***Nina Rossa Haddad*[1], *Beita Badiei*[1], *S. Leigh Curvin-Aquilla*[1], *Arieana Y. Johnson*[1], *Aiden Willis*[1], *Hana B. Minsky*[1], *Kaitlin L. Williams*[1] *and Luis A. Garza*[1,2,3,*]**

[1]*Department of Dermatology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA;* [2]*Department of Cell Biology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; and* [3]*Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA*
[*]*Corresponding author e-mail: lag@jhmi.edu*

REFERENCES

French SA, Tangney CC, Crane MM, Wang Y, Appelhans BM. Nutrition quality of food purchases varies by household income: the SHoPPER study. BMC Public Health 2019;19:231.

González-Castro MI, Olea-Serrano MF, Rivas-Velasco AM, Medina-Rivero E, Ordoñez-Acevedo LG, De León-Rodríguez A. Phthalates and bisphenols migration in Mexican food cans and plastic food containers. Bull Environ Contam Toxicol 2011;86:627–31.

Haddad NR, Badiei B, Williams KL, Garza LA. Positive correlation of hidradenitis suppurativa and ultra-processed foods consumption. Arch Dermatol Res 2024;316:172.

Lee AG, Kang S, Yoon HJ, Im S, Oh SJ, Pak YK. Polystyrene microplastics exacerbate systemic inflammation in high-fat diet-induced obesity. Int J Mol Sci 2023;24:12421.

Lin Q, Zhao S, Pang L, Sun C, Chen L, Li F. Potential risk of microplastics in processed foods: preliminary risk assessment concerning polymer types, abundance, and human exposure of microplastics. Ecotoxicol Environ Saf 2022;247:114260.

Liu Z, Lu Y, Zhong K, Wang C, Xu X. The associations between endocrine disrupting chemicals and markers of inflammation and immune responses: a systematic review and meta-analysis. Ecotoxicol Environ Saf 2022;234:113382.

Maarouf M, Platto JF, Shi VY. The role of nutrition in inflammatory pilosebaceous disorders: implication of the skin-gut axis. Australas J Dermatol 2019;60:e90–8.

Preda-Naumescu A, Ahmed HN, Mayo TT, Yusuf N. Hidradenitis suppurativa: pathogenesis, clinical presentation, epidemiology, and comorbid associations. Int J Dermatol 2021;60:e449–58.

Singh R, Fathy R, Kassamali B, Noe MH, Barbieri JS, LaChance A, et al. Increased ambient outdoor temperatures are associated with increased disease flaring in hidradenitis suppurativa. Arch Dermatol Res 2023;316:49.

van Straalen KR, Prens EP, Gudjonsson JE. Insights into hidradenitis suppurativa. J Allergy Clin Immunol 2022;149:1150–61.

Enhancer of *TRPS1* rs12549956 Influences Hair Thickness in Chinese Populations

Journal of Investigative Dermatology (2025) **145,** 1202–1205; doi:10.1016/j.jid.2024.10.601

TO THE EDITOR

Hair thickness is a critical parameter in dermatology and cosmetology, particularly for hair loss treatments. The efficacy of procedures such as follicular unit transplantation and extraction is highly contingent upon the individual's hair thickness, necessitating tailored treatment approaches. Despite its clinical importance, the genetic basis of hair thickness remains poorly explored. To date, only Fujimoto et al (2009, 2008) have examined hair thickness in Asian populations using cross-sectional hair images. Their candidate gene analysis of hair morphogenesis identified genetic loci in *EDAR*, involved in ectodermal development, and *FGFR2*, associated with epithelial proliferation, both of which are closely related to hair thickness (Fujimoto et al, 2009, 2008).

We conducted a GWAS on hair thickness, involving 3682 individuals across 2 Chinese cohorts using a discovery-replication design. The discovery phase included 2961 Chinese participants (64.2% female) with a wide age range (31–87 years) from the Taizhou longitudinal (TZL) cohort, with ethical approval granted by the Ethics Committee of Fudan University (Shanghai, China) (ethics research approval 85). All participants provided written, informed consent. The replication phase involved 721 young Uyghur (UYG) individuals (mean age = 20.1 ± 1.3 years, 60.9% female) recruited by Xinjiang Medical University, with approval from the Shanghai Institutes for Biological Sciences (ERSIBS-261410). Written, informed consent was obtained from all participants.

Hair thickness was quantitatively assessed in the vertex region using ×4 microscopic images, processed through our image processing algorithm (Supplementary Materials and Methods and Supplementary Figure S1). The digital measurements of hair thickness were highly concordant with assessments independently performed by a dermatologist, who analyzed a random subset of 400 microscopic images from both the TZL (Pearson's $r = 0.96$) (Supplementary Figure S2) and UYG ($r = 0.94$) cohorts.

Hair thickness exhibited a largely normal distribution in both the TZL and UYG cohorts, with TZL participants displaying significantly thicker hair than those in UYG (mean thickness = 84 vs 80 μm; $P < 7.4 \times 10^{-12}$) (Supplementary Figure S2). Age was inversely correlated with hair thickness, showing a decrease of 3.12 μm per decade ($P = 1.9 \times 10^{-30}$). In addition, men had significantly thinner hair than women, with a difference of 1.83 μm ($P = 1.8 \times 10^{-4}$). Notably, individuals with straight hair had thicker hair than those with curly hair, especially in the TZL cohort, where the difference was 3.14 μm ($P = 1.6 \times 10^{-3}$) (Supplementary Figure S3).

Genomic principal component analysis, integrated with global samples from the 1000 Genomes Project, affirmed the genetic closeness of our TZL samples to East Asians, while positioning our UYG samples between East Asians and Europeans (Supplementary Figure S4). SNP-based heritability for hair thickness in the TZL cohort was estimated at 0.17 using GCTA (genome-wide complex trait analysis) (standard error = 0.08, $P = .03$) (Supplementary Table S1), with no evidence of genome-wide inflation observed

Abbreviations: TZL, Taizhou longitudinal; UYG, Uyghur

Accepted manuscript published online 14 November 2024; corrected proof published online 5 December 2024

Table1. Summary of the Putative Causal SNPs in Our Study and Previous Studies

					TZL			UYG			Meta-Analysis	
SNP	CHR	BP[1]	Gene	EA[2]	EAF	Beta	*P*-Value	EAF	Beta	*P*-Value	Beta	*P*-Value
rs56022216	8	117326907	*TRPS1*	C	0.57	2.13	9.57×10^{-9}	0.70	2.70	1.36×10^{-2}	2.22	3.82×10^{-11}
rs12549956	8	117306816		C	0.57	2.05	2.66×10^{-8}	0.69	2.75	1.01×10^{-2}	2.16	1.18×10^{-10}
rs3827760[3]	2	109257152	*EDAR*	G	0.95	1.47	7.41×10^{-2}	0.38	2.37	4.53×10^{-3}	1.92	4.27×10^{-3}
rs4752566[3]	10	181511951	*FGFR2*	T	0.88	−1.32	1.77×10^{-2}	0.77	0.84	2.96×10^{-1}	−0.62	9.62×10^{-2}

Abbreviations: CHR, chromosome; BP, base pair; EA, effect allele; EAF, effect allele frequency; TZL, Taizhou longitudinal; UYG, Uyghurs.
[1]BP positions are according to human reference hg19.
[2]The allele is the genotype that increases hair thickness.
[3]These SNPs are previously reported to be associated with hair thickness.

