

DATA REPORT

Distribution of two *OCA2* polymorphisms associated with pigmentation in East-Asian populationsNicole Murray¹, Heather L Norton² and Esteban J Parra¹

Two *OCA2* polymorphisms (rs1800414 and rs74653330) have been associated with pigmentation in East Asians. We explored the distribution of these markers in a panel of samples from populations around the world. The derived allele of rs1800414 has high frequencies in a broad East-Asian region, whereas the derived allele of rs74653330 is primarily restricted to northern East Asia. Our data suggest that these polymorphisms may have been selected independently in different regions of East Asia.

Human Genome Variation (2015) 2, 15058; doi:10.1038/hgv.2015.58; published online 10 December 2015

Although pigmentation varies globally, it has been more thoroughly studied and is therefore better understood in European populations. This has led to a research gap, especially in East-Asian populations. The *OCA2* gene, which is thought to be responsible for maintaining pH levels within melanosomes,¹ has been shown to be under positive selection in both European and East-Asian populations.^{2,3} However, the variants and haplotypes favored by selection are different in each population.^{4–6} For example, a variant located within the *HERC2* gene is known to affect the expression of the nearby *OCA2* gene, and it is strongly associated with blue eyes in European populations.^{7–9} The *HERC2* rs12913832 allele associated with blue eyes has a high frequency in Europe but is not present in East-Asian populations.^{7–9} In addition, two non-synonymous polymorphisms, rs1800414 and rs74653330, have been associated with pigmentation in East Asians^{5,10,11} and are not found at high frequencies in any population outside of East Asia.¹² It has been suggested that the phenotype of lighter skin is a result of convergent evolution in Europe and East Asia.^{2,6,13}

Available population data indicate that the rs1800414 and rs74653330 polymorphisms show a distinct geographical distribution. The highest frequencies of the derived rs1800414 G allele are found in Japan, China and Korea, whereas the derived rs74653330 A allele has the highest frequencies in northern East Asia, including Mongolia.^{12,14}

In this report, we provide further data on the global distribution of rs1800414 and rs74653330, with a primary focus on the allelic frequencies observed in East Asia. Briefly, the two polymorphisms were genotyped in the Human Genome Diversity Project–Centre d'Étude du Polymorphisme Humain (HGDP–CEPH) samples (http://www.cephb.fr/en/hgdp_panel.php) by LCG Genomics (Beverly, MA, USA) by using KASP genotyping technology. The HGDP–CEPH panel includes samples for more than 1,000 individuals from 52 populations around the world. Supplementary Table 1 shows the allelic frequencies of both markers in the HGDP–CEPH panel. In agreement with previous data, both polymorphisms are primarily restricted to East-Asian populations. The derived rs1800414 G allele has a broad distribution in East Asia, with the highest frequencies observed in the Japanese population (79%) and several populations from China

(Dai, Miaozi, Han, Hezhen, Tujia and Xibo, with frequencies between 65 and 50%). In contrast, the distribution of the derived rs74653330 A allele is more restricted, with the highest frequencies found in Altaic speaking populations from northern East Asia and Mongolia, such as the Yakut from Siberia (36%), the Daur (33%), the Oroqen (28%), the Hezhen (22%) and the Mongola (20%). Figure 1 shows a map of East Asia with the frequencies of both polymorphisms. The derived rs1800414 G and rs74653330 A alleles are not present in any of the samples from Africa, the Middle East or Oceania. In the Americas, the rs1800414 G allele is also absent, and one Maya individual is heterozygous for rs74653330. Both derived alleles are present at very low frequencies in Central–South Asia (rs1800414 G: 4.4%; rs74653330 A: 2.1%) and Europe (rs1800414 G: 0.3%; rs74653330 A: 1%). Within Central–South Asia, the derived alleles are primarily present in the Hazara (Pakistan) and Uygur (China). Within Europe, the derived alleles are observed only in Russia. The presence of the two derived alleles in some of the populations from Central–South Asia and Europe seems to be the consequence of gene flow from East-Asian groups.

