IJC
International Journal of Cancer

Integrated pathway and epistasis analysis reveals interactive effect of genetic variants at *TERF1* and *AFAP1L2* loci on melanoma risk

Myriam Brossard[1,2]*, Shenying Fang[3]*, Amaury Vaysse[1,2], Qingyi Wei[4], Wei V. Chen[5], Hamida Mohamdi[1,2], Eve Maubec[1,2,6], Nolwenn Lavielle[1,2], Pilar Galan[7], Mark Lathrop[8], Marie-Françoise Avril[9], Jeffrey E. Lee[3], Christopher I. Amos[10]** and Florence Demenais[1,2]**

[1] INSERM, UMR-946, Genetic Variation and Human Diseases Unit, Paris, France
[2] Institut Universitaire d'Hématologie, Université Paris Diderot, Sorbonne Paris Cité, Paris, France
[3] Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX
[4] Duke University Medical Center and Department of Medicine, Duke Cancer Institute, Duke University School of Medicine, Durham, NC
[5] Laboratory Informatics System, Department of Clinical Applications & Support, The University of Texas M. D. Anderson Cancer Center, Houston, TX
[6] AP-HP, Service de Dermatologie, Hôpital Bichat, Université Paris Diderot, Paris, France
[7] INSERM, UMR U557, Institut National de la Recherche Agronomique,U1125, Conservatoire National des arts et Métiers, Centre de Recherche en Nutrition Humaine, Ile de France, Bobigny, France
[8] McGill University and Genome Quebec Innovation Centre, Montreal, QC, Canada
[9] AP-HP, Hôpital Cochin et Université Paris Descartes, Paris, France
[10] Department of Community and Family Medicine, Geisel College of Medicine, Dartmouth College, Hanover, NH

Genome-wide association studies (GWASs) have characterized 13 loci associated with melanoma, which only account for a small part of melanoma risk. To identify new genes with too small an effect to be detected individually but which collectively influence melanoma risk and/or show interactive effects, we used a two-step analysis strategy including pathway analysis of genome-wide SNP data, in a first step, and epistasis analysis within significant pathways, in a second step. Pathway analysis, using the gene-set enrichment analysis (GSEA) approach and the gene ontology (GO) database, was applied to the outcomes of MELARISK (3,976 subjects) and MDACC (2,827 subjects) GWASs. Cross-gene SNP–SNP interaction analysis within melanoma-associated GOs was performed using the INTERSNP software. Five GO categories were significantly enriched in

Key words: genome-wide association studies, melanoma, pathway analysis, gene–gene interaction

Abbreviations: ACD: adrenocortical dysplasia homolog (mouse); AFAP1L2: actin filament associated protein 1-like 2; AIF: apoptosis-inducing factor; AIFM3: apoptosis-inducing factor, mitochondrion-associated, 3; Akt: v-akt murine thymoma viral oncogene; ATM: ataxia telangiectasia mutated serine/threonine kinase; CASP8: caspase 8, apoptosis-related cysteine peptidase; CDK4: cyclin-dependent kinase 4; CDKN2A: cyclin-dependent kinase inhibitor 2A; CDKN2B: cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4); CEU: Northern Europeans from Utah; CMTM7: CKLF-like MARVEL transmembrane domain containing 7; kb: kilobase; MC1R: melanocortin 1 receptor (alpha melanocyte stimulating hormone receptor); NCBI: National Center for Biotechnology Information; PI3K: phosphoinositide 3-kinase; POT1: protection of telomeres 1; PRKCE: protein kinase C, epsilon; SNP: single nucleotide polymorphism; TERF1: telomeric repeat binding factor (NIMA-interacting) 1; TERF2IP: telomeric repeat binding factor 2, interacting protein; TERT: telomerase reverse transcriptase; TIAL1: TIA1 cytotoxic granule-associated RNA binding protein-like 1; TNFSF4: tumor necrosis factor ligand superfamilly member 4; TSI: Tuscans from Italy

Additional Supporting Information may be found in the online version of this article.

*M.B. and S.F. contributed equally to this work

**F.D. and C.I.A. contributed equally to this work

Grant sponsor: Programme Hospitalier de Recherche Clinique (PHRC); **Grant number:** AOM-07-195; **Grant sponsor:** Institut National du Cancer (INCa); **Grant number:** INCa_5982; **Grant sponsor:** Ligue Nationale Contre le Cancer; **Grant sponsor:** PRE 09/FD; **Grant sponsor:** Fondation pour la Recherche Médicale (FRM); **Grant sponsor:** FDT20130928343; **Grant sponsor:** the European Commission under the 6th Framework Programme; **Grant number:** LSH-CT-2006-018702; **Grant sponsor:** National Institutes of Health (NIH); **Grant numbers:** R01CA100264, 2P50CA093459, P30CA016672, R01CA133996; **Grant sponsor:** National Cancer Institute; **Grant numbers:** SPORE P50 CA093459, Developmental Research Program P50 CA093459; **Grant sponsors:** Miriam and Jim Mulva Research Fund, Marit Peterson Fund for Melanoma Research

DOI: 10.1002/ijc.29570

History: Received 17 Dec 2014; Accepted 30 Mar 2015; Online 17 Apr 2015

Correspondence to: Florence Demenais, UMR-946, INSERM-Université Paris Diderot, 27 rue Juliette Dodu, 75010 Paris, France, Tel.: +33153725002, Fax: +33153725049, E-mail: florence.demenais@inserm.fr

genes associated with melanoma (false discovery rate $\leq$ 5% in both studies): response to light stimulus, regulation of mitotic cell cycle, induction of programmed cell death, cytokine activity and oxidative phosphorylation. Epistasis analysis, within each of the five significant GOs, showed significant evidence for interaction for one SNP pair at *TERF1* and *AFAP1L2* loci ($p_{meta\text{-}int} = 2.0 \times 10^{-7}$, which met both the pathway and overall multiple-testing corrected thresholds that are equal to 9.8×10^{-7} and 2.0×10^{-7}, respectively) and suggestive evidence for another pair involving correlated SNPs at the same loci ($p_{meta\text{-}int} = 3.6 \times 10^{-6}$). This interaction has important biological relevance given the key role of TERF1 in telomere biology and the reported physical interaction between TERF1 and AFAP1L2 proteins. This finding brings a novel piece of evidence for the emerging role of telomere dysfunction into melanoma development.

