GENETICS

Genome-wide association study in almost 195,000 individuals identifies 50 previously unidentified genetic loci for eye color

Mark Simcoe[1,2], Ana Valdes[1,3], Fan Liu[4,5,6], Nicholas A. Furlotte[7], David M. Evans[8,9], Gibran Hemani[9,10], Susan M. Ring[9,10], George Davey Smith[9,10], David L. Duffy[11], Gu Zhu[11], Scott D. Gordon[11], Sarah E. Medland[11], Dragana Vuckovic[12,13,14], Giorgia Girotto[12,13], Cinzia Sala[15], Eulalia Catamo[12], Maria Pina Concas[13], Marco Brumat[12], Paolo Gasparini[12,13], Daniela Toniolo[15], Massimiliano Cocca[13], Antonietta Robino[13], Seyhan Yazar[16], Alex Hewitt[16,17,18], Wenting Wu[19], Peter Kraft[20], Christopher J. Hammond[1,2], Yuan Shi[21], Yan Chen[4,5,6], Changqing Zeng[5], Caroline C. W. Klaver[22,23,24], Andre G. Uitterlinden[23,25], M. Arfan Ikram[23], Merel A. Hamer[26], Cornelia M. van Duijn[23,27], Tamar Nijsten[26], Jiali Han[19], David A. Mackey[16], Nicholas G. Martin[11], Ching-Yu Cheng[21,28]; the 23andMe Research Team; the International Visible Trait Genetics Consortium, David A. Hinds[7], Timothy D. Spector[1]*, Manfred Kayser[4]*†, Pirro G. Hysi[1,2]*†

Human eye color is highly heritable, but its genetic architecture is not yet fully understood. We report the results of the largest genome-wide association study for eye color to date, involving up to 192,986 European participants from 10 populations. We identify 124 independent associations arising from 61 discrete genomic regions, including 50 previously unidentified. We find evidence for genes involved in melanin pigmentation, but we also find associations with genes involved in iris morphology and structure. Further analyses in 1636 Asian participants from two populations suggest that iris pigmentation variation in Asians is genetically similar to Europeans, albeit with smaller effect sizes. Our findings collectively explain 53.2% (95% confidence interval, 45.4 to 61.0%) of eye color variation using common single-nucleotide polymorphisms. Overall, our study outcomes demonstrate that the genetic complexity of human eye color considerably exceeds previous knowledge and expectations, highlighting eye color as a genetically highly complex human trait.

INTRODUCTION

Eye color is primarily determined by melanin abundance within the iris pigment epithelium, which is greater in brown than in blue eyes (*1*), and both the density and distribution of stromal melanocyte cells (*2*). Ratios of the two forms of melanin, eumelanin and pheomelanin, within the iris as well as light absorption and scattering by extracellular components (Tyndall scattering) are additional factors that give irises their color (*3*). Absolute melanin quantity and the eumelanin:pheomelanin ratio are higher in brown irises (*4*), while blue or green irises have very little of both pigments and relatively more pheomelanin.

European populations, or those with partial European origin, display the largest diversity of iris color, varying from the lightest blue to darkest brown. The prevalence of blue eyes correlates with geographic latitude across Europe and neighboring areas (*5*), likely as a result of human migration, sexual, and possibly natural selection (*6*–*8*). Similarly, eye color variation with varying degrees of brown irises is seen in Asian populations (*9*), although with a much reduced range compared to brown eye color variation in Europeans.

Iris color is highly heritable (*10*). Previous genome-wide association studies (GWASs) have identified various single-nucleotide polymorphisms (SNPs) in and around 10 genes significantly associated

with eye color (*11–14*), highlighting the polygenic nature of a trait that, in the past, was assumed to be genetically simple (*15*). The strongest genetic influence on eye color is exerted by the neighboring *HERC2* and *OCA2* genes (*13*, *14*, *16*, *17*), where a long distance enhancer effect of an intronic SNP in *HERC2* was demonstrated interacting with the *OCA2* promoter functioning as a molecular switch between light and dark pigmentation (*18*). Previously available genetic knowledge allows accurate prediction of blue and brown eye color, for instance, with a DNA test system based on six SNPs from six genes, including *HERC2* and *OCA2* (*19*), that has been used in anthropological (*20*) and forensic (*21*, *22*) applications. However, nonblue and nonbrown eye color can be genetically predicted with considerably lower accuracy (*19*, *21*, *22*), likely because of unknown predictive SNPs and responsible genes. Moreover, the phenotypic variance of eye color not previously explained by GWASs ranged between 26% for a blue versus a brown scale (*10*) and 50% for a scale using three categories (*11*) across studies. This illustrates the likely presence of yet unknown genes responsible for the noted missing genetic predicting accuracy (*19*, *21*, *22*) and missing heritability of human eye color (*23*).

To overcome these limitations and better understand the genetics of human eye color, we carried out the largest eye color GWAS to date. Our study involved 157,485 individuals of European ancestry in the discovery stage and an additional 35,501 ancestral European individuals in the replication stage, as well as 1636 Asians (of Han Chinese and Indian ancestry) additionally used for replication purposes, in total 194,622 individuals of different ancestries.

RESULTS

Linear regression results were obtained for 11,532,091 SNPs from 157,485 individuals of European ancestry in the discovery dataset from the personal genetics company 23andMe Inc. The Devlin's genomic factor (*24*) was λ = 1.13, while a linkage disequilibrium (LD) score regression intercept of 1.095 with an [(intercept − 1)/(mean(χ^2) − 1)] ratio of 0.19 was obtained, which is consistent with expectations based on the large sample size and polygenic architecture (*25*). In total, 12,192 SNPs were associated at genome-wide significance ($P < 5 \times 10^{-8}$) with eye color. These SNPs clustered in 52 distinct genomic regions, 50 across the autosomes and 2 on chromosome X (Table 1 and Fig. 1). While confirming the 10 genomic regions previously associated with human eye color (*11*, *26*, *27*), the remaining 42 genomic regions identified represent novel discoveries.

As expected, many of the strongest associations were observed for SNPs within *HERC2* (rs1129038, $P < 10^{-330}$), the strongest eye color–associated region known previously, and genes involved in five of the seven known types of oculocutaneous albinism (OCA): OCA1 (*TYR*, rs1126809, $P = 1.8 \times 10^{-255}$), OCA2 (*OCA2*, rs1800407, $P < 10^{-330}$), OCA3 (*TYRP1*, rs13297008; $P = 5.0 \times 10^{-250}$), OCA4 (*SLC45A2*, rs16891982; $P = 2.0 \times 10^{-255}$), and OCA6 (*SLC24A5*, rs1426654; $P = 2.7 \times 10^{-26}$). *HERC2*, *TYR*, *OCA2*, *TYRP1*, and *SLC45A2* have previously been associated with eye color (*13*, *14*, *16*, *17*, *19*), while previous studies found association with hair and skin pigmentation for *SLC24A5* (*27*) and eye color in a South Asian population (*28*), but not with eye color in Europeans as we showed here. No significant eye color associations were found for either the OCA5 (the 4q24 region) or the OCA7 (*C10ORF11*) locus, the other two genes involved in OCA, in our large discovery sample; however,

C10ORF11 has recently been associated with human eyebrow color (*29*). The remaining four previously reported loci (*11*) were also strongly associated with eye color in the present study: *LYST* (rs2385028, $P = 2.4 \times 10^{-12}$), *IRF4* (rs12203592, $P = 1.6 \times 10^{-321}$), *SLC24A4* (rs17184180, $P = 1.3 \times 10^{-284}$), *TSPAN10* (previously reported as *NPLOC4*, rs6420484; $P = 7.5 \times 10^{-90}$), and *TTC3* (rs2835660, $P = 1.6 \times 10^{-17}$).

Among the 41 novel genetic loci identified in our discovery analysis, we detected significant eye color association for SNPs within *TPCN2* (rs72928978, $P = 6.42 \times 10^{-32}$) and *MITF* (rs75114713, $P = 8.11 \times 10^{-24}$). These two genetic loci and five others (near the *DTL*, *AP3M2*, *SOX5*, *DCT*, and *SIK1* genes; Table 1) were associated with hair and skin pigmentation in previous GWASs (highlighted with asterisk in Table 1) (*27*, *30–32*), while their eye color association is reported here. A comparison of our results with those from the GWAS Catalog (*33*) revealed that 34 (81%) of the 41 novel eye color–associated loci are unique to eye color and not shared with other pigmentation traits (Table 1), which is 36 (69%) when considering all 52 eye color–associated loci we identified here including the 11 previously known loci. The unique novel eye color loci include SNPs in *TRAF3IP1* (rs74409360, $P = 8.1 \times 10^{-20}$) and *SEMA3A* (rs6944702, $P = 1.7 \times 10^{-10}$). *TRAF3IP1* has been previously associated with iris furrows, while *SEMA3A* was associated with iris crypt variation (*3*). These findings suggest that eye color may, at least in part considered by the eye color phenotyping done here, be mediated through structural effects within the iris. Significant association was also observed for SNPs within *HIVEP3* (rs6696511, $P = 1.3 \times 10^{-45}$), which was previously associated with refractive error and myopia (*34*). Novel eye color association within *PPARGC1A* (rs4521336, $P = 1.3 \times 10^{-9}$) and within *MAP2K6* (rs3809761, $P = 5.1 \times 10^{-9}$) likely arises from their participation in pigmentation pathways, the PGC1-α protein, the product of the *PPARGC1A* gene. PGC1-α activates the *MITF* promoter (*35*) acting as a regulator in the pigment pathway of the tanning response, while overexpression of *MAP2K6* increases melanocyte dendricity (*36*), allowing greater transportation of melanosomes. We report here eye color associations for genetic loci located on the X chromosome. DNA variants clustering around the *SSX1* (rs78542430, $P = 5.93 \times 10^{-53}$) and *TMEM255A* (rs5957354, $P = 5.43 \times 10^{-23}$) genes were significantly associated with eye color. Little is published about the *TMEM255A* and *SSX1* genes and their potential functional roles relating to the eye, and further studies are required to understand the role of these genes in iris pigmentation. Notably, a recently published large GWAS on hair color in Europeans was the first to discover X-chromosomal loci to be involved in human pigmentation (*31*), albeit with another gene (*COL46A*, 12 Mbp upstream from nearest associated locus in our study) than we found to be associated with eye color here. Notably, SNPs in or near the *MC1R* gene, known to be involved in light skin and red hair color, did not show eye color association in our study.

