# Genome-wide association study of pigmentary traits (skin and iris color) in individuals of East Asian ancestry

**Lida Rawofi** [1] , **Melissa Edwards** [1] , **S Krithika** [1] , **Phuong Le** [1] , **David Cha** [1] , **Zhaohui Yang** [2] , **Yanyun Ma** [3] , **Jiucun Wang** [4] , **Bing Su** [5] , **Li Jin** [4] , **Heather L Norton** [6] , **Esteban J Parra** Corresp. [1]

[1] Department of Anthropology, University of Toronto at Mississauga, Mississauga, Canada

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

[3] MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China

[4] State Key laboratory of Genetic Engineering, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, China

[5] State Key Laboratory of Genetic Resources and Evolution, Kumming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

[6] Department of Anthropology, University of Cincinnati, Cincinnati, United States

Corresponding Author: Esteban J Parra
Email address: esteban.parra@utoronto.ca

Background. Currently, there is limited knowledge about the genetics underlying pigmentary traits in East Asian populations. Here, we report the results of the first genome-wide association study of pigmentary traits (skin and iris color) in individuals of East Asian ancestry.

Methods. We obtained quantitative skin pigmentation measures (M-index) in the inner upper arm of the participants using a portable reflectometer (N=305). Quantitative measures of iris color (expressed as L*, a* and b* CIELab coordinates) were extracted from high-resolution iris pictures (N=342). We also measured the color differences between the pupillary and ciliary regions of the iris (e.g. iris heterochromia). DNA samples were genotyped with Illumina's Infinium Multi-Ethnic Global Array (MEGA) and imputed using the 1000 Genomes Phase 3 samples as reference haplotypes.

Results. For skin pigmentation, we did not observe any genome-wide significant signal. We followed-up in three independent Chinese samples the lead SNPs of five regions showing multiple common markers (minor allele frequency ≥ 5%) with good imputation scores and suggestive evidence of association (p-values < 10-5). One of these markers, rs2373391, which is located in an intron of the ZNF804B gene on chromosome 7, was replicated in one of the Chinese samples (p=0.003). For iris color, we observed genome-wide signals in the OCA2 region on chromosome 15. This signal is driven by the non-synonymous rs1800414 variant, which explains 11.9%, 10.4% and 6% of the variation observed in the b*, a* and L* coordinates in our sample, respectively. However, the OCA2 region was not associated with iris heterochromia.

Discussion. Additional genome-wide association studies in East Asian samples will be

necessary to further disentangle the genetic architecture of pigmentary traits in East Asian populations.

1 **Genome-wide association study of pigmentary traits (skin and iris color) in individuals of**
2 **East Asian ancestry**

4 Lida Rawofi[1], Melissa Edwards[1], S. Krithika[1]. Phuong Le[1], David Cha[1], Zhaohui Yang[2], Yanyun
5 Ma[3], Jiucun Wang[4], Bing Su[5], Li Jin[4], Heather L. Norton[6], Esteban J. Parra[1]

7 [1]Department of Anthropology, University of Toronto at Mississauga, Mississauga, ON, Canada
8 [2]Yunnan Key Laboratory of Primate Biomedical Research, Institute of Primate Translational
9 Medicine, Kunming University of Science and Technology, Kunming, China
10 [3]MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan
11 University, Shanghai, China
12 [4]State Key Laboratory of Genetic Engineering, Collaborative Innovation Center for Genetics and
13 Development, School of Life Sciences, Fudan University, Shanghai, China.
14 [5]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology,
15 Chinese Academy of Sciences, Kunming, China
16 [6]Department of Anthropology, University of Cincinnati, Cincinnati, USA

18     **Abstract**
19
20 **Background.** Currently, there is limited knowledge about the genetics underlying pigmentary
21 traits in East Asian populations. Here, we report the results of the first genome-wide association
22 study of pigmentary traits (skin and iris color) in individuals of East Asian ancestry.
23 **Methods.** We obtained quantitative skin pigmentation measures (M-index) in the inner upper
24 arm of the participants using a portable reflectometer (N=305). Quantitative measures of iris
25 color (expressed as L*, a* and b* CIELab coordinates) were extracted from high-resolution iris
26 pictures (N=342). We also measured the color differences between the pupillary and ciliary
27 regions of the iris (e.g. iris heterochromia). DNA samples were genotyped with Illumina's
28 Infinium Multi-Ethnic Global Array (MEGA) and imputed using the 1000 Genomes Phase 3
29 samples as reference haplotypes.
30 **Results**. For skin pigmentation, we did not observe any genome-wide significant signal. We
31 followed-up in three independent Chinese samples the lead SNPs of five regions showing
32 multiple common markers (minor allele frequency $\geq$ 5%) with good imputation scores and
33 suggestive evidence of association (p-values < $10^{-5}$). One of these markers, rs2373391, which is
34 located in an intron of the *ZNF804B* gene on chromosome 7, was replicated in one of the
35 Chinese samples (p=0.003). For iris color, we observed genome-wide signals in the *OCA2* region
36 on chromosome 15. This signal is driven by the non-synonymous rs1800414 variant, which
37 explains 11.9%, 10.4% and 6% of the variation observed in the b*, a* and L* coordinates in our
38 sample, respectively. However, the *OCA2* region was not associated with iris heterochromia.
39 **Discussion**. Additional genome-wide association studies in East Asian samples will be necessary
40 to further disentangle the genetic architecture of pigmentary traits in East Asian populations.

41

42     **Introduction.**
43
44     Human pigmentation diversity is primarily driven by the type, amount and distribution of
45 melanin in the skin, hair and iris.  Pigmentation is a polygenic trait, and the last decade has
46 witnessed numerous efforts to elucidate the genetic architecture of pigmentation through

47   association and functional studies (Lamason et al., 2005; Kayser et al., 2008; Visser, Kayser &
48   Palstra 2012; Lui et al., 2015). There have also been important advances in the development of
49   methods based on reflectance and bioimaging technologies, which have made it possible to
50   obtain quantitative measurements of skin, hair and iris pigmentation (Liu et al., 2010; Walsh et
51   al., 2011; Edwards et al., 2012; Beleza et al., 2013; Andersen et al., 2013; Norton et al., 2015;
52   Edwards et al., 2015; Wollstein et al., 2017). As a result of these efforts, dozens of genetic
53   markers have been associated with pigmentary phenotypes. However, the overwhelming
54   majority of studies have focused on European populations, and there are still substantial gaps in
55   our understanding of the genetic basis of pigmentation in other population groups.

56          Nonetheless, available research has shown skin lightening in Europe and East Asia to
57   have occurred independently through convergent evolution. In Europe, markers within the
58   *HERC2* gene (e.g. rs12913832), which is close to *OCA2*, are strongly associated with blue eyes,
59   and have also been associated with light skin pigmentation in GWAS. (Kayser et al., 2008; Lui et
60   al., 2015; Sturm et al., 2008; Visser et al., 2012). The haplotype defined by rs12913832 is
61   primarily restricted to Europe. In East Asia however, two non-synonymous variants in the *OCA2*
62   gene; rs1800414 (His615Arg) and rs74653330 (Ala481Thr) have been associated with light skin
63   and eye color (Edwards et al., 2010; Abe et al., 2013; Eaton et al., 2015; Edwards et al., 2015;
64   Norton et al., 2015; Yang et al 2016). Both of these polymorphisms are predicted to have a
65   deleterious effect on the protein (Eaton et al., 2015). Functional studies have recently confirmed
66   the role of rs1800414 in East Asian pigmentation (Yang et al., 2016). The haplotypes defined by
67   rs1800414 and rs74653330 are restricted to East Asia but they have a very different geographic
68   distribution. The SNP rs1800414 is very frequent across East Asia while rs74653330 is primarily
69   restricted to the Altaic speaking populations from Northern East Asia and Mongolia (Murray et
70   al., 2015). There is general agreement that these two non-synonymous mutations arose long after
71   the split of European and East Asian populations (Chen et al., 2015; Murray et al., 2015; Yang et
72   al., 2016). In addition to variants in the *OCA2* gene, a non-synonymous mutation at rs885479
73   (Arg163Gln) in *MC1R* has also been association with pigmentation in East Asia (Yamaguchi et
74   al., 2012). Interestingly, the derived 163Gln allele is present in very high frequencies in East
75   Asian populations (>60%), but very low frequencies in European and African populations.

76          Most of the studies conducted in East Asian populations have been candidate gene
77   studies. Here, we report the results of the first genome wide association study of pigmentary
78   traits in an East Asian population sample. Skin pigmentation was measured with a portable
79   reflectometer while eye color was measured quantitatively (CIELab color space) from high-
80   resolution pictures of the iris. The samples were genotyped with Illumina's MEGA array, and
81   imputed with the 1KG Phase 3 reference samples. We followed up the main signals observed in
82   our GWAS in two independent East Asian samples. This study provides important insights about
83   the architecture of skin pigmentation in East Asian populations.

84

85   **Materials and methods.**

86   *Sample collection.*

87    Recruitment of study participants was carried out at the University of Toronto (Canada). All
88    participants ranged between 18 and 35 yr of age and were recruited using online and print
89    advertisements directed towards the University of Toronto student community. A personal
90    questionnaire was administered to each participant to determine their age, sex and whether or not
91    they had been diagnosed with any pigmentation-related diseases or disorders.

