Hum Genet (2016) 135:1279–1286
DOI 10.1007/s00439-016-1718-y

ORIGINAL INVESTIGATION

Genome-wide scans reveal variants at *EDAR* predominantly affecting hair straightness in Han Chinese and Uyghur populations

Sijie Wu[1] · Jingze Tan[2] · Yajun Yang[2,3] · Qianqian Peng[1] · Manfei Zhang[1,2] · Jinxi Li[1] · Dongsheng Lu[1] · Yu Liu[1] · Haiyi Lou[1] · Qidi Feng[1] · Yan Lu[1] · Yaqun Guan[4] · Zhaoxia Zhang[4] · Yi Jiao[4] · Pardis Sabeti[5,6] · Jean Krutmann[7] · Kun Tang[1,2] · Li Jin[1,2] · Shuhua Xu[1,2,8] · Sijia Wang[1,2]

Received: 12 June 2016 / Accepted: 23 July 2016 / Published online: 3 August 2016

Abstract Hair straightness/curliness is one of the most conspicuous features of human variation and is particularly diverse among populations. A recent genome-wide scan found common variants in the Trichohyalin (*TCHH*) gene that are associated with hair straightness in Europeans, but different genes might affect this phenotype in other populations. By sampling 2899 Han Chinese, we performed the first genome-wide scan of hair straightness in East Asians, and found *EDAR* (rs3827760) as the predominant gene ($P = 4.67 \times 10^{-16}$), accounting for 3.66 % of the total variance. The candidate gene approach did not find further significant associations, suggesting that hair straightness may be affected by a large number of genes with subtle effects. Notably, genetic variants associated with hair straightness in Europeans are generally low in frequency in Han Chinese, and vice versa. To evaluate the relative contribution of these variants, we performed a second genome-wide scan in 709 samples from the Uyghur, an admixed population with both eastern and western Eurasian ancestries. In Uyghurs, both *EDAR* (rs3827760: $P = 1.92 \times 10^{-12}$) and *TCHH* (rs11803731: $P = 1.46 \times 10^{-3}$) are associated with hair straightness, but *EDAR* (OR 0.415) has a greater effect than *TCHH* (OR 0.575). We found no significant interaction between *EDAR* and *TCHH* ($P = 0.645$), suggesting that these two genes affect hair straightness through different mechanisms. Furthermore, haplotype analysis indicates that *TCHH* is not subject to selection. While *EDAR* is under strong selection in East Asia, it does not appear to be subject to selection after the admixture in Uyghurs. These suggest that hair straightness is unlikely a trait under selection.

S. Wu, J. Tan, and Y. Yang contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-016-1718-y) contains supplementary material, which is available to authorized users.

✉ Li Jin
lijin.fudan@gmail.com

✉ Shuhua Xu
xushua@picb.ac.cn

✉ Sijia Wang
wangsijia@picb.ac.cn

[1] Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, University of Chinese Academy of Sciences, 320 YueYang Road, Shanghai 200031, China

[2] 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 200438, China

[3] Fudan-Taizhou Institute of Health Sciences, Taizhou, Jiangsu 225300, China

[4] Department of Biochemistry, Preclinical Medicine College, Xinjiang Medical University, Urumqi 830011, China

[5] Department of Organismic and Evolutionary Biology, Center for Systems Biology, Harvard University, Cambridge, MA 02138, USA

[6] The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[7] IUF-Leibniz Research Institute for Environmental Medicine, Dusseldorf, Germany

[8] School of Life Science and Technology, ShanghaiTech University, Shanghai 200031, China

Introduction

Hair straightness/curliness is one of the most conspicuous features of human variation. It is particularly diverse within populations and between individuals (Loussouarn et al. 2007), and shows high heritability (Medland et al. 2009b). Recent genome-wide scans in Caucasians found that common variants in the Trichohyalin (*TCHH*) gene are associated with hair straightness in Europeans (Eriksson et al. 2010; Medland et al. 2009a). Another genome-wide scan in admixed Latin Americans identified variants in *EDAR*, *TCHH*, *GATA3,* and *PRSS53* associated with hair straightness (Adhikari et al. 2016). *TCHH* affects hair straightness in both populations while *EDAR*, *GATA3,* and *PRSS53* are specific to Latin Americans. These findings suggest substantial genetic heterogeneity between diverse populations and it is possible that hair straightness is affected by different mechanisms in other populations, such as Africans and Asians. Therefore, it is of significant value to explore the genetics of hair straightness in populations other than Europeans and Latin Americans. Candidate gene approaches found a variant in the *EDAR* gene to be associated with hair straightness in East Asians, but no genome-wide scan of hair straightness has been performed in East Asians. A number of interesting questions remain to be answered: Are the genes found in other populations also associated with hair straightness in East Asians? Is there any gene specific to East Asians' straight hair? What is the relative contribution of these genes? Do these genes interact with each other while affecting hair straightness? To answer these questions, here we performed the first genome-wide scan of hair straightness in East Asians (Han Chinese).