(lambda = 1.01) (Supplementary Figure S5). Our GWAS, which adjusted for sex, age, hair curliness, and genomic principal components, identified 20 SNPs within the intronic region of *LINC00536* at locus 8q23.3, displaying genome-wide significant associations with hair thickness; the most pronounced signal was at rs56022216 ($P = 9.57 \times 10^{-9}$) (Supplementary Figure S5 and Table 1). Repeating the GWAS while conditioning on rs56022216 or excluding adjustments for hair curliness did not reveal any additional loci. These 20 SNPs showed high linkage disequilibrium ($0.5 < r^2 \leq 1.0$) and consistently demonstrated nominally significant associations in the UYG cohort in the same direction of effect. A meta-analysis of the TZL and UYG data further enhanced the significance at 8q23.3 (rs56022216 $P = 3.82 \times 10^{-11}$) (Figure 1b), without uncovering new loci.

Although the signal at 8q23 demonstrated a consistent effect across Chinese Han and UYG populations, it likely does not account for major global variations in hair thickness, given the allele frequencies and effect size observed. Analysis of the 1000 Genomes Project dataset revealed a trend in the frequencies of the derived A allele of rs56022216, which is associated with thinner hair, decreasing gradually from East Asians (0.52) (Supplementary Figure S6) to Europeans (0.20), whereas Africans display frequencies similar to those of East Asians (0.45). This distribution differs from the global patterns of hair thickness, where East Asians possess the thickest hair. Furthermore, no significant positive selection signals were detected at 8q23 (Supplementary Figure S7), suggesting a widespread fixation of the major allele across diverse continental groups.

A lookup of 2 previously reported hair thickness–associated SNPs in our GWAS results showed that rs3827760 in *EDAR* reached a nominally significant association with hair thickness in the same effect direction, whereas rs4752566 in *FGFR2* was not significant (Table 1), thus partially replicating previous findings. In addition, rs3827760 has also been reported to be significantly associated with hair curliness (Wu et al, 2016). We then looked up 12 other previously reported hair curliness–associated SNPs (Liu et al, 2018; Medland et al, 2009) in our GWAS, and none was nominally significant (Supplementary Table S2).

To elucidate the role of 8q23.3 in hair morphogenesis, we conducted a series of functional annotations using multiple databases, including Combined Annotation Dependent Depletion, DeepSEA, EIGEN, FunSeq2, GWAVA, and REMM. Except for GWAVA, all identified rs12549956 as the most likely functionally causal variant (Supplementary Table S3). This SNP demonstrated strong enhancer activities in Roadmap's histone modification tracks, notably H3K4me1, in fibroblasts and keratinocytes (Figure 1e). The 3DIV HI-C database located the associated region within the same topologically associating domain as *LINC00536* and *TRPS1*, with significant chromatin interactions exclusively observed between rs12549956 and *TRPS1* (Figure 1f). Single-cell sequencing data from human hair follicles (Ober-Reynolds et al, 2023) showed that *TRPS1* is most highly expressed in the dermal papilla (Figure 1g). Pathogenic variants of *TRPS1* are known to cause trichorhinophalangeal syndrome types I and III, characterized by sparse and slow-growing scalp hair (Momeni et al, 2000). Similarly, in mouse models (Ge et al, 2020), *Trps1* shows peak expression in dermal papilla from embryonic day 16.5. Skin from *Trps1*-knockout newborn mice, when transplanted onto the backs of nude mice, exhibits smaller hair follicles and thinner hair shafts than wild-type skin, indicating a critical role for *Trps1* in hair follicle development (Fantauzzo and Christiano, 2012; Zhang et al, 2019).

In conclusion, our pioneering GWAS on hair thickness in Chinese populations identified 8q23 as a crucial genetic susceptibility locus. Comprehensive functional annotation analyses pinpointed rs12549956 as an enhancer of *TRPS1*, a gene that plays a significant role in regulating hair thickness in humans. Although our findings provide insights into the genetic basis of hair thickness, we hope that future research will expand upon these results by performing functional validation to further elucidate the underlying mechanisms.

DATA AVAILABILITY STATEMENT

The GWAS summary statistics are available from the National Omics Data Encyclopedia (http://www.biosino.org/node/) under the project identification document OEP005375. Data usage must be in full compliance with the Regulations on Management of Human Genetic Resources in China. Individual genotype and phenotype data cannot be shared owing to Institutional Review Board restrictions on privacy concerns. Other relevant data supporting the key findings of this study are available within the letter and Supplementary Materials or from the corresponding author on reasonable request.

