Research Article

A Genome-Wide Association Study of Cutaneous Squamous Cell Carcinoma among European Descendants

Satu J. Siiskonen[1], Mingfeng Zhang[2], Wen-Qing Li[3,4], Liming Liang[5,6], Peter Kraft[5,6], Tamar Nijsten[1], Jiali Han[5,7,8], and Abrar A. Qureshi[3,4,9,10]

Abstract

Background: No GWAS on the risk of cutaneous squamous cell carcinoma (SCC) has been published. We conducted a multistage genome-wide association study (GWAS) to identify novel genetic loci for SCC.

Methods: The study included 745 SCC cases and 12,805 controls of European descent in the discovery stage and 531 SCC cases and 551 controls of European ancestry in the replication stage. We selected 64 independent loci that showed the most significant associations with SCC in the discovery stage (linkage disequilibrium $r^2 < 0.4$) for replication.

Results: Rs8063761 in the *DEF8* gene on chromosome 16 showed the strongest association with SCC ($P = 1.7 \times 10^{-9}$ in the combined set; $P = 1.0 \times 10^{-6}$ in the discovery set and $P = 4.1 \times 10^{-4}$ in the replication set). The variant allele of rs8063761 (T allele) was associated with a decreased expression of *DEF8*

($P = 1.2 \times 10^{-6}$). Besides, we validated four other SNPs associated with SCC in the replication set, including rs9689649 in *PARK2* gene ($P = 2.7 \times 10^{-6}$ in combined set; $P = 3.2 \times 10^{-5}$ in the discovery; and $P = 0.02$ in the replication), rs754626 in the *SRC* gene ($P = 1.1 \times 10^{-6}$ in combined set; $P = 1.4 \times 10^{-5}$ in the discovery and $P = 0.02$ in the replication), rs9643297 in *ST3GAL1* gene ($P = 8.2 \times 10^{-6}$ in combined set; $P = 3.3 \times 10^{-5}$ in the discovery; and $P = 0.04$ in the replication), and rs17247181 in *ERBB2IP* gene ($P = 4.2 \times 10^{-6}$ in combined set; $P = 3.1 \times 10^{-5}$ in the discovery; and $P = 0.048$ in the replication).

Conclusion: Several genetic variants were associated with risk of SCC in a multistage GWAS of subjects of European ancestry.

Impact: Further studies are warranted to validate our finding and elucidate the genetic function of these variants. *Cancer Epidemiol Biomarkers Prev; 25(4); 714–20. ©2016 AACR.*

Introduction

A number of genome-wide association studies (GWAS) have identified dozens of common genetic variants associated with

[1]Department of Dermatology, Erasmus Medical Center, Rotterdam, the Netherlands. [2]Department of Dermatology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts. [3]Department of Dermatology, Warren Alpert Medical School, Brown University, Providence, Rhode Island. [4]Department of Epidemiology, School of Public Health, Brown University, Providence, Rhode Island. [5]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts. [6]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts. [7]Department of Epidemiology, Richard M. Fairbanks School of Public Health, Simon Cancer Center, Indiana University, Indianapolis, Indiana. [8]Melvin and Bren Simon Cancer Center, Indiana University, Indianapolis, Indiana. [9]Department of Dermatology, Rhode Island Hospital, Providence, Rhode Island. [10]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts.

human pigmentation traits and skin cancers (1, 2). However, the published GWAS on skin cancers exclusively focused on melanoma and basal cell carcinoma (BCC), while no GWAS on squamous cell carcinoma (SCC) has been published to date. The candidate gene–based studies have identified several genes associated with the risk of SCC, including genes regulating pigmentation traits, such as melanocortin 1 receptor (*MC1R*; ref. 3–5), IFN regulatory factor 4 (*IRF4*; ref. 6), and tyrosinase (*TYR*; ref. 7), genes irrelevant to pigmentation traits, such as X-ray repair cross-complementing protein 1 (*XRCC1*; refs. 8, 9), cytotoxic lymphocyte-associated antigen-4 (*CTLA-4*; ref. 10), as well as ubiquitin-associated domain-containing protein 2 (*UBAC2*) and exocyst complex component 2 (*EXOC2*), which were identified in a candidate genetic study of SCC based on the GWAS significant loci of BCC (11). Aiming to identify novel genetic loci for SCC risk, we conducted a multi-stage GWAS on SCC in the populations of European descent.

Materials and Methods

Study population

SCC GWAS. In the discovery stage, we combined three component studies: the Harvard studies, the Rotterdam Study (RS)-I and the RS-II.

In the Harvard GWAS, we combined data from four case–control studies nested within the Nurses' Health Study (NHS) and the Health Professionals Follow-up Study (HPFS): a type II diabetes case–control study nested within the NHS (T2D_NHS), a type II diabetes case–control study nested within the HPFS

AACR

(T2D_HPFS), a coronary heart disease case–control study nested within the NHS (CHD_NHS), and a coronary heart disease case–control study nested within the HPFS (CHD_HPFS). Overlapped samples among the component datasets were excluded. Detailed descriptions of the study population were presented in the Supplementary Materials and Methods and were published previously (12, 13).

