A genome-wide scan on individual typology angle found variants at 24A2</i>associated with skin color variation in Chinese populations

Journal Item

Journal Pre-proof

A genome-wide scan on individual typology angle found variants at *SLC24A2* associated with skin color variation in Chinese populations

Fudi Wang, Qi Luo, Yan Chen, Yu Liu, Ke Xu, Kaustubh Adhikari, Xiyang Cai, Jialin Liu, Yi Li, Xuyang Liu, Luis-Miguel Ramirez-Aristeguieta, Ziyu Yuan, Yong Zhou, Fu-Feng Li, Binghua Jiang, Li Jin, Andres Ruiz-Linares, Zhaohui Yang, Fan Liu, Sijia Wang

PII: S0022-202X(21)02226-0

DOI: https://doi.org/10.1016/j.jid.2021.07.186

Reference: JID 3130

To appear in: *The Journal of Investigative Dermatology*

Received Date: 4 December 2020

Revised Date: 1 July 2021

Accepted Date: 9 July 2021

Please cite this article as: Wang F, Luo Q, Chen Y, Liu Y, Xu K, Adhikari K, Cai X, Liu J, Li Y, Liu X, Ramirez-Aristeguieta L-M, Yuan Z, Zhou Y, Li F-F, Jiang B, Jin L, Ruiz-Linares A, Yang Z, Liu F, Wang S, A genome-wide scan on individual typology angle found variants at *SLC24A2* associated with skin color variation in Chinese populations, *The Journal of Investigative Dermatology* (2021), doi: https://doi.org/10.1016/j.jid.2021.07.186.

A genome-wide scan on individual typology angle found variants at *SLC24A2* associated with skin color variation in Chinese populations

Fudi Wang[1,2,3,#], Qi Luo[2,#], Yan Chen[4,5,6], Yu Liu[2], Ke Xu[7], Kaustubh Adhikari[8,9], Xiyang Cai[2], Jialin Liu[4,5,6], Yi Li[4,5,6], Xuyang Liu[10], Luis-Miguel Ramirez-Aristeguieta[11], Ziyu Yuan[12,13], Yong Zhou[14], Fu-Feng Li[15], Binghua Jiang[7], Li Jin[2,12,13], Andres Ruiz-Linares[16,17,9], Zhaohui Yang[7,#], Fan Liu[4,5,6,#], Sijia Wang[2,18,#,$]

[1] Department of Skin & Cosmetic Research, Shanghai Skin Disease Hospital, China

[2] CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences

[3] Department of Science, Inertia Shanghai Biotechnology Co., Ltd, Shanghai, PR China

[4] Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

[5] University of Chinese Academy of Sciences, Beijing, China

[6] China National Center for Bioinformation, Beijing, China

[7] Academy of Medicine Science, Zhengzhou University, Zhengzhou, China

[8] School of Mathematics and Statistics, Faculty of Science, Technology, Engineering and Mathematics, The Open University, Milton Keynes, MK7 6AA, United Kingdom.

[9] Department of Genetics, Evolution and Environment, and UCL Genetics Institute, University College London, London, WC1E 6BT, UK

[10] Yunnan Key Laboratory of Primate Biomedical Research, Institute of Primate Translational Medicine, Kunming University of Science and Technology, Kunming, China

[11] QST Lab. Faculty of Odontology, University of Antioquia, Medellin, Colombia

[12] Fudan University Taizhou Institute of Health Sciences, Taizhou, China.

[13] State Key Laboratory of Genetic Engineering, Human Phenome Institute, and School of Life Sciences, Fudan University, Shanghai, China

[14] Clinical Research Institute, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[15] Basic Medical Department, Shanghai University of Traditional Chinese Medicine, Shanghai, China

[16] Ministry of Education Key Laboratory of Contemporary Anthropology and Collaborative Innovation Center of Genetics and Development, School of Life Sciences and Human Phenome Institute, Fudan University, Yangpu District, Shanghai, China.

[17] UMR 7268 ADES, CNRS, Aix-Marseille Université, EFS, Faculté de Médecine Timone, Marseille, 13005, France.

[18] Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China

[#] Fudi Wang and Qi Luo contributed equally to this work; Zhaohui Yang, Fan Liu and Sijia Wang co-supervised the study.

[$] Correspondence: Sijia Wang (wangsijia@picb.ac.cn)

Running title: Skin color GWAS in Chinese populations

Key words: Skin color; GWAS; *SLC24A2*; East Asian populations; individual typology angle

ORCIDs

Fudi Wang: 0000-0002-0208-3343

Qi Luo: 0000-0002-6988-2051

Yan Chen: 0000-0002-6873-8648

Yu Liu: 0000-0001-7110-127X

Ke Xu: 0000-0002-0408-1659

Kaustubh Adhikari: 0000-0001-5825-4191

Xiyang Cai: 0000-0002-3868-5426

Jialin Liu: 0000-0003-4065-5397

Yi Li: 0000-0002-6988-2051

Xuyang Liu: 0000-0002-0871-8979

Luis-Miguel Ramirez-Aristeguieta: 0000-0002-0961-5824

Ziyu Yuan: 0000-0003-1434-0383

Yong Zhou: 0000-0001-5221-8026

Fu-feng Li: 0000-0002-0566-3589

Binghua Jiang: 0000-0003-4526-2031

Li Jin: 0000-0001-9201-2321

Andrés Ruiz-Linares: 0000-0001-8372-1011

Zhaohui Yang: 0000-0003-0958-4439

Fan Liu: 0000-0001-9241-8161

Sijia Wang: 0000-0001-6961-7867

TO THE EDITOR,

Skin pigmentation functions as a shield to prevent UV damage on the DNA of epidermal cells. A good number of genome-wide association studies (GWASs) on skin color variation or skin photosensitivity have been conducted that discovered a large number of associated loci (Ganguly et al., 2019). Polymorphisms/mutations at these loci have been used to molecular type skin color for individuals from different continental groups (Chen et al., 2020), and have been associated with various forms of Albinism (Marcon and Maia, 2019), loss of photoprotection, and increased rates of photoaging (Liu et al., 2016). However, there has been surprisingly few GWASs conducted in East Asian populations, likely due to the presumption of a low skin color variation in East Asians (Rawofi et al., 2017). The question regarding the presence of East Asian-specific skin color alleles remains to be answered. Individual typology angle (ITA˚) is a quantitative skin color measurement based on the colorimetric parameters of the L*a*b* system (Chardon et al., 1991). The validity of ITA˚ in serving as an objective skin color measurement has been extensively evaluated with regard to its correlations with constitutive pigmentation, geographical distribution of skin pigmentation, and biological markers of UV-induced erythema (Del Bino and Bernerd, 2013). Despite the wide use of ITA˚ in clinical dermatology, few genetic studies had adopted it as the measurement of skin color.

Here, we report a GWAS of ITA˚ in Chinese, followed by replications in Chinese and Latin Americans. The discovery sample included a total of 6,964 individuals of Chinese origin from two cohorts: the Jidong cohort (JD, *N*=5,034) and the National Survey of Physical Traits cohort (NSPT, *N*=1,930). Studies in these cohorts were approved by the Ethics

Committee of Kailuan General Hospital of Tangshan City and the Medical Ethics Committee, Staff Hospital, Jidong Oilfield Branch, China National Petroleum Corporation in July, 2013 (approval No. 2013 YILUNZI1), as well as the Ethics Committees of Fudan University (14117) and the Shanghai Institutes for Biological Sciences (ER-SIBS-261410). The replication sample included a total of 2,053 individuals from two cohorts: Taizhou Longitudinal cohort of Chinese origin (TZL, *N*=1,787) and the Colombian cohort of Latin American origin (*N*=266). The TZL study was conducted with the approval of the Ethics Committee of Fudan University (Ethics Research Approval 85), Shanghai, China and the Colombian cohort study with the approval of the bioethics committee of the Odontology Faculty at the University of Antioquia (CONCEPTO 01-2013). All participants provided written informed consent. For the discovery cohorts, sample characteristics and phenotyping details are provided in **Supplementary Materials (Table S1, Figure 1a-b)**. In brief, we derived ITA° from high resolution portrait photos, which were processed using an in-house pipeline, involving a face detector, automated facial landmarking, cheek segmentation, and color analysis. In a subgroup of 50 randomly selected samples, the ITA° derived from portrait photos was highly correlated with the ITA° measured by a spectrophotometer (*r*=0.94, $P = 4.42 \times 10^{-24}$, **Figure S1**). The Chardon skin color type (Chardon et al., 1991) as defined by ITA° cutoffs was fairly concordant with human perception (*Kappa*=0.47). Although dark skins were not seen and very light (1.51%) and brown (0.27%) skins were rarely seen in our Chinese samples, ITA° did show a substantial variance (Var=8.03, mean=35.95, min=-26.36, max=68.42) across light (23.41%), intermediate (59.69%), and tan (15.12%) categories.

Females had a significant lighter skin than males (β=6.18, $P = 9.22 \times 10^{-270}$) and aging increased coloration (β=-0.22, $P = 8.08 \times 10^{-244}$), which were consistent with previous observations (Tan et al., 2020). Z transformed ITA° (z-ITA°) was used in the subsequent genetic association analysis (Figure S2).

In the discovery stage of the study, we conducted a genome-wide inverse variance, fixed-effect meta-analysis of two GWASs, which were independently conducted in JD and NSPT, totaling 6,964 Chinese individuals. No evidence of population sub-stratification or genome inflation was detected (λ=0.99, **Figure 1c**). The meta-analysis identified a total of 19 SNPs at 3 genomic loci showing genome-wide significant association with z-ITA° (**Table S2**), including one previously unreported locus on 9p21.3 (*SLC24A2*), and two previously known loci on 15q12.6 (*OCA2*) and 15p21.1 (*SLC24A5*).