What's new?

Only a small percentage of melanoma risk is known to be linked to genetic variants identified by genome-wide association studies. In this study, the authors sought to determine whether variants within multiple pathways might interact to increase this risk. They identified five pathways associated with melanoma, and showed significant interaction between variants at *TERF1* and *AFAP1L2* loci. Given the key role of TERF1 in telomere biology and its physical interaction with AFAP1L2, these results support a role for telomere dysfunction in melanoma development.

Cutaneous melanoma (CM) is the most aggressive form of skin cancer and has shown a consistent increase in incidence among fair-skinned populations over the past four decades.[1] CM risk likely results from the joint effects and interactions of environmental factors, including exposure to ultraviolet radiation, and many genetic factors. There is a broad spectrum of genetic variation that underlies genetic susceptibility to CM. This variation extends from rare high-penetrance mutations influencing familial CM risk [such as the well-established cyclin-dependent kinase inhibitor 2A (*CDKN2A*)[2] and cyclin-dependent kinase 4 (*CDK4*)][3] to common low-risk variants involved in more common forms of CM, as those recently identified by genome-wide association studies (GWASs). GWASs of CM have been successful in identifying 13 loci associated with CM risk[4,5] but, as for other complex diseases, genetic variants at these loci account for only a small portion of CM risk.

Typically, GWASs focus on the analysis of individual SNPs, and only the strongest evidence for associations for a limited number of top-ranked SNPs is reported. One major limitation of this approach is that disease susceptibility is likely to result from the combined and possibly interactive effects of many genetic factors, each making a small contribution to overall disease risk, and the effect of such factors may be missed if they are examined individually. Recently, a number of complementary approaches to GWAS have been proposed to prioritize genes involved in complex diseases, including GWAS pathway analysis and epistasis analysis.

Pathway analysis is based on the premise that genes do not work in isolation; instead, genes that belong to biological and functional units (pathways or gene sets) can harbor markers which might be detectable when examined jointly. This approach can also provide novel insights into disease biology. A number of methods for performing genome-wide pathway analysis of SNP data have been proposed, of which the most popular is the gene-set enrichment analysis[6] (GSEA) method. This approach uses a competitive test to search for pathways enriched in genes associated with disease risk. GSEA has successfully identified a number of pathways associated with risk of complex diseases,[7] including cancers such as basal cell carcinoma of the skin[8] and bladder cancer.[9]

Another way to increase power of detecting association of markers with disease risk is to conduct statistical analysis that considers gene–gene interactions; this is especially true for low marginal-effect SNPs. Although advances in computational methods[10,11] have considerably reduced the enormous computing time required to examine interactions at the genome-wide level, the multiple hypothesis testing issue remains a major limitation of genome-wide epistasis analysis. Statistical and biological filtering pipelines have been proposed to limit the search for pairwise interactions among a subset of genetic markers.[12] Although these two filtering procedures most often have been used independently, combining them may further limit the interaction search and thus increase statistical power.

In this study, we propose a two-step analysis strategy aimed at characterizing pathways associated with melanoma risk, in the first step, and identifying interactions among all cross-gene SNP pairs within significant pathways, in the second step. This analysis strategy was applied to two large case–control studies of melanoma with genome-wide SNP data: the French MELARISK study, which comprised of 3,976 subjects of European ancestry, and served as the discovery dataset, and the North-American MD Anderson Cancer Center (MDACC) study, which included 2,827 European-ancestry subjects and was used as the replication dataset.

Material and Methods

Study populations

Protocols of both MELARISK and MDACC studies have been approved by the ethical committees of Paris-Saint-Louis, Paris-Necker and Ile-de-France II (Paris, France) for

the MELARISK study and by the MD Anderson Cancer Center Institutional Review Board (Houston, Texas) for the MDACC study. All subjects participating to these studies gave their written informed consent.

Discovery study. The discovery dataset comprised of 1,244 CM cases from the MELARISK study and 3,144 population-based controls. The protocol of recruitment and data collection of CM patients has been described in detail elsewhere.[13] Briefly, melanoma patients were recruited between 1992 and 2011 through a nationwide network of French Dermatology Departments and Oncogenetic clinics that constitute the French Melanoma Study Group and MELARISK collection. MELARISK patients were histologically confirmed melanoma cases of European ancestry who had been enrolled for the presence of family history of melanoma, multiple primaries and/or diagnosis before 40 years of age, in order to enrich the case series for genetic susceptibility. Information on demographic and clinical characteristics, family history of melanoma and known melanoma risk factors was collected through interview and clinical examination, extraction of medical records and by completing a questionnaire. Confirmation of each reported case of melanoma was sought through review of medical records, review of pathological material and/or from pathological reports. In case of family history, only one CM patient per family was included in our study. Confirmed melanoma cases included in this study had been tested negative for germline mutations of *CDKN2A*, the major known locus predisposing to familial melanoma.[2] Population-based controls were recruited from the same various parts of France as the cases and had genotyping data provided by the Centre National de Génotypage (CNG, Evry, France). After quality control (QC) of genotypic data, 1,179 cases and 2,797 controls were included in the analysis.

Replication study. The MDACC study population has been described in the published GWAS.[14] Briefly, this replication population consisted of 1,804 European-ancestry patients with newly diagnosed CM and 1,026 controls, who were recruited from The University of Texas MD Anderson Cancer Center between March 1998 and August 2008. Of these subjects, 931 CM patients and 1,026 age- and sex-matched cancer-free controls had completed a lifestyle questionnaire to provide information about their demographic and the known risk factors for CM. The remaining 873 CM patients without questionnaire data were recruited, as a part of a case series study for survival analysis. After quality control of genotypic data, 1,801 cases and 1,026 controls were included in the analysis.

Genotype data

Cases and controls of the French study were genotyped in three stages. The first two stages of genotypic data contributed to the GenoMEL consortium GWASs.[15,16] Genotyping was based on Illumina HumanHap300 Beadchip version 2 duo array and Illumina Humancnv 370K array for stage 1 cases and controls, respectively (575 cases, 1,923 controls) and on Illumina Human610-Quad array for stage 2 and stage 3 case–control samples (669 cases, 1,221 controls). Genotyping was performed at Centre National de Génotypage (Evry, France) for 93% of all samples and at Service XS (Leiden, the Netherlands) for the remainder. Standard quality controls measures were applied to samples and SNPs separately for each genotyping stage. Detailed information about the QC procedure can be found in the published GWAS papers for stage 1 and stage 2 samples.[15,16] The same QC criteria were applied to stage 3 samples. In addition, principal component analysis (PCA) was applied to all pooled cases and controls who had passed QC to remove a few additional outliers and duplicates.

