Identification of Susceptibility Loci for Cutaneous Squamous Cell Cancer

Supplementary Materials

Maryam M Asgari[1,5], Wei Wang[2], Nilah M Ioannidis[2,3], Jacqueline Itnyre[2], Thomas J Hoffmann[4], Eric Jorgenson[5], Alice S Whittemore[2]

[1]Department of Dermatology, Massachusetts General Hospital, Boston, MA 02114, USA; [2]Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA 94305, USA; [3]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA; [4]Department of Epidemiology and Biostatistics and Institute for Human Genetics, University of California, San Francisco, CA 94158, USA; [5]Division of Research, Kaiser Permanente Northern California, Oakland, CA 94612, USA. Correspondence: Alice S. Whittemore, Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA 94305, USA. E-mail: alicesw@stanford.edu

Case verification. Potentially eligible cases were identified from electronic pathology records of all pathology specimens from all KPNC medical facilities. All pathology specimens are assigned Systematized Nomenclature of Medicine (SNOMED) codes to classify diagnoses by organ (topography code) and morphological alteration (morphology code). The pathology records of all potential cases were queried for any report with a skin SNOMED topology code and cutaneous SCC morphology code. All pathology reports for potential SCC cases were reviewed and assigned case status (definite SCC, possible SCC, no SCC) by the study dermatologist (MA) based on the microscopic description of the pathology specimen in the electronic report. Only cases assigned definitive SCC status were included in the analysis.

Subject Quality Control (QC). We used an ordered filtering process (**Table S8**) to exclude potentially eligible subjects with the following properties:

a) Ambiguous genetic sex (n=65), as determined by X-chromosome homozygosity rates between 0.2 and 0.8.

b) Ambiguous SCC diagnosis (n = 1753).

c) SNP call rate (CR) < 0.97 or SNP heterozygosity > 0.4 (n=59).

d) Close genetic relationship to another eligible subject (n=3,696), defined by having expected number of alleles shared IBD greater than or equal to 0.25, as determined using 99,679 weakly correlated (pairwise $R^2 < 0.5$) SNPs typed on all four arrays. We selected the case in discordant case-control pairs and the subject with greater CR in concordant pairs.

e) Absence of EUR principal components of ancestry (PCAs) (n = 329).

We implemented this filtering using PLINK v1.9 (https://www.cog-genomics.org/plink2/) (Chang et al., 2015) and custom programming in the R programming language. A total of 5,955 subjects failed the filter, leaving 67,867 eligible subjects, whom we classified in two array-based groups: the 61,457 subjects typed on the EUR array with the Axiom v1 reagent (screening phase), and the remaining 6,410 subjects typed on other arrays (replication phase). (**Table S9**).

Genotype QC. Genotype quality control has been described previously [Kvale 2015, http://www.ncbi.nlm.nih.gov/pubmed/26092718]. In addition, we excluded typed SNPs with CR<0.9.

SNP Imputation. Genotype imputation was performed by first pre-phasing the data with SHAPE-IT v2.r644 (Delaneau et al., 2012), using the family structure of first-degree cryptically related subjects. We then imputed a total of 31,081,353 variants from the 1000 Genomes Project (phase I integrated release, March 2012, with Aug 2012 chromosome X update, with singletons removed) as a cosmopolitan reference panel with IMPUTE2 v2.2.2 (Howie et al., 2009; Howie et al., 2011; Howie et al; 2012). We excluded 21,445,450 SNPs with minor allele frequency (MAF) <1% in either screening or replication phase and 25,435 SNPs with low imputation accuracy (defined as Information ≤ 0.3, where Information is the metric used in Impute2 to estimate the imputation accuracy (Marchini and Howie, 2010)). These exclusions left 9,610,468 variants for analysis.

Population stratification. We determined principal components of ancestry (PCAs) for the combined data from screening and replication phases using the smartpca program in the EIGENSOFT4.2 software package (Patterson et al., 2006), as has been described (Banda 2015, http://www.ncbi.nlm.nih.gov/pubmed/26092716). For each of the screening and replication phases, we then used PLINK to evaluate SCC-association with each SNP's allelic count using the Armitage-Cochran trend statistic, adjusted for gender, the first 10 principal components of ancestry, and for the replication phase, the genotyping array. We combined regression coefficients and P-values obtained from screening and replication phases using Cochran's method (Cochran 1954) implemented in R, and identified as significant those SNPs with combined P-values < $5x10^{-8}$ (Pe'er et al., 2008). We checked for residual population stratification by examining the inflation factor λ_{GC} (Devlin and Roeder, 1999), and quantile-quantile plots of percentiles of the observed distribution of test statistics versus those expected under the global null hypothesis. We found little evidence for confounding by population stratification, as shown by the plot in **Figure S5,** and as indicated by an inflation factor of λ = 1.05.

Data visualization and SNP annotation. We visualized results using Manhattan plots created in R with QQman package (Turner, 2014) and plots generated by LocusZoom (Pruim et al., 2010) and HaploView (Barrett et al., 2005). To identify independent associations within a region, we used step-wise multiple logistic regressions, adjusted for the most strongly associated SNPs in the region. We also used the Annovar package (Wang et al., 2010) to perform RefSeq gene-based annotations of SCC-associated SNPs. We also used Annovar to map SNPs to the RegulomeDB (Boyle et al., 2012), GWAScatlog (Welter et al., 2014), and ClinVar (Landrum et al., 2014) databases to identify potential regulatory SNPs, and/or SNPs that have been previously associated with skin-pigmentation phenotypes or skin cancers, in addition to a PubMed search. We used the UCSC Coordinate Conversion (LiftOver, https://genome.ucsc.edu/cgi-bin/hgLiftOver) to convert genome coordinates in GWAScatlog from hg38 to hg19.

Joint effects of pigment-related SCC genotypes and pigmentation phenotypes. We used bivariate analyses to examine whether the pigment-related SNPs in Table 1A of the text are associated with SCC risk purely via their effects on skin pigmentation type (skin color, sun sensitivity and tanning ability), or through a mechanism independent of skin type (**Figure S6**). Since we lacked subjects' self-reported or clinically assessed skin phenotypes, we estimated them using their total counts of "risk alleles" for each skin phenotype, where the "risk allele" is the one associated positively with fair skin, sun sensitivity or inability to tan**.** Specifically, we compiled sets of SNPs associated ($P<10^{-5}$) with each of three skin pigmentation phenotypes, using the GWASs cited in the GWAScatalog (http://www.ebi.ac.uk/gwas/). To exclude highly correlated SNPs ($R^2>0.85$), we estimated pairwise R^2 based on the genotype data of the 1000 Genomes Project phase III subjects with European ancestry (CEU, TSI, FIN, GBR and IBS). The genotype data were extracted using the data slicer [http://browser.1000genomes.org/Homo_sapiens/UserData/SelectSlice] and R^2 was estimated using HaploView (Barrett et al., 2005). We used PHASE, version 2.1, (Stephens et al., 2001, 2003) to estimate the ASIP haplotypes based on genotype data. **Table S4** shows the SNPs used for each of the skin phenotypes. We also assigned each subject a polygenetic SCC risk score calculated as the total count of SCC risk alleles of the six pigment-related SNPs in text **Table 1**.

