


# SCIENTIFIC REPORTS

OPEN

## Positive selection in Europeans and East-Asians at the *ABCA12* gene

Roberto Sirica<sup>1</sup>, Marianna Buonaiuto<sup>1</sup>, Valeria Petrella<sup>2</sup>, Lucia Sticco<sup>3</sup>, Donatella Tramontano<sup>7</sup>, Dario Antonini<sup>2,8</sup>, Caterina Missero<sup>2,8</sup>, Ombretta Guardiola<sup>1</sup>, Gennaro Andolfi<sup>1</sup>, Heerman Kumar<sup>4,5</sup>, Qasim Ayub<sup>4,5,6</sup>, Yali Xue<sup>6</sup>, Chris Tyler-Smith<sup>6</sup>, Marco Salvemini<sup>2</sup>, Giovanni D'Angelo<sup>3</sup> & Vincenza Colonna<sup>1</sup> 

Received: 16 August 2018

Accepted: 7 February 2019

Published online: 19 March 2019

Natural selection acts on genetic variants by increasing the frequency of alleles responsible for a cellular function that is favorable in a certain environment. In a previous genome-wide scan for positive selection in contemporary humans, we identified a signal of positive selection in European and Asians at the genetic variant rs10180970. The variant is located in the second intron of the *ABCA12* gene, which is implicated in the lipid barrier formation and down-regulated by UVB radiation. We studied the signal of selection in the genomic region surrounding rs10180970 in a larger dataset that includes DNA sequences from ancient samples. We also investigated the functional consequences of gene expression of the alleles of rs10180970 and another genetic variant in its proximity in healthy volunteers exposed to similar UV radiation. We confirmed the selection signal and refine its location that extends over 35 kb and includes the first intron, the first two exons and the transcription starting site of *ABCA12*. We found no obvious effect of rs10180970 alleles on *ABCA12* gene expression. We reconstructed the trajectory of the T allele over the last 80,000 years to discover that it was specific to *H. sapiens* and present in non-Africans 45,000 years ago.

ATP binding cassette (ABC) transporters are trans-membrane ubiquitous proteins, that translocate natural substrates across plasma membranes. In humans, there are 49 genes coding for ABC transporters, arranged in eight subfamilies extensively studied because at least 11 of 49 genes are known to cause severe inherited diseases<sup>1</sup>. The ATP-binding cassette, sub-family A member 12 (*ABCA12*) gene was discovered from cDNA of human placenta<sup>2</sup>. *ABCA12* is 207 kb long with fifty-three exons and two very long introns at its beginning (26.5 kb and 47.3 kb, respectively, Fig. 1a). *ABCA12* is a keratinocyte transmembrane transporter that binds and hydrolyzes ATP to transport lipids in the lamellar granules<sup>3</sup>. This activity is required to form the extracellular lipid barrier in the outermost layer of the skin, the *stratum corneum* of the epidermis<sup>3</sup>. The lipid barrier is composed of three lipid classes (cholesterol, free fatty acids, and ceramides) and acts as a primary barrier between the body and the environment to prevent excessive water loss and to avoid penetration of pathogens<sup>4</sup>. *ABCA12* has also a role in immunity: in macrophages it regulates the cellular cholesterol metabolism via an LXRb-dependent post-transcriptional mechanism<sup>5</sup>.

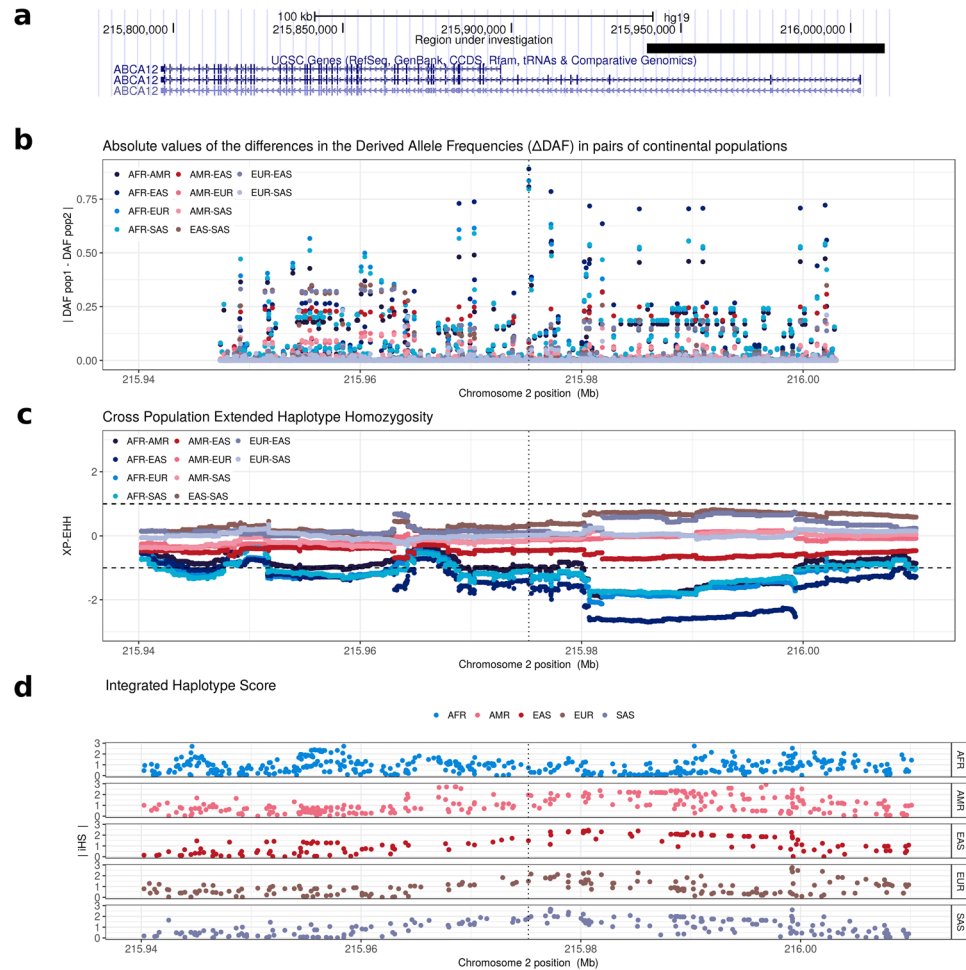
UV-radiation has a major effect on skin and keratinocytes and it is one of the most studied environmental stressors of the epidermal homeostasis. In keratinocytes, UV-radiation induces mutagenesis, apoptosis, proliferation, and metabolic changes. UV-radiation has a major role in determining skin pigmentation<sup>6,7</sup>, but it also can have a detrimental effect on the lipid barrier reducing the intercellular lipid cohesion<sup>8</sup>. Earlier studies suggest that UVB radiation down-regulates epidermal *ABCA12* gene expression<sup>9</sup> and in general UV light influences other ABC transporters activity in lymphocytes<sup>10</sup>.

