Loci associated with skin pigmentation identified in African populations

Nicholas G. Crawford[1], **Derek E. Kelly**[1,2,*], **Matthew E. B. Hansen**[1,*], **Marcia H. Beltrame**[1,*], **Shaohua Fan**[1,*], **Shanna L. Bowman**[3,4,*], **Ethan Jewett**[5,6,*], **Alessia Ranciaro**[1], **Simon Thompson**[1], **Yancy Lo**[1], **Susanne P. Pfeifer**[7], **Jeffrey D. Jensen**[7], **Michael C. Campbell**[1,8], **William Beggs**[1], **Farhad Hormozdiari**[9,10], **Sununguko Wata Mpoloka**[11], **Gaonyadiwe George Mokone**[12], **Thomas Nyambo**[13], **Dawit Wolde Meskel**[14], **Gurja Belay**[14], **Jake Haut**[1], **NISC Comparative Sequencing Program**[†], **Harriet Rothschild**[15], **Leonard Zon**[15,16], **Yi Zhou**[15,17], **Michael A. Kovacs**[18], **Mai Xu**[18], **Tongwu Zhang**[18], **Kevin Bishop**[19], **Jason Sinclair**[19], **Cecilia Rivas**[20], **Eugene Elliot**[20], **Jiyeon Choi**[18], **Shengchao A. Li**[21,22], **Belynda Hicks**[21,22], **Shawn Burgess**[19], **Christian Abnet**[21], **Dawn E. Watkins-Chow**[20], **Elena Oceana**[23], **Yun S. Song**[5,6,24,25,26], **Eleazar Eskin**[27], **Kevin M. Brown**[18], **Michael S. Marks**[3,4,‡], **Stacie K. Loftus**[20,‡], **William J. Pavan**[20,‡], **Meredith Yeager**[21,22,‡], **Stephen Chanock**[21,‡], and **Sarah Tishkoff**[1,25,§]

[1]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

[2]Genomics and Computational Biology Graduate Program, University of Pennsylvania, Philadelphia, PA 19104, USA

[3]Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia Research Institute, Philadelphia, PA 19104, USA

[4]Department of Pathology and Laboratory Medicine and Department of Physiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

[5]Department of Electrical Engineering and Computer Sciences, University of California–Berkeley, Berkeley, CA 94704, USA

[6]Department of Statistics, University of California–Berkeley, Berkeley, CA 94704, USA

[7]School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA

§Corresponding author. tishkoff@mail.med.upenn.edu.
*These authors contributed equally to this work.
†NISC Comparative Sequencing Program collaborators and affiliations are listed in the supplementary materials.
‡These authors contributed equally to this work.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/cgi/content/full/science.aan8433/DC1)

Materials and Methods
Figs. S1 to S21
Tables S1 to S8
NISC Comparative Sequencing Program Collaborator List
References (103–130)

[8]Department of Biology, Howard University, Washington, DC 20059, USA

[9]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

[10]Program in Medical and Population Genetics, Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard, Cambridge, MA 02142, USA

[11]Department of Biological Sciences, University of Botswana, Gaborone, Botswana

[12]Department, of Biomedical Sciences, University of Botswana School of Medicine, Gaborone, Botswana

[13]Department of Biochemistry, Muhimbili University of Health and Allied Sciences, Dares Salaam, Tanzania

[14]Department of Biology, Addis Ababa University, Addis Ababa, Ethiopia

[15]Stem Cell Program, Division of Hematology and Oncology, Pediatric Hematology Program, Boston Children's Hospital and Dana Farber Cancer Institute, Harvard Medical School, Boston, MA 02115, USA

[16]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA

[17]Harvard Stem Cell Institute, Harvard University, Cambridge, MA 02138, USA

[18]Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

[19]Translational and Functional Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

[20]Genetic Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

[21]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD 20892, USA

[22]Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick, MD 21701, USA

[23]Department of Molecular Pharmacology, Physiology and Biotechnology, Brown University, Providence, RI 02912, USA

[24]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA

[25]Department of Biology, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA 19104, USA

[26]Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104, USA

[27]Department of Computer Science, Department of Human Genetics, University of California–Los Angeles, Los Angeles, CA 90095, USA

Abstract

Despite the wide range of skin pigmentation in humans, little is known about its genetic basis in global populations. Examining ethnically diverse African genomes, we identify variants in or near *SLC24A5, MFSD12, DDB1, TMEM138, OCA2* and *HERC2* that are significantly associated with skin pigmentation. Genetic evidence indicates that the light pigmentation variant at *SLC24A5* was introduced into East Africa by gene flow from non-Africans. At all other loci, variants associated with dark pigmentation in Africans are identical by descent in southern Asian and Australo-Melanesian populations. Functional analyses indicate that *MFSD12* encodes a lysosomal protein that affects melanogenesis in zebrafish and mice, and that mutations in melanocyte-specific regulatory regions near *DDB1/TMEM138* correlate with expression of UV response genes under selection in Eurasians.

Variation in epidermal pigmentation is a striking feature of modern humans. Human pigmentation is correlated with geographic and environmental variation (Fig. 1). Populations at lower latitudes have darker pigmentation than populations at higher latitudes, suggesting that skin pigmentation is an adaptation to differing levels of ultraviolet radiation (UVR) (1). Because equatorial regions receive more UVR than temperate regions, populations from these regions (including sub-Saharan Africans, South Asians, and Australo-Melanesians) have darker pigmentation (Fig. 1), which likely mitigates the negative impact of high UVR exposure such as skin cancer and folate degradation (1). In contrast, the synthesis of vitamin D_3 in response to UVR, needed to prevent rickets, may drive selection for light pigmentation at high latitudes (1).

The basal layer of human skin contains melanocytes, specialized pigment cells that harbor subcellular organelles called melanosomes in which melanin pigments are synthesized and stored and then transferred to keratinocytes (2). Melanosome morphology and content differs between melanocytes that synthesize mainly eumelanins (black-brown pigments) or pheomelanins (pigments which range from yellow to reddish-brown) (3). Variation in skin pigmentation is due to the type and quantity of melanins generated, melanosome size, and the manner in which keratinocytes sequester and degrade melanins (4).

While over 350 pigmentation genes have been identified in animal models, only a subset of these genes have been linked to normal variation in humans (5). Of these, there is limited knowledge about loci that affect pigmentation in populations with African ancestry (6, 7).

Skin pigmentation is highly variable within Africa

To identify genes affecting skin pigmentation in Africa, we used a DSM II ColorMeter to quantify light reflectance from the inner arm as a proxy for melanin levels in 2,092 ethnically and genetically diverse Africans living in Ethiopia, Tanzania, and Botswana (table S1 and figs. S1 and S2) (8). Skin pigmentation levels vary extensively among Africans, with darkest pigmentation observed in Nilo-Saharan speaking pastoralist populations in Eastern Africa and lightest pigmentation observed in San hunter-gatherer populations from southern Africa (Fig. 2 and table S1).

A locus associated with light skin color in Europeans is common in East Africa

We genotyped 1,570 African individuals with quantified pigmentation levels using the Illumina Infinium Omni5 Genotyping array. After quality control, we retained ~4.2 million biallelic single nucleotide polymorphisms (SNPs) for analysis. A GWAS analysis with linear mixed models, controlling for age, sex, and genetic relatedness (9), identified four regions with multiple significant associations (p-value $< 5 \times 10^{-8}$) (Fig. 1, fig. S3, and tables S2 and S3).

We then performed fine-mapping using local imputation of high coverage sequencing data from a subset of 135 individuals and data from the Thousand Genomes Project (TGP) (10) (Fig. 3 and table S3). We ranked potential causal variants within each locus using CAVIAR, a fine-mapping method that accounts for linkage disequilibrium (LD) and effect sizes (11) (Table 1). We characterized global patterns of variation at these loci using whole genome sequences from West African, Eurasian and Australo-Melanesian populations (10,12,13).

The SNPs with strongest association with skin color in Africans were on chromosome 15 at or near the *Solute Carrier Family 24 Member 5 (SLC24A5)* gene (Figs. 1 and 3 and tables S2 and S3). A functional non-synonymous mutation within *SLC24A5* (rsl426654) (14) was significantly associated with skin color (F-test, p-value $= 5.5 \times 10^{-62}$) and was identified as potentially causal by CAVIAR (Table 1). The rsl426654 (A) allele is at high frequency in European, Pakistani, and Indian populations (Fig. 1) and is a target of selection in Europeans, Central Asians and North Indians (15-17). In Africa this variant is common (28-50% frequency) in populations from Ethiopia and Tanzania with high Afroasiatic ancestry (18, 19) and is at moderate frequency (5-11%) in San and Bantu populations from Botswana with low levels of East African ancestry and recent European admixture (20, 21) (Fig. 1 and figs. S2 and S4). We observe a signature consistent with positive selection at *SLC24A5* in Europeans on the basis of extreme values of Tajima's D statistic (fig. S5).

Based on coalescent analysis with sequence data from the Simons Genomic Diversity Project (SGDP) (13), the time to most recent common ancestor (TMRCA) of most Eurasian lineages containing the rsl426654 (A) allele is 29 kya (95% credible interval (CI) 28-31 kya), consistent with prior studies (6, 16) (Fig. 4). Haplotype analysis indicates that the rsl426654 (A) variant in Africans is on the same extended haplotype background as Europeans (Fig. 5 and fig. S6), likely reflecting gene flow from western Eurasia over at least the past 3-9 ky (22). The rsl426654 (A) variant is at high frequency (28%) in Tanzanian populations, suggesting a lower bound (~*5* kya) for introduction of this allele into East Africa, the time of earliest migration from Ethiopia into Tanzania (23). Further, the frequency of the rsl426654 (A) variant in eastern and southern Africans exceeds the inferred proportion of non-African ancestry (figs. S2 and S4). Estimates of genetic differentiation (F_{ST}) at the rsl426654 SNP between the West African Yoruba (YRI) and Ethiopian Amhara populations is 0.76, among the top 0.01% of values on chromosome 15 (table S4). These results are consistent with selection for the rsl426654 (A) allele in African populations following introduction, although complex models of demographic history cannot be ruled out.

A lysosomal transporter protein associated with skin pigmentation

The region with the second strongest genetic association with skin pigmentation contains the *Major Facilitator Superfamily Domain Containing 12 (MFSD12)* gene on chromosome 19 (Figs. 1 and 3 and tables S2 and S3). *MFSD12* is homologous to other genes containing MFS domains, conserved throughout vertebrates, which function as transmembrane solute transporters (24). *MFSD12* mRNA levels are low in de-pigmented skin of vitiligo patients (25), likely due to autoimmune related destruction of melanocytes.

The *MFSD12* locus is in a region with extensive recombination, enabling us to fine-map eight potentially causal SNPs (Table 1 and table S3) that cluster in two regions: one within *MFSD12* and the other ~7,600–9,000 bp upstream of *MFSD12* (Fig. 3). Many SNPs are in predicted regulatory regions active in melanocytes and/or keratinocytes (Table 1 and Fig. 3) and show enhancer activity in luciferase expression assays in a WM88 melanoma cell line (Table 1, table S5, and fig. S7). Within *MFSD12*, the two SNPs that CAVIAR identifies as having highest probability of being causal are rs56203814 (F-test, p-value = 3.6×10^{-18}), a synonymous variant within exon 9, and rsl0424065 (F-test, p-value = 5.1×10^{-20}), located within intron 8. They are 130 bp apart, in strong LD, and impact gene expression in luciferase expression assays ($1.5 - 2.7 \times$ higher expression than the minimal promoter; fig. S7). The SNPs upstream of *MFSD12* with highest probability of being causal are rsll2332856 (F-test, p-value = 3.8×10^{-16}) and rs6510760 (F-test, p-value = 6.5×10^{-15}). They are 346 bp apart, in strong LD, and impact gene expression in luciferase expression assays ($4.0 - 19.7 \times$ higher expression than the minimal promoter; fig. S7).

The derived rs56203814 and rsl0424065 (T) alleles associated with dark pigmentation are present only in African populations (or those of recent African descent) and are most common in East African populations with Nilo-Saharan ancestry (Fig. 1 and fig. S4). Coalescent analysis of the SGDP dataset indicates that the rsl0424065 (T) allele predates the 300 kya origin of modern humans (26) (estimated TMRCA of 612 kya, 95% CI 515-736 kya) (Fig. 4).

At rs6510760 and rsll2332856, the ancestral (G) and (T) alleles, respectively, associated with light pigmentation, are nearly fixed in Europeans and East Asians and are common in San as well as Ethiopian and Tanzanian populations with Afroasiatic ancestry (Fig. 1 and fig. S4). The derived rs6510760 (A) and rs112332856 (C) alleles (associated with dark pigmentation) are common in all sub-Saharan Africans except the San, as well as in South Asian and Australo-Melanesian populations (Fig. 1 and fig. S4). Haplotype analysis places the rs6510760 (A) allele (and linked rsll2332856 (C) allele) in Australo-Melanesians on similar haplotype backgrounds relative to central and eastern Africans (Fig. 5 and fig. S6), suggesting they are identical by descent from an ancestral African population. Coalescent analysis of the SGDP dataset indicates that the TMRCA for the derived rs6510760 (A) allele is 996 kya (95% CI 0.82-1.2 mya; Fig. 4).

We do not detect evidence for positive selection at *MFSD12* using Tajima's D and iHS statistics [figs. S5 and S8, as expected if selection were ancient (27)]. However, levels of genetic differentiation are elevated when comparing East African Nilo-Saharan and western

European (CEU) populations (e.g., F_{ST}=0.85 for rsll2332856, top 0.05% on chromosome 19), consistent with differential selection at this locus (28) (table S4).

MFSD12 is within a cluster of 10 genes with high expression levels in primary human melanocytes relative to primary human keratinocytes (29), with *MFSD12* the most differentially expressed (90X; table S6). The genomic region (chr19:3541782-3581062) encompassing *MFSD12* and neighboring gene *HMG20B* (a transcription factor common in melanocytes) has numerous DNase I hypersensitive sites and is enriched for H3K27ac enhancer marks in melanocytes (top 0.1% genome-wide; Fig. 3), suggesting this region may regulate expression of genes critical to melanocyte function (30).

Analyses of gene expression using RNA-sequencing data from 106 primary melanocyte cultures (table S7), indicates that African ancestry is significantly correlated with decreased *MFSD12* gene expression (Pearson Correlation Coefficient (PCC) p-value = 5.0×10^{-2}; fig. S9). We observed significant associations between genotypes at rs6510760 and rsll2332856 with expression of *HMG20B* (Bonferroni adjusted p-value (P_{adj}) $< 4.9 \times 10^{-3}$) and *MFSD12* ($P_{adj} < 3.4 \times 10^{-2}$) (fig. S9). In each case, the alleles associated with dark pigmentation correlate with decreased gene expression. Allele-specific expression (ASE) analysis indicates that individuals heterozygous for either rs6510760 or rsll2332856 show increased allelic imbalance, relative to homozygotes, for *MFSD12* (Mann-Whitney-Wilcoxon (MWW) test, p-value = 4.9×10^{-3} and 1.3×10^{-2}, respectively), consistent with regulation of gene expression in *cis.* A haplotype containing the rs6510760(A)/rsll2332856(C) variants associated with dark pigmentation showed 4.9 times lower expression in luciferase assays than the haplotype containing rs6510760(G)/rs112332856(T) variants associated with light pigmentation [Kruskal-Wallis Rank Sum (KWRS) Test, p- value=7.7×10^{-7}; fig. S7 and table S5]. We did not have power to detect an association between expression of *MFSD12* and rs56203814 or rsl0424065 due to low frequency (~2%) of the alleles associated with dark pigmentation in the primary melanocyte cultures.

MFSD12 suppresses eumelanin biogenesis in melanocytes from lysosomes

We silenced expression of the mouse ortholog of MFSD12 *(Mfsdl2)* using small hairpin RNAs (shRNAs) in immortalized melan-Ink4a mouse melanocytes derived from C57BL/6J-Ink4a$^{-/-}$ mice (31) which almost exclusively make eumelanin (Fig. 6). Reduction of *Mfsd12* mRNA by ~80% with two distinct lentivirally encoded shRNAs (Fig. 6A) caused a 30-50% increase in melanin content compared to control cells (Fig. 6B), and a higher percentage of melanosomes per total cell area in most cells compared to cells transduced with nontarget shRNA (Fig. 6, C and D). A fraction of MFSD12-depleted cells harbored large clumps of melanin in autophagosome-like structures (fig. S10). These data suggest that MFSD12 suppresses eumelanin content in melanocytes and may offset autophagy.