Gene expression associations. We examined whether the SCC-associated SNPs with univariate P-values meeting the genome-wide threshold $5x10^{-8}$ are also associated with expression of nearby genes in both sun-exposed (lower leg) and non-sun-exposed (supra-pubic) skin tissue, in the version 6 analysis release (phs000424.v6.p1) of the Genotype-Tissue Expression project (GTEx Consortium, 2015). Gene expression association P-values were computed for SNP-gene pairs based on 302 samples of sun-exposed skin and 196 samples of non-sun-exposed skin using the interface available at http://www.gtexportal.org/home/testyourown

Allelic effects of SNPs in Table 1. For each of the SNPs at the 10 susceptibility loci in text Table 1, we used a likelihood ratio statistic (LRS) to evaluate how well allelic effects on SCC risk were

captured by a one-degree-of-freedom (DF) log-additive logistic model relative to a co-dominant (two DF) model. The LRS has a chi-square distribution with one DF under the null hypothesis that the log-additive allelic model fits well. We found significant improvement in fit for the co-dominant model relative to the log-additive model only for SNP rs1126809 in TYR at 11q24 ($P = 3.6x10^{-3}$). SCC odds-ratios for heterozygote and homozygote carriers of the A-allele of this SNP were 1.13 (95% confidence interval: 1.07-1.19) and 1.51 (1.39-1.64), respectively. The co-dominant model also provided significant improvement in fit relative to a recessive model ($P = 2.6x10^{-6}$).

Potential interaction between pairs of significant SNPs. We evaluated departures from an additive model for the dependence of SCC log odds on subjects' genotypes (g_1,g_2) at pairs of significant SNPs, for each of the 45 possible pairs of the 10 SNPs in Table 1 of the text. Specifically, we modeled the log odds of SCC as $\alpha + \beta_1 g_1 + \beta_2 g_2 + \delta g_1 g_2$, and tested the null hypothesis $\delta = 0$, using a likelihood ratio test. We found significant (P<0.05) evidence for interaction between the five SNP pairs shown in **Table S7.** Note that the interaction term is negative for four of the five pairs, which may relate to the roles of the genes in synthesizing and transporting melanin (**Figure S7).**

REFERENCES

Banda Y, Kvale MN, Hoffmann TJ, Hesselson SE, Ranatunga D, Tang H, et al. Characterizing Race/Ethnicity and Genetic Ancestry for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. Genetics. 2015;200(4):1285-95.

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21(2):263-5.

Bastiaens MT, ter Huurne JA, Kielich C, Gruis NA, Westendorp RG, Vermeer BJ, et al. Melanocortin-1 receptor gene variants determine the risk of nonmelanoma skin cancer independently of fair skin and red hair. Am J Hum Genet. 2001;68(4):884-94.

Branicki W, Brudnik U, Kupiec T, Wolanska-Nowak P, Szczerbinska A, Wojas-Pelc A. Association of polymorphic sites in the OCA2 gene with eye colour using the tree scanning method. Ann Hum Genet. 2008;72(Pt 2):184-92.

Branicki W, Brudnik U, Wojas-Pelc A. Interactions between HERC2, OCA2 and MC1R may influence human pigmentation phenotype. Ann Hum Genet. 73. England2009. p. 160-70.

Box NF, Duffy DL, Irving RE, Russell A, Chen W, Griffyths LR, et al. Melanocortin-1 receptor genotype is a risk factor for basal and squamous cell carcinoma. J Invest Dermatol. 2001;116(2):224-9.

Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012;22(9):1790-7.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.

Cochran WG. The combination of estimates from different experiments. Biometrics. 1954;10:101-29.

Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat Methods. 2012;9(2):179-81.

Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999;55(4):997-1004.

Duffy DL, Montgomery GW, Chen W, Zhao ZZ, Le L, James MR, et al. A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. Am J Hum Genet. 2007;80(2):241-52.

Edwards M, Bigham A, Tan J, Li S, Gozdzik A, Ross K, et al. Association of the OCA2 polymorphism His615Arg with melanin content in east Asian populations: further evidence of convergent evolution of skin pigmentation. PLoS Genet. 2010;6(3):e1000867.

Eiberg H, Troelsen J, Nielsen M, Mikkelsen A, Mengel-From J, Kjaer KW, et al. Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. Hum Genet. 2008;123(2):177-87.

GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348(6235):648-60.

Han J, Kraft P, Nan H, Guo Q, Chen C, Qureshi A, et al. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. PLoS Genet. 2008;4(5):e1000074.

Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009;5(6):e1000529.

Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012;44(8):955-9.

Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3 (Bethesda). 2011;1(6):457-70.

Kayser M, Liu F, Janssens AC, Rivadeneira F, Lao O, van Duijn K, et al. Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. Am J Hum Genet. 2008;82(2):411-23.

Kennedy C, ter Huurne J, Berkhout M, Gruis N, Bastiaens M, Bergman W, et al. Melanocortin 1 receptor (MC1R) gene variants are associated with an increased risk for cutaneous melanoma which is largely independent of skin type and hair color. J Invest Dermatol. 2001;117(2):294-300.

Kvale MN, Hesselson S, Hoffmann TJ, Cao Y, Chan D, Connell S, et al. Genotyping informatics and quality control for 100,000 subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. Genetics. 2015;200(4):1051-60.

Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014;42(Database issue):D980-5.

Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet. 2010;11(7):499-511.

Mengel-From J, Borsting C, Sanchez JJ, Eiberg H, Morling N. Human eye colour and HERC2, OCA2 and MATP. Forensic Sci Int Genet. 2010;4(5):323-8.

Nan H, Kraft P, Qureshi AA, Guo Q, Chen C, Hankinson SE, et al. Genome-wide association study of tanning phenotype in a population of European ancestry. J Invest Dermatol. 2009;129(9):2250-7.

Palmer JS, Duffy DL, Box NF, Aitken JF, O'Gorman LE, Green AC, et al. Melanocortin-1 receptor polymorphisms and risk of melanoma: is the association explained solely by pigmentation phenotype? Am J Hum Genet. 2000;66(1):176-86.

Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006;2(12):e190.

Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol. 2008;32(4):381-5.

Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26(18):2336-7.

