



Published in final edited form as:

*Eur J Cancer*. 2020 September ; 136: 84–94. doi:10.1016/j.ejca.2020.04.049.

## Genetic variants in *TKT* and *DERA* in the NADPH pathway predict melanoma survival

Ning Gu<sup>1,2,3,4,†</sup>, Wei Dai<sup>3,4,5,†</sup>, Hongliang Liu<sup>3,4</sup>, Jie Ge<sup>3,4</sup>, Sheng Luo<sup>6</sup>, Eunyoung Cho<sup>7,8</sup>, Christopher I. Amos<sup>9</sup>, Jeffrey E. Lee<sup>10</sup>, Xin Li<sup>11</sup>, Hongmei Nan<sup>11</sup>, Hua Yuan<sup>1,2,12,\*</sup>, Qingyi Wei<sup>3,4,13,\*</sup>

<sup>1</sup>Jiangsu Key Laboratory of Oral Diseases, Nanjing Medical University, Nanjing, Jiangsu, 210029, China

<sup>2</sup>Department of Oral and Maxillofacial Surgery, The Affiliated Hospital of Stomatology, Nanjing Medical University, Nanjing, Jiangsu, 210029, China

<sup>3</sup>Duke Cancer Institute, Duke University Medical Center, Durham, NC, 27710, USA

<sup>4</sup>Department of Population Health Sciences, Duke University School of Medicine, Durham, NC, 27710, USA

<sup>5</sup>Department of Dermatology, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong, 510515, China

<sup>6</sup>Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC, 27710, USA

<sup>7</sup>Department of Dermatology, Warren Alpert Medical School, Brown University, Providence, RI, 02912, USA

<sup>8</sup>Department of Epidemiology, Brown University School of Public Health, Providence, RI, 02912, USA

<sup>9</sup>Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, TX, 77030, USA

<sup>10</sup>Department of Surgical Oncology, The University of Texas M. D. Anderson Cancer Center, Houston, TX, 77030, USA

<sup>11</sup>Department of Epidemiology, Fairbanks School of Public Health, Indiana University, Indianapolis, IN, 46202, USA

\*Correspondence authors: Hua Yuan, Jiangsu Key Laboratory of Oral Diseases, Nanjing Medical University, Nanjing, Jiangsu, 210029, China, Tel: (86) 2585031867; Fax: (86) 2586516414; yuanhua@njmu.edu.cn; or Qingyi Wei, Duke Cancer Institute, Duke University Medical Center and Department of Medicine, Duke School of Medicine, 905 S LaSalle Street, Durham, NC, 27710, USA, Tel.: (919) 660-0562, Fax: (919) 681-7386; qingyi.wei@duke.edu.

†These authors contributed equally to this work.

Conflict of interest statement

The authors declare no conflict of interest.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

<sup>12</sup>Collaborative Innovation Center For Cancer Personalized Medicine, Nanjing Medical University, Nanjing, Jiangsu, 210029, China

<sup>13</sup>Department of Medicine, Duke University School of Medicine, Durham, NC, 27710, USA

## Abstract

**Background:** Cutaneous melanoma (CM) is the most lethal type of skin cancers. Nicotinamide adenine dinucleotide phosphate (NADPH) plays an important role in anabolic reactions and tumorigenesis, but many genes are involved in the NADPH system.

**Methods:** We used 10,912 SNPs (2,018 genotyped and 8,894 imputed) in 134 NADPH-related genes from a genome-wide association study (GWAS) of 858 patients from The University of Texas MD Anderson Cancer Center (MDACC) in a single-locus analysis to predict CM survival. We then replicated the results in another GWAS dataset of 409 patients from the Nurses' Health Study (NHS) and Health Professionals Follow-up Study (HPFS).

**Results:** There were 95/858 (11.1%) and 48/409 (11.7%) patients who died of CM, respectively. In multivariable Cox regression analyses, we identified two independent SNPs (*TKT*rs9864057 G>A and *DERA* rs12297652 A>G) to be significantly associated with CM-specific survival [hazards ratio (HR) of 1.52, 95% confidence interval (CI)=1.18-1.96,  $P=1.06\times 10^{-3}$  and 1.51 (1.19-1.91,  $5.89\times 10^{-4}$ )] in the meta-analysis, respectively. Furthermore, an increasing number of risk genotypes of these two SNPs was associated with a higher risk of death in the MDACC, the NHS/HPFS, and their combined datasets ( $P_{\text{trend}}<0.001$ , =0.004, and  $<0.001$ , respectively). In the expression quantitative trait loci (eQTL) analysis, *TKT*rs9864057 G>A and *DERA* rs12297652 A>G were also significantly associated with higher mRNA expression levels in sun-exposed lower-leg skin ( $P=0.043$  and 0.006, respectively).

**Conclusions:** These results suggest that these two potentially functional SNPs may be valuable prognostic biomarkers for CM survival, but larger studies are needed to validate these findings.

## Keywords

cutaneous melanoma; nicotinamide adenine dinucleotide phosphate; single-nucleotide polymorphism; genome-wide association study; melanoma-specific survival

## 1. Introduction

Cutaneous melanoma (CM) is the most lethal type of skin cancers. Although CM is much less common than basal or squamous cell carcinomas, it accounts for the majority of skin cancer deaths, because of its tendency to invade and spread during the course of the disease [1,2]. It has been estimated that 96,480 new CM cases, about 5.5% of all of new cancer cases, will be diagnosed in the United States in 2019, and 7,230 patients will die of this disease [3]. Between 2008 and 2014 in the US, the overall five-years survival rate after diagnosis was 91.8% [4], while patients with a late or distant-stage CM generally have a much poorer prognosis with a five-year survival rate of only approximately 18% [2]; survival is particularly poor for those patients with visceral metastasis (a median survival of 6 months) [5]. To better identify high-risk subgroups with a poor survival, it is urgent to identify additional factors, such as genetic variants, that are involved in CM prognosis. Such

investigations are likely to help identify clinically relevant mechanisms, allowing for more accurate selection of CM patients for the most effective management and treatment.

Nicotinamide adenine dinucleotide phosphate (NADPH) plays a key role in anabolic reactions in all cellular organisms as a reducing agent [6]. The NADPH system generates free radicals in biological immune cells to destroy pathogens through a process called the respiratory burst [7]. Through the NADPH system, cells produce reactive oxygen species (ROS) that cause cellular and tissue damage by oxidizing both DNA and membranes, thus contributing to cellular dysfunction, aging, neurodegeneration, cell death and cancer [8]. The NADPH system also promotes migration of inflammatory cells and modulates radiation-induced senescent cells by producing ROS [9]. In cancer cells, defects in the mitochondrial oxidative metabolism provide reduced equivalents through NADPH for metabolizing hydroperoxides to increase production of superoxide, hydrogen peroxide, and hydroperoxide [10]. The NADPH oxidases are an important site of ROS generation that is regulated in cancer cells and T cells in the context of antitumor immunity [11]. Several publications have highlighted the impact of NADPH on CM. For example, melanoma cells are characterized by altered redox signaling, especially in the form of higher levels of ROS, than those required for normal cell signals [12]. Two isoforms of NADPH (NOX1 and NOX4) are overexpressed, producing high ROS levels in melanoma cells, and the abundant ROS then stimulates the proliferation of tumor cells [13-15]. Some studies have shown that ultraviolet-induced chemical excitation of melanin fragments leads to DNA damage by initiating the ROS-dominated NADPH [16-18].

