

Genome-wide association study identifies three loci associated with melanoma risk

D Timothy Bishop¹, Florence Demenais², Mark M Iles¹, Mark Harland¹, John C Taylor¹, Eve Corda^{2,3}, Juliette Randerson-Moor¹, Joanne F Aitken⁴, Marie-Francoise Avril⁵, Esther Azizi⁶, Bert Bakker⁷, Giovanna Bianchi-Scarrà⁸, Brigitte Bressac-de Paillerets⁹, Donato Calista¹⁰, Lisa A Cannon-Albright¹¹, Thomas Chin-A-Woeng¹², Tadeusz Dębniak¹³, Gilli Galore-Haskel⁶, Paola Ghiorzo⁸, Ivo Gut¹⁴, Johan Hansson¹⁵, Marko Hočvar¹⁶, Veronica Höiom¹⁵, John L Hopper¹⁷, Christian Ingvar¹⁸, Peter A Kanetsky¹⁹, Richard F Kefford²⁰, Maria Teresa Landi²¹, Julie Lang²², Jan Lubiński¹³, Rona Mackie²³, Josep Malvehy²⁴, Graham J Mann²⁰, Nicholas G Martin²⁵, Grant W Montgomery²⁵, Frans A van Nieuwpoort²⁶, Srdjan Novakovic¹⁶, Håkan Olsson¹⁸, Susana Puig²⁴, Marjan Weiss⁷, Wilbert van Workum¹², Diana Zelenika¹⁴, Kevin M Brown²⁷, Alisa M Goldstein²¹, Elizabeth M Gillanders²⁸, Anne Boland¹⁴, Pilar Galan²⁹, David E Elder³⁰, Nelleke A Gruis²⁶, Nicholas K Hayward²⁵, G Mark Lathrop^{3,14}, Jennifer H Barrett¹ & Julia A Newton Bishop¹

We report a genome-wide association study of melanoma conducted by the GenoMEL consortium based on 317K tagging SNPs for 1,650 selected cases and 4,336 controls, with replication in an additional two cohorts (1,149 selected cases and 964 controls from GenoMEL, and a population-based case-control study in Leeds of 1,163 cases and 903 controls). The genome-wide screen identified five loci with genotyped or imputed SNPs reaching $P < 5 \times 10^{-7}$. Three of these loci were replicated: 16q24 encompassing *MC1R* (combined $P = 2.54 \times 10^{-27}$ for rs258322), 11q14-q21 encompassing *TYR* ($P = 2.41 \times 10^{-14}$ for rs1393350) and 9p21 adjacent to *MTAP* and flanking *CDKN2A* ($P = 4.03 \times 10^{-7}$ for rs7023329). *MC1R* and *TYR* are associated with pigmentation, freckling and

cutaneous sun sensitivity, well-recognized melanoma risk factors. Common variants within the 9p21 locus have not previously been associated with melanoma. Despite wide variation in allele frequency, these genetic variants show notable homogeneity of effect across populations of European ancestry living at different latitudes and show independent association to disease risk.

Cutaneous melanoma is almost entirely a disease of fair-skinned individuals. Increased risk for melanoma is associated with a family history of the disease¹, pigmentation phenotypes, the number of melanocytic nevi^{2,3} and iatrogenic immuno-suppression⁴. Pigmentation risk factors include fair skin, blue or green eyes, blond or red hair, sun sensitivity or inability to tan⁵⁻⁸, each associated with an approximate

¹Section of Epidemiology and Biostatistics, Leeds Institute of Molecular Medicine, Cancer Research UK Clinical Centre at Leeds, St James's University Hospital, Leeds, UK. ²INSERM, U946, Fondation Jean-Dausset-CEPH, 75010 Paris, France. ³Fondation Jean Dausset-CEPH, 75010, Paris, France. ⁴Viertel Centre for Research in Cancer Control, The Cancer Council, Queensland, Spring Hill, Brisbane, Australia. ⁵AP-HP, Hôpital Cochin, Service de Dermatologie, Faculté Paris 5, Paris, France. ⁶Department of Dermatology, Sheba Medical Center, Tel Hashomer, Sackler Faculty of Medicine, Tel Aviv University, Israel. ⁷Department of Clinical Genetics, Center of Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands. ⁸Department of Oncology, Biology and Genetics, University of Genoa, Genova, Italy. ⁹Service de Génétique et FRE 2939 CNRS, Institut de Cancérologie Gustave Roussy, Villejuif, France. ¹⁰Dermatology Unit, Maurizio Bufalini Hospital, Cesena, Italy. ¹¹Genetic Epidemiology, Biomedical Informatics, University of Utah, Salt Lake City, Utah, USA. ¹²ServiceXS, Plesmanlaan 1d, 2333 BZ Leiden, The Netherlands. ¹³International Hereditary Cancer Center, Pomeranian Medical University, Szczecin, Poland. ¹⁴Commissariat à l'énergie Atomique, Institut de Génétique, Centre National de Génotypage, Evry, France. ¹⁵Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden. ¹⁶Institute of Oncology Ljubljana, Zaloška 2, Ljubljana, Slovenia. ¹⁷Centre for Molecular, Environmental, Genetic and Analytic (MEGA) Epidemiology, School of Population Health, University of Melbourne, Melbourne, Australia. ¹⁸Department of Oncology, University Hospital Lund, Barngatan 2B, Lund, Sweden. ¹⁹Centre for Clinical Epidemiology & Biostatistics and Department of Biostatistics & Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA. ²⁰Westmead Institute for Cancer Research, University of Sydney at Westmead Millennium Institute and Melanoma Institute Australia, Westmead, Australia. ²¹Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA. ²²Department of Medical Genetics, University of Glasgow, UK. ²³Departments of Medical Genetics and of Public Health & Health Policy, 1 Lilybank Gardens, University of Glasgow, UK. ²⁴Melanoma Unit, Dermatology Department, Hospital Clinic, IDIBAPS, Barcelona, Spain and CIBER de Enfermedades Raras, Barcelona, Spain. ²⁵Queensland Institute of Medical Research, Herston, QLD, Australia. ²⁶Department of Dermatology, Leiden University Medical Centre, Leiden, The Netherlands. ²⁷Melanoma Genomics Laboratory, The Translational Genomics Research Institute (TGen), Phoenix, Arizona, USA. ²⁸Inherited Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, Maryland, USA. ²⁹UMR U557 Inserm, U1125 Inra, Cnam, Paris 13, CRNH IdF, Bobigny, France. ³⁰Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA. Correspondence should be addressed to D.T.B. (d.t.bishop@leeds.ac.uk).