<sup>1</sup>Institute of Genetics and Biophysics, National Research Council, Naples, Italy. <sup>2</sup>Department of Biology, University of Naples Federico II, Napoli, Italy. <sup>3</sup>Institute of Protein Biochemistry, National Research Council, Naples, Italy.

<sup>4</sup>Monash University Malaysia Genomics Facility, Tropical Medicine and Biology Multidisciplinary Platform, Bandar Sunway, Selangor Darul Ehsan, Malaysia. <sup>5</sup>School of Science, Monash University Malaysia, Bandar Sunway, Selangor Darul Ehsan, Malaysia. <sup>6</sup>Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK.

<sup>7</sup>Department of Molecular Medicine and Medical Biotechnology, University of Naples Federico II, Napoli, Italy.

<sup>8</sup>CEINGE Biotecnologie Avanzate, Napoli, 80145, Italy. Roberto Sirica, Marianna Buonaiuto, Valeria Petrella and Lucia Sticco contributed equally. Correspondence and requests for materials should be addressed to V.C. (email: vincenza.colonna@igb.cnr.it)



**Figure 1.** Positive selection at *ABCA12*. (a) The *ABCA12* gene has 55 exons and two long introns at its beginning. The black rectangle indicates the 70 kb region surrounding rs10180970 considered in this project. (b) rs10180970 is the most differentiated variant between Africans and non-Africans as shown by the absolute difference of the derived allele frequency ( $\Delta$ DAF), however, other variants seem to be also extremely differentiated. (c) The Cross Population Extended Haplotype Homozygosity statistic (XP-EHH), measured between pairs of continental populations, shows a signal of positive selection in non-Africans over 35 kb downstream rs10180970, especially in East-Asians. (d) The signal is confirmed by the Integrated Haplotype Score (iHS) within continental populations. In panels b–d the dashed line indicates the genomic position of rs10180970.

The always growing availability of ancient DNA information allows us to directly reconstruct the history of genetic mutations and understand if consequences of past adaptations are relevant for contemporary humans<sup>11</sup>. Past events of positive selection can be tested as detection of a major shift in allele frequencies<sup>12–14</sup>.

In a genome-wide scan for positive selection in contemporary humans, we identified a signal of positive selection in Europeans and Asians at rs10180970 C/T, located in the second intron of *ABCA12*<sup>15</sup>. In the study, we demonstrated that the derived allele T is highly frequent in Asians and Europeans, compared to Africans, where the ancestral C variant prevails. This evidence, together with the identification of patterns typical of positive selection in the genomic region of rs10180970 let us speculate that ancestor of current Europeans and Asians should have had an advantage in carrying the T allele and that it would be interesting to determine the functional consequences and the trajectory of the T allele through time.

Here we present results from investigating the genomic region surrounding rs10180970. We confirm that this region underwent positive selection and investigate one possible selected phenotype. We have extended the set of individuals used to investigate genetic variation and natural selection at rs10180970, including also DNA sequences from ancient human samples. We confirm the signal of positive selection at rs10180970, investigate linkage with genetic variants in its proximity and the effect of the two alleles on *ABCA12* gene expression.

## Results

**Natural selection signal at the *ABCA12* gene extends 10 kb downstream rs10180970.** We identified rs10180970 as a possible candidate for positive selection in Europeans and Asians during a genome-wide scan for positive selection conducted on populations from Phase I 1000 Genomes Project<sup>15</sup>. In this study, we considered a 70 kb region surrounding rs10180970 including the first two exons, the first and most of the second

		T/T	T/C	C/C
Complete Genomics full dataset	Africans Europeans & Asians	0	0.09	0.91
		0.89	0.10	0.01
Complete Genomics unique	Africans Europeans & Asians	0	0.05	0.95
		0.88	0.11	0.01
1000 Genomes Phase III	Africans Europeans & Asians	0.01	0.14	0.84
		0.85	0.14	0.01

**Table 1.** Genotypes frequencies at rs10180970 in Phase III 1000 Genome and Complete Genomics datasets both full and non overlapping with 1000 Genomes Phase I (unique).

introns, and the transcription starting site of *ABCA12* (Fig. 1a). Compared to Phase I, the 1000 Genomes Phase III dataset<sup>16</sup> includes 1,412 more samples belonging to 12 more populations (Supplementary Table S1) and in the region under study includes 1,612 variants for which the ancestral state is known and that we used to calculate statistics suggestive of population differentiation and positive selection.