Sturm RA, Duffy DL, Zhao ZZ, Leite FP, Stark MS, Hayward NK, et al. A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. Am J Hum Genet. 2008;82(2):424-31.

Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet. 2003;73(5):1162-9.

Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet. 2001;68(4):978-89.

Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat Genet. 2007;39(12):1443-52.

Turner SD. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. 2014; Doi: http://dx.doi.org/10.1101/005165.

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.

Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 2014;42(Database issue):D1001-6.

Supplementary Tables

Table S1. SNPs in HERC2 and OCA2

A. Pigment-related SNPs

Gene	SNP	Traits	Reference	Position	Info*	Major/ Minor allele	OR (95% CI) †	P value
HERC2	rs12913832	Hair color, tanning, eye color	Han et al., 2008; Eiberg et al., 2008; Branicki et al., 2009; Nan et al., 2009; Sturm et al., 2008	28365618	0.93	G/A	0.90 (0.86-0.94)	7.28×10^{-7}
HERC2	rs1129038	Eye color	Mengel-From et al., 2010; Donnelly et al., 2012	28356859	1.00	T/C	0.90 (0.86-0.93)	4.86×10^{-7}
HERC2	rs916977	Eye color	Kayser et al., 2008	28513364	1.00	C/T	0.91 (0.86-0.95)	5.97×10^{-5}
HERC2	rs1667394	Eye color	Sulem et al., 2007	28530182	1.00	T/C	0.9 (0.86-0.95)	2.25×10^{-5}
OCA2	rs7495174	Eye color	Sulem et al., 2007; Duffy et al., 2007	28344238	0.93	A/G	0.82 (0.76-0.89)	9.92×10^{-7}
OCA2	rs6497268	Eye color	Duffy et al., 2007	28338713	0.87	C/A	0.99 (0.95-1.04)	7.35×10^{-1}
OCA2	rs11855019	Eye color	Duffy et al., 2007	28335820	0.94	A/G	0.87 (0.82-0.92)	8.02×10^{-7}
OCA2	rs1800407	Eye color	Branicki et al., 2008; Sturm et al., 2008	28230318	0.83	C/T	1.13 (1.05-1.21)	5.58×10^{-4}
OCA2	rs1800414	Skin color	Donnelly et al., 2012; Edwards et al., 2010	28197037	0.81	T/C	1.02 (0.8-1.3)	9.02×10^{-1}

B. SNPs achieving genome-wide significance after adjusting for ten SNPs in Table 1

Gene	SNP	Position (bp)	Info*	Major/Minor allele	mOR‡	P-value
HERC2	rs150280398	28460081	0.880	C/T	1.29	4×10^{-8}
HERC2	rs186790432	28400732	0.889	A/G	1.31	4×10^{-9}
HERC2	rs189893622	28400060	0.868	C/A	1.32	2×10^{-9}
OCA2	rs149906873	28333710	0.777	G/A	1.31	6×10^{-10}
OCA2	rs7168800	28341575	0.797	G/A	1.28	2×10^{-9}
OCA2	rs72712694	28307735	0.773	T/C	1.31	6×10^{-10}
OCA2	rs72714118	28333196	0.795	G/A	1.28	2×10^{-9}
OCA2	rs72714121	28334889	0.795	G/T	1.28	2×10^{-9}
OCA2	rs77542847	28330373	0.794	C/A	1.27	3×10^{-9}

* Impute-2 measure of imputation accuracy;
† Odds ratio per minor allele associated with SCC risk and 95% confidence interval;
‡ mOR = odds-ratio per minor allele, adjusted for gender, ancestry, array/reagent and genotypes at 10 SNPs in Table 1 of text.

Table S2. Nonsynonymous SNPs in MC1R previously associated with pigmentation and skin cancer

SNP	rs#	Previously associated phenotypes	Info*	Risk† allele	ORs (95%CI) ‡	p
Val60Leu	rs1805005	Sunburn, NMSC, CMM (Bastiaens et al., 2001, Kennedy et al., 2001)	0.617	T	0.89 (0.83-0.95)	9.01E-04
Asp84Glu	rs1805006	red hair, fair skin, NMSC (Bastiaens et al., 2001, Kennedy et al., 2001)		A	MAF<0.01	
Val92Met	rs2228479	Sunburn, NMSC, CMM (Bastiaens et al., 2001, Kennedy et al., 2001)	0.932	A	1.08 (1.02-1.15)	1.31E-02
Arg142His	rs11547464	red hair, fair skin, BCC, CMM (Bastiaens et al., 2001, Kennedy et al., 2001)		A	MAF<0.01	
Arg151Cys	rs1805007	red hair, fair skin, BCC, SCC, CMM, NMSC (Box et al., 2001, Palmer et al., 2000, Bastiaens et al., 2001, Kennedy et al., 2001)	0.809	T	1.50 (1.41-1.6)	5.69E-37
Arg160Trp	rs1805008	red hair, fair skin, BCC, SCC, CMM, NMSC (Box et al., 2001, Palmer et al., 2000, Bastiaens et al., 2001, Kennedy et al., 2001)	0.885	T	1.31 (1.23-1.39)	1.17E-18
Arg163Gln	rs885479	blonde/brown hair, CMM, BCC (Palmer et al., 2000, Kennedy et al., 2001)	0.790	A	0.98 (0.9-1.07)	6.80E-01
His260Pro	N/A	red hair, fair skin, NMSC, CMM (Bastiaens et al., 2001, Kennedy et al., 2001)		C		
Asp294His	rs1805009	red hair, fair skin, BCC, SCC, CMM, NMSC (Box et al., 2001, Palmer et al., 2000, Bastiaens et al., 2001, Kennedy et al., 2001)	0.476	C	1.1 (0.88-1.37)	3.90E-01

* Impute-2 measure of imputation accuracy;
† The allele that was associated with skin cancer-predisposing pigmentation traits;
‡ Odds ratios (and 95% confidence intervals) associated with the risk of SCC in the present study.

Table S3. Odds-ratios (ORs) for chromosome 16 SNPs, after adjustment for the total dosage of the MC1R nonsynonymous SNPs*

SNP	risk_allele	OR	95%CI	P values
rs4268748	C	1.22	1.17-1.28	1.96E-16
rs35063026	T	1.33	1.26-1.42	1.17E-20
rs78703231	C	1.05	0.99-1.1	8.21E-02

* SNPs listed in Table S.4 with MAF≥0.01 and with an rs number.

