RESEARCH ARTICLE SUMMARY

HUMAN EVOLUTION

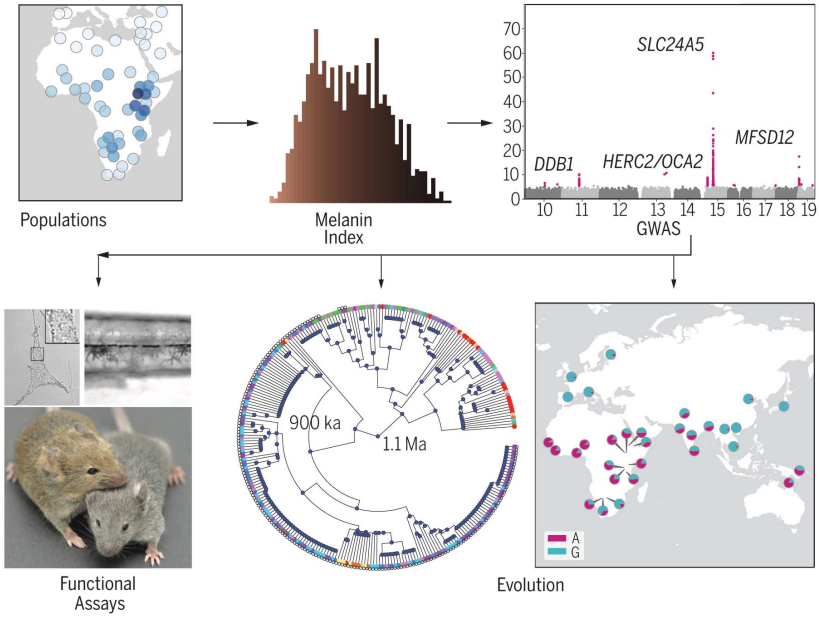
Loci associated with skin pigmentation identified in African populations

Nicholas G. Crawford, Derek E. Kelly,* Matthew E. B. Hansen,* Marcia H. Beltrame,* Shaohua Fan,* Shanna L. Bowman,* Ethan Jewett,* Alessia Ranciaro, Simon Thompson, Yancy Lo, Susanne P. Pfeifer, Jeffrey D. Jensen, Michael C. Campbell, William Beggs, Farhad Hormozdiari, Sununguko Wata Mpoloka, Gaonyadiwe George Mokone, Thomas Nyambo, Dawit Wolde Meskel, Gurja Belay, Jake Haut, NISC Comparative Sequencing Program, Harriet Rothschild, Leonard Zon, Yi Zhou, Michael A. Kovacs, Mai Xu, Tongwu Zhang, Kevin Bishop, Jason Sinclair, Cecilia Rivas, Eugene Elliot, Jiyeon Choi, Shengchao A. Li, Belynda Hicks, Shawn Burgess, Christian Abnet, Dawn E. Watkins-Chow, Elena Oceana, Yun S. Song, Eleazar Eskin, Kevin M. Brown, Michael S. Marks,† Stacie K. Loftus,† William J. Pavan,† Meredith Yeager,† Stephen Chanock,† Sarah A. Tishkoff‡

INTRODUCTION: Variation in pigmentation among human populations may reflect local adaptation to regional light environments, because dark skin is more photoprotective, whereas pale skin aids the production of vitamin D. Although genes associated with skin pigmentation have been identified in European populations, little is known about the genetic basis of skin pigmentation in Africans.

RATIONALE: Genetically and phenotypically diverse African populations are informative for mapping genetic variants associated with skin pigmentation. Analysis of the genetics of skin pigmentation in Africans informs upon melanocyte biology and the evolution of skin pigmentation in humans.

RESULTS: We observe extensive variation in skin pigmentation in Africa, with lowest melanin levels observed in southern African San hunter-gatherers and highest levels in East African Nilo-Saharan pastoralists. A genome-wide association study (GWAS) of 1570 Africans identified variants significantly associated with skin pigmentation, which clustered in four genomic regions that together account for almost 30% of the phenotypic variation.



GWAS and functional assays illuminate the genetic basis of pigmentation in Africa. A GWAS identified four genomic regions associated with skin pigmentation in Africa. Functional assays in melanocytes, zebrafish, and mice characterized their impact on skin pigmentation. Evolutionary genetic analyses revealed that most derived variants evolved before the origin of modern humans. Ma, million years ago.

The most significantly associated single-nucleotide polymorphisms were at *SLC24A5*, a gene associated with pigmentation in Europeans. We show that *SLC24A5* was introduced into East Africa >5 thousand years ago (ka) and has risen to high frequency.

The second most significantly associated region is near the gene *MFSD12*. Using in vitro and in vivo analyses, we show that *MFSD12* codes for a lysosomal protein that modifies pigmentation in human melanocytes, with decreased *MFSD12* expression associated with darker pigmentation. We also show that genetic knockouts of *MFSD12* orthologs affect pigmentation in both zebrafish and mice.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aan8433

A third highly associated region encompasses a cluster of genes that play a role in ultraviolet (UV) response and DNA damage repair. We find the strongest associations in a regulatory region upstream of *DDB1*, the gene encoding damage-specific DNA binding protein 1, and that these variants are associated with increased expression of *DDB1*. The alleles associated with light pigmentation swept to near fixation outside of Africa due to positive selection, and we show that these lineages coalesce ~60 ka, corresponding with the time of migration of modern humans out of Africa.

The fourth significantly associated region encompasses the *OCA2* and *HERC2* loci. We identify previously uncharacterized variants at *HERC2* associated with the expression of *OCA2*. These variants arose independently from eye and skin pigmentation–associated variants in non-Africans. We also identify variants at *OCA2* that are correlated with alternative splicing; alleles associated with light pigmentation are correlated with a shorter transcript, which lacks a transmembrane domain.

CONCLUSION: We identify previously uncharacterized genes and variants associated with skin pigmentation in ethnically diverse Africans. These genes have diverse functions, from repairing UV damage to playing important roles in melanocyte biology. We show that both dark and light pigmentation alleles arose before the origin of modern humans and that both light and dark pigmented skin has continued to evolve throughout hominid history. We show that variants associated with dark pigmentation in Africans are identical by descent in South Asian and Australo-Melanesian populations. This study sheds light on the evolutionary history, and adaptive significance, of skin pigmentation in humans. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†These authors contributed equally to this work.
‡Corresponding author. Email: tishkoff@pennmedicine.upenn.edu
Cite this article as N. G. Crawford *et al.*, *Science* **358**, eaan8433 (2017). DOI: 10.1126/science.aan8433

Erratum 14 December 2017. See erratum.

RESEARCH ARTICLE

HUMAN EVOLUTION

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]‡ Sarah A. Tishkoff[1,25]§

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 South 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 ultraviolet 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 those 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 differ between melanocytes that synthesize mainly eumelanins (black-brown pigments) or pheomelanins (pigments that 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*).

Although more than 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 2101 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 1570 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 genome-wide association study (GWAS) analysis with linear mixed models, controlling for age, sex, and genetic relatedness (*9*), identified four regions with multiple significant associations ($P < 5 \times 10^{-8}$) (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) (Fig. 3 and table S3) (*10*). We ranked potential causal variants

[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. [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, Dar es 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 and Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA 90095, USA.
*These authors contributed equally to this work. †National Institutes of Health Intramural Sequencing Center (NISC) Comparative Sequencing Program collaborators and affiliations are listed in the supplementary materials. ‡These authors contributed equally to this work. §Corresponding author. Email: tishkoff@pennmedicine.upenn.edu

within each locus using CAVIAR, a fine-mapping method that accounts for linkage disequilibrium (LD) and effect sizes (Table 1) (*11*). 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 nonsynonymous mutation within *SLC24A5* (rs1426654) (*14*) was significantly associated with skin color (*F* test, $P = 5.5 \times 10^{-62}$) and was identified as potentially causal by CAVIAR (Table 1). The rs1426654 (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*–*18*). In Africa, this variant is common (28 to 50% frequency) in populations from Ethiopia and Tanzania with high Afro-Asiatic ancestry (*19*, *20*) and is at moderate frequency (5 to 11%) in San and Bantu-speaking populations from Botswana with low levels of East African ancestry and recent European admixture (Fig. 1 and figs. S2 and S4) (*21*, *22*). We observe a signature consistent with positive selection at *SLC24A5* in Europeans based on extreme values of Tajima's *D* statistic (fig. S5).

