



Published in final edited form as:

Nat Genet. ; 43(11): 1114–1118. doi:10.1038/ng.958.

Genome-wide association study identifies a new melanoma susceptibility locus at 1q21.3

A full list of authors and affiliations appears at the end of the article.

Introductory paragraph

We performed a genome-wide association study of melanoma in a discovery cohort of 2,168 Australian melanoma cases and 4,387 controls, confirming several previously characterised melanoma-associated loci and identifying two novel susceptibility loci on chromosome 1. The most significant genotyped SNPs in the novel loci were at 1q21.3 nearby several genes including *ARNT* and *SETDB1* (rs7412746, $P=2.5 \times 10^{-7}$, OR=0.82) and at 1q42.12 in the DNA repair gene *PARP1* (rs3219090, $P=9.5 \times 10^{-7}$, OR=0.82). Both new findings were replicated in three independent case-control studies (Europe: 2,804 cases, 7,618 controls; United States 1: 1,804 cases, 1,026 controls; United States 2: 585 cases, 6,500 controls). Estimates of the ORs in the combined replication cohorts were 0.89 for rs7412746 ($P=1.5 \times 10^{-5}$) and 0.91 for rs3219090 ($P=3.4 \times 10^{-3}$). Meta-analysis of all case-control studies combined showed genome-wide significance ($P=9.0 \times 10^{-11}$) for rs7412746 and suggestive significance ($P=9.3 \times 10^{-8}$) for rs3219090.

Users may view, print, copy, download and text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence should be addressed to: Stuart MacGregor, Queensland Statistical Genetics, Queensland Institute of Medical Research, Royal Brisbane Hospital, 300 Herston Road, Brisbane 4029, Australia, Mailing Address: Queensland Institute of Medical Research, Locked Bag 2000, Herston, QLD 4029, Australia, tel. +61 7 3845 3563, fax. +61 7 3362 0101, stuart.macgregor@qimr.edu.au.

AUTHOR CONTRIBUTIONS

S.M., N.K.H. and K.M.B. wrote the paper. G.W.M., N.K.H., K.M.B., J.M.T., A.K.H., Z.Z.Z and M.S. designed, analyzed, and managed the sample preparation and genotyping aspects of the study. S.M. and J.Z.L. performed data analysis. D.C.W., D.L.D., G.W.M., N.K.H., S.M., J.N.P., D.R.N. and N.G.M. oversaw collection of the Queensland samples and contributed to statistical analyses, data interpretation, and manuscript preparation. G.J.M., A.E.C., E.A.H., H.S., J.A.M., J.J., M.F., M.A.J., R.F.K., G.G.G., B.K.A., J.F.A., and J.L.H. oversaw collection, genotyping and analysis of the AMFS study. D.T.B., J.A.N-B., M.M.I., H.O., S.P., G.B-S., J.H., F.D., M.T.L., T.D., R.M., E.A., B.B-de P., A.M.G., P.A.K., N.A.G., P.H. and D.E.E. contributed to collection, genotyping and analysis of the GenoMEL samples. C.I.A., Q.W., L-E.W. and J.E.L. contributed to collection, genotyping and analysis of the United States 1 - MD Anderson Cancer Center samples. A.A.Q., M.Z. and J.H. contributed to collection, genotyping and analysis of the United States 2 - Harvardsamples.

CONFLICTS OF INTEREST:

All authors have no conflict of interest or financial interest in this work.

URLs

1000 Genomes: <http://www.1000genomes.org>

EIGENSOFT: <http://genepath.med.harvard.edu/~reich/Software.htm>

MACH2: <http://www.sph.umich.edu/csg/abecasis/MACH/index.html>

LOCUSZOOM: <http://csg.sph.umich.edu/locuszoom/>

PLINK: <http://pngu.mgh.harvard.edu/~purcell/plink/>

Chicago EQTL browser: <http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/>

Sibpair: <http://www.qimr.edu.au/davidD>

Main Text

To date, genome-wide association studies (GWAS) for melanoma^{1,2}, pigmentation³ and nevocogenesis^{4,5} have identified a small number of low-penetrance melanoma susceptibility variants. These variants appear to exert their effect on melanoma risk through their role in the known melanoma-associated risk phenotypes of pigmentation and nevus count. In contrast to other cancers, these variants have shown relatively large effects on disease risk (odds ratios (OR) >1.5), with previous melanoma GWASs underpowered to detect variants of small effect. Here we describe a large melanoma GWAS with sufficient power to detect the small effects typically observed for other cancers ($1.1 < \text{OR} < 1.5$).

Melanoma cases of European descent (n=2,168) were selected from the Queensland study of Melanoma: Environment and Genetic Associations (Q-MEGA)⁶ and the Australian Melanoma Family Study⁷ (AMFS). Three Australian samples of European descent were used as controls (n=4,387)⁶⁻⁸. Samples were genotyped on Illumina SNP arrays (Cases: Omni1-Quad or HumanHap610; Controls: Omni1-Quad or HumanHap610 or HumanHap670, Table 1). Cases and controls were combined into a single data set for quality control (including principal component analysis for outlier removal) and imputation (Supplementary Note). Imputation based on 1000 Genomes Project⁹ data allowed association testing for 5,480,804 well imputed SNPs, which helped recover the full sample size for SNPs only typed on a subset of the arrays. After cleaning, the genomic inflation factor (λ) for those SNPs directly genotyped in all individuals in these discovery samples was 1.04 (Supplementary Fig. 1).

