Human Molecular Genetics, 2018, Vol. 27, No. 23 4145–4156

doi: 10.1093/hmg/ddy282
Advance Access Publication Date: 30 July 2018
Association Studies Article

ASSOCIATION STUDIES ARTICLE

Combining common genetic variants and non-genetic risk factors to predict risk of cutaneous melanoma

Fangyi Gu[1,2,†], Ting-Huei Chen[3,†], Ruth M. Pfeiffer[2], Maria Concetta Fargnoli[4], Donato Calista[5], Paola Ghiorzo[6], Ketty Peris[7], Susana Puig[8], Chiara Menin[9], Arcangela De Nicolo[10], Monica Rodolfo[11], Cristina Pellegrini[4], Lorenza Pastorino[6], Evangelos Evangelou[12,13], Tongwu Zhang[2], Xing Hua[2], Curt T. DellaValle[2], D. Timothy Bishop[14], Stuart MacGregor[15], Mark I. Iles[14], Matthew H. Law[15], Anne Cust[16], Kevin M. Brown[2], Alexander J. Stratigos[17], Eduardo Nagore[18], Stephen Chanock[2], Jianxin Shi[2], Melanoma Meta-Analysis Consortium, MelaNostrum Consortium and Maria Teresa Landi[2,*]

[1]Department of Cancer Prevention and Control, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA, [2]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA, [3]Department of Mathematics and Statistics, Laval University, Quebec, Canada, [4]Department of Dermatology, University of L'Aquila, L'Aquila, Italy, [5]Department of Dermatology, Maurizio Bufalini Hospital, Cesena, Italy, [6]Department of Internal Medicine and Medical Specialties, University of Genoa and Genetics of Rare Cancers, Ospedale Policlinico San Martino, Genoa, Italy, [7]Institute of Dermatology, Catholic University, Rome, Italy, [8]Dermatology Department, Melanoma Unit, Hospital Clínic de Barcelona, IDIBAPS, Universitat de Barcelona, Barcelona, Spain and Centro de Investigación Biomédica en Red en Enfermedades Raras (CIBERER), Valencia, Spain, [9]Department of Immunology and Molecular Oncology, Veneto Institute of Oncology IOV–IRCCS, Padua, Italy, [10]Cancer Genomics Program, Veneto Institute of Oncology IOV–IRCCS, Padua, Italy, [11]Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy, [12]Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece, [13]Department of Epidemiology and Biostatistics, Imperial College London, London, UK, [14]Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, University of Leeds, UK, [15]Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia, [16]Sydney School of Public Health, and Melanoma Institute Australia, The University of Sydney, Sydney, Australia, [17]1st Department of Dermatology–Venereology, National and Kapodistrian University of Athens School of Medicine, Andreas Sygros Hospital, Athens, Greece and [18]Department of Dermatology, Instituto Valenciano de Oncología, València, Spain

*To whom correspondence should be addressed at: Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, 9609 Medical Center Drive, Room 7E106, Bethesda, MD 20892-9769, USA. Tel: (240) 276-7236; Email: landim@mail.nih.gov

†Co-first authors.

Received: April 20, 2018. **Revised:** June 14, 2018. **Accepted:** July 24, 2018

Abstract

Melanoma heritability is among the highest for cancer and single nucleotide polymorphisms (SNPs) contribute to it. To date, only SNPs that reached statistical significance in genome-wide association studies or few candidate SNPs have been included in melanoma risk prediction models. We compared four approaches for building polygenic risk scores (PRS) using 12 874 melanoma cases and 23 203 controls from Melanoma Meta-Analysis Consortium as a training set, and newly genotyped 3102 cases and 2301 controls from the MelaNostrum consortium for validation. We estimated adjusted odds ratios (ORs) for melanoma risk using traditional melanoma risk factors and the PRS with the largest area under the receiver operator characteristics curve (AUC). We estimated absolute risks combining the PRS and other risk factors, with age- and sex-specific melanoma incidence and competing mortality rates from Italy as an example. The best PRS, including 204 SNPs (AUC = 64.4%; 95% confidence interval (CI) = 63–65.8%), developed using winner's curse estimate corrections, had a per-quintile OR = 1.35 (95% CI = 1.30–1.41), corresponding to a 3.33-fold increase comparing the 5th to the 1st PRS quintile. The AUC improvement by adding the PRS was up to 7%, depending on adjusted factors and country. The 20-year absolute risk estimates based on the PRS, nevus count and pigmentation characteristics for a 60-year-old Italian man ranged from 0.5 to 11.8% (relative risk = 26.34), indicating good separation.

Introduction

The incidence of cutaneous melanoma is increasing in western countries (1–3), with about 132 000 new cases worldwide each year. Melanoma is highly curable when detected in its earliest stages, with a 5-year survival rate of 98%. However, notwithstanding improved treatments in recent years (4–6), survival rates decline to 62% and 18% for regional and distant stage disease, respectively (2,7). Identifying subjects at high risk for melanoma is critical to provide targeted screening and early detection, and numerous melanoma risk prediction models have been built to facilitate this aim (8–20). Previous models mainly included environmental or host risk factors, such as age, family history, sun exposure, sunburns, number of melanocytic nevi and/or pigmentation characteristics. Several of these risk factors have a strong genetic component, and genetic factors are strongly implicated in the etiology of melanoma. Heritability for melanoma has been estimated to be 58%, among the highest for cancer (21). Rare high-risk variants in a few genes, such as *CDKN2A* (22), *CDK4* (23), *BAP1* (24), *TERT* (25), *POT1* (26,27), *ACD* (28) and *PARK2* (29) and variants with intermediate allele frequency (~1–5%), including variants in *MITF* (30), explain ~40% of familial melanoma, but account for a very small proportion of melanoma in the general population.

A large proportion of missing heritability is due to common genetic variants (31), which, when combined, may confer substantial risk. Genome-wide association studies (GWAS) of cutaneous melanoma have identified 20 genetic loci associated with melanoma risk to date (32), some of which are near genes related to pigmentation (*ASIP*, *SLC45A2*, *HERC2/OCA2*, *MC1R* and *TYR*) (33,34) and/or are associated with nevus count (*TERT*, *PLA2G6*, *CDKN2A-MTAP*, *IRF4*) (32,35,36). Building on these findings, a few previous reports of melanoma risk prediction models have combined 11 to 19 single nucleotide polymorphisms (SNPs) that reached genome-wide significance (37–39) or a few candidate SNPs with biological relevance (38).

A considerable proportion of phenotypic variation can be explained by the combination of genetic loci not achieving GWAS significance (40). In this study, we thoroughly explored models that included SNPs selected based on different criteria to build polygenic risk scores (PRS) that could capture the underlying genetic risk for melanoma. We used the largest meta-analysis of melanoma GWAS data to date from the Melanoma Meta-Analysis Consortium (MMAC) (32) as a training set and validated the performance of the PRS in newly genotyped subjects from Southern Europe, a population typically underrepresented in melanoma studies, from the MelaNostrum Consortium. We assessed the association of the PRS with melanoma risk, also adjusting for host/environmental melanoma risk factors. Finally, we built an absolute risk model for melanoma risk by combining relative risks for the PRS and other risk factors using the age- and gender-specific melanoma incidence rates and competing mortality rates from Italy as example. We identified a PRS including 204 SNPs that reached an area under the receiving operator characteristics curve (AUC) of 64.4%. The combination of this PRS and the traditional risk factors for melanoma (light hair color, light eye color, high sun sensitivity, large number of nevi as well as older age and male sex) strongly stratified subjects based on melanoma risk.