Table S4. SNPs determining polygenic scores for: A. skin color, B. sun sensitivity and C. tanning ability*

Gene	SNP	Locus	Major/minor allele	Risk allele†	OR (95% CI)‡
A. Skin Color					
SLC45A2	rs16891982	5p13.3	C/G	G	1.91 (1.7-2.14)
TYR	rs1042602	11q14.3	C/A	A	0.98 (0.94-1.01)
SLC24A5	rs1834640	15q21.1	A/G	A	1.04 (0.85-1.27)
IRF4 §	rs12203592	6p25.3	C/T	T	1.56 (1.49-1.62)
B. Sun Sensitivity					
NTM	rs12421680	11q25	G/A	A	0.99 (0.96-1.03)
IRF4	rs12203592	6p25.3	C/T	T	1.56 (1.49-1.62)
TYR	rs1126809	11q14.3	G/A	A	1.19 (1.15-1.24)
MC1R	rs1805007	16q24.3	C/T	T	1.5 (1.41-1.6)
MC1R ¶	rs1805008	16q24.3	C/T	T	1.31 (1.24-1.4)
ASIP haplotype	rs4911414_T, rs1015362_G	20	G/T	T	1.30 (1.22-1.38)
C. Tanning ability					
PPARGC1B	rs32579	5q32	G/A	G	1.06 (1.02-1.11)
IRF4	rs12203592	6p25.3	C/T	T	1.56 (1.49-1.62)
TYR	rs1393350	11q14.3	G/A	A	1.18 (1.14-1.23)
CALCOCO1/HOXC13	rs7969151	12q13.13	G/A	A	0.99 (0.95-1.04)
PAPOLA/VRK1	rs17094273	14q32.2	G/A	A	1.02 (0.97-1.08)
HERC2	rs12913832	15q13.1	G/A	G	1.12 (1.07-1.16)
MC1R	rs1805007	16q24.3	C/T	T	1.5 (1.41-1.6)
DBNDD1	rs11648785	16q24.3	C/T	C	1.17 (1.12-1.22)
PRDM15	rs7279297	21q22.3	A/G	A	0.98 (0.94-1.02)
SLC45A2	rs35391	5p13.2	C/T	C	1.83 (1.59-2.1)
ASIP haplotype	rs4911414_T, rs1015362_G	20q11.22	G/A	G	1.30 (1.22-1.38)

* all SNPs obtained from GWAS catalogue (with $p<5x10^{-6}$), excluding redundant ones either due to high LD with other SNPs in the table or deemed non-independent signal from other SNPs;
† allele associated with high-risk pigment phenotype;
‡ OR and 95% CI associated with SCC in the present study;
§ Replacing rs1540771 in GWAS catalogue, based on Han et al., 2008;
¶ Not in GWAS catalog, but a significant predictor of skin sensitivity to sun in Sulem et al., 2007.

Table S5. SCC risk in relation to joint tertiles of SCC risk allele count and score for: A. skin color; B. sun sensitivity and C. tanning ability

Number of SCC risk alleles *	Cases	Controls	OR †	CI ‡	Cases	Controls	OR	CI	Cases	Controls	OR	CI	Cases	Controls	OR	CI
	A. Skin Color §															
	Dark				Medium				Fair				Total			
Low	431	6981	1.00	REF	650	8415	1.21	1.06-1.37	404	4485	1.40	1.21-1.61	1485	19881	1.00	-
Moderate	533	5516	1.41	1.23-1.61	779	7034	1.57	1.39-1.78	620	4438	1.99	1.75-2.27	1932	16988	1.37	1.27-1.47
High	960	6434	2.08	1.84-2.34	1665	9078	2.50	2.23-2.79	1604	7301	2.94	2.63-3.3	4229	22813	2.08	1.95-2.22
Total	1924	18931	1.00	REF	3094	24527	1.18	1.11-1.25	2628	16224	1.41	1.32-1.5	7646	59682	-	-
	B. Sun Sensitivity ¶															
	Low				Moderate				High				Total			
Low	889	12055	1.00	REF	483	6127	1.07	0.95-1.20	113	1699	0.93	0.76-1.14	1485	19881	1.00	-
Moderate	588	5534	1.31	1.18-1.47	807	7122	1.39	1.26-1.54	537	4332	1.54	1.37-1.72	1932	16988	1.31	1.22-1.41
High	303	2318	1.62	1.41-1.86	1020	6674	1.82	1.65-2.01	2906	13821	2.43	2.24-2.63	4229	22813	1.86	1.72-2.00
Total	1780	19907	1.00	REF	2310	19923	1.05	0.98-1.13	3556	19852	1.29	1.20-1.39	7646	59682	-	-
	C. Tanning Ability \\															
	High				Moderate				Low				Total			
Low	839	12125	1.00	REF	499	5931	1.15	1.02-1.29	147	1825	1.07	0.89-1.29	1485	19881	1.00	-
Moderate	518	5135	1.33	1.18-1.49	813	7128	1.47	1.32-1.62	601	4725	1.58	1.41-1.76	1932	16988	1.30	1.21-1.41
High	371	2625	1.79	1.57-2.04	1174	7255	2.02	1.84-2.23	2684	12933	2.47	2.27-2.69	4229	22813	1.90	1.77-2.04
Total	1728	19885	1.00	REF	2486	20314	1.11	1.04-1.19	3432	19483	1.28	1.19-1.37	7646	59682	-	-

* risk alleles of top SNPs at six pigment-related loci;
† OR=odds-ratio, adjusted for gender, PCAs and genotyping array/reagent. Marginal OR's for each score are also adjusted for tertiles of the other score;
‡ CI=95% confidence interval;
§ polygenic skin color score: higher allele counts associated with lighter skin color;
¶ number of SNP alleles associated with burning;
\\ number of SNP alleles associated with tanning.

Table S6. Effects of SNP rs4455710 at 6p21, by history of smoking and immunosuppression

		#cases	#controls	OR (allele C)	95% CI	P-value
Smoking status	never	3463	32189	0.84	0.80-0.88	1.356e-11
	ever	3786	25586	0.87*	0.83-0.91	2.964e-08
	unknown	452	2391	---	---	---
Immuno -suppressed	yes	111‡	355§	0.75	0.54-1.04	0.0807
	no	7590	59811	0.86†	0.83-0.89	2.069e-18

* Interaction P-value = 0.3833
† Interaction P-value = 0.4483
‡ 42 with organ transplant, 52 with chronic lymphocytic leukemia (CLL), and 17 with HIV infection.
§ 49 with organ transplant, 99 with CLL and 207 with HIV infection.

