HGG Advances

ARTICLE

Uncovering the genetic architecture and evolutionary roots of androgenetic alopecia in African men

Rohini Janivara,[1] Ujani Hazra,[1] Aaron Pfennig,[1] Maxine Harlemon,[1,2] Michelle S. Kim,[1,3] Muthukrishnan Eaaswarkhanth,[1] Wenlong C. Chen,[4,5,6] Adebola Ogunbiyi,[7] Paidamoyo Kachambwa,[8,9] Lindsay N. Petersen,[8,9] Mohamed Jalloh,[10,11] James E. Mensah,[12] Andrew A. Adjei,[13] Ben Adusei,[14] Maureen Joffe,[4] Serigne M. Gueye,[15] Oseremen I. Aisuodionoe-Shadrach,[16] Pedro W. Fernandez,[17] Thomas E. Rohan,[18] Caroline Andrews,[19] Timothy R. Rebbeck,[19,20] Akindele O. Adebiyi,[7] Ilir Agalliu,[18] and Joseph Lachance[1,21,*]

Summary

Androgenetic alopecia is a highly heritable trait. However, much of our understanding about the genetics of male-pattern baldness comes from individuals of European descent. Here, we examined a dataset comprising 2,136 men from Ghana, Nigeria, Senegal, and South Africa that were genotyped using the Men of African Descent and Carcinoma of the Prostate Array. We first tested how genetic predictions of baldness generalize from Europe to Africa and found that polygenic scores from European genome-wide association studies (GWASs) yielded area under the curve statistics that ranged from 0.513 to 0.546, indicating that genetic predictions of baldness generalized poorly from European to African populations. Subsequently, we conducted an African GWAS of androgenetic alopecia, focusing on self-reported baldness patterns at age 45. After correcting for age at recruitment, population structure, and study site, we identified 266 moderately significant associations, 51 of which were independent ($p < 10^{-5}$, $r^2 < 0.2$). Most baldness associations were autosomal, and the X chromosome does not seem to have a large impact on baldness in African men. Although Neanderthal alleles have previously been associated with skin and hair phenotypes, within the limits of statistical power, we did not find evidence that continental differences in the genetic architecture of baldness are due to Neanderthal introgression. While most loci that are associated with androgenetic alopecia do not have large integrative haplotype scores or fixation index statistics, multiple baldness-associated SNPs near the *EDA2R* and *AR* genes have large allele frequency differences between continents. Collectively, our findings illustrate how population genetic differences contribute to the limited portability of polygenic predictions across ancestries.

Introduction

Most complex traits are polygenic, and their genetic architectures have been shaped by evolutionary history.[1,2] Here, genetic architecture refers to the specific genetic variants that influence a trait, as well as their allele frequencies and effect sizes.[3] Neutral phenomena such as population bottlenecks and geographical isolation can cause the relative importance of different genetic risk factors to vary by population.[4,5] Similarly, natural selection can also impact population-specific genetic architectures.[5–7] Differences in the genetic architectures of complex traits between genetically distant populations can arise from differences in causal allele frequencies, epistasis, and gene-environment interactions.[8] Because of this, genetic predictors that are developed in one part of the world often transfer poorly to other parts of the world.[9–11] Furthermore, the generalizability of polygenic predictions is known to vary by trait.[11,12]

Androgenetic alopecia, commonly known as male-pattern baldness (MPB), is a model trait due to its variable prevalence across populations, relevance to health, and high heritability. This condition typically involves progressive loss of hair above the temples and at the vertex of the scalp due to the progressive miniaturization of hair follicles.[13] Notably, hair morphology varies by genetic ancestry,[14,15] and the prevalence of androgenetic alopecia varies across the world.[16–18] Men of non-European descent

[1]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA; [2]Department of Biology, Morgan State University, Baltimore, MD, USA; [3]Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA; [4]Strengthening Oncology Services Research Unit, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; [5]Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; [6]National Cancer Registry, National Institute for Communicable Diseases a Division of the National Health Laboratory Service, Johannesburg, South Africa; [7]College of Medicine, University of Ibadan, Ibadan, Nigeria; [8]Centre for Proteomic and Genomic Research, Cape Town, South Africa; [9]Mediclinic Precise Southern Africa, Cape Town, South Africa; [10]Université Cheikh Anta Diop de Dakar, Dakar, Senegal; [11]Université Iba Der Thiam de Thiès, Thiès, Senegal; [12]Korle-Bu Teaching Hospital and University of Ghana Medical School, Accra, Ghana; [13]Department of Pathology, University of Ghana Medical School, Accra, Ghana; [14]37 Military Hospital, Accra, Ghana; [15]Université Cheikh Anta Diop de Dakar, Dakar, Senegal; [16]College of Health Sciences, University of Abuja, University of Abuja Teaching Hospital and Cancer Science Centre, Abuja, Nigeria; [17]Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa; [18]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA; [19]Dana-Farber Cancer Institute, Boston, MA, USA; [20]Harvard T.H. Chan School of Public Health, Boston, MA, USA
[21]Lead contact
*Correspondence: joseph.lachance@biology.gatech.edu
https://doi.org/10.1016/j.xhgg.2025.100428.

are more likely to retain their frontal hair lines than men of European descent and the onset of MPB tends to be later in Japanese men.[19] Rates of alopecia areata, an inflammatory condition of hair follicles that results in patchy bald spots, also vary by ancestry.[20] Disparities exist for other forms of alopecia, including scarring alopecia (i.e., central centrifugal cicatricial alopecia), which is disproportionally observed in women of African descent.[21] Importantly, baldness is not only a cosmetic trait—it has also been associated with endocrine, metabolic, and cardiovascular diseases, likely due to the myriad effects of sex hormones.[22] Baldness has long been associated with heredity,[23] and men whose fathers and/or maternal grandfathers are bald are more likely to experience hair loss.[24] The pedigree heritability of MPB in the UK Biobank (UKBB) was estimated to be 62%.[25] Genetic risk factors for MPB in European populations include polymorphisms at 20p11 and Xq12, including alleles near the *EDA2R* (ectodysplasin A2 receptor; OMIM: 300276) and *AR* (androgen receptor; OMIM: 313700) genes.[26,27] Intriguingly, introgressed DNA from Neanderthals is enriched for genes that are involved in keratin formation, i.e., skin and hair phenotypes.[28,29] Because Neanderthal DNA is almost exclusively found in non-African genomes,[30] ancient introgression may contribute to continental differences in the genetic architecture of MPB.

During the past few years several genome-wide association studies (GWASs) have expanded our understanding of the genetic architecture of MPB.[31–34] Analyzing data from more than 22,000 men of European descent, Heilmann-Heimbach et al.[32] conducted an early meta-analysis of MPB. European individuals in the top quartile of a polygenic score (PGS) built from this study were more likely to be bald than individuals in the lowest quartile (odds ratio = 4.16; 63 independent GWAS loci).[32] In a study of more than 70,000 men, Pirastu et al.[33] identified 71 susceptibility loci that were enriched for *Wnt* signaling and apoptosis pathways. In an analysis of more than 52,000 UKBB participants, Hagenaars et al.[34] identified more than 270 independent autosomal and X-linked variants associated with hair loss. The Hagenaars et al.[34] PGS using these variants was able to distinguish between European men who had severe hair loss vs. no hair loss for prediction models that included age as a covariate (area under the curve [AUC] = 0.79). A more recent study by Chen et al.[35] that used feature selection before training a polygenic predictor on the UKBB dataset reported slightly boosted AUC statistics of severe hair loss vs. no hair loss in European populations (AUC = 0.81 when age was included as an explanatory variable). MPB-associated loci from the UKBB are enriched for weak signatures of negative selection, perhaps due to pleiotropic effects on other traits.[25] Additional analyses of more than 72,000 exomes from the UKBB suggest that rare variants make only a small contribution to the prevalence of MPB in Europe.[36] However, one limitation of these large-scale genetic studies is that they largely focused on European cohorts, thereby skewing our understanding of variants that contribute to MPB. Consequently, it is unknown if the genetic architecture of MPB differs for African men.

To overcome existing knowledge gaps, we tested how well European-ascertained polygenic risk scores for MPB generalize to sub-Saharan African populations. We then inferred the population genetic architecture of this complex trait by performing an African GWAS of androgenetic alopecia. Our subsequent analyses explored multiple evolutionary hypotheses for the poor portability of genetic predictions. These potential causes include private alleles that are due to recent mutations, Neanderthal alleles that are only informative about baldness in Europeans, and natural selection acting on baldness-associated loci.

Subjects and methods

Baldness phenotyping

Participants included in this study were African men without a diagnosis of prostate cancer or any other cancers who were recruited as controls for a large case-control study by the Men of African Descent and Carcinoma of the Prostate (MADCaP) Network.[37,38] These men were recruited from seven study sites: Hôpital Général de Grand Yoff/Institut de Formation et de Recherche en Urologie in Dakar, Senegal; 37 Military Hospital in Accra, Ghana; Korle-Bu Teaching Hospital in Accra, Ghana; University College Hospital in Ibadan, Nigeria; University of Abuja Teaching Hospital in Abuja, Nigeria; Wits Health Consortium/National Health Laboratory Services in Johannesburg, South Africa; and Stellenbosch University in Cape Town, South Africa. Informed consent was obtained from all study participants. At each study site, participants were recruited using protocols approved by that site's institutional review board/ethics review board and the central data management center at the Dana-Farber Cancer Institute. Participants completed an interviewer-administered questionnaire that queried them about their hair patterns at age 30 and at age 45 using the Hamilton-Norwood baldness scale. To better integrate our data with previous work (e.g., data field 2395 in the of UKBB data) and facilitate an ordinal GWAS of androgenetic alopecia, we rescaled Hamilton-Norwood scores into four different categories: no hair loss, slight hair loss, moderate hair loss, severe hair loss (Figure 1). We then excluded all individuals who reported incongruent baldness scores (i.e., more hair at age 45 than at age 30). To avoid the potential for reporting bias and/or misclassification, our subsequent analyses focus on self-reported baldness scores at age 45, while using age at study recruitment to correct for recall errors. Table S1 contains a full list of the number of individuals per study site in each baldness class and Figure S1 is a PCA plot showing the population genetic structure of this dataset. A total of 2,136 African individuals were studied.