Since Europeans and Han Chinese may share few or no genetic variants associated with hair straightness, genome-wide scans in Han Chinese or Europeans may not have enough power to evaluate the relative contribution of variants. We therefore performed a second genome-wide scan in Uyghurs, an admixed population with both East Asian and European ancestries. Genetic variants that are nearly fixed in the ancestral populations could remain at a moderate frequency in the admixed population, giving us an opportunity to evaluate the relative contribution of these genes on hair straightness, as well as the interaction between them.

Materials and methods

Populations and samples

The Han Chinese samples were collected in Taizhou, Jiangsu Province in 2014, as part of the Taizhou Longitudinal Study (Wang et al. 2009). In total, 2899 individuals (including 1038 males and 1861 females, with an age range of 31–87) were enrolled. We completed the study under the approved protocols from the Institutional Research Board at Fudan University. The Uyghur samples were collected at Xinjiang Medical University in 2013–2014. In total, 709 individuals (including 276 males and 433 females, with an age range of 17–25) were enrolled. The research was conducted with the official approval from the Ethics Committee of the Shanghai Institutes for Biological Sciences, Shanghai, China. All participants had provided written consent. In both studies, we collected blood samples, from which DNA was extracted.

Phenotype measures

We rated hair straightness by eye on a three-point scale (straight, wavy, and curly), following the established standard (Medland et al. 2009a; Tan et al. 2013). In each case, we confirmed that the volunteers had never had their hair treated. Two investigators independently rated the hair straightness for each case. If the rating was inconsistent, a third investigator would re-evaluate the phenotyping and make the final decision. Since the frequency of curly hair was very low in our samples (11 cases in total from the 2 populations), we merged wavy and curly hair into one category "non-straight" in further analyses (Table S1).

Genotype quality control and imputation

All samples were genotyped using the Illumina HumanOmniZhongHua-8 chips, which interrogates 894,517 SNPs. We excluded individuals with more than 5 % missing data, related individuals, and the ones that failed the X-chromosome sex concordance check or had ethnic information incompatible with their genetic information. We also excluded SNPs with more than 2 % missing data, with a minor allele frequency smaller than 1 %, and the ones that failed the Hardy–Weinberg deviation test ($P < 1\times10^{-5}$). After applying these filters, we obtained a dataset of 2899 samples with 776,213 SNPs for the Han Chinese, and 709 samples with 810,648 SNPs for the Uyghurs.

The chip genotype data were firstly phased using SHAPEIT (O'Connell et al. 2014). IMPUTE2 (Howie et al. 2009) was then used to impute genotypes at ungenotyped SNPs using the 1000 Genomes Phase 3 data as reference, which included haplotype information for 2504 individuals across the world for 81,706,044 variant positions. SNPs with an impute info score (imputation certainty) of <80 %, MAF <1 % or a missing rate of >2 % of genotypes were eliminated from further analyses. Finally, for the Uyghur sample, a total of 6,414,304 imputed SNPs passed quality control and were combined with 810,648 genotyped SNPs for further analyses. For the Han Chinese sample, a total of 6,343,243 imputed SNPs passed quality control and were

combined with 776,213 genotyped SNPs for association analysis.

Statistical analysis

Population stratification analysis

We used the EIGENSTRAT utility from the EIGENSOFT package (Price et al. 2006) to correct for possible population stratification. First, we combined our data sets with 1000 Genomes Phase 3 and chose 102,284 SNPs in low-linkage equilibrium ($r^2 < 0.2$). Principal component (PC) analysis did not find outliers in Han Chinese and Uyghur (Fig. S1a). The top 4 PCs were later used in the regression model in association analysis. The number of PCs to include was based on the proportion of variance explained and the degree of genomic inflation (lambda).

Association test

We used PLINK 1.7 (Purcell et al. 2007) to perform the primary genome-wide association tests using multiple linear regression with an additive genetic model incorporating gender, batch effect, and four genetic PCs as covariates. The quantile–quantile plots and the degree of genomic inflation (lambda) were used for all association tests to test for signs of inflation. To validate the PLINK results, association analyses were also performed on the imputed dataset using the IMPUTE2 genotype probabilities in SNPTEST v2.5.2 (Marchini and Howie 2010) and results from both methods were consistent. The *P* values were adjusted by applying a Bonferroni correction (genome-wide significant threshold: $P = 6.02 \times 10^{-8}$). An Epistasis test was also performed using PLINK 1.7 (Purcell et al. 2007). Quantile–quantile and Manhattan plots were created in R. Regional association and linkage disequilibrium plots were generated using LocusZoom (Pruim et al. 2010).

To evaluate the relative effects of *TCHH* and *EDAR* variants on hair straightness, we applied a logistic regression model:

$$\log \frac{\pi}{1-\pi} = \beta_0 + \beta_1 \mathrm{cov} + \beta_2 TCHH + \beta_3 EDAR$$

in which, the top four principle components, gender, and batch effects are included as covariates.