92    Biogeographical ancestry was determined using information from the personal questionnaire,
93    which inquired about the ancestry, place of birth and first language of each participant's maternal
94    and paternal grandparents. The sample used in this study comprised individuals who had
95    grandparents from China, Japan, South Korea or Taiwan. In the few cases where information
96    about grandparents was not available, we used information about both parents to assess
97    biogeographical ancestry. The total number of individuals included in the study was 425. The
98    study was approved by the University of Toronto Research and Ethics Board (Protocol
99    Reference #27015), and all participants provided written informed consent. A 2-ml saliva sample
100    was obtained from each participant using the Oragene·DNA (OG-500) collection kit (DNA
101    Genotek, Canada). All participants were instructed not to eat, drink or smoke for at least 30 min
102    prior to obtaining the sample to ensure maximal sample purity. DNA was isolated from each
103    sample using the protocol provided by DNA Genotek and eluted in 500 ml of TE (10 mM Tris-
104    HCl, 1 mM EDTA, pH 8.0) Buffer. Prior to genotyping, 23 DNA samples were excluded due to
105    poor DNA quality. The final number of samples genotyped was 402.

106    *Measurement of pigmentary traits.*

107    Skin pigmentation was measured quantitatively using the DSMII Dermaspectrometer (Cortex
108    Technologies, Hadsund, Denmark), and melanin levels were reported as M index (individuals
109    with higher M-values have darker pigmentation). Measurements were taken three times on the
110    inner skin of the upper right arm and pigmentation was reported as the mean of the three
111    measures, after excluding outliers. High-resolution pictures of the right iris of each participant
112    were taken with a Fujifilm Finepix S3 Pro 12-megapixel DSLR mounted on a Nikor 105-mm
113    macro lens. To control for lighting and exposure, photographs were taken with a coaxial
114    biometric illuminator to deliver a constant and uniform source of light to each iris at 5,500 K
115    (D55 illuminant).  All photographs were taken under the same setting Delaneau et al., 2012)
116    conditions, with an aperture of f/19, exposure sensitivity (ISO) set at 200 and a shutter speed of
117    1/125 s. Iris pigmentation was digitally scored using a custom program designed to crop out both
118    the pupil and sclera to retain only the iris.  A wedge of the iris was then extracted, and color
119    scores in CIELab coordinates were calculated from the pupillary and ciliary zones. In addition to
120    the L*, a* and b* coordinates for the iris wedge, the program calculated the parameter delta,
121    which describes color differences in the pupillary and ciliary regions of the iris. Detailed
122    information about this program has been described in Edwards et al., (2015).

123    *Genotyping, phasing and imputation*

124    Genotyping was carried out with Illumina's Infinium Multi-Ethnic Global Array (MEGA) at the
125    Clinical Genomics Centre (Mount Sinai Hospital, Toronto, Ontario, Canada) using standard
126    protocols. The MEGA array, which includes approximately 1.7 million markers, was designed to

127   capture common genome variation in diverse population groups. Four samples were included as
128   blind duplicates, and the concordance rate was in all samples higher than 99.99%. We used the
129   program GenomeStudio to carry out the basic QC steps recommended by Illumina. After this
130   initial QC step, approximately 1.4 million were retained for further analyses. The number of
131   autosomal markers included was approximately 1.36 million. We performed additional QC steps
132   to remove samples and markers, according to the following criteria, Sample QC: 1/ removal of
133   samples with missing call rates < 0.9, 2/ removal of samples that were outliers in Principal
134   Component Analysis (PCA) plots, 3/ removal of samples with sex discrepancies, 4/ removal of
135   samples that were outliers for heterozygosity, and 5/ removal of related individuals (pi-hat >
136   0.2). Marker QC: 1/ removal of markers with genotype call rate < 0.95, 2/ removal of markers
137   with Hardy-Weinberg p-values < $10^{-6}$, 3/ removal of Insertion/Deletion (Indel) markers, 4/
138   removal of markers with allele frequencies < 0.01, 5/ removal of markers not present in the 1000
139   Genomes reference panel, or that do not match on chromosome, position and alleles, 6/ removal
140   of A/T or G/C SNPs with MAF > 40% in the 1000 Genomes East Asian reference samples, and
141   7/ removal of SNPs with allele frequency differences > 20% between the study sample and the
142   1000 Genomes East Asian reference sample. After these QC steps, we retained 377 samples and
143   520,076 markers.

144   After performing the QC steps described above, the samples were phased using the program
145   SHAPEIT2 (and imputed at the Sanger Imputation Service, using the Positional Burrows-
146   Wheeler Transform (PBWT) algorithm (Durbin, 2014), and the samples of the 1000 Genomes as
147   reference haplotypes.

148   *Population structure*

149   We used the program EIGENSOFT to perform PCA and evaluate population stratification after
150   pruning markers in high LD and removing regions showing high LD or genomic complexity.

151   *Statistical analyses.*

152   As a first step of the statistical analyses, we carried out a linear regression with M-values as the
153   dependent variable, and sex and the first four Principal Component Axis as independent
154   variables and saved the standardized residuals. A similar process was carried out for the L*, a*,
155   b* and delta iris values, but in this case, due to deviations from normality, the unstandardized
156   residuals were transformed using the rank-based inverse normal transformation. The M-value
157   residuals and the L*, a*, b* and delta transformed residuals were used as input for the
158   association tests with the program SNPTEST v2 (Marchini and Howie, 2010), using an additive
159   model and the expected test (e.g. using genotype dosages) in order to control for genotype
160   uncertainty. For the L*, a*, b* coordinates that define iris color, we also run a Bayesian Multiple
161   Phenotype test implemented in the program SNPTEST (-mpheno option). This test evaluates the
162   three coordinates jointly and provides a $\log^{10}$ Bayes Factor reporting the ratio of two
163   probabilities: the probability of the data under an unconstrained model ($M_1$), and the probability
164   of the data under a null model ($M_0$) in which there is no effect. For example, a $\log^{10}$ Bayes Factor
165   of 3 indicates that the probability of the data under the model M1 is 1000-fold higher than the
166   probability of the data under the null model with no genotype effects.

167   Of the 377 samples that were retained after the post-genotyping QC step, some samples had
168   missing phenotype data. The final number of samples with valid skin pigmentation data was 305,
169   and the final number of samples with valid iris color data was 342.

170   *Annotation of genome-wide significant and suggestive signals.*

171   The genome-wide significant ($p<10^{-8}$) and suggestive signals ($p<10^{-5}$) identified in the statistical
172   analyses were annotated using the online SNP-Nexus tool (http://snp-nexus.org/), which provides
173   extensive annotations, including potential effects of non-synonymous coding SNPs on protein
174   function (e.g. SIFT and Polyphen), potential regulatory effects (e.g. conserved transcription
175   factor binding sites, microRNAs, Enhancers and CpG islands), evidence of evolutionary
176   conservation (e.g. PHAST and GERP++) and evidence of association with complex diseases and
177   disorders (e.g. GAD and NHGRI Catalogue of Published Genome-Wide Association studies).

178   *Replication in independent East Asian samples.*

179   The markers showing the strongest evidence of association with skin pigmentation were
180   followed up in three independent Chinese samples, for which skin pigmentation measures (M-
181   index or L* values) were also available.  The first two samples were collected by the Laboratory
182   of Contemporary Anthropology at Fudan University between 2013 and 2015. Constitutive
183   pigmentation was measured with a DSMII colormeter (Cortex Technology, Hadsund, Denmark)
184   and pigmentation levels were reported as M-index values (darker pigmentation corresponds to
185   higher M-index values). Pigmentation was estimated as the mean of three measures, after
186   excluding outliers. The first sample was a cohort from the Jinan Military Hospital (N=559).  All
187   the participants were male, and the mean age of this sample was 21.13 years. Constitutive
188   pigmentation was measured in the buttocks during the month of September. The second sample
189   was a cohort from Taizhou (N=568), comprising 404 females and 164 males, with an average
190   age of 44.32 years. Constitutive pigmentation was measured in the upper inner arm in the month
191   of April. Genotyping of the SNPs was carried out using SNaPshot. In both samples, the
192   association of the relevant markers with M-index values was carried out using linear regression,
193   including age, rs1800414 and sex (only for the Taizhou cohort) as covariates. The third sample
194   comprises 346 Han Chinese individuals (College students) for whom constitutive pigmentation
195   was measured in the buttocks and the upper inner arm. The study took place in the month of
196   September. In this case, pigmentation was estimated as the mean of three measurements and
197   reported using L* values (darker pigmentation corresponds to lower L* values). Genotyping was
198   done using Sanger sequencing. The association of the variants with pigmentation was tested with
199   the program PLINK, using sex and age as covariates. Unfortunately, to our knowledge, there
200   have been no studies in East Asia reporting quantitative iris color estimates using the CIELab
201   color space, so it was not possible to follow-up the signals identified for iris color in our East
202   Asian samples.