Received 1 January; accepted 10 June; published online 5 July 2009; doi:10.1038/ng.411

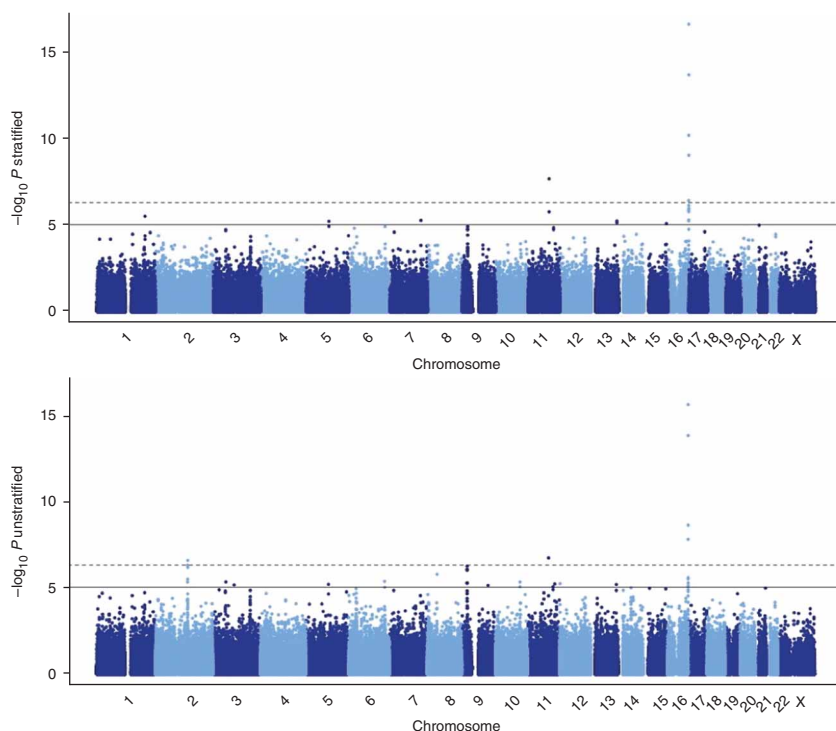


Figure 1 Results of Cochran-Armitage (CA) trend test for all SNPs passing quality control reported as $-\log_{10} P$ values. The analysis stratified by geographic region is on the top with the unstratified analysis on the bottom (see Online Methods for details). The solid horizontal line indicates P value of 10^{-5} and the dashed line indicates P value of 5×10^{-7} .

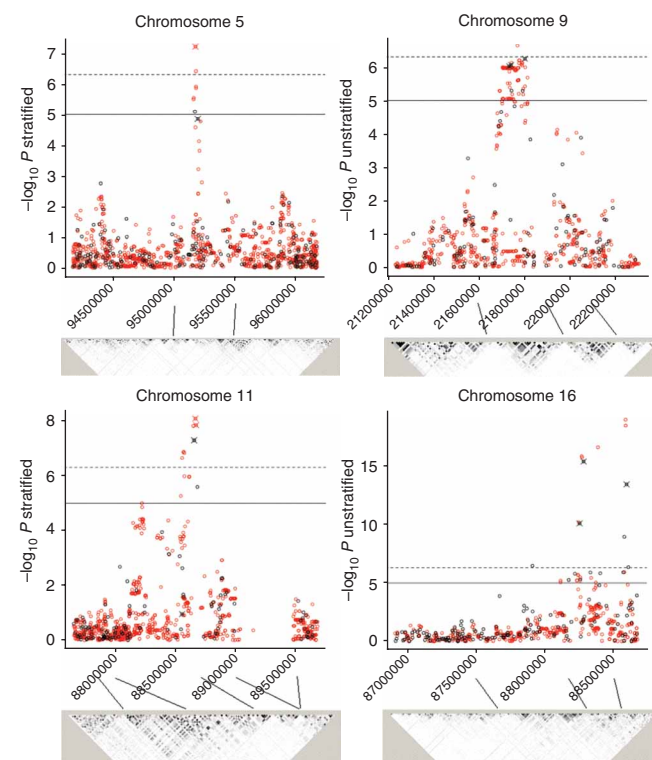
doubling of risk. Variants of the *MC1R* (melanocortin-1 receptor) gene are associated with the combination of red hair, freckling and sun sensitivity, but *MC1R* may also have a nonpigmentation effect^{7–10}. A detailed genetic screen of pigmentation loci found that *TYR* (tyrosinase) gene variants and a haplotype spanning the *ASIP* (agouti signaling protein) locus were also associated with melanoma¹¹. Increased intermittent exposure to UV radiation, rather than chronic exposure, is thought to be responsible for the continued increase in incidence in many populations¹².

The most common high-penetrance melanoma susceptibility gene (*CDKN2A*) maps to 9p21 (ref. 13). Germline *CDKN2A* mutations are carried by about 2% of all melanoma cases across populations¹⁴. The more cases of melanoma in a family, the more prevalent are *CDKN2A* germline mutations¹⁵, which seem to be associated with increased numbers of nevi in at least some families¹⁶.

This study was carried out by the GenoMEL Consortium (**Supplementary Note**), which focuses its efforts on studies of the genetic susceptibility to melanoma. The current genome-wide association study (GWAS) is based on population samples collected by GenoMEL

Figure 2 Stratified CA trend tests. The $-\log_{10} P$ values are from the CA trend test stratified (by geographical region) for genotyped and imputed SNPs in candidate regions for follow-up. Genotyped SNPs are shown in black and imputed SNPs in red; those SNPs that were followed up in the replication stage are indicated by a cross. The solid horizontal line indicates a P value of 10^{-5} and the dashed line indicates a P value of 5×10^{-7} . The greyscale plots indicate extent of pairwise linkage disequilibrium (measured by r^2) between SNPs, estimated from our data using Haploview²³. Imputed data were not used for this. The darkest shading (black) indicates $r^2 \geq 0.9$, the next darkest $0.8 \geq r^2 > 0.9$, and so forth. The position of the SNPs relative to the graph is indicated by the lines between the two.

To confirm the effectiveness of our methods for dealing with stratification, we produced quantile-quantile plots (**Supplementary Note**). We selected regions with multiple SNPs (within 50 kb) that each had a



participants to identify common genetic variants contributing to melanoma risk across populations of European ancestry living at different latitudes. In total, ten GenoMEL groups contributed 1,650 cases and 1,065 controls of European ancestry for this GWAS (**Supplementary Table 1**). Cases were selected in an effort to enrich for individuals that may have greater genetic predisposition to melanoma, through selection of cases with family history, multiple primaries or early onset of disease (see Online Methods). Genotyping results were also available from 1,824 French controls held by Centre National de Génotypage (CNG) and 1,447 UK controls from the 1958 Birth Cohort genotyped by the Wellcome Trust Case Control Consortium (WTCCC)¹⁷ (**Supplementary Fig. 1**).

Following quality control and exclusion of persons of apparent non-European ancestry through principal components analysis (**Supplementary Fig. 2**), case-control analyses were conducted both (i) stratified (the primary analysis) and (ii) unstratified by geographical region (**Fig. 1**). The unstratified analysis should provide greater power for associations common to multiple sample sets, whereas the stratified analysis protects against population stratification but is likely conservative in the absence of stratification.

