Quantifying the Polygenic Contribution to Cutaneous Squamous Cell Carcinoma Risk

Joanne E. Sordillo[1], Peter Kraft[2], Ann Chen Wu[1,4] and Maryam M. Asgari[1,3,4]

Genetic factors play an important role in cutaneous squamous cell carcinoma risk. Genome-wide association studies have identified 21 single nucleotide polymorphisms associated with cutaneous squamous cell carcinoma risk. Yet no studies have attempted to quantify the contribution of heritability to cutaneous squamous cell carcinoma risk by calculating the population attributable risk using a combination of all discovered genetic variants. Using an additive multi-locus linear logistic model, we determined the cumulative association of these 21 genetic regions to cutaneous squamous cell carcinoma population attributable risk. We computed a multi-locus population attributable risk of 62%, suggesting that if the effects of all the risk alleles were removed from a population, the cutaneous squamous cell carcinoma risk would drop by 62%. Using stratified analysis, we also examined the impact of sex on polygenic risk score, and found that men have an increased relative risk throughout the spectrum of the polygenic risk score. Quantifying the impact of genetic predisposition on the proportion of cancer cases can guide future research decisions and public health policy planning.

Journal of Investigative Dermatology (2018) ■, ■–■; doi:10.1016/j.jid.2018.01.031

INTRODUCTION

Cutaneous squamous cell carcinoma (cSCC) is a common form of skin cancer responsible for a substantial public health burden and significant health care costs (Housman et al., 2003). Non-Hispanic whites have a high lifetime prevalence of cSCC, ranging from 7% to 11% (Kallini et al., 2015). Previous studies have shown that cSCC risk clusters in families, suggesting a strong heritable component to this common form of skin cancer (Asgari et al., 2015; Hussain, 2009), but the contribution of genetics to cSCC risk has not been quantified. Recently, genome-wide association studies (GWAS) among non-Hispanic whites have identified variants in 21 genetic loci that are associated with cSCC risk (Asgari et al., 2016; Chahal et al., 2016; Siiskonen et al., 2016). But the degree to which these genetic variants impact the overall burden of cSCC is unclear.

One way to measure the contribution of genetic variants to disease risk is to examine the population attributable risk (PAR), which approximates the reduction in incidence that would be achieved if the risk factor was eliminated from that population (Rothman, 2008). Although risk variants cannot be removed from the population, genetic PARs are often used in epidemiologic research to estimate the degree to which a disease can be attributed to the risk variant. The validity of the PAR for measuring genetic contributions to disease risk has been established (Witte et al., 2014) and has bounds that range from 0% to 100%. We computed a multi-locus PAR for cSCC using 21 published single nucleotide polymorphisms (SNPs) from large-scale GWAS. To examine how PAR can be modified by a patient variable, we explored the effects of sex on PAR by stratifying our population by sex. Finally, we used data from the 21 published variants to calculate a polygenic risk score. In other cancers, such as breast cancer, risk prediction models that combine a polygenic risk score with epidemiologic risk factors provide substantial risk stratification of the general population (Maas et al., 2016; Meads et al., 2012). Quantifying the genetic contribution of these SNPs to cSCC risk, in combination with clinical and environmental risk factors, could help identify individuals at greatest risk of developing cSCC, who would benefit most from enhanced monitoring programs.

RESULTS

Cutaneous SCC risk was associated with 21 SNPs, all of which were used to compute a multi-locus PAR (Kraft et al., 2009; Rockhill et al., 1998). Not all risk alleles are minor alleles. For SNPs associated with a reduction of cSCC risk ("protective" SNP), we took the inverse odds ratio and the major allele frequencies for the association. This way, all variants could be added to the population attributable risk calculation. The multi-locus PAR (Table 1) suggests that if the effects of all risk alleles were removed from a population, there would be a 62% reduction in cSCC cases. SNPs in *IRF4, TYR, RALY, FOXP1*, and *MCIR* used to compute the multi-locus PAR were identified in more than one GWAS, and had similar odds ratios, showing a consistency across study populations. Prevalent SNPs with moderate to high odds ratios contributed the most the multi-locus PAR, and

[1]*Precision Medicine Translational Research Center, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts, USA;* [2]*Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; and* [3]*Department of Dermatology, Massachusetts General Hospital, Boston, Massachusetts, USA*

[4]*These authors contributed equally to this work.*

Correspondence: Maryam M. Asgari, Department of Dermatology, Massachusetts General Hospital, 50 Staniford Street, Suite 270, Boston, Massachusetts 02114, USA. E-mail: pores@mgh.harvard.edu

Abbreviations: cSCC, cutaneous squamous cell carcinoma; GWAS, genome-wide association studies; SNP, single nucleotide polymorphism; PAR, population attributable risk