203

204   **Results**

205   *Distribution of pigmentary traits in East Asian sample.*

206  Supplementary Figure 1 shows the distribution of skin pigmentation values, expressed as the M-
207  index. The average M-value was 37.83 (SD=2.843). Supplementary Figures 2A, 2B and 2C
208  show the scatterplots of a* and b*, L* and a* and L* and b* iris color coordinates, respectively.
209  The CIELab plots show that there is a substantial amount of variation in the three iris color
210  coordinates, emphasizing the importance to use quantitative estimates of iris color, instead of
211  categorical classifications (which in the case of this sample, would be restricted to "brown"
212  color).  Finally, the distribution of the delta values (difference in color coordinates between the
213  pupillary and ciliary regions of the iris) are depicted in Supplementary Figure 3.

214  *Population structure*

215  Supplementary Figure 4 shows the representation of the first two axes of a PCA analysis of the
216  East Asian sample included in this study, as well as the East Asian samples of the 1000 Genomes
217  project (CDX: Chinese Dai in Xishuangbanna, KHV: Kinh in Ho Chi Minh City, Vietnam, CHB:
218  Han Chinese in Beijing, CHS: Han Chinese South and JPT: Japanese in Tokyo). Most of the East
219  Asian individuals included in this study overlap with the two Han Chinese samples (CHB and
220  CHS) and a few individuals with the Japanese samples. Interestingly, the East Asian individuals
221  who reported ancestry from Korea in our sample form a discrete cluster that is located between
222  the Han Chinese and the Japanese 1000 Genomes samples.

223  *Results of Genome-Wide Association Study*

224  Figure 1 shows a Manhattan plot reporting the association results for skin pigmentation. We did
225  not observe any genome-wide significant signal ($p<5x10^{-8}$). Supplementary Figure 5 shows the
226  QQ plot corresponding to skin pigmentation. There was no evidence of genomic inflation
227  (lambda: 1.00). Table 1 reports the list of the common markers (maf>5%) with the lowest p-
228  values identified in the GWA study. All the markers listed in the Table have good imputation
229  info scores (info> 0.8), and correspond to regions with multiple significant markers
230  (Supplementary Excel File). The regional plots for these regions are depicted in Supplementary
231  Figure 6. Figure 2 shows the Manhattan plots reporting the association results for iris color (L*,
232  a*, b* and delta). The QQ plots for these traits are depicted in Supplementary Figure 7. Again,
233  there is no evidence of genomic inflation for any of the traits. We observed a genome-wide
234  signal in the well-known *OCA2* region. In this region, the top signal was rs76930569, which had
235  p-values of $9.27x10^{-11}$ and $5.30x10^{-12}$ for the a* and b* color dimensions, respectively. The
236  Bayesian multiple phenotype test implemented in SNPTEST, based on the L*, a* and b*
237  dimensions, also provided strong support for the *OCA2* region, with a $log^{10}$ Bayes Factor of 7. In
238  addition to the *OCA2* region, we identified numerous regions showing multiple markers with
239  suggestive associations ($p<10^{-5}$) and good imputation scores. The lead signals for these regions
240  are listed on Table 2, for each of the traits analyzed (L*, a*, b* and delta). In order to facilitate
241  interpretation, we provide in the Table the P-values for all the traits. It can be seen that all the
242  regions show nominal significance in more than one iris color dimension. Additional information
243  about relevant markers within each region is provided in the Supplementary Excel File. The
244  regional plots are depicted in Supplementary Figures 8, 9, 10 and 11, respectively.

245  *Replication of suggestive skin pigmentation signals.*

246    The lead SNPs showing suggestive evidence of association ($p<10^{-5}$) were followed up in three
247    independent Chinese samples, for which skin pigmentation measures in the buttocks or inner arm
248    were available (M-index or L* values). The results of the replication effort are shown in Table 3.
249    One of the markers, rs2373391 located on the *ZNF804B* gene on chromosome 7, was replicated
250    in the Physical Examination cohort. In this sample, the A allele was significantly associated with
251    lighter skin pigmentation (p=0.003, Bonferroni-corrected significance threshold p=0.01). This
252    marker also had a concordant effect on the College student sample (A allele associated with
253    lighter skin pigmentation –higher L* values), but the p-values were not nominally significant. In
254    the Taizhou cohort, in which melanin values were measured in the upper arm, rs2373391 was
255    also not significant and the regression coefficient was close to 0. None of the remaining markers
256    was nominally significant in any of the three replication samples.

257    *Association results for loci identified in European populations.*

258    We explored in our East Asian sample the allele frequencies and effect sizes of loci identified in
259    European populations. These results are depicted in Supplementary Table 1. The table includes
260    two non-synonymous variants in the genes *SLC24A5* (rs1426654) and *SLC45A2* (rs16891982),
261    which have very high frequencies in Europe and very strong effects on skin pigmentation in
262    admixed groups (Lamason et al., 2005; Norton et al., 2007). The table also includes variants that
263    reached genome-wide significance in the largest GWAS of skin pigmentation in populations of
264    European ancestry, which included more than 17,000 samples (Liu et al. 2015). Five of the six
265    variants are either absent or present at very low frequencies in the East Asian sample and have
266    relatively low imputation scores (info < 0.8). The only exception is the polymorphism rs4268748
267    located near the *MC1R* gene, in which the minor allele (allele C) has a frequency of 27.4% and a
268    good imputation score (info = 0. 983). None of the variants reached nominal significance in our
269    East Asian sample, except the *SLC24A5* rs1426654 non-synonymous variant, which is present at
270    a very low frequency in the sample (frequency of the derived A allele 1.6%).

271

**272    Discussion**

273    In this paper, we describe the results of a genome-wide association study of pigmentary traits
274    (skin pigmentation and iris color) in East Asian populations. We were able to confirm the
275    important role that the gene *OCA2* plays in East Asian populations. In our iris color GWA, we
276    observed a genome-wide significant signal in the *OCA2* region. The lead SNP in this region was
277    rs76930569, and this marker showed particularly strong associations with the a* and b* color
278    dimensions (p-values of $9.27 \times 10^{-11}$ and $5.30 \times 10^{-12}$, respectively). However, rs76930569 was not
279    associated with delta, which is a measure that captures iris heterochromia (e.g. the color
280    differences between the pupillary and ciliary regions of the iris). This indicates that the genetic
281    architecture of iris color is different from the genetic architecture of iris heterochromia. It is
282    important to note that rs76930569 is in very strong linkage disequilibrium with the non-
283    synonymous SNP rs1800414 ($R^2$=1 in CHB, CHS and JPT 1KG samples). Not surprisingly, the
284    p-values of rs1800414 were very similar to those of rs76930569 in our sample, and most
285    probably this non-synonymous SNP is the causal polymorphism driving the association. A recent

286  study (Yang et al., 2016) using cultured melanocytes, and transgenic and targeted gene
287  modification analyses on zebrafish and mouse have shown that the rs1800414 G variant
288  (Arg615) is functional and leads to skin lightening. In our skin pigmentation analysis, the
289  rs1800414 G variant was also associated with lighter skin pigmentation, but did not reach
290  genome-wide significance (beta=-0.81, p=5.5x10$^{-4}$). The magnitude of the skin pigmentation
291  effect observed in our study is quite similar to the effects described in previous reports in which
292  the M-index was used to describe constitutive pigmentation (e.g. Edwards et al., 2010, beta=-
293  1.26 in a sample of individuals of East Asian ancestry living in Canada, and beta=--0.86 in a Han
294  Chinese sample; Eaton et al. 2015, beta=-0.91 in a sample of individuals of East Asian ancestry
295  living in Canada).

296  Aside from the eye color signal in the *OCA2* region, no other regions surpassed the genome-wide
297  significance threshold for skin pigmentation or iris color. However, there were several regions
298  harboring multiple common markers with good imputation scores and suggestive p-values (p<10$^{-5}$
299  , Table 1 and Table 2; Supplementary Excel File). For skin pigmentation, we followed up the
300  lead SNPs for each region in three independent Chinese samples (Table 3). One of the SNPs,
301  rs2373391, which is located in an intron of the gene *ZNF804B*, replicated in one of the Chinese
302  samples (beta=-1.132, p=0.003).  The gene *ZNF804B* encodes a zinc finger protein, but it has
303  been poorly characterized. Variants within this gene have been nominally associated with a
304  number of traits in previous GWA studies (Anorexia nervosa, Wang et al., 2011; Heschl's gyrus
305  morphology, Cai et al., 2014; and IgG glycosylation, Lauc et al., 2013, among others), but to our
306  knowledge, no associations with pigmentary traits have been reported for this or nearby genes.
307  Given that we were able to replicate the association only in one of the three East Asian samples,
308  it will be important to investigate the potential role of this region in skin pigmentation in larger
309  samples from East Asia. None of the other regions showing suggestive associations with skin
310  pigmentation were replicated in the follow-up samples.

311  Unfortunately, we could not follow up any of the suggestive regions identified for iris color,
312  because of the lack of studies in East Asia reporting quantitative measures of iris color. In this
313  study, we show that there is a substantial amount of variation in iris color in East Asia
314  (Supplementary Figure 2), and that the *OCA2* region (and more particularly, the non-
315  synonymous variant rs1800414) is a major determinant of the variation observed. We estimated
316  that rs1800414 explains 11.9%, 10.4% and 6% of the variation observed in the b*, a* and L*
317  coordinates in our sample, respectively. Therefore, there is a substantial amount of variation that
318  remains unexplained in our sample. This highlights the need to carry out more studies using
319  quantitative measures of iris color in East Asia. Categorical definitions of iris color would not be
320  useful in East Asian populations, because in such a classification most of the irises would have
321  been categorized as "brown", but there is a substantial range of L*, a* and b* values within this
322  category.

