

# Genome-wide association studies identify several new loci associated with pigmentation traits and skin cancer risk in European Americans

Mingfeng Zhang<sup>1</sup>, Fengju Song<sup>1,3</sup>, Liming Liang<sup>4,5</sup>, Hongmei Nan<sup>2,6</sup>, Jiangwen Zhang<sup>7</sup>, Hongliang Liu<sup>8</sup>, Li-E. Wang<sup>8</sup>, Qingyi Wei<sup>8</sup>, Jeffrey E. Lee<sup>9</sup>, Christopher I. Amos<sup>10</sup>, Peter Kraft<sup>4,5</sup>, Abrar A. Qureshi<sup>1,2</sup> and Jiali Han<sup>1,2,3,4,\*</sup>

<sup>1</sup>Clinical Research Program, Department of Dermatology and <sup>2</sup>Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA, <sup>3</sup>Department of Epidemiology, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China, <sup>4</sup>Department of Epidemiology and <sup>5</sup>Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA, <sup>6</sup>Division of Cancer Epidemiology, Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD, USA, <sup>7</sup>FAS Research Computing, Harvard University, Cambridge, MA, USA, <sup>8</sup>Department of Epidemiology and <sup>9</sup>Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA and <sup>10</sup>Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Hanover, NH, USA

Received December 29, 2012; Revised March 4, 2013; Accepted March 22, 2013

**Aiming to identify novel genetic loci for pigmentation and skin cancer, we conducted a series of genome-wide association studies on hair color, eye color, number of sunburns, tanning ability and number of non-melanoma skin cancers (NMSCs) among 10 183 European Americans in the discovery stage and 4504 European Americans in the replication stage (for eye color, 3871 males in the discovery stage and 2496 males in the replication stage). We targeted novel chromosome regions besides the known ones for replication. As a result, we identified a new region downstream of the *EDNRB* gene on 13q22 associated with hair color and the strongest association was the single-nucleotide polymorphism (SNP) rs975739 ( $P = 2.4 \times 10^{-14}$ ;  $P = 5.4 \times 10^{-9}$  in the discovery set and  $P = 1.2 \times 10^{-6}$  in the replication set). Using blue, intermediate (including green) and brown eye colors as co-dominant outcomes, we identified the SNP rs3002288 in *VASH2* on 1q32.3 associated with brown eye ( $P = 7.0 \times 10^{-8}$ ;  $P = 5.3 \times 10^{-5}$  in the discovery set and  $P = 0.02$  in the replication set). Additionally, we identified a significant interaction between the SNPs rs7173419 and rs12913832 in the *OCA2* gene region on brown eye color ( $P$ -value for interaction =  $3.8 \times 10^{-3}$ ). As for the number of NMSCs, we identified two independent SNPs on chr6 and one SNP on chromosome 14: rs12203592 in *IRF4* ( $P = 7.2 \times 10^{-14}$ ;  $P = 1.8 \times 10^{-8}$  in the discovery set and  $P = 6.7 \times 10^{-7}$  in the replication set), rs12202284 between *IRF4* and *EXOC2* ( $P = 5.0 \times 10^{-8}$ ;  $P = 6.6 \times 10^{-7}$  in the discovery set and  $P = 3.0 \times 10^{-3}$  in the replication set) and rs8015138 upstream of *GNG2* ( $P = 6.6 \times 10^{-8}$ ;  $P = 5.3 \times 10^{-7}$  in the discovery set and  $P = 0.01$  in the replication set).**

## INTRODUCTION

Natural hair color, eye color, skin color and sun sensitivity reflect the variation in human basal and transient pigmentation, which has been known to be highly heritable (1,2).

More recently, dozens of common genetic variants have been identified to be associated with pigmentary phenotypes from the genome-wide association studies (GWASs) and were reviewed by Gerstenblith *et al.* (3). These GWASs identified genetic variants in the genes already known to be

\*To whom correspondence should be addressed at: 181 Longwood Avenue, Boston, MA 02120, USA. Tel: +1 6175252098; Fax: +1 6175252088; Email: jiali.han@channing.harvard.edu

associated with pigmentation, such as melanocortin 1 receptor (*MC1R*) (4,5), oculocutaneous albinism II (*OCA2*) (4), solute carrier family 45, member 2 (*SLC45A2*) (4), tyrosinase (*TYR*) (5,6), tyrosinase-related protein 1 (*TYRP1*) (6), kit ligand (*KITLG*) (5) and the agouti signaling protein (*ASIP*) (6), and also identified novel potential pigmentation genes, such as solute carrier family 24, member 4 (*SLC24A4*) (4,5), two-pore segment channel 2 (*TPCN2*) (6) and interferon regulatory factor 4 (*IRF4*) (4). Of interest, some of the genetic determinants are shared across different pigmentation traits, whereas others are distinct, suggesting various mechanisms of pigmentation across different organs (3). Europeans have the majority of variation in pigmentation characteristics compared with other populations, especially in hair and eye colors (5). In this study, we conducted a series of GWASs on hair color, eye colors, tanning ability and the number of sunburns in European Americans, aiming to identify novel genetic loci for pigmentation and to understand the variability of complex pigmentation characteristics in the general population to a greater extent. In addition, given that human pigmentation traits are susceptibility factors for skin cancer (3,7), we further evaluated the newly identified pigmentation loci in association with the risk of melanoma and non-melanoma skin cancer (NMSC). We also conducted a GWAS on the number of NMSCs in this study to identify new loci for the risk of NMSCs (8–10).

## RESULTS

The samples used in each GWAS of the present study are summarized in Supplementary Material, Table S1. The quantile–quantile plots based on the meta-analysis did not demonstrate a systematic deviation from the expected distribution, with the

overall genomic control inflation factors ( $\lambda_{GC}$ ) of 0.99–1.04 (Supplementary Material, Fig. S1). The Manhattan plot of each GWAS is presented in Supplementary Material, Figure S2. We focused on the novel chromosome regions suggesting promising associations besides the previously known ones and selected the single-nucleotide polymorphisms (SNPs) with the strongest associations in these regions for replication. A total of 16 SNPs were finally selected for replication, with detailed information presented in Supplementary Material, Table S2. Details of the SNP selection for replication are presented in Materials and Methods.

As a result, for the GWAS of natural hair color, the SNP rs975739, which ranked the top besides the known pigmentation SNPs in the discovery set ( $P = 5.4 \times 10^{-9}$  in the discovery set), revealed a nominal significance in both of the replication sets ( $P = 0.05$  in the skin cancer set,  $P = 3.3 \times 10^{-6}$  in the RF\_NHS set and  $P = 1.2 \times 10^{-6}$  in the combined replication set, Table 1). This association reached a  $P$ -value of  $2.4 \times 10^{-14}$  after combining the discovery set and the replication set. The variant allele (G allele) of this SNP was associated with darker hair color with the regression parameter beta of 0.08 and a standard error (SE) of 0.01. Another tag-SNP rs1146927 in this region, which was in linkage disequilibrium (LD) with rs975739 ( $r^2 = 0.84$  in this study), was also replicated in the replication set ( $P = 2.9 \times 10^{-8}$  in the discovery set,  $P = 0.03$  in the skin cancer replication set and  $P = 2.8 \times 10^{-9}$  in the combined set, Supplementary Material, Table S3). Both of the SNPs (rs975739 and rs1146927) reached the GWAS significance level and revealed a similar association pattern (beta = 0.8 and SE = 0.01 for rs975739; beta = 0.7 and SE = 0.01 for rs1146927), suggesting the same signal represented by them. The association of the other tag-SNP rs58188699 in this region (LD  $r^2 = 0.42$  with rs975739 in this study) with natural hair color ( $P = 7.5 \times$

**Table 1.** Association of the SNP rs975739 downstream of the *EDNRB* gene with non-red hair colors<sup>a</sup>

Data set	rs975739 T>G Sample size	Blond hair color	Beta (SE)	$P$ -value <sup>b</sup>	$P$ -value for heterogeneity
Discovery set					
BC_NHS (female) <sup>c</sup>	1741	223	0.07 (0.02)	4.8E – 03	
T2D_NHS (female) <sup>d</sup>	2377	302	0.07 (0.02)	1.6E – 03	
T2D_HPFS (male) <sup>e</sup>	1529	165	0.10 (0.03)	8.0E – 04	
CHD_NHS (female) <sup>f</sup>	783	97	0.06 (0.04)	0.09	
CHD_HPFS (male) <sup>g</sup>	640	87	0.08 (0.04)	0.05	
All (meta-analysis)	7070	874	0.07 (0.01)	5.4E – 09	0.92
Replication set					
Skin cancer (female and male) <sup>h</sup>	1639	183	0.05 (0.03)	0.05	
RF_NHS (female) <sup>i</sup>	2516	314	0.10 (0.02)	3.3E – 06	
All (meta-analysis)	4155	497	0.08 (0.02)	1.2E – 06	0.16
Combined set (meta-analysis)	11 225	1371	0.08 (0.01)	2.4E – 14	0.74

<sup>a</sup>Non-red hair colors were coded as: 1, blond; 2, light brown; 3, dark brown; 4, black; beta > 0 means association with darker hair color.