To identify whether there was any interaction between *TCHH* and *EDAR* variants, we introduced the interaction effect of these two genes into our logistic regression model:

$$\log \frac{\pi}{1-\pi} = \beta_0 + \beta_1 \mathrm{cov} + \beta_2 TCHH + \beta_3 EDAR + \beta_4 \mathrm{interaction}$$

in which, the top four principle components, gender, and batch effects are included as covariates.

The proportion of variance in hair straightness explained by the associated genetic variants was estimated using GCTA (Yang et al. 2010, 2011), assuming a prevalence of 0.1 in Han Chinese and 0.5 in Uyghurs samples with age and gender as covariates.

We firstly used METAL (Willer et al. 2010) to perform meta-analysis of Han Chinese, Uyghur and Latin American data sets by the inverse-weighted variance method. Heterogeneity of associated SNP across studies was tested via the Cochran's *Q* statistic (Pereira et al. 2010) and its magnitude expressed by I^2 (Higgins and Thompson 2002). For SNPs with significant heterogeneity, a random effects model was used for meta-analysis by METASOFT (Han and Eskin 2011).

Inference of local ancestral proportion

We used HAPMIX (Pritchard et al. 2009) to perform local ancestry inference, which was defined as the estimated number of European ancestral genes for each locus in each sample. We first chose 792,884 SNPs which were well aligned with the 1000 Genomes Project. We used CHB (97) and CEU (85) as input reference populations. Then we selected an LD-pruned ($r^2 < 0.2$) set of 102,284 autosomal SNPs to estimate European ancestral proportions in Uyghurs using STRUCTURE (Hubisz et al. 2009) (Fig. S1b). The estimated European ancestral proportion and 30 generations since admixture for Uyghur were used as prior hypotheses when we ran HAPMIX to infer local ancestry proportion. HAPMIX provided local ancestral proportion estimates as the expected probability of zero, one or two copies of European ancestral proportion at each SNP. We then used MIXSCORE (Pasaniuc et al. 2011) to combine both SNP information and local ancestry estimates to perform association tests.

Haplotype analysis and tests of natural selection

We used the R package rehh (Gautier and Vitalis 2012) to perform positive selection analysis (Sabeti et al. 2002). Firstly, we calculated extended haplotype homozygosity (EHH) for all SNPs until EHH <0.05 in CEU, CHB, and our Uyghur samples. Next, the integrated haplotype score (iHS) was calculated on all SNPs, with an allele frequency bin of 0.05 to standardize iHS scores against other SNPs of its frequency class within the region. Finally, we calculated *P* values assuming a Gaussian distribution of iHS scores under the neutral model, which was checked by plotting the values against a Gaussian distribution (Voight et al. 2006). We also performed genome-wide CMS analysis in African (YRI), European (CEU), and East Asian (JPT + CHB) populations from the 1000 Genomes Project (Grossman et al. 2010) to verify the result. We used HaploPS (Liu et al. 2013) to

detect the founder haplotype of *EDAR* and *TCHH* in different populations using our data and other populations from the 1000 Genomes data, and to identify the common set of SNPs that are present in all haplotypes. We then calculated the proportion of loci carrying the same alleles as the similarity $S = 1 - d/L$, where d represents the number of sites (out of L) carrying different alleles on two haplotypes.

For admixed populations, genomic regions showing excessive or reduced ancestry proportions are likely to be signatures of natural selection (Bhatia et al. 2014). In this study, we chose those loci showing strong deviation from genome average (three SDs above or below the genome-wide average) as candidates for natural selection.

Results

Genome-wide scan of hair straightness in Han Chinese

A genome-wide scan found *EDAR* as the predominant gene associated with hair straightness in Han Chinese (rs3827760: $P = 4.67 \times 10^{-16}$; Fig. 1; see also Fig. S2 for a regional zoom-in plot). Rs3827760 SNP accounted for 3.36 % of the total variance. An additional analysis performed using the same methodology and conditioning on rs3827760 resulted in substantial attenuation of statistical significance for the other significant SNPs (Fig. S3a), suggesting that the vast majority of the signal associated with *EDAR* at this locus can be attributed to rs3827760. No further significant SNPs were found after imputation (Fig. S3b). We also found no evidence for epistasis between the *EDAR* SNPs and any other SNP across the genome.

We performed stratification analysis on gender because it had a significant effect on hair straightness ($P = 4.13 \times 10^{-5}$). We found that rs3827760 still bore the most significant signal in females ($P = 8.71 \times 10^{-13}$), while no genome-wide significant signals were found in males (rs3827760: $P = 4.60 \times 10^{-5}$) (Fig. S3c,d). The previously reported *TCHH* SNP (rs11803731) associated with hair straightness in European populations was nearly fixed in our Han Chinese samples (MAF 0.0008), so we were not able to evaluate its effect on hair straightness in East Asians. The other two SNPs rs17143387 and rs11150606, which showed significant association in Latin Americans (rs17143387: $P = 4\times10^{-8}$; rs11150606: $P = 7\times10^{-9}$) were not replicated in Han Chinese (rs17143387: $P = 0.735$; rs11150606: $P = 0.618$).