Table 1 Genome-wide association and replication for melanoma risk variants

| Locus | Genome-wide phase | | | | Replication phase (GenoMEL + Leeds) | | | | Combined (Genome-wide + Replication) | | | |
|--|-------------------------|--------------|------|------------------------|-------------------------------------|------------------------|-------------------|------------------------|--------------------------------------|---|---------|--|
| | SNP | Minor allele | MAF | P | OR (CI) | P | OR (CI) | P | OR (CI) | P | OR (CI) | |
| Loci considered for follow-up on basis of this study | | | | | | | | | | | | |
| Chromosome 9 (21737803–21806528) | | | | | | | | | | | | |
| | rs4636294 | G | 0.51 | 1.77×10^{-5} | 0.83 (0.76, 0.90) | 0.032 | 0.90 (0.83, 0.99) | 1.97×10^{-6} | 0.86 (0.81, 0.92) | | | |
| | rs2218220 | T | 0.51 | 1.85×10^{-5} | 0.83 (0.76, 0.90) | 0.046 | 0.91 (0.83, 1.00) | 6.40×10^{-6} | 0.87 (0.82, 0.92) | | | |
| | rs1335510 | G | 0.41 | 1.14×10^{-4} | 0.84 (0.77, 0.92) | – | – | – | – | | | |
| | rs935053 ^a | A | 0.49 | 4.72×10^{-6} | 0.81 (0.74, 0.89) | – | – | – | – | | | |
| | rs10757257 | A | 0.41 | 3.39×10^{-5} | 0.83 (0.76, 0.91) | – | – | – | – | | | |
| | rs7023329 | G | 0.50 | 1.14×10^{-5} | 0.82 (0.75, 0.90) | 0.023 | 0.90 (0.82, 0.99) | 4.03×10^{-7} | 0.85 (0.80, 0.91) | | | |
| Chromosome 11 (88551344–88667691) | | | | | | | | | | | | |
| | rs1042602 | A | 0.39 | 1.13×10^{-1} | 0.93 (0.85, 1.02) | 0.054 | 0.91 (0.83, 1.00) | 1.07×10^{-2} | 0.92 (0.87, 0.98) | | | |
| | rs1393350 | A | 0.27 | 4.28×10^{-8} | 1.30 (1.19, 1.43) | 1.38×10^{-7} | 1.30 (1.18, 1.44) | 2.41×10^{-14} | 1.29 (1.21, 1.38) | | | |
| | rs1126809 | A | 0.27 | – | – | 1.17×10^{-6} | 1.27 (1.16, 1.40) | 1.17×10^{-6} | 1.27 (1.16, 1.40) | | | |
| | rs1847142 ^a | A | 0.32 | 8.35×10^{-9} | 1.33 (1.21, 1.47) | 6.48×10^{-4} | 1.26 (1.10, 1.44) | 2.24×10^{-11} | 1.31 (1.21, 1.41) | | | |
| | rs1806319 | C | 0.35 | 2.49×10^{-6} | 1.24 (1.13, 1.35) | – | – | – | – | | | |
| | rs10830253 | G | 0.31 | – | – | 2.81×10^{-6} | 1.26 (1.14, 1.39) | 2.81×10^{-6} | 1.26 (1.14, 1.39) | | | |
| Chromosome 16 (87913062–88607035) | | | | | | | | | | | | |
| | rs2353033 | C | 0.41 | 3.32×10^{-7} | 1.26 (1.15, 1.38) | – | – | – | – | | | |
| | rs352935 | C | 0.46 | 5.78×10^{-6} | 1.22 (1.12, 1.33) | – | – | – | – | | | |
| | rs164741 | A | 0.29 | 1.50×10^{-6} | 1.26 (1.15, 1.39) | – | – | – | – | | | |
| | rs7188458 | A | 0.42 | 7.99×10^{-11} | 1.34 (1.23, 1.46) | 1.20×10^{-3} | 1.25 (1.09, 1.42) | 1.16×10^{-12} | 1.30 (1.21, 1.40) | | | |
| | rs459920 | C | 0.47 | 5.41×10^{-6} | 0.82 (0.75, 0.89) | – | – | – | – | | | |
| | rs12918773 ^a | A | 0.08 | 1.34×10^{-16} | 1.87 (1.62, 2.16) | – | – | – | – | | | |
| | rs258322 | A | 0.09 | 7.54×10^{-17} | 1.81 (1.58, 2.08) | 1.07×10^{-10} | 1.55 (1.36, 1.77) | 2.54×10^{-27} | 1.67 (1.52, 1.83) | | | |
| | rs1800359 | A | 0.42 | 1.41×10^{-6} | 0.80 (0.74, 0.88) | – | – | – | – | | | |
| | rs11861084 | A | 0.43 | 1.64×10^{-6} | 0.81 (0.74, 0.88) | – | – | – | – | | | |
| | rs4408545 | T | 0.51 | 1.29×10^{-6} | 0.81 (0.74, 0.88) | – | – | – | – | | | |
| | rs4238833 | G | 0.36 | 1.04×10^{-9} | 1.32 (1.21, 1.44) | – | – | – | – | | | |
| | rs4785763 | A | 0.32 | 2.84×10^{-14} | 1.42 (1.30, 1.56) | 5.13×10^{-8} | 1.30 (1.18, 1.43) | 5.96×10^{-22} | 1.36 (1.28, 1.45) | | | |
| | rs8059973 | A | 0.17 | 6.81×10^{-7} | 0.74 (0.65, 0.83) | – | – | – | – | | | |
| Loci considered for follow-up on basis of other studies | | | | | | | | | | | | |
| Chromosome 20 (32635433–33040650) | | | | | | | | | | | | |
| | rs910873 | A | 0.07 | – | – | 1.92×10^{-8} | 1.55 (1.33, 1.81) | 1.92×10^{-8} | 1.55 (1.33, 1.81) | | | |
| | rs17305573 | C | 0.09 | – | – | 3.95×10^{-5} | 1.53 (1.25, 1.87) | 3.95×10^{-5} | 1.53 (1.25, 1.87) | | | |
| | rs4911442 | G | 0.12 | – | – | 1.80×10^{-5} | 1.48 (1.24, 1.77) | 1.80×10^{-5} | 1.48 (1.24, 1.77) | | | |
| | rs1885120 | C | 0.07 | – | – | 1.25×10^{-8} | 1.58 (1.35, 1.85) | 1.25×10^{-8} | 1.58 (1.35, 1.85) | | | |
| Chromosome 22 (36874244–36875565) | | | | | | | | | | | | |
| | rs2284063 | G | 0.37 | 5.60×10^{-5} | 0.83 (0.76, 0.91) | 6.45×10^{-5} | 0.82 (0.75, 0.90) | 2.40×10^{-9} | 0.83 (0.78, 0.88) | | | |
| | rs6001027 | G | 0.37 | 4.50×10^{-5} | 0.83 (0.76, 0.91) | 2.26×10^{-3} | 0.86 (0.78, 0.95) | 1.94×10^{-8} | 0.83 (0.78, 0.89) | | | |

SNPs in regions showing the highest significance levels are listed with their chromosome and position (in base-pair range), minor allele, and Minor allele frequencies (MAF) computed from genome-wide phase controls. The replication data sets consist of the Leeds case-control study and GenoMEL case-control samples. The following SNPs were examined because of published literature. rs1042602 was reported to be associated with hair color and skin phenotype, so was genotyped in both replication sets¹¹. rs1126809 is a functional SNP in 77R, so was genotyped in both replication sets. rs187142 was genotyped in the genome-wide data set and in the Leeds case-control samples but not genotyped in the GenoMEL replication set. rs10830253 is a replacement for rs1806319 and rs188458 was genotyped in the genome-wide data set and in the GenoMEL replication set but was not genotyped in the Leeds case-control samples. Two other regions containing putative melanoma-susceptibility genes were examined. The chromosome 20 region contains *ASIP*, a pigmentation gene, identified by an analysis of pigmentation genes¹¹, which is in proximity to a region identified in a pooled genome-wide analysis²¹ (none of these SNPs were on the genome-wide array); the chromosome 22 region was identified in an adjoining paper involving a GWAS of nevus count variation²². ^aImputed SNPs.

