Cell-type–specific eQTL of primary melanocytes facilitates identification of melanoma susceptibility genes

Tongwu Zhang,[1,7] Jiyeon Choi,[1,7] Michael A. Kovacs,[1] Jianxin Shi,[2] Mai Xu,[1] NISC Comparative Sequencing Program,[9] Melanoma Meta-Analysis Consortium,[10] Alisa M. Goldstein,[3] Adam J. Trower,[4] D. Timothy Bishop,[4] Mark M. Iles,[4] David L. Duffy,[5] Stuart MacGregor,[5] Laufey T. Amundadottir,[1] Matthew H. Law,[5] Stacie K. Loftus,[6] William J. Pavan,[6,8] and Kevin M. Brown[1,8]

[1]*Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA;* [2]*Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA;* [3]*Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA;* [4]*Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, LS9 7TF, United Kingdom;* [5]*Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4006, Australia;* [6]*Genetic Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA*

Most expression quantitative trait locus (eQTL) studies to date have been performed in heterogeneous tissues as opposed to specific cell types. To better understand the cell-type–specific regulatory landscape of human melanocytes, which give rise to melanoma but account for <5% of typical human skin biopsies, we performed an eQTL analysis in primary melanocyte cultures from 106 newborn males. We identified 597,335 *cis*-eQTL SNPs prior to linkage disequilibrium (LD) pruning and 4997 eGenes (FDR < 0.05). Melanocyte eQTLs differed considerably from those identified in the 44 GTEx tissue types, including skin. Over a third of melanocyte eGenes, including key genes in melanin synthesis pathways, were unique to melanocytes compared to those of GTEx skin tissues or TCGA melanomas. The melanocyte data set also identified *trans*-eQTLs, including those connecting a pigmentation-associated functional SNP with four genes, likely through *cis*-regulation of *IRF4*. Melanocyte eQTLs are enriched in *cis*-regulatory signatures found in melanocytes as well as in melanoma-associated variants identified through genome-wide association studies. Melanocyte eQTLs also colocalized with melanoma GWAS variants in five known loci. Finally, a transcriptome-wide association study using melanocyte eQTLs uncovered four novel susceptibility loci, where imputed expression levels of five genes (*ZFP90*, *HEBP1*, *MSC*, *CBWD1*, and *RP11–383H13.1*) were associated with melanoma at genome-wide significant *P*-values. Our data highlight the utility of lineage-specific eQTL resources for annotating GWAS findings, and present a robust database for genomic research of melanoma risk and melanocyte biology.

[Supplemental material is available for this article.]

Expression quantitative trait locus (eQTL) analysis is a powerful method to study gene expression and regulatory profiles in human populations. Early studies mainly focused on eQTLs for whole blood or blood-derived cells due to sample accessibility (Stranger et al. 2007; Pickrell et al. 2010), and more recently, numerous eQTL data sets derived from normal human tissues have been made publicly available. Perhaps most notable are those from the Genotype-Tissue Expression (GTEx) (The GTEx Consortium 2015) project representing >44 tissue types of hundreds of postmortem donors. These studies have collectively emphasized the cell-type–specific nature of eQTLs, where 29%–80% of eQTLs are cell type specific (Dimas et al. 2009; Nica et al. 2011; Fairfax et al. 2012; The GTEx Consortium 2015). While eQTLs from normal tissues provide valuable insights, tissues are constituted of multiple distinct cell types with specific gene regulatory profiles as exemplified by eQTLs of different blood-isolated cell types (Fairfax et al. 2012). Moreover, the collection and sampling process of tissue samples from organs does not allow precise control over cell representation, adding a major source of biological variability in addition to other technical variation (McCall et al. 2016). However, other than for immune cells (Kim-Hellmuth et al. 2017), induced Pluripotent Stem Cells (iPSC) (Kilpinen

[7]These authors are co-first authors and contributed equally to this work.
[8]These authors are co-senior authors and contributed equally to this work.
[9]A complete list of the NISC Comparative Sequencing Program authors appears at the end of this paper.
[10]A complete list of the Melanoma Meta-Analysis Consortium authors appears at the end of this paper.
Corresponding authors: bpavan@mail.nih.gov, kevin.brown3@nih.gov
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.233304.117.

et al. 2017), or smooth muscle cells (Liu et al. 2018), eQTL data sets representing single primary cell types and direct comparison of these to the tissue type of origin have been lacking.

eQTLs may be particularly useful for annotating variants associated with complex traits, as such variants are likely enriched for eQTLs (Nicolae et al. 2010). A recent study suggested that two-thirds of candidate common trait susceptibility genes identified as eQTLs are not the nearest genes to the GWAS lead SNPs, highlighting the utility of this approach in annotating GWAS loci (Zhu et al. 2016). Importantly, GWAS variants are enriched in eQTLs in a tissue-specific manner. For instance, whole blood eQTLs are enriched with autoimmune disorder-associated SNPs but not with GWAS SNPs for bipolar disease or type 2 diabetes (The GTEx Consortium 2015). These findings highlight the importance of using eQTL data sets from relevant cell types when following up GWAS loci for a specific disease. In addition to providing functional insights for known GWAS loci, eQTL data may be useful for identification of novel trait-associated loci via imputation of genotype-correlated gene expression levels into GWAS data sets (Gamazon et al. 2015; Gusev et al. 2016). Such approaches, usually referred to as transcriptome-wide association studies (TWASs), enable assignments of potentially disease-associated loci via estimations of their genetically regulated expression.

