Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma

Thirteen common susceptibility loci have been reproducibly associated with cutaneous malignant melanoma (CMM). We report the results of an international 2-stage meta-analysis of CMM genome-wide association studies (GWAS). This meta-analysis combines 11 GWAS (5 previously unpublished) and a further three stage 2 data sets, totaling 15,990 CMM cases and 26,409 controls. Five loci not previously associated with CMM risk reached genome-wide significance ($P < 5 \times 10^{-8}$), as did 2 previously reported but unreplotted loci and all 13 established loci. Newly associated SNPs fall within putative melanocyte regulatory elements, and bioinformatic and expression quantitative trait locus (eQTL) data highlight candidate genes in the associated regions, including one involved in telomere biology.

CMM primarily occurs in fair-skinned individuals; the major host risk factors for CMM include pigmentation phenotypes1–4, number of melanocytic nevi5,6 and a family history of melanoma7. Six population-based GWAS of CMM have been published8–13, identifying 12 regions with associations that reached genome-wide significance. Some of these regions were already established risk loci for melanoma, for example, through candidate gene studies14 (for a review, see ref. 15). A thirteenth region at 1q42.12, tagged by rs3219090 in SLC45A2, was confirmed as genome-wide significant by a recent study ($P = 1.03 \times 10^{-8}$)16. As might be expected for common variants influencing CMM risk, many of these loci contain genes that are implicated in one of the two well-established heritable phenotypes associated with risk for melanoma: pigmentation (SLC45A2, TYR, MC1R and ASIP) and nevus count (CDKN2A-MTAP, PLA2G6 and TERT) (Supplementary Table 1)17. The presence of DNA repair genes such as PARP1 and ATM at two loci suggests a role for DNA maintenance pathways, leaving four loci where the functional mechanism is less clear (ARNT-SETDB1, CASP8, FTO and MX2).

Of particular interest is TERT, which is involved in telomere maintenance; SNPs in this region have been associated with a variety of cancers11,18–22. Further, the DNA repair functions of ATM and PARP1 extend to telomere maintenance and response to telomere damage23,24. Longer telomeres have been associated with higher nevus counts, and it has been proposed that longer telomeres delay the onset of cell senescence, allowing further time for mutations leading to malignancy to occur20,25. There is evidence that longer telomeres increase melanoma risk20,26,27 and that other telomere-related genes are likely involved in the etiology of melanoma, but none of these loci has yet reached genome-wide significance (or even $P < 1 \times 10^{-6}$)28.

In addition, two independent SNPs at 11q13.3 (near CCND1) and 15q13.1 (adjacent to the pigmentation gene OCA2) have been associated previously with melanoma but did not meet the strict requirements for genome-wide significance, either not reaching $P = 5 \times 10^{-8}$ in the initial report or not replicating in additional studies10,11,29. The meta-analysis described here has resolved the status of these two loci as well as identified new melanoma susceptibility loci.

RESULTS

Study overview

We conducted a two-stage genome-wide meta-analysis. Stage 1 consisted of 11 GWAS data sets totaling 12,874 cases and 23,203 controls from Europe, Australia and the United States; this stage included all 6 published CMM GWAS and 5 unpublished ones (Supplementary Table 2). In stage 2, we genotyped 3,116 CMM cases and 3,206 controls from 3 additional data sets (consisting of 1,692 cases and 1,592 controls from Cambridge, UK, 639 cases and 823 controls from Breakthrough Generations, UK, and 785 cases and 791 controls from Athens, Greece; Online Methods) for the most significantly associated SNP from each region reaching $P < 1 \times 10^{-6}$ in stage 1 and included these results in an overall meta-analysis of both stages, totaling 15,990 melanoma cases and 26,409 controls. Details of these studies can be found in the Supplementary Note. Given that the previous single largest melanoma GWAS was of 2,804 cases and 7,618 controls11, the current meta-analysis represents a 4-fold increase in sample size in comparison to previous efforts to identify the genetic determinants of melanoma risk. Unless otherwise indicated, we report the $P$ values from the overall meta-analysis combining the two stages (Supplementary Table 3). Forest plots of the results from the individual GWAS can be found in Supplementary Figure 1.

All stage 1 studies underwent similar quality control procedures and were imputed using the same reference panel, and the results were analyzed in the same way, with the exception of the Harvard and MD Anderson Cancer Center (MDACC) studies (Online Methods). A fixed-effects ($P_{\text{fixed}}$) or random-effects ($P_{\text{random}}$) meta-analysis was conducted as appropriate, depending on the extent of between-study heterogeneity. A total of 9,470,333 imputed variants passed quality control in at least 2 studies, of which 3,253 reached $P_{\text{fixed}} < 1 \times 10^{-6}$ and 2,543 reached $P_{\text{fixed}} < 5 \times 10^{-8}$. For reference, we provide a list of SNPs that reached $P_{\text{fixed}} < 1 \times 10^{-7}$ (or $P_{\text{random}}$, if $I^2$ measure of
heterogeneity was >31% (Supplementary Table 4). The stage 1 meta-analysis genome-wide inflation value (λ) was 1.032, and, as λ increases with sample size, we also adjusted λ to a population of 1,000 cases and 1,000 controls. The resulting λ,000 value of 1.002 suggested minimal inflation. Quantile-quantile plots for the stage 1 meta-analysis and individual GWAS can be found in Supplementary Figures 2 and 3. To further confirm that our results were not influenced by inflation, the stage 1 meta-analysis was repeated correcting for the λ values from individual studies; P values were essentially unchanged (Online Methods and Supplementary Table 3).

All 13 previously reported genome-wide significant loci (most initially identified in one of the studies included here) reached P < 5 × 10−8 in stage 1 (Fig. 1 and Supplementary Table 4). In addition to confirming the two previously reported loci with associations below genome-wide significance at 11q13.3 (rs498136; 89 kb from CDKN2A) and 15q13.1 (rs4778138; in OCA2), we found three previously unreported loci reaching genome-wide significance at 6p22.3, 7p21.1 and 9q31.2 (Fig. 2 and Table 1). SNPs in another 11 regions reached P < 1 × 10−6 (Supplementary Table 3): notably, 3 were close to known telomere-related genes (rs2995264 is in OBFC1 (ref. 31) at 10q24.33, rs11779437 is 1.1 Mb from TERF1 (ref. 32) at 8q13.3, and rs4731207 is 66 kb from POT1 at 7q31.33, in which loss-of-function variants occur in some families with melanoma3,34). Given the importance of telomeres in melanoma, we additionally genotyped two SNPs that did not quite reach our P-value threshold of P < 1 × 10−6 but are close to telomere-related genes (rs12696304 at 3q26.2 (P fixed = 1.6 × 10−5) is 1.1 kb from TERF1, and rs75691080 at 20q13.33 (P fixed = 1.0 × 10−6) is 19.4 kb from RTEL1). In total, 18 SNPs were carried through to stage 2 (Online Methods).

Including the stage 2 results in the overall meta-analysis led to the identification of two new genome-wide significant regions, 2p22.2 and 10q24.33 (Fig. 2, Table 1 and Supplementary Table 3). The stage 2 data also served the purpose of independently confirming with genotype data the meta-analysis results from imputed SNPs. Five SNPs—rs4778138 (OCA2 in 15q13.1), rs498136 (CDKN2A in 11q13.3), and the new rs10739221 (9q31.2), rs6750047 (2p22.2) and rs2995264 (10q24.33) SNPs—all reached P < 0.05 in the genotyped stage 2 samples. We estimated the power to reach P < 0.05 in the stage 2 samples for all SNPs that reached genome-wide significance in the stage 1 meta-analysis (Online Methods and Supplementary Table 5). The association signals for 2p22.2 (Fig. 2) span the 3′ UTR of RMDN2 rs1636744 was well powered (>90%), the probability that all four of these well-powered SNPs would reach P < 0.05 in the analysis of stage 2 data was only 0.916 × 0.736 × 0.787 × 0.955 = 0.51, so it is not surprising that one was not significant. The SNPs at 7p21.1 (rs1636744) and 6p22.3 (rs6914598) did not reach nominal significance in stage 2, but for both SNPs the confidence intervals for the effect estimates overlapped those from the stage 1 meta-analysis.