We assessed the localization of human MFSD12 isoform c (RefSeq NM_174983.4) tagged at the C terminus with the HA epitope (MFSD12-HA). By immunofluorescence microscopy, MFSD12-HA localized to punctate structures throughout the cell. Surprisingly, these puncta, like those labeled by the endogenous lysosomal membrane protein LAMP2, but not the melanosomal enzyme TYRP1, overlapped only weakly with pigmented melanosomes (Fig.

6, E to G; quantified in Fig. 6H). Instead, MFSD12-HA co-localized with LAMP2 (Fig. 6E, quantified in Fig. 6H), indicating that MFSD12 protein localizes to late endosomes and/or lysosomes in melanocytes and not to eumelanosomes.

MFSD12 influences pigmentation in zebrafish xanthophore pigment cells

We targeted transmembrane domain 2 (TMD2) in the highly conserved zebrafish ortholog of *mfsdl2a* with CRISPR/Cas9 (Fig. 7). We focused on *mfsd12a* as its paralog *mfsd12b* is predicted to be a pseudogene (32). Although pigmentation was not qualitatively altered in melanophores, the cells that make eumelanin, compound heterozygotes of *mfsd12a* alleles exhibited reduced staining of xanthophores, the cells responsible for pteridine-based yellow pigmentation in wild type zebrafish (Fig. 7, A and B) (33, 34). This was not due to a failure of the xanthophores to develop in *mfsd12a* mutants, since GFP labeled xanthophores were robust along the lateral line in both wild-type and *mfsd12a* mutant zebrafish (Fig. 7, C and D). Taken together, these results suggest that *MFSD12* influences xanthophore pigment production in pterinosomes.

Functional characterization of *MFSD12* in mice

CRISPR/Cas9 was used to generate a *Mfsd12* null allele in a wild-type mouse background (Fig. 7E and fig. S11). Four founders were observed with a uniformly gray coat color, rather than the expected agouti coat color (fig. S11, A and B). These four gray founders harbored deletions at the targeted site (fig. S11C). Microscopic observation revealed a lack of pheomelanin, resulting in white, rather than yellow, banding of hairs in *Mfsd12* mutants (Fig. 7F).

The *Mfsd12* knockout coat color appeared phenotypically similar to that of *grizzled (gr)* mice, an allele previously mapped to a syntenic ~2 Mb region overlapping *Mfsd12* (35). Like our CRISPR/Cas9 *Mfsd12* knockout, homozygous *gr/gr* mice are characterized by a gray coat resulting from dilution of yellow pheomelanin pigment from the sub-terminal agouti band of the hair shaft. Exome sequencing of an archived *gr/gr* DNA sample, subsequently confirmed by Sanger sequencing in an independent colony, identified a 9 bp in-frame deletion within exon2 of *Mfsd12* (fig. S12) as the sole mutation affecting a coding sequence in this mapped candidate region. The deleted amino acids for the *gr/gr* allele, *Mfsd12* p.Leu163_Ala165del, are in the cytoplasmic loop between the transmembrane domains TM4 and TM5 within a highly conserved MFS domain (fig. S13). These results indicate that mutation of *Mfsd12* is responsible for the gray coat color of *gr/gr* mutant mice, and that loss of *Mfsd12* reduces pheomelanin within the hairs of agouti mice.

Taken together, these results indicate that *MFSD12* plays a conserved role in vertebrate pigmentation. Depletion of MFSD12 increases eumelanin content in a cell-autonomous manner in skin melanocytes, consistent with the lower levels of *MFSD12* expression observed in melanocytes from individuals with African ancestry. Since MFSD12 localizes to lysosomes and not to eumelanosomes, this may reflect an indirect effect through modified lysosomal function. By contrast, loss of MFSD12 has the opposite effect on pheomelanin production, reflecting a more direct effect on function of pheomelanosomes, which have a

distinct morphology (3), gene expression profile (36), and, like zebrafish pterinosomes, a potentially different intracellular origin from eumelanosomes (37). While disruption of *MFSD12* alone accounts for changes in pigmentation, the role of neighboring loci such as *HMG20B* on pigmentation remains to be explored.

Skin pigmentation associated loci that play a role in UV response are targets of selection

Another genomic region associated with pigmentation encompasses a ~195 kb cluster of genes on chromosome 11 that play a role in UV response and melanoma risk including the *Damage Specific DNA Binding Protein 1 (DDB1)* gene (Figs. 1 and 3 and table S3). DDB1 (complexed with DDB2 and XPC) functions in DNA repair (38); levels of DDB1 are regulated by UV exposure and MC1R signaling, a regulatory pathway of pigmentation (39). DDB1 is a component of CUL4-RING E3 ubiquitin ligases that regulate several cellular and developmental processes (40); it is critical for follicle maintenance and female fertility in mammals (41) and for plastid size and fruit pigmentation in tomatoes (42). Knockouts of *DDB1* orthologs are lethal in both mouse and fruit fly development (43), and *DDB1* only exhibits rare (<1% frequency) non-synonymous mutations in the TGP dataset. Genetic variants near *DDB1* were associated with human pigmentation in an African population with high levels of European admixture (7).

Due to extensive LD in this region, CAVIAR identified 33 SNPs predicted to be causal (Table 1). The most strongly associated SNPs are located in a region conserved across vertebrates flanked by *TMEM138* and *TMEM216* (44) ~36-44 kb upstream of *DDB1*, and are in high LD within this cluster ($r^2 > 0.7$ in East Africans) (Fig. 3, Table 1, and table S3). Among these, the most significantly associated SNP is rs7948623 (F-test, p-value = 2.2×10^{-11}), located 172 bp downstream of *TMEM138*, which shows enhancer activity in WM88 melanoma cells (91.9 - 140.8 × higher than the minimal promoter; fig. S7 and table S5).

A second group of tightly linked SNPs (LD $r^2 > 0.7$ in East Africans) with predicted high probability of containing causal variants spans a ~195 kb region encompassing *DDB1* and *TMEM138* (Table 1 and Fig. 3). Two SNPs that tag this LD block are rs1377457 (F-test, p-value = 1.5×10^{-9}), located ~7600 bp downstream of *TMEM138*, and rs148172827 (F-test, p-value = 1.8×10^{-9}), an insertion/deletion polymorphism at *TKFC* (Triokinase And FMN Cyclase) located in an enhancer active in WM88 melanoma cells (67.6 - 76.2 × higher than the minimal promoter; fig. S7 and table S5) which overlaps a *MITF* binding site in melanocytes (29, 45); both SNPs interact with the promoters of *DDB1* and neighboring genes in MCF-7 cells (46, 47) (Table 1 and Fig. 3). SNPs within introns of *DDB1* (rs12289370, rs7934735, rs11230664, rs12275843, and rs7120594) also tag this LD block (Table 1 and Fig. 3).

RNA-Seq data from 106 primary melanocyte cultures indicate that African ancestry is significantly correlated with increased *DDB1* gene expression (PCC, p-value = 2.6×10^{-5}; fig. S9). Association tests using a permutation approach indicated that, of the 35 protein-coding genes with a transcription start site within 1Mb of rs7948623, expression of *DDB1* is most strongly associated with a SNP in an intron of *DDB1*, rs7120594, at marginal

statistical significance after correction for ancestry and multiple testing (P_{adj} = 0.06; fig. S9). The allele associated with dark pigmentation at rs7120594 correlates with increased *DDB1* expression. We did not have power to detect an association between expression of *DDB1* and SNPs in LD with rs7948623 due to low minor allele frequencies (~2%). The role of *DDB1* and neighboring loci on human pigmentation remains to be further explored.

The derived rs7948623 (T) allele near *TMEM138* (associated with dark pigmentation) is most common in East African Nilo-Saharan populations and is at moderate to high frequency in South Asian and Australo-Melanesian populations (Fig. 1 and fig. S4). At SNP rs11230664, within *DDB1*, the ancestral (C) allele (associated with dark pigmentation) is common in all sub-Saharan African populations, having the highest frequency in East African Nilo-Saharan, Hadza, and San populations (88-96%), and is at moderate to high frequency in South Asian and Australo-Melanesian populations (12 - 66%) (Fig. 1 and fig. S4). The derived (T) allele (associated with light pigmentation) is nearly fixed in European, East Asian, and Native American populations.

In South Asians and Australo-Melanesians, the alleles associated with darker pigmentation reside on closely related, or identical, haplotypes to those observed in Africa (Fig. 5 and fig. S6), suggesting that they are identical by descent. The TMRCAs for the derived dark allele at rs7948623 and the derived light allele at rs11230664 are estimated to be older than 600 kya and 250 kya, respectively (Fig. 4).

Consistent with a selective sweep, we see an excess of rare alleles (and extreme negative Tajima's D values) and high levels of homozygosity extending ~350-550 kb in Europeans and Asians, respectively (figs. S5 and S14). We observe extreme negative Tajima's D values in East African Nilo-Saharans and San over a shorter distance (115 kb and 100 kb, respectively) (fig. S5). A haplotype extending greater than 195 kb is common in Eurasians and rare in Africans (Fig. 5) and tags the alleles associated with light skin pigmentation. The TMRCA of a large number of haplotypes carrying the rs7948623 (A) allele in non-Africans, associated with light pigmentation, is 60 kya (95% CI: 58-62 kya), close to the inferred time of the migration of modern humans out of Africa (48, 49) (Fig. 4). These results, combined with large F_{ST} values between Africans and Europeans at SNPs tagging the extended haplotype near *DDB1* (e.g., F_{ST} = 0.98 between Nilo-Saharans and CEU at rs7948623, within the top 0.01% of values on chromosome 11, table S4) are consistent with differential selection of alleles associated with light and dark pigmentation in Africans and non-Africans at this locus.

Identification of variation at *OCA2* and *HERC2* impacting skin pigmentation

Another region of significantly associated SNPs encompasses the *OCA2* and *HERC2* loci on chromosome 15 (Fig. 3 and table S3). *HERC2* was identified in GWAS for eye, hair, and skin pigmentation traits (6, 7, 50-52). The oculocutaneous albinism II gene (*OCA2*, formerly called the P gene) encodes a 12- transmembrane domain-containing chloride transporter protein and affects pigmentation by modulating melanosomal pH (53). The most common types of albinism in Africans are caused by mutations in *OCA2* (54).

Due to extensive LD in the *OCA2* and *HERC2* region, CAVIAR predicted 10 potentially causal SNPs (Table 1) that cluster within three regions. We order these clusters on the basis of physical distance; region 1 is located within *OCA2*, and regions 2 and 3 are located within introns of *HERC2* (Fig. 3).

The SNP with highest probability of being causal from CAVIAR analysis is rs1800404 (F-test, p-value = 1.0), a synonymous variant located in region 1 within exon 10 of *OCA2* (Fig. 3, Table 1, and table S3). The ancestral rs1800404 (C) allele, associated with dark pigmentation, is common in most Africans as well as southern/eastern Asians and Australo-Melanesians, whereas the derived (T) allele, associated with light pigmentation, is most common (frequency >70%) in Europeans and San (Fig. 1 and fig. S4). Haplotype (Fig. 5) and coalescent analyses (Fig. 4 and fig. S6) show two divergent clades, one enriched for the rs1800404 (C) allele and the other for the rs1800404 (T) allele. Coalescent analysis indicates that the TMRCA of all lineages is 1.7 mya (95% CI: 1.52.0 mya) and the TMRCA of lineages containing the derived (T) allele is 629 kya (95% CI 426-848 kya) (Fig. 4). The deep coalescence of lineages, and the positive Tajima's *D* values in this region in both African and non-African populations (fig. S5), is consistent with balancing selection acting at this locus.

The SNP with highest probability of being causal in region 3 is rs4932620 (F-test, p-value = 3.2×10^{-9}) located within intron 11 of *HERC2* (Fig. 3, Table 1, and table S3). This SNP is 917 bp from rs916977, a SNP associated with blue-eye color in Europeans (55, *56*), and is in strong LD ($r^2 = 1.0$ in most East African populations) with SNPs extending into region 2 of *HERC2* (Table 1). The derived rs4932620 (T) allele associated with dark skin pigmentation is most common in Ethiopian populations with high levels of Nilo-Saharan ancestry and is at moderate frequency in other Ethiopian, Hadza, and Tanzania Nilo-Saharan populations (Fig. 1 and fig. S4). Haplotype analysis indicates that the rs4932620 (T) allele in South Asians and Australo-Melanesians is on the same or similar haplotype background as in Africans (Fig. 5 and fig. S6), suggesting it is identical by descent. The TMRCA of haplotypes containing the rs4932620 (T) allele is 247 kya (95% CI: 158-345 kya) (Fig. 4).

We also observe an LD block of SNPs within *HERC2* that are associated with skin pigmentation independently of the SNPs described above, though they do not reach genome-wide significance (table S3). These are in a region with enhancer activity in Europeans (50). For example, SNP rs6497271 (F-test, p-value = 1.8×10^{-6}), which is located 437 bp from SNP rs12913832, has been associated with skin color in Europeans (50) and is in a consensus SOX2 motif (a transcription factor which modulates levels of *MITF* in melanocytes) (57) (Fig. 3). The ancestral rs6497271 (A) allele associated with dark pigmentation is on haplotypes in South Asians and Australo-Melanesians similar or identical to those in Africans (Fig. 5 and fig. S6), suggesting they are identical by descent. The derived (G) allele associated with light skin pigmentation is most common in Europeans and San and dates to 921 kya (CI: 700 kya-1.3 mya) (Figs. 1 and 4 and figs. S4 and S6). SNPs associated with pigmentation at all three regions show high allelic differentiation when comparing East African Nilo-Saharans and CEU ($F_{ST} = 0.72 - 0.85$, top 0.5% on chromosome 15) (table S4).

Analyses of RNA-Seq data from 106 primary melanocyte cultures indicates that African ancestry is significantly correlated with increased *OCA2* gene expression (PCC, $P_{adj} = 6.1 \times 10^{-7}$) (fig. S9). A permutation approach identified significant associations between *OCA2* expression and SNPs within an LD block tagged by rs4932620 extending across regions 2 and 3 ($P_{adj} = 2.2 \times 10^{-2}$). Alleles in this LD block associated with dark pigmentation correlate with increased *OCA2* expression. We did not observe associations between the candidate causal variants in region 1 and *OCA2* expression despite a high minor allele frequency (34%). However, we observe a significant association between a haplotype tagged by rs1800404 and alternative splicing resulting in inclusion/exclusion of exon 10 (linear regression *t* test, p-value = 9.1×10^{-40}). Exon 10 encodes the amino acids encompassing the third transmembrane domain of *OCA2* and is the location of several albinism-associated *OCA2* mutations (58, 59), raising the possibility that the shorter transcript encodes a non-functional channel. Comparing splice junction usage across individuals, we estimate that each additional copy of the light rs1800404 (T) allele reduces inclusion of exon 10 by ~20% (95% CI 17.9-21.5%; fig. S9). Homozygotes for the light rsl800404 (T) allele are, therefore, expected to produce ~60% functional OCA2 protein (compared to individuals with albinism who produce no functional OCA2 protein).

Skin pigmentation is a complex trait

To estimate the proportion of pigmentation variance explained by the top eight candidate SNPs at *SLC24A5, MFSD12, DDB1/TMEM138* and *OCA2/HERC2*, we used a linear mixed model with two genetic random effect terms, one based on the genome-wide kinship matrix, and the other based on the kinship matrix derived from the set of significant variants. Approximately 28.9% (S.E. 10.6%) of the pigmentation variance is attributable to these SNPs. Considering each locus in turn and all significantly associated variants (p- value $< 5 \times 10^{-8}$), the trait variation attributable to each locus is: *SLC24A5* (12.8%, S.E. 3.5%), *MFSD12* (4.5%, S.E. 2.1%), *DDB1/TMEM138* (2.2%, S.E. 1.5%), and *OCA2/HERC2* (3.9%, S.E. 2.9%). Thus ~29% of the additive heritability of skin pigmentation in Africans is due to variation at these four regions. This observation indicates that the genetic architecture of skin pigmentation is simpler (i.e., fewer genes of stronger effect) than other complex traits, such as height (60). Additionally, most candidate causal variants are in non-coding regions, indicating the importance of regulatory variants influencing skin pigmentation phenotypes.