GWAS for melanoma risk, nevus count, and multiple pigmentation traits have identified numerous associated genetic loci (Stokowski et al. 2007; Sulem et al. 2007, 2008; Brown et al. 2008; Gudbjartsson et al. 2008; Han et al. 2008; Bishop et al. 2009; Falchi et al. 2009; Nan et al. 2009, 2011; Duffy et al. 2010; Eriksson et al. 2010; Amos et al. 2011; Barrett et al. 2011; Macgregor et al. 2011; Candille et al. 2012; Zhang et al. 2013; Jacobs et al. 2015; Law et al. 2015; Liu et al. 2015; Hysi et al. 2018; Visconti et al. 2018), with melanoma GWAS alone identifying 20 regions associated with risk. Trait-associated variation explaining many of these loci could reasonably be expected to be reflected in the biology of the melanocyte, the pigment-producing cell in human skin and the cellular origin of melanoma. Melanocytes are the cells in the skin that function to produce the melanin pigments, eumelanin and pheomelanin, in response to neuroendocrine signals and UV-exposure (Costin and Hearing 2007). These melanin pigments are contained in lysosome-related organelles called melanosomes, are shuttled to the melanocyte dendrites, and transferred to neighboring keratinocytes, thus protecting skin from UV radiation (Sitaram and Marks 2012). The process of pigmentation is complex and multigenic, and it is regulated by genes with diverse cellular functions including those within MAPK, PI3K, Wnt/beta catenin signaling pathways (Liu et al. 2014), as well as those involved in lysosome-related functions and vesicular trafficking (Sitaram and Marks 2012).

While several skin-related eQTL data sets are available, the largest ones (GTEx [The GTEx Consortium 2015], MuTHER [Nica et al. 2011], EUROBATS [Buil et al. 2015]) are derived from heterogeneous skin tissues, of which melanocytes only represent a small fraction. The Cancer Genome Atlas Project (TCGA) also offers a considerable set of tumor tissue expression data accompanied by genotype information providing a platform for tumor-type relevant eQTL data including melanoma (https://cancergenome.nih.gov/), but these tumor tissues contain a high burden of somatic aberrations, are heterogeneous and may reflect multiple disease subtypes, and may not represent the underlying biology associated with cancer risk and/or pigmentation. Given these limitations, we took advantage of the accessibility of primary melanocytes obtained from foreskin tissues and built a cell-type–specific eQTL data set to study the lineage-specific regulatory function of melanoma- and pigmentation-associated common variants.

Results

Melanocyte eQTLs are distinct from those of other tissue types

In order to create a melanocyte-specific eQTL resource, we obtained primary melanocyte cultures isolated from foreskin of 106 healthy newborn males predominantly of European descent (Supplemental Table S1). We then cultured all 106 lines following a uniform procedure to harvest RNA and DNA, for RNA sequencing and genotyping, respectively (see Methods). Given the relatively small size of our sample set, we initially focused our analysis on local eQTLs (*cis*-eQTL), where we assessed the association between expression of each gene with common variants within ±1 Mb of transcription start sites (TSSs), following the best practices from the GTEx project (see Methods). In all, we identified 4997 "eGenes" (genes exhibiting association with genotypes of at least one SNP at FDR < 0.05) (Supplemental Table S2) and 597,335 genome-wide "significant eQTLs" (unique SNP-gene pairs showing FDR < 0.05; SNPs were not LD-pruned), which are higher numbers than any GTEx tissue type of similar sample size (Supplemental Table S3). Melanocyte eGenes were enriched with Gene Ontology (GO) terms including metabolic process, mitochondrial translation, biosynthetic process, catalytic activity, and ion-binding, as well as lysosome and metabolic pathways (Supplemental Table S4). Further, melanocyte eGenes included 46% of genes categorized with GO terms as containing "melanin" (*OCA2*, *TRPC1*, *CTNS*, *DCT*, *MCHR1*, *SLC45A2*, *TYR*, *BCL2*, *WNT5A*, *MC1R*, and *MYO5A*) (http://amigo.geneontology.org) and 20% of curated pigmentation genes (based on human and mouse phenotype, OMIM, MGI) such as *IRF4*, *TRPM1*, and *MC1R* (Supplemental Tables S5, S6), reflecting pigmentation-related biology of melanocytes.

Direct comparison of significant melanocyte eQTLs with 44 GTEx tissue types indicated that the shared eQTL proportion (π_1) between melanocytes and each of the GTEx tissue types was 0.74 (vs. transformed fibroblasts) or lower, suggesting relatively low levels of sharing even with two types of skin samples ($\pi_1 = 0.67$ with Skin_Sun_Exposed, and 0.58 with Skin_Not_Sun_Exposed) (Fig. 1). This contrasts with the considerably higher levels of sharing between the two types of skin samples ($\pi_1 = 0.91$) or among brain tissues (average $\pi_1 = 0.87$) in GTEx. We further focused the comparison of our melanocyte data set to three tissue types that are directly relevant to melanoma and pigmentation phenotypes: the two above-mentioned GTEx skin types, as well as skin cutaneous melanomas (SKCM) collected through TCGA (adding an adjustment for local DNA copy number) (see Supplemental Material). Collectively, these four eQTL data sets identified 12,136 eGenes, with 382 eGenes shared among all four data sets. Notably, 1801 eGenes (36% of melanocyte eGenes) were entirely private to melanocytes, and a total of 6187 eGenes (51% of eGenes from all four data sets) were specific to only one of four data sets (Supplemental Fig. S1; Supplemental Table S2). eGenes from these four data sets collectively accounted for 150 of 379 (40%) curated pigmentation genes, with the majority specific to one data set (Supplemental Fig. S2).

Melanocyte eQTLs are enriched in *cis*-regulatory signatures and supported by allelic imbalance

We next sought to determine whether melanocyte eQTLs were corroborated by allelic imbalance variants in heterozygous