Genotyping, quality control, and imputation

Genotype data were inferred using the MADCaP Array, a custom genotyping platform that is optimized for detecting genetic associations in sub-Saharan Africa.[39] This array is known commercially as the Axiom Pan-African Cancer Research Array. Details about SNP calling can be found in other MADCaP Network publications.[39,40] We performed standard quality control (QC) procedures using PLINK. Samples were included if call rates exceeded 98%, they were not related to other samples (kinship coefficient <0.0884, i.e., second degree of closer relationships

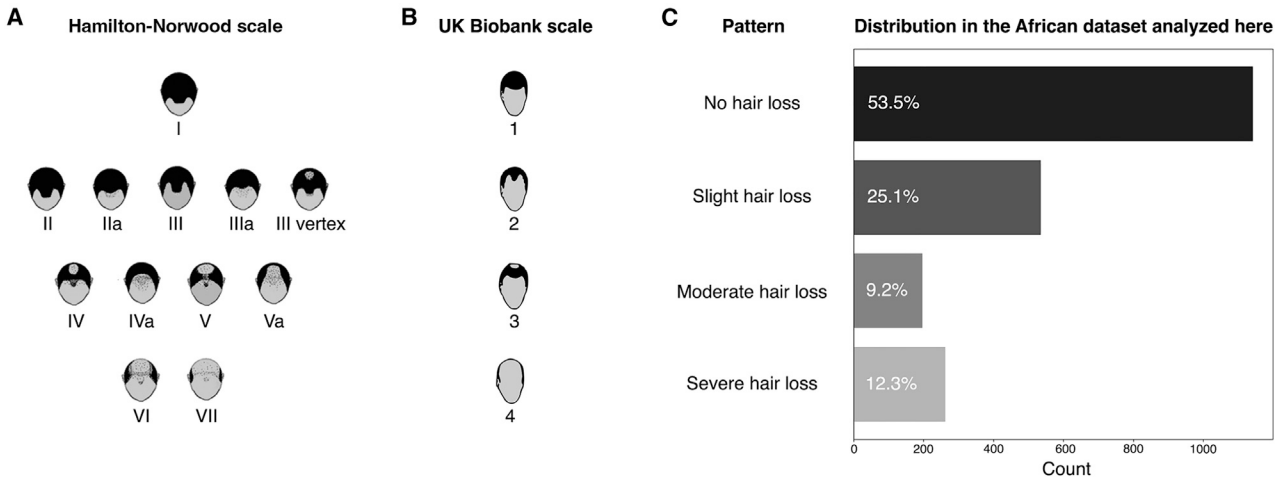

Figure 1. Baldness phenotypes

(A) The Hamilton-Norwood baldness scale partitioned into four categories.

(B) Baldness scale from question 2395 of the UKBB questionnaire.

(C) Proportions of each baldness pattern in the African dataset that was analyzed here. Individuals were sampled from Senegal, Ghana, Nigeria, and South Africa. Study-site-specific counts of each baldness phenotype can be found in Table S1.

were excluded), and they had an African ancestry percentage of more than 70% (as per previous work by Janivara et al.[40]). Markers were included if genotype missingness was less than 5%, a minor allele frequency (MAF) filter was passed (MAF >0.01 in our pooled African dataset), and there were no detectable departures from Hardy-Weinberg proportions ($p < 10^{-5}$). After these QC filters, 2,136 unrelated samples with 1,092,609 markers remained. We then imputed additional markers using the TOPMed Imputation Panel and Server.[41] Previous work has found that the TOPMed imputation panel performs well on African datasets in general,[42] and the MADCaP Network dataset in particular.[40] Post-imputation QC filtering included imputation quality (r^2 score >0.8) and MAF cutoffs (MAF >0.01), yielding a total of 15,378,257 autosomal and 583,558 X chromosome polymorphisms.

Tests of PGS performance

To assess the portability of genetic predictors of MPB, we extracted lists of baldness-associated SNPs from the Hagenaars et al.[34] PGS and the Heilmann-Heimbach et al.[32] PGS. Using the allele dosage and effect sizes of each trait-associated SNP, we calculated PGS for 2,136 African men. After correcting for covariates (present age, sample site, and population structure—the first three PCs in PCA), the prediction accuracy of each predictor was estimated using the nonparametric Bayesian inference of the covariate-adjusted receiver operating characteristic curve package in R.[43] Thus, the AUC statistics reported here isolate the effects of genetic risk factors without being inflated by the inclusion of other predictors. Because PGS performance can be more effective for individuals at the extreme ends of distributions,[44] we performed additional tests of PGS performance by comparing the relative proportions of individuals in each phenotypic class for individuals in the top 10% of each PGS distribution to individuals in the middle 20% of each PGS distribution. These analyses were repeated for both the Hagenaars et al.[34] PGS and the Heilmann-Heimbach et al.[32] PGS.

African GWAS analyses

We performed an ordinal GWAS of MPB using the MADCaP Network dataset, focusing on the four ordered phenotypes in Figure 1: no hair loss, slight hair loss, moderate hair loss, and severe hair loss (i.e., baldness scores range from 1 to 4, as per question 2395 of the UKBB questionnaire). The proportional odds logistic mixed model package[45] was used to estimate p values and effect sizes of each genetic variant. Our ordinal GWAS also corrected for multiple covariates: current age (to correct for recollection errors in self-reported baldness phenotypes), recruitment study site (to correct for possible batch effects in recruitment and referral bias), and the first three principal components (to correct for population structure, Figure S1). LDlink[46] was used to identify if African GWAS hits were in linkage disequilibrium (LD) with previous baldness hits. We used the LDexpress tool[47] to examine whether any of our African associations with MPB were in LD with GTEx eQTLs, i.e., variants that affect gene expression (thresholds, $r^2 \geq 0.2$; GTEx $p < 10^{-10}$; genomic window, ±500 kb). GWAS summary statistics and population genetic statistics for our top 51 independent marginal hits are reported in Table S2. A gene set enrichment analysis was performed using snpXplorer.[48] This analysis used gene sets from Gene Ontology: Biological Processes, Kyoto Encyclopedia of Genes and Genomes, Reactome, and WikiPathways.

To examine regional differences in the genetic architecture of MPB across Africa, we also conducted a mini GWAS of 1,058 West African individuals and a mini GWAS of 1,078 South African individuals. We report regional association statistics (p values and effect sizes), regional allele frequencies, and genetic heterogeneity statistics (i.e., I^2 statistics)[49] for our top 51 independent marginal hits in Table S3.

The total SNP heritability of MPB for each chromosome was inferred using a modified version of the LD adjusted kinships (LDAK)[50] approach, specifically Basic Protocol 2 as described by Srivastava et al.[51] This protocol included additional MAF (MAF ≥ 0.05) and LD-pruning (–indep-pairwise 100 50 0.1) filters in PLINK, as well as pruning out individuals with crypted relatedness using a genetic relatedness matrix (GRM) generated from autosomal data (max relatedness = 0.05). SNP heritability was calculated using REML for chromosomes 1–22 together using 22 GRMs, while SNP heritability for the X chromosome was computed separately. We estimated 95% confidence intervals

(CIs) for the proportion of SNP heritability due to each chromosome using a resampling approach. Chromosome-specific values of SNP heritability were resampled using estimates of the mean and standard deviation reported by LDAK. Proportions were generated by dividing resampled values for each chromosome by the summed resampled values of all chromosomes, i.e., we generated normalized 23-element vectors of chromosomal SNP heritability proportions. This procedure was repeated 10,000 times to generate 95% CIs for each chromosome. We performed linear regression of the chromosome size against the proportion of heritability explained in our African GWAS using the *linregress* function in scipy v1.10.0.[52] The impact of rare vs. common variants on MPB in Africa was quantified using genetic variance contribution (*gvc*) statistics of individual polymorphisms,[53] where $gvc = 2\beta^2 p(1-p)$ given an effect size β and an allele frequency p. LD clumping ($r^2 < 0.2$) was used to reduce the 162,576 polymorphisms with GWAS p values of less than 0.01 to an independent set of 29,251 associations. We then divided these polymorphisms into two bins, corresponding with rare (MAF < 0.05) and common (MAF ≥ 0.05) variants, and found what fraction of the overall summed *gvc* belonged to each bin.

Assessing statistical power

To assess whether a sample size of 2,136 individuals was sufficient to distinguish continental differences in the genetic architecture of MPB, we quantified how well the 42 autosomal predictors of MPB from Heilmann-Heimbach et al.[32] performed in a downsampled UKBB cohort and our MADCaP Network dataset. The downsampled UKBB analysis focused on 2,136 European ancestry individuals aged 45–50 years with baldness patterns and proportions matching those in the MADCaP dataset (data field 2395 from their first visit). A downsampled European ancestry GWAS was conducted using the same procedures as our African ancestry GWAS of MPB. Subsequent analyses included (1) quantifying PGS performance using AUC statistics, (2) statistical power calculations, (3) examining whether MPB-associated alleles are enriched for low p values in replication cohorts, and (4) comparisons of minor allele frequencies and effect sizes in different replication cohorts. Additional details about these approaches can be found in the supplemental methods subsection of the supplemental information.

Evolutionary genetics analyses

To perform enrichment analyses of MPB-associated SNPs, we generated comparable sets of control SNPs that were not associated with the trait in our study. This involved obtaining 1,000 independent ($r^2 < 0.5$) autosomal control SNPs from the 1000 Genomes Project[54] phase 3 African populations (1KGP AFR) for every MPB-associated SNP. Control SNPs were matched to MPB associations for allele frequency (±5% in 1KGP AFR for African MPB associations, ±5% in 1KGP EUR for Heilmann-Heimbach et al.[32] and Hagenaars et al.[34] MPB associations). Because SNPs can vary in the number of polymorphisms they tag, control SNPs were also matched in terms of the number of SNPs in LD with the target SNP (±10%, $r^2 < 0.5$ in 1KGP AFR). For instance, if an MPB-associated SNP had 20 SNPs in LD ($r^2 < 0.5$), a control SNP would be select that between 18 and 22 SNPs in LD. The proximity of control SNPs to coding DNA was also matched to that of MPB-associated SNPs: control SNPs for exonic variants were required to also be exonic, and control SNPs for intronic and intergenic variants were selected to have a similar distance in base pairs to the nearest exonic position (±5%). Finally, control SNPs were also matched to MPB SNPs with respect to local gene density, defined here as the number of genes located within 500 kb of each SNP. Specifically, control SNPs were chosen to be near similar (±10%) numbers of genes around the SNP ($r^2 < 0.5$, physical distance of 500 kb).