Results

Comparison of four models to estimate PRS using MMAC as a training dataset and MelaNostrum as the testing dataset

The characteristics of the MMAC training dataset are reported in Law *et al.* (32). The genotyping testing set from the MelaNostrum Consortium included 5599 subjects (3124 cases and 2475 controls) from Greece, Cyprus, Italy and Spain. Of this set, all the 194 subjects from Cyprus and two additional subjects had no phenotypic covariates and thus were excluded from the analyses including traditional melanoma risk factors. Thus, the MelaNostrum population (Table 1) included 775 melanoma cases and 752 controls from Greece, 1266 cases and 361 controls from Italy and 1061 cases and 1188 controls from Spain. Cases included more women than controls, were older, had lighter eye color and hair color, had lower skin photo-type and had more nevi. Subjects' characteristics by country and study site are presented in Supplementary Tables 1A and 1B.

The PRS in Model 1, with 17 genome-wide significant SNPs in MMAC (32) plus rs4778138 as proxy for rs7164220, achieved AUC = 62.8% (95% confidence interval (CI) = 61.4–64.3%) in the testing dataset. In model 2, the best AUC = 63.9% (62.5–65.4%) was achieved with the *P*-value threshold = 5×10^{-8} and $r^2 = 0.01$ for clumping. This model included 23 SNPs, comprising the 18 SNPs included in Model 1 plus five additional SNPs: four on chr.16 in the *MC1R* region and one on chr.9 in the *CDKN2A/MTAP*

Table 1. Characteristics of the MelaNostrum study population (n = 5403)

		Case N = 3102	%	Control N = 2301	%
Study site	Greece	775	25.0	752	32.7
	Italy	1266	40.8	361	15.7
	Spain	1061	34.2	1188	51.6
Sex[a]	Male	1453	46.8	1241	53.9
	Female	1649	53.2	1060	46.1
Age[b]	≤29	241	7.8	544	23.6
	30–39	494	15.9	576	25.0
	40–49	652	21.0	528	22.9
	50–59	636	20.5	319	13.9
	60–77	870	28.0	242	10.5
	≥78	143	4.6	26	1.1
	Missing	66	2.1	66	2.9
Family history of melanoma	No	1919	61.9	853	37.1
	Yes	227	7.3	159	6.9
	Missing	956	30.8	1289	56.0
Eye color[a]	Dark	1198	38.6	1262	54.8
	Medium	1065	34.3	644	28.0
	Light	575	18.5	241	10.5
	Missing	264	8.5	154	6.7
Hair color[a]	Black	323	10.4	342	14.9
	Dark brown/light/reddish brown	1874	60.4	1607	69.8
	Blond	486	15.7	147	6.4
	Red	126	4.1	37	1.6
	Missing	293	9.4	168	7.3
Skin phototype[a]	III-VI	1521	49.0	1250	54.3
	I-II	1349	43.5	779	33.9
	Missing	232	7.5	272	11.8
Nevi[a]	≤50	816	26.3	1143	49.7
	>50	1702	54.9	631	27.4
	Missing	584	18.8	527	22.9
Acute sun damage[c]	No	465	43.8	782	65.8
	Yes	521	49.1	334	28.1
	Missing	75	7.1	72	6.1
Chronic sun damage[c]	No	822	77.5	1089	91.7
	Yes	180	17.0	44	3.7
	Missing	59	5.6	55	4.6
Sunburns[c]	No	604	26.0	238	10.2
	Yes	1458	62.7	1123	48.2
	Missing	265	11.4	188	8.1
Intermittent sun exposure[c]	No/some	740	58.5	248	68.7
	High	429	33.9	83	23.0
	Missing	97	7.7	30	8.3
Chronic sun exposure[c]	No	1288	55.4	1108	71.5
	Yes	483	20.8	298	19.2
	Missing	556	23.9	143	9.2
Melanoma body site	Head/neck	347	11.2		
	Trunk	1254	40.4		
	Upper limbs	383	12.3		
	Lower limbs	703	22.7		
	Hands/feet	154	5.0		
	Unknown	212	6.8		
	Missing	49	1.6		
Melanoma type	SSM	1733	55.9		
	NM	365	11.8		
	LM	162	5.2		
	Acral	88	2.8		
	Mucosal	2	0.1		
	Undetermined	298	9.6		
	Missing	454	14.6		

(Continued)

Table 1. (Continued)

		Case N = 3102	%	Control N = 2301	%
Multiple melanoma	No	2564	82.7		
	Yes	342	11.0		
	Missing	196	6.3		
Thickness according to Breslow (mm)	<1.00	1060	34.2		
	1.01–2.00	440	14.2		
	2.01–4.00	335	10.8		
	>4.00	215	6.9		
	Undetermined	176	5.7		
	Missing	876	28.2		

[a]Variables included in all analyses.
[b]Age at diagnosis for cases and age at study enrollment for controls.
[c]Due to high missing rates in some studies, these variables were only evaluated, and therefore presented here, in subgroups of studies: acute and chronic sun damage is included in the Spanish study; intermittent sun exposure is included in the Italian study; sunburn and chronic sun exposure are included in both the Spanish and Italian studies.

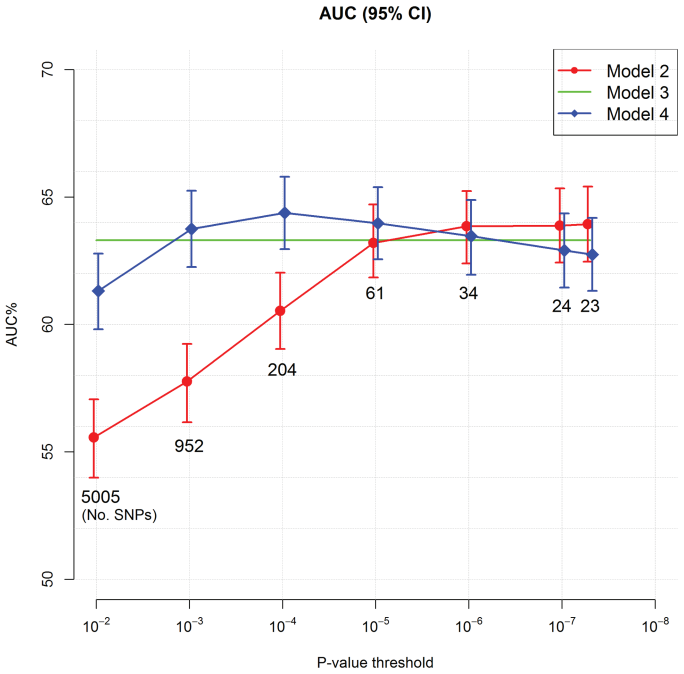

Figure 1. AUC and 95% CIs for three different models. Model 2 (see Methods section for details) used LD clumping $r^2 = 0.01$, and different P-value thresholds for SNP inclusion. Model 3 was constructed using LDPred (47). Model 4 is similar to Model 2 but corrects the effect size estimation for winner's curse (43). Model 1 is not represented in the Figure; it has AUC = 62.8%.

region. While keeping the linkage disequilibrium (LD) clumping criteria at $r^2 = 0.01$ and changing P-value thresholds from 5×10^{-8} up to 10^{-2} (Model 2), the corresponding AUC decreased steadily down to 55.6% (95% CI = 54.1–57.1%) for P-value = 10^{-2} (Fig. 1). Using LDPred (Model 3), the best AUC was 63.3% (95% CI = 60.8–65.4%). Model 4, correcting for the winner's curse bias and using LD clumping $r^2 = 0.01$, provided the PRS with the best performance at P-value threshold 10^{-4}. It included 204 SNPs and had AUC = 64.4% (95% CI = 63.0–65.8%) (Fig. 1). In the country-specific validation, the AUCs corresponding to the P-value 10^{-4} were 61.3%, 60.9% and 63.7% (95% CI = 61.4–66.0%) for the Greek, Italian and Spanish samples, respectively (Supplementary Table 2). As a sensitivity analysis, we reran the validation excluding all 196 subjects with missing phenotypic covariates to match the population used for the overall analyses and obtained the same 204 SNPs. The 204 SNPs in the PRS with P-value < 10^{-4} are listed in Supplementary Table 3, and the corresponding genotyping data can be found on github at this link: https://github.com/xtmgah/Melanoma_PRS.