In terms of heritability, the 13 loci that were genome-wide significant before this meta-analysis explained 16.9% of the familial relative risk (FRR) for CMM, with MCIR explaining 5.3% alone (Online Methods). Including the seven loci confirmed or reported here (2p22.2, 6p22.3, 7p21.1, 9q31.2, 10q24.33, 11q13.3 and 15q13.1) explained an additional 2.3% of FRR. In total, all 20 loci explained 19.2% of the FRR for CMM; this is a conservative estimate given the assumption of a single associated SNP per locus.

We tested all new and known CMM risk loci for association with nevus count or pigmentation (Supplementary Table 1). Aside from the known association between OCA2 and pigmentation, none of the newly identified loci were associated (P > 0.05) with these phenotypes. After confirmation of the loci in the stage 2 analysis, we performed conditional analysis on the stage 1 meta-analysis results to determine whether there were additional association signals within 1 Mb on either side of the top SNP at each locus using Genome-wide Complex Trait Analysis (GCTA) software36 (Online Methods and Supplementary Table 6). This analysis found that, although there were additional SNPs associated with CMM at each locus, for all but chromosomes 7 and 11 the additional signals were not strongly associated with melanoma (P < 1 × 10−7; for more detail, see Supplementary Figs. 4 and 5, and the Supplementary Note). We then conducted a comprehensive bioinformatic assessment of the top SNP from each of the seven new genome-wide significant loci using a range of annotation tools, databases of functional and eQTL resources, and previously published GWAS results (Online Methods and Supplementary Tables 7–9). We applied the same analyses to each locus, but, to limit repetition, where nothing was found for a given resource (for example, the National Human Genome Research Institute (NHGRI) GWAS catalog), we do not explicitly report this.

2p22.2

Although rs6750047 at 2p22.2 was not genome-wide significant in the stage 1 meta-analysis, it reached genome-wide significance (P fixed = 7.0 × 10−9, odds ratio (OR) = 1.10, P = 0.00; Table 1 and Supplementary Table 3) in stage 2 and the overall meta-analysis. The association signals for 2p22.2 (Fig. 2) span the 3′ UTR of RMDN2...
(also known as FAM82A1) and the entirety of the CYP1B1 gene, and as such there is a wealth of bioinformatic annotation for SNPs associated with CMM risk. Considering the 26 SNPs with $P$ values within 2 orders of magnitude of that for rs6750047 at 2p22.2 (Supplementary Tables 7–9), HaploReg37 reported a significant enrichment of strong enhancers in epidermal keratinocytes (4 observed, 0.6 expected; $P = 0.003$). The paired rs162329 and rs162330 SNPs (linkage disequilibrium (LD) $r^2 = 1.0$, 98 bp apart, $P_{\text{fixed}} = 3.91 \times 10^{-6}$, $I^2 = 11.23$) lie

Figure 2 Regional association plots for the new genome-wide significant loci at 2p22.2, 6p22.3, 7p21.1, 9q31.2 and 10q24.33 and the newly confirmed region at 15q13.1 (OCA2). $-\log_{10}(P_{\text{fixed}})$ values for SNPs from the stage 1 meta-analysis of 12,874 cases and 23,203 controls have been plotted against their genomic positions using LocusZoom. The rsID is listed for the peak SNP in each region (purple diamond). The $P$ values and effect sizes for the listed SNPs can be found in Supplementary Table 3. For the remaining SNPs, the color indicates LD ($r^2$) with the peak SNP. Neither rs2995264 at 10q24.33 nor rs6750047 at 2p22.2 were genome-wide significant in stage 1, but they reached this level of significance in the overall meta-analysis. The plot for 11q13.3 (CCND1) can be found in Supplementary Figure 6.
approximately 10 kb upstream of the CYP1B1 transcription start site in a potential enhancer in keratinocytes and other cell types37–40. These two SNPs are eQTLs for CYP1B1 in three independent liver sample sets41,42. In addition, several SNPs, including the peak SNP for 2p22.2 (rs6750047), are strong CYP1B1 eQTLs in lymphoblastic cell lines (LCLs) in the Multiple-Tissue Human Expression Resource37 (MuTHER; P < 5 × 10−7). It is worth noting that the overlap between the liver and LCL eQTLs is incomplete; rs162330 and rs162331 are only weak eQTLs in MuTHER data (P = 0.01). In terms of functional annotation, the most promising SNP near rs6750047 was rs1374191 (Pfixed = 5.4 × 10−5, OR = 1.07, P = 0.00); in addition to being a CYP1B1 eQTL in LCLs (MuTHER, P = 6.9 × 10−8), this SNP is positioned in a stronger enhancer region in multiple cell types, including melanocytes and keratinocytes37–40. In summary, SNPs at 2p22.2 associated with melanoma lie in putative melanocyte and keratinocyte enhancers and are also cross-tissue eQTLs for CYP1B1.

CYP1B1 metabolizes endogenous hormones, having a role in hormone-associated cancers, including breast and prostate cancers (reviewed in ref. 44). CYP1B1 also metabolizes exogenous chemicals, resulting in pro-cancer (for example, polycyclic aromatic hydrocarbons) and anti-cancer (for example, tamoxifen) outcomes44. The former is of interest as CYP1B1 is regulated by ARNT, a gene at the melanoma-associated 1q21 locus32. The CYP1B1 promoter is methylated in melanoma cell lines and tumor samples45. CYP1B1 missense protein variants have been associated with cancers, including squamous cell carcinoma and hormone-associated cancers34,46. Of these variants, only rs1800440 (p.Asn453Ser) was moderately associated with melanoma (Pfixed = 1.83 × 10−5; OR = 0.90, P = 0.00), and it was included in the bioinformatic annotation (Supplementary Tables 7–9). rs1800440 is not in LD with the peak SNP for 2p22.2 from the meta-analysis for CMM risk, rs6750047 (r2 = 0.04), and adjusting for rs6750047 only slightly reduced its association with CMM (P = 4.3 × 10−4; Online Methods). Truncating mutations in CYP1B1 are implicated in primary congenital glaucoma57. As glaucoma cases were used as controls in the contributing Western Australian Melanoma Health Study (WAMHS) melanoma GWAS, we considered the impact of excluding glaucoma cases; the SNP remained associated with CMM with genome-wide significance even after such exclusions (Supplementary Table 10 and Supplementary Note).

6p22.3

rs6914598 (Pfixed = 3.5 × 10−5, OR = 1.11, P2 = 0.00) lies at 6p22.3 in intron 12 of CDKAL1, a gene that modulates the expression of a range of genes, including proinsulin, via tRNA methylthiolation48,49. Bioinformatic assessment of the 35 SNPs with P values within 2 orders of magnitude of that of the peak SNP at 6p22.3, rs6914598, by HaploReg37 indicated that the most functionally interesting SNP was rs7776158 (stage 1 Pfixed = 3.8 × 10−5, P = 0, in complete LD with rs6914598, r2 = 1.0), which lies in a predicted melanocyte enhancer that binds IRF4 (refs. 38,39). IRF4 binding is of interest given the existence of a functional SNP, rs12203592, in the IRF4 gene50 associated with nevus count, skin pigmentation and tanning response51–54.

7p21.1

rs1636744 (Pfixed = 7.1 × 10−9, OR = 1.10, P = 0.00; Fig. 2) is in an intergenic region at 7p21.1 and lies 63 kb from AGR3. rs1636744 is an eQTL for AGR3 in lung tissue (Genotype-Tissue Expression (GTEx) Project, P = 1.6 × 10−6)55,56. AGR3 is a member of the protein disulfide isomerase family, which generates and modifies disulfide bonds during protein folding57. AGR3 expression has been associated with breast cancer risk58 and poor survival in ovarian cancer59. GTEx data confirmed that AGR3 is expressed in human skin samples. The region containing rs1636744 is not conserved in primates (UCSC Genome Browser60), and RegulomeDB30 indicated that there is little functional activity at this SNP. More promising are rs847377 and rs847404, which, in addition to both being AGR3 eQTLs in lung tissue55 and associated with CMM risk (stage 1 Pfixed = 3.89 × 10−5 and 1.72 × 10−7, respectively), are in putative weak enhancers in a range of cell types, including melanocytes and keratinocytes37–40. Adjusting for rs1636744 rendered the associations for rs847377 and rs847404 non-significant (P > 0.6), indicating that these SNPs are tagging a common signal. rs1636744, rs847377 and rs847404 are not eQTLs for AGR3 in sun-exposed skin.