To get the same set of SNPs across samples and increase SNP density, imputations were carried out in stage 1, stage 2 and stage 3 samples separately using MaCH[17] (version 4.4.3) and Hapmap3 CEU+TSI populations as reference panel (release 2). A total of 1,032,745 SNPs that had imputation quality score (rsq) $\geq$0.8 and minor allele frequency (MAF) $\geq$0.05 and were shared by stage 1, stage 2 and stage 3 samples were retained for subsequent analysis.

MDACC samples were genotyped with the Illumina HumanOmni1-Quad_v1-0_B array at the Johns Hopkins University Center for Inherited Disease Research (CIDR). QC was applied to samples and SNPs, as described previously.[14] Genome-wide imputation, in MDACC, was performed using MaCH (version 1.0) program and Hapmap3 CEU panel (release 2). After imputation 1,067,258 SNPs with rsq $\geq$0.80 and MAF $\geq$0.05 were retained for analysis.

Statistical methods

Pathway analysis. Pathway analysis was applied to outcomes of MELARISK and MDACC genome-wide association analyses, respectively. Association between individual SNPs and CM risk in the MELARISK dataset was assessed by logistic regression using allele dosage for SNP effect and adjusting for stage of genotyping using STATA V12 (distributed by Stata Corporation, College Station, TX). Test of SNP effect was based on the Wald test. We checked that further inclusion of the first two principal components, obtained from PCA of genotyped SNP data, in the regression model did not change the results. In the MDACC dataset, the initial GWAS was conducted using the ProbABEL[18] software, which includes logistic regression analysis for disease status onto genotype probabilities assuming an additive model (which is equivalent to allele dosage). The effect of each SNP on risk of melanoma was examined *via* a Wald test, as in MELARISK.

Pathway analysis of each dataset was performed using the GSEA approach implemented in the GenGen package.[6] GSEA derives an enrichment score to detect gene sets significantly enriched in genes associated with disease compared to the whole genome. It uses the single SNP association test statistic for the most significant SNP in each gene to represent

the gene. All genes are ranked in descending order of their test statistic value. A weighted Kolmogorov–Smirnov-like running sum statistic is calculated to determine, within each pathway, overrepresentation of highly ranked genes from the ranked list of all genes. The Kolmogorov–Smirnov-like statistic is normalized to account for differences in the number of genes across the pathways. Statistical significance for pathways is calculated by permutation. SNPs between the start site and the 3′-untranslated region were assigned to genes using NCBI dbSNP Build 132 and human Genome Build 37.1. To map genes to pathways, we used gene ontology (GO) categories. To reduce the number of overlapping GO categories, we considered biological process and molecular function level-4 ontologies, as provided by the GenGen package. Our analysis was also restricted to GOs that contained at least 20 and at most 200 genes to avoid testing overly narrow or broad functional categories. Statistical significance of pathway enrichment scores was determined by 100,000 SNP statistics permutations. We used SNP statistics permutation across the genome instead of case–control status permutation as it is more computationally efficient and does not require access to raw data. Simulations have shown that SNP-statistics permutations perform much better than gene-statistics permutations when compared to the gold-standard approach of permuting the case–control status.[19] We computed empirical *p*-values and the false discovery rate (FDR) to correct for multiple testing. We used a FDR ≤ 0.05 as a stringent criterion for statistical significance in the discovery dataset and a FDR ≤ 0.05 in the replication dataset as a criterion for a validated result.

Cross-gene SNP–SNP interaction analysis within melanoma-associated pathways. In MELARISK, we analyzed all cross-gene SNP–SNP interactions within each melanoma-associated pathway. Pairwise SNP–SNP interactions were evaluated by logistic regression assuming an additive model for SNP main effects and interaction using the INTERSNP[20] software (version 1.14). The additive effect of a SNP was represented by a variable that was coded -1, 0 and 1 for homozygote for the major allele, heterozygote and homozygote for the minor allele, respectively. The interaction term was modeled by the multiplication of variables between SNPs. Test of interaction was performed by comparing the full model, which included the additive effects of the two SNPs plus the interaction term to the restricted model without interaction, using a likelihood-ratio test which follows a χ^2 distribution with one degree of freedom.

For each gene, we examined all SNPs passing QC and lying from 50 kb upstream to 50 kb downstream of the gene (Build 37.1). As required by INTERSNP, we used the best guess genotype provided by MaCH when its maximum posterior probability was greater than or equal to 0.90 (otherwise it was set to missing). We discarded strongly correlated SNPs ($r^2 \geq 0.8$) and all SNP pairs for which one or more of the nine genotypic combinations appeared in fewer than five subjects in cases and controls, respectively. All pairs of SNPs showing suggestive evidence for interaction ($p_{\text{int}} \leq 10^{-4}$) in MELARISK were subjected to replication in MDACC. We selected SNP pairs that replicated in MDACC at the nominal 5% level and showed the same direction of interaction as in MELARISK. We then meta-analyzed the discovery and replication results using a fixed-effects model.

To correct for multiple testing, we used a hierarchical bottom-up procedure. We first corrected for multiple interaction tests for each gene pair, then for multiple gene pairs within a pathway and finally across all pathways found significantly associated with melanoma. Specifically, for each gene pair, we computed the effective number of independent interaction tests from the eigenvalues of the correlation matrix of products of SNPs allele dosages, similarly to the method proposed by Li and Ji.[21] The effective number of independent tests in a pathway was estimated by the sum of the effective number of independent tests for a gene pair over all gene pairs tested within that pathway in the discovery dataset. The corrected critical threshold for the number of tests in a pathway (T_{pathway}) was thus equal to the 5% type I error divided by the effective number of independent tests in that pathway. Finally, to correct for overall statistical significance across all melanoma-associated pathways, we applied a Bonferroni correction to the pathway-corrected threshold (T_{pathway}) to get the overall critical threshold (T_{overall}).