Many genetic variants have been found to be associated with cancer risk or survival patients with CM [1, 19]. However, the biological relevance of most identified genetic variants, such as single nucleotide polymorphisms (SNPs), remains unknown. Understanding of genetic factors that contribute to the NADPH pathway offers opportunities for prediction and identification of high-risk CM patients for personalized management and treatment.

In the present study, therefore, we systematically examined associations between genetic variants in NADPH-related genes and CM survival, by using publicly available genotyping datasets and also evaluated correlations between significant SNPs and their gene expression levels to identify biological mechanisms that may underlie the observed associations.

## 2. Patients and methods

### 2.1. Study populations

In the present study, we used two publically available genome-wide association study (GWAS) genotyping datasets: a discovery dataset from The University of Texas MD Anderson Cancer Center (MDACC) and a replication dataset by from the Nurses' Health Study (NHS) and the Health Professionals Follow-up Study (HPFS). There were 95/858 (11.1%) and 48/409 (11.7%) patients who died of CM, respectively. All patients from the discovery dataset gave written informed consent, and the protocol was approved by the MDACC Institutional Review Board. For the replication dataset, a written informed consent was also obtained from each subject, and the study protocol was approved by the

institutional review boards of Brigham and Women's Hospital and the Harvard T.H. Chan School of Public Health, and those of participating registries as required.

## 2.2. Discovery Dataset

The discovery dataset included 858 non-Hispanic white patients with CM who were recruited between March 1993 and August 2008 to be participants in a hospital-based case-control study [20]. All CM patients were classified according to the American Joint Committee on Cancer (AJCC) staging system [21] and followed using standardized methods and guidelines [22]. Demographic and clinical variables such as age, sex, Breslow thickness, tumor stage, ulceration, and mitotic rate were available in the datasets obtained from the dbGaP database (accession: phs000187.v1.p1) [23] (Supplementary Table 1). The details of genotyping information and data quality control have previously been reported [20]. Using the MACH software program and the 1000 Genomes Project CEU population (March 2010 release) as the reference, data from the MDACC study have been imputed (imputation quality  $r^2 = 0.8$ ) [24].

## 2.3. Replication Dataset

The replication dataset from the NHS/HPFS was produced by merging two subdatasets from two other studies: one having 317 female cases from the NHS and the other having 177 male cases from the HPFS [25]. There were 409 non-Hispanic white patients in the NHS/HPFS GWAS dataset. Participants were enrolled in the NHS in 1976 and the HPFS in 1986, and were diagnosed after the enrollment up to the 2008 follow-up cycle for both cohorts. Clinical information on age, sex, survival outcome, and genotype data were available. Genotype data were generated using the Illumina HumanHap610 array and imputed using the MACH software program by using the 1000 Genomes Project CEU population (Northern Europeans from Utah) database (phase I v3, March 2012, with imputation quality  $r^2 = 0.8$ ) as reference panel [26, 27]. The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required.

## 2.4. Gene and SNP extraction

Because females carry two copies of the X chromosome but males are heterozygous, and there is no standard statistical data based on sex-specific analysis, five genes in the X chromosome (but no pseudogenes and none on Y chromosome) were excluded.

Pseudogenes were also excluded because they have no biological function. A total of 134 NADPH pathway genes located on autosomes were extracted from the online Molecular Signatures Database, which includes gene sets extracted from original research publications as well as and the entire collections from the online resources, such as **GO** and **KEGG** (<http://software.broadinstitute.org/gsea/msigdb/index.jsp>) (Supplementary Table 1). All selected genes were expanded with  $\pm 2$ -kb flanking regions and mapped to all the SNPs identified in the MDACC GWAS dataset following standard quality-control criteria, including minor allele frequency (MAF)  $\geq 0.05$ , genotyping rate  $\geq 95\%$ , and Hardy-Weinberg equilibrium (HWE)  $P$  value  $\geq 1 \times 10^{-5}$ . Consequently, 10,912 [with 2,018 (18.5%)

genotyped and 8,894 (81.5%) imputed] common SNPs in the NADPH pathway genes were extracted from the MDACC GWAS dataset. A detailed flowchart is shown in Figure 1.

## 2.5. Statistical analysis

CM-specific survival (CMSS) was the outcome of interest in the present study, calculated based on the date of CM diagnosis to the last follow-up or CM-related death date. We used multivariate Cox proportional hazards regression models and the GenABEL package of R software to compute the hazards ratio (HR) and 95% confidence interval (CI) for associations between SNPs and CMSS in an additive genetic model. Multivariate analysis of the MDACC dataset was adjusted by age, sex, Breslow thickness, tumor stage, ulceration, and mitotic rate. Multivariate analysis of the NHS/HPFS dataset as well as the combination of the two datasets were adjusted only for age and sex, because clinical factors were not available in the NHS/HPFS dataset.

Because most of the SNPs in the present study were imputed based on linkage disequilibrium (LD) among them, false-positive report probability (FPRP) has been suggested as a better probability threshold for these types of datasets for multiple testing correction than the false discovery rate (FDR) [28]. In the FPRP calculation, a prior probability of 0.10 was selected to detect an HR of 2.0 for an association with variant genotypes or minor alleles of each SNP. SNPs with  $FPRP < 0.2$  were selected for replication in the NHS/HPFS dataset. We also used a meta-analysis to combine the results from the discovery and replication sets. When there was no heterogeneity between the MDACC and the NHS/HPFS datasets (Cochran's Q test  $P$ -value  $> 0.100$  and the heterogeneity statistic ( $I^2$ )  $< 50\%$ ), we used a fixed-effects model; otherwise, a random-effects model was used. To identify independent predictive SNPs for CMSS, the validated SNPs together with clinical prognostic variables were included in a multivariate stepwise Cox model only using the MDACC dataset that had more covariates available for further adjustment. We used Kaplan-Meier survival curves and log-rank tests to evaluate the effect of selected SNPs on the cumulative probability of CMSS. We also used the Chi-square-based Q-test with  $P < 0.05$  to evaluate effect difference in the stratified analyses. We used the receiver operating characteristic (ROC) curve, which relies on the value of area under the curve (AUC), to illustrate the CMSS prediction with the sensitivity and specificity. Time-dependent AUC and ROC analyses were performed using the two R packages of survival and timeROC [29]. The expression quantitative trait loci (eQTL) analysis for associations with the alleles and genotypes of the significant SNPs was assessed by linear regression analysis using the R software. All analyses were performed with SAS (version 9.3.3; SAS Institute, Cary, NC, USA) unless otherwise specified. All reported  $P$  values were two-sided.

