Combining multiple populations and risk phenotypes greatly expands our knowledge of the genetic architecture of melanoma

Maria Teresa Landi[1,*,@], D. Timothy Bishop[2,*], Stuart MacGregor[3,*], Mitchell J. Machiela[1,*], Alexander J. Stratigos[4,#], Paola Ghiorzo[5,#], Myriam Brossard[6,7], Donata Calista[8], Jiyeon Choi[1], Maria Concetta Fargnoli[9], Tongwu Zhang[1], Monica Rodolfo[10], Adam J. Trower[11], Chiara Menin[12], Jacobo Martinez[13], Andreas Hadjisavvas[8], Lei Song[1], Irene Stefanaki[14], Rose Yang[1], Alisa M. Goldstein[1], Miriam Potrony Mateu[15], Katerina P. Kypreou[14], Lorenza Pastorino[5], Paola Queirolo[16], Cristina Pellegrini[9], Laura Cattaneo[8], Matthew Zawistowski[17], Pol Giminez Xavier[8], Arantxa Rodriguez[8], Lisa Elefanti[12], Siranoush Manoukian[10], Licia Rivoltini[10], Blair H. Smith[18], Maria Loizidou[19], Laura Del Regno[20], Daniela Massi[8], Mario Mandala[8], Kiarash Khosrotehani[21,22], Lars A. Akslen[23,24], Christopher I. Amos[25], Per A. Andresen[26], Marie-Françoise Avril[27], Esther Azizi[28,29], Duffy_melanoma_TBC B[30], Veronique Bataille[31], Giovanna Bianchi Scarrà[32,33], Lisa M. Bowdler[34], Kathryn P. Burdon[35], Wei V. Chen[36], Veryan Codd[37], Jamie E. Craig[38], Tadeusz Dębniak[39], Mario Falchi[31], Shenying Fang[40], Eitan Friedman[29], Sarah Simi[8], Pilar Galan[41], Zaida Garcia-Casado[8], Elizabeth M. Gillanders[42], Scott Gordon[43], Adele Green[44,45], Nelleke A. Gruis[46], Johan Hansson[47], Mark Harland[48], Jessica Harris[49], Per Helsing[50], Anjali Henders[51], Marko Hočevar[52], Veronica Höiom[47], David Hunter[53], Christian Ingvar[54], Rajiv Kumar[55], Julie Lang[56], G. Mark Lathrop[57], Jeffrey E. Lee[40], Xin Li[53], Jan Lubiński[39], Rona M. Mackie[58,56], Maryrose Malt[59], Josep Malvehy[8], Kerrie McAloney[43], Hamida Mohamdi[60], Anders Molven[24,61], Eric K. Moses[62], Rachel E. Neale[63], Srdjan Novaković[64], Dale R. Nyholt[65,66], Håkan Olsson[67,68], Nicholas Orr[69], Joan Anton Puig-Butille[15], Abrar A. Qureshi[70], Graham L. Radford-Smith[71,72,73], Juliette Randerson-Moor[48], Celia Requena[8], Casey Rowe[21], Nilesh J. Samani[37], Marianna Sanna[31], Dirk Schadendorf[74,75], Hans-Joachim Schulze[76], Lisa A. Simms[71], Mark Smithers[77,78], Fengju Song[79], Anthony J. Swerdlow[80,81], Nienke van der Stoep[82], Remco van Doorn[46], Alessia Visconti[31], Leanne Wallace[51], Sarah V. Ward[62], Lawrie Wheeler[49], Duffy_melanoma_TBC Z[30], Amy Hutchinson[1,83], Kristine Jones[1,83], Michael Malasky[1,83], Aurelie Vogt[1,83], Weiyin Zhou[1,83], Vasiliki Mamakou[84], Karen A. Pooley[85], David E. Elder[86], Jiali Han[87], Belynda Hicks[1,83], Nicholas K. Hayward[88], Peter A. Kanetsky[89], Chad Brummett[90], Grant W. Montgomery[91], Catherine M Olsen[92], Caroline Hayward[93], Alison M. Dunning[94], Nicholas G. Martin[43], Vangelis Evangelou[8], Graham J. Mann[95], Paul D. P.[94], Douglas F. Easton[85], Jennifer H. Barrett[11], Anne E. Cust[96], Goncalo Abecasis[97], David L. Duffy[43], David C. Whiteman[92], Helen Gogas[98], Arcangela De Nicolo[99], Margaret A. Tucker[1], Julia A. Newton Bishop[48], Ketty Peris[20], Stephen J. Chanock[1], Kevin M. Brown[1], Florence Demenais[6,7], Susana Puig[15,+], Eduardo Nagore[8,+], Jianxin Shi[1,&], Mark M. Iles[11,&,@], Matthew H. Law[3,&,@], GenoMEL Consortium[100], Q-MEGA and QTWIN Investigators[100], ATHENS Melanoma Study Group[100], 23andMe[100], The SDH Study Group[100], Essen-Heidelberg Investigators[100], AMFS Investigators[100], MelaNostrum Consortium[100]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA
[2]Leeds Institute of Medical Research at St James's, University of Leeds, Leeds, UK & Leeds Institute for Data Analytics, University of Leeds, Leeds, UK
[3]Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia
[4]Department of Dermatology, Andreas Syggros Hospital, Medical School, National and Kapodistrian University of Athens, Athens, Greece
[5]Genetics of Rare Cancers, Department of Internal Medicine (DiMI), University of Genoa and Policlinico San Martino Research Hospital Genoa, Genoa, Italy
[6]Institut National de la Santé et de la Recherche Médicale (INSERM), UMR-946, Genetic Variation and Human Diseases Unit, Paris, France
[7]Institut Universitaire d'Hématologie, Université Paris Diderot, Sorbonne Paris Cité, Paris, France.
[8](No affiliation data provided)
[9]Department of Dermatology, Department of Biotechnological and Applied Clinical Sciences, University of L'Aquila, L'Aquila, Italy
[10]Unit of Immunotherapy, Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS Istituto Nazionale Tumori, Milan, Italy
[11]Leeds Institute for Data Analytics, University of Leeds, Leeds, UK

[12]Immunology and Molecular Oncology Unit, Veneto Institute of Oncology IOV - IRCCS, Padua, Italy
[13]Red Valenciana de Biobancos, FISABIO
[14]Department of Dermatology, University of Athens School of Medicine, Andreas Sygros Hospital, Athens, Greece
[15]Dermatology Department, Melanoma Unit, Hospital Clínic de Barcelona, IDIBAPS, Universitat de Barcelona, Centro de Investigación Biomédica en Red en Enfermedades Raras (CIBERER), Barcelona, Spain
[16]Medical Oncology, Policlinico San Martino Research Hospital Genoa, Genoa, Italy
[17]Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI 48109, USA; Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI 48109, US
[18]Division of Population Health Sciences, Ninewells Hospital and Medical School, University of Dundee, Dundee DD1 9SY, United Kingdom
[19]Unknown_Cyprus_TBC
[20]Institute of Dermatology, Catholic University, Rome, Italy
[21]The University of Queensland, UQ Diamantina institute, Brisbane Australia
[22]Department of Dermatology, Princess Alexandra Hospital, Brisbane, Australia
[23]Centre for Cancer Biomarkers CCBIO, Department of Clinical Medicine, University of Bergen, Bergen, Norway
[24]Department of Pathology, Haukeland University Hospital, Bergen, Norway
[25]Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Hanover, New Hampshire, USA
[26]Department of Pathology, Molecular Pathology, Oslo University Hospital, Rikshospitalet, Oslo, Norway
[27]Assistance Publique–Hôpitaux de Paris, Hôpital Cochin, Service de Dermatologie, Université Paris Descartes, Paris, France
[28]Department of Dermatology, Sheba Medical Center, Tel Hashomer, Sackler Faculty of Medicine, Tel Aviv, Israel
[29]Oncogenetics Unit, Sheba Medical Center, Tel Hashomer, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
[30]Duffy_melanoma_TBC
[31]Department of Twin Research and Genetic Epidemiology, King's College London, London, SE1 7EH, UK Plus, for Veronique: Department of Dermatology, West Herts NHS Trust, Herts, HP2 4AD, UK
[32]Department of Internal Medicine and Medical Specialties, University of Genoa, Genoa, Italy
[33]Laboratory of Genetics of Rare Cancers, Istituto di ricovero e cura a carattere scientifico Azienda Ospedaliera Universitaria (IRCCS AOU) San Martino-IST Istituto Nazionale per la Ricerca sul Cancro, Genoa, Italy
[34]?Molecular Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Australia?
[35]Menzies Institute for Medical Research, University of Tasmania, Hobart, Tasmania, Australia
[36]Department of Genetics, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA
[37]Telomere_consortium_TBC
[38]Department of Ophthalmology, Flinders University, Adelaide, Australia
[39]International Hereditary Cancer Center, Pomeranian Medical University, Szczecin, Poland
[40]Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA
[41]Université Paris 13, Equipe de Recherche en Epidémiologie Nutritionnelle (EREN), Centre de Recherche en Epidémiologie et Statistiques, Institut National de la Santé et de la Recherche Médicale (INSERM U1153), Institut National de la Recherche Agronomique (INRA U1125), Conservatoire National des Arts et Métiers, Communauté d'Université Sorbonne Paris Cité, F-93017 Bobigny, France
[42]Inherited Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, Maryland, USA
[43]Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Australia
[44]Cancer and Population Studies, QIMR Berghofer Medical Research Institute, Brisbane, Australia
[45]CRUK Manchester Institute and Institute of Inflammation and Repair, University of Manchester, Manchester, UK
[46]Department of Dermatology, Leiden University Medical Centre, Leiden, The Netherlands
[47]Department of Oncology-Pathology, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden
[48]Leeds Institute of Medical Research at St James's, University of Leeds, Leeds, UK
[49]Translational Research Institute, Institute of Health and Biomedical Innovation, Princess Alexandra Hospital, Queensland University of Technology, Brisbane, Australia
[50]Department of Dermatology, Oslo University Hospital, Rikshospitalet, Oslo, Norway
[51]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia.
[52]Department of Surgical Oncology, Institute of Oncology Ljubljana, Ljubljana, Slovenia
[53]Harvard_nevus_TBC
[54]Department of Surgery, Clinical Sciences, Lund University, Lund, Sweden
[55]Division of Molecular Genetic Epidemiology, German Cancer Research Center, Im Neuenheimer Feld 580, Heidelberg Germany
[56]Department of Medical Genetics, University of Glasgow, Glasgow, UK
[57]McGill University and Genome Quebec Innovation Centre, Montreal, Canada
[58]Department of Public Health, University of Glasgow, Glasgow UK
[59]QIMR Berghofer, confirm dep
[60]Melarisk-Florence - TBC
[61]Gade Laboratory for Pathology, Department of Clinical Medicine, University of Bergen, Bergen, Norway