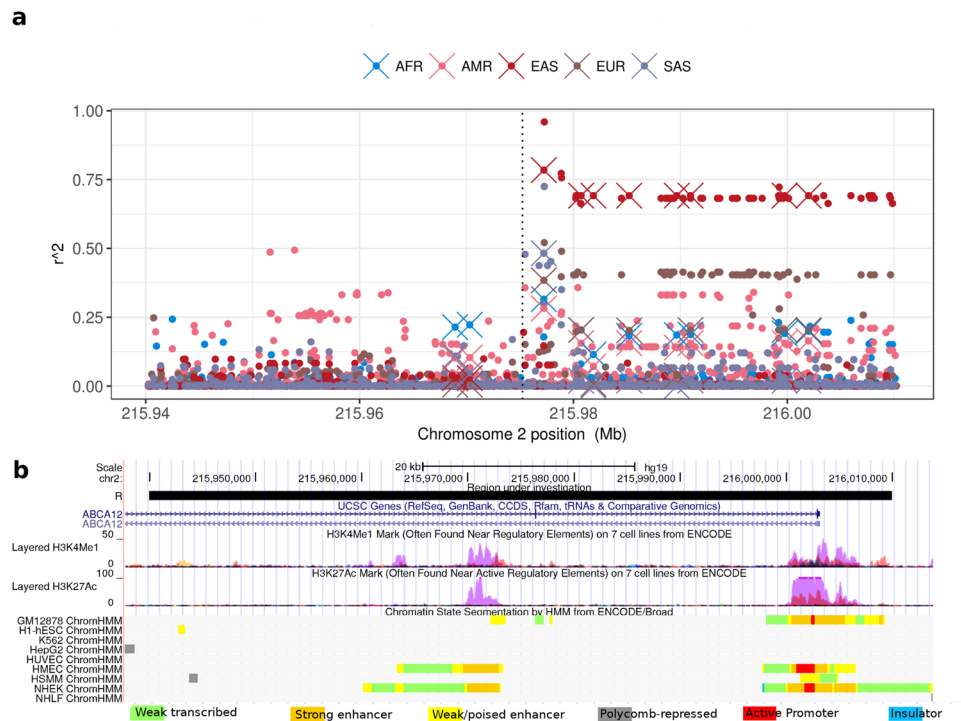
Between pairs of continental populations, we evaluated the absolute difference of the derived allele frequency ( $\Delta$ DAF) and the Cross Population Extended Haplotype Homozygosity (XP-EHH<sup>17</sup>).  $\Delta$ DAF approaches 1 when the derived allele prevails in one population and is almost absent in the other. According to  $\Delta$ DAF, rs10180970 is still the most differentiated variant in comparisons of Africans and non-Africans in 70 kb surrounding rs10180970 (Fig. 1b). In Phase III the T/T genotype frequency in the combined sample of Europeans and Asians is significantly higher than Africans (Table 1, Fisher exact test p-value < 2.2e-16). Similarly, in the Complete Genomics dataset (CG)<sup>18</sup>, the T/T genotype is significantly more frequent among Asians and Europeans compared to Africans. This is true both when considering samples partially overlapping with Phase I, and samples unique to CG (Table 1, Fisher exact test p-value < 2.2e-16), suggesting that the observed pattern of higher derived allele frequencies in Europeans and Asians compared to Africans was not due to the specific set of sample used but is reproducible. XP-EHH detects recent positive selection highlighting selective sweeps in which the selected allele has approached or achieved fixation in one population but remains polymorphic in the overall population. In the 1000 Genomes Phase III dataset, we observe a stretch of negative XP-EHH outside the expected variance range (i.e., -1,1) in pairs of Africans with non-Africans. The signal is located approximately at 10 kb downstream rs10180970, suggesting recent positive selection in non-Africans in this region (Fig. 1c), and it is particularly strong in East-Asians. Within continental populations we calculated the Integrated Haplotype Score (iHS<sup>19</sup>). iHS detects the classic signal of strong directional selection through the identification of unusually long haplotypes with low diversity surrounding core SNPs. As for XP-EHH, also for iHS we observe peaks of unusually high iHS 10 kb downstream rs10180970 in non-African populations and especially in East-Asians (Fig. 1d).

Altogether these results show that while rs10180970 is the most differentiated variant in the *ABCA12* gene, haplotypes downstream rs10180970 differ between Africans and non-Africans. Haplotypes carrying the T allele of rs10180970 are very common in Europeans and Asians and tend to be longer and less diverse in these two populations compared to the others, suggesting that this region might have been affected by recent positive selection in these two populations.

As linkage between variants in a long haplotype can be responsible for spurious signals, we computed linkage disequilibrium (LD) in continental populations between rs10180970 and 2,034 genetic variants in 70 kb surrounding rs10180970. For all populations but Africans, we observe LD within 10 kb downstream rs10180970 (Fig. 2a). In East-Asians and Europeans LD extends over 35 kb downstream rs10180970, in a region inclusive of the first intron and the first exon. Within this block, LD is high in East-Asians ( $r^2$  around 0.75) and moderate in Europeans ( $r^2$  around 0.50). The genomic region in high LD contains functional features such as chromatin regulation (H3K4Me1 and H3K27Ac), enhancers and a promoter located in the proximity of the first intron and includes SNPs with  $\Delta$ DAF > 0.70 (Fig. 2b). Therefore we conclude that other SNPs are in high LD with rs10180970 in East-Asians and part of a long haplotype that is located in a genomic region with regulatory features.

**No significant association between *ABCA12* expression and genotype at rs10180970.** Having confirmed the signal of natural selection at rs10180970 we next asked if the two alleles have a different effect on gene expression. Furthermore, in the hypothesis that another variant in linkage with rs10180970 can be responsible for the functional changes, we decided to include also rs2970968 in the analysis. Besides being in linkage disequilibrium with rs10180970 in East-Asians and having high  $\Delta$ DAF like other variants, rs2970968 has two further features that make it an interesting variant. First, we predicted the functional consequences of the variants with high  $\Delta$ DAF and in linkage disequilibrium with rs10180970 in East-Asians using Funseq<sup>20</sup> (Supplementary Table S2). We confirmed that rs2970968 (leftmost dark red cross in Fig. 2a) is located in a region predicted to be an enhancer<sup>21</sup> in lymphoblastoid, HeLa and HUVEC cell lines and that it is located in transcription factor binding peak of several transcription factors, although not in transcription factor motifs within peaks (Supplementary Fig. S1). One of the transcription factors is implicated in mammalian skin tumors (ATF2<sup>22,23</sup>) and three others are regulated by UVB radiation (STAT3, GATA3, and POU2F2)<sup>24-26</sup>. Secondly, rs2970968 is the closest SNP to a region that has shown to be essential for the functioning of the promoter of *ABCA12*<sup>27</sup> (Supplementary Fig. S1).

We investigated *ABCA12* expression in relation to the three genotypes of rs10180970 (C/C, C/T, T/T) and rs2970968 (A/A, A/G, G/G). Human skin would be the ideal tissue to test allele-specific gene expression, however sampling technique is invasive. It is instead possible to quantify *ABCA12* expression in total mRNA extracted from pulled hairs without the need for invasive skin biopsies<sup>28</sup>. Because *ABCA12* expression is influenced by sun



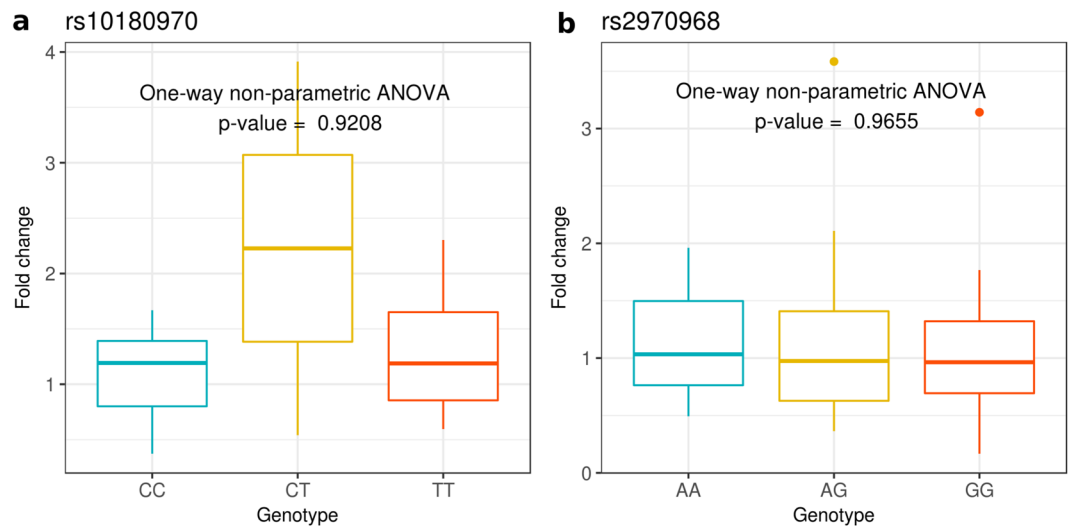
**Figure 2.** Linkage disequilibrium between rs10180970 and variants within 70 kb. (a) Vertical dotted line indicates the genomic position of rs10180970. Circles represent other variants at their genomic position on the x-axis and the  $r^2$  measures the non-random association of the alleles of each variant with rs10180970 on the y-axis. Crosses distinguish pairs in which the genetic variants have  $\Delta\text{DAF} > 0.70$  as shown in Fig. 1. (b) Screenshot from Genome Browser showing chromatin modification and state segmentation for several cell types in the genomic region under investigation.

