Genetic variants in the folate metabolic pathway genes predict melanoma-specific survival

Running Head: Genetic variants in MTHFD1 and ALPL predict melanoma-specific survival

W. Dai,1,2,3,† H. Liu,2,3,† Y. Liu,2,3,† X. Xu,2,3 D. Qian,2,3 S. Luo,5 E. Cho,6,7,8 D. Zhu,9 C. I. Amos,9 S. Fang,10 J. E. Lee,10 X. Li,8,11 H. Nan,8,11 C. Li4,* and Q. Wei2,3,12,*

1Department of Dermatology, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong, 510515, China
2Duke Cancer Institute, Duke University Medical Center, Durham, NC, 27710, USA
3Department of Medicine, Duke University School of Medicine, Durham, NC, 27710, USA
4Department of Dermatology, Xijing Hospital, Fourth Military Medical University, Xi’an, Shaanxi, 710032, China
5Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC, 27710, USA
6Department of Dermatology, Warren Alpert Medical School, Brown University, Providence, RI, 02912, USA

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/BJD.18878

This article is protected by copyright. All rights reserved
7Department of Epidemiology, Brown University School of Public Health, Providence, RI, 02912, USA
8Channing Division of Network Medicine, Department of Medicine, Brigham and Women’s Hospital, Boston, MA, 02115, USA
9Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, TX, 77030, USA
10Department of Surgical Oncology, The University of Texas M. D. Anderson Cancer Center, Houston, TX, 77030, USA
11Department of Epidemiology, Fairbanks School of Public Health, Indiana University, Indianapolis, IN, 46202, USA
12Department of Population Health Sciences, Duke University School of Medicine, Durham, NC, 27710, USA
†These authors contributed equally to this work.

Corresponding author: Chunying Li

E-mail: qingyi.wei@duke.edu
What's already known about this topic?

- Wide range of survival rates exits among melanoma patients with similar clinical characteristics, therefore development of complementary biomarkers with specific prognostic potential is needed.
- Hypothesis-driven approach by pooling the effects of single nucleotide polymorphisms in a specific biological pathway as genetic risk scores can increase their prognostic utility and genetic polymorphisms in folate metabolism have been associated with cancer risk.

What does this study add?

- Two genetic variants in the folate metabolic pathway genes, \textit{MTHFD1} rs1950902 and \textit{ALPL} rs1091706, are significantly associated with cutaneous melanoma-specific survival.
- The risk prediction model enables estimation of cutaneous melanoma-specific survival with the addition of risk genotypes in genes of folate metabolic pathway, and can provide personalized patient education.

Key words

genome-wide association study, cutaneous melanoma-specific survival, folate metabolism, single-nucleotide polymorphism.

Funding sources

The MD Anderson Study was supported by National Institutes of Health/National Cancer Institute (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 Harvard NHS/HPFS Study was in part supported by National Institutes of Health/National Cancer Institute (R01 CA49449, P01 CA87969,
This work was also partly supported by National Natural Science Foundation of China (No.81625020, 81402736). Qingyi Wei was supported by start-up funds from Duke Cancer Institute, Duke University Medical Center and also in part supported by the Duke Cancer Institute as part of the P30 Cancer Center Support Grant (Grant ID: NIH CA014236). Wei Dai was sponsored by the China Postdoctoral Science Foundation funded project (2019M662982).

Conflicts of Interest None declared.

Manuscript word count 3427 words

Table and figure count 5

Abbreviations
ALPL, alkaline phosphatase; AUC, area under curve; BFDP, Bayesian false-discovery probability; CI, confidence interval; CM, cutaneous melanoma; CMSS, cutaneous melanoma-specific survival; GTEx, Genotype-Tissue Expression; GWAS, genome-wide association study; HR, hazards ratio; HR_{adj}, adjusted hazards ratio; LD, linkage disequilibrium; MDACC, The University of Texas MD Anderson Cancer Center; MTHFD1, methylenetetrahydrofolate dehydrogenase 1; NHS, the Nurses’ Health Study; HPFS, the Health Professionals Follow-up Study; ROC, receiver operating characteristic; SNP, single-nucleotide polymorphism.
Summary (abstract)

Background: Folate metabolism plays an important role in DNA methylation and nucleic acid synthesis and thus may function as a regulatory factor in cancer development. Genome-wide association studies (GWAS) have identified some single-nucleotide polymorphisms (SNPs) associated with cutaneous melanoma-specific survival (CMSS), but no SNPs were found in genes involved in the folate metabolic pathway.

Objective: To examine associations between SNPs in folate metabolic pathway genes and CMSS.

Methods: We comprehensively evaluated 2,645 (422 genotyped and 2,223 imputed) common SNPs in folate metabolic pathway genes from a published GWAS of 858 patients from The University of Texas M.D. Anderson Cancer Center and performed the validation in another GWAS of 409 patients from the Nurses’ Health Study and Health Professionals Follow-up Study, in which 95/858 (11.1%) and 48/409 (11.5%) patients died of cutaneous melanoma, respectively.