<sup>b</sup>Adjusted for the top four principal components of genetic variance in the discovery sets and adjusted for gender in the skin cancer replication set.

<sup>c</sup>Postmenopausal invasive breast cancer case–control study nested within the NHS.

<sup>d</sup>Type 2 diabetes case–control study nested within the NHS.

<sup>e</sup>Type 2 diabetes case–control study nested within the HPFS.

<sup>f</sup>Coronary heart disease case–control study nested within the NHS.

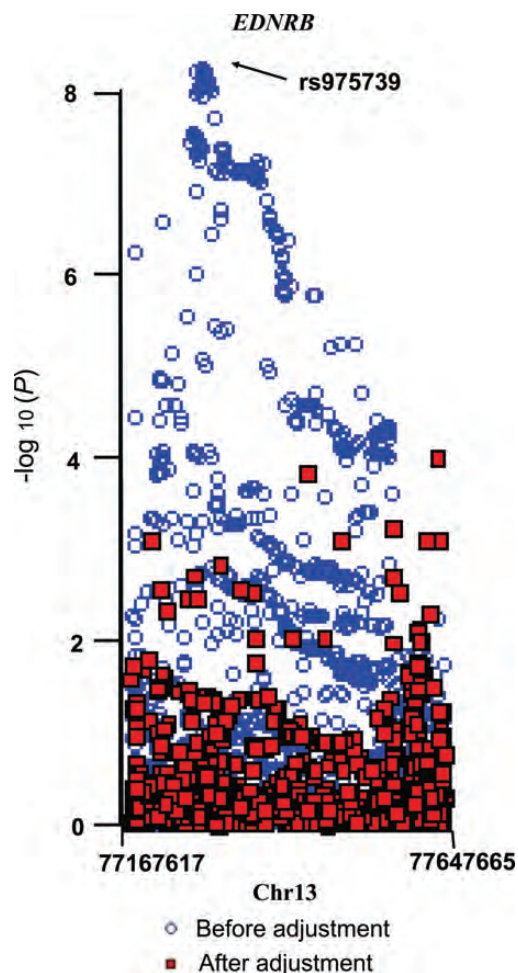
<sup>g</sup>Coronary heart disease case–control study nested within the HPFS.

<sup>h</sup>Skin cancer case–control study nested within the NHS and HPFS.

<sup>i</sup>Renal function study nested within the NHS.

$10^{-8}$  in the discovery set) was not replicated in the replication set ( $P = 0.27$  in the skin cancer replication set, Supplementary Material, Table S3).

These novel hair-color-related SNPs identified in this study are located in the region downstream of the endothelin receptor type B (*EDNRB*) gene. We additionally evaluated the association of all the SNPs within 200 kb surrounding the gene region with hair color. The regional association plot in the discovery set is presented in Figure 1. After adjustment for the SNP rs975739, the associations of 1611 SNPs in this region were substantially attenuated, and none of them remained regionally significant ( $0.05/1611 = 3.1 \times 10^{-5}$ , the smallest  $P = 1.5 \times 10^{-4}$  after adjustment, Fig. 1). We also constructed the haplotypes based on the three tag-SNPs (rs975739, rs1146927 and rs58188699). When we used the most prevalent haplotype (the combination of the common allele of rs58188699 with the rare alleles of both rs975739 and rs1146927, 38.7% of the population) as a reference, the haplotypes with the common alleles of both rs975739 and rs1146927, together with either the rare allele (37.8% of the population) or the common allele (18.9% of the population) of rs58188699, were marginally associated with lighter hair



**Figure 1.** Regional plot of the *EDNRB* gene. Manhattan plots for the GWAS of hair color in the region within 200 kb surrounding *EDNRB* on chromosome 13, before and after adjustment by the SNP rs975739.

color ( $P = 0.08$ ,  $\beta = 0.07$ ,  $SE = 0.04$  with the rare allele of rs58188699, and  $P = 0.07$ ,  $\beta = 0.05$ ,  $SE = 0.03$  with the common allele of rs58188699), suggesting that these associations were originally from the same signal as rs975739 and rs1146927. Additionally, all of the 117 SNPs with  $P < 1 \times 10^{-7}$  in the discovery stage were in LD ( $r^2 > 0.6$ ).

The SNP rs975739 was not associated with eye colors ( $P = 0.78$  for brown eye color,  $P = 0.33$  for intermediate eye color,  $P = 0.46$  for blue eye color), tanning ability ( $P = 0.72$ ) or the number of sunburns ( $P = 0.67$ ). Similar results were detected for the other two SNPs (rs1146927 and rs58188699). We also evaluated the association between the two significant SNPs and red hair color phenotype, and no significant association was found ( $P = 0.95$  for rs975739 and  $P = 0.90$  for rs1146927).

We also replicated the association between the SNP rs12421680 in the neurotrimin (*NTM*) gene on chromosome 11 and the number of sunburns ( $P = 5.3 \times 10^{-5}$  in the discovery set and  $P = 0.02$  in the replication set, Table 2). After combining the discovery and replication sets,  $P$ -value for the association of this SNP reached  $5.6 \times 10^{-6}$ . The variant allele (A allele) was associated with an increased number of sunburns ( $\beta = 0.41$ ,  $SE = 0.09$ ). Additionally, based on the transcript expression profiling data of 79 HapMap CEU cell lines (NCBI GEO database, accession GSE7792) (11), we found that the AA genotype was associated with a lower expression of *NTM* than the GG/GA genotypes ( $P < 0.001$ , Supplementary Material, Fig. S3). There was no association between this SNP and eye colors ( $P = 0.54$  for brown eye color,  $P = 0.69$  for intermediate eye color,  $P = 0.80$  for blue eye color), natural hair color ( $P = 0.55$ ) or tanning ability ( $P = 0.49$ ). Because the imputation quality of this SNP was not particularly high in four subsets in the discovery stage (imputation  $r^2$ : 0.61–0.62 for T2D\_NHS, CHD\_NHS, T2D\_HPFS and CHD\_HPFS), we re-genotyped this SNP in the subset of T2D\_HPFS using TaqMan assay to validate the result. The association of this SNP with the number of sunburns was comparable by using the re-genotyped data ( $\beta = 0.82$ ,  $SE = 0.30$ ,  $P = 0.01$ ) and the imputed data ( $\beta = 0.88$ ,  $SE = 0.38$ ,  $P = 0.02$ ).

As for the GWASs on eye color, we replicated the association of the SNP rs3002288 in vasohibin 2 (*VASH2*) with brown eye color ( $P = 7.0 \times 10^{-8}$  in the combined set;  $P = 4.3 \times 10^{-5}$  in the discovery set and  $P = 3.3 \times 10^{-5}$  in the replication set), as well as the association of the SNP rs12520016 upstream of DNA polymerase sigma (*POLS*) with intermediate (including green) eye color ( $P = 9.1 \times 10^{-7}$  in the combined set;  $P = 4.7 \times 10^{-5}$  in the discovery set and  $P = 0.03$  in the replication set, Table 3). The variant allele of rs3002288 (A allele) showed an inverse association with brown eye color ( $\beta = -0.22$ ,  $SE = 0.04$ ), and the variant allele of rs12520016 (G allele) showed a positive association with intermediate eye color ( $\beta = 0.42$ ,  $SE = 0.08$ ). The association between the SNP rs3002288 and brown eye color was approaching the GWAS significance level of  $10^{-8}$ . Neither of the two SNPs revealed an association with natural hair color ( $P = 0.37$  for rs3002288 and  $P = 0.58$  for rs12520016), tanning ability ( $P = 0.93$  for rs3002288 and  $P = 0.71$  for rs12520016) or the number of sunburns ( $P = 0.57$  for rs3002288 and  $P = 0.91$  for rs12520016).

**Table 2.** Association of the SNP rs12421680 in the *NTM* gene with the number of sunburns<sup>a</sup>

Data set	rs12421680 G>A Sample size	Sunburns 10+	Beta (SE)	<i>P</i> -value <sup>i</sup>	<i>P</i> -value for heterogeneity
Discovery set					
BC_NHS (female) <sup>b</sup>	2153	852	0.37 (0.19)	0.05	
T2D_NHS (female) <sup>c</sup>	2915	1430	0.55 (0.22)	0.01	
T2D_HPFS (male) <sup>d</sup>	2170	1075	0.88 (0.38)	0.02	
CHD_NHS (female) <sup>c</sup>	1056	668	0.41 (0.35)	0.24	
CHD_HPFS (male) <sup>i</sup>	989	631	0.67 (0.56)	0.23	
All (meta-analysis)	9283	4656	0.49 (0.12)	5.3E - 05	0.79
Replication set					
Skin cancer (female and male) <sup>e</sup>	1655	558	0.44 (0.23)	0.05	
RF_NHS (female) <sup>h</sup>	2581	891	0.24 (0.17)	0.16	
All (meta-analysis)	4236	1449	0.31 (0.14)	0.02	0.49
Combined set (meta-analysis)	13 519	6105	0.41 (0.09)	5.6E - 06	0.32

<sup>a</sup>The number of sunburns was coded as a continuous variable; beta > 0 means association with increased number of sunburns.