Results of tests of association for SNPs directly genotyped in all discovery samples are displayed in Figure 1 (a similar pattern was seen for imputed SNPs, data not shown). Three of the previously reported melanoma susceptibility loci (*MC1R*, *ASIP*, *MTAP/CDKN2A*)¹⁻⁴ reached genome-wide significance. Two additional regions were noteworthy at chromosome 1q42.12 and 1q21.3; for both loci there was at least one SNP directly genotyped in all discovery samples with $P < 1 \times 10^{-6}$ (Table 2, Supplementary Table 1, Supplementary Fig. 2A and 2B) as well as at least one imputed SNP with $P < 5 \times 10^{-7}$ (Fig. 2A and 2B, Supplementary Fig. 3A and 3B).

For replication, we selected nine novel genomic regions and evaluated them *in silico* from array data in two additional case-control studies from Europe¹ and the United States¹⁰ (Table 1). In each of the nine regions, we selected the most strongly associated SNPs present on both the Omni1-Quad and HumanHap610 arrays (since such SNPs were directly genotyped in all our samples, as well as in both sets of replication samples). We further limited follow-up region choice to those with at least two SNPs with $P < 10^{-4}$ (i.e. there must be a supporting SNP in addition to the primary SNP). Both chromosome 1 regions show significant associations in the replication samples whilst the other seven regions did not (Table 2, Supplementary Fig. 4 and 5). We sought further replication of the two chromosome 1 regions in an additional set of cases and controls from the United States (Table 1, Supplementary Table 2); rs7412746 clearly replicated (OR=0.86, $P=0.0076$, one-sided; meta-analysis of all three replication cohorts OR=0.89, $P=1.5 \times 10^{-5}$), with rs3219090 showing a trend toward significance in the expected direction (OR=0.95, $P=0.20$,

one-sided; meta-analysis of all three replication cohorts $OR=0.91$, $P = 3.4 \times 10^{-3}$). Based on the ORs seen in the replication cohorts, rs7412746 and rs3219090 each explain 0.1% of the genetic variance in melanoma risk. The meta-analysis P -values for all case-control studies combined were $P=9.0 \times 10^{-11}$ (genome-wide significant) for rs7412746 and $P=9.3 \times 10^{-8}$ (suggestively significant) for rs3219090.

We tested for association of rs7412746 and rs3219090 with pigmentation and nevus phenotypes, available on a subset of our discovery sample (up to 1,146 cases and 1,080 controls, Supplementary Note). SNP rs7412746 showed nominally significant association with blue versus non-blue eye colour ($P=0.02$), fair versus non-fair hair colour ($P=0.01$) and dark brown versus non-dark brown hair colour ($P=0.02$), as well as borderline association with nevus count ($P=0.06$). The direction of effect of rs7412746 on blue eye colour, fair hair colour, dark hair colour and nevus count was the same in the case and control subsets of our discovery sample. No association was seen between rs7412746 and skin colour or freckling. SNP rs3219090 was not associated with any pigmentation or nevus traits. Adjusting for pigmentation or nevus traits did not appreciably change the association of either locus with melanoma (rs7412746 OR before correction 0.82, after correction 0.84, $P=0.33$ for difference; rs3219090 OR before correction 0.82, after correction 0.83, $P=0.61$ for difference).

We also tested for differences in the strength of the associations of rs7412746 and rs3219090 with melanoma in early versus late onset (≤ 40 compared with >40 years at age of onset) and *in situ* versus invasive (79% of cases were invasive) subsets of the Australian cases. We found no differences in the association OR for these subsets. For early onset versus controls rs7412746 yielded $OR=0.83$, 95% CI 0.75,0.91 ($P=0.79$ for difference in frequency between early and late) and rs3219090 yielded $OR=0.81$, 95% CI 0.73,0.90 ($P=0.63$ for difference in frequency between early and late). For invasive versus controls rs7412746 yielded $OR=0.80$, 95% CI 0.73,0.88 ($P=0.60$ for difference in frequency between invasive and *in situ*) and rs3219090 yielded $OR=0.84$, 95% CI 0.76,0.93 ($P=0.38$ for difference in frequency between invasive and *in situ*).