exposure<sup>29</sup>, to discriminate between the genetic and the environmental effects, we performed the experiment in a group of individuals of different geographical origin but all resident in the same city. We determined the genotypes at rs10180970 and rs2970968 through DNA sequencing and measured *ABCA12* expression through quantitative PCR of reverse transcribed total mRNA extracted from pulled hair roots. We considered as reference the average *ABCA12* gene expression of genotypes of the ancestral alleles (C/C for rs10180970, and AA for rs2970968), which are also the allele prevalent in Africans, and calculated the fold change of the other observed genotypes. After quality controls, we had data on both genotype and *ABCA12* expression for 36 and 39 samples at rs10180970 and rs2970968, respectively. As shown in Fig. 3, there is no significant correlation between *ABCA12* gene expression and genotypes at both loci (One-way non-parametric ANOVA p-values  $> 0.9$  for both sites), suggesting no effect of rs10180970 or rs2970968 genotypes on *ABCA12* expression in hair roots. Therefore, we conclude that the effect of the selection at these two loci on gene expression is not obvious when individuals are exposed to similar UV radiation.

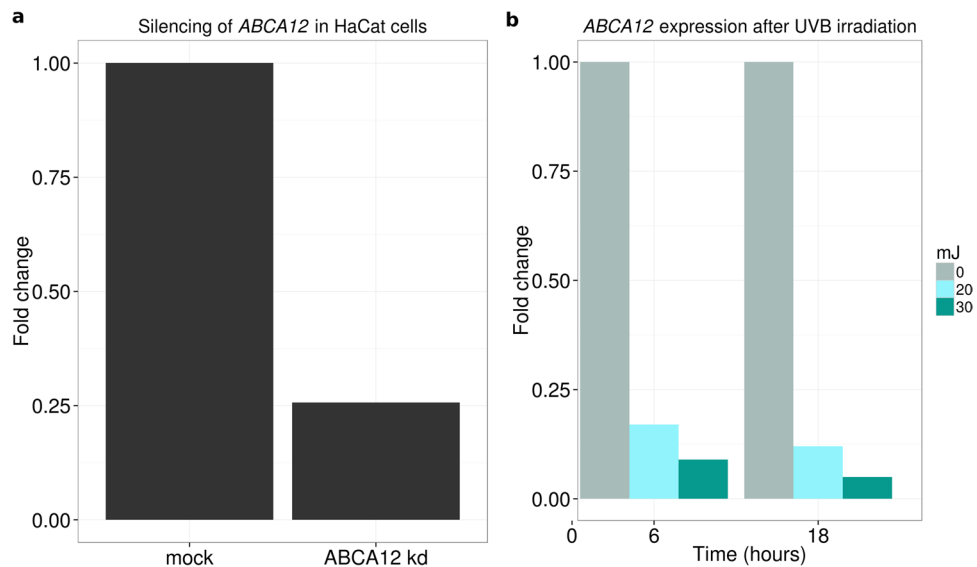
While it is not possible to model the gene effect in mouse because there is no homologous region in mouse, we found that *ABCA12* is well expressed in HaCaT cells, a spontaneously transformed aneuploid immortal keratinocyte cell line from adult human skin<sup>30</sup>. We determined by Sanger sequencing that HaCaT cells are heterozygous at rs10180970 and therefore can be used for allele-specific expression assessment at this locus. As a preliminary work, we silenced *ABCA12* in HaCaT cells using small interfering RNAs and confirmed that *ABCA12* expression is reduced after UVB radiation (20 mJ and 30 mJ), with downregulation remaining sustained up to 18 hours after treatment (Fig. 4).

**The T variant of rs10180970 is specific to *H. sapiens* and rapidly increased in frequency in non-Africans.** We determined the pattern of genetic variation at rs10180970 in time and space in 88 ancient and 2,764 contemporary samples ranging the last 80,000 years with the aim to understand when the T allele first appeared and how its frequency evolved through time. Among contemporary samples, 41 were collected for this study and are from 9 geographically diverse populations (Supplementary Table S1). Their genotype at rs10180970 was determined by Sanger sequencing a 600 bp region containing the variant. For the rest of the contemporary samples genotypes at rs10180970 were publicly available from the 1000 Genomes Phase III<sup>16</sup> and a study on West Africans<sup>31</sup>. For ancient samples, we used publicly available raw sequencing data (Supplementary Table S3) to perform variant calling. Beside *H. sapiens*, ancient samples include also one *Denisova* sample and two *Neanderthals*.

The allele frequency of the derived allele T ( $f_T$ ) through time was determined grouping individuals in bins of 2000 years (Fig. 5). Some pools are made of very few individuals ( $n = 1$  in extreme cases). Individuals were subdivided according to species and within *H. sapiens* in Africans and non-Africans. In our data, the T allele appeared



**Figure 3.** *ABCA12* expression in relation to the three genotypes of (a) rs10180970 and (b) rs2970968. Fold change refers to average *ABCA12* gene expression of ancestral genotypes, which are also the genotypes prevalent in Africans (C/C for rs10180970, and AA for rs2970968).