Next, we conducted a conditional analysis to identify SNPs associated with eye color after adjusting for the effect of the main association signal at a respective locus. This analysis highlighted 115 conditionally associated SNPs distributed across the novel and previously known 52 genomic regions (table ST1; a secondary list of 489 high-quality control (QC) scoring, significantly associated SNPs is provided in table ST2) that improve the phenotypic variability explained by genetic factors compared to the lead SNPs. Of the 115 independently associated SNPs, 9 were missense mutations (table ST3) (*37*), which may suggest their causal role.

Table 1. Strongest associated SNPs from the 52 genomic regions independently associated with eye color in the European discovery cohort (*N* = 157,485). A full list of 115 independently associated SNPs from conditional analysis at these loci can be found in table ST1. SNPs with novel associations for eye color identified here are highlighted in bold. SNPs that were associated with other pigmentation traits, but not iris color, in previous studies are indicated by an asterisk (*). Chr is the chromosome for the given SNP, genome position (Pos) refer to the HG build 37, RS is the rsid for the given SNP, Freq is the allele frequency of the stated reference allele, Beta is the effect size for the stated reference allele, SE is the respective standard error for the beta, Alt allele is the alternative allele for the given SNP, and *N* is the sample size tested for the respective SNP.

Region	Chr	Pos	RS	Nearest gene	Ref allele	Alt allele	*N*	Freq	Beta	SE	*P*
1	**1**	**9166344**	**rs6693258**	***GPR157***	**T**	**C**	**157483**	**0.216**	**−0.056**	**0.0091**	**9.86 × 10⁻¹⁰**
2	**1**	**42110888**	**rs6696511**	***HIVEP3***	**T**	**C**	**157483**	**0.382**	**0.109**	**0.0077**	**1.35 × 10−45**
3	1	212421629	rs351385	**DTL*	G	A	157483	0.580	−0.083	0.0075	2.31 × 10−28
4	1	236035805	rs2385028	LYST	T	C	157483	0.245	0.061	0.0086	2.43 × 10⁻¹²
5	**2**	**46233381**	**rs13016869**	***PRKCE***	**G**	**C**	**157483**	**0.208**	**−0.058**	**0.0092**	**2.23 × 10⁻¹⁰**
6	**2**	**206950236**	**rs112747614**	***INO80D***	**G**	**C**	**157483**	**0.696**	**0.054**	**0.0081**	**3.69 × 10⁻¹¹**
7	**2**	**219755011**	**rs121908120**	***WNT10A***	**T**	**A**	**157483**	**0.976**	**0.140**	**0.0247**	**1.54 × 10⁻⁸**
8	**2**	**223483670**	**rs12614022**	***FARSB***	**G**	**A**	**157483**	**0.636**	**0.048**	**0.0078**	**4.26 × 10⁻¹⁰**
9	**2**	**239276278**	**rs74409360**	***TRAF3IP1***	**T**	**C**	**157483**	**0.076**	**−0.128**	**0.0140**	**8.09 × 10⁻²⁰**
10	**3**	**42762488**	**rs3912104**	***CCDC13***	**T**	**A**	**157483**	**0.541**	**0.045**	**0.0075**	**1.47 × 10⁻⁹**
11	3	69980177	rs116359091	**MITF*	G	A	157483	0.974	−0.229	0.0246	1.02 × 10⁻²⁰
12	**4**	**23939399**	**rs4521336**	***PPARGC1A***	**T**	**C**	**157483**	**0.359**	**0.048**	**0.0079**	**1.28 × 10⁻⁹**
13	**4**	**59359559**	**rs141318671**	**No gene**	**G**	**A**	**157483**	**0.993**	**−0.318**	**0.0568**	**2.09 × 10⁻⁸**
14	**4**	**90059434**	**rs6828137**	***TIGD2***	**T**	**G**	**157483**	**0.457**	**−0.041**	**0.0075**	**3.93 × 10⁻⁸**
15	5	311902	rs62330021	*PDCD6, AHRR*	G	A	157483	0.948	−0.325	0.0171	1.36 × 10⁻⁸⁰
16	5	33951693	rs16891982	**SLC45A2*	G	C	157483	0.958	−0.654	0.0191	1.97 × 10⁻²⁵⁵
17	**5**	**40273518**	**rs348613**	***DAB2***	**G**	**A**	**157483**	**0.001**	**0.731**	**0.1227**	**2.54 × 10⁻⁹**
18	**5**	**123896988**	**rs72777200**	***ZNF608***	**T**	**C**	**157483**	**0.828**	**0.074**	**0.0100**	**1.22 × 10⁻¹³**
19	**5**	**148216187**	**rs11957757**	***ADRB2***	**G**	**A**	**157483**	**0.547**	**0.042**	**0.0075**	**2.94 × 10⁻⁸**
20	6	396321	rs12203592	**IRF4*	T	C	157483	0.171	−0.385	0.0100	1.61 × 10⁻³²¹
21	**6**	**10538183**	**rs6910861**	***GCNT2***	**G**	**A**	**157483**	**0.461**	**0.078**	**0.0076**	**1.20 × 10⁻²⁴**
22	**6**	**158841725**	**rs341147**	***TULP4***	**G**	**A**	**157483**	**0.343**	**−0.045**	**0.0078**	**6.45 × 10⁻⁹**
23	**7**	**45960645**	**rs2854746**	***IGFBP3***	**G**	**C**	**157483**	**0.587**	**−0.043**	**0.0076**	**1.24 × 10⁻⁸**
24	**7**	**83653553**	**rs6944702**	***SEMA3A***	**T**	**C**	**157483**	**0.512**	**0.048**	**0.0075**	**1.71 × 10⁻¹⁰**
25	**8**	**12690997**	**rs6997494**	***LONRF1***	**T**	**G**	**157483**	**0.480**	**0.048**	**0.0074**	**1.07 × 10⁻¹⁰**
26	8	42003663	rs12543326	**AP3M2*	G	A	157483	0.875	0.065	0.0113	9.07 × 10⁻⁹
27	**8**	**81350433**	**rs147068120**	***ZBTB10***	**T**	**C**	**157483**	**0.038**	**−0.122**	**0.0198**	**8.34 × 10⁻¹⁰**
28	9	12677471	rs13297008	**TYRP1*	G	A	157483	0.388	0.260	0.0077	4.99 × 10⁻²⁵⁰
29	**9**	**27366436**	**rs12552712**	***MOB3B***	**T**	**C**	**157483**	**0.604**	**−0.056**	**0.0077**	**2.79 × 10⁻¹³**
30	**9**	**132001056**	**rs12335410**	***IER5L***	**T**	**C**	**157483**	**0.221**	**−0.058**	**0.0090**	**1.28 × 10⁻¹⁰**
31	11	68831364	rs72928978	**TPCN2*	G	A	157483	0.877	0.142	0.0121	6.42 × 10⁻³²
32	11	89017961	rs1126809	**TYR*	G	A	157483	0.721	0.285	0.0083	1.82 × 10⁻²⁵⁵
33	12	23979791	rs9971729	**SOX5*	C	A	157483	0.566	0.082	0.0076	5.03 × 10⁻²⁷
34	**12**	**92567833**	**rs790464**	***BTG1***	**T**	**C**	**157483**	**0.189**	**0.060**	**0.0096**	**4.50 × 10⁻¹⁰**
35	**13**	**74178399**	**rs2095645**	***KLF12***	**G**	**A**	**157483**	**0.672**	**0.050**	**0.0079**	**3.26 × 10⁻¹⁰**
36	13	95189401	rs9301973	**DCT*	G	A	157483	0.671	0.066	0.0080	1.48 × 10⁻¹⁶
37	**14**	**69236136**	**rs138777265**	***ZFP36L1***	**C**	**A**	**157483**	**0.047**	**−0.109**	**0.0176**	**6.21 × 10⁻¹⁰**
38	14	92780387	rs17184180	**SLC24A4*	T	A	157483	0.567	0.271	0.0075	1.34 × 10⁻²⁸⁴
39	15	28211758	rs4778218	**OCA2*	G	A	157483	0.834	−0.520	0.0100	<1 × 10⁻³³⁰
40	15	28356859	rs1129038	**HERC2*	T	C	157483	0.734	−2.605	0.0058	<1 × 10⁻³³⁰
41	**15**	**48426484**	**rs1426654**	***SLC24A5***	**G**	**A**	**157483**	**0.102**	**0.684**	**0.0644**	**2.72 × 10⁻²⁶**