Evolution of skin pigmentation in modern humans

Skin pigmentation is highly variable within Africa. Populations such as the San from southern Africa are almost as lightly pigmented as Asians (1), while the East African Nilo-Saharan populations are the most darkly pigmented in the world (Fig. 1). Most alleles associated with light and dark pigmentation in our dataset are estimated to have originated prior to the origin of modern humans ~300 ky ago (26). In contrast to the lack of variation at *MC1R*, which is under purifying selection in Africa (61), our results indicate that both light and dark alleles at *MFSD12, DDB1, OCA2,* and *HERC2* have been segregating in the hominin lineage for hundreds of thousands of years (Fig. 4). Further, the ancestral allele is associated with light pigmentation in approximately half of the predicted causal SNPs;

Neanderthal and Denisovan genome sequences, which diverged from modern human sequences 804 kya (62), contain the ancestral allele at all loci. These observations are consistent with the hypothesis that darker pigmentation is a derived trait that originated in the genus *Homo* within the past ~2 million years after human ancestors lost most of their protective body hair, though these ancestral hominins may have been moderately, rather than darkly, pigmented (63, 64). Moreover, it appears that both light and dark pigmentation has continued to evolve over hominid history.

Individuals from South Asia and Australo-Melanesia share variants associated with dark pigmentation at *MFSD12, DDB1/TMEM138, OCA2* and *HERC2* that are identical by descent from Africans. This raises the possibility that other phenotypes shared between Africans and some South Asian and Australo-Melanesian populations may also be due to genetic variants identical by descent from African populations rather than convergent evolution (65). This observation is consistent with a proposed southern migration route out of Africa ~80 kya (66). Alternatively, it is possible that light and dark pigmentation alleles segregated in a single African source population (13, 48) and that alleles associated with dark pigmentation were maintained outside of Africa only in the South Asian and Australo-Melanesian populations due to selection.

By studying ethnically, genetically, and phenotypically diverse Africans, we identify novel pigmentation loci that are not highly polymorphic in other populations. Interestingly, the loci identified in this study appear to affect multiple phenotypes. For example, DDB1 influences pigmentation (42), cellular response to the mutagenic effect of UVR (39) and female fertility (41). Thus, some of the pigmentation-associated variants identified here may be maintained due to pleiotropic effects on other aspects of human physiology.

It is important to note that genetic variants that do not reach genome-wide significance in our study might also impact the pigmentation phenotype. Indeed, the 1000 most strongly associated SNPs exhibit enrichment for genes involved in pigmentation and melanocyte physiology in the mouse phenotype database and in ion transport and pyrimidine metabolism in humans (table S8). Future research in larger numbers of ethnically diverse Africans may reveal additional loci associated with skin pigmentation and will further shed light on the evolutionary history, and adaptive significance, of skin pigmentation in humans.

Materials and Methods

Individuals in the study were sampled from Ethiopia, Tanzania and Botswana. Written informed consent was obtained from all participants, and research/ethics approval and permits were obtained from all relevant institutions (67). To measure skin pigmentation we used a DSM II ColorMeter to quantify reflectance from the inner under arm. Red reflectance values were converted to a standard melanin index score (68). DNA was extracted from whole blood using a salting out procedure (PureGen).

A total of 1,570 samples were genotyped on the Illumina Omni5M SNP array (5M dataset) that includes ~4.5 million SNPs. Genotypes were clustered and called in *Genome Studio* software. Variant positions are reported in hg19/37 coordinates. The overall completion rate

was 98.8%. Each individual's sex was verified based on X chromosome inbreeding coefficients. We used Beagle 4.0 (69) to phase the Illumina 5M SNP array data merged with SNPs from the TGP dataset that were filtered to exclude related individuals.

High coverage (>30 X) Illumina Sequencing was performed on a subset of the genotyped individuals (N=135). Variants were called following the approach described in (13). Adapter sequences were trimmed with *trimadap.* Reads were aligned using *bwa mem* to the human reference sequence build 37 (hg19). After alignment we marked duplicate reads prior to calling variants with *GATK HaplotypeCaller* (70). To select high quality variants we employed a two-set filtering strategy. First, we used the GATK variant quality score recalibration to score variant sites. We used TGP, OMIM, and our curated genotypes from the Illumina Omni 5M SNP array as training data. After recalibration we discarded sites with the lowest scores. In addition, we discarded sites in low-complexity regions listed in (71) and duplicate regions identified with *Delly* (72).

We performed local imputation around each of the regions showing significant associations with skin pigmentation from GWAS using the Illumina Omni 5M SNP dataset. We extracted array genotypes within 1Mb (500 kb upstream and 500 kb downstream) of top GWAS variants from each region and phased them using *SHAPEIT2* (73, 74). The reference panel came from two datasets: filtered variants from the 135 African genomes and TGP (10). After phasing, imputation was performed using *Minimac3* (75). Imputation performed very well at most loci ($R^2 > 0.91$ with MAF 0.05) (table S3).

To identify SNPs associated with pigmentation, GWAS was performed first on the Illumina Omni 5M SNP dataset, and independently with imputed variants at candidate regions, using linear mixed models implemented in *EMMAX* software (9). Age and sex were included as covariates, and we corrected for genetic relatedness with an IBS kinship matrix. We used *CAVIAR* to identify variants in the imputed dataset most likely to be causal (11). Ontology enrichment for genes near the top 1000 most strongly associated variants from the 5M dataset was obtained using the annotation tool.

GREAT (76)

We estimated the contribution to the variation in melanin index from the top candidate causal variants with a restricted maximum likelihood (REML) analysis implemented in the Genome-wide Complex Trait Analysis (GCTA) software (77). The variance parameters for two genetic relationship matrices (GRMs) are estimated: one GRM is constructed (78) from genome-wide background variants with MAF>0.01, and one GRM is constructed from the set of 8 top pigmentation-associated variants. The contribution of each locus to the melanin index variation is estimated similarly, using all genome-wide significant (p-value $<5 \times 10^{-8}$) variants within each locus to construct the pigmentation-associated GRM (table S3). REML iterations are based on maximizing the Average Information matrix.

To test for neutrality in the regions flanking our top GWAS variants we calculated Tajima's D, F_{ST}, and extended haplotype homozygosity using iHS (79-81). Tajima's D was measured along chromosomes 11 and 15 using 50 kb sliding windows. Due to a high recombination rate observed near the *MFSD12* locus, we used 10 kb windows in that region (chromosome

19). Vcftools was used to calculate both Tajima's D and F_{ST} (82). To calculate extended haplotype homozygosity (iHS) we used *Selscan* (83). Unstandardized iHS scores were normalized within 100kb bins according to the frequency of the derived allele. We then identified the signals of positive selection by calculating the proportion of SNPs with |iHS| >2 in non-overlapping windows (80). To identify outlier windows we calculated 5th and 95th percentiles. Population differentiation was assessed with Weir and Cockerham's fixation index F_{ST} (79) between each pair of populations. Outliers were identified using empirical p-values.

Median joining haplotype networks (84) for the Illumina Omni 5M SNP dataset were constructed and visualized at genomic regions of interest using *NETWORK* (85). In addition, we constructed genealogies of regions flanking candidate causal SNPs using a hierarchical clustering approach with sequence data from the Simons Genome Diversity Project (13). Briefly, we considered a single copy of each chromosome from each of the 279 individuals from (13). We inferred recombination breakpoints within a symmetric window surrounding each locus using the program *kwarg* (86) and identified the longest shared haplotype between each pair of sequences in which no recombination events occurred. We then computed the expected coalescence time between each pair of sequences, conditional on the observed number of mutations in the non-recombining region. Genealogies were constructed by applying the WPGMA hierarchical clustering algorithm to the estimated pairwise coalescence times. Our estimator accounts for recombination events and the population size history. However, simulation studies indicate that accounting for time-varying population size has relatively little effect on our estimates when the size changes according to previously inferred histories for human populations (67, 87). Because the true population sizes and relationships among the populations we considered are complex and imprecisely known, we assumed a constant population size of $N = 10^4$ in our analyses. The robustness analysis presented in (67) describes how our time estimates would change under different demographic histories and selective pressures.

To identify candidate causal GWAS variants altering gene expression we visualized and intersected variants with chromHMM tracks (88), DNase-1 hypersensitivity peaks, H3K27ac signal tracks for keratinocytes and melanocytes (29), CTCF signal tracks from keratinocytes (89), and ChIP-seq signal tracks from MITF (45). Variants were intersected with chromatin annotations using *bedtools* (90). Functional consequences of variants were also assessed using deepSEA (91) and deltaSVM (92). The effect of genetic variants on transcription factor binding was predicted using the *MEME* suite (93) for all transcription factors in the JASPAR 2016 CORE Vertebrate motif set (94).

To test for associations between gene expression, genetic variation, and ancestry we used eQTL and allele-specific expression (ASE) analyses on transcriptomes and genotype data from primary cultures of human melanocytes, isolated from foreskin of 106 individuals of assorted ancestries. All 106 individuals were genotyped on Illumina OmniExpress arrays and genotypes were subsequently imputed using the Michigan Imputation Server (75) based on the TGP reference panel and using *SHAPEIT* for phasing (73). RNA sequencing was performed to a mean depth of 87 million reads per sample. *STAR* (95) was used for aligning reads, and *RSEM* (96) was used to quantify the gene expression. Quantile normalization was

applied in all samples to get the final *RSEM* value. To account for hidden factors driving expression variabilities, a set of covariates were further identified using the PEER method (97 and applied to calculate the normalized expression matrices. Principal components analysis was performed using genotypic data to capture population structure and ancestry using the *struct.pea* module of GLU. Using the normalized expression values and principal components, Pearson correlation between gene expression levels and ancestry was calculated, and associations between GWAS variant genotypes and gene expression levels were evaluated using ordinary least squares regression.

To identify associations between our GWAS candidate causal variants and expression of nearby genes (using the 106 melanocyte transcriptomes), we first found all protein-coding genes with transcription start site (TSS) within 1 Mb of the top GWAS variant for each locus and RSEM values greater than 0.5 in the primary melanocyte cultures. Pearson correlation was used to measure the association between ancestry and gene expression.

For each locus, we tested whether any genes with a transcription start site within 1Mb of the top SNP had an eQTL amongst the set of pigmentation QTLs using an additive linear model with the first two principal components of ancestry as covariates. To identify significant variant-gene associations we used a permutation approach for each locus independently (67). This was repeated for each gene and focal variants across all genes were adjusted for multiple testing using the Bonferroni correction (98).

We also carried out allele-specific expression (ASE) analyses for each significant eQTL SNP. Sites with at least 30 mapped reads, <5% mapping bias, and ENCODE 125 bp mappability score >=1 were retained for further analysis. For genes with a heterozygous coding variant amongst the melanocyte transcriptomes, allelic expression (**AE**) was computed as **AE**= |0.5 - N_A/(N_A+ N_R)|, where N_R is the number of reads carrying the reference allele and N_A is the number of reads containing the alternative allele. For each GWAS variant, differences in gene **AE** between GWAS heterozygotes and homozygotes was evaluated by Wilcoxon rank-sum test. This was repeated for all possible genes and GWAS variants and Bonferroni-corrected p-values less than 0.05 were considered significant. For several genes, including *HMG20B* and *DDB1*, ASE could not be measured for some or all variants of interest because no individuals were heterozygous for both the test- variant and a coding variant.

For *OCA2* we tested for an association between inclusion rates of exon 10, which contains our top candidate causal variant in the region, rs1800404, and individual genotypes at rs1800404. For each melanocyte transcriptome, reads spanning the exon9-exon10 and exon9-exonll junctions were extracted from the splice-junction files output by *STAR.* For each individual, a percent spliced in *(PSI)* value was calculated. To estimate the effect of variation at rs1800404 on exon 10 inclusion, ordinary least squares regression was carried out between *PSI* and dosage of the alternative allele for rs1800404 across individuals. A two-sided *t* test was used to calculate a p-value.

To test the functional impact of a subset of GWAS variants on gene expression, predicted regulatory sequences containing variants were cloned into a pGL*A*.23 firefly luciferase

reporter vector. Vectors were transfected into a WM88 melanocytic melanoma cell line and co-transfected with renilla luciferase control vector (pRL-CMV) in a dual luciferase assay. Relative luciferase activity (firefly/renilla luminescence ratio) is presented as fold change compared to cells transfected with the empty pGL4.23 vector. Data were analyzed with a modified Kruskal-Wallis Rank Sum test and pairwise comparisons between groups were performed using the Conover method. P values were corrected for multiple comparisons with the Benjamini-Hochberg method using the R package PMCMR, and p-values less than 0.05 were considered significant.

We characterized the function of *MFSD12* in vitro in immortalized melanocytes and in vivo in both zebrafish and mice. Immortalized melan-Ink4a melanocytes from C57BL/6 *Ink4a-Arf1*$^{-/-}$ mice were cultured as described (31). To deplete MFSD12, cells were infected with recombinant lentiviruses – generated by transient transfection in HEK293T cells – to express *Mfsd12*-targeted shRNAs or non-target controls. Cells resistant to puromycin (also encoded by the lentiviruses) were analyzed 5-7 days after infection. Knockdown efficiency of *Mfsd12* mRNA in cells expressing *Mfsd12*-specific shRNAs relative to non-target shRNA was quantified by reverse transcription/quantitative PCR (detected with SYBR Green; Applied Biosystems) relative to *Tubb4b* (encoding β-tubulin) as a reference gene. Melanin content in cell lysates was determined by a spectrophotometric assay as described (99), and melanin coverage in intact cells was determined by bright field microscopy and analysis using the "Analyze Particles" plug-in in ImageJ (National Institutes of Health). To analyze MFSD12-HA localization, melan-Ink4a melanocytes were transiently transfected with MFSD12-HA expression plasmids and analyzed 48 hours later by bright field and immunofluorescence microscopy as described (100) using the TA99 monoclonal antibody to TYRP1 (American Type Culture Collection) to detect melanosomes, rabbit anti-LAMP2A (Abcam) to detect lysosomes, and rat anti-HA (Roche) to detect the transgene. Percent signal overlap in the cell periphery was determined from background-subtracted, thresholded, binary images using the "Analyze Particles" plug-in in ImageJ. Statistical significance was determined using unpaired, twotailed student's *t* tests: $p < 0.05$: *, $p < 0.01$: **, $p < 0.001$: ***, $p < 0.0001$: ****. Details are provided in (67).

Zebrafish mutagenesis using CRISPR/Cas9 was performed to target mfsd12a. Compound heterozygous mutant fish for analysis were generated from F1 incrosses of mutant founder fish. For methylene blue staining, embryos were collected following fertilization and placed in zebrafish system water containing 0.0002% methylene blue and analyzed at 6 dpf. For GFP analysis, embryos were injected with 25 pg Tg(aox5:PALM-GFP) and 80 pg tol2 mRNA and GFP expression was evaluated in mosaic injected fish at 5 dpf. Larvae were anesthetized in sub-lethal 1× tricane solution and placed in 100ul of a low melt agarose solution (0.8%).

In mice two targets for CRISPR/Cas9 cleavage were selected within Exon 2 of *Mfsd12* to generate a 134bp deletion resulting in a null allele of *Mfsd12. A* mixture of Cas9 mRNA (TriLink BioTechnologies) and each of the two synthesized gRNAs was used for pronuclear injection into C57BL/6J × FVB/N F1 hybrid zygotes. Mutation carrying mice were viable and presented with grey coat color distinct from littermates. Hairs were plucked from postnatal day 18 mice and individual awl hairs were mounted in permount and imaged with

a stereomicroscope (Zeiss SteREO Discovery.V12) at the base of the sub apical yellow band where the switch from eumelanin to pheomelanin is visible.