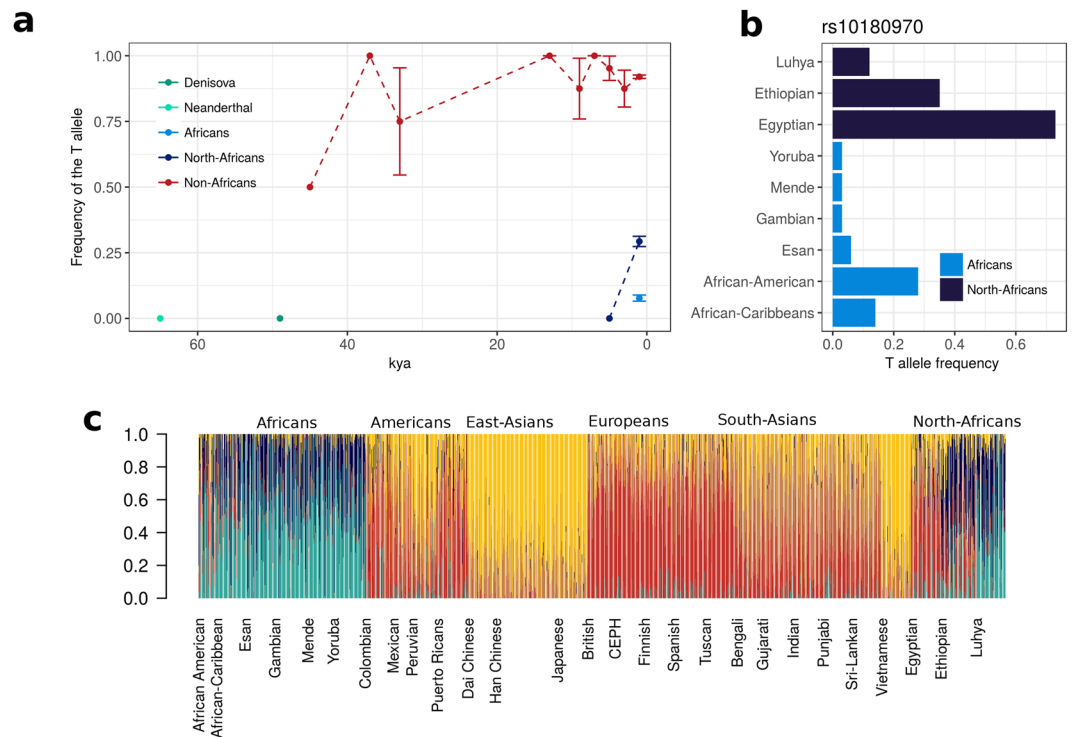


**Figure 4.** *ABCA12* expression in HaCaT cells and UVB irradiation (a) *ABCA12* is expressed in HaCaT cells and can be silenced using small interfering RNA. (b) *ABCA12* expression is reduced after Ultraviolet B (UVB) radiation and cells do not recover even after 18 hours. T6 and T18 indicate 6 and 18 hours after UVB radiation; mJ = millijoule.

for the first time around 45 kya in a *H. sapiens* sample from Russia<sup>32</sup> and rapidly  $f_T$  reached 0.9 in non-Africans remaining stable until the present day.

Of the 16 African ancient samples available to date, only three have reliable data in the rs10180970 region (Supplementary Table S3). The most ancient African data is from Ethiopia 4.5 kya<sup>33</sup>, while the other two are from 1.3–2 kya<sup>34</sup>. All ancient Africans have genotype C/C. In contemporary Africans  $f_T$  is very low but not completely absent (0.09), while in North-Africans  $f_T$  is on average 0.4 and T is the major allele in Egyptians, within North-Africans (Fig. 5b). Admixture analysis shows that the Africans and North-Africans with higher  $f_T$ , in fact, have quite high European and South Asian components probably reflecting recent admixture (Fig. 5c).

Both Neanderthals and the one Denisova sample available are homozygous for the ancestral C allele, suggesting that the T allele might have been specific to *H. sapiens*, as far as we know about the other two species. In support of this hypothesis, two recent publications find that there is a very little probability of admixture in the region surrounding rs10180970 between *H. sapiens* and Neanderthals<sup>35</sup> (Supplementary Fig. S2) and *H. sapiens* and Denisova<sup>36</sup>.



**Figure 5.** Prevalence of the T allele of rs10180970 in time and space. **(a)** T allele trajectory in the last 80,000 years in 88 ancient and 2,764 contemporary samples. The T allele is observed for the first time 45,000 years ago in non-Africans and rapidly rose in frequency. **(b)** T allele frequency is very low among African populations with the exception of North-Africans (especially Egyptians and Ethiopians) and African American. **(c)** T allele was probably recently introgressed in African American, as admixture analysis show a shared genetic component of African American with Europeans and East-Asians in the region under investigation. Vertical bars represent single individuals and colors reflect membership to admixture clusters in the most likely hypothesis of five clusters.

We conclude that the T variant of rs10180970 is specific to *H. sapiens* and rapidly increased in frequency in non-Africans, while the low prevalence of T in contemporary Africans might be due to recent admixture with Europeans and South-Asians.

## Discussion

Humans are exposed to environmental agents through the naked skin, and in fact, strong natural selection has been detected in a number of skin-related phenotypic traits such as hair and eye pigmentation<sup>6,37,38</sup>, and hair thickness<sup>39,40</sup>. The lipid barrier is the most external layer of the skin and it acts as an interface with the environment, providing protection against UV, dehydration, and pathogens. The formation of lipid barrier is a complex process and mutations in genes involved in this process result in skin diseases. Nevertheless, to date, to our knowledge, no studies investigated natural selection of genes that contribute to build the lipid barrier such as *ABCA12*.

The genetic variant rs10180970 located on chromosome 2 in the *ABCA12* gene, is predicted to be under positive selection in Europeans and Asians, as we discovered in a genome-wide scan based on population differentiation<sup>15</sup>. Here we investigated the genomic region containing rs10180970 to ask whether the variant is truly under positive selection and investigated the functional consequences of the selection event.

We confirmed the signal of selection in Europeans and Asians and discovered high linkage disequilibrium in a 35 kb region downstream rs10180970 (Fig. 2). In the 35 kb region the signals of positive selection are as strong or stronger than those initially found at rs10180970 (Fig. 1) and include the promoter of the gene (Fig. 2). By using a larger dataset of DNA sequences and a more fine population subdivision compared to the one used to discover the signal, we observe that the signal is stronger in East-Asians compared to South-Asians, possibly because South-Asians exposure to UV radiation is more similar to that of Africans. Despite confirming the positive selection signal we could not observe obvious allele-specific effect of rs10180970 on *ABCA12* gene expression (Fig. 3) in a sample of individuals exposed to the same level of UV radiation. Finally, we traced the history through time of the allele under selection using publicly available ancient human sequences and established that the mutation is specific to *H. sapiens* (Fig. 5, Supplementary Fig S2). While our findings strongly corroborate the signal of selection they also suggest that a different functional approach should be used to decipher the functional consequences of the event of selection in terms of gene expression.

We suggest that further work is required to fully decipher the effect of selection on *ABCA12* and identified the keratinocyte-derived HaCaT cells as a model system for *in vitro* experiments. We determined that *ABCA12*

is expressed and can be silenced in the HaCaT cells, and more importantly HaCaT cells are heterozygous at rs10180970, therefore ideal for investigating allele-specific effects. On the other hand, because of the possibility that other sites in linkage with rs10180970 are involved, it would be probably worth while to consider sequence information at other sites as well. More in general, overall indication for future research implicate the study of the non-coding regions of *ABCA12*, like sequence/haplotype-specific enhancer assay, even if the dimensions of the gene can be a limitation.