## 3. Results

### 3.1. Characteristics of study populations

The analyses included 858 patients from the MDACC dataset and 409 patients from the NHS/HPFS dataset. The age range at diagnosis was between 17 and 94 years with a mean age of 52.4 years ( $52.4 \pm 14.4$ ) in the MDACC dataset, compared with 34-87 years and 61.1 years ( $61.1 \pm 10.8$  years) in the NHS/HPFS dataset. The male/female ratio was 57.8% (496/

42.2% (362) in the MDACC dataset, compared with 33.7% (138)/66.3% (271) in the NHS/HPFS dataset. Univariate analysis revealed that age, sex, tumor stage, Breslow thickness, ulceration, and mitotic rate were all significantly associated with CMSS in the MDACC dataset, while age but not sex was statistically significantly associated with CMSS in the NHS/HPFS dataset. The median follow-up time for patients was 81.1 months in the MDACC dataset and 179.0 months in the NHS/HPFS dataset. Death rates in the MDACC dataset (95/858, 11.1%) and the NHS/HPFS dataset (48/409, 11.7%) were similar (Supplementary Table 2).

### 3.2. Gene and SNP extraction

After the exclusion of five genes in the X chromosome, we included 134 NADPH-related genes from the online database (MSigDB) (Supplementary Table 1). Initially, we extracted 10,912 common SNPs from the MDACC GWAS dataset (2,018 genotyped SNPs and 8,894 imputed SNPs). The associations between all these SNPs and CMSS were presented in a Manhattan plot (Supplementary Figure 1). Using  $P < 0.05$  as a threshold, we identified 834 SNPs as significantly associated with CMSS in the discovery phase. After further screening for false-positive findings with a FPRP  $< 0.2$ , 341 SNPs were selected for further replication (Figure 1).

In the replication analyses, we continued to use the Cox regression analysis (including only the two available variables of age and sex) to verify the 341 SNPs in the NHS/HPFS dataset. Five SNPs in two genes [rs9864057, rs17234092, rs17306163, rs62255994 in *TKT* (transketolase) and rs12297652 in *DERA* (*deoxyribose-phosphate aldolase*)] remained statistically significant ( $P < 0.05$ ) for CMSS (Figure 1). Subsequently, LD analysis showed a high LD (i.e., the same LD plots) among the four SNPs in *TKT* gene. After meta-analysis of these five SNPs, the same associations remained statistically significant, and there was no statistically significant heterogeneity in the results for these five SNPs between the two datasets (Table 1). However, in the univariate analysis as shown in Supplementary Table 2, sex was a risk factor for the MDACC datasets but not for the NHS/HPFS dataset, but in the multivariable modeling, sex was no longer a risk factor for the MDACC dataset (Table 2).

### 3.3. Genetic variants in NADPH-related genes as independent survival predictors

The four SNPs in the same LD plot of *TKT* were subsequently annotated and filtered using SNPinfo (<http://snpinfo.niehs.nih.gov/snpinfo/snpfunchtml>), RegulomeDB (<http://www.regulomedb.org/>) and F-SNP (<http://compbio.cs.queensu.ca/F-SNP>) for *in silico* functional prediction. As shown in Supplementary Table 3, out of the four LD SNPs on *TKT*, *TKT* rs9864057 is located at the potential enhancer regions of 16 tissues and the DNase I sensitive site, while *DERA* rs12297652 may modify the binding activity of LXR. Therefore, we selected these two SNPs as the representative SNPs for further analysis. In stepwise Cox regression analysis, including the two tagging SNPs, with adjustment for other clinical covariates in the MDACC dataset, these two SNPs (*TKT* rs9864057 G>A and *DERA* rs12297652 A>G) remained significant and independent predictors of CMSS (Table 2). All SNPs, both genotyped and imputed, in *TKT* and *DERA* are shown in regional association plots containing 200-kb up and downstream of rs9864057 and rs12297652, respectively (Supplementary Figure 2).

### 3.4. Two independent SNPs as CM survival predictors in the MDACC dataset, the NHS/HPFS dataset, and the combination of both datasets

We further examined associations between the identified independent SNPs and CM survival in the MDACC dataset, the NHS/HPFS dataset, and their combination in multivariate analysis. The risk alleles of these two independent SNPs were rs9864057 A in *TKT* and rs12297652 G in *DERA* (Table 3).

Under an additive genetic model, the *TKT* rs9864057 A allele and the *DERA* rs12297652 G allele were significantly associated with CMSS in the MDACC dataset (trend test:  $P=0.010$  and  $0.005$ , respectively). Under a dominant genetic model, the *TKT* rs9864057 A and *DERA* rs12297652 G genotypes were both associated with elevated risk of CMSS (HR=1.82, 95% CI =1.20-2.77 and  $P=0.005$  for rs9864057 GA+AA vs. GG and HR=2.09, 95% CI =1.28-3.41 and  $P=0.003$  for rs12297652 AG+GG vs. AA). Similar results were obtained in the NHS/HPFS dataset (trend test:  $P=0.05$  and  $P=0.05$ , respectively, and HR=1.76, 95% CI =0.99-3.11,  $P=0.053$  and HR=2.14, 95% CI =1.07-4.30,  $P=0.03$ , respectively). Furthermore, when these two datasets were combined, the associations persisted (trend test: rs9864057,  $P=0.001$  and rs12297652,  $P=0.001$ ; rs9864057 GA+AA vs. GG: HR=1.66, 95% CI = 1.19-2.30,  $P=0.003$ ; rs12297652 AG+GG vs. AA: HR=1.93, 95% CI =1.32-2.84,  $P=0.001$ ) (Table 3).

### 3.5. Analysis of combined genotypes of the two independent SNPs

To assess the combined effect of the two SNPs, we combined the risk genotypes of rs9864057 GA+AA and rs12297652 AG+GG into one variable as the number of risk genotypes. The trend test consistently showed that the risk increased with the number of risk genotypes in the MDACC dataset ( $P<0.001$ ), the NHS/HPFS dataset ( $P=0.004$ ), and the combined dataset ( $P<0.001$ ). Furthermore, when we dichotomized all patients into 0-1 risk genotypes and 2 risk genotypes, patients with 2-risk genotypes had a higher risk of death than those with 0-1 risk genotypes in the MDACC dataset (HR=2.12, 95% CI=1.38-3.26 and  $P<0.001$ ), the NHS/HPFS dataset (HR=2.38, 95% CI=1.35-4.20 and  $P=0.003$ ), and the combined dataset (HR=2.02, 95% CI=1.45-2.83,  $P<0.001$ ). These results were further evaluated by Kaplan-Meier plot to visualize the associations between the number of risk genotypes and CMSS (Figure 2a-2f).

### 3.6. Stratified analyses for the combined efficacy of risk genotypes on CMSS

We then performed stratified analyses by age and sex for the associations between CMSS and the number of risk genotypes in the MDACC dataset and the NHS/HPFS dataset using multivariate Cox regression analysis with adjustment where appropriate, looking for possible interaction between these covariates and SNPs. The risk of CM death associated with the 2-risk genotype was statistically higher for males  $\geq 50$  years in the MDACC dataset and for females  $> 50$  years in the NHS/HPFS dataset. We found no evidence for any interactions between the subgroups (Supplementary Table 4).

