Genome-wide association study in 176,678 Europeans reveals genetic loci for tanning response to sun exposure

Alessia Visconti1, David L. Duffy2, Fan Liu3,4,5, Gu Zhu2, Wenting Wu6, Yan Chen3,4, Pirro G. Hysi1, Changqing Zeng3, Marianna Sanna1, Mark M. Iles7, Peter A. Kanetsky8, Florence Demenais9,10, Merel A. Hamer11, Andre G. Uitterlinden12,13, M. Arfan Ikram13, Tamar Nijsten11, Nicholas G. Martin2, Manfred Kayser5, Tim D. Spector1, Jiali Han6,14, Veronique Bataille1,15 & Mario Falchi1

The skin's tendency to sunburn rather than tan is a major risk factor for skin cancer. Here we report a large genome-wide association study of ease of skin tanning in 176,678 subjects of European ancestry. We identify significant association with tanning ability at 20 loci. We confirm previously identified associations at six of these loci, and report 14 novel loci, of which ten have never been associated with pigmentation-related phenotypes. Our results also suggest that variants at the AHR/AGR3 locus, previously associated with cutaneous malignant melanoma the underlying mechanism of which is poorly understood, might act on disease risk through modulation of tanning ability.
un exposure has consistently been associated with increased risk of all skin cancers, including cutaneous malignant melanoma (CMM), basal cell carcinoma and squamous cell carcinoma. These are the most common types of cancer in European populations of which the incidence rate is higher in fair-skinned people, rather than darker-skinned people. The tanning response after exposure to sunlight is mainly determined by melanin pigmentation, and aims at protecting the skin from DNA photodamage. Tanning response shows large variability within and between populations. The heritability of ease of skin tanning attributable to common genetic variation in the UK Biobank sample has been estimated to be 0.454 ± 0.006.

Results

Genome-wide association analyses. Ease of skin tanning and genotype data were available for 121,296 individuals of European ancestry from the UK Biobank (UKBB), which were divided in two groups according to their skin’s ability to tan, with 38.6% of the individuals reporting that they never tan and only burn, or get mildly or occasionally tanned (Methods; Supplementary Table 1). We carried out a genome-wide association analysis at 8,351,141 SNPs, assuming an additive genetic model, with sex included as covariate, and applying PCA-based correction to address potential population stratification (Methods; Supplementary Fig. 1). We identified 10,834 SNPs passing genome-wide significance ($P < 5.0 \times 10^{-8}$) that mapped to 30 distinct loci (Fig. 1, Supplementary Table 2, Supplementary Data 1 and Supplementary Fig. 2 and 3). The genetic inflation factor $\lambda_{GC}$ was 1.10. Therefore, to rule out the possibility that some associations could be driven by confounding biases, we used the LD score regression approach, which confirmed the presence of an underlying polygenic architecture (intercept = 1.04 ± 0.01).

We attempted to replicate these 30 loci in five additional cohorts of European ancestry for which data on tanning ability were available ($N = 55,382$; Methods; Supplementary Tables 3–8). Meta-analysis of the results in these five replication cohorts confirmed, at a Bonferroni-corrected threshold of $0.05/30 = 1.67 \times 10^{-3}$, the association at 20 of the 30 top-associated SNPs at each locus. The replicated loci encompass six genes previously identified several DNA variants involved in tanning ability or in skin sensitivity to sunlight, encompassing seven genes, namely ASIP, EXOC2, HERC2, IRF4, MC1R, SLC45A2 and TYR – that are already known to be associated with both pigmentation-related traits, such as hair, eye or skin colour, and skin cancer. However, it has already been observed that while some loci exert an effect on both pigmentation and tanning ability, others have a more specific effect. This also suggests two pathways for skin cancer development: via pigmentation or independent of pigmentation.

To further investigate the genetic basis of skin tanning in Europeans and their effect on skin cancer susceptibility, we performed a large-scale GWAS using data from the UK Biobank, identifying ten novel associations, and replicating ten genes previously associated with ease of skin tanning or pigmentation-related phenotypes. Additionally, we show a genetic correlation between ease of skin tanning and non-melanoma skin cancer, and highlight shared genetic effects between variants at AHR/AGR3 and tanning ability and CMM risk.
associated with tanning ability, four genes previously associated with pigmentation-related traits, and ten novel associations (Table 1, Supplementary Tables 2 and 9, and Supplementary Fig. 4–23). Conditional association analysis highlighted 14 further independent genome-wide significant associations, which were replicated in the additional cohorts (Supplementary Table 10).