Results

Pathway analysis

GSEA was applied to GWAS results from single SNP association analysis conducted in 1,179 melanoma cases and 2,797 controls from the MELARISK study and in 1,801 melanoma cases and 1,026 controls from the MDACC study. In MELARISK, 459,637 of 1,032,745 Hapmap3-imputed SNPs were located within genes and mapped to 21,810 genes. Of these, 6,873 genes were assigned to 316 level-4 GO categories. In MDACC, 475,093 of 1,067,258 SNPs were located within genes and mapped to 22,096 genes. Of these, 6,909 genes were assigned to 319 level-4 GO categories. A total of 15 GOs achieved the threshold for statistical significance (FDR ≤ 0.05) in MELARISK. Five of these GOs were successfully replicated in MDACC (FDR ≤ 0.05) and were the following: response to light stimulus (GO:0009416), regulation of mitotic cell cycle (GO:0007346), induction of programmed cell death (GO:0012502), cytokine activity (GO:0005125) and oxidative phosphorylation (GO:0006119) (Table 1). These GO categories included from 52 to 193 genes. The proportion of genes that drive the enrichment score of a pathway varied from 25% (regulation of mitotic cell cycle) to 43.2% (cytokine activity) in MELARISK and from 25% (regulation of mitotic cell cycle) to 45.3% (response to light stimulus) in MDACC (Table 1). These proportions were quite similar between the two studies for the significant gene sets, except for cytokine activity where that proportion was higher in

Table 1. Pathways associated with melanoma risk in MELARISK and MDACC studies

	MELARISK				MDACC			
Pathway (GO class)	No. of genes[1]	p[2]	FDR[2]	No. of genes driving a pathway (%)[3]	No. of genes[1]	p[2]	FDR[2]	No. of genes driving a pathway (%)[3]
Response to light stimulus (GO:0009416)	75	$<10^{-5}$	0.001	31 (41.3)	75	$<10^{-5}$	0.002	34 (45.3)
Regulation of mitotic cell cycle (GO:0007346)	52	$<10^{-5}$	0.001	13 (25.0)	52	5.6×10^{-3}	0.05	13 (25.0)
Induction of programmed cell death (GO:0012502)	193	5.0×10^{-5}	0.004	71 (36.8)	193	3.0×10^{-4}	0.02	76 (39.4)
Cytokine activity (GO:0005125)	183	3.6×10^{-3}	0.05	79 (43.2)	188	9.1×10^{-3}	0.05	63 (33.5)
Oxidative phosphorylation (GO:0006119)	54	1.6×10^{-3}	0.04	20 (37.0)	54	6.5×10^{-4}	0.02	18 (33.3)

[1]Number of genes in a pathway having at least one GWAS SNP mapped to a gene.
[2]Empirical p-value and FDR of a pathway estimated by 100,000 permutations of SNP statistics.
[3]Number of genes driving the enrichment score of a pathway and proportion (%) over total number of genes in that pathway.

MELARISK (43.2%) than in MDACC (33.5%). The genes driving the five significant gene sets showed small degree of overlap as there were at most four and nine overlapping genes between induction of programmed cell death and response to light stimulus GOs in MELARISK and MDACC, respectively (Supporting Information Figs. 1a and 1b).

Cross-gene SNP–SNP interaction analysis within melanoma-associated pathways

We examined 518 distinct genes that belonged to the five pathways identified by GSEA. This constituted a total of 7,975 analyzed SNPs, which resulted in 5,577,994 within-pathway cross-gene SNP–SNP interaction tests in the discovery MELARISK dataset. The effective number of independent tests per pathway and corresponding pathway and overall multiple-testing corrected thresholds are shown in Supporting Information Table 1. There were four SNP pairs related to four gene pairs that had $p_{\text{int}} \leq 10^{-4}$ in MELARISK and replicated at the nominal 5% level in MDACC (Table 2). Two of these pairs involved two SNPs within the same gene (*PRKCE*), which were independent [linkage disequilibrium (LD) measures being $r^2 = 0.001$ and $D' = 0.05$]. An additional SNP pair, related to one of the four replicated gene pairs (*TERF1*/*AFAP1L2*), was found to have a p values of 2.1×10^{-4} in MELARISK (close to the 10^{-4} threshold chosen to select pairs for replication) and to strongly replicate in MDACC ($p_{\text{int}} = 2.4 \times 10^{-4}$) (Table 2). The two SNP pairs at *TERF1* and *AFAP1L2* loci were sharing one SNP (rs3863241) at *TERF1* locus and had their two SNPs at *AFAP1L2* locus in moderate LD ($r^2 = 0.56$; $D' = 0.86$). All five SNP pairs showed increased evidence for interaction in the meta-analysis of MELARISK and MDACC. The regression coefficients and standard errors associated with each SNP–SNP interaction effect are shown in Table 2. The strongest interaction was found for rs3863241 (*TERF1*)/rs649785 (*AFAP1L2*) pair with $p_{\text{meta-int}} = 2.0 \times 10^{-7}$, which remained significant after correction for multiple testing, provided T_{pathway} and T_{overall} are equal to 9.8×10^{-7} and 2.0×10^{-7}, respectively (Table 2 and Supporting Information Table 1). The other pair at *TERF1* and *AFAP1L2* loci had $p_{\text{meta-int}} = 3.6 \times 10^{-6}$, which was higher by only 0.6 order of magnitude than the pathway-corrected threshold (Table 2). The other three pairs had $p_{\text{meta-int}}$ ranging from 2.2×10^{-6} for rs10189339 (*PRKCE)*/rs17098973 (*TIAL1*) to 1.0×10^{-5} for rs6743144 (*PRKCE)*/rs178295 (*AIFM3)* pair. These p-values were from 1.5 to 2.4 and from 2.2 to 3.1 orders of magnitude higher than their corresponding pathway and overall multiple-testing corrected thresholds (Table 2 and Supporting Information Table 1). We noted that for all five SNP pairs, no single SNP had a nominally significant marginal effect and, in the presence of interaction in the model, the SNP main effects were also not nominally significant in either MELARISK or MDACC, except for rs6743144 in *PRKCE* (p = 0.03) and rs614004 in *CMTM7* ($p = 1.1 \times 10^{-3}$) in MELARISK (Supporting Information Table 2).

