Multiplex melanoma families are enriched for polygenic risk

Published in:
Human Molecular Genetics
Publication date:
2020

Human Molecular Genetics, 2020, Vol. 29, No. 17 2976–2985

doi: 10.1093/hmg/ddaa156
Advance Access Publication Date: 27 July 2020
Association Studies Article

ASSOCIATION STUDIES ARTICLE

Multiplex melanoma families are enriched for polygenic risk

Matthew H. Law[1,*,†], Lauren G. Aoude[2,3,‡], David L. Duffy[4], Georgina V. Long[5,6,7,8], Peter A. Johansson[2,¶], Antonia L. Pritchard[2,9,||], Kiarash Khosrotehrani[10,11,††], Graham J. Mann[5], Grant W. Montgomery[12,‡‡], Mark M. Iles[13,§], Anne E. Cust[5,6,14,§§], Jane M. Palmer[2,¶¶], Melanoma GWAS Consortium*, Kerwin F. Shannon[5,6,|||], Andrew J. Spillane[5,6,†††], Jonathan R. Stretch[5,6], John F. Thompson[5,6,‡‡‡], Robyn P.M. Saw[5,6,§§§], Richard A. Scolyer[5,6,15,¶¶¶], Nicholas G. Martin[4], Nicholas K. Hayward[2] and Stuart MacGregor[1,$$$]

[1]Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia, [2]Oncogenomics, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia, [3]Surgical Oncology Group, The University of Queensland Diamantina Institute, The University of Queensland, Brisbane, QLD 4102, Australia, [4]Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia, [5]Melanoma Institute Australia, The University of Sydney, Sydney, NSW 2065, Australia, [6]Faculty of Medicine and Health, The University of Sydney, Sydney, NSW 2050, Australia, [7]Department of Medical Oncology, Mater Hospital, North Sydney, NSW 2060, Australia, [8]Department of Medical Oncology, Royal North Shore Hospital, St Leonards, NSW 2065, Australia, [9]Genetics and Immunology, University of the Highlands and Islands, Inverness IV2 5NA, UK, [10]Experimental Dermatology Group, The University of Queensland Diamantina Institute, The University of Queensland, Brisbane, QLD 4102, Australia, [11]Department of Dermatology, Princess Alexandra

[†]Matthew H. Law, http://orcid.org/0000-0002-4303-8821
[‡]Lauren G. Aoude, http://orcid.org/0000-0003-1448-3923
[¶]Peter A. Johansson, http://orcid.org/0000-0001-7015-5452
[||]Antonia L. Pritchard, http://orcid.org/0000-0001-5336-0454
[††]Kiarash Khosrotehrani, http://orcid.org/0000-0002-6406-4076
[‡‡]Grant W. Montgomery, http://orcid.org/0000-0002-4140-8139
[§]Mark M. Iles, http://orcid.org/0000-0002-2603-6509
[§§]Anne E. Cust, http://orcid.org/0000-0002-5331-6370
[¶¶]Jane M. Palmer, http://orcid.org/0000-0002-2702-522X
[|||]Kerwin F. Shannon, http://orcid.org/0000-0001-8538-3132
[†††]Andrew J. Spillane, http://orcid.org/0000-0002-8354-434X
[‡‡‡]John F. Thompson, http://orcid.org/0000-0002-2816-2496
[§§§]Robyn P.M. Saw, http://orcid.org/0002-3354-806X
[¶¶¶]Richard A. Scolyer, http://orcid.org/0000-0002-8991-0013
[$$$]Stuart MacGregor, http://orcid.org/0000-0001-6731-8142
*See Supplementary note for full author list.

Received: February 18, 2020. **Revised:** July 10, 2020. **Accepted:** July 14, 2020

Hospital, Brisbane, QLD 4102, Australia, [12]Molecular Biology, The University of Queensland, Brisbane, QLD 4072, Australia, [13]Leeds Institute for Medical Research, University of Leeds, Leeds LS2 9JT, UK, [14]School of Public Health, The University of Sydney, Sydney, NSW 2006, Australia and [15]Department of Tissue Oncology and Diagnostic Pathology, Royal Prince Alfred Hospital and NSW Health Pathology, Sydney, NSW 2050, Australia

*To whom correspondence should be addressed at: 300 Herston Road, Herston, QLD 4006, Australia. Tel: +61 733620213; Fax: +61 733620101; Email: matthew.law@qimrberghofer.edu.au

Abstract

Cancers, including cutaneous melanoma, can cluster in families. In addition to environmental etiological factors such as ultraviolet radiation, cutaneous melanoma has a strong genetic component. Genetic risks for cutaneous melanoma range from rare, high-penetrance mutations to common, low-penetrance variants. Known high-penetrance mutations account for only about half of all densely affected cutaneous melanoma families, and the causes of familial clustering in the remainder are unknown. We hypothesize that some clustering is due to the cumulative effect of a large number of variants of individually small effect. Common, low-penetrance genetic risk variants can be combined into polygenic risk scores. We used a polygenic risk score for cutaneous melanoma to compare families without known high-penetrance mutations with unrelated melanoma cases and melanoma-free controls. Family members had significantly higher mean polygenic load for cutaneous melanoma than unrelated cases or melanoma-free healthy controls (Bonferroni-corrected *t*-test $P = 1.5 \times 10^{-5}$ and 6.3×10^{-45}, respectively). Whole genome sequencing of germline DNA from 51 members of 21 families with low polygenic risk for melanoma identified a *CDKN2A* p.G101W mutation in a single family but no other candidate high-penetrance melanoma susceptibility genes. This work provides further evidence that melanoma, like many other common complex disorders, can arise from the joint action of multiple predisposing factors, including rare high-penetrance mutations, as well as via a combination of large numbers of alleles of small effect.

Introduction

Cutaneous melanoma (hereafter melanoma) is an aggressive cancer that originates in melanocytes, the pigment-producing cells of the skin. Melanoma can cluster in families; one formal definition of this is when the number of familial cases is greater than expected based on total members, their ages and the melanoma incidence in their birth cohort (1). Using this definition, Aitken *et al.* found that 4.7% of 1116 families had a case count higher than expected given the number, age and cohort of family members. Other, related, approaches have estimated that up to 9% of melanoma cases have a high familial risk. Within melanoma-dense families, high-penetrance segregating mutations have been found in *CDKN2A* (2,3), *CDK4* (4,5), *BAP1* (6), *MITF* (7,8), *TERT* (9,10), *POT1* (11,12), *ACD*, and *TERF2IP* (13). However, these rare high-penetrance mutations underpin melanoma development in only approximately 50% of densely affected families (14).

