nature genetics

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

Stuart MacGregor[1], Grant W Montgomery[1], Jimmy Z Liu[1], Zhen Zhen Zhao[1], Anjali K Henders[1], Mitchell Stark[1], Helen Schmid[2], Elizabeth A Holland[2], David L Duffy[1], Mingfeng Zhang[3], Jodie N Painter[1], Dale R Nyholt[1], Judith A Maskiell[4], Jodie Jetann[5], Megan Ferguson[5], Anne E Cust[4,6], Mark A Jenkins[4], David C Whiteman[1], Håkan Olsson[7,33], Susana Puig[8,9], Giovanna Bianchi-Scarrà[10], Johan Hansson[11], Florence Demenais[12,13], Maria Teresa Landi[14], Tadeusz Dębniak[15], Rona Mackie[16], Esther Azizi[17], Brigitte Bressac-de Paillerets[12,18], Alisa M Goldstein[14], Peter A Kanetsky[19,20], Nelleke A Gruis[21], David E Elder[22], Julia A Newton-Bishop[23], D Timothy Bishop[23], Mark M Iles[23], Per Helsing[24], Christopher I Amos[25], Qingyi Wei[25], Li-E Wang[25], Jeffrey E Lee[26], Abrar A Qureshi[3,27], Richard F Kefford[2], Graham G Giles[28], Bruce K Armstrong[6], Joanne F Aitken[5], Jiali Han[3,27,29], John L Hopper[4], Jeffrey M Trent[30,31], Kevin M Brown[32], Nicholas G Martin[1], Graham J Mann[2] & Nicholas K Hayward[1]

© 2011 Nature America, Inc. All rights reserved.

We performed a genome-wide association study of melanoma in a discovery cohort of 2,168 Australian individuals with melanoma and 4,387 control individuals. In this discovery phase, we confirm several previously characterized melanoma-associated loci at *MC1R*, *ASIP* and *MTAP–CDKN2A*. We selected variants at nine loci for replication in three independent case-control studies (Europe: 2,804 subjects with melanoma, 7,618 control subjects; United States 1: 1,804 subjects with melanoma, 1,026 control subjects; United States 2: 585 subjects with melanoma, 6,500 control subjects). The combined meta-analysis of all case-control studies identified a new susceptibility locus at 1q21.3 (rs7412746, $P = 9.0 \times 10^{-11}$, OR in combined replication cohorts of 0.89 (95% CI 0.85–0.95)). We also show evidence suggesting that melanoma associates with 1q42.12 (rs3219090, $P = 9.3 \times 10^{-8}$). The associated variants at the 1q21.3 locus span a region with ten genes, and plausible candidate genes for melanoma susceptibility include *ARNT* and *SETDB1*. Variants at the 1q21.3 locus do not seem to be associated with human pigmentation or measures of nevus density.

To date, genome-wide association studies (GWAS) for melanoma[1,2], pigmentation[3] and nevogenesis[4,5] have identified a small number of low-penetrance melanoma susceptibility variants. These variants seem to affect melanoma risk through their involvement in the known melanoma-associated risk phenotypes of pigmentation and nevus count. In contrast to variants identified in other cancers, these variants have been shown to have relatively large effects on disease risk (odds ratios (OR) > 1.5); previous melanoma GWAS were 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 < OR <1.5).

Melanoma cases in individuals 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 Study[7] (AMFS). Samples from Australian individuals of European descent from three different sources were used as controls ($N = 4,387$)[6–8]. Samples were genotyped on Illumina single-nucleotide polymorphism (SNP) arrays (samples from subjects with melanoma, Omni1-Quad or HumanHap610; control samples, Omni1-Quad, HumanHap610 or HumanHap670; **Table 1**). Samples from subjects with melanoma and control subjects were combined into a single data set for quality control (including principal component analysis for outlier removal) and imputation (**Supplementary Note**). Imputation based on data from the 1000 Genomes Project[9] 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 that were directly genotyped in all individuals in these discovery samples was 1.04 (**Supplementary Fig. 1**).