323  To our knowledge, this is the first genome-wide association study of pigmentary traits in East
324  Asian populations. The main weakness of this study is the small sample size, which limits our
325  statistical power to identify variants with small effects on skin and iris pigmentation. Another
326  limitation is that constitutive pigmentation was measured in different body sites (e.g. buttocks or

327   inner arm) in the discovery and replication samples. Ideally, constitutive pigmentation should be
328   measured in the buttocks, because there is less exposure to UV in this body site. However, often
329   it is not possible to obtain pigmentation measures in the buttocks, and constitutive pigmentation
330   is measured in the inner upper arm. In this site, there may be some exposure to UV light,
331   particularly at certain times of the year, and several studies have reported that measurements of
332   pigmentation in the inner arm are not interchangeable with estimates in the buttocks (Johansen et
333   al. 2016; Bieliauskiene et al. 2017). In the discovery sample from Canada, inner upper arm
334   measures were collected either in the winter/early spring (January to April) or the fall (October
335   and November), so it would not be expected that UVR exposure would have a substantial effect
336   on melanin levels. Similarly, in the two replication samples in which pigmentation measures
337   were taken in the inner arm, Taizhou cohort and College students, the measures were collected in
338   April, and September, respectively. Additionally, it is important to note that in one of the
339   replication samples, pigmentation was reported as the CIELAB L* value, instead of the M-index.
340   Although these two measures are highly correlated (Shriver and Parra, 2000), ideally all the
341   discovery and replication samples should have been measured using the same index. Given that
342   measurements were taken in different body sites, and for one replication sample using different
343   pigmentation units, we could not carry out a meta-analysis of the discovery and replication
344   samples. Instead, in our replication effort, we evaluated if the SNPs that were followed up in the
345   replication samples reached significance after Bonferroni-correction based on the number of
346   independent tests. Finally, it is important to note that for this GWAS we used both directly
347   genotyped and imputed data. This is the standard protocol used in this type of studies, because it
348   increases the power to identify associated variants (Spencer et al., 2009; Marchini and Howie,
349   2010). We implemented two different strategies to minimize potential problems related to
350   imputation: 1) We used tests implemented in the program SNPTEST that take into account
351   uncertainty in the imputed genotypes, and 2) We only followed up markers with very high
352   imputation scores (info > 0.8).

353   In spite of these limitations, this study provides important insights about the genetic architecture
354   of skin and iris color in East Asian populations. Using the program Quanto
355   (http://biostats.usc.edu/Quanto.html) we estimated that our skin pigmentation association study
356   (N=305) had good power (>0.8) to identify variants explaining approximately 12.5% of the
357   variance of the trait. This implies that there are no polymorphic variants in our samples with
358   large effects on skin melanin levels. For iris color (N=342), our study had good power to identify
359   variants explaining approximately 11% of the variance of the trait, and in fact we identified a
360   SNP with an effect of approximately that size. It will be critical to expand the number of
361   genome-wide association studies in East Asian populations in order to be able to identify genetic
362   markers with smaller effects on skin and iris pigmentation. We cannot exclude the possibility
363   that there are variants with strong effects on pigmentation that have been fixed in East Asian
364   populations due to the action of positive selection.  This is something that happened in Europe
365   with the well-known *SLC24A5* rs1426654 A allele, which is the variant with the largest effect on
366   melanin levels reported in human populations (Lamason et al., 2005). Identifying this type of
367   variants, if they exist in East Asian populations, would require association studies in admixed
368   samples with a substantial East Asian contribution.

369   We followed up in our East Asian sample genetic markers in 6 loci that have been reported to
370   influence skin pigmentation in European populations (*SLC45A5, IRF4, HERC2, SLC24A5,*
371   *DEF8/MC1R* and *RALY/ASIP,* Supplementary Table 1). Most of these variants are absent
372   (*RALY/ASIP* rs6059655) or present at very low frequencies in the East Asian sample (*SLC45A2*
373   rs16891982, *IRF4* rs12203592, *HERC2* rs12913832 and *SLC24A5* rs4268748, Supplementary
374   Table 1). Because of their low frequencies, these SNPs explain little variation of skin
375   pigmentation in East Asian populations. However, it is important to note that the estimated effect
376   size of the derived *SLC24A5* rs1426654 A allele in our GWAS is quite large (beta=-2.46), in
377   agreement with numerous reports indicating that this is the locus with the strongest effect on skin
378   pigmentation so far described in human populations (Lamason et al. 2005; Basu Mallick et al.
379   2013; Beleza et al., 2013). Only one of the variants reaching genome-wide significance in a large
380   GWAS study in a European sample (*DEF8/MC1R* rs4268748; Liu et al., 2015) is present with
381   relatively high frequencies in our East Asian sample. In fact, the allele reported to decrease
382   pigmentation in the European sample (C allele) is slightly more frequent in East Asian
383   populations than in European populations (28.5% in EAS 1000 Genomes Project sample vs.
384   22.8% in EUR 1000 Genomes project). However, the rs4268748 polymorphism is not nominally
385   significant in our sample. One potential explanation is that rs4268748 is in linkage
386   disequilibrium (LD) with causal loci in Europe, and the pattern of LD between rs4268748 and
387   the causal loci, or the allele frequencies of the causal loci are different in European and East
388   Asian populations. Supporting this explanation is the detailed analysis of the *MC1R* region
389   reported for one of the European samples (Rotterdam study) included in Liu et al. (2015)
390   GWAS. These authors described that 3 known high-penetrance variants located within the *MC1R*
391   gene (rs1805007 (R151C), rs1805008 (R160W), rs1805009 (D294H)) showed the most
392   significant association with skin pigmentation in the Rotterdam study sample, and replacing the
393   SNP rs4268748 by the high penetrance variants marginally increased the amount of the
394   phenotypic variance explained. These three *MC1R* high-penetrance variants, two of which have
395   frequencies higher than 5% in Europe (rs1805007 and rs1805008), are not found or are found in
396   extremely low frequencies in East Asia, and this may explain the absence of an association signal
397   in our study. Yamaguchi et al. (2012) reported that a low-penetrance *MC1R* variant, rs885479
398   (R163Q), which is common in East Asia (frequency derived 163Q allele in EAS 1000 Genomes
399   project sample = 61.6%) but much less frequent in Europe was nominally associated with light
400   skin in a Japanese sample. As expected, the rs885479 polymorphism was present in our East
401   Asian sample, with a frequency of the derived 163Q allele of 60.1%. However, this variant was
402   not nominally associated with skin pigmentation in our sample (beta = -0.046, p= 0.843).

403   Overall, our study confirm previous research indicating that the evolution of light skin in East
404   Asia and Europe took place, at least to a large extent, independently in both groups after the split
405   of the ancestral East Asian and European populations following the Out-of-Africa migration. In
406   Europe, functional derived variants in the genes *SLC24A5, SLC45A2* and *OCA2/HERC2*
407   increased in frequency as a result of positive selection (Lamason et al., 2005; Izaguirre et al.,
408   2006; Norton et al., 2007; Lao et al., 2007; Wilde et al., 2014). In East Asia, the non-
409   synonymous polymorphism rs1800414 located within the *OCA2* gene dramatically increased in
410   frequency, also as a result of an independent selective event (Lao et al. 2007; Edwards et al.,
411   2010; Donnelly et al., 2012; Hider et al., 2013). This constitutes, with lactase persistence and

412 adaptation to altitude, one of the most fascinating examples of convergent evolution in human
413 populations.

414

415

**Conclusion**

417 Here we report the first genome-wide association of pigmentary traits (skin pigmentation and iris
418 color) in East Asian populations. We measured these traits using quantitative methods, and we
419 show that there is considerable variation not only in skin pigmentation but also in iris color, in
420 spite of the fact that using categorical classifications, all the irises in this sample would have
421 been categorized as "brown". We confirmed the important role that the *OCA2* gene plays in
422 normal pigmentation variation in East Asian populations. In particular, the non-synonymous
423 variant rs1800414 explains a substantial amount of variation in iris color. We did not observe
424 any genome-wide significant variant for skin pigmentation, but one of the lead SNPs showing
425 suggestive significance in our study, rs2373391, which is located in an intron of the ZNF804B
426 gene on chromosome 7, was replicated in an independent Chinese sample (p=0.003). It will be
427 critical to carry out additional association studies in East Asian populations in order to uncover
428 additional variants with smaller effects on skin pigmentation and iris color.