On the basis of 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 rs1426654 (A) allele is 29 thousand years ago (ka) [95% critical interval (CI), 28 to 31 ka], consistent with previous studies (*15*, *17*) (Fig. 4). Haplotype analysis indicates that the rs1426654 (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 to 9 ky (*23*). The rs1426654 (A) variant is at high frequency (28%) in Tanzanian populations, suggesting a lower bound (~5 ka) for introduction of this allele into East Africa, the time of earliest migration from Ethiopia into Tanzania (*24*). Furthermore, the frequency of the rs1426654 (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 rs1426654 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 rs1426654 (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 (*25*). *MFSD12* mRNA levels are low in depigmented skin of vitiligo patients (*26*), 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 ~7600 to 9000 base pairs (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 the highest probability of being causal are rs56203814 (*F* test, $P = 3.6 \times 10^{-18}$), a synonymous variant within exon 9, and rs10424065 (*F* test, $P = 5.1 \times 10^{-20}$), located within intron 8. They are 130 bp apart, are in strong LD, and affect gene expression in luciferase expression assays (1.5 to 2.7× higher

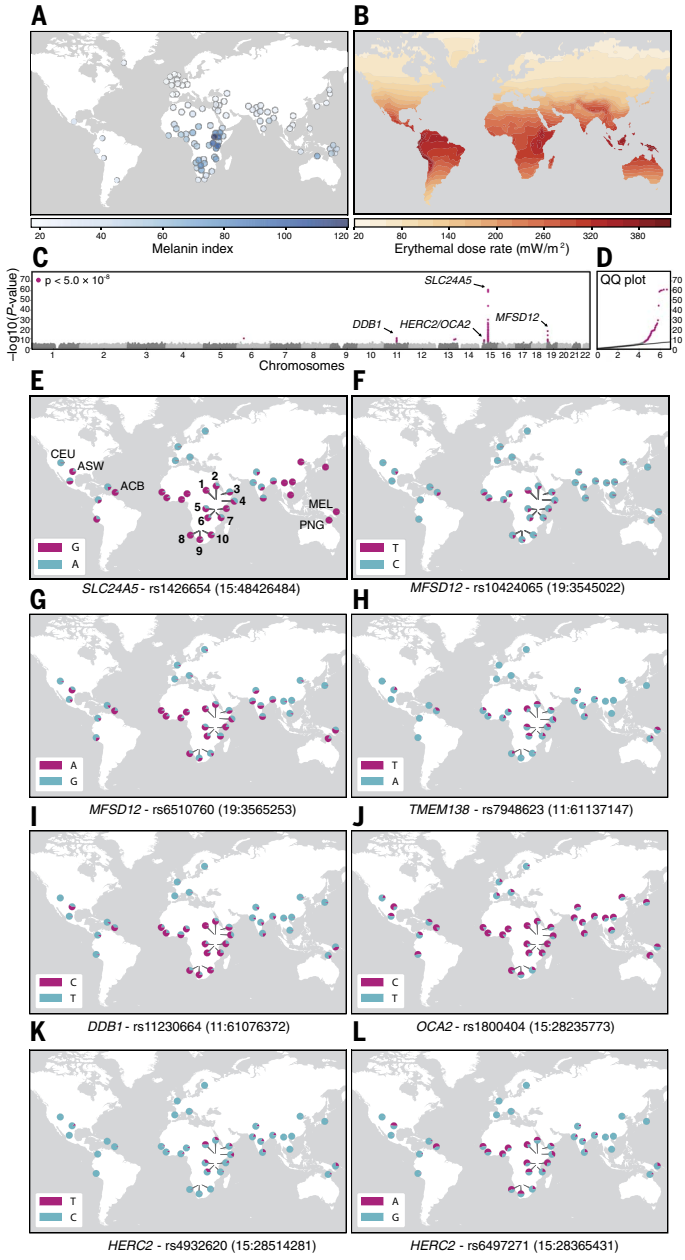


Fig. 1. Correlations between allele frequencies at loci associated with pigmentation and UV exposure in global populations. (**A**) Global variation in skin pigmentation indicated by melanin index (MI). These data were integrated with MI data for global populations from (*1*, *105*). (**B**) Mean erythemal dose rate. (**C**) Manhattan plot of $-\log_{10}$ transformed *P* values from GWAS of skin pigmentation with the Illumina Omni5 SNP array. (**D**) Quantile-quantile (QQ) plot of observed versus expected *P* values from the GWAS. In both (C) and (D), significant SNPs at $P < 5 \times 10^{-8}$ are highlighted in purple. (**E** to **L**) Allele frequencies of genetic variants associated with skin pigmentation in global populations. African populations included are Ethiopia Nilo-Saharan (1), Ethiopia Omotic (2), Ethiopia Semitic (3), Ethiopia and Tanzania Cushitic (4), Tanzania Nilo-Saharan (5), Tanzania Hadza (6), Tanzania Sandawe (7), Botswana Bantu (8), Botswana San/Bantu admixed (9), and Botswana San (10). 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*). Non-Aboriginal populations in the Americas are indicated: CEU (European ancestry), ASW (African-American Southwest U.S.), and ACB (African Caribbean in Barbados).

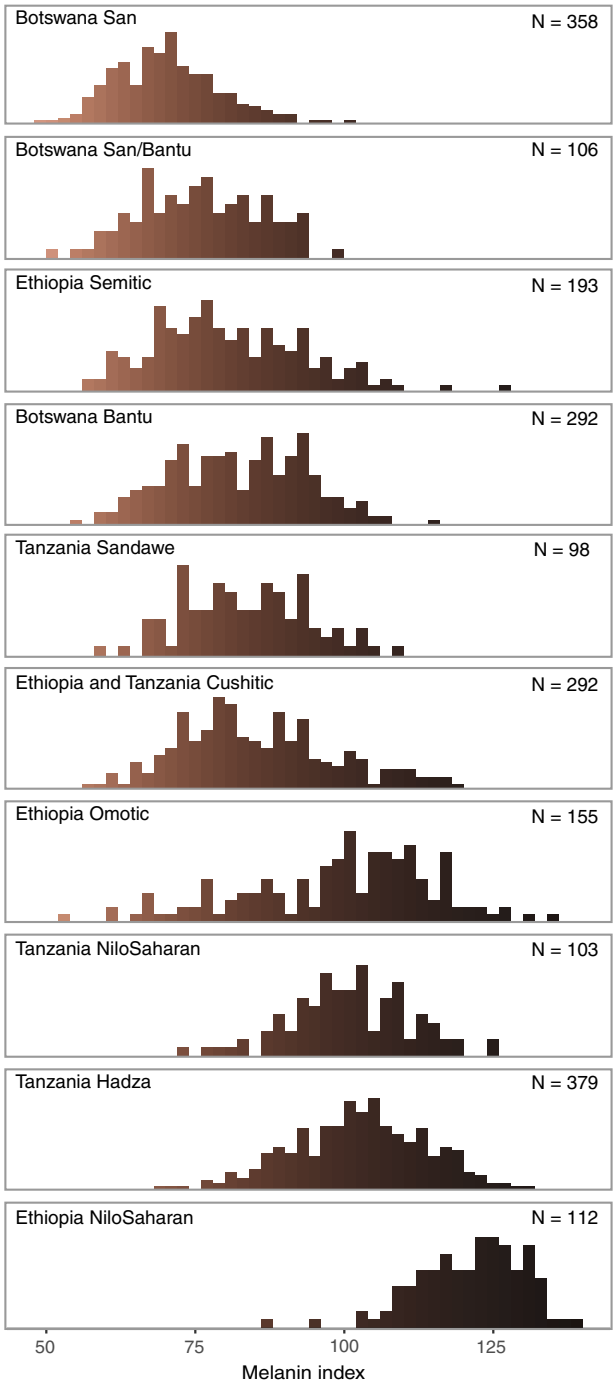


Fig. 2. Melanin distributions. Histograms of melanin index computed from underarm measurements with a DSM II ColorMeter for all individuals in each population as described in (*70*). Skin tones were visualized by displaying the scaled mean red, green, and blue values from the ColorMeter for individuals binned by melanin index.

expression than the minimal promoter; fig. S7). The SNPs upstream of *MFSD12* with highest probability of being causal are rs112332856 (*F* test, $P = 3.8 \times 10^{-16}$) and rs6510760 (*F* test, $P = 6.5 \times 10^{-15}$). They are 346 bp apart, are in strong LD, and affect gene expression in luciferase expression assays (4.0 to 19.7× higher expression than the minimal promoter; fig. S7).

The derived rs56203814 (T) and rs10424065 (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 data set indicates that the rs10424065 (T) allele predates the 300-ka origin of modern humans (estimated TMRCA of 612 ka; 95% CI, 515 to 736 ka) (Fig. 4) (*27*).

At rs6510760 and rs112332856, 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 Afro-Asiatic 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 rs112332856 (C) allele] in Australo-Melanesians on similar haplotype backgrounds relative to central and eastern Africans (Fig. 5 and fig. S6), suggesting that they are identical by descent from an ancestral African population. Coalescent analysis of the SGDP data set indicates that the TMRCA for the derived rs6510760 (A) allele is 996 ka [95% CI, 0.82 to 1.2 million years ago (Ma); 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 (*28*)]. However, levels of genetic differentiation are elevated when comparing East African Nilo-Saharan and western European (CEU) populations (for example, $F_{ST} = 0.85$ for rs112332856, top 0.05% on chromosome 19), consistent with differential selection at this locus (table S4) (*29*).

MFSD12 is within a cluster of 10 genes with high expression levels in primary human melanocytes relative to primary human keratinocytes (*30*), with *MFSD12* as the most differentially expressed (90×; table S6). The genomic region (chr19:3541782-3581062) encompassing *MFSD12* and neighboring gene *HMG20B* (a transcription factor common in melanocytes) has numerous deoxyribonuclease (DNase) I hypersensitive sites (DHS) and is enriched for H3K27ac enhancer marks in melanocytes (top 0.1% genome-wide; Fig. 3), suggesting that this region may regulate expression of genes critical to melanocyte function (*31*).