The ratio of males to females was similar in cases and two of the control samples but the third control sample was all-female (samples from an endometriosis study⁸). We repeated our analysis without the all-female sample set and the results were similar; rs7412746 $OR=0.82$ in full data set, $OR=0.84$ with the all-female control set removed ($P=0.42$ for difference in frequency between endometriosis control set and remaining controls); rs3219090 $OR=0.82$ in full data set, $OR=0.82$ with the all-female control set removed ($P=0.96$ for difference in frequency between endometriosis control set and remaining controls). In the full Australian sample, there were no differences in the strength of association in only male or only female cases and controls; rs7412746 $OR=0.82$ and rs3219090 $OR=0.81$ in male only samples; rs7412746 $OR=0.84$ and rs3219090 $OR=0.81$ in female only samples ($P=0.83$, $P=0.90$ for OR difference between sexes for rs7412746 and rs3219090, respectively).

The associated region at 149 Mb on chromosome 1 spans approximately 450 Kb and harbours ten genes. The peak imputed SNP at this locus in the Australian case control

sample, rs267735 ($P=5.5 \times 10^{-8}$) maps 1 Kb upstream of the transcription start site (TSS) of *LASS2* (genome build 36 position 149,215,120), although there is substantial linkage disequilibrium (LD) that spans several genes in the region. All but one (*ANXA9*) of these genes are expressed in normal cultured human melanocytes, and most are also expressed across the vast majority of melanoma cell lines examined¹¹. Several of the genes in the region have been implicated in cancer or cancer-related processes, including *MCL1* (anti-apoptotic protein), *ARNT* (hypoxia-inducible factor 1 beta), and *LASS2* (ceramide synthase 2). The SNP rs7412746 significantly influences the expression (i.e. is an expression quantitative trait loci or eQTL) of several genes in the region including *CTSK* (Chicago eQTL browser). Perhaps the strongest candidate in the region is *SETDB1*; a recent study in zebrafish has shown a role for variation in this gene in melanoma development¹². Further study will be required to determine which gene or genes at this locus mediate melanoma risk.

In contrast to the 149 Mb region, the associated region at 224 Mb spans only 70 Kb and encompasses a single gene in its entirety (45 Kb), poly (ADP-ribose) polymerase 1 (*PARP1*). The peak imputed SNP is rs2695238 ($P=3.8 \times 10^{-7}$ in the Australian case-control sample, genome build 36 position 224,671,142) and lies ~9 Kb upstream of the TSS of *PARP1* with several highly correlated SNPs lying within the gene. *PARP1* encodes a chromatin-associated enzyme which modifies various nuclear proteins by poly-ADP-ribosylation. PARP1 plays a key role in multiple cellular processes such as differentiation, proliferation, and tumor transformation and plays a key role in the repair of single-strand DNA breaks. Interestingly, a recent candidate gene study¹³ reported a nominally significant association between the intronic *PARP1* SNP rs3219125 and melanoma in a set of 585 melanoma cases and 585 controls (OR 1.89, 95% CI 1.34–2.68), with stronger effect in patients with melanoma of the head and neck. SNPs rs3219090 and rs3219125 are not highly correlated in 1000 Genomes CEU samples ($r^2=0.042$). SNP rs3219125 was not genotyped in our Australian discovery cohort but was well imputed (imputation $r^2=0.70$) and showed marginal evidence for association ($P=0.053$). While no strongly-associated imputed or genotyped SNPs within the *PARP1* locus alter the protein-coding sequence of the gene, two SNPs directly adjacent to each other and located within a nuclear factor 1 (NF1) transcription factor binding site were strongly associated (rs3754376: imputed $P=7.39 \times 10^{-7}$, OR=1.22; rs3754375: imputed $P=3.0 \times 10^{-3}$, OR=1.16). Both SNPs are in complete LD with each other and rs2695238 (pairwise $D'=1$ for all 3 pairs, pairwise r^2 in the range 0.39 to 0.83). Further study will be required to assess whether these or other variants within this region directly mediate melanoma risk.

In our Australian discovery cohort, there remains an excess of positive results in the Q-Q plot after the removal of SNPs located within previously identified melanoma susceptibility regions (Supplementary Fig. 1). A small proportion of this excess was explained by the two novel chromosome 1 regions described here. Work examining the distribution of effect sizes obtained from GWAS suggests that for a wide range of traits, there are many more loci that will be found by conducting GWAS on larger samples¹⁴. Our data are consistent with there being further common SNPs influencing melanoma risk and we expect that further studies of additional melanoma samples will allow us to identify and characterize further loci.