P value $< 10^{-5}$ from either the stratified or unstratified analysis and conducted imputation on these regions. We found five regions, on chromosomes 2, 5, 9, 11 and 16, in which at least one imputed or genotyped SNP achieved genome-wide significance ($P < 5 \times 10^{-7}$, **Table 1**, **Supplementary Table 2** and **Fig. 2**). The chromosome 2 region showed strong evidence of association in the unstratified analysis only and was not followed up as it contains the *LCT* (lactase) gene, whose SNPs are known to vary in frequency across Europe (**Supplementary Table 2**)¹⁸. To replicate the other loci harboring genome-wide significant SNPs, we analyzed an independent set of 1,149 cases and 964 controls from GenoMEL. As in the first cohort, cases were selected for enrichment of family history, multiple primaries or early onset of disease (see Online Methods and

Supplementary Note). A further replication set comprised 1,163 cases and 903 controls from population-based studies from Leeds, UK. Samples were genotyped at the most significant SNPs from the GWAS, and also at any imputed SNPs that showed more significant association. If assays could not be produced or analyzed, alternative SNPs were genotyped.

Of the four regions that we attempted to replicate, only one (chromosome 5) showed no evidence of replication (**Table 1** and **Supplementary Table 3**). The strongest association was with a region near the telomere of chromosome 16q: the three SNPs chosen for replication covered a 340-kb region with P values ranging between 7.99×10^{-11} and 7.54×10^{-17} in the GWAS sample and between 1.16×10^{-12} and 2.54×10^{-27} in the combined (genome-wide and

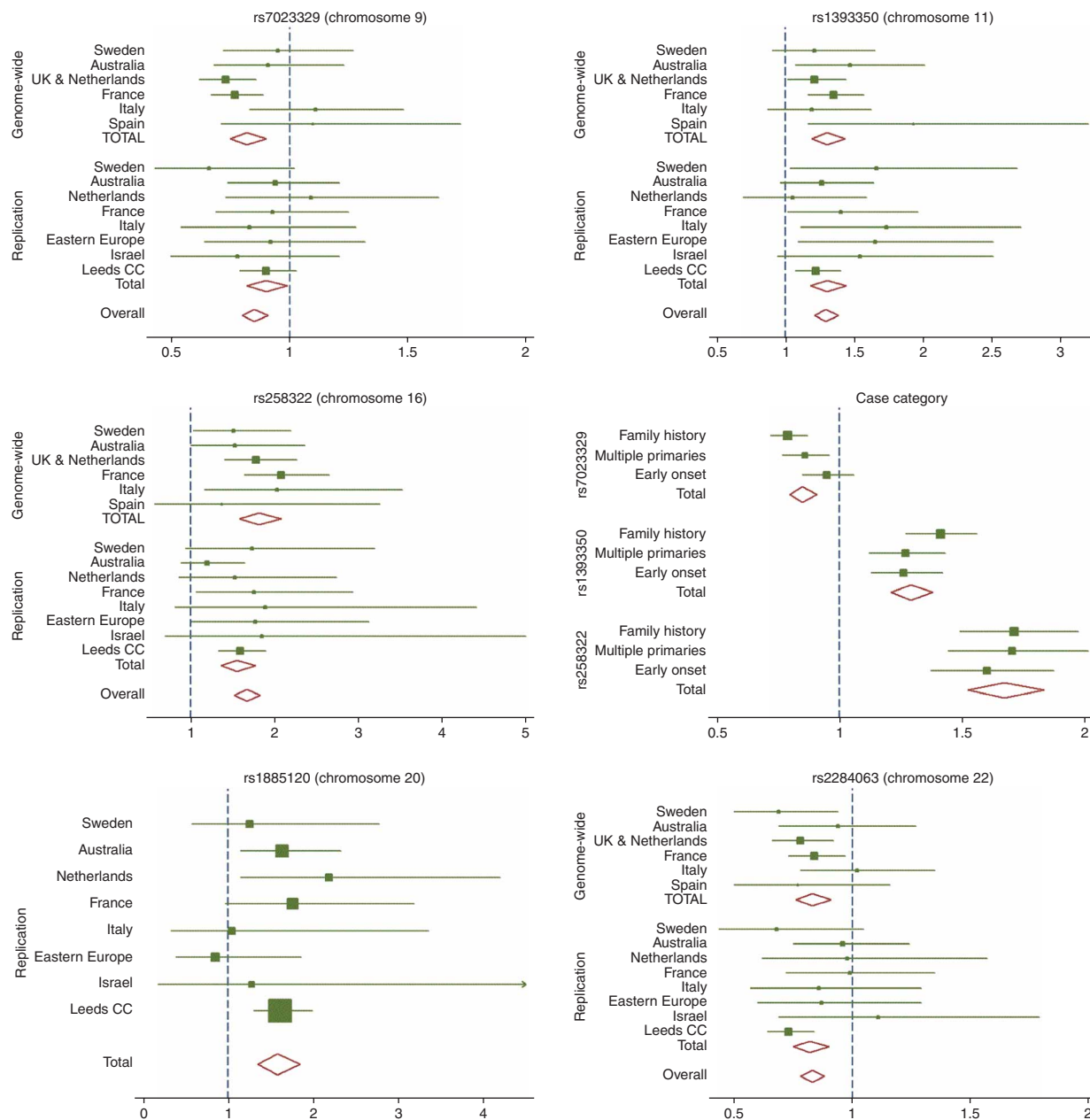


Figure 3 Forest plot of associations by geography and case category. The forest plot of the estimated per-allele OR from the CA trend test across geographical regions and overall for (i) chromosome 9 (rs7023329, top left), (ii) chromosome 11 (rs1393350, top right), (iii) chromosome 16 (rs258322, middle left), (iv) by case-category for these three SNPs (middle right), (v) chromosome 20 (rs1885120, only genotyped in replication set, bottom left) and (vi) chromosome 22 (rs2284063, bottom right).

replication phases) sample. This region includes several candidate genes, notably *MC1R*, *CDK10* (involved in cell cycle regulation) and *FANCA* (which regulates genomic stability). A number of the 16q SNPs were reported earlier by Han *et al.*¹⁹, who in a GWAS examining hair color and skin pigmentation identified rs258322, which showed the strongest effect in both their and our study. Han *et al.*¹⁹ showed that the rs258322 signal was explained by functional *MC1R* variants. In this study, stepwise regression analysis of 30 tag SNPs covering 1.4 Mb on 16q24 (Supplementary Table 4) showed evidence for association with three SNPs (rs258322, rs4785763, rs8059973); one of these (rs8059973) was not reported by the Han *et al.* study¹⁹ and seems to have an association independent of the other two on melanoma risk. The magnitude of the association with rs258322 (per-allele odds ratio (OR) of 1.67, 95% confidence interval (CI) = 1.52–1.83, Table 1) is notable, as it is comparable to that found by a recent meta-analysis of *MC1R* variants⁹, even though this SNP is distant from *MC1R*.