Results: We identified two independent SNPs (\textit{MTHFD1} rs1950902 G>A and \textit{ALPL} rs10917006 C>T) to be associated with CMSS in both datasets, and their meta-analysis yielded an allelic hazards ratio of 1.75 (95% confidence interval=1.32-2.32, \(P=9.96\times10^{-5}\)) and 2.05 (1.39-3.01, \(P=2.84\times10^{-4}\)), respectively. The genotype-phenotype correlation analyses provided additional support for biologic plausibility of these two variants’ roles in tumour progression, suggesting that variation in SNP-related mRNA expression levels is likely to be the mechanism underlying the observed associations with CMSS.

Conclusion: Two possibly functional genetic variants, \textit{MTHFD1} rs1950902 and \textit{ALPL} rs10917006, were likely to be independently or jointly associated with CMSS, which may add to personalized treatment in the future, once further validated.
Introduction

Cutaneous melanoma (CM) is the most lethal form of skin cancers because of its high metastatic potential. In 2019, there will be an estimated 96,480 new cases and 7,230 deaths from CM in the United States. The CM incidence rate has risen rapidly over the past 30 years, although the death rate has declined by 1% per year in adults 50 years of age and older between 2006 and 2015. While the conventional staging system can subdivide patients with clinically localized CM into those with low risk of recurrence and death, about 10-20% of the patients will defy the prediction. Therefore, it remains a critical need to identify prognostic biomarkers for CM.

Folate metabolism comprises an interlinked network of reactions that transfer one-carbon units for numerous cellular functions, including the biosynthesis of purines and pyrimidines for DNA and RNA as well as methyl groups for DNA methylation. Numerous studies have investigated alterations in DNA methylation and disruption of DNA integrity and repair, all of which have been observed with folate depletion. These aberrations may lead to tumour initiation and progression, linking folate metabolism to tumorigenesis, which associated with patients’ survival. However, results from epidemiological studies on folate intake and CM risk are controversial. These studies varied between in method of folate status determination, and did not account for the association between effect of genetic factors and survival of CM. Since folate can either suppress the development of early lesions in normal tissue or facilitate the proliferation of neoplastic cells. This duality, coupled with the observation that folic acid (a synthetic form of folate) can affect cellular phototoxicity and photogenotoxicity, provides a strong rationale for evaluating the relationship between folate intake and skin cancer risk. Furthermore, the folate receptor-targeting ligand molecule in the liposomal formulation may alter the mode of anticancer action of the encapsulated drug, suggesting a new possibility for folate-mediated cancer treatment. It has also been shown that immunoglobulin conjugated folate complexes can enhance the immune response by natural killer cells against folate receptor-positive CM tumour cells in vitro, suggesting their potential role in the host-rejection of tumour cells.

Most of the published genome-wide association studies (GWASs) have mainly focused on single-nucleotide polymorphisms (SNPs) that reached genome-wide significance, most of which was associated with risk susceptibility and did not have very long follow-up period. We take a targeted pathway-based, multigene approach to identify genetic variation in a metabolism-related signaling pathway and its association with cancer survival. Since the number of SNPs to be tested is greatly reduced, the typically highly stringent GWAS significance threshold could be much relaxed for such a targeted pathway-based approach. Given the role of folate metabolism in melanoma development, we hypothesized that genetic variants, i.e., SNPs, in genes involved in
the folate metabolic pathway are associated with the survival of CM patients. We tested this hypothesis by using two independently published GWAS datasets. We focused our analysis on SNPs that may have potential biological functions and thus are most likely provide additional insights about the complex mechanisms of melanoma progression.

Materials and methods

Study populations

The analysis used the GWAS dataset from The University of Texas M.D. Anderson Cancer Center (MDACC) study as a discovery dataset and the Nurses’ Health Study (NHS) and Health Professionals Follow-up Study (HPFS) as a validation dataset. All participants were from a hospital-based case-control study of CM, for which cases were recruited from among non-Hispanic white patients. A detailed description of subject selection and data collection for these two GWAS datasets has been published elsewhere \(^{15,16}\), and all patients provided a written informed consent under a protocol approved by Institutional Review Boards at both MD Anderson and Brigham and Women’s Hospital.