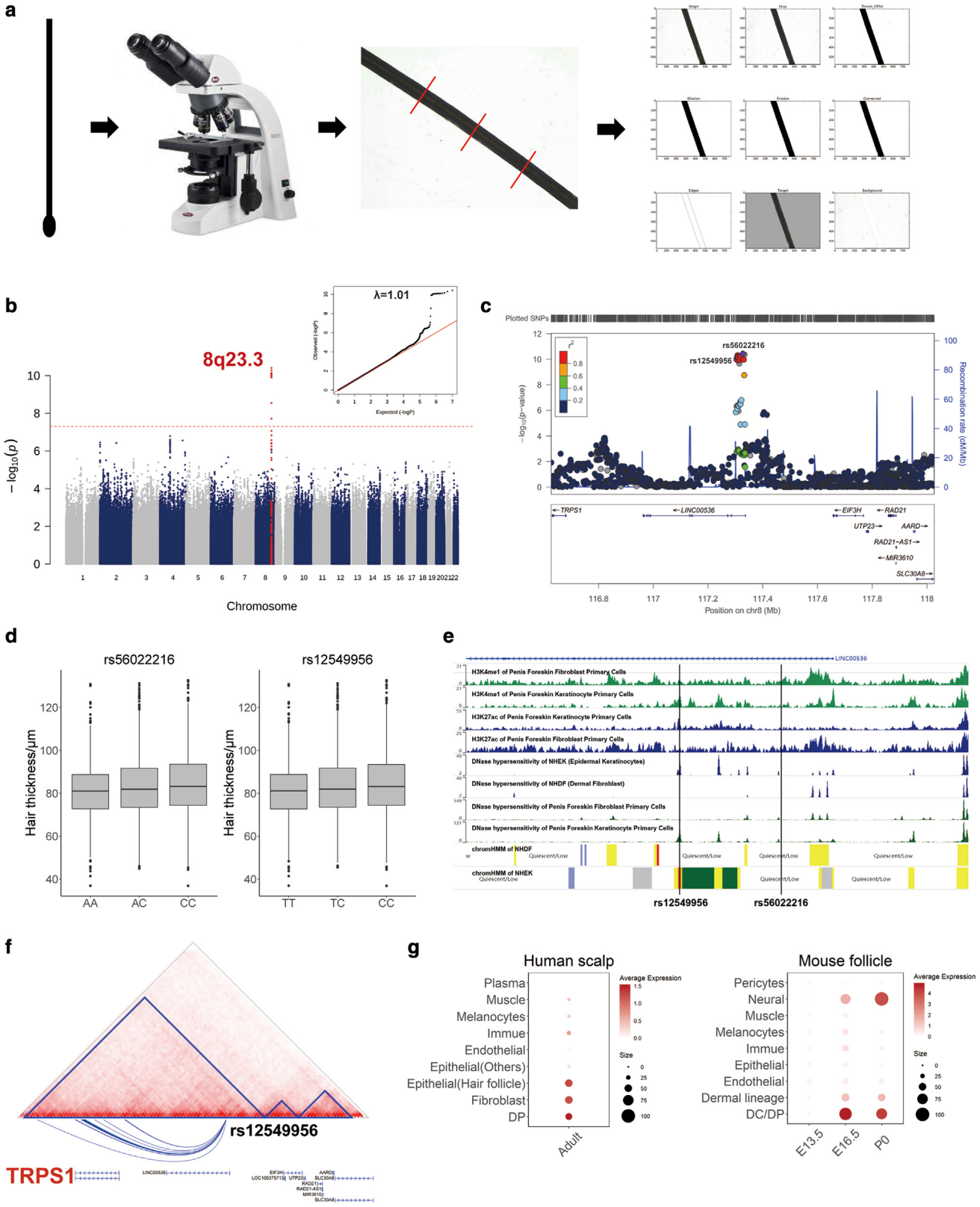


Figure 1. GWAS of hair thickness identified a significant signal at 8q23.3 in Chinese populations. (**a**) Schematic diagram of phenotyping process. (**b**) Manhattan plot and quantile−quantile plot of the meta-analysis result. The red line indicates the threshold for genome-wide significance ($P = 5 \times 10^{-8}$). (**c**) Region association plot for the significant region at 8q23.3. Increasing color intensities represented increasing linkage disequilibrium (r^2) with the lead SNP rs56022216. (**d**) Effect of derived alleles shows a contribution to hair thickness. (**e**) Visualization of epigenetic tracks H3K4me1, H3K27ac, and DNase hypersensitivity at the region around rs56022216 and rs12549956 annotated by Roadmap. (**f**) Visualization of chromatin interaction by Hi-C data on NHEKs. Blue connections represented significant 3D interactions between candidate region and *TRPS1*. (**g**) Expression of *TRPS1* in different cell types of human scalp skin and mouse dorsal skin. 3D, 3-dimensional; DC, dermal condensate; DP, dermal papilla; E13.5, embryonic day 13.5; E16.5, embryonic day 16.5; NHEK, normal human epidermal keratinocyte; P0, postnatal day 0.

KEYWORDS
GWAS; Hair thickness; Hair morphology; *TRPS1*

ORCIDs
Qili Qian: http://orcid.org/0009-0005-8797-5242
Sijie Wu: http://orcid.org/0000-0003-4087-4274
Junyu Luo: http://orcid.org/0000-0003-2206-1931
Yajun Yang: http://orcid.org/0000-0001-5713-0103
Li Jin: http://orcid.org/0000-0001-9201-2321
Sijia Wang: http://orcid.org/0000-0001-6961-7867

CONFLICT OF INTEREST
The authors state no conflict of interest.

ACKNOWLEDGMENTS
This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (grant number XDB38020400 to SW), National Natural Science Foundation of China (grant number 32325013), CAS Project for Young Scientists in Basic Research (grant number YSBR-077), Shanghai Science and Technology Commission Excellent Academic Leaders Program (22XD1424700), Science and Technology Innovation Team (Tianshan Innovation Team) (20221100619 to WZ), and the Human Phenome Data Center of Fudan University. Correspondences regarding the phenotype measurement should be addressed to WZ (zwx2020@126.com).

AUTHOR CONTRIBUTIONS
Conceptualization: SWa; Data Curation: QQ, SWu, JL, YG, WZ; Formal Analysis: QQ, SWu, JL; Funding Acquisition: SWa; Investigation: QQ, SWu, JL; Methodology: QQ, SWu; Project Administration: SWa; Resources: SWa, WZ, LJ, YY; Software: QQ, SWu; Supervision: SWa; Validation: QQ, SWu, SWa; Visualization: QQ, SWu, SWa; Writing - Original Draft Preparation: QQ, SWu, SWa; Writing - Review and Editing: QQ, SWu, JL, YY, LJ, SWa

***Qili Qian*[1,8], *Sijie Wu*[1,8], *Junyu Luo*[1,2], *Yaqun Guan*[3], *Yajun Yang*[4,5], *Li Jin*[5,6], *Wenxin Zheng*[7] and *Sijia Wang*[1,*]**

[1]*CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing, China;* [2]*Guangzhou National Laboratory, Guangzhou International Bio Island, Guangzhou, China;* [3]*Department of Biochemistry, Preclinical Medicine College, Xinjiang Medical University, Urumqi, China;* [4]*Fudan-Taizhou Institute of Health Sciences, Taizhou, China;* [5]*State Key Laboratory of Genetic Engineering and Ministry of Education, Key Laboratory of Contemporary Anthropology, Collaborative Innovation, Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, China;* [6]*Human Phenome Institute, Fudan University, Shanghai, China; and* [7]*Quality Standards, Institute of Animal Husbandry of Xinjiang Academy Animal Science (Xinjiang Breeding Sheep and Wool Cashmere Quality Safety Supervision and Inspection Center), Urumqi, China*

[8]*These authors contributed equally to this work.*

[*]*Corresponding author. e-mail: wangsijia@sinh.ac.cn*

SUPPLEMENTARY MATERIAL
Supplementary material is linked to the online version of the paper at www.jidonline.org, and at https://doi.org/10.1016/j.jid.2024.10.601.

REFERENCES

Fantauzzo KA, Christiano AM. Trps1 activates a network of secreted Wnt inhibitors and transcription factors crucial to vibrissa follicle morphogenesis. Development 2012;139:203–14.

Fujimoto A, Kimura R, Ohashi J, Omi K, Yuliwulandari R, Batubara L, et al. A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. Hum Mol Genet 2008;17:835–43.

Fujimoto A, Nishida N, Kimura R, Miyagawa T, Yuliwulandari R, Batubara L, et al. FGFR2 is associated with hair thickness in Asian populations. J Hum Genet 2009;54:461–5.

Ge W, Tan SJ, Wang SH, Li L, Sun XF, Shen W, et al. Single-cell transcriptome Profiling reveals Dermal and Epithelial cell fate decisions during embryonic Hair Follicle Development. Theranostics 2020;10:7581–98.

Liu F, Chen Y, Zhu G, Hysi PG, Wu S, Adhikari K, et al. Meta-analysis of genome-wide association studies identifies 8 novel loci involved in shape variation of human head hair. Hum Mol Genet 2018;27:559–75.

Medland SE, Nyholt DR, Painter JN, McEvoy BP, McRae AF, Zhu G, et al. Common variants in the trichohyalin gene are associated with straight hair in Europeans. Am J Hum Genet 2009;85:750–5.

Momeni P, Glöckner G, Schmidt O, von Holtum D, Albrecht B, Gillessen-Kaesbach G, et al. Mutations in a new gene, encoding a zinc-finger protein, cause tricho-rhino-phalangeal syndrome type I. Nat Genet 2000;24:71–4.

Ober-Reynolds B, Wang C, Ko JM, Rios EJ, Aasi SZ, Davis MM, et al. Integrated single-cell chromatin and transcriptomic analyses of human scalp identify gene-regulatory programs and critical cell types for hair and skin diseases. Nat Genet 2023;55:1288–300.

Wu S, Tan J, Yang Y, Peng Q, Zhang M, Li J, et al. Genome-wide scans reveal variants at EDAR predominantly affecting hair straightness in Han Chinese and Uyghur populations. Hum Genet 2016;135:1279–86.

Zhang Y, Nakamura T, Furukawa F, Muragaki Y. Trps1-deficient transplanted skin gave rise to a substantial amount of hair: Trps1 is unnecessary for hair development. Dermatol Reports 2019;11:7853.

