nature genetics

SUPPLEMENTARY INFORMATION

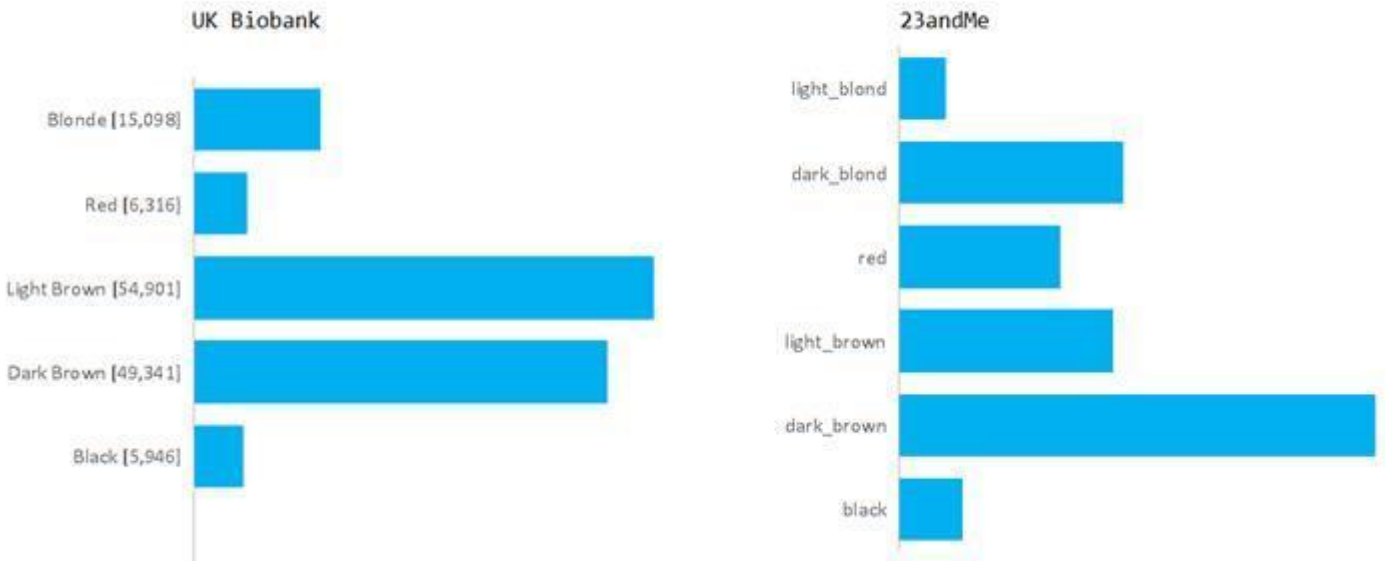
https://doi.org/10.1038/s41588-018-0100-5

In the format provided by the authors and unedited.

Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability

Pirro G. Hysi[1,2,24], Ana M. Valdes[1,3,4,24], Fan Liu[5,6,7,24], Nicholas A. Furlotte[8], David M. Evans[9,10], Veronique Bataille[1], Alessia Visconti[1], Gibran Hemani[10], George McMahon[10], Susan M. Ring[10], George Davey Smith[10], David L. Duffy[11], Gu Zhu[11], Scott D. Gordon[11], Sarah E. Medland[11], Bochao D. Lin[12], Gonneke Willemsen[12], Jouke Jan Hottenga[12], Dragana Vuckovic[13], Giorgia Girotto[13,14], Ilaria Gandin[13], Cinzia Sala[13], Maria Pina Concas[14], Marco Brumat[13], Paolo Gasparini[13,14], Daniela Toniolo[15], Massimiliano Cocca[14], Antonietta Robino[14], Seyhan Yazar[16,17], Alex W. Hewitt[16,18,19], Yan Chen[5,6], Changqing Zeng[5], Andre G. Uitterlinden[20,21], M. Arfan Ikram[21], Merel A. Hamer[22], Cornelia M. van Duijn[21], Tamar Nijsten[22], David A. Mackey[16,18,19], Mario Falchi[1], Dorret I. Boomsma[12], Nicholas G. Martin[11], The International Visible Trait Genetics Consortium[23], David A. Hinds[8], Manfred Kayser[7,25*] and Timothy D. Spector[1,25*]