The results of the association tests 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* and *MTAP–CDKN2A*)[1–4] reached genome-wide significance. Two additional noteworthy regions were identified 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** and **Supplementary Fig. 2a**,**b**), as well as at least one imputed SNP with $P < 5 \times 10^{-7}$ (**Fig. 2a**,**b** and **Supplementary Fig. 3a**,**b**).

A full list of affiliations appears at the end of the paper.

Received 18 February; accepted 8 September; published online 9 October 2011; doi:10.1038/ng.958

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–M.D. Anderson Cancer Center	Omni1-Quad	1,804	1,026
Replication sample 3: United States 2–Harvard	N.A.[a]	585	6,500

[a]United States 2 samples were typed only for rs7412746 and rs3219090 using the OpenArray SNP Genotyping System.

For replication analyses, we selected nine of the new loci and evaluated them *in silico* using array data in two additional case-control studies from Europe[1] and the United States[10] (**Table 1**). In each of the nine regions, we selected the most strongly associated SNPs present on both the Omni1-Quad and HumanHap610 arrays (as 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 regions with at least two SNPs with $P < 10^{-4}$ (that is, there had to be a supporting SNP in addition to the primary SNP). Both chromosome 1 regions showed significant association with melanoma in the replication samples, whereas the other seven regions did not (**Table 2** and **Supplementary Figs. 4** and **5**). We looked for further replication of the two chromosome 1 regions by examining an additional set of subjects with melanoma and control subjects from the United States (**Table 1** and **Supplementary Table 2**). In this analysis, rs7412746 clearly replicated (OR = 0.86, $P = 0.0076$, one-sided; meta-analysis of all three replication cohorts: OR = 0.89, 95% confidence interval (CI) 0.85–0.95, $P = 1.5 \times 10^{-5}$), and rs3219090 showed 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, 95% CI 0.85–0.97, $P = 3.4 \times 10^{-3}$). Based on the ORs seen in the replication cohorts, rs7412746 and rs3219090 each explain 0.1% of the genetic variance associated with melanoma risk. The meta-analysis *P* values for all the case-control studies combined were: $P = 9.0 \times 10^{-11}$ for rs7412746 and $P = 9.3 \times 10^{-8}$ (not meeting our genome-wide significance threshold of 5.0×10^{-8}) for rs3219090.

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

We also tested for differences in the strength of the associations of the rs7412746 and rs3219090 SNPs with melanomas of early versus late onset (≤40 years of age compared with >40 years of age at onset) and with an *in situ* versus invasive phenotype (79% of cases were invasive) in the Australian cohort. We found no difference in the association ORs of 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 the groups of individuals with melanoma and in two of the control subject groups, but the third control subject group was all female (samples taken from an endometriosis study[8]). We repeated our analysis without the all-female sample set, and the results were similar: for 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); for 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 cohort, there were no differences in the strength of association with melanoma between groups of subjects with melanoma composed of only males or only females and control individuals: 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 encodes ten genes. The peak imputed SNP at this locus in samples from Australian case-control sample, rs267735 ($P = 5.5 \times 10^{-8}$), maps to 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 in the vast majority of the melanoma cell lines examined[11]. Several of the genes in the region have been implicated in cancer or cancer-related processes, including *MCL1* (encoding anti-apoptotic protein), *ARNT* (hypoxia-inducible factor 1-β) and *LASS2* (ceramide sythase 2). The rs7412746 SNP significantly influences the expression (is an expression quantitative trait locus or eQTL) of several genes in the region (best $P = 6.2 \times 10^{-7}$ for *CTSK* using the eQTL browser (http://eqtl.uchicago.edu/). Perhaps the strongest candidate gene in the region is *SETDB1*; a recent study in zebrafish has shown a role for variation