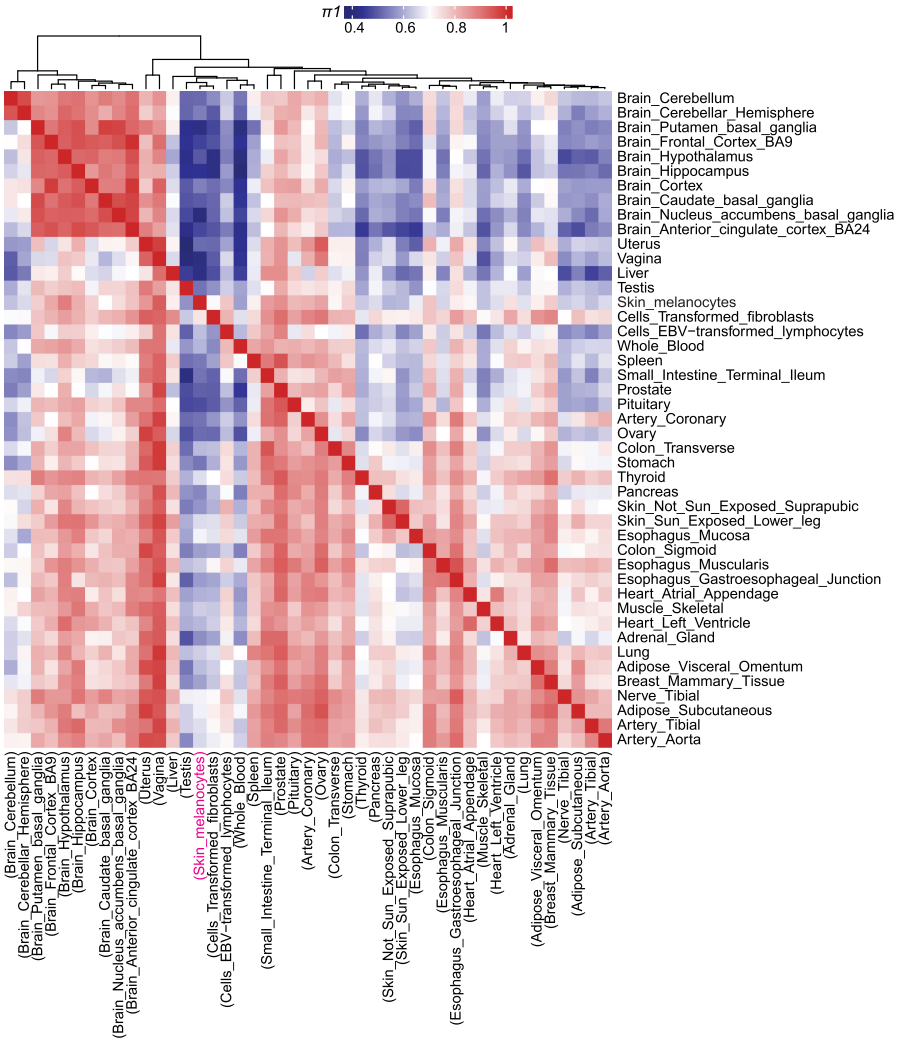

Figure 1. Melanocyte eQTLs display a distinct pattern from those of 44 GTEx tissue types. Dendrogram and heat map presenting the sharing of eQTLs between human primary melanocytes and 44 other GTEx tissue types. Pairwise π_1 statistics were calculated from single-tissue eQTL discoveries in each tissue using all the genome-wide significant eQTL SNP-gene pairs. π_1 is only calculated when the gene is expressed and testable both in discovery (columns) and replication (rows) tissues. Higher π_1 values indicate an increased replication of eQTLs between two tissue types. π_1 values range between ~0.41 and 1 and are color-coded from blue (low sharing) to red (high sharing). Tissues are clustered using the Spearman's correlation of π_1 values. Note that π_1 values are not symmetrical, since each entry in row *i* (replication tissue) and column *j* (discovery tissue) is an estimate of $\pi_1 = Pr$ (eQTL in tissue *i* given an eQTL in tissue *j*). Discovery tissue names are shown in parentheses on the *bottom*. The position of the skin melanocyte eQTL data set from the discovery tissues is shown in pink.

individuals from the same data set. To determine genome-wide allele-specific expression (ASE), we performed binomial tests at the single-sample level, identifying 48,038 unique allelic imbalance variants (FDR < 0.05 or effect size >0.15) (Supplemental Table S7). Of these unique variants, 38.6% (18,532 of 48,038 variants) were in the coding region of significant melanocyte eGenes, demonstrating an enrichment of ASE in melanocyte eGenes (Fisher's exact test $P = 2.34 \times 10^{-73}$, odds ratio = 1.82) (Supplemental Table S7). Further, the average allelic effects of 48,038 ASE variants from all the heterozygous individuals were significantly larger in the eGene group (Wilcoxon signed rank test $P = 1.67 \times 10^{-34}$; average |Mean AE| = 0.046 for eGenes vs. 0.035 for non-eGenes; effect size = 0.115) (Supplemental Fig. S3A). Similarly, the proportions of heterozygous individuals displaying allelic imbalance at each locus was significantly higher in the eGene group (Wilcoxon signed rank test $P = 1.27 \times 10^{-81}$; mean % = 13.4 for eGenes vs. 8.4 for non-eGenes; effect size = 0.195) (Supplemental Fig. S3B).

We then further examined if melanocyte eQTLs were enriched within epigenetic signatures marking melanocyte *cis*-regulatory elements. We specifically examined regions of open chromatin (marked by DNase I hypersensitivity sites, DHS), as well as promoter and enhancer histone marks (H3K27ac, H3K4me1, and H3K4me3) generated from primary cultured human melanocytes by the ENCODE and Epigenome Roadmap Projects (www.encodeproject.org; www.roadmapepigenomics.org) (The ENCODE Project Consortium 2012; Roadmap Epigenomics Consortium et al. 2015). Indeed, higher proportions of melanocyte eQTL SNPs were localized to melanocyte DHS, H3K27ac, H3K4me1, and H3K4me3 peaks compared to all tested SNPs (i.e., *cis*-SNPs ±1 Mb of TSSs of all the genes tested for eQTL) (Supplemental Fig. S4A). Enrichment of melanocyte eQTL SNPs for each of the melanocyte *cis*-regulatory signatures was statistically significant ($P < 1 \times 10^{-4}$, 10,000 permutations; 1.81- to 5.48-fold) (Table 1) and mostly more pronounced than that observed in GTEx skin tissues or melanoma tumors (Supplemental Fig. S4B; Table 1). Enrichment of melanocyte eQTLs was also observed in additional genomic features (Supplemental Figs. S4, S5; Supplemental Material).