[62]Centre for Genetic Origins of Health and Disease, Faculty of Medicine, Dentistry and Health Sciences, The University of Western Australia, Western Australia, Australia
[63]Cancer Aetiology & Prevention, QIMR Berghofer Medical Research Institute, Brisbane, Australia
[64]Department of Molecular Diagnostics, Institute of Oncology Ljubljana, Ljubljana, Slovenia
[65]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia
[66]Molecular Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Australia
[67]Department of Oncology/Pathology, Clinical Sciences, Lund University, Lund; Sweden
[68]Department of Cancer Epidemiology, Clinical Sciences, Lund University, Lund, Sweden
[69]Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, UK
[70]Department of Dermatology, The Warren Alpert Medical School of Brown University, Rhode Island, USA
[71]Inflammatory Bowel Diseases, QIMR Berghofer Medical Research Institute, Brisbane, Australia
[72]Department of Gastroenterology and Hepatology, Royal Brisbane & Women's Hospital, Brisbane, Australia
[73]University of Queensland School of Medicine, Herston Campus, Brisbane, Australia
[74]Department of Dermatology, University Hospital Essen, Essen, Germany
[75]German Consortium Translational Cancer Research (DKTK), Heidelberg, Germany
[76]Department of Dermatology, Fachklinik Hornheide, Institute for Tumors of the Skin at the University of Münster, Münster, Germany
[77]Queensland Melanoma Project, Princess Alexandra Hospital, The University of Queensland, Australia
[78]Mater Research Institute, The University of Queensland, Australia.
[79]Departments of Epidemiology and Biostatistics, Key Laboratory of Cancer Prevention and Therapy, Tianjin, National Clinical Research Center of Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin, P. R. China
[80]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK
[81]Division of Breast Cancer Research, The Institute of Cancer Research, London, UK
[82]Department of Clinical Genetics, Center of Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands
[83]Cancer Genome Research Laboratory, Leidos Biomedical Research Inc., Bethesda, Maryland, USA
[84]National and Kapodistrian University of Athens, Dromokaiteio Psychiatric Hospital, Athens, Greece
[85]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
[86]Department of Pathology and Laboratory Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA
[87]Department of Epidemiology, Richard M. Fairbanks School of Public Health, Melvin and Bren Simon Cancer Center, Indiana University, Indianapolis, USA
[88]Oncogenomics, QIMR Berghofer Medical Research Institute, Brisbane, Australia
[89]Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA
[90]Department of Anesthesiology, University of Michigan, Ann Arbor, MI, USA
[91]Molecular Biology, the University of Queensland, Brisbane, Australia.
[92]Cancer Control Group, QIMR Berghofer Medical Research Institute, Brisbane, Australia
[93]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH42 XU, United Kingdom
[94]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK
[95]Centre for Cancer Research, University of Sydney at Westmead, Millennium Institute for Medical Research and Melanoma Institute Australia, Sydney, Australia
[96]Cancer Epidemiology and Services Research, Sydney School of Public Health, University of Sydney, Sydney, New South Wales, Australia.
[97]Department of Biostatistics, University of Michigan, Ann Arbor, 48109, MI, USA
[98]First Department of Internal Medicine, Laikon General Hospital Greece, National and Kapodistrian University of Athens, Athens, Greece
[99]Cancer Genomics Program, Veneto Institute of Oncology IOV - IRCCS, Padua, Italy
[100]A full list of members and affiliations appears in the Supplementary Note

[*]Co-first authors
[#]Co-second authors
[+]Co-second to last authors
[&]Co-last authors
[@]Corresponding authors

Summary

Genetic susceptibility to cutaneous melanoma is poorly characterized. Leveraging multiple cutaneous melanoma genome-wide association studies we identified 56 significant loci which show enrichment for melanocytic enhancer sites. We explore risk estimates across geographical regions and varying host factors. The acral lentiginous subtype was uniquely unrelated to pigmentation. Combining this meta-analysis with large nevus and pigmentation genome-wide association studies, and leveraging gene and transcriptome wide association approaches identified a further 53 loci, for a total of 109 new genetic loci. These findings describe in unprecedented scale the genetic architecture of cutaneous melanoma, highlighting the role of nevus, pigmentation, immune, and telomere pathways.

Introduction

Cutaneous melanoma (CM) is a deadly malignancy with increasing incidence in fair-skinned populations worldwide [1]. Increased risk for CM is associated with ultraviolet light exposure [2], as well as host factors including family history [3], certain pigmentary phenotypes (notably fair skin, blue or green eyes, blonde or red hair, and sun sensitivity or inability to tan) [4–7], number of melanocytic nevi [8,9], longer telomeres [10,11], and immunosuppression [12].

Identified melanoma genetic risk variants include rare, highly penetrant mutations in genes such as *CDKN2A* [13,14] and *POT1* [15,16], as well more common variation e.g. low-penetrance variants in *MC1R* [17,18]. Population-based genome-wide association studies (GWAS) of CM susceptibility in populations of European ancestry have identified 21 genetic loci reaching genome-wide significance ($P < 5 \times 10^{-8}$) [19–26] (**Table 1**). Additional approaches, including family-based analyses of CM [27], combining CM and nevus count GWASs [28] and TWAS [29] have identified further loci that despite not reaching $P < 5 \times 10^{-8}$ in a CM-only GWAS, very likely influence melanoma risk (**Table 1**).

This meta-analysis of CM susceptibility is more than three times the effective sample size of previous CM GWAS, providing unprecedented power to identify CM susceptibility variants as well as improved ability to map independent variants in known CM susceptibility regions. We report here 68 independent CM associated variants across 56 loci that confirm the importance of key functional pathways and highlight previously unknown CM etiologic routes (**Table 1, Table 2**). Stratified analyses revealed differences in the relative importance of melanoma risk pathways by subtype, particularly lack of involvement of the pigmentation pathway for acral melanoma. Combined analysis of CM, nevus and pigmentation GWAS data, and leveraging expression data through transcriptome wide association studies (TWAS) uncovered an additional 53 loci; of note a number of loci were pleiotropically associated with multiple traits.

Results

Study overview

We performed an international GWAS meta-analysis of CM susceptibility with 30,134 cases where CM status was clinically confirmed (**Online Methods**), 6,626 self-reported CM cases and 375,188 CM-free controls from the United Kingdom, United States, Australia, Northern and Western Europe as well as the Mediterranean, a highly sun exposed population often under-represented in CM studies (**Supplementary Table 1**). Following quality controls and sample exclusions (**Online Methods**), we imputed the previously and newly genotyped samples and ran the association and meta-analysis.

We performed separate total (clinically confirmed cases + self-reported cases from the UK Biobank and 23andMe, Inc.) and confirmed CM meta-analyses as sensitivity analyses to explore the utility of, and power gain from, including self-reported CM cases. Risk loci were deemed genome-wide significant when variants had fixed effects meta-analysis P-values $< 5 \times 10^{-8}$ (P_{meta}); where variants exhibited notable heterogeneity ($I^2 > 31\%$) random effects P-values (P_{meta_r}) needed to be $< 5 \times 10^{-8}$ as well (**Online Methods**). Q-Q plots (**Supplementary Figure 1)** and LD Score Regression [30] (LDSC; **Online Methods**) intercepts for contributing data showed little to no inflation for individual studies (majority of studies intercept < 1.04; **Supplementary Table 1)**, indicating our large study that included a diverse set of CM case populations adequately controlled for case differences by ancestry.

Before combining the two self-report sets and the clinically confirmed GWAS data, we used LDSC [30] to investigate the genetic correlation (Rg) between self-reported CM cases and the confirmed cases only meta-analysis GWAS data (**Supplementary Table 2**). Correlations with clinically confirmed CM GWAS results were high, with the self-report 23andMe, Inc. and UK BioBank (UKBB) GWAS having Rg of 0.99 (SE = 0.13) and 0.70 (SE = 0.15), respectively (**Supplementary Table 2)**. The Rg for the UKBB self-reported cases was expected to be lower since all UKBB self-reported cases that could be confirmed by cancer registries were moved into the UKBB confirmed set (**Online Methods, Supplementary Table 1**). Similarly, LDSC SNP-heritability (h^2) were comparable between confirmed CM cases ($h^2_{confirmed} = 0.13$, 95% CI = 0.09-0.17) and self-reported CM cases ($h^2_{23andMe} = 0.10$, 95% CI = 0.03-0.16; $h^2_{UKBB\_SR} = 0.18$, 95% CI = 0.04-0.31) (**Supplementary Table 2**). Based on the high Rg and similarity in h^2 estimates for self-report and clinically confirmed CM cases, we merged both sets into an overall total CM meta-analysis (h^2_{total}= 0.11, 95% CI = 0.08-0.15). The lambda and LDSC intercept for the total CM meta-analysis indicated the majority of apparent inflation is due to polygenic signal ($\lambda = 1.165$, Intercept 1.051, ratio 0.16; **Supplementary Table 2**). A similar h^2_{total} (12%) was estimated using genetic effect-size distribution inference from summary level data (GENESIS; **Online methods**) [31].

Conditional joint analysis of the total CM meta-analysis using GCTA [32] identified a total of 56 loci that met our requirements for genome-wide significance (**Online Method**s; **Table 1, Figure 1, Supplementary Figure 2**). In addition to the 56 sentinel variants, an additional 12 independent variants with linkage disequilibrium (LD) $r^2_{EUR} < 0.05$ with sentinel variants at or near 5 loci (*TERT*, *CDKN2A*, *OCA2*, *MC1R*, and *TP53*) were identified (**Supplementary Table 3**). Individual regional association plots for the association signals have been provided as a **Supplementary File**. Conditional analysis identified a further 6 variants at or near *SLC45A2*, *IRF4*, *CCND1*, *GPRC5A*, and *MC1R)*; however, these 6 variants were not carried forward due to either $P_{meta} > 5 \times 10^{-8}$ in the single variant analysis or notable heterogeneity ($I^2 > 31\%$) and not genome-wide significant under the random effects model (**Supplementary Table 4)**. This significant expansion of discovered CM risk loci highlights the utility of a large, international sample of CM cases and demonstrates the increases in power from including a series of self-reported CM cases. In addition, we used GENESIS (**Online methods**), which enables a reformulation of the variance explained by associated SNPs to estimate a theoretical optimal area under the curve, rather than formally testing this via a training and prediction set [31] to estimate the potential AUC derived from the lead SNPs from the 56 identified loci. The estimated AUC was 68%; for comparison the 2015 CM meta-analysis [25] estimate was ~63%. This estimate does not include any host factors, and would require benchmarking in a prospective study for validation.

Previous CM GWAS have identified 21 genome-wide significant loci [19–26], while family-based methods or the combination of CM with nevus count have identified a further 10 loci including *IRF4*, *MITF, HDAC4,* and *GPRC5A* [27,28,33] (**Table 1**). Many of these SNPs are involved in pigmentation, tanning response, or nevus count. Others are in or near genes associated with DNA repair, telomere length or regulation of senescence [34]; some are associated with more than one trait (**Table 1**). Our analysis directly replicates nearly all of these previously reported loci. Of note, the peak variant near *IRF4* is rs12215602 and not the previously reported rs12203592, for which

substantial heterogeneity was observed (I^2 = 81%, P_{meta_R} = 0.1, P_{meta} = 1.9 × 10^{-13}). The previously reported peak variant for the locus in intron 8 of *FTO* rs16953002 did not formally replicate (P_{meta_r} = 1.2 × 10^{-6}, I^2 = 35%, P_{meta} = 3.8 ×10^{-8}) nor did any SNPs in LD, with the m6ost strongly associated SNP at this region being rs62034121 (**Table 1**). rs16953002 and rs62034121 are in LD (r^2_{EUR} = 0.96), and these variants are genome-wide significant in the confirmed cases only meta-analysis, and the combined CM plus nevus analysis (see below and **Online Methods**). A recent melanoma GWAS [26] reported rs187843643 near *BASP1* as a melanoma locus; while we include the contributing GWAS this locus does not reach genome-wide significance (I^2 = 35%, P_{meta_r} = 0.3, P_{meta} = 0.00071; Table 1). In addition, previously reported CM susceptibility variants centromeric to *AGR3* (rs1636744, rs819368 and rs2389832) do not exhibit association $P_{meta} < 5 \times 10^{-8}$; instead a different independent variant (rs117132860, $r^2_{EUR} < 0.03$ with previously reported variants) located between rs819368 and rs2389832 is observed at genome-wide significance (P_{meta}= 3.8 × 10^{-21}). This region exhibits a complex LD structure and more focused investigations are needed to disentangle the CM association signal in this region.