There are two main findings from our work. First, we reconstructed the demographic history of the T allele using the currently available information to discover that it is specific to *H. sapiens* and it was apparently very frequent in humans that migrated out of Africa prior to 50kya. The second main finding is that while rs10180970 might act as a sentinel for the signal of selection, the signal extends over 35 kb and includes the first intron, the first two exons and the transcription starting site of *ABCA12*. Linkage disequilibrium in this region is high in East-Asians and Europeans, there are several regions predicted to be enhancers and two regions predicted to be sites of chromatin modifications (methylation and acetylation), thus possibly influencing gene expression.

Our study endorses and refines previous studies. Similar to *ABCA12*, other members of the ABC family have also been identified in previous scans for positive selection<sup>41</sup>. The genomic region containing *ABCA12* was identified as one of the most differentiated between continental populations<sup>42</sup>, and in fact *ABCA12* hosts one of the top ten SNPs in the genome, useful to discriminate continental populations in clustering analyses<sup>43</sup>. *ABCA12* was identified as one of the top candidate genes in a study investigating local adaptation in Asians<sup>44</sup> and stands out in a selection scan performed with LD-based statistics on HapMap Phase II data<sup>45</sup>. However, none of these studies provides an indication of the exact genomic location of the signal within the gene. In this study, we corroborate the signal of selection at rs10180970 that was localized in our previous study<sup>15</sup>, and provide initial evidence that will help understand why the T allele of rs10180970 was favored in the ancestors of current Asians, in particular, East-Asians, and Europeans. Our work also shows that with the exception of few cases (e.g.<sup>46,47</sup>) it is surprisingly difficult to identify the selected allele and selected phenotype, even with abundant sequence data and functional tests.

## Methods

**Samples and genotyping.** Forty samples were collected from healthy fully informed, consenting, adult volunteers from 9 countries (Supplementary Table S1). Saliva samples were collected using the Oragene DNA (OG-500) kit. Genomic DNA was phenol–chloroform extracted from saliva<sup>48</sup>. All experiments were carried out in accordance with relevant guidelines and regulations. The study was approved by the ethics committee of the University of Naples Federico II (protocol number 106/15) and conducted in accordance with the relevant guidelines and regulations. Written informed consent was obtained from all participants.

Genotypes at rs10180970 and rs2970968 were obtained by Sanger-sequencing a 325 bp and 400 bp regions, respectively, of the *ABCA12* gene where the variants are located. The fragments were amplified from genomic DNA by PCR in a final volume of 50  $\mu$ l using oligos designed with Primer3<sup>49,50</sup> and the following conditions: 1 cycle 95 °C for 5'; 35 cycles at 95 °C for 1', 62 °C for 1', 72 °C for 45"; 1 cycle at 72 °C for 10'. PCR products were purified using PCR purification kit from QIAGEN.

All other contemporary samples in the study are part of public collections and genotype data at rs10180970 was available in form of vcf files<sup>16,18,31</sup>. For ancient samples we screened publicly available sequence reads data of 400 samples spanning 80,000–200 years before present<sup>32,51–64</sup>, however filtering for coverage >2 in the region of interest resulted in 89 samples listed in Supplementary Table S3. Variant calling in ancient samples was performed with samtools/bcftools<sup>65</sup>.

**Derived allele frequencies, linkage disequilibrium, admixture, introgression, natural selection analyses.** Derived allele frequencies were computed using vcftools<sup>66</sup> using the information on derived allele from the 1000 Genomes Project<sup>16</sup>. Linkage disequilibrium (LD) between rs10180970 and any other SNP in  $\pm$ 2 Mb was computed with PLINK1.9<sup>67</sup>. Clustering analysis in a region of 2 Mb surrounding rs10180970 including 55,683 Single Nucleotide Polymorphisms (SNPs) was done with ADMIXTURE<sup>68</sup>. ADMIXTURE was run under the hypotheses of 2 to 10 clusters with cross-validation computation.

Data for checking if the region of rs10180970 is among the ones putatively introgressed from Neanderthal were downloaded from [http://genetics.med.harvard.edu/reichlab/Reich\\_Lab/Datasets\\_-\\_Neandertal\\_Introgression.html](http://genetics.med.harvard.edu/reichlab/Reich_Lab/Datasets_-_Neandertal_Introgression.html).

Integrated haplotype score (iHS) and Cross-Population Extended Haplotype Homozygosity (XP-EHH) were calculated with Hapbin<sup>69</sup>. Parameters for iHS calculation were  $-\text{minmaf}$  0.01,  $-\text{cutoff}$  0.01. iHS and XP-EHH log-ratio were normalized to have zero mean and unit variance using the standard normalization implemented in Hapbin that groups variants in bins of 2% frequency.

**RNA extraction and processing from hair follicles.** Hair follicles (5 to 7 bulbs) were grasped as near to the scalp as possible without damaging hair roots. Total RNA was extracted from bulbs on the day of collection, or following storage at 4 °C for 1–5 days in RNA Later and placed into a 1.5 ml microcentrifuge tube containing 1 ml of Trizol Reagent (Fisher Molecular Biology). RNA extraction was then conducted according to manufacturer's instructions. 500 ng RNA were reverse-transcribed using SuperScript<sup>®</sup> VILO<sup>™</sup> MasterMix (Invitrogen). qPCR was performed on technical triplicates per each sample using Light Cycler 480<sup>®</sup> II (Roche), on 2  $\mu$ l of previously diluted cDNA (1:5) template using LightCycler<sup>®</sup> 480 SYBR Green I Master Mix (Roche), according to the manufacturer instructions. The thermal profile consisted of 1 cycle at 95 °C for 10'; 40 cycles at 95 °C for 10'', 60 °C for 10'', 72 °C for 10''. Fold changes were calculated relative to Hypoxanthine Phospho Ribosyl Transferase (*HPRT1*) expression as in<sup>70</sup>.

**HaCat cells culturing, genotyping, silencing and UV radiation.** HaCat cells were cultured in 75 cm<sup>2</sup> tissue culture flasks in Dulbecco's Modified Eagle's Medium, containing 10% fetal bovine serum and 1% penicillin and streptomycin. Cells were maintained in a humidified atmosphere (5% CO<sub>2</sub>) at 37 °C. Genomic DNA was extracted from HaCat cells<sup>71</sup> and genotype at rs10180970 was determined by Sanger sequencing.