The locus at 9p21.3, as the most significant signal over the genome, contained a total of 14 significant SNPs, where the lead SNP (rs10122939) was an intron variant of *SLC24A2* (**Figure 2a**). The derived G allele showed a skin darkening effect reaching genome-wide significance in JD (β=-0.10, $P = 1.29 \times 10^{-9}$), nominal significance in NSPT (β=-0.07, $P = 4.71 \times 10^{-3}$), and further enhanced significance in the meta-analysis (β=-0.09, $P = 3.61 \times 10^{-11}$). This association was also genome-wide significant for the original ITA° without z-transformation (β=-1.89, $P = 7.28 \times 10^{-10}$). No genome-wide significant association signal was observed at 9p21.3 after conditioning on the genotype of the lead SNP rs10122939 (**Figure S3**). The G allele of rs10122939 was highly prevalent in East Asians (our sample: f=0.34, in CHB of the 1000 Genomes Project: f=0.29), but was rare in Europeans (in

EUR of 1000 Genomes Project f=0.004, **Figure 2b**), which may explain the failure of previous European GWASs in detecting its effect (other SNPs in LD showed the same pattern, **Table S2**). A series of population genetic analyses did not reveal significant evidence for positive selection surrounding rs10122939 at 9p21.3 (**Figure S4**).

The other two loci have been previously associated with skin color. On chromosome 15q12.6, two well-known East Asian-specific missense variants rs1800414 (His615Arg) and rs74653330 (Ala481Thr) of *OCA2* (Edwards et al., 2010, Yang et al., 2016) showed genome-wide significant association with ITA˚, where the derived C allele of rs1800414 and T allele of rs74653330 had a skin lightening effect ($P = 1.11 \times 10^{-10}$ and 3.05×10^{-8}, respectively; **Table S2** and **Figure S5a-b**). Both alleles were nearly absent outside of East Asia (**Figure S5c-d**). On chromosome 15p21.1, we confirmed the well-known effect of a missense variant rs1426654 of *SLC24A5*, of which the derived A allele had low frequency but a large skin lightening effect ($P = 4.64 \times 10^{-10}$, **Table S2 and Figure S6**).

We then looked up a total of 9,183 SNPs from four recently published large GWASs of skin/hair/eye pigmentation traits (Visconti et al., 2018), (Hysi et al., 2018), Adhikari et al., 2019, (Simcoe et al., 2021). A total of 151 SNPs at 13 loci showed nominally significant (p<0.01) association, but except those in or close to *OCA2* and *SLC24A5* genes, none survived strict Bonferroni correction of multiple testing (p<0.05/9,183; **Table S4**).

To replicate the finding at 9p21.3, we tested the association between rs10122939 and ITA˚ in an additional Chinese cohort from Taizhou (TZL, N=1,787) and in a Latin American cohort from Colombia (N=266). The effect of rs10122939 was successfully replicated at

nominal significance and the G allele also showed a skin darkening effect in both replication cohorts (TZL: β=-0.09, $P = 5.96 \times 10^{-3}$, and Colombian: β=-0.26, $P = 3.89 \times 10^{-2}$). A meta-analysis of all samples further enhanced the significance level of this association ($P = 2.13 \times 10^{-13}$).

To further investigate the potential functions of rs10122939, we conducted a function annotation analysis using the quantitative scoring system imbedding the 3DSNP database (Lu et al., 2017), which revealed significant evidence for rs10122939 serving as an enhancer of *SLC24A2* (Figure 2c). We further performed the luciferase report assays to experimentally validate this finding. The transcriptional activity of the enhancer containing the rs10122939 ancestral A allele was significantly higher than that of the corresponding derived G allele in the A375 cell line (t-test *P*=0.012, **Figure 2d**). This pattern would predict less *SLC24A2* production for the derived G allele carriers. Future studies may consider in vivo gene editing experiments to further investigate the pigmentation effect of *SLC24A2*. Unlike the well-known skin color gene *SLC24A5* and *SLC45A2*, this solute carrier family (SLC) gene *SLC24A2* is not expressed in skin, but rather in brain and adrenal, mainly involved in neuronal activity (Haque and Moon, 2018, Zhou et al., 2020). It has been recently reported that neuronal activity induced by acute stress can drive a rapid and permanent loss of melanocyte stem cells, which leads up to the stress-induced hair greying (Zhang et al., 2020). A possible hypothesis could be that *SLC24A2* moderates neuronal activity which influences melanocyte stem cells, eventually causing changes in skin color. Alternatively, *SLC24A2* may change the innervation of the skin and affect ITA˚ via altered skin thickness or other relevant

characteristics. This hypothesis is particularly attractive, as we did observe a positive correlation between ITA˚ and trans-epidermal water loss (P=4.12×10^{-9}), a phenotype thought to be strongly associated with skin thickness (Bargo et al., 2013). The rs10122939 SNP is also associated with TEWL (P=0.01, Figure S8, Supplementary materials).

In conclusion, this is a meaningful skin color GWAS in well-sized Chinese populations. We identified an intron variant of *SLC24A2* (rs10122939) as a previously unreported East Asian-European differentiating polymorphism involved in skin color variation. The underlying cellular mechanism is to be explored.

Conflict of interests

The authors declare that they have no competing interests.

Acknowledgements

This work has received funding from Supported by the "Strategic Priority Research Program" of the Chinese Academy of Sciences (XDB38020400 and XDB38010400), the National Key Research and Development Project (2018YFC0910403), Shanghai Municipal Science and Technology Major Project (2017SHZDZX01), the National Basic Research Program (2015FY111700), the "CAS Interdisciplinary Innovation Team" project, Beijing Advanced Discipline Fund, the National Natural Science Foundation of China (81930056, 91651507, 32070579, 31771393 and 31601016), the Open Project of Key Laboratory of Genomic and Precision Medicine of the CAS, and State Key Laboratory of Genetic Resources and Evolution grant (GREKF20-13), the Leverhulme Trust (F/07 134/DF), BBSRC

(BB/I021213/1), the Excellence Initiative of Aix-Marseille University - A*MIDEX (a French "Investissements d'Avenir" programme, 2RUIZLRE/RHRE/ID18HRU201 and 20-07874), the Scientific and Technology Committee of Shanghai Municipality (18490750300), Ministry of Science and Technology of China (2020YFE0201600) and the 111 Project (B13016), Santander Research and Scholarship Award, Bogue Fellowship from University College London. Correspondences regarding the luciferase assays and population comparative analyses should be addressed to Zhaohui Yang (yangzhaohui415@163.com) and Fan Liu (liufan@big.ac.cn), respectively.

Data availability

Summary statistics for all analyzed variants for the Jidong Study and the National Survey of Physical Traits can be viewed at NODE under accession number OEP001341, or directly at http://www.biosino.org/node/project/detail/OEP001341, and accessed by submitting a request for data-access. Data usage shall be in full compliance with the Regulations on Management of Human Genetic Resources in China. Individual genotype and phenotype data cannot be shared due to IRB restrictions on privacy concerns. All other relevant data supporting the key findings of this study are available within the letter and supplementary Information files, or from the corresponding author upon reasonable request.

CRediT statement

Conceptualization: Fudi Wang, Sijia Wang; Data Curation: Fudi Wang; Formal Analysis: Fudi Wang, Qi Luo, Yan Chen, Yu Liu, Xiyang Cai, Jialin Liu, Yi Li, Zhaohui Yang;

Funding Acquisition: Zhaohui Yang, Fan Liu, Sijia Wang; Investigation: Fudi Wang, Qi Luo, Ke Xu, Xuyang Liu, Zhaohui Yang; Methodology: Fudi Wang, Qi Luo; Project Administration: Fudi Wang, Fan Liu, Sijia Wang; Resources: Luis-Miguel Ramirez-Aristeguieta, Ziyu Yuan, Fu-feng Li, Yong Zhou, Binghua Jiang, Li Jin, Andres Ruiz-Linares, Sijia Wang; Software: Fudi Wang, Qi Luo; Supervision: Fan Liu, Sijia Wang; Validation: Yan Chen, Kaustubh Adhikari; Visualization: Fudi Wang, Fan Liu, Sijia Wang; Writing – Original Draft Preparation: Fudi Wang, Fan Liu; Writing – Review & Editing: Qi Luo, Yan Chen, Yu Liu, Kaustubh Adhikari, Xiyang Cai, Jialin Liu, Yi Li, Zhaohui Yang, Sijia Wang.

Reference

Bargo PR, Walston ST, Chu M, Seo I, Kollias N. Non-invasive assessment of tryptophan fluorescence and confocal microscopy provide information on skin barrier repair dynamics beyond TEWL. Exp Dermatol 2013;22(1):18-23.

Chardon A, Cretois I, Hourseau C. Skin colour typology and suntanning pathways. Int J Cosmet Sci 1991;13(4):191-208.

Chen Y, Branicki W, Walsh S, Nothnagel M, Kayser M, Liu F, et al. The impact of correlations between pigmentation phenotypes and underlying genotypes on genetic prediction of pigmentation traits. Forensic Sci Int Genet 2020;50:102395.

Del Bino S, Bernerd F. Variations in skin colour and the biological consequences of ultraviolet radiation exposure. Br J Dermatol 2013;169 Suppl 3:33-40.

Edwards M, Bigham A, Tan J, Li S, Gozdzik A, Ross K, et al. Association of the OCA2 polymorphism His615Arg with melanin content in east Asian populations: further evidence of convergent evolution of skin pigmentation. PLoS Genet 2010;6(3):e1000867.

Ganguly K, Saha T, Saha A, Dutta T, Banerjee S, Sengupta D, et al. Meta-analysis and prioritization of human skin pigmentation-associated GWAS-SNPs using ENCODE data-based web-tools. Arch Dermatol Res 2019;311(3):163-71.

Haque MN, Moon IS. Stigmasterol upregulates immediate early genes and promotes neuronal cytoarchitecture in primary hippocampal neurons as revealed by transcriptome analysis. Phytomedicine 2018;46:164-75.

Hysi PG, Valdes AM, Liu F, Furlotte NA, Evans DM, Bataille V, et al. Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability. Nat Genet 2018;50(5):652-6.

Liu F, Hamer MA, Deelen J, Lall JS, Jacobs L, van Heemst D, et al. The MC1R Gene and Youthful Looks. Curr Biol 2016;26(9):1213-20.

Lu Y, Quan C, Chen H, Bo X, Zhang C. 3DSNP: a database for linking human noncoding SNPs to their three-dimensional interacting genes. Nucleic Acids Res 2017;45(D1):D643-D9.

Marcon CR, Maia M. Albinism: epidemiology, genetics, cutaneous characterization, psychosocial factors. An Bras Dermatol 2019;94(5):503-20.

Rawofi L, Edwards M, Krithika S, Le P, Cha D, Yang Z, et al. Genome-wide association study of pigmentary traits (skin and iris color) in individuals of East Asian ancestry. PeerJ 2017;5:e3951.