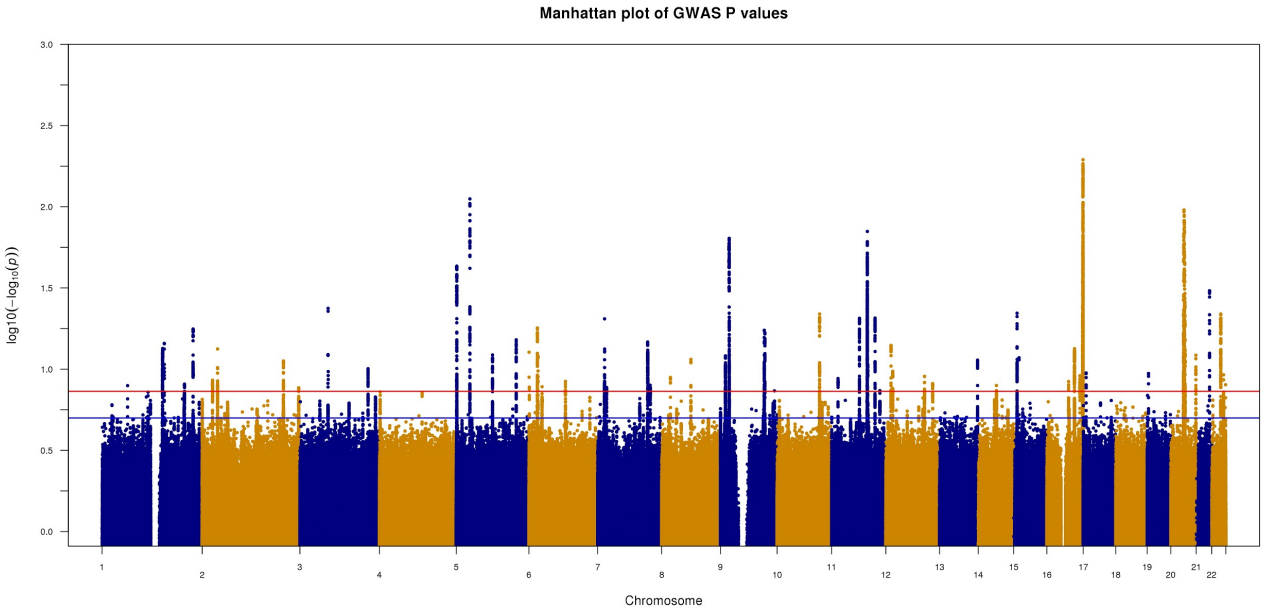

Figure 1. *Manhattan plot for the total CM meta-analysis.*

Figure 1 Legend: *Log_{10} of -log_{10}(P_{meta}) is displayed to truncate strong signals at loci such as MC1R and ASIP.*

Table 1. *Loci previously associated with CM susceptibility.*

CHR:BP	rsID	Pub	Gene	EA/ NEA	Freq	P	OR	Nevi	Hair	Tan	Tel
1:150,938,571	rs8444	[35]	Multiple	G/A	0.645	3.89 × 10^{-14}	1.08	-	-	Y	-
1:226,603,635	rs2695237	[28,35,36]	*PARP1*	T/C	0.628	2.12 × 10^{-18}	1.10	Y	-	-	-
2:38,298,139	rs1800440	[25,28]	*CYP1B1*	T/C	0.824	4.84 × 10^{-14}	1.10	Y	-	Y	-
2:202,122,995	rs3769823	[22]	*CASP8*	A/G	0.305	5.90 × 10^{-12}	1.08	-	-	-	-
2:240,081,540	rs3791512	[28]	*HDAC4*	C/G	0.814	2.09 × 10^{-8}	1.08	Y	-	-	-
3:70,014,091	rs149617956	[27a]	*MITF*	G/A	0.998	1.74 × 10^{-13}	0.38	-	Y	Y	-
3:169,493,283	rs3950296	[28]	*TERC*	C/G	0.747	8.51 × 10^{-11}	1.08	Y	-	-	Y
5:1,323,212	rs13178866	[22,37,7]	*TERT*	C/T	0.554	1.22 × 10^{-18}	0.87	-	Y	-	Y
5:17,454,083	rs187843643	[6]	*BASP1*	C/T	0.995	0.26[c]	0.90	-	-	-	-
5:33,951,693	rs16891982	[22,38,7]	*SLC45A2*	C/G	0.122	1.96 × 10^{-28}	0.51	-	Y	Y	-
5:149,211,868	rs32578	[28,39]	*PPARGC1B*	G/A	0.658	6.91 × 10^{-16}	1.09	Y	-	Y	-

6:1,145,265	rs12215602[b]	28,38,40[a]	*IRF4*	G/A	0.721	1.78×10^{-8}	0.94	Y	-	Y	-
6:21,163,919	rs6914598	25	*CDKAL1*	T/C	0.683	1.18×10^{-18}	0.91	-	-	Y	-
7:17,134,708	rs117132860[b]	25,41	*AHR*	G/A	0.981	3.83×10^{-21}	0.71	Y	-	Y	-
9:21,803,880	rs871024	20,28	*MTAP, CDKN2A*	C/A	0.477	6.58×10^{-23}	1.18	Y	Y	-	-
9:109,054,417	rs10739220	25,28	*TMEM38B*	C/T	0.260	4.40×10^{-18}	1.10	Y	Y	-	-
9:110,711,586	rs1339759	28	*KLF4*	C/G	0.666	1.87×10^{-17}	1.09	Y	-	-	-
10:105,694,301	rs7902587	25,28	*OBFC1*	C/T	0.904	1.38×10^{-22}	0.86	Y	-	-	Y
11:69,380,898	rs4354713	22,25	*CCND1*	A/G	0.356	2.60×10^{-21}	1.10	-	Y	-	-
11:89,017,961	rs1126809	20	*TYR*	G/A	0.757	4.63×10^{-43}	0.82	-	Y	Y	-
11:108,175,462	rs1801516	22	*ATM*	G/A	0.856	2.33×10^{-21}	1.14	Y	-	-	-
12:13,070,752	rs1056927	28	Multiple	A/G	0.561	3.33×10^{-9}	0.93	Y	-	-	-
14:64,390,030	rs10873172	28	*SYNE2*	G/C	0.290	1.18×10^{-8}	1.06	Y	Y	-	-
15:28,365,618	rs12913832	21,25	*OCA2*	A/G	0.335	1.54×10^{-11}	0.88	-	Y	Y	-
15:33,277,710	rs117648907	28	*FNM1*	C/T	0.983	1.77×10^{-12}	0.80	Y	-	-	-
16:54,118,132	rs62034121	24	*FTO*	C/A	0.822	4.87×10^{-7}[b]	0.91	Y	-	-	-
16:89,986,117	rs1805007	20	*MC1R*	C/T	0.937	3.15×10^{-52}	0.57	Y	Y	Y	-
19:3,540,539	rs12984831	28	*MFSD12*	G/C	0.984	3.86×10^{-10}	0.65	Y	-	Y	-
20:32,665,748	rs6059655	19,20	*ASIP*	A/G	0.061	5.12×10^{-41}	1.45	-	Y	Y	-
21:42,743,496	rs408825	22	*MX2*	C/T	0.413	4.10×10^{-31}	0.89	-	-	Y	-
22:38,602,140	rs11914181	20,28,40	Multiple	T/C	0.457	1.38×10^{-22}	0.91	Y	-	Y	-

Table 1 footnote: ***CHR, BP:*** *hg19 positional information.* ***rsID****: dbSNP142 rs number.* ***Pub****lications: 7 (PMID: 21693730). We also summarise* ***Supplementary Table 3****;* ***Gene*** *prioritises the functional target if known, followed by melanocyte or all tissue TWAS data, or finally the closest protein coding gene; multiple indicates three or more genes. The effect allele (****EA****) and non effect allele (****NEA****) are listed, as are the effect allele frequency in the HRC reference panel [42]; total meta-analysis* ***P****-value and Odds Ratio (****OR****) are with respect to the EA. We also indicate if this locus is associated with other traits*: ***Nevi****: Pleiotropically associated with CM and nevus count (****Online methods****;* ***Supplementary Table 7****);* ***Hair****: Pleiotropically associated with CM and hair colour (****Online methods****;* ***Supplementary Table 8****). Tanning response (****Tan****) and Telomere length (****Telo****) indicates the lead SNP is associated with these traits when corrected for multiple testing (****Online methods. Supplementary Table 5****).* [a]*Associated with CM by non-GWAS based approaches.* [b]*While this locus overlaps the previously reported IRF4 or AGR3 locus, the lead variants are independent.* [c]*Variant meta-analysis results are heterogenous ($I^2 > 31\%$) and random effects estimates are presented.*

The total CM meta-analysis also confirmed 8 regions previously identified by combining nevus count and CM GWAS data [28] (**Table 1**). While the peak variant for the 8th region, rs12984831, near *MFSD12* and *FZR1*, is independent from the previously reported peak rs34466956 ($r^2_{EUR} = 0.001$), rs34466956 is also significantly associated in the combined analysis ($P_{CM+nevus} = 2.5 \times 10^{-9}$; **Supplementary Table 7**). These results highlight the ability of combined trait GWAS (e.g., CM and nevus count) to identify loci associated with CM; a method we implement below for the expanded CM meta-analysis. The remaining 27 loci have not previously been reported as CM susceptibility loci (**Table 2**; full results in **Supplementary Table 3**).

The meta-analysis for only pathologically confirmed CM cases (N = 30,134; **Supplementary Table 1**) identified a total of 47 loci associated with CM susceptibility (**Supplementary Table 6, Supplementary Figures 3-4**). Only two loci were significant in the confirmed-only CM and not the total CM analysis. The first, on chromosome 4, rs2301293 (confirmed $P_{meta} = 3.2 \times 10^{-9}$, OR = 0.89, $I^2 = 0\%$) while not significant in the total CM meta-analysis, showed negligible difference in its estimate ($P_{meta} = 5.8 \times 10^{-8}$, $I^2 = 10\%$ OR = 0.91) . The second locus is the previously reported locus on chromosome 16 containing *FTO* with SNP rs16953002 meeting our significance requirements ($P_{meta} = 3.6 \times 10^{-9}$, $I^2 = 27.2$).

Thus the total meta-analysis, which included 6,626 self-reported CM cases and over 290,000 controls (**Supplementary Table 1**) allowed identification of 11 loci beyond those found in the confirmed GWAS meta-analysis alone, demonstrating the advantage of including self-reported CM cases. Results for SNPs with a fixed or random $P < 5 \times 10^{-7}$, as appropriate, from the total meta-analysis are reported in **Supplementary Table 15**.