Analyses of gene expression using RNA sequencing (RNA-seq) data from 106 primary melanocyte cultures (table S7) indicate that African ancestry is significantly correlated with decreased *MFSD12* gene expression [Pearson correlation coefficient (PCC), $P = 5.0 \times 10^{-2}$; fig. S9]. We observed significant associations between genotypes at rs6510760 and rs112332856 with expression of *HMG20B* [Bonferroni-adjusted *P* (P_{adj}) < 4.9×10^{-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 rs112332856 show increased allelic imbalance, relative to homozygotes, for *MFSD12* (Mann-Whitney-Wilcoxon test, $P = 4.9 \times 10^{-3}$ and 1.3×10^{-2}, respectively), consistent with regulation of gene expression in cis. A haplotype containing the rs6510760 (A)/rs112332856 (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 test, $P = 7.7 \times 10^{-7}$; fig. S7 and table S5). We did not have power to detect an association between expression of *MFSD12* and rs56203814 or rs10424065 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 (*Mfsd12*) using short hairpin RNAs (shRNAs) in immortalized melan-Ink4a mouse melanocytes derived from C57BL/6J-Ink4a$^{-/-}$ mice (*32*), which almost exclusively make eumelanin (Fig. 6). Reduction of *Mfsd12* mRNA by ~80% with two distinct lentivirally encoded shRNAs (Fig. 6A) caused a 30 to 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 hemagglutinin (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 colocalized with LAMP2 (Fig. 6E; quantified in Fig. 6H), indicating that the 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 *mfsd12a* with CRISPR-Cas9 (Fig. 7). We focused on *mfsd12a* because its paralog *mfsd12b* is predicted to be a pseudogene (*33*). 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) (*34*, *35*). This was not due to a failure of the xanthophores to develop in *mfsd12a* mutants, because green fluorescent protein (GFP)–labeled xanthophores were robust along the lateral line in both wild-type and *mfsd12a* mutant zebrafish (Fig. 7, C and D). Together, these results suggest that *MFSD12* influences xanthophore pigment production in pterinosomes.

Functional characterization of *MFSD12* in mice

CRISPR-Cas9 was used to generate an *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

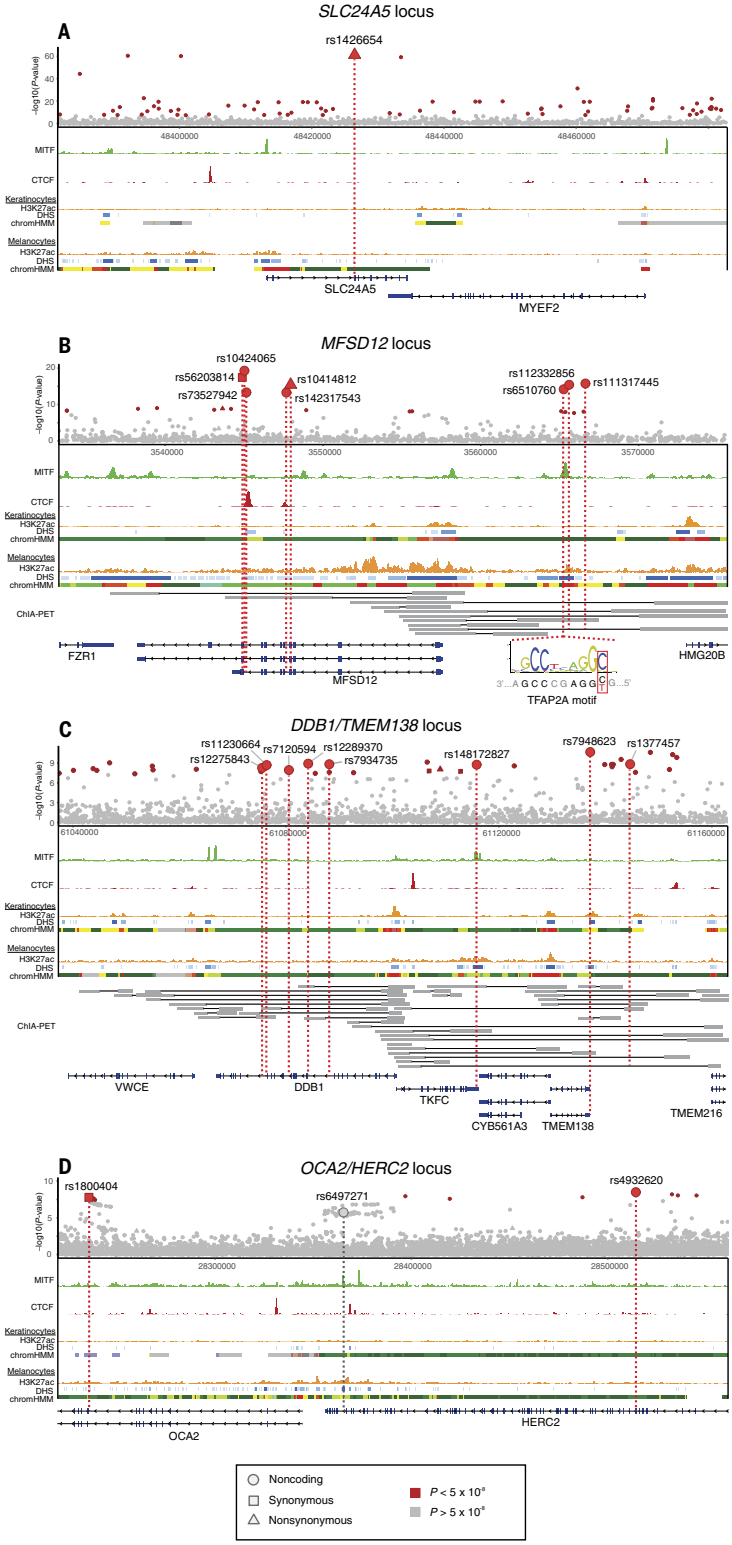


Fig. 3. Genomic context of GWAS loci. Plot of $-\log_{10}(P \text{ value})$ versus genomic position for variants near the four regions with most strongly associated SNPs from GWAS, including annotations for genes, MITF ChIP-seq (chromatin immunoprecipitation sequencing) data for melanocytes (*48*), a CTCF ChIP-seq track for NHEK keratinocytes, and H3K27ac, DNase I hypersensitive sites (DHS), and chromHMM tracks for melanocytes and keratinocytes from the Roadmap Epigenomics data set (*30*). Genome-wide significant variants are highlighted in red. Circles, squares, and triangles denote noncoding, synonymous, and nonsynonymous variants, respectively. (**A**) *SLC24A5* locus. (**B**) *MFSD12* locus. (**C**) *DDB1/TMEM138* locus. (**D**) *OCA2/HERC2* locus.

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* (*36*). 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 subterminal 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 exon 2 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.

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. Because 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 (*37*), and, like zebrafish pterinosomes, a potentially different intracellular origin from eumelanosomes (*38*). Although disruption of *MFSD12* alone accounts for changes in pigmentation, the role of neighboring loci such as *HMG20B* on pigmentation remains to be explored.

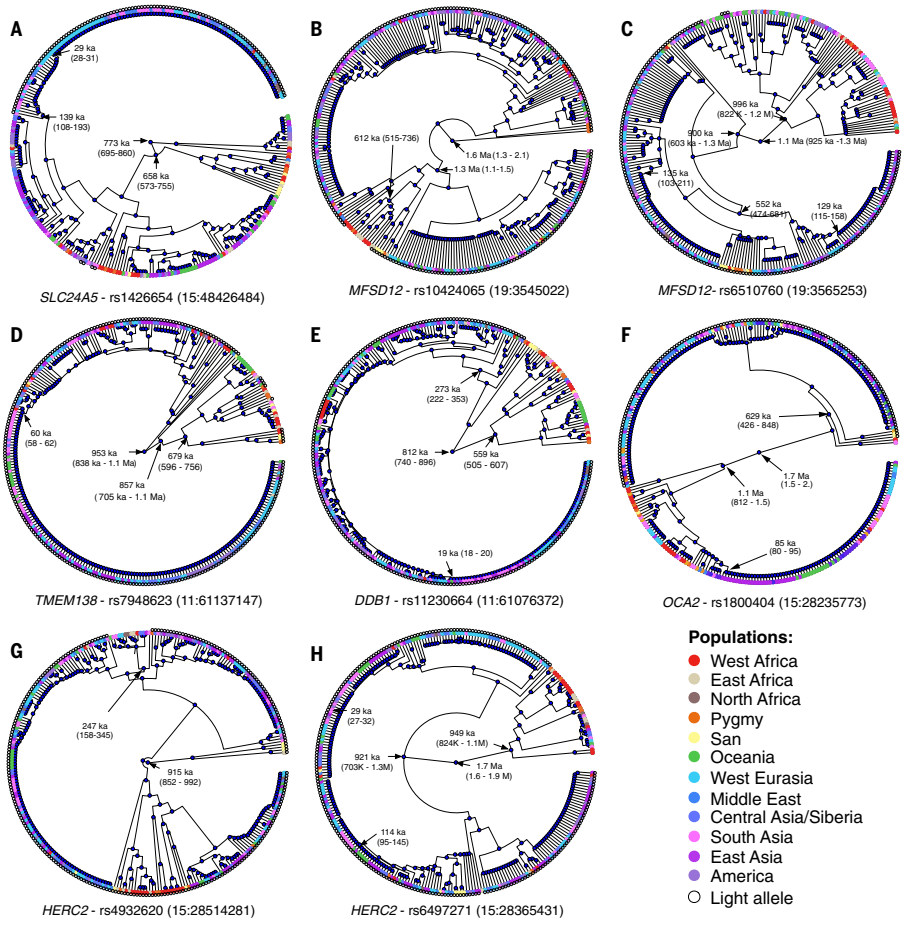


Fig. 4. Coalescent trees and TMRCA dating. Inferred genealogies for regions flanking candidate causal loci. Each leaf corresponds to a single sampled chromosome from 1 of 278 individuals in the Simons Genome Diversity Project (*13*). Leaf nodes are colored by the population of origin of the individual, and sequences carrying the light allele are indicated with an open circle, located next to the leaf node. Node heights and 95% CI 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).