Tests of enrichment of Neanderthal DNA used Skov's introgression map (Supplementary dataset 1 in Skov et al.[55]), which was originally generated from 27,566 Icelandic genomes. This introgression map constitutes the most comprehensive characterization of the introgression landscape in Europeans to date. We calculated Neanderthal introgression frequencies for non-overlapping 10-kb windows across the genome. In two separate sub-analyses, we intersected MPB-associated variants from Hagenaars et al.[34] and Heilmann-Heimbach et al.[32] with the tiled map of introgression frequencies. Empirical cumulative distribution functions (ECDFs) of introgression frequencies of MPB-associated variants were then compared with the ECDFs of introgression frequencies of 1,000 sets of bootstrapped matched control SNPs, i.e., we tested whether Europeans are more likely to have Neanderthal DNA in genomic regions that are associated with MPB compared with other parts of the genome. Statistical significance was assessed using 1,000 Kolmogorov-Smirnov tests. The harmonic mean p value[56] of these 1,000 tests was used to summarize whether MPB-associated variants are in genomic regions that are enriched for Neanderthal DNA. For these analyses, we only included associations from Hagenaars et al.[34] and Heilmann-Heimbach et al.[32] that had at least one matched control SNP (i.e., 185 associations and 38 associations, respectively). This approach has approximately 80% power to detect a Neanderthal enrichment at a significance level of 0.05 if at least 49 of 185 Hagenaars et al.[34] and 19 of 38 Heilmann-Heimbach et al.[32] MPB associations are within introgressed fragments in Skov's introgression map (see supplemental information and Figure S2). CrossMap[57] v0.6.4 was used to harmonize hg19 and hg38 genomic positions.

Population genetic analyses used genetic data from the 1KGP. Joint site frequency spectrum plots were generated using individuals of African ancestry (population codes: ACB, ASW, ESN, GWD, LWK, MSL, and YRI) and European ancestry (CEU, FIN, IBS, GBR, and TSI). A full list of 1KGP population codes can be found at: https://www.internationalgenome.org. Trait-level tests of polygenic selection leveraged integrative haplotype scores (iHSs) of autosomal variants.[58] A full description of this approach has been described elsewhere.[59] Briefly, this involved inferring whether sets of MPB-associated SNPs are more likely to have outlier values of iHS statistics (iHS < -1.96 or iHS > 1.96) than sets of matched control SNPs. We then obtained empirical percentiles for Hagenaars et al.[34] Heilmann-Heimbach et al.,[32] and African MPB-associations by comparing outlier proportions for each set of MPB-associated SNPs with 1,000 matched sets of control SNPs. Fixation index (F_{ST}) calculations used the following equation: $F_{ST} = Var(p)/(\overline{p}(1-\overline{p}))$, where $Var(p)$ refers to the variance in allele frequencies across Europe and Africa and $\overline{p}$ refers to the mean allele frequency across both continental populations.[60] African allele frequencies in these calculations were generated by pooling ESN, GWD, LWK, MSL, and YRI individuals together, while European frequencies were generated by pooling CEU, FIN, IBS, GBR, and TSI. individuals together. Continental frequencies of haplotypes at Xq12 were calculated from 1KGP data using LDlink[46] and genotype data at three SNPS: rs12558842, rs2497911, and rs1204041. Ancestral and derived state information for all SNPs was obtained from the 1KGP phase 3 dataset. As described in

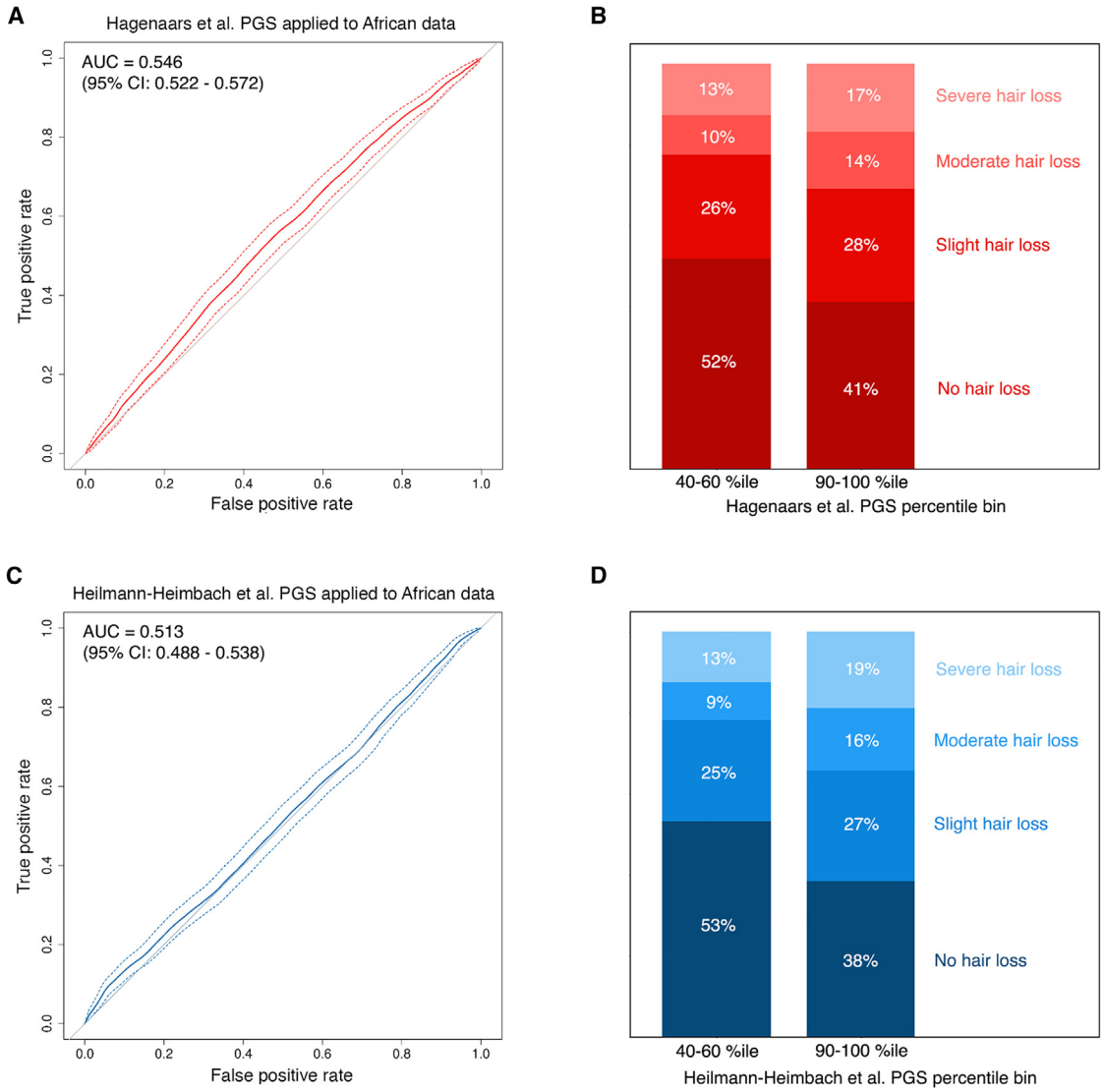

Figure 2. Performance of PGSs in Africa
ROC curves focus on comparisons between individuals who have any baldness vs. no baldness. The 95% CIs are indicated by dotted lines. AUC statistics corrected for present age, study site, and population structure.
(A) AUC statistics of the Hagenaars et al.[34] PGS applied to African data.
(B) Baldness patterns of African men in the middle 20% and top 10% of the Hagenaars et al.[34] PGS distribution.
(C) AUC statistics of the Heilmann-Heimbach et al.[32] PGS applied to African data.
(D) Baldness patterns of African men in the middle 20% and top 10% of the Heilmann-Heimbach et al.[32] PGS distribution.

section 8.3 of the supplementary material of the 1KGP phase 3 paper,[54] ancestral alleles were inferred using the six-way EPO alignments from Ensemble v71.

Results

Portability of baldness predictions to sub-Saharan Africa

Although multiple polygenic predictors of MPB exist, they have mostly been ascertained in European populations. To find out whether these genetic predictors are informative regarding MPB in Africa, we applied two different PGS to the MADCaP Network dataset. Although the Hagenaars et al.[34] PGS performed well when tested on a European ancestry cohort, it had only a limited ability to distinguish between any hair loss vs. no hair loss in African men (Figure 2A) (AUC = 0.546; 95% CI: 0.522–0.572). Indeed, it performed poorly for slight hair loss vs. no hair loss (AUC = 0.533; 95% CI: 0.503–0.562), moderate hair loss vs. no hair loss (AUC = 0.558; 95% CI: 0.514–0.601), and severe hair loss vs. no hair loss (AUC = 0.565; 95% CI: 0.527–0.603). Despite these low AUC statistics, individuals in the top 10% of the Hagenaars et al.[34] PGS distribution were 33% more likely to have severe baldness at age 45 than individuals in the middle 20% of the PGS distribution (Figure 2B). Similarly, although the Heilmann-Heimbach et al.[32] PGS also performed reasonably well on a European cohort, its ability to distinguish between any hair loss vs.