<sup>b</sup>Postmenopausal invasive breast cancer case-control study nested within the NHS.

<sup>c</sup>Type 2 diabetes case-control study nested within the NHS.

<sup>d</sup>Type 2 diabetes case-control study nested within the HPFS.

<sup>e</sup>Coronary heart disease case-control study nested within the NHS.

<sup>f</sup>Coronary heart disease case-control study nested within the HPFS.

<sup>g</sup>Skin cancer case-control study nested within the NHS and HPFS.

<sup>h</sup>Renal function study nested within the NHS.

<sup>i</sup>Adjusted for age and the top four principal components of genetic variance in the discovery sets and adjusted for age and gender in the skin cancer replication set.

**Table 3.** Association of the SNP rs3002288 in the *VASH2* gene with brown eye color and the SNP rs12520016 upstream of *POLS* gene with intermediate eye color<sup>a</sup>

Data set	Sample size	rs3002288 G>A			rs12520016 T>G				
		Brown eye color	Beta (SE)	<i>P</i> -value <sup>f</sup>	<i>P</i> -value for heterogeneity	Intermediate eye color	Beta (SE)	<i>P</i> -value <sup>f</sup>	<i>P</i> -value for heterogeneity
Discovery set									
T2D_HPFS (male) <sup>b</sup>	2292	717	-0.25 (0.07)	4.0E - 04		761	0.52 (0.14)	3.4E - 04	
CHD_HPFS (male) <sup>c</sup>	1064	316	-0.12 (0.11)	0.24		348	0.58 (0.19)	2.8E - 03	
KS_HPFS (male) <sup>d</sup>	515	165	-0.50 (0.15)	5.1E - 04		146	0.15 (0.31)	0.62	
All (meta-analysis)	3871	1033	-0.25 (0.06)	4.3E - 06	0.11	1109	0.49 (0.11)	4.7E - 06	0.48
Replication set									
PrCa_HPFS (male) <sup>e</sup>	2496	760	-0.18 (0.06)	3.3E - 03		805	0.29 (0.14)	0.03	
Combined set (meta-analysis)	6367	1793	-0.22 (0.04)	7.0E - 08	0.38	1914	0.42 (0.08)	9.1E - 07	0.25

<sup>a</sup>Eye colors were coded as co-dominant outcomes (brown versus non-brown eye color and green versus non-green eye color).

<sup>b</sup>Type 2 diabetes case-control study nested within the HPFS.

<sup>c</sup>Coronary heart disease case-control study nested within the HPFS.

<sup>d</sup>Kidney stone case-control study nested within the HPFS.

<sup>e</sup>Prostate cancer case-control study nested within the HPFS.

<sup>f</sup>Adjusted for the top four principal components of genetic variance in the discovery sets.

Of interest, we found that the SNP rs7173419 located in the intron region of the *OCA2* gene revealed a significant association with brown eye color (T allele, beta = 0.33, SE = 0.06;  $P = 1.3 \times 10^{-7}$  in the discovery set), whereas this association was inverted (T allele, beta = -0.48, SE = 0.08;  $P = 1.5 \times 10^{-8}$  in the discovery set) after adjustment by the top SNP in this region (rs12913832, A allele, beta = 2.82, SE = 0.10;  $P = 1.2 \times 10^{-177}$  in the discovery stage of brown eye color GWAS). The SNP rs12913832 is located upstream of *OCA2* (in the intron of *HERC2*) and known as a pigmentation SNP by controlling the expression of *OCA2* (12,13). We replicated this finding in the replication set (rs7173419, T allele; beta = 0.56, SE = 0.07,  $P = 4.1 \times 10^{-15}$  before adjustment by

rs12913832; and beta = -0.21, SE = 0.10,  $P = 0.04$  after adjustment by rs12913832). We further conducted a stratification analysis of the SNP rs12913832 by the genotypes of rs7173419 among the pooled population of the discovery set and the replication set. As shown in Table 4, the effect of the SNP rs12913832 on brown eye color decreased with more counts of the variant allele (T allele) of rs7173419, and the interaction of these two SNPs was significant ( $P = 3.8 \times 10^{-3}$ ).

As for the number of NMSCs, we identified two independent SNPs on chromosome 6 (LD  $r^2 = 0.12$ ): rs12203592 in *IRF4* with a *P*-value of  $7.2 \times 10^{-14}$  in the combined set (T allele, beta = 0.21, SE = 0.03;  $P = 1.8 \times 10^{-8}$  in the

**Table 4.** Interaction between the SNPs rs7173419 and rs12913832 on brown eye color<sup>a</sup> and melanoma risk

rs7173419	Eye color		rs12913832 G>A <sup>b</sup>		P-value <sup>c</sup>	P-value for interaction
	Sample size	Brown eye color, % (SD)	Beta	SD		
CC	3755	26.3 (0.7)	3.92	0.13	6.9E - 215	3.8E - 03
CT	2178	36.1 (1.0)	2.54	0.12	6.9E - 103	
TT	337	45.4 (2.7)	2.18	0.26	6.6E - 17	
rs7173419	Melanoma risk		rs12913832 G>A <sup>b</sup>		P-value <sup>d</sup>	P-value for interaction
	Cases	Controls	OR	95% CI		
CC	1134	610	0.64	0.53–0.77	3.6E - 06	0.06
CT	592	363	0.76	0.61–0.94	0.01	
TT	78	53	1.05	0.63–1.73	0.86	

<sup>a</sup>Eye colors were coded as brown versus non-brown eye color.

<sup>b</sup>rs12913832 was coded in an additive model, with G as wild-type allele and A as variant allele.

<sup>c</sup>Adjusted for the top four principal components.

<sup>d</sup>Adjusted for age, gender and the top two principal components.

discovery stage and  $P = 6.7 \times 10^{-7}$  in the replication stage) and rs12202284 between *IRF4* and exocyst complex component 2 (*EXOC2*) with a  $P$ -value of  $5.0 \times 10^{-8}$  in the combined set (A allele, beta = 0.15, SE = 0.03;  $P = 6.6 \times 10^{-7}$  in the discovery stage and  $P = 3.0 \times 10^{-3}$  in the replication stage; Table 5). After mutual adjustment by each other, the association of each SNP with the number of NMSCs remained significant ( $P = 6.0 \times 10^{-4}$  for rs12203592 and  $P = 5.8 \times 10^{-3}$  for rs12202284), suggesting independent effects of these two SNPs. We previously reported an association of the SNP rs12203592 with hair color (4) and tanning ability (14). In this study, we additionally found a significant association of this SNP with the number of sunburns (T allele, beta = 1.13, SE = 0.24;  $P = 2.2 \times 10^{-6}$ ; Table 6). We also identified a novel SNP rs8015138 upstream of guanine nucleotide-binding protein, gamma 2 (*GNG2*), on chromosome 14 associated with the number of NMSCs with a  $P$ -value of  $6.6 \times 10^{-8}$  in the combined set (C allele, beta = -0.11, SE = 0.02;  $P = 5.3 \times 10^{-7}$  in the discovery stage and  $P = 0.01$  in the replication stage; Table 5).

We did not find novel susceptibility loci for tanning ability, but replicated the previous findings, including rs12203592 in *IRF4* ( $P = 3.3 \times 10^{-23}$ ), rs1126809 in *TYR* ( $P = 5.0 \times 10^{-21}$ ), rs12913832 in *HERC2* ( $P = 1.4 \times 10^{-22}$ ) and rs1805007 ( $P = 1.1 \times 10^{-65}$ ) and rs1805008 ( $P = 1.3 \times 10^{-13}$ ) in *MC1R* (Table 6).

Based on the knowledge that human pigmentation traits are host susceptibility factors for skin cancer (7), we further evaluated the association between the pigmentation determining loci and melanoma risk using the melanoma GWAS of the MD Anderson Cancer Center with 1804 melanoma cases and 1026 controls. As shown in Table 7, the SNP rs12520016 associated with intermediate eye color was associated with decreased risk of melanoma (G allele; OR, 0.74; 95% CI, 0.58–0.95;  $P = 0.02$ ). The SNP rs12421680 associated with increased numbers of sunburns and lower expression of the gene *NTM* was associated with increased risk of melanoma (A allele; OR 1.12; 95% CI 1.00–1.26;  $P = 0.05$ ). Consistent with the interaction between the SNP rs7173419 and the SNP rs12913832 on brown eye color, the protective effect of the SNP rs12913832 on melanoma risk

was most significant among the carriers of CC genotype of rs7173419 (OR 0.64; 95% CI 0.53–0.77;  $P = 3.6 \times 10^{-6}$ ), compared with the carriers of CT genotype (OR 0.76; 95% CI 0.61–0.94;  $P = 0.01$ ) and TT genotype (OR 1.05; 95% CI 0.63–1.73;  $P = 0.86$ ). The interaction was borderline significant ( $P$ -value for interaction = 0.07, Table 4). Besides, the SNP rs12203592 associated with increased number of NMSCs was also associated with increased risk of melanoma (T allele; OR 1.18; 95% CI 1.02–1.34;  $P = 0.02$ , Table 7).