To define SCC in the Harvard GWAS, the participants in the NHS and HPFS reported newly diagnosed cancers biennially. With their permission, medical records were obtained and reviewed by physicians to confirm their self-reported diagnosis. We included the pathologically confirmed invasive SCC cases with no previously diagnosed cancer as the eligible cases in the SCC GWAS. The controls were defined as those who did not report any type of major cancers. The study protocol was approved by the Institutional Review Board of Brigham and Women's Hospital and the Harvard School of Public Health (Boston, MA).

The Rotterdam study. The Rotterdam Study (RS) is a prospective population-based follow-up study of the determinants and prognosis of chronic diseases in the elderly including skin diseases and cancer in participants living in Rotterdam, the Netherlands. The design has been prescribed in detail (14). In brief, the RS consists of a major cohort (RS-I) and two extensions (RS-II and RS-III). The cohort is predominantly (95%) Caucasian and the overall response rate for all three cycles at baseline was 72.0%. The Medical Ethics Committee of the Erasmus Medical Center and the review board of the Dutch Ministry of Health, Welfare and Sports have ratified the RS. From each participant, written informed consent was obtained. For this study, we included participants from the RS-I and RS-II cohorts, as the number of SCC cases in RS-III was very small.

To ascertain histologically confirmed SCC cases, all participants from the RS were linked with the nationwide network and registry of histo- and cytopathology in the Netherlands (PALGA; up to 23rd September 2011; ref. 15). PALGA was founded in 1971 and achieved complete national coverage in 1991. Every obtained pathology excerpt contains encrypted patient data, a report identifier, a conclusion of the pathologist (often differentiating between biopsy and excision and stating the localization of the SCC), and a PALGA diagnosis line derived from Systematized Nomenclature of Medicine (15). We used a previously published approach to identify unique SCC cases based on date of diagnosis, biopsy/excision and/or tumor localization (16). If the diagnosis or the number of unique SCC remained unsure, the medical files were hand searched and cases that remained dubious were excluded.

Replication study. In the replication stage, a fast-track replication was conducted among 531 SCC cases and 551 healthy controls in the skin cancer case–control study nested within NHS and HPFS (skin cancer study). All the cases and controls in this study were from the subcohorts of NHS and HPFS who had given a blood specimen. Eligible cases consisted of pathologically confirmed invasive SCC cases diagnosed after the baseline up to 2006 follow-up cycle for both cohorts, who had no previously diagnosed cancer. Controls were randomly selected from participants who were free of diagnosed cancer up to and including the questionnaire cycle in which the case was diagnosed. One or two controls

were matched to each case by age (± 1 year). Cases and their matched controls were selected in the same cohort.

Information on pigmentation traits were collected from prospective questionnaires in both NHS and HPFS using similar wording. We used categorical variables to indicate the natural hair color (red, blonde, light brown, dark brown, and black), tanning ability (practically none, light tan, average tan and deep tan in NHS; painful burn and peel, burn then tan, tan without burn in HPFS), and the total number of lifetime severe sunburns (none, 1–2, 3–5, 6, and more).

Genotyping, imputation, and quality control

SCC GWAS. We previously performed genotyping in the component sets of Harvard GWAS using the Affymetrix 6.0 array (12, 13). We used the MACH program (17) to impute 2,543,887 autosomal SNPs based on haplotypes from the HapMap (18) database phase II data build 35 (CEU) in all of the four component sets, including T2D_NHS, T2D_HPFS, CHD_NHS, and CHD_HPFS (19). Samples from the four studies were imputed separately. We observed high imputation quality in each cohort's imputation. SNPs with imputation $r^2 > 0.95$ and minor allele frequency (MAF) > 0.01 in each study were included in meta-analysis. Finally, a total of 1,777,244 SNPs were included in the Harvard GWAS.

Genotyping from participants of the RS has been described before (14). In brief, DNA from whole blood was extracted following standard protocols. The Infinium II HumanHap550K Genotyping BeadChip version 3 was used to genotype both RS-I and RS-II cohorts. Next, the RS-I and RS-II cohorts were imputed separately using the HapMap Phase II CEU reference panel (Build 36) as the reference panel and using a two-step procedure imputation algorithm implemented in the program MACH. SNPs with imputation $r^2 < 0.03$, MAF < 0.02 for RS-I, and MAF < 0.08 for RS-II were excluded. After quality control, a total of 2,356,032 SNPs from RS-I and 1,956,891 SNPs from RS-II were available for GWAS and meta-analysis.

Replication study. We selected 64 independent SNPs [linkage disequilibrium (LD) $r^2 < 0.4$] showing the strongest associations with SCC in the discovery stage for replication in the skin cancer study. We genotyped these SNPs using TaqMan OpenArray system 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. We excluded five SNPs (rs4980694, rs12210050, rs13156707, rs11263585, and rs3099065) that failed genotyping in the replication set. The rest 59 SNPs were successfully genotyped with call rate >85% and Hardy–Weinberg *P* value > 0.01. Detailed information of these SNP was presented in Supplementary Table S1.