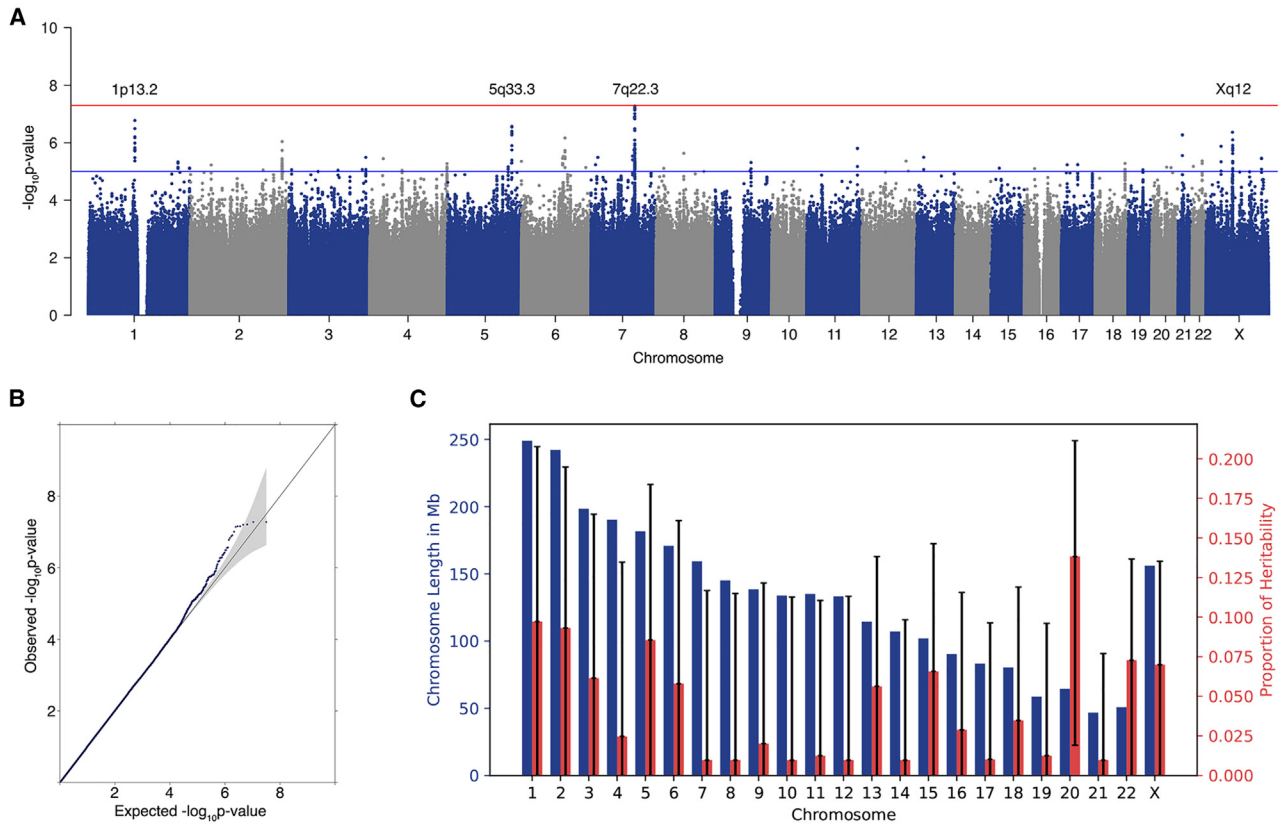

Figure 3. African GWAS of androgenetic alopecia
(A) Manhattan plot for the ordinal GWAS of MPB. Phenotypes were scored using a 4-point scale. Sample size: 2,136 African men.
(B) QQ plot for the ordinal GWAS of MPB in Africa ($\lambda_{GC} = 0.998$).
(C) Physical sizes (blue) and relative contribution to SNP heritability of each chromosome. Error bars indicate 95% CIs.

no hair loss was effectively no better than chance when applied to African men (Figure 2C) (AUC = 0.513; 95% CI: 0.488–0.538), and it performed poorly for slight hair loss vs. no hair loss (AUC = 0.504; 95% CI: 0.474–0.533), moderate hair loss vs. no hair loss (AUC = 0.538; 95% CI: 0.496–0.582), and severe hair loss vs. no hair loss (AUC = 0.516; 95% CI: 0.477–0.555). However, individuals in the top 10% of the Heilmann-Heimbach et al.[32] PGS distribution were 50% more likely to be severely bald at age 45 than individuals in the middle 20% of the PGS distribution (Figure 2D). Overall, our results reveal that the genetic predictions of androgenetic alopecia developed from European populations do not transfer well to sub-Saharan Africa.

An African GWAS of MPB

Given the poor portability of genetic predictors from European populations, we performed a GWAS of MPB in Africa to identify continental differences in the genetic architecture of this complex trait (Figure 3A). For this we used the MADCaP dataset, which contains 2,136 samples from Senegal, Ghana, Nigeria, and South Africa. This GWAS used a proportional odds logistic mixed model which grouped the different intensities of baldness into four ordinal classes. Following adjustments for population structure, study site, and participant age, the genomic inflation factor in this analysis was minimal ($\lambda_{GC} = 0.998$), indicating the absence of batch effects and confounding. Although our sample size lacked the statistical power to capture any genome-wide significant associations, we found 51 independent marginally significant associations ($p < 10^{-5}$; LD pruning threshold, $r^2 < 0.2$) (Figure 3A). The QQ plot shown in Figure 3B provides additional support that these marginal associations are not due to population stratification. A full list of these African hits, including their effect sizes, can be found in Table S2. Among the 51 marginal associations, 36 are not in LD with previously reported GWAS hits for any trait in the NHGRI-EBI GWAS Catalog[61] ($r^2 < 0.2$). Key loci that were associated with MPB in Africa include 7q22.2, 1p13.2, and Xq12 (Figure 3A). The 7q22.2 locus contains moderately significant variants in the intronic regions of *COG5* (OMIM: 606821) and *HBP1* genes (OMIM: 616714). The lead African SNP at the 1p13.2 locus (rs116494345) is monomorphic in Europe and Asia, i.e., its impact on MPB is Africa specific. The lead SNP at the Xq12 locus (rs1204041) is in the intronic region of the *AR* (androgen receptor) gene. Other notable baldness associations are rs191219783 (proximal to *CCR7*, a chemokine receptor gene that is expressed by dendritic cells in hair

follicles[62]) (OMIM: 600242) and rs143451223 (proximal to *SLC30A10*, a manganese transporter gene) (OMIM: 611146).

We also examined how well our top marginal significant associations with MPB replicate in West Africa and South Africa, and whether any individual SNPs make an outsized contribution to the genetic architecture of MPB in different regions of Africa. 45 of our 51 marginal associations had p values of less than 0.05 in both West Africa and South Africa, while the remaining six had p values of greater than 0.05 in West Africa and less than 0.05 in South Africa (Table S3). Similarly, 45 of the 51 marginal associations have alleles with the same direction of effect for both West and South Africa. Regional effect sizes are also positively correlated (Pearson's correlation coefficient, 0.738). Thus, detectable African associations with MPB are largely driven by moderate evidence from both regions of Africa, as opposed to strong evidence from only one region. Despite these general patterns, I^2 statistics reveal evidence of genetic heterogeneity across Africa for multiple MBP-associated variants, including rs543425158 (I^2 = 90.7%), rs1583026902 (I^2 = 85.1%), rs140088608 (I^2 = 80.8%), and rs7305258 (I^2 = 80.8%). Allele frequency differences between West and South African populations also contribute to regional differences in the genetic architecture of MPB (Table S3). For example, the T allele at rs138885958 is approximately twice as common in South Africa as it is in West Africa. Last, although low-frequency private alleles can contribute the genetic architecture of complex traits, the modest sample sizes of each regional GWAS (1,058 for West Africa and 1,078 for South Africa) limit our statistical power to detect associations between MPB and rare alleles that are found in only one part of Africa.

We further explored the genetic architecture of MPB in Africa by testing whether X chromosomes are enriched for SNP heritability and by quantifying the relative importance of rare vs. common alleles. To test whether the X chromosome makes a disproportionally large contribution to the heritability of MPB, we compared relative estimates of the total SNP heritability to chromosome size. We note that larger chromosomes need not make larger contributions to variation in MPB (Figure 3C) (linear regression of chromosome length vs. proportion of SNP heritability on each chromosome, $R^2 = 0.07$ and $p = 0.21$). Interestingly, the X chromosome contributes only modestly to the genetic variance of MPB in African men; i.e., it does not seem to make an outsized contribution to the genetic variance of this complex trait. Although this pattern may reflect the underlying biology, we note that a sample size of 2,136 individuals limits our ability to accurately infer the proportion of the total SNP heritability that is due to each chromosome. Furthermore, the lower statistical power to detect X-linked associations compared with autosomal associations may also contribute to the patterns seen in Figure 3C. Note that, if a chromosome has large numbers of common, independent SNPs that are moderately associated with a trait, LDAK-REML can aggregate these effects, resulting in high heritability estimates for a chromosome despite a lack of strong associations (e.g., chromosome 20). We also found that 42% of the observed genetic variance for MPB in Africa was due to rare alleles (MAF < 0.05) and the remaining 58% was due to common alleles (MAF ≥ 0.05). These proportions are roughly comparable with what has been observed for other complex traits.[63]

Functional annotation of private alleles tends to be limited, especially when these polymorphisms are Africa specific. One byproduct of this is that 37 of our African associations with MPB are not in LD with GTEx eQTLs that affect gene expression. Of the remaining 14 African associations, 5 are in LD with skin eQTLs and 3 are in LD with testis eQTLs. We did not find evidence that the set of 51 African MPB associations was enriched for any specific biological pathway (Table S4).

Sample size, statistical power, and replication

We examined sample size considerations by focusing on how well 42 autosomal variants from Heilmann-Heimbach et al.[32] replicated in European and African ancestry cohorts of 2,136 individuals. First, we quantified the relative performance of a European-ancestry PGS on a downsampled UKBB cohort and our MADCaP Network dataset. As seen in Figure S3, the downsampled UKBB cohort has an AUC statistic of 0.616, i.e., 2,136 samples are sufficient to infer that the polygenic predictor performs better than random chance. By contrast, the AUC statistic for the MADCaP Network cohort is 0.511. The separation between the UKBB and MADCaP ROC curves is statistically significant (Delong's test, $p = 1.58 \times 10^{-9}$), consistent with the existence of continental differences in the genetic architecture of MPB. Second, we used power calculations to test whether 2,136 samples are sufficient to detect enrichment for low p values in a replication cohort. As seen in Figure S4A, this was indeed the case for power calculations that used European and African allele frequencies. Third, we used empirical data to test whether PGS variants were enriched for low p values in replication cohorts. Focusing on the downsampled UKBB GWAS, we found that PGS variants from Heilmann-Heimbach et al.[32] were more likely to have low p values than null expectations (Figure S4B, orange curve), i.e., sample size did not present a major limitation to replication in an ancestry-matched cohort. By contrast, we found that PGS variants were only slightly enriched for low p values in a MADCaP Network dataset that was not ancestry matched to the Heilmann-Heimbach et al.[32] study (Figure S4B, green curve). Thus, the genetic architecture of MPB in the European ancestry Hellmann-Heimbach et al. study seems to be more similar to that of the UKBB than the MADCaP Network dataset. Fourth, we examined summary statistics of individual MPB-associated loci. Sample sizes of 2,136 yielded only moderately strong p values for individual MPB associations in each replication cohort (Table S5). We found that 3 of 42 autosomal

associations from Heilmann-Heimbach et al.[32] had p values of less than 0.05 in a 2,136-sample African ancestry GWAS, and 15 of 42 autosomal associations from Heilmann-Heimbach et al.[32] had p values of less than 0.05 in a 2,136-sample European ancestry GWAS (Figure S5). MPB associations with the strongest replication p values did not always have large minor allele frequencies in replication cohorts (Figures S5A and S5B), nor were they always variants with large Heilmann-Heimbach et al.[32] effect sizes (Figure S5C). Finally, we note that 19 of the 42 autosomal variants had overlapping effect sizes in the downsampled UKBB cohort and our MADCaP Network dataset (Figure S6).