For silencing experiments, HaCaT cells were transfected with a *ABCA12*-targeting siRNA (Sigma Aldrich). Transfection was performed with Oligofectamine following the manufacturer's protocol. Cells were collected 72 hours after transfection. Silencing efficiency was evaluated by Real Time Quantitative PCR (RT-qPCR) following total RNA extraction using RNeasy Mini Kit-QIAGEN and 1 µg and reverse-transcription using SuperScript<sup>®</sup> VILO<sup>™</sup> MasterMix (Invitrogen).

For UVB radiation HaCaT cells were plated in 60 mm plates and grown to 70–90% confluence. Subsequently, cells were irradiated with 20 and 30 mJ/cm<sup>2</sup> using three lamps (Philips Ultraviolet 8 TL 20 W/01 RS lamps; Philips, Eindhoven, Netherlands) generating UVB light in the range of 290–320 nm with an emission peak at 312 nm. The intensity of UVB irradiation was measured using a phototherapy radiometer (International Light, Newburyport, MA) of UVB. After irradiation cells were incubated for 0, 6 and 18 hours and *ABCA12* expression was assessed by Real Time PCR (RT-PCR). Fold changes were calculated relative to Ribosomal Protein Large P0 (*RPLP0*) expression as in<sup>70,72</sup>.

## Data Availability

The datasets generated during and analysed during the current study are available from the corresponding author on request.

## References

- Vasiliou, V., Vasiliou, K. & Nebert, D. W. Human atp-binding cassette (abc) transporter family. *Human genomics* **3**, 281 (2009).
- Annilo, T. *et al.* Identification and characterization of a novel *abca* subfamily member, *abca12*, located in the lamellar ichthyosis region on 2q34. *Cytogenetic and genome research* **98**, 169–176 (2002).
- Akiyama, M. The roles of *abca12* in epidermal lipid barrier formation and keratinocyte differentiation. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* **1841**, 435–440 (2014).
- van Smeden, J. & Bouwstra, J. A. Stratum corneum lipids: their role for the skin barrier function in healthy subjects and atopic dermatitis patients. In *Skin Barrier Function*. **49**, 8–26 (Karger Publishers (2016).
- Fu, Y. *et al.* *Abca12* regulates *abca1*-dependent cholesterol efflux from macrophages and the development of atherosclerosis. *Cell metabolism* **18**, 225–238 (2013).
- Jablonski, N. G. & Chaplin, G. The colours of humanity: the evolution of pigmentation in the human lineage. *Phil. Trans. R. Soc. B* **372**, 20160349 (2017).
- Jablonski, N. G. & Chaplin, G. Human skin pigmentation as an adaptation to uv radiation. *Proceedings of the National Academy of Sciences* **107**, 8962–8968 (2010).
- Biniak, K., Levi, K. & Dauskardt, R. H. Solar uv radiation reduces the barrier function of human skin. *Proceedings of the National Academy of Sciences* **109**, 17111–17116 (2012).
- Markó, L. *et al.* Keratinocyte atp binding cassette transporter expression is regulated by ultraviolet light. *Journal of Photochemistry and Photobiology B: Biology* **116**, 79–88 (2012).
- Dumitriu, I. *et al.* Uv irradiation inhibits abc transporters via generation of adp-ribose by concerted action of poly (adp-ribose) polymerase-1 and glycohydrolase. *Cell death and differentiation* **11**, 314 (2004).
- Marciniak, S. & Perry, G. H. Harnessing ancient genomes to study the history of human adaptation. *Nature Reviews Genetics* **18**, nrg-2017 (2017).
- Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient eurasians. *Nature* **528**, 499 (2015).
- Ye, K., Gao, F., Wang, D., Bar-Yosef, O. & Keinan, A. Dietary adaptation of fads genes in europe varied across time and geography. *Nature ecology & evolution* **1**, 0167 (2017).
- Gelabert, P., Olalde, I., de Dios, T., Civit, S. & Lalueza-Fox, C. Malaria was a weak selective force in ancient europeans. *Scientific reports* **7**, 1377 (2017).
- Colonna, V. *et al.* Human genomic regions with exceptionally high levels of population differentiation identified from 911 whole-genome sequences. *Genome biology* **15**, R88 (2014).
- Consortium, G. P. *et al.* A global reference for human genetic variation. *Nature* **526**, 68 (2015).
- Sabeti, P. C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832 (2002).
- Drmanac, R. *et al.* Human genome sequencing using unchained base reads on self-assembling dna nanoarrays. *Science* **327**, 78–81 (2010).
- Voight, B. E., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *Plos biology* **4**, e72 (2006).
- Khurana, E. *et al.* Integrative annotation of variants from 1092 humans: application to cancer genomics. *Science* **342**, 1235587 (2013).
- Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43 (2011).
- Leslie, M. C. & Bar-Eli, M. Regulation of gene expression in melanoma: new approaches for treatment. *Journal of cellular biochemistry* **94**, 25–38 (2005).
- Papassava, P. *et al.* Overexpression of activating transcription factor-2 is required for tumor growth and progression in mouse skin tumors. *Cancer research* **64**, 8573–8584 (2004).
- Bito, T. *et al.* Ultraviolet light induces stat3 activation in human keratinocytes and fibroblasts through reactive oxygen species and dna damage. *Experimental dermatology* **19**, 654–660 (2010).
- Rácz, E. *et al.* Gata3 expression is decreased in psoriasis and during epidermal regeneration; induction by narrow-band uvb and il-4. *Plos One* **6**, e19806 (2011).
- Heinloth, A. N. *et al.* Identification of distinct and common gene expression changes after oxidative stress and gamma and ultraviolet radiation. *Molecular carcinogenesis* **37**, 65–82 (2003).
- Shimizu, Y. *et al.* A palindromic motif in the -2084 to -2078 upstream region is essential for *abca12* promoter function in cultured human keratinocytes. *Scientific reports* **4**, 6737 (2014).
- Takeichi, T., Sugiura, K., Matsuda, K., Kono, M. & Akiyama, M. Novel *abca12* splice site deletion mutation and *abca12* mrna analysis of pulled hair samples in harlequin ichthyosis. *Journal of dermatological science* **69**, 259–261 (2013).
- Lonsdale, J. *et al.* The genotype-tissue expression (gtex) project. *Nature genetics* **45**, 580 (2013).
- Boukamp, P. *et al.* Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *The Journal of cell biology* **106**, 761–771 (1988).