Seven common *MC1R* variants (Val60Leu, Val92Met, Arg151Cys, Ile155Thr, Arg160Trp, Arg163Gln, and Asp294His) were previously genotyped among a subgroup of the skin cancer case–control study (257 SCC cases and 282 controls). Detailed descriptions of this subgroup study were previously published (5).

Statistical analysis

We used logistic regression to test associations between minor allele counts and SCC risk in the discovery and replication sets. In

Siiskonen et al.

Table 1. The component datasets of the SCC GWAS

	Sample size		Genotype	
Dataset	Total	SCC cases/controls	Platform	Number of imputed SNPs[a]
Discovery set				
Harvard GWAS	5,509	424/5,085	Affymetrics 6.0	1,777,244[b]
RS-I	5,892	258/5,634	Affymetrics 6.0	2,356,032[c]
RS-II	2,149	63/2,086	Affymetrics 6.0	1,956,891[d]
All (meta-analysis)	13,550	745/12,805		2,392,512
Replication set				
Skin cancer study	1,082	531/551	TaqMan OpenArray	59[e]

[a]Imputated on the basis of the genotyped SNPs and haplotype information in the HapMap phase II data build 35 (CEU).
[b]Filtered by imputation $r^2 > 0.95$ and MAF > 0.01.
[c]Filtered by imputation $r^2 > 0.3$ and MAF > 0.02.
[d]Filtered by imputation $r^2 > 0.3$ and MAF > 0.08.
[e]64 SNPs were selected for replication but five of them failed in genotyping.

the discovery stage, we used the imputed genotype data based on HapMap phase II data build 35 (CEU) for the analyses. We analyzed the three component GWASs in the discovery set separately. We adjusted for age, gender, and the three largest principal components of genetic variation of each GWAS in the regression model. These principal components were calculated for all individuals on the basis of ca. 10,000 unlinked markers using the EIGENSTRAT software (20). We additionally adjusted for the component study set in the Harvard GWAS. We used PLINK for the GWAS analyses in this study. Associations in each GWAS were combined in an inverse variance weighted meta-analysis using METAL (21). Age and gender were adjusted in the skin cancer replication set. The same meta-analysis was conducted to combine the discovery set and the replication set.

Results

We combined three component studies in the discovery stage of this SCC GWAS, which included the Harvard GWAS, the RS-I, and the RS-II. We used the skin cancer case–control study nested in the Harvard cohorts in the replication stage. The details of each component study of this multi-stage SCC GWAS were summarized in Table 1. In the discovery stage, we analyzed a total of 2,392,512 autosomal SNPs imputed from HapMap phase II data build 35 (CEU) after quality control for their associations with SCC. The quantile-quantile (Q–Q) plot of the SCC GWAS did not demonstrate a systematic deviation from the expected distribution (Supplementary Fig. S1). The overall genomic control inflation factor (λ_{GC}) was 1.00. The Manhattan plot of this SCC GWAS was presented in Supplementary Fig. S2. We selected 64 independent SNPs (LD $r^2 < 0.4$) that showed the strongest associations (by P values) with SCC in the discovery stage for replication. Detailed information of these selected SNP was presented in Supplementary Table S1. Five SNPs (rs4980694, rs12210050, rs13156707, rs11263585, and rs3099065) that failed genotyping in the replication set were excluded.

After combining the discovery set and the replication set, we identified the SNP rs8063761 on chromosome 16 most significantly associated with SCC risk ($P = 1.7 \times 10^{-9}$ in the combined set; $P = 1.0 \times 10^{-6}$ in the discovery set; and $P = 4.1 \times 10^{-4}$ in the replication set). The variant allele of this SNP (T allele; MAF, 0.33) was associated with an increased risk of SCC with odds ratio (OR) of 1.34 [95% confidence interval (CI) 1.22–1.47] compared with the wild allele (A allele; Table 2). This SNP was located in the *intron* region of the differentially expressed in FDCP 8 homolog (*DEF8*) gene. We additionally evaluated the association of all the SNPs within 200 kb surrounding this gene with SCC risk and presented the regional plot of *DEF8* gene in Fig. 1. As shown in the regional plot, the SNP rs8051733 was in high LD with the identified SNP rs8063761 (LD $r^2 > 0.8$). The SNP rs8051733 was also in the *intron* region of *DEF8* gene, and the variant allele of the SNP rs8051733 (G allele; MAF, 0.30) was associated with an increased risk of SCC with OR of 1.37 (95% CI, 1.16–1.61; $P = 2.0 \times 10^{-4}$ in the discovery set) compared with the wild allele (A allele).