In Table 6, we summarize the pigmentation SNPs identified in the previous GWASs (4–6) and the present study. We presented the associations of them with natural hair color, eye colors, tanning ability, the number of sunburns, mole counts, as well as the risk of melanoma, basal cell carcinoma (BCC), squamous cell carcinoma (SCC) and the number of NMSCs in our study population. In addition, we found that the associations of the SNPs newly identified in the present study remained similar after further adjusting for the pigmentation SNPs identified in the previous studies (rs16891982, rs12203592, rs1408799, rs35264875, rs3829241, rs1126809, rs1042602, rs12821256, rs12896399, rs12913832, rs1805007, rs1805008, rs4911414 and rs1015362, as listed in Table 6), suggesting they are independent from the previously known pigmentation SNPs (Supplementary Material, Table S4). We further screened the pair-wise interactions between these SNPs (including the ASIP haplotypes composed of rs4911414 and rs1015362) on pigmentation traits. None of them revealed significant interactions with each other on any pigmentation traits after the Bonferroni correction ( $P = 0.05/78 = 6.4 \times 10^{-4}$ ).

## DISCUSSION

Previously, we reported the GWASs of hair color and tanning response among 2287 women with 528 173 SNPs genotyped by the Illumina HumanHap550 (4,14), in which we identified novel loci associated with hair color and tanning response in the gene regions of *IRF4* and *SLC24A4* in addition to the known ones, *MC1R*, *OCA2* and *SLC45A2* (4). In the present study, we enlarged the sample size to 10 183 men and women and analyzed 8 221 074 autosomal SNPs

**Table 5.** Association of the SNP rs12203592 in the *IRF4* gene, rs12202284 between *IRF4* and *EXOC2* and rs8015138 upstream of *GNCG2* with the number of NMSCs.<sup>a</sup>

Data set	Sample size	NMSCs, 5+	rs12203592 C>T Beta (SE)	P-value <sup>b</sup>	rs12202284 C>A Beta (SE)	P-value <sup>b</sup>	rs8015138 A>C Beta (SE)	P-value <sup>b</sup>	P-value for heterogeneity
<b>Discovery set</b>									
BC_NHS (female) <sup>b</sup>	1620	45	0.20 (0.07)	0.0037	0.21 (0.08)	0.005	-0.09 (0.05)	0.081	
T2D_NHS (female) <sup>c</sup>	2427	65	0.21 (0.08)	0.011	0.14 (0.07)	0.039	-0.12 (0.04)	0.005	
T2D_HPFS (male) <sup>d</sup>	1308	83	0.24 (0.15)	0.11	0.17 (0.12)	0.16	-0.07 (0.08)	0.41	
CHD_NHS (female) <sup>e</sup>	760	23	0.21 (0.14)	0.13	0.36 (0.11)	0.0015	-0.23 (0.08)	0.002	
CHD_HPFS (male) <sup>f</sup>	632	39	0.47 (0.25)	0.056	0.49 (0.21)	0.017	-0.32 (0.13)	0.01	
All (meta-analysis)	6747	255	0.22 (0.03)	1.8E-8	0.21 (0.04)	6.6E-7	-0.14 (0.03)	5.3E-7	0.26
<b>Replication set</b>									
Skin cancer (female and male) <sup>g</sup>	2078	118	0.20 (0.04)	6.7E-7	0.10 (0.03)	0.003	-0.08 (0.03)	0.01	
Combined set (meta-analysis)	8825	373	0.21 (0.03)	7.2E-14	0.15 (0.03)	5.0E-8	-0.11 (0.02)	6.6E-8	0.11

<sup>a</sup>The number of NMSCs was coded as: 1-0, 2-1, 3-2-4, 4-5+; beta > 0 means association with increased number of NMSCs.

<sup>b</sup>Postmenopausal invasive breast cancer case-control study nested within the NHS.

<sup>c</sup>Type 2 diabetes case-control study nested within the NHS.

<sup>d</sup>Type 2 diabetes case-control study nested within the HPFS.

<sup>e</sup>Coronary heart disease case-control study nested within the NHS.

<sup>f</sup>Coronary heart disease case-control study nested within the HPFS.

<sup>g</sup>Skin cancer case-control study nested within the NHS and HPFS.

<sup>h</sup>Adjusted for the top four principal components of genetic variance in the discovery sets and adjusted for gender in the skin cancer replication set.

imputed from 1000 Genomes Projects on hair color, tanning ability, the number of sunburns and the number of NMSCs. We additionally conducted a GWAS on eye color among 3871 males with 2 543 887 HapMap imputed autosomal SNPs. Using a multistage study design, we identified several novel genetic variants associated with pigimentary phenotypes and the risk of skin cancer. A novel region in the *EDNRB* gene was most significantly associated with natural hair color. The *EDNRB* gene encodes a G protein-coupled receptor that activates a phosphatidylinositol-calcium second messenger system. Studies suggest that mutations in *EDNRB* can cause Waardenburg syndrome, which is characterized by pigmentation abnormalities, including depigmented patches of the skin and hair, vivid blue eyes or heterochromia irides and sensorineural hearing loss (15). The requirement for *EDNRB* signaling pathway during melanocyte development has been reported (16,17). However, the previous studies on genetic variants were limited to the nonsynonymous mutations (15). In the present study, we identified a region downstream of this gene associated with natural human hair color, and the SNP rs975739 revealed the strongest association ( $P = 2.4 \times 10^{-14}$ ). Our finding was consistent with a previous study presented at the 2010 meeting of the American Society of Human Genetics (18), reporting an association between another *EDNRB* SNP rs1668619 (in complete LD with rs975739) and hair color with a  $P$ -value of  $1 \times 10^{-8}$ . The SNP rs1668619 reached a  $P$ -value of  $6.8 \times 10^{-9}$  in our hair color GWAS. We did not find significant association between variants in this region and melanoma risk, and the conclusion of the association between *EDNRB* and melanoma risk was still controversial (19,20). No association was identified for tanning ability or the number of sunburns either, suggesting the involvement of this gene in the basal pigmentation but not the response to sun exposure.

In consistent with our previous findings on the association of the *IRF4* SNP rs12203592 with increased risk of all three types of skin cancer (21), we further found that this SNP was associated with an increased number of NMSCs. We also confirmed the association of this SNP with melanoma risk among an independent population from our previous report in the melanoma GWAS of the MD Anderson Cancer Center. *IRF4* regulates gene expression in response to interferon and other cytokines (22) and is involved in tumor immunity (23), which is plausible to play an important role in cancer development. The other SNP rs12202284 associated with the number of NMSCs was located between *IRF4* and *EXOC2*. We previously reported another SNP rs12210050 in this region associated with an increased risk of both BCC and SCC (24). These two SNPs (rs12202284 and rs12210050) were in modest LD ( $r^2 = 0.77$ ) and revealed similar effects on the number of NMSCs in the present study (rs12202284, A allele, beta = 0.15, SE = 0.03;  $P = 5.0 \times 10^{-8}$ ; rs12210050, T allele, beta = 0.19, SE = 0.04,  $P = 8.5 \times 10^{-8}$ ), which may represent the same causal signal. Neither of the two SNPs was in LD with the SNP rs12203592 in this study ( $r^2 = 0.10$  and  $0.12$ , respectively).

An *NTM* polymorphism was associated with the number of sunburns in this study. *NTM* is a member of the IgLON family of immunoglobulin domain-containing glycosylphosphatidylinositol-anchored cell adhesion molecules (25), which is

**Table 6.** Summary of pigimentary SNPs identified from the previous and the present GWASs and their associations with pigmentation traits, melanoma risk and NMSCs in this study<sup>a</sup>