The chromosome 11 region (overall strongest evidence for rs1393350, $P = 2.41 \times 10^{-14}$) includes the *TYR* gene, which has previously been associated with melanoma¹¹; the coding variant (rs1126809) is included in the replication panel and shows a strong association in terms of OR for genotyped SNPs (OR = 1.27, 95% CI = 1.16–1.40, $P = 1.17 \times 10^{-6}$; Table 1 and Supplementary Table 3). Stepwise regression analysis indicates that, after accounting for rs1393350, no other SNPs in the region were independently associated with melanoma (Online Methods and Supplementary Table 4).

The third locus replicated is on 9p21, with the combined P value reaching 4.03×10^{-7} for rs7023329, which falls within the *MTAP* gene (Table 1 and Supplementary Table 3). Stepwise regression analysis of 11 tag SNPs covering a 2-Mb region on 9p21 showed evidence for the independent associations with one replicated SNP (rs7023329) plus another one, rs1011970, which is 246 kb centromeric and is within *ANRIL*, an antisense noncoding mRNA starting in the promoter of the p14ARF transcript of the *CDKN2A* gene and overlapping *CDKN2B*²⁰ (Supplementary Table 4, Supplementary Fig. 3). Notably, antisense transcripts in the *MTAP* region have been suggested for *CDKN2A* and *CDKN2B*³.

We also attempted to replicate findings from two other related genome-wide association studies^{21,22}. SNPs in these regions did not achieve the critical significance level for follow-up in our GWAS study, but the available evidence suggested these regions as strong candidates for containing melanoma susceptibility genes. A further pigmentation gene has been reported to be associated with melanoma (*ASIP*¹¹), adjacent to a reported melanoma locus recently identified by a pooling study²¹; this chromosomal region is confirmed by the present data (Table 1 and Supplementary Table 3). In an accompanying article in this issue of the journal, Falchi *et al.* report a GWAS of nevus count variation based on volunteer twins from the UK, with replication in an Australian cohort²². Falchi *et al.* report two loci associated with nevus count and also show these are each associated with increased risk of melanoma. One of these loci overlaps with the chromosome 9 region identified here (Supplementary Note). The second nevus locus reported is on chromosome 22 and was also found to be associated with melanoma risk in our study (Table 1 and Supplementary Table 3).

Multiple logistic regression analysis of the three loci on chromosomes 9, 11 and 16 showed significant evidence for six SNPs being associated with melanoma risk (Supplementary Table 4). No pairwise interactions between the most significant SNPs at the three loci were significant (all $P > 0.19$), showing that the pattern of association with the three loci is consistent with a multiplicative model (Supplementary Table 4). Furthermore, examination of the associations by case

phenotype category showed similar associations for family history, multiple primaries and early onset (Fig. 3), although for some SNPs at the 9q21 and 16q24 loci the effect size was marginally weaker for early-onset melanoma.

The strength of our study, which combines cases and controls from geographically distinct sites, lies in its ability to identify genetic variants that affect disease risk within diverse populations of European origin and to estimate the effects by population. Despite the variation in allele frequencies between groups (Supplementary Table 2) and the large differences in sun exposure between sites, the effect sizes for all three identified regions show marked homogeneity across the groups (Fig. 3).

Given that the current study was well powered to detect variants with OR > 1.5, it seems unlikely that there exist many additional common SNPs with a large effect on melanoma risk. Some of the genetic variants found to date to be associated with melanoma risk have also been found to be associated with pigmentation phenotypes^{11–15} and nevus numbers²², but other variants may be involved. Further, Falchi *et al.*²² found that adjustment of the SNP–melanoma association for nevus count reduced the strength of the SNP association with melanoma, suggesting that the SNP and nevus count are measuring in part the same disease-associated factor(s). Further studies are required to assess whether the loci characterized by the present study act mainly through melanoma-associated phenotypes and/or have independent associations with melanoma risk.

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 study 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 RO1 ROI CA83115. Work at CNIG was supported by the Ministère de l'Enseignement Supérieur et de la Recherche and Institut National du Cancer (INCa).

This study was also supported by the following: research grant no. C-20070050 from the Israel Cancer Association, the Intramural Research Program of the NIH, National Cancer Institute, Division of Cancer Epidemiology and Genetics, and by National Cancer Institute grants RO1 CA5558 to M.T.L. Grants 03/0019 and 06/0265 from Fondo de Investigaciones Sanitarias; grant RO-1 CA 83115 (fund 538226 from National Cancer Institute); Francisco Cuellar from CONACYT, Mexico, 2007 Italian Ministry of Health DGRST.4/4235-P1.9.A.B to G.B.-S., the Swedish Cancer Society, Kamprad Foundation, Lund Hospital Fund, National Institutes of Health (#CA88363), The Swedish Cancer Society (Grant No: 4860-B06-04XBC), The Swedish Research Council (Grant No: K2006-74X-20141-01-3) and the Radiumhemmet Research Funds. Grants from Institut National du Cancer (INCa-PL016) and Ligue Nationale contre le Cancer (PRE2005.LNCC/FD) to E.D., PHRC2007 (AOM-07-195) to M.-E.A. and E.D., grant from Institut National du Cancer (Melanoma Network RS Number 13) to B.B.-deP. This work was developed in the Melanoma Unit of Hospital Clinic I Provincial de Barcelona, IDIBAPS. Ascertainment and data collection in Sydney and the Australian Melanoma Family Study (AMFS) was supported by NHMRC grants 107359, 200071 and 402761 and grants from Cancer Councils of NSW, Victoria and Queensland and the Cancer Institute NSW. J.H. is an Australia Fellow of the NHMRC. This work was supported by the Intramural Program at the National Human Genome Research Institute, National Institutes of Health. Funding was provided by NIH National Cancer Institute grant RO1 CA102422 (to L.A.C.-A.).

The Principal Investigators of Q-MEGA would like to thank D. Statham, M. de Nooyer, I. Gardner and B. Haddon for project management; A. Henders, M. Campbell and M. Stark for managing sample processing and preparation; D. Smyth and H. Beeby for data management and J. Palmer and J. Simmons for ascertainment of clinical records. We also thank the numerous interviewers who collected questionnaire data. Most of all we thank the individuals with melanoma and their families for their cooperation.

The Sydney and AMFS investigators are grateful to all members of the recruitment, data collection and laboratory team, especially C. Watts, R. Dalziel, K. Mahendran, G. St. George and C. Agha-Hamilton and AMFS coordinators J. Maskiell, M. Ferguson and J. Jettan.

This study makes use of data generated by the Wellcome Trust Case Control Consortium. A full list of the investigators who contributed to the generation of the data are available from the Wellcome Trust website. Funding for the project was provided by the Wellcome Trust under award 076113.

The authors thank the EGEA cooperative group for having given access to data on the EGEA (Epidemiological study on the Genetics and Environment of Asthma) study. Biological specimens of the French Family study group were obtained from IGR Biobank.

The authors are grateful for the contributions of P.A. Van Belle to the work of GenoMEL. The authors would also like to thank D. Seminara.