Sex differences. It has been repeatedly observed that adult females are fairer-skinned than males1,2. In this study, given the large sample size and the highly significant effects identified, we further investigated the sex-by-SNP interaction for the 20 loci associated with tanning ability. We identified significant interactions ($P < 2.5 \times 10^{-10}$) at five different loci in the UKBB data set, which, however, were not confirmed in the additional cohorts (Supplementary Table 11).

Genome-wide significant loci. Six of the identified loci included genes associated with tanning ability in previous GWASs (NHGRI–EBI GWAS Catalog, release 26 June 2017): HERC2/ OCA2, IRF4, MC1R, RALY/ASIP, SLC45A2 and TYR3–6. Additionally, we identified four other genes whose evidence of association with pigmentation-related traits had previously been limited only to skin pigmentation (BNC229), hair colour (TPCN2) or both hair and eye colour (SLC24A4 and TYRP1)4,18. The high statistical power provided by the large UKBB sample size (>80% for a common variant with minor allele frequency of 0.5 and odds ratio >1.06 at α-level 5 × 10^{-8}) allowed for the identification of low-penetrance common DNA variants of small effect size at ten novel genes. Five of these novel genes are known for their involvement in melanin synthesis: DCT (TYRP2) which catalyses melanin production19, EMX2 and PPARGC1B, regulators of the melanocyte-specific transcription factor (MITF)19,20, and PDE4B and RIPK5, which have been suggested to be MITF targets21,22. Five associations do not harbour any obvious candidate for ease of skin tanning nor pigmentation: AHR/AGR3 ATP11A, TRPS1, KIAA0930 and an intergenic region flanking the PA2G4P pseudogene, whose functional role in tanning ability would require further investigation. Interestingly, KIAA0930 encodes the Q6IICG6 protein that interacts with the tyrosine 3-monoxygenase/tryptophan 5-monoxygenase activation protein, which have been suggested to be important in melanocyte development in a mouse model23.

Among the genes identified by the conditional analysis, namely CHM1PA, FANCA, SPIRE2, DEFB, AFG3L1P, DBND1 and PRDM7, DBND1 has been previously associated with tanning ability2, while mutations in the FANCA gene are responsible for Fanconi anaemia, a disease characterised by short stature, bone marrow failure and increased risk of multiple cancers. Affected individuals present areas of skin hypopigmentation and café au lait spots with abnormal melanin deposition.

Effect of ease of skin tanning variants on skin cancer. Several of the identified genes have previously been associated with increased susceptibility to skin cancer. Common DNA variants at the AFG3L1P, AGR3, HERC2/OCA2, MC1R, RALY/ASIP, SLC45A2 and TYR genes have been linked to CMM11,24, and common DNA variants at the AHR, ASIP, BNC2, DEFB, IRF4, MC1R, OCA2, TYR and SLC45A2 genes have been associated with non-melanoma skin cancer risk12,13,25. ATP11A was significantly associated in the first stage of a large GWAS including 12,945 self-reported basal cell carcinoma cases and 274,252 controls of European ancestry from the 23andMe database (but failed replication in the final meta-analysis13, with top association at rs1765871 (P = 4.9 × 10^{-9}) in strong linkage disequilibrium with our top-associated SNP (rs1046793; $r^2 = 0.90$). Suggestive association with skin cancer has also been reported for TRPS122. Additionally, common variants at the FANCA gene have been suggested to predict melanoma survival26.

While the number of melanoma cases in the UKBB data set was not sufficient to generate a reliable estimate of the genetic
correlation between ease of skin tanning and CMM, we could estimate the genetic correlation of ease of skin tanning with non-melanoma skin cancer, which amounted to $\rho = 0.33$ (SD = 0.16, $p = 0.04$; Methods, Supplementary Tables 12 and 13, Supplementary Figs. 24 and 25, and Supplementary Data 2).