9q31.2

The melanoma-associated variants at 9q31.2, with the peak signal at rs10739221 (overall Pfixed = 7.1 × 10−11, P = 0.00; Fig. 2) are intergenic. The nearest genes are TMEM38B, ZNF462 and the nucleotide excision repair gene RAD23B61. Although bioinformatic annotation did not identify any putative functional SNPs, on the basis of the importance of DNA repair in melanoma, RAD23B is of particular interest. rs10739221 is 635 kb from the leukemia-associated

### Table 1: Genome-wide significant results from a two-stage meta-analysis of GWAS of CMM from Europe, the United States and Australia

| SNP          | Region | Gene            | Minor allele | MAF (min INFO) | Stage 1 meta-analysis | Stage 2 meta-analysis | Overall meta-analysis |
|--------------|--------|-----------------|--------------|----------------|-----------------------|-----------------------|-----------------------|
| rs6750047    | 2p22.2 | RMDN2 (CYP1B1)  | A            | 0.43 (0.96)    | 0.088                 | 0.113                 | 0.092                 |
| rs6914598    | 6p22.3 | CDKAL1          | C            | 0.32 (0.88)    | 0.11                  | 0.26                  | 0.037                 |
| rs1636744    | 7p21.1 | AGR3            | T            | 0.40 (0.96)    | 0.11                  | 1.8 × 10−9            | 0.032                 |
| rs10739221   | 9q31.2 | TMEM38B (RAD23B, TAL2) | T          | 0.24 (0.94)    | 0.12                  | 9.6 × 10−9            | 0.145                 |
| rs2995264    | 10q24.3| OBF1            | G            | 0.088 (0.94)   | 0.14                  | 8.5 × 10−7            | 0.206                 |
| rs498136     | 11q13.3| CCND1           | A            | 0.32 (0.97)    | 0.12                  | 1.0 × 10−10           | 0.124                 |
| rs4778138    | 15q13.1| OCA2            | G            | 0.16 (0.82)    | −0.18                 | 3.1 × 10−9            | −0.156                |

For each region, we report the chromosomal location, nearest gene and any other promising candidate gene in parentheses for the top SNP. We also report the 1000 Genomes Project European population minor allele frequency (MAF) and the minimum imputation quality across all studies (min INFO). The stage 1 meta-analysis field reports the effect size estimate (β) and P value for the minor allele from the meta-analysis of 11 CMM GWAS, totaling 12,874 cases and 23,203 controls. Following the genotyping of these loci in 3 additional data sets (totaling 3,116 cases and 3,206 controls), we provide the stage 2 meta-analysis results. Finally, we provide the overall meta-analysis of all available data. The results are shown for the top SNP in each region that reached P < 1 × 10−6 in stage 1 and so was genotyped in stage 2; per-study results and evidence of heterogeneity of effect estimates across studies (P) can be found in Supplementary Table 3. Where P values were below 31%, fixed-effects meta-analysis was used; otherwise, random-effects meta-analysis was used. All genome-wide significant SNPs had low heterogeneity (I2 < 31%) in both stage 1 and the overall analysis. Regions previously confirmed as associated with melanoma (for example, MC1R) are not shown. As indicated by an asterisk, rs6914598 in 6p22.3 is not genome-wide significant after full multiple-testing correction, for example, P < 3.06 × 10−6 as in Li et al.71.
TAL2 gene and 1.2 Mb from KLF4, which regulates both telomerase activity and the melanoma-associated TERT gene.

10q24.33

Although not genome-wide significant in stage 1, rs2995264 at 10q24.33 is strongly associated with telomere length and was genotyped in stage 2. rs2995264 was significantly associated with CMM in the Cambridge study ($P = 0.046$) and showed strong association in the Breakthrough data set ($P = 8.0 \times 10^{-4}$); in the overall meta-analysis, this SNP reached genome-wide significance ($P_{\text{fixed}} = 2.2 \times 10^{-9}$, $I^2 = 27.14$). The melanoma association signal at 10q24.33 (Fig. 2) spans the OBFC1 gene and the promoter of SH3PD2A. Given the strong association with telomere length at this locus, the most promising candidate is OBFC1, a component of the telomere maintenance complex.

HaploReg reported that SNPs with $P$ values within two orders of magnitude of that for rs2995264 at 10q24.33 were significantly more likely to fall in putative enhancers in keratinocytes than would be expected by chance ($P < 0.001$). Promising candidate functional SNPs included the conserved rs1594668 and rs11191827 markers, which lie in putative melanocyte and keratinocyte enhancers and bind transcription factors. The association observed at rs2995264 in 10q24.33 was independent of a melanoma association at 10q25.1 in a recent report. Our peak SNP for 10q24.33 (rs2995264) and the 10q25.1 SNPs (rs17119434, rs17119461 and rs17119490) reported in Teerlink et al. are not in linkage equilibrium ($I^2 < 0.01$), and, in turn, these SNPs were not associated with CMM in our meta-analysis ($P > 0.2$).

11q13.3

The peak signal for the CMM-associated variants at 11q13.3 was at rs498136 (overall $P_{\text{fixed}} = 1.5 \times 10^{-12}$, OR = 1.13, $I^2 = 0.00$; Supplementary Fig. 6) to the promoter of CCND1. In the initial report of CCND1 (ref. 11), rs11263498 was borderline significant in its association with melanoma ($P = 3.2 \times 10^{-7}$) and, although supported ($P = 0.017$) by the two replication studies, exhibited significant heterogeneity and did not reach genome-wide significance (overall $P_{\text{random}} = 4.6 \times 10^{-4}$, $P = 45.00$). The previously reported rs11263498 SNP and the meta-analysis peak at rs498136 in 11q13.3 are in strong LD ($r^2 = 0.95$).

Bioinformatic assessment of the CCND1 region indicated that the peak SNP at 11q13.3, rs498136, is in a putative enhancer in keratinocytes in both Encyclopedia of DNA Elements (ENCODE) and Roadmap Epigenomics data. Considering other SNPs strongly associated with CMM, both the previously reported rs11263498 (stage 1 $P_{\text{fixed}} = 1.8 \times 10^{-9}$, OR = 1.12, $I^2 = 0.00$) and rs868089 (stage 1 $P_{\text{fixed}} = 2.0 \times 10^{-9}$, OR = 1.12, $I^2 = 0.00$) lie in putative melanocyte enhancers.

Somatic CCND1 amplification in CMM tumors positively correlates with markers of reduced overall survival, including Breslow thickness and ulceration. The CCND1 association with breast cancer has been extensively fine mapped, identifying three independent association signals. The two strongest functional associations with breast cancer but were not themselves associated with CMM risk (stage 1 $P_{\text{fixed}} > 0.1$, $I^2 = 0.00$). Although the third signal in breast cancer was not functionally characterized, its tag SNP rs494406 was modestly associated with CMM (stage 1 $P_{\text{fixed}} > 0.0002$, $I^2 = 0.00$, $r^2 = 0.47$ with rs498136 in 11q13.3). rs494406 was no longer significant after adjustment for rs498136 ($P = 0.53$; Supplementary Table 6), suggesting that SNPs that are in LD in this region are associated with risk of both melanoma and breast cancer.

15q13.1

Both OCA2 and the nearby HERC2 at the 15q13.1 locus have long been associated with pigmentation traits, rs12913832 in HERC2, also known as rs11855019, is the major determinant of eye color in Europeans, making this region a strong candidate for CMM risk. Of the studies contributing to this meta-analysis previously reported a genome-wide significant association between melanoma and rs1129038 and rs1291382 in HERC2 (in strong LD, reported as $r^2 = 0.985$), but this association was not supported ($P > 0.05$) by any of the three replication GWAS (final $P = 2.5 \times 10^{-4}$). Stratification might be an issue for this locus, as eye color frequencies vary markedly across European populations. Indeed, in our meta-analysis, which included all four of these GWAS, both rs1129038 and rs1291382 showed highly heterogeneous effects in the CMM risk meta-analysis ($P_{\text{random}} = 0.037$ and 0.075, respectively; $P > 0.77$).