Evolutionary genetics of MPB

We investigated multiple evolutionary hypotheses to explain the poor portability of polygenic predictions and continental differences in the genetic architecture of MPB. Possible explanations involve the presence of private alleles, ancient introgression of Neanderthal DNA, and natural selection acting on genomic regions that have been associated with baldness. Comparisons between the allele frequencies of baldness-associated loci in different continents reveal the existence of both ascertainment bias and population-specific variation (Figure S7). Asymmetries in joint site frequency spectrum plots arise because genetic associations can only be detected for markers that are polymorphic in the discovery population.[64] In Figure S7, European ascertainment yields an excess of SNPs on the left and right sides of Figures S7A and S7B and African ascertainment yields an excess of SNPs on the top and bottom sides of Figure S7C. Notably, 34% of the variants in the Hagenaars et al.[34] PGS and 21% of the variants in the Heilmann-Heimbach et al.[32] PGS are near monomorphic in Africa (MAF < 0.01). Similarly, 54% of independent African associations implicated in this present study are near monomorphic in Europe (MAF < 0.01).

While previous studies have noted that Neanderthal introgression impacts genes that affect skin and hair morphology-related traits in European genomes,[29,65,66] none so far have tested for a direct association between introgressed alleles and MPB. Importantly, African genomes are largely free of Neanderthal DNA. This difference between European and African genomes could be a potential source of divergence in the genetic architecture of MPB. Hence, we tested for enrichment for Neanderthal introgression in autosomal MPB-associated variants that were ascertained in European and African populations. Using Skov's introgression map,[55] we calculated the introgression frequencies of Neanderthal DNA in non-overlapping 10-kb windows around baldness-associated SNPs and compared these frequencies against 1,000 sets of matched control SNPs. If Neanderthal introgression is a driver of differences in the genetic architecture of MPB, then we would expect to find that European-ascertained baldness loci are enriched for Neanderthal DNA compared with the rest of the genome. However, neither MPB associations from Hagenaars et al.[34] ($p = 0.79$) nor from Heilmann-Heimbach et al.[32] ($p = 0.81$) were enriched for Neanderthal DNA (Figures 4A and 4B). However, we caution that these analyses have limited power. Simulations revealed that we have 80% power to detect Neanderthal enrichment at a significance level of 0.05 if at least approximately 26% and approximately 50% of the MPB associations from Hagenaars et al.[34] and Heilmann-Heimbach et al.[32] were truly introgressed, respectively (Figure S2).

Natural selection can contribute to poor portability of polygenic predictions.[67] If local adaptation has been a driver of differences in the genetic architecture of MPB, then we would expect baldness-associated loci to be enriched for outliers of selection scans. Here, we tested for evidence of polygenic selection by examining whether sets of baldness-associated alleles have greater extended haplotype homozygosity than sets of control SNPs in 26 global populations from the 1KGP. Regardless of ascertainment scheme, sets of baldness-associated SNPs were not enriched for outlier iHS statistics compared with the rest of the genome (Figure 4C) (see Table S2 for iHS statistics of individual MPB associations from our African GWASs).

To further understand the source of variation in the genetic architecture of MPB, we investigated whether individual baldness-associated loci have divergent allele frequencies between Europe and Africa. Highly divergent loci were identified by calculating F_{ST} statistics for baldness-associated SNPs that were ascertained in European and African populations. Prior work has revealed that an F_{ST} of 0.4 corresponds with the 98.5th percentile of demographic simulations that assume neutral evolution of European and African populations.[68] Consistent with the polygenic scans of selection described above, most baldness-associated SNPs have F_{ST} statistics that resemble the rest of the genome (Figure 5). Although some SNPs are F_{ST} outliers, this need not indicate that baldness-associated SNPs are direct targets of natural selection, as genetic hitchhiking can cause allele frequencies to differ greatly across populations. High F_{ST} SNPs from European studies of baldness include rs4649041 at 1p36, rs13092705 at 3q26, rs17053607 at 4q32, rs9300169 at 12p12, and three SNPs at Xq12 (rs5965561, rs12558842, and rs2497911). High F_{ST} SNPs from our African GWAS include rs143451223 at 1q41 and rs1204041 at Xq12.

Multiple X-linked variants that are associated with MPB are evolutionary outliers.[69,70] Key associations near the *EDA2R* and *AR* genes include rs12558842 in Hagenaars et al.,[34] rs2497911 in Heilmann-Heimbach et al.,[32] and rs1204041 in our study. Although the ancestral A allele at rs12558842 is nearly fixed in East Asian genomes, it is uncommon in Africa (1KGP allele frequencies of 0.998 in East Asia, 0.836 in Europe, and 0.060 in Africa). The derived C allele at rs2497911 has an allele frequency of 0.998 in East Asia, 0.827 in Europe, and 0.041 in Africa. By contrast, the derived T allele at rs1204041 is common in Africa and all but absent from Eurasian populations (1KGP allele frequencies of 0.488 in Africa, 0.002 in Europe, and 0.000

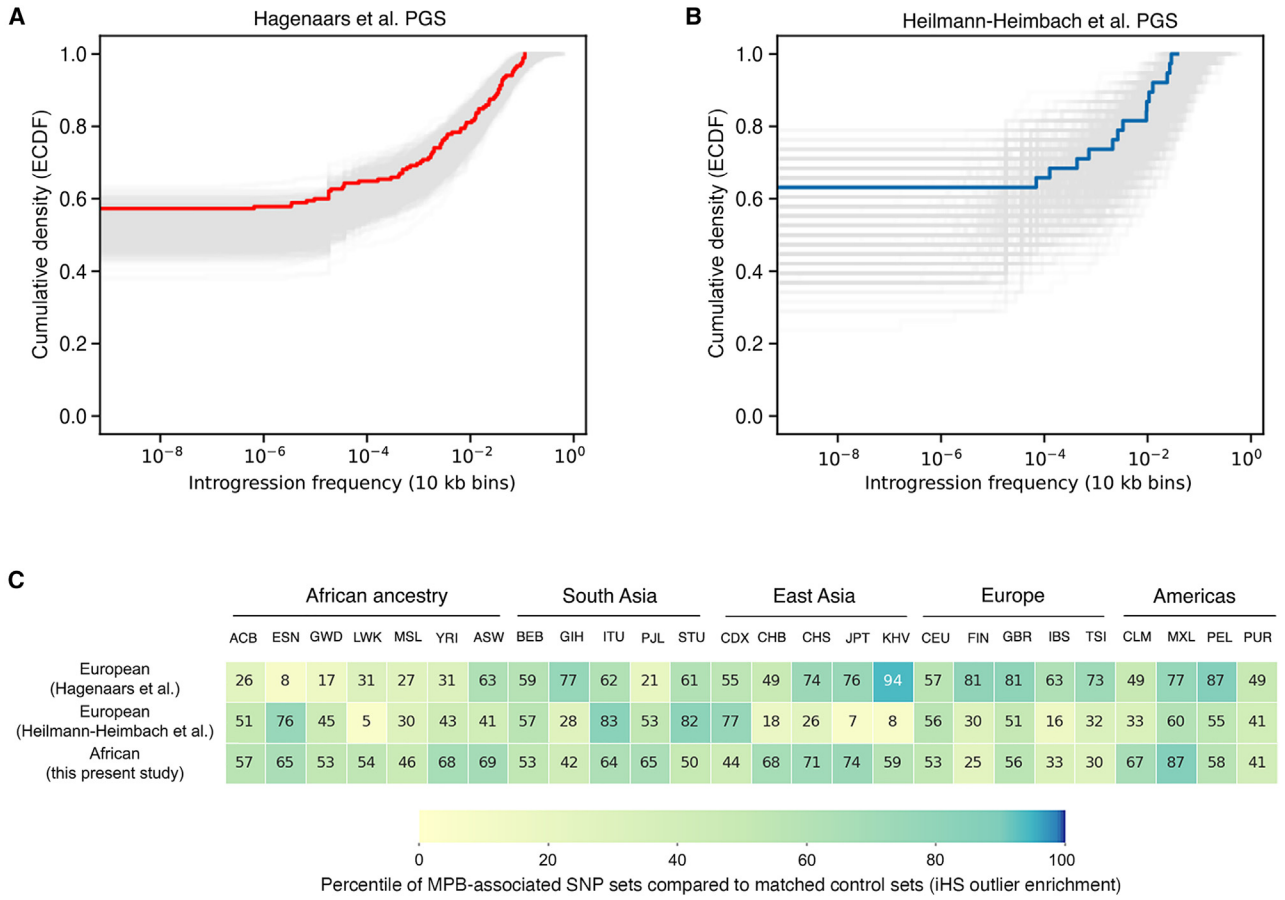

Figure 4. Tests of enrichment of Neanderthal DNA and recent polygenic selection

ECDFs of baldness-associated SNPs were compared with 1,000 sets of matched control SNPs, shown in gray. Most genomic loci have low introgression frequencies.

(A) Loci from the Hagenaars et al.[34] PGS, shown in red, are not enriched for introgressed Neanderthal DNA.

(B) Loci from the Heilmann-Heimbach et al.[32] PGS, shown in blue, are not enriched for introgressed Neanderthal DNA.

(C) Tests of iHS outlier enrichment for baldness-associated variant lists (1KGP data). Percentile ranks were generated from comparisons with 1,000 sets of matched control SNPs. Overall, autosomal SNPs that are associated with MPB are not enriched for signatures of recent positive selection.

in East Asia). These three variants are within 304 kb of each other and in LD, and additional insights can be gleaned from haplotype frequencies at Xq12 in different continental populations. We note that that the ancestral AAC Xq12 haplotype is all but absent from global populations (Figure 5E) (SNP order: rs12558842, rs2497911, and rs1204041). Intriguingly, multiple divergent haplotypes have risen to high frequency, including CAT in Africa and ACC in Eurasia (Figure 5E). These patterns are consistent with the Xq12 region having undergone multiple independent selection events. Intriguingly, the *EDA2R* gene at Xq12 is functionally related to the *EDAR* gene at 2q13 (OMIM: 604095), which has previously been associated with scalp hair thickness and implicated in a genome-wide scan of selection.[71] The *AR* gene at Xq12 encodes a steroid-hormone activated transcription factor that has many downstream effects.[72] Furthermore, we note that Xq12 is in a genomic region with low recombination rates, and divergent allele frequencies at this locus may be the byproduct of genetic hitchhiking.