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 (*39*); levels of DDB1 are regulated by UV exposure and MC1R signaling, a regulatory pathway of pigmentation (*40*). DDB1 is a component of CUL4-RING E3 ubiquitin ligases that regulate several cellular and developmental processes (*41*); it is critical for follicle maintenance and female fertility in mammals (*42*) and for plastid size and fruit pigmentation in tomatoes (*43*). Knockouts of *DDB1* orthologs are lethal in both mouse and fruitfly development (*44*), and *DDB1* only exhibits rare (<1% frequency) nonsynonymous mutations in the TGP data set. Genetic variants near *DDB1* were associated with human pigmentation in an African population with high levels of European admixture (*7*).

Because of 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* (*45*) ~36 to 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, $P = 2.2 \times 10^{-11}$), located 172 bp downstream of *TMEM138*, which shows enhancer activity in WM88 melanoma cells (91.9 to 140.8× higher than the minimal promoter; fig. S7 and table S5) and interacts with the promoters of *DDB1* and neighboring genes in MCF-7 cells (Table 1 and Fig. 3) (*46*, *47*).

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, $P = 1.5 \times 10^{-9}$), located ~7600 bp downstream of *TMEM138*, and rs148172827 (*F* test, $P = 1.8 \times 10^{-9}$), an insertion/deletion polymorphism at *TKFC* (triokinase and FMN cyclase) located in an enhancer active in WM88 melanoma cells (67.6 to 76.2× higher than the minimal promoter; fig. S7 and table S5), which overlaps an *MITF* binding site in melanocytes (*30*, *48*); both SNPs interact with the promoters of *DDB1* and neighboring genes in MCF-7 cells (Table 1 and Fig. 3) (*46*, *47*). 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

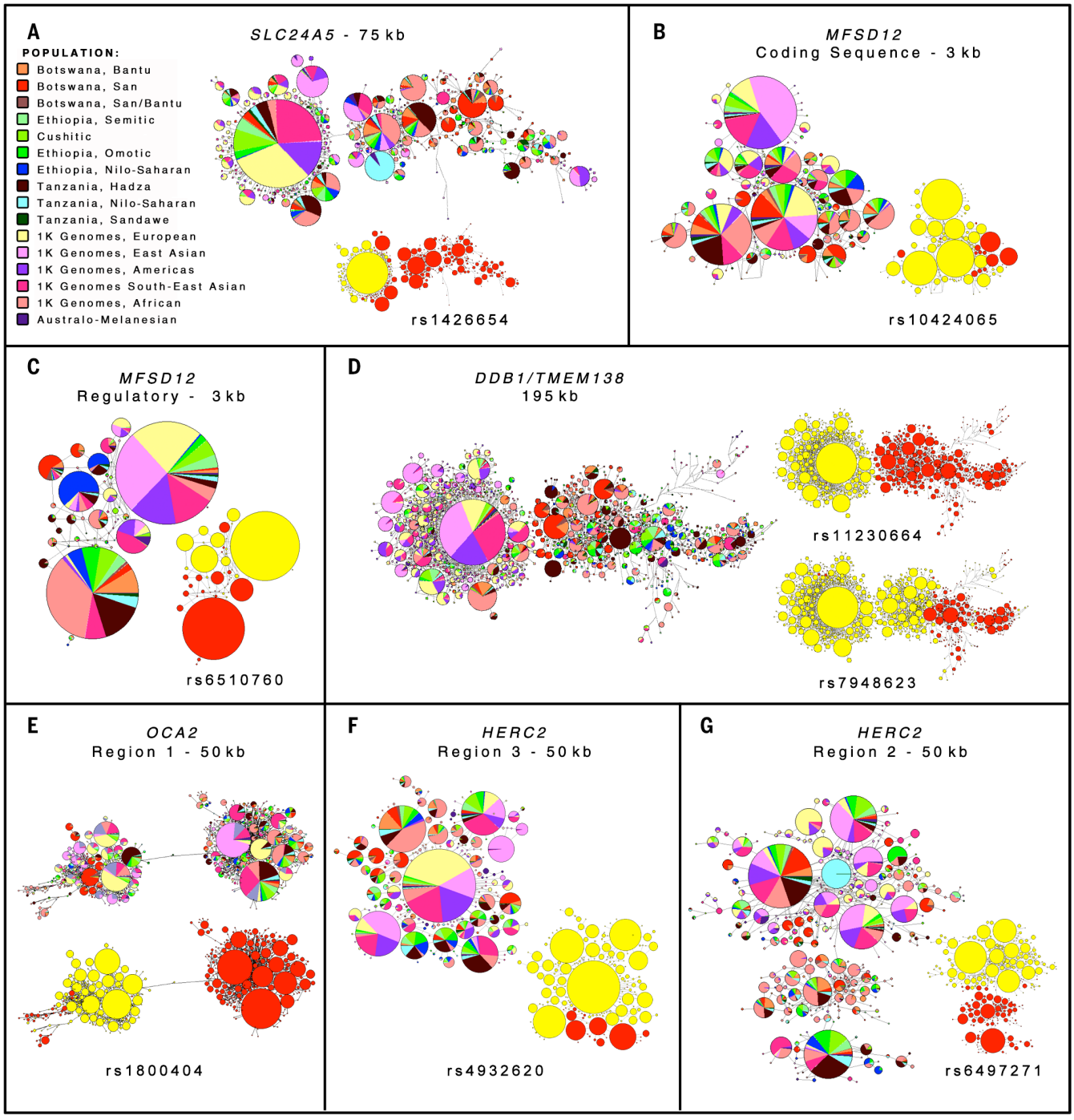


Fig. 5. Haplotype networks at *SLC24A5*, *MFSD12*, *DDB1/TMEM138*, and *OCA2/HERC2*. Median-joining haplotype networks of regions containing candidate causal variants. Connections between circles indicate genetic relatedness, whereas size is relative to the frequency of haplotypes. Ancestry proportions are displayed as pie charts. Yellow and red subfigures indicate which haplotypes contain the allele associated with dark pigmentation (red) or light pigmentation (yellow). (**A**) Region (75 kb) flanking the causal variant at *SLC24A5*. (**B** and **C**) Regions (3 kb) flanking rs10424065 in *MFSD12* and rs6510760 upstream of *MFSD12*. (**D**) Region (195 kb) flanking *DDB1* extending from *PGA5* to *SDHAF2*. (**E** to **G**) Regions 1, 3, and 2 (50 kb) at *OCA2* and *HERC2* (ordered based on highest to lowest probability of being causal from CAVIAR analysis).

correlated with increased *DDB1* gene expression (PCC, $P = 2.6 \times 10^{-5}$; fig. S9). Association tests using a permutation approach indicated that, of the 35 protein-coding genes with a transcription start site within 1 Mb 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 the 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