To characterize *Mfsd12* in grizzled mice, Illumina generated whole genome sequences of *grizzled*, JIGR/DN (*gr/gr*) reads were mapped using *bwa mem* to GRCm38/mm10 (available at SRA Accession SRR5571237). Sequence variants between JIGR/DN *gr/gr* and C57BL/6J reference genome within the *gr/gr* candidate region were identified using *SnpEff*(101). Validation of a 12 bp deletion within *Mfsd12* was performed using samples from an independently maintained *gr/gr* colony provided by the laboratory of Dr. Margit Burmeister.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

REFERENCES AND NOTES

1. Jablonski NG, Chaplin G. The evolution of human skin coloration. J Hum Evol. 2000; 39:57–106. DOI: 10.1006/jhev.2000.0403 [PubMed: 10896812]

2. Marks MS, Seabra MC. The melanosome: Membrane dynamics in black and white. Wat Rev Mol Cell Biol. 2001; 2:738–748. DOI: 10.1038/35096009

3. Mover FH. Genetic variations in the fine structure and ontogeny of mouse melanin granules. Am Zool. 1966; 6:43–66. DOI: 10.1093/icb/6.1.43 [PubMed: 5902512]

4. Ebanks JP, Koshoffer A, Wickett RR, Schwemberger S, Babcock G, Hakozaki T, Boissy RE. Epidermal keratinocytes from light vs. dark skin exhibit differential degradation of melanosomes. J Invest Dermatol. 2011; 131:1226–1233. DOI: 10.1038/jid.2011.22 [PubMed: 21326292]

5. Liu F, Visser M, Duffy DL, Hysi PG, Jacobs LC, Lao O, Zhong K, Walsh S, Chaitanya L, Wollstein A, Zhu G, Montgomery GW, Henders AK, Mangino M, Glass D, Bataille V, Sturm RA, Rivadeneira F, Hofman A, van IJcken WFJ, Uitterlinden AG, Palstra R-JTS, Spector TD, Martin NG, Nijsten TEC, Kayser M. Genetics of skin color variation in Europeans: Genome-wide association studies with functional follow-up. Hum Genet. 2015; 134:823–835. DOI: 10.1007/s00439-015-1559-0 [PubMed: 25963972]

6. Beleza S, Johnson NA, Candille SI, Absher DM, Coram MA, Lopes J, Campos J, Araújo II, Anderson TM, Vilhjálmsson BJ, Nordborg M, Correia E Silva A, Shriver MD, Rocha J, Barsh GS, Tang H. Genetic architecture of skin and eye color in an African-European admixed population. PLOS Genet. 2013; 9:e1003372.doi: 10.1371/journal.pgen.1003372 [PubMed: 23555287]

7. Lloyd-Jones LR, Robinson MR, Moser G, Zeng J, Beleza S, Barsh GS, Tang H, Visscher PM. Inference on the genetic basis of eye and skin color in an admixed population via Bayesian linear mixed models. Genetics. 2017; 206:1113–1126. DOI: 10.1534/genetics.116.193383 [PubMed: 28381588]

8. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O, Ibrahim M, Juma AT, Kotze MJ, Lema G, Moore JH, Mortensen H, Nyambo TB, Omar SA, Powell K, Pretorius GS, Smith MW, Thera MA, Wambebe C, Weber JL, Williams SM. The genetic structure and history of Africans and African Americans. Science. 2009; 324:1035–1044. DOI: 10.1126/science.1172257 [PubMed: 19407144]

9. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E. Variance component model to account for sample structure in genome-wide association studies. Nat Genet. 2010; 42:348–354. DOI: 10.1038/ng.548 [PubMed: 20208533]

10. 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015; 526:68–74. DOI: 10.10.38/nature15393 [PubMed: 26432245]

11. Hormozdiari F, Kostem E, Kang EY, Pasaniuc B, Eskin E. Identifying causal variants at loci with multiple signals of association. Genetics. 2014; 198:497–508. DOI: 10.1534/genetics.114.167908 [PubMed: 25104515]

12. Vernot B, Tucci S, Kelso J, Schraiber JG, Wolf AB, Gittelman RM, Dannemann M, Grote S, McCoy RC, Norton H, Scheinfeldt LB, Merriwether DA, Koki G, Friedlaender JS, Wakefield J, Pääbo S, Akey JM. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. Science. 2016; 352:235–239. DOI: 10.1126/science.aad9416 [PubMed: 26989198]

13. Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A, Skoglund P, Lazaridis I, Sankararaman S, Fu Q, Rohland N, Renaud G, Erlich Y, Willems T, Gallo C, Spence JP, Song YS, Poletti G, Balloux F, van Driem G, de Knijff P, Romero IG, Jha AR, Behar DM, Bravi CM, Capelli C, Hervig T, Moreno-Estrada A, Posukh OL, Balanovska E, Balanovsky O, Karachanak-Yankova S, Sahakyan H, Toncheva D, Yepiskoposyan L, Tyler-Smith C, Xue Y, Abdullah MS, Ruiz-Linares A, Beall CM, Di Rienzo A, Jeong C, Starikovskaya EB, Metspalu E, Parik J, Villems R, Henn BM, Hodoglugil U, Mahley R, Sajantila A, Stamatoyannopoulos G, Wee JTS, Khusainova R, Khusnutdinova E, Litvinov S, Ayodo G, Comas D, Hammer MF, Kivisild T, Klitz W, Winkler CA, Labuda D, Bamshad M, Jorde LB, Tishkoff SA, Watkins WS, Metspalu M, Dryomov S, Sukernik R, Singh L, Thangaraj K, Pääbo S, Kelso J, Patterson N, Reich D. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature. 2016; 538:201–206. DOI: 10.1038/nature18964 [PubMed: 27654912]

14. Lamason RL, Mohideen MA, Mest JR, Wong AC, Norton HL, Aros MC, Jurynec MJ, Mao X, Humphreville VR, Humbert JE, Sinha S, Moore JL, Jagadeeswaran P, Zhao W, Ning G, Makalowska I, McKeigue PM, O'donnell D, Kittles R, Parra EJ, Mangini NJ, Grunwald DJ, Shriver MD, Canfield VA, Cheng KC. SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science. 2005; 310:1782–1786. DOI: 10.1126/science.1116238 [PubMed: 16357253]

15. Jonnalagadda M, Bharti N, Patil Y, Ozarkar S, Joshi SMKR, Norton H. Identifying signatures of positive selection in pigmentation genes in two South Asian populations. Am J Hum Biol. 2017; 29:e23012.doi: 10.1002/ajhh.23012

16. Basu Mallick C, Iliescu FM, Möls M, Hill S, Tamang R, Chaubey G, Goto R, Ho SYW, Gallego Romero I, Crivellaro F, Hudjashov G, Rai N, Metspalu M, Mascie-Taylor CGN, Pitchappan R, Singh L, Mirazon-Lahr M, Thangaraj K, Villems R, Kivisild T. The light skin allele of *SLC24A5* in South Asians and Europeans shares identity by descent. PLOS Genet. 2013; 9:e1003912.doi: 10.1371/journal.pgen.1003912 [PubMed: 24244186]

17. Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M, Sirak K, Gamba C, Jones ER, Llamas B, Dryomov S, Pickrell J, Arsuaga JL, de Castro JMB, Carbonell E, Gerritsen F, Khokhlov A, Kuznetsov P, Lozano M, Meller H, Mochalov O, Moiseyev V, Guerra MAR, Roodenberg J, Vergès JM, Krause

J, Cooper A, Alt KW, Brown D, Anthony D, Lalueza-Fox C, Haak W, Pinhasi R, Reich D. Genome-wide patterns of selection in 230 ancient Eurasians. Nature. 2015; 528:499–503. DOI: 10.1038/nature16152 [PubMed: 26595274]

18. Pagani L, Kivisild T, Tarekegn A, Ekong R, Plaster C, Gallego Romero I, Ayub Q, Mehdi SQ, Thomas MG, Luiselli D, Bekele E, Bradman N, Balding DJ, Tyler-Smith C. Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. Am J Hum Genet. 2012; 91:83–96. DOI: 10.1016/j.ajhg.2012.05.015 [PubMed: 22726845]

19. Tekola-Ayele F, Adeyemo A, Chen G, Hailu E, Aseffa A, Davey G, Newport MJ, Rotimi CN. Novel genomic signals of recent selection in an Ethiopian population. Eur J Hum Genet. 2015; 23:1085–1092. DOI: 10.1038/ejhg.2014.233 [PubMed: 25370040]

20. Schlebusch CM, Skoglund P, Sjödin P, Gattepaille LM, Hernandez D, Jay F, Li S, De Jongh M, Singleton A, Blum MGB, Soodyall H, Jakobsson M. Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. Science. 2012; 338:374–379. DOI: 10.1126/science.1227721 [PubMed: 22997136]

21. Pickrell JK, Patterson N, Barbieri C, Berthold F, Gerlach L, Güldemann T, Kure B, Mpoloka SW, Nakagawa H, Naumann C, Lipson M, Loh PR, Lachance J, Mountain J, Bustamante CD, Berger B, Tishkoff SA, Henn BM, Stoneking M, Reich D, Pakendorf B. The genetic prehistory of southern Africa. Nat Commun. 2012; 3:1143–1146. DOI: 10.10.38/ncomms2140 [PubMed: 23072811]

22. Pagani L, Schiffels S, Gurdasani D, Danecek P, Scally A, Chen Y, Xue Y, Haber M, Ekong R, Oljira T, Mekonnen E, Luiselli D, Bradman N, Bekele E, Zalloua P, Durbin R, Kivisild T, Tyler-Smith C. Tracing the route of modern humans out of Africa by using 225 human genome sequences from Ethiopians and Egyptians. Am J Hum Genet. 2015; 96:986–991. DOI: 10.1016/j.ajhg.2015.04.019 [PubMed: 26027499]

23. Ehret C. An African Classical Age: Eastern and Southern Africa in World History 1000 BC to AD 400 (Oxford. 1998

24. Madej MG, Dang S, Yan N, Kaback HR. Evolutionary mix-and-match with MFS transporters. Proc Natl Acad Sci USA. 2013; 110:5870–5874. DOI: 10.1073/pnas.1303538110 [PubMed: 23530251]

25. Yu R, Broady R, Huang Y, Wang Y, Yu J, Gao M, Levings M, Wei S, Zhang S, Xu A, Su M, Dutz J, Zhang X, Zhou Y. Transcriptome analysis reveals markers of aberrantly activated innate immunity in vitiligo lesional and non-lesional skin. PLOS ONE. 2012; 7:e51040.doi: 10.1371/journal.pone.0051040 [PubMed: 23251420]

26. Richter D, Grün R, Joannes-Boyau R, Steele TE, Amani F, Rué M, Fernandes P, Raynal JP, Geraads D, Ben-Ncer A, Hublin JJ, McPherron SP. The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age. Nature. 2017; 546:293–296. DOI: 10.1038/nature22335 [PubMed: 28593967]

27. Przeworski M. The signature of positive selection at randomly chosen loci. Genetics. 2002; 160:1179–1189. [PubMed: 11901132]

28. Le Corre V, Kremer A. The genetic differentiation at quantitative trait loci under local adaptation. Mol Ecol. 2012; 21:1548–1566. DOI: 10.1111/j.1365-294X.2012.05479.x [PubMed: 22332667]

29. Roadmap Epigenomics Consortium. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu YC, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh KH, Feizi S, Karlic R, Kim AR, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJ, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai LH, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M. Integrative analysis of 111 reference human epigenomes. Nature. 2015; 518:317–330. DOI: 10.1038/nature14248 [PubMed: 25693563]

30. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, Hoke HA, Young RA. Super-enhancers in the control of cell identity and disease. Cell. 2013; 155:934–947. DOI: 10.1016/j.cell.2013.09.053 [PubMed: 24119843]

31. Sviderskaya EV, Hill SP, Evans-Whipp TJ, Chin L, Orlow SJ, Easty DJ, Cheong SC, Beach D, DePinho RA, Bennett DC. p16^{Ink4a} in melanocyte senescence and differentiation. J Natl Cancer Inst. 2002; 94:446–454. DOI: 10.1093/jnci/94.6.446 [PubMed: 11904317]

32. Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch G-J, White S, Chow W, Kilian B, Quintais LT, Guerra-Assunção JA, Zhou Y, Gu Y, Yen J, Vogel J-H, Eyre T, Redmond S, Banerjee R, Chi J, Fu B, Langley E, Maguire SF, Laird GK, Lloyd D, Kenyon E, Donaldson S, Sehra H, Almeida-King J, Loveland J, Trevanion S, Jones M, Quail M, Willey D, Hunt A, Burton J, Sims S, McLay K, Plumb B, Davis J, Clee C, Oliver K, Clark R, Riddle C, Elliot D, Threadgold G, Harden G, Ware D, Begum S, Mortimore B, Kerry G, Heath P, Phillimore B, Tracey A, Corby N, Dunn M, Johnson C, Wood J, Clark S, Pelan S, Griffiths G, Smith M, Glithero R, Howden P, Barker N, Lloyd C, Stevens C, Harley J, Holt K, Panagiotidis G, Lovell J, Beasley H, Henderson C, Gordon D, Auger K, Wright D, Collins J, Raisen C, Dyer L, Leung K, Robertson L, Ambridge K, Leongamornlert D, McGuire S, Gilderthorp R, Griffiths C, Manthravadi D, Nichol S, Barker G, Whitehead S, Kay M, Brown J, Murnane C, Gray E, Humphries M, Sycamore N, Barker D, Saunders D, Wallis J, Babbage A, Hammond S, Mashreghi-Mohammadi M, Barr L, Martin S, Wray P, Ellington A, Matthews N, Ellwood M, Woodmansey R, Clark G, Cooper J, Tromans A, Grafham D, Skuce C, Pandian R, Andrews R, Harrison E, Kimberley A, Garnett J, Fosker N, Hall R, Garner P, Kelly D, Bird C, Palmer S, Gehring I, Berger A, Dooley CM, Ersan-Ürün Z, Eser C, Geiger H, Geisler M, Karotki L, Kirn A, Konantz J, Konantz M, Oberländer M, Rudolph-Geiger S, Teucke M, Lanz C, Raddatz G, Osoegawa K, Zhu B, Rapp A, Widaa S, Langford C, Yang F, Schuster SC, Carter NP, Harrow J, Ning Z, Herrero J, Searle SM, Enright A, Geisler R, Plasterk RH, Lee C, Westerfield M, de Jong PJ, Zon LI, Postlethwait JH, Nüsslein-Volhard C, Hubbard TJ, Roest Crollius H, Rogers J, Stemple DL. The zebrafish reference genome sequence and its relationship to the human genome. Nature. 2013; 496:498–503. DOI: 10.1038/nature12111 [PubMed: 23594743]

33. Kelsh RN, Brand M, Jiang YJ, Heisenberg CP, Lin S, Haffter P, Odenthal J, Mullins MC, van Eeden FJ, Furutani-Seiki M, Granato M, Hammerschmidt M, Kane DA, Warga RM, Beuchle D, Vogelsang L, Nüsslein-Volhard C. Zebrafish pigmentation mutations and the processes of neural crest development. Development. 1996; 123:369–389. [PubMed: 9007256]

34. Le Guyader S, Jesuthasan S. Analysis of xanthophore and pterinosome biogenesis in zebrafish using methylene blue and pteridine autofluorescence. Pigment Cell Res. 2002; 15:27–31. DOI: 10.1034/j.1600-0749.2002.00045.x [PubMed: 11837453]

35. Bloom JL, Falconer DS. "Grizzled," a mutant in linkage group X of the mouse. Genet Res. 1966; 7:159–167. DOI: 10.1017/S0016672300009587

36. Kobayashi T, Vieira WD, Potterf B, Sakai C, Imokawa G, Hearing VJ. Modulation of melanogenic protein expression during the switch from eu- to pheomelanogenesis. J Cell Sci. 1995; 108:2301–2309. [PubMed: 7673350]

37. Raposo G, Tenza D, Murphy DM, Berson JF, Marks MS. Distinct protein sorting and localization to premelanosomes, melanosomes, and lysosomes in pigmented melanocytic cells. J Cell Biol. 2001; 152:809–824. DOI: 10.1083/jcb.152.4.809 [PubMed: 11266471]

38. Chu G, Chang E. Xeroderma pigmentosum group E cells lack a nuclear factor that binds to damaged DNA. Science. 1988; 242:564–567. DOI: 10.1126/science.3175673 [PubMed: 3175673]

39. Kadekaro AL, Leachman S, Kavanagh RJ, Swope V, Cassidy P, Supp D, Sartor M, Schwemberger S, Babcock G, Wakamatsu K, Ito S, Koshoffer A, Boissy RE, Manga P, Sturm RA, Abdel-Malek ZA. Melanocortin 1 receptor genotype: An important determinant of the damage response of melanocytes to ultraviolet radiation. EASES J. 2010; 24:3850–3860. DOI: 10.1096/fj.10-158485