The SNP pair with significant evidence for interaction, after taking into account multiple testing, showed a pattern of interaction in which the odds ratios (ORs) associated with GG (or CC) genotype of rs649785 (*AFAP1L2*) had an inverse effect depending on the genotype, TT or CC, at rs3863241 (*TERF1*) (Fig. 1*a*). A similar pattern was observed for the other SNP pair (rs597371/rs3863241) involving the same loci (Fig. 1*b*). Further haplotype analysis of the two SNPs at *AFAP1L2* locus by stratifying on genotype at rs3863241 (*TERF1* locus), using the THESIAS[22] program, showed that a single haplotype bearing the minor C allele of rs649785 and minor G allele of rs597371 had an inverse effect on melanoma risk depending on the genotype TT (OR = 1.31, 95% CI = 1.13–1.52; $p = 2.9 \times 10^{-4}$) or CC (OR = 0.71, 95% CI = 0.59–0.85, $p = 2.7 \times 10^{-4}$) at rs3863241 (Table 3). Regarding the other SNP pairs, the pair related to *PRKCE*/*AIFM3* showed a similar pattern of interaction as the *TERF1*/

Table 2. SNP pairs showing evidence for interaction on the risk of melanoma

						MELARISK		MDACC		META-ANALYSIS		Multiple-testing corrected thresholds[5]	
						Interaction Effect		*Interaction Effect*		*Interaction Effect*			
Pathway (GO class)	SNP	Chr[1]	Genes[2]	Alleles[3]	MAF[4]	β_{int} (se_{int})[5]	p_{int}[6]	β_{int} (se_{int})[5]	p_{int}[6]	$\beta_{meta\text{-}int}$ ($se_{meta\text{-}int}$)[7]	$p_{meta\text{-}int}$[8]	$T_{pathway}$	$T_{overall}$
Regulation of mitotic cell cycle													
	rs3863241 rs649785[10]	8 10	*31kU TERF1* *18kD AFAP1L2*	C/T C/G	0.46 0.48	−0.27 (0.07)	2.1×10^{-4}	−0.29 (0.08)	2.4×10^{-4}	−0.28 (0.05)	2.0×10^{-7}	9.8×10^{-7}	2.0×10^{-7}
	rs3863241 rs597371[10]	8 10	*31kU TERF1* *9kD AFAP1L2*	C/T G/A	0.46 0.42	−0.32 (0.07)	1.3×10^{-5}	−0.17 (0.08)	0.03	−0.25 (0.05)	3.6×10^{-6}	9.8×10^{-7}	2.0×10^{-7}
Induction of programmed cell death													
	rs10189339[11] rs17098973	2 10	*PRKCE (Intron)* *45kU TIAL1*	G/A A/G	0.32 0.26	0.38 (0.09)	4.3×10^{-5}	0.25 (0.10)	0.009	0.31 (0.07)	2.2×10^{-6}	3.6×10^{-8}	7.1×10^{-9}
	rs6743144[11] rs178295	2 22	*PRKCE (Intron)* *15kU AIFM3*	A/G C/T	0.48 0.28	−0.34 (0.08)	2.8×10^{-5}	−0.17 (0.09)	0.04	−0.26 (0.06)	1.0×10^{-5}	3.6×10^{-8}	7.1×10^{-9}
Cytokine activity													
	rs10912551 rs614004	1 3	*33kD TNFSF4* *CMTM7 (Intron)*	T/G G/A	0.21 0.48	−0.39 (0.09)	7.7×10^{-6}	−0.20 (0.09)	0.05	−0.31 (0.07)	3.5×10^{-6}	1.2×10^{-7}	2.5×10^{-8}

[1]Chr is the chromosome number where the SNP is located.
[2]Gene symbol; k indicates distance to gene in kilobases (Build 37.1); U indicates "upstream of"; D indicates "downstream of".Upstream and downstream respectively refer to the 5′ and 3′ ends of the coding strand of each gene.
[3]Minor allele/Major allele.
[4]MAF, minor allele frequency estimated in controls.
[5]β_{int} is the regression coefficient of interaction effect in logistic regression assuming an additive model; se_{int} is the standard error of β_{int}.
[6]p_{int} is the *p*-value of the likelihood-ratio test for interaction (one degree of freedom test assuming an additive model).
[7]$\beta_{meta\text{-}int}$ and $se_{meta\text{-}int}$ are the regression coefficient and standard error of interaction effect obtained from the meta-analysis of MELARISK and MDACC results (using a fixed-effects model).
[8]$p_{meta\text{-}int}$ is the *p*-value associated with the Wald test of meta-analyzed interaction effect.
[9]$T_{pathway}$ and $T_{overall}$ are respectively the pathway-corrected and overall-corrected critical thresholds to take into account multiple testing (see material and methods for details).
[10]rs649785 and rs597371 are in moderate linkage disequilibrium (r^2=0.56; D'=0.86).
[11]rs10189339 and rs6743144 are independent (r^2=0.001; D'=0.05).