429

**Acknowledgments**

431 We would like to thank all the individuals that participated in this study.

432

**References**

434 Abe Y, Tamiya G, Nakamura T, Hozumi Y, and Suzuki T. 2013. Association of melanogenesis genes
435        with    skin color variation among Japanese females. *Journal of Dermatological Science* 69:167-
436        172. 10.1016/j.jdermsci.2012.10.016

437 Andersen JD, Johansen P, Harder S, Christoffersen SR, Delgado MC, Henriksen ST, Nielsen MM,
438        Sørensen E, Ullum H, Hansen T, Dahl AL, Paulsen RR, Børsting C, Morling N. 2013. Genetic
439        analyses of the human eye colours using a novel objective method for eye colour classification.
440        *Forensic Science International* 7:508-515. 10.1016/j.fsigen.2013.05.003

441 Basu Mallick C, Iliescu FM, Möls M, Hill S, Tamang R, Chaubey G, Goto R, Ho SY, Gallego Romero I,
442        Crivellaro F, Hudjashov G, Rai N, Metspalu M, Mascie-Taylor CG, Pitchappan R, Singh L,
443        Mirazon-Lahr M, Thangaraj K, Villems R, Kivisild T. 2013. The light skin allele of SLC24A5 in
444        South Asians and Europeans shares identity by descent. *PLoS Genetics* 9:e1003912.
445        10.1371/journal.pgen.1003912

446 Beleza S, Santos AM, McEvoy B, Alves I, Martinho C, Cameron E, Shriver MD, Parra EJ, and Rocha J.
447        2013. The Timing of Pigmentation Lightening in Europeans. *Molecular Biology and*
448        *Evolution* 30:24-35. 10.1093/molbev/mss207

449   Beleza S, Johnson NA, Candille SI, Absher DM, Coram MA, Lopes J, Campos J, Araújo II, Anderson
450       TM, Vilhjálmsson BJ, Nordborg M, Correia E Silva A, Shriver MD, Rocha J, Barsh GS, Tang H.
451       2013. Genetic architecture of skin and eye color in an African-European admixed population.
452       *PLoS Genetics* 9:e1003372. 10.1371/journal.pgen.1003372

453   Bieliauskiene G, Philipsen PA, Wulf HC. 2017. Can constitutive pigmentation be measured on upper
454       inner arm? Correlation between arm and buttocks pigmentation. *Photodermatology,*
455       *Photoimmunology and Photomedicine* 33:233-236. 10.1111/phpp.12314

456   Cai DC, Fonteijn H, Guadalupe T, Zwiers M, Wittfeld K, Teumer A, Hoogman M, Arias-Vasquez A,
457       Yang Y, Buitelaar J, Fernandez G, Brunner HG, van Bokhoven H, Franke B, Hegenscheid K,
458       Homuth G, Fisher SE, Grabe HJ, Francks C, and Hagoort P. 2014. A genome-wide search for
459       quantitative trait loci affecting the cortical surface area and thickness of Heschl's gyrus. *Genes*
460       *Brain and Behavior* 13:675-685. 10.1111/gbb.12157

461   Chen H, Hey J, and Slatkin M. 2015. A hidden Markov model for investigating recent positive selection
462       through haplotype structure. *Theoretical Population Biology* 99:18-30. 10.1016/j.tpb.2014.11.001

463   Delaneau O, Marchini J, and Zagury JF. 2012. A linear complexity phasing method for thousands of
464       genomes. *Nature Methods* 9:179-181. 10.1038/nmeth.1785

465   Donnelly MP, Paschou P, Grigorenko E, Gurwitz D, Barta C, Lu RB, Zhukova OV, Kim JJ, Siniscalco
466       M, New M, Li H, Kajuna SL, Manolopoulos VG, Speed WC, Pakstis AJ, Kidd JR, Kidd KK.
467       2012. A global view of the OCA2-HERC2 region and pigmentation. *Human Genetics* 131:683-
468       696. 10.1007/s00439-011-1110-x

469   Durbin R. 2014. Efficient haplotype matching and storage using the positional Burrows-Wheeler
470       transform (PBWT). *Bioinformatics* 30:1266-1272. 10.1093/bioinformatics/btu014

471   Eaton K, Edwards M, Krithika S, Cook G, Norton H, and Parra EJ. 2015. Association study confirms the
472       role of two OCA2 polymorphisms in normal skin pigmentation variation in East Asian
473       populations. *American Journal of Human Biology* 27:520-525. 10.1002/ajhb.22678

474   Edwards M, Bigham A, Tan JZ, Li SL, Gozdzik A, Ross K, Jin L, and Parra EJ. 2010. Association of the
475       OCA2 Polymorphism His615Arg with Melanin Content in East Asian Populations: Further
476       Evidence of Convergent Evolution of Skin Pigmentation. *Plos Genetics* 6.
477       10.1371/journal.pgen.1000867

478   Edwards M, Cha D, Krithika S, Johnson M, Cook G, and Parra EJ. 2016. Iris pigmentation as a
479       quantitative trait: variation in populations of European, East Asian and South Asian ancestry and
480       association with candidate gene polymorphisms. *Pigment Cell & Melanoma Research* 29:141-
481       162. 10.1111/pcmr.12435

482   Edwards M, Gozdzik A, Ross K, Miles J, and Parra EJ. 2012. Technical note: Quantitative measures of
483       iris color using high resolution photographs. *American Journal of Physical*
484       *Anthropology* 147:141-149. 10.1002/ajpa.21637

485   Hider JL, Gittelman RM, Shah T, Edwards M, Rosenbloom A, Akey JM, Parra EJ. 2013. Exploring
486       signatures of positive selection in pigmentation candidate genes in populations of East Asian
487       ancestry. *BMC Evolutionary Biology* 13:150. 10.1186/1471-2148-13-150

488    Izagirre N, García I, Junquera C, de la Rúa C, Alonso S. 2006. A scan for signatures of positive selection
489          in candidate loci for skin pigmentation in humans. *Molecular Biology and Evolution* 23:1697-
490          1706.

491    Johansen P, Andersen JD, Madsen LN, Ullum H, Glud M, Børsting C, Gniadecki R, Morling N. 2016.
492          Pigmentary Markers in Danes--Associations with Quantitative Skin Colour, Nevi Count, Familial
493          Atypical Multiple-Mole, and Melanoma Syndrome. *PLoS One* 11:e0150381.
494          10.1371/journal.pone.0150381

495    Kayser M, Liu F, Janssens A, Rivadeneira F, Lao O, van Duijn K, Vermeulen M, Arp P, Jhamai MM, van
496          Ijcken WFJ, den Dunnen JT, Heath S, Zelenika D, Despriet DDG, Klaver CCW, Vingerling JR,
497          De Jong P, Hofman A, Aulchenko YS, Uitterlinden AG, Oostra BA, and van Duijn CM. 2008.
498          Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris
499          color gene. *American Journal of Human Genetics* 82:411-423. 10.1016/j.ajhg.2007.10.003

500    Lamason RL, Mohideen M, Mest JR, Wong AC, Norton HL, Aros MC, Jurynec MJ, Mao XY,
501          Humphreville VR, Humbert JE, Sinha S, Moore JL, Jagadeeswaran P, Zhao W, Ning G,
502          Makalowska I, McKeigue PM, O'Donnell D, Kittles R, Parra EJ, Mangini NJ, Grunwald DJ,
503          Shriver MD, Canfield VA, and Cheng KC. 2005. SLC24A5, a putative cation exchanger, affects
504          pigmentation in zebrafish and humans. *Science* 310:1782-1786. 10.1126/science.1116238

505    Lao O, de Gruijter JM, van Duijn K, Navarro A, Kayser M. 2007. Signatures of positive selection in
506          genes associated with human skin pigmentation as revealed from analyses of single nucleotide
507          polymorphisms. *Annals of Human Genetics* 71:354-369.

508    Lauc G, Huffman JE, Pucic M, Zgaga L, Adamczyk B, Muzinic A, Novokmet M, Polasek O, Gornik O,
509          Kristic J, Keser T, Vitart V, Scheijen B, Uh HW, Molokhia M, Patrick AL, McKeigue P, Kolcic
510          I, Lukic IK, Swann O, van Leeuwen FN, Ruhaak LR, Houwing-Duistermaat JJ, Slagboom PE,
511          Beekman M, de Craen AJM, Deelder AM, Zeng Q, Wang W, Hastie ND, Gyllensten U, Wilson
512          JF, Wuhrer M, Wright AF, Rudd PM, Hayward C, Aulchenko Y, Campbell H, and Rudan I.
513          2013. Loci Associated with N-Glycosylation of Human Immunoglobulin G Show Pleiotropy with
514          Autoimmune Diseases and Haematological Cancers. *Plos Genetics* 9.
515          10.1371/journal.pgen.1003225

516    Liu F, Visser M, Duffy DL, Hysi PG, Jacobs LC, Lao O, Zhong KY, Walsh S, Chaitanya L, Wollstein A,
517          Zhu G, Montgomery GW, Henders AK, Mangino M, Glass D, Bataille V, Sturm RA, Rivadeneira
518          F, Hofman A, van Ijcken WFJ, Uitterlinden AG, Palstra R, Spector TD, Martin NG, Nijsten TEC,
519          and Kayser M. 2015. Genetics of skin color variation in Europeans: genome-wide association
520          studies with functional follow-up. *Human Genetics* 134:823-835. 10.1007/s00439-015-1559-0

521    Liu F, Wollstein A, Hysi PG, Ankra-Badu GA, Spector TD, Park D, Zhu G, Larsson M, Duffy DL,
522          Montgomery GW, Mackey DA, Walsh S, Lao O, Hofman A, Rivadeneira F, Vingerling JR,
523          Uitterlinden AG, Martin NG, Hammond CJ, and Kayser M. 2010. Digital Quantification of
524          Human Eye Color Highlights Genetic Association of Three New Loci. *Plos Genetics* 6.
525          10.1371/journal.pgen.1000934

526    Marchini J, and Howie B. 2010. Genotype imputation for genome-wide association studies. *Nature*
527          *Reviews Genetics* 11:499-511. 10.1038/nrg2796

528    Murray N, Norton H, and Parra EJ. 2015. Distribution of two OCA2 polymorphisms associated with
529          pigmentation in East Asian populations. *Human Genome Variation* 2:15058

530  Norton HL1, Kittles RA, Parra E, McKeigue P, Mao X, Cheng K, Canfield VA, Bradley DG, McEvoy B,
531      Shriver MD. **2007**. Genetic evidence for the convergent evolution of light skin in Europeans and
532      East Asians. *Molecular Biology and Evolution* 24:710-722.