Table 2: *Novel loci associated with CM.*

CHR:BP	rsID	Gene	EA/ NEA	FREQ	P	OR	Nevi	Hair	Tan	Telo
1:63,727,542	rs670318	*ATG4C*	T/C	0.047	1.21×10^{-8}	0.865	-	-	Y	-
1:154,994,978	rs76798800	*DCST2, ADAM15*	G/T	0.753	3.86×10^{-15}	0.917	Y	-	Y	-
1:205,181,062	rs2369633	*RBBP5*	T/C	0.083	8.13×10^{-9}	1.098	-	Y	Y	-
2:25,778,637	rs12473635	*DTNB*	T/C	0.776	2.86×10^{-9}	0.931	Y	-	-	-
5:90,262,612	rs12523094	*GPR98*	T/C	0.567	5.85×10^{-13}	1.074	-	Y	Y	-
6:22,719,379	rs72834823	*HDGFL1*	T/A	0.819	1.04×10^{-12}	1.097	Y	-	Y	-
6:32,748,953	rs28986343	*SKIV2L*	C/T	0.952	1.61×10^{-8}	1.153	-	-	-	-
6:91,005,743	rs6908626	*BACH2*	G/T	0.844	3.92×10^{-9}	1.086	-	-	-	-
7:22,115,454	rs12539524	*RAPGEF5*	C/T	0.846	1.93×10^{-8}	0.926	-	-	-	-
7:124,396,645	rs4731207	*POT1*	G/A	0.540	1.99×10^{-15}	0.926	Y	-	-	Y
7:130,742,066	rs7803075	*MKLN1*	A/G	0.317	1.08×10^{-8}	0.937	Y	Y	-	-
8:21,951,009	rs6994183	*DMTN, HR*	A/T	0.866	1.29×10^{-9}	0.918	-	-	-	-
8:72,864,240	rs13263376	*RP11-383H13.1, MSC*	G/A	0.364	4.85×10^{-8}	0.929	Y	-	Y	-
9:12,600,284	rs10809803	*LURAP1L,*	G/C	0.367	2.11×10^{-12}	0.930	-	Y	Y	-
9:134,457,580	rs3780269	*POMT1*	G/A	0.691	4.43×10^{-8}	0.945	Y	-	-	-
11:16,041,305	rs7941496	*SOX6*	G/T	0.516	1.75×10^{-9}	1.061	Y	-	Y	-
11:120,195,702	rs12290699	*TMEM136*	T/C	0.745	4.17×10^{-8}	0.938	-	-	-	-
12:17,275,460	rs4237963	*LMO3*	T/A	0.207	1.99×10^{-10}	0.923	-	-	-	-
12:96,378,807	rs10859996	*RP11-256L6.3*	C/T	0.635	9.35×10^{-10}	1.065	-	-	-	-
12:116,580,291	rs113469387	*MED13L*	G/A	0.907	7.05×10^{-9}	0.915	-	Y	Y	-
13:113,533,594	rs1278763	*ATP11A*	G/C	0.466	4.31×10^{-12}	0.934	-	-	Y	-
16:68,822,971	rs4420522	Multiple, *CDH1*	A/G	0.690	4.25×10^{-14}	0.924	Y	Y	-	-
16:82,188,801	rs2911423	*MPHOSPH6*	C/T	0.217	4.75×10^{-9}	1.070	-	-	-	Y
17:7,571,752	rs78378222	*TP53*	T/G	0.989	3.33×10^{-10}	0.757	Y	-	-	-
20:62,291,767	rs143190905	*RTEL1*	G/T	0.907	6.54×10^{-13}	1.145	-	-	-	Y
22:45,622,684	rs5766565	*KIAA0930*	A/G	0.647	5.93×10^{-10}	1.066	Y	Y	Y	-
22:50,722,408	rs79966207	*PLXNB2*	T/C	0.849	9.56×10^{-9}	0.923	-	Y	-	-

Table 2 footnote: ***CHR, BP:*** *hg19 positional information.* ***rsID****: dbSNP142 rs number. We also summarise Supplementary Table 3;* ***Gene*** *prioritises the functional target if known, followed by melanocyte or all tissue TWAS data, or finally the closest protein coding gene; multiple indicates three or more genes. The effect allele (****EA****) and non effect allele (****NEA****) are listed, as are the effect allele frequency in the HRC reference panel [42]; total meta-analysis* ***P****-value and Odds Ratio (****OR****) are with respect to the EA. We also indicate is this locus is associated with other traits*: ***Nevi****: Pleiotropically associated with CM and nevus count (****Online methods****;* ***Supplementary Table 7****);* ***Hair****: Pleiotropically associated with CM and hair colour (****Online methods****;* ***Supplementary Table 8****). Tanning response (****Tan****) and Telomere length (****Telo****) indicates the lead SNP is associated with these traits when corrected for multiple testing (****Online methods, Supplementary Table 5)***

Melanoma associations by sex, age at diagnosis and subtype

We performed separate GWAS by sex, age at CM diagnosis (≤40, between 40 and 60, ≥60) and major CM subtypes (superficial spreading (SS), lentigo maligna (LM), nodular melanoma (NM), and acral lentiginous (AL)) to identify variants associated with select subgroups (**Supplementary Table 16**). Our analysis did not identify any additional variants reaching statistical significance after adjustment for multiple testing (5×10^{-8} / 13), suggesting that if variants are associated with only one analytic subgroup they are undetectable at our current sample size and may require other methods for uncovering associations.

We also performed polygenic risk score (PRS) analyses defined based on lead independent genome-wide SNPs for nevi count (10 variants; **Online methods**) and pigmentation (276 variants; **Online methods**) to further explore if either traits' association with CM is different across the sub phenotypes (significance threshold = 0.05/28; **Online methods**). We observed no significant differences in the distribution of the tested PRSs for the sex and age at CM diagnosis subgroups. We did, however, detect differences in the distribution of pigmentation PRS for acral lentiginous subtype compared to all non-acral subtypes ($P = 2.1 \times 10^{-4}$). Our analyses indicated darker genetically-inferred pigmentation levels in AL cases compared to SS, LM and NM cases ($P = 5.3 \times 10^{-5}$, 0.01, 4.8×10^{-4}, respectively) as well as no differences in genetically-inferred pigmentation between AL cases and controls (P = 0.65, **Supplementary Figure 5**). These findings provide strong genetic evidence that the pigmentation pathway is less important for risk of AL melanoma than for other subtypes of CM and suggests that public health preventative measures for AL melanoma may be unique from other CM subtypes. No significant differences were observed by subtype for the nevi count PRS.

Annotation and discovery by utilising risk phenotypes for CM

To determine their possible mode of influencing risk of CM, variants independently associated with CM in the total meta-analysis were looked up in GWAS of telomere length, tanning response, pigmentation and nevus count (**Online Methods**, **Table 1** and **2**, **Supplementary Table 5**, **7**, **8**). Using a Bonferroni corrected threshold of phenotype P-value < 0.000735 (0.05/68 independent SNPs), 14 of the 27 novel loci are associated with tanning response or pigmentation (**Table 2**, **Supplementary Table 5**), further indicating the importance of pigmentation pathways in CM susceptibility. A number of the novel loci including rs12473635 near *DTNB* and rs78378222 near *TP53* are associated with nevus count, reinforcing the role of nevi in CM susceptibility. Furthermore, three novel loci have been previously associated with telomere length (rs4731207/*POT1*, rs2911423/*MPHOSPH6*, and rs143190905/*RTEL1* [43] (**Supplementary Table 5**) providing additional support for telomere involvement in CM susceptibility. The remainder of the newly discovered sentinel variants are not associated with these phenotypes and suggest novel pathways unrelated to these phenotypes may be important for CM susceptibility.

Leveraging additional phenotypes to identify and annotate CM risk loci

To identify further loci influencing CM risk, and provide a more nuanced annotation of discovered CM risk loci, we utilised a range of approaches. To explore the overlap between CM loci and risk phenotypes, we combined our total CM GWAS meta-analysis with a nevus count GWAS meta-analysis (N = 65,777; **Online Methods**) and separately with a UKBB hair colour GWAS (N = 352,662; **Online Methods**). Pairwise GWAS (GWAS-PW) [44] was used to determine whether loci identified were associated with only one trait or pleiotropically associated with both CM and either nevus count or pigmentation (**Online Methods**). As noted above, a previous reported combination of a CM and nevus GWAS [28] identified loci that were subsequently confirmed in this larger CM GWAS meta-analysis (**Table 1**), providing support for this approach. Together these analyses identified an additional 13 loci pleiotropically associated with CM and nevus count, 17 loci pleiotropically associated with CM and pigmentation, and 3 loci pleiotropically associated with CM, nevus count and pigmentation that were not identified in the total CM GWAS meta-analysis (**Table 3**, **Supplementary Table 7**, **Supplementary Table 8**). To test for enrichment of multiple

independent genetic variants associated with CM around specific genes, we used Multi-marker Analysis of GenoMic Annotation (MAGMA) [45] to perform a gene-based association test in the total CM meta-analysis data (**Methods**; **Supplementary Table 14**). The MAGMA analysis identified 11 loci not identified in the per-SNP analysis.

In parallel we examined data from a recently established cell-type specific melanocyte eQTL dataset [29] as well as tissue-based datasets available through GTEx [46] to identify candidate susceptibility genes from all independent CM SNPs. Colocalization analyses and integrative transcriptome-wide association studies (TWAS) utilising these expression datasets enabled formal testing for significant *cis* genetic correlations between gene expression and CM risk, allowing both characterisation of genes at loci detected in the total CM meta-analysis, and the identification of additional loci (**Online Methods**). This analysis builds on previous melanocyte TWAS which analyzed data from a prior CM GWAS meta-analysis [29] which identified imputed gene expression of five genes at four loci associated with CM. Importantly, the *CBWD1* locus on Chr9 was later identified as a genome-wide significant CM+nevus count pleiotropic locus [28] and confirmed again here (**Table 1**, **Supplementary Table 7**), and the other three loci (*ZFP90* on Chr16, *HEBP1* on Chr12, and *MSC* and *RP11-383H13.1* on Chr8) are now genome-wide significantly associated with CM in this larger GWAS meta-analysis (**Table 2**). This confirmation of previous TWAS loci supports the TWAS approach for both identifying new loci and pinpointing the likely functional genes at GWAS discovered loci (**Table 1, 2).**

To empirically identify the target tissues for CM risk variants, we used partitioned LD score regression [47] to determine the proportion of total CM GWAS meta-analysis heritability captured by gene expressed in melanocytes and in 50 GTEx tissue types. This demonstrated that partitioned CM heritability was significantly enriched in genes specifically expressed in melanocytes (2.76-fold, $P = 3.12 \times 10^{-6}$ for top 4,000 genes; **Supplementary Figure 6**). Importantly, melanocyte-specific genes displayed the most significant enrichment P-value with the largest enrichment fraction, when compared with 50 GTEx tissue types, while three skin derived tissues ranked near the top. Given tissues other than melanocytes were also significantly enriched for partitioned CM heritability we performed additional discovery and annotation of CM risk variants using expression data from melanocytes, GTEx skin tissues, and from all tissues types

Conducting a TWAS of all tissue types and using the total CM GWAS meta-analysis, we identified the expression of 170 genes associated with CM risk ($P_{TWAS} < 1.48 \times 10^{-5}$; **Supplementary Table 10)**. When restricted to skin tissue (**Online Methods**), 80 genes were significant, and when further restricting results to melanocytes alone, 41 genes demonstrated evidence for an association (**Supplementary Table 9)**. Of the melanocyte-specific gene associations, 33 overlapped with loci identified in the total CM GWAS meta-analysis, providing functional evidence as to which genes may underlie the association at these region (**Supplementary Table 3, Supplementary Table 9)**. Combining genes within 1 Mb of each other into discrete loci, TWAS analysis identified an additional 25 regions not identified in the total CM GWAS meta-analysis (**Table 4**, **Table 5**).

In aggregate these complementary approaches identified a total of 109 discrete loci (**Figure 2**); 56 formally significant at $P < 5\times 10^{-8}$ in the total CM meta-analysis (**Table 1**, **Table 2**, **Supplementary Table 3**), and the remainder supported by one or more of the additional analyses (**Table 3-5**, **Supplementary Tables 7-10**,**14**) and likely represent additional risk loci for CM, albeit ones requiring a larger CM GWAS meta-analysis to formally confirm.