In summary, our GWAS of melanoma identified two novel melanoma risk loci on chromosome 1 and replicated findings from previous melanoma GWASs. The observed effect size for the two novel loci was smaller than that observed for previously reported loci. Neither appears to be strongly correlated with human pigmentation or measures of nevus density, suggesting they may influence melanoma risk through distinct mechanisms. Identification of the causal variants at these loci will help refine estimates of risk for this increasingly common cancer.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Stuart MacGregor¹, Grant W. Montgomery¹, Jimmy Z. Liu¹, Zhen Zhen Zhao¹, Anjali K. Henders¹, Mitchell Stark¹, Helen Schmid², Elizabeth A Holland², David L. Duffy¹, Mingfeng Zhang³, Jodie N. Painter¹, Dale R Nyholt¹, Judith A. Maskiell⁴, Jodie Jetann⁵, Megan Ferguson⁵, Anne E. Cust^{4,21}, Mark A. Jenkins⁴, David C. Whiteman¹, Håkan Olsson⁶, Susana Puig^{7,31}, Giovanna Bianchi-Scarrà⁸, Johan Hansson⁹, Florence Demenais^{10,30}, Maria Teresa Landi¹¹, Tadeusz Dębniak²⁵, Rona Mackie²⁶, Esther Azizi²⁷, Brigitte Bressac - de Paillerets^{10,28}, Alisa M. Goldstein¹¹, Peter A. Kanetsky^{12,32}, Nelleke A. Gruis¹³, David E. Elder¹⁴, Julia A. Newton Bishop¹⁵, D. Timothy Bishop¹⁵, Mark M. Iles¹⁵, Per Helsing¹⁶, Christopher I. Amos¹⁷, Qingyi Wei¹⁷, Li-E Wang¹⁷, Jeffrey E. Lee¹⁸, Abrar A. Qureshi^{3,19}, Richard F. Kefford², Graham G. Giles²⁰, Bruce K. Armstrong²¹, Joanne F. Aitken⁵, Jiali Han^{3,19,22}, John L. Hopper⁴, Jeffrey M. Trent^{23,29}, Kevin M. Brown²⁴, Nicholas G. Martin¹, Graham J. Mann², and Nicholas K. Hayward¹

Affiliations

¹Queensland Institute of Medical Research, Brisbane, QLD 4029, Australia
²Westmead Institute of Cancer Research, University of Sydney at Westmead Millennium Institute and Melanoma Institute Australia, Westmead, NSW 2145 Australia
³Department of Dermatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115
⁴Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, School of Population Health, University of Melbourne, VIC 3052 Australia
⁵Viertel Centre for Research in Cancer Control, Cancer Council Queensland, Brisbane, QLD 4004 Australia
⁶Lund Cancer Center Department of Oncology, University Hospital, Lund, Sweden, on behalf of Lund Melanoma Group
⁷Dermatology Department, Melanoma Unit, Hospital Clínic, Institut de Investigació Biomèdica August Pi Suñe, Universitat de Barcelona, Barcelona, Spain
⁸Department of Internal Medicine, University of Genoa, Italy
⁹Department of Oncology-Pathology, Karolinska Institute and Karolinska University Hospital, Solna, Stockholm, Sweden
¹⁰INSERM, U946, Fondation Jean-Dausset-CEPH, F-75010, Paris, France
¹¹Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
¹²Centre for Clinical Epidemiology & Biostatistics, 219 Blockley Hall, University

of Pennsylvania, Philadelphia, PA, USA ¹³Department of Dermatology, Leiden University Medical Center, Leiden, The Netherlands ¹⁴Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA ¹⁵Section of Epidemiology and Biostatistics, Leeds Institute of Molecular Medicine, Cancer Research UK Clinical Centre at Leeds, St James's University Hospital, Leeds, UK ¹⁶Department of Dermatology, Oslo University Hospital, Rikshospitalet, N-0027 Oslo, Norway ¹⁷Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA ¹⁸Department of Surgical Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA ¹⁹Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115 ²⁰Cancer Epidemiology Centre, The Cancer Council Victoria, Carlton Victoria 3053 Australia ²¹Cancer Epidemiology and Services Research, Sydney School of Public Health, University of Sydney, NSW 2006 Australia ²²Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA ²³Genetic Basis of Human Disease Division, The Translational Genomics Research Institute, Phoenix, AZ, USA ²⁴Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA ²⁵International Hereditary Cancer Center, ul. Polabska 4, Szczecin, Poland ²⁶University of Glasgow, Department of Public Health, Glasgow, UK ²⁷Department of Dermatology / Dermato-Oncology Unit, Sheba Medical Center, Tel Hashomer, 52621, Israel ²⁸Service de Génétique, Institut de Cancérologie Gustave Roussy, Villejuif, France ²⁹Center for Cancer Genomics and Computational Biology Van Andel Research Institute, Grand Rapids, MI 49503, USA ³⁰Université Paris Diderot, Sorbonne Paris Cité, Institut Universitaire d'Hématologie, F-75010, Paris, France ³¹Centre of Biomedical Research on Rare Diseases (CIBERER), ISCIII, Barcelona, Spain ³²Department of Biostatistics & Epidemiology, 219 Blockley Hall, University of Pennsylvania, Philadelphia, PA, USA