It is interesting to note that the frequency distribution of the rs74653330 A allele reflects the present genetic structure at a genome-wide level in East Asia. We used the program PLINK¹⁵ to perform principal component analysis (PCA) of the East-Asian CEPH–HGDP populations by using genome-wide data (Affymetrix Axiom Human Origins Array) available in the HGDP–CEPH website (http://www.cephb.fr/en/hgdp_panel.php). We pruned SNPs based on linkage disequilibrium (LD) and removed five known areas of long-range LD. Figure 2 shows a visualization of the first two axes of the PCA using the program PAST (<http://folk.uio.no/ohammer/past/>). There is a clear geographic pattern with the northern populations (Yakut, Oroqen, Mongola, Daur and Hezhen) present on the left side of the plot. As described above, it is precisely in these populations in which the highest frequencies of the derived rs74653330 A allele are observed.

We explored the haplotype structure of the *OCA2* region in East Asia in detail. To do this, we merged the genotype data of the two markers of interest with the Affymetrix Human Origin data set for chromosome 15 plus the Illumina (San Diego, CA, USA) 650K data set for chromosome 15. The *OCA2* gene was extracted from this

¹Department of Anthropology, University of Toronto at Mississauga, Mississauga, ON, Canada and ²Department of Anthropology, University of Cincinnati, Cincinnati, OH, US. Correspondence: EJ Parra (esteban.parra@utoronto.ca)

