

# Uncovering the complex relationship between balding, testosterone and skin cancers in men

Received: 28 September 2022

Accepted: 24 August 2023

Published online: 03 October 2023

 Check for updates

Jue-Sheng Ong <sup>1</sup>✉, Mathias Seviiri<sup>1,2</sup>, Jean Claude Dusingize<sup>1</sup>, Yeda Wu <sup>1</sup>, Xikun Han <sup>1,3</sup>, Jianxin Shi <sup>4</sup>, Catherine M. Olsen <sup>1,5</sup>, Rachel E. Neale<sup>1</sup>, John F. Thompson <sup>6,7</sup>, Robyn P. M. Saw<sup>6,7</sup>, Kerwin F. Shannon <sup>6,7</sup>, Graham J. Mann <sup>6,8,9</sup>, Nicholas G. Martin <sup>10</sup>, Sarah E. Medland <sup>10</sup>, Scott D. Gordon <sup>10</sup>, Richard A. Scolyer <sup>6,8,11,12</sup>, Georgina V. Long <sup>6,8,13</sup>, Mark M. Iles <sup>14,15</sup>, Maria Teresa Landi <sup>4</sup>, David C. Whiteman <sup>1</sup>, Stuart MacGregor <sup>1</sup> & Matthew H. Law <sup>1,2</sup>✉

Male-pattern baldness (MPB) is related to dysregulation of androgens such as testosterone. A previously observed relationship between MPB and skin cancer may be due to greater exposure to ultraviolet radiation or indicate a role for androgenic pathways in the pathogenesis of skin cancers. We dissected this relationship via Mendelian randomization (MR) analyses, using genetic data from recent male-only meta-analyses of cutaneous melanoma (12,232 cases; 20,566 controls) and keratinocyte cancers (KCs) (up to 17,512 cases; >100,000 controls), followed by stratified MR analysis by body-sites. We found strong associations between MPB and the risk of KC, but not with androgens, and multivariable models revealed that this relationship was heavily confounded by MPB single nucleotide polymorphisms involved in pigmentation pathways. Site-stratified MR analyses revealed strong associations between MPB with head and neck squamous cell carcinoma and melanoma, suggesting that sun exposure on the scalp, rather than androgens, is the main driver. Men with less hair covering likely explains, at least in part, the higher incidence of melanoma in men residing in countries with high ambient UV.

The most common three types of skin cancers are named for their presumed cell of origin: the keratinocyte derived cancers (KC) basal cell carcinoma (BCC) and squamous cell carcinoma (SCC), and the melanocyte-derived melanoma. Exposure to ultraviolet radiation (UVR) is a key risk factor for skin cancer, and cutaneous melanocytes produce and distribute protective pigment-containing melanosomes to surrounding skin keratinocyte cells to protect against UV damage [For review see ref. 1]. In some Western countries such as Australia and the United States, cutaneous melanoma incidence is higher in men compared to women<sup>2–5</sup>. While this may in part derive from differences in sun exposure behaviours and reduced healthcare engagement

amongst men, these are insufficient to fully explain the difference<sup>2</sup>. Differences between men and women in immune responses and hormone levels may also influence the risk of melanoma<sup>2</sup> and KCs<sup>6</sup>.

Male-pattern baldness (MPB), also known as androgenetic alopecia, is characterised by progressive hair follicle miniaturisation, leading to hair loss in men. This condition is often linked with changes in dihydrotestosterone levels, with several studies showing that patients with MPB had higher endogenous testosterone levels than controls<sup>7–9</sup>. Recent genetic studies<sup>10–12</sup> of endogenous testosterone levels revealed genes that are associated with both testosterone levels and MPB. MPB has been associated with a greatly increased risk of scalp melanoma

A full list of affiliations appears at the end of the paper. ✉ e-mail: [JueSheng.Ong@qimrberghofer.edu.au](mailto:JueSheng.Ong@qimrberghofer.edu.au); [matthew.law@qimrberghofer.edu.au](mailto:matthew.law@qimrberghofer.edu.au)

(hazard ratio [HR] = 7.2, 95% confidence interval [CI]: 1.3–39.4) and SCC (HR = 7.1, 95% CI: 3.8–13.1)<sup>13</sup>. This is of interest as the numbers of head and neck melanomas have increased by over 50% between 1994 and 2015, primarily in males of European descent in the United States and Canada<sup>14</sup>. Melanomas of the head and neck, and particularly the scalp, are associated with higher mortality than other sites<sup>15,16</sup>. Across all studied populations, men experience higher rates of head and neck melanoma<sup>5</sup>. Cutaneous melanoma proliferation may be influenced by androgens either directly or through immune system suppression<sup>2,17,18</sup>, providing a potential link between MPB and cutaneous melanoma risk. A recent observational study reported a positive association between free testosterone and melanoma risk<sup>19</sup>, but whether this relationship is potentially mediated through MPB remains unknown. For KCs, a previous Canadian study revealed that the incidence of KC is higher in men<sup>20</sup>. Another retrospective cohort study from Australia showed potential sex differences in the incidence of KC in different body sites, where men are more likely to develop SCC in the scalp region<sup>21</sup>. To the best of our knowledge, there have been no large-scale studies to date evaluating the link between testosterone and KCs to dissect sex differences in KC incidences in these populations.

The increased incidence of skin cancers amongst men with MPB may: a) reflect higher chronic UV damage to the exposed scalp; b) result from a direct causal role for male-specific factors, including testosterone (consistent with the proposed androgen basis of melanoma hypothesis<sup>13</sup>); or c) be driven by other factors related to androgenic regulation of immune response. Leveraging the availability of large-scale genetic data Mendelian randomization (MR)<sup>22</sup>, a powerful genetic-based instrumental variable technique relying on the use of strong genetic proxies, can be used to detect and dissect causal pathways between MPB and skin cancers. As the allocation of genetic variants (namely single nucleotide polymorphisms, SNPs) is randomized at meiosis, the measured causal effect is less likely to be biased by reverse causality and confounding factors<sup>23</sup>. When used in combination with multivariable techniques<sup>24</sup>, we can extend MR frameworks

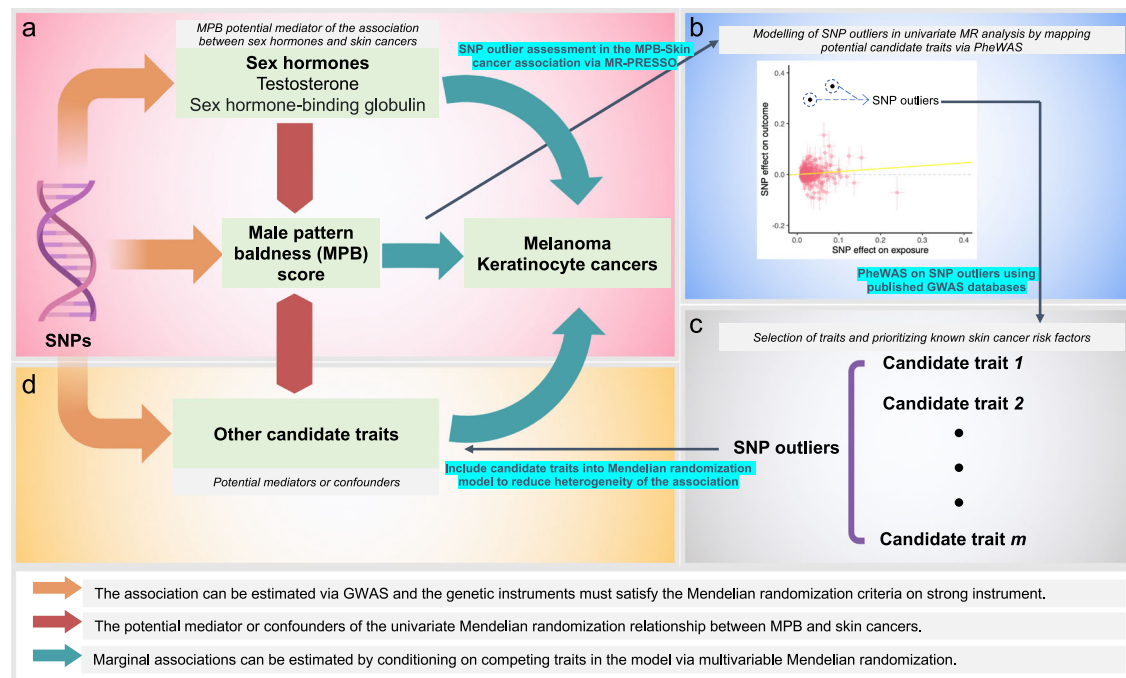
to disentangle the complex (and potentially mediated) relationship between endogenous testosterone, MPB and the risk of skin cancers.

In this study, we examined the link between endogenous testosterone levels, MPB, and skin cancers through a series of univariable and multivariable MR (MVMR) analyses. Using skin cancer outcome data from the UK Biobank (UKB)<sup>25</sup>, the QSkin Sun and Health Study (QSkin)<sup>26</sup>, and a recent large-scale cutaneous melanoma meta-analysis<sup>27</sup>, we comprehensively assessed the relationship between genetic instruments for MPB, sex hormones, and skin cancer susceptibility, followed by an exploratory analysis of whether these associations differed by anatomical location (body site) of the primary cancer site.