DDI-2: A Diverse Skin Condition Image Dataset Representing Self-Identified Asian Patients

Journal of Investigative Dermatology (2025) **145,** 1205–1208; doi:10.1016/j.jid.2024.09.018

TO THE EDITOR

Artificial intelligence (AI) holds promise for improving access to dermatological care and augmenting provider workflows (Gui et al, 2024). Several skin cancer diagnostic algorithms have received the Conformité Européenne mark, and the United States Food and Drug Administration recently authorized DermaSensor, an AI-enabled skin cancer device-based diagnostic (Gui et al, 2024) using elastic scattering spectroscopy (Chang and Daneshjou, 2024; Venkatesh et al, 2024). However, algorithm performance is often inflated by evaluation under idealized conditions (eg, on high-resolution images taken by specialized cameras, lack of artifacts such as surgical marker or secondary morphology changes) and is frequently significantly reduced when evaluated on datasets reflecting more real-world conditions (Combalia et al, 2022; Daneshjou et al, 2022). Furthermore, medical algorithms are often trained on private, siloed, sparsely annotated datasets, leading to one-off publications of limited generalizability (Daneshjou et al, 2021). Leading public

Abbreviations: AI, artificial intelligence; DDI, Diverse Dermatology Images; FST, Fitzpatrick skin type

Accepted manuscript published online 29 October 2024; corrected proof published online 22 November 2024

SUPPLRMENTARY MATERIALS AND METHODS

Populations and samples

The discovery set included 2961 healthy Chinese (1060 males and 1901 females, aged 31−87 years) sampled before 2014 in the Taizhou longitudinal (TZL) cohort. The TZL cohort is a long-term observational cohort study to explore the environmental and genetic risk factors for common and noncommunicable diseases (Wang et al, 2009), approved by the Ethics Committee of Fudan University (number 85) (Shanghai, China). All participants provided written informed consent.

The replication set was recruited from Xinjiang Medical University (Uyghur [UYG] cohort). This research program was conducted with the approval of Shanghai Institutes for Biological Sciences (Shanghai, China) (ER-SIBS-261410). Written, informed consent was obtained from all participants. Our replication set included 721 individuals (282 males and 439 females, aged 17−25 years), which were collected in 2013 and 2014 (Wu et al, 2018; Xiong et al, 2019).

Phenotyping

We utilized 2 methods to measure hair thickness: manual microscope measurement and automated quantization strategy.

Manual microscope measurement was conducted at Institute of Animal Husbandry of Xinjiang Academy Animal Science. The measurement process involved a professional operator selecting a hair segment of approximately 5 cm near the hair follicle on the vertex region for each person, cutting the segment into small pieces of around 5 mm with scissors, and placing them on a slide. These hair segments were then measured under a ×4 microscope. During measurement, the operator randomly selected 3 locations to measure the diameter and took the average value to determine the person's hair thickness. After measurement, the manually measured thickness data and microscope images were saved.

To obtain phenotypes faster in more images, we further developed an automated method for measuring hair thickness on the basis of microscopic images. The specific process is as follows:

- Convert the hair microscope images to gray-scale images (Supplementary Figure S1b).
- Binary threshold the gray-scale images. By observing the gray-scale distribution of the images, we found that the gray-scale values mainly have 2 peaks: the first peak has low gray scale (black) and consists of the hair part, whereas the second peak is skewed toward white and consists of the background. Therefore, we decided to use a threshold value to distinguish these 2 peaks and separate the hair part from the background. Because some images may have an overall gray-scale shift (caused by external light), a fixed threshold may not handle all images well. We decided to use the Otsu's binarization method, which assumes that the chromaticity of the image is mainly composed of 2 peaks and then selects a middle value to distinguish these 2 peaks as much as possible, thus separating the hair part from the background (Supplementary Figure S1c).
- Remove background noise. After binarizing the image, most of the hair can be distinguished from the background, but there are some background noises in the image: (i) the edges of the air bubbles covering the glass slide and stains are also classified as hair owing to their deep grayness. (ii) Some hair medullas are classified as white background owing to their light color. To deal with this, we use the cyclic dilation−erosion method to remove these 2 types of noise. First, we used dilation to blacken the white medulla wrapped by black hair (Supplementary Figure S1d). Then, we used erosion to whiten the black impurities floating in the white background (Supplementary Figure S1e). After iterating this cycle a few times, the hair part can be captured, and the noise can be removed. It should be noted that the value of dilation and erosion should be kept consistent so that the hair part remains the same size after dilation and erosion processing.
- Capture hair. Because some images may contain other parts of hair fragments in addition to the main target hair segment in the field of view as well as extra-large stains (difficult to remove with dilation−erosion), we used the method of connected components to detect all black blocks in the image and then only retained the largest connected component (Supplementary Figure S1f).
- Extract hair boundaries. We first used Canny to extract the boundary lines of the hair part. Some boundary lines are the 2 parallel edges of the hair, which can be conveniently used to extract the thickness (Supplementary Figure S1g).
- Obtain the uncorrected hair thickness value. We randomly selected 13 positions on boundary 1 and boundary 2 of the hair and then drew a perpendicular line on the other boundary to calculate the distance between the point and the line. Finally, the average value of the 13 thickness segments is calculated, and the SD is obtained (Supplementary Figure S1j).
- Phenotype correction. Owing to the lack of dimensional units in the automatically measured hair thickness values, we randomly sampled a subset of individuals and manually measured their hair thickness in micrometers. By performing linear regression between the manual measurements and the automatic readings, we obtained a correction coefficient (1.45 in both the TZL and UYG cohorts). Multiplying the automatically measured values by this correction coefficient allows us to convert the phenotypic values to units of micrometers, which provides a more accurate representation of the actual hair thickness.

Unlike previous studies (Fujimoto et al, 2009, 2008), we derived the hair thickness phenotype from planar images rather than cross-sectional images. Because the hair cross-section is elliptical rather than perfectly circular, our measurements primarily reflect the long diameter of the hair shaft. Given that East Asians have the roundest hair cross-sections (with an ellipticity of 82%, compared with 76% in Europeans and 57% in Africans) (Franbourg et al, 2003), we consider this method to be both appropriate and reliable. This approach is particularly well-suited for our East Asian cohort, ensuring the accuracy and relevance of our findings.

Genotyping

For both the TZL and UYG cohorts, DNA was extracted from blood samples using the Illumina HumanOmniZhongHua-8 chip (Illumina), which interrogates 894,517 SNPs. We used PLINK, version 1.9 (Purcell et al, 2007), to exclude individuals with >5% missing data, discordant sex information, or discordant ancestral information. We also discarded SNPs with >2% missing data, those with <1% minor allele frequency, or any SNPs that failed the Hardy–Weinberg deviation test ($P < 1 \times 10^{-5}$). We then used SHAPEIT (Williams et al, 2012) to phase the genotype data. Next, we used IMPUTE2 (Howie et al, 2009) to impute genotypes with the 1000 Genomes Project phase 3 data as the imputation reference panel. In the TZL cohort, we finally got 6,343,243 imputed SNPs with >0.8 imputation quality score and 776,213 unimputed SNPs. In the UYG cohort, we also got 6,414,304 imputed SNPs and 810,648 unimputed SNPs.

Statistical analyses

Population stratification analysis. The TZL cohort belongs to the East Asian population, whereas the UYG cohort is a Eurasian admixed population. To avoid interference from obvious population structure differences in our dataset, we first used principal component analysis with EIGENSTRAT (Price et al, 2006) to study the population structure differences between the East Asian Han and the Eurasian admixed populations. We selected 97 CHB (East Asian Han), 86 CEU (West European Caucasians), and 88 YRI (African Yoruba) individuals from the 1000 Genomes phase 3 dataset (Sudmant et al, 2015) and combined these individuals with our discovery and replication populations. We then selected 102,284 common autosomal SNPs with weak linkage disequilibrium (r2 < 0.2) to perform principal component analysis. We found that the TZL population clustered well with the East Asian population, and the UYG population, as a Eurasian admixed population, was located between the Eurasian populations and did not show any obvious outliers (Supplementary Figure S4). In addition, we performed the same principal component analysis separately in the TZL and UYG populations, and we chose the top 4 principal components as relevant covariates in subsequent statistical analyses.