Discussion

Our analyses of an African dataset revealed that genetic predictions of MPB generalize poorly across continental populations. To explore whether this was due to differences in the genetic architecture of MPB, we conducted a GWAS of this trait in sub-Saharan Africa. We identified 51 marginally significant baldness associations, many of which are due to polymorphisms that are only found in Africa. To the extent that we have statistical power to detect enrichment of introgressed DNA, we did not find evidence that Neanderthal introgression is a significant source of continental differences in the genetic basis of MPB. In general, recent positive selection does not seem to have been the primary driver of trait-level differences in the genetic architecture of MPB. Regardless, we note that some MPB-associated variants may be exceptions to this general pattern of neutral or nearly neutral evolution.

The complex evolutionary history of X-linked variants near the *EDA2R* and *AR* genes has implications for the

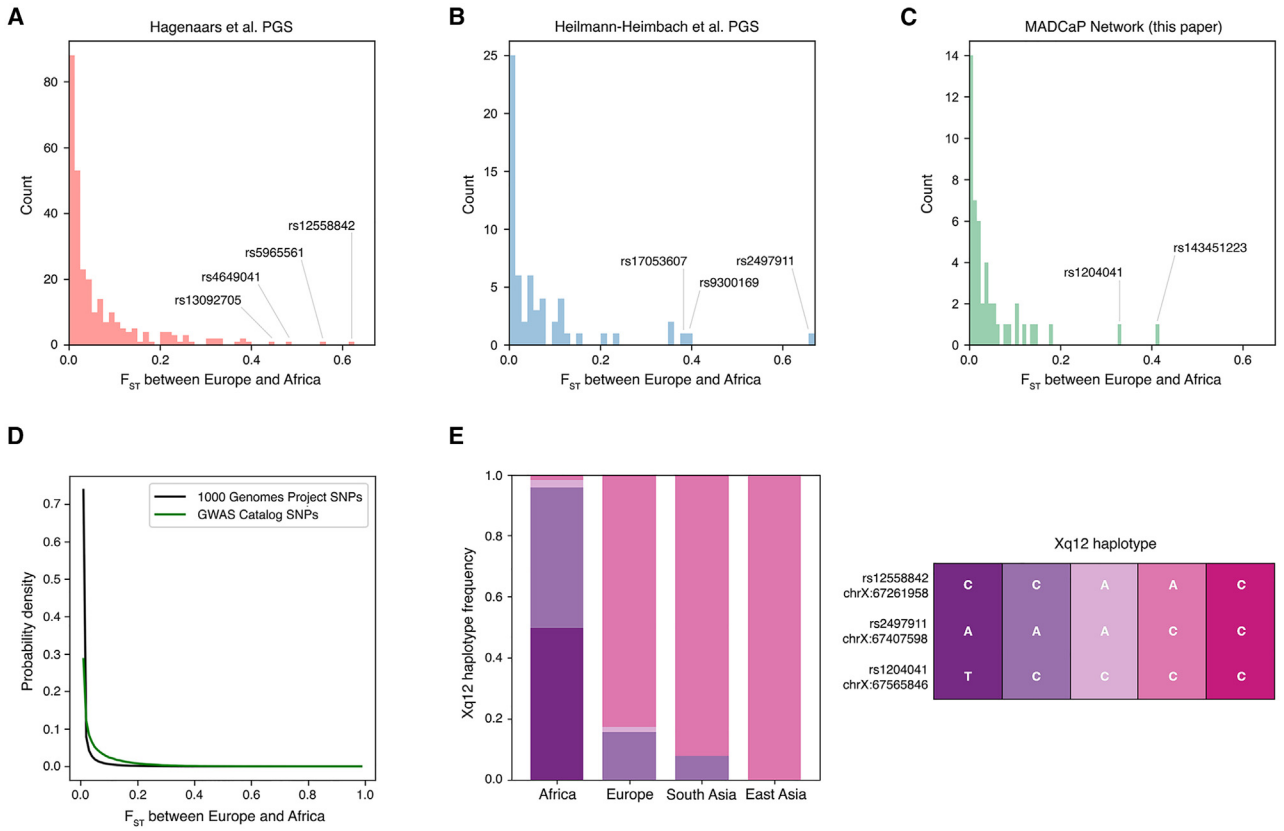

Figure 5. F_{ST} distributions and haplotype frequencies

Higher F_{ST} statistics are indicative of greater amounts of population structure, i.e., larger allele frequency differences between Europe and Africa (1KGP data).

(A) Distribution of F_{ST} statistics of baldness-associated loci from the Hagenaars et al.[34] PGS.

(B) Distribution of F_{ST} statistics of baldness-associated loci from the Heilmann-Heimbach et al.[32] PGS.

(C) Distribution of F_{ST} statistics of African baldness-associated loci.

(D) Genome-wide distributions of F_{ST} statistics from the 1KGP and SNPs in the NHGRI-EBI GWAS Catalog (all traits).

(E) Haplotype frequencies at Xq12 differ across the globe. Genomic positions refer to build hg38 and the ancestral haplotype is AAC.

prediction of androgenetic alopecia in different populations. We note that the X-linked rs12558842 polymorphism has the largest effect size in the Hagenaars et al.[34] PGS ($\beta = -0.54$ and $p < 5.1 \times 10^{-178}$) and rs2497911 has the greatest effect size in the Heilmann-Heimbach et al.[32] PGS ($\beta = 0.80$ and $p = 1 \times 10^{-320}$). Both rs12558842 and rs2497911 have minor alleles that are more common in Europe than in Africa. The African rs1204041 polymorphism has only a modest effect size ($\beta = 0.2168$ and $p = 4.60 \times 10^{-7}$). Collectively, these details suggest that knowing whether someone's maternal grandfather was bald is more likely to be informative for European men than African men.

This study adds to the growing body of evidence that genetic predictions generalize poorly across populations. Potential genetic causes of the poor portability of polygenic predictions can include allele frequency differences (including private alleles) and effect size differences. African populations also tend to have smaller LD blocks than non-African populations, and GWAS hits that tag causal variants in one population need not tag these same variants well in other populations. Although GWASs can yield valuable insights about complex traits,[12,73–75] the poor portability of findings across genetic ancestries can impact downstream applications like Mendelian randomization.[76,77] Similarly, although some authors have suggested that PGS for traits like baldness may be useful in crime scene investigations and reconstructing the phenotypes of archaic humans,[35] care must be taken when doing so, as the genetic architectures of complex traits can differ across populations.

Our study of androgenetic alopecia is not without its limitations. One caveat is that our sample sizes were relatively limited, which means that some of the marginally significant associations identified in this study are likely to be false positives. Although limited sample sizes can contribute to the limited replication of findings across studies, our analyses demonstrate that a sample size of 2,136 is sufficient to identify continental differences in the genetic architecture of MPB (Figures S3 and S4). However, we caution against overinterpreting effect size estimates for individual MPB-associated variants. African baldness patterns were obtained via a self-reported questionnaire using the Hamilton-Norwood scale. Although

this instrument has been widely used in various European populations, no classifier is perfect and phenotypic misclassification is an issue that can hamper any study,[78] including our analysis of African men and the European studies by Hagenaars et al.[34] and Heilmann-Heimbach et al.[32] We note that the UKBB cohort in Hagenaars et al.[34] had a median age of 57, while phenotypes in the MADCaP Network dataset refer to age 45. Furthermore, the African men included in our study were recruited from different hospital departments and there is potential that some of their comorbidities may have had a potential impact on the baldness score and pattern. Genotype-environment interactions may also contribute to the relatively poor portability of polygenic predictions. We note that the study sites analyzed here contain individuals from a wide range of sociolinguistic groups and genetic ancestries, and there may substantial heterogeneity in the genetic architecture of MPB and other complex traits within Africa.[79,80]

Going forward, there is a clear need to conduct additional studies of diverse African ancestry populations (including functional genomics experiments to infer molecular mechanisms that are due to private alleles and generation of large African reference panels). Future studies with larger sample sizes will have greater statistical power to detect genetic associations with population-specific rare alleles. Because early-onset MPB is indicative of the levels of sex hormones,[81] future studies combining baldness phenotypes with genetic data may be informative about the risks of other hormonal diseases.

Data and code availability

Genotype data presented in this analysis are publicly available via dbGAP (accession: phs002718.v1.p1). GWAS summary statistics can be found in the figshare repository: https://doi.org/10.6084/m9.figshare.26282122. Additional data are available through controlled access upon request from the MADCaP network (https://madcapnetwork.org/). Code is available via GitHub at https://github.com/LachanceLab/baldness_genetics.

Acknowledgments

We thank the three anonymous reviewers and members of the Center for Integrative Genomics at Georgia Institute of Technology for their helpful comments and suggestions. We also thank the individuals who participated in this study. This work was supported by NIGMS Grant R35GM133727, and it is a product of the MADCaP Network (NCI grants U01CA184374 and R01CA259200).

Declaration of interests

The authors declare no competing interests.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2025.100428.

Web resources

1000 Genomes Project (1KGP): https://www.internationalgenome.org/
GAS Power Calculator: https://csg.sph.umich.edu/abecasis/cats/gas_power_calculator/
Hagenaars et al. PGS: https://doi.org/10.1371/journal.pgen.1006594.s002
Heilmann-Heimbach et al. PGS: https://www.nature.com/articles/ncomms14694
LDlink: https://ldlink.nih.gov/
MADCaP Network: https://madcapnetwork.org/
NHGRI-EBI GWAS Catalog: https://www.ebi.ac.uk/gwas/
OMIM: https://www.omim.org/
POLMM: https://github.com/WenjianBI/POLMM/
Skov's introgression map: http://tinyurl.com/5dbwfpvk
snpXplorer: https://snpxplorer.net/
UK Biobank: https://www.ukbiobank.ac.uk