Melanoma GWAS signal is enriched in melanocyte-specific genes and eQTLs

Next, we sought to determine if melanocyte eQTLs were enriched with common risk variants from the most recent melanoma GWAS meta-analysis (Law et al. 2015). A quantile-quantile plot demonstrated an enrichment of significant GWAS *P*-values for eQTL SNPs compared to non-eQTL SNPs (Fig. 2A), which was the most pronounced in melanocyte eQTLs (estimated Lambda = 1.51) compared to three related tissue types as well as all the other GTEx tissue types (Supplemental Fig. S6).

To further assess the enrichment of melanoma heritability in melanocyte-specific expressed genes, we performed LD score regression analysis (Finucane et al. 2015). The results indicated that partitioned melanoma heritability was significantly enriched (2.54-fold; $P = 2.45 \times 10^{-6}$) in melanocyte-specific genes (top 4000 genes compared to 47 GTEx tissue types) as well as in those of three "skin" category GTEx tissue types ($P = 3.11 \times 10^{-6}$, 8.62×10^{-6}, and 4.37×10^{-5}, with 2.52-, 2.58-, and 2.34-fold for not sun-exposed

Table 1. Enrichment of eQTL SNPs in melanocyte *cis*-regulatory signatures

Epigenetic mark[a]		DHS	H3K27ac	H3K4Me1	H3K4Me3
Melanocyte eQTLs	Fold enrichment[b]	1.81	5.48	1.99	3.34
	P-value[c]	<0.0001	<0.0001	<0.0001	<0.0001
TCGA SKCM eQTLs	Fold enrichment[b]	1.72	2.69	1.86	3.44
	P-value[c]	<0.0001	<0.0001	<0.0001	<0.0001
Skin not sun-exposed eQTLs	Fold enrichment[b]	1.78	2.77	1.92	3.5
	P-value[c]	<0.0001	<0.0001	<0.0001	<0.0001
Skin sun-exposed eQTLs	Fold enrichment[b]	1.75	2.66	1.88	3.29
	P-value[c]	<0.0001	<0.0001	<0.0001	<0.0001

[a]DNase I hypersensitivity sites (DHS) and gene regulatory histone marks (H3K27ac, H3K4me1, and H3K4me3) of primary melanocytes (Epigenome Roadmap database; www.roadmapepigenomics.org).
[b]Mean fold enrichment of eQTL SNPs over control SNP sets (with similar distribution of MAF and LD) from 10,000 permutations that are overlapping with each epigenetic mark.
[c]See Methods section.

skin, sun-exposed skin, and transformed fibroblasts, respectively) (Fig. 2B; Supplemental Table S8; Supplemental Fig. S7).

A functional pigmentation SNP at the *IRF4* locus is a significant *trans*-eQTL for four genes in melanocytes

While the modest size of this data set limits power, we also performed *trans*-eQTL analyses for the SNPs that are located over 5 Mb away from the TSS of each gene or on a different chromosome. In all, we identified 15 genome-wide significant *trans*-eQTL genes (excluding genes of mappability < 0.8 or overlapping low complexity regions) (Supplemental Table S9). Of these, eight *trans*-eQTL SNPs were also *cis*-eQTLs for local genes within 1 Mb. Notably, rs12203592 (Chr 6: 396321), among these, is a genome-wide significant *trans*-eQTL SNP for four different genes on four separate chromosomes (*TMEM140*, *MIR3681HG*, *PLA1A*, and *NEO1*) and is also the strongest *cis*-eQTL SNP for the *IRF4* gene ($P = 7.9 \times 10^{-16}$, slope = −1.14), which encodes the transcription factor, interferon regulatory factor 4. All four genes displayed the same direction of allelic gene expression as *IRF4* levels relative to rs12203592 (Fig. 3). rs12203592 has previously been associated with human pigmentation phenotypes (Han et al. 2008). This variant was also shown to be a functional SNP mediating transcription of *IRF4* in melanocytes via C allele-preferential binding of the transcription factor, TFAP2, by collaborating with melanocyte lineage-specific transcription factor, MITF, in turn activating the melanin synthesis enzyme, *TYR*. The rs12203592-C allele (prevalent in African populations) is correlated with high *IRF4* levels in our melanocyte data set, validating the findings observed in a smaller sample set (Praetorius et al. 2013). Expression correlation analyses in melanocytes indicated that expression levels of *TMEM140*, *MIR3681HG*, *PLA1A*, and *NEO1* are significantly correlated with those of *IRF4* in the same direction as shown by *trans*-eQTLs (Pearson $r = 0.54$, 0.65, 0.53, and 0.58; $P = 2.67 \times 10^{-9}$, 5.34×10^{-14}, 4.28×10^{-9}, and 6.00×10^{-11}, respectively) (Supplemental Fig. S8). To assess if *IRF4* expression levels mediate the observed *trans*-eQTL effect for these four genes, we performed mediation analyses using regression-based methods (Supplemental Material; Supplemental Table S10), as well as a recently published Genomic Mediation analysis with Adaptive Confounding adjustment (GMAC) (Yang et al. 2017). We applied the GMAC method to 455 eSNP - *cis*-eGene - *trans*-gene trios (*trans*-eQTL cutoff: $P < 1 \times 10^{-5}$), 84 of which include rs12203592. A total of 121 trios displayed a suggestive mediation ($P < 0.05$), and 32 of them were by *IRF4 cis*-eQTL including those with *TMEM140* and *NEO1* (Supplemental Table S11). In contrast, another *cis*-eQTL gene, *RPS14*, sharing two SNPs with three *trans*-eQTL genes (Supplemental Table S9), did not show suggestive mediation (Supplemental Table S11). These results are consistent with *IRF4* expression levels mediating at least part of the observed *trans*-eQTL effect. We then sought to determine if IRF4 is predicted to bind to the genomic regions encompassing the rs12203592 *trans*-eQTL genes. Sequence motif enrichment analyses indicated that IRF4 binding motifs were enriched in the genomic regions of *TMEM140*, *MIR3681HG*, *PLA1A*, and *NEO1* (±2 kb of gene boundary; $P = 1.52 \times 10^{-2}$) (Supplemental Table S12), as well as in the above-mentioned 84 *trans*-eQTL genes ($P = 7.25 \times 10^{-26}$). Together, our data suggest a melanocyte-specific *trans*-eQTL network potentially regulated by the transcription factor, IRF4.