AUTHOR CONTRIBUTIONS

D.T.B. led the analysis group, contributed to the design of the study and was a member of the writing team; F.D. advised on analysis, contributed to the design of the study and was a member of the writing team; M.M.I. and J.H.B. advised on and carried out analyses, contributed to the design of the study and were members of the writing team; M.H. contributed to the design of the study and contributed genotyping information; J.C.T. and E.C. carried out analyses and were members of the writing team; J.R.-M., T.C.-a-W. and D.Z. contributed to the interpretation of genotyping data; J.F.A., M.-F.A., D.C., G.G.-H., V.H., C.I., J.Lang, J.M., N.G.M., G.W.M. and S.N. contributed to the identification of suitable samples for the study; E.A., P.G., M.T.L. and R.M. contributed to the design of the study and contributed to the identification of suitable samples for the study; B.B. contributed to the design of the study and supervised the initial processing of samples; G.B.-S., B.B.-deP., L.A.C.-A., T.D., I.G., J.H., M.H., J.L.H., R.F.K., J.Lubiński, G.J.M., F.A.vanN., H.O., S.P., W.vanW., K.M.B., P.G. and D.E.E. contributed to the design of the study; P.A.K., N.K.H. contributed to the design of the study, contributed to the interpretation of genotyping data and advised the analysis group; M.W. contributed to the design of the sample handling process; A.M.G. and E.M.G. advised on study design and statistical analysis; A.B. contributed genotyping information; N.A.G. contributed to the design of the study and was project deputy lead; G.M.L. contributed to the design of the study and the interpretation of genotyping data; J.A.N.B. led the study design group and was overall project leader.

Published online at <http://www.nature.com/naturegenetics/>.

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>.

- Cannon-Albright, L.A., Bishop, D.T., Goldgar, C. & Skolnick, M.H. Genetic predisposition to cancer. in *Important Advances in Oncology* (eds. DeVita, V.T., Hellman, S. & Rosenberg, S.A.) 39–55 (J. B. Lippincott & Co., Philadelphia, 1991).
- 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).
- Chang, Y.M. *et al.* A pooled analysis of melanocytic nevus phenotype and the risk of cutaneous melanoma at different latitudes. *Int. J. Cancer* **124**, 420–428 (2009).
- Hollenbeak, C.S. *et al.* Increased incidence of melanoma in renal transplantation recipients. *Cancer* **104**, 1962–1967 (2005).
- 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).
- Titus-Ernstoff, L. *et al.* Pigmentary characteristics and moles in relation to melanoma risk. *Int. J. Cancer* **116**, 144–149 (2005).
- 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).
- 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).
- Raimondi, S. *et al.* MC1R variants, melanoma and red hair color phenotype: a meta-analysis. *Int. J. Cancer* **122**, 2753–2760 (2008).
- Kanetsky, P.A. *et al.* Population-based study of natural variation in the melanocortin-1 receptor gene and melanoma. *Cancer Res.* **66**, 9330–9337 (2006).
- Gudbjartsson, D.F. *et al.* ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. *Nat. Genet.* **40**, 886–891 (2008).
- Elwood, J.M. & Jopson, J. Melanoma and sun exposure: an overview of published studies. *Int. J. Cancer* **73**, 198–203 (1997).
- 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).
- 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).
- Goldstein, A.M. *et al.* Features associated with germline *CDKN2A* mutations: a GenoMEL study of melanoma-prone families from three continents. *J. Med. Genet.* **44**, 99–106 (2007).
- Bishop, J.A. *et al.* Genotype/phenotype and penetrance studies in melanoma families with germline *CDKN2A* mutations. *J. Invest. Dermatol.* **114**, 28–33 (2000).
- Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
- Novembre, J. *et al.* Genes mirror geography within Europe. *Nature* **456**, 98–101 (2008).
- Han, J. *et al.* A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet.* **4**, e1000074 (2008).
- Pasmant, E. *et al.* Characterization of a germ-line deletion, including the entire *INK4/ARF* locus, in a melanoma-neural system tumor family: identification of *ANRIL*, an antisense noncoding RNA whose expression coclusters with *ARF*. *Cancer Res.* **67**, 3963–3969 (2007).
- Brown, K.M. *et al.* Common sequence variants on 20q11.22 confer melanoma susceptibility. *Nat. Genet.* **40**, 838–840 (2008).
- Falchi, M. *et al.* Genome-wide association study identifies loci at 9p21 and 22q13 associated with development of cutaneous nevi. *Nat. Genet.* advance online publication, doi:10.1038/ng.410 (5 July 2009).
- Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).

ONLINE METHODS

Study design. The study design is based on a GWAS of melanoma cases and controls, contributed by GenoMEL participating groups. **Supplementary Figure 1** contains an overview of the study. Groups were asked to identify, preferentially, melanoma cases either with a family history (but confirmed as not having a germline *CDKN2A* mutation), multiple primaries or onset before age 40 years. We argued that selecting for these criteria would 'enrich' for individuals with greater genetic predisposition and thereby increase power to identify genetic associations²⁴. Family history was restricted to three cases for families from European populations and four cases for Australian participants to reduce the chance of including individuals with high-penetrance mutations. Controls were recruited by the same research groups from the same populations. Initially, the study design involved minimally 100 cases and 100 controls from each of the 10 centers, but this was expanded subsequently. Genotyping was conducted in two phases. The first phase involved genotyping through ServiceXS in Leiden, The Netherlands, using the Illumina HumanHap300 BeadChip version 2 duo array (with 317K tagging SNPs). Subsequent genotyping of cases was conducted at CNG in Paris using the Illumina Humancnv370k array. Control data were available from two further groups: French controls from CNG genotyped on the Illumina HumanHap300 BeadChip version 2 duo array and UK controls from the WTCCC (described in ref. 17 but based for this analysis on Illumina HumanHap300 BeadChip version 2 duo array genotyping).

We categorized the research groups by their geographical locations, but to enhance power we identified six geographical regions within which the data from individual groups were pooled. These geographical regions were Sweden (Lund and Stockholm), Australia (Brisbane, Sydney and Australian Melanoma Family Study (AMFS) sites), Italy (Genoa and Emilia-Romagna), UK/Leiden, France and Spain. Replication was carried out in a further independent set of enriched cases and controls supplemented by a population-based case-control study conducted in Leeds, UK (**Supplementary Fig. 1**).

Sample sets. The cases and controls were selected from sets of samples accrued by the various contributing groups; these samples were recruited in different ways as described in the relevant publications. The contributing groups for the genome-wide genotyping were Barcelona²⁵, Brisbane (cases²⁶, controls²⁷), Emilia-Romagna²⁸, Genoa²⁹, Leiden, Leeds^{16,30}, Lund³¹, Paris³², Stockholm³³ and Sydney^{34,35}. The samples from Sydney included population-based early-onset cases and matching controls from the Australian Melanoma Family Study²¹ (AMFS, recruited in Sydney, Brisbane and Melbourne). Genotyping within the replication studies included samples from Poland³⁶, Slovenia³⁷ and Tel Aviv. Ethical consent was obtained from all participants.