## Results

### Methodology overview

In brief, we adopted a two-sample MR framework<sup>22</sup> to investigate the relationship between testosterone, MPB and risk of KCs (all KC, SCC only and BCC only) and melanoma. The skin cancer genome-wide association studies (GWASs) were derived from a recent melanoma GWAS meta-analysis<sup>27</sup> (participating studies in Supplementary Table 1) and the Australian QSkin cohort<sup>26</sup> and UKB<sup>25</sup> for KCs. GWAS analyses of risk factors (sex hormones) were conducted in subsets of UKB independent of those used in the melanoma GWAS meta-analysis to increase the robustness of MR findings. We first applied standard univariable MR techniques to examine the direct association between hormonal risk factors, MPB and skin cancers, followed by a multivariable approach to fit all risk factors simultaneously whilst including proxy traits captured by heterogeneous genetic outliers (see Fig. 1 for the complete study schematic). We finally evaluated site-specific cancers in a subset of cases from UKB and Melanoma Institute Australia (MIA) cohorts to assess whether these MR associations differed by primary tumour anatomic site (i.e. head and neck) (distribution of cases by anatomic site provided in Supplementary Tables 2–4; classifications in Supplementary Fig. 1), as reported in previous observational studies<sup>21</sup>.



**Fig. 1 | Schematic diagram outlining the overall study approach of modelling genetic outliers via MVMR.** Each panel (a), (b), (c) and (d) is listed in chronological order of the analysis procedure. **a** Schematic MR diagram. **b** MR scatter plot. **c** Selection of candidate traits for inclusion into MVMR via PheWAS findings.

**d** Modelling the candidate traits into the MVMR analysis to obtain the marginal effect of MPB on skin cancer risk, by conditioning on endogenous testosterone levels and other candidate traits.

### Assessment of statistical power for Mendelian randomization

We identified a total of 444 SNP instruments for male-pattern baldness explaining 13.95% of the phenotypic variance. Similarly, 118, 83 and 204 SNPs were used as genetic instruments for endogenous totalT, (estimated) freeT and SHBG in the univariable MR analyses, with these SNPs cumulatively explaining, 8.13, 4.94 and 14.5% of the phenotypic variances on these traits respectively. The comparison of our testosterone SNP associations (estimated in the non-overlapping samples) with those obtained in the testosterone GWAS reported in Ruth et al.<sup>11</sup> is shown in Supplementary Figs. 2–4, revealing highly concordant genetic effect sizes. Even when we restrict the analysis to SNPs with an association of  $p < 1 \times 10^{-5}$  with the exposure of interest within our independent UKB subset (halved the original discovery GWAS sample size), the cumulative variances explained by SNPs were largely unaffected (Table 1). With the relatively high proportion of variance tagged by our SNPs, our MR study has very good power (at least 90%) to detect associations at OR > 1.2 for a one SD change in the aforementioned risk factors (Supplementary Fig. 5). The conditional F-statistics (to quantify instrument strength of our combined instrument) in the MVMR setting for each tested combination of models is shown in Supplementary Table 5.

### Primary MR analyses evaluating the MPB association with risk of skin cancers

For melanoma, none of the risk factors were associated with the risk of melanoma, with OR point estimates close to one (estimated IVW OR between 0.96 to 0.99) (Table 2). For KCs, the estimated OR per SD increase in MPB score was 1.17 (95% confidence interval [CI] 1.08–1.27) for risk of all KC, 1.15 (95% CI 1.05–1.26) for BCC and 1.31 (95% CI 1.17–1.46) for SCC in the UKB. Estimates from QSkin yielded slightly smaller effect sizes, albeit with largely overlapping confidence intervals (Table 2). The fixed-effect meta-analysed MR estimates for MPB on KC outcomes across both cohorts were: KC 1.15 (95% CI 1.06–1.23), BCC 1.15 (95% CI 1.06–1.25), and SCC 1.28 (95% CI 1.15–1.43). We detected no strong association between endogenous sex hormone levels (totalT, freeT, SHBG) and any KC outcomes (Table 2). Results derived using alternative MR models are shown in Supplementary Tables 6–9 for each exposure, respectively. Estimates derived using only SNP instruments that were more robustly associated ( $p < 1 \times 10^{-5}$ ) with the exposure of interest measured in the independent UKB subset were not meaningfully different (Table 2; Scatter plot in Fig. 2).

MVMR analyses combining MPB and testosterone levels revealed no evidence that the marginal association between MPB and KCs is influenced by endogenous testosterone levels; marginal OR on KC: 1.16 (95% CI 1.06–1.28); BCC: 1.16 (95% CI 1.04–1.29); SCC: 1.30 (95% CI 1.13–1.48) per SD increase in MPB score (Supplementary Table 10).

### Sensitivity analyses to model the potential pleiotropic role of pigmentation-related SNPs

It is possible that some MPB SNPs are pleiotropically associated with skin cancer risk factors. To address this, we first adopted a non-parametric approach to identify potential pleiotropic variants captured by the MR-PRESSO outlier test<sup>28</sup>. We detected four potential SNP-outliers in the association between MPB and KC phenotypes (rs2669871 [near *KRT75*], rs3847069 [near *CUX1*], rs1805007 [non-synonymous functional SNP in *MC1R*], rs12203592 [functional SNP in an enhancer for *IRF4*<sup>29</sup>]); outliers were also confirmed via manual inspection of MR scatter plots (Fig. 3) and funnel plots (Supplementary Figs. 6–9).

PheWAS revealed strong associations between these SNPs and pigmentation traits, including skin colour and ease of skin tanning (Supplementary Table 11; Supplementary Figs. 10–13). Univariable MR analyses support a positive relationship between lighter skin or hair colour and MPB (beta for lighter skin colour on MPB Score=0.28 [95% CI 0.14–0.42] per SD unit increase in MPB; beta for lighter hair colour

**Table 1 | Assessment of instrument strength on major exposures of interest for the two-sample MR analyses**

Phenotype	unit measurement	Using all variants as SNP instruments			Using only variants associated with exposure at $p < 1e-5$ in indep. subset			Reference
		F-stat	nSNPs	Prop. of variance explained	nSNPs	Prop. of variance explained		
Male pattern baldness (MPB)	per SD change in self-reported MPB score	1800.72	473	0.140	287	0.123	–	
Total testosterone (totalT)	per SD change in measured endogenous total testosterone	981.79	174	0.081	117	0.074	Ruth et al.	
Predicted free testosterone (freeT)	per SD change in derived endogenous free testosterone	576.95	101	0.049	78	0.047	Ruth et al.	
Sex hormone-binding globulin (SHBG)	per SD change in measured endogenous SHBG	1880.57	274	0.145	266	0.144	Ruth et al.	

nSNPs refer to the total number of variants considered in the MR analyses after the harmonisation process. F-stat refers to the F-statistics of the association between the combined SNP instrument and the exposure (phenotype) of interest. See supplementary Data 1 for the individual SNP effect estimates for each exposure of interest.

**Table 2 | Validation of MR association between risk factors and skin cancer risk using SNP instruments derived from independent non-overlapping subset of UKB**

Cancer outcome	Risk factors	With all SNP instruments identified from entire UKB cohort				With only SNP instruments that has $p < 1e-5$ in independent UKB dataset							
		$P_{UKB}$	$OR_{UKB}$	$P_{OSKIN}$	$OR_{OSKIN}$	$P_{META}$	$OR_{META}$	$P_{UKB}$	$OR_{UKB}$	$P_{OSKIN}$	$OR_{OSKIN}$	$P_{META}$	$OR_{META}$
all KC	MPB	1.45E-04	1.17 (1.08 to 1.27)	0.78	1.03 (0.85 to 1.24)	2.94E-04	1.15 (1.06 to 1.23)	3.86E-03	1.16 (1.05 to 1.28)	0.35	1.11 (0.89 to 1.37)	2.61E-03	1.15 (1.05 to 1.26)
	TotalIT	0.83	1.01 (0.93 to 1.1)	0.3	1.15 (0.88 to 1.5)	0.61	1.02 (0.94 to 1.10)	0.81	1.01 (0.93 to 1.10)	0.32	1.15 (0.87 to 1.52)	0.6	1.02 (0.94 to 1.11)
	FreeT	0.61	0.96 (0.83 to 1.12)	0.89	0.97 (0.66 to 1.44)	0.6	0.96 (0.84 to 1.11)	0.61	0.96 (0.81 to 1.13)	0.74	0.94 (0.63 to 1.39)	0.55	0.95 (0.82 to 1.11)
	SHBG	0.22	1.05 (0.97 to 1.12)	0.59	1.05 (0.88 to 1.24)	0.18	1.05 (0.98 to 1.12)	0.22	1.05 (0.97 to 1.12)	0.59	1.05 (0.88 to 1.25)	0.18	1.05 (0.98 to 1.12)
BCC	MPB	1.84E-03	1.15 (1.05 to 1.26)	0.49	1.1 (0.84 to 1.43)	1.32E-03	1.15 (1.06 to 1.25)	0.02	1.14 (1.02 to 1.28)	0.31	1.17 (0.87 to 1.58)	0.01	1.15 (1.03 to 1.27)
	TotalIT	0.37	1.04 (0.95 to 1.15)	0.29	1.18 (0.86 to 1.62)	0.25	1.06 (0.96 to 1.15)	0.45	1.04 (0.94 to 1.15)	0.5	1.12 (0.81 to 1.55)	0.36	1.05 (0.95 to 1.15)
	FreeT	0.85	0.98 (0.83 to 1.16)	0.75	0.93 (0.57 to 1.5)	0.78	0.98 (0.84 to 1.14)	0.77	0.97 (0.81 to 1.17)	0.54	0.87 (0.55 to 1.37)	0.62	0.96 (0.81 to 1.14)
	SHBG	0.13	1.07 (0.98 to 1.16)	0.39	1.11 (0.88 to 1.39)	0.08	1.07 (0.99 to 1.15)	0.12	1.07 (0.98 to 1.16)	0.35	1.11 (0.89 to 1.40)	0.08	1.07 (0.99 to 1.16)
SCC	MPB	2.35E-06	1.31 (1.17 to 1.46)	0.65	1.08 (0.77 to 1.53)	4.25E-06	1.28 (1.15 to 1.43)	1.33E-04	1.29 (1.13 to 1.48)	0.48	1.15 (0.77 to 1.72)	1.19E-04	1.28 (1.13 to 1.45)
	TotalIT	0.24	0.92 (0.81 to 1.05)	0.31	1.24 (0.82 to 1.87)	0.41	0.95 (0.84 to 1.08)	0.49	0.95 (0.83 to 1.09)	0.42	1.19 (0.77 to 1.84)	0.68	0.97 (0.85 to 1.11)
	FreeT	0.46	0.92 (0.75 to 1.14)	0.82	1.07 (0.58 to 1.99)	0.54	0.94 (0.77 to 1.15)	0.59	0.94 (0.77 to 1.16)	0.79	1.09 (0.57 to 2.09)	0.67	0.96 (0.78 to 1.17)
	SHBG	0.92	0.99 (0.9 to 1.1)	0.99	1.00 (0.74 to 1.36)	0.93	1.00 (0.90 to 1.10)	0.86	0.99 (0.90 to 1.10)	0.96	1.01 (0.74 to 1.37)	0.88	0.99 (0.90 to 1.09)
Melanoma	MPB	-	-	-	-	0.84	0.99 (0.91 to 1.08)	-	-	-	-	0.38	0.96 (0.88 to 1.05)
	TotalIT	-	-	-	-	0.29	0.94 (0.84 to 1.05)	-	-	-	-	0.34	0.94 (0.84 to 1.06)
	FreeT	-	-	-	-	0.91	0.99 (0.81 to 1.20)	-	-	-	-	0.84	0.98 (0.78 to 1.22)
	SHBG	-	-	-	-	0.22	0.95 (0.88 to 1.03)	-	-	-	-	0.25	0.95 (0.88 to 1.03)