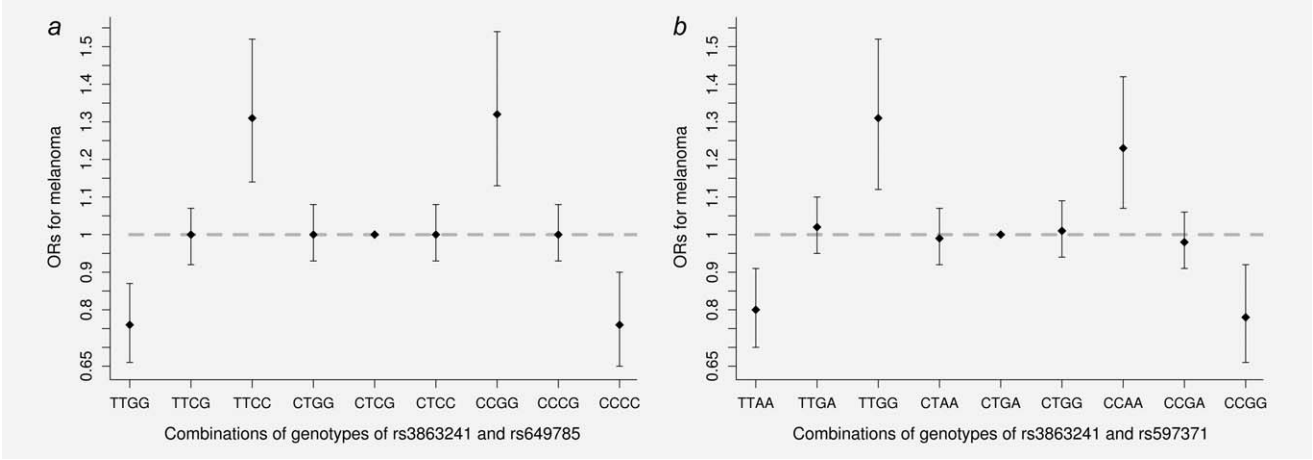

Figure 1. Two-locus odds ratios (ORs) and 95% confidence intervals for melanoma for each genotypic combination of *TERF1* and *AFAP1L2* SNPs. On the left graph (*a*), subjects are grouped by genotypes at both *TERF1* rs3863241 and *AFAP1L2* rs649785. On the right graph (*b*), subjects are grouped by genotypes at both *TERF1* rs3863241 and *AFAP1L2* rs597371. The *x* and *y* axes represent genotypic combinations of *TERF1* and *AFAP1L2* SNPs and odds ratios for melanoma, respectively. For each SNP pair, the odds ratios were computed using the estimates of SNP main effects and interaction obtained from the combined analysis of MELARISK and MDACC. Note that for each SNP pair, the double heterozygous genotype has an odds ratio of 1 using our coding scheme (as explained in the statistical methods section).

Table 3. Effect of haplotypes bearing the two *AFAP1L2* SNPs on the risk of melanoma by stratifying on genotypes at *TERF1* rs3863241 in pooled MELARISK and MDACC datasets

AFAP1L2 haplotypes			Genotypes of rs3863241 (*TERF1*)[3]					
			TT (844 cases/1,073 controls)		CT (1,404 cases/1,839 controls)		CC (607 cases/737 controls)	
rs649785[1]	rs597371[1]	Freq[2]	OR (95% CI)[4]	p[5]	OR (95% CI)[4]	p[5]	OR (95% CI)[4]	p[5]
G	A	0.49	–	–	–	–	–	–
C	G	0.38	1.31 (1.13–1.52)	2.9×10^{-4}	1.05 (0.93–1.17)	0.44	0.71 (0.59–0.85)	2.7×10^{-4}
C	A	0.10	1.05 (0.82–1.36)	0.70	0.97 (0.80–1.17)	0.73	0.78 (0.58–1.04)	0.09
G	G	0.03	0.91 (0.61–1.36)	0.65	0.73 (0.53–1.01)	0.06	1.19 (0.72–1.98)	0.49

[1]C is rs649785 minor allele; G is rs597371 minor allele.
[2]Haplotypes frequencies estimated in controls of the pooled sample. This frequency was similar in each stratum defined by *TERF1* rs3863241 genotypes.
[3]Haplotype analysis was conducted in each stratum of subjects according to their genotype at rs3863241 *(TERF1)*.
[4]OR (95% CI) is the odds ratio (95% confidence interval) for each haplotype with respect to the reference haplotype (GA).
[5]*p*-Value associated with Wald-test statistic.

AFAP1L2 pairs while the other two pairs (*PRKCE*/*TIAL1* and *TNFSF4*/*CMTM7*) showed a synergistic effect (Supporting Information Figs. 2a–2c).

Discussion

By testing for association on the basis of functional units such as GO categories, we identified five gene sets that were significantly enriched with genes associated with melanoma, the vast majority of which have not yet emerged from conventional single SNP analysis of GWAS. Epistasis analysis within these melanoma-associated pathways showed significant evidence for interaction for one SNP pair related to *TERF1* and *AFAP1L2* loci and suggestive evidence for another pair involving correlated SNPs at the same loci.

Three of the five GO categories that were identified in the two melanoma studies, response to light stimulus, regulation of mitotic cell cycle and induction of programmed cell death, included at most four loci previously reported by published GWASs[15,16] and that ranked high (top 1%) in MELARISK and/or MDACC GWASs. These loci included 16q24 (*MC1R*), 9p21 (*CDKN2B*, *CDKN2A*), 11q22 (*ATM*) and 2q33 (*CASP8*). However, all other genes driving the enrichment of these pathways in MELARISK and/or MDACC had not been reported to be robustly associated with melanoma risk. These three GOs involve biological processes that are known to be linked to melanoma development and progression. The other two GOs, cytokine activity and oxidative phosphorylation, represent novel pathways for melanoma risk. Only few candidate gene studies and one candidate pathway analysis have focused on immune-related genes and often have led to inconsistent results.[23] At the tumor level, a number of cytokines have been found to be highly expressed in human

melanomas.[24] Identifying the cytokine activity pathway as being associated with melanoma risk is of prime interest given the role of immune-related mechanisms in melanoma progression and the growing development of immunotherapy regimens.[25,26] Oxidative phosphorylation is an important energy-producing process in the mitochondria, and a key role of mitochondria in melanoma formation and progression has been highlighted.[27] However, polymorphisms of molecules involved in oxidative phosphorylation have not yet been reported to be associated with melanoma risk.

Our pathway analysis was based on the GSEA approach which has the major advantage of using a "no-cutoff" strategy that considers all genes from the genome without selecting significant genes according to a predetermined threshold as done by other methods such as ALIGATOR[28] and MAGENTA.[29] The choice of the boundaries within which SNPs are assigned to genes is not clearly defined. We strictly chose gene boundaries to decrease noise. However, when we extended the boundaries to 20 kb on each side of each gene, similar results were obtained (results not shown).

Epistasis analysis within the five melanoma-associated pathways uncovered two statistically significant SNP pairs at *TERF1* and *AFAP1L2* loci and three other pairs that did not meet the multiple-testing corrected thresholds but nominally replicated in MDACC. This is notable as very few of the increasing number of studies focusing on gene–gene interactions have been successful in detecting statistically significant interactions and/or showing replication. Although our analysis was based on imputed SNPs, we maximized imputation accuracy by using stringent QC criteria to retain SNPs for analysis. Moreover, the imputation quality scores of all SNPs showing interaction were greater than or equal to 0.9, and for those SNPs that were genotyped, the correlation between genotyped and imputed SNPs was at least 0.99. We assumed an additive model for both SNP main effects and interaction to increase power. Further analysis using a more general model including additive and dominance terms showed no significant departure from additivity for both main effects and interaction for all five SNP pairs that replicated in MDACC (results not shown).