While melanoma rates globally increase with ultraviolet radiation exposure, indicating a substantial environmental component to risk, even within a uniformly high-risk environment, there will still be genetic variation in risk. Using twin and family studies, melanoma's heritable component has been estimated to be approximately 55% (15,16). In addition to rare high-penetrance mutations, genome-wide association studies (GWASs) have identified many loci harboring common genetic variants with a low to moderate impact on melanoma risk (17,18). The polygenic disease burden arising from these moderate and common GWAS variants can be estimated from genome-wide single nucleotide polymorphism (SNP) heritability (19). The SNP heritability for melanoma has been estimated to be 30% (20). The difference between the SNP and twin heritability estimates is likely due to poorly tagged SNPs and is likely to diminish in time as SNP genotyping arrays and coverage (e.g. through whole genome-sequencing) improves. Although not all of the genes underpinning SNP heritability have been identified, Lu *et al.* estimated that approximately a third of the SNP heritability can be explained by the specific genes identified in GWASs. These variants can be aggregated in a weighted sum for predicting risk—a polygenic risk score (PRS) (21,22).

When compared to non-high-risk melanoma families, high-risk families from Aitken *et al.*'s study had significantly higher proportions of melanoma risk phenotypes that are themselves polygenic traits (higher nevus count, poor tanning and fair skin) suggesting that these families may be enriched for melanoma polygenic risk (1,23,24). Thus, in addition to the carriage of high-penetrance mutations, and shared environment/behaviors, familial aggregation of melanoma could arise from high polygenic risk within the family. For example, if the familial clustering is due solely to an unidentified high-penetrance variant, the PRS would be expected to be lower than in sporadic cases and more similar to the melanoma-free population. In contrast, if the clustering is due to the low penetrance variants in the PRS, the PRS would be expected to be higher than in the sporadic cases. Finally, it is possible that having a high polygenic load may increase the penetrance from a melanoma risk mutation. However, the data may not always be clear cut. For example, application of a PRS for bipolar disorder to a family densely enriched for this disease found that affected family members were closer to the mean of unrelated GWAS cases than to healthy controls (25). That is, family members had a higher polygenic risk for bipolar disorders than healthy controls, suggesting that it may, in part, explain their familial aggregation. A recent study reported significantly lower melanoma PRS scores in families carrying high-penetrance mutations compared to those families without such a mutation (26). Here we report the distribution of polygenic risk for cutaneous melanoma within densely affected melanoma families compared to unrelated melanoma cases and healthy controls.

Results

The summed PRS was not significantly different between the three unrelated melanoma groups from Melanoma Institute Australia (MIA) and Princess Alexandra Hospital (PAH) (Supplementary Material, Fig. S1). Hence, for subsequent analyses, we combined the unrelated melanoma cases into a single group; the distribution of PRS for each group is reported in raw values Supplementary Material, Table S1a and as deltaPRS relative to controls in Supplementary Material, Table S1b. Family members with melanoma from the Queensland Familial Melanoma Project (QFMP) had significantly higher scores compared to both unrelated melanoma cases and healthy controls (Fig. 1). All genotyped members of the families were included in these analyses; to confirm that inclusion of multiple family members was not impacting our interpretation, we also restricted the analysis to one case sample per family, and used the average of the PRS of genotyped family cases. In this sensitivity analysis, melanoma family cases were still enriched for polygenic risk for melanoma compared to the unrelated cases (deltaPRS = 0.505, SE = 0.036, $N = 658$; Holm corrected pairwise t-test P-value = 0.0015). Unaffected members of melanoma families had significantly higher PRS than healthy controls and were not significantly different from either the unrelated or family-based melanoma cases, possibly due to the relatively small numbers in this group (Fig. 1).

Scores by familial risk class

The distribution of PRS for affected family members from medium-risk families (two family members with melanoma) did not differ significantly when compared to those from high-risk families (three or more members with melanoma; Bonferroni-corrected t-test $P = 0.94$; unadjusted $P = 0.11$, Supplementary Material, Table S3a and b, Fig. S2). In turn, the distribution of PRS in unaffected family members did not differ significantly to melanoma cases from either medium or high-risk families (unadjusted t-test $P = 0.45$ and $P = 0.07$ respectively, corrected $P = 1.00$ and $P = 0.70$).

Whole genome sequencing of melanoma families

Whole genome sequencing of 51 people from 21 families with melanoma and a relatively low PRS (Materials and Methods) was also performed to investigate whether a high- or medium-penetrance mutation might be identified. In one family for whom three individuals with melanoma were available for sequencing, we identified a pathogenic founder *CDKN2A* variant, p.G101W. This variant, first described in 1994 (27), is one of the most frequently reported deleterious founder missense variants reported in *CDKN2A* (28) and results in impaired binding of p16 to CDK4 (29). No further segregating mutations or near segregating (to take into account phenocopies in the high UVR exposed Australian population) were identified in the remaining 20 families.

Discussion

In a previous bipolar disorder family study, affected members had disease PRS similar to the PRS of cases recruited from clinics without regard to their family history of disease (25). In contrast, in our melanoma study, we found that melanoma cases from melanoma-prone families had higher mean PRS than unrelated melanoma cases from clinics and hospitals (Fig. 2). This difference may reflect differences in the genetic architecture of two diseases or may simply be due to the use of only a single pedigree in the bipolar disorder study, and a larger sample of familial bipolar cases may find polygenic risk is enriched in both diseases. In terms of genetic architecture, melanoma has a number of common variants with a relatively large odds ratio (OR) for a complex trait (Supplementary Material, Table S4). However, our results were not driven by GWAS variants of large effect, with the mean of affected family members' PRS higher than that of unrelated controls in the absence of genetic variants from the *MC1R* or *CDKN2A* GWAS loci (data not shown).

Melanoma-free QFMP members had a mean score that was higher than healthy, unrelated controls (deltaPRS 0.347, SE 0.104, $P = 0.0085$; Fig. 2, Supplementary Material, Table S1b). The mean of melanoma-free family members was essentially the same ($P > 0.05$) as the unrelated melanoma cases (deltaPRS 0.354, SE 0.0222), and not significantly different to affected family members (deltaPRS 0.523, SE 0.0282). This suggests that even (currently) unaffected family members are at an elevated genetic risk for melanoma (as might be expected).