We additionally tested the association between the SNP rs8063761 and human pigmentation traits as well as the risk of other skin cancers based on our published GWAS data (2). As summarized in Supplementary Table S2, we found that the variant allele of this SNP (T allele) was associated with lighter hair color ($P = 6.9 \times 10^{-9}$), poorer tanning ability

Table 2. Association between the SNP rs8063761 in the intron region of *DEF8* gene and the risk of SCC

	Genotype (cases/controls)					
Dataset	AA	AT	TT	OR (95% CI)	P	$P_{heterogeneity}$
Discovery set[a]						
Harvard GWAS	178/2,427	192/2,167	54/490	1.26 (1.09–1.47)	2.23E−03	
RS-I	96/2,710	125/2,405	37/519	1.42 (1.18–1.71)	1.66E−04	
RS-II	26/189	29/908	8/989	1.25 (0.86–1.71)	0.24	
All (meta-analysis)	300/5,326	346/5,480	99/1,998	1.32 (1.18–1.47)	1.01E−06	0.60
Replication set[b]						
Skin cancer study	211/267	229/222	79/50	1.38 (1.16–1.65)	4.09E−04	
Combined set (meta-analysis)				1.34 (1.22–1.47)	1.74E−09	0.66

[a]Adjusted for age, gender, and the first three principal components in each study.
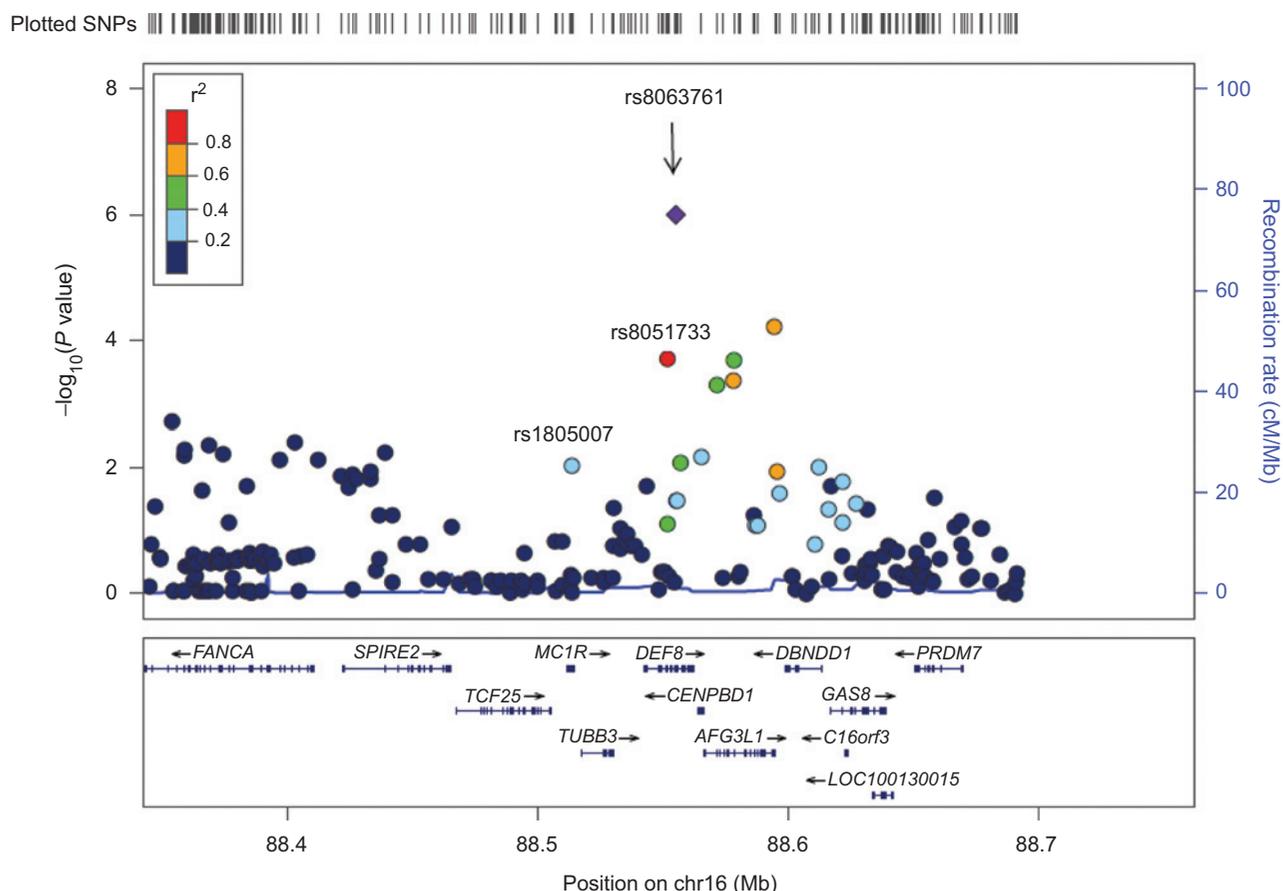[b]Adjusted for age and gender.