SNP	Gene	Chr.	Location	Ref.	A1	Hair color			Tanning ability			Number of sunburns			Blue eye color					
						Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value			
rs16891982 <sup>b</sup>	<i>SLC45A2</i>	5	33 987 450	G (major)	C (minor)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
rs12203592	<i>IRF4</i>	6	341 321	C	T	0.35	0.03	1.0E - 28	-0.35	0.03	3.3E - 23	1.13	0.24	2.2E - 06	0.08	0.18	0.66	-	-	-
rs1408799	<i>TYRP1</i>	9	12 662 097	C	T	0.04	0.01	5.6E - 03	0.04	0.01	4.3E - 04	-0.17	0.10	0.09	-0.27	0.05	4.9E - 07	-	-	-
rs35264875 <sup>c</sup>	<i>TPCN2</i>	11	68 602 975	T (minor)	A (major)	0.06	0.02	3.8E - 04	0.02	0.02	0.21	0.14	0.13	0.26	-	-	-	-	-	-
rs3829241	<i>TPCN2</i>	11	68 611 939	G	A	-0.04	0.01	2.1E - 03	-0.02	0.01	0.13	0.09	0.10	0.37	-0.03	0.05	0.54	-	-	-
rs1126809 <sup>c</sup>	<i>TYR</i>	11	88 657 609	G	A	-0.03	0.01	0.04	-0.12	0.01	5.0E - 21	0.60	0.11	1.8E - 08	-	-	-	-	-	-
rs1042602 <sup>b</sup>	<i>TYR</i>	11	88 551 344	C	A	-	-	-	-	-	-	-	-	-	-0.08	0.05	0.14	-	-	-
rs12821256	<i>KITLG</i>	12	87 852 466	C	T	0.18	0.02	6.8E - 19	0.04	0.02	0.04	-0.21	0.16	0.17	-0.18	0.08	0.03	-	-	-
rs12896399	<i>SLC24A4</i>	14	91 843 416	G	T	-0.16	0.01	1.5E - 36	-0.02	0.01	0.07	0.08	0.10	0.43	0.33	0.05	4.1E - 11	-	-	-
rs12913832	<i>HERC2</i>	15	26 039 213	G	A	0.39	0.01	1.4E - 167	0.14	0.01	1.4E - 22	-0.48	0.11	2.1E - 05	-2.84	0.11	1.2E - 158	-	-	-
rs1805007	<i>MC1R</i>	16	88 513 618	C	T	-0.16	0.03	2.9E - 09	-0.39	0.02	1.1E - 65	1.66	0.18	1.5E - 19	0.01	0.10	0.91	-	-	-
rs1805008 <sup>c</sup>	<i>MC1R</i>	16	88 513 645	C	T	-0.09	0.02	7.2E - 05	-0.16	0.02	1.3E - 13	0.71	0.18	4.8E - 05	-	-	-	-	-	-
rs4911414	<i>ASIP Haplotype</i>	20	32 193 105	G	T	-0.05	0.01	1.3E - 04	-0.07	0.01	3.8E - 09	0.26	0.10	0.01	0.05	0.05	0.32	-	-	-
rs1015362	<i>ASIP Haplotype</i>	20	32 202 273	C	T	0.01	0.01	0.50	0.02	0.01	0.20	-0.12	0.11	0.26	0.00	0.05	1.00	-	-	-
rs975739 <sup>d</sup>	<i>EDNRB</i>	13	77 279 147	G	T	-0.07	0.01	5.4E - 09	0.00	0.01	0.72	-0.04	0.10	0.67	0.04	0.05	0.46	-	-	-
rs12421680 <sup>d</sup>	<i>NTM</i>	11	130 856 178	G	A	-0.01	0.02	0.55	-0.01	0.02	0.49	0.49	0.12	5.3E - 05	0.01	0.06	0.80	-	-	-
rs3002288 <sup>d</sup>	<i>VASH2</i>	1	211 193 188	G	A	-0.01	0.01	0.37	0.00	0.01	0.93	-0.06	0.10	0.57	0.11	0.05	0.04	-	-	-
rs12520016 <sup>d</sup>	<i>POLS</i>	5	6 820 312	G	T	-0.02	0.03	0.58	0.01	0.03	0.71	0.03	0.21	0.91	0.36	0.12	2.1E - 03	-	-	-

Intermediate/green eye color			Brown eye color			Melanoma risk			Mole count			BCC risk			SCC risk			Number of NMSCs		
Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value
0.26	0.13	0.05	0.11	0.13	0.41	-0.52	0.19	0.01	-	-	-	-0.12	0.15	0.42	-0.49	0.34	0.15	-	-	-
0.12	0.19	0.54	-0.21	0.20	0.28	0.16	0.07	0.02	0.05	0.03	0.05	0.23	0.07	7.0E - 04	0.52	0.09	9.5E - 09	0.22	0.03	1.8E - 08
0.15	0.05	4.8E - 03	0.13	0.05	0.02	-0.02	0.06	0.68	0.01	0.02	0.70	0.10	0.04	0.01	-0.05	0.13	0.71	0.04	0.03	0.15
-	-	-	-	-	-	-0.02	0.07	0.77	-	-	-	-	-	-	0.16	0.10	0.11	0.04	0.04	0.24
0.02	0.05	0.70	0.01	0.05	0.79	0.00	0.06	0.96	-0.01	0.02	0.34	0.04	0.04	0.26	0.07	0.07	0.33	0.04	0.03	0.12
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.05	0.13	0.69	0.09	0.03	6.0E - 03
0.02	0.05	0.73	0.07	0.05	0.22	-0.10	0.06	0.09	-0.02	0.02	0.29	0.05	0.04	0.23	0.21	0.13	0.09	-	-	-
0.03	0.08	0.75	0.18	0.09	0.05	0.00	0.09	1.00	0.00	0.02	0.92	-0.08	0.06	0.20	-0.02	0.11	0.87	-0.06	0.05	0.16
-0.27	0.05	2.5E - 07	-0.08	0.05	0.12	0.05	0.06	0.39	0.02	0.02	0.11	-0.06	0.04	0.13	-0.09	0.12	0.43	-0.03	0.03	0.24
-0.24	0.06	2.1E - 05	2.82	0.10	1.2E - 177	-0.35	0.07	1.6E - 07	-0.03	0.02	0.07	-0.15	0.04	6.6E - 04	-0.01	0.08	0.86	-0.08	0.03	0.02
0.09	0.10	0.33	-0.10	0.10	0.31	0.55	0.09	8.0E - 09	-0.03	0.03	0.28	0.39	0.07	3.8E - 09	0.42	0.13	1.1E - 03	0.34	0.05	3.40E - 10
-	-	-	-	-	-	0.32	0.10	1.2E - 03	-	-	-	-	-	-	0.24	0.20	0.23	0.00	0.05	0.99
-0.09	0.05	0.08	0.05	0.05	0.40	0.10	0.06	0.09	0.01	0.02	0.72	0.11	0.04	0.01	0.26	0.13	0.04	0.05	0.03	0.11
-0.10	0.05	0.07	0.11	0.06	0.06	-0.07	0.06	0.28	-0.01	0.02	0.75	0.01	0.04	0.74	0.21	0.14	0.13	-0.01	0.03	0.87
-0.05	0.05	0.33	0.01	0.05	0.78	0.06	0.06	0.26	0.00	0.01	0.99	-0.03	0.04	0.37	-0.09	0.09	0.32	-0.07	0.03	9.0E - 03
0.02	0.06	0.69	-0.04	0.06	0.54	0.12	0.06	0.05	0.01	0.02	0.62	0.06	0.04	0.18	-0.10	0.08	0.18	0.02	0.04	0.5
0.13	0.05	0.02	-0.25	0.06	4.3E - 06	0.03	0.06	0.63	0.01	0.02	0.55	-0.03	0.04	0.52	0.05	0.07	0.45	0.00	0.03	0.88
-0.49	0.11	4.7E - 06	0.17	0.12	0.16	0.30	0.13	0.02	0.03	0.03	0.39	-0.01	0.08	0.90	0.12	0.16	0.44	-0.04	0.06	0.47

<sup>a</sup>The SNP rs1805009 in the *MC1R* gene was not in 1000 Genome or HapMap imputed data.

<sup>b</sup>Not in 1000 Genome imputed data.

<sup>c</sup>Not in HapMap imputed data.

<sup>d</sup>Newly identified pigmentation SNPs in this present GWAS.

**Table 7.** Associations of the pigmentation SNPs identified in this study with melanoma risk<sup>a</sup>

SNP	Reference allele	Test allele	OR (95% CI)	P-value <sup>b</sup>
rs975739	T	G	0.93 (0.83–1.04)	0.25
rs12421680	G	A	1.14 (1.01–1.27)	0.02
rs3002288	G	A	1.02 (0.92–1.14)	0.63
rs12520016	T	G	0.75 (0.59–0.96)	0.02
rs7173419	C	T	0.88 (0.77–1.00)	0.06
rs12203592	C	T	1.16 (1.01–1.34)	0.03

<sup>a</sup>The SNPs rs12202284 and rs8015138 were not available in the melanoma GWAS data.

<sup>b</sup>Adjusted for age, gender and the top two principal components.

involved in a variety of specific cell–cell interactions in the developing nervous system (26). As for the skin, NTM was involved in the aging process as a marker of cell adhesion, and a down-regulation of NTM was detected during the replicative senescence of human dermal fibroblasts (21). Although the studies of NTM have been mostly limited to the neural system, our finding of the association between an *NTM* polymorphism and the number of sunburns, along with the evidence from a previous study on skin (21), suggested a promising role in cell adhesion that NTM might play in the skin besides the neural system. The lack of association between the *NTM* SNP with natural hair color and tanning ability further suggested a pigmentation-independent mechanism underlying sunburn reaction.

The presence of melanin pigment within the iris is responsible for the visual impression of human eye colors (12). The function of the two genes identified in eye color GWASs (*VASH2* and *POLS*) remained unknown in terms of pigmentation. *VASH2* was known to play a role in the regulation of angiogenesis (27), and *POLS* was involved in DNA synthesis (28). Further studies on the two genes are required to validate the involvement in pigmentation process, especially in the eyes. The association of the *POLS* polymorphism and melanoma risk is upon further replication, and functional studies are warranted to investigate the biological mechanisms.