The Leeds-based case-control study recruited 1,274 population-based incident melanoma cases diagnosed between September 2000 and December 2006 from a geographically defined area of Yorkshire and the Northern region of the UK (63% response rate). Cases were identified by clinicians, pathology registers and via the Northern and Yorkshire Cancer Registry and Information Service to ensure overall ascertainment. For all but 18 months of the study period, recruitment was restricted to individuals with Breslow thickness of at least 0.75 mm. Some of these cases were genotyped genome-wide if they satisfied the criteria for case selection as described in Online Methods; the remainder (1,163 cases) were genotyped in the replication set.

Controls were ascertained by contacting general practitioners to identify eligible individuals. These controls were frequency-matched with cases for age and sex from general practitioners who had also had cases as part of their patient register. Overall there was a 55% response for controls (496 subjects); a subset of these controls was genotyped genome-wide and hence is excluded from the replication set. Controls were supplemented by a population-based group of 574 women who, following informed consent, agreed to participate in a study recording their history of sun exposure and sun bathing and for whom various measures of skin aging were recorded. DNA samples were also provided. A total of 903 controls was genotyped in the replication phase.

Sample handling and DNA preparation. Genome-wide genotyping for this study was conducted in Leiden by ServiceXS (SXS in **Supplementary Table 1**) and in Paris by Centre National de Génotypage (CNG in **Supplementary Table 1**) (**Supplementary Fig. 1**). Samples genotyped at ServiceXS were

processed by the Department of Human and Clinical Genetics, Leiden University Medical Centre (LUMC). Sample lists were provided by the contributing GenoMEL centres and sample tubes and barcodes were returned to the centers from LUMC. Samples returned from the GenoMEL centers were then cross-checked against the manifest list. At the same time, phenotypic information was sent to the analysis center in Leeds. Processed samples were then checked for quality control by performing a single PCR test and examining DNA concentration and quantities; a round of sample replacement proceeded for samples that were considered to have failed quality control. A similar process applied to samples genotyped at CNG.

Quality control was performed for all samples before processing on the Illumina arrays. DNAs with a minimum concentration of 30 ng/μl (measured by NanoDrop(r) ND-1000) were subjected to a test PCR and checked on an agarose gel.

Replication genotyping was carried out using Taqman SNP genotyping assay (Applied Biosystems), with the exception of rs4911442, which was genotyped using the KASPar competitive allele-specific PCR system (KBiosciences) for the Leeds case-control study.

Quality control. Genotypes were called using the proprietary software supplied by Illumina (BeadStudio, version 3.2), with imported cluster centers based on HapMap samples (supplied by Illumina) and call threshold set at 0.15 as recommended by Illumina. Initially, no SNP was excluded from the analysis on the basis of quality control, but SNPs showing some evidence of association were screened intensively for signs of poor quality control. Some problems with poor chip quality were identified, and where possible samples with low (<97%) call rates were re-genotyped.

Sample exclusions. Samples were excluded for any of the following reasons: (i) call rate of less than 97% (of the total number of SNPs on the array); (ii) evidence of non-European origin from principal components analysis (PCA); (iii) sex as inferred from genotyping not matching reported sex; (iv) evidence of first-degree relationship or genetic identity with another sample. Sex was ascertained from genotype data by calculating the heterozygosity rate on the 9035 X-chromosome markers within Beadstudio; persons with >10% heterozygosity were classified as female, otherwise male. Relationship analysis was carried out in PLINK³⁸ using estimated identity-by-descent sharing.

SNP quality control. SNP quality control was assessed by considering a range of measures: (i) test of Hardy-Weinberg equilibrium for the different control groups; (ii) call rate; (iii) minor allele frequency (MAF); (iv) differences in MAF between genotyping centers; (v) test for homogeneity of ORs across regions; (vi) concordance of results with neighboring SNPs; (vii) where possible review of clusters from Bead Studio. Rather than using fixed thresholds, we considered these measures for all SNPs showing evidence of association at 10^{-5} . We had no formal exclusion quality control criteria for SNPs; most SNPs either clearly failed quality control on multiple measures or raised no concern on any measure.

Principal components analysis and population stratification. To identify individuals of non-European ancestry, we thinned the SNPs to reduce linkage disequilibrium (LD) to a set of 67,315 SNPs such that no pair had $r^2 > 0.2$. The data were then combined with the HapMap data from 270 individuals of European, Asian and African origin. Before thinning, a quality control step was implemented; SNPs with call rate less than 97% were excluded as were SNPs with a Hardy-Weinberg P value < 0.0001 in any case-by-genotyping-laboratory (SXS or CNG) or control-by-genotyping-laboratory analysis. PCA was applied to the remaining SNPs using EIGENSTRAT^{39,40} with the first two principal components (PCs) clearly separating the HapMap data into three distinct clusters according to ancestry. Most of our samples were clustered with the HapMap European samples (**Supplementary Fig. 2b**); those that were not were excluded from subsequent analysis. The first two PCs captured 72% of the variation in the first 20 PCs (and 83% of the variation in the first 10 PCs).

The remaining European ancestry-only data were analyzed similarly by PCA implemented in EIGENSTRAT. After applying quality control, we thinned the SNPs such that pairwise r^2 never exceeded 0.5 (167,517 SNPs). Despite the fact that PCA makes no use of the geographical origin of the samples, plotting the

first two PCs clearly grouped the samples by the center from which they were collected, with little overlap between centers (**Supplementary Fig. 2a**). The first PC corresponded roughly to latitude and the second PC to longitude as has been predicted elsewhere¹⁸. The third PC separated those groups from the NE or SW from those groups from the NW or SE (also predicted¹⁸). This demonstrated that most of the individuals had ancestry reflective of the regions they had been collected from, although a few were clearly outliers from elsewhere in Europe.

For instance, although the majority of the Australian samples were clustered with the UK samples, several appeared to be more similar to the Italian samples, as did several of the French samples, suggesting that some had a southern European origin. Here the first three PCs capture 77% of the variation in the first 20 PCs (87% of the variation from the first 20 PCs is in the first 10). We were thus able to confirm that our groupings were roughly correct and we could account for stratification within Europe by adjusting for the first two or three PCs. Furthermore, by examining the loadings on the PCs, we could identify particular SNPs or regions with particularly strong loadings suggesting that these showed considerable variation across Europe. The regions with notably high loading were around *LCT*, *OCA2* and *HLA*.

Association analysis. The primary analysis was a Cochran-Armitage (CA) trend test, stratified by region (see Study design, above). In addition, we carried out an unstratified CA trend test. Two further analyses were conducted to assist in interpretation of results: logistic regression analysis adjusted for region and the first three PCs (to adjust for any residual within-region population stratification) and a CA trend test stratified by study group defined by geographical center and genotyping laboratory. Equivalent 1-degree-of-freedom stratified and unstratified trend tests were carried out for the X chromosome, in which males are treated as equivalent to homozygous females and a variance estimate that allows for the different variance of male and female contributions is used⁴¹.

Subgroup analyses of the different subtypes of case against all controls were also conducted for the most interesting SNPs.

Imputation. Imputation of ungenotyped SNPs was conducted using IMPUTE⁴², which predicts the genotypes of unobserved SNPs by means of a hidden Markov model using the genotype data at our observed markers and a set of known haplotypes (in this case the HapMap European samples). As the method is computationally intensive, we applied it only to those regions in which at least one SNP reached a *P* value of $<10^{-5}$ in either the CA trend test, the test stratified by group or the test stratified by region and further SNPs within 50 kb supported this result. We imputed 1 Mb on either side of any SNPs that reached the required *P* value as well as a 250-kb buffer on either side to avoid end effects. We assumed an effective population size of 11,400 and used a calling threshold of 0.9. We then applied both the CA trend test and a logistic regression adjusting for region.