Melanocyte eQTLs identified candidate melanoma susceptibility genes from GWAS loci

To assess colocalization of causal variants for melanoma GWAS and melanocyte eQTL, we applied the previously described eCAVIAR methodology (Hormozdiari et al. 2016). At a colocalization posterior probability (CLPP) cutoff of 1%, five of 20 known melanoma loci displayed colocalization of GWAS and melanocyte eQTL signal, with colocalization of eQTL signal for nine genes overall (Table 2; Fig. 4). The same analysis with two GTEx skin data sets observed colocalization at combined four loci and 21 genes (Supplemental Table S13). The union of all three data sets totaled 29 genes from six loci, indicating that these eQTL data sets complement each other rather than being redundant. Consistent with a previous report (The GTEx Consortium 2017), only 66% (four of six loci) but not all of melanoma GWAS signal colocalized with the nearest expressed gene in one or more of the three data sets. Importantly, melanocyte eQTLs (but not the skin data sets) validated *PARP1* as a target gene on the locus at Chr1q42.12, which was previously characterized as a melanoma susceptibility gene displaying melanocyte lineage-specific function (Fig. 4; Choi et al. 2017). Melanocyte eQTLs also uniquely identified a known pigmentation gene, *SLC45A2*, on the locus at 5p13.2 as a target gene, reflecting a melanin synthesis pathway uniquely captured in melanocyte eQTLs. Consistent with previous findings, eCAVIAR colocalization was observed for multiple genes in most of the loci, and genes with the highest CLPP scores from different eQTL data sets did not overlap for a given melanoma locus. In addition, we also performed eCAVIAR analyses for GWAS of melanoma-associated traits (number of melanocytic nevi, skin

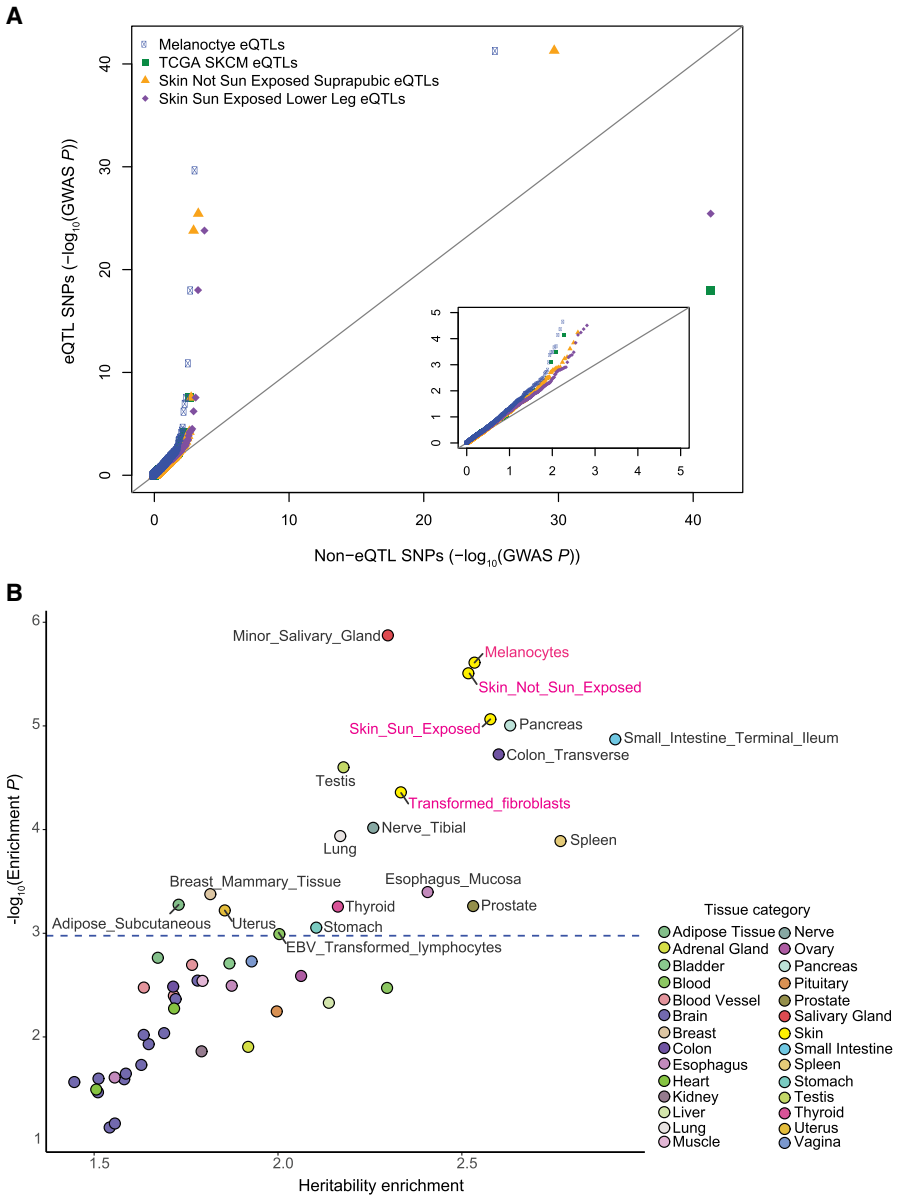

Figure 2. Melanoma GWAS signal is enriched in melanocyte-specific genes and eQTLs. (*A*) QQ plot presents melanoma GWAS LD-pruned *P*-values of significant eQTL SNPs versus non-eQTL SNPs for the melanocyte data set compared to those for sun-exposed skin, non-sun-exposed skin, and melanoma tumors. SNPs were classified as eQTL SNPs if they were significant eQTLs or in strong LD ($r^2 > 0.8$) with an eQTL SNP (eQTL SNPs threshold: FDR < 0.05) in each data set. The *inset* displays a zoomed-in view of a lower $-\log_{10}$ GWAS *P*-value range (0–5 range for *x*- and *y*-axes). (*B*) Melanoma heritability enrichment levels and *P*-values in top 4000 tissue-specific genes from LD score regression analysis are displayed. The dashed horizontal line marks FDR = 0.05 on the *y*-axis. Names of significantly enriched individual tissue types are shown next to the data points, and the others are color-coded based on GTEx tissue category. Tissue types from the "Skin" category including melanocytes are highlighted in pink.

pigmentation, ease of tanning, and hair color), and identified target genes from two of four nevus count loci and six of 11 pigmentation loci using melanocyte and skin eQTL data sets (Supplemental Material; Supplemental Tables S14–S16).