The present study also shows association with ease of skin tanning for the locus encompassing AGR3/AHR, which has previously been associated with CMM risk11 but not with tanning ability nor pigmentation. Our association at AGR3/AHR overlaps with previous findings for CMM risk, with rs1721028 (secondary association at AGR3/AHR) showing $r^2 > 0.8$ with rs1636744, identified by Law et al.11. We further investigate whether variants at this locus exert a shared effect on tanning ability and CMM using a Bayesian bivariate approach27 (Methods). The AGR3/AHR posterior probability for shared genetic effects between tanning ability and CMM was 1, thus indicating that the same genetic variants are involved in both decreased tanning ability and increased risk of melanoma.

### Effects of ease of skin tanning variants on hair colour

We further investigated the effects of loci associated with tanning ability in our study on hair colour, which was also characterised in the UKBB data set. Specifically, we separately assessed the association with non-red hair colour, and with red versus non-red hair colour (Methods; Supplementary Tables 14 and 15). Most of the genes that have been previously associated with non-red hair colour4,6,8,18 (i.e., IRF4, HERC2/OCA2, SLC24A4, SLC45A2, TPCN2 and TYRP1) showed a stronger association with non-red hair colour than with ease of skin tanning, with the exception of MC1R, RALY/ASIP and TYR (Supplementary Table 16). We additionally observed three genome-wide significant associations between non-red hair colour and BNC2 (previously associated with freckles8 and facial pigmentation9), DCT, and RIPK5, which were confirmed and replicated in a large meta-analysis study28. A marginal association was observed at KIAA0930. Red hair showed significant association with IRF4, HERC2/OCA2, MC1R and RALY/ASIP, while a marginal association was observed at SLC45A2 (Supplementary Table 17). Only MC1R was more strongly associated with red hair colour than with ease of skin tanning. Overall, seven loci were exclusively associated with tanning ability (AHR/AGR3, ATP11A, EMX2, PA2G4P4, PDE4B, PPARGC1B and TRPS1).

### Discussion

In this large-scale GWAS of ease of skin tanning in 176,678 Europeans, we have replicated ten loci previously associated with ease of skin tanning or pigmentation-related phenotypes (BNC2, HERC2/OCA2, IRF4, MC1R, RALY/ASIP, SLC24A4, SLC45A2, TPCN2, TYR and TYRP1), and identified ten novel associations (AHR/AGR3, ATP11A, DCT, EMX2, KIAA0930, PA2G4P4, PDE4B, PPARGC1B, RIPK5 and TRPS1).

The tanning response is determined by an increase in melanin production in melanocytes stimulated by ultraviolet radiation. We identified five novel associations between ease of skin tanning and genes previously suggested to be involved in the melanin synthesis pathway (DCT, EMX2, PDE4B, PPARGC1B and RIPK5)19–22. The relative proportion of eumelanin (dark/brown pigment) and pheomelanin (red/yellow pigment) is regulated by the MC1R gene, which is highly polymorphic in the European population31. Different low/intermediate-frequency variants in this gene have been associated with multiple pigmentation-related phenotypes3,8,9 and both melanoma11,24,30 and non-melanoma skin cancer12,13. Taking advantage of the large UKBB sample size, we tested the association between the nine most studied variants at MC1R29 and ease of skin tanning. Seven of these variants were highly significantly associated with ease of skin tanning, and they could fully explain our top GWAS association at the MC1R locus.

Non-melanoma skin cancer is the most common type of cancer in the United Kingdom and is well represented in the UKBB data set. Our analyses showed genetic correlation between non-melanoma skin cancer and ease of skin tanning in the UKBB data set. Overall, eight loci associated with ease of skin tanning in our study have been previously shown to influence the susceptibility to either non-melanoma skin cancer (BNC2 and IRF4)12,13 or both non-melanoma skin cancer and CMM (AHR/AGR3, HERC2/OCA2, MC1R, RALY/ASIP, SLC45A2 and TYR)11–13. Additionally, suggestive associations have been previously reported between common variants at ATP11A and TRPS1 and non-melanoma skin cancer12,13. Our association for ease of skin tanning at AHR/AGR3 overlapped with the association with CMM identified by Law et al.11. Combined analysis of our and Law et al. GWASs results indicated that the same genetic variants at the AGR3/AHR locus were driving the associations at both
phenotypes, thus suggesting that common variants at this locus might increase CMM susceptibility through modulation of ease of skin tanning. Taken together, these results broaden the spectrum of DNA variants involved in ease of skin tanning, and potentially modulating skin cancer risk.