Genome-wide scan of hair straightness in Uyghurs

A genome-wide scan in Uyghurs also found *EDAR* as the predominant gene associated with hair straightness (rs3827760: $P = 1.75 \times 10^{-12}$; Fig. 2; see also Fig. S4 for a regional zoom-in). From the quantile–quantile plot and the degree of genomic inflation, lambda ($\lambda = 1.014$), we found that the test correctly controlled for population stratification and there was no sign of inflation (Fig. 2). No further significant signals were found, after controlling the top SNP rs3827760 (Fig. S5a) and after imputation (Fig. S5b), suggesting that the vast majority of the signals associated with *EDAR* at this locus can be attributed to rs3827760. No evidence for epistasis between rs3827760 and other SNPs across the genome was found.

We also performed genome-wide scans in females and males separately and still found the SNP rs3827760 hit the

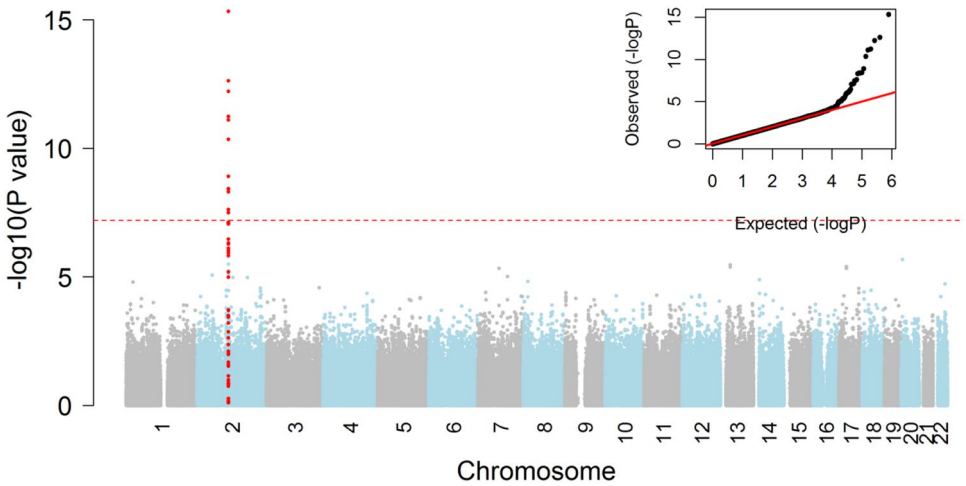

Fig. 1 Manhattan plot and quantile–quantile plot showing a genome-wide scan of hair straightness in Han Chinese. Manhattan (*main panel*) and quantile–quantile (QQ) (embedded) plots illustrating the results of the genome-wide scan in 2899 Han Chinese samples adjusted for top four PCs, gender, and age. The *red line* indicates the threshold for genome-wide statistical significance ($P < 6.02 \times 10^{-8}$). *Red dots* represent SNPs that are close (<5 kb) to the genome-wide significant signals. The quantile–quantile plot illustrates the deviation of association test statistics (*black dots*) from the distribution expected under the null hypothesis (*red line*) (color figure online)

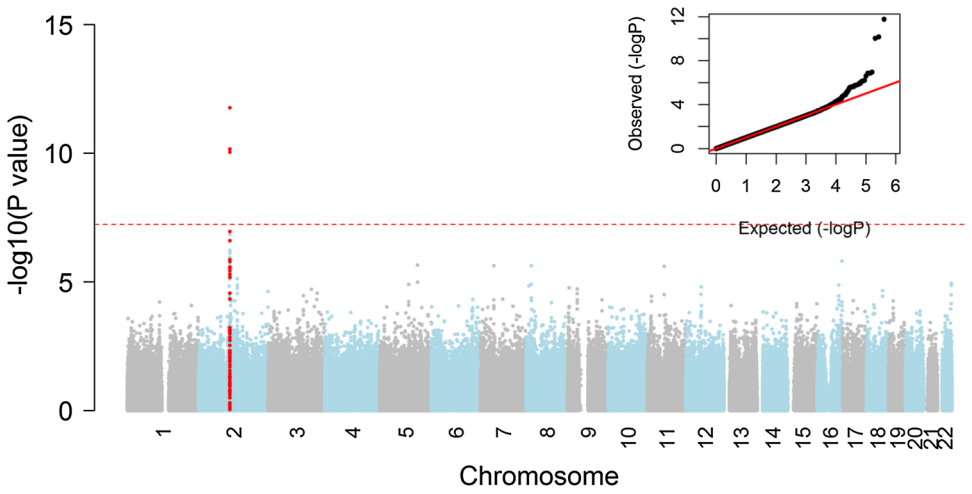

Fig. 2 Manhattan plot and quantile–quantile plot showing a genome-wide scan of hair straightness in Uyghurs. Manhattan (*main panel*) and quantile–quantile (QQ) (embedded) plots illustrating the results of the genome-wide scan in 709 Uyghur samples adjusted for top four PCs, gender and age. The *red line* indicates the threshold for genome-wide statistical significance ($P < 6.02 \times 10^{-8}$). Red dots represent SNPs that are close (<5 kb) to the genome-wide significant signals. The quantile–quantile plot illustrates the deviation of association test statistics (*black dots*) from the distribution expected under the null hypothesis (*red line*) (color figure online)

most significant signal (female: $P = 1.16 \times 10^{-7}$; male: $P = 1.39 \times 10^{-6}$; Fig. S5c,d). No significant heterogeneity was found when we compared the effect of *EDARV370A* between Han Chinese and Uyghurs for males and females (Fig. S6).