### 3.7. ROC curve and time-dependent AUC to estimate CMSS survival

We also used inverse probability of censoring weighting estimators for the ROC curve and time-dependent AUC to estimate the improvement in prediction by adding risk genotypes to

the model. In the combined dataset of both MDACC and NHS/HPFS, we found five-year risk of CMSS with clinical variables (age and sex) had an AUC=63.58%, while the addition of risk genotypes to the model significantly increased the AUC to 69.72% ( $P=0.005$ ) (Figure 3a-3b). However, such prediction was not statistically significant for both 5-year and 10-year survival prediction in the MDACC dataset with adjustment for age, sex, Breslow thickness, distant/regional metastasis and ulceration (Supplementary Figure 3a-3b) and in the NHS/HPFS dataset with adjustment for age and sex (Supplementary Figure 3c-3d). These results suggest that the sample sizes of the datasets are not large enough or more SNPs in different pathway genes need to be identified.

### 3.8. Expression quantitative trait loci (eQTL) analysis

To study the correlation between the two independent SNPs and mRNA expression levels of their corresponding genes, we used data on mRNA expression levels of these two SNPs for further statistical analysis using three *in silico* eQTL databases: 373 European descendants from the 1000 Genomes Project, The Cancer Genome Atlas (TCGA) database, and the Genotype-Tissue Expression (GTEx) project.

Through the GTEx project, we found that both *TKT* rs9864057 and *DERA* rs12297652 were significantly correlated with elevated mRNA expression levels in the skin of sun-exposed lower leg ( $P=0.043$  and  $0.006$ , respectively), but not significantly associated with mRNA expression levels in the whole blood or in non-sun-exposed skin (Supplementary Table 5). The lymphoblastoid cell-line data from the 373 European descendants and the TCGA database did not demonstrate any correlation between these two SNPs and mRNA expression levels (data not shown).

## 4. Discussion

NADPH is a key factor in controlling redox reactions in human tissue, including tumor tissue, and NADPH is closely related to cancer cell growth, metabolism, migration, invasion, and metastasis [10]. However, few studies have investigated the roles of genetic variants in the NADPH-related genes in predicting outcomes of cancers, including melanoma. In the present study, we found that two-independent SNPs (*TKT* rs9864057 G>A and *DERA* rs12297652 A>G) were significantly associated with CMSS. Additional analyses suggested that their variant alleles were correlated with elevated mRNA expression levels in sun-exposed skin, a possible biological mechanism underlying the associations between the variants of NADPH-related genes and CM progression that led to poor survival.

*TKT* is located on chromosome 3p21.1, and this gene encodes a thiamine-dependent enzyme that plays a role in the pentose phosphate pathway. In mammals, transketolase joins the pentose phosphate pathway and glycolysis, and then transfers sugar phosphates into the main carbohydrate metabolism, an important factor in the production of NADPH for biosynthesis [30]. Because alteration in cellular metabolism is one of the hallmarks of cancers [31], several studies have reported the clinical significance of *TKT* in energy regulatory mechanism, especially in the survival of patients with cancers. For example, it was reported that *TKT* expression was correlated with tumor size in breast cancer; in addition, *TKT* expression was higher in lymph node metastases than in primary tumor or



normal tissues of patients in whom high *TKT* levels were associated with poor survival after breast cancer [32]. *TKT* has also been associated with metastasis of ovarian and esophageal cancers, predicting poor survival [33, 34], but there are no such reports about melanoma survival. Other studies have reported significant associations between SNPs of the *TKT* gene and the morbidity/mortalities of some non-cancer diseases, particularly diabetic nephropathy [35-37]. Because melanoma is a cancer derived from neuroectoderm [38], a link between *TKT* genetic variants and nerve function has been reported [36]. Taken together, the findings in the present study support a significant role for *TKT* genetic variants in predicting CMSS.

*DERA* is located on chromosome 12p12.3 and encodes the human deoxyribose phosphate aldolase, which is involved in the glycosaminoglycan metabolism and the innate immune system. Deoxyribose phosphate aldolase activity levels have been correlated with *DERA* expression in most cell lines tested, and cells with high *DERA* activity can use deoxy nucleotide as a source of energy [39]. Through *DERA*, thymidine-derived 2-deoxy-D-ribose 5-phosphate enters the glycolytic pathway, affecting cancer cell growth, invasion, and metastasis [40]. Thymidine, via the glycolytic pathway, relies on *DERA in vitro* and *in vivo* to convert it into multiple thymidine-derived intermediate metabolites required for the survival of cells under low-glucose conditions; therefore, the number of viable cancer cells is thereby reduced, while *DERA* is also reduced in a microenvironment in which glucose was insufficient [41]. Inhibition of thymidine phosphorylase activity suppresses tumor growth by increasing the proportion of apoptotic cells and probably inhibiting angiogenesis in human epidermoid carcinoma cells [42]. These findings suggest that *DERA* plays a central role in the survival and growth of cancer cells, particularly under starvation conditions. Furthermore, enhanced *DERA*-dependent thymidine catabolism has been observed in human gastric cancer [40, 43]. *DERA* has not been previously reported to be associated with CM progression and prognosis.

While we found consistent evidence that two novel SNPs (*TKT* rs9864057 G>A and *DERA* rs12297652 A>G) in two NADPH-related gene predicted CMSS, the present study has some limitations. First, the available clinical variables in the two datasets were different, and the fact there were only two variables (age and sex) available in the NHS/HPFS dataset limited adjustment for other potentially relevant clinical covariates. The second limitation is that the predictive model was built on a non-Hispanic white population in the United States, which may limit generalization of our findings to other populations. Third, SNPs in other relevant genes may have been omitted due to limited knowledge regarding identification of NADPH-related genes. Finally, while differential expression is suggested by the *in silico* analyses, the exact molecular mechanisms of these two SNPs underlying the observed associations remain to be determined. Additional functional studies are needed to explore these newly identified SNPs in order to confirm their potential utility as biomarkers for CM prognosis.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We wish to thank all of the investigators and funding agencies that enabled the deposition of data in dbGaP that we used in this study. We would like to thank the Channing Division of Network Medicine at Brigham and Women's Hospital for their support. We also thank the John Hopkins University Center for Inherited Disease Research for conducting high-throughput genotyping for this study. We would like to thank the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data. We also thank all the individuals who participated in this project. The results published here are in whole or part based upon data generated by The Cancer Genome Atlas pilot project established by the NCI and NHGRI. Information about TCGA and the investigators and institutions that constitute the TCGA research network can be found at "<http://cancergenome.nih.gov>". The authors assume full responsibility for analyses and interpretation of these data.

### Funding

The MD Anderson Study was supported by NIH/NCI R01 CA100264, 2P50CA093459 and R01CA133996 as well as by The University of Texas MD Anderson Cancer Center Various Donors Melanoma and Skin Cancers Priority Program Fund; the Miriam and Jim Mulva Research Fund; the McCarthy Skin Cancer Research Fund and the Marit Peterson Fund for Melanoma Research. The NHS and the HPFS were in part supported by NIH/NCI R01 CA49449, P01 CA87969, UM1 CA186107 and UM1 CA167552. Qingyi Wei was supported by start-up funds from Duke Cancer Institute, Duke University Medical Center and in part supported by the Duke Cancer Institute as part of the P30 Cancer Center Support Grant (Grant ID: NIH CA014236).