Table S7. Significant (P<0.05) evidence for epistasis between pairs of SNPs in text Table 1

Locus 1					**Locus 2**					**Interaction‡**		
Locus	**SNP**	**Risk Allele**	**log OR***	**SD†**	**Locus**	**SNP**	**Risk Allele**	**log OR**	**SD**	**log OR**	**SD**	**P-value§**
6p21	rs4455710	T	0.26	0.04	15q13	rs12916300	T	0.18	0.03	-0.07	0.03	0.023
6p25	rs12203592	T	0.90	0.18	5p13	rs16891982	G	0.76	0.08	-0.23	0.09	0.013
20q11	rs6059655	A	0.86	0.27	5p13	rs16891982	G	0.70	0.07	-0.30	0.14	0.032
			0.21	0.04	11q14	rs1126809	A	0.15	0.02	0.11	0.04	0.017
			0.33	0.04	16q24	rs4268748	C	0.31	0.02	-0.10	0.05	0.041

* log odds-ratio;
† SD = standard deviation;
‡ on log-additive scale;
§ interaction based on chi-squared likelihood ratio test.

Table S8. Subject Quality Control

Step	Description	Subjects		
		Cases	Controls	Total
0	All self-reported non-Hispanic White RPGEH subjects	9,604	64,218	73,822
1	Exclude subjects with ambiguous sex (n = 65)	9,598	64,159	73,757
2	Exclude subjects with ambiguous SCC status (n = 1806)	7,792	64,159	71,951
3	Exclude subjects with CR* <0.97 & het† > 0.4 (n=59)	7,782	64,110	71,892
4	Prune related subject pairs‡ (n = 3,696)	7,719	60,477	68,196
5	Exclude subjects lacking European PCAs (n = 329)	7,701	60,166	67,867

* CR = call rate;
† het = heterozygosity (observed heterozygotes/total number of SNPs with no missing genotypes);
‡ expected number of alleles shared IBD >= 0.25.

Table S9. Distribution of Eligible Subjects by Study Phase, Case-control Status and Gender

	Screening Phase		Replication Phase		Total	
	Cases	Controls	Cases	Controls	Cases	Controls
Males	3825	21298	595	3026	4420	24324
Females	3066	33268	215	2574	3281	35842
Total	6891	54566	810	5600	7701	60166

Supplementary Figures

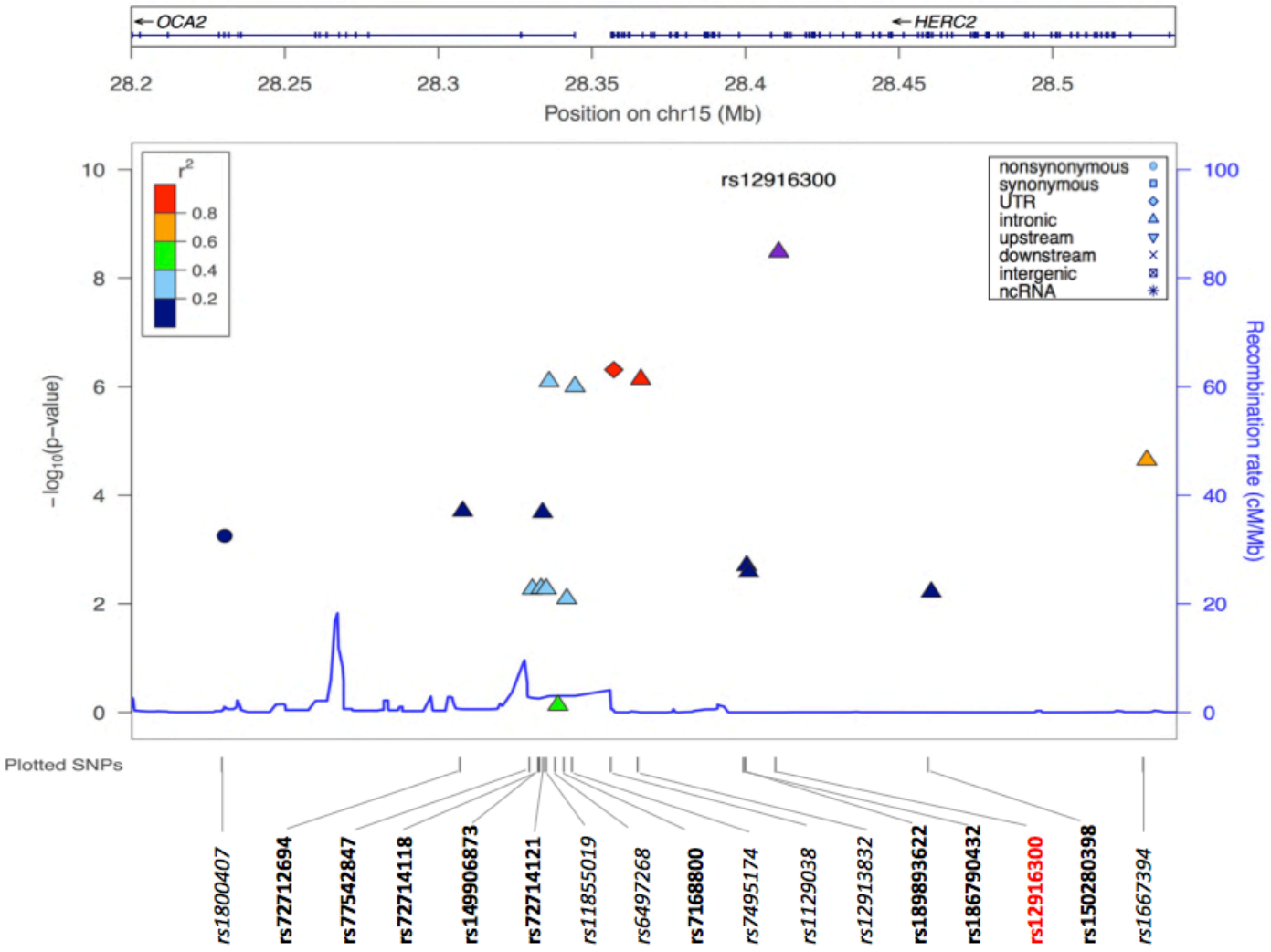

Figure S1. Manhattan plot enlargement showing significance levels for selected SNPs in 0.35 Mb region at Chr15q13. SNPs listed include top SNP (red), those previously associated with pigmentation *traits* (italics) and those that achieved significance after adjusting for all ten SNPs in Table 1 (bold).