To increase the power of the genome-wide scan in the admixed population, we estimated the local ancestry proportion at each SNP, and further controlled local ancestry proportion in the association analysis. The results nevertheless remained largely unchanged (Fig. S7).

The previously reported *TCHH* variant (rs11803731) was associated with hair straightness in our Uyghur samples ($P = 1.46 \times 10^{-3}$) while the other SNPs found in Latins Americans were not replicated in Uyghurs (rs17143387 (*GATA3*): $P = 0.087$; rs11150606 (*PRSS53*): $P = 0.227$). We found that rs3827760 (*EDAR*) had a greater effect on hair straightness than rs11803731 (*TCHH*) in Uyghur (rs3827760: OR 0.415, 95 % CI 0.327–0.532; rs11803731: OR 0.575, 95 % CI 0.409–0.808), and explained more variation (rs3827760: 5.51 %; rs11803731: 1.08 %). Logistic regression analysis showed no significant interaction effect between rs3827760 (EDAR) and rs11803731 (*TCHH*) ($P = 0.363$). A Breslow–Day test confirmed that the interaction effect is not significant ($P = 0.645$). The ratios of straight vs non-straight hair are shown in Table S2 (Fig. S8). Finally, we performed meta-analysis for Han Chinese and Uyghurs. Again, we found only the predominant *EDAR* gene to show a highly significant signal in the genome-wide scan (rs3827760: $P = 1.75 \times 10^{-30}$; rs11803731: $P = 1.49 \times 10^{-4}$; Fig. S9).

From the genome-wide scans, the SNP rs3827760 in *EDAR* gene accounts for 3.66 and 5.51 % of the total variance of hair straightness in Han Chinese and Uyghurs, respectively. This is only a small fraction of the heritability of hair straightness, estimated as 64.15 and 79.72 % in Han Chinese and Uyghurs, respectively. To explore the remaining heritability, we selected 23 genes previously shown to contribute to hair formation as suggestive functional candidate genes (Table S3). We chose 692 SNPs located within 20 kb of these candidate genes and performed an association study of these SNPs with hair straightness. The association *P* values for these 692 SNPs in 23 genes are summarized in Table S3. After multiple-test correction (threshold: $P < 7.23 \times 10^{-5}$), we found no significant association signals in Uyghur and Han Chinese.

Signatures of natural selection on *TCHH* and *EDAR*

We found that *TCHH* is not subject to selection in CHB, UYG, and CEU populations (most SNPs in *TCHH* region |iHS| <2), while *EDAR* is under strong positive selection in CHB but not in CEU and UYG (there is an apparent enrichment for high |iHS| values >3.56 (top 0.1 %) in CHB but none for CEU and UYG) (Fig. 3). We also obtained similar results from CMS scores (Fig. S10).

Haplotype analysis showed that the haplotype of *EDAR370A* in Uyghur is very similar to the haplotype of *EDAR370A* in East Asian and American populations (Fig. S11a), while the *TCHH* haplotype is similar to the haplotype of the European population (Fig. S11b). These