continued on next page

| Region	Chr	Pos	RS	Nearest gene	Ref allele	Alt allele	N	Freq	Beta	SE	P
42	**16**	**1376386**	**rs761063**	*UBE2I*	**G**	**C**	**157483**	**0.277**	**−0.050**	**0.0084**	2.07×10^{-9}
43	**17**	**1966889**	**rs4790309**	*SMG6*	**T**	**C**	**157483**	**0.456**	**0.041**	**0.0075**	3.71×10^{-8}
44	**17**	**67497367**	**rs3809761**	*MAP2K6*	**G**	**A**	**157483**	**0.811**	**0.059**	**0.0101**	5.08×10^{-9}
45	17	79612397	rs6420484	**TSPAN10*	G	A	157483	0.645	−0.161	0.0080	7.53×10^{-90}
46	**19**	**7581625**	**rs73488486**	*ZNF358*	**T**	**G**	**157483**	**0.097**	**0.086**	**0.0128**	2.40×10^{-11}
47	**20**	**4948248**	**rs2748901**	*SLC23A2*	**G**	**A**	**157483**	**0.463**	**−0.047**	**0.0075**	3.37×10^{-10}
48	21	38568882	rs2835660	*TTC3*	G	C	157483	0.545	−0.064	0.0075	1.62×10^{-17}
49	**21**	**44783287**	**rs622330**	**SIK1*	**G**	**A**	**157483**	**0.512**	**0.111**	**0.0075**	3.87×10^{-49}
50	**22**	**46369657**	**rs35051352**	*WNT7B*	**G**	**C**	**147875**	**0.549**	**−0.062**	**0.0081**	3.33×10^{-14}
51	**23**	**48118832**	**rs78542430**	*SSX1*	**T**	**A**	**157483**	**0.450**	**−0.095**	**0.0062**	5.93×10^{-53}
52	**23**	**119439335**	**rs5957354**	*TMEM255A*	**T**	**A**	**157483**	**0.861**	**−0.090**	**0.0092**	5.43×10^{-23}
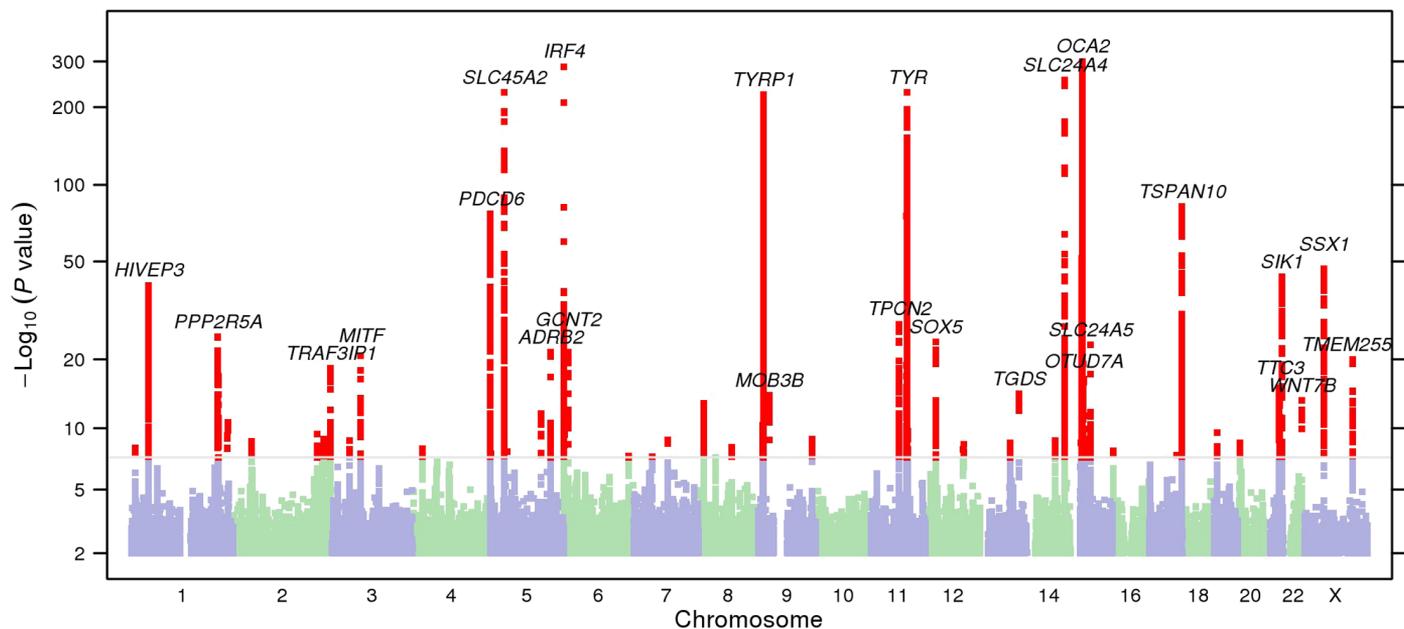
Fig. 1. Manhattan plot of eye color GWAS results in the European discovery cohort (N = 157,485). All results with P values $<5 \times 10^{-8}$ are indicated in red.

Using additional independent data from 35,501 individuals of European ancestry enrolled in nine different studies collected by the International Visible Trait Genetics (VisiGen) Consortium and its study partners, we sought to replicate the findings from the discovery analysis. For this, we tested the strongest associated SNP for each of the 52 regions (lead SNPs) identified in the discovery GWAS. To minimize data fragmentation and therefore maximize statistical power, results of regression analyses from each of the nine VisiGen studies were meta-analyzed together, and the outcome of this meta-analysis was used to replicate the discovery stage findings. Because of the differences in sample size, genotyping platforms, and imputation methods across the VisiGen studies, several SNPs highlighted by the discovery GWAS were not available or did not pass quality control in the replication dataset, particularly variants of low minor allele frequency (MAF). Some VisiGen studies had no availability of X chromosome data. Therefore, replication was attempted

for 48 lead SNPs from 50 autosomal regions highlighted in the discovery analysis (table ST4). Overall, for 47 (98%) of the 48 lead SNPs, the direction of effect was the same in the replication analysis as previously seen in the discovery analysis (Fig. 2 and table ST4). For 27 (56%) lead SNPs, we obtained significant replication after applying a strict Bonferroni adjustment for multiple testing (0.05/48 = 1.04×10^{-3}), including 18 (47%) of the 38 novel lead SNPs. An additional nine (24%) of the novel lead SNPs were associated at nominal (P < 0.05) significance. There was minimal heterogeneity between cohorts for these SNPs; a histogram of the heterogeneity I-scores is provided in fig. S1.

Results from both the discovery and replication studies were then meta-analyzed including a total of 192,986 Europeans from 10 populations. The Devlin's genomic factor was relatively unchanged in this analysis (λ = 1.14). With the added statistical power, we found genome-wide significant associations for SNPs in an additional
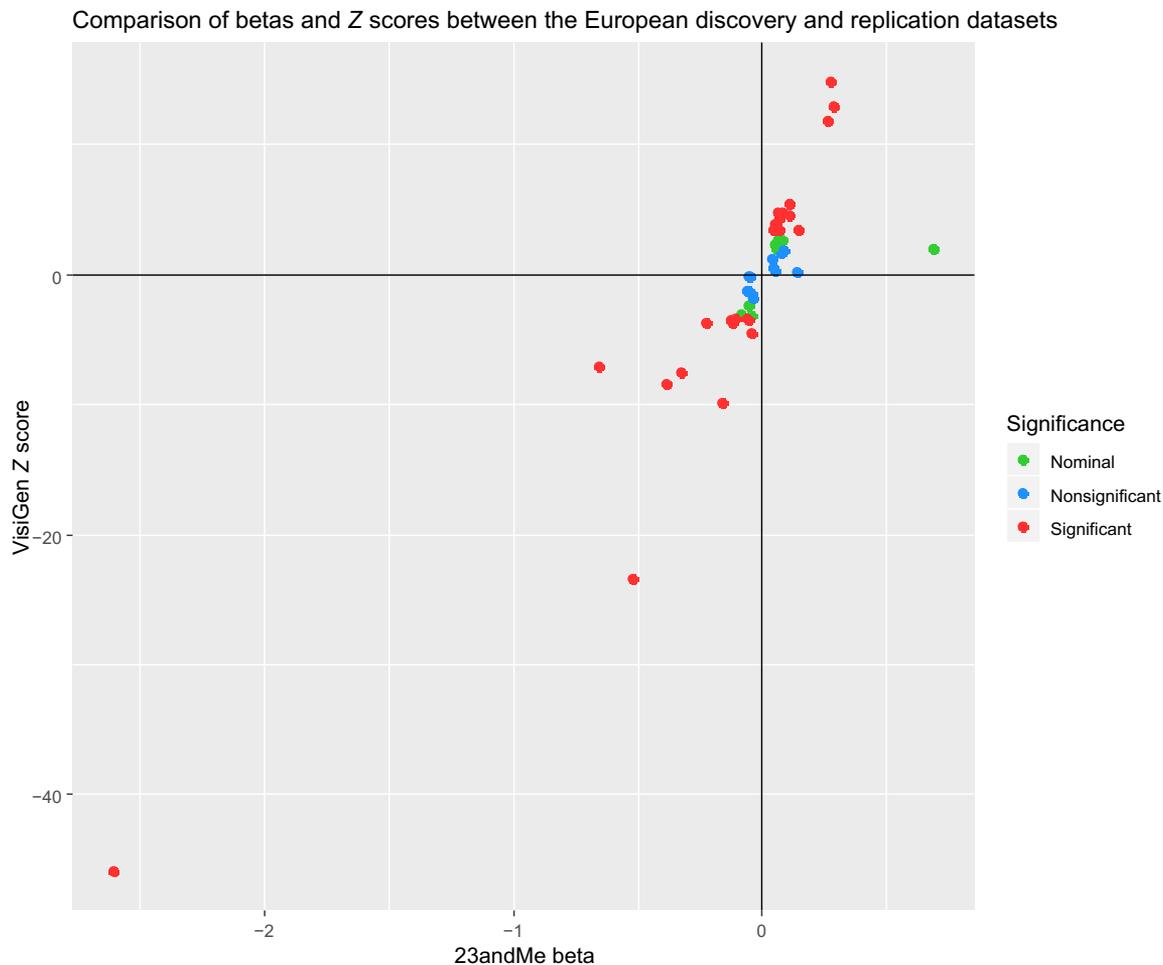
Fig. 2. Comparison of SNP beta and *Z* scores between the European discovery (23andMe, *N* = 157,485) and the European replication (VisiGen, *N* = 35,501) cohorts. Color codes represent significance in the replication cohort: red = $P < 1.04 \times 10^{-3}$; green = $P < 0.05$; blue = $P > 0.05$.