Simcoe M, Valdes A, Liu F, Furlotte NA, Evans DM, Hemani G, et al. Genome-wide association study in almost 195,000 individuals identifies 50 previously unidentified genetic loci for eye color. Sci Adv 2021;7(11).

Tan Y, Wang F, Fan G, Zheng Y, Li B, Li N, et al. Identification of factors associated with minimal erythema dose variations in a large-scale population study of 22 146 subjects. J Eur Acad Dermatol Venereol 2020;34(7):1595-600.

Visconti A, Duffy DL, Liu F, Zhu G, Wu W, Chen Y, et al. Genome-wide association study in 176,678 Europeans reveals genetic loci for tanning response to sun exposure. Nat Commun 2018;9(1):1684.

Yang Z, Zhong H, Chen J, Zhang X, Zhang H, Luo X, et al. A Genetic Mechanism for Convergent Skin Lightening during Recent Human Evolution. Mol Biol Evol 2016;33(5):1177-87.

Zhang B, Ma S, Rachmin I, He M, Baral P, Choi S, et al. Hyperactivation of sympathetic nerves drives depletion of melanocyte stem cells. Nature 2020;577(7792):676-81.

Zhou XG, He H, Wang PJ. A critical role for miR135a5pmediated regulation of SLC24A2 in neuropathic pain. Mol Med Rep 2020;22(3):2115-22.

Figure legends

Figure 1. GWASs for skin color in Chinese populations.

a) The face detection module returned the locations of 68 facial landmarks (white points), and a fixed area on cheek (white rectangle) was detected based on 7 facial landmarks. The participant has consented to publication of his image. b) Individual typology angle (ITA˚) was derived from the fixed area. Samples were classified into five skin color types based on ITA˚ according to Chardon's definition. c) Manhattan plot of the meta-analysis results for facial skin color from the GWASs (JD and NSPT, total N=6,964). The $-\log_{10} P$ values for association were plotted for all SNPs according to their physical positions (genome-build GRCh38.p13). The red line was corresponding to the threshold for genome-wide significance ($P = 5 \times 10^{-8}$). Genome-wide significant and nominal significant ($P = 1 \times 10^{-6}$) SNPs were shown in Tables S2 and S3, respectively.

Figure 2. A locus at 9p21.3 was associated with skin color in Chinese populations.

a) Regional association plot for the significantly associated region on chromosome 9p21.3 consisting of *SLC24A2*. Increasing color intensities represented higher levels of linkage disequilibrium (r^2) with rs10122939. b) Geographical distribution of the minor allele frequencies for rs10122939 in 2,504 multi-ethnic individuals from the 1000 Genome Project. The effect allele (G) showed a skin darkening effect. c) The top SNP rs10122939 was annotated as enhancer using 3DSNP. d) Luciferase reporter assays confirmed that the transcriptional activity of the enhancer containing the rs10122939 ancestral A allele was

significantly higher than that of the corresponding derived G allele in the A375 cell line. A similar trend was observed but did not reach the significant level in HEK293 (Figure S7).

Table S1. Sample characteristics

Characteristic[1]	JD	NSPT	*P*[2]
Subjects, n	5,034	1,930	
Male sex, n (%)	2,501 (49.68)	705 (36.53)	7.09×10^{-23}
Age in years, median (IQR)	44.00(34.00-56.00)	51.00(40.00-58.00)	3.09×10^{-16}
ITA˚, median (IQR)	36.59(32.42-40.89)	41.69(36.06-47.80)	1.31×10^{-151}
Skin color type, n (%)			
Very light (ITA˚ > 55˚)	3 (0.06)	138 (7.15)	
Light (41˚ < ITA˚ < 55˚)	1,214 (24.12)	897 (46.48)	
Intermediate (28˚ < ITA˚ < 41˚)	3,439 (68.32)	804 (41.66)	2.63×10^{-233}
Tan (10˚ < ITA˚ < 28˚)	378 (7.51)	73 (3.78)	
Brown (-30˚ < ITA˚ < 28˚)	0 (0)	18 (0.93)	

Abbreviations: IQR, interquartile range; ITA, individual topology angle;
[1]Baseline characteristics were compared between discovery dataset and JD dataset;
[2]*P*-value of t tests for continuous variables and χ^2 tests for categorical variables.

Table S2. SNPs showing genome-wide significant association (p<5e-8) with skin color ITA° in 6,964 Chinese individuals

							JD (N = 5,034)				NSPT (N = 1,930)				META (N = 6,964)		EAF 1KGP		
SNP	**Gene**	**CHR**	**MB**	**EA**	**OA**	**AA**	**EAF**	**Beta**	**SE**	**P**	**EAF**	**Beta**	**SE**	**P**	**Beta**	**P**	**EAS**	**EUR**	**AFR**
rs10122939	SLC24A2	9	20.30	G	A	A	0.34	-0.10	0.02	1.29E-09	0.33	-0.07	0.02	4.71E-03	-0.09	3.61E-11	0.30	0.00	0.27
rs10116858	SLC24A2	9	20.30	T	G	T	0.34	-0.10	0.02	6.35E-09	0.32	-0.07	0.03	3.61E-03	-0.09	1.12E-10	0.30	0.00	0.24
rs10117511	SLC24A2	9	20.30	A	G	G	0.33	-0.10	0.02	5.03E-09	0.32	-0.07	0.03	3.65E-03	-0.09	9.34E-11	0.29	0.00	0.13
rs7033540	SLC24A2	9	20.30	T	C	C	0.29	-0.10	0.02	4.19E-08	0.28	-0.07	0.03	0.01	-0.09	2.21E-09	0.26	0.00	0.21
rs7037385	SLC24A2	9	20.30	A	G	G	0.30	-0.09	0.02	9.59E-08	0.28	-0.07	0.03	5.23E-03	-0.09	2.06E-09	0.27	0.00	0.26
rs12551950	SLC24A2	9	20.31	G	A	A	0.31	-0.09	0.02	6.37E-08	0.30	-0.08	0.03	2.62E-03	-0.09	6.62E-10	0.29	0.01	0.18
rs7851959	SLC24A2	9	20.31	C	G	G	0.31	-0.10	0.02	4.44E-08	0.30	-0.08	0.03	3.10E-03	-0.09	5.68E-10	0.71	0.99	0.62
rs58951095	SLC24A2	9	20.31	A	G	G	0.30	-0.09	0.02	8.63E-08	0.29	-0.08	0.03	2.74E-03	-0.09	9.29E-10	0.27	0.00	0.19
rs57500604	SLC24A2	9	20.31	A	C	C	0.30	-0.10	0.02	5.77E-08	0.29	-0.08	0.03	3.78E-03	-0.09	8.98E-10	0.27	0.01	0.31
rs16938009	SLC24A2	9	20.31	G	A	A	0.31	-0.09	0.02	6.54E-08	0.30	-0.08	0.03	2.87E-03	-0.09	7.54E-10	0.29	0.01	0.32
rs12555331	SLC24A2	9	20.31	G	C	C	0.30	-0.09	0.02	1.56E-07	0.29	-0.07	0.03	5.04E-03	-0.09	3.17E-09	0.73	1.00	0.81
rs1456966	SLC24A2	9	20.31	G	A	A	0.33	-0.08	0.02	2.37E-06	0.31	-0.07	0.03	5.23E-03	-0.08	4.27E-08	0.31	0.01	0.49
rs2383137	SLC24A2	9	20.32	C	G	G	0.32	-0.08	0.02	1.27E-06	0.30	-0.07	0.03	6.71E-03	-0.08	3.04E-08	0.70	0.99	0.82
rs7023559	SLC24A2	9	20.32	C	A	A	0.33	-0.08	0.02	2.00E-06	0.32	-0.07	0.03	5.02E-03	-0.08	3.48E-08	0.31	0.07	0.31
rs1800414	OCA2	15	28.19	T	C	T	0.48	-0.08	0.02	1.96E-06	0.46	-0.10	0.02	9.55E-06	-0.09	1.11E-10	0.40	1.00	1.00
rs76930569	OCA2	15	28.20	C	T	C	0.48	-0.08	0.02	1.96E-06	0.43	-0.10	0.02	9.55E-06	-0.09	1.11E-10	0.40	1.00	1.00
rs76470826	OCA2	15	28.22	AT	A	AT	0.48	-0.07	0.02	5.19E-06	0.43	-0.10	0.02	1.32E-05	-0.08	4.22E-10	0.59	0.00	0.00
rs74653330	OCA2	15	28.23	T	C	C	0.04	0.18	0.04	7.93E-06	0.02	0.21	0.06	9.88E-04	0.19	3.05E-08	0.03	0.01	0.00
rs1426654	SLC24A5	15	48.42	A	G	G	0.03	0.27	0.03	2.55E-09	0.03	0.16	0.03	3.00E-02	0.24	4.64E-10	0.01	1.00	0.07
rs2413887	CTXN2	15	48.49	C	T	T	0.04	0.22	0.04	2.93E-07	0.02	0.15	0.07	0.03	0.20	3.36E-08	0.02	1.00	0.08

EA: effect allele; OA: other allele; AA: ancestral allele; EAF: effect allele frequency, 1KGP: 1000 Genome Project.

Table S3. SNPs showing suggestive association (5e-8<p<1e-6) with ITA° in the meta analysis of 6,964 Chinese individuals.