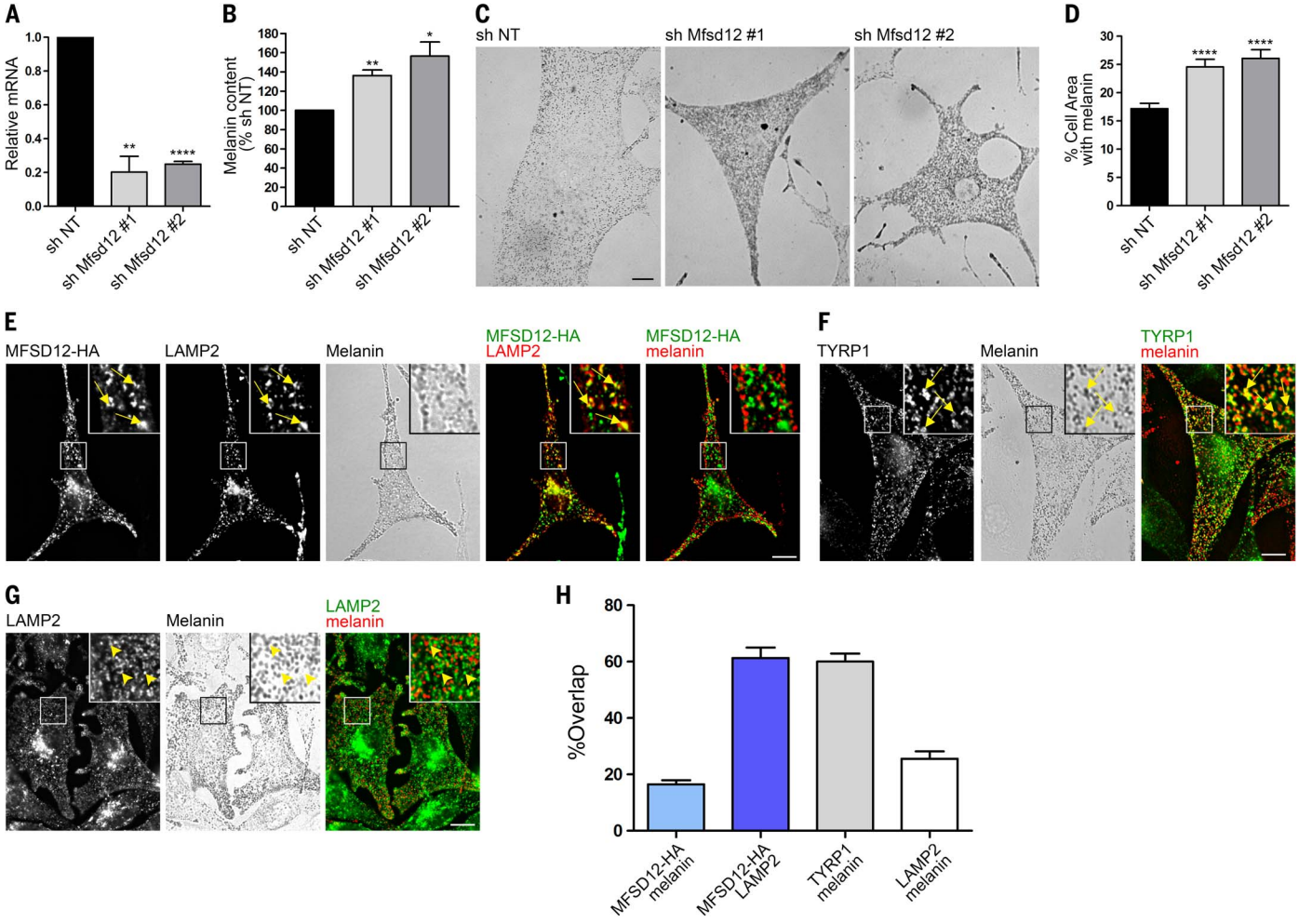


Fig. 6. MFSD12 suppresses eumelanin production but localizes to lysosomes. Immortalized melan-Ink4a melanocytes expressing nontarget (sh NT) shRNA or either of two shRNA plasmid clones (#1 and 2) targeting *Mfsd12* were analyzed for (**A**) *Mfsd12* mRNA content by quantitative reverse transcription polymerase chain reaction (qRT-PCR), (**B**) melanin content by spectrophotometry, or (**C**) percentage of cell area containing melanin by bright-field microscopy. (**D**) Quantification. Data in (A) to (C) represent means ± SEM, normalized to sh NT samples, from three separate experiments. In (C), *n* (sh NT) = 97 cells, *n* (shMfsd12 #1) = 68 cells, and *n* (shMfsd12 #2) = 71 cells. Scale bar, 10 μm (C). (**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 (F), and analyzed by immunofluorescence and bright-field microscopy. Bright-field (melanin) images show pigmented melanosomes (pseudo-colored red in the merged images). Insets, 4× magnification of boxed regions. Arrows, MFSD12-containing structures that overlap LAMP2 (E) or TYRP1-containing structures that overlap melanosomes (F); arrowheads, structures that do not overlap (G). Scale bars, 10 μm. (**H**) Quantification of overlap for structures labeled by MFSD12, TYRP1, LAMP2, and pigment. Data represent means ± SEM 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).

in 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 to 96%), and is at moderate to high frequency in South Asian and Australo-Melanesian populations (12 to 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 haplotypes closely related, or identical, 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 and 250 ka, 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 to 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 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 ka (95% CI, 58 to 62 ka), close to the inferred time of the migration of modern humans out of Africa (Fig. 4) (*49*). These results, combined with large F_{ST} values between Africans and Europeans at SNPs tagging the extended haplotype near *DDB1* (for example, F_{ST} = 0.98 between Nilo-Saharans and CEU at rs7948623, within the top 0.01% of values

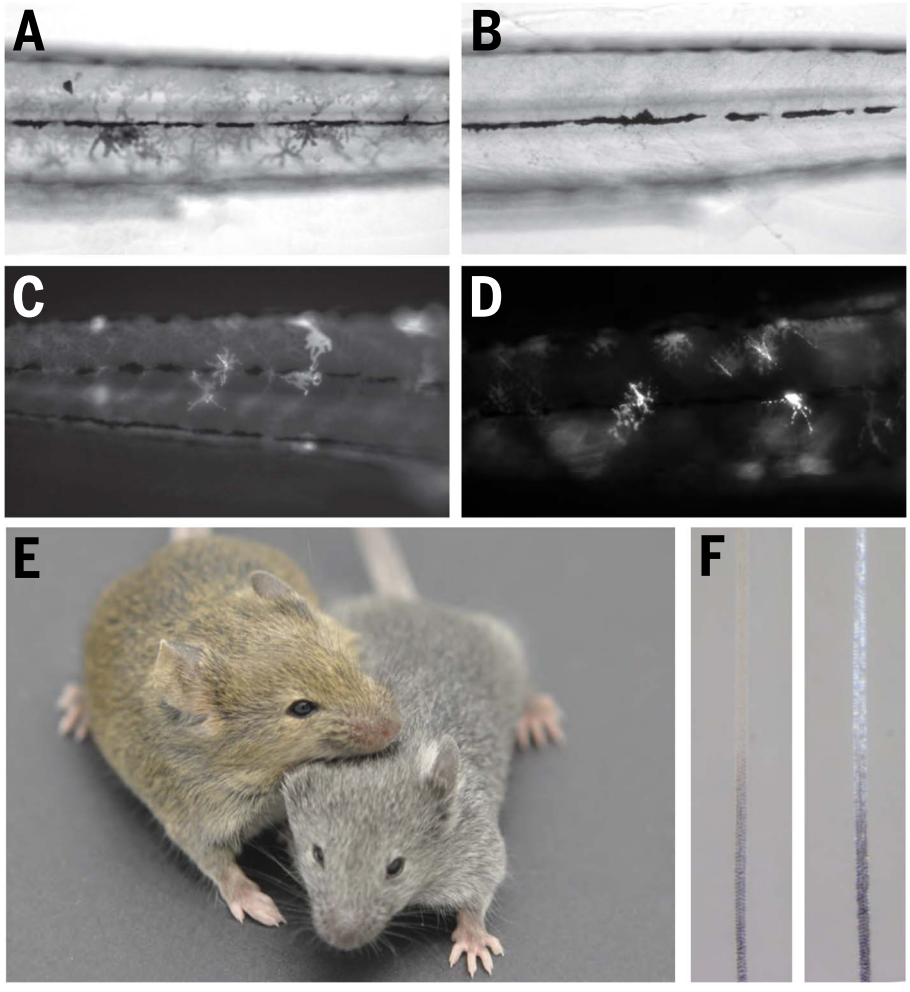


Fig. 7. In vivo zebrafish and mouse models of MFSD12 deficiency. (**A** and **B**) Representative images of methylene blue staining in wild-type TAB5 (A) and compound heterozygous mutant *mfsd12a* zebrafish (6 days postfertilization) (B). Note the absence of stained xanthophores in the *mfsd12a* mutant (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 (5 days postfertilization). (**E**) Wild-type agouti mouse (left) with a gray *Mfsd12*-targeted littermate (right). (**F**) Hair from the *Mfsd12*-targeted mouse has grossly normal eumelanin (lower black region of the hair shaft); however, the upper subapical yellow band in wild-type (E, left) appears white in the *Mfsd12* mutant (E, right) due to a reduction in pheomelanin.

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* affecting 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 (*5*–*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*).

Because of 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 based on 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, $P = 1.0$), a synonymous variant located in region 1 within exon 10 of *OCA2* (Fig. 3, Table 1, and table S3) associated with eye color in Europeans (*55*). The ancestral rs1800404 (C) allele, associated with dark pigmentation, is common in most Africans as well as southern and 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), consistent with a previous observation (*56*). 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 Ma (95% CI, 1.5 to 2.0 Ma), and the TMRCA of lineages containing the derived (T) allele is 629 ka (95% CI, 426 to 848 ka) (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, $P = 3.2 \times 10^{-9}$) 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 (*57*, *58*), 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 that it is identical by descent. The TMRCA of haplotypes containing the rs4932620 (T) allele is 247 ka (95% CI, 158 to 345 ka) (Fig. 4).

We also observe an LD block of SNPs within region 2 of *HERC2* that are associated with skin pigmentation, although 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, $P = 1.8 \times 10^{-6}$), 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 that modulates levels of *MITF* in melanocytes) (Fig. 3) (*59*). 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 that 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 ka (95% CI, 703 ka to 1.2 Ma) (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 to 0.85, top 0.5% on chromosome 15) (table S4).

Analyses of RNA-seq data from 106 primary melanocyte cultures indicate 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 ($P_{adj} = 2.2 \times 10^{-2}$).