nine separate genomic regions (table ST5). All nine genetic loci were novel, not previously associated with human eye color. Three of these, *PDE4D* (rs62370541, $P = 5.8 \times 10^{-9}$), *JAZF1* (rs849142, 3.0×10^{-9}), and *SOX6* (rs2351061, $P = 4.0 \times 10^{-8}$), have recently been associated with hair color (*31*). *PDE4D* inhibitors have been shown to increase melanin pigment in the skin of mouse models (*38*), as *PDE4D* is a target of the melanin-stimulating hormone/cyclic adenosine mono-phosphate (AMP)/melanocyte-inducing transcription factor (MITF) pathway (*38*).

Next, we compared results obtained from European subjects with association observations from a meta-analysis of 1636 Asian indi-viduals (959 Han Chinese and 677 Indians from Singapore), for which quantitative measurements of the iris color were available (see the Supplementary Methods for information about quantitative phenotyping). Reliable SNP data were available for 44 of the lead autosomal SNPs from the 52 genetic loci identified by conditional analysis in the European discovery cohort. The remaining eight lead SNPs had a MAF smaller than 1% in this cohort and thus were ex-cluded on account of the much smaller sample size. Thirty-one SNPs (70%) had the same direction of effect as the European analysis, and 5 (11%) of these 44 SNPs were significantly associ-ated with eye color in Asians, after adjustment for multiple testing

(Bonferroni-adjusted *P* value: $0.05/44 = 1.13 \times 10^{-3}$). This included SNPs from two of the newly identified genes *GPR157* (rs6693258, $P = 3.6 \times 10^{-5}$) and *SIK1* (rs622330, $P = 1.7 \times 10^{-4}$) (table ST6), as well as from the previously known *HERC2* gene (rs1129038, $P = 2.2 \times 10^{-4}$), which is the most strongly eye color–associated SNP in Europeans (*17*). For this marker, we observed considerable MAF differences between Europeans and Asians (T allele: 0.05 in Asians and 0.73 in Europeans). The strongest association in the Singaporean cohorts, however, was found within *SLC24A5* (rs1426654, $P = 1.9 \times 10^{-7}$) representing the OCA type 6 (OCA6) locus. Notably, OCA6 was first reported in a Chinese family (*39*). However, the *HERC2* and *SLC24A5* SNPs were the only two variants to display heterogeneity between the two Singaporean cohorts (table ST6), with association primarily being driven by the cohort of South Asian ancestry. This is likely a due to South Asian and European populations being more closely related than East Asian and European populations. The MAFs for rs1426654 were considerably different between Asians (0.80 over-all, 0.99 in East Asians, and 0.52 in South Asians) and Europeans (0.10) for the G allele, explaining why this SNP is often used as a DNA marker for biogeographic ancestry (*40*). Previously, *SLC24A5* rs1426654 not only explained a considerable proportion of the vari-ation in skin pigmentation between different continental populations

(41) but also was associated with skin and eye color variation within a South Asian population (28, 42). Despite such drastic differences in allele frequency for some SNP alleles, the shared eye color effect between populations from two different continents demonstrates the value of our multiethnic study.

DNA variants identified via GWAS are markers of statistical association and not necessarily causative. Because of the differences in LD across continental populations, association may not be universally replicable. Therefore, we next assessed the presence of association not just for the lead SNPs in each region but also by testing other SNPs located within the 52 genomic regions identified in the European discovery analysis. Several SNPs within these regions (table ST7) showed evidence for eye color association in both European and Asian populations, with genome-wide significance in Europeans and suggestive levels of genome-wide association ($P < 1 \times 10^{-5}$) in Asians, despite a much smaller sample size. These SNPs also showed no significant heterogeneity between the two Asian cohorts. It is therefore possible that, against the background of much stronger effects of European-only alleles or due to population-specific differences in MAF, some polymorphisms contributing to European eye color variation are also relevant for eye color variability in non-European populations, such as variation of brown eye color in Asians tested here.

In Europeans, the 112 autosomal SNPs identified through conditional analysis (all autosomal SNPs shown in table S1) explained 99.96% (SE = 6.5%, $P = 4.8 \times 10^{-279}$) of the liability scale for blue eyes (against brown eyes) and 38.5% (SE = 5.7%, $P = 2.2 \times 10^{-130}$) for intermediate eyes in the TwinsUK cohort, which was one of the VisiGen cohorts used for replication. Using the same linear scale as the GWAS analysis, these autosomal SNPs explained 53.2% (SE = 4.0%, $P = 1.2 \times 10^{-322}$) of the total phenotypic variation in eye color in TwinsUK.

Last, we performed in silico analyses to explore the putative function of the genetic loci our study highlighted with significant eye color association using the conditional SNPs. Gene set enrichment analysis identified multiple pathways with significant enrichment (table ST8). As expected, this included several pigmentation process pathways, with "Developmental Pigmentation" the most significant ($P = 7 \times 10^{-6}$), followed by "Frizzled Binding" ($P = 2.0 \times 10^{-4}$) and "Melanin Metabolic Process" ($P = 2.6 \times 10^{-4}$). We also examined the potential effects of the identified eye color–associated SNPs on gene expression using data from the GTEx Consortium (43). Despite the lack of iris tissue in the GTEx repository, many SNPs showed significant eQTL (expression quantitative trait loci) effects in multiple cell types and tissues (44), as seen for the associated SNPs across 38 (79%) of the 48 tissues in the GTEx dataset (table ST9). Most of the strongest effects were seen in nerve and sun-exposed skin tissue ($P = 8.47 \times 10^{-62}$ and $P = 4.84 \times 10^{-49}$, respectively) for rs2835660, where the C allele is significantly associated with a decrease in *TTC3* expression. *TTC3* is in proximity to *DSCR9*, a gene whose polymorphisms were previously associated with eye color (11). These results implicate *TTC3* as a more likely candidate gene influencing eye color at this genetic locus than *DSCR9* and that its effect on eye color is likely mediated through variation in gene expression. The lack of iris tissue information in GTEx likely explains the absence of stronger eQTL effects, such as for SNPs with regulatory effects over gene transcription (16).

DISCUSSION

We report the results of the largest GWAS for human eye color to date. In addition to confirming the association of SNPs in 11 previously known eye color genes (11, 13, 14, 17, 28), the identification of 50 novel eye color–associated genetic loci helps explain previously missing heritability of eye color variability in European populations. Moreover, because of the multiethnic design of our study, we demonstrate that several of the genetic loci discovered in Europeans also have an effect on eye color in Asians.

Eight of the genes in or near the loci newly associated with eye color in our study were previously reported for genetic associations with other pigmentation traits, such as hair and skin color, for instance, *TPCN2*, *MITF*, and *DCT* (27, 30, 32, 45). The commonality of associated DNA variants across the three pigmentation traits helps explain why the different pigmentation traits frequently (but not completely) intercorrelate in European populations. While many significant genetic associations are shared between iris color and other pigmentation traits, there are also notable differences. Although DNA variants within the *MC1R* gene are strongly associated with light skin and red hair color (27), no detectable association with eye color was found in our large GWAS, in line with previous albeit smaller-sized GWASs of more limited statistical power (11, 12, 14). Similarly, other DNA variants strongly associated with skin and hair color within genes, such as *SILV*, *ASIP*, and *POMC* (30), showed no statistically significant effect on eye color in this study, nor in previous studies. Moreover, we also identified 34 genetic loci that were significantly associated with eye color, but for which there is no report of significant association with hair and/or skin color. This is remarkable as the statistical power of the recent GWASs on hair color (31, 46) and sun sensitivity (32) were similar to that of our current eye color GWAS. Significant associations for SNPs in/near genes involved in iris structure, such as *TRAF3IP1* and *SEMA3A*, suggest that they exert their effects with changes in Tyndall scattering, rather than through alterations of melanin metabolism. Overall, this demonstrates that although many genes overlap between eye, hair, and skin color, the different human pigmentation traits are not completely determined by the same genes as we showed.