ACKNOWLEDGEMENTS

This work was supported by the Melanoma Research Alliance, the National Institutes of Health/National Cancer Institute (CA88363, CA83115, CA122838, CA87969, CA055075, CA100264, CA133996 and CA49449), the National Health and Medical Research Council of Australia (NHMRC) (107359, 200071, 241944, 339462, 380385, 389927, 389875, 389891, 389892, 389938, 402761, 443036, 442915, 442981, 496610, 496675, 496739, 552485, 552498), the Cancer Councils NSW, Victoria and Queensland, the Cancer Institute New South Wales, the Cooperative Research Centre for Discovery of Genes for Common Human Diseases (CRC), Cerylid Biosciences (Melbourne), the Australian Cancer Research Foundation, The Wellcome Trust and donations from Neville and Shirley Hawkins. The endometriosis sample genotyping was funded by a grant from the Wellcome Trust (WT084766/Z/08/Z). The Australian Twin Registry is supported by an Australian National Health and Medical Research Council (NHMRC) Enabling Grant (2004-2009). DLD, NKH and GWM are supported by the NHMRC Fellowships scheme and JLH is an Australia Fellow of the NHMRC. SM is the recipient of a Career Development Award from the NHMRC (496674, 613705). DCW is a Future Fellow of the Australian Research Council. KMB is supported by the National Institutes of Health/National Cancer Institute. JMT is supported by the National Institutes of Health/National Cancer Institute (CA109544). BKA is supported by the University of Sydney Medical Foundation. AEC is the recipient of a NHMRC public health postdoctoral fellowship (520018) and a Cancer Institute NSW Early Career Development Fellowship (10/ECF/2-06). AMG and MTL were supported by the Intramural Research Program of the NIH, NCI and DCEG. The AMFS and Q-MEGA gratefully acknowledges all its participants, the hard work of all its research interviewers and examiners and of Chantelle Agha-Hamilton in managing the AMFS biospecimens. Q-MEGA thanks Amanda Baxter, Monica de Nooyer, Isabel Gardner, Dixie Statham, Barbara Haddon, Margaret J. Wright, Jane Palmer, Judith Symmons, Belinda Castellano, Lisa Bardsley, Sara Smith, David Smyth, Leanne Wallace, Megan J. Campbell, Anthony Caracella, Marina Kvaskoff, Barbara

Haddon, Olivia Zheng, Brett Chapman and Harry Beeby for their input into project management, sample processing and database development. We are grateful to the many research assistants and interviewers for assistance with the studies contributing to the QIMR and AMFS collections.

We acknowledge with appreciation all the participants in the QIMR and endometriosis studies. We thank Endometriosis Associations for supporting study recruitment. We thank Sullivan Nicolaides and Queensland Medical Laboratory for pro bono collection and delivery of blood samples and other pathology services for assistance with blood collection.

We thank the following U.S. state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY.

The GenoMEL replication sample was funded by the European Commission under the 6th Framework Programme, contract no: LSHC-CT-2006-018702, by Cancer Research UK Programme Award (C588/A4994) and by US National Institutes of Health R01 CA83115. Research at M.D. Anderson Cancer Center was partially supported by NIH grants R01CA100264, 2P50CA093459-, P30CA016672, and R01CA133996. The Center for Inherited Disease Research performed genotyping for MDACC and is supported by contract HHSN268200782096C. GENEVA performed data cleaning and is supported by NIH grant HG004446. The Barcelona team research has also been supported by Fondo de Investigaciones Sanitarias grant 09/1393.

APPENDIX

METHODS

Statistical Analysis

Genotyping and data quality control

Australian Discovery Sample: DNA was extracted from peripheral blood or saliva samples. Australian twin and endometriosis sample controls were genotyped at deCODE Genetics (Reykjavik, Iceland) on the Illumina HumanHap610W Quad and Illumina HumanHap670 Quad Beadarrays, respectively. AMFS controls were genotyped by Illumina (San Diego) on Illumina Omni1-Quad arrays. Cases were genotyped by Illumina (San Diego) on Illumina Omni1-Quad (568 AMFS cases, 699 Q-MEGA cases) and HumanHap610W Quad arrays (998 Q-MEGA cases). All genotypes were called with the Illumina BeadStudio software. SNPs with a mean BeadStudio GenCall score < 0.7 were excluded from the control data sets. All samples had successful genotypes for $> 95\%$ of SNPs. SNPs with call rates either < 0.95 (minor allele frequency, $MAF > 0.05$) or < 0.99 ($MAF > 0.01$), Hardy-Weinberg equilibrium in controls $P < 10^{-6}$, and/or $MAF < 0.01$ were excluded. Cryptic relatedness between individuals was assessed through the production of a full identity-by-state matrix. Ancestry outliers were identified by principal component (PC) analysis, using data from 11 populations of the HapMap 3 Project and five Northern European populations genotyped by the GenomeEUtwin consortium, using the EIGENSOFT package¹⁵. Individuals lying 2 standard deviations from the mean PC1 and PC2 scores were excluded from subsequent analyses. Following these exclusions, there were 2,168 case samples (1,242 typed on the Omni1-Quad and 926 typed on the Hap610 arrays, respectively) and 4,387 controls (431 typed on the Omni1-Quad and 3,956 typed on the Hap610 arrays) retained for subsequent analyses (Table 1). Individuals typed on the Omni1-Quad array had genotypes for up to 816,169 SNPs, while individuals typed on the Hap610/670 arrays had genotypes for up to 544,483 SNPs. There were 299,394 SNPs passing QC and overlapping between these arrays (and hence directly genotyped on all Australian samples).