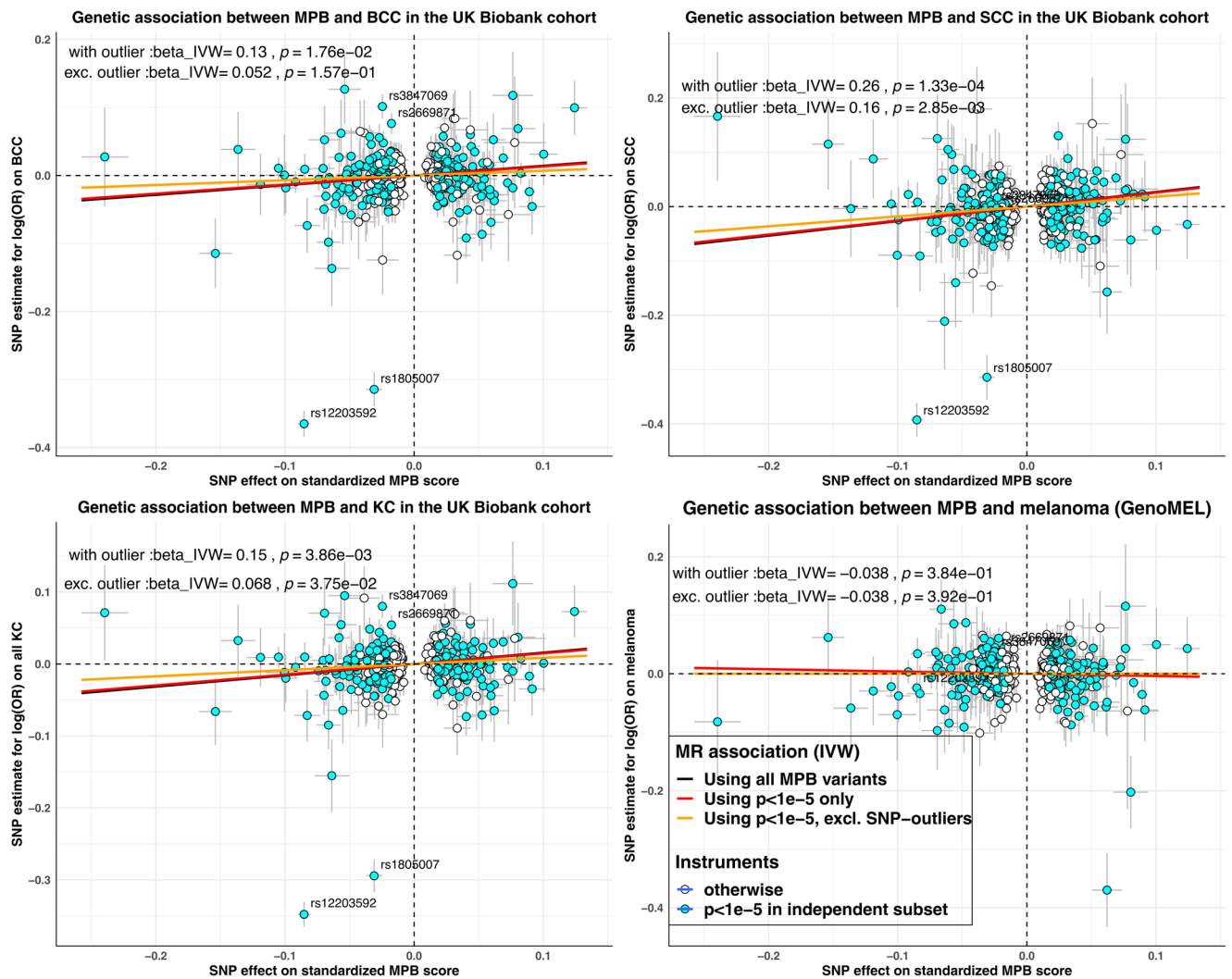
on MPB = 0.13 [95% CI 0.08–0.18], Supplementary Table 12), suggesting a potential common biological pathway. In our MVMR model incorporating pigmentation variables (hair and skin colour), the marginal association between MPB and KCs (including BCC and SCC separately) showed signs of attenuation towards the null (e.g. OR on KC 1.05 [95% CI 0.98–1.12] adjusted for pigmentation compared with original MVMR OR 1.16 [1.06–1.28]; see Fig. 3). These results suggest that the association between genetic variants associated with MPB and skin cancer body-wide might be driven by a pleiotropic effect on pigmentation. MVMR results for all tested combinations of the five traits (each model satisfying the minimal conditional-F > 10 requirement<sup>30</sup>) are shown in Supplementary Tables 13–15, each revealing similar marginal effect sizes from MPB.

In our second approach, we removed the four pleiotropic SNPs detected via MR-PRESSO from our instrument set and repeated our analyses. Both the revised findings from univariable, and MVMR models, revealed no clear evidence for an overall genetic association between MPB and skin cancers, apart from the association with SCC (MPB-SCC univariate MR estimate: OR 1.17 [95% CI 1.07–1.28],  $p = 8.06 \times 10^{-4}$ ; MVMR estimate: OR 1.15 [95% CI 1.02–1.28],  $p = 0.02$ ), see Figs. 2 and 3. We observed a large reduction in the univariable MR model's Cochran's Q statistics following exclusion of the four pleiotropic SNPs, indicating that the genetic heterogeneity among the SNP effect sizes on KC was much smaller in the revised model (see Supplementary Tables 16 and 17). The revised MR estimates from alternative MR models had largely overlapped 95% C.I. with estimates from the [IVW MR model excluding the SNP-outliers detected by MR-PRESSO] (Supplementary Table 18), indicating minimal evidence of horizontal pleiotropy biases on our findings.

**Stratified MR analysis between MPB and skin cancers by body sites**

**Melanoma.** For cutaneous melanoma, we obtained primary body-site-specific cancer data from the UKB and the Melanoma Institute Australia to investigate evidence for a primary site-specific association between MPB and melanoma. A higher genetic predisposition towards MPB was associated with increased risk of head and neck melanoma (meta-analysed OR 1.31 [95% CI 1.07–1.61] in the original model; OR 1.23 [95% CI 1.00–1.50] in the outlier-robust model). MPB was not associated with cutaneous melanoma at other body sites (Table 3). Similar positive findings were obtained when we evaluate melanoma at the scalp region specifically: estimated OR 1.66 [95% C.I. 0.99–2.80], but with signs of attenuation towards the null in the outlier-robust model estimate (OR 1.33 [0.79 – 2.25]) (Supplementary Table 19). Further splitting the association analysis by evaluating scalp melanoma and head and neck melanoma excluding the scalp region separately yielded very similar findings—indicating that the overall MR association between MPB and head and neck melanoma is primarily driven by melanoma on the scalp region (Supplementary Table 19). We also found no evidence that the MPB-melanoma association differed by Breslow thickness with the 95% confidence intervals on the OR estimates for both thick (OR 1.12 [0.89–1.39]) and thin (OR 1.04 [0.82–1.33]) melanoma largely overlapping (see Supplementary Table 20 for the ORs derived from alternative MR models).

**Keratinocyte cancers.** Combining data for site-specific BCC and SCC separately from the UKB and the QSkin cohort, we found limited evidence for an association between MPB and head/neck BCCs (OR 1.08 [0.99–1.17]) or SCCs (OR 0.92 [0.80–1.05]) using pleiotropy-robust MR models. The estimates for all other body sites were largely consistent with a null effect. MR estimates derived using alternative MR models show consistent effect sizes with the IVW estimate, albeit with much lower precision (Supplementary Table 19).



**Fig. 2 | Comparison of MR effect sizes for MPB on skin cancer risk.** beta\_IVW refers to the (fixed effect) inverse variance weighted MR effect estimate;  $p$ = corresponding two-sided z-test  $P$ -value of the beta\_IVW estimate; MPB Male pattern baldness. Each panel illustrates the MR scatter plot for the association between genetically predicted MPB score and the log(OR) on various types of skin cancer. Each point represents a single MPB SNP instrument, with the corresponding horizontal and vertical error bars reflecting its standard error on the genetic association with MPB and skin cancers, respectively. Points in light blue are MPB variants

that showed stronger evidence of association in the UKB subset independent of those used to derive the SNP-skin cancer association. The annotated SNPs on each panel refers to the SNPs identified as outliers via the MR-PRESSO outlier test. Apart from melanoma (right-bottom panel; which shows predominantly null findings), the IVW effect estimates derived using MPB variants after excluding the genetic outliers show strong attenuation of effect sizes towards the null. Source data are provided as a Source Data file.