The major strengths of our study compared with previous eye color GWASs arise from the larger sample size, which translated into increased statistical power and also the ability to lower the threshold of MAF for which sufficient power to detect association is available. Rare SNPs are often a source of considerable phenotypic variation (47). For instance, seven (6%) of the independently associated SNPs identified by conditional analysis in the discovery cohort had a MAF between 0.1 and 1%. Despite their low frequency, however, five (71%) of these rare SNPs were in the same region as other, more common conditional SNPs that did replicate. The remaining two loci (*DAB2* and an intronic region on chromosome 4) that were not formally replicated should therefore be considered only as strong candidates with respect to their association with eye color, pending independent validation in future studies.

Another strength of this work is the inclusion of European and non-European populations. Non-European populations are underrepresented in the GWAS literature in general, including in pigmentation GWASs, but their study is important for the understanding of the genetic basis of human phenotypes (48). Although eye color variation is typically attributed to individuals of (at least partial) European descent, or those originating from areas nearby Europe, more subtle variation in brown eyes is also observed in Asian populations without European admixture (9). Our results from the Asian cohorts showed remarkable consistence in the genetic architecture of eye color among individuals of different continental ancestries

with Asian replication for the two major European genes *OCA2* and *HERC2*. Moreover, our findings also suggest that while a single regulatory variant in *HERC2* is responsible for most blue/brown variation in Europeans (*16*), many additional DNA variants across both *OCA2* and *HERC2* seem to have independent effects. This hypothesis is further supported by our conditional analysis in the European discovery cohort, identifying independent associations spanning ~14 mbp across both genes rather than a concentrated cluster centered at *HERC2* rs1129038. This is remarkable given the large eye color variation from the lightest blue to the darkest brown in Europeans, compared with the more limited variation within brown eye color in Asians.

In conclusion, our work has identified numerous novel genetic loci associated with human eye color in Europeans, of which a subset also shows effects in Asians, despite their largely reduced phenotypic eye color variation compared with Europeans. The genetic loci we identified explain the majority (53.2%) of eye color phenotypic variation (classified using a three-category scale) in Europeans and a large proportion of the previously noted missing heritability of eye color. Our findings clearly demonstrate that eye color is a genetically highly complex human trait, similar to hair (*31*) and skin color (*32*), as highlighted recently in large European GWASs. The large number of novel eye color–associated genetic loci identified here provide a valuable resource for future functional studies, aiming to understand the molecular mechanisms that explain their eye color association, and for future genetic prediction studies, aiming to improve DNA-based eye color prediction in anthropological and forensic applications.

METHODS

We performed a GWAS using information from two sources: 157,485 research participants of European ancestry recruited among the customer base of the personal genomics company 23andMe Inc. (Sunnyvale, CA, USA) (*49*) and a meta-analysis of 35,501 European individuals from nine populations collected by members of the International Visible Trait Genetics (VisiGen) Consortium (*45*) and their study partners. As a secondary analysis and for comparative purposes, we included an additional set of 1636 individuals of Asian ancestry from two populations (959 Han Chinese and 677 Indians from Singapore).

Populations and participants
23andMe
Participants. Participants included for this analysis were 23andMe consumers who consented to participate in research and were determined to have greater than 97% European ancestry using genotype clustering. A segmented identity-by-descent algorithm (*50*) was then applied to obtain the maximal subset (157,485) of unrelated individuals to be used as our discovery cohort (full details on ancestry and relatedness calculations are provided in the Supplementary Methods, and principal component plots are provided in fig. S2).

Genotyping. Participants were genotyped on one of 23andMe's own custom-designed SNP arrays. The V4 platform is a fully customized array containing a subset of SNPs from previous versions, while the V3 is a customized platform based on the OmniExpress+ BeadChip, and the V1 and V2 platforms are customized variants of the Illumina HumanHap550+ BeadChip (full descriptions of platforms and genotyping procedure are provided in the Supplementary Methods). Imputation of additional variants was completed using 1000 Genomes phase 1 as reference haplotypes (*51*).

Phenotyping. Phenotyping for iris color in this cohort was derived from questionnaires. Participants were asked to self-categorize eye color into one of seven distinct groups ranging from blue to dark brown based on color matching (full details are provided in the Supplementary Methods). This categorization allowed greater phenotypic resolution than previous three-color categories while maintaining groups distinctive enough to reduce misclassification. These groups were converted into a numerical scale ranging from 0 (blue) to 6 (dark brown) and used as the outcome variable for linear regression–based GWAS analysis.

VisiGen consortium and study partners
Participants. Specific recruitment criteria varied for each cohort (Supplementary Materials), with participants from the United Kingdom, The Netherlands, Italy, and European descendants in the United States and Australia.

Principal components analyses examined the genetic ancestry for each cohort and confirmed each was of nonadmixed European ancestry (specific details for each cohort are provided in the Supplementary Methods).

Phenotyping. Eye color categorization varied between cohorts, with most of the European cohorts adopting a 3- or 5-point categorical scale (details for each cohorts' classification are provided in the Supplementary Methods).

Singaporean cohorts
Participants. Full GWAS summary statistics were available for 959 participants of Chinese descent recruited at the Singapore Polyclinic and for 677 participants in the Singapore Indian Eye Study (SINDI).

Phenotyping. Phenotypes in these cohorts were ascertained using image analysis of digital photographs, better suited to capture variation within these populations (full details are provided in the Supplementary Methods).

Statistical analyses
Association tests. We performed an independent linear regression for each cohort under the assumption of an additive model for allelic effects, with adjustments made for age, sex, and the first five principal components. Adjustment was made also for the genotyping platform (V1 to V4), and additional measures controlling for relatedness were implemented for the VisiGen family cohorts (full details for each cohort are provided in the Supplementary Methods). The use of a linear scale may be limited by the assumption that there is an equal difference between each group; however, as eye color phenotypically follows linear scale across the color spectrum, this choice of statistical model is appropriate and in accordance with previous studies on eye color (*11, 27, 49*).

Meta-analyses. The QC'ed summary statistics from the VisiGen and Singaporean cohorts were pooled into two separate meta-analyses: the first was conducted using data from the nine VisiGen European cohorts, and the second using data from the two Singaporean cohorts. Both meta-analyses were performed using METAL (*52*) applying a weighted Z-score approach, given the differing phenotypic scales used across individual cohorts, with genomic control adjustments (*24*) made for the VisiGen meta-analysis (see the Supplementary Methods). A final meta-analysis was then conducted between the 23andMe cohort and the European VisiGen cohorts following the same procedures with a weighted Z score approach and genomic control adjustments.

Definition of genomic region and significance. We adopted the customary definition for a GWAS significance threshold of 5×10^{-08}.

For this work, an "associated region" was a genomic region demarcated by consecutive significantly associated genomic markers, separated by a nonassociated region greater than 1 million base pairs.

LD score regression. LD Hub (*53*) was used to perform LD score regression (*25*) on the summary statistics from our GWASs to test for possible inflation. This method is more suited to our study than the Devlin (*24*) genomic inflation lambda because of high trait polygenicity and the large cohort sizes (*25*).

Conditional analysis. Conditional analysis was conducted using summary statistics with genome-wide complex trait analysis (GCTA) (*54*) following the same procedure described by Yang *et al.* (*55*). Genotypic data from unrelated participants in TwinsUK were used as a reference for LD structure, a distance of 1000 kb was set as an assumption of complete linkage equilibrium, and a collinearity threshold was set at $R^2 > 0.9$.

Variant effect prediction. The predicted effects of variants were performed using the Ensembl Variant Effect Predictor (VEP) (*37*).

Heritability explained by the independent SNPs. Population liability scale heritability (*56*) explained by associated SNPs identified in the 23andMe discovery cohort was calculated using an unrelated sample from the TwinsUK cohort using restricted maximum likelihood analysis (*54*) under the assumption that the population prevalence (taken from the 23andMe cohort) is 30.3% for blue eyes and 38.3% for intermediate colors.

Pathway analysis. Genetic pathway analysis for iris color was performed with MAGENTA (*57*) using the 23andMe discovery SNP association results as input (full details are provided in the Supplementary Methods). Gene set definitions defined by Gene Ontology (*58*) were obtained from the Molecular Signatures Database (version MSigDB v6.1) (*59*) for this analysis.

Members of consortia and affiliations
23andMe research team

23andMe Inc., Sunnyvale, California, USA—M. Agee, A. Auton, R. K. Bell, K. Bryc, S. L. Elson, P. Fontanillas, K. E. Huber, A. Kleinman, N. K. Litterman, M. H. McIntyre, J. L. Mountain, E. S. Noblin, C. A. M. Northover, S. J. Pitts, O. V. Sazonova, J. F. Shelton, S. Shringarpure, C. Tian, J. Y. Tung, and V. Vacic.

International Visible Trait Genetics Consortium

Department of Twins Research and Genetic Epidemiology, King's College London, London, United Kingdom—P. G. Hysi and T. D. Spector.

Department of Genetic Identification, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands—M. Kayser and F. Liu.

QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia—D. L. Duffy and N. G. Martin.

University of Queensland Diamantina Institute, University of Queensland, Brisbane, Queensland, Australia—D. M. Evans.