To investigate potential effects of population stratification in the cleaned data set, we used 160,000 randomly selected SNPs (culled from the set of 299,394 directly-genotyped SNPs) to generate first 10 principal components (as well as the first two, with similar results, data not shown) for the case and control samples combined using EIGENSOFT.

Genotyping and data quality control details for the replication samples are given in the Online Supplementary Note.

Genomic imputation—Imputation for the Australian samples was performed using MACH2¹⁶ with 1000 Genomes Project (June 2010 release)⁹ data obtained from people of northern and western European ancestry collected by the Centre d'Etude du Polymorphisme Humain. Imputation was based on a set of autosomal SNPs common to all melanoma case-control samples (n=292,043). Imputation was run in two stages. First, data from a set of representative Australian sample individuals was compared to the phased haplotype data from the 1000 Genomes data to generate recombination and error maps. For the second stage, data were imputed for all individuals using the phased 1000 Genomes data as the reference panel and the recombination as well as error files generated in stage 1. In total 5,480,804 1000 Genomes SNPs could be imputed with imputation $r^2 > 0.5$.

Association analysis

Australian Discovery Sample: Association analysis of genotyped SNPs was performed using the PLINK --assoc option¹⁷.

Analysis of dosage scores from the imputation analysis was done using mach2dat¹⁶. Analysis was done both with and without the first 10 principal components included as covariates (mach2dat for imputed SNPs, PLINK --logistic option for genotyped SNPs). Results are presented in the main text without including principle components as covariates. Adjusting for principal components did not change any of the *P*-values by more than a factor of 10 (Supplementary Figures 2A, 2B). Meta-analysis of discovery and replication cohorts was done using PLINK (--meta-analysis option), with ORs weighted by the inverse of their variance (fixed effects model). Tests for heterogeneity of ORs between studies was tested using Cochran's Q statistic (neither rs3219090 nor rs7412746 showed evidence for heterogeneity of ORs between studies). The proportion of variance explained by rs3219090 and rs7412746 was derived assuming a population prevalence of 0.05¹⁸ and sibling relative risk of 3¹⁹. Given the small ORs observed and assuming a large number of similar small effects, the proportion of genetic variance explained was computed as the ratio of the log of the locus specific population relative risk (PRR) in siblings to the log of the overall relative risk in siblings²⁰. PRR was estimated from the ORs and allele frequencies using output from the GRR function in Sibpair. Tests for heterogeneity with the early/late AAO and in situ/invasive subsets of the cases were done by computing the test of association in one subgroup against the other (PLINK --assoc option). For pigmentation traits, each subgroup was compared with all remaining subgroups using the PLINK --assoc option. Association analysis of nevus count was done using linear regression, with permutation used to compute empirical p-values. Association analysis of melanoma corrected for pigmentation and nevus count was done using logistic regression with factors for each level for the pigmentation and

nevus count variables (pigmentation and nevus factors included simultaneously). Tests for differences in ORs with and without covariates were conducted by generating 1000 bootstrap replicates from the data: the actual data difference in OR with and without covariates was compared with bootstrap replicates to compute empirical P values. Association plots were created using LocusZoom²¹.

Replication Sample 1 - GenoMEL: A trend test was applied to each SNP in turn stratified by broad geographic region (8 regions pre-specified).

Replication Sample 2 - MD Anderson Cancer Center: Association analysis of genotyped SNPs was done using the PLINK --logistic option¹³. The first 2 principal components were included to adjust for population structure.

Replication Sample 3 - Harvard: Association analysis of genotyped SNPs was done using SAS V9.1 (SAS Institute, Cary, NC). Unconditional logistic regression was employed to calculate odds ratios (ORs) and 95% confidence intervals (CIs) adjusting for age and gender.

REFERENCES

1. Bishop DT, et al. Genome-wide association study identifies three loci associated with melanoma risk. *Nat Genet.* 2009; 41:920–5. [PubMed: 19578364]
2. Brown KM, et al. Common sequence variants on 20q11.22 confer melanoma susceptibility. *Nat Genet.* 2008; 40:838–40. [PubMed: 18488026]
3. Gudbjartsson DF, et al. ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. *Nat Genet.* 2008; 40:886–91. [PubMed: 18488027]
4. Falchi M, et al. Genome-wide association study identifies variants at 9p21 and 22q13 associated with development of cutaneous nevi. *Nat Genet.* 2009; 41:915–9. [PubMed: 19578365]
5. Duffy DL, et al. IRF4 variants have age-specific effects on nevus count and predispose to melanoma. *Am J Hum Genet.* 2010; 87:6–16. [PubMed: 20602913]
6. Baxter AJ, et al. The Queensland Study of Melanoma: environmental and genetic associations (Q-MEGA); study design, baseline characteristics, and repeatability of phenotype and sun exposure measures. *Twin Res Hum Genet.* 2008; 11:183–96. [PubMed: 18361720]
7. Cust AE, et al. Population-based, case-control-family design to investigate genetic and environmental influences on melanoma risk: Australian Melanoma Family Study. *Am J Epidemiol.* 2009; 170:1541–54. [PubMed: 19887461]
8. Painter JN, et al. Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. *Nat Genet.* 2010
9. Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature.* 2010; 467:1061–73. [PubMed: 20981092]
10. Li C, et al. Genetic variants of the ADPRT, XRCC1 and APE1 genes and risk of cutaneous melanoma. *Carcinogenesis.* 2006; 27:1894–901. [PubMed: 16621887]
11. Johansson P, Pavey S, Hayward N. Confirmation of a BRAF mutation-associated gene expression signature in melanoma. *Pigment Cell Res.* 2007; 20:216–21. [PubMed: 17516929]
12. Ceol CJ, et al. The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. *Nature.* 2011; 471:513–7. [PubMed: 21430779]
13. Zhang M, Qureshi AA, Guo Q, Han J. Genetic variation in DNA repair pathway genes and melanoma risk. *DNA Repair (Amst).*
14. Park JH, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet.* 2010; 42:570–5. [PubMed: 20562874]
15. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics.* 2006; 38:904–909. [PubMed: 16862161]