							JD (N = 5,034)				NSPT (N = 1,930)				META (N = 6,964)		EAF 1KGP		
SNP	**Gene**	**CHR**	**MB**	**EA**	**OA**	**AA**	**EAF**	**Beta**	**SE**	**P**	**EAF**	**Beta**	**SE**	**P**	**Beta**	**P**	**EAS**	**EUR**	**AFR**
rs201197089	GABBR1	6	29.57	C	CT	CT	0.14	-0.10	0.02	1.08E-05	0.16	-0.08	0.04	0.02	-0.10	5.81E-07	0.30	0.00	0.27
rs11304792	SLC24A2	9	20.31	C	CT	CT	0.40	-0.07	0.02	8.16E-06	0.38	-0.07	0.02	5.02E-03	-0.07	1.34E-07	0.37	0.16	0.53
rs6475429	SLC24A2	9	20.31	C	A	C	0.47	-0.06	0.02	2.17E-04	0.43	-0.08	0.02	6.77E-04	-0.07	6.40E-07	0.41	0.50	0.94
rs28641053	SLC24A2	9	20.31	A	C	C	0.40	-0.07	0.02	1.85E-05	0.37	-0.08	0.02	9.09E-04	-0.07	6.15E-08	0.35	0.01	0.47
rs7021984	SLC24A2	9	20.31	A	T	T	0.33	-0.08	0.02	2.40E-06	0.31	-0.06	0.03	0.02	-0.08	1.47E-07	0.69	0.98	0.76
rs7023747	SLC24A2	9	20.32	T	G	T	0.34	-0.08	0.02	3.73E-06	0.34	-0.06	0.03	0.02	-0.07	3.05E-07	0.35	0.08	0.38
rs3915945	RAB11FIP2	10	119.75	A	G	G	0.15	0.10	0.02	4.03E-06	0.12	0.08	0.03	0.02	0.10	2.70E-07	0.13	0.01	0.00
rs116921893	RAB11FIP2	10	119.76	C	T	T	0.15	0.10	0.02	4.03E-06	0.12	0.08	0.03	0.02	0.10	2.70E-07	0.13	0.01	0.00
rs74638723	RAB11FIP2	10	119.77	G	A	A	0.15	0.10	0.02	5.03E-06	0.12	0.08	0.03	0.02	0.10	3.04E-07	0.13	0.01	0.01
rs3740550	RAB11FIP2	10	119.77	A	G	G	0.15	0.10	0.02	5.03E-06	0.12	0.08	0.03	0.02	0.10	3.04E-07	0.13	0.01	0.01
rs60424173	RAB11FIP2	10	119.78	G	T	T	0.16	0.10	0.02	9.13E-06	0.12	0.07	0.03	0.03	0.09	8.20E-07	0.13	0.01	0.01
rs79852633	CASC2	10	119.84	A	G	G	0.15	0.10	0.02	8.87E-06	0.12	0.10	0.03	3.28E-03	0.10	9.64E-08	0.13	0.01	0.00
rs79250499	CASC2	10	119.86	A	C	C	0.15	0.10	0.02	1.03E-05	0.12	0.10	0.03	4.22E-03	0.10	1.42E-07	0.13	0.01	0.01
rs7077320	CASC2	10	119.87	C	T	T	0.15	0.10	0.02	1.02E-05	0.11	0.10	0.03	3.23E-03	0.10	1.09E-07	0.13	0.01	0.01
rs61507658	CASC2	10	119.88	G	A	A	0.14	0.10	0.02	1.44E-05	0.10	0.11	0.04	1.67E-03	0.11	8.48E-08	0.11	0.01	0.00
rs201658840	CASC2	10	119.90	T	TG	TG	0.15	0.09	0.02	4.51E-05	0.12	0.10	0.03	3.15E-03	0.10	4.71E-07	0.13	0.01	0.01
rs75295597	OCA2	15	28.23	A	T	A	0.49	-0.06	0.02	1.88E-04	0.43	-0.09	0.02	6.76E-05	-0.07	8.67E-08	0.39	1.00	1.00
rs12442916	OCA2	15	28.23	G	A	G	0.37	-0.06	0.02	2.81E-04	0.35	-0.09	0.02	1.82E-04	-0.07	3.00E-07	0.30	0.94	1.00
rs10775262	OCA2	15	28.23	T	C	C	0.36	-0.06	0.02	6.46E-04	0.34	-0.09	0.02	2.37E-04	-0.07	9.21E-07	0.29	0.87	0.80
rs10775263	OCA2	15	28.23	C	T	C	0.36	-0.06	0.02	6.46E-04	0.34	-0.09	0.02	2.37E-04	-0.07	9.21E-07	0.29	0.87	0.80
rs730502	OCA2	15	28.23	G	T	G	0.36	-0.06	0.02	5.01E-04	0.34	-0.09	0.02	1.70E-04	-0.07	5.39E-07	0.29	0.87	0.80
rs2267434	ACO2	22	41.88	T	C	C	0.41	0.07	0.02	1.79E-05	0.51	0.06	0.02	0.01	0.07	8.59E-07	0.51	0.01	0.26
rs5751114	ACO2	22	41.88	G	A	G	0.41	0.07	0.02	1.24E-05	0.51	0.06	0.02	0.02	0.07	7.27E-07	0.51	0.02	0.52
rs2267435	ACO2	22	41.88	G	A	A	0.41	0.07	0.02	1.85E-05	0.51	0.06	0.02	0.01	0.07	8.83E-07	0.51	0.01	0.21
rs2284080	ACO2	22	41.89	C	G	G	0.41	0.07	0.02	1.85E-05	0.51	0.06	0.02	0.01	0.07	8.83E-07	0.49	0.99	0.79
rs5751123	POLR3H	22	41.93	A	C	C	0.42	0.07	0.02	2.83E-05	0.52	0.06	0.02	9.28E-03	0.07	8.36E-07	0.52	0.01	0.22
rs5758384	POLR3H	22	41.93	T	G	G	0.42	0.07	0.02	2.83E-05	0.52	0.06	0.02	9.28E-03	0.07	8.36E-07	0.52	0.01	0.22
rs5758386	POLR3H	22	41.94	A	G	G	0.42	0.07	0.02	3.01E-05	0.52	0.06	0.02	9.82E-03	0.07	9.42E-07	0.52	0.01	0.57
rs5758388	POLR3H	22	41.94	T	C	C	0.42	0.07	0.02	2.83E-05	0.52	0.06	0.02	9.28E-03	0.07	8.36E-07	0.52	0.01	0.22
rs8138232	POLR3H	22	41.94	T	C	C	0.42	0.07	0.02	2.83E-05	0.52	0.06	0.02	9.28E-03	0.07	8.36E-07	0.52	0.01	0.23

EA: effect allele; OA: other allele; AA: ancestral allele; EAF: effect allele frequency, 1KGP: 1000 Genome Project.

SUPPLEMENTARY MATERIALS FOR:

A genome-wide scan on individual typology angle found variants at *SLC24A2* associated with skin color variation in Chinese populations

MATERIALS AND METHODS

Study populations

The Jidong cohort (JD)

The Jidong cohort (JD) is a community-based, long-term observational cohort study to evaluate health related risk factors. The baseline data were collected from 2013 to 2014 in the Staff Hospital, Jidong Oilfield Branch, China. Approval was obtained from the Ethics Committee of Kailuan General Hospital of Tangshan City and the Medical Ethics Committee, Staff Hospital, Jidong Oilfield Branch, China National Petroleum Corporation in July, 2013 (approval No. 2013 YILUNZI1). The approval had been renewed in 2018. To date, 9,078 individuals over 18 years old have been enrolled after excluding individuals who were unable or unwilling to participate. Written informed consent was obtained from all participants. This study included a total of 5,601 individuals (2,512 men and 3,089 women, aged 31-87) who paid the return visit in 2018. The facial photograph and blood samples were collected in the Staff Hospital at the same time.

The National survey of physical traits cohort (NSPT)

The National survey of physical traits cohort (NSPT) is a population-based prospective cohort study to explore environmental and genetic factors associated with physical traits and diseases. The NSPT cohort study was conducted with the official approval of the Ethics Committees of Fudan University (14117) and the Shanghai Institutes for Biological Sciences (ER-SIBS-261410). The NSPT totally collected samples of 3,565 Han Chinese individuals (1,320 men and 2,245 women, aged 17-83 years) in 2015-2018 from three sites (i.e. Taizhou, Nanning and Zhengzhou). All individuals provided written informed consent. Portrait photos of 1,930 individuals (705 men and 1,225 women, aged 18-79) were collected in accordance with phenotyping standard operating procedure. Therefore, only 1,930 individuals were included in this study.

The Taizhou longitudinal cohort (TZL)

The Taizhou longitudinal cohort study (TZL) is a long-term observational cohort study to explore the environmental and genetic risk factors for common and non-communicable diseases. This research program was conducted with the approval of the Ethics Committee of Fudan University (Ethics Research Approval 85), Shanghai, China. The detailed characteristics have been described before (Wang et al., 2009). Our replication set included 1,787 health Han Chinese with portrait photos, ranging from 31-85 years old.

Colombian cohort

The Colombian cohort collected data from participants of several Latin American countries to study the genetics of normal variation in physical appearance. For the genetic analysis of ITA°, 266 participants were recruited in the city of Medellin, Colombia. Ethical approval was obtained from the bioethics committee of the Odontology Faculty at the University of Antioquia (CONCEPTO 01-2013). Age of the participants was between 18 to 50 years, with an average of 27 years.

Phenotyping

All participants were asked not to take part in vigorous exercise an hour before their study visit, not to wear make-up, and to refrain from alcohol and tobacco use 24 hours before the visit. All photographs were taken in a confined space with stabilized LED light source. Besides, all participants wore a shawl to help give consistent light illumination. A Canon 70D digital camera (lens: Canon EF 40mm f/2.8) was used for all subjects without the flash. The facial photograph for each participant consisted of a frontal facial shot with the eyes closed and no facial expression. The resolution of photographs was 300dpi.

This study adopted individual typology angle (ITA°) as a quantitative measurement of skin color, which was derived from high resolution portrait photos. ITA° is a quantitative variable designed for measuring skin pigmentation based on colorimetric parameters. ITA° could be used to classify skin types, i.e., very light (ITA°>55), light (41~55), intermediate (28~41), tan (10~28), brown (-30~10), and dark (<-30) (Chardon et al., 1991). It has been demonstrated

that the ITA˚-based skin color classification is physiologically relevant (Del Bino and Bernerd, 2013).

Portrait photos were processed using an in-house developed facial skin color quantification pipeline, which involves a face detector, automated facial landmarking, cheek segmentation, and color analysis.

The face detection was achieved by the get_frontal_face_detector function from the C++ library Dlib (King, 2009). This function returns an object detector that is configured to find human faces that are looking more or less towards the camera, and the detector is composed of a linear classifier combined with classic Histogram of Oriented Gradients (HOG) features (Dalal, 2005), an image pyramid, and the sliding window detection scheme (Sullivan, 2014). The face landmarking was achieved by the shape_predictor function from the C++ library Dlib based on the face region detected by the get_frontal_face_detector. This predictor was created by training an ensemble of regression trees for face alignment on the iBUG 300-W face landmark dataset (C. Sagonas, 2016, C. Sagonas, 2013a, 2013b). It takes an image of a human face as input and identifies the locations of 68 facial landmarks including the corners of the mouth and eyes, tip of the nose, and edges of cheeks (Figure 1a).