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 (*30*, *92*). Variants intersecting enhancer regions tested by luciferase assays were labeled with Y (significant enhancer activity) or N (no enhancer activity) (fig. S7). Chromatin interactions with nearby genes measured in MCF-7 or K562 cell lines as identified by ChIA-PET are listed with gene names (Chromatin interactions) (*46*, *47*). 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*	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			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*
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*	
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*

continued on next page

Location	RSID	Ancestral>Derived	Beta(SE)	P	DHS melanocytes	DHS other	Luciferase activity	LD block	Nearest gene	Chromatin interactions
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*	
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*	
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*	

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, $P = 9.1 \times 10^{-40}$). Exon 10 encodes the amino acids encompassing the third transmembrane domain of *OCA2* and is the location of several albinism-associated *OCA2* mutations (*60*, *61*), raising the possibility that the shorter transcript encodes a nonfunctional 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 to 21.5%; fig. S9). Therefore, homozygotes for the light rs1800404 (T) allele are 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. About 28.9% (SE, 10.6%) of the pigmentation variance is attributable to these SNPs. Considering each locus in turn and all significantly associated variants ($P < 5 \times 10^{-8}$), the trait variation attributable to each locus is as follows: *SLC24A5* (12.8%; SE, 3.5%), *MFSD12* (4.5%; SE, 2.1%), *DDB1/TMEM138* (2.2%; SE, 1.5%), and *OCA2/HERC2* (3.9%; SE, 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 (that is, fewer genes of stronger effect) than other complex traits, such as height (*62*). In addition, most candidate causal variants are in noncoding 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 the most lightly pigmented among Africans, whereas 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 data set are estimated to have originated before the origin of modern humans ~300 ka (*27*). In contrast to the lack of variation at *MC1R*, which is under purifying selection in Africa (*63*), 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). Furthermore, the ancestral allele is associated with light pigmentation in about half of the predicted causal SNPs; Neandertal and Denisovan genome sequences, which diverged from modern human sequences 804 ka (*64*), 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 (My) after human ancestors lost most of their protective body hair, although these ancestral hominins may have been moderately, rather than darkly, pigmented (*65*, *66*). Moreover, it appears that both light and dark pigmentation have 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 (*67*). This observation is consistent with a proposed southern migration route out of Africa ~80 ka (*68*). Alternatively, it is possible that light and dark pigmentation alleles segregated in a single African source population (*13*, *69*) 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 European populations. The loci identified in this study appear to affect multiple phenotypes. For

example, *DDB1* influences pigmentation (*43*), cellular response to the mutagenic effect of UVR (*40*), and female fertility (*42*). Thus, some of the pigmentation-associated variants identified here may be maintained because of 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 affect the pigmentation phenotype. 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 (*70*). 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 (*70*, *71*). DNA was extracted from whole blood using a salting out procedure (PureGen).

A total of 1570 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 (*72*) 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×) 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* (*73*). 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 (*74*) and duplicate regions identified with *Delly* (*75*).

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 1 Mb (500 kb upstream and 500 kb downstream) of top GWAS variants from each region and phased them using *SHAPEIT2* (*76*, *77*). The reference panel came from two datasets: filtered variants from the 135 African genomes and TGP (*10*). After phasing, imputation was performed using *Minimac3* (*78*). 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 (*79*).

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 (*80*). The variance parameters for two genetic relationship matrices (GRMs) are estimated: one GRM is constructed (*81*) 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 ($P < 5 \times 10^{-8}$) 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 (*82*–*84*). 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} (*85*). To calculate extended haplotype homozygosity (iHS) we used *Selscan* (*86*). Unstandardized iHS scores were normalized within 100 kb 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 (*83*). To identify outlier windows we calculated 5th and 95th percentiles. Population differentiation was assessed with Weir and Cockerham's fixation index F_{ST} (*82*) between each pair of populations. Outliers were identified using empirical P values.

Median joining haplotype networks (*87*) for the Illumina Omni 5M SNP dataset were constructed and visualized at genomic regions of interest using *NETWORK* (*88*). 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* (*89*) 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 (*70*, *90*). 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 (*70*) 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 (*91*), DNase I hypersensitivity peaks, H3K27ac signal tracks for keratinocytes and melanocytes (*30*), CTCF signal tracks from keratinocytes (*92*), and ChIP-seq signal tracks from MITF (*48*). Variants were intersected with chromatin annotations using *bedtools* (*93*). Functional consequences of variants were also assessed using deepSEA (*94*) and deltaSVM (*95*). The effect of genetic variants on transcription factor binding was predicted using the *MEME* suite (*96*) for all transcription factors in the JASPAR 2016 CORE Vertebrate motif set (*97*).

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 (*78*) based on the TGP reference panel and using *SHAPEIT* for phasing (*76*). RNA sequencing was performed to a mean depth of 87 million reads per sample. *STAR* (*98*) was used for aligning reads, and *RSEM* (*99*) 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 (*100*) 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.pca* module of *GLU* (https://github.com/bioinformed/glu-genetics). 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 1 Mb 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 (*70*). This was repeated for each gene, and focal variants across all genes were adjusted for multiple testing using the Bonferroni correction (*101*).

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 exon 9 to exon 10 and exon 9 to exon 11 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 pGL4.23 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 (*32*). 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 to 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 (*102*), 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 (*103*) 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, two-tailed student's *t* tests: $P < 0.05$: *, $P < 0.01$: **, $P < 0.001$: ***, $P < 0.0001$: ****. Details are provided in (*70*).

Zebrafish mutagenesis using CRISPR-Cas9 was performed to target mfsd12a (*70*). 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 1x tricane solution and placed in 100 μl 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 134 bp 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 gray 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* (*104*). 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.

REFERENCES AND NOTES

1. N. G. Jablonski, G. Chaplin, The evolution of human skin coloration. *J. Hum. Evol.* **39**, 57–106 (2000). doi: 10.1006/jhev.2000.0403; pmid: 10896812
2. M. S. Marks, M. C. Seabra, The melanosome: Membrane dynamics in black and white. *Nat. Rev. Mol. Cell Biol.* **2**, 738–748 (2001). doi: 10.1038/35096009; pmid: 11584301
3. F. H. Moyer, Genetic variations in the fine structure and ontogeny of mouse melanin granules. *Am. Zool.* **6**, 43–66 (1966). pmid: 5902512
4. J. P. Ebanks *et al*., Epidermal keratinocytes from light vs. dark skin exhibit differential degradation of melanosomes. *J. Invest. Dermatol.* **131**, 1226–1233 (2011). doi: 10.1038/jid.2011.22; pmid: 21326292
5. F. Liu *et al*., Genetics of skin color variation in Europeans: Genome-wide association studies with functional follow-up. *Hum. Genet.* **134**, 823–835 (2015). doi: 10.1007/s00439-015-1559-0; pmid: 25963972
6. S. Beleza *et al*., Genetic architecture of skin and eye color in an African-European admixed population. *PLOs Genet.* **9**, e1003372 (2013). doi: 10.1371/journal.pgen.1003372; pmid: 23555287
7. L. R. Lloyd-Jones *et al*., Inference on the genetic basis of eye and skin color in an admixed population via Bayesian linear mixed models. *Genetics* **206**, 1113–1126 (2017). doi: 10.1534/genetics.116.193383; pmid: 28381588
8. S. A. Tishkoff *et al*., The genetic structure and history of Africans and African Americans. *Science* **324**, 1035–1044 (2009). doi: 10.1126/science.1172257; pmid: 19407144
9. H. M. Kang *et al*., Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348–354 (2010). doi: 10.1038/ng.548; pmid: 20208533
10. The 1000 Genomes Project Consortium, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). doi: 10.1038/nature15393; pmid: 26432245
11. F. Hormozdiari, E. Kostem, E. Y. Kang, B. Pasaniuc, E. Eskin, Identifying causal variants at loci with multiple signals of association. *Genetics* **198**, 497–508 (2014). doi: 10.1534/genetics.114.167908; pmid: 25104515
12. B. Vernot *et al*., Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science* **352**, 235–239 (2016). doi: 10.1126/science.aad9416; pmid: 26989198
13. S. Mallick *et al*., The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016). doi: 10.1038/nature18964; pmid: 27654912
14. R. L. Lamason *et al*., SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **310**, 1782–1786 (2005). doi: 10.1126/science.1116238; pmid: 16357253