Association between PRS and melanoma risk in the testing dataset considering well-established melanoma risk factors

Melanoma traditional risk factors were associated with melanoma risk in MelaNostrum data (Supplementary Table 4). The PRS with 204 SNPs was weakly, but significantly, correlated with nevus count and pigmentation variables in MelaNostrum controls overall and in country-specific analyses (Table 2).

Table 2. Correlation of PRS[a] and phenotypes in the MelaNostrum control population, overall and by country of residence

Phenotype	Corr	P	N
Overall			
Sex: 0 = male 1 = female	−0.01	0.55	2301
Age[b]	0.04	0.07	2235
Nevus count: 1 = ≤50; 2 = >50	0.13	<0.0001	1774
Eye color: 0 = dark, 1 = medium, 2 = light	0.09	<0.0001	2147
Hair color: 1 = black,2 = dark brown/light/reddish brown, 3 = blond,4 = red	0.14	<0.0001	2144
Skin phototype: 0 = III-VI; 1 = I-II	0.15	<0.0001	2029
Greece			
Sex: 0 = male 1 = female	−0.05	0.14	752
Age[b]	0.003	0.94	692
Nevus count: 1 = ≤50; 2 = >50	0.16	0.006	313
Eye color: 0 = dark, 1 = medium, 2 = light	0.1	0.02	634
Hair color: 1 = black,2 = dark brown/light/reddish brown, 3 = blond,4 = red	0.18	<0.0001	636
Skin phototype: 0 = III-VI; 1 = I-II	0.17	<0.0001	623
Italy			
Sex: 0 = male 1 = female	−0.02	0.74	361
Age[b]	−0.04	0.41	358
Nevus count: 1 = ≤50; 2 = >50	0.07	0.22	304
Eye color: 0 = dark, 1 = medium, 2 = light	0.09	0.10	354
Hair color: 1 = black,2 = dark brown/light/reddish brown, 3 = blond,4 = red	0.14	0.008	345
Skin phototype: 0 = III-IV; 1 = I-II	0.23	<0.0001	355
Sunburns: 0 = no; 1 = yes	0.04	0.53	249
Intermittent sun exposure: 0 = none/some; 1 = high	−0.04	0.53	331
Chronic sun exposure: 0 = no; 1 = yes	−0.04	0.50	249
Spain			
Sex: 0 = male 1 = female	0.01	0.83	1188
Age[b]	0.02	0.50	1185
Nevus count: 1 = ≤50; 2 = >50	0.07	0.02	1157
Eye color: 0 = dark, 1 = medium, 2 = light	0.03	0.26	1159
Hair color: 1 = black,2 = dark brown/light/reddish brown, 3 = blond,4 = red	0.11	0.0001	1152
Skin phototype: 0 = III-VI; 1 = I-II	0.11	0.0002	1095
Acute sun damage: 0 = no; 1 = yes	−0.01	0.81	1116
Chronic sun damage (actinic keratoses): 0 = no; 1 = yes	0.05	0.06	1133
Sunburns: 0 = no; 1 = yes	−0.007	0.81	1112
Chronic sun exposure: 0 = no; 1 = yes	−0.001	0.95	1157

[a]Continuous score based on the best winner's curse model.
[b]Age at diagnosis for cases and age at study enrollment for controls.

No correlation was observed with age, sex and sun exposure. The PRS was significantly associated with melanoma risk in the overall population and in each country separately (Table 3). The odds ratio (OR) per PRS quintile was 1.35 (95% CI = 1.30–1.41) in the overall population, which corresponds to a 3.3-fold increased melanoma risk comparing the highest versus the lowest PRS quintile. The ORs per PRS quintile were 1.31 (95% CI: 1.22–1.42) in Greece, 1.32 (95% CI: 1.21–1.43) in Italy and 1.40 (95% CI: 1.31–1.48) in Spain, corresponding to a 2.98-, 3.04- and 3.79-fold risk increase in the highest versus lowest PRS quintile, respectively. Adjusting for demographic factors did not substantially change the ORs, while adjusting for pigmentation factors and nevus count decreased the per quintile OR of PRS to 1.23 (95% CI = 1.13–1.35) in the overall population, and 1.29, 1.23 and 1.26 in Greece, Italy and Spain, respectively. Additionally adjusting for sun exposure–related variables for the Italian and Spanish samples did not affect the results (Table 3). There were no major differences in PRS–melanoma associations by categories of age, sex, nevus count, pigmentation or tumor characteristics (data not shown).

The AUC differences from models without and with PRS varied by country (Table 4). Adding the PRS improved the AUC by 7.3% in Italy and 2.0% in Spain (model with demographic factors); this reflects the age distribution: cases and controls had similar age in the Italian study, while controls were younger than cases in the Spanish study.

Absolute risk of developing melanoma in the Italian population

Absolute melanoma risk considering competing mortality risk showed substantial risk separation by different risk profiles in the Italian population aged 50, 60 and 70 years; risks ranged from 0.15% (0.16%) to 7.20% (3.66%) at 10 years and from 0.35% (0.29%) to 11.85% (7.10%) at 20 years in men (women) across different combinations of PRS and phenotype risk factors (Fig. 2A and B and Supplementary Table 5). For example, a 60-year old Italian man in the highest risk category (light eye color, red hair, I-II skin photo-type, 50+ nevi, 5th PRS quintile) had estimated 10-year and 20-year absolute melanoma risks of 5.38% and 11.76%, respectively, compared to 0.21% and 0.48% for a man of the same age in the lowest risk category (dark eye color, brown hair, III-VI skin photo-type, <50 nevi, 1st PRS quintile). Similar patterns were observed for women. The attributable risk of the PRS based on the relative risk estimates from the cases was 0.26 in the Italian population.

Table 3. ORs between PRS and melanoma risk, adjusting for different melanoma risk factors

	$OR_{per\ quintile}$	L95	U95	P	$OR_{5th\ versus\ 1st\ quintile}$
Overall					
PRS	1.35	1.30	1.41	<0.0001	3.33
PRS + Demographics[a]	1.35	1.29	1.41	<0.0001	3.30
PRS + Demographics + pigmentation[b] + nevi	1.23	1.13	1.35	<0.0001	2.32
Greece					
PRS	1.31	1.22	1.42	<0.0001	2.98
PRS + Demographics[a]	1.33	1.23	1.44	<0.0001	3.11
PRS + Demographics + pigmentation[b] + nevi	1.29	1.19	1.40	<0.0001	2.76
Italy					
PRS	1.32	1.21	1.43	<0.0001	3.04
PRS + Demographics[a]	1.32	1.21	1.44	<0.0001	3.02
PRS + Demographics + pigmentation[b] + nevi	1.23	1.13	1.35	0.0003	2.32
PRS + Fully adjusted[c]	1.23	1.12	1.35	<0.0001	2.29
Spain					
PRS	1.40	1.31	1.48	<0.0001	3.79
PRS + Demographics[a]	1.38	1.29	1.48	<0.0001	3.63
PRS + Demographics + pigmentation[b] + nevi	1.26	1.16	1.37	<0.0001	2.55
PRS + Fully adjusted[d]	1.27	1.17	1.38	<0.0001	2.62

[a]Demographic includes age, sex and country (for overall population).
[b]Pigmentation includes eye color, hair color and skin phototype.
[c]Full model in the Italian population additionally adjusted for chronic sun exposure, intermittent sun exposure and history of sunburns.
[d]Full model in the Spanish population additionally adjusted for chronic sun exposure, chronic sun damage, acute sun damage and history of sunburns.