Association test. GWASs were separately conducted in both the TZL and UYG cohorts on hair thickness using PLINK, version 1.9, where additive allele effects were tested in linear models adjusted for covariates (age, sex, and the top 4 genomic principal components). Meta-analysis was conducted to combine these 2 GWASs using the fixed-effects model in METAL (Willer et al, 2010). Genetic inflation factor lambda and Q–Q (quantile–quantile) plots were used to evaluate whether population stratification had affected our association analysis. Owing to the required sample size, the heritability of hair thickness was estimated only in the discovery set in GCTA (genome-wide complex trait analysis) (Yang et al, 2011). All plots were generated using R software.

Functional annotation. After identifying multiple related genetic loci using GWAS, it is common to find high linkage disequilibrium between these candidate SNPs. To find the most likely disease-causing mutation sites, we performed fine mapping using functional prediction methods. Specifically, we used 6 functional annotation databases from SNPnexus (Oscanoa et al, 2020), including Combined Annotation Dependent Depletion (Kircher et al, 2014), DeepSEA (Zhou and Troyanskaya, 2015), EIGEN (Ionita-Laza et al, 2016), FunSeq2 (Fu et al, 2014), GWAVA (Ritchie et al, 2014), and REMM (Smedley et al, 2016). These methods mainly rely on existing manually annotated mutation–disease data as training data and use machine learning and deep learning methods to construct models that predict the probability of other mutation sites being pathogenic or having biological functional effects.

We used existing epigenetic databases such as RoadMap (Roadmap Epigenomics Consortium et al, 2015) and ENCODE (ENCODE Project Consortium et al, 2007) to annotate the significant locus region in greater detail and searched for obvious epigenetic modifications in the candidate site region to determine its potential function. For example, H3K4me1, H3K27ac, and DNAse modifications often suggest that the region is a potential enhancer, whereas H3K4me3 modification indicates the promoter region. We used UCSC genome browser on Human February 2009 (GRCh37/hg19) Assembly (Kent et al, 2002) to visualize these annotations.

To further explore the functional genes regulated by the candidate region, we visualized the 3-dimensional chromatin structure around rs56022216 in 3DIV (Yang et al, 2018).

Tests for natural selection. Natural selection often leads to reduced polymorphism and extensive linkage disequilibrium in selected genomic regions. We used the rehh package (Gautier and Vitalis, 2012) to calculate integrated haplotype score (Liu et al, 2013) in CEU, CHB + JPT, and YRI populations. We selected the top 0.1% of integrated haplotype score values across the genome as a threshold to identify regions under selection. To further improve statistical power, we applied the Composite of Multiple Signals (Grossman et al, 2010) method to test the likelihood of significant signals being under natural selection.

SUPPLEMENTARY REFERENCES

ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 2007;447:799–816.

Franbourg A, Hallegot P, Baltenneck F, Toutain C, Leroy F. Current research on ethnic hair. J Am Acad Dermatol 2003;48:S115–9.

Fu Y, Liu Z, Lou S, Bedford J, Mu XJ, Yip KY, et al. FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. Genome Biol 2014;15:480.

Fujimoto A, Kimura R, Ohashi J, Omi K, Yuliwulandari R, Batubara L, et al. A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. Hum Mol Genet 2008;17:835–43.

Fujimoto A, Nishida N, Kimura R, Miyagawa T, Yuliwulandari R, Batubara L, et al. FGFR2 is associated with hair thickness in Asian populations. J Hum Genet 2009;54:461–5.

Gautier M, Vitalis R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics 2012;28:1176–7.

Grossman SR, Shlyakhter I, Karlsson EK, Byrne EH, Morales S, Frieden G, et al. A composite of multiple signals distinguishes causal variants in regions of positive selection. Science 2010;327:883–6.

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:e1000529.

Ionita-Laza I, McCallum K, Xu B, Buxbaum JD. A spectral approach integrating functional genomic annotations for coding and noncoding variants. Nat Genet 2016;48:214–20.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res 2002;12:996–1006.

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46:310–5.

Liu X, Ong RT, Pillai EN, Elzein AM, Small KS, Clark TG, et al. Detecting and characterizing genomic signatures of positive selection in global populations. Am J Hum Genet 2013;92:866–81.

Marcus JH, Novembre J. Visualizing the geography of genetic variants. Bioinformatics (Oxford, England) 2017;33:594–5.

Oscanoa J, Sivapalan L, Gadaleta E, Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). Nucleic Acids Res 2020;48:W185–92.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–9.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75.

Ritchie GR, Dunham I, Zeggini E, Flicek P. Functional annotation of noncoding sequence variants. Nat Methods 2014;11:294–6.

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015;518:317–30.

Smedley D, Schubach M, Jacobsen JOB, Köhler S, Zemojtel T, Spielmann M, et al. A whole-genome analysis framework for effective identification of pathogenic regulatory variants in Mendelian disease. Am J Hum Genet 2016;99:595–606.

Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. Nature 2015;526:75–81.

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;26:2190–1.

Williams AL, Patterson N, Glessner J, Hakonarson H, Reich D. Phasing of many thousands of genotyped samples. Am J Hum Genet 2012;91:238–51.