40. Zhang Y, Feng S, Chen F, Chen H, Wang J, McCall C, Xiong Y, Deng XW. *Arabidopsis* DDB1-CUL4 ASSOCIATED FACTOR1 forms a nuclear E3 ubiquitin ligase with DDB1 and CUL4 that is involved in multiple plant developmental processes. Plant Cell. 2008; 20:1437–1455. DOI: 10.1105/tpc.108.058891 [PubMed: 18552200]

41. Yu C, Zhang YL, Pan WW, Li XM, Wang ZW, Ge ZJ, Zhou JJ, Cang Y, Tong C, Sun QY, Fan HY. CRL4 complex regulates mammalian oocyte survival and reprogramming by activation of TET proteins. Science. 2013; 342:1518–1521. DOI: 10.1126/science.1244587 [PubMed: 24357321]

42. Lieberman M, Segev O, Gilboa N, Lalazar A, Levin I. The tomato homolog of the gene encoding UV-damaged DNA binding protein 1 (DDB1) underlined as the gene that causes the *high pigment-1* mutant phenotype. Theor Appl Genet. 2004; 108:1574–1581. DOI: 10.1007/s00122-004-1584-1 [PubMed: 14968305]

43. Takata K, Yoshida H, Yamaguchi M, Sakaguchi K. Drosophila damaged DNA-binding protein 1 is an essential factor for development. Genetics. 2004; 168:855–865. DOI: 10.1534/genetics.103.025965 [PubMed: 15514059]

44. Lee JH, Silhavy JL, Lee JE, Al-Gazali L, Thomas S, Davis EE, Bielas SL, Hill KJ, Iannicelli M, Brancati F, Gabriel SB, Russ C, Logan CV, Sharif SM, Bennett CP, Abe M, Hildebrandt F, Diplas BH, Attié-Bitach T, Katsanis N, Rajab A, Koul R, Sztriha L, Waters ER, Ferro-Novick S, Woods CG, Johnson CA, Valente EM, Zaki MS, Gleeson JG. Evolutionarily assembled cis- regulatory module at a human ciliopathy locus. Science. 2012; 335:966–969. DOI: 10.1126/science.1213506 [PubMed: 22282472]

45. Laurette P, Strub T, Koludrovic D, Keime C, Le Gras S, Seberg H, Van Otterloo E, Imrichova H, Siddaway R, Aerts S, Cornell RA, Mengus G, Davidson I. Transcription factor MITF and remodeller BRG1 define chromatin organisation at regulatory elements in melanoma cells. eLife. 2015; 4:e06857.doi: 10.7554/elife.06857

46. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, Sim HS, Peh SQ, Mulawadi FH, Ong CT, Orlov YL, Hong S, Zhang Z, Landt S, Raha D, Euskirchen G, Wei CL, Ge W, Wang H, Davis C, Fisher-Aylor KI, Mortazavi A, Gerstein M, Gingeras T, Wold B, Sun Y, Fullwood MJ, Cheung E, Liu E, Sung WK, Snyder M, Ruan Y. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 2012; 148:84–98. DOI: 10.1016/j.cell.2011.12.014 [PubMed: 22265404]

47. Teng L, He B, Wang J, Tan K. 4DGenome: A comprehensive database of chromatin interactions. Bioinformatics. 2015; 31:2560–2564. DOI: 10.1093/bioinformatics/btv158 [PubMed: 25788621]

48. Malaspinas AS, Westaway MC, Muller C, Sousa VC, Lao O, Alves I, Bergström A, Athanasiadis G, Cheng JY, Crawford JE, Heupink TH, Macholdt E, Peischl S, Rasmussen S, Schiffels S, Subramanian S, Wright JL, Albrechtsen A, Barbieri C, Dupanloup I, Eriksson A, Margaryan A, Moltke I, Pugach I, Korneliussen TS, Levkivskyi IP, Moreno-Mayar JV, Ni S, Racimo F, Sikora M, Xue Y, Aghakhanian FA, Brucato N, Brunak S, Campos PF, Clark W, Ellingvåg S, Fourmile G, Gerbault P, Injie D, Koki G, Leavesley M, Logan B, Lynch A, Matisoo-Smith EA, McAllister PJ, Mentzer AJ, Metspalu M, Migliano AB, Murgha L, Phipps ME, Pomat W, Reynolds D, Ricaut FX, Siba P, Thomas MG, Wales T, Wall CM, Oppenheimer SJ, Tyler-Smith C, Durbin R, Dortch J, Manica A, Schierup MH, Foley RA, Lahr MM, Bowern C, Wall JD, Mailund T, Stoneking M, Nielsen R, Sandhu MS, Excoffier L, Lambert DM, Willerslev E. A genomic history of Aboriginal Australia. Nature. 2016; 538:207–214. DOI: 10.1038/nature18299 [PubMed: 27654914]

49. Beltrame MH, Rubel MA, Tishkoff SA. Inferences of African evolutionary history from genomic data. Curr Opin Genet Dev. 2016; 41:159–166. DOI: 10.1016/j.gde.2016.10.002 [PubMed: 27810637]

50. Visser M, Kayser M, Palstra RJ. *HERC2* rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the *OCA2* promoter. Genome Res. 2012; 22:446–455. DOI: 10.1101/gr.128652.111 [PubMed: 22234890]

51. Kayser M, Liu F, Janssens ACJW, Rivadeneira F, Lao O, van Duijn K, Vermeulen M, Arp P, Jhamai MM, van Ijcken WFJ, den Dunnen JT, Heath S, Zelenika D, Despriet DDG, Klaver CCW, Vingerling JR, de Jong PTVM, Hofman A, Aulchenko YS, Uitterlinden AG, Oostra BA, van Duijn CM. Three genome-wide association studies and a linkage analysis identify *HERC2* as a human iris color gene. Am J Hum Genet. 2008; 82:411–423. DOI: 10.1016/j.ajhg.2007.10.003 [PubMed: 18252221]

52. Han J, Kraft P, Nan H, Guo Q, Chen C, Qureshi A, Hankinson SE, Flu FB, Duffy DL, Zhao ZZ, Martin NG, Montgomery GW, Flayward NK, Thomas G, Hoover RN, Chanock S, Hunter DJ. A genome-wide association study identifies novel alleles associated with hair color and skin

pigmentation. PLOS Genet. 2008; 4:e1000074.doi: 10.1371/journal.pgen.1000074 [PubMed: 18483556]

53. Bellono NW, Escobar IE, Lefkovith AJ, Marks MS, Oancea E. An intracellular anion channel critical for pigmentation. eLife. 2014; 3:e04543.doi: 10.7554/eLife.04543 [PubMed: 25513726]

54. Brilliant MH. The mouse *p* (*pink-eyed dilution*) and human P genes, oculocutaneous albinism type 2 (OCA2), and melanosomal pH. Pigment Cell Res. 2001; 14:86–93. DOI: 10.1034/j. 1600-0749.2001.140203.x [PubMed: 11310796]

55. Rafati A, et al. Association of rs12913832 in the *HERC2* gene affecting human iris color variation. ASJ. 2015; 12:9–16.

56. Eiberg H, Troelsen J, Nielsen M, Mikkelsen A, Mengel-From J, Kjaer KW, Hansen L. 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. 2008; 123:177–187. DOI: 10.1007/s00439-007-0460-x [PubMed: 18172690]

57. Seberg HE, Van Otterloo E, Loftus SK, Liu H, Bonde G, Sompallae R, Gildea DE, Santana JF, Manak JR, Pavan WJ, Williams T, Cornell RA. TFAP2 paralogs regulate melanocyte differentiation in parallel with MITF. PLOS Genet. 2017; 13:e1006636.doi: 10.1371/journal.pgen. 1006636 [PubMed: 28249010]

58. Oetting WS, Garrett SS, Brott M, King RA. P gene mutations associated with oculocutaneous albinism type II (OCA2). Hum Mutat. 2005; 25:323.doi: 10.1002/humu.9318

59. Kerr R, Stevens G, Manga P, Salm S, John P, Haw T, Ramsay M. Identification of P gene mutations in individuals with oculocutaneous albinism in sub-Saharan Africa. Hum Mutat. 2000; 15:166–172. DOI: 10.1002/SICI1098-100420000215:2<166:AID-HUMU5>3.0.CO;2-Z [PubMed: 10649493]

60. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan J, Kutalik Z, Amin N, Buchkovich ML, Croteau-Chonka DC, Day FR, Duan Y, Fall T, Fehrmann R, Ferreira T, Jackson AU, Karjalainen J, Lo KS, Locke AE, Mägi R, Mihailov E, Porcu E, Randall JC, Scherag A, Vinkhuyzen AAE, Westra H-J, Winkler TW, Workalemahu T, Zhao JH, Absher D, Albrecht E, Anderson D, Baron J, Beekman M, Demirkan A, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Fraser RM, Goel A, Gong J, Justice AE, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Lui JC, Mangino M, Mateo Leach I, Medina-Gomez C, Nalls MA, Nyholt DR, Palmer CD, Pasko D, Pechlivanis S, Prokopenko I, Ried JS, Ripke S, Shungin D, Stancáková A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Afzal U, Arnlöv J, Arscott GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Blüher M, Bolton JL, Böttcher Y, Boyd HA, Bruinenberg M, Buckley BM, Buyske S, Caspersen IH, Chines PS, Clarke R, Claudi-Boehm S, Cooper M, Daw EW, De Jong PA, Deelen J, Delgado G, Denny JC, Dhonukshe-Rutten R, Dimitriou M, Doney ASF, Dörr M, Eklund N, Eury E, Folkersen L, Garcia ME, Geller F, Giedraitis V, Go AS, Grallert H, Grammer TB, Gräßler J, Grönberg H, de Groot LCPGM, Groves CJ, Haessler J, Hall P, Haller T, Hallmans G, Hannemann A, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q, Hemani G, Henders AK, Hillege HL, Hlatky MA, Hoffmann W, Hoffmann P, Holmen O, Houwing-Duistermaat JJ, Illig T, Isaacs A, James AL, Jeff J, Johansen B, Johansson Å, Jolley J, Juliusdottir T, Junttila J, Kho AN, Kinnunen L, Klopp N, Kocher T, Kratzer W, Lichtner P, Lind L, Lindström J, Lobbens S, Lorentzon M, Lu Y, Lyssenko V, Magnusson PKE, Mahajan A, Maillard M, McArdle WL, McKenzie CA, McLachlan S, McLaren PJ, Menni C, Merger S, Milani L, Moayyeri A, Monda KL, Morken MA, Müller G, Müller- Nurasyid M, Musk AW, Narisu N, Nauck M, Nolte IM, Nöthen MM, Oozageer L, Pilz S, Rayner NW, Renstrom F, Robertson NR, Rose LM, Roussel R, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Schunkert H, Scott RA, Sehmi J, Seufferlein T, Shi J, Silventoinen K, Smit JH, Smith AV, Smolonska J, Stanton AV, Stirrups K, Stott DJ, Stringham HM, Sundstrom J, Swertz MA, Syvänen A-C, Tayo BO, Thorleifsson G, Tyrer JP, van Dijk S, van Schoor NM, van der Velde N, van Heemst DA, van Oort FVSH, Vermeulen N, Verweij JM, Vonk LL, Waite M, Waldenberger R, Wennauer LR, Wilkens C, Willenborg T, Wilsgaard MK, Wojczynski A, Wong AF, Wright Q, Zhang D, Arveiler SJL, Bakker J, Beilby RN, Bergman S, Bergmann R, Biffar J, Blangero DI, Boomsma SR, Bornstein P, Bovet P, Brambilla MJ, Brown H, Campbell MJ, Caulfield A, Chakravarti R, Collins FS, Collins DC, Crawford LA, Cupples J, Danesh U, de Faire HM, den Ruijter R, Erbel J, Erdmann JG, Eriksson M, Farrall E, Ferrannini J, Ferrières I, Ford NG, Forouhi

T, Forrester RT, Gansevoort PV, Gejman C, Gieger A, Golay O, Gottesman V, Gudnason U, Gyllensten DW, Haas AS, Hall TB, Harris AT, Hattersley AC, Heath C, Hengstenberg AA, Hicks LA, Hindorff AD, Hingorani A, Hofman GK, Hovingh SE, Humphries SC, Hunt E, Hypponen KB, Jacobs M-R, Jarvelin P, Jousilahti AM, Jula J, Kaprio JJP, Kastelein M, Kayser F, Kee SM, Keinanen-Kiukaanniemi LA, Kiemeney JS, Kooner C, Kooperberg S, Koskinen P, Kovacs AT, Kraja M, Kumari J, Kuusisto TA, Lakka C, Langenberg L, Le Marchand T, Lehtimäki S, Lupoli Madden PA, Männistö F, Manunta S, Marette P, Matise A, McKnight TC, Meitinger B, Moll T, Montgomery FL, Morris GW, Morris AD, Murray AP, Nelis JC, Ohlsson C, Oldehinkel AJ, Ong KK, Ouwehand WH, Pasterkamp G, Peters A, Pramstaller PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ritchie M, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schwarz PE, Sebert S, Sever P, Shuldiner AR, Sinisalo J, Steinthorsdottir V, Stolk RP, Tardif J-C, Tönjes A, Tremblay A, Tremoli E, Virtamo J, Vohl M-C, Amouyel P, Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle E, Bottinger EP, Bouchard C, Cauchi S, Chambers JC, Chanock SJ, Cooper RS, de Bakker PI, Dedoussis G, Ferrucci L, Franks PW, Froguel P, Groop LC, Haiman CA, Hamsten A, Hayes MG, Hui J, Hunter DJ, Hveem K, Jukema JW, Kaplan RC, Kivimaki M, Kuh D, Laakso M, Liu Y, Martin NG, März W, Melbye M, Moebus S, Munroe PB, Njølstad I, Oostra BA, Palmer CN, Pedersen A, Perola M, Pérusse L, Peters U, Powell JE, Power C, Quertermous T, Rauramaa R, Reinmaa E, Ridker PM, Rivadeneira F, Rotter JI, Saaristo TE, Saleheen D, Schlessinger D, Slagboom PE, Snieder H, Spector TD, Strauch K, Stumvoll M, Tuomilehto J, Uusitupa M, van der Harst P, Völzke H, Walker M, Wareham NJ, Watkins H, Wichmann H-E, Wilson JF, Zanen P, Deloukas P, Heid IM, Lindgren CM, Mohlke KL, Speliotes EK, Thorsteinsdottir U, Barroso I, Fox CS, North KE, Strachan DP, Beckmann JS, Berndt SI, Boehnke M, Borecki IB, McCarthy MI, Metspalu A, Stefansson K, Uitterlinden AG, van Duijn CM, Franke L, Wilier CJ, Price AL, Lettre G, Loos RJF, Weedon MN, Ingelsson E, O'Connell JR, Abecasis GR, Chasman DI, Goddard ME, Visscher PM, Hirschhorn JN, Frayling TM, Electronic Medical Records and Genomics (eMEMERGEGE) Consortium, MIGen Consortium, PAGEGE Consortium, LifeLines Cohort Study. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet. 2014; 46:1173–1186. DOI: 10.1038/ng.3097 [PubMed: 25282103]

61. Harding RM, Healy E, Ray AJ, Ellis NS, Flanagan N, Todd C, Dixon C, Sajantila A, Jackson IJ, Birch-Machin MA, Rees JL. Evidence for variable selective pressures at MC1R. Am J Hum Genet. 2000; 66:1351–1361. DOI: 10.1086/302863 [PubMed: 10733465]

62. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PLF, Maricic T, Good JM, Marques-Bonet T, Alkan C, Fu Q, Mallick S, Li H, Meyer M, Eichler EE, Stoneking M, Richards M, Talamo S, Shunkov MV, Derevianko AP, Hublin J-J, Kelso J, Slatkin M, Pääbo S. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature. 2010; 468:1053–1060. DOI: 10.1038/nature09710 [PubMed: 21179161]

63. Jablonski NG, Chaplin G. The colours of humanity: The evolution of pigmentation in the human lineage. Philos Trans R Soc Lond B Biol Sci. 2017; 372:20160349.doi: 10.1098/rsth.2016.0349 [PubMed: 28533464]

64. Rogers AR, Iltis D, Wooding S. Genetic variation at the MCIR locus and the time since loss of human body hair. Curr Anthropol. 2004; 45:105–108. DOI: 10.1086/381006

65. Ang KC, Ngu MS, Reid KP, Teh MS, Aida ZS, Koh DXR, Berg A, Oppenheimer S, Salleh H, Clyde MM, Md-Zain BM, Canfield VA, Cheng KC. Skin color variation in Orang Asli tribes of Peninsular Malaysia. PLOS ONE. 2012; 7:e42752.doi: 10.1371/journal.pone.0042752 [PubMed: 22912732]