Gene selection and SNP genotyping

According to the Molecular Signatures Database of GSEA website (http://software.broadinstitute.org/gsea/msigdb/search.jsp), we selected 27 folate metabolic pathway genes located on autosomes (Table S1). In the MDACC dataset, genomic DNA extracted from whole blood was genotyped by the Illumina HumanOmni-Quad_v1_0_B array, in which the National Center for Biotechnology Information Database of Genotypes and Phenotypes (dbGaP Study Accession: phs000187.v1.p1) was used. Based on the 1000 Genomes Project, we performed phase I v2 CEU Genome-wide imputation using the MACH software (March 2010 release). Briefly, the typed or imputed common SNPs (minor allele frequency \( \geq 0.05 \), genotyping success rate \( > 95\% \), and Hardy-Weinberg equilibrium \( P \) value \( > 1 \times 10^{-5} \), and from imputation for those SNPs with \( r^2 \geq 0.8 \) within genes in the folate metabolic pathway or their \( \pm 2 \) kilobase flanking regions were included in association analysis. In the NHS/HPFS dataset, genotyping was performed with the Illumina HumanHap550 array, HumanHap610 array and Affymetrix 6.0 array, and imputation analysis was based on genotyped SNPs and haplotype information from phase II HapMap CEU data using the program (MACH March 2012 release). Only SNPs with imputation quality \( r^2 > 0.8 \) were included.

Results

Patient characteristics

This article is protected by copyright. All rights reserved
The present study included a discovery dataset of 858 CM patients from MDACC and a validation dataset of 409 CM patients from NHS/HPFS. The baseline characteristics of both patient cohorts have been published elsewhere\textsuperscript{15,16}. In the MDACC study, CM patients were aged between 17 and 94 years at diagnosis (mean age ± standard deviation, 52.4 ± 14.4 years), with a median follow-up time of 81.1 months. There were more males (496, 57.8%) than females (362, 42.2%) and many more diagnosed with stages I/II (709, 82.6%) than with stages III/IV (149, 17.4%). In the NHS/HPFS study, patient age ranged between 34 and 87 years at diagnosis (61.1 ± 10.8 years), and 66.3% (271) were women. These patients had a relatively longer median follow-up time (179.0 months). However, death rates were similar between the MDACC (95/858, 11.1%) and NHS/HPFS (48/409, 11.5%) studies (Table S2). We did not adjust for principal components in both discovery and validation dataset, because none of principal components was significantly associated with melanoma survival, indicating that no detectable population stratification in both MDACC and NHS/HPFS datasets was observed.

**Multivariate analyses of associations between SNPs and CMSS**

Figure 1 provides a flowchart of the present study design. We first performed a single-locus analysis in the MDACC dataset with multivariate Cox proportional hazards regression to assess the associations between 422 genotyped (Table S3) and 2,223 imputed SNPs in the 27 folate metabolic pathway genes with CMSS, adjusting for age, sex, tumour stage, Breslow thickness, ulceration and mitotic rate. Manhattan plots of associations between these variants and CMSS and quantile-quantile plot of the observed \( P \) values were shown (Figure S1 and S2). Among these SNPs, 202 were significantly associated with CMSS at \( P < 0.05 \) in an additive genetic model. After multiple test correction (Table S4) by a Bayesian false-discovery probability (BFDP) method, 190 SNPs were still considered noteworthy. As a result, 181 SNPs were identified in the NHS/HPFS validation dataset, of which 21 SNPs in two genes were validated and remained significantly associated with CMSS after multiple-testing correction by BFDP < 0.8. On the basis of the \textit{in silico} functional prediction by using SNPinfo (http://SNPinfo.niehs.nih.gov/SNPinfo/SNPfunc.html) and RegulomeDB (http://www.regulomedb.org/), five of these 21 SNPs were predicted to be functional, including one SNP in methylenetetrahydrofolate dehydrogenase 1 (\textit{MTHFD1}) and four SNPs in alkaline phosphatase (\textit{ALPL}) (Table S5). The subsequent meta-analysis using both datasets for these five SNPs showed that the same associations remained statistically significant, without any evidence for heterogeneity between the two datasets (Table 1). By using the Versatile Gene-based Association Study method, we performed the gene-based test and found three genes (i.e., \textit{MTHFD1L}, \textit{QDPR} and \textit{ALPI}) with an empirical \( P \) value < 0.05. Although no significant gene-
based test statistic was found for \textit{ALPL} and \textit{MTHFD1}, the top SNPs in both of these two genes had a significant \( P \) value < 0.01 (Table S6).

Further analysis found all four SNPs in \textit{ALPL} to be in high linkage disequilibrium (LD) (Figure S3); therefore, we selected the genotyped rs10917006 in \textit{ALPL} as an independent tagSNP, having considered \( P \) values, LD and predicted functions, along with rs1950902 in \textit{MTHFD1} (also a genotyped SNP), for further analyses. These two SNPs together with other variables were included in a multivariable stepwise Cox model, in which both (rs1950902 and rs10917006) remained significantly associated with CMSS at \( P < 0.05 \) (Table S7).