Figure 1.
Regional plot of the *DEF8* gene. Regional Manhattan Plots for the SCC GWAS within 200 kb surrounding the *DEF8* gene on chromosome 6.

($P = 2.1 \times 10^{-44}$), and increased number of previous sunburns ($P = 3.7 \times 10^{-11}$). No significant association was found between this SNP and the risk of BCC ($P = 0.06$) or the number of non-melanoma skin cancer ($P = 0.47$).

We compared the association between the SNP rs8063761 and SCC risk before and after adjustment of human pigmentation traits in the skin cancer study. As shown in Supplementary Table S3, this association remained significant, but substantially attenuated after additionally adjusted for hair color, tanning ability, and the number of sunburns ($P = 4.1 \times 10^{-4}$ before adjustment and $P = 0.01$ after adjustment).

Of interest, we found that the SNP rs8063761 was associated with the expression of *DEF8* gene based on the published data of expression quantitative trait loci derived from a GWAS of global gene expression in the lymphoblastoid cell lines of 550 individuals of British descent (22). The variant allele of the SNP rs8063761 (T allele) was associated with a decreased expression of *DEF8* gene ($\beta = -0.14$; SE $= 0.03$; $P = 1.2 \times 10^{-6}$). These data were available online from the Real-time Engine of eQTLs uploaded by Dr. Liang (23).

However, we noticed that the *MC1R* gene (a previously known pigmentation gene associated with SCC risk) is located close to the *DEF8* gene. The SNP rs1805007 in the *MC1R* gene is in weak LD with the SNP rs8063761 in the *DEF8* gene (LD $r^2 = 0.3$). The variant allele of the SNP rs1805007 (T allele; MAF,

0.07) was associated with an increased risk of SCC with OR of 1.31 (95% CI, 1.07–1.62; $P = 9.6 \times 10^{-3}$ in the discovery set) compared with the common allele (C allele). After mutual adjustment by each SNP in the Harvard GWAS set, the association between the SNP rs8063761 and SCC risk remained significant ($P = 9.8 \times 10^{-3}$; OR, 1.23; 95% CI, 1.05–1.45), while the association of the SNP rs1805007 was no longer significant ($P = 0.34$). We further adjusted for all 7 common *MC1R* variants (Val60Leu, Val92Met, Arg151Cys, Ile155Thr, Arg160Trp, Arg163Gln, and Asp294His) among a subgroup of the skin cancer case–control study (257 SCC cases and 282 controls genotyped for all 7 *MC1R* SNPs) and the association of the *DEF8* SNP rs8063761 became nonsignificant ($P = 0.85$; OR, 1.04; 95% CI, 0.71–1.51 after adjustment vs. $P = 4.2 \times 10^{-3}$; OR, 1.44; 95% CI, 1.12–1.85 before adjustment).

In addition, we validated the association between four other SNPs and SCC risk in the replication set, which were rs9689649 in the *intron* of parkinson protein 2 (*PARK2*) gene on chromosome 6 ($P = 2.7 \times 10^{-6}$ in the combined set; $P = 3.2 \times 10^{-5}$ in the discovery set and $P = 0.02$ in the replication set), rs754626 in the *intron* of v-src avian sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (*SRC*) gene on chromosome 20 ($P = 1.1 \times 10^{-6}$ in the combined set; $P = 1.4 \times 10^{-5}$ in the discovery set and $P = 0.02$ in the replication set), rs9643297 in the *intron* of ST3 beta-galactoside alpha-2,3-sialyltransferase 1 (*ST3GAL1*) gene on

Siiskonen et al.

Table 3. Four SNPs replicated for the association with the risk of SCC

SNP	Gene[a]	Location	Chr.	Position[b]	Ref	Eff (minor)	MAF	Meta-analysis		Discovery stage		Replication stage	
								OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P
rs9689649	PARK2	Intron	6	162,021,168	A	C	0.22	1.38 (1.21–1.58)	2.73E−06	1.45 (1.22–1.73)	3.15E−05	1.28 (1.04–1.57)	0.02
rs754626	SRC	Intron	20	35,450,754	T	G	0.25	1.32 (1.18–1.47)	1.12E−06	1.34 (1.17–1.53)	1.44E−05	1.26 (1.03–1.55)	0.02
rs9643297	ST3GAL1	Intron	8	134,552,877	A	G	0.31	1.32 (1.17–1.50)	8.20E−06	1.42 (1.21–1.68)	3.26E−05	1.21 (1.01–1.45)	0.04
rs17247181	ERBB2IP	Intron	5	65,291,428	C	T	0.10	1.40 (1.21–1.62)	4.22E−06	1.42 (1.21–1.68)	3.08E−05	1.34 (1.003–1.79)	0.048

[a]Closest gene to the genetic variant.
[b]Based on HapMap build 35.

chromosome 8 ($P = 8.2 \times 10^{-6}$ in the combined set; $P = 3.3 \times 10^{-5}$ in the discovery set and $P = 0.04$ in the replication set), and rs17247181 in the *intron* of erbb2 interacting protein (*ERBB2IP*) gene on chromosome 5 ($P = 4.2 \times 10^{-6}$ in the combined set; $P = 3.1 \times 10^{-5}$ in the discovery set and $P = 0.048$ in the replication set; Table 3).