66. Pagani L, Lawson DJ, Jagoda E, Mörseburg A, Eriksson A, Mitt M, Clemente F, Hudjashov G, DeGiorgio M, Saag L, Wall JD, Cardona A, Mägi R, Wilson Sayres MA, Kaewert S, Inchley C, Scheib CL, Järve M, Karmin M, Jacobs GS, Antao T, Iliescu FM, Kushniarevich A, Ayub Q, Tyler-Smith C, Xue Y, Yunusbayev B, Tambets K, Mallick CB, Saag L, Pocheshkhova E, Andriadze G, Muller C, Westaway MC, Lambert DM, Zoraqi G, Turdikulova S, Dalimova D, Sabitov Z, Sultana GNN, Lachance J, Tishkoff S, Momynaliev K, Isakova J, Damba LD, Gubina M, Nymadawa P, Evseeva I, Atramentova L, Utevska O, Ricaut F-X, Brucato N, Sudoyo H, Letellier T, Cox MP, Barashkov NA, Škaro V, Mulahasanovi L, Primorac D, Sahakyan H, Mormina M, Eichstaedt CA, Lichman DV, Abdullah S, Chaubey G, Wee JTS, Mihailov E, Karunas A, Litvinov S, Khusainova R, Ekomasova N, Akhmetova V, Khidiyatova I, Marjanovi D,

Yepiskoposyan L, Behar DM, Balanovska E, Metspalu A, Derenko M, Malyarchuk B, Voevoda M, Fedorova SA, Osipova LP, Lahr MM, Gerbault P, Leavesley M, Migliano AB, Petraglia M, Balanovsky O, Khusnutdinova EK, Metspalu E, Thomas MG, Manica A, Nielsen R, Villems R, Willerslev E, Kivisild T, Metspalu M. Genomic analyses inform on migration events during the peopling of Eurasia. Nature. 2016; 538:238–242. DOI: 10.1038/nature19792 [PubMed: 27654910]

67. Materials and methods are available as supplementary materials.

68. Wagner JK, Jovel C, Norton HL, Parra EJ, Shriver MD. Comparing quantitative measures of erythema, pigmentation and skin response using reflectometry. Pigment Cell Res. 2002; 15:379–384. DOI: 10.1034/j.1600-0749.2002.02042.x [PubMed: 12213095]

69. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet. 2007; 81:1084–1097. DOI: 10.1086/521987 [PubMed: 17924348]

70. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297–1303. DOI: 10.1101/gr.107524.110 [PubMed: 20644199]

71. Li H. Toward better understanding of artifacts in variant calling from high-coverage samples. Bioinformatics. 2014; 30:2843–2851. DOI: 10.1093/bioinformatics/btu356 [PubMed: 24974202]

72. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: Structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012; 28:i333–i339. DOI: 10.1093/bioinformatics/bts378 [PubMed: 22962449]

73. O'Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocoa M, Traglia M, Huang J, Huffman JE, Rudan I, McQuillan R, Fraser RM, Campbell H, Polasek O, Asiki G, Ekoru K, Hayward C, Wright AF, Vitart V, Navarro P, Zagury J-F, Wilson JF, Toniolo D, Gasparini P, Soranzo N, Sandhu MS, Marchini J. A general approach for haplotype phasing across the full spectrum of relatedness. PLOS Genet. 2014; 10:e1004234.doi: 10.1371/journal.pgen.1004234 [PubMed: 24743097]

74. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat Methods. 2011; 9:179–181. DOI: 10.1038/nmeth.1785 [PubMed: 22138821]

75. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh PR, Iacono WG, Swaroop A, Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis GR, Fuchsberger C. Next-generation genotype imputation service and methods. Nat Genet. 2016; 48:1284–1287. [PubMed: 27571263]

76. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010; 28:495–501. DOI: 10.1038/nbt.1630 [PubMed: 20436461]

77. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet. 2011; 88:76–82. DOI: 10.1016/j.ajhg.2010.11.011 [PubMed: 21167468]

78. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010; 42:565–569. [PubMed: 20562875]

79. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984; 38:1358–1370. [PubMed: 28563791]

80. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLOS Biol. 2006; 4:e72.doi: 10.1371/journal.pbio.0040072 [PubMed: 16494531]

81. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989; 123:585–595. [PubMed: 2513255]

82. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. The variant call format and VCFtools. Bioinformatics. 2011; 27:2156–2158. DOI: 10.1093/bioinformatics/btr330 [PubMed: 21653522]

83. Szpiech ZA, Hernandez RD. selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. Mol Biol Evol. 2014; 31:2824–2827. DOI: 10.1093/molbev/msu211 [PubMed: 25015648]

84. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phytogenies. Mol Biol Evol. 1999; 16:37–48. DOI: 10.1093/oxfordjournals.molbev.a026036 [PubMed: 10331250]

85. Fluxus Engineering. 1999. www.fluxus-engineering.com

86. Lyngsø, RB., Song, YS., Hein, J. Algorithms in Bioinformatics. Casadio, R., Myers, G., editors. Vol. 3692. Springer; 2005. p. 239-250.Lecture Notes in Computer Science series

87. Terhorst J, Kamm JA, Song YS. Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat Genet. 2017; 49:303–309. [PubMed: 28024154]

88. Ernst J, Kellis M. ChromHMM: Automating chromatin-state discovery and characterization. Nat Methods. 2012; 9:215–216. DOI: 10.1038/nmeth.1906 [PubMed: 22373907]

89. Chadwick LH. The NIH Roadmap Epigenomics Program data resource. Epigenomics. 2012; 4:317–324. DOI: 10.2217/epi.12.18 [PubMed: 22690667]

90. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. DOI: 10.1093/bioinformatics/btq033 [PubMed: 20110278]

91. Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model. Nat Methods. 2015; 12:931–934. DOI: 10.1038/nmeth.3547 [PubMed: 26301843]

92. Lee D, Gorkin DU, Baker M, Straber BJ, Asoni AL, McCallion AS, Beer MA. A method to predict the impact of regulatory variants from DNA sequence. Nat Genet. 2015; 47:955–961. [PubMed: 26075791]

93. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: Tools for motif discovery and searching. Nucleic Acids Res. 2009; 37:W202–W208. DOI: 10.1093/nar/gkp335 [PubMed: 19458158]

94. Mathelier A, Fornes O, Arenillas DJ, Chen CY, Denay G, Lee J, Shi W, Shyr C, Tan G, Worsley-Hunt R, Zhang AW, Parcy F, Lenhard B, Sandelin A, Wasserman WW. JASPAR 2016: A major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2016; 44:D110–D115. DOI: 10.1093/nar/gkv1176 [PubMed: 26531826]

95. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29:15–21. DOI: 10.1093/bioinformatics/bts635 [PubMed: 23104886]

96. Li B, Dewey CN. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12:323.doi: 10.1186/1471-2105-12-323 [PubMed: 21816040]

97. Stegle O, Parts L, Durbin R, Winn J. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLOS Comput Biol. 2010; 6:e1000770.doi: 10.1371/journal.pcbi.1000770 [PubMed: 20463871]

98. Bonferroni, CE. Teoria statistica delle classi e calcolo delle probabilita. Vol. 8. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze; 1936.

99. Delevoye C, Hurbain I, Tenza D, Sibarita JB, Uzan-Gafsou S, Ohno H, Geerts WJC, Verkleij AJ, Salamero J, Marks MS, Raposo G. AP-1 and KIF13A coordinate endosomal sorting and positioning during melanosome biogenesis. J Cell Biol. 2009; 187:247–264. DOI: 10.1083/jcb.200907122 [PubMed: 19841138]

100. Calvo PA, Frank DW, Bieler BM, Berson JF, Marks MS. A cytoplasmic sequence in human tyrosinase defines a second class of di-leucine-based sorting signals for late endosomal and lysosomal delivery. J Biol Chem. 1999; 274:12780–12789. DOI: 10.1074/jbc.274.18.12780 [PubMed: 10212263]

101. Cingolani P, Platts A, Wang L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin). 2012; 6:80–92. DOI: 10.4161/fly.19695 [PubMed: 22728672]

102. Jonnalagadda M, Ozarkar S, Ashma R, Kulkarni S. Skin pigmentation variation among populations of West Maharashtra, India. Am J Hum Biol. 2016; 28:36–43. DOI: 10.1002/ajhb.22738 [PubMed: 26126512]

103. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. DOI: 10.1093/bioinformatics/btp352 [PubMed: 19505943]

104. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience. 2015; 4:7.doi: 10.1186/s13742-015-0047-8 [PubMed: 25722852]

105. Ayres DL, Darling A, Zwickl DJ, Beerli P, Holder MT, Lewis PO, Huelsenbeck JP, Ronquist F, Swofford DL, Cummings MP, Rambaut A, Suchard MA. BEAGLE: An application programming interface and high-performance computing library for statistical phylogenetics. Syst Biol. 2012; 61:170–173. DOI: 10.1093/syshio/syr100 [PubMed: 21963610]

106. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009; 19:1655–1664. DOI: 10.1101/gr.094052.109 [PubMed: 19648217]

107. 1000 Genomes Project Consortium. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. Nature. 2010; 467:1061–1073. DOI: 10.1038/nature09534 [PubMed: 20981092]

108. Gazal S, Sahbatou M, Babron MC, Génin E, Leutenegger AL. High level of inbreeding in final phase of 1000 Genomes Project. Sci Rep. 2015; 5:17453.doi: 10.1038/srep17453 [PubMed: 26625947]

109. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–909. DOI: 10.1038/ng1847 [PubMed: 16862161]

110. Cortex Technology. DSM II ColorMeter. 2017. www.cortex.dk

111. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012; 44:821–824. [PubMed: 22706312]

112. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. The UCSC Table Browser data retrieval tool. Nucleic Acids Res. 2004; 32:D493–D496. DOI: 10.1093/nar/gkh103 [PubMed: 14681465]

113. Hormozdiari F, van de Bunt M, Segrè AV, Li X, Joo JWJ, Bilow M, Sul JH, Sankararaman S, Pasaniuc B, Eskin E. Colocalization of GWAS and eQTL signals detects target genes. Am J Hum Genet. 2016; 99:1245–1260. DOI: 10.1016/j.ajhg.2016.10.003 [PubMed: 27866706]

114. Sloan CA, Chan ET, Davidson JM, Malladi VS, Strattan JS, Hitz BC, Gabdank I, Narayanan AK, Ho M, Lee BT, Rowe LD, Dreszer TR, Roe G, Podduturi NR, Tanaka F, Hong EL, Cherry JM. ENCODE data at the ENCODE portal. Nucleic Acids Res. 2016; 44:D726–D732. DOI: 10.1093/nar/gkv1160 [PubMed: 26527727]

115. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. Brief Bioinform. 2013; 14:178–192. DOI: 10.1093/bib/bbs017 [PubMed: 22517427]

116. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nat Biotechnol. 2011; 29:24–26. DOI: 10.1038/nbt.1754 [PubMed: 21221095]

117. Grant CE, Bailey TL, Noble WS. FIMO: Scanning for occurrences of a given motif. Bioinformatics. 2011; 27:1017–1018. DOI: 10.1093/bioinformatics/btr064 [PubMed: 21330290]

118. Castel SE, Levy-Moonshine A, Mohammadi P, Banks E, Lappalainen T. Tools and best practices for data processing in allelic expression analysis. Genome Biol. 2015; 16:195.doi: 10.1186/s13059-015-0762-6 [PubMed: 26381377]

119. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc B. 1995; 57:289–300.

120. Varshney GK, Carrington B, Pei W, Bishop K, Chen Z, Fan C, Xu L, Jones M, LaFave MC, Ledin J, Sood R, Burgess SM. A high-throughput functional genomics workflow based on CRISPR/ Cas9-mediated targeted mutagenesis in zebrafish. Nat Protoc. 2016; 11:2357–2375. DOI: 10.1038/nprot.2016.141 [PubMed: 27809318]

121. Varshney GK, Pei W, LaFave MC, Idol J, Xu L, Gallardo V, Carrington B, Bishop K, Jones M, Li M, Harper U, Huang SC, Prakash A, Chen W, Sood R, Ledin J, Burgess SM. High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. Genome Res. 2015; 25:1030–1042. DOI: 10.1101/gr.186379.114 [PubMed: 26048245]

122. Carrington B, Varshney GK, Burgess SM, Sood R. CRISPR-STAT: An easy and reliable PCR-based method to evaluate target-specific sgRNA activity. Nucleic Acids Res. 2015; 43:e157.doi: 10.1093/nar/gkv802 [PubMed: 26253739]

123. Chen GK, Marjoram P, Wall JD. Fast and flexible simulation of DNA sequence data. Genome Res. 2009; 19:136–142. DOI: 10.1101/gr.083634.108 [PubMed: 19029539]

124. Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, Walters GB, Jonasdottir A, Gylfason A, Kristinsson KT, Gudjonsson SA, Frigge ML, Helgason A, Thorsteinsdottir U, Stefansson K. Finescale recombination rate differences between sexes, populations and individuals. Nature. 2010; 467:1099–1103. DOI: 10.1038/nature09525 [PubMed: 20981099]

125. Chen H, Hey J, Slatkin M. A hidden Markov model for investigating recent positive selection through haplotype structure. Theor Popul Biol. 2015; 99:18–30. DOI: 10.1016/j.tpb.2014.11.001 [PubMed: 25446961]

126. Smith, J., Coop, G., Stephens, M., Novembre, J. Estimating time to the common ancestor for a beneficial allele. Aug 24. 2016 071241 [Preprint]https://doi.org/10.1101/071241

127. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science. 2015; 348:648–660. DOI: 10.1126/science.1262110 [PubMed: 25954001]

128. Haltaufderhyde KD, Oancea E. Genome-wide transcriptome analysis of human epidermal melanocytes. Genomics. 2014; 104:482–489. DOI: 10.1016/j.ygeno.2014.09.010 [PubMed: 25451175]

129. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010; 38:e164.doi: 10.1093/nar/gkq603 [PubMed: 20601685]

130. Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L. Natural selection has driven population differentiation in modern humans. Nat Genet. 2008; 40:340–345. DOI: 10.1038/ng.78 [PubMed: 18246066]
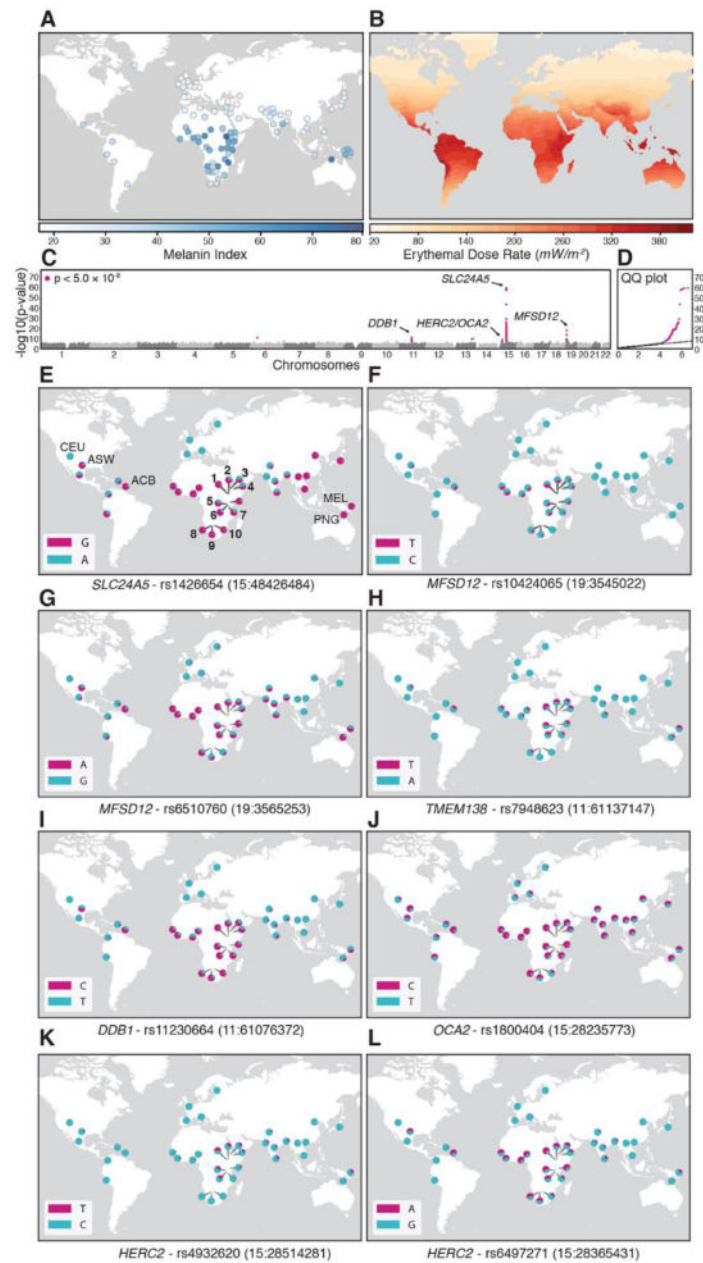
Fig. 1. Correlations between allele frequencies at genes associated with pigmentation and UV exposure in global populations