In multivariable Cox regression analyses using an additive model, we evaluated the effects of these two significant SNPs on risk of death, with adjustment for covariates where appropriate. In the MDACC dataset, we observed a significant risk effect of the \textit{MTHFD1} rs1950902 A allele (\( P_{\text{trend}} = 0.005 \)) and the \textit{ALPL} rs10917006 T allele (\( P_{\text{trend}} = 0.023 \)) on CM survival; these were also observed in the NHS/HPFS dataset (\( P_{\text{trend}} = 0.007 \) for the \textit{MTHFD1} rs1950902 A allele and \( P_{\text{trend}} = 0.004 \) for the \textit{ALPL} rs10917006 T allele) and in the MDACC and NHS/HPFS combined dataset (\( P_{\text{trend}} = 0.0003 \) for the \textit{MTHFD1} rs1950902 A allele and \( P_{\text{trend}} = 0.0007 \) for the \textit{ALPL} rs10917006 T allele) (Table 2). We presented Kaplan-Meier curves to illustrate the association of CMSS and risk genotypes of \textit{MTHFD1} rs1950902 and \textit{ALPL} rs10917006 (Figure 2a-2f). For illustrative purposes, regional association plots 50 kb up- and down-stream of the gene regions for \textit{MTHFD1} and \textit{ALPL} in the MDACC dataset were also generated (Figure S4).

\textbf{Survival of CM patients with combined risk genotypes}

To better estimate the joint association between the two independent SNPs and CMSS, we combined the risk genotypes (those associated with increased risk of death) of \textit{MTHFD1} rs1950902 GA+AA (1 vs. 0 for GG) and \textit{ALPL} rs10917006 CT+TT (1 vs 0 for CC) into one variable as a genetic score. First, we categorized all the patients into three groups according to the number of risk genotype (i.e., 0, 1 and 2). We found that the trend test indicated a risk-genotype dose-response effect. Specifically, the effect on CMSS increased as the number of risk genotypes increased in the MDACC dataset (\( P_{\text{trend}} = 0.0001 \)), the NHS/HPFS dataset (\( P_{\text{trend}} < 0.0001 \)) and the MDACC and NHS/HPFS combined dataset (\( P_{\text{trend}} < 0.0001 \)) after adjustment for covariates where appropriate. We then dichotomized all patients into the 0 risk genotype group and the 1-2 risk genotypes group. As shown in Table 2, compared with the 0 risk genotype group, the 1-2 risk genotypes group had a significant higher CM-death risk in the MDACC dataset (adjusted Hazards Ratio [HR\text{adj}] = 1.95; 95\% confidence interval [CI] = 1.28-2.96, \( P = 0.002 \)), the NHS/HPFS dataset (HR\text{adj} = 2.73; 95\% CI = 1.51-4.96, \( P = 0.0009 \)) and the MDACC and
NHS/HPFS combined dataset (HR_{adj} = 2.09; 95% CI = 1.49-2.92, \( P < 0.0001 \)). We also used Kaplan-Meier curves to illustrate the association between the number of risk genotypes and CMSS (Figure 2g-2i).

**Stratified analyses for the effect of combined risk genotypes on CMSS**

Next, we performed stratified analyses to investigate whether the joint effect of risk genotypes on CMSS was modified by other variables including age, sex, tumour stage, Breslow thickness, ulceration and mitotic rate in the MDACC dataset but only age and sex in the NHS/HPFS dataset. Compared with those with the 0 risk genotype, patients in the 1-2 risk genotypes group exhibited significantly poorer survival; this was evident in all the subgroups except for the no ulceration subgroup, those with Breslow thickness \( \leq 1 \) mm, and those with mitotic rate \( \leq 1 \) in the MDACC dataset and the subgroups age \( \leq 50 \) and male subjects in the NHS/HPFS dataset. However, no heterogeneity was found among these subgroups. (Table S8).

**Receiver operating characteristic (ROC) curve and internal validation**

To assess the predictive accuracy of the two independent SNPs, we generated the time-dependent area under receiver curve (AUC) of the ROC curve for CM patients by stage in the MDACC dataset in the presence of other variables. Consistently, the predictive performance of the ROC of 5-year CMSS was significantly improved by the addition of risk genotypes to the model (Figure 3a) with an increase in area from 86.70% to 89.15% (\( P = 0.036 \)). In the time-dependent AUC, we found that predictive performance was improved by the addition of risk genotypes from the beginning of follow-up and persisting over times in CM patients with stages I/II (Figure 3b) but not in those with stages III/IV (data not shown), indicating the observed effect in early stage CM patients. We did not perform further ROC analysis in the NHS/HPFS dataset, because only age and sex were available.