Table 4. Performance of risk prediction model with and without PRS

Traditional covariates in models	AUC (95% CI) Model without PRS	AUC (95% CI) Model with PRS	AUC difference	P-difference
Overall				
Demographic[a]	76.5% (75.2–77.8%)	78.2% (77.0–79.4%)	1.7% (1.1–2.2%)	<0.0001
Demographic + pigmentation[b] + nevi	80.1% (78.9–81.3%)	81.0% (79.8–82.2%)	0.8% (0.5–1.2%)	<0.0001
Greece				
Demographic[a]	67.9% (65.2–70.7%)	70.7% (68.0–73.4%)	2.7% (1.3–4.1%)	0.0002
Demographic + pigmentation[b] + nevi	69.8% (67.1–72.5%)	71.7% (69.1–74.4%)	1.9% (0.7–3.1%)	0.003
Italy				
Demographic[a]	53.9% (50.6–57.2%)	61.2% (57.8–64.5%)	7.3% (3.4–11.2%)	0.0001
Demographic + pigmentation[b] + nevi	64.8% (61.6–68.1%)	66.6% (63.4–69.8%)	1.7% (0.6–3.0%)	0.04
Fully adjusted[c]	67.0% (63.7–70.3%)	68.5% (65.4–71.7%)	1.4% (-0.1–2.9%)	0.07
Spain				
Demographic[a]	78.6% (76.7–80.5%)	80.6% (78.8–82.4%)	2.0% (1.2–2.8%)	<0.0001
Demographic + pigmentation[b] + nevi	87.7% (86.3–89.3%)	88.3% (86.8–89.7%)	0.5% (0.1–0.8%)	0.005
Fully adjusted[d]	88.7% (87.3–90.1%)	89.1% (87.6–90.5%)	0.4% (0.1–0.7%)	0.005

[a]Demographic includes age, sex and country (for overall population).
[b]Pigmentation includes eye color, hair color and skin phototype.
[c]Full model in the Italian population additionally adjusted for chronic sun exposure, intermittent sun exposure and history of sunburns.
[d]Full model in the Spanish population additionally adjusted for chronic sun exposure, chronic sun damage, acute sun damage and history of sunburns.

Discussion

We report on a PRS for melanoma risk that combines 204 common SNPs and had an AUC of 64.4%. This PRS was obtained using a model that corrected for the winner's curse bias in SNP effect size estimates. Based on the PRS, subjects in the highest quintile had ~2.5-fold risk of melanoma compared to those in the lowest quintile, after adjusting for other major melanoma risk factors. Although not directly comparable, a 2.5- to 3-fold increased risk of melanoma is equivalent or even stronger than the risk of very severe solar damage (10), family history, gender and many pigmentation and UV-related risk factors (10,41). This PRS, in combination with pigmentation characteristics and number of nevi, strongly differentiated melanoma risk in the Italian population and thus could be useful towards identifying high-risk subjects who could potentially benefit from increased surveillance.

Optimal *P*-value threshold to select SNPs for disease risk prediction depends on the number of causal SNPs and their effect size distribution, and the sample size of the training data set (40,42). Accordingly, we thoroughly explored models that included SNPs based on different selection criteria to build PRS that could capture the underlying genetic risk for melanoma. We used a very large training data to maximize the accuracy of the PRS. The AUC (64.4%) of the best PRS is larger than the PRS-based AUCs for other cancers using the largest GWAS summary data, such as the AUC for lung (56.4%), colorectal (57.4%), pancreatic (58.7%) (43) or breast (61.5%) (44) cancers. It is only slightly smaller than the AUC (65.4%) for prostate cancer (43), which was

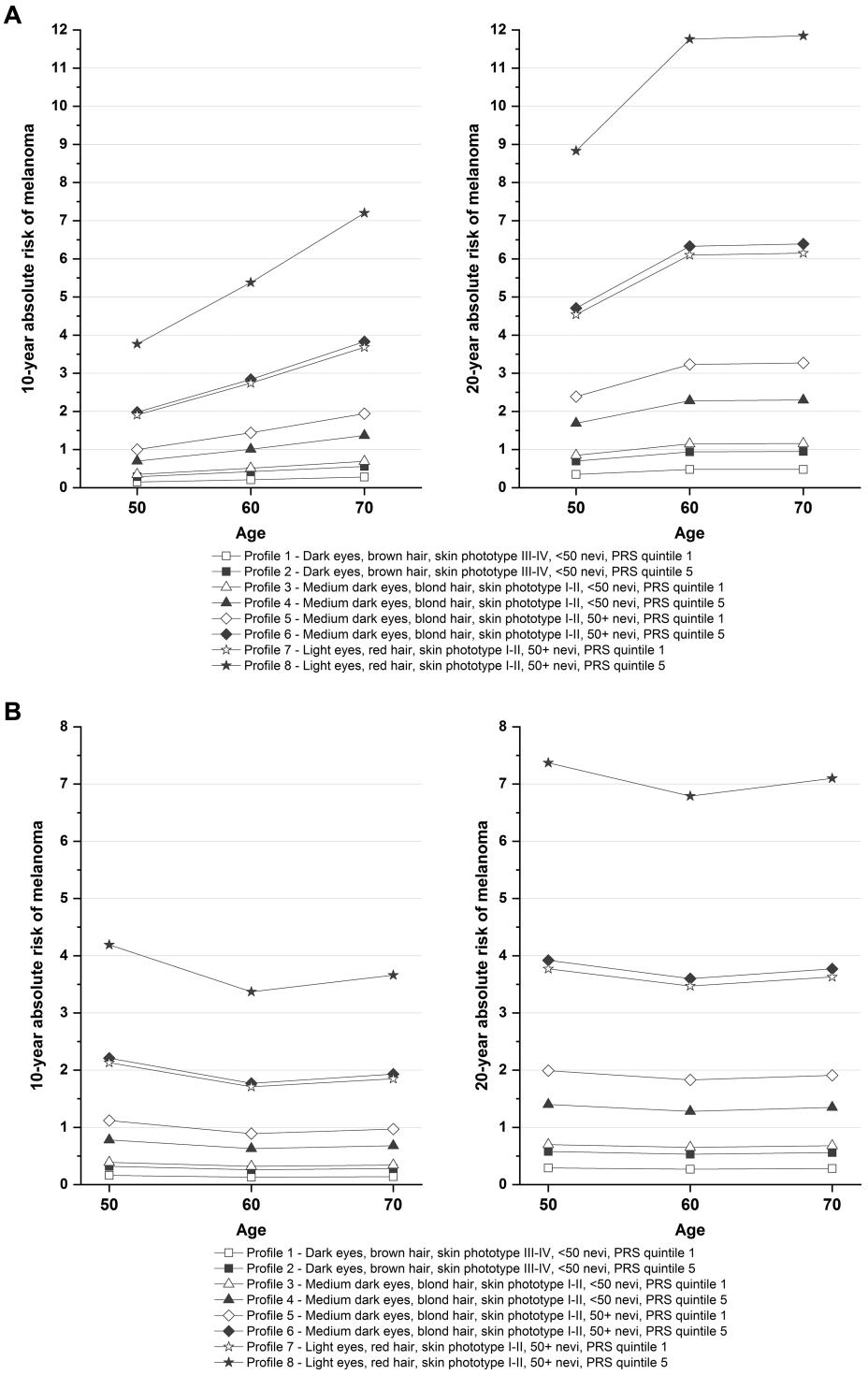

Figure 2. 10- and 20-year absolute risk of melanoma for Italian men (Fig. 2A) and women (Fig. 2B), by age and risk profile. The absolute risk was estimated in a model that includes the PRS and other established risk factors, using age- and sex-specific incidence rates of melanoma as well as death rates of other causes from the Italian population. Corresponding risks are also shown in Supplementary Table 5.