We also checked the associations between the previously reported SCC risk SNPs identified by candidate gene-based approach for their associations in this study and observed significant associations with consistent directions of effect for the SNPs in *IRF4* (rs12203592, OR = 1.54 for T allele, $P = 2.57 \times 10^{-4}$), *UBAC2* (rs7335046, OR = 1.19 for G allele, $P = 0.03$), and *EXOC2* (rs12210050, OR = 1.34 for T allele, $P = 2.33 \times 10^{-5}$). The SNP in *CTLA4* (rs3087243) was not associated with SCC risk in our study population ($P = 0.47$). The *XRCC1* SNPs (rs25487 and rs1799782) were not tested in this study.

Discussion

In this study, we identified several genetic loci associated with SCC risk by a multi-stage SCC GWAS. To the best of our knowledge, this is the first reported GWAS on SCC to date, which sheds new lights on the genetic basis of SCC development.

The most significant locus identified in this SCC GWAS was the SNP rs8063761 in the *intron* of *DEF8* gene. This SNP was also associated with human pigmentation traits, including hair color ($P = 6.9 \times 10^{-9}$), tanning ability ($P = 2.1 \times 10^{-44}$), and the number of sunburns ($P = 3.7 \times 10^{-11}$) in our cohort population. Given that human pigmentation traits are susceptibility factors for skin cancer (1, 24), we compared the association of this SNP with SCC before and after adjustment of pigmentation traits. We found that this association substantially attenuated after adjusted for hair color, tanning ability, and the number of sunburns in the skin cancer study in the Harvard cohorts ($P = 4.1 \times 10^{-4}$ before adjustment and $P = 0.01$ after adjustment), suggesting a mediated association of this SNP with SCC through pigmentation. Of interest, we found this SNP was also strongly associated with the expression level of *DEF8* gene based on the published eQTL database ($P = 1.2 \times 10^{-6}$; ref. 23).

However, the genetic function of *DEF8* gene was largely unknown. Furthermore, we noticed that the *DEF8* gene was close to the *MC1R* gene, which is a well-known pigmentation gene and has been associated with skin cancer risk (3–5). In this SCC GWAS, only one SNP in the *MC1R* gene, rs1805007, was in weak LD with the SNP rs8063761 (LD $r^2 = 0.3$; based on HapMap phase II data build 35, CEU). After adjusting for the SNP rs1805007 in the Harvard GWAS set, the association of the SNP rs8063761 with SCC risk remained significant ($P = 9.8 \times 10^{-3}$; OR, 1.23; 95% CI, 1.05–1.45). However, further adjustment of all 7 common *MC1R* variants (Val60Leu, Val92Met,

Arg151Cys, Ile155Thr, Arg160Trp, Arg163Gln, and Asp294His) in a subgroup of the replication study (257 SCC cases and 282 controls) eliminated such an association ($P = 0.85$ after adjustment vs. $P = 4.2 \times 10^{-3}$ before adjustment), suggesting such an association may be driven by the *MC1R* SNPs.

Other genetic loci identified in this study for the association with SCC include polymorphisms within the *ST3GAL1*, *SRC*, *ERBB2IP* and *PARK2* genes. Although their associations didn't reach GWAS significance, the replication of these findings suggested potential effects of these loci on SCC risk. The *ST3GAL1* gene encodes a sialyltransferase, β-galactoside α-2,3-sialyltransferase 1 (ST3Gal I), which has been reported to play a role in the development of multiple tumors (25–28). A previous study found high expression of ST3Gal I in human cutaneous SCC lesions (29), which further supported our findings. However, this is the first study to report an association between a genetic polymorphism in *ST3GAL1* gene and skin cancer risk. The *SRC* gene is as a well-known proto-oncogene. Genetic abnormalities in this gene have been largely reported in multiple cancers (30–32), including melanoma (33). However, no previous studies have found genetic variants in *SRC* associated with SCC risk. The *ERBB2IP* gene encodes the Erbin protein, which is a binding partner of the Erb-B2 protein and regulates Erb-B2 function and localization. Erbin has also been shown to affect the Ras signaling pathway by disrupting Ras–Raf interaction (34). Altered expression of Erbin and Erb-B2 has been found in skin BCC tumors in a previous study, but not in SCC (35). The Parkinson disease–associated gene, *PARK2*, has also been associated with human malignancies in previous studies, including lung cancer (36, 37), glioblastoma (38, 39), and breast cancer (40), but no studies have reported on skin cancer. The precise function of this gene remains largely unknown.