**Genotype-phenotype correlation analyses**

Two independent SNPs showed some evidence of functional relevance based on the online prediction tools, including SNPinfo and RegulomeDB. For example, \textit{MTHFD1} rs1950902 is predicted to be a missense variant, which might play a role in exonic splicing enhancer regulation. \textit{ALPL} rs10917006 is located in an intron and predicted to affect translation of protein USF2. To investigate possible correlations between the SNP genotypes and their corresponding mRNA expression levels, we conducted expression quantitative trait loci analysis with data from The Cancer Genome Atlas database. Although there were no significant associations between genetic variants of \textit{MTHFD1} rs1950902 and corresponding mRNA levels, another SNP rs4902284 in strong LD with rs1950902 (\( r^2 = 0.97 \) in HaploReg v4.1) was significantly associated with mRNA.
expression levels of MTHFD1 in primary CM tissue (P value in additive and dominant models were 0.041 and 0.035, respectively) (Figure S5a-b). We also found that the minor ALPL rs10917006 T allele was associated with an increased mRNA expression of ALPL in normal skin from the sun-exposed lower leg and unexposed suprapubic area, but a borderline in the whole blood from the Genotype-Tissue Expression (GTEx) Portal (http://www.gtexportal.org/home/) (P = 0.0007, 0.014 and 0.062, respectively, Table S9). However, no significant associations were observed between genotypes of these two SNPs and their mRNA expression levels in blood samples from 373 Europeans from the 1000 Genomes Project (data not shown). Moreover, according to experimental data from the ENCODE project, two SNPs (rs1950902 and rs10917006) are located in a DNase I hypersensitive site, where histone modification H3K27 acetylation predicted some active enhancer and promoter functions (Figure S6). These results support the possibility that these two independent SNPs have an effect on their gene expression at the transcription level.

Discussion
Folate, as a cofactor in de novo nucleotide synthesis, is critically required for cell division, growth and DNA repair. Mutations in folate metabolism genes, such as methionine synthase reductase, cause epigenetic instability on mouse development 17. In cancer cells, where DNA replication and cell division occur at a much faster rate, the blockade of folate metabolism causes inhibition of tumour proliferation and growth, forming the rational for using antifolate drugs in chemotherapy 9. Recently, genetic variants in the folate metabolism genes have been linked to risk of cancer, such as pancreatic cancer, meningioma and adult glioma 18,19. For example, lung cancer risk was also reported to be associated with genetic variants in the folate metabolism genes 20, and these associations may be modulated by dietary folate intake 21,22. Another study found that folate intake affected ovarian cancer survival 23. Although epidemiologic studies have generated contradictory results on the protection of folate intakes or concentrations against risk of cancers 24, higher intake of folate from food was found to be associated with a modestly increased risk of CM 7 and skin cancers (including melanoma and keratinocyte cancer) 25. These lines of evidence suggested that folate intake and its related metabolism might affect progression of melanoma, and this association is important for incorporating folate intervention into clinical practice, but no such studies have been reported. The present study, however, does provide the support for a link of genetic variants in MTHFD1 and ALPL to CM patient survival.

In the present study, we provided additional evidence for associations of CMSS with genetic variants in two folate metabolic genes, MTHFD1 and ALPL. Specifically, CM patients with a higher number of risk genotypes of these two genes had worse survival. It is worth mentioning
that the findings were consistent across different analyses and the majority of subgroup comparisons, suggesting a robust association between the number of risk genotypes and CMSS. In particular, the genotype-phenotype correlation demonstrates that MTHFD1 expression levels may be modulated by rs1950902 through another SNP in high LD and that ALPL expression levels may be regulated by rs10917006, suggesting that these findings are biologically plausible.

MTHFD1, located on chromosome 14q23.3, is a gene encoding a trifunctional enzyme involved in nuclear de novo thymidylate biosynthesis in the folate-dependent one-carbon metabolism. Recently, MTHFD1 has been reported to interact with the histone acetyl reader bromodomain-containing protein 4, and thus to maintain a pool of nuclear folate metabolites and to control gene expression in leukemia. Since the folate pool is maintained by MTHFD1, alterations in its activity due to genetic variants could modify disease susceptibility. More importantly, knockdown of MTHFD1 inhibited distant metastasis of melanoma cells in vivo, and metastasizing melanoma cells might reversibly increase expression of folate pathway enzymes, including MTHFD1, inhibition of which might selectively impair metastasis via oxidative stress. Furthermore, though GWASs have associated MTHFD1 variants with risk of colon cancer and postmenopausal breast cancer, MTHFD1 variants were also associated with event-free survival of patients with childhood acute lymphoblastic leukemia. However, no published reports have linked MTHFD1 variants to CM patient survival. Given that MTHFD1 is a key enzyme providing one-carbon units for thymidylate biosynthesis, the present study suggests that genetic variants in MTHFD1 may play a role in CM progression and prognosis.

ALPL, a gene that encodes a tissue non-specific isozyme, is located on chromosome 1p36.12. This enzyme is a defining marker of osteoblast activity and an important regulator of epithelial plasticity that influences biological processes including regulation of extracellular ATP levels and cell apoptosis. However, ALPL also acts as an oncogene in certain cancers. For example, a recent study showed that tumour-derived alkaline phosphatase, often elevated in metastatic cancer tissues, was associated with reduced disease-free survival in prostate cancer. Together with other stem-cell markers, a high expression of ALPL was significantly associated with shorter survival in refractory glioblastoma. To date, there is no report of altered ALPL expression in CM tumour tissues, and the present study suggests that genetic variants in ALPL may influence CM progression by regulating its mRNA expression levels, although the exact mechanisms underlying the observed association remain unknown.