15. S. Beleza *et al*., The timing of pigmentation lightening in Europeans. *Mol. Biol. Evol.* **30** 24–35 (2013). doi: 10.1093/molbev/mss207; pmid: 22923467
16. M. Jonnalagadda *et al*., Identifying signatures of positive selection in pigmentation genes in two South Asian populations. *Am. J. Hum. Biol.* **29**, e23012 (2017). doi: 10.1002/ajhb.23012; pmid: 28439965
17. C. Basu Mallick *et al*., The light skin allele of *SLC24A5* in South Asians and Europeans shares identity by descent. *PLOS Genet.* **9**, e1003912 (2013). doi: 10.1371/journal.pgen.1003912
18. I. Mathieson *et al*., Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015). doi: 10.1038/nature16152; pmid: 26595274
19. L. Pagani *et al*., Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *Am. J. Hum. Genet.* **91**, 83–96 (2012). doi: 10.1016/j.ajhg.2012.05.015; pmid: 22726845
20. F. Tekola-Ayele *et al*., Novel genomic signals of recent selection in an Ethiopian population. *Eur. J. Hum. Genet.* **23**, 1085–1092 (2015). doi: 10.1038/ejhg.2014.233; pmid: 25370040
21. C. M. Schlebusch *et al*., Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *Science* **338**, 374–379 (2012). doi: 10.1126/science.1227721; pmid: 22997136
22. J. K. Pickrell *et al*., The genetic prehistory of southern Africa. *Nat. Commun.* **3**, 1143 (2012). doi: 10.1038/ncomms2140; pmid: 23072811
23. L. Pagani *et al*., Tracing the route of modern humans out of Africa by using 225 human genome sequences from Ethiopians and Egyptians. *Am. J. Hum. Genet.* **96**, 986–991 (2015). doi: 10.1016/j.ajhg.2015.04.019; pmid: 26027499
24. C. Ehret, *An African Classical Age: Eastern and Southern Africa in World History, 1000 BC to AD 400* (Oxford, 1998).
25. M. G. Madej, S. Dang, N. Yan, H. R. Kaback, Evolutionary mix-and-match with MFS transporters. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5870–5874 (2013). doi: 10.1073/pnas.1303538110; pmid: 23530251
26. R. Yu *et al*., Transcriptome analysis reveals markers of aberrantly activated innate immunity in vitiligo lesional and non-lesional skin. *PLOS ONE* **7**, e51040 (2012). doi: 10.1371/journal.pone.0051040; pmid: 23251420
27. D. Richter *et al*., The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age. *Nature* **546**, 293–296 (2017). doi: 10.1038/nature22335; pmid: 28593967
28. M. Przeworski, The signature of positive selection at randomly chosen loci. *Genetics* **160**, 1179–1189 (2002). pmid: 11901132
29. V. Le Corre, A. Kremer, The genetic differentiation at quantitative trait loci under local adaptation. *Mol. Ecol.* **21**, 1548–1566 (2012). doi: 10.1111/j.1365-294X.2012.05479.x; pmid: 22332667
30. Roadmap Epigenomics Consortium, Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015). doi: 10.1038/nature14248; pmid: 25693563
31. D. Hnisz *et al*., Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013). doi: 10.1016/j.cell.2013.09.053; pmid: 24119843
32. E. V. Sviderskaya *et al*., $p16^{Ink4a}$ in melanocyte senescence and differentiation. *J. Natl. Cancer Inst.* **94**, 446–454 (2002). doi: 10.1093/jnci/94.6.446; pmid: 11904317
33. K. Howe *et al*., The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503 (2013). doi: 10.1038/nature12111; pmid: 23594743
34. R. N. Kelsh *et al*., Zebrafish pigmentation mutations and the processes of neural crest development. *Development* **123**, 369–389 (1996). pmid: 9007256
35. S. Le Guyader, S. Jesuthasan, Analysis of xanthophore and pterinosome biogenesis in zebrafish using methylene blue and pteridine autofluorescence. *Pigment Cell Res.* **15**, 27–31 (2002). doi: 10.1034/j.1600-0749.2002.00045.x; pmid: 11837453
36. J. L. Bloom, D. S. Falconer, 'Grizzled', a mutant in linkage group X of the mouse. *Genet. Res.* **7**, 159–167 (1966). doi: 10.1017/S0016672300009587
37. T. Kobayashi *et al*., Modulation of melanogenic protein expression during the switch from eu- to pheomelanogenesis. *J. Cell Sci.* **108**, 2301–2309 (1995). pmid: 7673350
38. G. Raposo, D. Tenza, D. M. Murphy, J. F. Berson, M. S. Marks, Distinct protein sorting and localization to premelanosomes, melanosomes, and lysosomes in pigmented melanocytic cells. *J. Cell Biol.* **152**, 809–824 (2001). doi: 10.1083/jcb.152.4.809; pmid: 11266471
39. G. Chu, E. Chang, Xeroderma pigmentosum group E cells lack a nuclear factor that binds to damaged DNA. *Science* **242**, 564–567 (1988). doi: 10.1126/science.3175673; pmid: 3175673
40. A. L. Kadekaro *et al*., *Melanocortin 1 receptor* genotype: An important determinant of the damage response of melanocytes to ultraviolet radiation. *FASEB J.* **24**, 3850–3860 (2010). doi: 10.1096/fj.10-158485; pmid: 20519635
41. Y. Zhang *et al*., *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* **20**, 1437–1455 (2008). doi: 10.1105/tpc.108.058891; pmid: 18552200
42. C. Yu *et al*., CRL4 complex regulates mammalian oocyte survival and reprogramming by activation of TET proteins. *Science* **342**, 1518–1521 (2013). doi: 10.1126/science.1244587; pmid: 24357321
43. M. Lieberman, O. Segev, N. Gilboa, A. Lalazar, I. Levin, 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.* **108**, 1574–1581 (2004). pmid: 14968305
44. K.-i. Takata, H. Yoshida, M. Yamaguchi, K. Sakaguchi, Drosophila damaged DNA-binding protein 1 is an essential factor for development. *Genetics* **168**, 855–865 (2004). doi: 10.1534/genetics.103.025965; pmid: 15514059
45. J. H. Lee *et al*., Evolutionarily assembled cis-regulatory module at a human ciliopathy locus. *Science* **335**, 966–969 (2012). doi: 10.1126/science.1213506; pmid: 22282472
46. G. Li *et al*., Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84–98 (2012). doi: 10.1016/j.cell.2011.12.014; pmid: 22265404
47. L. Teng, B. He, J. Wang, K. Tan, 4DGenome: A comprehensive database of chromatin interactions. *Bioinformatics* **31**, 2560–2564 (2015). doi: 10.1093/bioinformatics/btv158; pmid: 25788621
48. P. Laurette *et al*., Transcription factor MITF and remodeller BRG1 define chromatin organisation at regulatory elements in melanoma cells. *eLife* **4**, e06857 (2015). doi: 10.7554/eLife.06857; pmid: 25803486
49. M. H. Beltrame, M. A. Rubel, S. A. Tishkoff, Inferences of African evolutionary history from genomic data. *Curr. Opin. Genet. Dev.* **41**, 159–166 (2016). doi: 10.1016/j.gde.2016.10.002; pmid: 27810637
50. M. Visser, M. Kayser, R.-J. Palstra, *HERC2* rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the *OCA2* promoter. *Genome Res.* **22**, 446–455 (2012). doi: 10.1101/gr.128652.111; pmid: 22234890
51. M. Kayser *et al*., Three genome-wide association studies and a linkage analysis identify *HERC2* as a human iris color gene. *Am. J. Hum. Genet.* **82**, 411–423 (2008). doi: 10.1016/j.ajhg.2007.10.003; pmid: 18252221
52. J. Han *et al*., A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLOS Genet.* **4**, e1000074 (2008). doi: 10.1371/journal.pgen.1000074; pmid: 18483556
53. N. W. Bellono, I. E. Escobar, A. J. Lefkovith, M. S. Marks, E. Oancea, An intracellular anion channel critical for pigmentation. *eLife* **3**, e04543 (2014). doi: 10.7554/eLife.04543; pmid: 25513726
54. M. H. Brilliant, The mouse *p* (*pink-eyed dilution*) and human P genes, oculocutaneous albinism type 2 (OCA2), and melanosomal pH. *Pigment Cell Res.* **14**, 86–93 (2001). doi: 10.1034/j.1600-0749.2001.140203.x; pmid: 11310796
55. N. Eriksson *et al*., Web-based, participant-driven studies yield novel genetic associations for common traits. *PLOS Genet.* **6**, e1000993 (2010). doi: 10.1371/journal.pgen.1000993; pmid: 20585627
56. H. L. Norton *et al*., Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol. Biol. Evol.* **24**, 710–722 (2007). doi: 10.1093/molbev/msl203; pmid: 17182896
57. A. Rafati *et al*., Association of rs12913832 in the *HERC2* gene affecting human iris color variation. *ASJ* **12**, 9–16 (2015).
58. H. Eiberg *et al*., Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the *HERC2* gene inhibiting *OCA2* expression. *Hum. Genet.* **123**, 177–187 (2008). doi: 10.1007/s00439-007-0460-x; pmid: 18172690
59. H. E. Seberg *et al*., TFAP2 paralogs regulate melanocyte differentiation in parallel with MITF. *PLOS Genet.* **13**, e1006636 (2017). pmid: 10.1371/journal.pgen.1006636; pmid: 28249010
60. W. S. Oetting, S. S. Garrett, M. Brott, R. A. King, P gene mutations associated with oculocutaneous albinism type II (OCA2). *Hum. Mutat.* **25**, 323 (2005). doi: 10.1002/humu.9318; pmid: 15712365
61. R. Kerr *et al*., Identification of *P* gene mutations in individuals with oculocutaneous albinism in sub-Saharan Africa. *Hum. Mutat.* **15**, 166–172 (2000). doi: 10.1002/(SICI)1098-1004(200002)15:2<166::AID-HUMU5>3.0.CO;2-Z; pmid: 10649493
62. A. R. Wood *et al*., Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014). doi: 10.1038/ng.3097; pmid: 25282103
63. R. M. Harding *et al*., Evidence for variable selective pressures at MC1R. *Am. J. Hum. Genet.* **66**, 1351–1361 (2000). doi: 10.1086/302863; pmid: 10733465
64. D. Reich *et al*., Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010). doi: 10.1038/nature09710; pmid: 21179161
65. N. G. Jablonski, G. Chaplin, The colours of humanity: The evolution of pigmentation in the human lineage. *Philos. Trans. R. Soc. London B Biol. Sci.* **372**, 20160349 (2017). doi: 10.1098/rstb.2016.0349; pmid: 28533464
66. A. R. Rogers, D. Iltis, S. Wooding, Genetic variation at the MCIR locus and the time since loss of human body hair. *Curr. Anthropol.* **45**, 105–108 (2004). doi: 10.1086/381006
67. G. H. Perry, N. J. Dominy, Evolution of the human pygmy phenotype. *Trends Ecol. Evol.* **24**, 218–225 (2009). doi: 10.1016/j.tree.2008.11.008; pmid: 19246118
68. L. Pagani *et al*., Genomic analyses inform on migration events during the peopling of Eurasia. *Nature* **538**, 238–242 (2016). doi: 10.1038/nature19792; pmid: 27654910
69. A.-S. Malaspinas *et al*., A genomic history of Aboriginal Australia. *Nature* **538**, 207–214 (2016). doi: 10.1038/nature18299; pmid: 27654914
70. Materials and methods are available as supplementary materials.
71. J. K. Wagner, C. Jovel, H. L. Norton, E. J. Parra, M. D. Shriver, Comparing quantitative measures of erythema, pigmentation and skin response using reflectometry. *Pigment Cell Res.* **15**, 379–384 (2002). doi: 10.1034/j.1600-0749.2002.02042.x; pmid: 12213095
72. S. R. Browning, B. L. Browning, Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007). doi: 10.1086/521987; pmid: 17924348
73. A. McKenna *et al*., The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010). doi: 10.1101/gr.107524.110; pmid: 20644199
74. H. Li, Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30**, 2843–2851 (2014). doi: 10.1093/bioinformatics/btu356; pmid: 24974202
75. T. Rausch *et al*., DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012). doi: 10.1093/bioinformatics/bts378; pmid: 22962449
76. J. O'Connell *et al*., A general approach for haplotype phasing across the full spectrum of relatedness. *PLOS Genet.* **10**, e1004234 (2014). doi: 10.1371/journal.pgen.1004234; pmid: 24743097
77. O. Delaneau, J. Marchini, J.-F. Zagury, A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2011). doi: 10.1038/nmeth.1785; pmid: 22138821
78. S. Das *et al*., Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016). doi: 10.1038/ng.3656; pmid: 27571263
79. C. Y. McLean *et al*., GREAT improves functional interpretation of *cis*-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010). doi: 10.1038/nbt.1630; pmid: 20436461
80. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011). doi: 10.1016/j.ajhg.2010.11.011; pmid: 21167468