obtained using a training dataset three-times larger than the one for melanoma. These results are consistent with the heritability estimates across cancers, which are highest for melanoma (58%, 95% CI = 43–73%) and prostate cancer (57%, 95% CI = 51–63%) (21). Absolute risk estimates for melanoma combining PRS and the other melanoma risk factors stratified Italian subjects very well into high- and low-risk groups, suggesting potential application of PRS in melanoma precision prevention. We used the Italian population because we could obtain age- and sex-specific incidence and mortality rates from cancer registries (AIRTUM) (53,54), which were not available for Spain and Greece, and we had data on the traditional risk factors for

this study population. Moreover, we wanted to investigate the range of estimated absolute risks in a country without routine melanoma screening, where people are not perceived to be at high risk for the disease, and so this model could constitute an important tool for melanoma prevention. Similar calculation can be conducted for other countries using their own age- and gender-specific melanoma incidence and mortality rates. Since the absolute disease risk for short prediction intervals (e.g. 5 years) is proportional to the relative risk multiplied by the age-specific baseline incidence, the PRS effect on absolute risk estimates could be substantially stronger in populations with higher melanoma incidence rates, including Australia and Northern European countries.

Several melanoma risk factors have a genetic component, and the PRS, including SNPs at pigmentation- (e.g. *SLC45A2* or *MC1R*) or nevus-associated (e.g. *MTAP*) loci, was weakly correlated with both pigmentation characteristics and nevus count. Overall, the AUC improvement provided by the PRS over traditional risk factors ranged from 0.8 to 1.7% depending on the variables in the models, with some variability also due to the different study designs across the countries. When only age and sex were included in the models, adding the PRS improved the AUC, particularly in the Italian population where cases and controls were matched on age. However, when pigmentation and nevi variables were added, the improvement was reduced overall and for all countries. The impact of the PRS on absolute risk was more noticeable, leading to a doubling of absolute risk for each profile when changing the PRS quintile from the lowest to the highest. This was particularly meaningful for older men, who had the highest melanoma incidence rate in the Italian population. We could not test the effect of PRS in subjects with or without family history of melanoma since few studies collected this information. To avoid oversampling for family history that could bias the PRS effect estimates, we specifically excluded studies that were sampled based on family history.

Since the training data mostly included subjects from Northern Europe, Australia and the US and the validation set included subjects from Southern European countries (MelaNostrum), we evaluated whether the PRS could be useful across different populations. The model performance could be affected by the effect size (i.e. the OR) of the SNPs in the PRS and the variant allele frequency of the genes included in the PRS. We checked whether the effect sizes of each of the 204 SNPs in the best PRS differed between the training set and MelaNostrum subjects (Supplementary Table 6). The large majority of the SNPs had a similar effect size across populations; only three SNPs (rs75286671 at chr.4, rs187989493 at chr.7 and rs139791480 at chr.6) reached a statistically significant difference (P-value $< 2.45 \times 10^{-4}$). However, as expected, some SNPs in pigmentation-associated loci, such as rs7164220 around *HERC2*, rs250417 around *SCLC45A2* and rs1805008 around *MC1R*, had different minor allele frequencies between the training set and MelaNostrum (minor allele frequency = 0.119 versus 0.246; 0.03 versus 0.09; 0.08 versus 0.02, respectively). Thus, the PRS effect estimates can be transferred to other countries of European ancestry, but the ability to discriminate subjects at high or low risk for the disease could vary across different populations.

This study has many strengths. For building the PRS, we used the largest melanoma GWAS data to date as a training set, a major determinant of the accuracy of PRS prediction (40). We thoroughly explored different SNP selection criteria and statistical approaches, and chose one with the largest AUC to build the PRS. We genotyped for the first time many subjects from Mediterranean populations, typically underrepresented in melanoma studies, for independent validation. We also studied the impact of PRS with and without traditional risk factors for melanoma using various models. Finally, we estimated the absolute risk of melanoma for Italian subjects with different risk profiles and combinations of PRS.

Some limitations should also be noted: we lack prospective cohort data for model calibration, which would be ideal for the direct application of the risk prediction model to the public health or clinical setting. However, when we tested the fit of the relative risk model that was the basis of the absolute risk predictions using different approaches as proposed by Song *et al.* (45), none of the tests indicated lack of fit of the model (P-values ranging from 0.08 to 0.78, using 10 000 simulations). Thus, we conclude the relative risk model has adequate fit to the Italian case-control data. An additional limitation is that there was an upward bias for AUC estimate in Models 2, 3 and 4 with a single tuning parameter, because the validation dataset was used for both selecting the tuning parameter and calculating AUC. Such bias is minimal (typically less than 0.15%), as we have shown on simulation studies (43). Moreover, while we conducted imputation for missing data in pigmentation and nevi variables (about 10% and 20% of overall subjects), we had to exclude some traditional risk factors (e.g. family history) from the models because of larger missing data from some studies. Finally, there was heterogeneity among the contributing studies in study design and data collection, e.g. controls in some Spanish and Greek studies were younger than cases, while cases and controls from the Italian studies were matched on age. This discrepancy can explain the differences in the performance of the risk prediction model when including only the demographic variables with the PRS (Table 4). However, we saw no evidence of heterogeneity in SNPs' ORs among studies, suggesting that SNP and PRS estimates should be broadly applicable. Moreover, the absolute risk model is not affected by this issue because we only used the Italian studies which were age-matched.

Our study suggests that PRS, in combination with traditional melanoma risk factors, may help identify subjects who could benefit from heightened skin examination and sun avoidance. Prospective analyses of the PRS together with other melanoma risk factors are needed to validate the overall accuracy of risk prediction in Mediterranean and other countries. We expect that risk models combining genetic and non-genetic risk factors will be further improved when larger genetic studies become available in the future.

Materials and Methods

Study population and genotyping

Our PRS was constructed using summary level data from a GWAS meta-analysis from the MMAC (32), including 11 GWAS from Europe, Australia and the US, totaling 12 874 melanoma cases and 23 203 controls. The details of the study population, genotyping and quality control information are published previously (32).

We validated our PRS using independent GWAS data from the MelaNostrum consortium, formed by clinicians and researchers from institutions dedicated to melanoma management in Mediterranean countries. MelaNostrum included cases with histologically confirmed primary cutaneous melanoma and participants who were melanoma-free at study entry from Italy, Spain, Greece and Cyprus. Details of the design, data collection and genotyping methods are presented in the Online Data Supplement. All participants signed an informed consent, and the study was reviewed by institutional review boards of the

local hospitals and the National Cancer Institute. After quality control, 5599 subjects (3124 cases and 2475 controls) and 707 169 SNPs were used as a validation set for the PRS. Of the 5599 subjects, 194 subjects from Cyprus and two additional subjects had no phenotypic covariates and thus were excluded from the additional analyses including traditional melanoma risk factors. Thus, the total number of subjects for the overall analyses included 5403 subjects (3102 cases, 2301 controls) from Italy, Spain and Greece. Characteristics of the study population are summarized in Table 1 and Supplementary Tables 1A and 1B.