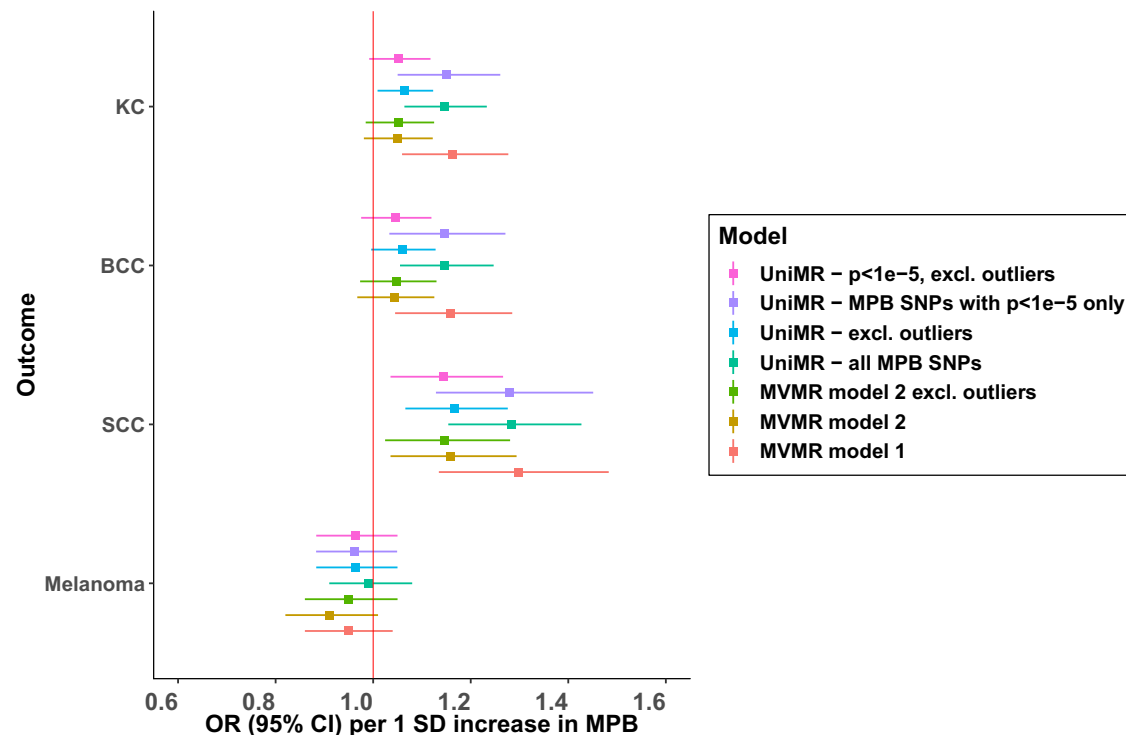
## Discussion

Melanomas and skin cancers occur with unequal distributions across the body surface, and depending on the country, at higher rates amongst men. While this distribution may be explained by patterns of sun exposure, it has also been postulated that differences in hair covering<sup>5,31</sup>, and possibly other hormonal factors<sup>5</sup>, may also contribute. We took advantage of several large genetic cohorts that collected site-specific melanoma and skin cancer data, and first showed that testosterone does not play a role on skin cancer susceptibility, a finding supported by our site-specific MR analyses indicating potential increased risk of scalp melanoma among individuals with a high genetic risk for balding. We presented evidence for an association between MPB and risk of SCC and finally revealed insights on mediation mechanisms on the link between MPB and these skin cancers through MVMR.

In Australia, the 2015 age-standardised rates of melanoma per 100,000 were higher for men (63.1) than for women (42.0)<sup>32</sup>. Whilst this trend is also observed in the United States, New Zealand and

Canada, the differences were lower or negligible in other countries (e.g. United Kingdom, Denmark and Sweden), and varied by primary body site<sup>5</sup>. Arguments for the involvement of sex hormones in melanoma have primarily stemmed from the observed difference in disease prevalence between men and women, prompting an investigation into potential sex-specific mechanisms that stimulate the proliferation of melanoma cells. Animal studies and in-vivo experiments have suggested that products of testosterone (i.e. progesterone) inhibit melanoma cell growth in a dose-dependent fashion, but there are very few large-scale epidemiological studies exploring this. Of note, our genetic-based approach failed to replicate the association between free testosterone and melanoma susceptibility reported in a recent cross-sectional study<sup>19</sup> using data from the UK Biobank. Here we found no evidence for a causal role between endogenous sex hormone levels (SHBG, freeT, totalT) and skin cancers. Interestingly, our multivariable MR model which enables the estimation of marginal effect sizes for these risk factors on cancer outcomes (upon conditioning on the other competing trait)

### Comparison for the MR association between genetically predicted MPB and risk of skin cancers derived from various univariate and multivariate MR models



**Fig. 3 | Comparison of MR-derived association between MPB and various univariable and MVMR models.** UniMR – MPB SNPs with  $p < 1e-5$  only: Univariable MR model for MPB on skin cancer risks using on MPB SNPs with association (z-score) two-tailed  $P$  value  $< 1e-5$  on MPB in the independent UK Biobank subset. UniMR – excl. outliers: Univariable MR model for MPB on skin cancer risks using all MPB SNPs excluding pleiotropic SNPs detected by MR-PRESSO. MVMR Model 1: MVMR model incorporating MPB, freeT and totalT. MVMR model 2: MVMR model incorporating MPB, totalT, skin colour and hair colour. The error bars reflect the

95% confidence intervals around each OR estimate. For all MVMR models, the reported OR estimates are the marginal OR estimates for skin cancer per 1 SD increase in MPB upon conditioning on the genetic effect sizes from other traits included in the model. Note that for MVMR model 2, including freeT into the model resulted in the weakening of the combined instrument for MVMR (conditional F-statistics  $< 10$  for some traits), which might result in weak instrument bias and hence were not included in the main analysis. However, these findings can be accessed in Supplementary Tables. Source data is provided as a Source Data file.

yielded similar conclusions of a null association between testosterone and skin cancers.

Among the three skin cancers evaluated, SCC showed the strongest association with MPB, followed by BCC. However, we found that the positive association between MPB and risk of KCs was almost completely driven by a functional SNP in the *IRF4* locus (rs12203592). This variant was also previously known to be associated with pigmentation and immune response<sup>29,33–36</sup>, thus may influence risk of skin cancers through pathways other than its impact on MPB. Disentangling the observed association between MPB with skin cancers requires careful consideration of both potential causal and pleiotropic mechanisms in play. While the practice of excluding heterogeneous variants and SNP outliers to mitigate potential horizontal pleiotropy bias in the outcomes is a valid approach, we adopted an alternative strategy. Specifically, we chose to model these pleiotropic associations with other traits using MVMR<sup>37</sup>. This decision was motivated by our aim to gain a more comprehensive understanding of potential mediators and genetic confounders that may underlie the relationship between MPB and the development of skin cancers.

In our investigation, we demonstrate how the pleiotropic variants, such as the detected SNP-outlier rs12203592 located in an enhancer for the *IRF4* gene may help explain the observed relationship between MPB and skin cancers (See Fig. 2). For instance, the role that *IRF4* plays in transforming pigmented terminal hair into unpigmented vellus hair is widely established<sup>29,34</sup>, suggesting a potential causal role for pigmentation on MPB. rs12203592 has been consistently identified by prior GWAS of nevus count, hair colour, development of freckles and

skin pigmentation, all of which are established risk factors for skin cancers<sup>10,29,38,39</sup>. Due to the strong magnitude of association between variants in *IRF4* and both hair loss and pigmentation, it is difficult to determine the type of genetic pleiotropy (vertical or horizontal) exerted by this variant on MPB and skin cancers (i.e. whether the association between *IRF4* on skin cancers were through change in baldness). When pigmentation variables were included in our MVMR model, the association between MPB and KC weakened; though this was also observed when we excluded *IRF4* from the revised univariable MR models. Hence, the association between MPB and overall KC is likely capturing an indirect influence of *IRF4* on both pigmentation-related variables (e.g. skin colour, nevus counts) and risk of balding (with the mode of pleiotropy cannot be reliably determined), through potential pleiotropic effects on autoimmune functions<sup>33</sup> which was not characterised in the present analysis.