For each side of the face, the pipeline automatically outlines a rectangle to segment the cheek part according to preselected anatomical landmarks in such a way that the rectangle contains the cheek-surrounding landmarks with the minimal horizontal and vertical distances between the landmarks (Figure 1a). After the segmentation, we manually examined the scope of the rectangles to exclude the confounding margins that incorporate photographing background and human hair. For the right cheek, the CIELAB (ISO, 2007) *L*, *a* and *b* values of all segmented pixels were obtained and converted to ITA˚ as $\mathrm{ITA}^\circ = \arctan\frac{L-50}{b} \times \frac{180}{\pi}$, and the mean ITA˚ was calculated to represent facial skin color of a portrait. As ITA˚ values were not subject to the normal distribution both in JD and NSPT (Figure S2a-b), we applied Z-transformed ITA˚ (z-ITA˚) in further analysis.

For quality control purposes, additional measures were obtained in a subgroup of 50 randomly selected photographs. These included accessing skin colorimetric parameters on the right cheek using a spectrophotometer (Skin Colorimeter CL 400, Courage+Khazaka Electronics, GmbH) and four skin color types (very light, light, intermediate and tan) perceived by an investigator according to the Chardon skin color type (Chardon et al., 1991). Note that dark skin color types were absent in our Chinese sample. Inter-method reliability was evaluated with the Kappa Statistic. No significant difference in ITA˚ was detected between the left and right cheeks (t-test P=0.13). Chardon skin color type showed a high degree of concordance with the perceived skin color type, as well as with the spectrophotometer values (Figure S1b-d).

For phenotyping in the Taizhou Longitudinal cohort, the photos were taken with a Canon 70D digital camera (lens: Canon EF 40mm f/2.8). The rest of the procedures (i.e. automated facial landmarking, cheek segmentation and color analysis) were the same as those used in the discovery cohorts.

For the Colombian cohort, ITA˚ was measured on the forehead. Due to the smaller sample size, outlier values on either tail of the ITA˚ were excluded (values ≤ -3 or ≥ 54). Z-transformed ITA˚ (z-ITA˚) was used for the analysis.

Among 1,887 subjects in the NSPT cohort, we also collected the trans-epidermal water loss (TEWL) measurement. TEWL was measured on cheek using Tewameter® TM 300 of Courage & Khazaka Electronics (median [IQR], g/h/m² = 10.60 [8.20-13.90]).

Genotyping

For both JD and NSPT cohorts, genomic DNA was extracted from blood samples using the MagPure Blood DNA KF Kit. All samples were genotyped using the Illumina Infinium Global Screening Array (GSA) consisting of about 710,000 SNPs. We implemented exclusion criteria for quality control using PLINK v1.9 (Chang et al., 2015). In detail, we excluded subjects with more than 5% missing data, the duplicated subjects, and subjects that failed the

X-chromosome gender concordance check or had ethnic information incompatible with their genetic information. We excluded SNPs that had more than 2% missing data, with a minor allele frequency (MAF) of less than 1%, and the ones that failed Hardy-Weinberg (HW) deviation test ($P < 1 \times 10^{-5}$). Imputation was performed using the 1000 Genomes Project Phase 3 as the reference panel. The chip genotype data was firstly phased using SHAPEIT3 (O'Connell et al., 2016). And IMPUTE2 was then used to impute genotypes (Howie et al., 2009). SNPs with an imputation quality score (INFO) of less than 0.6, MAF less than 0.01 or a missing rate more than 0.01 were eliminated from further analyses. Finally, 8,039,700 SNPs passed quality control and were used for further analyses.

For TZL, blood samples were collected, and DNA was extracted. All samples were genotyped using the Illumina HumanOmniZhongHua-8 chip, which interrogates 894,517 SNPs. After quality control with PLINK v1.9, the genotype data were phased using SHAPEIT and were imputed using IMPUTE2 with the 1000 Genomes Project Phase 3.

Blood samples were collected from the participants in the Colombian cohort, DNA was extracted, and genotyped on the Illumina Infinium Global Screening Array (GSA). After quality control using PLINK v1.9, 511,848 SNPs were retained. The genotype data was then phased using SHAPEIT2, and imputed using IMPUTE2 with the 1000 Genomes Project Phase 3 as the reference panel. Genotype data for SNP rs10122939 was obtained from the imputed dataset.

Statistical analysis

GWASs and meta-analysis

GWASs were separately conducted in JD and NSPT on Z-transformed ITA° values using software package PLINK (Chang et al., 2015) where additive allele effects were tested in linear models adjusted for covariates (age, gender, sampling locations and the top four genomic principal components). GWAS outputs were combined using inverse variance fixed-effect meta-analysis as implemented in METAL (Willer et al., 2010). P-values equal to

or smaller than $P=5\times10^{-8}$ from the meta-analysis were considered as genome-wide significant. The inflation factor was estimated close to 1.0 ($\lambda = 0.995$) and not further considered. GWAS results were visualized using Manhattan plots, Q-Q plots. Regional LD plots and association plots were generated using Haploview and LocusZoom (Barrett et al., 2005, Pruim et al., 2010). For the association analysis in the replication cohorts, additive models were used, with age, sex and 5 principal components of genotypes as covariates.

In the TZL and the Colombian cohort, Genetic Principal Components (PCs) were calculated from the chip genotype data. After QC, association test between SNP rs10122939 and Z-transformed ITA˚ values was conducted using PLINK, including age, sex, and 5 genetic PCs as covariates.

Candidate gene analysis

To compare our GWAS results with previous GWAS findings related to skin color, we conducted an examination for a list of 11,384 SNPs identified by recently large sample GWAS for tanning response to sun exposure (Visconti et al., 2018), skin pigmentation (Adhikari et al., 2019), hair color (Hysi et al., 2018) and eye color (Simcoe et al., 2021). Out of 11,384 SNPs, 9,183 were available in our GWAS results. SNPs with p-values smaller than Bonferroni-corrected threshold of 0.05/9,183 (1.09×10^{-6}) in our GWAS results were considered as significant.

Genetic diversity and signatures of positive selection

In order to detect the signatures of positive selection, we used two statistics commonly adopted in population genetics including the fixed index (F_{ST}) and integrated haplotype score (iHS).F_{ST} measures the proportion of genetic diversity caused by differences in allele frequencies between populations that is potentially resulted from the divergent natural selection (Holsinger and Weir, 2009). We used GCTA (Yang et al., 2011) to derive genome-wide F_{ST} analysis for 10 pairs of groups (pairwise among AFR, EAS, EUR, AMR, and SAS) based on the genotype data of samples from 1000 Genome Project. The empirical

significance cutoff was set to the values of the top 1% and top 5% of F_{ST} in the corresponding pairs of population in the whole genome. iHS is an integration of extended haplotype homozygosity (EHH) statistics to highlight the recent positive selection mutations that have not yet reached a fixed level. Larger values indicated adaptation to selection pressure in the most recent stage of evolution (Sabeti et al., 2002, Voight et al., 2006). Firstly, We used the selscan software (Szpiech and Hernandez, 2014) to calculate the iHS scores of SNPs based on genotype data from the 5 populations including AFR, EAS, EUR, AMR, and SAS of the 1000 Genome Project. The absolute iHS scores of the top 1% and 5% of all SNPs in these 22 randomly selected regions for each population were used as the empirical significance cutoff of corresponding population. Then, we calculated the iHS scores of SNPs within 200 kb upstream and downstream of the lead SNP we identified, and the iHS absolute score represents the selection intensity.

Functional Test Using Dual-Luciferase Reporter Assay

We amplified a 115 bp fragment (GRCh38.p13) of *SLC24A2* by PCR from genomic DNA of two homozygous individuals with respect to the corresponding genotypes rs10122939 (AA and GG) using primers tailed with *Kpn* I and *Xho* I restriction sites for rs10122939, and directionally subcloned them into the pGL3-promter expression vector (Promega). We verified all recombinant clones by sequencing. A375 Melanoma cells lines and HEK293 cell lines were routinely cultured, transfected and incubated overnight at 37℃ in 5% CO_2. A375 and HEK293 cells were cultured in high-glucose Dulbecco's modified Eagle's medium (DMEM) (Corning) with 10% fetal bovine serum (FBS) (Gibco). Transient transfection assays were conducted in these cells using the Lipofectamine 2000 transfection reagent (Invitrogen). All assays were performed with minimum of four replicates. After 48h incubation, we collected the cells and measured luciferase activity using the Dual-Luciferase Reporter Assay System (Beyotime).

References:

Adhikari K, Mendoza-Revilla J, Sohail A, Fuentes-Guajardo M, Lampert J, Chacon-Duque JC, et al. A GWAS in Latin Americans highlights the convergent evolution of lighter skin pigmentation in Eurasia. Nat Commun 2019;10(1):358.

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21(2):263-5.

C. Sagonas EA, G, Tzimiropoulos, S. Zafeiriou, M. Pantic. 300 faces In-the-wild challenge: Database and results. Image and Vision Computing (IMAVIS), Special Issue on Facial Landmark Localisation "In-The-Wild" 2016.

C. Sagonas GT, S. Zafeiriou, M. Pantic. 300 Faces in-the-Wild Challenge: The first facial landmark localization Challenge. Proceedings of IEEE Int'l Conf on Computer Vision (ICCV-W), 300 Faces in-the-Wild Challenge (300-W) 2013a.

C. Sagonas GT, S. Zafeiriou, M. Pantic. A semi-automatic methodology for facial landmark annotation. Proceedings of IEEE Int'l Conf Computer Vision and Pattern Recognition (CVPR-W), 5th Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2013) 2013b.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015;4:7.

Chardon A, Cretois I, Hourseau C. Skin colour typology and suntanning pathways. Int J Cosmet Sci 1991;13(4):191-208.

Dalal NaT, Bill. Histograms of Oriented Gradients for Human Detection. Comput Vision Pattern Recognit 2005;1:886-93.

Del Bino S, Bernerd F. Variations in skin colour and the biological consequences of ultraviolet radiation exposure. Br J Dermatol 2013;169 Suppl 3:33-40.

Holsinger KE, Weir BS. Genetics in geographically structured populations: defining, estimating and interpreting F(ST). Nat Rev Genet 2009;10(9):639-50.

Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 2009;5(6):e1000529.

Hysi PG, Valdes AM, Liu F, Furlotte NA, Evans DM, Bataille V, et al. Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability. Nat Genet 2018;50(5):652-6.

ISO. CIE standard colorimetric observers international organization for standardization. ISO 11664–1:2007 colorimetry – Part 1 2007.

King DE. Dlib-ml: A Machine Learning Toolkit. Journal of Machine Learning Research 2009;10:1755-8.

O'Connell J, Sharp K, Shrine N, Wain L, Hall I, Tobin M, et al. Haplotype estimation for biobank-scale data sets. Nat Genet 2016;48(7):817-20.

Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010;26(18):2336-7.

Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, et al. Detecting recent positive selection in the human genome from haplotype structure. Nature 2002;419(6909):832-7.

Simcoe M, Valdes A, Liu F, Furlotte NA, Evans DM, Hemani G, et al. Genome-wide association study in almost 195,000 individuals identifies 50 previously unidentified genetic loci for eye color. Sci Adv 2021;7(11).

Sullivan VKaJ. One millisecond face alignment with an ensemble of regression trees. 2014 IEEE Conference on Computer Vision and Pattern Recognition 2014:1867-74.

Szpiech ZA, Hernandez RD. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol Biol Evol 2014;31(10):2824-7.

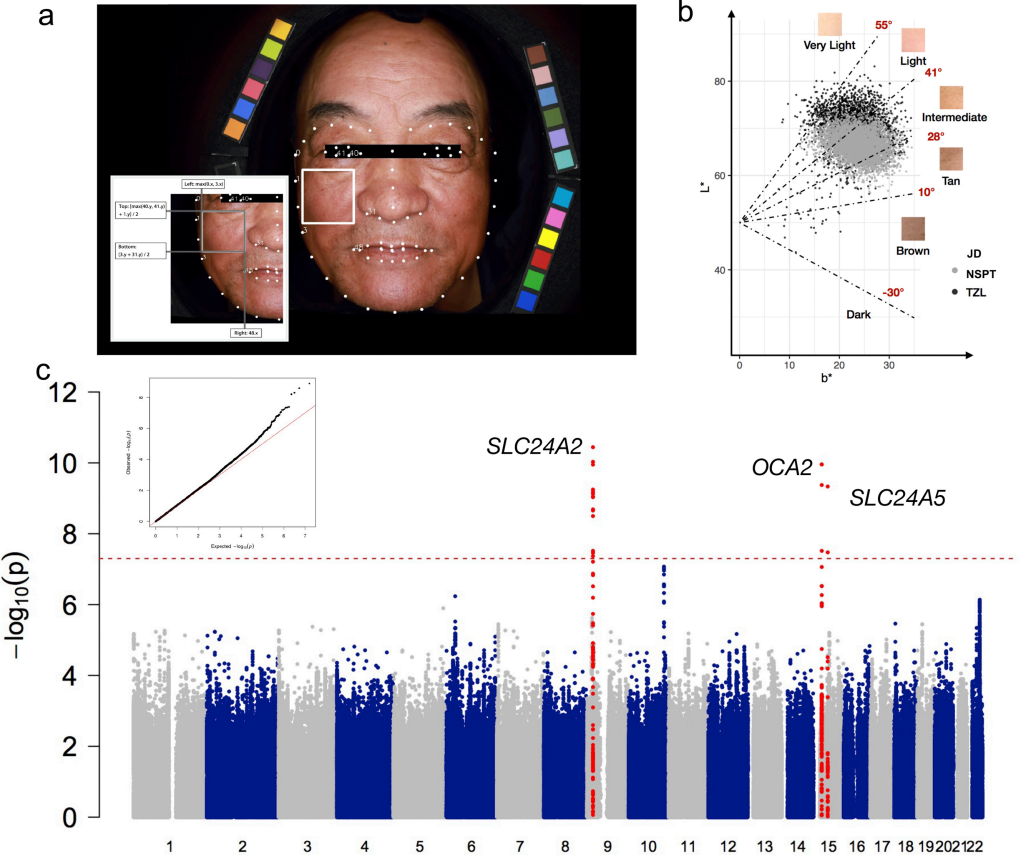
Visconti A, Duffy DL, Liu F, Zhu G, Wu W, Chen Y, et al. Genome-wide association study in 176,678 Europeans reveals genetic loci for tanning response to sun exposure. Nat Commun 2018;9(1):1684.

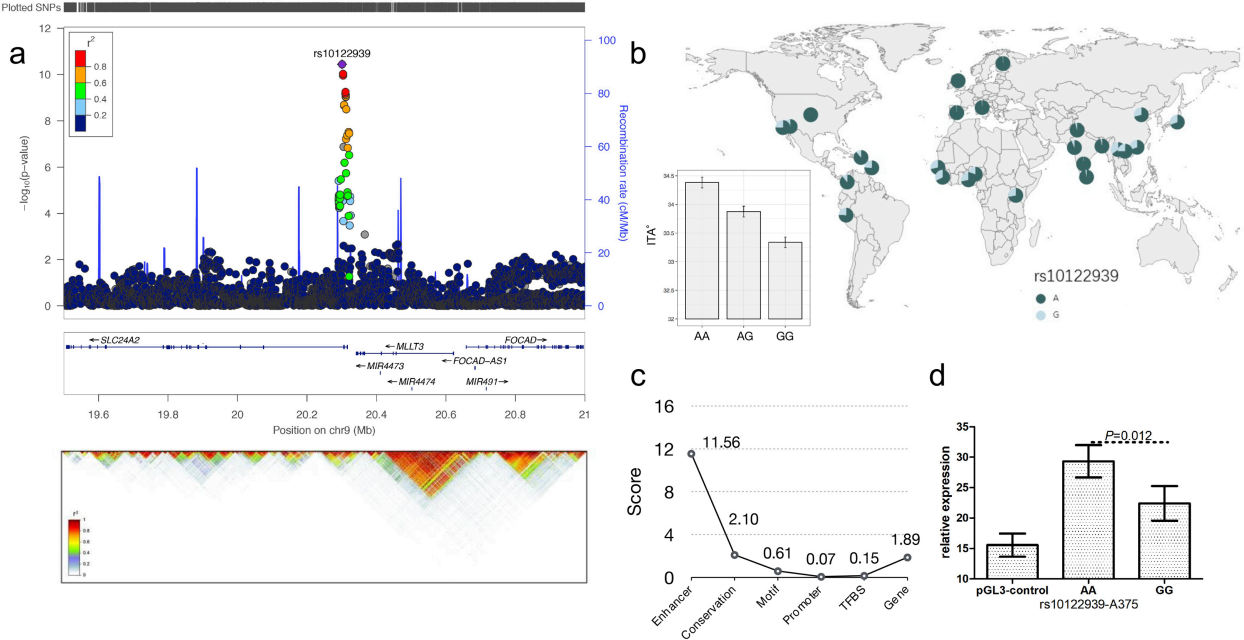
Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biol 2006;4(3):e72.

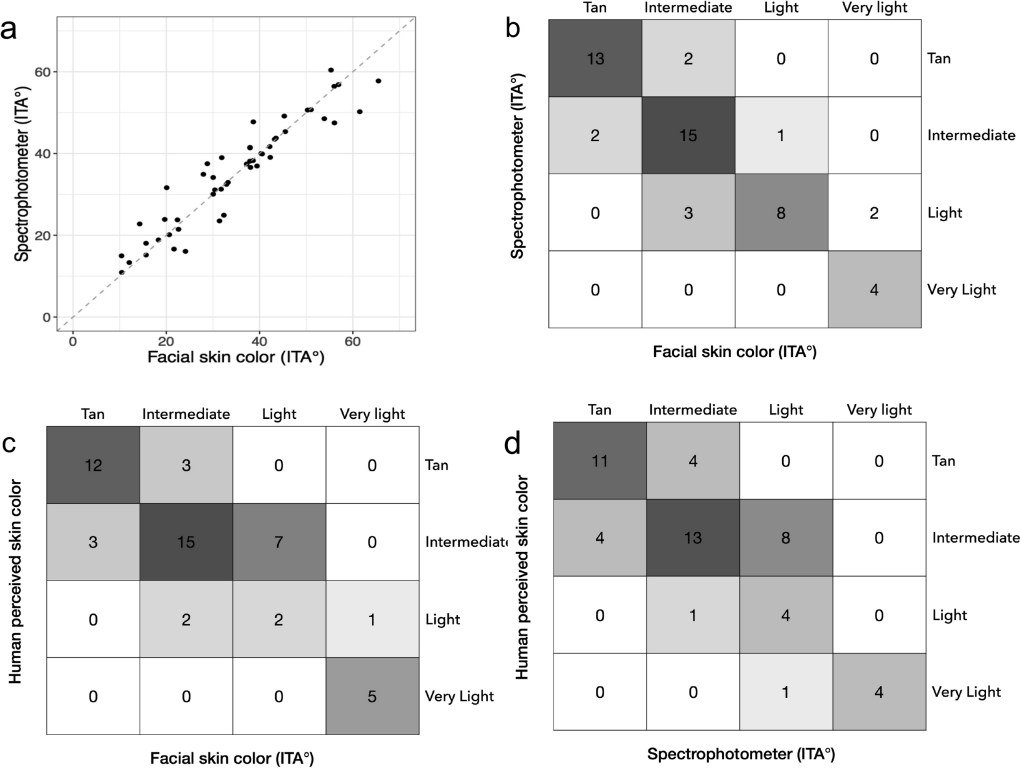
Wang X, Lu M, Qian J, Yang Y, Li S, Lu D, et al. Rationales, design and recruitment of the Taizhou Longitudinal Study. BMC Public Health 2009;9:223.