The *GNG2* gene was identified as a novel locus conferring susceptibility to the number of NMSCs in this study. *GNG2* encodes one of the gamma subunits of a G-protein, which is involved in signaling mechanisms across membranes and has been known to play important roles in the development and progression of various cancers. The G-protein  $\beta\gamma$  subunits play an important role in melanoma cell migration by crosstalk with Epac in  $Ca^{2+}$  signaling (29). However, there was still lack of specific studies of *GNG2* on NMSCs.

One limitation of this study is that we used self-reported pigmentation information. However, information on hair color and eye color is relatively objective, and a high reliability of self-reported tanning ability and sunburn history has been demonstrated in the previous studies (30,31). The high education level and interest in the health of cohort members further allows high quality and valid information to be collected on self-administered forms, and the self-reported pigmentation phenotypes in our cohorts significantly predicted the risk of skin cancers (including melanoma, SCC and BCC) in the previous studies (7,32,33). Another limitation

on the pigmentation information was that we had categorized information only for some continuous pigmentation traits, such as hair color, which may limit our power to reveal more new pigmentation genes.

In summary, we newly identified several genes and loci associated with human pigmentation traits and skin cancer risk by conducting a series of GWASs on hair color, eye colors, tanning ability, the number of sunburns and the number of NMSCs among individuals of European ancestry. The identification of these novel genetic determinants of pigmentary phenotypes and skin cancer may help further understand the complex mechanisms of human pigmentation and the genetic basis of skin cancer. The different variants identified for different traits may explain the highly variable pigmentation characteristics in European Americans. External replication is needed to extend our finding to other populations.

## MATERIALS AND METHODS

### Study population

#### Pigmentation GWASs

In the discovery stage of the GWASs on hair color, tanning ability, the number of sunburns and the number of NMSCs, we combined data from five case–control studies nested within the Nurses' Health Study (NHS) and the Health Professionals Follow-up Study (HPFS): a postmenopausal invasive breast cancer case–control study nested within the NHS (BC\_NHS,  $n = 2287$ ), a type 2 diabetes case–control study nested within the NHS (T2D\_NHS,  $n = 3116$ ), a type 2 diabetes case–control study nested within the HPFS (T2D\_HPFS,  $n = 2487$ ), a coronary heart disease case–control study nested within the NHS (CHD\_NHS,  $n = 1146$ ) and a coronary heart disease case–control study nested within the HPFS (CHD\_HPFS,  $n = 1147$ ). We excluded the breast cancer cases in BC\_NHS for the GWAS on the number of NMSCs. In the replication stage, a fast-track replication was conducted among 1782 healthy controls in the skin cancer case–control study nested within the NHS and HPFS (skin cancer). Two pigmentation SNPs successfully replicated in the skin cancer set were selected for a second replication among 2722 individuals from a renal function study nested within the NHS (RF\_NHS). Information on eye color was available in the HPFS only, so we combined data from three GWAS sets within the HPFS, T2D\_HPFS, CHD\_HPFS and a kidney stone case–control study nested within the HPFS (KS\_HPFS,  $n = 515$ ), with a total of 3871 men in the discovery stage of eye color GWASs. Two novel SNPs with promising association were selected for replication among 2496 participants from the prostate cancer case–control study nested within the HPFS (PrCa\_HPFS). Overlapped samples among the component data sets were excluded. Detailed descriptions of the study population were published previously (34–38). The study protocol was approved by the Institutional Review Board of Brigham and Women's Hospital and the Harvard School of Public Health.

Information on pigmentation traits was collected from prospective questionnaires in both the NHS and HPFS using similar wording, except for natural eye color, which was collected in the HPFS questionnaires only. We regressed ordinal



coding for natural hair color (1 = blonde, 2 = light brown, 3 = dark brown, 4 = black; excluding red hair color) and tanning ability (1 = practically none, 2 = light tan, 3 = average tan, 4 = deep tan in the NHS; 1 = pain: burn/peel, 2 = burn then tan, 3 = tan without burn in the HPFS). For the hair color phenotype, we excluded the individuals with red hair because red hair color has been known as a distinct trait from other hair colors and is predominantly determined by the *MC1R* gene. A total of 289 (3.9%) individuals with red hair were excluded in the hair color GWAS (77 in BC\_NHS, 41 in CHD\_NHS, 103 in T2D\_NHS, 19 in CHD\_HPFS and 49 in T2D\_HPFS). The number of lifetime severe sunburns was coded as a continuous variable, summarizing those on the face, limbs and back. Different eye colors have been hypothesized to be distinct phenotypes and we treated three categories of eye colors, blue, intermediate (including green) and brown eye colors, as co-dominant outcomes, comparing each with the other two, following a previous GWAS on eye color (5).

### NMSC case ascertainment

For all types of skin cancers, participants reported new cases biennially. With their permission, medical records were obtained and reviewed by physicians to confirm their self-reported diagnosis. Medical records were not obtained for self-reported cases of BCC, but the validity of BCC self-reports was >90% in validation studies in our cohorts in early years. Information on the cumulative number of NMSCs was collected through 2004 in the NHS and 2008 in the HPFS (39,40). The question asked was 'How many squamous or basal cell carcinoma lesions have you ever had removed by surgery, cryotherapy or other means? (Include only new primary cancers. Exclude melanoma and benign lesions like moles or actinic keratoses.)'. The possible choices were never, 1, 2–4, 5–10 and 11+. To minimize the misclassification of the outcome, the eligible participants in this study were Caucasians who reported at least one pathologically confirmed diagnosis of SCC or self-reported BCC in the cohort follow-up from 1984 up to 2004. We coded the number of NMSCs as a continuous variable by using the median number in each category.

### Melanoma risk association study

We evaluated the association between the pigimentary SNPs identified in this study with melanoma risk using the melanoma GWAS data from MD Anderson Cancer Center with 1804 melanoma cases and 1026 controls. Details of the study population have previously been described (41). Briefly, in this hospital-based case–control study, 1804 non-Hispanic patients with newly diagnosed, histologically confirmed, untreated melanoma were consecutively recruited at the University of Texas MD Anderson Cancer Center in Houston, TX, between April 1994 and April 2008. An additional 1026 cancer-free controls matched to the cases on age ( $\pm 5$  years), sex and ethnicity were recruited during the same period from among visitors to MD Anderson Cancer Center who were accompanying patients to outpatient clinics but neither seeking medical care nor biologically related to the patients included in this study. After informed consent

was obtained, each participant was asked to complete a questionnaire to provide information about their demographic and known risk factors for CM. A one-time blood sample (30 ml) was drawn from each study participant. All subjects included in this analysis were US non-Hispanic European descendants. The research protocol was approved by the Institutional Review Board at MD Anderson Cancer Center.

### SNP selection for replication

Four novel regions revealed the strongest associations with pigmentation traits in the discovery stage of this study besides the known pigmentation loci and were selected for replication: one region on chromosome 13 with hair color (smallest  $P = 5.4 \times 10^{-9}$ ), one region on chromosome 11 with the number of sunburns (smallest  $P = 5.3 \times 10^{-5}$ ), one region on chromosome 1 with brown eye color (smallest  $P = 4.2 \times 10^{-6}$ ) and one region on chromosome 5 with intermediate (including green) eye color (smallest  $P = 3.1 \times 10^{-6}$ ). For the region on chromosome 13 associated with natural hair color GWAS, we found 117 SNPs with  $P < 1 \times 10^{-7}$  in two LD blocks in this region (chr13:77 259 946–77 370 845). We forced in the top SNP in the regional Manhattan plot (rs975739) and selected other two tag-SNPs (rs1146927 and rs58188699) for replication using the Tagger program in Haploview 4.2 ( $r^2 > 0.8$ ) (42,43). A proxy SNP was selected for each of them to ensure the validity of genotyping. For each region associated with the number of sunburns, brown eye color and intermediate eye color, we selected the SNP with the strongest association as the representative. Besides, two novel regions revealed promising associations with the number of NMSCs in the discovery stage: one on chromosome 6 and the other on chromosome 14. We selected the five SNPs with the strongest associations as the representatives (two on chromosome 6 and three on chromosome 14). Additionally, the well-known pigmentation SNP rs12913832 in HECT and RLD domain containing E3 ubiquitin protein ligase 2 (*HERC2*) revealed the strongest association with brown eye color in our study and also a significant interaction with the SNP rs7173419 in *OCA2*. To validate this interaction, we further selected rs12913832 and rs7173419 for replication.