Statistical analyses

PRS computation. We built PRS using four methods based on ORs ($\widehat{OR}_t$) or equivalently $\widehat{\beta}_t = \log(\widehat{OR}_t)$, and P-values p_t from logistic regression analysis fit to each SNP individually in MMAC (32) (the training data).

The first PRS (Model 1) included only $K = 18$ SNPs achieving genome-wide significance in MMAC. Note that, for each locus, only the most significant SNPs were selected into the PRS. For each subject i in the validation dataset, the PRS was then calculated as

$$\mathrm{PRS_i} = \sum_{t=1}^{K} \widehat{\beta}_t g_{it},$$

where g_{it} is the genotypic value for SNP t for subject i.

The second PRS (Model 2) used different P-value thresholds for SNP inclusion (46). Briefly, we first performed LD clumping with PLINK (47) using correlation $r^2 = 0.01$ and window size 5 Mb, guided by the SNP P-values in the training data. Sensitivity analysis was performed using $r^2 = 0.1$, 0.2 and 0.3. Assuming there are M SNPs after LD clumping, the PRS for subject i with P-value threshold p is

$$\mathrm{PRS_i}(p) = \sum_{t=1}^{M} \widehat{\beta}_t g_{it} I\left(\mathrm{P_t} \le p\right),$$

where $I = 1$ if $P_t \le p$ and $I = 0$ otherwise, and the P-value threshold was chosen as 5×10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3} and 10^{-2}. The optimal P-value threshold was the one that maximized the prediction performance in the validation sample.

The third PRS (Model 3) was constructed using LDPred (48). LDPred includes all analyzed SNPs while re-estimating the effect size β_t as the posterior mean by conditioning on the marginal effect size estimates for all SNPs and LD information in a local region. Compared to the other models that require LD clumping, LDPred may have better performance when multiple SNPs in a local region are independently associated with the phenotype.

Finally, the fourth method (Model 4) is similar to Model 2 but corrects the effect size estimation for winner's curse, i.e. the fact that effect estimates for SNP selected based on having small P-values are upwardly biased. We recently demonstrated that correcting for this bias can improve the predictive performance of PRS (43). Following this approach, we bias-corrected the SNP specific estimates $\widehat{\beta}_t$, to obtain

$$\widehat{\beta}_t^{wcc}(p) = \mathrm{sign}\left(\widehat{\beta}_t\right)\left(\left|\widehat{\beta}_t\right| - \lambda(p)\right) I\left(\left|\widehat{\beta}_t\right| > \lambda(p)\right),$$

where $\lambda(p)$ depends on the P-value threshold p : $\lambda(p) = \Phi^{(-1)}(1 - \frac{p}{2})\widehat{\sigma}_t$. Here, $\Phi()$ is the probability distribution function for a standard normal distribution.

The rs4778138 SNP was reported as significant in MMAC but was not imputed well in MelaNostrum; thus, we included rs7164220 (LD $R^2 = 0.6$ with rs4778138) in all models even if it did not achieve the required significance level.

We evaluated the prediction performance of the four PRS scores in the MelaNostrum GWAS (the testing data) by calculating the AUC using the R package 'pROC' (49) with bootstrap CIs.

Contribution of PRS on melanoma risk prediction. We assessed the association of the PRS with the best predictive performance (coded in quintiles) with melanoma risk, alone and with additional risk factors, and evaluated its performance in risk prediction in the MelaNostrum data.

We imputed traditional risk factors, allowing for interactions with case status. The variables were assumed to be categorical and included: age at diagnosis for cases or at study enrollment for controls, eye color (dark, medium, light), hair color (black, dark brown/light/reddish brown, blond, red), intermittent sun exposure (none/some, high), sunlamp use (yes, no), actinic keratosis (yes, no), chronic sun exposure (yes, no), skin type (I-II, III-VI) and sunscreen use (yes, no). We did not impute missing family history and did not use this information in the model. The imputation was conducted using IVEware (50), and we analyzed the M = 5 imputed datasets, accounting for the random imputation in the variance computation using PROC MIANALYZE (Inc. SI. SAS 9.3. Cary, NC2011) (51). The largest amount of missingness was seen for sunscreen use (57.76%, excluded from the model); eye and hair color had ≤15% missing data. We observed no substantial differences in our findings when we excluded individuals with missing values in a sensitivity analysis (data not shown).

ORs and 95% CIs for association were calculated using logistic regression models (PROC Logistic, SAS 9.3). PRS quintiles were coded as an ordinal variable. We used data harmonized across the different studies and countries to adjust the PRS models. Specifically, models were 1) not adjusted; 2) adjusted for demographic factors only (age, sex, country of residence: Greece, Italy, Spain); 3) adjusted for demographic factors, pigmentation variables (eye color, hair color, skin phototype) and nevus count. Models adjusted for linear combinations of pigmentation characteristics obtained using factor analysis (52) yielded similar estimates and are thus not shown. We included an age × study site interaction term in the models to accommodate different age distributions across studies. We computed two-sided P-values using Wald tests; P-value < 0.05 was considered statistically significant.

We also stratified all analyses by country of residence. We further adjusted Italian models for chronic sun exposure, intermittent sun exposure and history of sunburns, and Spanish models for chronic sun exposure, chronic sun damage, acute sun damage and history of sunburns.

Contributions of PRS to prediction performance were evaluated by the difference of AUC between models with and without PRS, computed based on cross-validation, overall and by country.

Projecting probabilities (absolute risk) of developing melanoma in Italian subjects. The absolute risk r*(a,b) of melanoma in the age interval (a,b) is the probability of developing melanoma during that interval, given that one is alive and free of previous melanoma at age a,

$$r^*(a,b) = \int_a^b \lambda_m(t,x)\exp\left(-\int_a^t \lambda_m(u,x) + \lambda_D(u,x)\,du\right)dt. \quad (1)$$

The melanoma hazard rate λ_m was modeled as $\lambda_m(a,x) = (1\text{-}AR(x))\ \exp(\beta x)\ \lambda^*(a)$ as the product of one minus the age- and sex-specific attributable risk for all the risk factors in the model, the relative risk, $\exp(\beta x)$, that includes covariates x, and age- and sex-specific incidence rates from ITACAN, http://itacan.ispo.toscana.it, pooling data from 38 Italian cancer registries in 2009. For details see Pfeiffer and Gail (53), Chapter 5. The competing deaths hazard λ_D was estimated by subtracting 5-year age and sex-specific mortality rates for melanoma from 5 year-age and sex-specific all-cause mortality rates from ITACAN.

The attributable risk of the PRS was estimated using the Bruzzi formula (54).

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

Authors thank the study participants and the members of the GenoMEL and MelaNostrum consortia (Supplementary Notes) who made this study possible. We thank David Check for the help in developing figures for the manuscript. Details of the Melanoma Meta-Analysis Consortium and MelaNostrum Consortium can be found in Online Data Supplement.

Conflict of Interest statement. None declared.