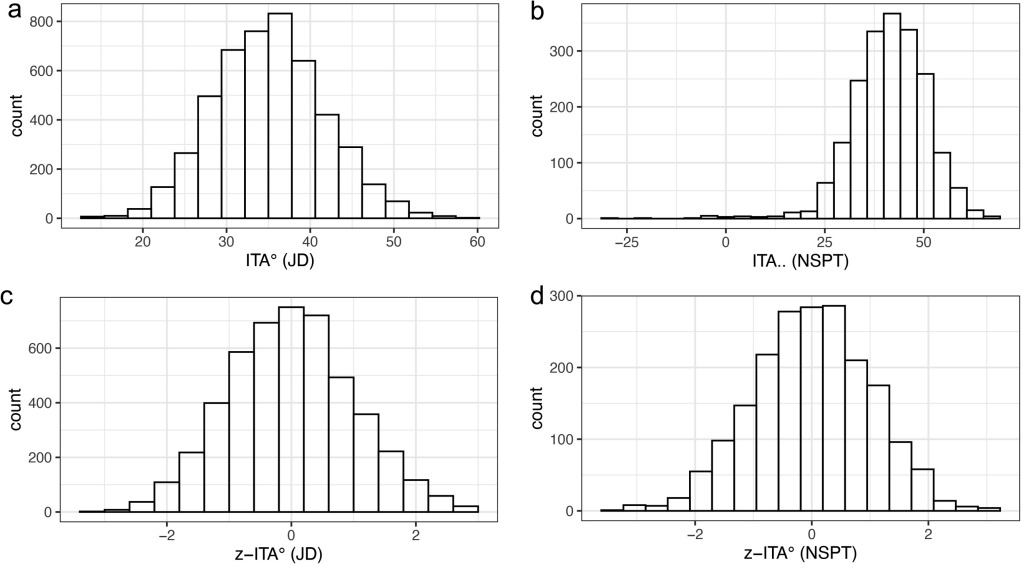
Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26(17):2190-1.

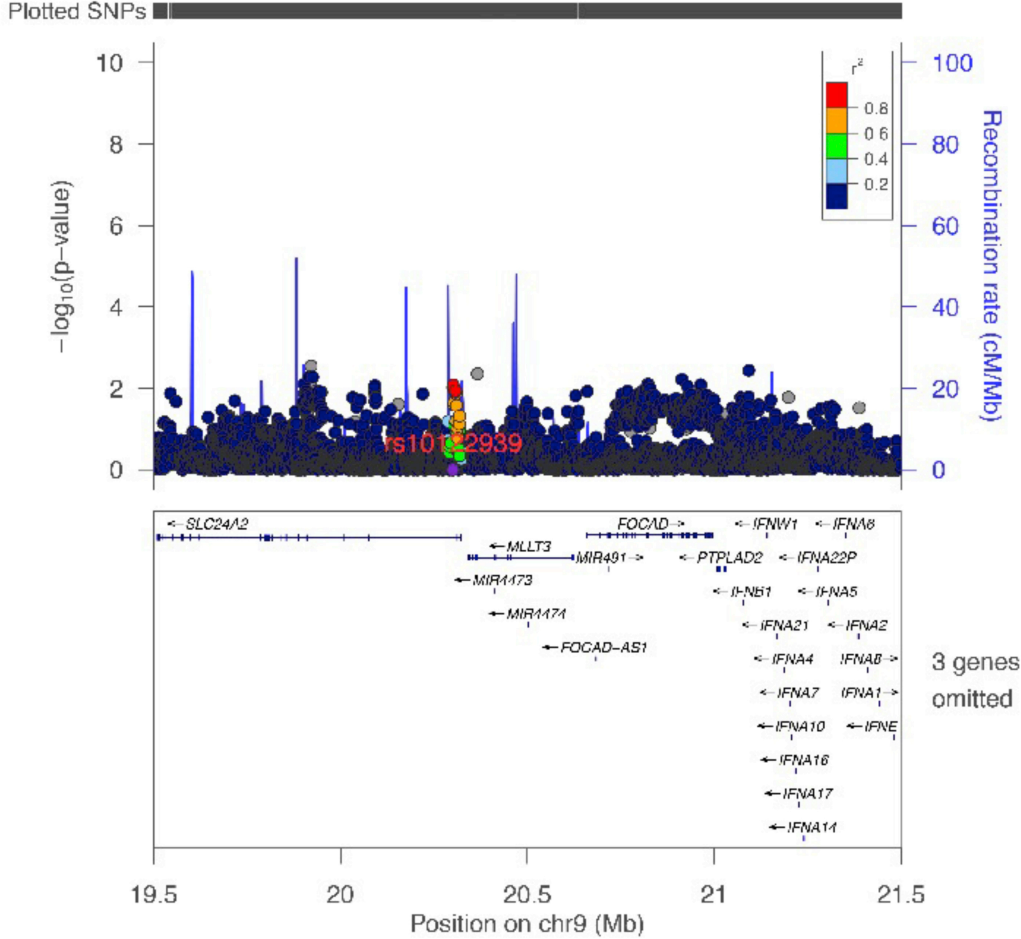
Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011;88(1):76-82.

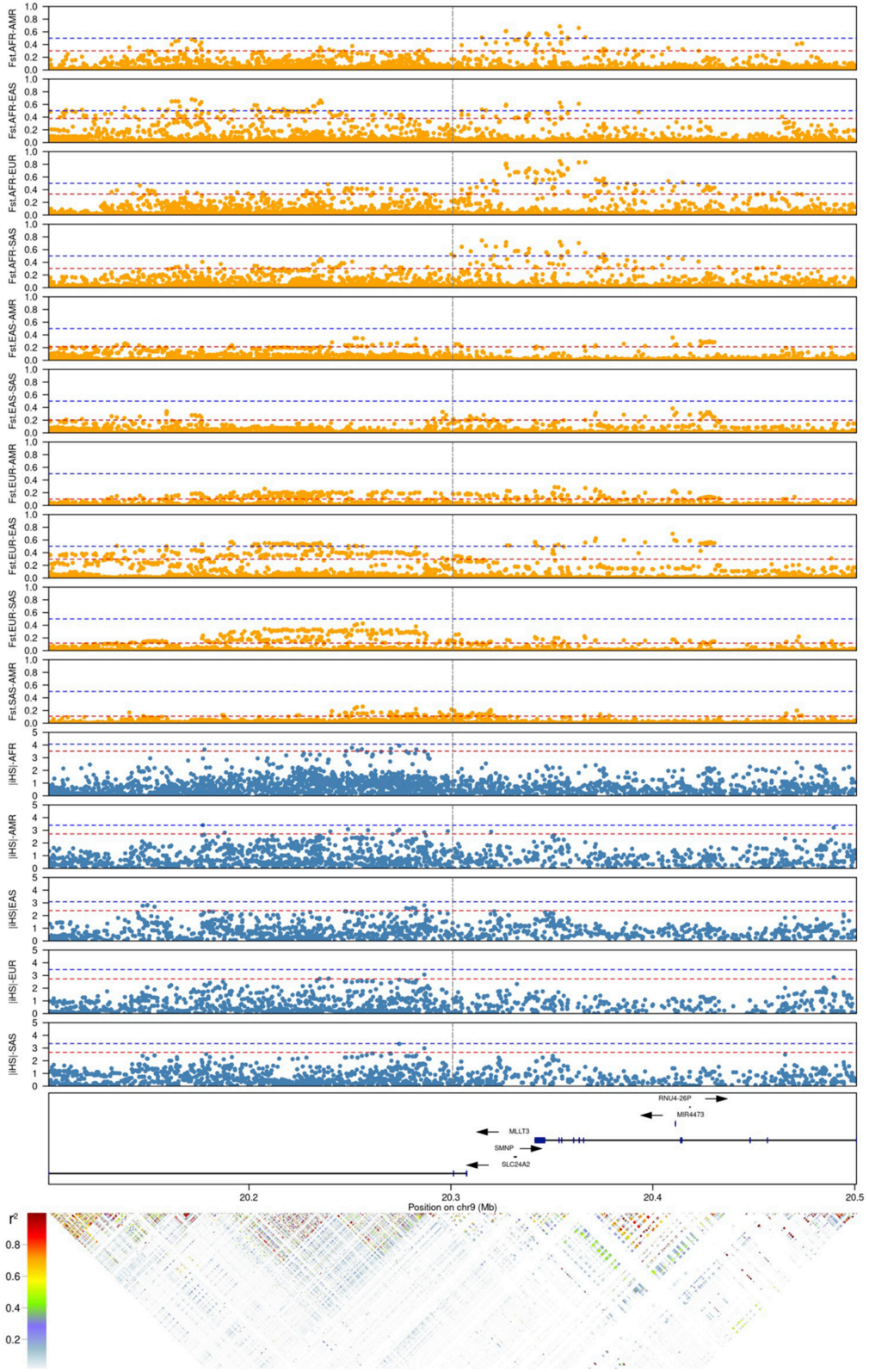

a
b
Very Light
Light
Intermediate
Tan
Brown
Dark
55°
41°
28°
10°
-30°
JD
NSPT
TZL
L*
b*
c
SLC24A2
OCA2
SLC24A5
−log10(p)
Observed −log10(p)
Expected −log10(p)

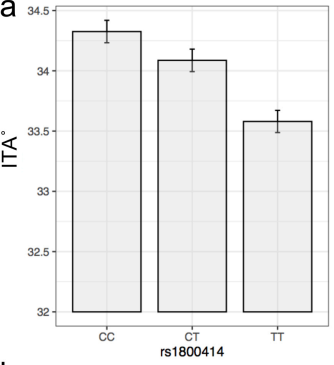

Plotted SNPs
a
rs10122939
r²
0.8
0.6
0.4
0.2
−log10(p-value)
Recombination rate (cM/Mb)
SLC24A2
MLLT3
FOCAD
FOCAD-AS1
MIR4473
MIR4474
MIR491
Position on chr9 (Mb)
b
ITA²
AA
AG
GG
rs10122939
A
G
c
Score
11.56
2.10
0.61
0.07
0.15
1.89
Enhancer
Conservation
Motif
Promoter
TFBS
Gene
d
relative expression
P=0.012
pGL3-control
AA
GG
rs10122939-A375

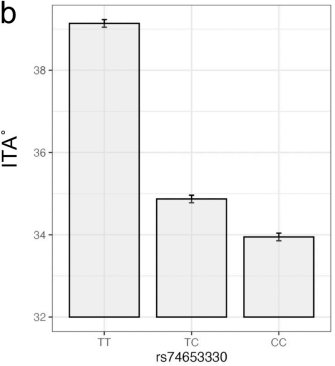

a
Spectrophotometer (ITA°)
Facial skin color (ITA°)
0
20
40
60
b
Tan
Intermediate
Light
Very light
Spectrophotometer (ITA°)
Facial skin color (ITA°)
13 2 0 0
2 15 1 0
0 3 8 2
0 0 0 4
Tan
Intermediate
Light
Very Light
c
Tan
Intermediate
Light
Very light
Human perceived skin color
Facial skin color (ITA°)
12 3 0 0
3 15 7 0
0 2 2 1
0 0 0 5
Tan
Intermediate
Light
Very Light
d
Tan
Intermediate
Light
Very light
Human perceived skin color
Spectrophotometer (ITA°)
11 4 0 0
4 13 8 0
0 1 4 0
0 0 1 4
Tan
Intermediate
Light
Very Light