The vast majority of melanoma families in the QFMP were pre-screened for known high-penetrance melanoma mutations in *CDKN2A* and *CDK4*. This study only includes those without identified mutations; 633 of the included 660 families were screened. We hypothesized that those families with a low polygenic risk for melanoma (i.e. similar to the distribution of PRS in healthy controls) might be more likely to harbor a (novel) high-penetrance mutation, and indeed recent work has shown that melanoma families with a high-penetrance mutation may have lower polygenic risk for melanoma compared to families without (26). Sequencing of 21 families selected from those with low mean-family PRS (that is, the average PRS across genotyped, affected, family members was more similar to healthy controls) identified a single family carrying a previously undetected high-penetrance deleterious *CDKN2A* mutation, p.G101W (27–29).

Possible explanations for these densely affected families with a low melanoma PRS but no detected high-penetrance mutation include the following: (i) there are shared environmental factors within the family, (ii) there are rare moderate-penetrance mutations that we were underpowered to find by sequencing, (iii) although the families were selected to have a low PRS, the PRS values are not very low when benchmarked against the PRS values seen in healthy individuals (Supplementary Material, Fig. S3) and (iv) since our PRS only indexed a minority of the genetic variation in melanoma risk, these families may be carrying a large number of common low-penetrance risk variants, which are not in our current PRS. As GWAS sample sizes increase, we will identify more low penetrance variants, and in the future, we will be able to better assess point (iv).

As a potential limitation, the unrelated melanoma cases were used to select the best PRS (score 3, Supplementary Material, Tables S4 and S5), which may have led to a slight inflation in the significance of their PRS compared to healthy controls. However, the R^2 difference between the chosen score and genome-wide significant SNPs alone was negligible, and the mean PRS of melanoma cases from melanoma families was still significantly higher than unrelated cases. To further explore this, we repeated the analyses using score 1, limited to SNPs in the PRS reaching $P < 5 \times 10^{-8}$. As shown in Supplementary Material, Fig. S4, cases from melanoma families still had significantly higher scores relative to unrelated melanoma cases (Bonferroni-corrected $P = 0.014$) indicating that our findings are not driven by selection of PRS in this manner. As a further limitation, pruning to genotyped SNPs across included arrays while reducing potential artefacts and biases may have reduced our power;

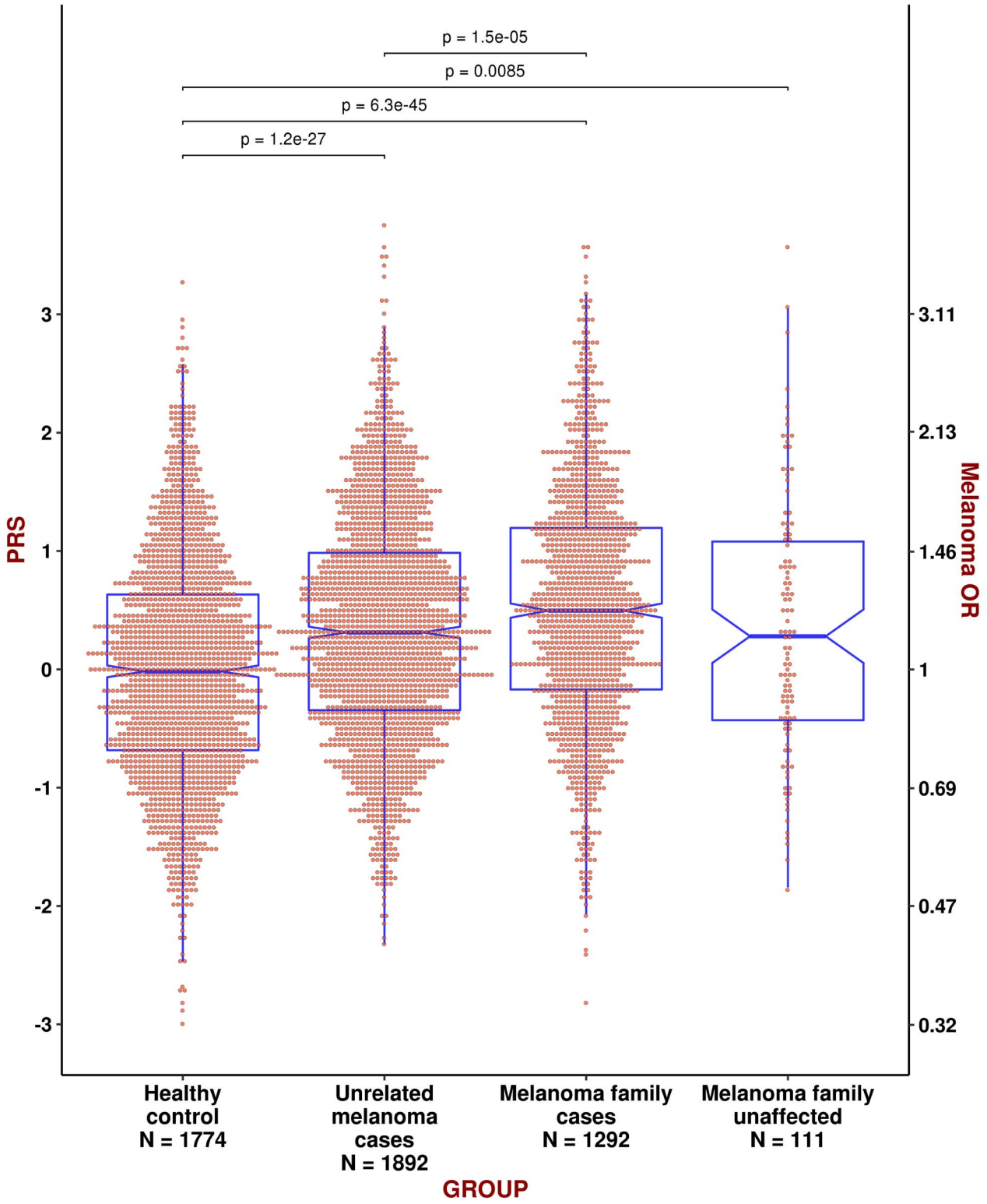

Figure 1. Distributions of PRSs by group PRS are plotted using dot-density binning, with each orange point representing a single person's PRS. The left y-axis reports the mean 0-centered PRS values while the right axis displays the melanoma OR for a 1 SD change in the PRS (see Materials and Methods). The distribution of the PRS have been summarized using a blue layered notched whisker plot, with the midpoint the median and the notches an approximation of the 95% confidence interval of the median (see Martials and Methods). Each boxed area covers the first to third quartile of the group, with the whiskers extending out to 1.5 × the interquartile range. Untransformed values relating to this distribution are tabulated in Supplementary Material, Table S1a and transformed values in Supplementary Material, Table S1b. Bonferroni-corrected pairwise *t*-tests *P*-values (6 comparisons in total) are displayed where the adjusted two-sided *P*-value is <0.05. Total sample numbers are reported for each group.