Received 3 April 2017; revised 9 January 2018; accepted 29 January 2018; accepted manuscript published online 13 February 2018; corrected proof published online XXX

Table 1. Cutaneous squamous cell carcinoma genome-wide association studies variants contribute to a multi-locus population attributable risk of 62%

Gene	Chromosome	SNP	Risk Allele	Risk Allele Frequency	Odds Ratio
Unknown	2	rs192481803	T	0.01	1.9
FOXP1	3	rs62246017	A	0.33	1.06
TPRG1/TP63	3	rs6791479	T	0.43	1.05
ERBB2IP	5	rs17247181	T	0.10	1.40
SLC45A2	5	rs35407[1]	G	0.96	1.69
IRF4	6	rs12203592	T	0.17	1.62
HLADQA1	6	rs4455710	T	0.38	1.17
PARK2	6	rs9689649	C	0.22	1.38
AHR	7	rs117132860	A	0.02	1.48
ST3GAL1	8	rs9643297	G	0.31	1.21
SEC16A	9	rs57994353	C	0.30	1.12
BNC2, CNTLN	9	rs10810657[1]	A	0.59	1.11
TYR	11	rs1126809	A	0.28	1.16
CADM1-BUD13	11	rs74899442	C	0.01	2.13
OCA2	15	rs1800407	T	0.07	1.2
HERC2	15	rs12916300[1]	T	0.74	1.14
DEF8	16	rs4268748	C	0.26	1.33
MC1R	16	rs1805007	T	0.07	1.46
DEF8	16	rs8063761	T	0.33	1.34
RALY-ASIP	20	rs6059655	A	0.07	1.27
SRC	20	rs754626	G	0.25	1.26

Abbreviation: SNP, single nucleotide polymorphism.
[1]These SNPs have been reported as protective SNPs, thus, the major risk allele and its corresponding odds ratio (the reciprocal of the protective odds ratio) are shown.

included SNPs in the genes *SLC45A2, SRC, HERC2, DEF8,* and *HLADQA1*.

Figure 1 shows the relative risk of cSCC increasing with higher percentiles of the polygenic risk score, based on 21 risk alleles. Sex-specific curves for males and females are shown, which reveal that males have higher risk as compared to females, particularly at polygenic risk scores that exceed the 40th percentile. This suggests a potential interaction between sex and genetic score. The risk of cSCC is twice the average population risk at the 84th percentile of the polygenic risk score for females and at the 77th percentile for males. A two-fold increased relative risk is a threshold widely used as a benchmark for clinically meaningful increased risk for common diseases, such as cSCC (Roberts et al., 2012).

DISCUSSION

Different measures can be used to assess how much known genetic factors contribute to overall cSCC risk including heritability on various scales, sibling relative risk, log relative risk genetic variance, the area under the receiver operating curve, and the PAR (Witte et al., 2014). The utility and limitations of these various metrics have been discussed extensively elsewhere (Witte et al., 2014). Here, we focused on the PAR and log relative risk variance (which is used to calculate the population distribution of the polygenic risk score) because of their direct relevance to the clinical utility of these genetic variants for cSCC prevention and screening.

Most polymorphisms that contribute to the multi-locus PAR are within biologically plausible gene candidates for cSCC etiology. *IRF4* (Han et al., 2008) is associated with pigmentation phenotypes. Variants in *DEF8* are associated with expression of a cell cycle progression gene *(CDK10)* in sun-exposed skin (GTEx Consortium, 2015). Expression of *CADM1*, a gene that modifies tumor interaction with cell-mediated immunity, is associated with survival in cSCC patients (Liu et al., 2013). Activation of *AHR* during UV radiation exposure may decrease apoptosis in keratinocytes (Frauenstein et al., 2013) with potential consequences for enhancing cancer risk. *ERBB2IP* has been shown to affect the Ras pathway (Kolch, 2003). Risk alleles in *SLC45A2*, a gene associated with skin pigmentation, were also linked to cSCC risk.

While the multi-locus PAR allows quantification of the joint impact of these variants on cSCC risk, this one statistic is of limited utility for screening purposes, as it does not consider known environmental risk factors, such as sun exposure and smoking. However, calculation of a polygenic risk score, using the combination of these risk alleles, helps us to visualize their relationship with cSCC risk, and may be a useful metric, when combined with relevant clinical variables, for screening. Although we have demonstrated the potential utility of the polygenic risk score here, this type of screening is not yet at a stage where it could be incorporated into clinical practice. The next phase of risk prediction models should incorporate both clinical and environmental risk factors, combined with a polygenic risk score to help optimize primary and secondary prevention strategies for cSCCs. Ideally, the model would be developed and calibrated in one population, and validated in an independent population. Identification of individuals with substantially elevated risk is of paramount importance for early detection and prevention efforts.