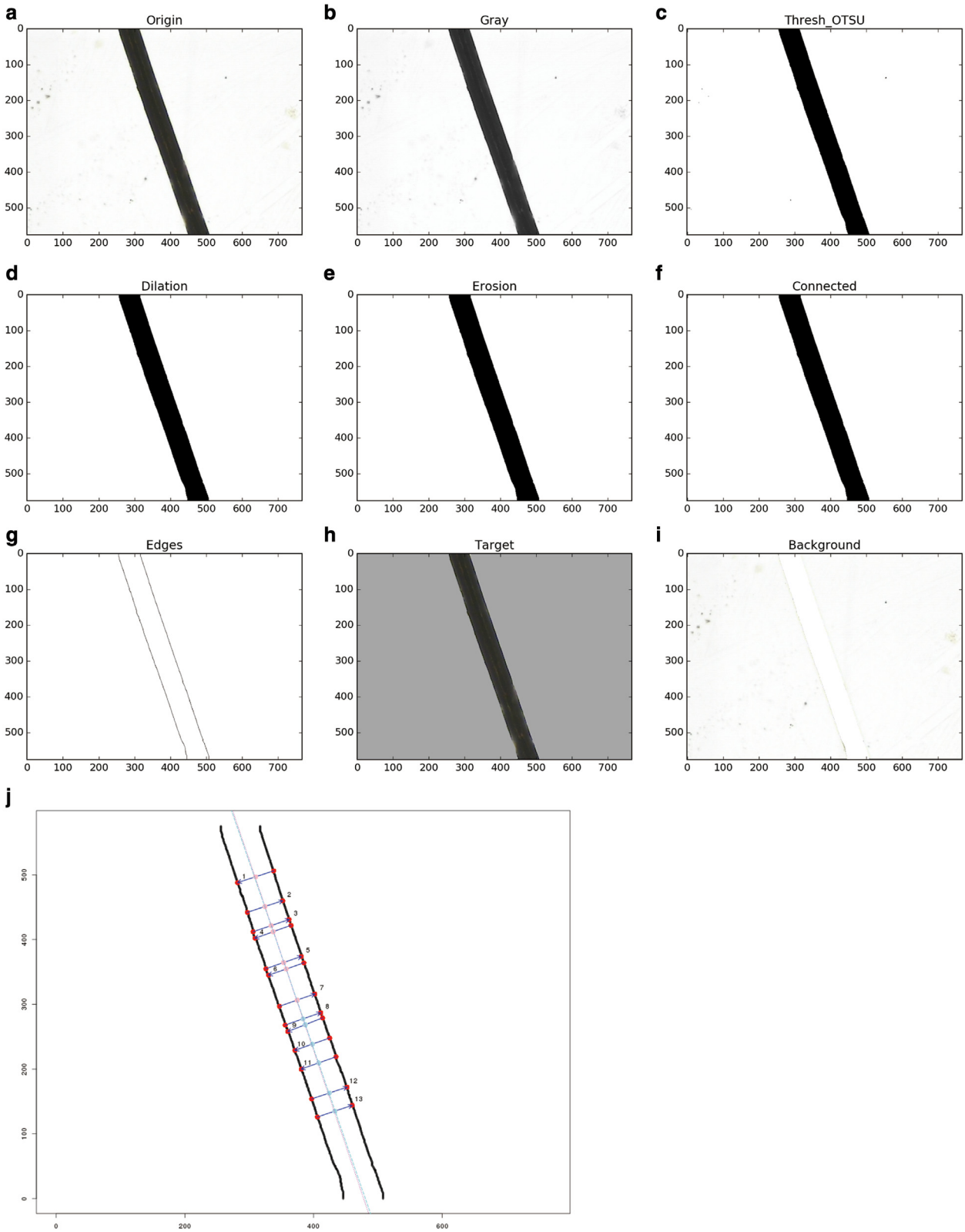
Wu S, Zhang M, Yang X, Peng F, Zhang J, Tan J, et al. Genome-wide association studies and CRISPR/Cas9-mediated gene editing identify regulatory variants influencing eyebrow thickness in humans. PLoS Genet 2018;14:e1007640.

Xiong Z, Dankova G, Howe LJ, Lee MK, Hysi PG, de Jong MA, et al. Novel genetic loci affecting facial shape variation in humans. eLife 2019;8:e49898.

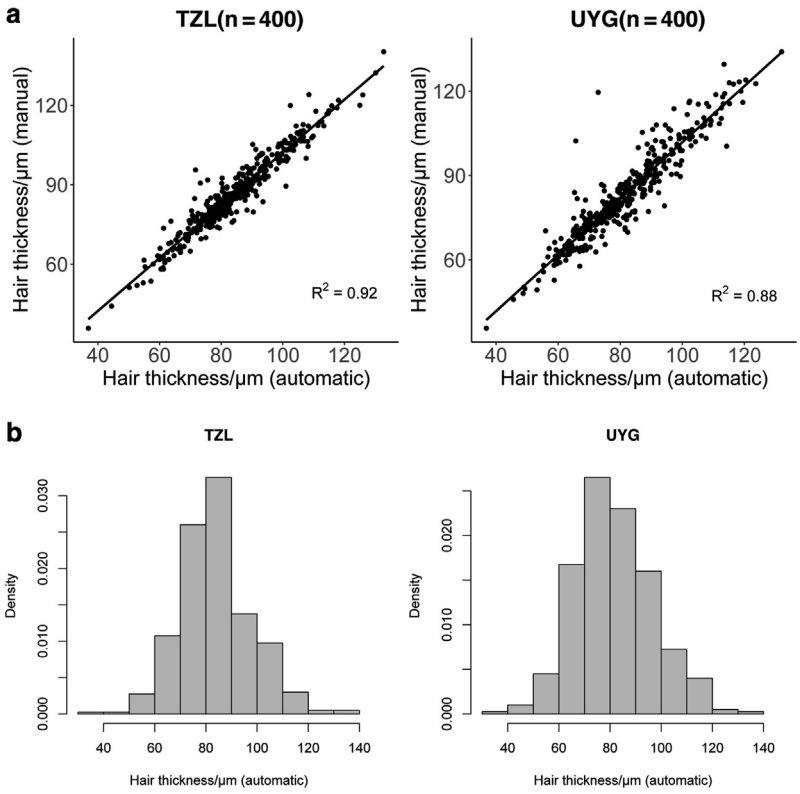
Yang D, Jang I, Choi J, Kim MS, Lee AJ, Kim H, et al. 3DIV: A 3D-genome Interaction Viewer and database. Nucleic Acids Res 2018;46:D52–7.

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

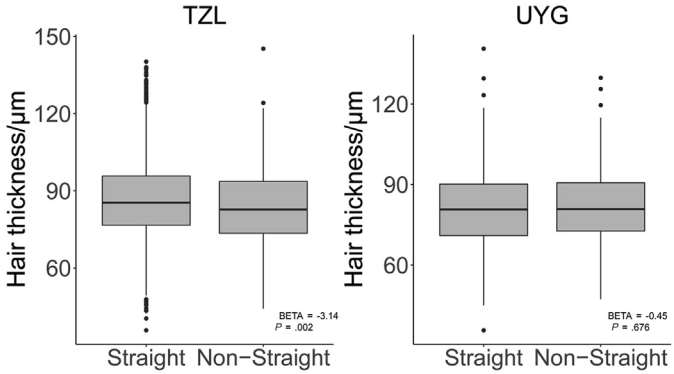
Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model. Nat Methods 2015;12:931–4.



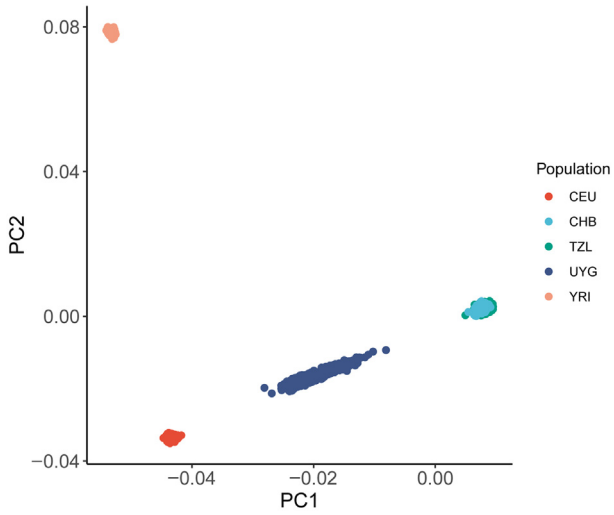
Supplementary Figure S1. Schematic diagram of automated quantization strategy. (**a**) Original hair microscopic image. (**b**) Gray-scale images. (**c**) Gray-scale images after Otsu's binarization. (**d, e**) Gray-scale images after cyclic dilation−erosion. (**f**) The largest connected hair component retained. (**g**) Extracted hair boundary after Canny. (**h, i**) Extracted hair object and the remaining background noise. (**j**) An example of hair thickness calculation; the red dots represent the locations of random sampling.