As our MR instruments for MPB explain large proportions of phenotypic variances, we had reasonable power to revisit previous observational findings on primary site-specific skin cancer associations with MPB, or at the very least, exclude very large OR effect sizes. Our site-specific MR analysis revealed two interesting key findings. Firstly, the estimated effect size on melanoma located in the head and neck region (based on IVW model excluding SNP outliers) was larger than those of other body sites for both the MIA and the UKB cohorts (Table 3), supporting the role of balding in exposing the scalp area of the skin to UVR radiation. Taken together with our MR findings on testosterone, this demonstrates that testosterone had no direct role in skin cancer formation (i.e. its only role is to remove the natural

**Table 3 | Comparison of estimated MR body-site-specific association between MPB and skin cancer in the UK and Australian population**

Cancer outcome by anatomical category	Study (UK)		Study (Aus)		Fixed-effect meta				
	Ncases <sub>UK</sub>	P <sub>UK</sub>	OR <sub>UK</sub>	Ncases <sub>Aus</sub>	P <sub>Aus</sub>	OR <sub>Aus</sub>	P <sub>META</sub>	OR <sub>META</sub>	
<i>Head and neck region</i>									
Melanoma (all MPB variant)	390	0.17	1.20 (0.92 to 1.57)	MIA	537	0.02	1.48 (1.07 to 2.04)	0.01	1.31 (1.07 to 1.61)
Melanoma (MPB exc. outlier)	390	0.24	1.18 (0.90 to 1.54)	MIA	537	0.10	1.30 (0.95 to 1.76)	0.05	1.23 (1.00 to 1.50)
SCC (all MPB variant)	1427	0.27	0.92 (0.80 to 1.07)	QSKIN	184	0.94	0.99 (0.65 to 1.49)	0.29	0.93 (0.81 to 1.06)
SCC (MPB exc. outlier)	1427	0.28	0.92 (0.80 to 1.07)	QSKIN	184	0.52	0.87 (0.57 to 1.32)	0.22	0.92 (0.80 to 1.05)
BCC (all MPB variant)	4768	0.12	1.07 (0.98 to 1.16)	QSKIN	439	0.12	1.25 (0.94 to 1.67)	0.05	1.08 (1.00 to 1.17)
BCC (MPB exc. outlier)	4768	0.34	1.11 (0.90 to 1.37)	QSKIN	439	0.36	1.14 (0.86 to 1.51)	0.08	1.08 (0.99 to 1.17)
<i>Upper limb/Arm region</i>									
Melanoma (all MPB variant)	446	0.01	0.72 (0.57 to 0.93)	MIA	280	0.59	1.10 (0.78 to 1.55)	0.08	0.84 (0.68 to 1.02)
Melanoma (MPB exc. outlier)	446	0.01	0.71 (0.56 to 0.92)	MIA	280	0.72	0.94 (0.68 to 1.31)	0.02	0.79 (0.65 to 0.97)
SCC (all MPB variant)	284	0.75	1.05 (0.77 to 1.44)	QSKIN	149	0.53	1.16 (0.73 to 1.86)	0.54	1.08 (0.84 to 1.41)
SCC (MPB exc. outlier)	284	0.91	1.02 (0.74 to 1.40)	QSKIN	149	0.92	1.02 (0.64 to 1.63)	0.88	1.02 (0.78 to 1.32)
BCC (all MPB variant)	636	0.47	0.93 (0.75 to 1.14)	QSKIN	248	0.49	1.15 (0.78 to 1.68)	0.77	0.97 (0.81 to 1.17)
BCC (MPB exc. outlier)	636	0.36	0.91 (0.73 to 1.12)	QSKIN	248	0.99	1.00 (0.69 to 1.46)	0.43	0.93 (0.77 to 1.12)
<i>Trunk region</i>									
Melanoma (all MPB variant)	708	0.20	0.89 (0.75 to 1.06)	MIA	1060	0.97	1.00 (0.78 to 1.30)	0.30	0.93 (0.80 to 1.07)
Melanoma (MPB exc. outlier)	708	0.15	0.88 (0.74 to 1.05)	MIA	1060	0.81	0.97 (0.76 to 1.24)	0.19	0.91 (0.79 to 1.05)
SCC (all MPB variant)	209	0.74	1.05 (0.77 to 1.43)	QSKIN	23	0.98	0.99 (0.32 to 3.03)	0.76	1.05 (0.78 to 1.41)
SCC (MPB exc. outlier)	209	0.92	1.02 (0.74 to 1.39)	QSKIN	23	0.71	0.81 (0.26 to 2.51)	1.00	1.00 (0.74 to 1.35)
BCC (all MPB variant)	1350	0.14	1.11 (0.97 to 1.28)	QSKIN	209	0.90	0.97 (0.66 to 1.45)	0.18	1.10 (0.96 to 1.25)
BCC (MPB exc. outlier)	1350	0.21	1.10 (0.95 to 1.26)	QSKIN	209	0.73	0.93 (0.63 to 1.39)	0.28	1.08 (0.94 to 1.23)
<i>Lower limb region</i>									
Melanoma (all MPB variant)	239	0.27	0.84 (0.62 to 1.14)	MIA	361	0.55	0.90 (0.65 to 1.26)	0.22	0.87 (0.69 to 1.09)
Melanoma (MPB exc. outlier)	239	0.33	0.86 (0.63 to 1.17)	MIA	361	0.55	0.93 (0.72 to 1.19)	0.17	0.85 (0.68 to 1.07)
SCC (all MPB variant)	109	0.60	0.88 (0.56 to 1.39)	QSKIN	102	0.48	1.24 (0.68 to 2.26)	1.00	1.00 (0.70 to 1.43)
SCC (MPB exc. outlier)	109	0.66	0.90 (0.57 to 1.42)	QSKIN	102	0.93	0.98 (0.55 to 1.75)	0.69	0.93 (0.65 to 1.33)
BCC (all MPB variant)	269	0.85	1.03 (0.76 to 1.39)	QSKIN	93	0.63	0.87 (0.49 to 1.54)	0.95	0.99 (0.76 to 1.30)
BCC (MPB exc. outlier)	269	0.90	1.02 (0.75 to 1.39)	QSKIN	93	0.27	0.72 (0.40 to 1.29)	0.69	0.95 (0.72 to 1.24)

For the MR analyses, 96620, 3176 and 1957 controls were used for the UKB, MIA and QSKIN analyses respectively. P-values here were derived from two-tailed z-tests on the MR association estimate (in log(OR)), unadjusted for multiple testing. SNP outliers for the repeated analyses excluding potentially pleiotropic/heterogenous SNPs) were determined using the MR-PRESSO outlier test at Bonferroni-corrected P value of 0.05. The model MPB exc. outlier refers to the model utilizing MPB variants excluding the four SNP outliers identified by MR-PRESSO.

protection from UV - hair). We observed the same trend for head and neck BCC (albeit with lower precision), where the magnitude of association was slightly higher for the QSkin cohort compared to the UK Biobank study (Table 3), making it unclear whether these MPB-BCC associations varied between populations in low-UVR (UK) and high-UVR (Australia) geographical regions. For melanoma, we found a strong association between MPB and melanoma at the head and neck region; combined with MPB being a trait that is specific towards men, might help explain the higher prevalence of head and neck melanoma in men as previously reported. However, the estimated effect size between balding and melanoma/KC at the head and neck region (predominantly scalp melanoma) in our datasets (e.g. equivalently head and neck melanoma OR 1.72 [1.14–2.59] per 2 SD increase in MPB score) appeared to be much more modest than those previously reported by Li and colleagues (scalp melanoma HR 7.15 [1.29–39.42] for highest balding category vs. none)<sup>13</sup>, though our estimates had much greater precision. Another potential explanation may be that observational associations between balding and skin cancers may be amplified by ascertainment and detection bias, where tumours at the scalp or forehead region are more apparent among patients with MPB.

This study has several notable strengths as compared to previous studies of similar nature. Firstly, the proportion of phenotypic variance explained by our genetic instruments were high ( $r^2$  estimates 0.04–0.15; Table 1), attributable to the largely heritable nature of both MPB and sex hormone phenotypes. Combined with the very large sample sizes for melanoma and KCs (approximately 2–4 times larger than those previously reported in any MR findings<sup>40–42</sup>), our statistical power to detect even subtle effect sizes (e.g. OR < 1.3 per SD change in exposure trait) were greatly improved. The polygenic basis of both exposures also enables better assessment of horizontal pleiotropy through various alternative MR techniques, ensuring that our main findings were not severely biased by violation of MR assumptions. In addition, the MVMR enables triangulation of a mediation effect for the association between testosterone or other putative biological pathways in our sensitivity analyses and skin cancers. Finally, we used MR to interrogate whether the association between MPB and skin cancers differed by body site, to compare with previous observational findings.

There are also some limitations to our study. The grading of MPB in the UKB was acquired through participant's self-report; hence we cannot exclude the possibility of self-reporting biases arising from negative social stigma<sup>43</sup>. In practice, the misclassification of MPB scores is unlikely to be directly associated with skin cancer outcomes, as that would generally just reduce the statistical power for (genetic) instrument identification. Hence, the technical difference arising from alternative classifications of MPB is not a potential concern in our MR study. The derived MR finding on melanoma using summary statistics from the melanoma GWAS meta-analysis consisted of participants from the UKB ( $n = 7782$  men), making up ~25% of the total number of cases included in the meta-analysis. However, given that the proportion of overlapping samples is less than 30%, any generated bias on the MR estimate is likely limited<sup>44</sup>. Moreover, we repeated our analyses using only subsets of the UKB participants not involved in the skin cancer GWASs and showed that our findings remain broadly unchanged. Our MR findings assume a linear relationship between balding and cancers, and hence might under-estimate the true effect size if the dose-response relationship violates the linearity assumption<sup>45</sup>. Whilst our analysis on the MIA cohort revealed no evidence that the MPB-melanoma association differed by Breslow thickness, we were unable to repeat this analysis in the UK Biobank as information on tumour thicknesses have not been recorded. Our subgroup analyses revealed evidence that this association is driven by melanoma in the scalp. Whilst this supports our inference on increased UV exposure, we did not have the necessary information/sample size to replicate this stratified MR analysis in the UK Biobank and QSkin cohort. Finally, further external validations will also be required to assess whether these

genetic instruments can be applied to probe the balding-skin cancer relationship in other non-European ancestries.