Received 30 September 2015; revised 27 October 2015; accepted 28 October 2015

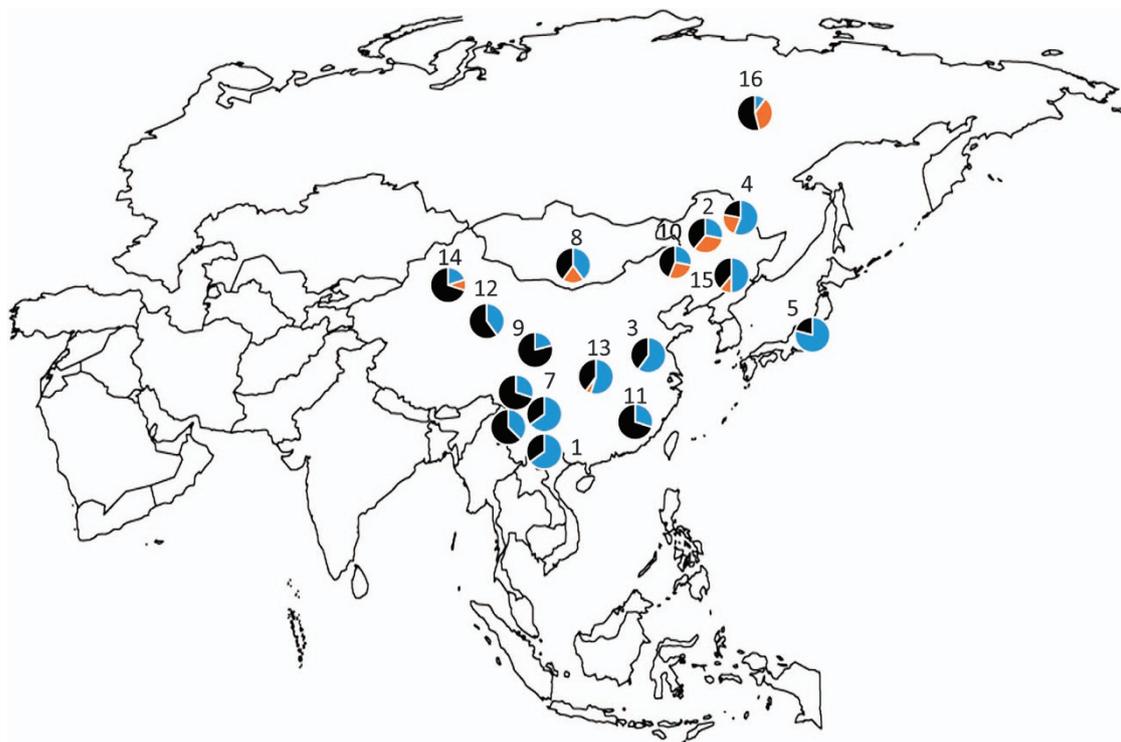


Figure 1. Distribution of allele frequencies for SNPs rs1800414 (blue) and rs74653330 (orange) in East-Asian populations: (1) Dia; (2) Daur; (3) Han; (4) Hezhen; (5) Japanese; (6) Lahu; (7) Miaozu; (8) Mongola; (9) Naxi; (10) Oroqen; (11) She; (12) Tu; (13) Tujia; (14) Uyghur; (15) Xibo; (16) Yakut; (17) Yizu; and (18) Cambodia.

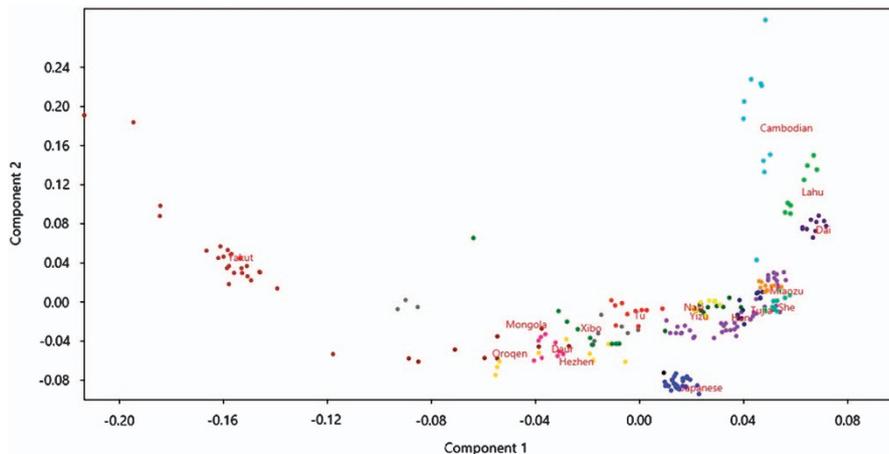


Figure 2. PCA (axes 1 and 2) showing population structure of East-Asian populations from the CEPH-HGDP panel.

data set by selecting markers from chromosome 15, position 25–26.5 Mb. On the basis of the north–south geographical gradient observed in the PCA output as well as the geographic distribution of the two polymorphisms, the haplotype analysis of East Asia was carried out separately in northern East Asia and the rest of East Asia. Populations that were included in the northern grouping included the Yakut from Siberia and the Oroqen, Mongola, Daur and Hezhen from northern China. The haplotype analyses were performed with the program Haploview.¹⁶ Figure 3 shows the haplotype structure surrounding the rs1800414 and rs74653330 polymorphisms. The two non-synonymous polymorphisms are located in the same LD block, but they are always found in different haplotypes. The haplotype analysis suggests that the haplotypes carrying the derived alleles for each polymorphism

arose independently from the same ancestral haplotype. Using the markers rs7170451–**rs1800414**–rs728405–rs728404–rs4778214–rs1448488–rs12903382–**rs74653330**–rs12910433–rs3794609–rs730502 to define the haplotype block (the relevant non-synonymous polymorphisms are labeled in bold), our results indicate that, from the ancestral haplotype ‘AAGAGCAGGTT’, a non-synonymous mutation at rs1800414 originated the haplotype ‘AGGAGCAGGTT’, and another non-synonymous mutation independently originated the haplotype ‘AAGAGCAAGTT’. Both derived haplotypes then increased in frequency in different regions of East Asia. The haplotype ‘AGGAGCAGGTT’ is now the most common haplotype in a broad region of East Asia, whereas the haplotype ‘AAGAGCAAGTT’ has become the most prevalent in northern East Asia. Several lines of evidence indicate that this