Funding

AQ1

Intramural Research Program of National Institutes of Health, National Cancer Institute, Division of Cancer Epidemiology and Genetics; Funding for the Melanoma Meta-Analysis Consortium study: Cancer Research UK Programme Award (C588/A19167) and the US National Institutes of Health (NIH; RO1-CA-083115-06); Roswell Park Comprehensive Cancer Center and National Cancer Institute (NCI) (P30CA016056 to F.G.); NHMRC and Cancer Institute NSW Fellowships (to A.C.); The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

References

1. Christenson, L.J., Borrowman, T.A., Vachon, C.M., Tollefson, M.M., Otley, C.C., Weaver, A.L. and Roenigk, R.K. (2005) Incidence of basal cell and squamous cell carcinomas in a population younger than 40 years. *JAMA*, **294**, 681–690.
2. Cancer Facts & Figures. (2017). American Cancer Society, pp. 24–25.
3. Weinstock, M.A. (2001) Epidemiology, etiology, and control of melanoma. *Med. Health R. I.*, **84**, 234–236.
4. Long, G.V., Hauschild, A., Santinami, M., Atkinson, V., Mandala, M., Chiarion-Sileni, V., Larkin, J., Nyakas, M., Dutriaux, C., Haydon, A. *et al.* (2017) Adjuvant dabrafenib plus trametinib in stage III BRAF-mutated melanoma. *New Engl. J. Med.*, **377**, 1813–1823.
5. Weber, J., Mandala, M., Del Vecchio, M., Gogas, H.J., Arance, A.M., Cowey, C.L., Dalle, S., Schenker, M., Chiarion-Sileni, V., Marquez-Rodas, I. *et al.* (2017) Adjuvant nivolumab versus ipilimumab in resected stage III or IV melanoma. *New Engl. J. Med.*, **377**, 1824–1835.
6. Wolchok, J.D., Chiarion-Sileni, V., Gonzalez, R., Rutkowski, P., Grob, J.J., Cowey, C.L., Lao, C.D., Wagstaff, J., Schadendorf, D., Ferrucci, P.F. *et al.* (2017) Overall survival with combined nivolumab and ipilimumab in advanced melanoma. *New Engl. J. Med.*, **377**, 1345–1356.
7. Gershenwald, J.E., Scolyer, R.A., Hess, K.R., Sondak, V.K., Long, G.V., Ross, M.I., Lazar, A.J., Faries, M.B., Kirkwood, J.M., McArthur, G.A. *et al.* (2017) Melanoma staging: evidence-based changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J. Clin.*, **67**, 472–492.
8. Cho, E., Rosner, B.A., Feskanich, D. and Colditz, G.A. (2005) Risk factors and individual probabilities of melanoma for whites. *J. Clin. Oncol.*, **23**, 2669–2675.
9. Davies, J.R., Chang, Y.M., Bishop, D.T., Armstrong, B.K., Bataille, V., Bergman, W., Berwick, M., Bracci, P.M., Elwood, J.M., Ernstoff, M.S. *et al.* (2015) Development and validation of a melanoma risk score based on pooled data from 16 case-control studies. *Cancer Epidemiol. Biomarkers Prev.*, **24**, 817–824.
10. Fears, T.R., Guerry, D.t., Pfeiffer, R.M., Sagebiel, R.W., Elder, D.E., Halpern, A., Holly, E.A., Hartge, P. and Tucker, M.A. (2006) Identifying individuals at high risk of melanoma: a practical predictor of absolute risk. *J. Clin. Oncol.*, **24**, 3590–3596.
11. Fortes, C., Mastroeni, S., Bakos, L., Antonelli, G., Alessandroni, L., Pilla, M.A., Alotto, M., Zappala, A., Manoorannparampill, T., Bonamigo, R. *et al.* (2010) Identifying individuals at high risk of melanoma: a simple tool. *Eur. J. Cancer Prev.*, **19**, 393–400.
12. Guther, S., Ramrath, K., Dyall-Smith, D., Landthaler, M. and Stolz, W. (2012) Development of a targeted risk-group model for skin cancer screening based on more than 100,000 total skin examinations. *J. Eur. Acad. Dermatol. Venereol.*, **26**, 86–94.
13. MacKie, R.M., Freudenberger, T. and Aitchison, T.C. (1989) Personal risk-factor chart for cutaneous melanoma. *Lancet*, **2**, 487–490.
14. Mar, V., Wolfe, R. and Kelly, J.W. (2011) Predicting melanoma risk for the Australian population. *Australas. J. Dermatol.*, **52**, 109–116.
15. Molinaro, A.M., Ferrucci, L.M., Cartmel, B., Loftfield, E., Leffell, D.J., Bale, A.E. and Mayne, S.T. (2015) Indoor tanning and the MC1R genotype: risk prediction for basal cell carcinoma risk in young people. *Am. J. Epidemiol.*, **181**, 908–916.
16. Olsen, C.M., Neale, R.E., Green, A.C., Webb, P.M., The, Q.S., The Epigene, S and Whiteman, D.C. (2015) Independent validation of six melanoma risk prediction models. *J. Investig. Dermatol.*, **135**, 1377–1384.
17. Quereux, G., Moyse, D., Lequeux, Y., Jumbou, O., Brocard, A., Antonioli, D., Dreno, B. and Nguyen, J.M. (2011) Development of an individual score for melanoma risk. *Eur. J. Cancer Prev.*, **20**, 217–224.
18. Schuchter, L., Schultz, D.J., Synnestvedt, M., Trock, B.J., Guerry, D., Elder, D.E., Elenitsas, R., Clark, W.H. and Halpern, A.C. (1996) A prognostic model for predicting 10-year survival in patients with primary melanoma. The Pigmented Lesion Group. *Ann. Intern. Med.*, **125**, 369–375.
19. Sneyd, M.J., Cameron, C. and Cox, B. (2014) Individual risk of cutaneous melanoma in New Zealand: developing a clinical prediction aid. *BMC Cancer*, **14**, 359.
20. Williams, L.H., Shors, A.R., Barlow, W.E., Solomon, C. and White, E. (2011) Identifying persons at highest risk of melanoma using self-assessed risk factors. *J. Clin. Exp. Dermatol. Res.*, **2**.
21. Mucci, L.A., Hjelmborg, J.B., Harris, J.R., Czene, K., Havelick, D.J., Scheike, T., Graff, R.E., Holst, K., Moller, S., Unger, R.H. *et al.* (2016) Familial risk and heritability