## Abbreviations:

<b>CM</b>	cutaneous melanoma
<b>SNP</b>	single-nucleotide polymorphisms
<b>GWAS</b>	genome-wide association studies
<b>MDACC</b>	The University of Texas MD Anderson Cancer Center
<b>NHS</b>	the Nurses' Health Study
<b>HPFS</b>	the Health Professionals Follow-up Study
<b>CMSS</b>	cutaneous melanoma-specific survival
<b>HR</b>	hazards ratio
<b>CI</b>	confidence interval
<b>eQTL</b>	expression quantitative trait loci
<b>FPRP</b>	false-positive report probability
<b>LD</b>	linkage disequilibrium
<b>NADPH</b>	nicotinamide adenine dinucleotide phosphate
<b>ROC</b>	receiver operating characteristic
<b>AUC</b>	area under the curve
<b>TKT</b>	transketolase
<b>DERA</b>	deoxyribose-phosphate aldolase

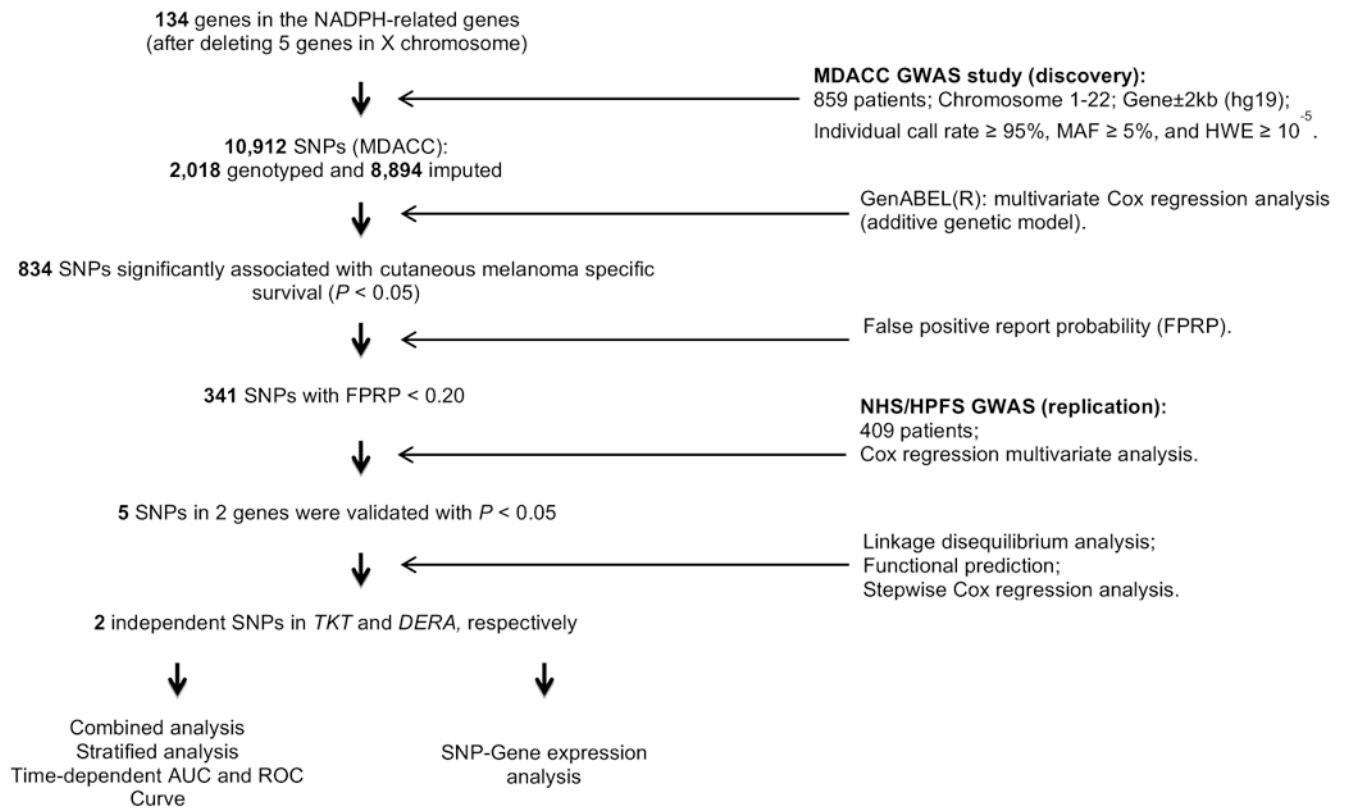
**ROS** reactive oxygen species**References**

1. Law MH, et al, Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. *Nat Genet*, 2015 47(9): p. 987–995. [PubMed: 26237428]
2. Guy GP Jr., et al., Vital signs: melanoma incidence and mortality trends and projections - United States, 1982–2030. *MMWR Morb Mortal Wkly Rep*, 2015 64(21): p. 591–6. [PubMed: 26042651]
3. Siegel RL, Miller KD, and Jemal A, Cancer statistics, 2019. *2019 69(1)*: p. 7–34.
4. Siegel RL, Miller KD, and Jemal A, Cancer statistics, 2018. *CA Cancer J Clin*, 2018 68(1): p. 7–30. [PubMed: 29313949]
5. Demierre MF, Epidemiology and prevention of cutaneous melanoma. *Curr Treat Options Oncol*, 2006 7(3): p. 181–6. [PubMed: 16615874]
6. Hanukoglu I and Rapoport R, Routes and regulation of NADPH production in steroidogenic mitochondria. *Endocr Res*, 1995 21(1-2): p. 231–41. [PubMed: 7588385]
7. Ogawa K, et al., The association of elevated reactive oxygen species levels from neutrophils with low-grade inflammation in the elderly. *Immun Ageing*, 2008 5: p. 13. [PubMed: 18950479]
8. Hervera A, et al., Reactive oxygen species regulate axonal regeneration through the release of exosomal NADPH oxidase 2 complexes into injured axons. *Nature Cell Biology*, 2018 20(3): p. 307–319. [PubMed: 29434374]
9. Sakai Y, et al., NADPH oxidase 4 mediates ROS production in radiation-induced senescent cells and promotes migration of inflammatory cells. *Free Radic Res*, 2018 52(1): p. 92–102. [PubMed: 29228832]
10. Hegedus C, et al., Redox control of cancer cell destruction. *Redox Biol*, 2018 16: p. 59–74. [PubMed: 29477046]
11. Banerjee R, Introduction to the Thematic Minireview Series: Redox metabolism and signaling. *J Biol Chem*, 2018 293(20): p. 7488–7489. [PubMed: 29622674]
12. Meierjohann S, Oxidative stress in melanocyte senescence and melanoma transformation. *Eur J Cell Biol*, 2014 93(1-2): p. 36–41. [PubMed: 24342719]
13. Rouaud F, et al., Mechanism of melanoma cells selective apoptosis induced by a photoactive NADPH analogue. *Oncotarget*, 2016 7(50): p. 82804–82819. [PubMed: 27756874]
14. Liu F, Gomez Garcia AM, and Meyskens FL Jr., NADPH oxidase 1 overexpression enhances invasion via matrix metalloproteinase-2 and epithelial-mesenchymal transition in melanoma cells. *J Invest Dermatol*, 2012 132(8): p. 2033–41. [PubMed: 22513785]
15. Yamaura M, et al., NADPH oxidase 4 contributes to transformation phenotype of melanoma cells by regulating G2-M cell cycle progression. *Cancer Res*, 2009 69(6): p. 2647–54. [PubMed: 19276355]
16. Yuan TA, et al., A Case-Control Study of the Genetic Variability in Reactive Oxygen Species-Metabolizing Enzymes in Melanoma Risk. *Int J Mol Sci*, 2018 19(1).
17. Raad H, et al., NADPH Oxidase-1 Plays a Key Role in Keratinocyte Responses to UV Radiation and UVB-Induced Skin Carcinogenesis. *J Invest Dermatol*, 2017 137(6): p. 1311–1321. [PubMed: 28132856]
18. Premi S, et al., Photochemistry. Chemiexcitation of melanin derivatives induces DNA photoproducts long after UV exposure. *Science*, 2015 347(6224): p. 842–7. [PubMed: 25700512]
19. Manolio TA, Genomewide association studies and assessment of the risk of disease. *N Engl J Med*, 2010 363(2): p. 166–76. [PubMed: 20647212]
20. Amos CI, et al., Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. *Hum Mol Genet*, 2011 20(24): p. 5012–23. [PubMed: 21926416]
21. Balch CM, et al., Final version of 2009 AJCC melanoma staging and classification. *J Clin Oncol*, 2009 27(36): p. 6199–206. [PubMed: 19917835]
22. Gershenwald JE and Ross MI, Sentinel-lymph-node biopsy for cutaneous melanoma. *N Engl J Med*, 2011 364(18): p. 1738–45. [PubMed: 21542744]