SUPPLEMENTARY MATERIALS

REFERENCES AND NOTES

1. A. R. Wielgus, T. Sarna, Melanin in human irides of different color and age of donors. *Pigment Cell Res.* **18**, 454–464 (2005).
2. R. A. Sturm, T. N. Frudakis, Eye colour: Portals into pigmentation genes and ancestry. *Trends Genet.* **20**, 327–332 (2004).
3. M. Larsson, D. L. Duffy, G. Zhu, J. Z. Liu, S. Macgregor, A. F. McRae, M. J. Wright, R. A. Sturm, D. A. Mackey, G. W. Montgomery, N. G. Martin, S. E. Medland, GWAS findings for human iris patterns: Associations with variants in genes that influence normal neuronal pattern development. *Am. J. Hum. Genet.* **89**, 334–343 (2011).
4. G. Prota, D. N. Hu, M. R. Vincensi, S. A. McCormick, A. Napolitano, Characterization of melanins in human irides and cultured uveal melanocytes from eyes of different colors. *Exp. Eye Res.* **67**, 293–299 (1998).
5. P. Frost, European hair and eye color. *Evol. Hum. Behav.* **27**, 85–103 (2006).
6. P. Frost, Human skin-color sexual dimorphism: A test of the sexual selection hypothesis. *Am. J. Phys. Anthropol.* **133**, 779–780 (2007).
7. N. G. Jablonski, G. Chaplin, The colours of humanity: The evolution of pigmentation in the human lineage. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **372**, 20160349 (2017).
8. M. P. Donnelly, P. Paschou, E. Grigorenko, D. Gurwitz, C. Barta, R. B. Lu, O. V. Zhukova, J. J. Kim, M. Siniscalco, M. New, H. Li, S. L. B. Kajuna, V. G. Manolopoulos, W. C. Speed, A. J. Pakstis, J. R. Kidd, K. K. Kidd, A global view of the OCA2-HERC2 region and pigmentation. *Hum. Genet.* **131**, 683–696 (2012).
9. M. Edwards, D. Cha, S. Krithika, M. Johnson, G. Cook, E. J. Parra, Iris pigmentation as a quantitative trait: Variation in populations of European, East Asian and South Asian ancestry and association with candidate gene polymorphisms. *Pigment Cell Melanoma Res.* **29**, 141–162 (2016).
10. G. Zhu, D. M. Evans, D. L. Duffy, G. W. Montgomery, S. E. Medland, N. A. Gillespie, K. R. Ewen, M. Jewell, Y. W. Liew, N. K. Hayward, R. A. Sturm, J. M. Trent, N. G. Martin, A genome scan for eye color in 502 twin families: Most variation is due to a QTL on chromosome 15q. *Twin Res.* **7**, 197–210 (2004).
11. F. Liu, A. Wollstein, P. G. Hysi, G. A. Ankra-Badu, T. D. Spector, D. Park, G. Zhu, M. Larsson, D. L. Duffy, G. W. Montgomery, D. A. Mackey, S. Walsh, O. Lao, A. Hofman, F. Rivadeneira, J. R. Vingerling, A. G. Uitterlinden, N. G. Martin, C. J. Hammond, M. Kayser, Digital quantification of human eye color highlights genetic association of three new loci. *PLOS Genet.* **6**, e1000934 (2010).
12. J. Han, P. Kraft, H. Nan, Q. Guo, C. Chen, A. Qureshi, S. E. Hankinson, F. B. Hu, D. L. Duffy, Z. Z. Zhao, N. G. Martin, G. W. Montgomery, N. K. Hayward, G. Thomas, R. N. Hoover, S. Chanock, D. J. Hunter, A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLOS Genet.* **4**, e1000074 (2008).
13. M. Kayser, F. Liu, A. C. J. W. Janssens, F. Rivadeneira, O. Lao, K. van Duijn, M. Vermeulen, P. Arp, M. M. Jhamai, W. F. J. van IJcken, J. T. den Dunnen, S. Heath, D. Zelenika, D. D. G. Despriet, C. C. W. Klaver, J. R. Vingerling, P. T. V. M. de Jong, A. Hofman, Y. S. Aulchenko, A. G. Uitterlinden, B. A. Oostra, C. M. van Duijn, Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. *Am. J. Hum. Genet.* **82**, 411–423 (2008).
14. P. Sulem, D. F. Gudbjartsson, S. N. Stacey, A. Helgason, T. Rafnar, K. P. Magnusson, A. Manolescu, A. Karason, A. Palsson, G. Thorleifsson, M. Jakobsdottir, S. Steinberg, S. Pálsson, F. Jonasson, B. Sigurgeirsson, K. Thorisdottir, R. Ragnarsson, K. R. Benediktsdottir, K. K. Aben, L. A. Kiemeney, J. H. Olafsson, J. Gulcher, A. Kong, U. Thorsteinsdottir, K. Stefansson, Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet.* **39**, 1443–1452 (2007).
15. G. C. Davenport, C. B. Davenport, F. Galton, Heredity of EYE-COLOR in man. *Science* **26**, 589–592 (1907).
16. H. Eiberg, J. Troelsen, M. Nielsen, A. Mikkelsen, J. Mengel-From, K. W. Kjaer, L. Hansen, Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. *Hum. Genet.* **123**, 177–187 (2008).
17. R. A. Sturm, D. L. Duffy, Z. Z. Zhao, F. P. N. Leite, M. S. Stark, N. K. Hayward, N. G. Martin, G. W. Montgomery, A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. *Am. J. Hum. Genet.* **82**, 424–431 (2008).
18. M. Visser, M. Kayser, R. J. Palstra, HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. *Genome Res.* **22**, 446–455 (2012).
19. F. Liu, K. van Duijn, J. R. Vingerling, A. Hofman, A. G. Uitterlinden, A. C. J. W. Janssens, M. Kayser, Eye color and the prediction of complex phenotypes from genotypes. *Curr. Biol.* **19**, R192–R193 (2009).
20. T. E. King, G. G. Fortes, P. Balaresque, M. G. Thomas, D. Balding, P. M. Delser, R. Neumann, W. Parson, M. Knapp, S. Walsh, L. Tonasso, J. Holt, M. Kayser, J. Appleby, P. Forster, D. Ekserdjian, M. Hofreiter, K. Schürer, Identification of the remains of King Richard III. *Nat. Commun.* **5**, 5631 (2014).
21. S. Walsh, F. Liu, K. N. Ballantyne, M. van Oven, O. Lao, M. Kayser, IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Sci. Int. Genet.* **5**, 170–180 (2011).
22. S. Walsh, L. Chaitanya, L. Clarisse, L. Wirken, J. Draus-Barini, L. Kovatsi, H. Maeda, T. Ishikawa, T. Sijen, P. de Knijff, W. Branicki, F. Liu, M. Kayser, Developmental validation

of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. *Forensic Sci. Int. Genet.* **9**, 150–161 (2014).

23. T. A. Manolio, F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. C. Mackay, S. A. McCarroll, P. M. Visscher, Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).

24. B. Devlin, K. Roeder, Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).

25. B. K. Bulik-Sullivan, P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang; Schizophrenia Working Group of the Psychiatric Genomics Consortium, N. Patterson, M. J. Daly, A. L. Price, B. M. Neale, LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

26. R. A. Sturm, M. Larsson, Genetics of human iris colour and patterns. *Pigment Cell Melanoma Res.* **22**, 544–562 (2009).

27. S. I. Candille, D. M. Absher, S. Beleza, M. Bauchet, B. McEvoy, N. A. Garrison, J. Z. Li, R. M. Myers, G. S. Barsh, H. Tang, M. D. Shriver, Genome-wide association studies of quantitatively measured skin, hair, and eye pigmentation in four European populations. *PLOS ONE* **7**, e48294 (2012).

28. M. Jonnalagadda, M. A. Faizan, S. Ozarkar, R. Ashma, S. Kulkarni, H. L. Norton, E. Parra, A genome-wide association study of skin and iris pigmentation among individuals of South Asian ancestry. *Genome Biol. Evol.* **11**, 1066–1076 (2019).

29. F. Peng, G. Zhu, P. G. Hysi, R. J. Eller, Y. Chen, Y. Li, M. A. Hamer, C. Zeng, R. L. Hopkins, C. L. Jacobus, P. L. Wallace, A. G. Uitterlinden, M. A. Ikram, T. Nijsten, D. L. Duffy, S. E. Medland, T. D. Spector, S. Walsh, N. G. Martin, F. Liu, M. Kayser, Genome-wide association studies identify multiple genetic loci influencing eyebrow color variation in Europeans. *J. Invest. Dermatol.* **139**, 1601–1605 (2019).

30. R. A. Sturm, R. D. Teasdale, N. F. Box, Human pigmentation genes: Identification, structure and consequences of polymorphic variation. *Gene* **277**, 49–62 (2001).

31. P. G. Hysi, A. M. Valdes, F. Liu, N. A. Furlotte, D. M. Evans, V. Bataille, A. Visconti, G. Hemani, G. M. Malhon, S. M. Ring, G. D. Smith, D. L. Duffy, G. Zhu, S. D. Gordon, S. E. Medland, B. D. Lin, G. Willemsen, J. J. Hottenga, D. Vuckovic, G. Girotto, I. Gandin, C. Sala, M. P. Concas, M. Brumat, P. Gasparini, D. Toniolo, M. Cocca, A. Robino, S. Yazar, A. W. Hewitt, Y. Chen, C. Zeng, A. G. Uitterlinden, M. A. Ikram, M. A. Hamer, C. M. van Duijn, T. Nijsten, D. A. Mackey, M. Falchi, D. I. Boomsma, N. G. Martin; International Visible Trait Genetics Consortium, D. A. Hinds, M. Kayser, T. D. Spector, Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability. *Nat. Genet.* **50**, 652–656 (2018).