533  Norton HL, Edwards M, Krithika S, Johnson M, Werren EA, and Parra EJ. 2016. Quantitative
534      Assessment of Skin, Hair, and Iris Variation in a Diverse Sample of Individuals and Associated
535      Genetic Variation. *American Journal of Physical Anthropology* 160:570-581. 10.1002/ajpa.22861

536  Shriver MD, Parra EJ. 2000. Comparison of narrow-band reflectance spectroscopy and tristimulus
537      colorimetry for measurements of skin and hair color in persons of different biological ancestry.
538      *American Journal of Physical Anthropology* 112:17-27.

539  Spencer CC, Su Z, Donnelly P, Marchini J. 2009. Designing genome-wide association studies: sample
540      size, power, imputation, and the choice of genotyping chip. *PLoS Genetics* 5:e1000477.
541      10.1371/journal.pgen.1000477

542  Visser M, Kayser M, and Palstra RJ. 2012. HERC2 rs12913832 modulates human pigmentation by
543      attenuating chromatin-loop formation between a long-range enhancer and the OCA2
544      promoter. *Genome Research* 22:446-455. 10.1101/gr.128652.111

545  Walsh S, Lindenbergh A, Zuniga SB, Sijen T, de Knijff P, Kayser M, and Ballantyne KN. 2011.
546      Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for
547      forensic intelligence. *Forensic Science International-Genetics* 5:464-471.
548      10.1016/j.fsigen.2010.09.008

549  Wang K, Zhang H, Bloss CS, Duvvuri V, Kaye W, Schork NJ, Berrettini W, Hakonarson H, and Price
550      Fdn Collaborative G. 2011. A genome-wide association study on common SNPs and rare CNVs
551      in anorexia nervosa. *Molecular Psychiatry* 16:949-959. 10.1038/mp.2010.107

552  Wilde S, Timpson A, Kirsanow K, Kaiser E, Kayser M, Unterländer M, Hollfelder N, Potekhina ID,
553      Schier W, Thomas MG, Burger J. 2014. Direct evidence for positive selection of skin, hair, and
554      eye pigmentation in Europeans during the last 5,000 y. *Proceedings of the National Academy of
555      Sciences USA* 111:4832-4837. 10.1073/pnas.1316513111

556  Wollstein A, Walsh S, Liu F, Chakravarthy U, Rahu M, Seland JH, Soubrane G, Tomazzoli L, Topouzis
557      F, Vingerling JR, Vioque J, Bohringer S, Fletcher AE, and Kayser M. 2017. Novel quantitative
558      pigmentation phenotyping enhances genetic association, epistasis, and prediction of human eye
559      colour. *Scientific Reports* 7. 10.1038/srep43359

560  Yamaguchi K, Watanabe C, Kawaguchi A, Sato T, Naka I, Shindo M, Moromizato K, Aoki K, Ishida H,
561      and Kimura R. 2012. Association of melanocortin 1 receptor gene (MC1R) polymorphisms with
562      skin reflectance and freckles in Japanese. *Journal of Human Genetics* 57:700-708.
563      10.1038/jhg.2012.96

564  Yang ZH, Zhong H, Chen J, Zhang XM, Zhang H, Luo X, Xu SH, Chen H, Lu DS, Han YL, Li JK, Fu
565      LJ, Qi XB, Peng Y, Xiang K, Lin Q, Guo Y, Li M, Cao XY, Zhang YF, Liao SY, Peng YM,
566      Zhang L, Guo XS, Dong SS, Liang F, Wang J, Willden A, Aun HS, Serey B, Sovannary T,
567      Bunnath L, Samnom H, Mardon G, Li QW, Meng AM, Shi H, and Su B. 2016. A Genetic
568      Mechanism for Convergent Skin Lightening during Recent Human Evolution.*Molecular Biology
569      and Evolution* 33:1177-1187. 10.1093/molbev/msw003