The SNPs of the two statistically significant SNP pairs belonged to the loci that extended from 50 kb upstream to 50 kb downstream of *TERF1* and *AFAP1L2*, two genes that were part of the melanoma-associated regulation of mitotic cell cycle GO and were thus selected for epistasis analysis according to our two-step analysis strategy. However, these loci spanned other genes. We first checked that these other genes did not belong to the GO class to which *TERF1* and *AFAP1L2* were part of or to any other melanoma-associated GO. Investigation of the LD pattern in the critical intervals clearly pointed toward *TERF1* gene, although this was less clear for *AFAP1L2* given the presence of recombination hot spots around that gene. Investigation of functional annotations of the interacting SNPs, using Haploregv2,[30] showed that rs3863241 (*TERF1* locus) was in LD with two *TERF1* SNPs, rs1482029 and rs1574690 ($r^2 = 0.96$ and 0.60 with rs3863241, respectively), that map to regulatory elements including enhancer-like elements in epidermal keratinocytes and to transcription factor binding sites. The two SNPs at *AFAP1L2* locus were both in LD with one SNP, rs10626058 ($r^2 = 0.71$ with rs649785 and $r^2 = 0.59$ with rs597371), that maps to binding sites of transcription factors, including the forkhead box P1 transcription factor (FOXP1), which is known to be involved in various cancers. Both *TERF1* [telomeric repeat binding factor (NIMA-interacting) 1] and *AFAP1L2* (actin filament associated protein 1-like 2) genes have relevant biological functions regarding cancer and melanoma, in particular. *TERF1* is one the six genes encoding proteins of the shelterin complex[31] that mediate the interaction of telomerase with telomeres and are involved in many telomere functions.[32] The identification of this gene by our study is a major interest because there is accumulating evidence that telomere dysfunction represents a novel mechanism for melanoma development. Variants of telomerase reverse transcriptase (*TERT*) gene have been identified by GWASs of melanoma.[4,16] Rare germline mutations in *TERT* promoter[33] and in three shelterin complex genes, *POT1*, *ACD* and *TERF2IP*, have been recently found to predispose to familial melanoma.[34–36] Risk of melanoma was also reported to be associated with a set of genes, other than *TERF1*, that influence telomere length.[37] *AFAP1L2* gene (also known as *XB130*) encodes an adaptor molecule, a member of the actin filament-associated protein (AFAP) family, that is a binding partner of tyrosine kinases and promotes cell proliferation and survival, in particular *via* the PI3K/Akt pathway, known to play a key role in melanoma tumors.[38] AFAP1L2 protein was also found to be involved in cell migration and invasion *via* association with Rac-GTPases including Rac1.[39] It is of note that recurrent *RAC1* mutations have been found in melanoma tumors and mutated Rac1 protein promotes melanocyte proliferation and migration.[40] Furthermore, our finding of statistical interaction at *TERF1* and *AFAP1L2* loci is supported by the physical interaction between TERF1 and AFAP1L2 proteins.[41] Indeed, AFAP1L2 is one of the 320 proteins that was found, by a screening experiment of 12,000 proteins in human cells, to interact with the proteins of the shelterin complex, especially TERF1.[41]

In conclusion, this study shows that the proposed two-step analysis strategy allowed not only identifying novel pathways associated with melanoma risk but also pointing out strong candidates for melanoma susceptibility. Of particular interest is the significant interaction found between *TERF1* and *AFAP1L2*, which has important biological relevance, given the key role of TERF1 in telomere biology and its physical interaction with AFAP1L2 protein. It also brings a new piece of evidence for the emerging role of telomere dysfunction into melanoma development. Further studies, including functional and experimental studies, are needed to confirm the current findings. Thus, integration of biological knowledge together with investigation of gene–gene interactions represents a powerful approach to disentangle the complex mechanisms underlying multifactorial diseases such as cancers.

Acknowledgements

The authors thank the French Familial Study Group for their contribution to the MELARISK collection [Pascale Andry-Benzaquen, Bertrand Bachollet, Frédéric Bérard, Pascaline Berthet, Françoise Boitier, Valérie Bonadona, Brigitte Bressac-de Paillerets, Jean-Louis Bonafé, Jean-Marie Bonnetblanc, Frédéric Cambazard, Olivier Caron, Frédéric Caux, Jacqueline Chevrant-Breton, Agnès Chompret (deceased), Stéphane Dalle, Liliane Demange, Olivier Dereure, Martin-Xavier Doré, Marie-Sylvie Doutre, Catherine Dugast, Laurence Faivre, Florent Grange, Philippe Humbert, Pascal Joly, Delphine Kerob, Christine Lasset, Marie Thérèse Leccia, Gilbert Lenoir, Dominique Leroux, Julien Levang, Dan Lipsker, Sandrine Mansard, Ludovic Martin, Tanguy Martin-Denavit, Christine Mateus, Jean-Loïc Michel, Patrice Morel, Laurence Olivier-Faivre, Jean-Luc Perrot, Caroline Robert, Sandra Ronger-Savle, Bruno Sassolas, Pierre Souteyrand, Dominique Stoppa-Lyonnet, Luc Thomas, Pierre Vabres, Eva Wierzbicka]. The authors thank the Epidemiological Study on the Genetics and Environment of Asthma (EGEA) cooperative group for giving access to data of the EGEA study (see https://egeanet.vjf.inserm.fr). We acknowledge that the biological specimens of the French MELARISK Study were obtained from the Institut Gustave Roussy and Fondation Jean Dausset–CEPH Biobanks. They thank the Centre National de Génotypage (CNG-CEA, Evry, France) and Service XS (Leiden, the Netherlands) for performing genome-wide genotyping in the MELARISK study. They thank the John Hopkins University Center for Inherited Disease Research for conducting high-throughput genotyping and the University of Washington for the performance of quality control of the high-density SNP data of the MD Anderson Cancer Center cohort. M. Brossard was supported by doctoral fellowships from Ligue Nationale Contre le Cancer (LNCC) and Fondation pour la Recherche Médicale (FRM).