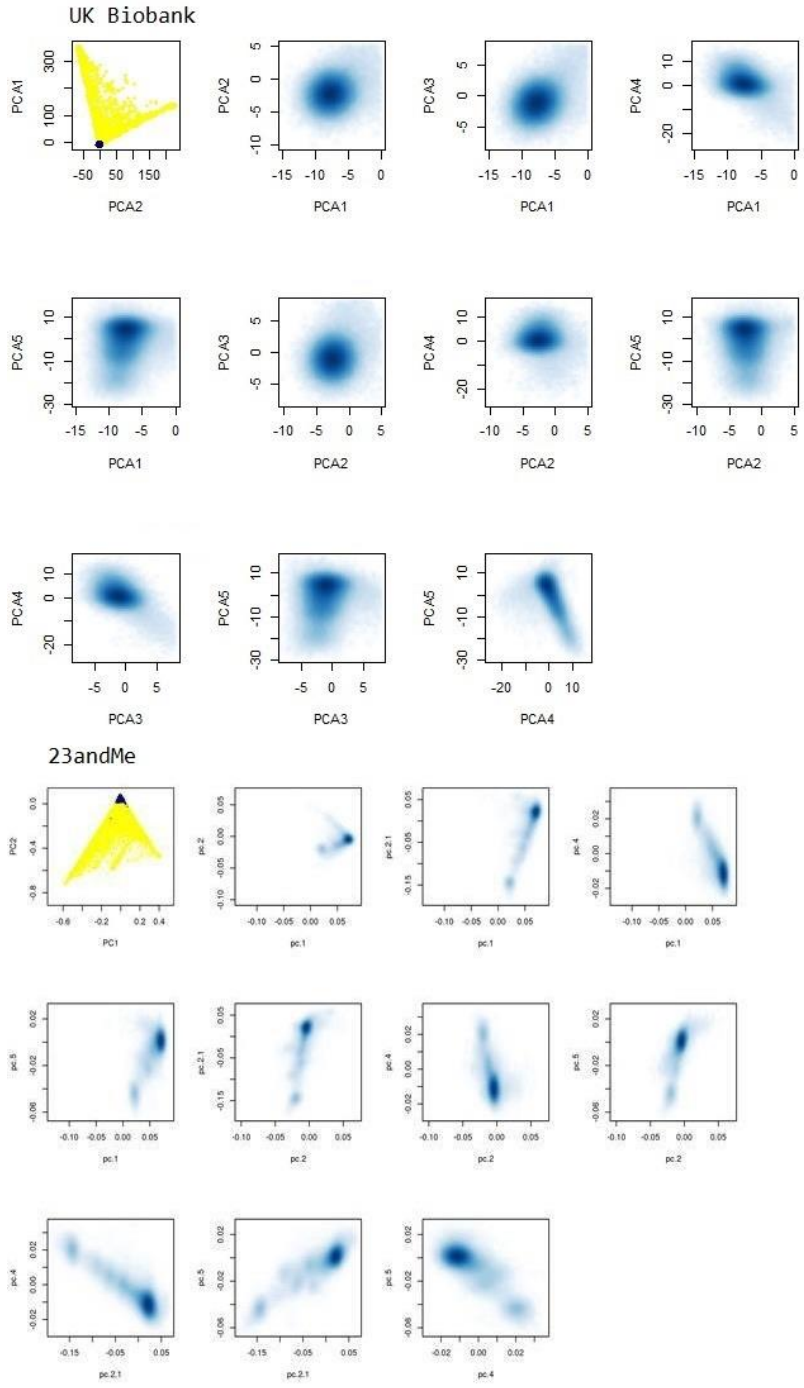
[1]King's College London Department of Twins Research and Genetic Epidemiology, London, UK. [2]Department of Ophthalmology, King's College London, London, UK. [3]Division of Rheumatology, Orthopaedics and Dermatology, School of Medicine, University of Nottingham, Nottingham, UK. [4]Nottingham NIHR Biomedical Research Centre, Nottingham, United Kingdom. [5]CAS Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, PR China, [6]University of Chinese Academy of Sciences, Beijing, PR China. [7]Department of Genetic Identification, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands. [8]23andMe, Inc., Mountain View, CA, USA. [9]University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, QLD, Australia. [10]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [11]QIMR Berghofer Medical Research Institute, Brisbane, Australia. [12]Netherlands Twin Register, Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands. [13]Department of Medicine, Surgery and Health Sciences, University of Trieste, Trieste, Italy. [14]Institute for Maternal and Child Health IRCCS "Burlo Garofolo", Trieste, Italy. [15]Division of Genetics and Cell Biology, San Faffaele Research Institute, Milano, Italy. [16]Centre for Ophthalmology and Visual Science, University of Western Australia, Lions Eye Institute, Perth, WA, Australia. [17]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, UK. [18]Centre for Eye Research Australia, University of Melbourne, Department of Ophthalmology, Royal Victorian Eye and Ear Hospital, Melbourne, Australia. [19]School of Medicine, Menzies Research Institute Tasmania, University of Tasmania, Hobart, Australia. [20]Department of Internal Medicine, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands. [21]Department of Epidemiology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands. [22]Department of Dermatology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands. [23]A list of members and affiliations appears in the Supplementary Note. [24]These authors contributed equally: Pirro G. Hysi, Ana Valdes, Fan Liu. [25]These authors jointly supervised this work: Manfred Kayser, Timothy D. Spector. *e-mail: m.kayser@erasmusmc.nl; tim.spector@kcl.ac.uk