16. Li Y, Abecasis G. Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. *American Journal of Human Genetics*. 2006; *S79*:2290.
17. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; *81*:559–75. [PubMed: 17701901]
18. Queensland Cancer Registry. *Cancer in Queensland, Incidence and Mortality 1982 to 2005*. The Cancer Council Queensland; 2008.
19. Do KA, Aitken JF, Green AC, Martin NG. Analysis of melanoma onset: assessing familial aggregation by using estimating equations and fitting variance components via Bayesian random effects models. *Twin Res*. 2004; *7*:98–113. [PubMed: 15053858]
20. James JW. Frequency in relatives for an all-or-none trait. *Ann Hum Genet*. 1971; *35*:47–9. [PubMed: 5106369]
21. Pruim RJ, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010; *26*:2336–7. [PubMed: 20634204]

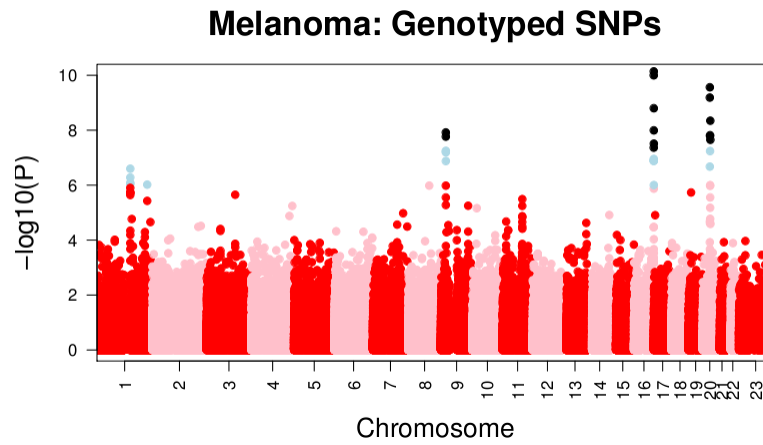


Figure 1.

Association results for SNPs directly genotyped in all Australian samples. SNPs with P -values exceeding genome-wide significance ($P < 5 \times 10^{-8}$) are shown in black, while SNPs with $5 \times 10^{-8} < P < 1 \times 10^{-6}$ are shown in blue. The y-axis is truncated at 1×10^{-9} , however, some SNPs from previously identified loci exceed this threshold (specifically at ~88 Mb on chromosome 16 near *MC1R* and at the *ASIP* locus 33 Mb on chromosome 20). The genome-wide significant signal on chromosome 9 is in the vicinity of the *MTAP/CDKN2A* region.