Finally, functional enrichment analysis at the associated DNA variants showed an enrichment for cis-eQTLs identified at 5% FDR in sun-exposed skin tissue and fibroblast from the GTX consortium project31, and for regulatory elements from the the Roadmap consortium project32, consistently with the observation that variants identified through GWASs are more likely to act through effects on gene regulation rather than directly altering protein-coding sequences.

Methods

Discovery cohort. The UK Biobank33 (UKBB) is a prospective cohort study of over 500,000 individuals from across the United Kingdom, aged between 40 and 70. Blood, urine and saliva samples were collected and each participant answered an extensive questionnaire on health and lifestyle. Written informed consent was obtained from each participant, in accordance with the Declaration of Helsinki.

Genotyping and imputation. A total of 152,736 samples belonging to the interim data release of UK Biobank genetic data (May 2015) were genotyped using a combination of two arrays: the UK Biobank Axiom array from Affymetrix (N = 102,754), that was specifically designed for the purpose of genotyping the UK Biobank participants, and the UK BILEVE array (N = 49,982), that was designed to study the genetics of lung health and disease. The UK10K haplotype reference and the 1000 Genomes Phase 3 reference panels were merged and used as reference panel in the IMPUTE2 software34. Kinship coefficients for all pairs were calculated using KING’s robust estimator35 and used to identify and remove related individuals. Imputation, relatedness assessment and quality control were performed by the analysis group at the Wellcome Trust Centre for Human Genetics, University of Oxford. Details are provided at the UK Biobank website (http://biobank.ccuu. ox.ac.uk). A total of 8,351,141 SNPs meeting the following conditions were included in our genome-wide association study: call rate ≥95%, minor allele frequency (MAF) ≥1% and Hardy–Weinberg equilibrium test with P ≥ 1 × 10−6.

Phenotyping. Ease of skin tanning was collected for 460,922 individuals with self-reported European ethnicity, 140,749 of whom had genotype data. Individuals that reported themselves to be of European ancestry but described their skin colour as ‘black’ were removed from the data set. As carried out in previous studies of similar phenotypes34 and in order to reduce misclassifications in the self-reported data, individuals were reporting that they never tanned only burn, or get mildly only occasionally tanned were included in the group of individuals showing a low tan response, while people who reported getting moderately or very tanned were included in the group of individuals showing a high tan response. Reliability of the reported tanning ability was cross-verified using self-reported sunburns. Furthermore, individuals reporting themselves to have red hair yet declaring that they get very or moderately tanned were included in the group of individuals showing a low tan response, while people who reported usually or always getting tanned were included in the group of individuals showing a high tan response. The study included 10,451 individuals via questionnaires, and individuals who reported getting easily burned while in the sun were considered as having a low tan response (Supplementary Table 4). All Rotterdam Study participants have provided written informed consent. The study has been approved by the medical ethics committee according to the World Medical Association’s Declaration of Helsinki (Population Study Act Rotterdam Study), executed by the Ministry of Health, Welfare and Sports of the Netherlands.

The Australian Study from the Queensland Institute of Medical Research (QIMR) comprises twins and their family members taking part in a long-standing study of melanoma risk factors44. Two independent samples were used for this analysis. The first included adolescent twins, their siblings and parents, and the second a collection of adult twins. These were genotyped in two phases. Phase 1 samples were genotyped using Illumina Human610-Quadv1_B arrays at deCODE Genetics, Iceland. For imputation, they were merged with a larger set of individuals genotyped on Human610-Quadv1_B and Human660W-Quad_v1_C arrays. Phase 2 samples were genotyped on HumanOmniExpress-12v1-1_A, HumanOmni25M-8v1-1_B and HumanCoreExome 12v1-0_C arrays from Illumina at the Diamantina Institute, University of Queensland. Self-reported ease of skin tanning was collected on 5509 individuals from the adolescents plus parents and on 2248 individuals from the adult twins, but only a total of 5149 individuals had genotype data and were included in this analysis (Supplementary Tables 5 and 6). Individuals reporting that they never tanned only burn, or usually burn and sometimes tanned were included in the group of individuals showing a low tan response, while people who reported of usually or always getting tanned were included in the group of individuals showing a high tan response. The study protocol was approved by appropriate institutional review boards, and all participants have provided written informed consent.