Received: January 22, 2024
Accepted: March 20, 2025

References

1. Benton, M.L., Abraham, A., LaBella, A.L., Abbot, P., Rokas, A., and Capra, J.A. (2021). The influence of evolutionary history on human health and disease. Nat. Rev. Genet. *22*, 269–283. https://doi.org/10.1038/s41576-020-00305-9.
2. O'Connor, L.J., Schoech, A.P., Hormozdiari, F., Gazal, S., Patterson, N., and Price, A.L. (2019). Extreme polygenicity of complex traits is explained by negative selection. Am. J. Hum. Genet. *105*, 456–476.
3. Timpson, N.J., Greenwood, C.M.T., Soranzo, N., Lawson, D.J., and Richards, J.B. (2018). Genetic architecture: the shape of the genetic contribution to human traits and disease. Nat. Rev. Genet. *19*, 110–124. https://doi.org/10.1038/nrg.2017.101.
4. Martin, A.R., Gignoux, C.R., Walters, R.K., Wojcik, G.L., Neale, B.M., Gravel, S., Daly, M.J., Bustamante, C.D., and Kenny, E.E. (2017). Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. Am. J. Hum. Genet. *100*, 635–649. https://doi.org/10.1016/j.ajhg.2017.03.004.
5. Sella, G., and Barton, N.H. (2019). Thinking about the evolution of complex traits in the era of genome-wide association studies. Annu. Rev. Genom. Hum. Genet. *20*, 461–493.
6. Durvasula, A., and Lohmueller, K.E. (2021). Negative selection on complex traits limits phenotype prediction accuracy between populations. Am. J. Hum. Genet. *108*, 620–631.
7. Hansen, T.F. (2006). The evolution of genetic architecture. Annu. Rev. Ecol. Evol. Syst. *37*, 123–157.
8. Lappalainen, T., Li, Y.I., Ramachandran, S., and Gusev, A. (2024). Genetic and molecular architecture of complex traits. Cell *187*, 1059–1075. https://doi.org/10.1016/j.cell.2024.01.023.
9. Privé, F., Aschard, H., Carmi, S., Folkersen, L., Hoggart, C., O'Reilly, P.F., and Vilhjálmsson, B.J. (2022). 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*, 12–23.
10. Ding, Y., Hou, K., Xu, Z., Pimplaskar, A., Petter, E., Boulier, K., Privé, F., Vilhjálmsson, B.J., Olde Loohuis, L.M., and Pasaniuc, B. (2023). Polygenic scoring accuracy varies across the genetic

ancestry continuum. Nature *618*, 774–781. https://doi.org/10.1038/s41586-023-06079-4.
11. Majara, L., Kalungi, A., Koen, N., Tsuo, K., Wang, Y., Gupta, R., Nkambule, L.L., Zar, H., Stein, D.J., Kinyanda, E., et al. (2023). Low and differential polygenic score generalizability among African populations due largely to genetic diversity. HGG Adv. *4*, 100184.
12. Carvalho, N.R., Harris, A.M., and Lachance, J. (2023). Different genetic architectures of complex traits and their relevance to polygenic score performance. Preprint at bioRxiv. https://doi.org/10.1101/2022.10.29.514295.
13. Lolli, F., Pallotti, F., Rossi, A., Fortuna, M.C., Caro, G., Lenzi, A., Sansone, A., and Lombardo, F. (2017). Androgenetic alopecia: a review. Endocrine *57*, 9–17. https://doi.org/10.1007/s12020-017-1280-y.
14. Lasisi, T. (2021). The Genetic Architecture and Evolutionary Function of Human Scalp Hair Morphology (The Pennsylvania State University).
15. Loussouarn, G., Lozano, I., Panhard, S., Collaudin, C., El Rawadi, C., and Genain, G. (2016). Diversity in human hair growth, diameter, colour and shape. An in vivo study on young adults from 24 different ethnic groups observed in the five continents. Eur. J. Dermatol. *26*, 144–154. https://doi.org/10.1684/ejd.2015.2726.
16. Oiwoh, S.O., Akinboro, A.O., Olasode, O.A., and Onayemi, E.O. (2021). Androgenetic alopecia: Prevalence and clinical characteristics in a South-West Nigerian population. Niger. J. Med. *30*, 507–513.
17. Severi, G., Sinclair, R., Hopper, J.L., English, D.R., McCredie, M.R.E., Boyle, P., and Giles, G.G. (2003). Androgenetic alopecia in men aged 40–69 years: prevalence and risk factors. Br. J. Dermatol. *149*, 1207–1213.
18. Pathomvanich, D., Pongratananukul, S., Thienthaworn, P., and Manoshai, S. (2002). A random study of Asian male androgenetic alopecia in Bangkok, Thailand. Dermatol. Surg. *28*, 804–807.
19. Asfour, L., Cranwell, W., and Sinclair, R. (2000). Male Androgenetic Alopecia. In Endotext, K.R. Feingold, B. Anawalt, M.R. Blackman, A. Boyce, G. Chrousos, E. Corpas, W.W. de Herder, K. Dhatariya, K. Dungan, and J. Hofland, et al., eds.
20. Sy, N., Mastacouris, N., Strunk, A., and Garg, A. (2023). Overall and racial and ethnic subgroup prevalences of alopecia areata, alopecia totalis, and alopecia universalis. JAMA Dermatol. *159*, 419–423.
21. Callender, V.D., and Onwudiwe, O. (2011). Prevalence and etiology of central centrifugal cicatricial alopecia. Arch. Dermatol. *147*, 972–974. https://doi.org/10.1001/archdermatol.2011.205.
22. Chen, S., Xie, X., Zhang, G., and Zhang, Y. (2022). Comorbidities in Androgenetic Alopecia: A Comprehensive Review. Dermatol. Ther. *12*, 2233–2247. https://doi.org/10.1007/s13555-022-00799-7.
23. Szasz, T.S., and Robertson, A.M. (1950). A theory of the pathogenesis of ordinary human baldness. Arch. Dermatol. Syphilol. *61*, 34–48. https://doi.org/10.1001/archderm.1950.01530080040004.
24. Chumlea, W.C., Rhodes, T., Girman, C.J., Johnson-Levonas, A., Lilly, F.R.W., Wu, R., and Guo, S.S. (2004). Family history and risk of hair loss. Dermatology *209*, 33–39. https://doi.org/10.1159/000078584.
25. Yap, C.X., Sidorenko, J., Wu, Y., Kemper, K.E., Yang, J., Wray, N.R., Robinson, M.R., and Visscher, P.M. (2018). Dissection of genetic variation and evidence for pleiotropy in male pattern baldness. Nat. Commun. *9*, 5407. https://doi.org/10.1038/s41467-018-07862-y.
26. Ellis, J.A., Stebbing, M., and Harrap, S.B. (2001). Polymorphism of the androgen receptor gene is associated with male pattern baldness. J. Invest. Dermatol. *116*, 452–455.
27. Hillmer, A.M., Brockschmidt, F.F., Hanneken, S., Eigelshoven, S., Steffens, M., Flaquer, A., Herms, S., Becker, T., Kortüm, A.-K., Nyholt, D.R., et al. (2008). Susceptibility variants for male-pattern baldness on chromosome 20p11. Nat. Genet. *40*, 1279–1281.
28. Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N., and Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. Nature *507*, 354–357. https://doi.org/10.1038/nature12961.
29. Reilly, P.F., Tjahjadi, A., Miller, S.L., Akey, J.M., and Tucci, S. (2022). The contribution of Neanderthal introgression to modern human traits. Curr. Biol. *32*, R970–R983. https://doi.org/10.1016/j.cub.2022.08.027.
30. Harris, D.N., Platt, A., Hansen, M.E., Fan, S., McQuillan, M.A., Nyambo, T., Mpoloka, S.W., Mokone, G.G., Belay, G., and Fokunang, C. (2023). Diverse African genomes reveal selection on ancient modern human introgressions in Neanderthals. Curr. Biol. *33*, 4905–4916.e5.
31. Li, R., Brockschmidt, F.F., Kiefer, A.K., Stefansson, H., Nyholt, D.R., Song, K., Vermeulen, S.H., Kanoni, S., Glass, D., Medland, S.E., et al. (2012). Six novel susceptibility Loci for early-onset androgenetic alopecia and their unexpected association with common diseases. PLoS Genet. *8*, e1002746. https://doi.org/10.1371/journal.pgen.1002746.
32. Heilmann-Heimbach, S., Herold, C., Hochfeld, L.M., Hillmer, A.M., Nyholt, D.R., Hecker, J., Javed, A., Chew, E.G.Y., Pechlivanis, S., Drichel, D., et al. (2017). Meta-analysis identifies novel risk loci and yields systematic insights into the biology of male-pattern baldness. Nat. Commun. *8*, 14694. https://doi.org/10.1038/ncomms14694.
33. Pirastu, N., Joshi, P.K., de Vries, P.S., Cornelis, M.C., McKeigue, P.M., Keum, N., Franceschini, N., Colombo, M., Giovannucci, E.L., Spiliopoulou, A., et al. (2017). GWAS for male-pattern baldness identifies 71 susceptibility loci explaining 38% of the risk. Nat. Commun. *8*, 1584. https://doi.org/10.1038/s41467-017-01490-8.
34. Hagenaars, S.P., Hill, W.D., Harris, S.E., Ritchie, S.J., Davies, G., Liewald, D.C., Gale, C.R., Porteous, D.J., Deary, I.J., and Marioni, R.E. (2017). Genetic prediction of male pattern baldness. PLoS Genet. *13*, e1006594. https://doi.org/10.1371/journal.pgen.1006594.
35. Chen, Y., Hysi, P., Maj, C., Heilmann-Heimbach, S., Spector, T.D., Liu, F., and Kayser, M. (2023). Genetic prediction of male pattern baldness based on large independent datasets. Eur. J. Hum. Genet. *31*, 321–328. https://doi.org/10.1038/s41431-022-01201-y.
36. Henne, S.K., Aldisi, R., Sivalingam, S., Hochfeld, L.M., Borisov, O., Krawitz, P.M., Maj, C., Nöthen, M.M., and Heilmann-Heimbach, S. (2023). Analysis of 72,469 UK Biobank exomes links rare variants to male-pattern hair loss. Nat. Commun. *14*, 5492.
37. Andrews, C., Fortier, B., Hayward, A., Lederman, R., Petersen, L., McBride, J., Petersen, D.C., Ajayi, O., Kachambwa, P., Seutloali, M., et al. (2018). Development, Evaluation, and Implementation of a Pan-African Cancer Research Network: Men

of African Descent and Carcinoma of the Prostate. J. Glob. Oncol. *4*, 1–14. https://doi.org/10.1200/JGO.18.00063.