(**A**) Global variation in skin pigmentation indicated by Melanin index (Ml). This data was integrated with Ml data for global populations from (1, 102). (**B**) Mean erythemal dose rate. (**C**) Manhattan plot of -log10 transformed p-values from GWAS of skin pigmentation with the Illumina Omni5 SNP array. (**D**) Quantile-Quantile plot of observed vs expected p-values from the GWASIn both (C) and (D) significant SNPs at p < 5 × 10^{-8} are highlighted in purple. (**E** to **L**) Allele frequencies of genetic variants associated with skin pigmentation in global populations. African populations included are 1. Ethiopia Nilosaharan, 2. Ethiopia Omotic, 3. Ethiopia and Tanzania Cushitic, 4. Ethiopia Semitic, 5. Tanzania Nilosaharan, 6. Tanzania Hadza, 7. Tanzania Sandawe, 8. Botswana Bantu 9. Botswana San/Bantu admixed,

and 10. Botswana San. The Melanesian (MEL) samples are from (12) and the Australian Aboriginal and Papua New Guinean samples (merged) are from the SGDP (PNG) (13). All other populations are from the TGP (10). NonAboriginal populations in the Americas are indicated: CEU (European ancestry), ASW (African-American Southwest US) and ACB (African Caribbean in Barbados).
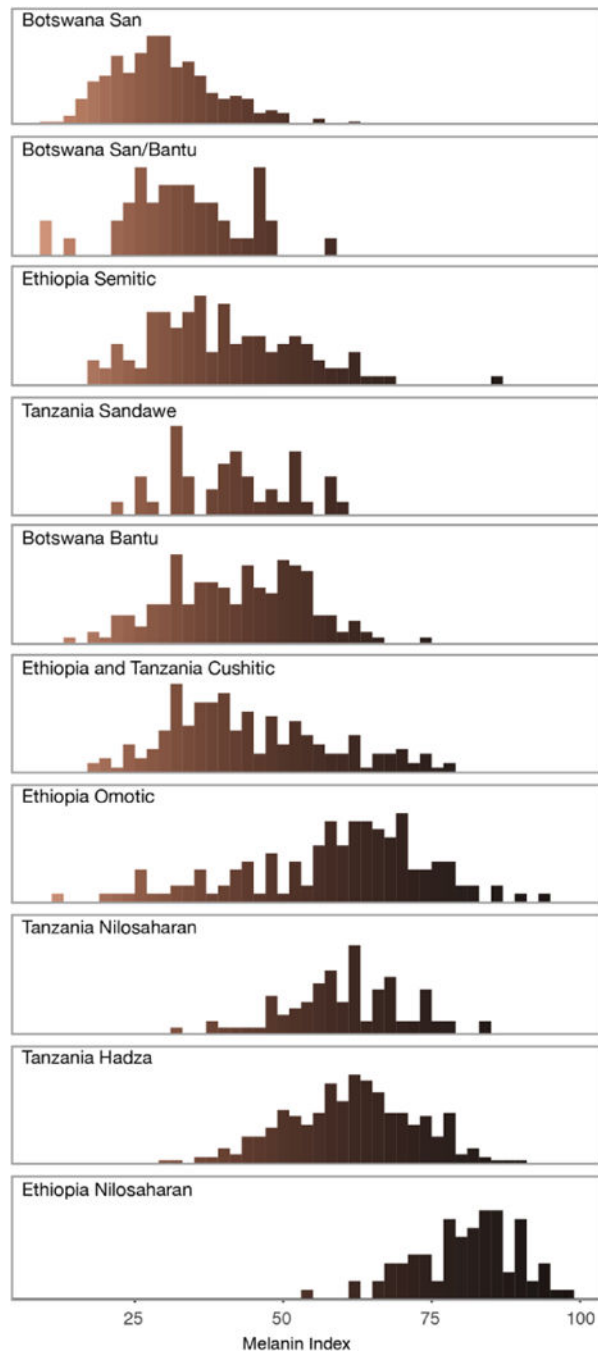
Fig. 2. Melanin distributions

Histograms of melanin index computed from under-arm measurements with a DSM II ColorMeter (68) for all individuals in each population as described in (67). Skin tones were visualized by displaying the mean red, green, and blue values from the ColorMeter for individuals binned by melanin index.

Fig. 3. Genomic context of GWAS loci

Plot of - log10(p-value) versus genomic position for variants near the four regions with most strongly associated SNPs from GWAS, including annotations for genes, MITF ChIP-seq data for melanocytes (45), a CTCF ChIP-seq track for NHEK keratinocytes, and H3K27ac, DNase hypersensitive sites (DHS), and chromHMM tracks for melanocytes and keratinocytes from the Roadmap Epigenomics dataset (29). Genome-wide significant variants are highlighted in red. Circles, squares, and triangles denote noncoding,

synonymous, and non-synonymous variants, respectively. **(A)** SLC24A5 locus, **(B)** MFSD12 locus, **(C)** DDB1/TMEM138, locus, and (D) OCA2/HERC2 locus.
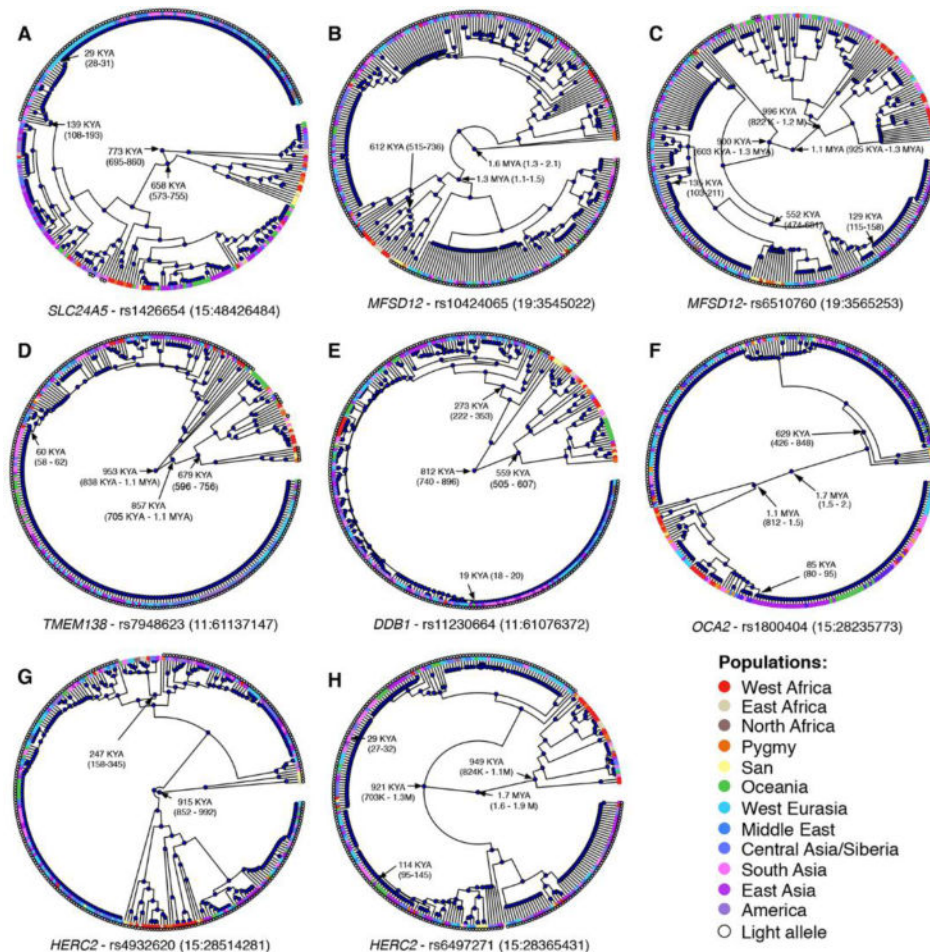
Fig. 4. Coalescent Trees and TMRCA dating

Inferred genealogies for regions flanking candidate causal loci. Each leaf corresponds to a single sampled chromosome from one of 278 individuals in the Simons Genome Diversity Project (13). Leaf nodes are colored by the population of origin of the individual and sequences carrying the light allele are indicated with an open dot. Node heights and 95% credible intervals are presented for a subset of internal nodes. Gene genealogies are shown for regions flanking **(A)** *SLC24A5*, rs1426654 (15:48426484), **(B)** *MFSD12*, rs10424065 (19:3545022), **(C)** *MFSD12*, rs6510760 (19:3565253), **(D)** *TMEM138*, rs7948623 (11:61137147), **(E)** *DDB1*, rs11230664 (11:61076372), **(F)** *OCA2*, rs1800404 (15:28235773), (G) *HERC2*, rs4932620 (15:28514281), and **(H)** *HERC2*, rs6497271 (15:28365431).
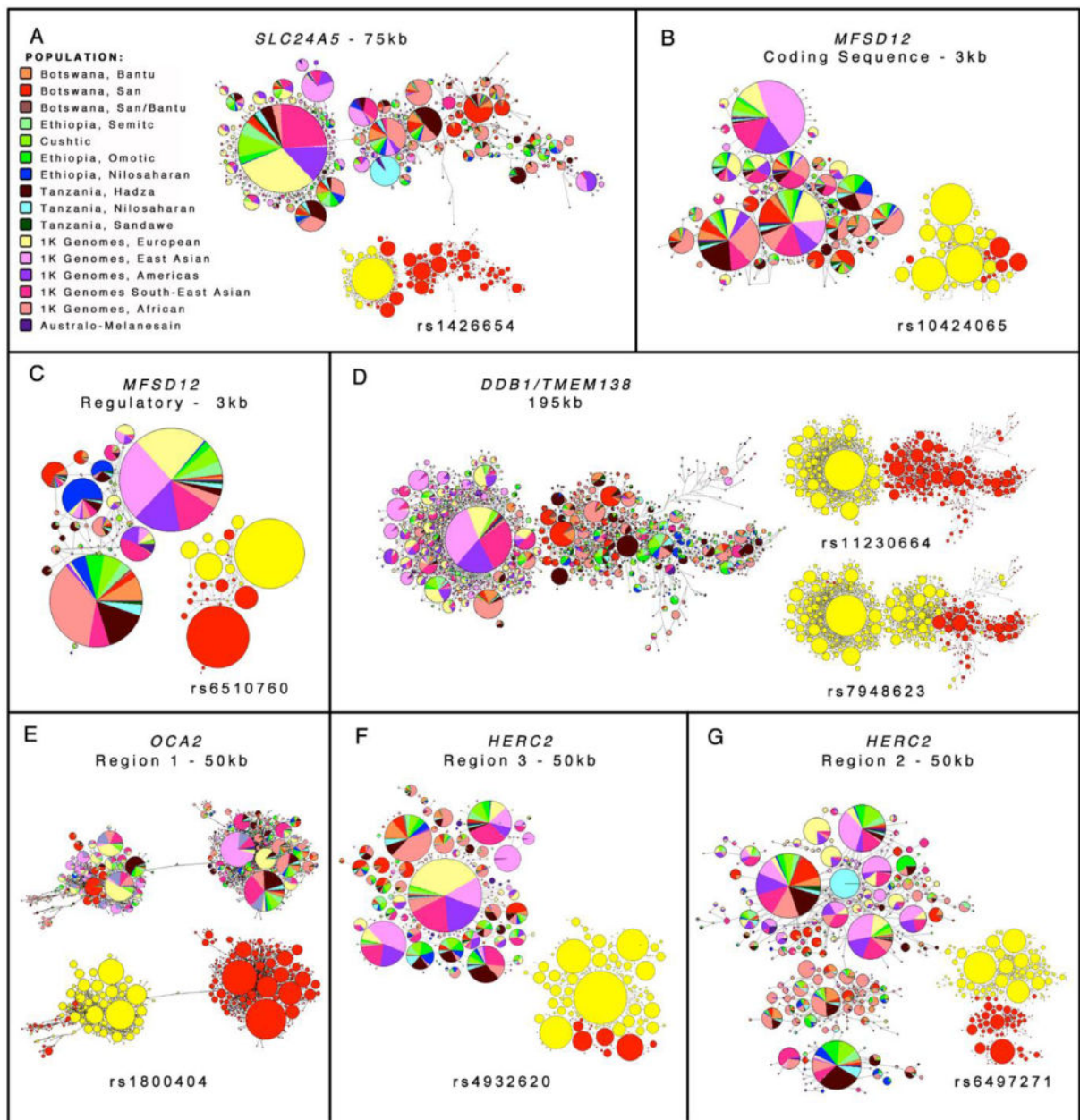
Fig. 5. Haplotype networks at *SLC24A5, MFSD12, DDB1/TMEM138*, and *OCA2/HERC2*
Median joining haplotype networks of regions containing candidate causal variants. Connections between circles indicate genetic relatedness while size is relative to the frequency of haplotypes. Ancestry proportions are displayed as pie charts. Yellow/red subfigures indicate which haplotypes contain the allele associated with dark pigmentation (red) or light pigmentation (yellow). (**A**) 75kb region flanking the causal variant at *SLC24A5.* (**B** and **C**) 3kb regions flanking rs10424065 in *MFSD12* and rs6510760 upstream of MFSD12. (**D**) 195 kb region flanking *DDB1* extending from *PGA5* through *SDHAF2.* (**E**, **F**, and **G**) 50 kb flanking regions 1, 3, and 2 at *OCA2* and *FIERC2* (ordered based on highest to lowest probability of being causal from CAVIAR analysis).
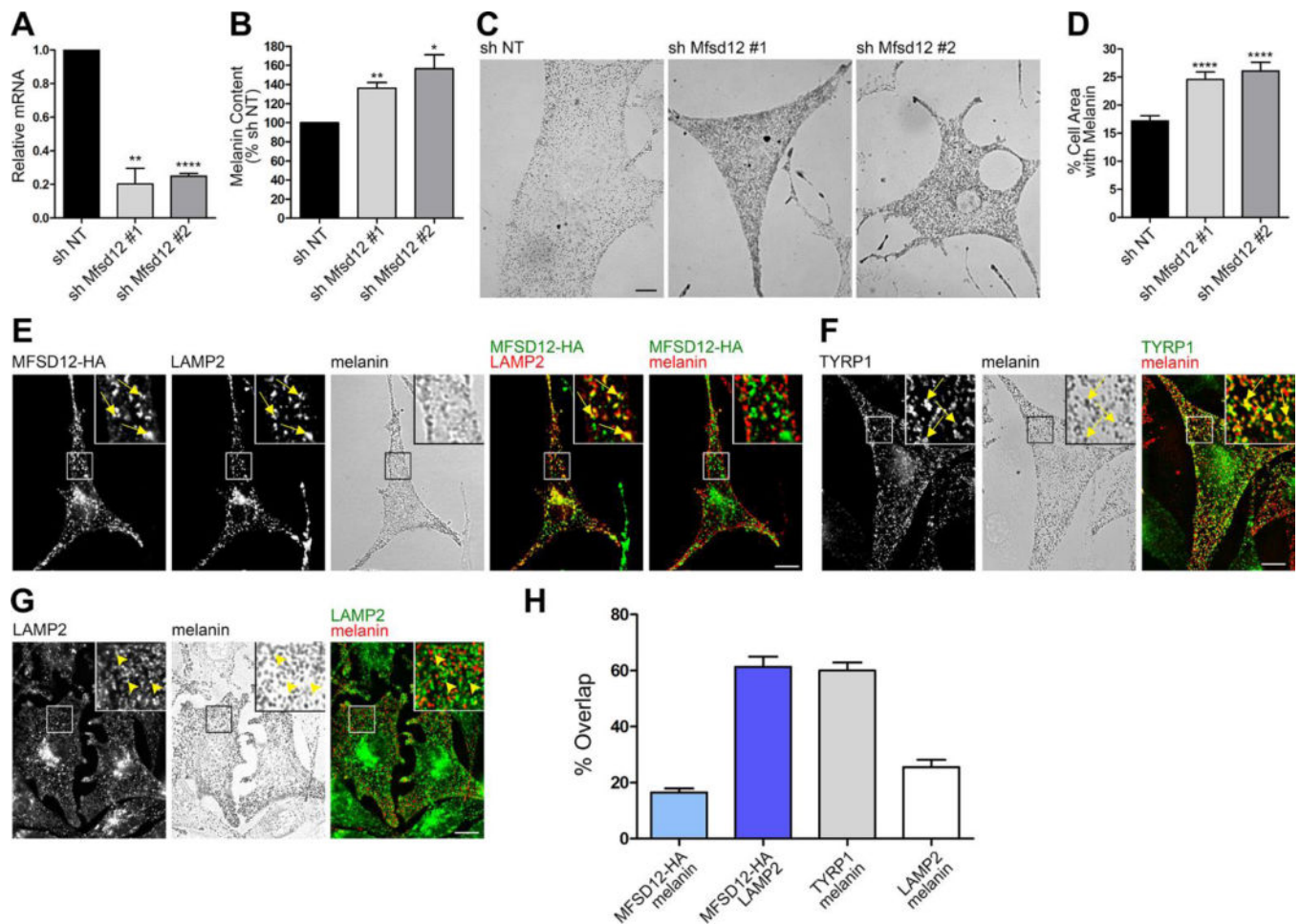
Fig. 6. MFSD12 suppresses eumelanin production but localizes to lysosomes