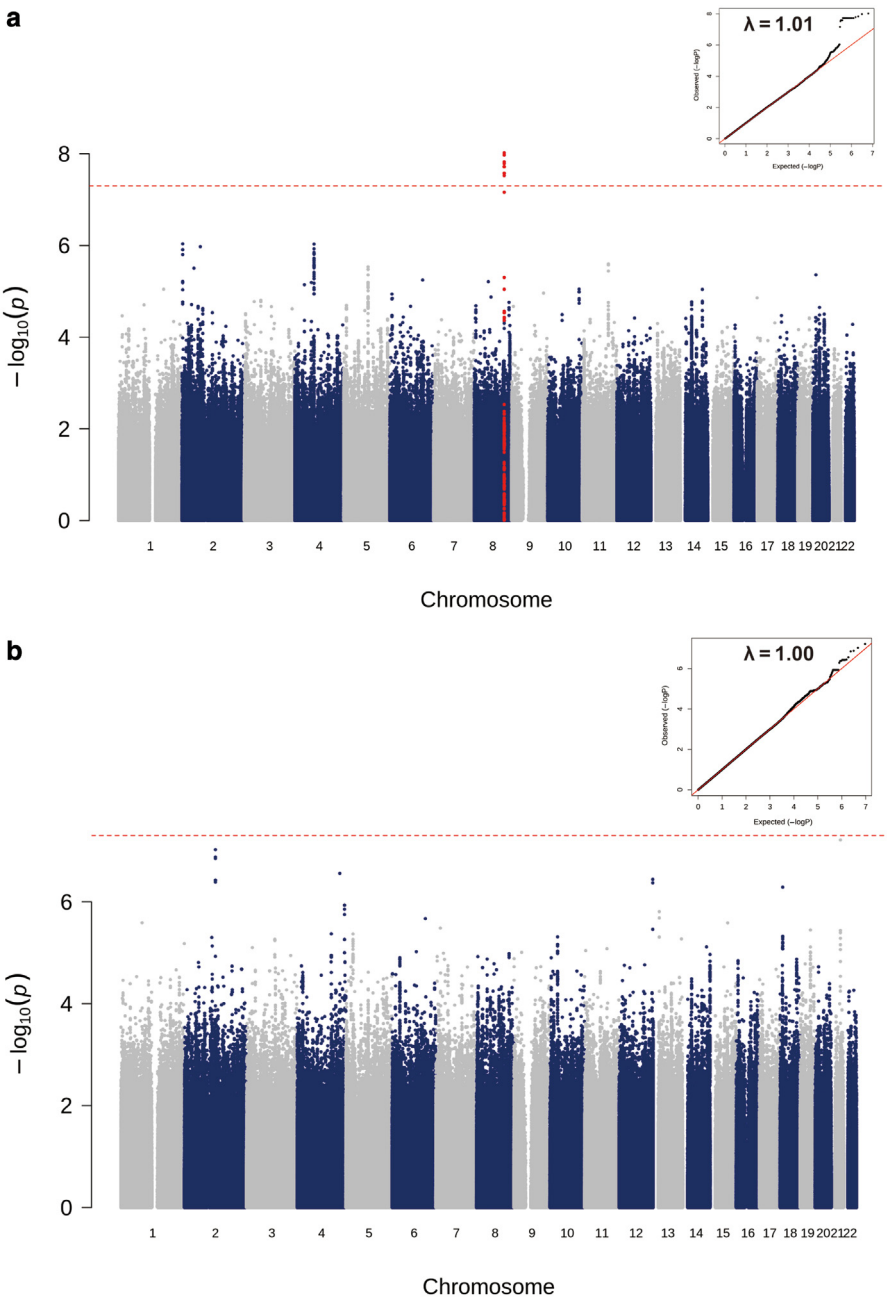
Supplementary Figure S2. Distribution of hair thickness in both the TZL and UYG cohorts. (**a**) Scatter plot of manually measured hair thickness and automatically measured hair thickness in both 2 cohorts. A total of 400 individuals were randomly selected from TZL and UYG cohorts, respectively. (**b**) Histogram of the automatically measured hair thickness in both the TZL and UYG cohorts. TZL, Taizhou longitudinal; UYG, Uyghur.



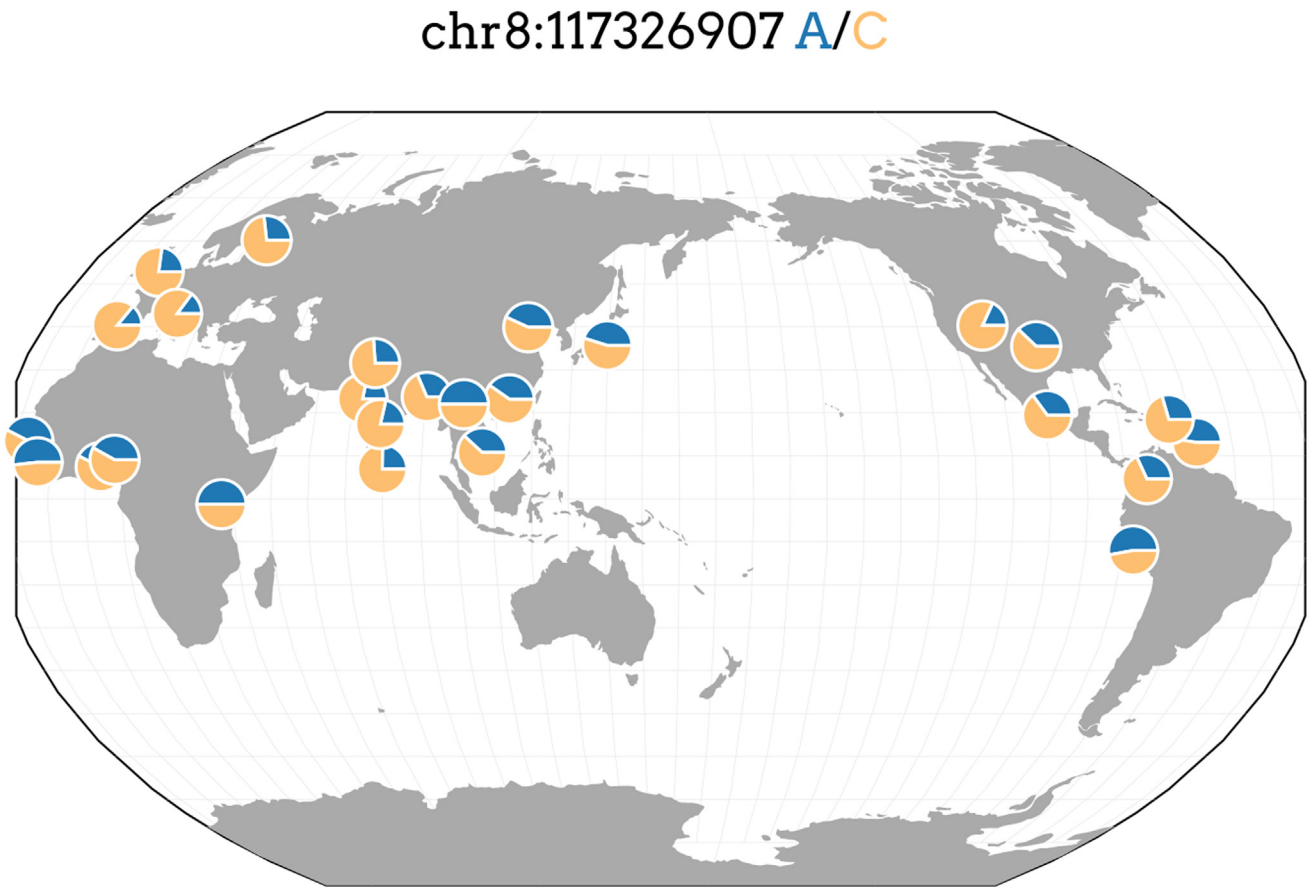
Supplementary Figure S3. Phenotypic correlation between hair thickness and hair curliness in both the TZL and UYG cohorts. TZL, Taizhou longitudinal; UYG, Uyghur.