increase in frequency may have been the result of positive selection favoring light skin in high-latitude regions. Both derived alleles are non-synonymous variants predicted to have a functional effect,¹¹ and both have been associated with lighter skin pigmentation in East-Asian populations.^{5,10,11} In addition, several studies have identified signatures of positive selection in the *OCA2* region in genome-wide scans in East-Asian populations.^{2,3} The geographic distribution of the variants strongly suggests that these two mutations arose after the separation of European and East-Asian populations. This is supported by a recent study that dated the derived G allele of the *OCA2* rs1800414 polymorphism to ~10,000 years ago.¹⁷ To our knowledge, there has been no attempt to date the polymorphism rs74653330. We used the dense, genome-wide SNP data available for the HGDP-CEPH panel to estimate the ages of the derived alleles at rs1800414 and rs74653330 in East-Asian populations. We used a method¹⁸ that relies on the decay of haplotype sharing of

the ancestral genomic segment on which the derived mutations occurred. Before the analysis, we removed individuals with pi-hat values exceeding 0.05 to minimize potential problems with cryptic relatedness. To account for the possibility that members of individual populations may have a most recent common ancestor (MRCA) that is more recent than the MRCA of the entire East-Asian sample, we calculated these age estimates assuming a correlated genealogy.¹⁸ Under these conditions, and assuming a generation time of 29 years,¹⁹ we estimated the age of the derived allele at rs74653330 to be 6,835 years (95% confidence interval (CI): 1,070–12,798). The estimated age of the derived allele at rs1800414 is quite similar at 6,397 years (95% CI: 1,183–11,446 years). This is slightly younger than a previous estimate of the age of the derived allele at rs1800414 using a different method (10,660 years; 95% CI of 8,070–15,780),¹⁷ although the CIs of our estimate overlap Chen's point estimate. The discrepancy in age may be explained by differences in the two methods as well as in