The present study has some limitations. One of the limitations is that the Q-Q plots deviated considerably between the observed and expected ones across the continuum of -log10(p), which
may indicate some selection bias in the MDACC dataset. One potential weakness was the incomplete adjustment in the NHS/HPFS dataset used in the validation, because unlike the MDACC dataset used in discovery, the NHS/HPFS GWAS dataset contained only age and sex. However, the results of significant association with risk of death across different datasets and strata were consistent, suggesting that our findings might not be affected by the incomplete adjustment. Another limitation is the lack of information about different treatments, which may have had an effect on outcomes. Nevertheless, it is highly unlikely that treatment plans varied systematically by germline genotype, nor did we observe differences in death risk by tumour stage, Breslow thickness, ulceration, or mitosis, which suggests minimal bias from different treatments, if any. Although participants in the NHS/HPFS validation dataset have folate intake information, the discovery dataset lack such detailed information. The validation dataset, which has a relative smaller sample size, was used as validation for genetic data, not for folate intake, which was also a potential study limitation. Another caveat is that the sample sizes in the two GWAS studies were not large enough to allow for the false discovery rate test, a more desirable multiple test correction method; however, the BFDP might be more appropriate for the highly correlated SNPs included in the analyses, as more than 84% of these SNPs were imputed based on the LD method. These imputed SNPs were also a limitation. Finally, because the GTEx portal and other biological function-prediction databases are insufficient in evaluating mechanistic function of SNPs, further functional investigation is needed to validate our findings.

In summary, variants of genes involved in folate metabolism may play an important role in determining cancer susceptibility, also affecting the response to therapy with antifolate drugs and eventually become useful in optimizing the treatment of CM patients. Given the importance of the folate metabolism pathway in biological processes and the number of risk genotypes of these two validated SNPs in folate metabolic pathway genes, these SNPs may be useful in personalized prognosis and treatment of CM patients, but it requires a large validation cohort to substantiate these findings in the future.
References


**Supporting Information**

Supplementary Figures and Tables and Acknowledgement can be found in the online version of this article at the publisher’s website.

**Figure Legends**

**Figure 1.** Research workflow for SNPs in the folate metabolic pathway genes. Abbreviations: ALPL, alkaline phosphatase; AUC, area under curve; BFDP, Bayesian false-discovery probability; CMSS, cutaneous melanoma-specific survival; GO, Gene Ontology; GWAS, genome wide association study; HWE, Hardy Weinberg equilibrium; KEGG, Kyoto Encyclopedia of Genes and Genomes; MAF, minor allele frequency; MDACC, The University of Texas M.D. Anderson Cancer Center; MTHFD1, methylenetetrahydrofolate dehydrogenase 1; NHS, the Nurses’ Health Study; HPFS, the Health Professionals Follow-up Study; PID, Pathway Interaction Database; ROC, receiver operating characteristic; SNP, single nucleotide polymorphism.

**Figure 2.** Selected SNPs and survival prediction. Kaplan-Meier curves of cutaneous melanoma-specific survival (CMSS) stratified by MTHFD1 rs1950902, assuming a dominant model in (a) the MDACC, (b) the NHS/HPFS and (c) the MDACC and NHS/HPFS combined dataset. Kaplan-Meier curves of CMSS stratified by ALPL rs10917006 in (d) the MDACC, (e) the NHS/HPFS and (f) the MDACC and NHS/HPFS combined dataset. Kaplan-Meier survival curves of the combined risk genotypes on CMSS: dichotomized 0 risk genotype group and 1-2 risk genotypes group in (g) the MDACC, (h) the NHS/HPFS and (i) the MDACC and NHS/HPFS combined dataset. Abbreviations: SNP, single nucleotide polymorphism; CMSS, cutaneous melanoma-specific survival; MTHFD1, methylenetetrahydrofolate dehydrogenase 1; MDACC, The University of Texas M.D. Anderson Cancer Center; NHS, the Nurses’ Health Study; HPFS, the Health Professionals Follow-up Study; ALPL, alkaline phosphatase.

**Figure 3.** ROC and AUC curves for CMSS prediction. ROC curve (a) and time-dependent AUC (b) estimation for five-year CMSS prediction in cutaneous melanoma patients with stages I/II from MDACC dataset. Clinical variables include age, sex, tumour stage, Breslow thickness, ulceration and mitotic rate. Abbreviations: ROC, receiver operating characteristic; AUC, area under receiver curve; CMSS, cutaneous melanoma-specific survival; MDACC, The University of Texas M.D Anderson Cancer Center.
Table 1. Meta-analysis of five validated SNPs in the folate metabolic pathway genes using two independently published melanoma GWAS datasets