Our approach of efficiently incorporating outlier information as candidate traits in our MVMR model is inspired by the MR-TReasure Your eXceptions (MR-TRYX) framework<sup>37</sup>. The key difference here, however, is that we restricted our candidate traits to those involved in pigmentation whilst the MR-TRYX model exhaustively examines all possible risk factors through a generic PheWAS platform. There may be more optimal candidates, such as immunological factors (e.g. eosinophil count, see Supplementary Table 11), which we did not consider in our approach. In practice, we are largely limited by the number of traits we can simultaneously model in an MVMR framework due to difficulty satisfying the conditional F-stat requirement as we increase the number of traits in the model, distribution of trait-specific conditional F-stat in all our tested MVMR models shown in Supplementary Table 5. Finally, our sample sizes in the body site-stratified analyses were limited and were only drawn from regions with two extreme ends of very low (UK) and very high ambient UV radiation (Australia). It remains unclear on whether our findings can be generalised onto other populations. Hence, replicating our site-specific findings on other populations with moderate UV radiation will help ensure our findings are generalisable.

In conclusion, genetic evidence in this study provides minimal support to the androgen-driven hypothesis linking sex hormones to the development of melanoma and keratinocyte cancers. Pigmentation-related factors very likely mediate the genetic relationship between balding and KCs at all body sites, evident through MVMR findings. Finally, we observed a modest body-site specific association between MPB and both the risk of melanoma and KCs involving the head and neck region greater than other body sites, suggesting that balding might increase susceptibility for melanoma around the head and neck region through reduced hair covering, a potential explanation for sex-differences in head and neck melanoma risk between men and women.

## Methods

### Ethical approval and patient consent

The UK biobank study has been formally approved by the UK Biobank Ethics Advisory Committee. The Qskin and AGDS study (used as controls for the MIA case-control GWAS) has been formally approved by the QIMR Berghofer Human Research Ethics Committee. The MIA study was approved by the Sydney Local Health District Ethics Review Committee at the Royal Prince Alfred Hospital in Sydney, Australia, respectively. This research project is approved by the QIMR Berghofer Human Research Ethics committee. The complete list of HREC involved in the individual studies contributing to the melanoma GWAS meta-analysis can be found in Landi et al.<sup>27</sup> (see Supplementary Notes 1). Written informed consent was obtained from all participants from all studies. The authors confirm that patient data has been obtained according to the terms and conditions of the databases where the data was sourced. More specifically, we have obtained specific permission from the MIA investigators to utilise anonymised genetic and patient data from the MIA study for this project.

### Construction of genetic instruments from the UKB

UKB is a large population-based cohort consisting of predominantly middle-aged (at the time of recruitment) white British participants recruited in the United Kingdom. Participants were genotyped (genetic QC and details available elsewhere<sup>25</sup>) and had extensive phenotypic information collected ranging from self-reported diet and lifestyle behaviour to measurements of disease biomarkers, mental health, medical history and cancer diagnosis. GWAS findings for both MPB and endogenous testosterone levels derived from UKB have been previously reported<sup>10–12</sup>.



**MPB.** Data for MPB (UKB-field ID 2395) were available for 227,354 male participants. In the UKB, MPB was self-reported and defined using a 4-point scale of increasing hair loss (see Supplementary Notes 2). We rank-transformed the ordinal MPB scales into standardised Z-scores. Analyses were restricted to only male participants of white British ancestry inferred through ancestral principal component (PC) clusters. We performed two GWASs on the rank-transformed MPB score using: i) all 227,354 available participants for instrument discovery; and ii) the subset of only 90,577 participants not overlapping the UKB KC GWASs.

**Sex hormones.** Endogenous serum total testosterone level (totalT) and sex-hormone-binding globulin (SHBG) were measured as part of the 2019 UKB biochemistry data release. Although publicly available GWAS summary statistics for total and free testosterone for UKB participants are available<sup>11</sup>, we adopted a similar phenotype definition to those from Ruth et al.<sup>11</sup> for the derivation of testosterone levels (including bioavailable/free-testosterone [freeT]). The derivation for serum freeT using serum totalT, albumin and SHBG in UKB is described in Supplementary Notes 3. Each GWAS on sex hormones was performed in individuals of white British ancestry using BOLT-LMM v2.3, a linear mixed model GWAS framework that accounts for population structure and cryptic relatedness among samples<sup>46</sup>. We fitted recruitment age, genotype array and the first 20 ancestral PCs as standard covariates for all analyses.

**Re-estimation of SNP effect sizes free from sample overlap.** To obtain unbiased MR estimates that minimize sample overlap in the 2-sample setting<sup>44</sup>, we repeated GWAS analyses for MPB and sex hormones using only a subset of the UKB participants independent of those used for the KC GWASs (see below, also Supplementary Notes 4).

### Sex-stratified GWAS summary statistics for melanoma

**All cutaneous melanoma.** Summary statistics from a recent fixed-effects GWAS meta-analysis of clinically confirmed cutaneous melanoma in men only were included<sup>27</sup>. All samples were clinically confirmed cutaneous melanoma. While the majority were invasive melanoma, specific histological subtype data was not available for all sets, and the GWAS meta-analysis will include a small subset of cases with in situ melanomas. Full details describing the participating studies, of analyses, collection of informed consent and ethical approval have been previously reported<sup>27</sup>. Briefly, the following standard GWAS cleaning procedures were performed: exclusion of samples (i) with >3% missingness (ii) with abnormal genotype heterozygosity (iii) was a European ancestry outlier or (iv) related to another sample at identity-by-descent (IBD)  $PI\_HAT > 0.15$ . SNPs were also filtered where we removed genotypes with a minor allele frequency (MAF)  $< 0.01$  or those with Hardy-Weinberg equilibrium (HWE)  $P < 5 \times 10^{-4}$  in controls or  $5 \times 10^{-10}$  in cases prior to imputation; our post-imputation analyses were restricted to SNPs with a MAF  $> 0.005$  and imputation quality score  $> 0.3$ . A fixed-effects inverse-variance weighted meta-analysis of melanoma risk GWAS was then performed (each GWAS was modelled via logistic regression, including PCs or equivalent control for residual population stratification). The final male-only melanoma risk GWAS included 12,232 cases and 20,566 controls. The distribution of cases across each study is shown in Supplementary Table 1.

**Site-specific primary cutaneous melanoma.** For the site-specific analyses, we obtained site-specific histopathologically confirmed primary cutaneous melanoma diagnoses for men from the UKB and the MIA. In UKB, we derived site-specific primary melanoma diagnoses from ICD-10 codes C43.0-9 (UKB field-ID: 41270). For MIA, melanoma cases with primary site-specific diagnoses and GWAS data were drawn from two complementary data repositories, the MIA Biospecimen Bank (protocol HREC/10/RPAH/530) and Melanoma Research Database (protocol HREC/11/RPAH/444)). Among the 2,236 men diagnosed

with melanoma, more than 99.5% (2227/2236) of the cases were of (or had shown progression into) invasive melanoma. Site-specific primary melanoma cases were then matched against 3,176 melanoma-free controls from the Australian Genetics of Depression Study (AGDS)<sup>47</sup> for the site-specific melanoma GWAS analysis. Details of the definition of anatomical primary site (body sites) categories, data consent and site-specific GWAS analysis in both the UKB and MIA study are available in Supplementary Notes 1, with case distribution across body sites reported in Supplementary Tables 2 and 3. To further investigate whether the association differed by Breslow thickness, we further performed a stratified MR analysis contrasting thick and thin melanoma (see Supplementary Notes 1 under study description on the MIA cohort for the adopted definitions for thick and thin melanoma).

### Sex-stratified GWAS summary statistics for keratinocyte cancers

**KCs in UK Biobank.** We performed sex-stratified GWAS analyses for KC using participants from UKB. Using the case definitions as per previous work<sup>48</sup>, we identified 3,483 invasive SCC and 10,718 invasive BCC cases confirmed through ICD-10 and histology records (UKB Field-ID: 40006, 40013) among UKB male participants ascertained through linkage of participant health records to national cancer registries. Men with no prior history of KC or any other cancer diagnosis ( $n = 96,620$ ) were used as controls. Our analyses were restricted to individuals of white British ancestry identified through self-report and clustering approaches on ancestral principal components<sup>49</sup>. We performed the GWAS analyses for BCC and SCC using SAIGE<sup>50</sup>, a recently developed software that implements linear mixed models for binary traits/outcomes, accounting for case-control imbalance and cryptic relatedness. We also performed a GWAS of a combined KC phenotype (cases=13,463; controls=96,620) to leverage the shared genetic architecture between SCC and BCC for improved power<sup>48</sup>.

**KCs in QSkin.** In total, 4,049 men with post-quality-control genetic data were clinically diagnosed with KC, with 1,064 and 502 cases identified to have invasive BCC and invasive SCC based on pathological records, respectively. Healthy controls were selected from participants screened/self-reported to have no history of KC or actinic keratoses at the time of recruitment<sup>26</sup>. The cleaning and quality control of the genetic data for the QSkin cohort have been previously described<sup>27</sup>. We performed a sex-stratified GWAS for KC, BCC and SCC, including only male participants of white European ancestry using SAIGE<sup>50</sup>, adjusting for recruitment age and the first ten ancestral principal components.

**Site-specific KCs.** Additional body-site-specific GWASs for KCs in men were conducted in the UKB and QSkin cohort. Similarly, these analyses were restricted to those of white European ancestry. For UKB, body-site-specific diagnoses of BCCs and SCCs were obtained through ICD-10 codes C44.0-7. For the QSkin cohort, site data of the KCs were manually recovered through histopathological records. To enable comparison between the two cohorts, we collapsed the site-specific cancers into 4 broad categories (similar to those used for the melanoma analysis): head and neck, upper limb, lower limb and trunk region. Details of the defined categories and the number of cases and controls included for each body site are also presented in Supplementary Notes 5 and Supplementary Tables 2–4.

### Univariable Mendelian randomisation analysis

Prior to the MR analysis, we harmonised the exposures [totalT, freeT, SHBG, MPB] and outcomes (skin cancers) datasets to align alleles and discard palindromic SNPs with non-strand-inferable allele frequencies (MAF  $> 0.3$ ). Multi-allelic variants and variants with inappropriate standard errors (i.e.  $> 1$  decimal place smaller than those approximated via sample size and minor allele frequency<sup>51</sup>, for outcome datasets) were also excluded.