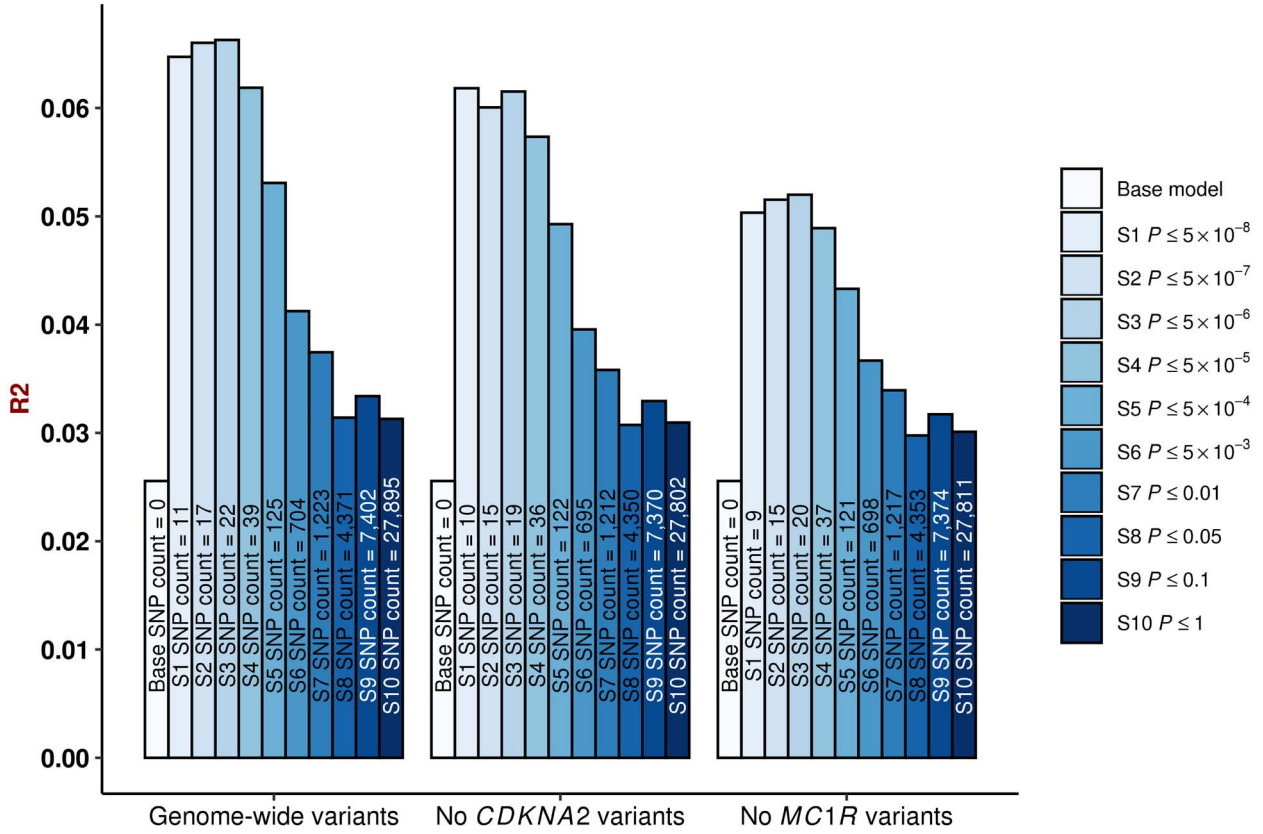

Figure 2. Model evaluation for PRS constructed from SNPs binned by α P-value threshold. The base model is a logistic regression of the first 10 principal components on unrelated melanoma cases versus controls. For each subsequent model, the PRS derived from the specified melanoma meta-analysis P-value range of SNPs was added to the regression (see Materials and Methods). For each PRS, we plot the model Nagelkerke R2 and report the number of included independent variants. The logistic regression P-values for all models including genome-wide variants were $<1 \times 10^{-4}$. We also repeat the analysis without the *MC1R* (no *MC1R*) or *CDKN2A* (no *CDKN2A*) regions (see Materials and Methods); logistic regression P-values for each model including a PRS with these regions dropped were all <0.001. This data are tabulated in Supplementary Material, Table S5.

however, we were still able to detect significant differences between analysis groups (e.g. Fig. 1).

Familial clustering in the context of a higher polygenic risk for melanoma than seen in unrelated melanoma cases from clinics and hospitals suggests that some familial clustering may be due to many common variants of low and moderate effect. Indeed, previous analysis of the QFMP found that high-risk families were more likely to have melanoma risk phenotypes that are themselves polygenic (e.g. high nevus count) (1,23,24). We explored if the number of affected family members was related to degree of polygenic risk for melanoma. The mean PRS for high-risk families (three or more affected members) while higher was not significantly different from those drawn from families with only two (Supplementary Material, Table S3a and b, Fig. S2). A larger dataset may be required to determine if there is a relationship between the number of affected family members and degree of enrichment for PRS. Regardless, high polygenic load in families does not rule out a role for segregating mutations as a high polygenic load may increase the penetrance of a rare variant (e.g. of a moderate-penetrance variant proposed above in example (ii)) such that a greater number of family members are affected.

This work provides further evidence that cutaneous melanoma, like other common complex disorders, can arise from multiple pathways including rare high-penetrance mutations as well as via a combination of large numbers of alleles of small effect.

Materials and Methods

Samples

Available to us were two large samples of unrelated melanoma cases, one large family study of melanoma and a sample of healthy controls (Table 1).