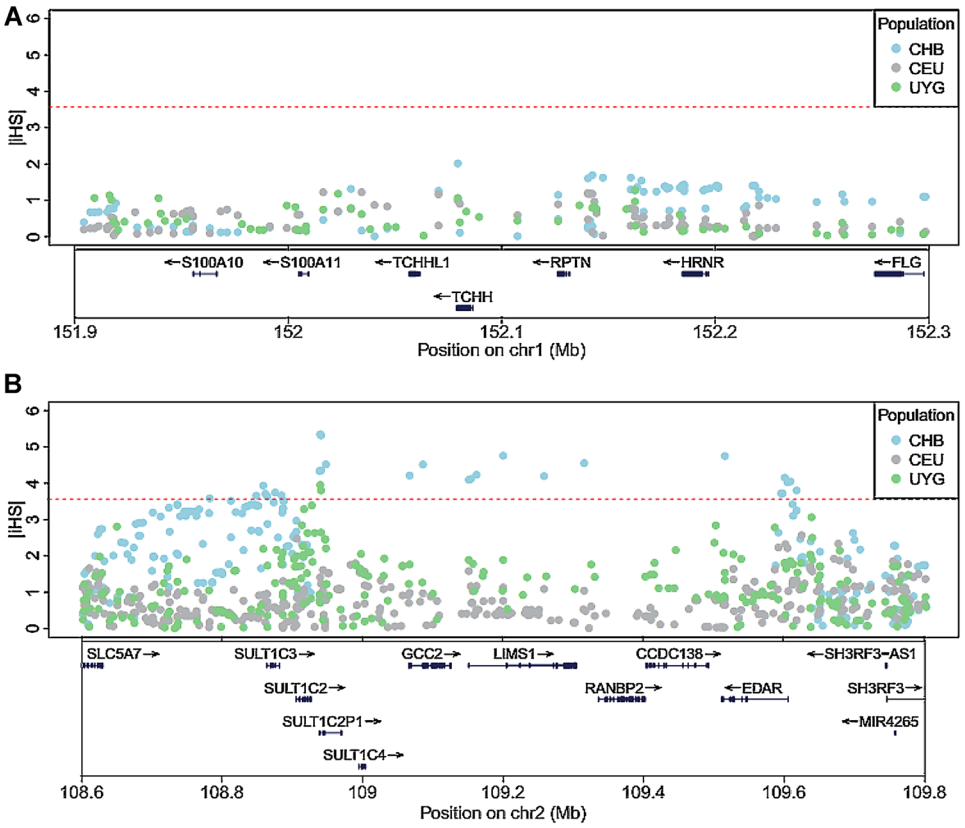

Fig. 3 Signatures of natural selection on *TCHH* and *EDAR*. Absolute iHS values are plotted against physical distance for *TCHH* and *EDAR* region in Han Chinese, European and Uyghur populations, which are denoted by *blue*, *gray,* and *green,* separately. A *red line* indicates the top 0.1 % chromosome-wide |iHS| threshold (>3.56). **a** The |iHS| results for the *TCHH* region. No enrichment for high iHS values (top 0.1 %) were found, which indicates that *TCHH* was not subject to natural selection. **b** The |iHS| results for the *EDAR* region. The enrichment for high |iHS| values (Top 0.1 %) found at *EDAR* region indicates that *EDAR* was subject to natural selection in Han Chinese populations (color figure online)

findings indicate that the causal variants in *EDAR* and *TCHH* arose as a single mutation event in the parental populations, respectively, and were subsequently passed to the Uyghurs through population admixture. Local ancestry inference analysis showed that the locus-specific European ancestral proportion across the genome of Uyghurs was estimated to be 51.08 ± 2.67 % (mean ± SD).We found that the European proportion of rs11803731 is 51.14 % and rs3827760 is 57.21 % (Fig. S12). The result showed that rs11803731 is not subject to selection after admixture in Uyghur populations while the ancestral *EDAR*370 V allele is over-represented (2.30 SD away from the genome average).

Discussion

EDAR as the only predominant gene affecting hair straightness in East Asians

Previous large-scale association studies of hair straightness have focused on European and Latin American populations. No genome-wide scan of hair straightness had been performed in East Asians prior to the study reported here, which included 2899 Han Chinese and 709 Uyghurs. We consistently found that *EDARV370A* (rs3827760) plays a predominant role in hair straightness. Since *EDARV370A* only explains a small fraction of the heritability of hair straightness, we used a candidate gene approach to explore the remaining heritability but found no further significant associations, suggesting that the hair straightness phenotype may be affected by a large number of genes with subtle effects. Future studies with larger sample sizes may uncover such variants with minor effects on hair straightness.

The impact of *EDAR* and *TCHH* on hair straightness in Uyghurs

The associated *EDAR* (rs3827760) and *TCHH* (rs11803731) variants are almost fixed in East Asian and European populations, respectively (Fig. S13). They are at moderate frequencies in the admixed Uyghur population, so we can evaluate the relative effect on hair straightness of *EDAR* and *TCHH*. Our results clearly show that *EDAR* has a greater effect and explains more variation than *TCHH* in the Uyghurs. We also found no interaction between these two genes in their effect on hair straightness. Given that Trichohyalin confers mechanical strength to the hair follicle inner root sheath (Medland et al. 2009a; Thibaut et al.

2007) while the Ectodysplasin A receptor gene *EDAR*, is part of an important pathway affecting the development of ectodermal derivatives (Fujimoto et al. 2008; Kamberov et al. 2013; Mou et al. 2008), it is probable that *EDAR* and *TCHH* affect hair straightness through different mechanisms.

Hair straightness may not be subject to natural selection

Hair color is reportedly subject to sexual selection, and it has been suggested that the shape of hair may be subject to sexual selection (Frost 2006). If that were the case, both *TCHH* and *EDAR* should show signatures of natural selection in European and East Asian populations, respectively. In our study, we found no signature of natural selection for *TCHH* in Europeans. Although we found that *EDAR* is under strong selection in East Asians, it has not been subject to selection after the admixture in Uyghurs. These findings do not support hair straightness to be a trait under selection. Since *EDARV370A* affects a range of ectodermal derived phenotypes, hair straightness might be a byproduct of strong selection on *EDAR* in East Asians for other traits.