32. A. Visconti, D. L. Duffy, F. Liu, G. Zhu, W. Wu, Y. Chen, P. G. Hysi, C. Zeng, M. Sanna, M. M. Iles, P. A. Kanetsky, F. Demenais, M. A. Hamer, A. G. Uitterlinden, M. A. Ikram, T. Nijsten, N. G. Martin, M. Kayser, T. D. Spector, J. Han, V. Bataille, M. Falchi, Genome-wide association study in 176,678 Europeans reveals genetic loci for tanning response to sun exposure. *Nat. Commun.* **9**, 1684 (2018).

33. J. MacArthur, E. Bowler, M. Cerezo, L. Gil, P. Hall, E. Hastings, H. Junkins, A. McMahon, A. Milano, J. Morales, Z. M. Pendlington, D. Welter, T. Burdett, L. Hindorff, P. Flicek, F. Cunningham, H. Parkinson, The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).

34. M. S. Tedja, R. Wojciechowski, P. G. Hysi, N. Eriksson, N. A. Furlotte, V. J. M. Verhoeven, A. I. Iglesias, M. A. Meester-Smoor, S. W. Tompson, Q. Fan, A. P. Khawaja, C.-Y. Cheng, R. Höhn, K. Yamashiro, A. Wenocur, C. Grazal, T. Haller, A. Metspalu, J. Wedenoja, J. B. Jonas, Y. X. Wang, J. Xie, P. Mitchell, P. J. Foster, B. E. K. Klein, R. Klein, A. D. Paterson, S. M. Hosseini, R. L. Shah, C. Williams, Y. Y. Teo, Y. C. Tham, P. Gupta, W. Zhao, Y. Shi, W.-Y. Saw, E.-S. Tai, X. L. Sim, J. E. Huffman, O. Polašek, C. Hayward, G. Bencic, I. Rudan, J. F. Wilson; CREAM Consortium; 23andMe Research Team; UK Biobank Eye; Vision Consortium, P. K. Joshi, A. Tsujikawa, F. Matsuda, K. N. Whisenhunt, T. Zeller, P. J. van der Spek, R. Haak, H. Meijers-Heijboer, E. M. van Leeuwen, S. K. Iyengar, J. H. Lass, A. Hofman, F. Rivadeneira, A. G. Uitterlinden, J. R. Vingerling, T. Lehtimäki, O. T. Raitakari, G. Biino, M. P. Concas, T.-H. Schwantes-An, R. P. Igo Jr., G. Cuellar-Partida, N. G. Martin, J. E. Craig, P. Gharahkhani, K. M. Williams, A. Nag, J. S. Rahi, P. M. Cumberland, C. Delcourt, C. Bellenguez, J. S. Ried, A. A. Bergen, T. Meitinger, C. Gieger, T. Y. Wong, A. W. Hewitt, D. A. Mackey, C. L. Simpson, N. Pfeiffer, O. Pärssinen, P. N. Baird, V. Vitart, N. Amin, C. M. van Duijn, J. E. Bailey-Wilson, T. L. Young, S.-M. Saw, D. Stambolian, S. M. Gregor, J. A. Guggenheim, J. Y. Tung, C. J. Hammond, C. C. W. Klaver, Genome-wide association meta-analysis highlights light-induced signaling as a driver for refractive error. *Nat. Genet.* **50**, 834–848 (2018).

35. J. Shoag, R. Haq, M. Zhang, L. Liu, G. C. Rowe, A. Jiang, N. Koulisis, C. Farrel, C. I. Amos, Q. Wei, J. E. Lee, J. Zhang, T. S. Kupper, A. A. Qureshi, R. Cui, J. Han, D. E. Fisher, Z. Arany, PGC-1 coactivators regulate MITF and the tanning response. *Mol. Cell* **49**, 145–157 (2013).

36. M. Y. Kim, T. Y. Choi, J. H. Kim, J. H. Lee, J. G. Kim, K. C. Sohn, K. S. Yoon, C. D. Kim, J. H. Lee, T. J. Yoon, MKK6 increases the melanocyte dendricity through the regulation of Rho family GTPases. *J. Dermatol. Sci.* **60**, 114–119 (2010).

37. W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, A. Thormann, P. Flicek, F. Cunningham, The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

38. M. Khaled, C. Levy, D. E. Fisher, Control of melanocyte differentiation by a MITF-PDE4D3 homeostatic circuit. *Genes Dev.* **24**, 2276–2281 (2010).

39. A. H. Wei, D. J. Zang, Z. Zhang, X. Z. Liu, X. He, L. Yang, Y. Wang, Z. Y. Zhou, M. R. Zhang, L. L. Dai, X. M. Yang, W. Li, Exome sequencing identifies SLC24A5 as a candidate gene for nonsyndromic oculocutaneous albinism. *J. Invest. Dermatol.* **133**, 1834–1840 (2013).

40. E. Giardina, I. Pietrangeli, C. Martinez-Labarga, C. Martone, F. Angelis, A. Spinella, G. Stefano, O. Rickards, G. Novelli, Haplotypes in SLC24A5 gene as ancestry informative markers in different populations. *Curr. Genomics* **9**, 110–114 (2008).

41. R. L. Lamason, M. A. Mohideen, J. R. Mest, A. C. Wong, H. L. Norton, M. C. Aros, M. J. Jurynec, X. Mao, V. R. Humphreville, J. E. Humbert, S. Sinha, J. L. Moore, P. Jagadeeswaran, W. Zhao, G. Ning, I. Makalowska, P. McKeigue, D. O'donnell, R. Kittles, E. J. Parra, N. J. Mangini, D. J. Grunwald, M. D. Shriver, V. A. Canfield, K. C. Cheng, SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **310**, 1782–1786 (2005).

42. R. P. Stokowski, P. V. K. Pant, T. Dadd, A. Fereday, D. A. Hinds, C. Jarman, W. Filsell, R. S. Ginger, M. R. Green, F. J. van der Ouderaa, D. R. Cox, A genomewide association study of skin pigmentation in a South Asian population. *Am. J. Hum. Genet.* **81**, 1119–1132 (2007).

43. G. T. Consortium Human genomics, The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).

44. D. L. Nicolae, E. Gamazon, W. Zhang, S. Duan, M. E. Dolan, N. J. Cox, Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. *PLOS Genet.* **6**, e1000888 (2010).

45. F. Liu, M. Visser, D. L. Duffy, P. G. Hysi, L. C. Jacobs, O. Lao, K. Zhong, S. Walsh, L. Chaitanya, A. Wollstein, G. Zhu, G. W. Montgomery, A. K. Henders, M. Mangino, D. Glass, V. Bataille, R. A. Sturm, F. Rivadeneira, A. Hofman, W. F. J. van IJcken, A. G. Uitterlinden, R. J. T. S. Palstra, T. D. Spector, N. G. Martin, T. E. C. Nijsten, M. Kayser, Genetics of skin color variation in Europeans: Genome-wide association studies with functional follow-up. *Hum. Genet.* **134**, 823–835 (2015).

46. M. D. Morgan, E. Pairo-Castineira, K. Rawlik, O. Canela-Xandri, J. Rees, D. Sims, A. Tenesa, I. J. Jackson, Genome-wide study of hair colour in UK Biobank explains most of the SNP heritability. *Nat. Commun.* **9**, 5271 (2018).

47. International Multiple Sclerosis Genetics Consortium, Low-frequency and rare-coding variation contributes to multiple sclerosis risk. *Cell* **175**, 1679–1687 (2018).

48. L. A. Hindorff, V. L. Bonham, L. C. Brody, M. E. C. Ginoza, C. M. Hutter, T. A. Manolio, E. D. Green, Prioritizing diversity in human genomics research. *Nat. Rev. Genet.* **19**, 175–185 (2018).

49. N. Eriksson, J. M. Macpherson, J. Y. Tung, L. S. Hon, B. Naughton, S. Saxonov, L. Avey, A. Wojcicki, I. Pe'er, J. Mountain, Web-based, participant-driven studies yield novel genetic associations for common traits. *PLOS Genet.* **6**, e1000993 (2010).

50. B. M. Henn, L. Hon, J. M. Macpherson, N. Eriksson, S. Saxonov, I. Pe'er, J. L. Mountain, Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. *PLOS ONE* **7**, e34267 (2012).

51. 1000 Genomes Project Consortium, A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).

52. C. J. Willer, Y. Li, G. R. Abecasis, METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

53. J. Zheng, A. M. Erzurumluoglu, B. L. Elsworth, J. P. Kemp, L. Howe, P. C. Haycock, G. Hemani, K. Tansey, C. Laurin; Early Genetics and Lifecourse Epidemiology (EAGLE) Eczema Consortium, B. S. Pourcain, N. M. Warrington, H. K. Finucane, A. L. Price, B. K. Bulik-Sullivan, V. Anttila, L. Paternoster, T. R. Gaunt, D. M. Evans, B. M. Neale, LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).

54. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

55. J. Yang, T. Ferreira, A. P. Morris, S. E. Medland; Genetic Investigation of ANthropometric Traits (GIANT) Consortium; DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium, P. A. F. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. N. Weedon, R. J. Loos, T. M. Frayling, M. I. McCarthy, J. N. Hirschhorn, M. E. Goddard, P. M. Visscher, Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, S361–S363 (2012).

56. S. H. Lee, N. R. Wray, M. E. Goddard, P. M. Visscher, Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* **88**, 294–305 (2011).

57. A. V. Segrè; DIAGRAM Consortium; MAGIC investigators, L. Groop, V. K. Mootha, M. J. Daly, D. Altshuler, Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLOS Genet.* **6**, e1001058 (2010).

58. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene Ontology: Tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

59. A. Liberzon, A. Subramanian, R. Pinchback, H. Thorvaldsdóttir, P. Tamayo, J. P. Mesirov, Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).