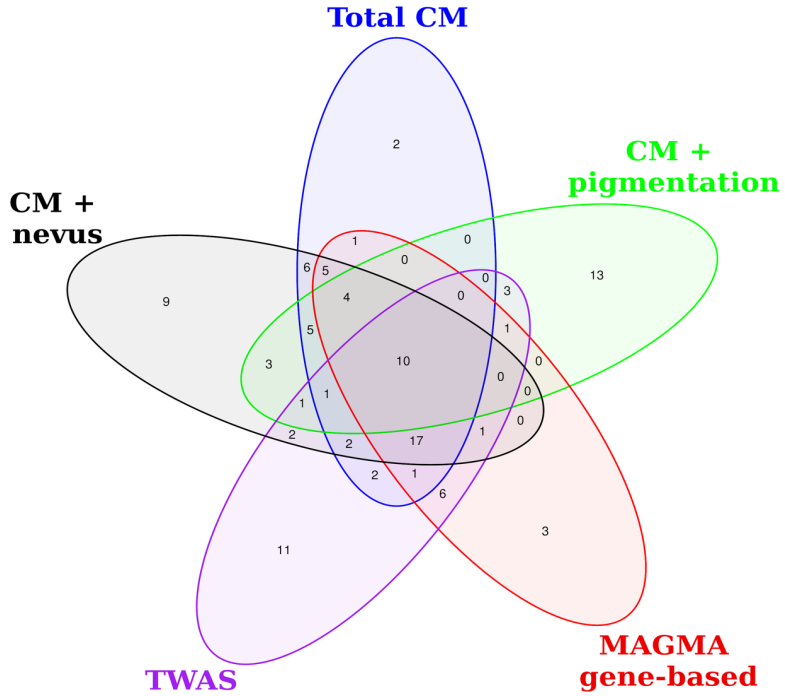

Figure 2. Overlap of loci identified by complementary approaches

Figure 2 legend: *Loci identified in the **total CM** meta-analysis (**Supplementary Table 3**), the pleiotropic analysis with nevus count (**CM + nevus, black**, **Supplementary Table 7**) and hair colour (**CM + pigmentation**, **green, Supplementary Table 8**), and the **MAGMA gene-based** test (**Red, Supplementary Table 14**) are plotted together with a merged count of loci identified by TWAS using melanocyte or all tissue expression (**TWAS merge**, **purple**, **Supplementary Table 9, Supplementary Table 10**).*

Table 3. *Novel loci pleiotropically associated with CM and nevus count or hair colour.*

CHR:BP	rsID	Gene	EA/ NEA	FREQ	CM P	CM + nevus P	CM + Hair P
1:24,787,947	rs195720	*STPG1*	A/G	0.389	2.26×10^{-5}	-	4.78×10^{-12}
1:78,450,517	rs34517439	*FUBP1*	C/A	0.900	3.71×10^{-4}	-	3.36×10^{-12}
1:120,528,932	rs2453042	*NOTCH2*	G/A	0.151	1.17×10^{-6}	2.07×10^{-8}	-
1:200,265,151	rs183661447	*RP11-532L16.3, ZNF281*	A/G	0.008	2.79×10^{-7}	3.25×10^{-8}	-
1:214,673,271	rs7533482	*PTPN14*	T/C	0.788	2.27×10^{-5}	-	2.09×10^{-13}
2:135,430,709	rs6745983	*TMEM163*	G/A	0.49	2.42×10^{-3}	-	4.85×10^{-13}
2:214,065,880	rs16849932	*IKZF2*	G/A	0.918	3.30×10^{-3}	-	2.49×10^{-10}
4:37,470,753	rs11730662	*C4orf19*	C/T	0.297	2.44×10^{-3}	1.77×10^{-8}	-
4:106,260,427	rs17321108	*RN7SL89P, PPA2*	C/T	0.040	1.05×10^{-7}	1.41×10^{-8}	-
5:56,011,357	rs7714232	*MAP3K1*	A/T	0.821	6.81×10^{-4}	-	3.21×10^{-22}
6:7,189,567	rs75818295	*RREB1*	C/T	0.955	1.87×10^{-3}	-	8.27×10^{-10}
6:11,637,483	rs548304	*ADTRP*	G/A	0.803	2.67×10^{-5}	-	1.46×10^{-10}

6:15,503,696	rs10949304	*DTNBP1*	G/C	0.451	0.00155	4.2×10^{-9}	-
6:50,790,642	rs2857482	*TFAP2B*	C/T	0.107	4.60×10^{-5}	4.59×10^{-10}	-
6:151,577,739	rs10434896[a]	*AKAP12*	A/T	0.524	2.13×10^{-7}	1.77×10^{-9}	6.36×10^{-42}
8:13,113,8979	rs111595456	*ASAP1*	C/T	0.033	7.88×10^{-4}	7.35×10^{-10}	-
9:235201	rs593179[a]	*CBWD1*	A/G	0.550	1.97×10^{-6}	6.16×10^{-12}[b]	3.03×10^{-43}
10:111,889,779	rs11194997	*MIX1*	G/A	0.808	3.45×10^{-6}	-	2.70×10^{-11}
11:419,994	rs12418575	ANO9, SIGIRR	C/A	0.809	8.75×10^{-6}	4.71×10^{-8}	-
11:7,543,519	rs11041426	*CTD-2516F10.2*	G/A	0.397	1.56×10^{-4}	-	3.72×10^{-33}
11:62,206,288	rs9645690	*AHNAK*	C/T	0.664	2.04×10^{-5}	2.55×10^{-8}	9.44×10^{-34}
11:91,616,691	rs12225068	*FAT3*	A/G	0.917	3.80×10^{-5}	-	6.48×10^{-10}
13:76,351,286	rs474240	*LMO7*	A/G	0.354	1.42×10^{-4}	-	6.18×10^{-9}
13:114,744,546	rs75414584	*RASA3*	C/T	0.900	6.31×10^{-3}	-	4.62×10^{-12}
14:69,226,931	rs11625064	*ZFP36L1*	T/C	0.634	6.27×10^{-6}	4.05 × 10-10	3.21×10^{-19}
14:92,795,912	rs4904871	*SLC24A4*	A/G	0.436	7.78×10^{-4}	-	7.31×10^{-276}
14:103,852,725	rs3825566	Multiple	C/T	0.366	1.51×10^{-4}	-	1.42×10^{-16}
15:26,106,257	rs12906552	*ATP10A*	T/G	0.250	1.01×10^{-4}	-	2.02×10^{-8}
15:48,426,484	rs1426654	*SLC24A5*	A/G	0.946	3.17×10^{-3}	-	1.43×10^{-9}

Table 3 Footnote: *Results for the lead variants from pleiotropic loci (lead SNP reaching $P < 5 \times 10^{-8}$ and GWAS-PW Model 3 PPA > 0.5,* ***Online methods****) distinct to those in the total CM meta-analysis (****Table 1****,* ***Table 2****). **CHR, BP:** hg19 positional information. **rsID**: dbSNP142 rs number. **Gene** prioritises genes that the variant is an eQTL for in GTEx skin datasets or otherwise is the closest protein coding gene; multiple indicates three or more genes. We report the total CM meta-analysis P (**CM P**), and the **CM+nevus** or **CM+hair** colour Stouffer's meta-analysis P-value. Full results can be found in Supplementary Tables 7 and 8. [a]Peak SNPs are in LD; rs10434896 (CM+nevus) and rs10434895 LD $r^2_{EUR} = 0.99$; rs593179 (CM+nevus) and rs520015 $r^2_{EUR} = 0.63$. [b]Locus previously reported as pleiotropically associated with CM and nevus count, but not significant for CM alone.*

Table 4. ***Genes identified by melanocyte TWAS outside of regions identified in the CM GWAS meta-analysis.***

		Locus Peak CM Variant			TWAS		
Gene	**TWAS P**	**rsID**	**CHR:BP**	**CM P**	**Z**	**P**	**Heritability**
NIPAL3	6.39×10^{-6}	rs2294524	1:24,770,594	9.58×10^{-7}	4.51	6.39×10^{-6}	0.674
NOTCH2	1.39×10^{-6}	rs2793830	1:120,466,108	3.77×10^{-7}	4.83	1.39×10^{-6}	0.265
PTPN14	1.35×10^{-6}	rs7533482	1:214,673,271	2.24×10^{-5}	-4.83	1.35×10^{-6}	0.374
CBWD1	3.85×10^{-6}	rs478882	9:205,964	3.84×10^{-6}	-4.62	3.85×10^{-6}	0.673
C9orf66	1.32×10^{-6}	rs478882	9:205,964	3.84×10^{-6}	4.84	1.32×10^{-6}	0.339
RIN3	1.32×10^{-5}	rs8010344	14:93,086,918	4.88×10^{-6}	-4.36	1.32×10^{-5}	0.753
IRX6	1.10×10^{-6}	rs12919110	16:55,319,789	2.14×10^{-6}	-4.87	1.10×10^{-6}	0.242
P2RY11	1.20×10^{-5}	rs3826785	19:10,227,149	8.15×10^{-5}	-4.38	1.20×10^{-5}	0.353

Table 4 Footnotes: *For each gene with a $P_{TWAS} < 1.48 \times 10^{-5}$ that does not overlap an existing CM region we report the local peak CM variant from the total confirmed plus self report GWAS meta-analysis, and TWAS Z score and P-value, as well as TWAS heritability result. Full results for all*

genes with a $P_{TWAS} < 1.48 \times 10^{-5}$ and FDR < 0.05 can be found in Supplementary Table 9. CBWD1 and C9orf66 are within 1 Mb of each other are are merged into a single locus.

Table 5. *Genes identified by all tissue TWAS outside of regions identified in the CM GWAS meta-analysis.*

Gene	CHR	Start	End	Max. TWAS P-value	Min. TWAS P-value	Locus
NIPAL3	1	24,742,284	24,799,466	6.39E-06	6.39E-06	1
RP11-109P14.10	1	38,326,369	38,327,252	2.75E-05	1.49E-05	2
PTPN22	1	114,356,433	114,414,381	2.05E-05	2.05E-05	3
AP4B1-AS1	1	114,399,257	114,443,859	5.12E-06	5.12E-06	3
AP4B1	1	114,437,370	114,447,823	1.58E-05	8.03E-06	3
NOTCH2	1	120,454,176	120,612,240	1.39E-06	1.39E-06	4
PTPN14	1	214,522,039	214,725,792	1.35E-06	1.35E-06	5
WDPCP	2	63,348,518	64,054,977	1.97E-05	1.65E-06	6
ST3GAL6	3	98,451,080	98,540,045	1.47E-05	3.34E-06	7
FGFR3	4	1,795,034	1,810,599	1.79E-06	1.79E-06	8
SLC22A5	5	131,705,444	131,731,306	1.18E-05	1.18E-05	9
AC000120.7	7	91,829,328	91,836,171	2.62E-06	2.62E-06	10
SRPK2	7	104,751,151	105,039,755	3.64E-06	2.86E-06	11
ASAH1	8	17,913,934	17,942,494	3.36E-05	3.36E-05	12
CBWD1	9	121,041	188,979	1.74E-05	2.90E-06	13
C9orf66	9	213,108	215,893	1.32E-06	1.32E-06	13
SMC2	9	106,856,541	106,903,698	7.10E-05	9.24E-06	14
PIP5KL1	9	130,683,158	130,693,076	9.59E-06	9.59E-06	15
DPM2	9	130,697,378	130,700,763	4.03E-05	4.03E-05	15
RP11-339B21.10	9	131,193,877	131,194,285	7.60E-06	3.83E-06	15
ODF2	9	131,217,465	131,263,571	4.35E-06	4.35E-06	15
GDI2	10	5,807,186	5,884,095	2.47E-05	2.22E-06	16
ADD3	10	111,756,126	111,895,323	1.92E-05	7.39E-06	17
NR1H3	11	47,269,851	47,290,396	2.13E-05	2.13E-05	18
GCN1L1	12	120,565,007	120,632,513	4.00E-05	4.00E-05	19
RIN3	14	92,980,118	93,155,339	1.47E-05	1.32E-05	20

IRX6	16	55,357,672	55,364,672	1.10E-06	1.10E-06	21
ARL17B	17	44,352,150	44,439,130	2.49E-05	7.34E-06	22
MAR2	19	8,478,154	8,503,901	4.58E-05	4.58E-05	23
C19orf66	19	10,196,798	10,203,928	2.08E-05	2.08E-05	24
P2RY11	19	10,222,214	10,226,048	1.20E-05	1.20E-05	25
ALDH16A1	19	49,956,426	49,974,305	3.38E-05	3.38E-05	25

Table 5 Footnotes: *For each gene with a $P_{TWAS} < 5.0 \times 10^{-5}$ that does not overlap an existing CM region we report the **gene** name and position, as well as the **minimum** and **maximum TWAS P-value**. Full results for all genes can be found in **Supplementary Table 10**. Genes within 1 Mb of each other are merged into a single **locus**.*

Colocalization analyses using a previously described method, eCAVIAR [48], and melanocyte eQTLs nominated 8 genes from 7 CM loci at a colocalization posterior probability (CLPP) cutoff of 1% (**Supplementary Table 11**). eQTLs from two types of skin tissue also added colocalization for 4 additional CM loci. While this conservative method provided additional support for our TWAS results by nominating the same genes for some of the novel CM GWAS loci (e.g. *MSC* on Chr8, and *MPHOSPH6* on Chr16), it also provided additional information such as identifying *MED13L* as the putative target gene at the chromosome 12 locus tagged by rs113469387. Combining the candidate susceptibility gene list from TWAS and colocalization analyses, we performed pathway enrichment analysis using the Ingenuity Pathway Analysis (IPA) tool. Among significantly enriched canonical pathways were those relevant to melanoma and cancer-related signaling including cell cycle regulation and apoptosis, melanocyte-specific metabolism, telomerase signaling, as well as immune response pathways (**Supplementary Table 12**). Notably, the melanocyte master regulator, MITF, was identified as the top upstream regulator of the candidate susceptibility genes ($P = 1.7 \times 10^{-5}$; **Supplementary Table 13**), further supporting the role of melanocyte-specific regulation underlying melanoma susceptibility.