Multiple regression analysis. Stepwise logistic regression was performed to identify those SNPs independently associated with melanoma within each of the validated regions (chromosomes 9, 11 and 16). We identified all SNPs that passed the quality control for imputation within the genome-wide analysis and that were within 1 Mb of the SNP giving the strongest evidence within each region. (The region covered on chromosome 16 is 1.4 Mb long because it is close to the telomere). Any SNP not associated with melanoma ($P > 0.01$) was eliminated, and SNPs were thinned such that no pair had $r^2 > 0.8$. This left 50 SNPs from a total set of 480. The model assumed an additive effect on the logistic scale at the locus of interest, and analyses were adjusted for geographical region (as previously defined). Chromosomal regions were analyzed separately and in combination, and the analysis either included every individual or was restricted to individuals with complete genotyping at the markers being analyzed. At each step of the forward-selection procedure, the global significance of the model was evaluated, as was the significance of the new marker. The criteria for accepting a new marker were $P < 0.001$ for each marker included in the model from the likelihood ratio test and a decrement in the global *P* value of the model. Of the 5,456 individuals in the dataset, 4,959 (90.9%) had complete genotyping. The stepwise regression was conducted using Stata. Including individuals without complete data or considering regions

separately made no qualitative difference to the results. Therefore, results are presented for analyses restricted to those with complete genotyping and a combined analysis of all three genetic regions.

Case category analysis. Cases were classified as having a family history if at least one first-degree relative had melanoma; individuals with multiple primary melanoma but no family history were classified in the 'multiple melanoma' category, and cases with onset before age 40 without either a family history or multiple primary melanoma were classified as 'early onset' (**Supplementary Table 1**). Using these categories, we performed classical logistic regression adjusted for geographical region to estimate ORs for a given case category (for example, cases with family history versus controls) and to test for homogeneity of the SNP effect across regions using a likelihood ratio test. The homogeneity of ORs by case category was tested using trichotomous regression (for example, controls, cases with family history, cases without family history) and a likelihood-ratio test with 1 d.f. (that is, comparing a model where the ORs with and without family history are equal to a model where the two ORs are estimated).

Power to detect associations. We calculated the power to detect disease-associated regions in our initial genome-wide analysis at a *P* value of 5×10^{-7} and a sample size of 1,539 cases and 3,917 controls. We used effect sizes estimated in the replication stage in order to avoid bias, although this may be conservative since the cases from Leeds were population based and not selected as other case series. We assumed the marker being tested was in complete LD with the causative locus and that the baseline risk of disease was 0.05.

Our calculations show that we have 97% and 84% power to detect the most significant SNPs on chromosomes 16 and 11, respectively. However, we have very low power ($\sim 1\%$) to detect the region on chromosome 9. Although power is good ($>95\%$) to detect the previously identified region found on chromosome 20 (ref. 21), the low frequency led to it not being tagged by our array. Considering the range of values we have the power to detect, we have at least 80% power to detect any SNP with an OR of 1.6 if $MAF > 0.05$ (assuming sufficient coverage), an OR of 1.5 if $MAF > 0.08$, an OR of 1.4 if $MAF > 0.12$ or an OR of 1.3 if $MAF > 0.25$. The lowest OR we have 80% to detect is 1.27 (when $MAF = 0.5$). Thus, it is unlikely that we have missed any common variants with effect sizes > 1.5 (unless coverage is poor in these areas). The study is not well powered to detect effects below 1.27, and thus there may well be other regions with SNPs of similar effect to those we found on chromosome 9 still to be discovered.

Replication analysis and Israeli samples. The genome-wide analysis involved European populations, and PCA was used to identify participants apparently not of European ancestry. Within the replication set, all samples were again derived from European populations with the exception of one population from Israel. Without genome-wide analyses to confirm comparability with the European genetic profile, we analyzed the replication data both with and without the Israeli samples. The analyses presented here include the Israeli samples but analyses excluding these samples are not qualitatively different from these.

URLS. GenoMEL, <http://www.genomel.org>; EGEA (Epidemiological study on the Genetics and Environment of Asthma) study, <http://ifr69.vjf.inserm.fr/~egeanet>; WTCCC (Wellcome Trust Case Control Consortium), <http://www.wtccc.org.uk>.

- Antoniou, A.C. & Easton, D.F. Polygenic inheritance of breast cancer: implications for design of association studies. *Genet. Epidemiol.* **25**, 190–202 (2003).
- Puig, S. *et al.* Role of the CDKN2A locus in patients with multiple primary melanomas. *J. Clin. Oncol.* **23**, 3043–3051 (2005).
- 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).
- Whiteman, D.C. *et al.* Combined effects of obesity, acid reflux and smoking on the risk of adenocarcinomas of the oesophagus. *Gut* **57**, 173–180 (2008).
- Landi, M.T. *et al.* Genetic susceptibility in familial melanoma from northeastern Italy. *J. Med. Genet.* **41**, 557–566 (2004).
- Pastorino, L. *et al.* CDKN2A mutations and MC1R variants in Italian patients with single or multiple primary melanoma. *Pigment Cell Melanoma Res.* **21**, 700–709 (2008).
- Newton Bishop, J.A. *et al.* Mutation testing in melanoma families: INK4A, CDK4 and INK4D. *Br. J. Cancer* **80**, 295–300 (1999).



31. Borg, A. *et al.* High frequency of multiple melanomas and breast and pancreas carcinomas in CDKN2A mutation-positive melanoma families. *J. Natl. Cancer Inst.* **92**, 1260–1266 (2000).
32. Chaudru, V. *et al.* Influence of genes, nevi, and sun sensitivity on melanoma risk in a family sample unselected by family history and in melanoma-prone families. *J. Natl. Cancer Inst.* **96**, 785–795 (2004).
33. Platz, A. *et al.* Screening of germline mutations in the CDKN2A and CDKN2B genes in Swedish families with hereditary cutaneous melanoma. *J. Natl. Cancer Inst.* **89**, 697–702 (1997).
34. Holland, E.A. *et al.* Analysis of the p16 gene, CDKN2, in 17 Australian melanoma kindreds. *Oncogene* **11**, 2289–2294 (1995).
35. Holland, E.A., Schmid, H., Kefford, R.F. & Mann, G.J. CDKN2A (P16(INK4a)) and CDK4 mutation analysis in 131 Australian melanoma probands: effect of family history and multiple primary melanomas. *Genes Chromosom. Cancer* **25**, 339–348 (1999).
36. Debniak, T. *et al.* Common variants of DNA repair genes and malignant melanoma. *Eur. J. Cancer* **44**, 110–114 (2008).
37. Peric, B. *et al.* Prevalence of variations in melanoma susceptibility genes among Slovenian melanoma families. *BMC Med. Genet.* **9**, 86 (2008).
38. 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).
39. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
40. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
41. Clayton, D. Testing for association on the X chromosome. *Biostatistics* **9**, 593–600 (2008).
42. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).