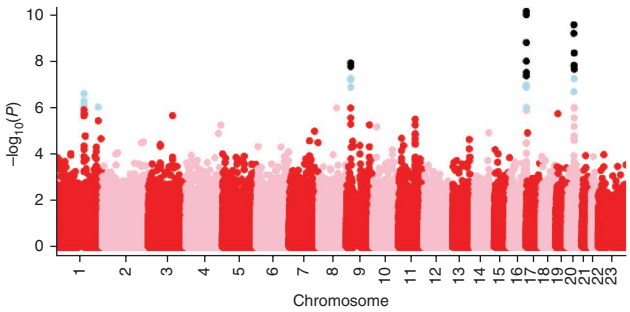

Figure 1 Association results for SNPs directly genotyped in all Australian samples. SNPs with *P* values exceeding genome-wide significance ($P < 5 \times 10^{-8}$) are shown in black, and SNPs with $5 \times 10^{-8} < P < 1 \times 10^{-6}$ are shown in blue. The *y* axis is truncated at 1×10^{-9}; however,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

© 2011 Nature America, Inc. All rights reserved.

Table 2 Results for nine loci selected from the discovery sample

			Australia		United States 1		Europe		Combined replication samples[a]		Discovery plus replication	
SNP	Chr	Coordinate	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		

OR, Odds ratio; Chr., Chromosome.
[a]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 and United States 2) are given in the text.

in this gene in melanoma development[12]. 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 *PARP1*. *PARP1* encodes a chromatin-associated enzyme that modifies various nuclear proteins by poly(ADP)-ribosylation. The PARP1 protein has a key role in multiple cellular processes, such as differentiation, proliferation and tumor transformation, and has important functions in the repair of single-strand DNA breaks. Of note, a recent candidate gene study[13] reported a nominally significant association between the intronic *PARP1* rs3219125 SNP and melanoma in a set of 585 individuals with melanoma and 585 control subjects (OR 1.89, 95% CI 1.34–2.68), with stronger effect seen in subjects with melanoma of the head and neck. The rs3219090 and rs3219125 SNPs were not highly correlated in the 1000 Genomes CEU samples ($r^2 = 0.042$). The rs3219125 SNP was not genotyped in our Australian discovery cohort but was well imputed (imputation $r^2 = 0.70$) and showed marginal evidence of association ($P = 0.053$). Although 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 three pairs, pairwise r^2 in the range 0.39–0.83). Further study will be required to establish this as a melanoma risk locus and 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 quantile-quantile 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 new 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 samples[14]. Our data are consistent with there being further common SNPs influencing melanoma risk, and we expect that further studies of additional individuals with melanoma will allow us to identify and characterize more loci.

In summary, our GWAS of melanoma identified one new melanoma susceptibility locus on chromosome 1 and replicated findings from