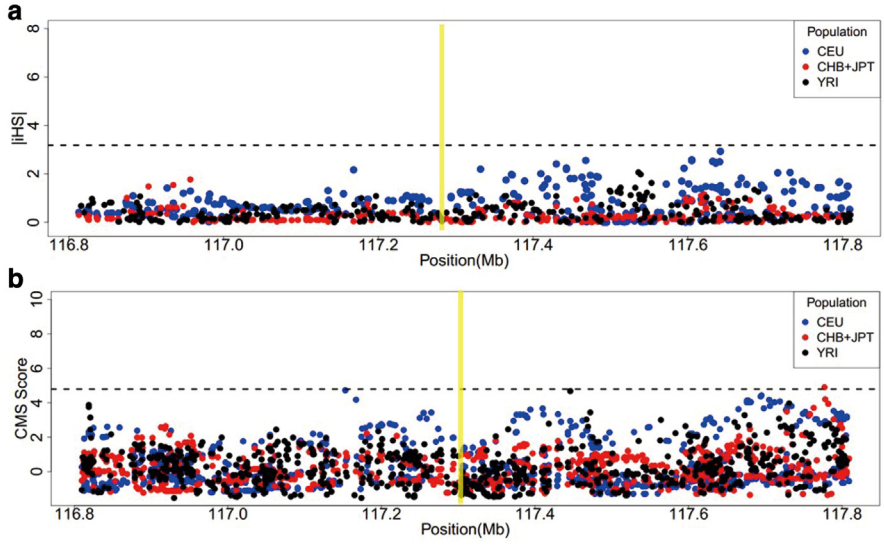
Supplementary Figure S4. Structural analysis of the samples from the TZL and UYG cohorts, combined with representative populations from the 1000 Genomes Project. YRI denotes Yoruba in Ibadan, Nigeria; CEU denotes Utah residents with Northern and Western European ancestry from the CEPH collection; and CHB denotes Han Chinese in Beijing, China. PC, principal component; TZL, Taizhou longitudinal; UYG, Uyghur.



Supplementary Figure S5. Manhattan plot and quantile–quantile plot showing the results of the genome-wide scan for hair thickness in TZL and UYG. **(a)** the GWAS in TZL adjusted for age, sex, and genetic PCs. **(b)** The GWAS in UYG adjusted for age, sex, and genetic PCs. PC, principal component; TZL, Taizhou longitudinal; UYG, Uyghur.



Supplementary Figure S6. The geography of allele frequency on the rs56022216. The effect alleles were marked blue. Allele frequencies were obtained from 1000 Genome data and visualized by the Geography of Genetic Variants Browser (Marcus and Novembre, 2017).



Supplementary Figure S7. Results of the tests for natural selection on chromosome 8q23. (**a**) Absolute value of iHSs for region 8q23 in CEU, CHB + JPT, and YRI populations. (**b**) CMS scores in the same region. CHB denotes Han Chinese in Beijing, China; JPT denotes Japanese in Tokyo, Japan; and YRI denotes Yoruba in Ibadan, Nigeria. CMS, composite of multiple signals; iHS, integrated haplotype score.

Supplementary Table S1. Heritability of Hair Thickness in TZL

Cohort	h^2	SE	*P*-Value
TZL	0.170	0.081	.025

Abbreviations: SE, standard error; TZL, Taizhou longitudinal; UYG, Uyghur.
Heritability was not calculated in UYG cohort owing to the small sample size.

Supplementary Table S2. Summary of Previously Reported SNPs Associated with Hair Curliness

					TZL (Hair Thickness)			UYG (Hair Thickness)		
SNP	Gene	CHR	BP	EA	EAF	Beta	*P*-Value	EAF	Beta	*P*-Value
rs80293268	*ERRFI1*	1p36.23	8207579	C		—		0.01	−0.50	.88
rs6658216	*PEX14*	1p36.22	10561604	C	0.39	−0.28	.46	0.33	0.15	.86
rs11203346	*PADI3*	1p36.13	17600822	G		—		0.05	−0.32	.86
rs17646946	*TCHHL1*	1q21.3	152062767	A				0.13	−0.97	.39
rs12997742	*TGFA*	2p13.3	70786598	C	0.37	−0.69	.07	0.36	0.54	.50
rs74333950	*WNT10A*	2q35	219746292	G	0.27	0.67	.10	0.19	−0.60	.53
rs506863	*FRAS1*	4q21.21	79255688	C	0.80	−0.29	.52	0.63	1.43	.07
rs1999874	*GATA3*	10p14	8353101	A	0.19	−0.30	.52	0.36	0.75	.37
rs2219783	*LGR4*	11p14.1	27411298	G	0.06	−0.62	.42	0.04	2.15	.28
rs11170678	*HOXC13*	12p13.13	54154174	G		—		0.11	−1.42	.27
rs11078976	*KRTAP*	17q21.2	39189360	T	0.73	0.07	.86	0.81	0.20	.84
rs310642	*PTK6*	20q13.33	62161998	C	0.06	0.26	.73	0.05	−1.86	.27

Abbreviations: BP, base pair; CHR, chromosome; EA, effect allele; EAF, effect allele frequency; TZL, Taizhou longitudinal; UYG, Uyghur.
EA is related to straight hair. GWAS of hair thickness here did not correct the hair curliness.

Supplementary Table S3. Prediction Functional Scores of the 21 SNPs in 8q23.3

SNP	CADD	DeepSEA	EIGEN	FunSeq2	GWAVA	REMM
rs12549956[1]	9.15[2]	7.05[2]	0.45[2]	0.81[2]	0.36	0.84[2]
rs4425786[1]	5.60	3.03	0.09	0	0.24	0.03
rs4876641[1]	2.24	4.69	0.09	0.81[2]	0.19	0.57
rs16888292[1]	2.70	2.61	−0.15	0	0.42[2]	0.01
rs12550516[1]	2.62	—	0.23	0.16	0.20	0.09
rs16888277[1]	0.94	3.59	−0.33	0.19	0.35	0.05
rs4876640[1]	0.83	3.44	−0.43	0.19	0.21	0.002
rs1918717[1]	1.14	2.61	−0.44	0.19	0.27	0
rs58810439	7.55	2.98	0.20	0	0.31	0.31
rs16888282	5.39	4.33	−0.01	0	0.31	0.003
rs1918719	4.47	3.00	0.15	0.62	0.29	0.53
rs1405982	1.06	4.41	0.14	0.62	0.31	0.01
rs10092355	3.96	2.17	−0.30	0	0.23	0.12
rs12547047	2.75	—	−0.27	0	0.21	0.001
rs1405981	0.85	4.07	0.07	0.62	0.19	0.003
rs140854377	3.10	2.20	−0.90	0.19	0.14	0.004
rs12542908	2.06	2.50	−0.33	0	0.23	0.01
rs10955771	1.08	1.74	−0.32	0	0.3	0.10
rs6992138	1.03	2.07	−0.60	0	0.28	0
rs140035482	0.70	1.90	−0.34	0	0.19	0.02
rs56022216	0.61	1.65	−1.10	0	0.34	0

Abbreviation: CADD, Combined Annotation Dependent Depletion.
These 21 SNPs displayed genome-wide significant associations with hair thickness in meta-analysis.
[1]SNPs are located on the enhancer regions of skin-related cell types annotated by Roadmap.
[2]The highest functional score in each functional annotation integration database.