<table>
<thead>
<tr>
<th>SNP</th>
<th>Allele</th>
<th>Gene</th>
<th>Discovery-MDACC (n=858)</th>
<th>Validation-NHS/HPFS (n=409)</th>
<th>Combined-Meta-analysis (n=1267)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1950902</td>
<td>G&gt;A</td>
<td>MTHFD1</td>
<td>EAF 0.19, HR (95% CI) 1.68 (1.17-2.41), P 0.005, BFDP 0.462</td>
<td>EAF 0.18, HR (95% CI) 1.87 (1.18-2.94), P 0.007, BFDP 0.555</td>
<td>P&lt;sub&gt;het&lt;/sub&gt; 0.718, HR (95% CI) 1.75 (1.32-2.32), P 9.96×10&lt;sup&gt;-5&lt;/sup&gt;</td>
</tr>
<tr>
<td>rs12086386</td>
<td>C&gt;T</td>
<td>ALPL</td>
<td>EAF 0.05, HR (95% CI) 2.07 (1.19-3.59), P 0.010, BFDP 0.649</td>
<td>EAF 0.06, HR (95% CI) 2.17 (1.19-3.95), P 0.012, BFDP 0.686</td>
<td>P&lt;sub&gt;het&lt;/sub&gt; 0.910, HR (95% CI) 2.12 (1.41-3.18), P 3.04×10&lt;sup&gt;-4&lt;/sup&gt;</td>
</tr>
<tr>
<td>rs56222534</td>
<td>T&gt;C</td>
<td>ALPL</td>
<td>EAF 0.06, HR (95% CI) 1.94 (1.16-3.25), P 0.011, BFDP 0.662</td>
<td>EAF 0.07, HR (95% CI) 2.34 (1.32-4.16), P 0.004, BFDP 0.541</td>
<td>P&lt;sub&gt;het&lt;/sub&gt; 0.634, HR (95% CI) 2.12 (1.44-3.09), P 1.32×10&lt;sup&gt;-4&lt;/sup&gt;</td>
</tr>
<tr>
<td>rs10917006</td>
<td>C&gt;T</td>
<td>ALPL</td>
<td>EAF 0.06, HR (95% CI) 1.83 (1.09-3.09), P 0.023, BFDP 0.746</td>
<td>EAF 0.07, HR (95% CI) 2.34 (1.32-4.16), P 0.004, BFDP 0.541</td>
<td>P&lt;sub&gt;het&lt;/sub&gt; 0.534, HR (95% CI) 2.05 (1.39-3.01), P 2.84×10&lt;sup&gt;-4&lt;/sup&gt;</td>
</tr>
<tr>
<td>rs11586977</td>
<td>C&gt;A</td>
<td>ALPL</td>
<td>EAF 0.06, HR (95% CI) 1.83 (1.09-3.09), P 0.023, BFDP 0.746</td>
<td>EAF 0.07, HR (95% CI) 2.34 (1.32-4.16), P 0.004, BFDP 0.541</td>
<td>P&lt;sub&gt;het&lt;/sub&gt; 0.534, HR (95% CI) 2.05 (1.39-3.01), P 2.84×10&lt;sup&gt;-4&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

1Reference allele/effect allele;
2Adjusted for age, sex, Breslow thickness, distant/regional metastasis, ulceration and mitotic rate in an additive genetic model;
3BFDP was used for multiple test correction with detected a highest HR of 2.0 and a prior probability of 0.1;
4Adjusted for age and sex in an additive genetic model;
5Meta-analysis in a fix-effects model;
6Genotyped SNPs in the MDACC dataset;
7Imputed SNPs in the MDACC dataset.

Abbreviations: SNP, single-nucleotide polymorphism; GWAS, genome-wide association study; MDACC, The University of Texas MD Anderson Cancer Center; NHS/HPFS, the Nurses’ Health Study/Health Professionals Follow-up Study; EAF, effect allele frequency; HR, hazards ratio; CI, confidence interval; BFDP, Bayesian false-discovery probability; P<sub>het</sub>, P value for heterogeneity by Cochrane’s Q test; I<sup>2</sup>, heterogeneity statistic; MTHFD1, methylenetetrahydrofolate dehydrogenase 1; ALPL, alkaline phosphatase.
Table 2. Associations between two independent SNPs in the folate metabolic pathway genes and CMSS of patients in the MDACC dataset, the NHS/HPFS dataset and the MDACC and NHS/HPFS combined dataset