38. Agalliu, I., Lin, W.-K.J., Zhang, J.S., Jacobson, J.S., Rohan, T.E., Adusei, B., Snyper, N.Y.F., Andrews, C., Sidahmed, E., Mensah, J.E., et al. (2022). Overall and central obesity and prostate cancer risk in African men. Cancer Causes Control *33*, 223–239.
39. Harlemon, M., Ajayi, O., Kachambwa, P., Kim, M.S., Simonti, C.N., Quiver, M.H., Petersen, D.C., Mittal, A., Fernandez, P.W., Hsing, A.W., et al. (2020). A Custom Genotyping Array Reveals Population-Level Heterogeneity for the Genetic Risks of Prostate Cancer and Other Cancers in Africa. Cancer Res. *80*, 2956–2966. https://doi.org/10.1158/0008-5472.CAN-19-2165.
40. Janivara, R., Chen, W.C., Hazra, U., Baichoo, S., Agalliu, I., Kachambwa, P., Simonti, C.N., Brown, L.M., Tambe, S.P., Kim, M.S., et al. (2024). Heterogeneous genetic architectures of prostate cancer susceptibility in sub-Saharan Africa. Nat. Genet. *56*, 2093–2103. https://doi.org/10.1038/s41588-024-01931-3.
41. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. Nat. Genet. *48*, 1284–1287. https://doi.org/10.1038/ng.3656.
42. Sengupta, D., Botha, G., Meintjes, A., Mbiyavanga, M., AWI-Gen Study; and H3Africa Consortium, Hazelhurst, S., Mulder, N., Ramsay, M., and Choudhury, A. (2023). Performance and accuracy evaluation of reference panels for genotype imputation in sub-Saharan African populations. Cell Genom. *3*, 100332. https://doi.org/10.1016/j.xgen.2023.100332.
43. de Carvalho, V.I., and Rodriguez-Alvarez, M.X. (2018). Bayesian nonparametric inference for the covariate-adjusted ROC curve. Preprint at arXiv. https://doi.org/10.48550/arXiv.1806.00473.
44. Torkamani, A., Wineinger, N.E., and Topol, E.J. (2018). The personal and clinical utility of polygenic risk scores. Nat. Rev. Genet. *19*, 581–590. https://doi.org/10.1038/s41576-018-0018-x.
45. Bi, W., Zhou, W., Dey, R., Mukherjee, B., Sampson, J.N., and Lee, S. (2021). Efficient mixed model approach for large-scale genome-wide association studies of ordinal categorical phenotypes. Am. J. Hum. Genet. *108*, 825–839.
46. Machiela, M.J., and Chanock, S.J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics *31*, 3555–3557. https://doi.org/10.1093/bioinformatics/btv402.
47. Lin, S.H., Thakur, R., and Machiela, M.J. (2021). LDexpress: an online tool for integrating population-specific linkage disequilibrium patterns with tissue-specific expression data. BMC Bioinf. *22*, 608. https://doi.org/10.1186/s12859-021-04531-8.
48. Tesi, N., Van Der Lee, S., Hulsman, M., Holstege, H., and Reinders, M.J.T. (2021). snpXplorer: a web application to explore human SNP-associations and annotate SNP-sets. Nucleic Acids Res. *49*, W603–W612.
49. Higgins, J.P.T., and Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. Stat. Med. *21*, 1539–1558. https://doi.org/10.1002/sim.1186.
50. Speed, D., Hemani, G., Johnson, M.R., and Balding, D.J. (2012). Improved heritability estimation from genome-wide SNPs. Am. J. Hum. Genet. *91*, 1011–1021. https://doi.org/10.1016/j.ajhg.2012.10.010.
51. Srivastava, A.K., Williams, S.M., and Zhang, G. (2023). Heritability Estimation Approaches Utilizing Genome-Wide Data. Curr. Protoc. *3*, e734. https://doi.org/10.1002/cpz1.734.
52. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.
53. Park, J.H., Wacholder, S., Gail, M.H., Peters, U., Jacobs, K.B., Chanock, S.J., and Chatterjee, N. (2010). Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat. Genet. *42*, 570–575. https://doi.org/10.1038/ng.610.
54. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74. https://doi.org/10.1038/nature15393.
55. Skov, L., Coll Macià, M., Sveinbjörnsson, G., Mafessoni, F., Lucotte, E.A., Einarsdóttir, M.S., Jonsson, H., Halldorsson, B., Gudbjartsson, D.F., Helgason, A., et al. (2020). The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes. Nature *582*, 78–83. https://doi.org/10.1038/s41586-020-2225-9.
56. Wilson, D.J. (2019). The harmonic mean p-value for combining dependent tests. Proc. Natl. Acad. Sci. USA *116*, 1195–1200. https://doi.org/10.1073/pnas.1814092116.
57. Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.P., and Wang, L. (2014). CrossMap: a versatile tool for coordinate conversion between genome assemblies. Bioinformatics *30*, 1006–1007. https://doi.org/10.1093/bioinformatics/btt730.
58. Johnson, K.E., and Voight, B.F. (2018). Patterns of shared signatures of recent positive selection across human populations. Nat. Ecol. Evol. *2*, 713–720. https://doi.org/10.1038/s41559-018-0478-6.
59. Hazra, U., and Lachance, J. (2023). Differences in disease burdens across human populations are governed more by neutral evolution than by natural selection. Preprint at medRxiv. https://doi.org/10.1101/2021.12.10.21267630.
60. Wright, S. (1943). Isolation by Distance. Genetics *28*, 114–138. https://doi.org/10.1093/genetics/28.2.114.
61. Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T., Güneş, O., Hall, P., Hayhurst, J., et al. (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. *51*, D977–D985. https://doi.org/10.1093/nar/gkac1010.
62. Nagao, K., Kobayashi, T., Moro, K., Ohyama, M., Adachi, T., Kitashima, D.Y., Ueha, S., Horiuchi, K., Tanizaki, H., Kabashima, K., et al. (2012). Stress-induced production of chemokines by hair follicles regulates the trafficking of dendritic cells in skin. Nat. Immunol. *13*, 744–752. https://doi.org/10.1038/ni.2353.
63. Wainschtein, P., Jain, D., Zheng, Z., TOPMed Anthropometry Working Group; and NHLBI Trans-Omics for Precision Medicine TOPMed Consortium, Cupples, L.A., Shadyab, A.H., McKnight, B., Shoemaker, B.M., Mitchell, B.D., et al. (2022). Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. Nat. Genet. *54*, 263–273. https://doi.org/10.1038/s41588-021-00997-7.
64. Kim, M.S., Naidoo, D., Hazra, U., Quiver, M.H., Chen, W.C., Simonti, C.N., Kachambwa, P., Harlemon, M., Agalliu, I., Baichoo, S., et al. (2022). Testing the generalizability of ancestry-specific polygenic risk scores to predict prostate cancer in sub-Saharan Africa. Genome Biol. *23*, 194. https://doi.org/10.1186/s13059-022-02766-z.

65. Dannemann, M., and Kelso, J. (2017). The Contribution of Neanderthals to Phenotypic Variation in Modern Humans. Am. J. Hum. Genet. *101*, 578–589. https://doi.org/10.1016/j.ajhg.2017.09.010.
66. Wei, X., Robles, C.R., Pazokitoroudi, A., Ganna, A., Gusev, A., Durvasula, A., Gazal, S., Loh, P.R., Reich, D., and Sankararaman, S. (2023). The lingering effects of Neanderthal introgression on human complex traits. Elife *12*, e80757. https://doi.org/10.7554/eLife.80757.
67. Yair, S., and Coop, G. (2022). Population differentiation of polygenic score predictions under stabilizing selection. Philos. Trans. R. Soc. Lond. B Biol. Sci. *377*, 20200416. https://doi.org/10.1098/rstb.2020.0416.
68. Quiver, M.H., and Lachance, J. (2022). Adaptive eQTLs reveal the evolutionary impacts of pleiotropy and tissue-specificity while contributing to health and disease. HGG Adv. *3*, 100083. https://doi.org/10.1016/j.xhgg.2021.100083.
69. Hillmer, A.M., Freudenberg, J., Myles, S., Herms, S., Tang, K., Hughes, D.A., Brockschmidt, F.F., Ruan, Y., Stoneking, M., and Nöthen, M.M. (2009). Recent positive selection of a human androgen receptor/ectodysplasin A2 receptor haplotype and its relationship to male pattern baldness. Hum. Genet. *126*, 255–264. https://doi.org/10.1007/s00439-009-0668-z.
70. Casto, A.M., Henn, B.M., Kidd, J.M., Bustamante, C.D., and Feldman, M.W. (2012). A tale of two haplotypes: the EDA2R/AR Intergenic region is the most divergent genomic segment between Africans and East Asians in the human genome. Hum. Biol. *84*, 641–694.
71. Kamberov, Y.G., Wang, S., Tan, J., Gerbault, P., Wark, A., Tan, L., Yang, Y., Li, S., Tang, K., Chen, H., et al. (2013). Modeling recent human evolution in mice by expression of a selected EDAR variant. Cell *152*, 691–702. https://doi.org/10.1016/j.cell.2013.01.016.
72. Davey, R.A., and Grossmann, M. (2016). Androgen Receptor Structure, Function and Biology: From Bench to Bedside. Clin. Biochem. Rev. *37*, 3–15.
73. Racimo, F., Berg, J.J., and Pickrell, J.K. (2018). Detecting Polygenic Adaptation in Admixture Graphs. Genetics *208*, 1565–1584. https://doi.org/10.1534/genetics.117.300489.
74. Johnson, R., Burch, K.S., Hou, K., Paciuc, M., Pasaniuc, B., and Sankararaman, S. (2021). Estimation of regional polygenicity from GWAS provides insights into the genetic architecture of complex traits. PLoS Comput. Biol. *17*, e1009483.
75. Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T.J.C., van der Sluis, S., Andreassen, O.A., Neale, B.M., and Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. Nat. Genet. *51*, 1339–1348.
76. Smith, G.D., Richmond, R.C., and Pingault, J.-B. (2022). Combining Human Genetics and Causal Inference to Understand Human Disease and Development (Cold Spring Harbor Laboratory Press).
77. Swerdlow, D.I., Kuchenbaecker, K.B., Shah, S., Sofat, R., Holmes, M.V., White, J., Mindell, J.S., Kivimaki, M., Brunner, E.J., Whittaker, J.C., et al. (2016). Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. Int. J. Epidemiol. *45*, 1600–1616. https://doi.org/10.1093/ije/dyw088.
78. Gupta, M., and Mysore, V. (2016). Classifications of Patterned Hair Loss: A Review. J. Cutan. Aesthetic Surg. *9*, 3–12. https://doi.org/10.4103/0974-2077.178536.
79. Gheybi, K., Mmekwa, N., Lebelo, M.T., Patrick, S.M., Campbell, R., Nenzhelele, M., Soh, P.X.Y., Obida, M., Loda, M., Shirindi, J., et al. (2023). Linking African ancestral substructure to prostate cancer health disparities. Sci. Rep. *13*, 20909. https://doi.org/10.1038/s41598-023-47993-x.
80. Pfennig, A., Petersen, L.N., Kachambwa, P., and Lachance, J. (2023). Evolutionary genetics and admixture in African populations. Genome Biol. Evol. *15*, evad054.
81. Bang, H.J., Yang, Y.J., Lho, D.S., Lee, W.Y., Sim, W.Y., and Chung, B.C. (2004). Comparative studies on level of androgens in hair and plasma with premature male-pattern baldness. J. Dermatol. Sci. *34*, 11–16. https://doi.org/10.1016/j.jdermsci.2003.11.005.