A major strength of this study is the large sample size of pathologically confirmed SCC cases and healthy controls in this study and it is the first reported SCC GWAS so far. With the sample size of 745 SCC cases and 12,805 healthy controls, we estimated the power of 80% to detect the effect size of 1.40, 1.30, 1.20, and 1.15 for the genetic variants with minor allele frequency of 0.05, 0.10, 0.25, and 0.5, respectively, based on a two-sided α of 0.05. Besides, we were able to collect information on human pigmentation traits prospectively in our cohort population for mediation analyses.

In summary, we identified several novel genes and genetic loci associated with SCC risk using a multi-stage GWAS design. The identification of these genetic loci may help us understand the complex mechanisms of developing SCC and suggest new therapeutic targets for this common skin cancer. Further studies are warranted to validate our findings and extend to other populations. Functional studies are needed to elucidate the genetic functions of these identified genes and loci in SCC development.

Authors' Contributions

Conception and design: S.J. Siiskonen, M. Zhang, T. Nijsten, J. Han, A.A. Qureshi
Development of methodology: M. Zhang, P. Kraft, T. Nijsten, J. Han, A.A. Qureshi
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): T. Nijsten, J. Han
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S.J. Siiskonen, M. Zhang, W.-Q. Li, L. Liang, P. Kraft, T. Nijsten, J. Han, A.A. Qureshi
Writing, review, and/or revision of the manuscript: S.J. Siiskonen, M. Zhang, W.-Q. Li, P. Kraft, J. Han, A.A. Qureshi
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): M. Zhang, W.-Q. Li, J. Han, A.A. Qureshi
Study supervision: M. Zhang, T. Nijsten, J. Han, A.A. Qureshi

References

1. Gerstenblith MR, Shi J, Landi MT. Genome-wide association studies of pigmentation and skin cancer: a review and meta-analysis. Pigment Cell Melanoma Res 2010;23:587–606.
2. Zhang M, Song F, Liang L, Nan H, Zhang J, Liu H, et al. Genome-wide association studies identify several new loci associated with pigmentation traits and skin cancer risk in European Americans. Hum Mol Genet 2013;22:2948–59.
3. 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;116:224–9.
4. 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;68:884–94.
5. Han J, Kraft P, Colditz GA, Wong J, Hunter DJ. Melanocortin 1 receptor variants and skin cancer risk. Int J Cancer 2006;119:1976–84.
6. 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:e1000074.
7. Nan H, Kraft P, Hunter DJ, Han J. Genetic variants in pigmentation genes, pigmentary phenotypes, and risk of skin cancer in Caucasians. Int J Cancer 2009;125:909–17.
8. Nelson HH, Kelsey KT, Mott LA, Karagas MR. The XRCC1 Arg399Gln polymorphism, sunburn, and non-melanoma skin cancer: evidence of gene-environment interaction. Cancer Res 2002;62:152–5.
9. Kang SY, Lee KG, Lee W, Shim JY, Ji SI, Chung KW, et al. Polymorphisms in the DNA repair gene XRCC1 associated with basal cell carcinoma and squamous cell carcinoma of the skin in a Korean population. Cancer Sci 2007;98:716–20.
10. Welsh MM, Applebaum KM, Spencer SK, Perry AE, Karagas MR, Nelson HH. CTLA4 variants, UV-induced tolerance, and risk of non-melanoma skin cancer. Cancer Res 2009;69:6158–63.
11. Nan H, Xu M, Kraft P, Qureshi AA, Chen C, Guo Q, et al. Genome-wide association study identifies novel alleles associated with risk of cutaneous basal cell carcinoma and squamous cell carcinoma. Hum Mol Genet 2011;20:3718–24.
12. Qi L, Cornelis MC, Kraft P, Stanya KJ, Linda Kao WH, Pankow JS, et al. Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. Hum Mol Genet 2010;19:2706–15.
13. Rimm EB, Giovannucci EL, Willett WC, Colditz GA, Ascherio A, Rosner B, et al. Prospective study of alcohol consumption and risk of coronary disease in men. Lancet 1991;338:464–8.
14. Hofman A, Darwish Murad S, van Duijn CM, Franco OH, Goedegebure A, Ikram MA, et al. The Rotterdam Study: 2014 objectives and design update. Eur J Epidemiol 2013;28:889–926.
15. Casparie M, Tiebosch AT, Burger G, Blauwgeers H, van de Pol A, van Krieken JH, et al. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Cell Oncol 2007;29:19–24.
16. Flohil SC, Koljenovic S, de Haas ER, Overbeek LI, de Vries E, Nijsten T. Cumulative risks and rates of subsequent basal cell carcinomas in the Netherlands. Br J Dermatol 2011;165:874–81.
17. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. Annu Rev Genomics Hum Genet 2009;10:387–406.
18. HapMap. Available from: http://www.hapmap.org.
19. Marchini J, Donnelly P, Cardon LR. Genome-wide strategies for detecting multiple loci that influence complex diseases. Nat Genet 2005;37:413–7.
20. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–9.
21. MATAL. Available from: http://www.sph.umich.edu/csg/abecasis/Metal/index.html.