New loci for CM shed light on the role of the immune system and DNA repair

The rs28986343 variant (OR = 1.15, $P_{meta} = 1.6 \times 10^{-8}$) in the *HLA* region supports the role of immunity in the etiology of CM. Interestingly, two variants in this *HLA* region, rs9275642 and rs1050529, have been associated with basal cell carcinoma (BCC) [49]; the CM variant rs28986343 is independent to both of these BCC variants ($r^2_{EUR} < 0.03$). Our TWAS analyses from 4 tissue types identified *SKIV2L* as a likely susceptibility gene in this locus (**Supplementary Table 10**). *SKIV2L* encodes an RNA helicase, which is a signature component of cytosolic 3'-to 5' RNA exosome. SKIV2L RNA exosome was shown to have roles in innate immunity by limiting the activation of a viral RNA sensing mechanism and potentially suppressing autoimmune response [50]. In addition, the rs6908626 variant on chromosome 9 (**Table 2**) is in LD r^2_{EUR} 0.97 with SNPs (e.g. rs72928038) associated with autoimmune disorders including vitiligo [51], and rs6908626 is near *BACH2*, a gene that plays a critical role in immune regulation [52],

A region on chromosome 17, with peak SNP rs78378222 (**Table 2**), is pleiotropically associated with nevus count (**Supplementary Table 7**) and spans the 3′ UTR of *TP53*. This gene further highlights the importance of DNA repair for CM susceptibility. *TP53* responds to cellular stresses to regulate target gene expression resulting in DNA repair, cell cycle arrest, apoptosis, and cellular senescence, which may explain the association with both CM and nevus count. Rare germline mutations in *TP53* lead to Li-Fraumeni syndrome [53] which is associated with early onset of cancer, including CM [54]. The rs78378222 *TP53* variant has few other variants in LD. Evidence suggests

rs78378222 disrupts the polyadenylation signal sequence from AATAAA to AATACA which impairs proper 3′ termination and polyadenylation of the *TP53* transcript resulting in reduced RNA expression levels [55]. Associations with rs78378222 are similarly observed for BCC [49,55] and glioma [56,57]. We also identify an independant additional CM associated variant at the *TP53* locus, rs1641548 (**Supplementary Table 3**; OR = 1.15, $P_{meta} = 4.1 \times 10^{-10}$), which has not been reported in association with other traits.

TWAS analysis using melanocyte-specific expression (**Supplementary Table 9**) of the the chromosome 16 locus tagged by rs4420522 (**Table 2**, OR = 1.08, $P_{meta}= 4.25\times10^{-14}$) indicates the likely target genes are *CDH1*, *ZFP90*, *FTLP14*, and *TANGO6*. Among these, *CDH1* has been best characterized for its roles in melanoma and other cancer types. *CDH1* is a calcium-dependent protein that regulates cell adhesion and motility, and may contribute to carcinogenesis by increasing cellular proliferation, invasion, and metastasis. Germline mutations in this gene are associated with a variety of tumors including gastric [58], breast [59], and potentially colorectal cancer [60]. HaploReg [61] indicates the tagging rs4420522 variant is located in a large LD block with many putatively functional variants, and is itself an enhancer in a range of cell types including melanocytes and keratinocytes. The peak CM SNP rs4420522 is in LD $r^2_{EUR} = 0.93$ with rs9929218, which is a GWAS hit for colorectal cancer [62].

A variant near *TERC* (rs3950296) was also newly associated with CM risk alone (OR = 1.08, $P_{meta} = 8.5 \times 10^{-11}$), having previously been reported as a pleiotropic locus for nevus count and CM [28]. *TERC* is a non-coding RNA that binds to *TERT* and serves as a TTAGGG template for telomere length repeat expansion. In addition, new and known CM loci are associated with telomere length including rs7705526/*TERT*, rs4731207/*POT1*, rs2911423/*MPHOSPH6*, rs7902587/*OBFC1,* and rs143190905/*RTEL1* (**Table 1-2**, **Supplementary Table 5**). Among these, expression of *POT1*, *MPHOSPH6*, and *OBFC1* were shown to be associated with CM risk by our TWAS data (**Supplementary Table 9, 10**). This provides further support for the role of telomere length in CM susceptibility. Longer telomeres delay cellular senescence and are associated with increased CM susceptibility [11,63,64]. Telomere length has been associated with nevus count [10]. This relationship may allow for more time for melanocytes to acquire damaging mutations. Our report, as well as prior CM GWAS identified variants near *CCND1* (rs4354713), *ATM* (rs1801516), and *PARP1* (rs2695237), all genes with roles in telomere maintenance, DNA repair, and regulation of senescence [34,65].

When considering the additional loci associated with CM (**Figure 2, Supplementary Table 7-10,14**) some interesting highlights stand out. For example, across known loci and those reported here for the first time there are 12 loci that are pleiotropic for CM, pigmentation and nevus count (**Figure 2**, **Tables 1-2**, **Supplementary Table 7-8**). Of these, the chromosome 6 locus with peak SNP rs2857482 spans the *TFAP2B* gene, which is involved in the proliferation and differentiation of melanocytes from neural crest cells [66]. Some loci receive support from multiple approaches, with variants near *NOTCH2* that may play a role in the maintenance and differentiation of melanocytes [67], reaching genome-wide significance in the combined CM and nevus analysis and the MAGMA gene-based test (**Supplementary Tables 7, 13**). TWAS identifies the expression of *NOTCH2* as significantly associated with CM, and the peak eQTL for *NOTCH2* in melanocytes, rs2641316, is in LD $r^2_{EUR} = 0.99$ with the peak CM+nevus variant rs2453042. *IRX6* overlaps a CM and nevus locus previously reported [28]on chromosome 16, with the lead GWAS variant rs12930459 in LD $r^2_{EUR} = 0.75$ with the local TWAS peak rs12919110 (**Table 3**, **Supplementary Table 7, 9**). Among the loci found only in the MAGMA gene-based analysis (**Supplementary Table 14)**, *LPP* is noteworthy for its association with vitiligo. The peak SNP for vitiligo, rs13076312 [51] is modestly associated with CM ($P_{meta} = 2.0 \times 10^{-6}$, $I^2 = 4.1\%$). As expected given it was detected via a gene based analysis, there are additional associations in this region, the strongest at rs73887368 ($P_{meta} = 1.8 \times 10^{-7}$, OR = 1.10, $I^2 = 3.1\%$), which is independent to the vitiligo association (LD $r^2_{EUR} < 0.01$).

Discussion

Our GWAS of CM is the largest to date with over three times the effective sample size of prior CM GWAS analyses (**Supplementary Table 1**). By harnessing self-reported CM cases in the total GWAS meta-analysis we identified a total of 56 CM susceptibility loci with 68 independent variants across these loci. In parallel our CM meta-analysis also confirms several loci previously identified by TWAS [29], supporting our use of TWAS in this analysis to identify additional genes associated with CM (**Table 4**, **Table 5**). The same TWAS analysis, as well as eQTL colocalization and multimarker genomic annotations identified promising gene candidates at these risk loci. Pairwise GWAS with CM-related nevi count and pigmentation traits, TWAS, and gene-based approaches identified a further 53 independent loci, that while not formally reaching genome-wide significance for CM alone, likely represent additional risk loci. In total, our integrative analysis of CM susceptibility identified over 100 genomic loci and genes important for CM susceptibility (**Tables 1-4, Figure 2**), constituting a substantial leap in progress in terms of understanding CM genetic architecture beyond the previously identified 21 CM GWAS susceptibility loci (**Table 1**).

With the inclusion of a large series of self-reported CM cases, our analyses showed high genetic correlation between self-reported and clinically confirmed cases (**Supplementary Table 2**), indicating self-reported CM cases are a valuable resource for genomic CM studies. Gains in power from the inclusion of self-reported CM cases allowed for the identification of 11 additional CM susceptibility loci than the analysis restricted to confirmed CM cases only (**Supplementary Table 3, 6**). Furthermore, by incorporating cases from a variety of geographic areas, including the highly sun-exposed with generally darker pigmentation and often underrepresented Mediterranean CM cases, allowed for assessment of CM genetic susceptibility across geographic regions. Interestingly, we find no evidence for differences in CM locus effect estimates by contributing GWAS (**Supplementary Figure 8**) or differences in effect size and allele frequency by geographic regions (**Supplementary Figure 9**) suggesting the genetic architecture of identified CM loci remains similar across geographic region despite differences in environmental UV light exposure. However, the stratified analysis based on CM histological subtypes, identified acral lentiginous melanomas, which were more frequent in Mediterranean cases, as uniquely unrelated to pigmentation loci. This suggests that acral lentiginous melanoma may not benefit from the same public health preventative measures as other CM subtypes. In contrast, the stratified analyses based on age at diagnosis and gender found no evidence for differences in distribution of nevi or pigmentation-related loci.

The expansion of new loci and genes discovered in our analysis augments past understanding of CM risk. Our results confirm the highly-interrelated relationship of CM with nevus count and pigmentation traits by showing that loci previously identified via nevus count[28] or pigmentation [41] can be directly associated with CM (**Table 1**). In turn we build on this relationship by performing expanded pleiotropic analyses, identifying additional loci associated with both these traits and CM, but not yet significantly associated with CM alone (**Table 3**). Interestingly, following these expanded pleiotropic analyses, many loci were shown not to be associated with either nevus count or pigmentation, indicating that we are also identifying risk variants that act outside of these classic CM risk phenotypes (**Table 1, Table 2)**. Insights as to which pathways these additional loci may act through can be derived by TWAS, which identified genes that emphasize the roles of immunity (*CDH1*, *SKIV2L*), as well as DNA repair and telomere length (*TP53*, *POT1*). For other loci (e.g. rs12539524 on chromosome 7, Table 2) we are unable to yet assign a putative function, suggesting additional biological mechanisms important for CM etiology remain to be discovered.

Our large, international genetic meta-analysis showcases the utility of including self-reported CM cases, complementary analytical approaches, and data from multiple sources to improve understanding of CM susceptibility and provide strong support for the role of nevi formation, pigmentation, telomere length and immunity in the etiology of CM.

Online Methods

Quality Control, Imputation and Association Analysis

Data cleaning was performed using Illumina GenomeStudio/BeadStudio (San Diego, CA, USA) and PLINK (v1.90b5.4) [68,69]. Prior to imputation any SNP with either minor allele frequency (MAF) < 0.01, Hardy-Weinberg Equilibrium (HWE) P-value $5 < 10^{-4}$ in controls or $5 < 10^{-10}$ in cases was removed. Similarly, any individual was removed who was missing > 0.03 of variants, had heterozygosity values either > 0.05 or < -0.05 or 3 sd from the mean, whose genetically-predicted sex did not match their recorded sex, or who was determined to be non-European based on principal component analysis (PCA) was removed. In addition one of any pair of individuals estimated to be related with identity by descent (IBD) pihat > 0.15 was removed.