MPB SNP instruments were clumped to remove variants in linkage-disequilibrium (LD) (window = 10 megabase [Mb], max LD  $r^2 = 0.001$  in PLINK v1.96b<sup>52</sup>) to ensure strict independence. To control for bias from potentially weak instruments, we additionally curated instrument sets for MPB, totalT, freeT and SHBG to consist only of SNPs that were robustly associated with the corresponding traits at  $p < 1 \times 10^{-5}$  in the independent UKB subset (i.e. no sample overlap with UKB KC GWAS). Statistical power for MR was assessed based on the proportion of variance explained by these instruments (Supplementary Notes 6). The inverse variance-weighted (IVW) estimator was then used to derive the log(odds ratio) (log[OR]) estimates for skin cancers for a standard deviation (SD) increase in the exposure (e.g. genetically predicted endogenous testosterone levels). For SCC, BCC and KC outcomes (and the corresponding site-stratified analyses), we combined the MR log(OR) estimate derived from QSkin and UKB through a fixed-effect inverse variance-weighted model, using the *rmeta* package (v3.0) in the statistical software R.

Alternative MR techniques (MR-PRESSO, MR weighted median, MR weighted mode, MR-Egger, MR-Robust) that relax assumptions regarding horizontal pleiotropy and invalid instruments were also applied to ensure the robustness of our MR inferences<sup>28,53–55</sup>. The specific strength and limitations for each of these methods are reported in the Supplementary Notes 7. All primary MR analyses were performed using the *MendelianRandomization* R package and *TwoSampleMR* R package v0.4.23 curated by the MR-Base platform<sup>56,57</sup>. MR scatter plots were used to inspect statistical outliers, alongside the outlier-test implemented directly via MR-PRESSO<sup>28</sup> (see below).

### Modelling the influence of androgenic and other pleiotropic pathways

We explored the possibility that some MPB-associated SNPs might exert pleiotropic effects that modify the risk of developing skin cancers. However, excluding variants with potentially large effect sizes on the outcome based on (horizontal) pleiotropy is less efficient if information linking the outlier and other pleiotropic pathways can be incorporated via MR. This allows us to understand potential mediators and/or genetic confounders of the exposure-outcome relationship in MR<sup>37</sup>. Here we outline two complementary approaches; full details of these models are in Supplementary Notes 5. We first applied the MR-PRESSO<sup>28</sup> outlier test on MPB vs. skin cancers to identify potential outliers that can bias the MR-IVW results. We then performed phenome-wide association studies (PheWAS) to evaluate the association between the set of potentially pleiotropic SNPs and a series of pigmentation/skin-related phenotypes likely to confound any associations between MPB and skin cancer (see Supplementary Notes 8). This was done by querying the candidate SNPs against two GWAS databases, MRC-IEU OpenGWAS and the OpenTarget platform<sup>56,58,59</sup>. Based on the potential candidate SNP-trait association, we then assessed putative mediation mechanisms driving the association between MPB and skin cancer via the candidate trait. We first attempted to validate the relationship in a univariable MR framework, followed by a multivariable MR (MVMR) analysis incorporating the candidate trait alongside testosterone and MPB on skin cancer (See Fig. 1).

In our second approach, we manually excluded the set of pleiotropic variants identified in the MR-PRESSO outlier test altogether and repeated the univariable and MVMR analyses. Curation of the genetic instruments for the MVMR analysis and the selection criteria for traits to achieve strong instrument strength for the analysis are described in Supplementary Notes 9. The MVMR association analyses were performed using the *mv\_multiple()* function from the *TwoSampleMR* R package<sup>24,56</sup>.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The genetic summary data for the risk factors presented in this manuscript are primarily derived from the UK Biobank cohort (Supplementary Data 1). The UK Biobank phenotype and genetic data can be obtained by application directly to the UK Biobank. The sex-specific melanoma GWAS meta-analysis summary statistics can be obtained through direct application to the study principal investigators (M. H. Law [matthew.law@qimrberghofer.edu.au](mailto:matthew.law@qimrberghofer.edu.au); M. M. Iles [M.M.Iles@leeds.ac.uk](mailto:M.M.Iles@leeds.ac.uk); M. T. Landi [landim@mail.nih.gov](mailto:landim@mail.nih.gov)). The complete GWAS summary statistics data for KC derived from the QSkin study is available under restricted access, and the data access can be obtained by contacting the following QSkin principal investigators (D. C. Whiteman [David.Whiteman@qimrberghofer.edu.au](mailto:David.Whiteman@qimrberghofer.edu.au); C. M. Olsen [Catherine.Olsen@qimrberghofer.edu.au](mailto:Catherine.Olsen@qimrberghofer.edu.au)). The GWAS findings from the site-stratified melanoma data in the MIA is available under restricted access, and data access can be obtained by contacting the MIA investigator R.A. Scolyer ([EAscolyer@melanoma.org.au](mailto:EAscolyer@melanoma.org.au)). The summary statistics for SNPs analysed in this study are also provided in Supplementary Data 1. Source data are provided with this paper.

### Code availability

Results from this work were generated using generic open-source MR software codes/functions implemented in R (such as those offered in the R package *TwoSampleMR* v0.4.22, available here <https://github.com/MRCIEU/TwoSampleMR>) and *MendelianRandomization* v0.5.0 (available here <https://github.com/cran/MendelianRandomization>), illustrated using the *ggplot2* v3.2.1.R package. Source data are provided with this paper. No custom software codes are used to generate the reported findings. Our analysis code and a test example can be accessed here <https://doi.org/10.5281/zenodo.7988335>.

### References

1. Sturm, R. A. & Duffy, D. L. Human pigmentation genes under environmental selection. *Genome Biol.* **13**, 248 (2012).
2. Schwartz, M. R., Luo, L. & Berwick, M. Sex Differences in Melanoma. *Curr. Epidemiol. Rep.* **6**, 112–118 (2019).
3. Rampen, F. H. Sex differences in survival from cutaneous melanoma. *Int. J. Dermatol.* **23**, 444–52 (1984).
4. Liu, F. et al. A unique gender difference in early onset melanoma implies that in addition to ultraviolet light exposure other causative factors are important. *Pigment Cell Melanoma Res.* **26**, 128–135 (2013).
5. Olsen, C. M., Thompson, J. F., Pandeya, N. & Whiteman, D. C. Evaluation of Sex-Specific Incidence of Melanoma. *JAMA Dermatol.* **156**, 553–560 (2020).
6. Kuklinski, L. F. et al. Sex hormones and the risk of keratinocyte cancers among women in the United States: A population-based case-control study. *Int. J. Cancer* **139**, 300–309 (2016).
7. Zhang, Y., Xu, J., Jing, J., Wu, X. & Lv, Z. Serum Levels of Androgen-Associated Hormones Are Correlated with Curative Effect in Androgenic Alopecia in Young Men. *Med. Sci. Monit.* **24**, 7770–7777 (2018).
8. Kaufman, K. D. Androgens and alopecia. *Mol. Cell. Endocrinol.* **198**, 89–95 (2002).
9. Piraccini, B. M. & Alessandrini, A. Androgenetic alopecia. *G. Ital. Dermatol. Venereol.* **149**, 15–24 (2014).
10. Pirastu, N. et al. GWAS for male-pattern baldness identifies 71 susceptibility loci explaining 38% of the risk. *Nat. Commun.* **8**, 1584 (2017).
11. Ruth, K. S. et al. Using human genetics to understand the disease impacts of testosterone in men and women. *Nat. Med.* **26**, 252–258 (2020).
12. Yap, C. X. et al. Dissection of genetic variation and evidence for pleiotropy in male pattern baldness. *Nat. Commun.* **9**, 5407 (2018).