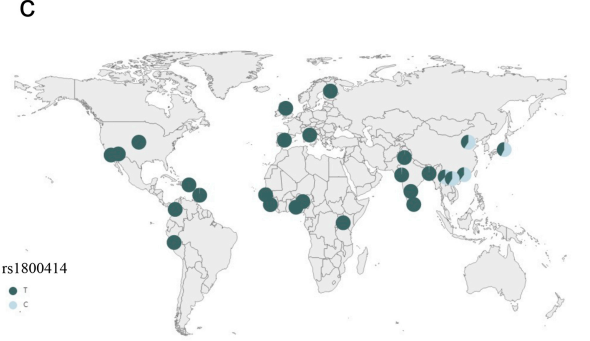

a
count
ITA° (JD)
b
count
ITA.. (NSPT)
c
count
z−ITA° (JD)
d
count
z−ITA° (NSPT)


Plotted SNPs
10
8
6
4
2
0
−log10(p−value)
100
80
60
40
20
0
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
rs10122939
SLC24A2
FOCAD
IFNW1
IFNA6
MLLT3
MIR491
PTPLAD2
IFNA22P
MIR4473
IFNB1
IFNA5
MIR4474
IFNA21
IFNA2
FOCAD-AS1
IFNA4
IFNA8
3 genes
omitted
IFNA7
IFNA1
IFNA10
IFNE
IFNA16
IFNA17
IFNA14
19.5
20
20.5
21
21.5
Position on chr9 (Mb)

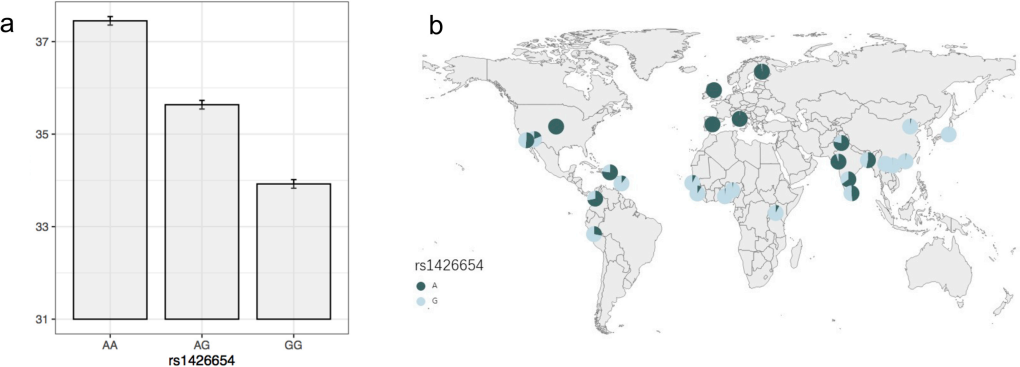

Fst:AFR:AMR
Fst:AFR:EAS
Fst:AFR:EUR
Fst:AFR:SAS
Fst:EAS:AMR
Fst:EAS:SAS
Fst:EUR:AMR
Fst:EUR:EAS
Fst:EUR:SAS
Fst:SAS:AMR
|iHS|:AFR
|iHS|:AMR
|iHS|:EAS
|iHS|:EUR
|iHS|:SAS
RNU4-26P
MIR4473
MLLT3
SMNP
SLC24A2
20.2
20.3
20.4
20.5
Position on chr9 (Mb)
r²
0.8
0.6
0.4
0.2

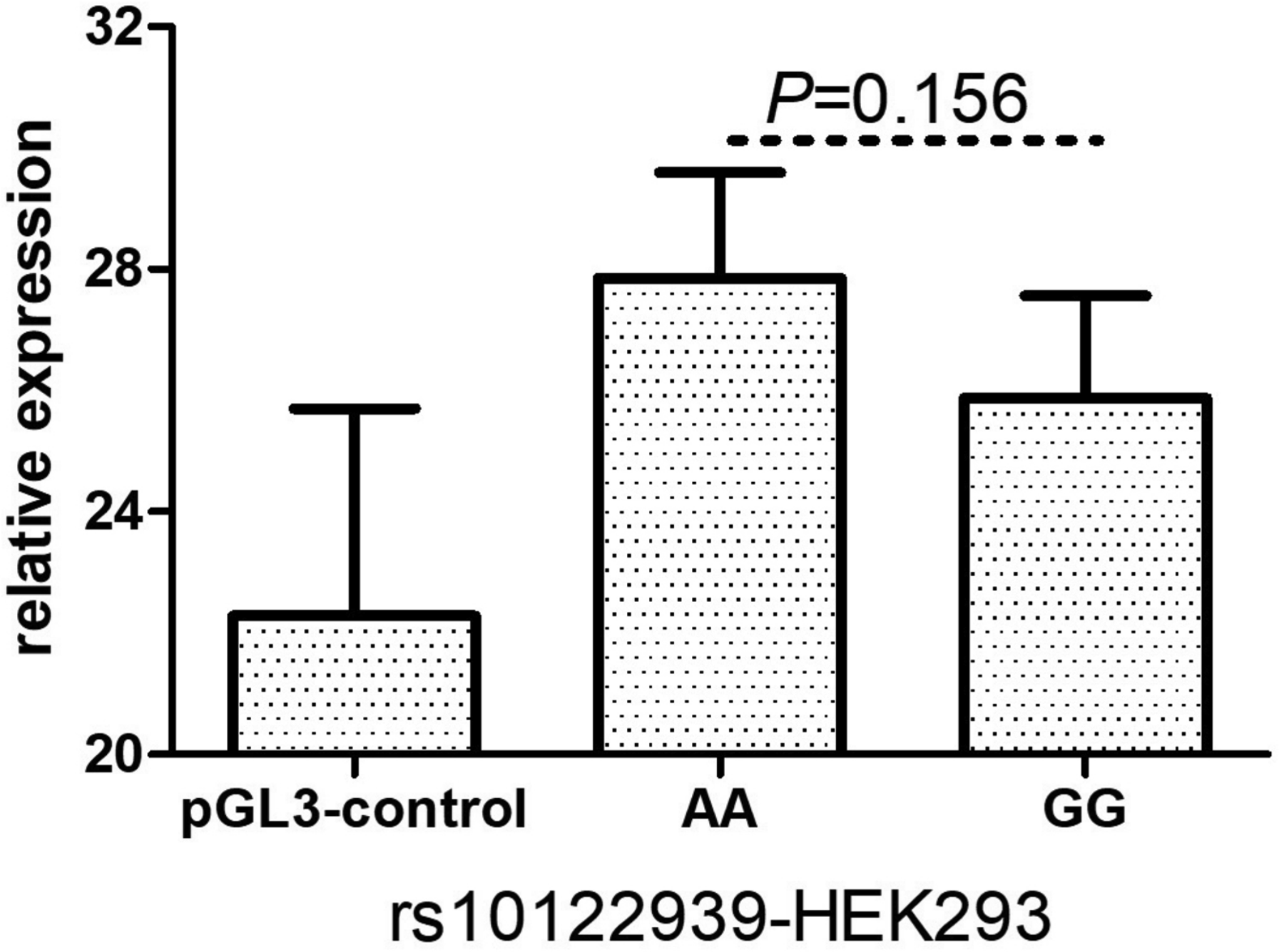

a
34.5
34
33.5
33
32.5
32
ITA°
CC
CT
TT
rs1800414

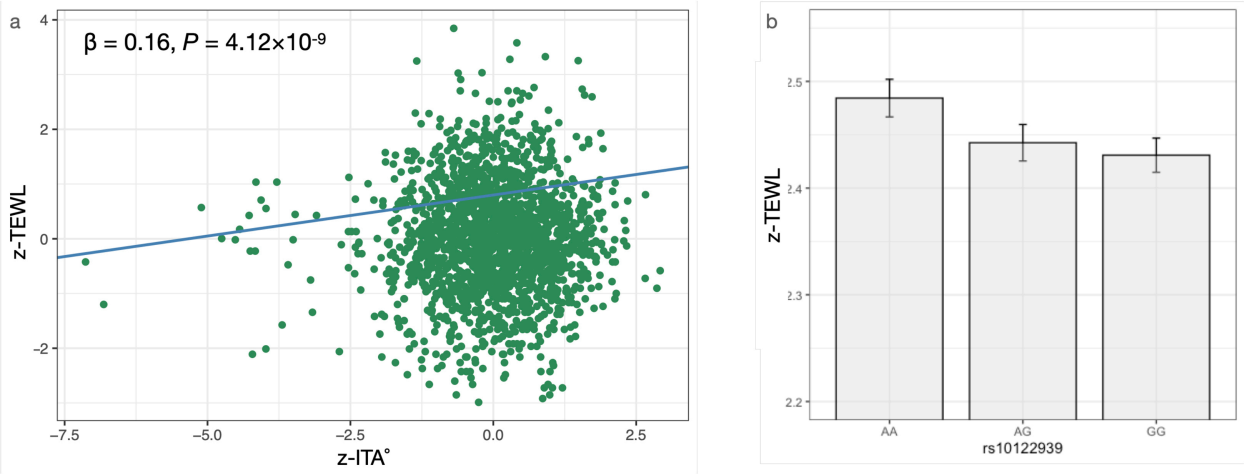

b
38
36
34
32
ITA°
TT
TC
CC
rs74653330


c
rs1800414
T
C


d
rs74653330
C
T


a
37
35
33
31
AA
AG
GG
rs1426654
b
rs1426654
A
G


32
28
24
20
relative expression
P=0.156
pGL3-control
AA
GG
rs10122939-HEK293


a
β = 0.16, P = 4.12×10-9
z-TEWL
z-ITA°
4
2
0
−2
−7.5
−5.0
−2.5
0.0
2.5
b
z-TEWL
2.5
2.4
2.3
2.2
AA
AG
GG
rs10122939