Imputation was conducted using the Michigan Imputation Server with the Haplotype Reference Consortium panel (HRC version 1) and run using Minimac3 [70]. Following imputation, any imputed variant with imputation quality score $r^2 < 0.5$ or MAF < 0.0001 was rejected. To handle variants with the same name (e.g. triplicate SNPs), variant IDs were converted to the format CHR:BP:A1A2 prior to meta-analysis.

Logistic regression was then conducted using PLINK (v1.90b5.4) [68,69] with either geographic region (in GenoMEL Phase 1 and 2 data) or principal components as covariates to account for potential population stratification. Individual studies were checked for evidence of inflation by producing QQ plots (**Supplementary Figure 1**) and calculating the corresponding inflation factor λ and LDSC intercept (**Supplementary Table 1**).

Where individual studies have deviated from this protocol, details are included in the study description in the Supplementary Material.

Meta-Analysis, conditional analysis and loci selection

Meta-analysis of the GWAS were conducted using both inverse-variance weighted fixed effects and random effects meta-analysis [71] as implemented in PLINK v1.90b5.4 [68,69]. Meta-analyses were conducted for confirmed only cases, and in the total set including self report sets (23andMe, Inc. and a portion of UK Biobank).

Genome-wide Complex Trait Analysis (GCTA, v1.26.0) was used to identify independently associated variants [32]. To ensure we were only detecting completely independent SNPs the collinearity threshold (--cojo-collinear) was set to $r^2 = 0.05$. The threshold for genome-wide significance 5×10^{-8} and fixed effect meta-analysis p-values and log(OR) effect sizes were analysed.

Linkage-disequilibrium between SNPs for conditional mapping, and where reported in the manuscript, was calculated using a reference population of 5,000 individuals selected randomly from the portion of the UK Biobank population determined to be European by PCA (LD_{EUR}). Variants were converted to best guess genotype (threshold 0.3). Best guess data were cleaned for missingness > 0.03, HWE $P < 1 \times 10^{-6}$, MAF < 0.001

To limit the chance of false positive claims of novel SNP/loci, we further filtered the list of 75 conditionally independent variants (**Supplementary Table 4**) to those (i) genome-wide significant ($P < 5 \times 10^{-8}$) in single SNP and joint conditional analysis, and (ii) as recommended [72] where there was evidence of heterogeneity between studies ($I^2 > 31\%$) the random effect P-value also needed to be $< 5 \times 10^{-8}$. Passing variant were further checked to ensure that MAFs and effect sizes were consistent across studies and that the result was not driven by a single study (**Supplementary**

Figures 8-9). The 68 retained variants were combined into 56 loci using a concatenating 1mb window (**Supplementary Table 3**). Regional association plots for all 56 loci were interactively plotted by LDassoc (https://ldlink.nci.nih.gov/) [73] and included as Supplementary Materials.

Joint analyses of CM and nevus count and pigmentation

Nevus GWAS meta-analysis

Using beta meta-analysis weighted by SE as implemented in PLINK 1.90b5.4 we combined the recently published nevus meta-analysis (N = 52,506) [28] which excluded samples with melanoma but may include a small portion of overlap with the controls used for some melanoma GWAS datasets; participants of the QSKIN study with nevus count that are are non overlapping and unrelated (IBD pihat < 0.15) to the Q-SKIN melanoma case control set (N = 12,930) and new nevus GWAS using Q-TWIN participants (N = 341). Total sample size was 65,597.

Pigmentation GWAS

A GWAS for hair colour was performed on 352,662 UK Biobank samples not included in the melanoma GWAS who self-reported having either blonde, light brown, dark brown or black hair (coded as 1,2,3,4). Hair colour was then treated as a continuous variable and regressed on imputed genotype adjusting for principal comments using the same approach as for the melanoma GWAS.

Joint analyses

The melanoma results were then jointly analysed first with nevus count and then with hair colour. Two approaches were to this were taken. Firstly the total confirmed plus self report CM GWAS meta-analysis results were combined with the separate nevus and pigmentation GWAS data using Stouffer's method (P-value weighted by per SNP sample N) as implemented in METAL [74]. FUMA [75] was used to identify independent SNPs with $P < 5 \times 10^{-8}$; independent SNPs within 1Mb were considered to be single loci. Secondly the melanoma and pigmentation/nevus GWAS results were analysed using GWAS-PW [44], which estimates the posterior probability of four possible models for each genetic region: (i) Association with CM only, (ii) association with the second trait only, (iii) association with both traits (pleiotropic), (iv) association with both traits, but co-located and independent (v) No association with either trait. Given that nevus count and pigmentation are believed to act directly on melanoma risk, model (iv) seemed unrealistic so we only considered models (i), (ii), (iii) and (v). For nevus count, SNPs were assigned to blocks using the recommended boundaries for GWAS-PW (https://bitbucket.org/nygcresearch/ldetect-data). For hair CM and hair colour, 50 SNP windows were used for blocks as the default LD blocks contained multiple independent hair colour loci. Following the approach taken by [28], any locus with a top SNP reaching $P < \times 10^{-8}$ for the combined CM and nevus/hair colour analysis and with a posterior probability > 0.5 that the locus is associated with both traits (model 3) to ensure that the association is not driven by a single trait was declared to be pleiotropically associated with both traits.

MAGMA gene-based analysis

Multi-marker Analysis of GenoMic Annotation (MAGMA) as implemented in FUMA to analyse summary GWAS data [75] is a gene-based approach that accounts for linkage disequilibrium between SNPs and genes by fitting a multiple regression model [45]. This was applied to the total (confirmed melanoma plus self report melanoma) GWAS meta-analysis, and gene-based P-values were deemed significant at a bonferroni corrected threshold adjusting for 18,732 genes ($P < 2.67 \times 10^{-6}$). Significant genes were assigned to a single locus where applicable using an expanding 1 Mb window as done for single SNP associations. Loci were deemed novel if they did not overlap with a loci identified in the single-SNP analysis. Where loci overlapped with genes/regions identified by

TWAS or combination with risk phenotypes (e.g. nevus count) this is noted in **Supplementary Table 14**.

Analysis of pigmentation and nevi polygenic risk score across melanoma subtypes

For each subject in our study, we calculated two polygenic scores (PRS), using 276 genetic variants associated with pigmentation and 10 genetic variants associated with nevi count. Nevi count SNPs were derived from the same Nevus GWAS meta-analysis used for the pleiotropic analysis (N = 65,597), with independant lead SNPs with $P < 5 \times 10^{-8}$ identified using FUMA [75], with the LD r^2 cut off for independence < 0.05. Pigmentation PRS SNPs were selected by XXXXX. PRS were calculated for each subject using the regression coefficient from the GWAS of pigmentation or nevi count and the genotypic value of the SNP for the subject. We then tested if PRS distribution differed between males and females, across age groups, and histology subtypes. In total, we performed 27 comparisons and thus any comparison with p-value less than 0.05/27 was declared as statistically significant.

GENESIS estimation of heritability and polygenic risk

We used GENESIS [31] to estimate the genetic architecture (number of causal SNPs and their effect size distribution) using the summary level statistics from the GWAS meta-analysis. Quantile-quantile plot comparing the p-values generated from this fitted distribution against the observed p-values suggested a three component Gaussian mixture model for the effect size distribution. Based on this estimated genetic architecture, we calculated the heritability at the observational scale and the number of SNP reaching genome-wide significance for a given GWAS with known sample size. Similarly, GENESIS calculated the AUC for an additive polygenic risk prediction model built based on a discovery GWAS of known sample size.

UK Biobank melanoma risk phenotype GWAS

Methods for the UK BB GWAS data included in sup table 5 e.g. ease of tanning etc.

Linkage Disequilibrium Score Regression

As LD score regression (LDSC) is sensitive to the quality of input SNPs, GWAS or meta-analysis variants were filtered to the list of high quality HapMap SNPs provided [76]. Using LD Score regression v1.0.0 genomic inflation (Lamdba), Intercept and SNP-heritability (h^2) was estimated. h^2 estimates were converted to the liability scale using the lifetime population prevalence for CM in Australia (0.0588) [77].

LD score regression of tissue-specific genes

Melanoma heritability enrichment for SNPs around tissue-specific genes was assessed using LD score regression as described previously (Finucane et al. 2018; Zhang et al. 2018). We used stratified LD score regression implemented in LDSC program (https://github.com/bulik/ldsc). First, to reduce batch effects, RNA-seq data for both GTEx tissues and primary melanocyte was quantified as RPKM using RNA-SeQC (v1.18) [78] followed by quantile normalization. To define the tissue-specific genes, we calculated the t-statistic of each gene for a given tissue, excluding all samples from the same tissue category (we treated the tissue category for melanocytes as "Skin" together with two types of skin and transformed skin fibroblasts). Tissue category assignment was based on the previous publications [29,79]. We selected the top 1,000, 2,000, and 4,000 tissue-specific genes by t-statistic, added a 100-kb window around their transcribed regions to define tissue-

specific genome annotation, and applied stratified LD score regression on a joint SNP annotation to estimate the heritability enrichment against the total CM GWAS data from the current study.

Colocalization of melanoma GWAS and eQTLs

For colocalization of melanoma GWAS signal and eQTLs from melanocytes and GTEx tissues, we used CAusal Variants Identification in Associated Regions (eCAVIAR, http://genetics.cs.ucla.edu/caviar/index.html) [48]. For each locus, both GWAS and eQTL summary statistics of 50 SNPs upstream and downstream of the GWAS lead SNP were extracted as the input for eCAVIAR. We computed the CLPP score with maximum number of two causal SNPs in each locus, and applied CLPP >1% (0.01) cutoff for positive co-localization. Thus, for a given GWAS variant, an eGene with a CLPP score above the colocalization cutoff is considered a target gene. To avoid reporting spurious effect, we applied a conservative criteria and only reported variants displaying LD $r^2 < 0.9$ with the CM GWAS lead SNP and genome-wide significant eQTL P-value.

TWAS

We performed 45 transcriptome-wide association studies (TWAS) by predicting the gene expression phenotypes using the total CM GWAS meta-analysis and both GTEx and melanocyte RNA-seq expression data. TWAS/FUSION (http://gusevlab.org/projects/fusion/) was used to perform the TWAS analysis, allowing for multiple prediction models, independent reference LD, additional feature statistics and cross-validation results [80]. The total CM GWAS meta-analysis summary statistics were included with no significance thresholding. The precomputed expression reference weights for GTEx gene expression (V6) RNA-seq across 44 tissue types were downloaded from TWAS/FUSION website (http://gusevlab.org/projects/fusion/). We computed functional weights from the primary melanocyte RNA-seq data [29] one gene at a time. Genes that failed quality control during the heritability check (using minimum heritability P-value 0.01) were excluded from the further analyses. We restricted the *cis*-locus to 500kb on either side of the gene boundary. A genome-wide significance cutoff (TWAS P-value < 0.05/number of genes tested) as well as FDR < 0.05 was applied to the final TWAS result.

URLs and Software versions

TBC

Acknowledgments

Please see the Supplementary Note for acknowledgments.

Author Contributions

MTL, MMI, TB, MHL, SM - Project design, data collection, analysis, funding support, article writing.