Unrelated melanoma cases—MIA. Patients with histopathologically confirmed melanoma ($N = 1740$) with germline DNA available from blood samples in the MIA Biospecimen Bank (protocol HREC/10/RPAH/530, HREC/11/RPAH/444) were identified. Included samples were collected between April 2000 and February 2016. Further inclusion and exclusion criteria are reported in the GWAS cleaning section. Ethical approval was managed by the Human Research Ethics Committee of the Royal Prince Alfred Hospital (Sydney, New South Wales, Australia), and all biospecimens and patient data were collected with written informed consent.

Unrelated melanoma cases—PAH. Patients whose locally invasive primary melanoma was confirmed by histopathology between 1994 and 2007 were recruited with informed consent from the Multidisciplinary Melanoma Clinic at the PAH and related clinics ($N = 236$). Ethical oversight and management were through the Metro South Hospital and Health Service Human Research Ethics Committee (protocol HREC/12/QPAH/349, Woolloongabba, Queensland, Australia).

Table 1. Sample counts pre quality control (QC) and post-QC for QFMP melanoma family cases and unaffected members

Samples	Family-based		Unrelated			
	Melanoma cases	Unaffected	PAH	MIA 1	MIA 2	Healthy controls
Pre-QC	1342	127	236	1439	301	1799
Post-QC	1292	111	232	1373	287	1774
Male	596	51	118	826	182	823
Female	696	60	114	547	105	951

Unrelated melanoma cases from the PAH and MIA are also reported. MIA samples were genotyped across two batches, with the counts shown separately.

QFMP. Index cases for 1912 melanoma families recruited to the QFMP were ascertained by contacting all individuals with a histologically confirmed melanoma in the Queensland Cancer Register diagnosed from 1 January 1982 to 31 December 1990 (30). An iterative and sequential sampling scheme was used where if a first degree relative had melanoma, then the first-degree relatives of that new case were also included. Informed consent was a requirement for study/cohort inclusion, and all work was overseen and approved by the Human Research Ethic Committee of the Queensland Institute of Medical Research (now QIMR Berghofer Medical Research Institute), University of Sydney, University of Melbourne and cancer registries of New South Wales, Victoria and Queensland.

At the time of recruitment, the majority of QFMP families (1403 of 1912 families) had only one case member. Recruited families were ranked as low-, medium- or high-risk using a risk index (*T*) that factored in the number of confirmed cases of melanoma versus number of unaffected family members, ages and year of birth (31). All high and medium *T*-index families were approached for blood samples, as were 250 of the low *T*-index families; in total, 738 families consented to providing blood. A total of 1469 participants from 680 families provided blood samples from which DNA was extracted. Further inclusion and exclusion criteria are reported in the GWAS cleaning section (Table 1). Subsequent data linkages identified additional cases of melanoma within these families, resulting in families increasing their *T* index class. As a result, we adopted a more simplified risk class based on simple counts of affected family members.

Genotyped QFMP families were classified for PRS analyses into three classes based on their current number of affected family members: low-risk (one affected member), medium-risk (two affected members) and high-risk (three or more affected members). A single family was dropped as it had missing total affected family count data. Following data cleaning (see GWAS cleaning section below), genotyped samples were from a single low-risk family (one unaffected family member genotyped), 207 medium-risk families and 453 high-risk families. The single low-risk family was dropped from all analyses. The final count of included unaffected family members was 111 (Table 1). For affected family members, 261 were from the 207 medium-risk families and 1031 from the 453 high-risk families. For the 111 unaffected family members, one was from a medium-risk family, while 110 were from 110 different high-risk families (i.e. one per family). A more detailed breakdown of the number of individuals genotyped from families is presented in Supplementary Material, Table S2.

Healthy controls without melanoma. A total of 1799 unrelated individuals who self-reported as being free from melanoma and had no family history of melanoma were drawn from the Brisbane Adolescent Twin Study (32,33). The majority (80%) of samples were parents of twins, with the remainder being a single individual from twin or sibling pairs. Ethics approval was overseen by the Queensland Institute of Medical Research (now QIMR Berghofer Medical Research Institute) Human Research Ethics Committee (Herston, Queensland, Australia).

GWAS genotyping

Unrelated melanoma cases from MIA and PAH and QFMP samples were genotyped in two batches utilizing the Illumina Oncoarray (San Diego, CA, USA). The first batch included the majority of the MIA samples ($N = 1439$) and the PAH samples ($N = 236$), and the second batch included the 1469 QFMP samples and a further 301 samples from MIA. Healthy melanoma-free controls were genotyped using the Illumina HumanHap610 arrays (San Diego, CA, USA) (34).

GWAS cleaning

Raw genotypes were cleaned and exported using Illumina GenomeStudio/BeadStudio (San Diego, CA, USA). Following protocols published by the Oncoarray Consortium (35), oncoarray genotype data were cleaned of poorly performing assays, and variant positions were updated and aligned where required. Subsequently, oncoarray and Illumina HumanHap610 GWAS data were separately aligned to 1000 genomes and then cleaned using PLINK (v1.90b5.4) (36,37). Variants were filtered out for having a minor allele frequency <0.01 or for Hardy–Weinberg equilibrium *P*-value $<5 \times 10^{-4}$ in controls or $<5 \times 10^{-10}$ in cases. Individuals were retained if they had calls for at least 97% of variants, and their heterozygosity values were between -0.05 and 0.05. Samples were also discarded if their genetically determined sex did not match phenotype records. Principal component analysis in PLINK including reference populations from the 1000 Genomes Project was used to identify and exclude individuals with non-European ancestry (>6 SD from the centroid of principal components 1 and 2). For the unrelated melanoma cases and healthy controls, identity by descent was used to remove related individuals such that no pair existed with pihat >0.15 within and across groups. Within families, identity by descent was used to confirm reported relationships.

Merging genotyped data

Following GWAS cleaning, there were 1892 unrelated melanoma cases genotyped on the Oncoarray, 1660 from MIA and 232 from PAH. From the melanoma families, there were 1292 cases with melanoma and 111 unaffected members, also genotyped on the Oncoarray. A total of 1774 healthy controls without melanoma genotyped on the Illumina HumanHap610 array remained after cleaning.

The genotype data for the Oncorray (SNPs = 397 083) and Illumina HumanHap610 (SNPs = 494 381) were filtered to the autosomal SNPs in common across the two arrays. The combined GWAS data were cleaned again (missing genotypes/calls <3%, Hardy–Weinberg equilibrium *P*-value $>1 \times 10^{-6}$) using PLINK v1.90b6.10. One further MIA case from the second batch was removed for missing genotype calls >3% in the merged set. The final, merged, cleaned set included 135 325 SNPs, 3184 individuals with melanoma, 113 unaffected family members, and 1774 healthy control samples (Table 1).