Acknowledgments This work was supported by the National Natural Science Foundation of China (31322030, 91331108 to S.W., 30890034, 31271338 to L.J., 91331204, 31525014 to S.X., 31501011 to Y.L., 31260263 to Y.G., 31401061 to Q.P., 31071102 to J.T., 31071096 to Y.Y.), the National Basic Research Program (2012CB944600 to L.J.), the Ministry of Science and Technology (2011BAI09B00 to L.J.), the Ministry of Education (311016 to L.J.), the Chinese Academy of Sciences (XDB13040100 to S.X., 2014KIP214 to Q.P.), the Science and Technology Commission of the Shanghai Municipality (14YF1406800 to Y.L., 16YF1413900 to H.L.), the National Program for Top-notch Young Innovative Talents of The "Wanren Jihua" Project to S.X., the National Thousand Young Talents Award to S.W., the Max Planck-CAS Paul Gerson Unna Independent Research Group Leadership Award to S.W., open projects from the State Key Laboratory of Genetic Engineering, Fudan University to S.W., and the CAS Key Laboratory of Computational Biology to J.T.

References

Adhikari K, Fontanil T, Cal S, Mendoza-Revilla J, Fuentes-Guajardo M, Chacon-Duque JC, Al-Saadi F, Johansson JA, Quinto-Sanchez M, Acuna-Alonzo V, Jaramillo C, Arias W, Barquera Lozano R, Macin Perez G, Gomez-Valdes J, Villamil-Ramirez H, Hunemeier T, Ramallo V, Silva de Cerqueira CC, Hurtado M, Villegas V, Granja V, Gallo C, Poletti G, Schuler-Faccini L, Salzano FM, Bortolini MC, Canizales-Quinteros S, Rothhammer F, Bedoya G, Gonzalez-Jose R, Headon D, Lopez-Otin C, Tobin DJ, Balding D, Ruiz-Linares A (2016) A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features. Nat Commun 7:10815. doi:10.1038/ncomms10815

Bhatia G, Tandon A, Patterson N, Aldrich MC, Ambrosone CB, Amos C, Bandera EV, Berndt SI, Bernstein L, Blot WJ, Bock CH, Caporaso N, Casey G, Deming SL, Diver WR, Gapstur SM, Gillanders EM, Harris CC, Henderson BE, Ingles SA, Isaacs W, De Jager PL, John EM, Kittles RA, Larkin E, McNeill LH, Millikan RC, Murphy A, Neslund-Dudas C, Nyante S, Press MF, Rodriguez-Gil JL, Rybicki BA, Schwartz AG, Signorello LB, Spitz M, Strom SS, Tucker MA, Wiencke JK, Witte JS, Wu X, Yamamura Y, Zanetti KA, Zheng W, Ziegler RG, Chanock SJ, Haiman CA, Reich D, Price AL (2014) Genome-wide scan of 29,141 African Americans finds no evidence of directional selection since admixture. Am J Hum Genet 95:437–444. doi:10.1016/j.ajhg.2014.08.011

Eriksson N, Macpherson JM, Tung JY, Hon LS, Naughton B, Saxonov S, Avey L, Wojcicki A, Pe'er I, Mountain J (2010) Web-based, participant-driven studies yield novel genetic associations for common traits. PLoS Genet 6:e1000993. doi:10.1371/journal.pgen.1000993

Frost P (2006) European hair and eye color—a case of frequency-dependent sexual selection? Evol Hum Behav 27:85–103. doi:10.1016/j.evolhumbehav.2005.07.002

Fujimoto A, Kimura R, Ohashi J, Omi K, Yuliwulandari R, Batubara L, Mustofa MS, Samakkarn U, Settheetham-Ishida W, Ishida T, Morishita Y, Furusawa T, Nakazawa M, Ohtsuka R, Tokunaga K (2008) A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. Hum Mol Genet 17:835–843. doi:10.1093/Hmg/Ddm355

Gautier M, Vitalis R (2012) rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics 28:1176–1177. doi:10.1093/bioinformatics/bts115

Grossman SR, Shylakhter I, Karlsson EK, Byrne EH, Morales S, Frieden G, Hostetter E, Angelino E, Garber M, Zuk O, Lander ES, Schaffner SF, Sabeti PC (2010) A composite of multiple signals distinguishes causal variants in regions of positive selection. Science 327:883–886. doi:10.1126/science.1183863

Han B, Eskin E (2011) Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. Am J Hum Genet 88:586–598. doi:10.1016/j.ajhg.2011.04.014

Higgins JPT, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. Stat Med 21:1539–1558. doi:10.1002/sim.1186

Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. doi:10.1371/journal.pgen.1000529

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322–1332. doi:10.1111/j.1755-0998.2009.02591.x