Amos et al. found that rs4778138 in OCA2, which is only in weak LD with rs1291382 ($r^2 = 0.12$), exhibited more consistent association across studies, albeit not with genome-wide significance. In our overall meta-analysis, we confirmed that rs4778138 at 15q13.1 is associated with CMM risk ($P_{\text{fixed}} = 2.2 \times 10^{-11}$, OR = 0.84, $P = 0.00$; Fig. 2). Following adjustment of the 15q13.1 signal by rs4778138, the effect size for the eye color–associated SNP rs1291382 was reduced from $\beta = 0.12$ to $\beta = 0.064$. Conversely, adjustment for rs1291382 reduced the strength of the association of rs4778138 with CMM ($\beta$ reduced from $-0.178$ to $-0.114$, corrected $P = 1.6 \times 10^{-4}$). rs1291382 was poorly imputed across studies, reaching INFO > 0.8 in only six studies, and we were unable to conclusively exclude a role for rs1291382 at this locus. HaploReg indicated that rs4778138 is within a putative melanocyte enhancer in Roadmap Epigenomics data. Although it is not clear which gene(s) at 15q13.1 are influenced by melanoma-associated SNPs, the fact that rs4778138 is associated with eye colors intermediate to blue and brown supports a role for OCA2.

Evidence of additional melanoma susceptibility loci

A further nine loci were associated with CMM risk at multiple SNPs with $P < 1 \times 10^{-6}$ in stage 1 but did not reach $P < 5 \times 10^{-8}$ in the overall meta-analysis (Supplementary Table 3). Given that genome-wide significance is based on Bonferroni correction assuming 1 million independent tests, we would expect only 1 locus to reach $P < 1 \times 10^{-6}$, and the probability that as many as 9 loci reach this threshold is $1.1 \times 10^{-6}$ (exact binomial probability), so it is highly likely that several of these are genuine.

Of the 16 regions that reached $P < 1 \times 10^{-6}$, 3 were near genes involved in telomere biology at 7q31.3 (rs4731207 near POT1), 8q13.3 (rs11779437 near TERFI) and 10q24.33 (rs2995264 near OBFC1) (Supplementary Table 3). Given the evidence for telomere biology in melanoma and the fact that previous genome-wide significant SNPs are near the telomere maintenance genes TERT, PARP1 and ATM, we included two further biological candidates: rs12696304 located 1.1 kb from TERC at 3q26.2 and rs75691080 at 20q13.33 near RTEL1. Of these five SNPs, rs2995264 (10q24.33 in OBFC1) attained genome-wide significance in the overall analysis, and rs12696304 (3q26.2, near TERC) was significant in stage 2 ($P = 4.0 \times 10^{-5}$) and reached $P = 2.8 \times 10^{-7}$ in the overall meta-analysis (Supplementary Table 3). Although falling short of genome-wide significance, this evidence is nonetheless suggestive of an association at this locus. Neither rs4731207 (66 kb from POT1 at 7q31.33) nor rs75691080 (19.4 kb from RTEL1 at 20q13.33) were significantly associated with melanoma risk in stage 2, but in neither case did the estimated effect differ significantly from that in stage 1. In addition, rs75691080 (20q13.3, near RTEL1) was marginally associated with...
nevoid count ($P = 0.058$; **Supplementary Table 1**). Of the SNPs near telomere-related genes, rs11779437 at 8q13.3 was the most distant (1.1 Mb from *TERF1*) and was the only one to show a significantly different effect in stage 2 (overall $P_{\text{random}} = 0.013$, OR = 0.93, $I^2 = 42.06$). This is most likely due to the initial signal being a false positive but may be due to a lack of power.

**DISCUSSION**

This 2-stage meta-analysis, representing a 4-fold increase in sample size in comparison to the previous largest single melanoma GWAS, has confirmed all 13 previously reported loci as well as resolved 2 likely associations at *CCND1* and *HERC2-OC2A*. The *CCND1* association with melanoma only partially overlaps the signal observed for breast cancer$^{68}$. The *HERC2-OC2A* association is with rs4778138 at 15q13.1, which may be a subtype modifier of eye color$^{70}$, but we cannot rule out the possibility that the association at this locus is influenced by the canonical blue/brown eye color variant rs12913832.

Our stage 1 meta-analysis of over 12,000 melanoma cases identified 3 new risk-associated regions, with only rs10739221 formally significant ($P < 0.05$) in stage 2 (**Table 1**). Two further loci (2p22.2 and 10q24.33) reached genome-wide significance with the addition of the stage 2 data (Fig. 2, **Table 1** and **Supplementary Table 3**). In total, our overall meta-analysis identified 20 genome-wide significant loci: 13 previously replicated, 2 previously reported but first confirmed here, and 5 new to this report. The new loci identified in this meta-analysis explain an additional 2.3% of the FR for CMM. Overall, 19.2% of the FR is explained by all 20 genome-wide significant loci combined.

Except for the association at 9q31.2, the reported loci contain SNPs that are both strongly associated with melanoma and fall within putative regulatory elements in keratinocyte or melanocyte cells, with the nearby nucleotide excision repair gene *RAD23B* at 9q31.2 identified as a promising candidate. eQTL data sets suggest that melanoma-associated SNPs at 7p21 regulate the expression of *AGR3*, albeit in lung tissue and not sun-exposed skin. *AGR3* expression has been implicated in breast and ovarian cancer outcome. SNPs at 2p22.3 are associated with the expression of *CYP1B1*. Although this gene is better known for its role in hormone-associated cancers, it may influence melanoma risk through metabolism of exogenous compounds, a process regulated by *ARNT* at the 1q21 melanoma-associated locus.

We have used the power of this large collection of CMM cases and controls to identify five new susceptibility loci, none of which are significantly associated with classical CMM risk factors and thus highlight new disease pathways. Interestingly, we now have genome-wide significant evidence for association between CMM risk and a SNP in the telomere-related gene *OBFC1* at 10q24.33, in addition to the established telomere-related associations at the *TERT-CLPTM1L*, *PARP1* and *ATM* loci. We also have support, albeit not genome-wide significant, for *TERC*, the most significant predictor of leukocyte telomere length in a recent study$^{75}$. Of the 20 loci that now reach genome-wide significance for CMM risk, 5 are in regions known to be related to pigmentation, 3 are in nevoid-related regions and 4 are in regions related to telomere maintenance. This gives further evidence that the telomere pathway, with its effect on the growth and senescence of cells, may be important in understanding the development of melanoma.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**ACKNOWLEDGMENTS**

Please see the **Supplementary Note** for acknowledgments.

**AUTHOR CONTRIBUTIONS**

M.M.I. and M.H.I. led, designed and carried out the statistical analyses and wrote the manuscript. M. Harland was involved in the Leeds genotyping design. J.C.T. carried out statistical analyses. J.R.-M. and N.v.d.S. carried out genotyping and contributed to the interpretation of genotyping data. J.A.N.B. led the GenoMEL Consortium and contributed to study design. N.A.G. was deputy lead of the consortium and contributed to study design. S.M., N.K.H., D.T.B. and J.H.B. designed and led the overall study. J. Han supervised and carried out statistical analysis of the Harvard GWAS data. F.S. and A.A.Q. carried out statistical analysis of the Harvard GWAS data. C.L.A. led and carried out statistical analysis of the MD Anderson GWAS data. W.V.C., J.E.L. and S.F. contributed to the analysis and interpretation of the MD Anderson GWAS data. F.D. led, designed and contributed to the sample collection, analysis and interpretation of the French MELARISK GWAS and advised on the overall statistical analysis. M.B. contributed to the analysis and interpretation of the French MELARISK GWAS data. M.-F.A. led, designed and contributed to the sample collection of the French MELARISK GWAS. G.M.I. led and contributed to the genotyping and interpretation in the French MELARISK GWAS. R.K. and D.S. led and contributed to the sample collection and analysis for the Heidelberg data set. H.-J.S. contributed to the sample collection and analysis for the Heidelberg data set. S.V.W. led and contributed to the sample collection for the WAMHS study. E.K.M. provided coordination and oversight for the WAMHS study. D.C.W. led, designed and contributed to the sample collection for the SDH data set. J.E.C. led and designed the Glaucoma study. K.P.B. contributed to the analysis and interpretation of the Glaucoma data set. G.L.R.-S. led and contributed to the analysis and interpretation of the IBD data set. I.A.S. contributed to the analysis and interpretation of the IBD data set. G.I.M. led and contributed to the sample collection, analysis and interpretation of the AMFS study. A.E.C. contributed to the sample collection, analysis and interpretation of the AMFS study. D.R.N. contributed to the sample collection and analysis of the Q-MEGA, Endometriosis and QTWIN data sets. N.G.M. led and contributed to the sample collection and analysis for the Q-MEGA, Endometriosis and QTWIN data sets. D.L.D. contributed to the sample collection and analysis for the Q-MEGA, Endometriosis and QTWIN data sets.