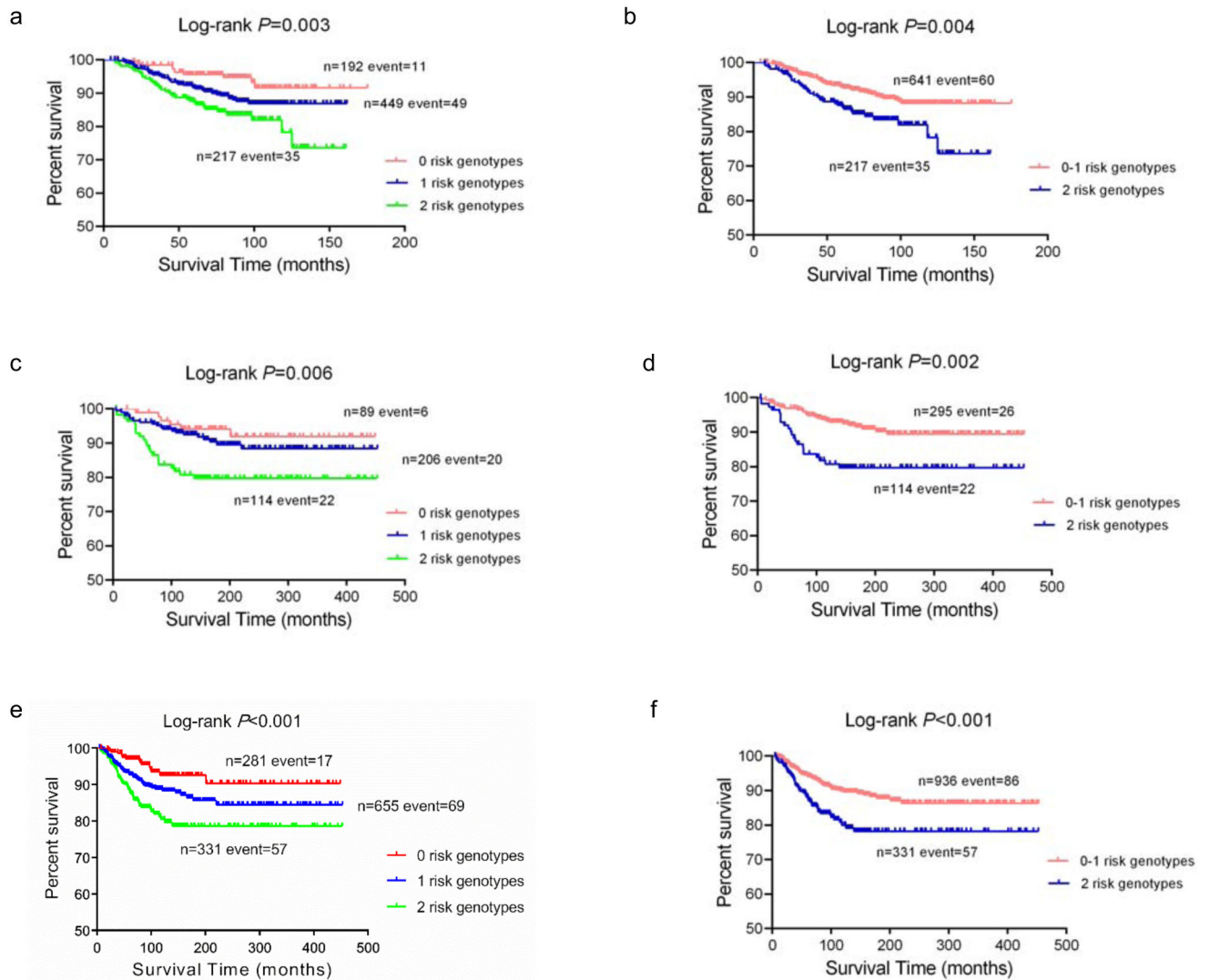
23. Mailman MD, et al., The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet*, 2007 39(10): p. 1181–6. [PubMed: 17898773]
24. Li Y, et al., MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol*, 2010 34(8): p. 816–34. [PubMed: 21058334]
25. Song F, et al., Exonuclease 1 (EXO1) gene variation and melanoma risk. *DNA Repair (Amst)*, 2012 11(3): p. 304–9. [PubMed: 22230721]
26. Riedel KG, et al., Ocular abnormalities in mucopolipidosis IV. *Am J Ophthalmol*, 1985 99(2): p. 125–36. [PubMed: 3918453]
27. Biernacka JM, et al., Assessment of genotype imputation methods. *BMC Proc*, 2009 3 Suppl 7: p. S5.
28. Wacholder S, et al., Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst*, 2004 96(6): p. 434–42. [PubMed: 15026468]
29. Chambless LE and Diao G, Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Stat Med*, 2006 25(20): p. 3474–86. [PubMed: 16220486]
30. Sax CM, et al., Transketolase gene expression in the cornea is influenced by environmental factors and developmentally controlled events. *Cornea*, 2000 19(6): p. 833–41. [PubMed: 11095059]
31. Vander Heiden MG, Cantley LC, and Thompson CB, Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science*, 2009 324(5930): p. 1029–33. [PubMed: 19460998]
32. Tseng CW, et al., Transketolase Regulates the Metabolic Switch to Control Breast Cancer Cell Metastasis via the alpha-Ketoglutarate Signaling Pathway. *Cancer Res*, 2018 78(11): p.2799–2812. [PubMed: 29599405]
33. Chao YK, et al., Transketolase Serves a Poor Prognosticator in Esophageal Cancer by Promoting Cell Invasion via Epithelial-Mesenchymal Transition. *J Cancer*, 2016 7(13): p. 1804–1811. [PubMed: 27698919]
34. Ricciardelli C, et al., Transketolase is upregulated in metastatic peritoneal implants and promotes ovarian cancer cell proliferation. *Clin Exp Metastasis*, 2015 32(5): p. 441–55. [PubMed: 25895698]
35. Tanhauserova V, et al., Genetic variability in enzymes of metabolic pathways conferring protection against non-enzymatic glycation versus diabetes-related morbidity and mortality. *Clin Chem Lab Med*, 2014 52(1): p. 77–83. [PubMed: 23492569]
36. Spallone V, Might genetics play a role in understanding and treating diabetic polyneuropathy? *Diabetes Metab Res Rev*, 2017 33(4).
37. Pacal L, et al., Role of thiamine status and genetic variability in transketolase and other pentose phosphate cycle enzymes in the progression of diabetic nephropathy. *Nephrol Dial Transplant*, 2011 26(4): p. 1229–36. [PubMed: 20826743]
38. Little EG and Eide MJ, Update on the current state of melanoma incidence. *Dermatol Clin*, 2012 30(3): p. 355–61. [PubMed: 22800543]
39. Salleron L, et al., DERA is the human deoxyribose phosphate aldolase and is involved in stress response. *Biochim Biophys Acta*, 2014 1843(12): p. 2913–25. [PubMed: 25229427]
40. Tabata S, et al., Thymidine Catabolism as a Metabolic Strategy for Cancer Survival. *Cell Rep*, 2017 19(7): p. 1313–1321. [PubMed: 28514652]
41. Furukawa T, et al., Thymidine phosphorylase in cancer aggressiveness and chemoresistance. *Pharmacol Res*, 2018 132: p. 15–20. [PubMed: 29604437]
42. Matsushita S, et al., The effect of a thymidine phosphorylase inhibitor on angiogenesis and apoptosis in tumors. *Cancer Res*, 1999 59(8): p. 1911–6. [PubMed: 10213500]
43. Hirayama A, et al., Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry. *Cancer Res*, 2009 69(11): p. 4918–25. [PubMed: 19458066]