Kamberov YG, Wang SJ, Tan JZ, Gerbault P, Wark A, Tan LZ, Yang YJ, Li SL, Tang K, Chen H, Powell A, Itan Y, Fuller D, Lohmueller J, Mao JH, Schachar A, Paymer M, Hostetter E, Byrne E, Burnett M, McMahon AP, Thomas MG, Lieberman DE, Jin L, Tabin CJ, Morgan BA, Sabeti PC (2013) Modeling recent human evolution in mice by expression of a selected EDAR variant. Cell 152:691–702. doi:10.1016/j.cell.2013.01.016

Liu X, Ong RT, Pillai EN, Elzein AM, Small KS, Clark TG, Kwiatkowski DP, Teo YY (2013) Detecting and characterizing genomic signatures of positive selection in global populations. Am J Hum Genet 92:866–881. doi:10.1016/j.ajhg.2013.04.021

Loussouarn G, Garcel AL, Lozano I, Collaudin C, Porter C, Panhard S, Saint-Leger D, de La Mettrie R (2007) Worldwide diversity of hair curliness: a new method of assessment. Int J Dermatol 46(Suppl 1):2–6. doi:10.1111/j.1365-4632.2007.03453.x

Marchini J, Howie B (2010) Genotype imputation for genome-wide association studies. Nat Rev Genet 11:499–511. doi:10.1038/nrg2796

Medland SE, Nyholt DR, Painter JN, McEvoy BP, Mcrae AF, Zhu G, Gordon SD, Ferreira MAR, Wright MJ, Henders AK, Campbell MJ, Duffy DL, Hansell NK, Macgregor S, Slutske WS, Heath AC, Montgomery GW, Martin NG (2009a) Common variants in the trichohyalin gene are associated with straight hair in Europeans. Am J Hum Genet 85:750–755. doi:10.1016/j.ajhg.2009.10.009

Medland SE, Zhu G, Martin NG (2009b) Estimating the heritability of hair curliness in twins of European Ancestry. Twin Res Hum Genet 12:514–518

Mou C, Thomason HA, Willan PM, Clowes C, Harris WE, Drew CF, Dixon J, Dixon MJ, Headon DJ (2008) Enhanced ectodysplasin-A receptor (EDAR) signaling alters multiple fiber characteristics to produce the east asian hair form. Hum Mutat 29:1405–1411. doi:10.1002/humu.20795

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

Pasaniuc B, Zaitlen N, Lettre G, Chen GK, Tandon A, Kao WH, Ruczinski I, Fornage M, Siscovick DS, Zhu X, Larkin E, Lange LA, Cupples LA, Yang Q, Akylbekova EL, Musani SK, Divers J, Mychaleckyj J, Li M, Papanicolaou GJ, Millikan RC, Ambrosone CB, John EM, Bernstein L, Zheng W, Hu JJ, Ziegler RG, Nyante SJ, Bandera EV, Ingles SA, Press MF, Chanock SJ, Deming SL, Rodriguez-Gil JL, Palmer CD, Buxbaum S, Ekunwe L, Hirschhorn JN, Henderson BE, Myers S, Haiman CA, Reich D, Patterson N, Wilson JG, Price AL (2011) Enhanced statistical tests for GWAS in admixed populations: assessment using African Americans from CARe and a Breast Cancer Consortium. PLoS Genet 7:e1001371. doi:10.1371/journal.pgen.1001371

Pereira TV, Patsopoulos NA, Salanti G, Ioannidis JPA (2010) Critical interpretation of Cochran's Q test depends on power and prior assumptions about heterogeneity. Res Synth Methods 1:149–161. doi:10.1002/jrsm.13

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909. doi:10.1038/ng1847

Pritchard JK, Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, Ruczinski I, Beaty TH, Mathias R, Reich D, Myers S (2009) Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. PLoS Genet 5:e1000519. doi:10.1371/journal.pgen.1000519

Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ (2010) LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26:2336–2337. doi:10.1093/bioinformatics/btq419

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575. doi:10.1086/519795

Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832–837. doi:10.1038/Nature01140

Tan JZ, Yang YJ, Tang K, Sabeti PC, Jin L, Wang SJ (2013) The adaptive variant EDARV370A is associated with straight hair in East Asians. Hum Genet 132:1187–1191. doi:10.1007/s00439-013-1324-1

Thibaut S, Barbarat P, Leroy F, Bernard BA (2007) Human hair keratin network and curvature. Int J Dermatol 46(Suppl 1):7–10. doi:10.1111/j.1365-4632.2007.03454.x

Voight BF, Kudaravalli S, Wen XQ, Pritchard JK (2006) A map of recent positive selection in the human genome. PLoS Biol 4:446–458. doi:10.1371/journal.pbio.0040072

Wang XF, Lu M, Qian J, Yang YJ, Li SL, Lu DR, Yu SZ, Meng W, Ye WM, Jin L (2009) Rationales, design and recruitment of the Taizhou Longitudinal Study. BMC Public Health. doi:10.1186/1471-2458-9-223

Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26:2190–2191. doi:10.1093/bioinformatics/btq340

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42:565–569. doi:10.1038/ng.608

Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88:76–82. doi:10.1016/j.ajhg.2010.11.011