81. J. Yang *et al*., Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010). doi: 10.1038/ng.608; pmid: 20562875
82. B. S. Weir, C. C. Cockerham, Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984). doi: 10.2307/2408641; pmid: 28563791
83. B. F. Voight, S. Kudaravalli, X. Wen, J. K. Pritchard, A map of recent positive selection in the human genome. *PLOS Biol.* **4**, e72 (2006). doi: 10.1371/journal.pbio.0040072; pmid: 16494531
84. F. Tajima, Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989). pmid: 2513255
85. P. Danecek *et al*., The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011). doi: 10.1093/bioinformatics/btr330; pmid: 21653522
86. Z. A. Szpiech, R. D. Hernandez, selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* **31**, 2824–2827 (2014). doi: 10.1093/molbev/msu211; pmid: 25015648
87. H. J. Bandelt, P. Forster, A. Röhl, Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999). doi: 10.1093/oxfordjournals.molbev.a026036; pmid: 10331250
88. Fluxus Engineering, www.fluxus-engineering.com.
89. R. B. Lyngsø, Y. S. Song, J. Hein, Minimum recombination histories by branch and bound, in *Algorithms in Bioinformatics*, R. Casadio, G. Myers, Eds. (Lecture Notes in Computer Science Series, Springer, 2005), vol. 3692, pp. 239–250.
90. J. Terhorst, J. A. Kamm, Y. S. Song, Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017). doi: 10.1038/ng.3748; pmid: 28024154
91. J. Ernst, M. Kellis, ChromHMM: Automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012). doi: 10.1038/nmeth.1906; pmid: 22373907
92. L. H. Chadwick, The NIH Roadmap Epigenomics Program data resource. *Epigenomics* **4**, 317–324 (2012). doi: 10.2217/epi.12.18; pmid: 22690667
93. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010). doi: 10.1093/bioinformatics/btq033; pmid: 20110278
94. J. Zhou, O. G. Troyanskaya, Predicting effects of noncoding variants with deep learning–based sequence model. *Nat. Methods* **12**, 931–934 (2015). doi: 10.1038/nmeth.3547; pmid: 26301843
95. D. Lee *et al*., A method to predict the impact of regulatory variants from DNA sequence. *Nat. Genet.* **47**, 955–961 (2015). doi: 10.1038/ng.3331; pmid: 26075791
96. T. L. Bailey *et al*., MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009). doi: 10.1093/nar/gkp335; pmid: 19458158
97. A. Mathelier *et al*., JASPAR 2016: A major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **44**, D110–D115 (2016). doi: 10.1093/nar/gkv1176; pmid: 26531826
98. A. Dobin *et al*., STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). doi: 10.1093/bioinformatics/bts635; pmid: 23104886
99. B. Li, C. N. Dewey, RSEM: Accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011). doi: 10.1186/1471-2105-12-323; pmid: 21816040
100. O. Stegle, L. Parts, R. Durbin, J. Winn, A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLOS Comput. Biol.* **6**, e1000770 (2010). doi: 10.1371/journal.pcbi.1000770; pmid: 20463871
101. C. E. Bonferroni, *Teoria statistica delle classi e calcolo delle probabilità* (Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 1936), vol. 8.
102. C. Delevoye *et al*., AP-1 and KIF13A coordinate endosomal sorting and positioning during melanosome biogenesis. *J. Cell Biol.* **187**, 247–264 (2009). doi: 10.1083/jcb.200907122; pmid: 19841138
103. P. A. Calvo, D. W. Frank, B. M. Bieler, J. F. Berson, M. S. Marks, 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* **274**, 12780–12789 (1999). doi: 10.1074/jbc.274.18.12780; pmid: 19505943
104. P. Cingolani *et al*., A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* **6**, 80–92 (2012). doi: 10.4161/fly.19695; pmid: 25722852
105. A. Ruiz-Linares *et al*., Admixture in Latin America: Geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLOS Genet.* **10**, e1004572 (2014). doi: 10.1371/journal.pgen.1004572; pmid: 21963610

ACKNOWLEDGMENTS

We thank J. Akey and R. McCoy for Melanesian genotype data; A. Clark, C. Brown, and Y. S. Park for critical review of the manuscript; members of the Tishkoff laboratory for helpful discussion; A. Weeraratna at Wistar Institute, Philadelphia, for providing the WM88 melanocytic cell line; D. Parichy for the aox5:palmGFP plasmid; G. Xu and R. Yang at University of Pennsylvania for technical assistance; M. Burmeister at University of Michigan for *grizzled* mouse samples; L. Garrett at the Embryonic Stem Cell and Transgenic Mouse Core [National Human Genome Research Institute (NHGRI)]; R. Sood in the Zebrafish Core (NHGRI); and the African participants. We acknowledge the contribution of the staff members of the Cancer Genomics Research Laboratory [National Cancer Institute (NCI)], the NIH Intramural Sequencing Center, the NCI Center for Cancer Research Sequencing Facility, the Yale University Skin SPORE Specimen Resource Core, and the Botswana–University of Pennsylvania Partnership. This work used computational resources of the NIH High-Performance Computing (HPC) Biowulf cluster. This research was funded by the following grants: NIH grants 1R01DK104339-0 and 1R01GM113657-01 and NSF grant BCS-1317217 to S.A.T., NIH grant R01 AR048155 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) to M.S.M, NIH grant R01 AR066318 from NIAMS to E.O., NIH grants 5R24OD017870-04 and 1U54DK110805-01 to L.Z. and Y.Z., NIH grant R01-GM094402 to Y.S.S., and NIH grant K12 GM081259 from NIGMS to S.B. M.H.B. was partly supported by a "Science Without Borders" fellowship from CNPq, Brazil. Y.S.S. is a Chan Zuckerberg Biohub investigator. This work was supported in part by the Center of Excellence in Environmental Toxicology (NIH P30-ES013508, T32-ES019851 to M.E.B.H.) (National Institute of Environmental Health Sciences), the Intramural Program of the NHGRI, and the Division of Cancer Epidemiology and Genetics, NCI, NIH, federal funds from NCI under contract HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products, or organizations imply endorsement by the U.S. government. L.Z. is a founder and stockholder of Fate Therapeutics, Marauder Therapeutics, and Scholar Rock. Data are available at dbGAP accession number phs001396.v1.p1 and SRA BioProject PRJNA392485. In memory of G. Lema and E. Kimaro, who made important contributions to this project.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/358/6365/eaan8433/suppl/DC1
Materials and Methods
Figs. S1 to S21
Tables S1 to S8
NISC Comparative Sequencing Program Collaborator List
References (*106*–*133*)

27 May 2017; accepted 3 October 2017
Published online 12 October 2017
10.1126/science.aan8433