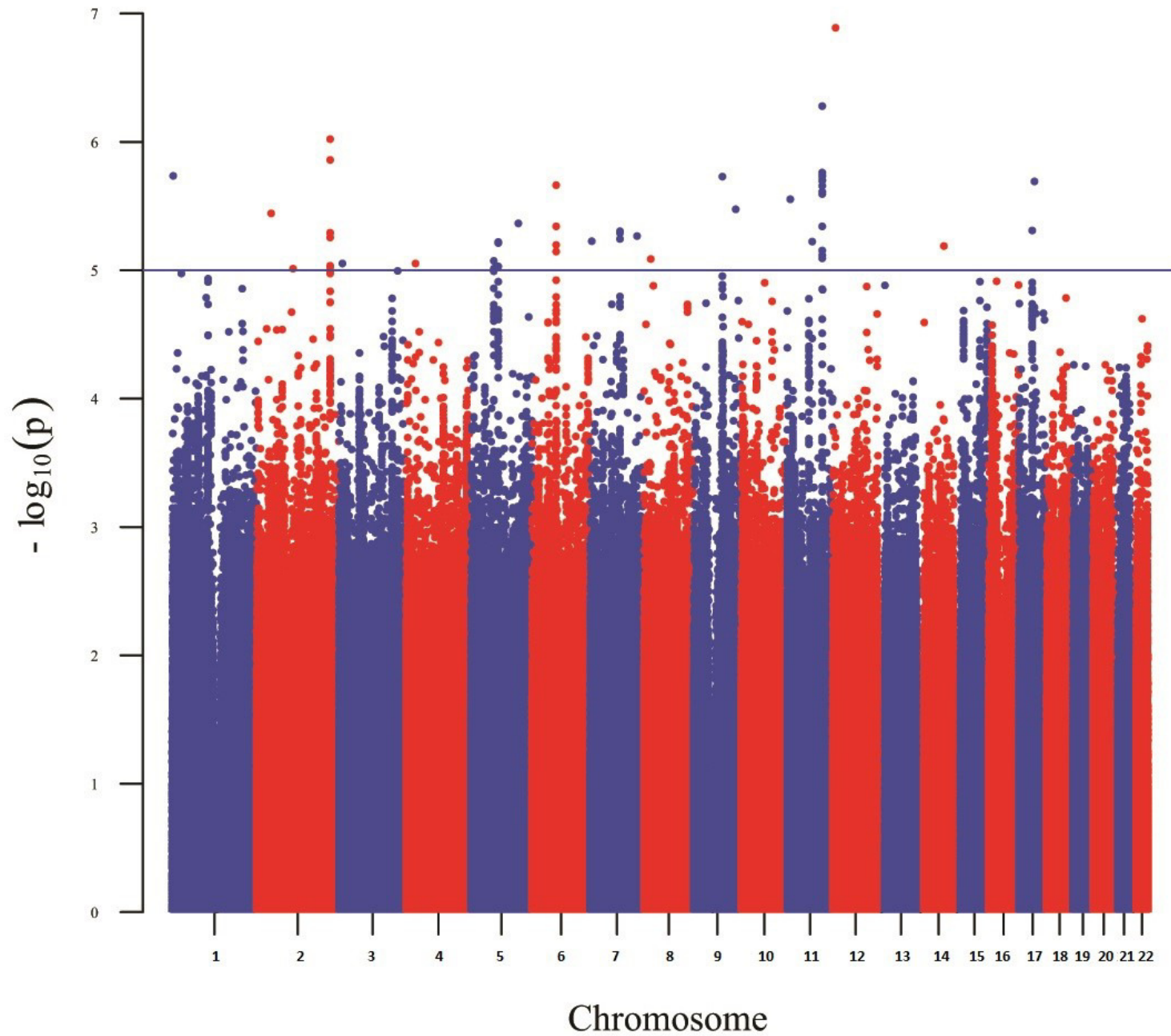
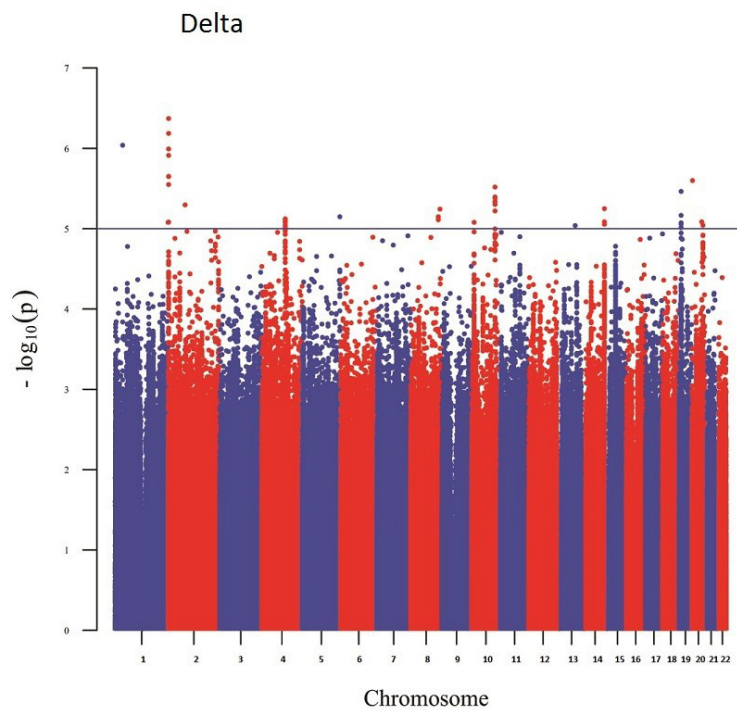
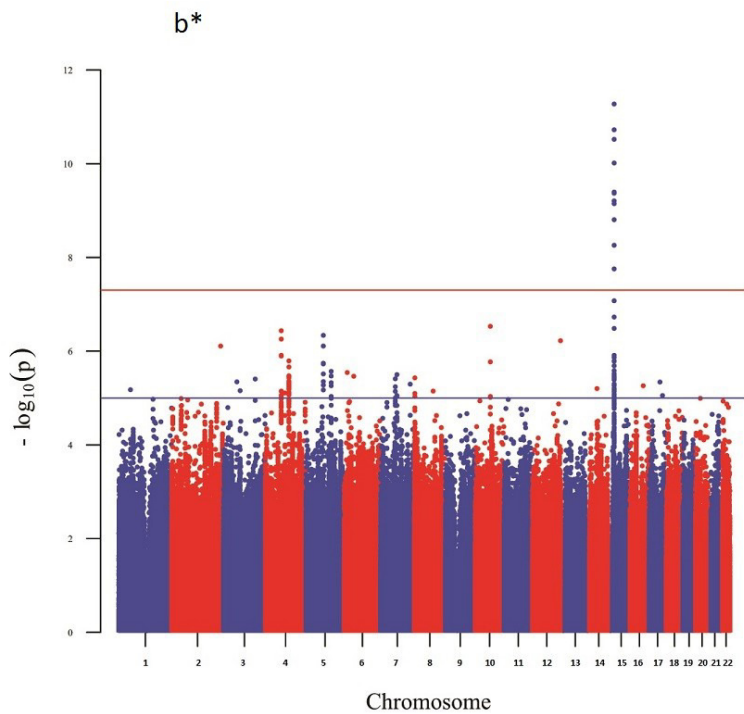
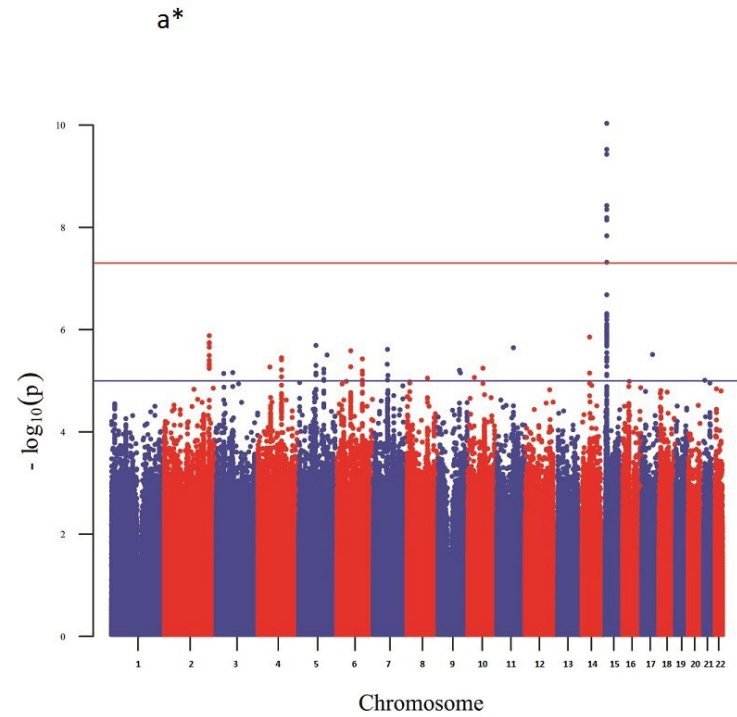
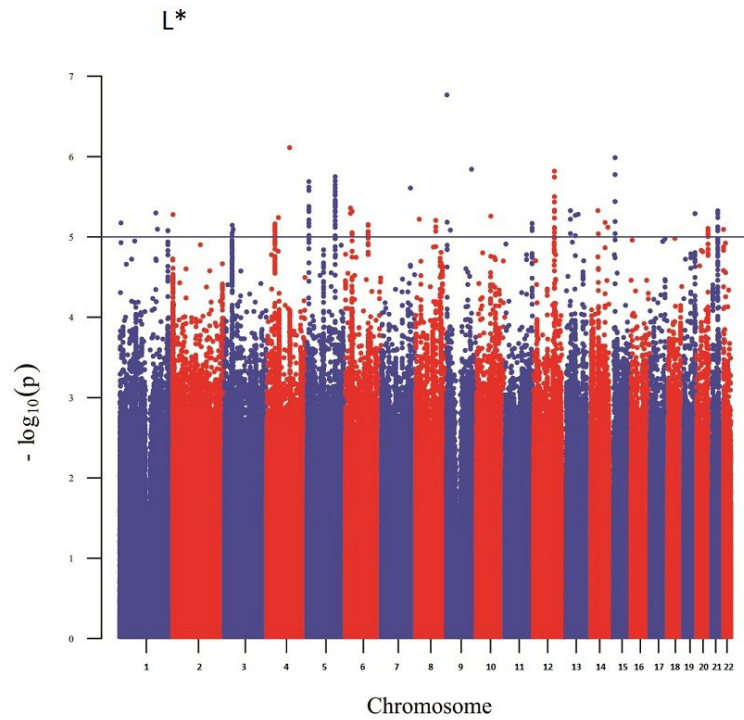
570

571

572

573

574
575
576
577

Manhattan plot reporting the association results for skin pigmentation.

Manhattan plot reporting the association results for iris color (L*, a*, b* and delta)

Main signals observed in the GWAS of skin pigmentation (M-values). We report the beta, SE and P-values obtained with the model using the standardized and unstandardized residuals.

1   Table 1. Main signals observed in the GWAS of skin pigmentation (M-values). We report the beta, SE and P-values obtained with the model using

2   the standardized and unstandardized residuals.

| SNP | CHR | POS | NEA/EA | Gene | Frq EA | INFO | Beta[1] | SE[1] | P[1] | Beta[2] | SE[2] | P[2] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rs2003589 | 2 | 217527465 | T/C | *IGFBP2* | 0.063 | 0.843 | 0.885 | 0.177 | 9.54E-07 | 2.464 | 0.494 | 1.05E-06 |
| rs853807 | 5 | 67752638 | C/T | | 0.485 | 0.956 | -0.347 | 0.077 | 8.45E-06 | -0.961 | 0.213 | 9.75E-06 |
| rs57836066 | 6 | 71304950 | T/G | *RP11-134K13.4* | 0.086 | 0.979 | -0.661 | 0.137 | 2.17E-06 | -1.838 | 0.382 | 2.40E-06 |
| rs2373391 | 7 | 88449300 | T/A | *ZNF804B* | 0.681 | 0.987 | -0.384 | 0.083 | 4.98E-06 | -1.076 | 0.232 | 5.09E-06 |
| rs7945369 | 11 | 103425586 | T/C | | 0.537 | 0.968 | -0.405 | 0.079 | 5.26E-07 | -1.159 | 0.224 | 3.97E-07 |

4   [1]Beta estimate using standardized residuals

5   [2]Beta estimate using unstandardized residuals

27
28
29
30
31
32
33

Main signals observed in the GWAS of iris color (L*, a*, b* and delta).

For each trait, we report the P-values obtained for all the other traits in the association study, as well as the log10 Bayes Factor obtained in the joint analysis of L*,a* and b*.