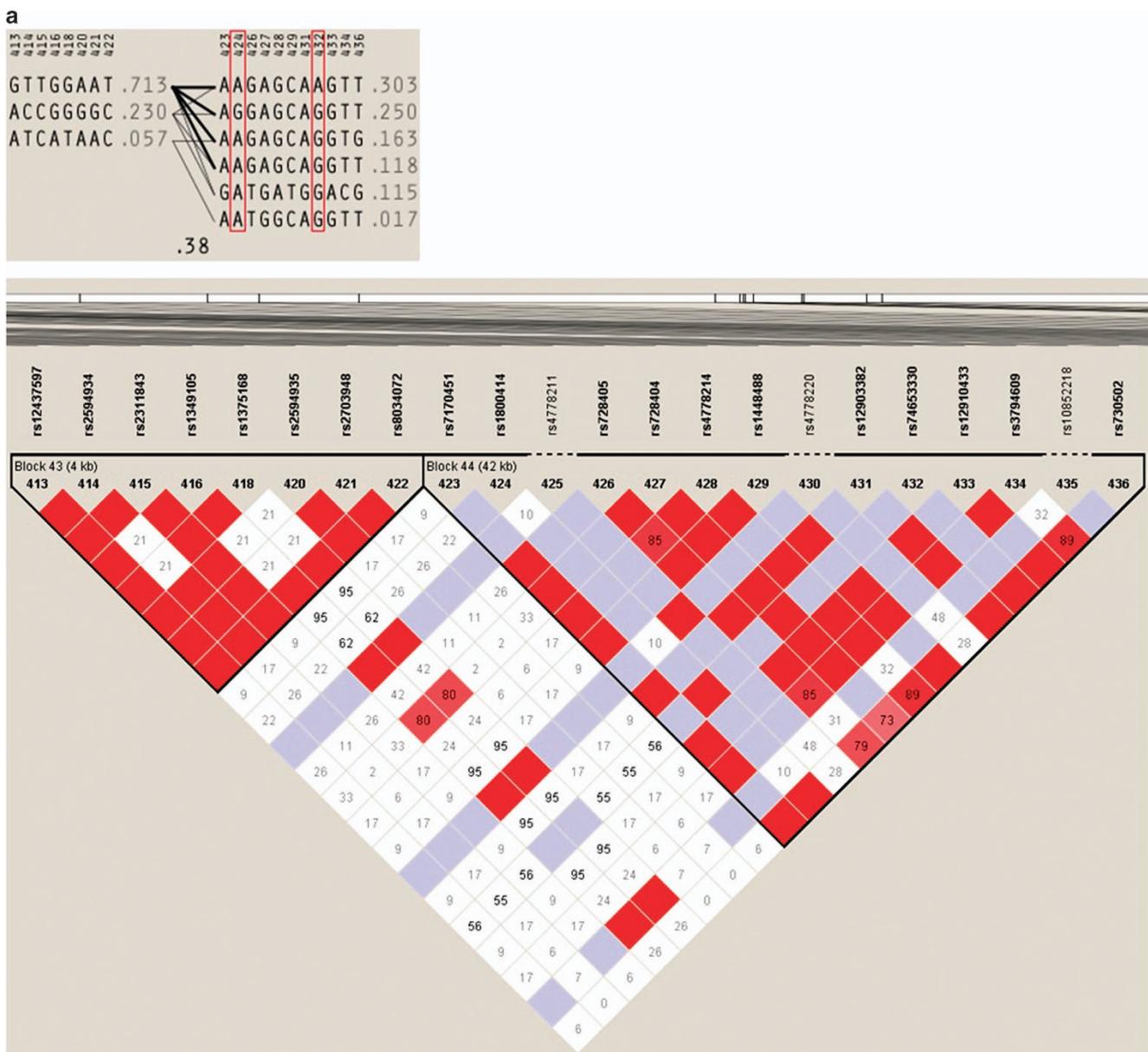


Figure 3. Haplotype block structure and pattern of LD of the *OCA2* region including markers rs1800414 (marker 424) and rs74653330 (marker 432). (a) Northern East Asia; (b) in the rest of Asia.