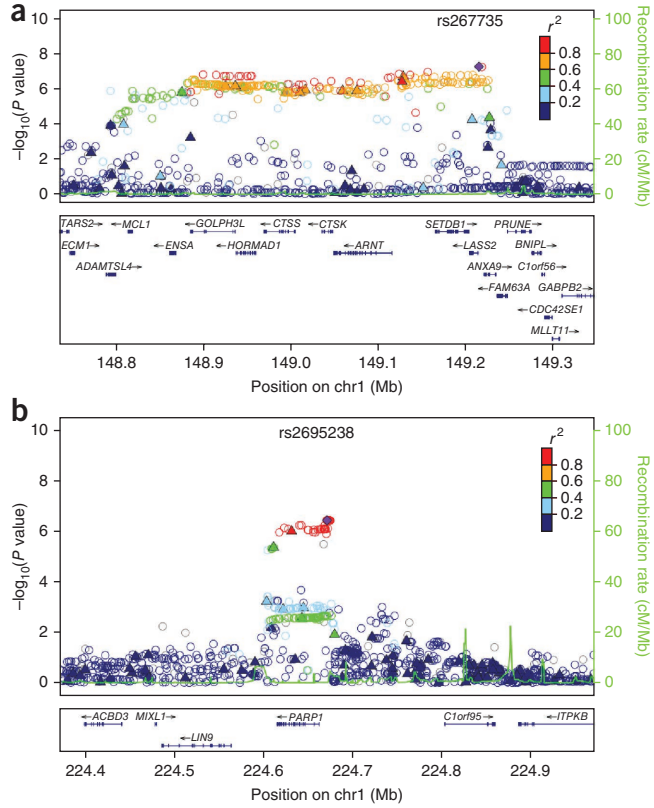

Figure 2 Discovery sample association results at two loci on chromosome 1 for both SNPs directly genotyped in all Australian samples and imputed SNPs. (**a**,**b**) Genotyped SNPs are indicated by filled-in triangles and imputed SNPs by empty circles. The top-ranked SNP at each locus is shown as a filled-in 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 was based on dosage scores, whereas with genotyped SNPs, hard genotype calls were used. Association results shown are for the chromosome 1 locus near 149 Mb (**a**) and SNPs in the vicinity of the *PARP1* association signal (**b**). The color scheme indicates linkage disequilibrium between the most strongly associated SNPs for the 149 Mb and *PARP1* regions (shown in purple, rs267735 and rs2695238, respectively) and other genotyped SNPs in the two regions.

© 2011 Nature America, Inc. All rights reserved.

previous melanoma GWAS. The observed effect size for the new locus was smaller than those observed for previously reported loci. Variants at the newly identified locus do not seem to be associated with human pigmentation or measures of nevus density, suggesting that they may influence melanoma risk through distinct mechanisms. Identification of the causal variants at this locus will help refine estimates of risk for this increasingly common cancer.

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/; Sib-pair, http://genepi.qimr.edu.au/staff/davidD/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

The AMFS and Q-MEGA gratefully acknowledge all their participants and the hard work of all their research interviewers and examiners and especially acknowledge C. Agha-Hamilton for managing the AMFS biospecimens. Q-MEGA thanks A. Baxter, M. de Nooyer, I. Gardner, D. Statham, B. Haddon, M.J. Wright, J. Palmer, J. Symmons, B. Castellano, L. Bardsley, S. Smith, D. Smyth, L. Wallace, M.J. Campbell, A. Caracella, M. Kvaskoff, O. Zheng, B. Chapman and H. Beeby for their input in 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 S. Nicolaides and the Queensland Medical Laboratory for the *pro bono* collection and delivery of blood samples and other pathology services for assistance with blood collection. We thank the following US state cancer registries for their help: Alabama, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Idaho, Illinois, Indiana, Iowa, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Nebraska, New Hampshire, New Jersey, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Virginia, Washington and Wyoming.

This work was supported by the Melanoma Research Alliance, the National Cancer Institute (NCI) of the US National Institutes of Health (NIH) (CA88363, CA83115, CA122838, CA87969, CA055075, CA100264, CA133996 and CA49449), the National Health and Medical Research Council (NHMRC) of Australia (107359, 200071, 241944, 339462, 380385, 389927, 389875, 389891, 389892, 389938, 402761, 443036, 442915, 442981, 496610, 496675, 496739, 552485 and 552498), the Cancer Councils New South Wales, Victoria and Queensland, the Cancer Institute New South Wales, the Cooperative Research Centre (CRC) for Discovery of Genes for Common Human Diseases, Cerylid Biosciences (Melbourne), the Australian Cancer Research Foundation, The Wellcome Trust and donations from Neville and Shirley Hawkins. 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). D.L.D., N.K.H. and G.W.M. are supported by the NHMRC Fellowships scheme and J.L.H. is an Australia Fellow of the NHMRC. S.M. is the recipient of a Career Development Award from the NHMRC (496674, 613705). D.C.W. is a Future Fellow of the Australian Research Council. K.M.B. is supported by the NCI of the NIH. J.M.T. is supported by the NCI of the NIH (CA109544). B.K.A. is supported by the University of Sydney Medical Foundation. A.E.C. is the recipient of a NHMRC public health postdoctoral fellowship (520018) and a Cancer Institute NSW Early Career Development Fellowship (10/ECF/2-06). A.M.G. and M.T.L. were supported by the Intramural Research Program of the NIH, NCI and DCEG. 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 the NIH (CA83115). Research at M.D. Anderson Cancer Center (MDACC) was partially supported by NIH grants (CA100264, 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 an NIH grant (HG004446). Research by the Barcelona team was supported by Fondo de Investigaciones Sanitarias (09/1393).