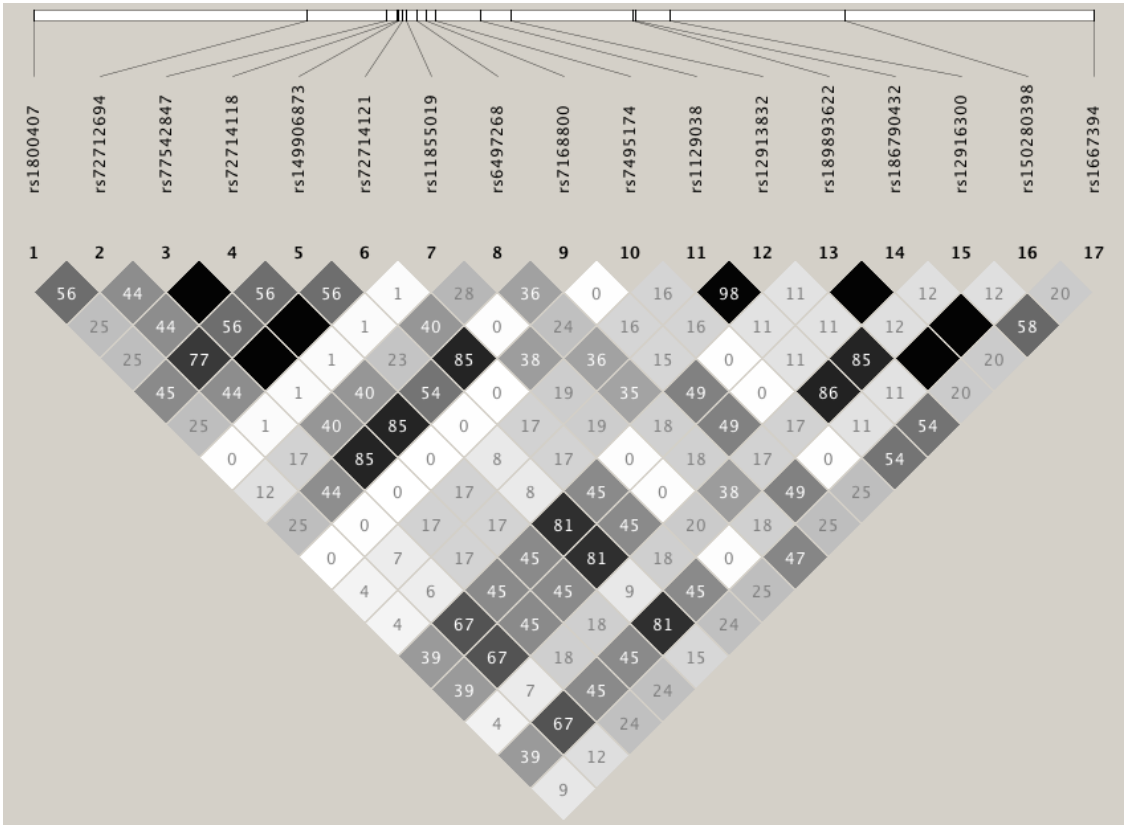

Figure S2. Squared correlation coefficients (R^2) between pairs of statistically significant and previously reported SNPs in the HERC2/OCA2 region on chromosome 15q13, based on Caucasian 1000 Genomes data. Black squares indicate perfect correlation.

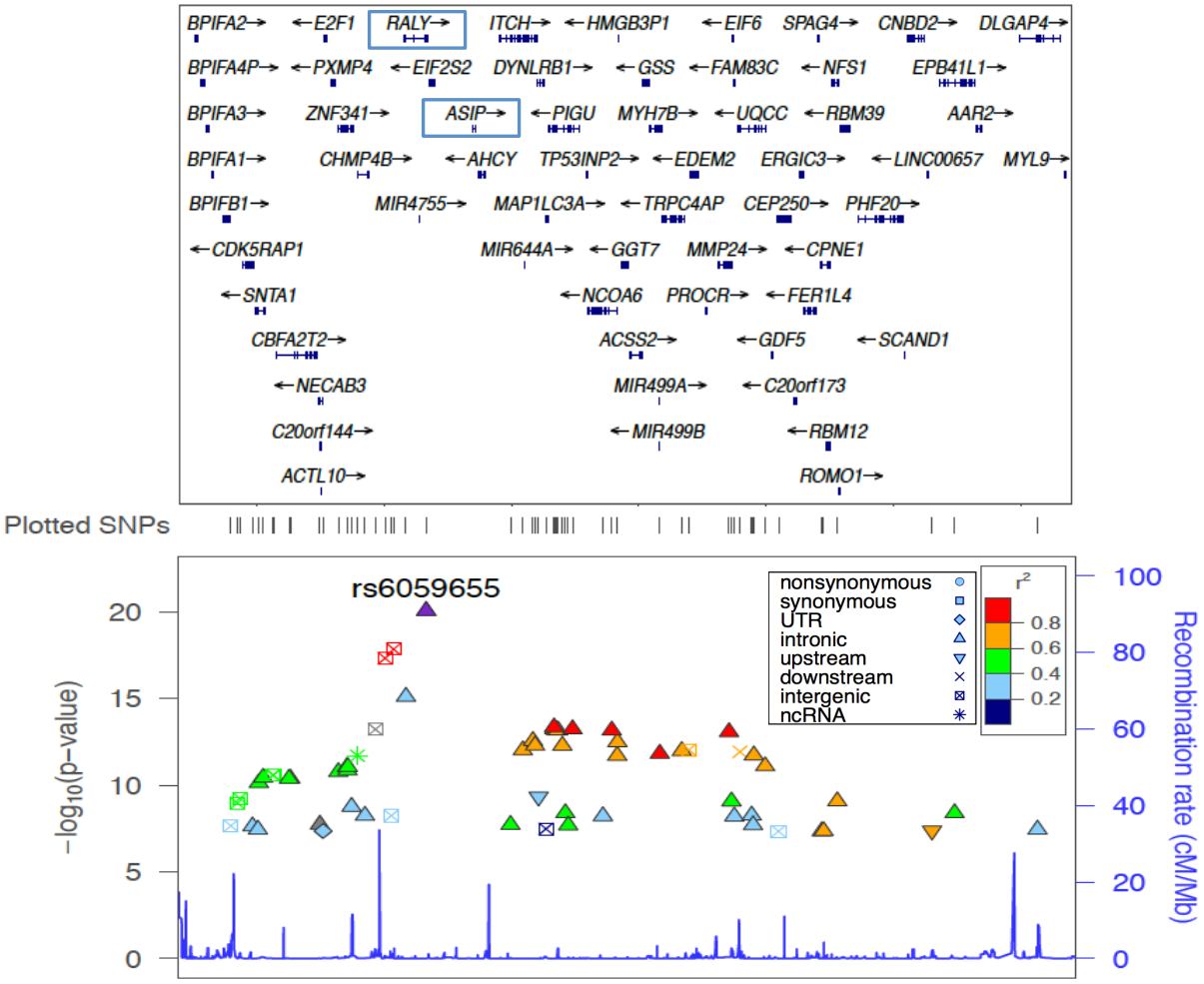

Figure S3. Manhattan plot enlargement showing 3.5 Mb region at the 20q11 locus. The upper panel shows the name and location of genes in the region, with an arrow indicating the transcribed strand of a gene and ticks indicating exons. The genes *RALY* (containing the top SNP rs6059655) and *ASIP* (associated with skin pigmentation and skin cancer) are shown in boxes. The lower panel shows the significance levels of the SNPs in this region.