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>Multivariate analysis</th>
<th>Frequency</th>
<th>Multivariate analysis</th>
<th>Frequency</th>
<th>Multivariate analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MDACC (n=858)</td>
<td>NHS/HPFS (n=409)</td>
<td>MDACC + NHS/HPFS (n=1267)</td>
<td></td>
<td>MDACC (n=858)</td>
<td>NHS/HPFS (n=409)</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>Death (%)</td>
<td>HR (95% CI)</td>
<td>P</td>
<td>All</td>
<td>Death (%)</td>
</tr>
<tr>
<td>MTHFD1 rs1950902 G&gt;A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>554</td>
<td>50 (9.03)</td>
<td>1.00</td>
<td></td>
<td>275</td>
<td>24 (8.73)</td>
</tr>
<tr>
<td>GA</td>
<td>281</td>
<td>43 (15.30)</td>
<td>1.92 (1.26-2.92)</td>
<td>0.002</td>
<td>123</td>
<td>22 (17.89)</td>
</tr>
<tr>
<td>AA</td>
<td>23</td>
<td>2 (8.70)</td>
<td>1.51 (0.36-6.25)</td>
<td>0.573</td>
<td>11</td>
<td>2 (18.18)</td>
</tr>
<tr>
<td>Trend test</td>
<td></td>
<td></td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GA+AA</td>
<td>304</td>
<td>45 (14.80)</td>
<td>1.90 (1.25-2.87)</td>
<td>0.003</td>
<td>134</td>
<td>24 (17.91)</td>
</tr>
<tr>
<td>ALPL rs10917006 C&gt;T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>751</td>
<td>77 (10.25)</td>
<td>1.00</td>
<td></td>
<td>353</td>
<td>35 (9.92)</td>
</tr>
<tr>
<td>CT</td>
<td>107</td>
<td>18 (16.82)</td>
<td>1.83 (1.09-3.09)</td>
<td>0.023</td>
<td>54</td>
<td>12 (22.22)</td>
</tr>
<tr>
<td>TT</td>
<td>0</td>
<td>0 (0.00)</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>1 (50.00)</td>
</tr>
<tr>
<td>Trend test</td>
<td></td>
<td></td>
<td>0.023</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CT+TT</td>
<td>107</td>
<td>18 (16.82)</td>
<td>1.83 (1.09-3.09)</td>
<td>0.023</td>
<td>56</td>
<td>13 (23.21)</td>
</tr>
<tr>
<td>Combined number of risk genotypes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>482</td>
<td>41 (8.51)</td>
<td>1.00</td>
<td></td>
<td>236</td>
<td>17 (7.20)</td>
</tr>
<tr>
<td>1</td>
<td>341</td>
<td>45 (13.20)</td>
<td>1.75 (1.13-2.71)</td>
<td>0.012</td>
<td>156</td>
<td>25 (16.03)</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>9 (25.71)</td>
<td>4.32 (2.02-9.23)</td>
<td>0.0002</td>
<td>17</td>
<td>6 (35.29)</td>
</tr>
<tr>
<td>Trend test</td>
<td></td>
<td></td>
<td>0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>482</td>
<td>41 (8.51)</td>
<td>1.00</td>
<td></td>
<td>236</td>
<td>17 (7.20)</td>
</tr>
<tr>
<td>1-2</td>
<td>376</td>
<td>54 (14.36)</td>
<td>1.95 (1.28-2.96)</td>
<td>0.002</td>
<td>173</td>
<td>31 (17.92)</td>
</tr>
</tbody>
</table>

1Adjusted for age, sex, Breslow thickness, distant/regional metastasis, ulceration and mitotic rate in the MDACC dataset;
2Adjusted for age and sex in the NHS/HPFS dataset;
3Adjusted for age and sex in the MDACC and NHS/HPFS combined dataset;
4Risk genotypes include MTHFD1 rs1950902 GA+AA and ALPL rs10917006 CT+TT.

Abbreviations: SNP, single-nucleotide polymorphism; CMSS, cutaneous melanoma-specific survival; MDACC, The University of Texas MD Anderson Cancer Center; NHS/HPFS, the Nurses’ Health Study/Health Professionals Follow-up Study; HR, hazards ratio; CI, confidence interval; MTHFD1, methylenetetrahydrofolate dehydrogenase 1; ALPL, alkaline phosphatase.
Figure 1

- 27 genes in the folate metabolism pathway (KEGG, GO, REACTOME and PID)
- 2,645 common SNPs: 422 genotyped and 2,223 imputed
- 202 SNPs significantly associated with cutaneous melanoma specific survival ($P < 0.05$)
- 190 SNPs with BFPD < 0.00
- 21 SNPs in two genes were validated with $P < 0.05$ and had the same direction of effects
- Two independent SNPs: MTHFD1 rs1950902 and ALPL rs18917006
- Combined analysis
  - Stratified analysis
  - Time-dependent AUC and ROC Curve
- SNP-Gene expression analysis

**MDACC GWAS study:** 858 patients; Individual call rate > 95%; MAF > 5%; INFO > 10%; Chromosome 1-22; Gene ± 2kb (hg19)

**GeneABELOR:**
- Cox proportional hazards regression analysis
  - Cutaneous melanoma specific survival
  - Additive genetic model

**Validation in NHGRI-HPFS GWAS study:** 499 patients; Cox proportional hazards regression analysis
- Additive genetic model
- Functional prediction;
  - Linkage disequilibrium analysis;
  - SNP geno analysis