AUTHOR CONTRIBUTIONS

S.M., N.K.H. and K.M.B. wrote the manuscript. 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 sample collection, genotyping and analysis in 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.-d.P., A.M.G., P.A.K., N.A.G., P.H. and D.E.E. contributed to the collection, genotyping and analysis of the GenoMEL samples. C.I.A., Q.W., L.-E.W. and J.E.L. contributed to the collection, genotyping and analysis of the United States 1–M.D. Anderson Cancer Center samples. A.A.Q., M.Z. and J.H. contributed to the collection, genotyping and analysis of the United States 2–Harvard samples.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Bishop, D.T. *et al.* Genome-wide association study identifies three loci associated with melanoma risk. *Nat. Genet.* **41**, 920–925 (2009).
2. Brown, K.M. *et al.* Common sequence variants on 20q11.22 confer melanoma susceptibility. *Nat. Genet.* **40**, 838–840 (2008).
3. Gudbjartsson, D.F. *et al.* ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. *Nat. Genet.* **40**, 886–891 (2008).
4. Falchi, M. *et al.* Genome-wide association study identifies variants at 9p21 and 22q13 associated with development of cutaneous nevi. *Nat. Genet.* **41**, 915–919 (2009).
5. Duffy, D.L. *et al.* IRF4 variants have age-specific effects on nevus count and predispose to melanoma. *Am. J. Hum. Genet.* **87**, 6–16 (2010).
6. Baxter, A.J. *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.* **11**, 183–196 (2008).
7. Cust, A.E. *et al.* Population-based, case-control-family design to investigate genetic and environmental influences on melanoma risk: Australian Melanoma Family Study. *Am. J. Epidemiol.* **170**, 1541–1554 (2009).
8. Painter, J.N. *et al.* Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. *Nat. Genet.* **43**, 51–54 (2010).
9. Durbin, R.M. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
10. Li, C. *et al.* Genetic variants of the ADPRT, XRCC1 and APE1 genes and risk of cutaneous melanoma. *Carcinogenesis* **27**, 1894–1901 (2006).
11. Johansson, P., Pavey, S. & Hayward, N. Confirmation of a BRAF mutation-associated gene expression signature in melanoma. *Pigment Cell Res.* **20**, 216–221 (2007).
12. Ceol, C.J. *et al.* The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. *Nature* **471**, 513–517 (2011).
13. Zhang, M., Qureshi, A.A., Guo, Q. & Han, J. Genetic variation in DNA repair pathway genes and melanoma risk. *DNA Repair (Amst).* **10**, 111–116 (2011).
14. Park, J.H. *et al.* Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat. Genet.* **42**, 570–575 (2010).

© 2011 Nature America, Inc. All rights reserved.