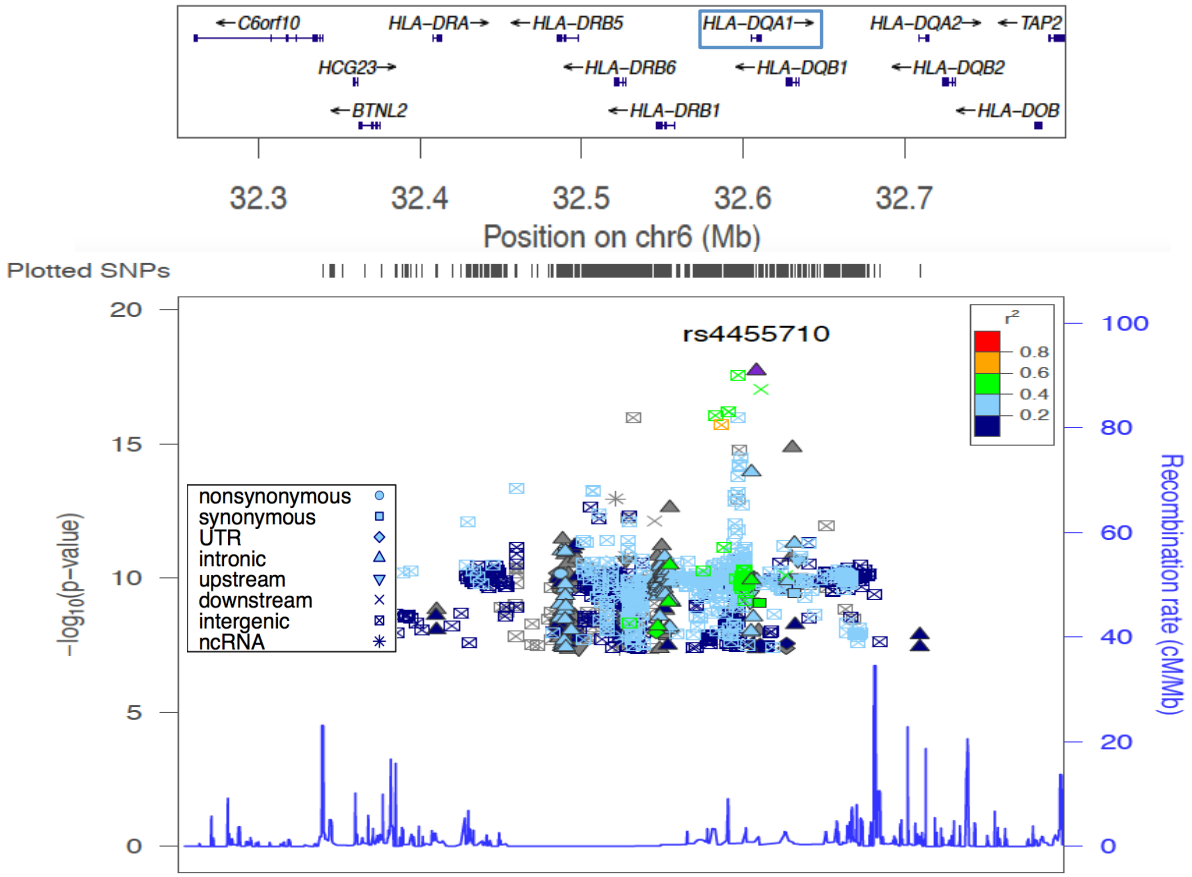

Figure S4. Manhattan plot enlargement showing 0.55 Mb region at the 6p21 locus. The upper panel shows the name and location of genes in the region, with an arrow indicating the transcribed strand of a gene and ticks indicating exons. The box encloses the gene *HLA-DQA1* containing the top SNP rs4455710. The lower panel shows the significance levels of the SNPs in this region.