Supplementary Figure 1

Distribution of self-reported hair color in the UK Biobank and 23andMe cohorts.

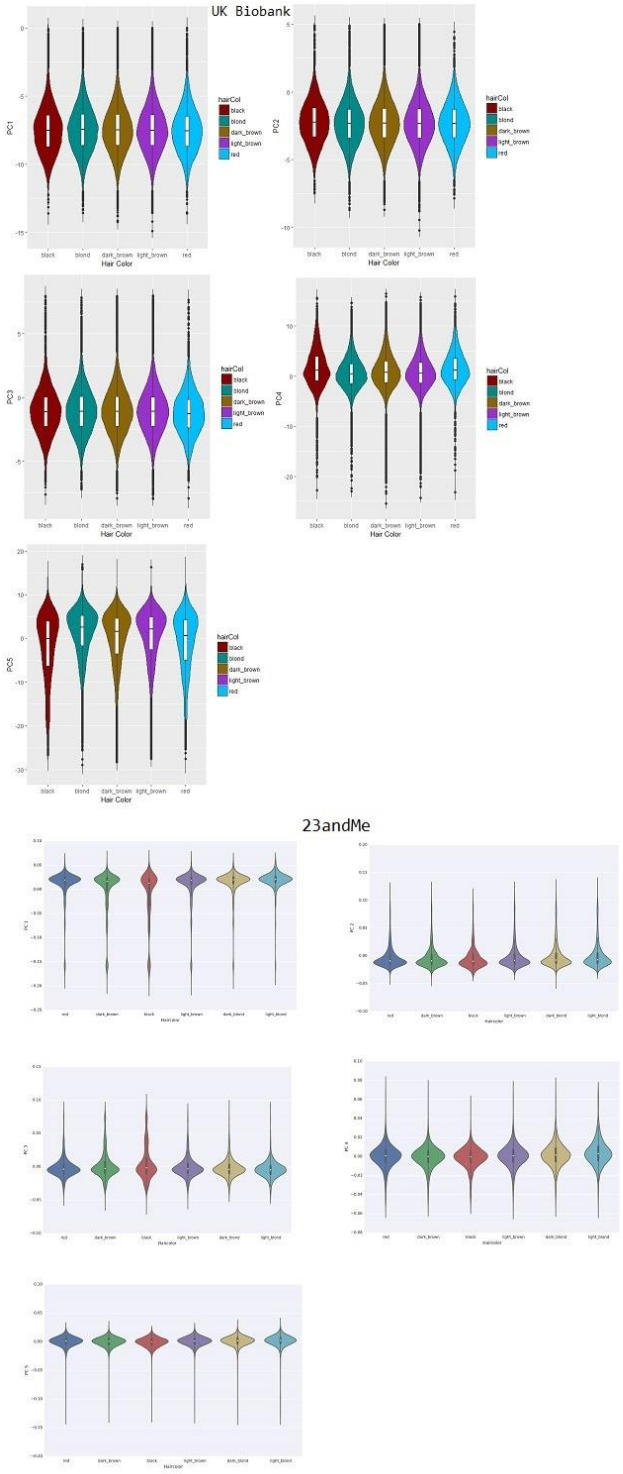
Plots show the densities for each hair color category in the UK Biobank (top) and 23andMe cohorts (bottom).