[1]Queensland Institute of Medical Research, Brisbane, Queensland, Australia. [2]Westmead Institute of Cancer Research, University of Sydney at Westmead Millennium Institute and Melanoma Institute Australia, Westmead, New South Wales, Australia. [3]Department of Dermatology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. [4]Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, School of Population Health, The University of Melbourne, Victoria, Australia. [5]Viertel Centre for Research in Cancer Control, Cancer Council Queensland, Brisbane, Queensland, Australia. [6]Cancer Epidemiology and Services Research, Sydney School of Public Health, The University of Sydney, New South Wales, Australia. [7]Lund Cancer Center Department of Oncology, University Hospital, Lund, Sweden. [8]Melanoma Unit, Dermatology Department, Hospital Clínic, Institut de Investigació Biomèdica August Pi Suñe, Universitat de Barcelona, Barcelona, Spain. [9]Centre of Biomedical Research on Rare Diseases (CIBERER), Instituto de Salud Carlos III, Barcelona, Spain. [10]Department of Internal Medicine, University of Genoa, Genoa, Italy. [11]Department of Oncology-Pathology, Karolinska Institute and Karolinska University Hospital, Solna, Stockholm, Sweden. [12]INSERM, U946, Fondation Jean-Dausset–CEPH, Paris, France. [13]Université Paris Diderot, Sorbonne Paris Cité, Institut Universitaire d'Hématologie Paris, France. [14]Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health, Bethesda, Maryland, USA. [15]International Hereditary Cancer Center, Szczecin, Poland. [16]University of Glasgow Department of Public Health, Glasgow, UK. [17]Department of Dermatology/Dermato-Oncology Unit, Sheba Medical Center, Tel Hashomer, Israel. [18]Service de Génétique, Institut de Cancérologie Gustave Roussy, Villejuif, France. [19]Center for Clinical Epidemiology & Biostatistics, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [20]Department of Biostatistics & Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [21]Department of Dermatology, Leiden University Medical Center, Leiden, The Netherlands. [22]Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [23]Section of Epidemiology and Biostatistics, Leeds Institute of Molecular Medicine, Cancer Research UK Clinical Centre at Leeds, St James's University Hospital, Leeds, UK. [24]Department of Dermatology, Oslo University Hospital, Oslo, Norway. [25]Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas, USA. [26]Department of Surgical Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas, USA. [27]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. [28]Cancer Epidemiology Centre, The Cancer Council Victoria, Carlton, Victoria, Australia. [29]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [30]Genetic Basis of Human Disease Division, The Translational Genomics Research Institute, Phoenix, Arizona, USA. [31]Center for Cancer Genomics and Computational Biology, Van Andel Research Institute, Grand Rapids, Michigan, USA. [32]Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health, Bethesda, Maryland, USA. [33]On behalf of the Lund Melanoma Group. Correspondence should be addressed to S.M. (stuart.macgregor@qimr.edu.au).

© 2011 Nature America, Inc. All rights reserved.

ONLINE METHODS

Statistical analysis, genotyping and data quality control for the Australian discovery sample. DNA was extracted from peripheral blood or saliva samples. Control samples from Australian twins and individuals with endometriosis were genotyped at deCODE Genetics on the Illumina HumanHap610W Quad and Illumina HumanHap670 Quad Beadarrays, respectively. AMFS control subjects were genotyped by Illumina on Illumina Omni1-Quad arrays. Samples from individuals with melanoma were genotyped by Illumina on Illumina Omni1-Quad (568 AMFS subjects, 699 Q-MEGA subjects) and HumanHap610W Quad arrays (998 Q-MEGA subjects). Approval for the Australian melanoma subjects and for AMES control subjects was obtained from the Human Research Ethics Committees of QIMR, the University of Sydney, The University of Melbourne and the cancer registric of NSW, Victoria and Queensland. Approval for the twin and endometriosis studies was obtained from the QIMR Human Research Ethics Committee and the Australian Twin Registry. The protocols for studies performed at the University of Texas M.D. Anderson Cancer Center and Harvard University were approved by the Institutional Review Boards of these institutions. Approval for the GenoMEL studies was obtained for each recruiting center. Informed consent was obtained from all participants.