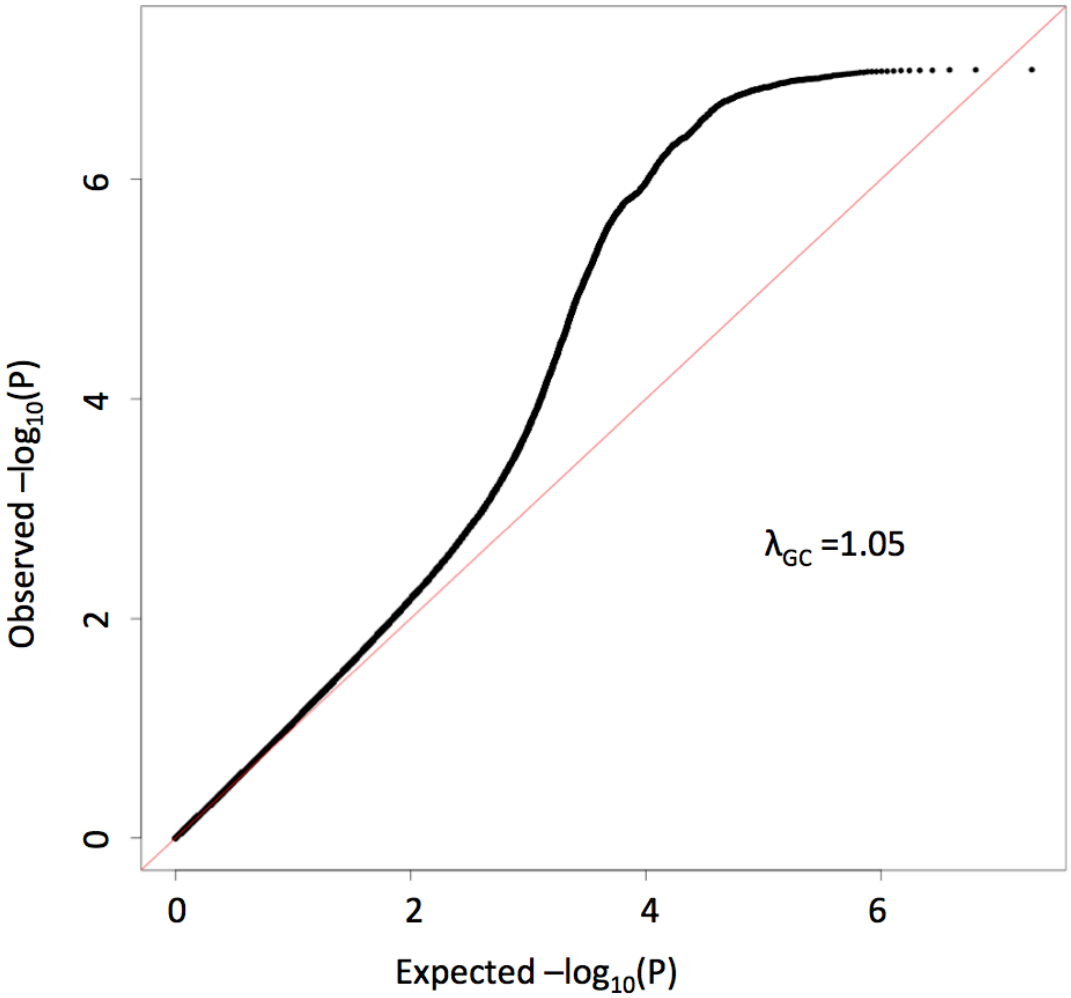

Figure S5. Plot of P-values > 10^{-8} corresponding to percentiles of the empirical distribution of 9,610,468 squared Cochran-Armitage trend statistics (vertical axis) versus P-values of percentiles of a central chi-squared variable with one degree of freedom (horizontal axis).

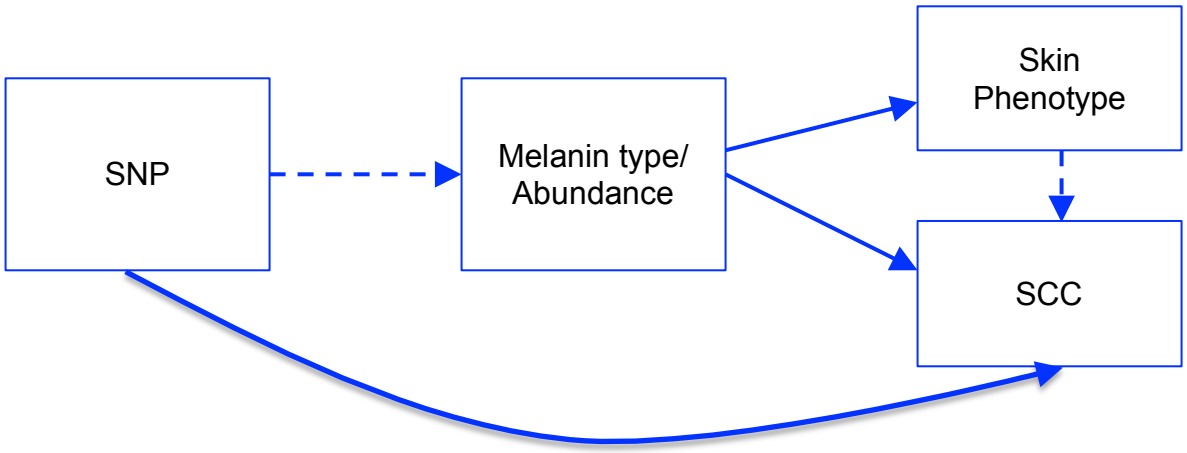

Figure S6. Possible mechanisms for association of SCC risk with SNPs at pigment-related loci

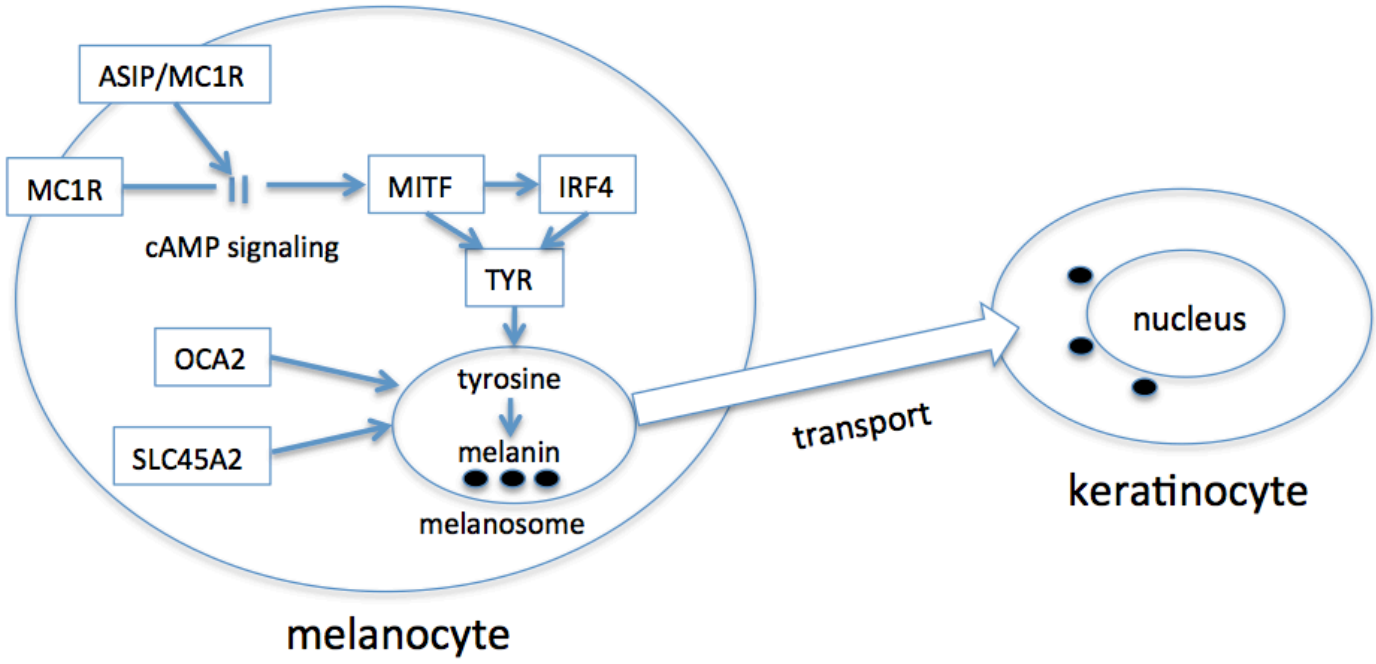

Figure S7. MC1R, located on the plasma membrane of melanocytes, activates the cAMP signaling pathway that increases MITF expression, while ASIP and other MC1R-antagonists block this pathway. MITF up-regulates the transcription of major melanogenic enzymes such as tyrosine kinase encoded by TYR, and can also activate expression of IRF4, which cooperates with MITF to activate TYR expression. The membrane-associated transporters SLC45A2 and OCA2 are involved in post-translation processing and transporting of melanosomal proteins into the melanosomes and thereby supporting melanin production. Melanin in the melanocytes is then transported to keratinocytes, where it forms a coat around their nuclei and protects them from UVR-induced DNA damage.