22. Liang L, Morar N, Dixon AL, Lathrop GM, Abecasis GR, Moffatt MF, et al. A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. Genome Res 2013;23:716–26.

23. eQTLs R-teo. Available from: http://eqtl.rc.fas.harvard.edu/mrce/.

24. Han J, Colditz GA, Hunter DJ. Risk factors for skin cancers: a nested case-control study within the Nurses' Health Study. Int J Epidemiol 2006;35:1514–21.

25. Picco G, Julien S, Brockhausen I, Beatson R, Antonopoulos A, Haslam S, et al. Over-expression of ST3Gal-I promotes mammary tumorigenesis. Glycobiology 2010;20:1241–50.

26. Videira PA, Correia M, Malagolini N, Crespo HJ, Ligeiro D, Calais FM, et al. ST3Gal.I sialyltransferase relevance in bladder cancer tissues and cell lines. BMC Cancer 2009;9:357.

27. Sproviero D, Julien S, Burford B, Taylor-Papadimitriou J, Burchell JM. Cyclooxygenase-2 enzyme induces the expression of the alpha-2,3-sialyltransferase-3 (ST3Gal-I) in breast cancer. J Biol Chem 2012;287:44490–7.

28. Drinnan NB, Halliday J, Ramsdale T. Inhibitors of sialyltransferases: potential roles in tumor growth and metastasis. Mini Rev Med Chem 2003;3:501–17.

29. Ferreira SA, Vasconcelos JL, Silva RC, Cavalcanti CL, Bezerra CL, Rego MJ, et al. Expression patterns of alpha2,3-sialyltransferase I and alpha2,6-sialyltransferase I in human cutaneous epithelial lesions. Eur J Histochem 2013;57:e7.

30. Russello SV, Shore SK. SRC in human carcinogenesis. Front Biosci 2004;9:139–44.

31. Hamilton SR. Molecular genetic alterations as potential prognostic indicators in colorectal carcinoma. Cancer 1992;69:1589–91.

32. Banck MS, Kanwar R, Kulkarni AA, Boora GK, Metge F, Kipp BR, et al. The genomic landscape of small intestine neuroendocrine tumors. J Clin Invest 2013;123:2502–8.

33. Vergani E, Vallacchi V, Frigerio S, Deho P, Mondellini P, Perego P, et al. Identification of MET and SRC activation in melanoma cell lines showing primary resistance to PLX4032. Neoplasia 2011;13:1132–42.

34. Kolch W. Erbin: sorting out ErbB2 receptors or giving Ras a break?Sci STKE 2003;2003:pe37.

35. Lebeau S, Masouye I, Berti M, Augsburger E, Saurat JH, Borradori L, et al. Comparative analysis of the expression of ERBIN and Erb-B2 in normal human skin and cutaneous carcinomas. Br J Dermatol 2005;152:1248–55.

36. Iwakawa R, Okayama H, Kohno T, Sato-Otsubo A, Ogawa S, Yokota J. Contribution of germline mutations to PARK2 gene inactivation in lung adenocarcinoma. Genes Chromosomes Cancer 2012;51:462–72.

37. Zhang M, Chan MH, Tu WJ, He LR, Lee CM, He M. Using the theory of coevolution to predict protein-protein interactions in non-small cell lung cancer. Chin J Cancer 2013;32:91–8.

38. Veeriah S, Taylor BS, Meng S, Fang F, Yilmaz E, Vivanco I, et al. Somatic mutations of the Parkinson's disease-associated gene PARK2 in glioblastoma and other human malignancies. Nat Genet 2010;42:77–82.

39. Mulholland PJ, Fiegler H, Mazzanti C, Gorman P, Sasieni P, Adams J, et al. Genomic profiling identifies discrete deletions associated with translocations in glioblastoma multiforme. Cell Cycle 2006;5:783–91.

40. Letessier A, Garrido-Urbani S, Ginestier C, Fournier G, Esterni B, Monville F, et al. Correlated break at PARK2/FRA6E and loss of AF-6/Afadin protein expression are associated with poor outcome in breast cancer. Oncogene 2007;26:298–307.

Cancer Epidemiology, Biomarkers & Prevention

AACR American Association for Cancer Research

A Genome-Wide Association Study of Cutaneous Squamous Cell Carcinoma among European Descendants

Satu J. Siiskonen, Mingfeng Zhang, Wen-Qing Li, et al.

Updated version	Access the most recent version of this article at: doi:10.1158/1055-9965.EPI-15-1070
Supplementary Material	Access the most recent supplemental material at: http://cebp.aacrjournals.org/content/suppl/2016/02/12/1055-9965.EPI-15-1070.DC1

Cited articles	This article cites 37 articles, 5 of which you can access for free at: http://cebp.aacrjournals.org/content/25/4/714.full#ref-list-1
Citing articles	This article has been cited by 4 HighWire-hosted articles. Access the articles at: http://cebp.aacrjournals.org/content/25/4/714.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cebp.aacrjournals.org/content/25/4/714. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.