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) or with Hardy-Weinberg equilibrium in controls of $P < 10^{-6}$ 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 5 Northern European populations genotyped by the GenomeEUtwin consortium, using the EIGENSOFT package[15]. Individuals lying ≥2 s.d. from the mean PC1 and PC2 scores were excluded from subsequent analyses. After these exclusions, 2,168 samples from individuals with melanoma (1,242 genotyped on the Omni1-Quad and 926 genotyped on the Hap610 arrays) and 4,387 samples from control subjects (431 genotyped on the Omni1-Quad and 3,956 genotyped on the Hap610 arrays) were retained for subsequent analyses (**Table 1**). Individuals genotyped on the Omni1-Quad array had genotypes for up to 816,169 SNPs, whereas those genotyped on the Hap610 or Hap670 arrays had genotypes for up to 544,483 SNPs. There were 299,394 SNPs that passed quality control and overlapped between these arrays (and hence directly genotyped in all Australian samples).

To investigate the 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 the first 10 principal components (as well as the first two, with similar results, data not shown) for the combined case and control samples using EIGENSOFT.

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

Genomic imputation. Imputation for the Australian samples was performed using MACH2 (ref. 16) with the 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 case-control samples ($N = 292,043$). Imputation was run in two stages. First, data from a set of representative Australian discovery sample individuals was compared to the phased haplotype data from the 1000 Genomes data to generate recombination and error maps. In the second stage, data were imputed for all individuals using the phased 1000 Genomes data as the reference panel and the recombination and error files generated in stage 1. In total, 5,480,804 SNPs from the 1000 Genomes data 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 option[17]. Analysis of dosage scores from the imputation analysis was performed using mach2dat[16]. Analysis was carried out 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 Fig. 2a**,**b**). Meta-analysis of discovery and replication cohorts was performed using PLINK (-meta-analysis option) with ORs weighted by the inverse of their variance (fixed-effects model). The heterogeneity of ORs between studies was tested using Cochran's *Q* statistic (neither rs3219090 nor rs7412746 showed evidence of heterogeneity of ORs between studies). The proportion of variance explained by rs3219090 and rs7412746 was derived assuming a population prevalence of 0.05 (ref. 18) and sibling relative risk of 3 (ref. 19). Given the small ORs observed and assuming a large number of similarly 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 siblings[20]. PRR was estimated from the ORs and allele frequencies using output from the GRR function in Sibpair. Tests for heterogeneity between early- and late-onset melanomas in AAO and between *in situ* and invasive melanoma subsets were performed 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 carried out using linear regression, with permutation used to compute empirical *P* values. Association analysis of melanoma corrected for pigmentation and nevus count was performed using logistic regression with factors for each level of 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 1,000 bootstrap replicates from the data, and the actual data differences in OR with and without covariates were compared with bootstrap replicates to compute empirical *P* values. Association plots were created using LocusZoom[21].

Replication sample 1: GenoMEL. A trend test was applied to each SNP in turn, which was stratified by broad geographic region (8 regions prespecified).

Replication sample 2: M.D. Anderson Cancer Center. Association analysis of genotyped SNPs was performed using the PLINK-logistic option[13]. The first two principal components were included to adjust for population structure.

Replication sample 3: Harvard. Association analysis of genotyped SNPs was performed using SAS V9.1 (SAS Institute). Unconditional logistic regression was employed to calculate ORs and 95% CIs, adjusting for age and gender.

15. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
16. Li, Y. & Mach Abecasis, G. 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. *Am. J. Hum. Genet.* **S79**, 2290 (2006).
17. 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).
18. Queensland Cancer Registry. Cancer in Queensland: incidence and mortality, 1982 to 2005 (The Cancer Council Queensland, 2008).
19. Do, K.A., Aitken, J.F., Green, A.C. & Martin, N.G. Analysis of melanoma onset: assessing familial aggregation by using estimating equations and fitting variance components via Bayesian random effects models. *Twin Res.* **7**, 98–113 (2004).
20. James, J.W. Frequency in relatives for an all-or-none trait. *Ann. Hum. Genet.* **35**, 47–49 (1971).
21. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).

© 2011 Nature America, Inc. All rights reserved.