13. Li, W. Q., Cho, E., Han, J., Weinstock, M. A. & Qureshi, A. A. Male pattern baldness and risk of incident skin cancer in a cohort of men. *Int. J. Cancer* **139**, 2671–2678 (2016).
14. Bray, H. N. et al. Head and Neck Melanoma Incidence Trends in the Pediatric, Adolescent, and Young Adult Population of the United States and Canada, 1995–2014. *JAMA Otolaryngol. Head. Neck Surg.* **145**, 1064–1072 (2019).
15. Xie, C. et al. Impact of scalp location on survival in head and neck melanoma: A retrospective cohort study. *J. Am. Acad. Dermatol.* **76**, 494–498.e2 (2017).
16. Ding, Y. et al. Comparing the characteristics and predicting the survival of patients with head and neck melanoma versus body melanoma: a population-based study. *BMC Cancer* **21**, 420 (2021).
17. Morvillo, V. et al. Androgen receptors in human melanoma cell lines IIB-MEL-LES and IIB-MEL-IAN and in human melanoma metastases. *Melanoma Res.* **12**, 529–538 (2002).
18. Rampen, F. H. & Mulder, J. H. Malignant melanoma: an androgen-dependent tumour? *Lancet* **1**, 562–564 (1980).
19. Watts, E. L. et al. Prospective analyses of testosterone and sex hormone-binding globulin with the risk of 19 types of cancer in men and postmenopausal women in UK Biobank. *Int. J. Cancer* **149**, 573–584 (2021).
20. Tang, E., Fung, K. & Chan, A.-W. Incidence and mortality rates of keratinocyte carcinoma from 1998–2017: a population-based study of sex differences in Ontario, Canada. *CMAJ* **193**, E1516–E1524 (2021).
21. Khalid, A., van Essen, P., Crittenden, T. A. & Dean, N. R. The anatomical distribution of non-melanoma skin cancer: a retrospective cohort study of 22 303 Australian cases. *ANZ J. Surg.* **91**, 2750–2756 (2021).
22. Evans, D. M. & Davey, S. G. Mendelian Randomization: New Applications in the Coming Age of Hypothesis-Free Causality. *Annu. Rev. Genom. Hum. Genet.* **16**, 327–350 (2015).
23. Davies, N. M., Holmes, M. V. & Davey Smith, G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* **362**, k601 (2018).
24. Burgess, S. & Thompson, S. G. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am. J. Epidemiol.* **181**, 251–260 (2015).
25. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
26. Olsen, C. M. et al. Cohort profile: The QSkin Sun and Health Study. *International Journal of Epidemiology* **41** 929–929i <https://doi.org/10.1093/ije/dys107> (2012).
27. Landi, M. T. et al. Genome-wide association meta-analyses combining multiple risk phenotypes provide insights into the genetic architecture of cutaneous melanoma susceptibility. *Nat. Genet.* **52**, 494–504 (2020).
28. Verbanck, M., Chen, C.-Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).
29. Praetorius, C. et al. A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. *Cell* **155**, 1022–1033 (2013).
30. Sanderson, E., Spiller, W. & Bowden, J. Testing and Correcting for Weak and Pleiotropic Instruments in Two-Sample Multivariable Mendelian Randomisation. *Stat Med.* **40**, 5434–5452 (2021).
31. Dabouz, F. et al. Clinical and histological features of head and neck melanoma: a population-based study in France. *Br. J. Dermatol.* **172**, 707–715 (2015).
32. Australian Institute of Health and Welfare; *Cancer in Australia 2019*. 147 <https://www.aihw.gov.au/reports/cancer/cancer-in-australia-2019> (2019).
33. Nam, S. & Lim, J. S. Essential role of interferon regulatory factor 4 (IRF4) in immune cell development. *Arch. Pharm. Res.* **39**, 1548–1555 (2016).
34. Han, J. et al. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet* **4**, e1000074 (2008).
35. Do, T. N., Ucisik-Akkaya, E., Davis, C. F., Morrison, B. A. & Dorak, M. T. An intronic polymorphism of IRF4 gene influences gene transcription in vitro and shows a risk association with childhood acute lymphoblastic leukemia in males. *Biochim. Biophys. Acta* **1802**, 292–300 (2010).
36. Heilmann-Heimbach, S. et al. Meta-analysis identifies novel risk loci and yields systematic insights into the biology of male-pattern baldness. *Nat. Commun.* **8**, 14694 (2017).
37. Cho, Y. et al. Exploiting horizontal pleiotropy to search for causal pathways within a Mendelian randomization framework. *Nat. Commun.* **11**, 1010 (2020).
38. Duffy, D. L. et al. Novel pleiotropic risk loci for melanoma and nevus density implicate multiple biological pathways. *Nat. Commun.* **9**, 4774 (2018).
39. Hernando, B. et al. Genetic determinants of freckle occurrence in the Spanish population: Towards ephelides prediction from human DNA samples. *Forensic Sci. Int. Genet.* **33**, 38–47 (2018).
40. Dusingize, J. C. et al. Body mass index and height and risk of cutaneous melanoma: Mendelian randomization analyses. *Int. J. Epidemiol.* **49**, 1236–1245 (2020).
41. Liyanage, U. E. et al. Is there a causal relationship between vitamin D and melanoma risk? A Mendelian randomization study. *Br. J. Dermatol.* **182**, 97–103 (2020).
42. Larsson, S. C. et al. Smoking, alcohol consumption, and cancer: A mendelian randomisation study in UK Biobank and international genetic consortia participants. *PLoS Med* **17**, e1003178 (2020).
43. Schielein, M. C. et al. Stigmatization caused by hair loss - a systematic literature review. *J. Dtsch. Dermatol. Ges.* **18**, 1357–1368 (2020).
44. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet. Epidemiol.* **40**, 597–608 (2016).
45. Burgess, S. & Thompson, S. G. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* **32**, 377–389 (2017).
46. Loh, P. R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018).
47. Byrne, E. M. et al. Cohort profile: the Australian genetics of depression study. *BMJ Open* **10**, e032580 (2020).
48. Liyanage, U. E. et al. Combined analysis of keratinocyte cancers identifies novel genome-wide loci. *Hum. Mol. Genet* **28**, 3148–3160 (2019).
49. Ong, J.-S. et al. Height and overall cancer risk and mortality: evidence from a Mendelian randomisation study on 310,000 UK Biobank participants. *Br. J. Cancer* **118**, 1262–1267 (2018).
50. Zhou, W. et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
51. Visscher, P. M. et al. Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet* **10**, e1004269 (2014).
52. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
53. Bowden, J., Davey, S. G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
54. Bowden, J., Davey, S. G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian Randomization with Some Invalid

- Instruments Using a Weighted Median Estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
55. Hartwig, F. P., Davey, S. G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**, 1985–1998 (2017).
56. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human genome. *Elife* **7**, e34408 (2018).
57. Yavorska, O. O. & Burgess, S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* **46**, 1734–1739 (2017).
58. Ochoa, D. et al. Open Targets Platform: supporting systematic drug-target identification and prioritisation. *Nucleic Acids Res* **49**, D1302–D1310 (2021).
59. Elsworth, B. et al. The MRC IEU OpenGWAS data infrastructure. *bioRxiv* <https://doi.org/10.1101/2020.08.10.244293> (2020).

## Acknowledgements

This work was conducted using data from the UK Biobank Resource (application number 25331), the melanoma GWAS meta-analysis consortium and the QSkin Sun and Health Study. The study was supported by a program grant (APP1073898) from the Australian National Health and Medical Research Council. SM, RAS and DCW are supported by Fellowships from the Australian National Health and Medical Research Council. MS was supported by the Australian Government Research Training Program (RTP) and the QUT Faculty of Health Scholarship. MMI was supported in part by the National Institute for Health and Care Research (NIHR) Leeds Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. The complete study acknowledgement and grant support for the use of the melanoma GWAS meta-analysis data can be found in Landi et al.<sup>27</sup>. For support on study illustrations, Fig. 1 and Supplementary Fig. 1 were created using BioRender under the number SO25U9BQAU and YD250OKBWU.

*Melanoma Institute Australia.* The creation, support, and analysis of this cohort would not have been possible without the effort and support of patients, staff, and researchers at Melanoma Institute Australia (MIA). The MIA cohort was supported by funding from MIA, the National Health and Medical Research Council (NHMRC), NSW Health Pathology, Cancer Institute New South Wales and infrastructure grants from Macquarie University and the Australian Cancer Research Foundation. Genotyping was also supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, and by a Worldwide Cancer Research grant 16–0101. We also recognise support from The Cameron Family through MIA.

## Author contributions

J.S.O. designed the study. J.S.O., M.S., X.H., J.C.D., C.M.O., R.E.N., M.T.L., J.S., M.M.I., M.H.L., R.A.S., G.V.L. involved in data preparation. J.S.O., M.S., M.H.L., M.M.I., M.T.L. and J.C.D. analysed the data. M.H.L., S.M., M.S., Y.W., M.M.I., R.A.S., G.V.L., M.I., R.E.N., C.M.O. and D.C.W.

provided critical feedback on study design. YW provided illustrations to support presentation study findings. S.M., D.W. and M.H.L. obtained funding for the study. J.S.O., M.S., S.M. and M.H.L. wrote the first draft of the manuscript. All authors contributed to the final draft of the manuscript. The corresponding authors J.S.O. and M.H.L. attest that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

## Competing interests

R.A.S. has received fees for professional services from F. Hoffmann-La Roche Ltd, Evaxion, Provectus Biopharmaceuticals Australia, Qbiotics, Novartis, Merck Sharp & Dohme, NeraCare, AMGEN Inc., Bristol-Myers Squibb, Myriad Genetics, GlaxoSmithKline. All other authors declare no conflict of interest.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-41231-8>.

**Correspondence** and requests for materials should be addressed to Jue-Sheng Ong or Matthew H. Law.

**Peer review information** *Nature Communications* thanks Wen-Qing Li and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

<sup>1</sup>Population Health Department, QIMR Berghofer Medical Research Institute, Herston, QLD, Australia. <sup>2</sup>School of Biomedical Sciences, Faculty of Health, and Institute of Health and Biomedical Innovation, Queensland University of Technology, Kelvin Grove, QLD, Australia. <sup>3</sup>Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. <sup>4</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. <sup>5</sup>Faculty of Medicine, University of Queensland, Herston, QLD, Australia. <sup>6</sup>Melanoma Institute Australia, The University of Sydney, Sydney, NSW, Australia. <sup>7</sup>Department of Melanoma and Surgical Oncology, Royal Prince Alfred Hospital, Sydney, NSW, Australia. <sup>8</sup>Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia. <sup>9</sup>John Curtin School of Medical Research, Australian National University, Canberra, ACT, Australia. <sup>10</sup>Department of Mental Health & Neuroscience, QIMR Berghofer Medical Research Institute, Herston, QLD, Australia. <sup>11</sup>Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital & NSW Health Pathology, Sydney, Australia. <sup>12</sup>Charles Perkins Centre, The University of Sydney, Sydney, NSW, Australia. <sup>13</sup>Department of Medical Oncology, Royal North Shore Hospital, St Leonards, NSW, Australia. <sup>14</sup>Leeds Institute of Medical Research & Leeds Institute for Data Analytics, University of Leeds, Leeds, UK. <sup>15</sup>NIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals NHS Trust, Leeds, UK. ✉ e-mail: [JueSheng.Ong@qimrberghofer.edu.au](mailto:JueSheng.Ong@qimrberghofer.edu.au); [matthew.law@qimrberghofer.edu.au](mailto:matthew.law@qimrberghofer.edu.au)