### Highlights

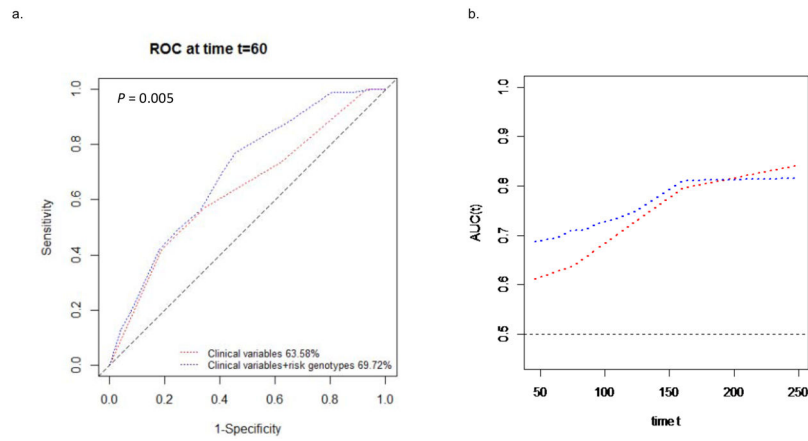
- In the pathway gene analysis, two novel and independent SNPs (*TKT* rs9864057 G>A and *DERA* rs12297652 A>G) were significantly associated with cutaneous melanoma-specific survival (HR=1.52, 95% CI=1.18-1.96,  $P=1.06\times 10^{-3}$ ) and 1.51 (1.19-1.91,  $5.89\times 10^{-4}$ )
- The associations were in an allele dose-response manner ( $P_{\text{trend}}<0.001$  for rs9864057 variant genotypes, 0.004 for rs12297652 variant genotypes, and  $<0.001$  for the number of their combined genotypes, respectively).
- *TKT* rs9864057 G>A and *DERA* rs12297652 A>G were significantly associated with higher mRNA expression levels in sun-exposed lower-leg skin ( $P=0.043$  and 0.006, respectively).

**Figure 1:**

Research flowchart. Abbreviations: NADPH: nicotinamide adenine dinucleotide phosphate; SNP: single-nucleotide polymorphism; MDACC: The University of Texas M.D. Anderson Cancer Center; FPRP: false-positive report probability; *TKT*: transketolase; *DERA*: deoxyribose-phosphate aldolase; AUC: area under the curve; ROC: receiver operating characteristic; GWAS: genome-wide association study; MAF: minor allele frequency; HWE: Hardy-Weinberg equilibrium; NHS: the Nurses' Health Study; HPFS: the Health Professionals Follow-up Study.

**Figure 2:**

The independent SNPs and CMSS. (a-f) Kaplan-Meier survival curves of the exact numbers of risk genotypes (a) in the MDACC dataset, (c) in the NHS/HPFS dataset and (e) in these two combined dataset; dichotomized groups of risk genotypes (b) in the MDACC dataset, (d) in the NHS/HPFS dataset and (f) in the combined dataset. Abbreviations: SNP: single-nucleotide polymorphism; CMSS: cutaneous melanoma-specific survival; MDACC: The University of Texas M.D. Anderson Cancer Center; NHS: the Nurses' Health Study; HPFS: the Health Professionals Follow-up Study.



**Figure 3:** ROC and time-dependent AUC estimation for prediction of CM survival in the combined dataset of MDACC and NHS/HPFS as estimated by inverse probability of censoring weighting approach. (a) Five-year CM survival prediction by ROC curve, (b) Time-dependent AUC estimation: based on age, sex and the combined risk genotypes of the two genes. Abbreviations: ROC: receiver operating characteristic; AUC: area under the curve; MDACC: The University of Texas M.D. Anderson Cancer Center; NHS: the Nurses' Health Study; HPFS: the Health Professionals Follow-up Study; CM: cutaneous melanoma.