Supplementary Figure 2

Plot of the first five principal components in UK Biobank and 23andMe subjects included in the analyses.

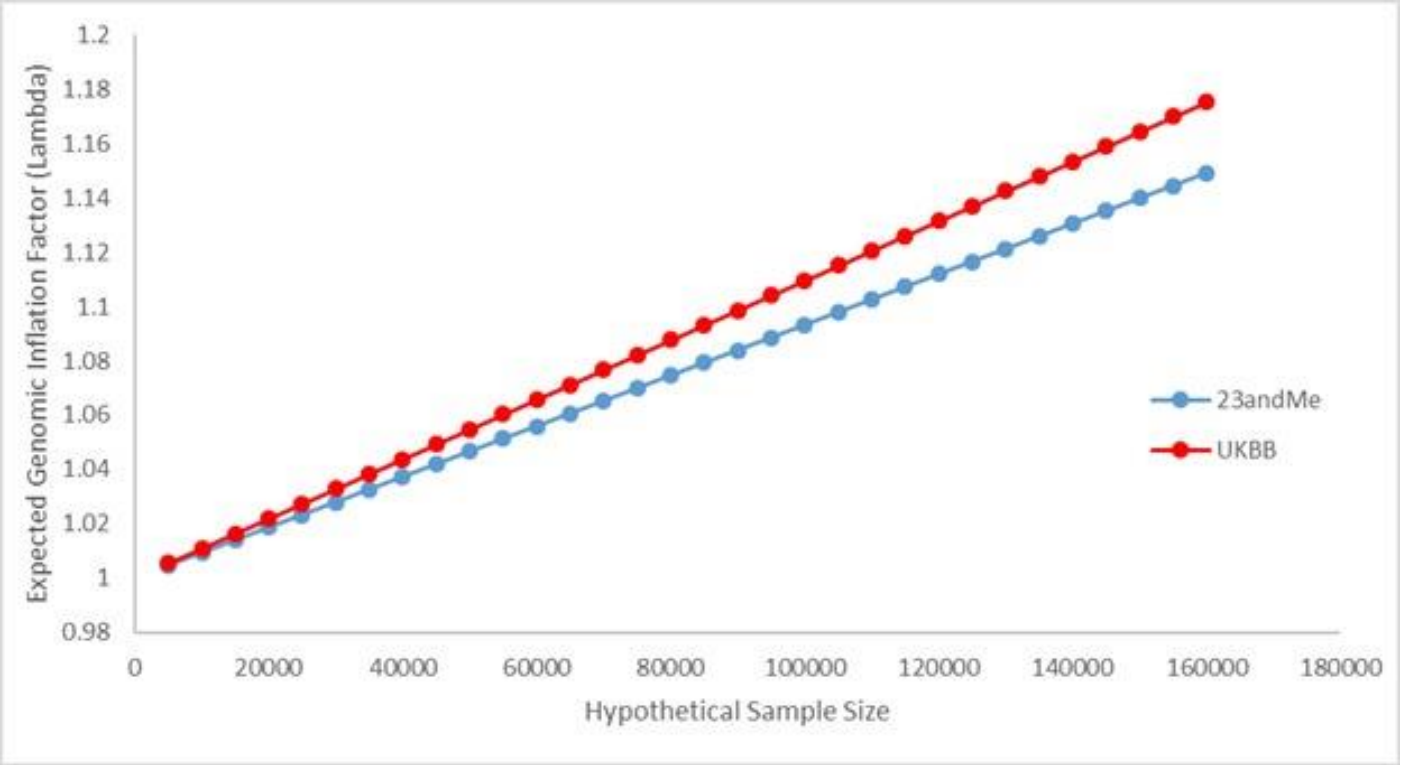
The top-left panel shows the participants of European descent that were included in the analyses (blue) in the backdrop of all multi-ethnic UK Biobank and 23andMe participants. The plots for any two of the first principal components for the subjects of European descent (included in the analyses) are also given.