We then performed permutation analyses to test for statistically significant enrichment of eQTLs from the four tissue types (including TCGA melanomas) in melanoma GWAS using four tiers of GWAS *P*-value thresholds (5×10^{-5}, 5×10^{-6}, 5×10^{-7}, and 5×10^{-8}) (Supplemental Table S17). The results indicated that melanoma-associated SNPs using all four thresholds are significantly enriched (at least twofold) in eQTLs. Notably, the number of GWAS loci displaying true overlap was much higher (8–12 loci) for melanocyte eQTLs than for two types of skin tissue or melanoma tumors (2–7 loci).

TWASs using melanocyte eQTL data identified four novel melanoma-associated loci

eQTL data can be utilized for transcriptome-wide association studies to impute gene expression levels into GWAS data sets. We performed a TWAS (Gusev et al. 2016) using summary statistics from the melanoma GWAS meta-analysis (Law et al. 2015) and the melanocyte eQTL data set as the reference data set (see Methods). Using 3187 eGenes passing a conservative cutoff for heritability estimates ($P < 0.01$) (Supplemental Table S18), the TWAS identified genes at three known melanoma loci at a genome-wide significant level (*MAFF* on Chr22q13.1, *CTSS* on Chr1q21.3, *CASP8* on Chr2q33-q34), with a fourth locus being suggestive (*PARP1* on Chr1q42.1) (Table 3). The TWAS further identified novel associations with melanoma at four genomic loci at a genome-wide significant level (*ZFP90* at Chr16q22.1, *HEBP1* at Chr12 p13.1, *MSC* and *RP11-383H13.1* at Chr8q13.3, and *CBWD1* at Chr9p24.3) (Table 3; Fig. 5).

We additionally performed a TWAS using each of the 44 GTEx tissue types as reference eQTL data sets. Forty-three GTEx tissue types identified one or more melanoma TWAS genes at a genome-wide significant level with a median of three genes per data set. Tibial nerve tissue identified the largest number of genes (11 genes), while melanocytes ranked third (Supplemental Table S19). Collectively, melanocyte and GTEx data sets identified 22 TWAS genes at six previously known melanoma GWAS loci (Chr1q 21.3, Chr1q42.1, Chr2q33-q34, Chr15q 13.1, Chr21q22.3, and Chr22q13.1) as well as nine TWAS genes at eight novel loci. Melanocyte eQTLs alone identified the majority of novel TWAS genes (five of nine), including the genes unique to the melanocyte data set (Supplemental Table S20). In contrast, none of the 44 GTEx tissue data sets produced more than one novel association for melanoma. Four novel melanoma TWAS genes added from 44 GTEx tissue types are *ERCC2* on Chr19q13.32, *KIF9* on Chr3p21.31, *MRAP2* on Chr6q14.2, and *ZBTB4* on Chr17p13.1. Finally, we conducted conditional analyses on the TWAS loci displaying marginally significant associations with multiple genes from melanocyte and GTEx tissue data