### Genotyping, imputation and quality control

We previously used either the Illumina/Affymetrix SNP chip for the genotyping of the component GWAS sets (34–36). We used Illumina HumanHap550 array to perform genotyping in the BC\_NHS (34), used Illumina 610Q array to perform genotyping in the KS\_HPFS and used Affymetrix 6.0 array in the other four GWAS sets, including T2D\_NHS, T2D\_HPFS (35), as well as CHD\_NHS, CHD\_HPFS (36). For five of the GWAS sets (BC\_NHS, T2D\_NHS, T2D\_HPFS, CHD\_NHS and CHD\_HPFS), we used the MACH program (44) to impute 8 221 074 autosomal SNPs based on haplotypes from the 1000 Genomes Project (using the August 2009 release of the Sanger Institute SNP calls based on the pilot 1 project). For the KS\_HPFS set, we used the same program to impute 2 543 887 autosomal SNPs based on haplotypes from the HapMap database phase II data build 35 (CEU) (45). The HapMap imputed data of

T2D\_HPFS and CHD\_HPFS were used for eye color GWASs to be combined with KS\_HPFS. Samples from the six GWASs were imputed separately. We observed high imputation quality in each cohort's imputation. In particular, the numbers of SNPs with MACH  $r^2 > 0.4$  were 7 438 552 (90.5%) for NHS\_BC, 7 399 703 (90.0%) for NHS\_T2D, 7 437 076 (90.5%) for NHS\_CHD, 7 370 611 (89.7%) for HPFS\_T2D and 7 417 449 (90.2%) for HPFS\_CHD. Most of the imputed SNPs were common [for all five cohorts, >98% SNPs have a minor allele frequency (MAF) > 0.01 and >79% SNPs have an MAF > 0.05]. SNPs with imputation  $r^2 > 0.4$  and an MAF > 0.01 in each study were included in meta-analysis. Finally, 7 588 169 SNPs were analyzed for natural hair color, tanning ability, the number of sunburns and the number of NMSCs, and 2 469 762 SNPs were analyzed for eye colors.

The promising SNPs selected from the discovery stage were genotyped in the replication sets using TaqMan/BioTrove assays at the Dana Farber/Harvard Cancer Center Polymorphism Detection Core. Laboratory personnel were blinded to the case-control status, and blinded quality control samples were inserted to validate genotyping procedures; concordance for the blinded samples was 100%. Primers, probes and conditions for genotyping assays are available upon request. All these SNPs were successfully genotyped with call rate >95%, except for the SNP rs1146926 (a proxy SNP of rs1146927), which failed in genotyping.

As for the melanoma risk association study, the samples were previously genotyped by the Illumina Omni 1M Quad V1-0\_B SNP chip (41). An imputation was conducted based on the genotyped SNPs and haplotype information in the HapMap phase II data build 35 (CEU) (45). For the SNPs used in this study, the SNPs rs12913832, rs12421680, rs12520016 and rs12203592 were directly genotyped, and the SNPs rs975739, rs3002288 and rs7173419 were imputed with imputation  $r^2$  of 1.00, 0.92 and 0.98, respectively.

### Statistical analysis

We used linear regression to test the association between minor allele counts and phenotypes of hair color, tanning ability, the number of sunburns and the number of NMSCs, and used logistic regression to test association separately for brown, intermediate and blue eye colors as co-dominant outcomes (comparing one with the other two). The imputed dosage data were used for the analysis. The model was adjusted for the first four principal components. Age was also adjusted for the GWASs of the number of sunburns and the number of NMSCs, but not for the other pigmentation traits, which were not likely to be confounded by age. We used ProbABEL for the GWAS analysis in this study. Associations in each component GWAS set were combined in an inverse-variance-weighted meta-analysis using the METAL software. We additionally adjusted for gender in the skin cancer replication set, which was mixed of males and females. The same meta-analysis was conducted to combine the component replication sets and to combine the discovery set and the replication set. We used logistic regression for the melanoma risk association study and adjusted for age, gender and the first two principal components. These principal components were calculated for all individuals on the basis of

approximately 10 000 unlinked markers using the EIGENSTRAT software (46). We used the algorithm developed by Kraft *et al.* (47) to infer the haplotypes from the unphased genotypic data. The description of this algorithm was previously published, and this approach has been proved by simulation studies as well as subsequent applications in several other studies (47–49).

### WEB RESOURCES

The URLs for data presented herein are as follows:

1000 Genomes Project, <http://www.1000genomes.org>  
International HapMap Project, <http://www.hapmap.org>  
METAL software, <http://www.sph.umich.edu/csg/abecasis/Metal/index.html>

### SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

### ACKNOWLEDGEMENTS

We thank Dr David Hunter, Dr Frank B. Hu, Dr Eric B. Rimm and Dr Gary Curhan for their work on the component GWAS sets. We thank Dr Mousheng Xu and Constance Chen for programming support, and thank Pati Soule and Dr Hardeep Ranu of the Dana Farber/Harvard Cancer Center High-Throughput Polymorphism Detection Core for sample handling and genotyping of the NHS and HPFS samples. We are grateful to Merck Research Laboratories for the funding of the GWAS of coronary heart disease. We thank Dr Jinyan Huang for improving the figure quality. We are also indebted to the participants in all of these studies. We thank the cancer registries in the following states for their help: Alabama, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Idaho, Illinois, Indiana, Iowa, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Nebraska, New Hampshire, New Jersey, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Virginia, Washington and Wyoming.

*Conflict of Interest statement.* None declared.

### FUNDING

This work is supported by National Institutes of Health grants CA122838, CA87969, CA055075, CA49449, CA100264 and CA093459.

### REFERENCES

1. Clark, P., Stark, A.E., Walsh, R.J., Jardine, R. and Martin, N.G. (1981) A twin study of skin reflectance. *Ann. Hum. Biol.*, **8**, 529–541.
2. Frisancho, A.R., Wainwright, R. and Way, A. (1981) Heritability and components of phenotypic expression in skin reflectance of Mestizos from the Peruvian lowlands. *Am. J. Phys. Anthropol.*, **55**, 203–208.
3. Gerstenblith, M.R., Shi, J. and Landi, M.T. (2010) Genome-wide association studies of pigmentation and skin cancer: a review and meta-analysis. *Pigment Cell Melanoma Res.*, **23**, 587–606.