Supplementary Figure 3

Violin plots of the first principal components for the UK Biobank and 23andMe participants.

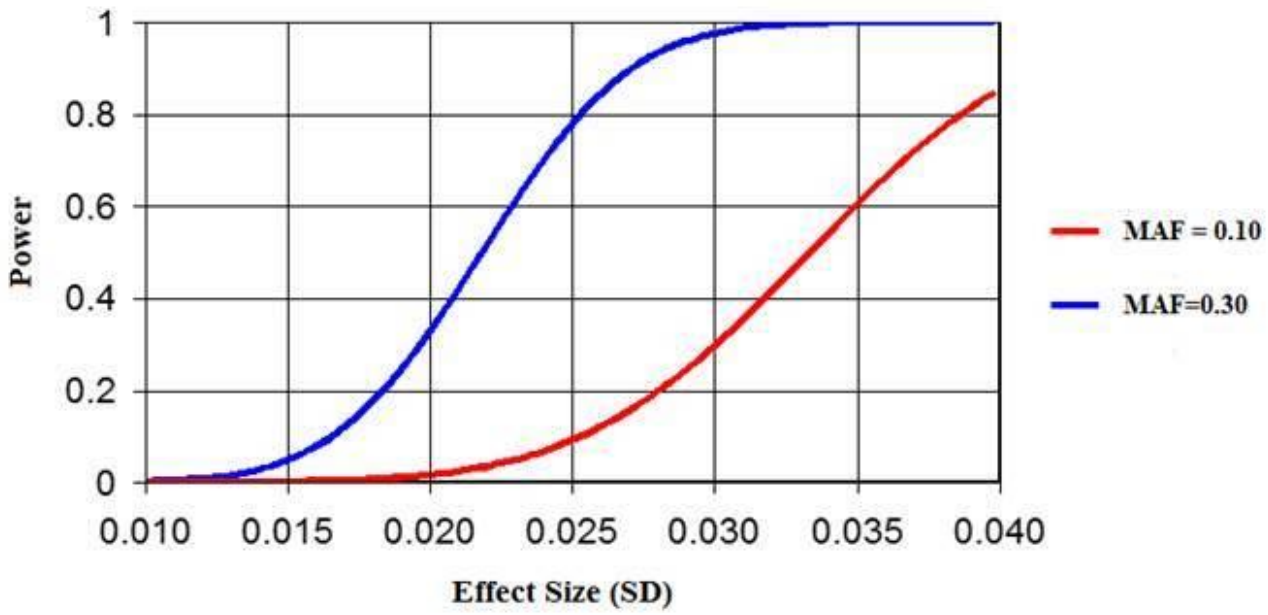
Each violin plot shows the distribution of principal components computed in the UK Biobank subjects (top) and the 23andMe cohort (bottom).



Supplementary Figure 4

Expected variation of genomic control inflation factor (λ) as a function of sample size.