References

1. Nikolaou V, Stratigos AJ. Emerging trends in the epidemiology of melanoma. *Br J Dermatol* 2014; 170:11–9.
2. Goldstein AM, Chan M, Harland M, et al. Features associated with germline CDKN2A mutations: a GenoMEL study of melanoma-prone families from three continents. *J Med Genet* 2007; 44:99–106.
3. Puntervoll HE, Yang XR, Vetti HH, et al. Melanoma prone families with CDK4 germline mutation: phenotypic profile and associations with MC1R variants. *J Med Genet* 2013;50:264–70.
4. Law MH, Macgregor S, Hayward NK. Melanoma genetics: recent findings take us beyond well-traveled pathways. *J Invest Dermatol* 2012;132: 1763–74.
5. Barrett JH, Taylor JC, Bright C, et al. Fine mapping of genetic susceptibility loci for melanoma reveals a mixture of single variant and multiple variant regions. *Int J Cancer* 2015;136:1351–60.
6. Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet* 2007;81:1278–83.
7. Wang K, Zhang H, Kugathasan S, et al. Diverse genome-wide association studies associate the IL12/IL23 pathway with Crohn Disease. *Am J Hum Genet* 2009;84:399–405.
8. Zhang M, Liang L, Xu M, et al. Pathway analysis for genome-wide association study of basal cell carcinoma of the skin. *PLoS One* 2011;6:e22760.
9. Menashe I, Figueroa JD, Garcia-Closas M, et al. Large-scale pathway-based analysis of bladder cancer genome-wide association data from five studies of European background. *PLoS One* 2012; 7:e29396.
10. Kam-Thong T, Czamara D, Tsuda K, et al. EPIBLASTER-fast exhaustive two-locus epistasis detection strategy using graphical processing units. *Eur J Hum Genet* 2011;19:465–71.
11. Kam-Thong T, Azencott C-A, Cayton L, et al. GLIDE: GPU-based linear regression for detection of epistasis. *Hum Hered* 2012;73:220–36.
12. Sun X, Lu Q, Mukheerjee S, et al. Analysis pipeline for the epistasis search—statistical versus biological filtering. *Front Genet* 2014;5:106.
13. Chaudru V, Chompret A, Bressac-de Paillerets B, et al. Influence of genes, nevi, and sun sensitivity on melanoma risk in a family sample unselected by family history and in melanoma-prone families. *J Natl Cancer Inst* 2004;96:785–95.
14. Amos CI, Wang L-E, Lee JE, et al. Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. *Hum Mol Genet* 2011;20:5012–23.
15. Bishop DT, Demenais F, Iles MM, et al. Genome-wide association study identifies three loci associated with melanoma risk. *Nat Genet* 2009;41:920–5.
16. Barrett JH, Iles MM, Harland M, et al. Genome-wide association study identifies three new melanoma susceptibility loci. *Nat Genet* 2011;43:1108–13.
17. Li Y, Willer C, Sanna S, et al. Genotype Imputation. *Annu Rev Genomics Hum Genet* 2009;10: 387−406.
18. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* 2010;11:134.
19. Guo Y-F, Li J, Chen Y, et al. A new permutation strategy of pathway-based approach for genome-wide association study. *BMC Bioinformatics* 2009; 10:429.
20. Herold C, Steffens M, Brockschmidt FF, et al. INTERSNP: genome-wide interaction analysis guided by a priori information. *Bioinformatics* 2009;25:3275–81.
21. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 2005;95:221–7.
22. Tregouet DA, Garelle V. A new JAVA interface implementation of THESIAS: testing haplotype effects in association studies. *Bioinformatics* 2007; 23:1038–9.
23. Schoof N, Iles MM, Bishop DT, et al. Pathway-based analysis of a melanoma genome-wide association study: analysis of genes related to tumour-immunosuppression. *PLoS One* 2011;6:e29451.
24. Elias EG, Hasskamp JH, Sharma BK. Cytokines and growth factors expressed by human cutaneous melanoma. *Cancers* 2010;2:794−808.
25. Flaherty KT, Hodi FS, Fisher DE. From genes to drugs: targeted strategies for melanoma. *Nat Rev Cancer* 2012;12:349–61.
26. Wargo JA, Cooper ZA, Flaherty KT. Universes collide: combining immunotherapy with targeted therapy for cancer. *Cancer Discov* 2014;4:1377–86.
27. Theodosakis N, Micevic G, Kelly DP, et al. Mitochondrial function in melanoma. *Arch Biochem Biophys* 2014;563C:56–9.
28. Holmans P, Green EK, Pahwa JS, et al. Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. *Am J Hum Genet* 2009;85:13–24.
29. Segrè AV, Groop L, Mootha VK, et al. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet* 2010;6: e1001058.
30. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012;40:D930–D934.
31. De Lange T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev* 2005;19:2100–10.
32. Robles-Espinoza CD, Velasco-Herrera M del C, Hayward NK, et al. Telomere-regulating genes and the telomere interactome in familial cancers. *Mol Cancer Res MCR* 2015;13:211–22.
33. Horn S, Figl A, Rachakonda PS, et al. TERT promoter mutations in familial and sporadic melanoma. *Science* 2013;339:959–61.
34. Robles-Espinoza CD, Harland M, Ramsay AJ, et al. POT1 loss-of-function variants predispose to familial melanoma. *Nat Genet* 2014;46:478–81.
35. Shi J, Yang XR, Ballew B, et al. Rare missense variants in POT1 predispose to familial cutaneous malignant melanoma. *Nat Genet* 2014;46:482–6.
36. Aoude LG, Pritchard AL, Robles-Espinoza CD, et al. Nonsense mutations in the Shelterin complex genes ACD and TERF2IP in familial melanoma. *J Natl Cancer Inst* 2015;107:dju408.
37. Iles MM, Bishop DT, Taylor JC, et al. The effect on melanoma risk of genes previously associated with telomere length. *J Natl Cancer Inst* 2014;106: dju267.
38. Shiozaki A, Liu M. Roles of XB130, a novel adaptor protein, in cancer. *J Clin Bioinforma* 2011;1:10.
39. Lodyga M, Bai XH, Kapus A, et al. Adaptor protein XB130 is a Rac-controlled component of lamellipodia that regulates cell motility and invasion. *J Cell Sci* 2010;123:4156–69.
40. Krauthammer M, Kong Y, Ha BH, et al. Exome sequencing identifies recurrent somatic *RAC1* mutations in melanoma. *Nat Genet* 2012;44: 1006–14.
41. Lee O-H, Kim H, He Q, et al. Genome-wide YFP fluorescence complementation screen identifies new regulators for telomere signaling in human cells. *Mol Cell Proteomics* 2011;10:M110.001628.