Melanoma risk effect sizes

The published melanoma risk meta-analysis included two GWASs that incorporated probands from QFMP families (17). These overlapping GWAS were removed from the meta-analysis, and it was re-run prior to PRS calculations; contributing datasets were otherwise as previously reported (17). A fixed effects meta-analysis was performed using GWAMA v2.1 (38). A total of 133 717 SNPs overlapped between the genotype data and meta-analysis results.

A total of 5000 randomly selected European Ancestry (as measured by principal component analysis) UK Biobank (39) individuals were used for linkage-disequilibrium (LD) calculations. Haplotype reference consortium-imputed best guess variants (imputation quality rsq and best guess threshold both ≥0.3) were retained if their call rate was ≥97%, had a minor allele frequency ≥0.001 and a Hardy–Weinberg equilibrium *P*-value $\geq 1 \times 10^{-6}$. Of the 133 717 variants overlapping between the genotype and meta-analysis data, 131 957 were in the LD reference panel. To explore the impact of *MC1R* (chromosome 16q) and *CDKN2A* (chromosome 9p) regions, which have common variants with relatively large effect sizes for a GWAS, on the PRS, we also generated SNP lists excluding all SNPs from 85 megabase (mb) to the end of chromosome 16 (328 variants) and 18–25 mb of chromosome 9 (505 variants).

LD clumping using the LD reference panel was performed in two stages using PLINK (36,37). The first used 250 kilobase windows at LD r^2 0.5 (100 373 clumps formed from the 131 957 top variants), and the second round used 10 mb windows at LD r^2 0.025 (27 895 clumps formed from 100 373 top variants). There were proportionally smaller counts of variants when *MC1R* or *CDKN2A* were excluded. The PRS for each sample was calculated using the score function in PLINK (36,37).

Selecting *P*-value ranges for PRSs

Clumped meta-analysis SNPs were binned by a *P*-value threshold to calculate a series of 10 scores. The best performing PRS (score 3, including independent SNPs with a *P*-value $\leq 5 \times 10^{-6}$) was identified using their Nagelkerke R^2 value from a logistic regression on melanoma status for unrelated melanoma cases versus healthy controls with the first 10 principal components included as covariates (Fig. 2, Supplementary Material, Table S5). The most strongly associated independent (lead) SNPs with a *P*-value $<5 \times 10^{-6}$ are reported in Supplementary Material, Table S4. Due to the use of variants only in common across the genotyping arrays, meta-analysis and LD panel, not all loci reported in (17) are present, e.g. *SLC45A2*.

Selection of melanoma families for sequencing

To explore whether families with melanoma and a relatively low PRS for melanoma (i.e. similar to the distribution of PRS in healthy controls) harbored a high-penetrance mutation, we calculated the mean PRS for each family, dividing the sum of the PRS of genotyped, affected, family members by the number of affected genotyped members. Starting at the lowest end of the mean PRS distribution (lowest being the least polygenic load for melanoma), we selected 51 individuals from 21 families who (i) had multiple members genotyped in this study and (ii) had sufficient excess DNA available for further analysis/sequencing in the future. The PRS for sequenced family members is displayed in Supplementary Material, Fig. S3.

Whole genome sequencing of melanoma families

Whole genome sequencing was performed by Macrogen (Seoul, South Korea) on the Illumina HiSeq 2000 platform (San Diego, USA), with a mean coverage of 60–96×. Sequence output was mapped to the UCSC human genome reference build 19 by the Burrows-Wheeler Aligner (40). SNPs were detected using bcftools and SAMtools mpileup with disabled BAQ computation (41), in/del were detected with pindel (42) and both were annotated to dbSNP144, including ExAc population frequencies (43) by ANNOVAR (44). Variants were then filtered on quality score (>70) and alternate read counts (>2 and >20% of all reads at a given position) for stringency of data. Sanger sequencing was performed to confirm identified variants of interest.

Additional methods

Unless otherwise noted, Bonferroni-corrected pairwise *t*-tests in *R* were used to test for significant differences between groups, with the significance threshold for multiple testing corrected *P*-values set to <0.05.

PRS distributions were plotted using the ggplot2 and ggpubr R packages (45,46). For plotting purposes, PRS distributions were rescaled to have mean 0 and with a standard deviation of 1 unit over the whole population by subtracting the population mean PRS from each PRS value, and then dividing by the standard deviation of the PRS for the entire dataset. The mean of the healthy controls was then subtracted from all transformed PRS values such that the plots are centered with the mean of the control population = 0. To summarize distributions, we used a notched whisker plot where the midpoint is the PRS median with the notches corresponding to 1.58 × the interquartile range divided by the square root of the sample number, which is an approximation of the 95% confidence interval of the median. The boxed area is the first to third quartile of the PRS distribution. Whiskers extend as far as 1.5 × the interquartile range with points beyond this plotted individually. In addition to plotting PRS values scaled to mean 0 as above on the left y-axis, on the right axis, we also display the OR for a 1 SD change in PRS. Transformed PRS together with the first 10 principal components were used in a logistic regression with unrelated melanoma samples versus healthy controls to determine a melanoma OR of 1.46 per SD of the PRS.

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

We gratefully acknowledge the donation of time and support by all the participants in this research project. We also wish to recognize the hard work of all research, support and admin staff not named elsewhere.

The Queensland Familial Melanoma Project and Brisbane Adolescent Twins Study could not have been performed without the support and hard work in terms of the collection and handling of samples, and database and project management, by A. Baxter, M. de Nooyer, I. Gardner, D. Statham, B. Haddon, M.J. Wright, J. Symmons, B. Castellano, L. Bowdler, S. Smith, D. Smyth, L. Wallace, M.J. Campbell, A. Caracella, M. Kvaskoff, O. Zheng, B. Chapman and H. Beeby. In addition to those named here, we acknowledge the work and contribution of the many research assistants and interviewers.

Recruitment and collection of the PAH dataset could not have taken place without the study participants, whom we gratefully thank. We acknowledge the hard work and support of Adele Green, Mark Smithers, Casey Rowe and Maryrose Malt for overall cohort management, funding applications, as well as consent, recruitment and sampling of participants.