of cancer among twins in Nordic countries. *JAMA*, **315**, 68–76.
22. Goldstein, A.M., Chan, M., Harland, M., Hayward, N.K., Demenais, F., Bishop, D.T., Azizi, E., Bergman, W., Bianchi-Scarra, G., Bruno, W. *et al.* (2007) Features associated with germline CDKN2A mutations: a GenoMEL study of melanoma-prone families from three continents. *J. Med. Genet.*, **44**, 99–106.
23. Zuo, L., Weger, J., Yang, Q., Goldstein, A.M., Tucker, M.A., Walker, G.J., Hayward, N. and Dracopoli, N.C. (1996) Germline mutations in the p16INK4a binding domain of CDK4 in familial melanoma. *Nat. Genet.*, **12**, 97–99.
24. Wiesner, T., Obenauf, A.C., Murali, R., Fried, I., Griewank, K.G., Ulz, P., Windpassinger, C., Wackernagel, W., Loy, S., Wolf, I. *et al.* (2011) Germline mutations in BAP1 predispose to melanocytic tumors. *Nat. Genet.*, **43**, 1018–1021.
25. Horn, S., Figl, A., Rachakonda, P.S., Fischer, C., Sucker, A., Gast, A., Kadel, S., Moll, I., Nagore, E., Hemminki, K. *et al.* (2013) TERT promoter mutations in familial and sporadic melanoma. *Science*, **339**, 959–961.
26. Robles-Espinoza, C.D., Harland, M., Ramsay, A.J., Aoude, L.G., Quesada, V., Ding, Z., Pooley, K.A., Pritchard, A.L., Tiffen, J.C., Petljak, M. *et al.* (2014) POT1 loss-of-function variants predispose to familial melanoma. *Nat. Genet.*, **46**, 478–481.
27. Shi, J., Yang, X.R., Ballew, B., Rotunno, M., Calista, D., Fargnoli, M.C., Ghiorzo, P., Bressac-de Paillerets, B., Nagore, E., Avril, M.F. *et al.* (2014) Rare missense variants in POT1 predispose to familial cutaneous malignant melanoma. *Nat. Genet.*, **46**, 482–486.
28. Aoude, L.G., Pritchard, A.L., Robles-Espinoza, C.D., Wadt, K., Harland, M., Choi, J., Gartside, M., Quesada, V., Johansson, P., Palmer, J.M. *et al.* (2015) Nonsense mutations in the shelterin complex genes ACD and TERF2IP in familial melanoma. *J. Natl. Cancer Ins.*, **107**.
29. Hu, H.H., Kannengiesser, C., Lesage, S., Andre, J., Mourah, S., Michel, L., Descamps, V., Basset-Seguin, N., Bagot, M., Bensussan, A. *et al.* (2016) PARKIN inactivation links Parkinson's disease to melanoma. *J. Natl. Cancer Ins.*, **108**.
30. Yokoyama, S., Woods, S.L., Boyle, G.M., Aoude, L.G., MacGregor, S., Zismann, V., Gartside, M., Cust, A.E., Haq, R., Harland, M. *et al.* (2011) A novel recurrent mutation in MITF predisposes to familial and sporadic melanoma. *Nature*, **480**, 99–103.
31. Lu, Y., Ek, W.E., Whiteman, D., Vaughan, T.L., Spurdle, A.B., Easton, D.F., Pharoah, P.D., Thompson, D.J., Dunning, A.M., Hayward, N.K. *et al.* (2014) Most common 'sporadic' cancers have a significant germline genetic component. *Hum. Mol. Genet.*, **23**, 6112–6118.
32. Law, M.H., Bishop, D.T., Lee, J.E., Brossard, M., Martin, N.G., Moses, E.K., Song, F., Barrett, J.H., Kumar, R., Easton, D.F. *et al.* (2015) Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. *Nat. Genet.*, **47**, 987–995.
33. Kenny, E.E., Timpson, N.J., Sikora, M., Yee, M.C., Moreno-Estrada, A., Eng, C., Huntsman, S., Burchard, E.G., Stoneking, M., Bustamante, C.D. *et al.* (2012) Melanesian blond hair is caused by an amino acid change in TYRP1. *Science*, **336**, 554.
34. Liu, F., Visser, M., Duffy, D.L., Hysi, P.G., Jacobs, L.C., Lao, O., Zhong, K., Walsh, S., Chaitanya, L., Wollstein, A. *et al.* (2015) Genetics of skin color variation in Europeans: genome-wide association studies with functional follow-up. *Hum. Genet.*, **134**, 823–835.
35. Duffy, D.L., Iles, M.M., Glass, D., Zhu, G., Barrett, J.H., Hoiom, V., Zhao, Z.Z., Sturm, R.A., Soranzo, N., Hammond, C. *et al.* (2010) IRF4 variants have age-specific effects on nevus count and predispose to melanoma. *Am. J. Hum. Genet.*, **87**, 6–16.
36. Falchi, M., Bataille, V., Hayward, N.K., Duffy, D.L., Bishop, J.A., Pastinen, T., Cervino, A., Zhao, Z.Z., Deloukas, P., Soranzo, N. *et al.* (2009) Genome-wide association study identifies variants at 9p21 and 22q13 associated with development of cutaneous nevi. *Nat. Genet.*, **41**, 915–919.
37. Fang, S., Han, J., Zhang, M., Wang, L.E., Wei, Q., Amos, C.I. and Lee, J.E. (2013) Joint effect of multiple common SNPs predicts melanoma susceptibility. *PloS One*, **8**, e85642.
38. Stefanaki, I., Panagiotou, O.A., Kodela, E., Gogas, H., Kypreou, K.P., Chatzinasiou, F., Nikolaou, V., Plaka, M., Kalfa, I., Antoniou, C. *et al.* (2013) Replication and predictive value of SNPs associated with melanoma and pigmentation traits in a Southern European case-control study. *PloS One*, **8**, e55712.
39. Kypreou, K.P., Stefanaki, I., Antonopoulou, K., Karagianni, F., Ntritsos, G., Zaras, A., Nikolaou, V., Kalfa, I., Chasapi, V., Polydorou, D. *et al.* (2016) Prediction of melanoma risk in a Southern European population based on a weighted genetic risk score. *J. Investig. Dermatol.*, **136**, 690–695.
40. Dudbridge, F. (2013) Power and predictive accuracy of polygenic risk scores. *PLoS Genet*, **9**, e1003348.
41. Ford, D., Bliss, J.M., Swerdlow, A.J., Armstrong, B.K., Franceschi, S., Green, A., Holly, E.A., Mack, T., MacKie, R.M., Osterlind, A. *et al.* (1995) Risk of cutaneous melanoma associated with a family history of the disease. The International Melanoma Analysis Group (IMAGE). *Int. J. Cancer*, **62**, 377–381.
42. Chatterjee, N., Wheeler, B., Sampson, J., Hartge, P., Chanock, S.J. and Park, J.H. (2013) Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nat. Genet.*, **45**, 400–405, 405e401–403.
43. Shi, J., Park, J.H., Duan, J., Berndt, S.T., Moy, W., Yu, K., Song, L., Wheeler, W., Hua, X., Silverman, D. *et al.* (2016) Winner's curse correction and variable thresholding improve performance of polygenic risk modeling based on genome-wide association study summary-level data. *PLoS Genet.*, **12**, e1006493.
44. Mavaddat, N., Pharoah, P.D.P., Michailidou, K., Tyrer, J., Brook, M.N., Bolla, M.K., Wang, Q., Dennis, J., Dunning, A.M., Shah, M. *et al.* (2015) Prediction of breast cancer risk based on profiling with common genetic variants. *J. Natl. Cancer Ins.*, **107**.
45. Song, M., Kraft, P., Joshi, A.D., Barrdahl, M. and Chatterjee, N. (2015) Testing calibration of risk models at extremes of disease risk. *Biostatistics*, **16**, 143–154.
46. Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F. and Sklar, P. (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, **460**, 748–752.
47. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
48. Vilhjalmsson, B.J., Yang, J., Finucane, H.K., Gusev, A., Lindstrom, S., Ripke, S., Genovese, G., Loh, P.R., Bhatia, G., Do, R. *et al.* (2015) Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.*, **97**, 576–592.

49. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C. and Muller, M. (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.*, **12**, 77.
50. Raghunathan TES, P.W. and Hoewyk, J.V. (2002) *IVEWARE: Imputation and Variance Estimation Software*. University of Michigan.
51. Little, R.J. and Rubin, D.B. (2000) Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Ann. Rev. Publ. Health*, **21**, 121–145.
52. Bartholomew, D. (1987) *Latent Variable Models and Factor Analysis*. Griffin, London.
53. Pfeiffer, R.M. and Gail, M.H. (2017) *Absolute Risk: Methods and Applications in Clinical Management and Public Health*. Chapman and Hall/CRC Reference.
54. Bruzzi, P., Green, S.B., Byar, D.P., Brinton, L.A. and Schairer, C. (1985) Estimating the population attributable risk for multiple risk factors using case-control data. *Am. J. Epidemiol.*, **122**, 904–914.