Bibliography

1. Lens, M. B. & Dawes, M. Global perspectives of contemporary epidemiological trends of cutaneous malignant melanoma. *Br. J. Dermatol.* **150**, 179–185 (2004).
2. Elwood, J. M. & Jopson, J. Melanoma and sun exposure: an overview of published studies. *Int. J. Cancer* **73**, 198–203 (1997).
3. Cannon-Albright, L. A., Bishop, D. T., Goldgar, C. & Skolnick, M. H. Genetic predisposition to cancer. *Important Adv. Oncol.* 39–55 (1991).
4. Naldi, L. *et al.* Cutaneous malignant melanoma in women. Phenotypic characteristics, sun exposure, and hormonal factors: a case-control study from Italy. *Ann. Epidemiol.* **15**, 545–550 (2005).
5. Titus-Ernstoff, L. *et al.* Pigmentary characteristics and moles in relation to melanoma risk. *Int. J. Cancer* **116**, 144–149 (2005).
6. Holly, E. A., Aston, D. A., Cress, R. D., Ahn, D. K. & Kristiansen, J. J. Cutaneous melanoma in women. I. Exposure to sunlight, ability to tan, and other risk factors related to ultraviolet light. *Am. J. Epidemiol.* **141**, 923–933 (1995).
7. Holly, E. A., Aston, D. A., Cress, R. D., Ahn, D. K. & Kristiansen, J. J. Cutaneous melanoma in women. II. Phenotypic characteristics and other host-related factors. *Am. J. Epidemiol.* **141**, 934–942 (1995).
8. Bataille, V. *et al.* Risk of cutaneous melanoma in relation to the numbers, types and sites of naevi: a case-control study. *Br. J. Cancer* **73**, 1605–1611 (1996).
9. Chang, Y. M. *et al.* A pooled analysis of Melanocytic nevus phenotype and the risk of cutaneous melanoma at different latitudes. *International Journal of Cancer* (2009). doi:10.1002/ijc.23869
10. Bataille, V. *et al.* Nevus size and number are associated with telomere length and represent potential markers of a decreased senescence in vivo. *Cancer Epidemiol. Biomarkers Prev.* (2007). doi:10.1158/1055-9965.EPI-07-0152
11. Han, J. *et al.* A prospective study of telomere length and the risk of skin cancer. *J. Invest. Dermatol.* (2009). doi:10.1038/jid.2008.238
12. Hollenbeak, C. S. *et al.* Increased incidence of melanoma in renal transplantation recipients. *Cancer* **104**, 1962–1967 (2005).
13. Kamb, A. *et al.* Analysis of the p16 gene (CDKN2) as a candidate for the chromosome 9p melanoma susceptibility locus. *Nat. Genet.* **8**, 23–26 (1994).
14. Berwick, M. *et al.* The prevalence of CDKN2A germ-line mutations and relative risk for cutaneous malignant melanoma: an international population-based study. *Cancer Epidemiol. Biomarkers Prev.* **15**, 1520–1525 (2006).
15. Robles-Espinoza, C. D. *et al.* POT1 loss-of-function variants predispose to familial melanoma. *Nat. Genet.* (2014). doi:10.1038/ng.2947
16. Shi, J. *et al.* Rare missense variants in POT1 predispose to familial cutaneous malignant melanoma. *Nat. Genet.* **46**, 482–486 (2014).
17. Palmer, J. S. *et al.* Melanocortin-1 receptor polymorphisms and risk of melanoma: is the association explained solely by pigmentation phenotype? *Am. J. Hum. Genet.* **66**, 176–186 (2000).
18. Landi, M. T. *et al.* MC1R, ASIP, and DNA repair in sporadic and familial melanoma in a mediterranean population. *J. Natl. Cancer Inst.* (2005). doi:10.1093/jnci/dji176
19. Brown, K. M. *et al.* Common sequence variants on 20q11.22 confer melanoma susceptibility. *Nat. Genet.* **40**, 838–840 (2008).
20. Bishop, D. T. *et al.* Genome-wide association study identifies three loci associated with melanoma risk. *Nat. Genet.* (2009). doi:10.1038/ng.411
21. Amos, C. I. *et al.* Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. *Hum. Mol. Genet.* **20**, 5012–5023 (2011).
22. Barrett, J. H. *et al.* Genome-wide association study identifies three new melanoma susceptibility loci. *Nat. Genet.* (2011). doi:10.1038/ng.959
23. Macgregor, S. *et al.* Genome-wide association study identifies a new melanoma susceptibility

927 locus at 1q21.3. *Nat. Genet.* **43**, 1114–1118 (2011).
928 24. Iles, M. M. *et al.* A variant in FTO shows association with melanoma risk not due to BMI. *Nat.*
929 *Genet.* **45**, 428–32, 432e1 (2013).
930 25. Law, M. H. *et al.* Genome-wide meta-analysis identifies five new susceptibility loci for
931 cutaneous malignant melanoma. *Nat. Genet.* (2015). doi:10.1038/ng.3373
932 26. Ransohoff, K. J. *et al.* Two-stage genome-wide association study identifies a novel
933 susceptibility locus associated with melanoma. *Oncotarget* **8**, 17586–17592 (2017).
934 27. Yokoyama, S. *et al.* A novel recurrent mutation in MITF predisposes to familial and sporadic
935 melanoma. (2011). doi:10.1038/nature10630
936 28. Duffy, D. L. *et al.* Novel pleiotropic risk loci for melanoma and nevus density implicate
937 multiple biological pathways. *Nat. Commun.* **9**, 4774 (2018).
938 29. Zhang, T. *et al.* Cell-type-specific eQTL of primary melanocytes facilitates identification of
939 melanoma susceptibility genes. *Genome Res.* **28**, 1621–1635 (2018).
940 30. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat.*
941 *Genet.* (2015). doi:10.1038/ng.3406
942 31. Zhang, Y., Qi, G., Park, J.-H. & Chatterjee, N. Estimation of complex effect-size distributions
943 using summary-level statistics from genome-wide association studies across 32 complex traits.
944 *Nat. Genet.* **50**, 1318–1326 (2018).
945 32. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics
946 identifies additional variants influencing complex traits. *Nat. Genet.* (2012).
947 doi:10.1038/ng.2213
948 33. Bertolotto, C. *et al.* A SUMOylation-defective MITF germline mutation predisposes to
949 melanoma and renal carcinoma. *Nature* **480**, 94–98 (2011).
950 34. Choi, J. *et al.* A common intronic variant of PARP1 confers melanoma risk and mediates
951 melanocyte growth via regulation of MITF. *Nat. Genet.* **49**, (2017).
952 35. MacGregor, S. *et al.* Two novel loci for melanoma susceptibility on chromosome bands
953 1q42.12 and 1q21.3. *Nat. Genet.* (2011).
954 36. Peña-Chilet, M. *et al.* Genetic variants in PARP1 (rs3219090) and IRF4 (rs12203592) genes
955 associated with melanoma susceptibility in a Spanish population. *BMC Cancer* **13**, (2013).
956 37. Law, M. H. *et al.* Meta-analysis combining new and existing data sets confirms that the TERT-
957 CLPTM1L locus influences melanoma risk. *J. Invest. Dermatol.* **132**, 485–487 (2012).
958 38. Duffy, D. L. *et al.* Multiple Pigmentation Gene Polymorphisms Account for a Substantial
959 Proportion of Risk of Cutaneous Malignant Melanoma. *J. Invest. Dermatol.* **130**, 520–528
960 (2010).
961 39. Visconti, A. *et al.* Genome-wide association study in 176,678 Europeans reveals genetic loci
962 for tanning response to sun exposure. doi:10.1038/s41467-018-04086-y
963 40. Duffy, D. L. *et al.* IRF4 variants have age-specific effects on nevus count and predispose to
964 melanoma. *Am. J. Hum. Genet.* (2010). doi:10.1016/j.ajhg.2010.05.017
965 41. Visconti, A. *et al.* Genome-wide association study in 176,678 Europeans reveals genetic loci
966 for tanning response to sun exposure. *Nat. Commun.* **9**, 1684 (2018).
967 42. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat.*
968 *Genet.* **48**, 1279–1283 (2016).
969 43. Delgado, D. A. *et al.* Genome-wide association study of telomere length among South Asians
970 identifies a second RTEL1 association signal. *J. Med. Genet.* **55**, 64–71 (2018).
971 44. Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human
972 traits. *Nat. Genet.* **48**, 709–717 (2016).
973 45. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set
974 Analysis of GWAS Data. *PLoS Comput. Biol.* (2015). doi:10.1371/journal.pcbi.1004219
975 46. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis:
976 multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
977 47. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide
978 association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
979 48. Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am.*
980 *J. Hum. Genet.* **99**, 1245–1260 (2016).

49. Chahal, H. S. *et al.* Genome-wide association study identifies 14 novel risk alleles associated with basal cell carcinoma. *Nat. Commun.* **7**, 12510 (2016).
50. Eckard, S. C. *et al.* The SKIV2L RNA exosome limits activation of the RIG-I-like receptors. *Nat. Immunol.* **15**, 839–845 (2014).
51. Jin, Y. *et al.* Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. *Nat. Genet.* (2016). doi:10.1038/ng.3680
52. Zhou, Y., Wu, H., Zhao, M., Chang, C. & Lu, Q. The Bach Family of Transcription Factors: A Comprehensive Review. *Clin. Rev. Allergy Immunol.* **50**, 345–356 (2016).
53. Li, F. P. & Fraumeni, J. F., Jr. Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? *Ann. Intern. Med.* **71**, 747–752 (1969).
54. Curiel-Lewandrowski, C., Speetzen, L. S., Cranmer, L., Warneke, J. A. & Loescher, L. J. Multiple primary cutaneous melanomas in Li-Fraumeni syndrome. *Arch. Dermatol.* **147**, 248–250 (2011).
55. Stacey, S. N. *et al.* A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat. Genet.* **43**, 1098–1103 (2011).
56. Ostrom, Q. T. *et al.* Sex-specific glioma genome-wide association study identifies new risk locus at 3p21.31 in females, and finds sex-differences in risk at 8q24.21. *Sci. Rep.* **8**, 7352 (2018).
57. Melin, B. S. *et al.* Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. *Nat. Genet.* **49**, 789–794 (2017).
58. Guilford, P. *et al.* E-cadherin germline mutations in familial gastric cancer. *Nature* **392**, 402–405 (1998).
59. Hansford, S. *et al.* Hereditary Diffuse Gastric Cancer Syndrome: CDH1 Mutations and Beyond. *JAMA Oncol* **1**, 23–32 (2015).
60. Kim, H. C. *et al.* The E-cadherin gene (CDH1) variants T340A and L599V in gastric and colorectal cancer patients in Korea. *Gut* **47**, 262–267 (2000).
61. Ward, L. D. & Kellis, M. HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* (2012). doi:10.1093/nar/gkr917
62. COGENT Study *et al.* Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat. Genet.* **40**, 1426–1435 (2008).
63. Iles, M. M. *et al.* The effect on melanoma risk of genes previously associated with telomere length. *J. Natl. Cancer Inst.* **106**, (2014).
64. Rachakonda, S. *et al.* Telomere length, telomerase reverse transcriptase promoter mutations, and melanoma risk. *Genes Chromosomes Cancer* (2018). doi:10.1002/gcc.22669
65. Derheimer, F. A. & Kastan, M. B. Multiple roles of ATM in monitoring and maintaining DNA integrity. *FEBS Lett.* **584**, 3675–3681 (2010).
66. Seberg, H. E. *et al.* TFAP2 paralogs regulate melanocyte differentiation in parallel with MITF. *PLoS Genet.* **13**, e1006636 (2017).
67. Kumano, K. *et al.* Both Notch1 and Notch2 contribute to the regulation of melanocyte homeostasis. *Pigment Cell Melanoma Res.* **21**, 70–78 (2008).
68. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
69. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
70. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
71. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control. Clin. Trials* **7**, 177–188 (1986).
72. Higgins, J. P. T. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* (2002). doi:10.1002/sim.1186
73. Machiela, M. J. & Chanock, S. J. LDassoc: an online tool for interactively exploring genome-wide association study results and prioritizing variants for functional investigation.

Bioinformatics **34**, 887–889 (2018).
74. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* (2010). doi:10.1093/bioinformatics/btq340
75. Watanabe, K., Taskesen, E., Van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* (2017). doi:10.1038/s41467-017-01261-5
76. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
77. Australian Institute of Health and Welfare. Cancer in Australia: Actual incidence data from 1982 to 2013 and mortality data from 1982 to 2014 with projections to 2017. *Asia Pac. J. Clin. Oncol.* **14**, 5–15 (2018).
78. DeLuca, D. S. *et al.* RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* **28**, 1530–1532 (2012).
79. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
80. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).