Conflict of Interest statement. RAS has received fees for professional services from Qbiotics Group Limited, Novartis Pharma AG, MSD Sharp & Dohme (Australia), NeraCare, AMGEN Inc., Bristol-Myers Squibb, Novartis Pharmaceuticals Australia Pty Limited, Myriad Genetics GmbH, GlaxoSmithKline Australia. RPMS has received honoraria for advisory board participation from MSD, Novartis and Qbiotics and speaking honoraria from BMS. JFT has received honoraria for advisory board participation from BMS Australia, MSD Australia, GSK and Provectus Inc and travel support from GlaxoSmithKline and Provectus Inc. GVL is a consultant advisor for Aduro, Amgen, Bristol-Myers Squibb, Highlight Therapeutics S.L., Mass-Array, Merck, MSD, Novartis, OncoSec Medical, Pierre Fabre, Roche, QBiotics, Skyline DX and Sandoz.

Funding

This work was funded by Worldwide Cancer Research grant 16-0101. This work was conducted using the UK Biobank Resource (application number 25331). Stuart MacGregor, Georgina V. Long, Lauren G. Aoude (APP1109048), Anne E. Cust (APP1147843), K. Khosrotehrani (APP1125290) and Nicholas K. Hayward (APP1117663) each acknowledge fellowship support from the Australian National Health and Medical Research Council (NHMRC) and from the Australian Research Council. Georgina V. Long is also supported by the University of Sydney Medical Foundation. Antonia L. Pritchard is supported by Highland Island Enterprise (HMS 9353763). Mark M. Iles acknowledges funding support from Cancer Research UK (c588/a19167) and the NIH (ca083115). Richard A. Scolyer is supported by a NHMRC Practitioner Fellowship and an NHMRC Program Grant. Robyn P.M. Saw is supported by Melanoma Institute Australia. John F. Thompson is supported by a NHMRC Program Grant. Support for specific cohorts is reported below:

Queensland Familial Melanoma Project and Brisbane Adolescent Twins Study

Sample and data collection was supported by funding from the NHMRC (200071, 241944, 339462, 380385, 389927, 389875, 389891, 389892, 389938, 443036, 442915, 442981, 496610, 496675, 496739, 552485, 552498, APP1049894), the Melanoma Research Alliance, the National Institute of Health's National Cancer Institute (CA88363, CA83115, CA122838, CA87969, CA055075, CA100264, CA133996 and CA49449) and the Australian Cancer Councils of New South Wales, Victoria and Queensland. Funding was also provided by the Cancer Institute New South Wales, the Cooperative Research Centre for Discovery of Genes for Common Human Diseases, Cerylid Biosciences (Melbourne), the Australian Cancer Research Foundation, The Wellcome Trust (WT084766/Z/08/Z) and donations from Neville and Shirley Hawkins.

Princess Alexandra Hospital

This collection was funded by the University of Queensland Diamantina Institute, the Meehan Foundation, a NHMRC fellowship (APP1125290), and Cancer Council Queensland (1125237).

Melanoma Institute Australia

The MIA biospecimen bank, sample consenting, recruitment and data collection was enabled by funding from MIA and Program Grants from the NHMRC. This work was also funded by Cancer Institute New South Wales grants and infrastructure grants from Macquarie University and the Australian Cancer Research Foundation.

Support from the the Cameron Family, Ainsworth Foundation and colleagues at Melanoma Institute Australia, Mater Hospital, Royal North Shore Hospital, Royal Prince Alfred Hospital and New South Wales Health Pathology is gratefully acknowledged.

References

1. Aitken, J.F., Duffy, D.L., Green, A., Youl, P., MacLennan, R. and Martin, N.G. (1994) Heterogeneity of melanoma risk in families of melanoma patients. *Am. J. Epidemiol.*, **140**, 961–973.
2. Kamb, A., Shattuck-Eidens, D., Eeles, R., Liu, Q., Gruis, N.A., Ding, W., Hussey, C., Tran, T., Miki, Y. and Weaver-Feldhaus, J. (1994) Analysis of the p16 gene (CDKN2) as a candidate for the chromosome 9p melanoma susceptibility locus. *Nat. Genet.*, **8**, 23–26.
3. Berwick, M., Orlow, I., Hummer, A.J., Armstrong, B.K., Kricker, A., Marrett, L.D., Millikan, R.C., Gruber, S.B., Anton-Culver, H., Zanetti, R. *et al.* (2006) The prevalence of CDKN2A germline mutations and relative risk for cutaneous malignant melanoma: an international population-based study. *Cancer Epidemiol. Biomark. Prev.*, **15**, 1520–1525.
4. 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.
5. Goldstein, A.M., Chidambaram, A., Halpern, A., Holly, E.A., Guerry, D., IV, Sagebiel, R., Elder, D.E. and Tucker, M.A. (2002) Rarity of CDK4 germline mutations in familial melanoma. *Melanoma Res.*, **12**, 51–55.
6. 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.
7. 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.
8. Bertolotto, C., Lesueur, F., Giuliano, S., Strub, T., de Lichy, M., Bille, K., Dessen, P., d'Hayer, B., Mohamdi, H., Remenieras, A. *et al.* (2011) A SUMOylation-defective MITF germline mutation predisposes to melanoma and renal carcinoma. *Nature*, **480**, 94–98.
9. 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.