**URLs**

GenoMEL, [http://www.genomel.org/](http://www.genomel.org/); Wellcome Trust Case Control Consortium, [http://www.wtccc.org.uk/](http://www.wtccc.org.uk/); RegulomeDB, [http://RegulomeDB.org/](http://RegulomeDB.org/); HaploReg, [http://www.broadinstitute.org/mammals/haploreg/](http://www.broadinstitute.org/mammals/haploreg/); GTEx, [http://www.gtexportal.org/](http://www.gtexportal.org/); MuTHER, [http://www.muther.ac.uk/](http://www.muther.ac.uk/); eQTL data accessed via Genevar, [http://www.sanger.ac.uk/resources/software/genevar/](http://www.sanger.ac.uk/resources/software/genevar/); eQTL Browser, [http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/](http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/); National Human Genome Research Institute (NHGRI) GWAS catalog, [http://www.genome.gov/gwastudies](http://www.genome.gov/gwastudies); Genome-wide Complex Trait Analysis (GCTA), [http://www.complextraitgenomics.com/software/gcta/](http://www.complextraitgenomics.com/software/gcta/); GTOOL, [http://www.well.ox.ac.uk/~c-freeman/software/gwas/gtool.html](http://www.well.ox.ac.uk/~c-freeman/software/gwas/gtool.html); Cancer Oncological Gene-environment Study (COGS), [http://www.nature.com/icogs/primer/common-variation-and-heritability-estimates-for-breast-ovarian-and-prostate-cancers/#/](http://www.nature.com/icogs/primer/common-variation-and-heritability-estimates-for-breast-ovarian-and-prostate-cancers/#/); VCFtools, [http://vcftools.sourceforge.net/](http://vcftools.sourceforge.net/); R scripts for Manhattan and quantile-quantile plots, [http://getoingeneticsdone.blogspot.com.au/2011/04/annotated-melanoma-plots-and-qq-plots.html](http://getoingeneticsdone.blogspot.com.au/2011/04/annotated-melanoma-plots-and-qq-plots.html); rmeta R package for forest plots, [https://cran.r-project.org/web/packages/rmeta/rmeta.pdf](https://cran.r-project.org/web/packages/rmeta/rmeta.pdf).
COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

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

1. 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).
2. Holly, E.A., Aston, D.A., Cress, R.D., Ahn, D.K. & Kristiansen, J.J. Cutaneous melanoma in women. II. Exposure to sunlight, ability to tan, and other risk factors related to ultraviolet light. Am. J. Epidemiol. 141, 923–933 (1995).
3. Naldi, L. et al. Cutaneous malignant melanoma in women: Phenotypic characteristics, sun exposure, and familial factors: a case-control study from Italy. Ann. Epidemiol. 15, 545–550 (2005).
4. Titus-Ernstoff, L. et al. Pigmentary characteristics and moles in relation to melanoma risk. Int. J. Cancer 116, 144–149 (2005).
5. 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).
6. 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, 420–429 (2009).
7. Cannon-Albright, L.A., Bishop, D.T., Goldgar, C. & Skolnick, M.H. Genetic predisposition to cancer. Important Adv. Oncol. 1991, 39–55 (1991).
8. Brown, K.M. et al. Genetic variants in 20q11.22 confer melanoma susceptibility. Nat. Genet. 40, 838–840 (2008).
9. Bishop, D.T. et al. Genome-wide association study identifies three loci associated with melanoma risk. Nat. Genet. 41, 920–925 (2009).
10. Amos, C.I. et al. Genome-wide association studies of melanoma in women. J. Natl. Cancer Inst. 100, 1519–1527 (2008).
11. Barrett, J.H. et al. Genome-wide association study identifies three new melanoma susceptibility loci. Nat. Genet. 43, 1108–1113 (2011).
12. Macgregor, S.G. et al. Genome-wide association study identifies a new melanoma susceptibility locus at 1q21.3. Nat. Genet. 43, 1114–1118 (2011).
13. Iles, M.M. et al. A variant in FTO shows association with melanoma risk not due to BMI. Nat. Genet. 45, 428–432 (2013).
14. Guillaumet-Longobardo, D. et al. A putative TERT promoter variant associates with cutaneous melanoma and basal cell carcinoma. Hum. Mol. Genet. 20, 5012–5023 (2011).
15. Barrett, J.H. et al. Genome-wide association study identifies three new melanoma susceptibility loci. Nat. Genet. 43, 1108–1113 (2011).
16. Peña-Chilet, M. et al. Genetic variants in PARP1 (rs3219090) and IRF4 (rs12203592) genes associated with melanoma susceptibility in a Spanish population. BMC Cancer 13, 160 (2013).
17. Falchi, M. et al. Genome-wide association study identifies variants at 9p21 and 22q13 associated with development of cutaneous nevi. J. Invest. Dermatol. 135, 1074–1079 (2015).
18. De Bakker, P.I. et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum. Mol. Genet. 17, R122–R128 (2008).
19. Miyake, Y. et al. RPA-like mammalian Ctc1-Stn1-Ten1 complex binds to single-stranded DNA and protects telomeres independently of the Pot1 pathway. Mol. Cell 36, 193–206 (2009).
20. van Steenbergh, B. & de Lange, T. Control of telomere length by the human telomeric protein TRF1. Nature 385, 743–747 (1997).
21. Robles-Espinoza, C.D. et al. POT1 loss-of-function variants predispose to familial melanoma. Nat. Genet. 46, 478–481 (2014).
22. Shi, J. et al. Rare missense variants in POT1 predispose to familial cutaneous malignant melanoma. Nat. Genet. 46, 482–486 (2014).
23. Codd, V. et al. Identification of seven loci affecting mean telomere length and their association with disease. Nat. Genet. 45, 422–427 (2013).
24. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat. Genet. 44, 369–375 (2012).
25. Ward, L.D. & Kellis, M. HaplReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 40, D930–D934 (2012).
26. Bernstein, B.E. et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat. Biotechnol. 28, 1045–1048 (2010).
27. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).
28. Boyle, A.P. et al. Renotation of functional variation in personal genomes using RegulomeDB. Genome Res. 22, 1790–1792 (2012).
29. Schadt, E.E. et al. Mapping the genetic architecture of human gene expression. PLoS Biol. 6, e1002078 (2011).
30. Grundberg, E. et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat. Genet. 44, 427–432 (2012).
31. Gajjar, K. et al. Mapping cardiovascular disease using fine-mapping of genetic association studies in cutaneous melanoma: the MelGene database. Tumour Biol. 35, 3891–3897 (2014).
32. Stollow, I., Akarsu, A.N. & Sarfarazi, M. Identification of three different truncating mutations in cytochrome P4501B1 (CYP1B1) as the principal cause of primary congenital glaucoma (Buphthalmos) in families linked to the GLC3A locus on chromosome 2p21. Hum. Mol. Genet. 6, 641–647 (1997).
33. Arrigan, S. et al. Identification of X-chromosome-linked PRR5 that influences hair color and pigmentation traits in European Americans. Hum. Mol. Genet. 15, 2294–2299 (2013).
34. Doherty, J.J. et al. The anterior gradient homolog 3 (AG3) is a tail-anchored protein in the endoplasmic reticulum (ER) of insulinoma cells. J. Biol. Chem. 287, 4108–4118 (2012).
35. Przeworski, M. et al. Diversity of the protein disulfide isomerase family: identification of breast tumor induced Hag2 and Hag3 as novel members of the protein family. Mol. Phylogenet. Evol. 36, 734–740 (2005).
36. Fletcher, G.C. et al. Hag2 and Hag3, human homologues of genes involved in differentiation, are associated with oestrogen receptor-positive breast tumours and interact with metastasis gene C4.4a and Drasyligncan. Br. J. Cancer 88, 579–585 (2003).
37. Brown, S.J. et al. The anterior gradient homolog 3 (AG3) gene is associated with differentiation and survival in ovarian cancer. Am. J. Surg. Pathol. 33, 904–912 (2011).
38. Kent, W.J. et al. The human genome browser at UCSC. Genome Res. 12, 996–1006 (2002).
39. del Maestro, F.R. et al. Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23. EMBO J. 13, 1831–1843 (1994).
40. Xia, X. et al. TAL2, a helix-loop-helix gene activated by the (7;12)(q26;p12) translocation in human T-cell leukemia. Proc. Natl. Acad. Sci. USA 98, 11416–11420 (1991).
41. Wong, C.W. et al. Kruppel-like transcription factor 4 contributes to maintenance of telomerase activity in stem cells. Stem Cells 28, 1510–1517 (2010).
42. Hentzoumeny, K. et al. Wnt3a-beta-catenin signaling regulates telomerase in stem cells and cancer cells. Science 336, 1549–1554 (2012).
43. Teerlinck, C. et al. A unique genome-wide association analysis in extended Utah high-risk pedigrees identifies a novel melanoma risk variant on chromosome arm 10q. Hum. Genet. 131, 77–85 (2012).
Matthew H Law1,6, D Timothy Bishop2,6, Jeffrey E Lee3,6, Myriam Brossard4,5,6, Nicholas G Martin6, Eric K Moses7, Fengju Song8, Jennifer H Barrett9, Rajiv Kumar9, Douglas F Easton10, Paul D P Pharao11, Anthony J Swedlow12,13, Katerina P Kypreou14, John C Taylor2, Mark Harland2, Juliette Randerson-Moor2, Lars A Akslen15,16, Per A Andreisen17, Marie-Francoise Avril18, Esther Azizi19,20, Giovanna Bianchi Scarrà21,22, Kevin M Brown23, Tadeusz Dekniak24, David L Duffy6, David E Elder25, Shening Fang3, Eitan Friedman20, Pilar Galan26, Paola Ghiorzo21,22, Elizabeth M Gillanders27, Alisa M Goldstein23, Nelleke A Gruis28, Johan Hansson29, Per Helsing30, Marko Hočevar31, Veronica Höiöm29, Christian Ingvar32, Peter A Kanetsky33, Wei V Chen34, GenoMELOnso155, Essen-Heidelberg Investigators35, The SHD Study Group35, Q-MEGA and QTWIN Investigators35, AMFS Investigators35, ATHENS Melanoma Study Group35, Maria Teresa Landi23, Julie Lang36, G Mark Lathrop37, Jan Lubinski38, Rona M Mackie36,38, Graham J Mann39, Anders Molven40, Grant W Montgomery41, Srdjan Novaković42, Håkan Olsson43,44, Susana Puig45,46, Joan Anton Puig-Butille45,46, Abrar A Qureshi47, Graham I. Radford-Smith48,50, Nienke van der Steep51, Remco van Doorn52, David C Whiteman52, Jamie E Craig53, Dirk Schadendorf54,55, Lisa A Simms48, Kathryn P Burdon56, Dale R Nyholt41,57, Karen A Pooley58, Alexander J Stratigos14, Anne E Curst59, Sarah V Ward7, Nicholas K Hayward60, Jiali Han61,62, Hans-Joachim Schulze63, Alison M Dunning11, Julia A Newton Bishop2, Florence Demenais4,5,6, Christopher I Amos64,65, Stuart MacGregor1,66 and Mark M Iles1,66

1Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. 2Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK. 3Department of Surgical Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas, USA. 4INSERM, UMR 946, Genetic Variation and Human Diseases Unit, Paris, France. 5Institut Universitaire d’Hématologie, Université Paris Diderot, Sorbonne Paris Cité, Paris, France. 6Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. 7Centre for Genetic Origins of Health and Disease, Faculty of Medicine, Dentistry and Health Sciences, University of Western Australia, Perth, Western Australia, Australia. 8Department of Epidemiology and Biostatistics, Key Laboratory of Cancer Prevention and Therapy, Tianjin, National Clinical Research Center of Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China. 9Division of Molecular Genetic Epidemiology, German Cancer Research Center, Heidelberg, Germany. 10Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. 11Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK. 12Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK. 13Division of Breast Cancer Research, The Institute of Cancer Research, London, UK. 14Department of Dermatology, University of Athens School of Medicine, Andreas Sygros Hospital, Athens, Greece. 15Centre for Cancer Biomarkers (CCBIO), Department of Clinical Medicine, University of Bergen, Bergen, Norway. 16Department of Pathology, Haukeland University Hospital, Bergen, Norway. 17Department of Pathology, Molecular Pathology, Oslo University Hospital, Rikshospitalet, Oslo, Norway. 18 Assistance Publique–Hôpitaux de Paris, Hôpital Cochin, Service de Dermatologie, Université Paris Descartes, Paris, France. 19Department of Dermatology, Sheba Medical Center, Tel Hashomer, Sackler Faculty of Medicine, Tel Aviv, Israel. 20Oncogenetics Unit, Sheba Medical Center, Tel Hashomer, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel. 21Department of Internal Medicine and Medical Specialties, University of Genoa, Genoa, Italy. 22Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health, Bethesda, Maryland, USA. 23International Hereditary and Laboratory Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. 24Université Paris 13, Equipe de Recherche en Épidémiologie Nutritionnelle (EREN), Centre de Recherche en Épidémiologie et Statistiques, INSERM U1153, Institut National de la Recherche Agronomique (INRA) U1125, Conservatoire National des Arts et Métiers, Communauté d’Université Sorbonne Paris Cité, Bobigny, France. 25Inherited Disease Research Branch, National Human Genome Research Institute, US National Institutes of Health, Baltimore, Maryland, USA. 26Department of Dermatology, Leiden University Medical Center, Leiden, the Netherlands. 27Department of Oncology-Pathology, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden. 28Department of Dermatology, Oslo University Hospital, Rikshospitalet, Oslo, Norway. 29Department of Surgical Oncology, Institute of Oncology Ljubljana, Ljubljana, Slovenia. 30Department of Surgery, Clinical Sciences, Lund University, Lund, Sweden. 31Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA. 32Department of Genetics, University of Texas MD Anderson Cancer Center, Houston, Texas, USA. 33A full list of members and affiliations appears in the Supplementary Note. 34Department of Medical Genetics, University of Glasgow, Glasgow, UK. 35McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada. 36Department of Public Health, University of Glasgow, Glasgow, UK. 37Centre for Cancer Research, University of Sydney at Westmead, Millennium Institute for Medical Research and Melanoma Institute Australia, Sydney, New South Wales, Australia. 38Gade Laboratory for Pathology, Department of Clinical Medicine, University of Bergen, Bergen, Norway. 39Molecular Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. 40Department of Molecular Diagnostics, Institute of Oncology Ljubljana, Ljubljana, Slovenia. 41Department of Oncology/Pathology, Clinical Sciences, Lund University, Lund, Sweden. 42Department of Cancer Epidemiology, Clinical Sciences, Lund University, Lund, Sweden. 43Department of Dermatology, Stanford University School of Medicine, Stanford, California, USA. 44Department of Dermatology, University of California, San Francisco, California, USA. 45Department of Dermatology, Stanford University School of Medicine, Stanford, California, USA. 46Centre de Investigación Biomédica en Red (CIBER) de Enfermedades Raras, Instituto de Salud Carlos III, Madrid, Spain. 47Department of Dermatology, Warren Alpert Medical School of Brown University, Providence, Rhode Island, USA. 48Inflammatory Bowel Diseases, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. 49Department of Gastroenterology and Hepatology, Royal Brisbane and Women’s Hospital, Brisbane, Queensland, Australia. 50University of Queensland School of Medicine, Herston Campus, Brisbane, Queensland, Australia. 51Department of Clinical Genetics, Center of Human and Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands. 52Cancer Control Group, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia.
These authors contributed equally to this work.
These authors jointly supervised this work. Correspondence should be addressed to M.H.L. (matthew.law@qimrberghofer.edu.au) or M.M.I. (m.m.iles@leeds.ac.uk).
ONLINE METHODS

Stage 1 array genotyping. The samples were genotyped on a variety of commercial arrays, as detailed in the Supplementary Note. All samples were collected with informed consent and with ethical approval (for details, see the Supplementary Note).

Stage 1 genome-wide imputation. Imputation was conducted across the genome, separately for each study, following a shared protocol. SNPs with MAF < 0.03 (MAF < 0.01 in AMFS, Q-MEGA_omni, Q-MEGA_610k, WAMHS and HEIDELBERG), control Hardy-Weinberg equilibrium $P < 1 \times 10^{-4}$ or missingness $> 0.03$ were excluded, as were any individuals who had call rates $<0.97$, were identified as first-degree relatives and/or were European-ancestry outliers by principal-components analysis using EIGENSTRAT$^{22}$. In addition, in each study where genotyping was conducted on more than one chip, any SNP not present on all the chips was removed before imputation to avoid bias. IMPUTE (v2.2)$^{23,24}$ was used for imputation for all studies but Harvard, which used MacH$^{25,26}$, and MDACC, which used MacH and Minimac$^{27}$. For GenoMEL, CIDRUK and MDACC samples, 1000 Genomes Project February 2012 data (build 37) were used as the reference panel, whereas, for the AMFS, Q-MEGA_omni, Q-MEGA_610k, WAMHS, MELARISK and HEIDELBERG data sets, 1000 Genomes Project April 2012 data (build 37) were the reference for imputation$^{25}$. In both cases, any SNP with MAF $<0.001$ in European (CEU) samples was dropped from the reference panel. The Harvard data were imputed using MACH with NCBI Build 35 of phase 2 HapMap CEU data as the reference panel, and only SNPs with imputation quality $R^2 > 0.95$ were included in the final analysis.