Immortalized melan-lnk4a melanocytes expressing non-target (sh NT) shRNA or either of two shRNA plasmid clones (#1 and 2) targeting *Mfsd12* were analyzed for *Mfsd12* mRNA content by (**A**) qRT-PCR, (**B**) melanin content by spectrophotometry, or (**C**) percent of cell area containing melanin by bright field microscopy; (**D**) shows quantification. Data in (A) to (C) represent mean ± s.e.m., normalized to sh NT samples, from three separate experiments; in (C), n(sh NT) = 97 cells, n (shMfsd12 #1) = 68 cells, n (shMfsd12 #2) = 71 cells. Bar in (C), 10 μm. (**E** to **G**) Melan-ink4a melanocytes transiently expressing MFSD12-HA (E) or not transfected (F) and (G) were fixed, immunolabeled for HA (E) and for LAMP2 to mark lysosomes E) and (G) or for TYRP1 to mark melanosomes (G), and analyzed by immunofluorescence and bright field microscopy. Bright field (melanin) images show pigmented melanosomes (pseudocolored red in the merged images). Insets, 4x magnification of boxed regions. Arrows, MFSD12-containing structures that overlap LAMP2 (E) or TYRP1-containing structures that overlap melanosomes; arrowheads, structures that do not overlap. Bars, 10 μm. (**H**) Quantification of overlap for structures labeled by MFSD12, TYRP1, LAMP2 and pigment. Data represent mean ± s.e.m. from three independent experiments, n = 17 cells (MFSD12 overlap with LAMP2 and melanin), 33 cells (TYRP1 overlap with melanin), or 23 cells (LAMP2 and melanin).
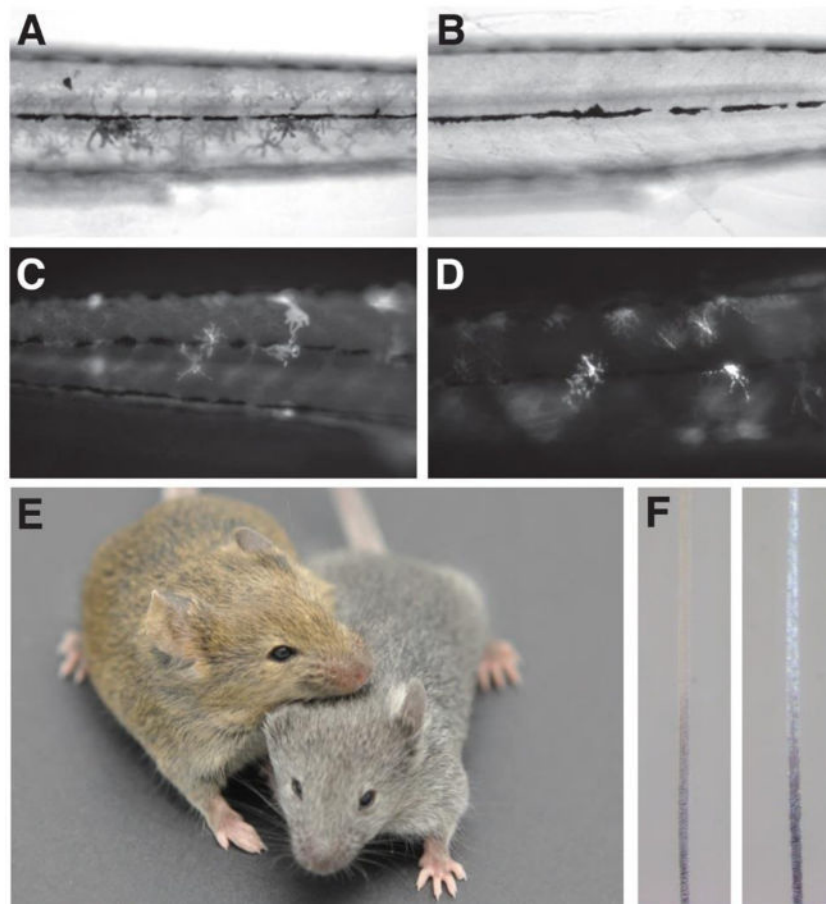
Fig. 7. In vivo zebrafish and mouse models of MFSD12 deficiency
(**A** and **B**) Representative images of methylene blue staining in wild-type TAB5. (A) and compound heterozygous *mfsd12a* zebrafish (6dpf). Note the absence of stained xanthophores in the *mfsd12a* mutant (B). (**C** and **D**) No difference was observed in the number or distribution of xanthophores detected by mosaic *Tg(aox5:PALM-GFP)* expression in injected wild-type TAB5 (C) or compound heterozygous mutant *mfsd12a* (D) zebrafish (5dpf). (E) A wild-type agouti mouse (left) is shown with a gray *Mfsd12* targeted littermate (right). (F) Flair from the *Mfsd12* targeted mouse has grossly normal eumelanin (lower black region of the hair shaft), however, the upper subapical yellow band in WT (E, left) appears white in the *Mfsd12* mutant (E, right) due to a reduction in pheomelanin.

Table 1

Annotations of candidate causal SNPs from GWAS

Top candidate causal variants for the four regions identified based on analysis with CAVIAR (11). For each variant, the genomic position (Location), RSID, and Ancestral>Derived alleles are shown, with the allele associated with dark pigmentation in bold. Beta and standard error (Beta(SE)) and the p-values from the GWAS (F test, linear mixed model) are given. For functional genomic data, nearest genes are given and variants overlapping DHS sites for melanocytes (E059) (DHS melanocytes) and/or other cell types (DHS other) available from Roadmap Epigenomics are indicated with X (29, 45). Variants intersecting regions with enhancer activity confirmed by luciferase expression assays are indicated with X (Luciferase Activity) as are chromatin interactions with nearby genes measured in MCF7 or K562 cell lines as identified by ChIA-PET (Chromatin Interactions) (46, 47). Those variants where luciferase activity was tested were labeled with Y (significant enhancer activity) or N (no enhancer activity) (fig. S7). SNPs that are in strong LD ($r^2 > 0.7$ in East Africans) are numerically labeled in the column titled LD Block.

Location	RSID	Ancestral> Derived	Beta(SE)	p-value	DHS Melanocytes	DHS Other	Luciferase Activity	LD Block	Nearest Gene	Chromatin Interactions
15:48485926	rs2413887	**T**>C	7.70(0.44)	4.9×10^{-62}				1	CTXN2	MYEF2
15:48426484	rs1426654	**G**>A	7.69(0.44)	5.5×10^{-62}	X	X	N	1	SLC24A5	
15:48392165	rs1834640	**G**>A	.56(0.44)	3.2×10^{-61}		X		1	SLC24A5	
15:48400199	rs2675345	**G**>A	7.62(0.44)	6.7×10^{-61}				1	SLC24A5	
15:48460188	rs8028919	**G**>A	-4.95(0.41)	5.0×10^{-32}				1	MYEF2	
19:3545022	rs10424065	**C**>T	4.48(0.48)	5.1×10^{-20}	X	X	Y	2	MFSD12	CACTIN,MFSD12
19:3544892	rs56203814	**C**>T	-4.38(0.50)	3.6×10^{-18}		X	Y	2	MFSD12	CACTIN,MFSD12
19:3566631	rs111317445	**C**>T	3.51(0.42)	1.7×10^{-16}			N	3	HMG20B	MFSD12
19:3547955	rs10414812	**C**>T	4.38(0.53)	3.8×10^{-16}	X		Y	2	MFSD12	CACTIN,FZR1,MFSD12
19:3565599	rs112332856	**T**>C	3.52(0.43)	3.8×10^{-16}	X	X	Y	3	HMG20B	MFSD12
19:3565253	rs6510760	**G**>A	3.54(0.45)	6.5×10^{-15}	X	X	Y	3	MFSD12	MFSD12
19:3545150	rs73527942	**T**>G	-3.58(0.47)	4.8×10^{-14}	X	X	Y	2	MFSD12	CACTIN,MFSD12
19:3547685	rs142317543	**C**>T	6.99(0.92)	5.0×10^{-14}	X	X	Y	2	MFSD12	CACTIN,FZR1,MFSD12
19:3566513	rs7254463	**C**>T	2.90(0.50)	9.0×10^{-9}		X	N	3	HMG20B	MFSD12
19:3565357	rs7246261	**C**>T	2.71(0.47)	1.1×10^{-8}	X	X		3	HMG20B	MFSD12
19:3565909	rs6510761	**T**>C	2.79(0.50)	2.2×10^{-8}	X	X		3	HMG20B	MFSD12
11:61137147	rs7948623	**A**>T	-2.94(0.44)	2.2×10^{-11}	X	X	Y	4	TMEM138	TKFC,DDB1, TMEM138
11:61148456	rs397709980	**G**/GA	-2.90(0.43)	2.4×10^{-11}		X	N	4	TMEM216	
11:61152630	rs4453253	**C**>T	-2.85(0.43)	5.4×10^{-11}	X	X	N	4	TMEM216	CYB561A3,TKFC,DDB1,TMEM138,TMEM216
11:61153401	rs4939520	**C**>T	-2.79(0.43)	1.4×10^{-10}		X	N	4	TMEM216	CYB561A3,TKFC,DDB1,TMEM138,TMEM216
11:61142943	rs4939519	**C**>T	-2.47(0.39)	2.8×10^{-10}		X		4	TMEM138	TMEM138

Location	RSID	Ancestral>Derived	Beta(SE)	p-value	DHS Melanocytes	DHS Other	Luciferase Activity	LD Block	Nearest Gene	Chromatin Interactions
11:61106525	rs2512809	C>T	−2.93(0.47)	7.4×10^{-10}			N	5	TKFC	TKFC,DDB1,TMEM138
11:61046876	rs11230658	T>C	3.01(0.49)	9.4×10^{-10}				5	VWCE	
11:61084180	rs12289370	G>A	2.99(0.49)	1.3×10^{-9}		X		5	DDB1	TKFC,DDB1
11:61144652	rs1377457	C>A	−3.01(0.49)	1.5×10^{-9}		X		5	TMEM138	CYB561A3;TKFC,DDB1,TMEM138
11:61088140	rs7934735	G>T	2.98(0.49)	1.5×10^{-9}				5	DDB1	DDB1
11:61141476	rs7394502	G>A	−2.41(0.40)	1.6×10^{-9}				5	TMEM138	TMEM138
11:61141164	rs10897155	C>T	−2.41(0.40)	1.6×10^{-9}			Y	5	TMEM138	TMEM138
11:61139869	rs11230678	G>A	−2.41(0.40)	1.7×10^{-9}				5	TMEM138	TMEM138
11:61115821	rs148172827	C/CATCAA	−2.95(0.49)	1.8×10^{-9}	X	X	Y	5	TKFC	CYB561A3;TKFC,DDB1,TMEM138
11:61144707	rs1377458	C>T	−2.40(0.40)	2.1×10^{-9}		X		5	TMEM138	CYB561A3;TKFC,DDB1,TMEM138
11:61076372	rs11230664	C>T	2.95(0.49)	2.1×10^{-9}		X	N	5	DDB1	DDB1
11:61122878	rs7951574	G>A	2.92(0.49)	2.8×10^{-9}				5	CYB561A3	
11:61054892	rs1108769	A>C	2.82(0.47)	3.0×10^{-9}		X		5	VWCE	TKFC,DDB1
11:61141259	rs57265008	T>C	2.34(0.39)	3.7×10^{-9}			Y	4	TMEM138	TMEM138
11:61222635	rs3017597	G>A	−2.77(0.47)	5.4×10^{-9}				5	SDHAF2	
11:61075524	rs12275843	T>C	2.64(0.45)	5.5×10^{-9}				5	DDB1	DDB1
11:61043773	rs73490303	G>C	2.67(0.46)	7.2×10^{-9}				5	VWCE	VWCE
11:61018855	rs653173	A>G	2.68(0.46)	8.2×10^{-9}		X		5	PGA5	
11:61063156	rs10897150	G>T	2.79(0.48)	8.8×10^{-9}		X		5	VWCE	TKFC,DDB1,VWCE
11:61108974	rs2260655	G>A	2.63(0.46)	9.0×10^{-9}				5	TKFC	CYB561A3;TKFC,DDB1,TMEM138
11:61152028	rs12791961	C>A	2.90(0.50)	9.7×10^{-9}		X		5	TMEM216	CYB561A3;TKFC,DDB1,TMEM138,TMEM216
11:61055014		GACTA/G	2.61(0.45)	1.1×10^{-8}		X		5	VWCE	TKFC,DDB1
11:61080557	rs7120594	T>C	2.58(0.45)	1.2×10^{-8}				5	DDB1	
11:61044470	rs9704187	G>C	2.58(0.45)	1.3×10^{-8}				5	VWCE	VWCE
11:61106892	rs2513329	G>C	−2.73(0.48)	1.6×10^{-8}			N	5	TKFC	CYB561A3;TKFC,DDB1,TMEM138,TMEM216
11:61033525	rs2001746	T>A	2.56(0.45)	1.7×10^{-8}				5	VWCE	VWCE
11:61112802	rs2305465	C>T	2.62(0.46)	1.8×10^{-8}	X	X		5	TKFC	CYB561A3;TKFC,DDB1,TMEM138
11:61037389		ATT/A	2.51(0.45)	3.5×10^{-8}				5	VWCE	
15:28514281	rs4932620	C>T	−2.85(0.48)	3.2×10^{-9}				6	HERC2	
15:28532639	rs1667393	C>T	−2.82(0.48)	6.3×10^{-9}		X	N	6	HERC2	

Location	RSID	Ancestral> Derived	Beta(SE)	p-value	DHS Melanocytes	DHS Other	Luciferase Activity	LD Block	Nearest Gene	Chromatin Interactions
15:28535675	rs1635167	C>T	−2.88(0.50)	8.9×10^{-9}				6	*HERC2*	
15:28545148	rs2905952	A>G	−3.16(0.55)	9.0×10^{-9}				6	*HERC2*	
15:28396894	rs12915877	T>G	−2.76(0.48)	1.1×10^{-8}	X			6	*HERC2*	
15:28487069	rs4932618	G>A	−2.69(0.47)	1.6×10^{-8}				6	*HERC2*	
15:28235773	rs1800404	C>T	2.54(0.45)	1.6×10^{-8}		X		7	*OCA2*	
15:28238158	rs1868333	G>A	−2.53(0.45)	2.2×10^{-8}				7	*OCA2*	
15:28419497	.	TA/T	−3.73(0.67)	2.6×10^{-8}				6	*HERC2*	
15:28238895	rs735066	A>G	−2.50(0.45)	3.5×10^{-8}	X			7	*OCA2*	