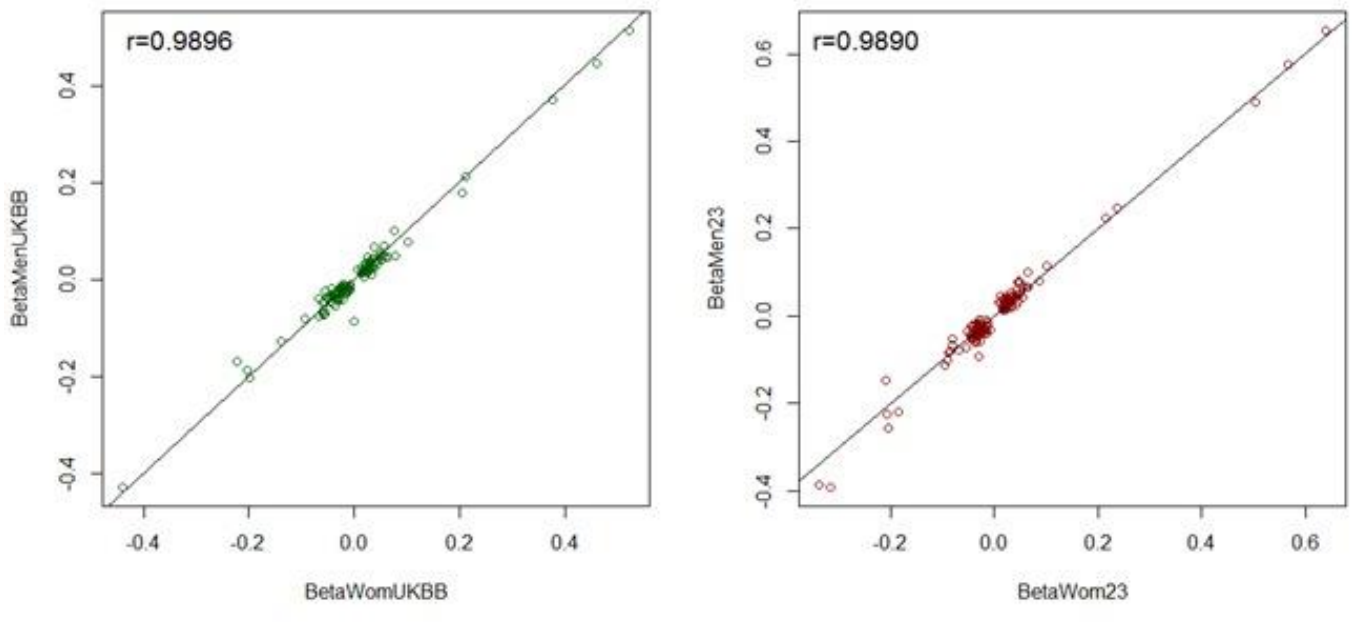
The expected λ were calculated using the same number of loci, LD structure and trait heritability observed in the 23andMe cohort (blue) and the UK Biobank cohort (red) in absence of factors artificially inflating association. These genomic inflation factors observed in the discovery cohort are a reflection of the power of these large sample analyzed in presence of polygenicity; they would be merely the equivalent of a λ_{GC} = 1.0186 for the 23andMe cohort and λ_{GC} = 1.0221 for the UKBB cohort had the effective sample sizes been smaller (*n* = 20,000) and the underlying genetic effects remained the same.



Supplementary Figure 5

Power to detect genome-wide significant associations.

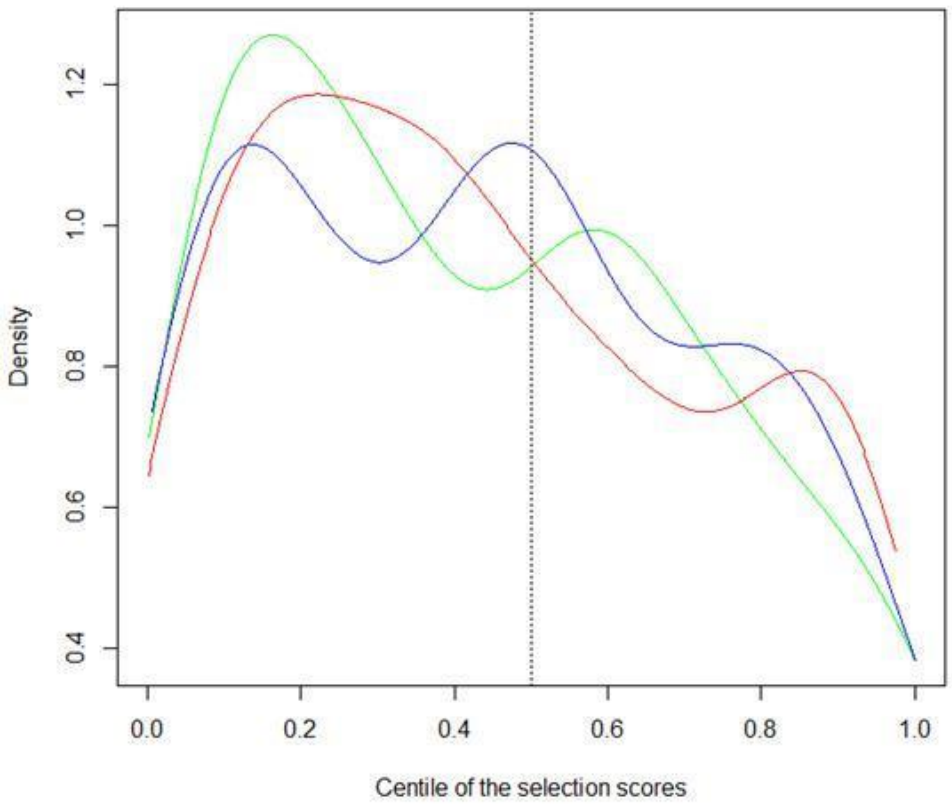
Plot generated using a sample of n = 140,000 subjects at $\alpha = 10^{-7}$. Power is shown for the most representative range of allele frequencies detected in GWAS (MAF = 0.10, red line; MAF = 0.30, blue line). The value on the *y*-axis denotes the probability that a true association would be detected at the pre-defined α; for example, for a locus with a MAF = 0.10 with an effect size of 0.035 s.d. per each copy of the risk allele over the trait, there is a 60% probability to replicate the same association at a genome-wide level of significance.