1    Table 2. Main signals observed in the GWAS of iris color (L*, a*, b* and delta). For each trait, we report the P-values obtained for all the other

2    traits in the association study, as well as the log10 Bayes Factor obtained in the joint analysis of L*,a* and b*.

| | CHR | POS | NEA/EA | *Gene* | Frq EA | INFO | Beta | SE | P(L*) | P(a*) | P(b*) | P(delta) | BF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **L\*signals** | | | | | | | | | | | | | |
| rs6664080 | 1 | 226797426 | T/C | C1orf95 | 0.297 | 0.993 | 0.362 | 0.08 | 8.36E-06 | 2.56E-03 | 2.88E-02 | 3.42E-03 | 2.774 |
| rs4425211 | 3 | 36951898 | A/G | TRANK1 | 0.072 | 0.959 | -0.676 | 0.148 | 7.14E-06 | 6.79E-05 | 6.33E-04 | 5.85E-01 | 1.678 |
| rs12501370 | 4 | 41043870 | G/C | APBB2 | 0.556 | 0.982 | -0.332 | 0.073 | 6.89E-06 | 3.14E-04 | 1.93E-03 | 7.97E-01 | 2.501 |
| rs2658084 | 5 | 10136135 | C/T | CTD-2199O4.1 | 0.835 | 0.97 | -0.502 | 0.104 | 2.05E-06 | 4.11E-03 | 3.32E-02 | 2.29E-01 | 2.976 |
| rs113633047 | 5 | 133402895 | G/T | | 0.282 | 0.95 | 0.419 | 0.087 | 2.00E-06 | 2.25E-04 | 5.94E-04 | 1.09E-01 | 2.735 |
| rs6924266 | 6 | 25505617 | C/T | LRRC16A | 0.233 | 0.98 | -0.408 | 0.087 | 4.38E-06 | 3.64E-05 | 1.05E-04 | 8.77E-02 | 2.583 |
| NA | 6 | 32626040 | C/T | | 0.075 | 0.996 | -0.64 | 0.138 | 4.81E-06 | 6.29E-05 | 1.79E-03 | 2.07E-01 | 2.003 |
| rs9373973 | 6 | 108079595 | T/A | SCML4 | 0.185 | 0.938 | -0.454 | 0.099 | 7.02E-06 | 2.19E-03 | 1.12E-02 | 9.11E-02 | 2.36 |
| rs145048184 | 8 | 97070341 | C/T | | 0.084 | 0.962 | 0.643 | 0.14 | 6.23E-06 | 8.89E-06 | 3.05E-05 | 5.74E-01 | 1.947 |
| rs610106 | 11 | 128935028 | T/C | ARHGAP32 | 0.814 | 0.97 | 0.44 | 0.096 | 6.79E-06 | 9.91E-04 | 5.40E-03 | 9.39E-01 | 2.318 |
| rs249625 | 12 | 98185248 | A/G | | 0.584 | 0.979 | -0.364 | 0.077 | 3.16E-06 | 5.72E-03 | 3.77E-03 | 9.27E-03 | 3.352 |
| rs76930569 | 15 | 28196145 | C/T | OCA2 | 0.623 | 1 | 0.38 | 0.076 | 1.03E-06 | 9.27E-11 | 5.30E-12 | 3.42E-01 | 7.007 |
| rs75161997 | 20 | 55701691 | C/T | | 0.06 | 0.992 | -0.732 | 0.161 | 7.84E-06 | 7.49E-04 | 1.77E-03 | 8.34E-01 | 1.485 |
| rs8131065 | 21 | 38011676 | C/T | | 0.215 | 0.979 | -0.419 | 0.09 | 4.75E-06 | 1.49E-02 | 1.25E-01 | 4.43E-02 | 3.478 |
| **a\*signals** | | | | | | | | | | | | | |
| rs55821297 | 4 | 111399598 | A/G | ENPEP | 0.771 | 0.926 | 0.435 | 0.092 | 2.96E-04 | 3.55E-06 | 2.19E-06 | 4.86E-05 | 2.710 |
| rs72763726 | 5 | 82044846 | C/T | | 0.090 | 0.950 | 0.652 | 0.135 | 1.32E-04 | 2.04E-06 | 4.60E-07 | 7.57E-01 | 2.349 |
| rs330203 | 5 | 119016414 | C/G | CTC-507E12.1 | 0.500 | 0.996 | -0.343 | 0.075 | 4.05E-03 | 7.02E-06 | 3.40E-06 | 8.74E-01 | 2.997 |
| rs9345521 | 6 | 65511281 | C/A | EYS | 0.467 | 0.976 | 0.352 | 0.074 | 1.64E-04 | 2.59E-06 | 6.92E-05 | 5.27E-01 | 2.879 |
| rs78001527 | 6 | 120168240 | G/C | | 0.074 | 0.973 | -0.686 | 0.146 | 1.65E-03 | 3.74E-06 | 6.68E-05 | 9.24E-01 | 1.892 |
| rs6977845 | 7 | 67216370 | T/A | | 0.093 | 0.881 | 0.666 | 0.139 | 2.58E-04 | 2.42E-06 | 5.04E-05 | 7.82E-01 | 1.979 |
| rs76930569 | 15 | 28196145 | C/T | OCA2 | 0.623 | 1.000 | 0.497 | 0.074 | 1.03E-06 | **9.27E-11** | **5.30E-12** | 3.42E-01 | 7.007 |
| **b\*signals** | | | | | | | | | | | | | |
| rs12510870 | 4 | 74358277 | T/C | AFM | 0.190 | 0.994 | -0.496 | 0.096 | 1.06E-01 | 3.37E-04 | 3.68E-07 | 1.60E-01 | 4.614 |
| rs1996603 | 4 | 111378362 | G/A | ENPEP | 0.779 | 0.960 | 0.451 | 0.092 | 1.32E-03 | 1.22E-05 | 1.61E-06 | 1.60E-03 | 2.728 |
| rs72763726 | 5 | 82044846 | C/T | | 0.090 | 0.950 | 0.691 | 0.134 | 1.32E-04 | 2.04E-06 | 4.60E-07 | 7.57E-01 | 2.349 |
| rs330203 | 5 | 119016414 | C/G | CTC-507E12.1 | 0.500 | 0.996 | -0.355 | 0.075 | 4.05E-03 | 7.02E-06 | 3.40E-06 | 8.74E-01 | 2.997 |
| rs141034411 | 8 | 3098640 | C/A | CSMD1 | 0.413 | 0.935 | 0.353 | 0.075 | 9.47E-05 | 3.20E-05 | 3.69E-06 | 5.71E-01 | 2.843 |
| rs2278745 | 10 | 71152091 | T/C | HK1 | 0.602 | 0.966 | -0.395 | 0.075 | 3.05E-03 | 5.70E-06 | 2.95E-07 | 1.51E-01 | 3.668 |
| rs76930569 | 15 | 28196145 | C/T | OCA2 | 0.623 | 1.000 | 0.527 | 0.074 | 1.03E-06 | **9.27E-11** | **5.30E-12** | 3.42E-01 | 7.007 |
| **Delta signals** | | | | | | | | | | | | | |
| rs72776813 | 2 | 2792075 | G/A | | 0.211 | 0.980 | -0.481 | 0.093 | 1.18E-03 | 8.21E-02 | 7.15E-02 | 4.25E-07 | 1.226 |
| rs243946 | 4 | 111300362 | C/T | ENPEP | 0.481 | 0.984 | -0.339 | 0.075 | 1.52E-04 | 3.60E-03 | 1.99E-03 | 7.74E-06 | 1.716 |
| rs16904127 | 8 | 130628606 | G/A | CCDC26 | 0.469 | 0.904 | -0.359 | 0.079 | 2.30E-01 | 8.01E-01 | 5.78E-01 | 7.13E-06 | 0.132 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rs7914735 | 10 | 110441149 | C/T | | 0.537 | 0.988 | -0.356 | 0.075 | 6.46E-02 | 3.40E-01 | 3.54E-01 | 3.03E-06 | -0.227 |
| NA | 14 | 106875131 | A/G | | 0.522 | 0.884 | -0.374 | 0.081 | | | | 5.63E-06 | |
| rs11667379 | 19 | 9126468 | G/C | | 0.411 | 0.954 | 0.348 | 0.074 | 1.70E-02 | 3.44E-01 | 2.92E-01 | 3.45E-06 | 0.532 |

Results of the replication effort for skin pigmentation signals in three independent Chinese samples.

Note that in the Physical Examination and Taizhou cohorts melanin was reported as M-index (higher M-values indicate darker skin), whereas in the sample of College students melanin was reported as L* (higher L* values indicate lighter skin).

1    Table 3. Results of the replication effort for skin pigmentation signals in three independent Chinese samples. Note that in the Physical
2    Examination and Taizhou cohorts melanin was reported as M-index (higher M-values indicate darker skin), whereas in the sample of College
3    students melanin was reported as L* (higher L* values indicate lighter skin).

| SNP | CHR | POS | NEA/EA | This study | | Physical Examination Cohort | | Taizhou cohort | | College Students | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Site | | | | Upper arm | | Buttocks | | Upper arm | | Buttocks | | Upper arm | |
| index | | | | M-index | | M-index | | M-index | | L* | | L* | |
| | | | | *Beta* | *P-value* | *Beta* | *P-value* | *Beta* | *P-value* | *Beta* | *P-value* | *Beta* | *P-value* |
| rs2003589 | 2 | 217527465 | T/C | 2.464 | 1.05E-06 | -0.104 | 0.864 | 0.171 | 0.575 | -1.927 | 0.259 | -0.659 | 0.635 |
| rs853807 | 5 | 67752638 | C/T | -0.961 | 9.75E-06 | -0.243 | 0.440 | -0.050 | 0.755 | 0.085 | 0.744 | -0.018 | 0.931 |
| rs57836066 | 6 | 71304950 | T/G | -1.838 | 2.40E-06 | 0.605 | 0.225 | 0.213 | 0.394 | 0.995 | 0.317 | 0.175 | 0.828 |
| rs2373391 | 7 | 88449300 | T/A | -1.076 | 5.09E-06 | **-1.132** | **0.003** | 0.024 | 0.889 | 0.120 | 0.739 | 0.273 | 0.349 |
| rs7945369 | 11 | 103425586 | T/C | -1.159 | 3.97E-07 | -0.614 | 0.051 | 0.211 | 0.190 | NA | NA | NA | NA |