The Nurses’ Health Study (NHS), the Nurses’ Health Study 2 (NHS2) and the Health Professionals Follow-Up Study (HPFS) are three large prospective cohort studies of US men and women of European ancestry (HS). Ease of skin tanning was assessed on 35,845 individuals by asking what kind of tan was developed after repeated sun exposures (e.g., a 2-week vacation outdoors) during childhood or adolescence, and categorised into a binary variable (Supplementary Table 7). Subjects were genotyped on multiple arrays (Affymetrix, Illumina HumanHap, Illumina OmniExpress, HumanCoreExome and HumanCoreExome II HumanHap550 K and Human 610 Quad Arrays of Illumina). Imputation was performed using the MaCH and minicmac software packages and the 1000 Genomes Project Phase 1, integrated variant set across 1092 individuals, v2, March 2012) as described elsewhere. Ease of skin tanning was collected on 10,451 individuals via questionnaires, and individuals who reported getting easily burned while in the sun were considered as having a low tan response (Supplementary Table 4). All Rotterdam Study participants have provided written informed consent. The study has been approved by the medical ethics committee according to the World Medical Association’s Declaration of Helsinki (Population Study Act Rotterdam Study), executed by the Ministry of Health, Welfare and Sports of the Netherlands.

Identification of independent signals within loci. We used a stepwise procedure to identify independent signals within the loci identified in the UKBB sample. Specifically, we extended each locus to include a 1 Mb flanking region on either side and fitted a new regression model, where the top-associated genome-wide significant SNP was included as a covariate (conditional model). We considered the genome-wide significant (P < 5 × 10−8) top-associated SNP resulting from the conditional model as an independent signal, and included it in the covariate set of a new conditional model. We stopped the stepwise procedure when we could not identify any additional genome-wide significant SNPs. Conditional models were build using PLINK36 (version 1.90 b3.38).

Replication cohorts. The TwinsUK cohort includes more than 13,000 monozygotic and dizygotic twin volunteers from all regions across the United Kingdom37. The phenotype was collected via nurse-administered questionnaires using a Fitzpatrick classification scheme. Individuals were dichotomised into two categories, low (sun-reactive skin type I and II) and high (sun-reactive skin type III and IV) tan response. Phenotypic and genotypic data were available for 3937 female individuals with European ancestry (Supplementary Table 3). Microarray genotyping was conducted using a combination of Illumina arrays (HumanHap300, HumanHap610, 1M-Duo and 12M-Duo 1 M) and imputation was performed using the IMPUTE2 software using haplotype information from the 1000 Genomes Project (Phase 1, integrated variant set across 1092 individuals, v2, March 2012), as previously described41. Guy’s and St Thomas’ Hospital NHS Trust Research Ethics Committee approved the study, and all twins provided informed written consent. The Australian Study from the Queensland Institute of Medical Research (QIMR) comprised twins and their family members taking part in a long-standing study of melanoma risk factors44. Two independent samples were used for this analysis. The first included adolescent twins, their siblings and parents, and the second a collection of adult twins. These were genotyped in two phases. Phase 1 samples were genotyped using Illumina Human610-Quadv1_B arrays at deCODE Genetics, Iceland. For imputation, they were merged with a larger set of individuals genotyped on Human610-Quadv1_B and Human660W-Quad_v1_C arrays. Phase 2 samples were genotyped on HumanOmniExpress-12v1-1_A, HumanOmni25M-8v1-1_B and HumanCoreExome 12v1-0_C arrays from Illumina at the Diamantina Institute, University of Queensland. Self-reported ease of skin tanning was collected on 5509 individuals from the adolescents plus parents and on 2248 individuals from the adult twins, but only a total of 5149 individuals had genotype data and were included in this analysis (Supplementary Tables 5 and 6). Individuals reporting that they never tanned only burn, or usually burn and sometimes tanned were included in the group of individuals showing a low tan response, while people who reported of usually or always getting tanned were included in the group of individuals showing a high tan response. The study protocol was approved by appropriate institutional review boards, and all participants have provided written informed consent.