31. Pagani, L. *et al.* Tracing the route of modern humans out of africa by using 225 human genome sequences from ethiopians and egyptians. *The American Journal of Human Genetics* **96**, 986–991 (2015).
32. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western siberia. *Nature* **514**, 445–449 (2014).
33. Llorente, M. G. *et al.* Ancient ethiopian genome reveals extensive eurasian admixture in eastern africa. *Science* **350**, 820–822 (2015).
34. Skoglund, P. *et al.* Reconstructing prehistoric african population structure. *Cell* **171**, 59–71 (2017).
35. Sankararaman, S. *et al.* The genomic landscape of neanderthal ancestry in present-day humans. *Nature* **507**, 354–357 (2014).
36. Sankararaman, S., Mallick, S., Patterson, N. & Reich, D. The combined landscape of denisovan and neanderthal ancestry in present-day humans. *Current Biology* **26**, 1241–1247 (2016).
37. Wilde, S. *et al.* Direct evidence for positive selection of skin, hair, and eye pigmentation in europeans during the last 5,000 y. *Proceedings of the National Academy of Sciences* **111**, 4832–4837 (2014).
38. Crawford, N. G. *et al.* Loci associated with skin pigmentation identified in african populations. *Science* (2017).
39. Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913 (2007).
40. Grossman, S. R. *et al.* A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* **327**, 883–886 (2010).
41. Pickrell, J. K. *et al.* Signals of recent positive selection in a worldwide sample of human populations. *Genome research* **19**, 826–837 (2009).
42. Tennessen, J. A. & Akey, J. M. Parallel adaptive divergence among geographically diverse human populations. *PLoS genetics* **7**, e1002127 (2011).
43. Lao, O., van Duijn, K., Kersbergen, P., de Knijff, P. & Kayser, M. Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. *The American Journal of Human Genetics* **78**, 680–690 (2006).
44. Qian, W., Deng, L., Lu, D. & Xu, S. Genome-wide landscapes of human local adaptation in asia. *PLoS One* **8**, e54224 (2013).
45. Jacobs, G. S., Sluckin, T. J. & Kivisild, T. Refining the use of linkage disequilibrium as a robust signature of selective sweeps. *Genetics* **203**, 1807–1825 (2016).
46. Fumagalli, M. *et al.* Greenlandic inuit show genetic signatures of diet and climate adaptation. *Science* **349**, 1343–1347 (2015).
47. Gerbault, P., Moret, C., Currat, M. & Sanchez-Mazas, A. Impact of selection and demography on the diffusion of lactase persistence. *PLoS One* **4**, e6369 (2009).
48. Sweet, D., Lorente, M., Valenzuela, A., Lorente, J. & Alvarez, J. C. Increasing dna extraction yield from saliva stains with a modified chelex method. *Forensic science international* **83**, 167–177 (1996).
49. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic acids research* **40**, e115–e115 (2012).
50. Koressaar, T. & Remm, M. Enhancements and modifications of primer design program primer3. *Bioinformatics* **23**, 1289–1291 (2007).
51. Prüfer, K. *et al.* The complete genome sequence of a neandertal from the altai mountains. *Nature* **505**, 43 (2014).
52. Meyer, M. *et al.* A high-coverage genome sequence from an archaic denisovan individual. *Science* **338**, 222–226 (2012).
53. Günther, T. *et al.* Ancient genomes link early farmers from atapuerca in spain to modern-day basques. *Proceedings of the National Academy of Sciences* **112**, 11917–11922 (2015).
54. Seguin-Orlando, A. *et al.* Genomic structure in europeans dating back at least 36,200 years. *Science* **346**, 1113–1118 (2014).
55. Raghavan, M. *et al.* The genetic prehistory of the new world arctic. *Science* **345**, 1255832 (2014).
56. Jones, E. R. *et al.* Upper palaeolithic genomes reveal deep roots of modern eurasians. *Nature communications* **6** (2015).
57. Rasmussen, M. *et al.* The genome of a late pleistocene human from a clovis burial site in western montana. *Nature* **506**, 225–229 (2014).
58. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day europeans. *Nature* **513**, 409 (2014).
59. Haak, W. *et al.* Massive migration from the steppe was a source for indo-european languages in europe. *Nature* **522**, 207 (2015).
60. Skoglund, P. *et al.* Genomic diversity and admixture differs for stone-age scandinavian foragers and farmers. *Science* **344**, 747–750 (2014).
61. Allentoft, M. E. *et al.* Population genomics of bronze age eurasia. *Nature* **522**, 167–172 (2015).
62. Rasmussen, M. *et al.* Ancient human genome sequence of an extinct palaeo-eskimo. *Nature* **463**, 757 (2010).
63. Malaspinas, A.-S. *et al.* Two ancient human genomes reveal polynesian ancestry among the indigenous botocudos of brazil. *Current Biology* **24**, R1035–R1037 (2014).
64. Raghavan, M. *et al.* Genomic evidence for the pleistocene and recent population history of native americans. *Science* **349**, aab3884 (2015).
65. Li, H. *et al.* The sequence alignment/map format and samtools. *Bioinformatics* **25**, 2078–2079 (2009).
66. Danecek, P. *et al.* The variant call format and vcfutils. *Bioinformatics* **27**, 2156–2158 (2011).
67. Chang, C. C. *et al.* Second-generation plink: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
68. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome research* **19**, 1655–1664 (2009).
69. Maclean, C. A., Chue Hong, N. P. & Prendergast, J. G. hapbin: An efficient program for performing haplotype-based scans for positive selection in large genomic datasets. *Molecular biology and evolution* **32**, 3027–3029 (2015).
70. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative pcr and the 2- $\delta\delta$ ct method. *methods* **25**, 402–408 (2001).
71. Laird, P. W. *et al.* Simplified mammalian dna isolation procedure. *Nucleic acids research* **19**, 4293 (1991).
72. Sirica, R. *Analysis of a genomic site in the ABCA12 gene under positive selection in Europeans and Asians*. Ph.D. thesis, Universita della Campania Luigi Vanvitelli (2017).

## Acknowledgements

We would like to thank all the anonymous volunteers that participated in the project. We also thank Cristina D’Aniello for helping with DNA extraction protocol and Erik Garrison for assisting sampling. A special thanks to Lucia De Martino for helping to submit the project to the ethical committee. C.T.S., Q.A., and Y.X. were supported by the Wellcome Trust grant 098051. C.M. and D.A. were supported by Telethon grant GGP16235. This article is based on chapters of the doctoral thesis of the first author.

## Author Contributions

V.C., C.M., Q.A., C.T.S., M.S. and G.D.A. conceived the experiment(s). V.C., R.S., M.B., V.P., L.S., D.T., D.A., O.G., G.A. and H.K. conducted the experiments. V.C., R.S., M.S. and G.D.A., analysed the results. V.C., R.S., M.B., Y.X., Q.A., C.T.S., M.S. and G.D.A. worked at the interpretation of the results. V.C., M.S. and G.D.A. wrote the manuscript. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-40360-9>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019