Stage 1 genome-wide association analysis. Imputed genotypes were analyzed as expected genotype counts on the basis of posterior probabilities (gene dosage) using SNPTEST2 (ref. 79), assuming an additive model with geographical region as a covariate (SNPTEST v2.5 for chromosome X). MACH imputed genotypes were analyzed using best-guess genotypes from MACH, and PLINK was used for logistic association tests adjusting by the top two principal components. Only markers with an imputation quality score (INFO/MacH $R^2$) $> 0.8$ were analyzed. Potential stratification was dealt with in the GenoMEL samples by including geographical region as a covariate (inclusion of principal components as covariates was previously found to make little difference$^{9}$ and elsewhere by including principal components as covariates)$^{22}$.

Meta-analysis. The heterogeneity of per-SNP effect sizes in studies contributing to the stage 1, stage 2 and overall meta-analyses was assessed using the $I^2$ metric$^{90}$. $P$ is commonly defined as the proportion of overall variance attributable to between-study variance, with values below 31% suggesting no more than mild heterogeneity. Where $F$ was less than 31%, a fixed-effects model was used, with fixed-effects $P$ values indicated by $P_{\text{fixed}}$; otherwise, a random-effects model was applied ($P_{\text{random}}$). The method of Dersimonian and Laird$^{81}$ was used to estimate the between-study variance $\hat{\tau}^2$. An overall random-effects estimate was then calculated using the weights $1/(\nu_i + \hat{\tau}^2)$, where $\nu_i$ is the variance of the estimated effect. $\hat{\tau}^2 = 0$ for the fixed-effects analyses. We report loci that reached significance at more than one marker, incorporating information from more than one study. Results for rs186133190 at 2p15 were only available from four studies; all other SNPs reported here used data from at least eight studies (Supplementary Table 3).

Per-study quantile-quantile plots of GWAS $P$ values are provided (Supplementary Fig. 3) as well as the quantile-quantile plot for the stage 1 meta-analysis (Supplementary Fig. 3a). We also provide the stage 1 quantile-quantile plot with previously reported regions removed (Supplementary Fig. 2b). Although there was minimal inflation remaining after correction by principal components or region of origin, to ensure that the residual genomic inflation was not biasing our results, the meta-analysis was repeated using the genome-wide association meta-analysis software GWAMA (v2.1)$^{92}$. Included studies were corrected by inflating SNP variance estimates by the genomic inflation ($\lambda$) value for each study. As expected, given the low level of residual inflation, the corrected and uncorrected results were very similar. Genome-wide inflation factor ($\lambda$)-corrected $P$ values are provided in Supplementary Table 3.

Stage 2 genotyping. A single SNP for each of the 16 new regions reaching $P < 1 \times 10^{-4}$ in stage 1 was subsequently genotyped in 3 additional melanoma case-control sets in stage 2 (Supplementary Table 3). Any region that only showed evidence for association with CMM at a single imputed SNP and in only one study was not followed up. Included in the stage 2 genotyping were rs75691080 at 20q13.33, which, although not quite reaching $P < 1 \times 10^{-4}$, lies 20 kb from RTELI, and rs12969304 at 3q26.2, which lies 1 kb from TERC. Both these genes are known to be telomere related and have been associated with leukocyte telomere length$^{35}$. The 16 new regions included rs2390419 at 11q13.33, which is 450 kb from our primary hit in the region of CCND1 (rs498136; Supplementary Figs. 5 and 6) and is in linkage equilibrium with the genome-wide significant hit in this region ($\hat{\tau}^2 = 0.002$) and so may represent an independent effect.

The first stage 2 data set of 1,797 cases and 1,709 controls was from 2 studies based in Cambridge, UK (see the Supplementary Note for details on the samples). These samples were genotyped using TaqMan assays (Applied Biosystems). We performed 2-μl PCR reactions in 384-well plates using 10 ng of DNA (dried) with 0.05 μl of assay mix and 1 μl of Universal Master Mix (Applied Biosystems), according to the manufacturer’s instructions. End-point reading of the genotypes was performed using the ABI 7900HT Real-Time PCR system (Applied Biosystems).

The second stage 2 data set comprised 711 cases and 890 controls from the Breakthrough Generations study. These samples were genotyped in the same way as the Cambridge stage 2 samples.

The third stage 2 data set comprised 800 cases and 800 controls from Athens, Greece. Genomic DNA was isolated from 200 μl of peripheral blood using the QIAGen DNA Blood mini kit (Qiagen). DNA concentration was quantified in samples before genotyping using the Quant-iT dsDNA HS Assay kit (Invitrogen). The concentration of the DNA was adjusted to 5 μg/μl. Selected SNPs were genotyped using the Sequenom iPLEX assay. Allele detection in this assay was performed using matrix-assisted laser desorption/ionization–time-of-flight (MALDI-TOF) mass spectrometry$^{83}$. Because genotyping was performed by Sequenom, specific reaction details are not available. As described by Gabriel et al.$^{85}$, the assay consists of an initial locus-specific PCR reaction followed by single-base extension using mass-modified deoxyxycytidate terminators of an oligonucleotide primer that anneals immediately upstream of the polymorphic site of interest. Using MALDI-TOF mass spectrometry, the distinct mass of the extended primer identifies the SNP allele.

Genotyping of 18 SNPs was attempted in stage 2; the rs186133190 (2p15), rs6750047 (2p22.2), rs498136 (11q13.3) and rs4731742 (7q32.3) assays failed in one or more stage 2 data sets (Supplementary Table 3). After quality control (excluding individuals missing $>1$ genotype call, SNPs missing in $>3\%$ of samples and SNPs with Hardy-Weinberg equilibrium $P < 5 \times 10^{-4}$), there were 1,692 cases and 1,592 controls from Cambridge, UK, 639 cases and 823 controls from Athens, Greece, and 1,142 cases and 1,142 controls from Athens, Greece, available for stage 2.

Statistical power for stage 2. We estimated the power to reach $P < 0.05$ in the stage 2 samples for all SNPs that reached genome-wide significance in the stage 1 meta-analysis (Supplementary Table 3). We converted odds ratios to genotype relative risks (as the SNPs are relatively frequent, this is a reasonable assumption) and estimated power by simulating cases and controls (10,000 iterations) and conducting a Cochran-Armitage trend test (Supplementary Table 5).

Conditional analysis. GCTA$^{96}$ was used to perform conditional/joint GWAS analysis of newly identified or confirmed loci. GCTA allows for the conditional analysis of summary meta-analysis statistics if provided with a sufficiently large reference population (2,000–5,000 samples) used in the meta-analysis to estimate LD. We used the QMega-610k set as a reference population to determine LD. QMega-610k imputation data for well-imputed SNPs (INFO $> 0.8$) was converted to best-guess genotypes using GTOOL software (see URLs).

After conversion to best-guess genotypes (genotype probability threshold of 0.5), SNPs with Hardy-Weinberg equilibrium $P < 1 \times 10^{-4}$, MAF $< 0.01$ and $>3\%$ missingness were removed. As described by Yang
36, we further cleaned the QMEGA-610k data set to include only completely unrelated individuals (identity-by-descent score ≤ 0.025 versus the standard 0.2 used in the meta-analysis), leaving a total of 4,437 individuals and 7.24 million autosomal SNPs in the reference panel.

Within each new locus, stage 1 fixed-effects summary meta-analysis data for SNPs within 1 Mb on either side of the top SNP were adjusted for the top SNP using the --cojo-cond option. As described by Yang et al. 36, we used the genomic control–corrected GWAS meta-analysis results. Where an additional SNP remained with P < 5 × 10^{-8} after adjustment for the top SNP, we performed an additional round including both SNPs. If the remaining SNPs had P values greater than 5 × 10^{-8}, no further analysis was performed. The results of this analysis are reported in Supplementary Table 6.