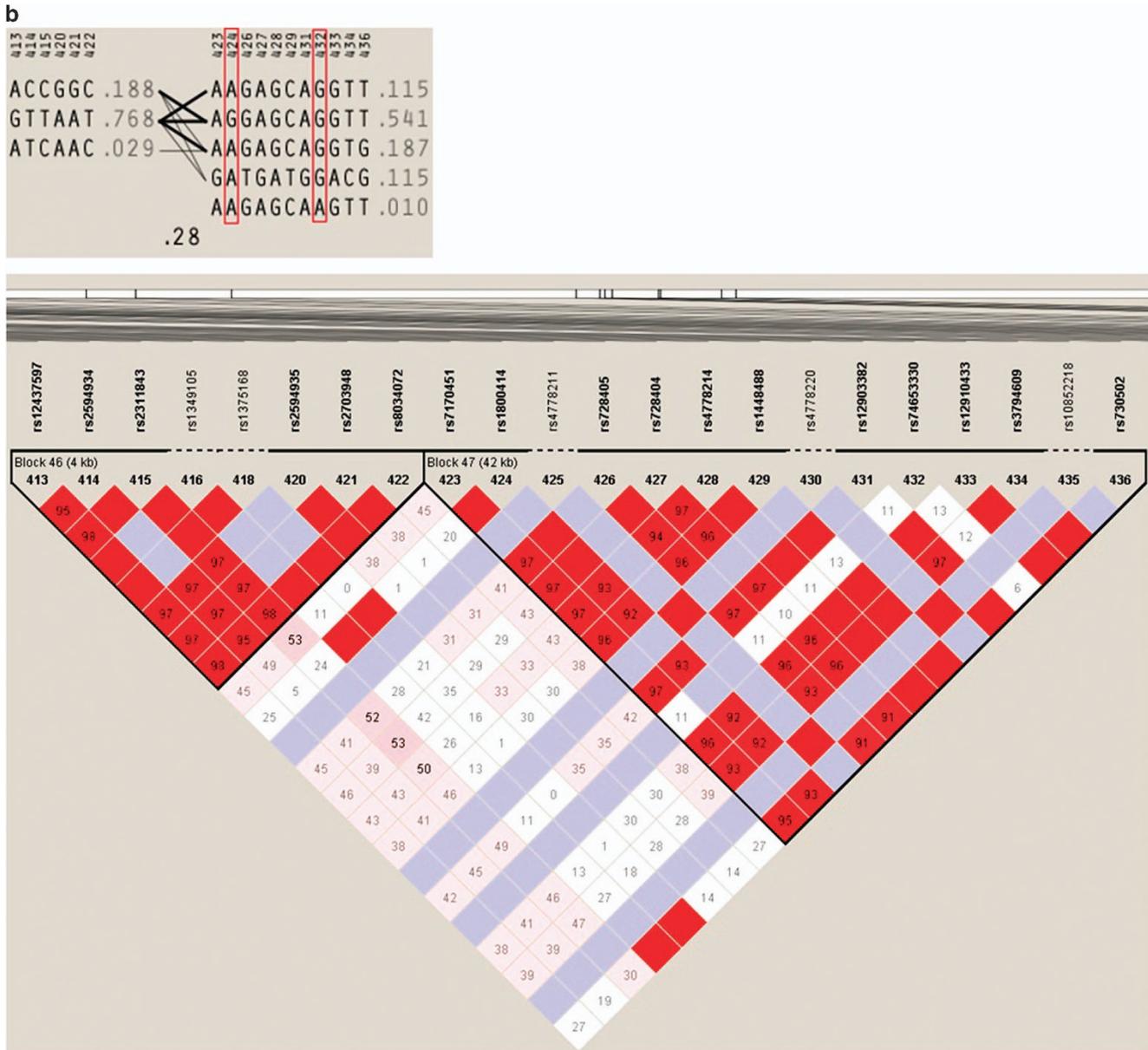


Figure 3. Continued.

differences among the East-Asian populations and the data sets used in each study.

Recent ancient DNA studies, which have characterized dense genomic data in Eurasian individuals spanning a broad archaeological period (e.g., from hunter gatherers to individuals living in the Bronze Age), have provided important information about the temporal distribution of genetic markers associated with pigmentation variation in Europe and have strengthened the case for selection operating in pigmentation-related genes in this region.^{20–22} Similar studies in East Asia have the potential to clarify the major events that have shaped the interesting distribution of the two non-synonymous variants of the *OCA2* gene in this vast area. In this respect, it will be important to consider not only potential selective effects but also the major population movements that have taken place in this region during the past 15,000 years.

HGV DATABASE

The relevant data from this Data Report are hosted at the *Human Genome Variation Database* at <http://dx.doi.org/10.6084/m9.figshare.hgv.756>, <http://dx.doi.org/10.6084/m9.figshare.hgv.759>.

COMPETING INTERESTS

The authors declare no conflict of interest.