The Nurses’ Health Study (NHS), the Nurses’ Health Study 2 (NHS2) and the Health Professionals Follow-Up Study (HPFS) are three large prospective cohort studies of US men and women of European ancestry (HS). Ease of skin tanning was assessed on 35,845 individuals by asking what kind of tan was developed after repeated sun exposures (e.g., a 2-week vacation outdoors) during childhood or adolescence, and categorised into a binary variable (Supplementary Table 7). Subjects were genotyped on multiple arrays (Affymetrix, Illumina HumanHap, Illumina OmniExpress, HumanCoreExome and HumanCoreExome II HumanHap550 K and Human 610 Quad Arrays of Illumina). Imputation was performed using the MaCH and minicmac software packages and the 1000 Genomes Project Phase 1, integrated variant set across 1092 individuals, v2, March 2012) as described elsewhere. Ease of skin tanning was collected on 10,451 individuals via questionnaires, and individuals who reported getting easily burned while in the sun were considered as having a low tan response (Supplementary Table 4). All Rotterdam Study participants have provided written informed consent. The study has been approved by the medical ethics committee according to the World Medical Association’s Declaration of Helsinki (Population Study Act Rotterdam Study), executed by the Ministry of Health, Welfare and Sports of the Netherlands.

Meta-analysis. Association studies in the TwinsUK and QIMR cohorts were conducted for independent SNPs using a linear mixed model as implemented in Merlin38 and GEMMA46, respectively, in order to take into account the non-independence of twins data. Association studies in the HK cohort were conducted using a logistic regression model adjusting for sex, age and the first five genotype principal components. Specifically, in the HS cohort, associations in
each component GWAS set (Affymetrix, Illumina HumanHap series, Illumina OmniExpress, HumanCore Exome, and OncoArray) were combined in an inverse-variance-weighted meta-analysis using the METAL software\textsuperscript{42}. Meta-analyses of the results obtained in the replication cohorts were carried out using a weighted $Z$-score method based on sample size, $P$ value and direction of effect in each study as implemented in the METAL software\textsuperscript{43}. We considered an association replicated when the meta-analysis $P$ value reached a Bonferroni-adjusted significance threshold of $P < 0.05/30 = 1.67 \times 10^{-3}$.

**SNP-by-sex interaction.** We assessed SNP-by-sex interaction in the UKBB data set for all replicated SNPs using PLINK\textsuperscript{36} (version 1.90 b38). The first five principal components assessed on the genomic data were included as covariates to control for potential stratification issues. We considered an interaction term significant when its $P$ value reached a Bonferroni-adjusted significance threshold of $P < 0.05/20 = 2.5 \times 10^{-3}$. We then attempted replication using four of our replication cohorts (the TwinsUK cohort included only female individuals). The meta-analysis in the replication cohorts was performed using a weighted $Z$-score method based on sample size, $P$ value and direction of effect in each study as implemented in the METAL software. We considered the interaction replicated when the meta-analysis $P$ value was smaller than the Bonferroni-adjusted significance threshold of $P < 0.05/5 = 0.01$.

**Power calculations.** Power was assessed with Genetic Power Calculator\textsuperscript{38} assuming a prevalence of 40% for increased tanning ability—as observed in both the UKBB (42%) and TwinsUK (38%) data sets.

**Identification of known loci.** We interrogated the NHGRI-EBI GWAS catalogue\textsuperscript{39} (release 26 June 2017; association $P < 5 \times 10^{-8}$) to identify overlap between SNPs identified in our study (or in tight linkage disequilibrium with them; $r^2 > 0.8$) and previously reported associations for ease of skin tanning, pigmentation-related phenotypes and skin cancer.