Supplementary Figure 6

Correlation of effect sizes between men and women participating in the UK Biobank and the 23andMe cohorts.

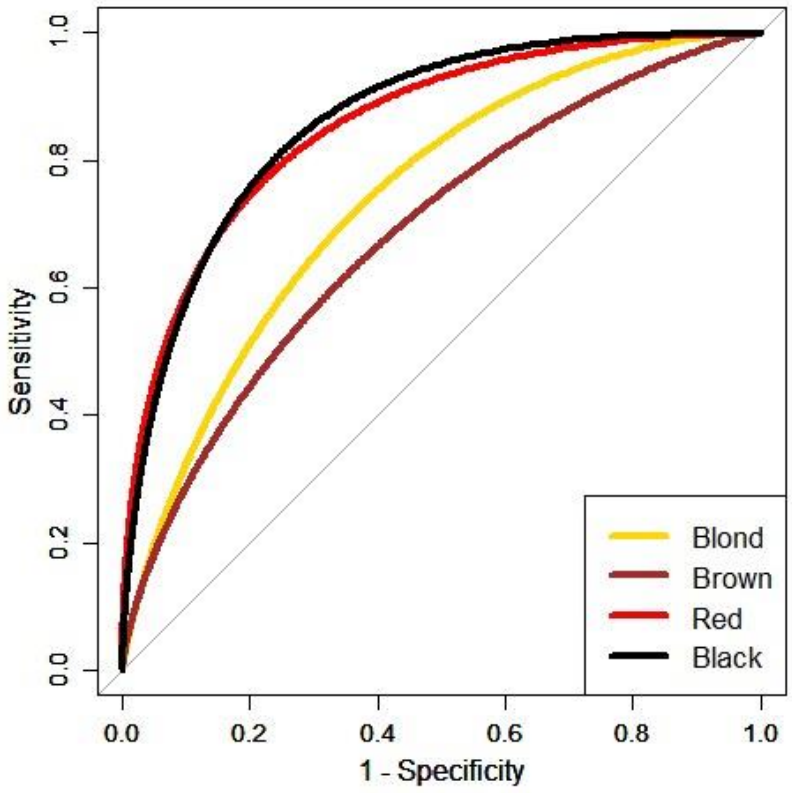
The black line denotes parity of effects, and each point represents one of the GWAS associated SNPs shown in **Supplementary Table 2**.



Supplementary Figure 7

Evidence of natural selection for the SNPs significantly associated with hair color.