REFERENCES

- Liu F, Wen B, Kayser M. Colorful DNA polymorphisms in humans. *Semin Cell Dev Biol* 2013; **24**: 562–575.
- Lao O, de Groot JM, van Duijn K, Navarro A, Kayser M. Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. *Ann Hum Genet* 2007; **71**: 354–369.
- Hider JL, Gittelman RM, Shah T, Edwards M, Rosenbloom A, Akey JM *et al*. Exploring signatures of positive selection in pigmentation candidate

- genes in populations of East Asian ancestry. *BMC Evol Biol* 2013; **13**: 150.
- 4 Anno S, Abe T, Yamamoto T. Interactions between SNP alleles at multiple loci contribute to skin color differences between caucasoid and mongoloid subjects. *Int J Biol Sci* 2008; **4**: 81–86.
- 5 Edwards M, Bigham A, Tan J, Li S, Gozdzik A, Ross K *et al*. Association of the OCA2 polymorphism His615Arg with melanin content in east Asian populations: further evidence of convergent evolution of skin pigmentation. *PLoS Genet* 2010; **6**: e1000867.
- 6 Donnelly MP, Paschou P, Grigorenko E, Gurwitz D, Barta C, Lu RB *et al*. A global view of the OCA2-HERC2 region and pigmentation. *Hum Genet* 2012; **131**: 683–696.
- 7 Eiberg H, Troelsen J, Nielsen M, Mikkelsen A, Mengel-From J, Kjaer KW *et al*. Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. *Hum Genet* 2008; **123**: 177–187.
- 8 Kayser M, Liu F, Janssens CJW, Rivadeneira F, Lao O, van Duijn K *et al*. Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. *Am J Hum Genet* 2008; **82**: 411–423.
- 9 Sturm RA, Duffy DL, Zhao ZZ, Leite FPN, Stark MS, Hayward NK *et al*. A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. *Am J Hum Genet* 2008; **82**: 424–431.
- 10 Abe Y, Tamiya G, Nakamura T, Hozumi Y, Suzuki T. Association of melanogenesis genes with skin color variation among Japanese females. *J Dermatol Sci* 2013; **69**: 167–172.
- 11 Eaton K, Edwards M, Krithika S, Cook G, Norton H, Parra EJ. Association study confirms the role of two OCA2 polymorphisms in normal skin pigmentation variation in East Asian populations. *Am J Hum Biol* 2015; **00**: 1–6.
- 12 Yuasa I, Harihara S, Jin F, Nishimukai H, Fujihara J, Fukumori Y *et al*. Distribution of OCA2*481Thr and OCA2*615Arg, associated with hypopigmentation, in several additional populations. *Leg Med* 2011; **13**: 215–217.
- 13 Norton HL, Kittles RA, Parra E, McKeigue P, Mao X, Cheng K *et al*. Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol Biol Evol* 2007; **24**: 710–722.
- 14 Yuasa I, Umetsu K, Harihara S, Miyoshi A, Saitou N, Park KS *et al*. OCA2*481Thr, a hypofunctional allele in pigmentation, is characteristic of northeastern Asian populations. *J Hum Genet* 2007; **52**: 690–693.
- 15 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
- 16 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
- 17 Chen H, Hey J, Slatkin M. A hidden Markov model for investigating recent positive selection through haplotype structure. *Theor Popul Biol* 2015; **99**: 18–30.
- 18 Gandolfo LC, Bahlo M, Speed TP. Dating rare mutations from small samples with dense marker data. *Genetics* 2014; **197**: 1315–1327.
- 19 Fenner JN. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am J Phys Anthropol* 2005; **128**: 415–423.
- 20 Wilde S, Timpson A, Kirsanow K, Kaiser E, Kayser M, Unterländer M *et al*. Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proc Natl Acad Sci USA* 2014; **111**: 4832–4837.
- 21 Mathieson I, Lazaridis I, Rohland N, Mallick S, Llamas B, Pickrell J *et al*. Eight thousand years of natural selection in Europe. *Biorxiv* 2015. Preprint at <http://dx.doi.org/10.1101/016477>.
- 22 Allentoft ME, Sikora M, Sjögren KG, Rasmussen S, Rasmussen M, Stenderup J *et al*. Population genomics of Bronze Age Eurasia. *Nature* 2015; **522**: 167–172.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>