**Genetic correlation between tanning ability and skin cancer.** The 2017 release of the UK Biobank genetic data included a further 367,186 individuals. We removed 968 individuals flagged because of low genetic data quality, 120,502 individuals who were estimated to be genetically related and 57,157 individuals affected by any cancer, either malignant or in situ, resulting in 907 CMM (ICD-10 code: C43) and 5912 non-melanoma skin cancer (ICD-10 code: C44) cases, and 181,740 controls (Supplementary Table 12). Genotype data were processed as described in ‘Genotype data processing’. We assessed SNP-by-sex interaction in the UKBB data set using a logistic regression approach, as implemented in PLINK (version 2.00), to estimate the genetic correlation between ease of skin tanning and cancer. We applied a Bayesian bivariate analysis as implemented in GWAS-PW\textsuperscript{27} to investigate whether loci here associated with ease of skin tanning and previously involved in melanoma risk were associated with both traits or (iv) that the genomic block includes two genetic variants, (i) associated only with CMM, (ii) associated only with tanning ability, (iii) associated with both traits or (iv) that the genomic block includes two genetic variants, associating independently with each of the two traits.

**Association study with natural hair colour.** Self-reported hair colour was assessed via questionnaire in 118,777 out of 121,296 individuals used in this study (Supplementary Tables 14 and 15). To assess association between hair colour and the loci replicated in our study of ease of skin tanning, we used (a) a linear regression model to test association with non-red-haired individuals where blonde = 1, light brown = 2, dark brown and black = 3, and (b) a logistic regression model to test association with red versus non-red hair colour. Both linear and logistic regression models were performed using PLINK\textsuperscript{39} (version 1.90 b38) assuming an additive genetic model and including age, sex and the first five principal components assessed on the genomic data as covariates.

**GTeX cis-eQTL analysis.** To study whether the replicated SNPs have a regulatory effect on gene expression levels, we used expression quantitative trait loci (eQTLs) data in three skin tissues from the GTeX consortium project\textsuperscript{15} (data release v6), namely fibroblasts, sun-exposed skin (lower leg) and sun-non-exposed skin (suprapubic). cis-eQTLs were defined by the GTeX consortium as SNPs 1 MB around the transcript start site passing 5% FDR\textsuperscript{11}. We extended the set of 34 independent SNPs replicated for ease of skin tanning by including any SNP in high linkage disequilibrium ($r^2 \geq 0.8$) with them using SNAP Proxy Search\textsuperscript{31} and data from both the HapMap 3 (release 2) and 1000 Genome (pilot 1) projects. We then evaluated an empirical enrichment $P$ value by comparing the overlap between the set of cis-eQTLs in the GTeX project database and the original extended set of SNPs with the overlap obtained using 1000 random sets of SNPs created using a cyclic permutation procedure\textsuperscript{32}.

**Analysis of epigenetic marks.** By following the same procedure described above for the cis-eQTL enrichment analysis, we additionally assessed enrichment and depletion of epigenetic markers by using histone marks and DNA accessibility peak data in epithelial foreskin melanocyte primary cells from the Roadmap consortium project\textsuperscript{38} (data release v9). We focused on a core set of histone modifications (namely: H3K4me1, H3K27me3, H3K36me3, H3K9me3 and H3K27ac) and DNA-seq accessibility data. Since the histone modification data for the studied cell line was available from two donors, we averaged the overlaps among samples.

**Data availability.** Supplementary data that support the findings of this study have been deposited in Zenodo with the https://doi.org/10.5281/zenodo.1194289 at the address https://doi.org/10.5281/zenodo.1194289. The association summary statistics for the 10,834 genome-wide significant SNPs are provided as Supplementary Data 1. Genomic coordinates are reported in GRCh37.p13. The association summary statistics for the SNPs associated in the UKBB non-melanoma skin cancer GWAS ($P < 1 \times 10^{-8}$) are provided as Supplementary Data 2. Genomic coordinates are reported in GRCh37.p13. The cis-eQTLs identified in the three studied skin tissues, and available within the GTeX project, are provided as Supplementary Data 3.

Received: 14 December 2016 Accepted: 3 April 2018

Published online: 08 May 2018

**References**