**Table 1.**

Meta-analysis of five validated SNPs using two published melanoma GWAS datasets of the MDACC and NHS/HPFS datasets

SNP	Allele <sup>1</sup>	Gene	Chr	MDACC (n=858)			NHS/HPFS (n=409)			Meta-analysis			
				EAF	HR (95% CI) <sup>2</sup>	<i>P</i> <sup>2</sup>	EAF	HR (95% CI) <sup>3</sup>	<i>P</i> <sup>3</sup>	<i>P</i> <sub>het</sub>	<i>I</i> <sup>2</sup>	HR (95% CI) <sup>4</sup>	<i>P</i> <sup>4</sup>
rs12297652	A/G	DERA	12p12.3	0.38	1.51 (1.13-2.01)	0.005	0.40	1.51 (1.00-2.27)	0.048	1.000	0.000	1.51 (1.19-1.91)	5.89×10 <sup>-4</sup>
rs9864057	G/A	TKT	3 p21.1	0.23	1.51 (1.10-2.06)	0.003	0.23	1.55 (1.00-2.40)	0.048	0.924	0.000	1.52 (1.18-1.96)	1.06×10 <sup>-3</sup>
rs17234092	C/T	TKT	3p21.1	0.24	1.53 (1.13-2.09)	0.006	0.24	1.65 (1.08-2.52)	0.021	0.778	0.000	1.57 (1.22-2.01)	3.84×10 <sup>-4</sup>
rs17306163	T/A	TKT	3p21.1	0.24	1.53 (1.13-2.09)	0.006	0.24	1.65 (1.08-2.52)	0.021	0.778	0.000	1.57 (1.22-2.14)	3.84×10 <sup>-4</sup>
rs62255994	C/T	TKT	3p21.1	0.23	1.60 (1.18-2.18)	0.010	0.23	1.58 (1.03-2.43)	0.037	0.963	0.000	1.59 (1.24-2.05)	2.90×10 <sup>-4</sup>

Abbreviations: SNP: single nucleotide polymorphism; EAF: effect allele frequency; HR: hazards ratio; CI: confidence interval; *P*<sub>het</sub>: *P* value for heterogeneity by Cochrane's Q test; GWAS: genome-wide association study; MDACC: The University of Texas M.D. Anderson Cancer Center; NHS: the Nurses' Health Study; HPFS: the Health Professionals Follow-up Study.

<sup>1</sup> Reference allele/effect allele.

<sup>2</sup> Adjusted for age, sex, Breslow thickness, distant/regional metastasis, ulceration and mitotic rate in the MDACC dataset.

<sup>3</sup> Adjusted for age and sex in the NHS/HPFS dataset.

<sup>4</sup> Meta-analysis in a fix-effects model, when the heterogeneity statistic (*I*<sup>2</sup>) < 50% or 0.5.

**Table 2.**

Predictors of CMSS selected by stepwise Cox regression analysis in the MDACC study

Parameter <sup>1</sup>	Category <sup>2</sup>	Frequency <sup>3</sup>	HR (95% CI)	P
Age	Continuous	836	1.02 (1.00-1.03)	0.062
Sex	Female/Male	353/483	1.36 (0.85-2.19)	0.198
Regional/distant metastasis	No/Yes	693/143	3.98 (2.59-6.10)	<.001
Breslow thickness(mm)	Continuous	836	1.16 (1.10-1.22)	<.001
Ulceration	No/Yes	681/155	2.51 (1.63-3.88)	<.001
Mitotic rate (mm <sup>2</sup> )	1/>1	271/565	2.83 (1.37-5.87)	0.005
<i>TKT</i> rs9864057 G>A	GG/GA/AA	491/303/42	1.54 (1.12-2.11)	0.008
<i>DERA</i> rs12297652 A>G	AA/AG/GG	316/394/126	1.52 (1.14-2.02)	0.004

Abbreviations: CMSS: cutaneous melanoma-specific survival; MDACC: MD Anderson cancer Center; HR: hazards ratio; CI: confidence interval;

<sup>1</sup> Stepwise analysis included age, sex, regional/distant metastasis, Breslow thickness, ulceration, mitotic rate and two SNPs in two genes (rs9864057 in *TKT*; rs12297652 in *DERA*);

<sup>2</sup> The leftmost was used as the reference.

<sup>3</sup> Twenty-two missing observations of ulceration were excluded; 836 patients remained for the stepwise analysis.

**Table 3.** Associations between two independent SNPs and CMSS in the MDACC dataset, the NHS/HPFS dataset, and the combined dataset

Genotype	MDACC (n=836)			NHS/HPFS (n=409)			MDACC+ NHS/HPFS (n=1,267)		
	Frequency	HR (95% CI) <sup>1</sup>	P <sup>1</sup>	Frequency	HR (95% CI) <sup>2</sup>	P <sup>2</sup>	Frequency	HR (95% CI) <sup>3</sup>	P <sup>3</sup>
<b>TKT rs9864057 G&gt;A</b>									
GG	491	1.00		239	1.00		746	1.00	
GA	303	1.84 (1.19-2.86)	0.006	148	1.71 (0.94-3.09)	0.077	457	1.58 (1.12-2.22)	0.009
AA	42	1.74 (0.78-3.88)	0.177	22	2.11 (0.72-6.19)	0.176	64	2.23 (1.21-4.13)	0.011
Trend test			0.010			0.049			0.001
GA+AA	345	1.82 (1.20-2.77)	0.005	170	1.76 (0.99-3.11)	0.053	521	1.66 (1.19-2.30)	0.003
<b>DERA rs12297652 A&gt;G</b>									
AA	316	1.00		145	1.00		471	1.00	
AG	394	2.03 (1.22-3.40)	0.007	200	2.11 (1.03-4.34)	0.042	604	1.86 (1.24-2.78)	0.002
GG	126	2.24 (1.20-4.15)	0.011	64	2.23 (0.91-5.50)	0.081	192	2.18 (1.33-3.57)	0.002
Trend test			0.005			0.049			0.001
AG+GG	520	2.09 (1.28-3.41)	0.003	264	2.14 (1.07-4.30)	0.033	796	1.93 (1.32-2.84)	0.001
<b>Number of risk genotypes<sup>4</sup></b>									
0	184	1.00		89	1.00		281	1.00	
1	439	2.46 (1.20-5.05)	0.014	206	1.42 (0.57-3.55)	0.458	655	1.78 (1.05-3.03)	0.033
2	213	4.24 (2.02-8.89)	<.001	114	3.08 (1.24-7.62)	0.015	331	3.12 (1.81-5.36)	<.001
Trend test			<.001			0.004			<.001
0-1	623	1.00		295	1.00		936	1.00	
2	213	2.12 (1.38-3.26)	<.001	114	2.38 (1.35-4.20)	0.003	331	2.02 (1.45-2.83)	<.001

Abbreviations: SNP: single-nucleotide polymorphisms; CMSS: cutaneous melanoma-specific survival; MDACC: The University of Texas M.D. Anderson Cancer Center; NHS: the Nurses' Health Study; HPFS: the Health Professionals Follow-up Study; HR: hazards ratio; CI: confidence interval.

<sup>1</sup> Adjusted for age, sex, Breslow thickness, distant/regional metastasis, ulceration and mitotic rate in Cox models of SNPs and CMSS in the MDACC dataset;

<sup>2</sup> Adjusted for age and sex in the NHS/HPFS dataset;

<sup>3</sup> Adjusted for age and sex in the combined dataset of MDACC and NHS/HPFS;

<sup>4</sup> Risk genotypes were rs9864057 GA+AA, rs12297652 AG+GG;

Twenty-two missing observations of ulceration were excluded in the MDACC dataset.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript