor on an outcome which uses genetic variants to approximate the average exposure to the exposure. This is a form of instrumental variable analysis – genetic variants associated with the risk factor are used as proxies (or instrumental variables) for the risk factor in question.

Population stratification – case-control GWAS attempts to detect allele frequency differences between cases and controls to identify the genetic determinants of disease risk. Population stratification refers to influences on the allele frequencies in the groups studies not due to case/control status. If not accounted for, population stratification can disguise true genetic associations and create false genetic associations by distorting the allele frequencies in the study groups. Differences in the genetic ancestry of the case and control populations is an important potential source of population stratification in GWAS.