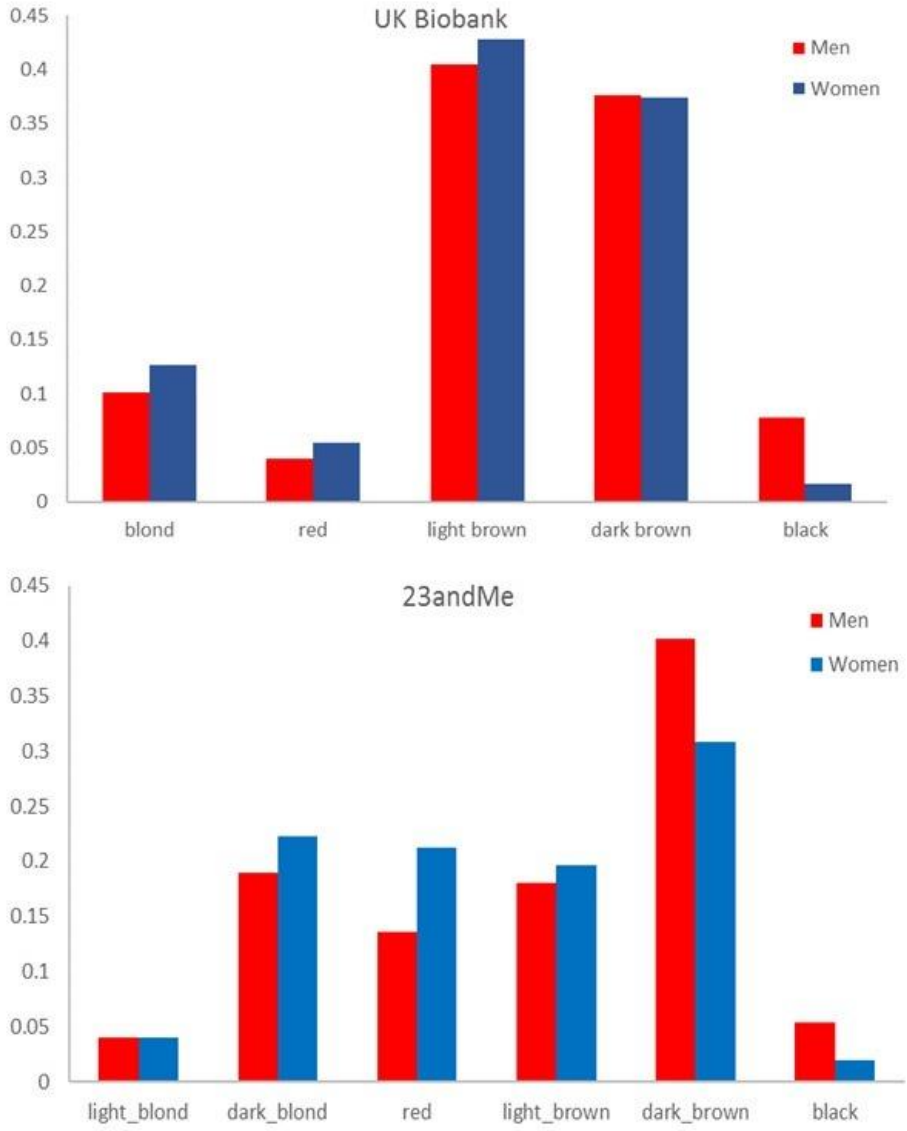
Three different natural selection tests were selected (see Online Methods), and the values plotted here represent the centile rank of the natural selection score within European populations (iHS test, red), compared to YRI Africans (XP-EHH CEU vs. YRI, green) and to CHB Chinese populations (blue) using 1000 Genomes data. The plot shows an enrichment for higher ranks (lower centiles) of selection for some SNPs within European populations (two tailed Wilcoxon test $P = 0.04$), evidence for stronger selection for these SNPs in Europeans compared to Africans (two-tailed Wilcoxon $P = 0.014$), but less significant compared to Chinese (two tailed Wilcoxon $P = 0.056$).



Supplementary Figure 8

Hair color prediction in participants of the Rotterdam and QIMR studies.

The prediction model is based on multinomial logistic regression including 20 HIrisPlex SNPs, a polygenic score for blond-brown-black (233 SNPs), and a polygenic score for red (25 SNPs); see **Supplementary Note** for marker and model specifications.



Supplementary Figure 9

Sex-specific prevalence of hair color categories in the UK Biobank and 23andMe cohorts.

In both cohorts, there is a higher prevalence of blond, red and light brown hair colors among women and a higher prevalence of black and dark brown hair colors among men.