10. Heidenreich, B., Nagore, E., Rachakonda, P.S., Garcia-Casado, Z., Requena, C., Traves, V., Becker, J., Soufir, N., Hemminki, K. and Kumar, R. (2014) Telomerase reverse transcriptase promoter mutations in primary cutaneous melanoma. *Nat. Commun.*, **5**, 3401.
11. 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.
12. 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.
13. 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 Inst.*, **107**.
14. Aoude, L.G., Wadt, K.A.W., Pritchard, A.L. and Hayward, N.K. (2015) Genetics of familial melanoma: 20 years after CDKN2A. *Pigment Cell Melanoma Res.*, **28**, 148–160.
15. Shekar, S.N., Duffy, D.L., Youl, P., Baxter, A.J., Kvaskoff, M., Whiteman, D.C., Green, A.C., Hughes, M.C., Hayward, N.K., Coates, M. *et al.* (2009) A population-based study of Australian twins with melanoma suggests a strong genetic contribution to liability. *J. Invest. Dermatol.*, **129**, 2211–2219.
16. Mucci, L.A., Hjelmborg, J.B., Harris, J.R., Czene, K., Havelick, D.J., Scheike, T., Graff, R.E., Holst, K., Möller, S., Unger, R.H. *et al.* (2016) Familial risk and heritability of cancer among twins in Nordic countries. *JAMA*, **315**, 68–76.
17. 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.
18. Ransohoff, K.J., Wu, W., Cho, H.G., Chahal, H.C., Lin, Y., Dai, H.-J., Amos, C.I., Lee, J.E., Tang, J.Y., Hinds, D.A. *et al.* (2017) Two-stage genome-wide association study identifies a novel susceptibility locus associated with melanoma. *Oncotarget*, **8**, 17586–17592.
19. Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.
20. 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.
21. Wray, N.R., Lee, S.H., Mehta, D., Vinkhuyzen, A.A.E., Dudbridge, F. and Middeldorp, C.M. (2014) Research review: polygenic methods and their application to psychiatric traits. *J. Child Psychol. Psychiatry*, **55**, 1068–1087.
22. Khera, A.V., Chaffin, M., Aragam, K.G., Haas, M.E., Roselli, C., Choi, S.H., Natarajan, P., Lander, E.S., Lubitz, S.A., Ellinor, P.T. *et al.* (2018) Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.*, **50**, 1219–1224.
23. Duffy, D.L., Zhu, G., Li, X., Sanna, M., Iles, M.M., Jacobs, L.C., Evans, D.M., Yazar, S., Beesley, J., Law, M.H. *et al.* (2018) Novel pleiotropic risk loci for melanoma and nevus density implicate multiple biological pathways. *Nat. Commun.*, **9**, 4774.
24. Visconti, A., Duffy, D.L., Liu, F., Zhu, G., Wu, W., Chen, Y., Hysi, P.G., Zeng, C., Sanna, M., Iles, M.M. *et al.* (2018) Genome-wide association study in 176,678 Europeans reveals genetic loci for tanning response to sun exposure. *Nat. Commun.*, **9**, 1684.
25. Collins, A.L., Kim, Y., Szatkiewicz, J.P., Bloom, R.J., Hilliard, C.E., Quackenbush, C.R., Meier, S., Rivas, F., Mayoral, F., Cichon, S. *et al.* (2013) Identifying bipolar disorder susceptibility loci in a densely affected pedigree. *Mol. Psychiatry*, **18**, 1245–1246.
26. Schlafly, A., Pfeiffer, R.M., Nagore, E., Puig, S., Calista, D., Ghiorzo, P., Menin, C., Fargnoli, M.C., Peris, K., Song, L. *et al.* (2019) Contribution of common genetic variants to familial aggregation of disease and implications for sequencing studies. *PLoS Genet.*, **15**, e1008490.
27. Hussussian, C.J., Struewing, J.P., Goldstein, A.M., Higgins, P.A., Ally, D.S., Sheahan, M.D., Clark, W.H., Jr., Tucker, M.A. and Dracopoli, N.C. (1994) Germline p16 mutations in familial melanoma. *Nat. Genet.*, **8**, 15–21.
28. Ciotti, P., Struewing, J.P., Mantelli, M., Chompret, A., Avril, M.F., Santi, P.L., Tucker, M.A., Bianchi-Scarrà, G., Bressac-de Paillerets, B. and Goldstein, A.M. (2000) A single genetic origin for the G101W CDKN2A mutation in 20 melanoma-prone families. *Am. J. Hum. Genet.*, **67**, 311–319.
29. Ruas, M., Brookes, S., McDonald, N.Q. and Peters, G. (1999) Functional evaluation of tumour-specific variants of p16INK4a/CDKN2A: correlation with protein structure information. *Oncogene*, **18**, 5423–5434.
30. Aitken, J.F., Green, A.C., MacLennan, R., Youl, P. and Martin, N.G. (1996) The Queensland familial melanoma project: study design and characteristics of participants. *Melanoma Res.*, **6**, 155–165.
31. Aitken, J., Welch, J., Duffy, D., Milligan, A., Green, A., Martin, N. and Hayward, N. (1999) CDKN2A variants in a population-based sample of Queensland families with melanoma. *J. Natl. Cancer Inst.*, **91**, 446–452.
32. Zhu, G., Duffy, D.L., Eldridge, A., Grace, M., Mayne, C., O'Gorman, L., Aitken, J.F., Neale, M.C., Hayward, N.K., Green, A.C. *et al.* (1999) A major quantitative-trait locus for mole density is linked to the familial melanoma gene CDKN2A: a maximum-likelihood combined linkage and association analysis in twins and their sibs. *Am. J. Hum. Genet.*, **65**, 483–492.
33. McGregor, B., Pfitzner, J., Zhu, G., Grace, M., Eldridge, A., Pearson, J., Mayne, C., Aitken, J.F., Green, A.C. and Martin, N.G. (1999) Genetic and environmental contributions to size, color, shape, and other characteristics of melanocytic naevi in a sample of adolescent twins. *Genet. Epidemiol.*, **16**, 40–53.
34. Macgregor, S., Montgomery, G.W., Liu, J.Z., Zhao, Z.Z., Henders, A.K., Stark, M., Schmid, H., Holland, E.A., Duffy, D.L., Zhang, M. *et al.* (2011) Genome-wide association study identifies a new melanoma susceptibility locus at 1q21. *Nat. Genet.*, **43**, 1114–1118.
35. Amos, C.I., Dennis, J., Wang, Z., Byun, J., Schumacher, F.R., Gayther, S.A., Casey, G., Hunter, D.J., Sellers, T.A., Gruber, S.B. *et al.* (2017) The OncoArray consortium: a network for understanding the genetic architecture of common cancers. *Cancer Epidemiol. Biomark. Prev.*, **26**, 126–135.
36. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., 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.

37. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. and Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, **4**, 7.
38. Mägi, R. and Morris, A.P. (2010) GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics*, **11**, 288.
39. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J. *et al.* (2018) The UK biobank resource with deep phenotyping and genomic data. *Nature*, **562**, 203–209.
40. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, **25**, 1754–1760.
41. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
42. Ye, K., Schulz, M.H., Long, Q., Apweiler, R. and Ning, Z. (2009) Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*, **25**, 2865–2871.
43. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B. *et al.* (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, **536**, 285–291.
44. Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.*, **38**, e164.
45. Kassambara, A. (2018) *ggpubr:'ggplot2'Based Publication Ready Plots, R package version 0.2.*
46. Wickham, H. (2011) ggplot2. *WIREs Comp Stat*, **3**, 180–185.