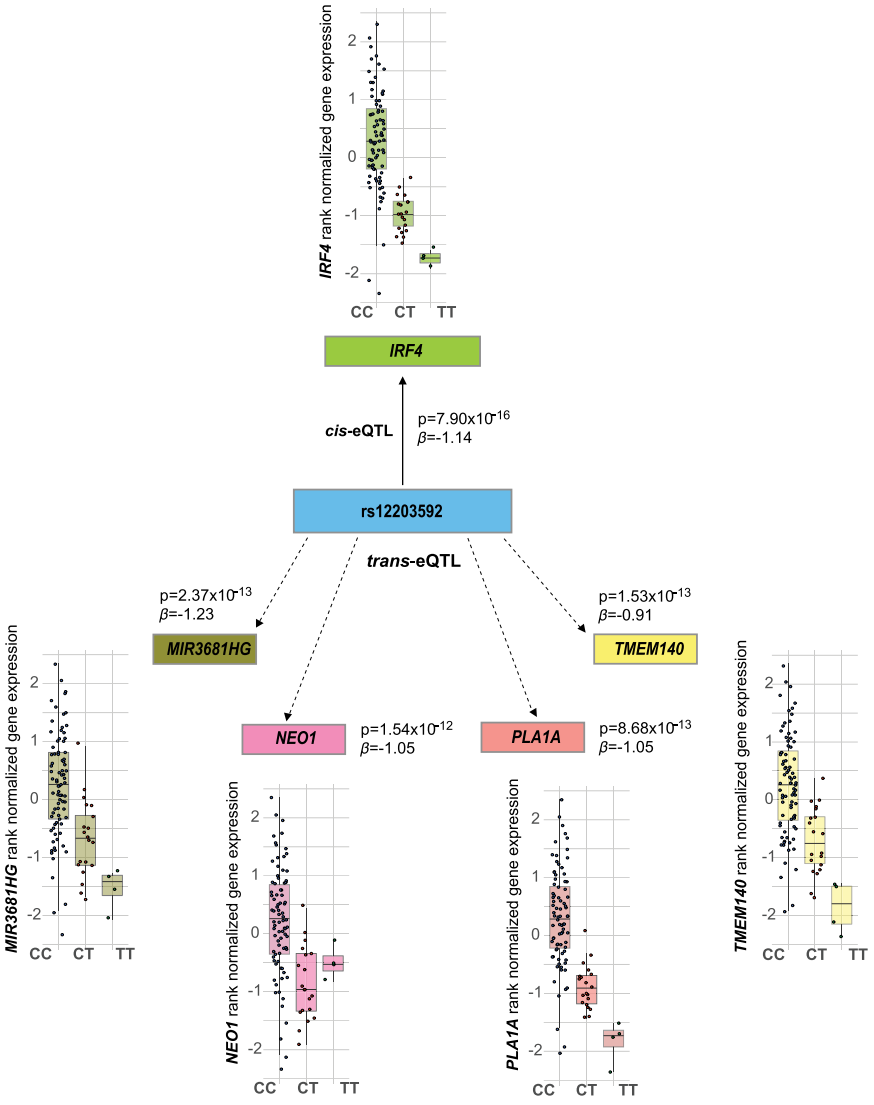

Figure 3. The pigmentation trait-associated variant, rs12203592, in *IRF4* is a *trans*-eQTL for four genes in melanocytes. *Cis*- or *trans*-eQTL *P*-values and effect sizes (β) are shown between rs12203592 and *IRF4* or rs12203592 and four genome-wide significant *trans*-eQTL genes (*TMEM140*, *MIR3681HG*, *PLA1A*, and *NEO1*). β values are shown relative to alternative alleles (T). Boxplots display gene expression levels based on rs12203592 genotypes (CC, CT, and TT).

sets. The analyses identified 15 jointly significant genes from 14 loci (Supplemental Table S20; Supplemental Material).

Discussion

In this study, we established a cell-type–specific eQTL data set using primary cultures of human melanocytes. Our data set identified a unique set of *cis*- and *trans*-eQTLs that are distinct from eQTLs of skin tissues. Melanocyte eQTLs are enriched in melanocyte-specific *cis*-regulatory elements and considerably improved melanoma GWAS annotation. Using this data set, we further identified novel melanoma TWAS loci. Our data highlight the utility of building even a modestly sized cell-type–specific data set.

Over a third of melanocyte eGenes were unique to melanocytes and not present in skin tissue data sets. GO analyses suggested that genes directly involved in melanin synthesis as well as those in lysosome and metabolic pathways were enriched in melanocyte eGenes among others. These observations are consistent with broad-based pleiotropic cell functions for genes expressed in melanocytes, including lysosome-related functions of melanin synthesis and transfer process (Sitaram and Marks 2012). Our data set was built with newborn males of primarily European descent aiming to align with the most relevant population for melanoma incidence. As there are gender differences observed in melanoma risk and mortality among others (Scoggins et al. 2006; Wendt et al. 2018; https://seer.cancer.gov/faststats [accessed on Oct. 10, 2018]), the current male-only data set cannot address gender-specific risk and related questions, which warrants future study.

Through *trans*-eQTL analysis, the melanocyte data set identified IRF4, or interferon regulatory factor 4, as a potential regulator of melanocytic lineage-specific gene expression for a set of putative downstream genes. *Trans*-eQTLs were shown to be more cell type specific than *cis*-eQTLs, and cell composition heterogeneity was proposed as a potential reason for low number of *trans*-eQTLs observed in bulk tissue data (Westra and Franke 2014; The GTEx Consortium 2017), suggesting that our single-cell-type data set might have facilitated the identification of the IRF4 *trans*-eQTL network in melanocytes. rs12203592 is a *cis*-eQTL in several other GTEx tissue types including whole blood, perhaps reflecting a better known function of *IRF4* in immune responses (Huber and Lohoff 2014). However, IRF4 has a documented role in melanocyte development, regulating expression of an enzyme essential in the production of melanin, tyrosinase (Praetorius et al. 2013). Given that IRF4 appears to function in regulation of distinct cell-type–specific processes, four rs12203592 *trans*-eQTL genes identified in melanocytes, including an interferon-stimulated gene, *TMEM140* (Kane et al. 2016), as well as perhaps a considerably larger subset of marginal *trans*-eQTL genes, could be good candidates for direct targets of IRF4. Further experimental assessment of IRF4 binding on the genomic regions of these *trans*-genes will provide additional support of this finding.

Through colocalization and TWAS, melanocyte eQTL identified unique candidate melanoma susceptibility genes for some known loci and also corroborated other data sets, including skin, in identifying candidate genes for other loci. Three melanoma loci displaying colocalization with melanocyte eQTLs were also supported by TWASs from one or more eQTL data sets (*PARP1*, *ARNT*, and *MX2*). On the other hand, some loci with larger LD

Table 2. Colocalization of melanoma GWAS and melanocyte eQTL signal

Melanoma GWAS locus	Gene	SNP ID	GWAS *P*-value	CLPP[a]	Melanoma GWAS lead SNP[b]	GWAS *P*-value[c]	r^2 [d]	Nearest expressed gene[e]
1q21.3	***ARNT***	**rs12410869**	$\mathbf{5.21 \times 10^{-13}}$	**0.07**	**rs12410869**	$\mathbf{5.21 \times 10^{-13}}$	**1.000**	*ARNT*
1q21.3	*ARNT*	rs36008098	8.55×10^{-13}	0.02	rs12410869	5.21×10^{-13}	1.000	*ARNT*
1q42.12	***PARP1***	**rs9426568**	$\mathbf{3.53 \times 10^{-13}}$	**1.00**	**rs1858550**	$\mathbf{1.68 \times 10^{-13}}$	**0.996**	*PARP1*
1q42.12	***MIXL1***	**rs9426568**	$\mathbf{3.53 \times 10^{-13}}$	**0.57**	**rs1858550**	$\mathbf{1.68 \times 10^{-13}}$	**0.996**	*PARP1*
1q42.12	***MIXL1***	**rs1858550**	$\mathbf{1.68 \times 10^{-13}}$	**0.25**	**rs1858550**	$\mathbf{1.68 \times 10^{-13}}$	**1.000**	*PARP1*
1q42.12	*ADCK3*	rs9426568	3.53×10^{-13}	0.01	rs1858550	1.68×10^{-13}	0.996	*PARP1*
1q42.12	*PSEN2*	rs9426568	3.53×10^{-13}	0.01	rs1858550	1.68×10^{-13}	0.996	*PARP1*
5p13.2	***SLC45A2***	**rs250417**	$\mathbf{2.30 \times 10^{-12}}$	**0.09**	**rs250417**	$\mathbf{2.30 \times 10^{-12}}$	**1.000**	*SLC45A2*
21q22.3	***MX2***	**rs408825**	$\mathbf{3.21 \times 10^{-15}}$	**0.12**	**rs408825**	$\mathbf{3.21 \times 10^{-15}}$	**1.000**	*MX2*
21q22.3	***MX2***	**rs443099**	$\mathbf{3.50 \times 10^{-15}}$	**0.09**	**rs408825**	$\mathbf{3.21 \times 10^{-15}}$	**1.000**	*MX2*
21q22.3	*BACE2*	rs364525	3.35×10^{-15}	0.04	rs408825	3.21×10^{-15}	0.996	*MX2*
21q22.3	*BACE2*	rs416981	3.28×10^{-15}	0.03	rs408825	3.21×10^{-15}	1.000	*MX2*
21q22.3	*BACE2*	rs408825	3.21×10^{-15}	0.02	rs408825	3.21×10^{-15}	1.000	*MX2*
21q22.3	*BACE2*	rs443099	3.50×10^{-15}	0.02	rs408825	3.21×10^{-15}	1.000	*MX2*
22q13.1	***APOBEC3G***	**rs132941**	$\mathbf{1.61 \times 10^{-12}}$	**0.12**	**rs132941**	$\mathbf{1.61 \times 10^{-12}}$	**1.000**	*PLA2G6*