4. Han, J., Kraft, P., Nan, H., Guo, Q., Chen, C., Qureshi, A., Hankinson, S.E., Hu, F.B., Duffy, D.L., Zhao, Z.Z. *et al.* (2008) A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet.*, **4**, e1000074.
5. Sulem, P., Gudbjartsson, D.F., Stacey, S.N., Helgason, A., Rafnar, T., Magnusson, K.P., Manolescu, A., Karason, A., Palsson, A., Thorleifsson, G. *et al.* (2007) Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet.*, **39**, 1443–1452.
6. Sulem, P., Gudbjartsson, D.F., Stacey, S.N., Helgason, A., Rafnar, T., Jakobsdottir, M., Steinberg, S., Gudjonsson, S.A., Palsson, A., Thorleifsson, G. *et al.* (2008) Two newly identified genetic determinants of pigmentation in Europeans. *Nat. Genet.*, **40**, 835–837.
7. Han, J., Colditz, G.A. and Hunter, D.J. (2006) Risk factors for skin cancers: a nested case-control study within the Nurses' Health Study. *Int. J. Epidemiol.*, **35**, 1514–1521.
8. Stacey, S.N., Gudbjartsson, D.F., Sulem, P., Bergthorsson, J.T., Kumar, R., Thorleifsson, G., Sigurdsson, A., Jakobsdottir, M., Sigurgeirsson, B., Benediksdottir, K.R. *et al.* (2008) Common variants on 1p36 and 1q42 are associated with cutaneous basal cell carcinoma but not with melanoma or pigmentation traits. *Nat. Genet.*, **40**, 1313–1318.
9. Stacey, S.N., Sulem, P., Masson, G., Gudjonsson, S.A., Thorleifsson, G., Jakobsdottir, M., Sigurdsson, A., Gudbjartsson, D.F., Sigurgeirsson, B., Benediksdottir, K.R. *et al.* (2009) New common variants affecting susceptibility to basal cell carcinoma. *Nat. Genet.*, **41**, 909–914.
10. Nan, H., Xu, M., Kraft, P., Qureshi, A.A., Chen, C., Guo, Q., Hu, F.B., Curhan, G., Amos, C.I., Wang, L.E. *et al.* (2011) Genome-wide association study identifies novel alleles associated with risk of cutaneous basal cell carcinoma and squamous cell carcinoma. *Hum. Mol. Genet.*, **20**, 3718–3724.
11. Huang, R.S., Duan, S., Bleibel, W.K., Kistner, E.O., Zhang, W., Clark, T.A., Chen, T.X., Schweitzer, A.C., Blume, J.E., Cox, N.J. *et al.* (2007) A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity. *Proc. Natl Acad. Sci. USA*, **104**, 9758–9763.
12. Sturm, R.A. and Larsson, M. (2009) Genetics of human iris colour and patterns. *Pigment Cell Melanoma Res.*, **22**, 544–562.
13. Sturm, R.A., Duffy, D.L., Zhao, Z.Z., Leite, F.P., Stark, M.S., Hayward, N.K., Martin, N.G. and Montgomery, G.W. (2008) 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.*, **82**, 424–431.
14. Nan, H., Kraft, P., Qureshi, A.A., Guo, Q., Chen, C., Hankinson, S.E., Hu, F.B., Thomas, G., Hoover, R.N., Chanock, S. *et al.* (2009) Genome-wide association study of tanning phenotype in a population of European ancestry. *J. Invest. Dermatol.*, **129**, 2250–2257.
15. Pingault, V., Ente, D., Dastot-Le Moal, F., Goossens, M., Marlin, S. and Bondurand, N. Review and update of mutations causing Waardenburg syndrome. *Hum. Mutat.*, **31**, 391–406.
16. Hou, L., Pavan, W.J., Shin, M.K. and Arnheiter, H. (2004) Cell-autonomous and cell non-autonomous signaling through endothelin receptor B during melanocyte development. *Development*, **131**, 3239–3247.
17. Stanchina, L., Baral, V., Robert, F., Pingault, V., Lemort, N., Pachnis, V., Goossens, M. and Bondurand, N. (2006) Interactions between Sox10, Edn3 and Ednrb during enteric nervous system and melanocyte development. *Dev. Biol.*, **295**, 232–249.
18. Hinds, D.A., Macpherson, J.M., Tung, J., Naughton, B., Kiefer, A., Do, C., Wojcicki, A., Mountain, J. and Eriksson, N. (2010) Genetics of infectious disease susceptibility and morphological traits in a participant driven cohort. *Abstract presented at the 60th Annual Meeting of The American Society of Human Genetics*, 2–6 November 2010, Washington, DC; <http://www.ashg.org/2010meeting/abstracts/fulltext/f22581.htm>.
19. Fernandez, L.P., Milne, R.L., Pita, G., Floristan, U., Sendagorta, E., Feito, M., Aviles, J.A., Martin-Gonzalez, M., Lazaro, P., Benitez, J. *et al.* (2009) Pigmentation-related genes and their implication in malignant melanoma susceptibility. *Exp. Dermatol.*, **18**, 634–642.
20. Soufir, N., Meziani, R., Lacapere, J.J., Bertrand, G., Fumeron, F., Bourillon, A., Gerard, B., Descamps, V., Crickx, B., Ollivaud, L. *et al.* (2005) Association between endothelin receptor B nonsynonymous variants and melanoma risk. *J. Natl Cancer Inst.*, **97**, 1297–1301.
21. Yoon, I.K., Kim, H.K., Kim, Y.K., Song, I.H., Kim, W., Kim, S., Baek, S.H., Kim, J.H. and Kim, J.R. (2004) Exploration of replicative senescence-associated genes in human dermal fibroblasts by cDNA microarray technology. *Exp. Gerontol.*, **39**, 1369–1378.
22. Lu, R. (2008) Interferon regulatory factor 4 and 8 in B-cell development. *Trends Immunol.*, **29**, 487–492.
23. Shaffer, A.L., Emre, N.C., Romesser, P.B. and Staudt, L.M. (2009) IRF4: Immunity. Malignancy! Therapy? *Clin. Cancer Res.*, **15**, 2954–2961.
24. Struyk, A.F., Canoll, P.D., Wolfgang, M.J., Rosen, C.L., D'Eustachio, P. and Salzer, J.L. (1995) Cloning of neurotrimin defines a new subfamily of differentially expressed neural cell adhesion molecules. *J. Neurosci.*, **15**, 2141–2156.
25. Sellar, G.C., Watt, K.P., Rabiash, G.J., Stronach, E.A., Li, L., Miller, E.P., Massie, C.E., Miller, J., Contreras-Moreira, B., Scott, D. *et al.* (2003) OPCML at 11q25 is epigenetically inactivated and has tumor-suppressor function in epithelial ovarian cancer. *Nat. Genet.*, **34**, 337–343.
26. Liu, J., Li, G., Peng, X., Liu, B., Yin, B., Tan, X., Fan, M., Fan, W., Qiang, B. and Yuan, J. (2004) The cloning and preliminary functional analysis of the human neurotrimin gene. *Sci. China C Life Sci.*, **47**, 158–164.
27. Kimura, H., Miyashita, H., Suzuki, Y., Kobayashi, M., Watanabe, K., Sonoda, H., Ohta, H., Fujiwara, T., Shimosegawa, T. and Sato, Y. (2009) Distinctive localization and opposed roles of vasohibin-1 and vasohibin-2 in the regulation of angiogenesis. *Blood*, **113**, 4810–4818.
28. Wang, Z., Castano, I.B., De Las Penas, A., Adams, C. and Christman, M.F. (2000) Pol kappa: a DNA polymerase required for sister chromatid cohesion. *Science*, **289**, 774–779.
29. Baljinnayam, E., Umemura, M., De Lorenzo, M.S., Xie, L.H., Nowycky, M., Iwatsubo, M., Chen, S., Goydos, J.S. and Iwatsubo, K. (2011) Gβγ subunits inhibit Epac-induced melanoma cell migration. *BMC Cancer*, **11**, 256.
30. Westerdaal, J., Anderson, H., Olsson, H. and Ingvar, C. (1996) Reproducibility of a self-administered questionnaire for assessment of melanoma risk. *Int. J. Epidemiol.*, **25**, 245–251.
31. Berwick, M. and Chen, Y.T. (1995) Reliability of reported sunburn history in a case-control study of cutaneous malignant melanoma. *Am. J. Epidemiol.*, **141**, 1033–1037.
32. Qureshi, A.A., Zhang, M. and Han, J. (2011) Heterogeneity in host risk factors for incident melanoma and non-melanoma skin cancer in a cohort of US women. *J. Epidemiol.*, **21**, 197–203.
33. Cho, E., Rosner, B.A., Feskanich, D. and Colditz, G.A. (2005) Risk factors and individual probabilities of melanoma for whites. *J. Clin. Oncol.*, **23**, 2669–2675.
34. Hunter, D.J., Kraft, P., Jacobs, K.B., Cox, D.G., Yeager, M., Hankinson, S.E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A. *et al.* (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.*, **39**, 870–874.
35. Qi, L., Cornelis, M.C., Kraft, P., Stanya, K.J., Linda Kao, W.H., Pankow, J.S., Dupuis, J., Florez, J.C., Fox, C.S., Pare, G. *et al.* (2010) Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. *Hum. Mol. Genet.*, **19**, 2706–2715.
36. Rimm, E.B., Giovannucci, E.L., Willett, W.C., Colditz, G.A., Ascherio, A., Rosner, B. and Stampfer, M.J. (1991) Prospective study of alcohol consumption and risk of coronary disease in men. *Lancet*, **338**, 464–468.
37. Taylor, E.N., Stampfer, M.J. and Curhan, G.C. (2005) Obesity, weight gain, and the risk of kidney stones. *JAMA*, **293**, 455–462.
38. Forman, J.P., Fisher, N.D., Schopick, E.L. and Curhan, G.C. (2008) Higher levels of albuminuria within the normal range predict incident hypertension. *J. Am. Soc. Nephrol.*, **19**, 1983–1988.
39. Lin, W., Qureshi, A.A., Kraft, P., Nan, H., Guo, Q., Hu, F.B., Jensen, M.K. and Han, J. (2011) ASIP genetic variants and the number of non-melanoma skin cancers. *Cancer Causes Control*, **22**, 495–501.
40. Wei-Passanese, E.X., Han, J., Lin, W., Li, T., Laden, F. and Qureshi, A.A. (2012) Geographical variation in residence and risk of multiple nonmelanoma skin cancers in US women and men. *Photochem. Photobiol.*, **88**, 483–489.
41. Amos, C.I., Wang, L.E., Lee, J.E., Gershenwald, J.E., Chen, W.V., Fang, S., Kosoy, R., Zhang, M., Qureshi, A.A., Vattathil, S. *et al.* (2011) Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. *Hum. Mol. Genet.*, **20**, 5012–5023.
42. Carlson, C.S., Eberle, M.A., Rieder, M.J., Yi, Q., Kruglyak, L. and Nickerson, D.A. (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am. J. Hum. Genet.*, **74**, 106–120.

43. Stram, D.O., Leigh Pearce, C., Bretsky, P., Freedman, M., Hirschhorn, J.N., Altshuler, D., Kolonel, L.N., Henderson, B.E. and Thomas, D.C. (2003) Modeling and E-M estimation of haplotype-specific relative risks from genotype data for a case-control study of unrelated individuals. *Hum. Hered.*, **55**, 179–190.
44. Li, Y., Willer, C., Sanna, S. and Abecasis, G. (2009) Genotype imputation. *Ann. Rev. Genomics Hum. Genet.*, **10**, 387–406.
45. Marchini, J., Donnelly, P. and Cardon, L.R. (2005) Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat. Genet.*, **37**, 413–417.
46. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.
47. Kraft, P., Cox, D.G., Paynter, R.A., Hunter, D. and De Vivo, I. (2005) Accounting for haplotype uncertainty in matched association studies: a comparison of simple and flexible techniques. *Genet. Epidemiol.*, **28**, 261–272.
48. Zhai, R., Gong, M.N., Zhou, W., Thompson, T.B., Kraft, P., Su, L. and Christiani, D.C. (2007) Genotypes and haplotypes of the VEGF gene are associated with higher mortality and lower VEGF plasma levels in patients with ARDS. *Thorax*, **62**, 718–722.
49. Feigelson, H.S., Cox, D.G., Cann, H.M., Wacholder, S., Kaaks, R., Henderson, B.E., Albanes, D., Altshuler, D., Berglund, G., Berrino, F. *et al.* (2006) Haplotype analysis of the HSD17B1 gene and risk of breast cancer: a comprehensive approach to multicenter analyses of prospective cohort studies. *Cancer Res.*, **66**, 2468–2475.