60. J. K. Pickrell, T. Berisa, J. Z. Liu, L. Ségurel, J. Y. Tung, D. A. Hinds, Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 709–717 (2016).

61. O. Delaneau, J. F. Zagury, J. Marchini, Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).

62. S. R. Browning, B. L. Browning, Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).

63. C. Fuchsberger, G. R. Abecasis, D. A. Hinds, minimac2: Faster genotype imputation. *Bioinformatics* **31**, 782–784 (2015).

64. X. Zheng, J. Shen, C. Cox, J. C. Wakefield, M. G. Ehm, M. R. Nelson, B. S. Weir, HIBAG—HLA genotype imputation with attribute bagging. *Pharm. J.* **14**, 192–200 (2014).

65. Q. S. Li, C. Tian, G. R. Seabrook, W. C. Drevets, V. A. Narayan, Analysis of 23andMe antidepressant efficacy survey data: Implication of circadian rhythm and neuroplasticity in bupropion response. *Transl. Psychiatry* **6**, e889 (2016).

66. C. B. Do, J. Y. Tung, E. Dorfman, A. K. Kiefer, E. M. Drabant, U. Francke, J. L. Mountain, S. M. Goldman, C. M. Tanner, J. W. Langston, A. Wojcicki, N. Eriksson, Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLOS Genet.* **7**, e1002141 (2011).

67. N. Eriksson, J. Y. Tung, A. K. Kiefer, D. A. Hinds, U. Francke, J. L. Mountain, C. B. Do, Novel associations for hypothyroidism include known autoimmune risk loci. *PLOS ONE* **7**, e34442 (2012).

68. A. K. Kiefer, J. Y. Tung, C. B. Do, D. A. Hinds, J. L. Mountain, U. Francke, N. Eriksson, Genome-wide analysis points to roles for extracellular matrix remodeling, the visual cycle, and neuronal development in myopia. *PLOS Genet.* **9**, e1003299 (2013).

69. J. Y. Tung, C. B. Do, D. A. Hinds, A. K. Kiefer, J. M. Macpherson, A. B. Chowdry, U. Francke, B. T. Naughton, J. L. Mountain, A. Wojcicki, N. Eriksson, Efficient replication of over 180 genetic associations with self-reported medical data. *PLOS ONE* **6**, e23473 (2011).

70. B. S. Hromatka, J. Y. Tung, A. K. Kiefer, C. B. Do, D. A. Hinds, N. Eriksson, Genetic variants associated with motion sickness point to roles for inner ear development, neurological processes and glucose homeostasis. *Hum. Mol. Genet.* **24**, 2700–2708 (2015).

71. A. Boyd, J. Golding, J. Macleod, D. A. Lawlor, A. Fraser, J. Henderson, L. Molloy, A. Ness, S. Ring, G. Davey Smith, Cohort Profile: The 'Children of the 90s'-the index offspring of the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.* **42**, 111–127 (2013).

72. A. Fraser, C. Macdonald-Wallis, K. Tilling, A. Boyd, J. Golding, G. Davey Smith, J. Henderson, J. Macleod, L. Molloy, A. Ness, S. Ring, S. M. Nelson, D. A. Lawlor, Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int. J. Epidemiol.* **42**, 97–110 (2013).

73. M. Zhang, F. Song, L. Liang, H. Nan, J. Zhang, H. Liu, L.-E. Wang, Q. Wei, J. E. Lee, C. I. Amos, P. Kraft, A. A. Qureshi, J. Han, Genome-wide association studies identify several new loci associated with pigmentation traits and skin cancer risk in European Americans. *Hum. Mol. Genet.* **22**, 2948–2959 (2013).

74. H. S. Chahal, W. Wu, K. J. Ransohoff, L. Yang, H. Hedlin, M. Desai, Y. Lin, H.-J. Dai, A. A. Qureshi, W.-Q. Li, P. Kraft, D. A. Hinds, J. Y. Tang, J. Han, K. Y. Sarin, Genome-wide association study identifies 14 novel risk alleles associated with basal cell carcinoma. *Nat. Commun.* **7**, 12510 (2016).

75. S. Yazar, H. Forward, C. M. McKnight, A. Tan, A. Soloshenko, S. K. Oates, W. Ang, J. C. Sherwin, D. Wood, J. A. Mountain, C. E. Pennell, A. W. Hewitt, D. A. Mackey, Raine Eye Health Study: Design, Methodology and Baseline Prevalence of Ophthalmic Disease in a Birth-cohort Study of Young Adults. *Ophthalmic Genet.* **34**, 199–208 (2013).

76. A. Hofman, G. G. O. Brusselle, S. D. Murad, C. M. van Duijn, O. H. Franco, A. Goedegebure, M. A. Ikram, C. C. W. Klaver, T. E. C. Nijsten, R. P. Peeters, B. H. C. Stricker, H. W. Tiemeier, A. G. Uitterlinden, M. W. Vernooij, The Rotterdam Study: 2016 objectives and design update. *Eur. J. Epidemiol.* **30**, 661–708 (2015).

77. A. Hofman, D. E. Grobbee, P. T. de Jong, F. A. van den Ouweland, Determinants of disease and disability in the elderly: The Rotterdam Elderly Study. *Eur. J. Epidemiol.* **7**, 403–422 (1991).

78. K. Estrada, M. Krawczak, S. Schreiber, K. van Duijn, L. Stolk, J. B. J. van Meurs, F. Liu, B. W. J. H. Penninx, J. H. Smit, N. Vogelzangs, J. J. Hottenga, G. Willemsen, E. J. C. de Geus, M. Lorentzon, H. von Eller-Eberstein, P. Lips, N. Schoor, V. Pop, J. de Keijzer, A. Hofman, Y. S. Aulchenko, B. A. Oostra, C. Ohlsson, D. I. Boomsma, A. G. Uitterlinden, C. M. van Duijn, F. Rivadeneira, M. Kayser, A genome-wide association study of northwestern Europeans involves the C-type natriuretic peptide signaling pathway in the etiology of human height variation. *Hum. Mol. Genet.* **18**, 3516–3524 (2009).

79. A. Moayyeri, C. J. Hammond, A. M. Valdes, T. D. Spector, Cohort Profile: TwinsUK and Healthy Ageing Twin Study. *Int. J. Epidemiol.* **42**, 76–85 (2013).

80. B. N. Howie, P. Donnelly, J. Marchini, A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLOS Genet.* **5**, e1000529 (2009).

81. O. Delaneau, J. Marchini, J. F. Zagury, A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2012).

82. X. Zhou, M. Stephens, Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).

83. T. A. Tun, J. Chua, Y. Shi, E. Sidhartha, S. G. Thakku, W. Shei, M. C. L. Tan, J. H. M. Quah, T. Aung, C.-Y. Cheng, Association of iris surface features with iris parameters assessed by swept-source optical coherence tomography in Asian eyes. *Brit. J. Ophthalmol.* **100**, 1682–1685 (2016).

84. R. Lavanya, V. S. E. Jeganathan, Y. Zheng, P. Raju, N. Cheung, E. S. Tai, J. J. Wang, E. Lamoureux, P. Mitchell, T. L. Young, H. Cajucom-Uy, P. J. Foster, T. Aung, S. M. Saw, T. Y. Wong, Methodology of the Singapore Indian Chinese Cohort (SICC) eye study: Quantifying ethnic variations in the epidemiology of eye diseases in Asians. *Ophthalmic Epidemiol.* **16**, 325–336 (2009).

85. R. Dorajoo, A. I. F. Blakemore, X. Sim, R. T.-H. Ong, D. P. K. Ng, M. Seielstad, T.-Y. Wong, S.-M. Saw, P. Froguel, J. Liu, E.-S. Tai, Replication of 13 obesity loci among Singaporean Chinese, Malay and Asian-Indian populations. *Int. J. Obes.* **36**, 159–163 (2012).

Science Advances

Genome-wide association study in almost 195,000 individuals identifies 50 previously unidentified genetic loci for eye color

Mark Simcoe, Ana Valdes, Fan Liu, Nicholas A. Furlotte, David M. Evans, Gibran Hemani, Susan M. Ring, George Davey Smith, David L. Duffy, Gu Zhu, Scott D. Gordon, Sarah E. Medland, Dragana Vuckovic, Giorgia Girotto, Cinzia Sala, Eulalia Catamo, Maria Pina Concas, Marco Brumat, Paolo Gasparini, Daniela Toniolo, Massimiliano Cocca, Antonietta Robino, Seyhan Yazar, Alex Hewitt, Wenting Wu, Peter Kraft, Christopher J. Hammond, Yuan Shi, Yan Chen, Changqing Zeng, Caroline C. W. Klaver, Andre G. Uitterlinden, M. Arfan Ikram, Merel A. Hamer, Cornelia M. van Duijn, Tamar Nijsten, Jiali Han, David A. Mackey, Nicholas G. Martin, Ching-Yu Cheng, the 23andMe Research Team, the International Visible Trait Genetics Consortium, David A. Hinds, Timothy D. Spector, Manfred Kayser and Pirro G. Hysi

ARTICLE TOOLS	http://advances.sciencemag.org/content/7/11/eabd1239
SUPPLEMENTARY MATERIALS	http://advances.sciencemag.org/content/suppl/2021/03/08/7.11.eabd1239.DC1
REFERENCES	This article cites 85 articles, 6 of which you can access for free http://advances.sciencemag.org/content/7/11/eabd1239#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service