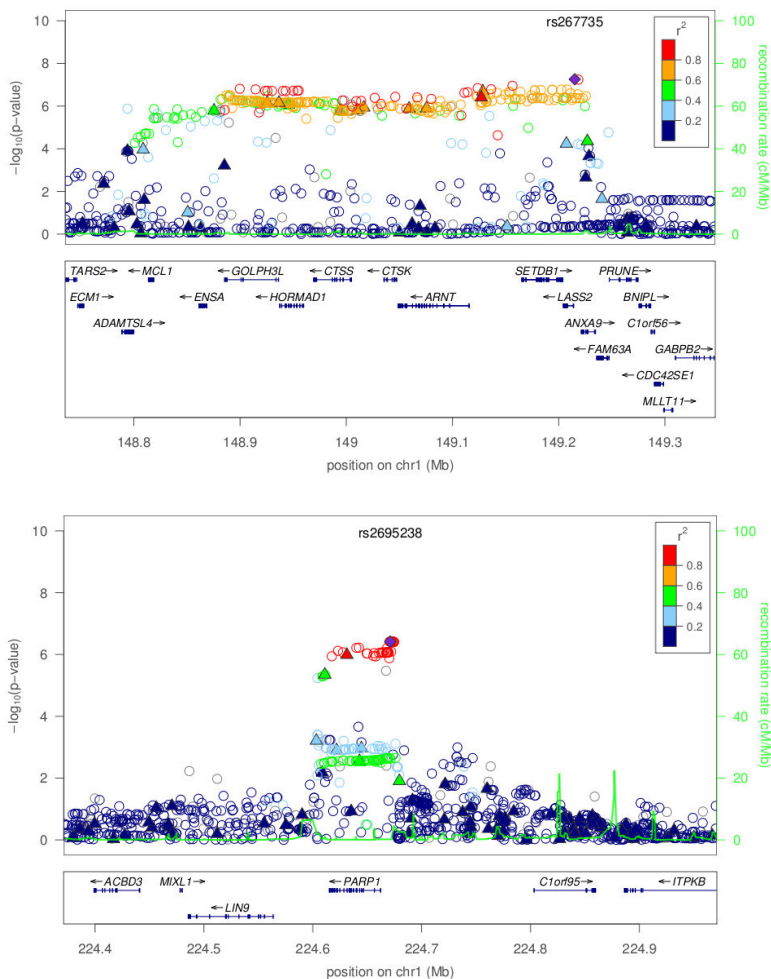


Figure 2. Discovery sample association results at two novel melanoma susceptibility loci on chromosome 1 for both SNPs directly genotyped in all Australian samples and imputed SNPs. Genotyped SNPs are indicated by solid triangles and imputed SNPs by hollow circles. The top ranked SNP at each locus is shown as a solid purple diamond (this SNP is an imputed SNP at both loci). Imputation p-values for all SNPs are plotted. Note imputed and genotyped p-values for genotyped SNPs differ slightly because for the imputed result, analysis is based on dosage scores whereas with genotyped SNPs hard genotype calls are used. Association results shown are for (A) the chromosome 1 locus near 149 Mb, and (B) SNPs in the vicinity of the *PARP1* association signal. The color scheme indicates linkage disequilibrium between the most strongly-associated SNPs for the 149 Mb and *PARP1* region (shown in purple, rs267735 and rs2695238, respectively) and other genotyped SNPs in the two regions.

Table 1
Study Samples

| Sample | Array | Cases | Controls |
|--|------------|-------|----------|
| Discovery sample: Australia | Omni1-Quad | 1,242 | 431 |
| | 610/670 | 926 | 3,956 |
| Replication Sample 1: Europe -GenoMEL | 610 | 2,804 | 7,618 |
| Replication Sample 2: United States 1 - MD Anderson Cancer Center | Omni1-Quad | 1,804 | 1,026 |
| Replication Sample 3: United States 2 - Harvard | NA* | 585 | 6,500 |

* United States 2 samples were only typed for rs7412746 and rs3219090 using the OpenArray™ SNP Genotyping System

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Results for nine loci selected from the discovery sample

| SNP | Chr | Coordinate | Australia | | United States 1 | | Europe | | Combined Replication samples* | | Discovery plus replication | |
|------------|-----|-------------|-----------|----------------------|-----------------|----------------------|--------|-------|-------------------------------|----------------------|----------------------------|-----------------------|
| | | | OR | P | OR | P | OR | P | OR | P | OR | P |
| rs7412746 | 1 | 149,127,095 | 0.82 | 2.5×10^{-7} | 0.85 | 2.7×10^{-3} | 0.92 | 0.014 | 0.90 | 2.6×10^{-5} | 0.87 | 9.0×10^{-11} |
| rs3219090 | 1 | 224,631,314 | 0.82 | 9.5×10^{-7} | 0.88 | 0.028 | 0.91 | 0.048 | 0.90 | 3.5×10^{-3} | 0.87 | 9.3×10^{-8} |
| rs10170188 | 2 | 205,757,059 | 1.19 | 3.3×10^{-5} | 0.99 | 0.86 | 0.99 | 0.74 | 0.99 | 0.70 | | |
| rs17065828 | 3 | 62,017,865 | 0.83 | 3.9×10^{-5} | 1.00 | 1.00 | 1.00 | 0.95 | 0.99 | 0.96 | | |
| rs13177645 | 5 | 115,031,773 | 0.83 | 2.1×10^{-5} | 1.01 | 0.80 | 0.96 | 0.50 | 0.99 | 0.76 | | |
| rs7811987 | 7 | 136,176,803 | 1.19 | 1.1×10^{-5} | 1.01 | 0.82 | 1.00 | 0.95 | 1.00 | 0.87 | | |
| rs6478444 | 9 | 121,721,401 | 1.19 | 5.6×10^{-6} | 0.90 | 0.067 | 1.01 | 0.92 | 0.96 | 0.41 | | |
| rs10766295 | 11 | 16,061,966 | 1.17 | 2.1×10^{-5} | 1.02 | 0.70 | 1.03 | 0.32 | 1.03 | 0.30 | | |
| rs1584186 | 11 | 25,137,541 | 1.21 | 4.4×10^{-5} | 1.04 | 0.54 | 0.96 | 0.37 | 0.98 | 0.65 | | |

* Results for Europe plus United States 1 samples only. The results for the two chromosome 1 SNPs in all three replication samples (Europe, United States 1, United States 2) are given in the text