Proportion of familial relative risk. We have used the formula for calculating the proportion of FRR as outlined by the Cancer Oncological Gene-environment Study (COGS; see URLs). Given that CMM incidence is low and that the odds ratios reported were small, we have assumed that the odds ratios derived from the stage 1 meta-analysis are equivalent to relative risks. With this assumption, we estimated the proportion of the FRR explained by each SNP (FRR_{SNP}) as

\[ \text{FRR}_{\text{SNP}} = \frac{(p^2 + q^2)/((p + q)^2)}{r} \]

where the risk allele and alternative allele frequencies are p and q, respectively, and r is the odds ratio for the risk allele.

Assuming an FRR_{mela}, for CMM of 2.19 (ref. 84) and using the combined effect of all SNPs (assuming a multiplicative effect and a single SNP per locus), we computed the proportion of FRR explained by a set of SNPs as

log_e (product of FRR_{SNP})/log_e (FRR_{mela}).

Association with nevus count or pigmentation. Pigmentation and nevus phenotype data were available for 980 melanoma cases and 499 control individuals from the Leeds case-control study. 85,86 Additional individuals from the Leeds melanoma cohort study 87 had pigmentation data available, giving a total of 1,458 subjects with melanoma and 499 control subjects. For the most significant SNP in each region reaching P < 1 × 10^{-8} in the initial meta-analysis, log-transformed age- and sex-adjusted total nevus count was regressed on the number of risk alleles, adjusting for case-control status. A sun sensitivity score was calculated for all subjects on the basis of factor analysis of six pigmentation variables (hair color, eye color, self-reported freckling as a child, propensity to burn, ability to tan and skin color on the inside upper arm) 89. This score was similarly regressed on the number of risk alleles and adjusted for case-control status. Full results can be found in Supplementary Table 1.

Bioinformatic annotation. As the SNP most associated with the phenotype is quite likely not the underlying functional variant 88, we performed a comprehensive bioinformatic assessment of SNPs with P_{fixed} (if I 31%) or P_{random} (if I ≥ 31%) values within a factor of 100 of the P value for the peak SNP. To ensure that we were not missing potentially interesting functional candidates, HaploReg was used to identify additional SNPs with r^2 > 0.8 and no more than 200 kb away from those SNPs with P values within a factor of 100 of the peak SNP using 1000 Genomes Project pilot data 78,79. GCTA was used to confirm that the SNPs carried forward for bioinformatic assessment were derived from a common signal. After adjustment for the top SNP for each locus, the SNPs selected for bioinformatic annotation at 6p22.3, 7p21.1, 10q24, 11q13.3 and 15q13.1 had CMM association P > 0.01. At 9q31.2, a single SNP, rs1484384, retained a modest melanoma association (P = 0.008) after adjustment for rs10739221; the rest had association P > 0.01. At 2p22.2, SNPs with P values within two orders of magnitude of the P value for the peak SNP rs6750047 included rs1800440, a nonsynonymous SNP with limited LD with rs6750047 (r^2 = 0.04). After adjustment for rs6750047, the significance of rs1800440 was essentially unchanged (P = 4.3 × 10^{-4}), and a second SNP, rs163092, remained weakly associated with melanoma (P = 0.008); all other SNPs had P > 0.01. Adjustment for both rs6750047 and rs1800440 removed the association between rs163092 and CMM (P > 0.01).

HaploReg 37 and RegulomeDB 40 were cross-checked to explore data reflecting transcription factor binding, open chromatin and the presence of putative enhancers. These tools summarize and collate data from the ENCODE 39 and Roadmap Epigenomics 33 public databases and from a range of other functional tools. The ENCODE and Roadmap Epigenomics projects have assayed a large number of different cell types, including primary keratinocyte and melanocyte cells and, for a limited number of assays, melanoma cell lines; predicted functional activity in these cell types was given priority over that from cell types less likely to be involved in CMM risk. The summary results reported by HaploReg and RegulomeDB assign regions a putative function on the basis of the combined results of multiple functional experiments and the position of the regions relative to known genes 38,39. For example, ENCODE assigns the label of ‘predicted enhancer’ to areas of open chromatin that overlap a signal for monomethylation of histone H3 at lysine 4 (H3K4me1) and bind transcription factors 39. Roadmap Epigenomics uses a similar ranking system to ENCODE, which is summarized in the documentation for HaploReg 37. For example, a weak enhancer will have only a weak signal for trimethylation of histone H3 at lysine 36 (H3K36me3), whereas an active enhancer will have strong signals for H3K36me3, H3K4me1 and acetylation of histone H3 at lysine 27 (H3K27ac). These labels are further divided into weak and strong ones depending on the quality of evidence. Although these labels are predicted or putative, ENCODE reports that >50% of predicted enhancers are confirmed by follow-up assays 39, and these serve as a useful guide in interrogating CMM-associated SNPs. Results from these tools were followed up in more detail using the UCSC Genome Browser 40 to explore data from the ENCODE 39 and the Roadmap Epigenomics 39 projects.

In addition, HaploReg uses genome-wide SNPs to estimate the background frequency of SNPs occurring in putative enhancer regions; this analysis was used to test for enrichment of CMM-associated SNPs with an uncorrected binomial test threshold of P = 0.05 (ref. 37).

The eQTL browser, the GTEx data set 55 and MuTHER 38,39 were further interrogated to attempt to resolve potential genes influenced by disease-associated SNPs. For these databases, we report only cis results; details on cell types and the definition of cis boundaries can be found in Supplementary Tables 7–9. The peak SNP for each locus, as well as other functionally interesting SNPs identified by HaploReg and RegulomeDB, were used to search listed eQTL databases. As the SNP coverage can differ for each database, where SNPs of interest were not present in the eQTL data sets, we searched using proxies in high LD (r^2 > 0.95). Although priority was given to cell types more likely to be involved in CMM biology (for example, sun-exposed skin in GTEx or skin in MuTHER), we report eQTLs from other tissue types to highlight the potential functional impact of the identified SNPs.

Regional plots of −log_{10} (P values) were generated using LocusZoom 90. Where pairwise LD measures are given, these were estimated from the 379 European-ancestry 1000 Genomes Project Phase 1 April 2012 samples using PLINK 91 or the --hap-r2 command in VCFTools 92 unless otherwise indicated.

To test for any overlap with published GWAS association, results reported in the NHGRI GWAS catalog for reported loci were extracted on 24 July 2014 and cross-checked against the stage 1 meta-analysis results.

Additional methods. Manhattan plots were generated in R using scripts written by S. Turner (see URLs). Forest plots were generated using the R rmeta package (see URLs).
79. 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).
80. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. Stat. Med. 21, 1539–1558 (2002).
81. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. Control. Clin. Trials 7, 177–188 (1986).
82. Mägi, R. & Morris, A.P. GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics 11, 288 (2010).
83. Gabriel, S., Ziaugra, L. & Tabbaa, D. SNP genotyping using the Sequenom MassARRAY iPLEX platform. Curr. Protoc. Hum. Genet. Chapter 2, Unit 2.12 (2009).
84. Cho, E., Rosner, B.A., Feskanich, D. & Colditz, G.A. Risk factors and individual probabilities of melanoma for whites. J. Clin. Oncol. 23, 2669–2675 (2005).
85. Newton-Bishop, J.A. et al. Melanocytic nevi, nevus genes, and melanoma risk in a large case-control study in the United Kingdom. Cancer Epidemiol. Biomarkers Prev. 19, 2043–2054 (2010).
86. Newton-Bishop, J.A. et al. Relationship between sun exposure and melanoma risk for tumours in different body sites in a large case-control study in a temperate climate. Eur. J. Cancer 47, 732–741 (2011).
87. Newton-Bishop, J.A. et al. Serum 25-hydroxyvitamin D3 levels are associated with Breslow thickness at presentation and survival from melanoma. J. Clin. Oncol. 27, 5439–5444 (2009).
88. Edwards, S.L., Beesley, J., French, J.D. & Dunning, A.M. Beyond GWASs: illuminating the dark road from association to function. Am. J. Hum. Genet. 93, 779–797 (2013).
89. Nica, A.C. et al. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genet. 7, e1002003 (2011).
90. Pruim, R.J. et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336–2337 (2010).
91. 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).
92. Danecek, P. et al. The variant call format and VCFtools. Bioinformatics 27, 2156–2158 (2011).