eCAVIAR (Hormozdiari et al. AJHG 2016) was used for testing colocalization of melanocyte eQTL and melanoma GWAS signal. Fifty SNPs upstream of and downstream from GWAS lead SNP in each locus were chosen to quantify the probability of the variant to be causal both in GWAS and eQTL studies.

[a]Colocalization posterior probability (CLPP): probability that the same variant is causal in both GWAS and eQTL. Only the genes of CLPP >0.01 and SNPs in perfect LD ($r^2 > 0.99$ in 1KG EUR population) with the GWAS lead SNP are presented. Genes of CLPP >0.05 are shown in bold.

[b]The lowest *P*-value SNP in the locus based on fixed effect model from Law et al. (2015) study.

[c]Melanoma GWAS *P*-value (fixed model) of the SNP in the left column.

[d]r^2 between the SNP from the eCAVIAR analysis and the melanoma GWAS lead SNP of the given locus (1000 Genomes, EUR).

[e]Gene whose gene body is closest to the melanoma GWAS SNP and is expressed in melanocytes at median TPM > 0.

blocks displayed variability in target gene prediction across different data sets as well as between eCAVIAR and TWAS approaches. For the melanoma locus at Chr1q21.3, a total of eight genes were colocalized from three eQTL data sets, and TWASs nominated nine genes. Each top CLPP score gene from melanocyte and skin data sets (*ARNT*, *CERS2*, and *SETDB1*) were also supported by TWASs in more than one tissue type, and TWAS joint/conditional analyses identified *CTSS* as explaining the most of the effect in this locus (Supplemental Material; Supplemental Fig. S9). These data imply that multiple statistical approaches using diverse tissue types including the cell type of disease origin is beneficial to robust target gene prediction. Collaboration of single-cell-type and whole tissue eQTL was also exemplified in *ASIP* for hair color GWAS colocalization (Supplemental Material).

TWASs also identified novel melanoma loci by leveraging tissue-specific eQTL data sets and reducing the multiple testing burden associated with GWASs. While identification of trait-associated gene expression differences via TWASs cannot be taken to imply causality for a specific gene, TWASs may nonetheless nominate plausible candidate risk genes at significant loci. Here, we identified five genes at four new melanoma susceptibility loci (*ZFP90* on Chr16q22.1, *HEBP1* on Chr12p13.1, *MSC* and *RP11-383H13.1* on Chr8q13.3, and *CBWD1* on Chr9p24.3) using melanocyte eQTLs as a reference set and four additional new genes/loci (*ERCC2*, *KIF9*, *MRAP2*, and *ZBTB4*) using 44 GTEx tissue types.

While most of these genes have known functions that might have relevance in melanomagenesis (Supplemental Material), *ERCC2* on Chr19q13.32, among them, is a nucleotide excision repair gene targeting UV-induced DNA damage and implicated in Xeroderma pigmentosum (Taylor et al. 1997). *MRAP2* on Chr6q14.2 encodes melanocortin-2-receptor accessory protein 2, which interacts with all melanocortin receptor proteins (MCRs) together with MRAP1 to regulate cell surface expression of MCRs (Ramachandrappa et al. 2013). Seemingly relevant functions of these new candidate genes warrant further studies on their roles in melanomagenesis. In all, our primary melanocyte eQTL data set considerably advanced identification of candidate melanoma susceptibility genes from known and new melanoma loci through multiple approaches, which highlights the unique value of cell-type–specific eQTL data sets.

Methods

Melanocyte culture

We obtained frozen aliquots of melanocytes isolated from foreskin of 106 healthy newborn males, mainly of European descent, following an established protocol (Halaban et al. 2000) from the SPORE in Skin Cancer Specimen Resource Core at Yale University. Cells were grown in lot-matched Dermal Cell Basal Medium (ATCC PCS-200-030) supplemented with lot-matched Melanocyte Growth Kit (ATCC PCS-200-041) and 1% amphotericin B/penicillin/streptomycin (120-096-711, Quality Biological) at 37°C with 5% CO_2. Every step of cell culture, DNA/RNA isolation, and sequencing/genotyping processes was performed in rerandomized batches. Before harvesting the cells, media was taken and tested for mycoplasma contamination using MycoAlert PLUS mycoplasma detection kit (LT07-710, Lonza). All 106 samples were negative for mycoplasma contamination.

Genotyping and imputation

Genomic DNA was isolated from melanocytes following a standard procedure while minimizing melanin carry-over and genotyped on the Illumina OmniExpress arrays (HumanOmniExpress-24-v1-1-a) at the Cancer Genomics Research Laboratory of the Division of Cancer Epidemiology and Genetics (NCI/NIH). Following quality control, genotypes were imputed using Michigan Imputation Server (Das et al. 2016) based on 1000 Genomes (Phase 3, v5) reference panel (The 1000 Genomes Project

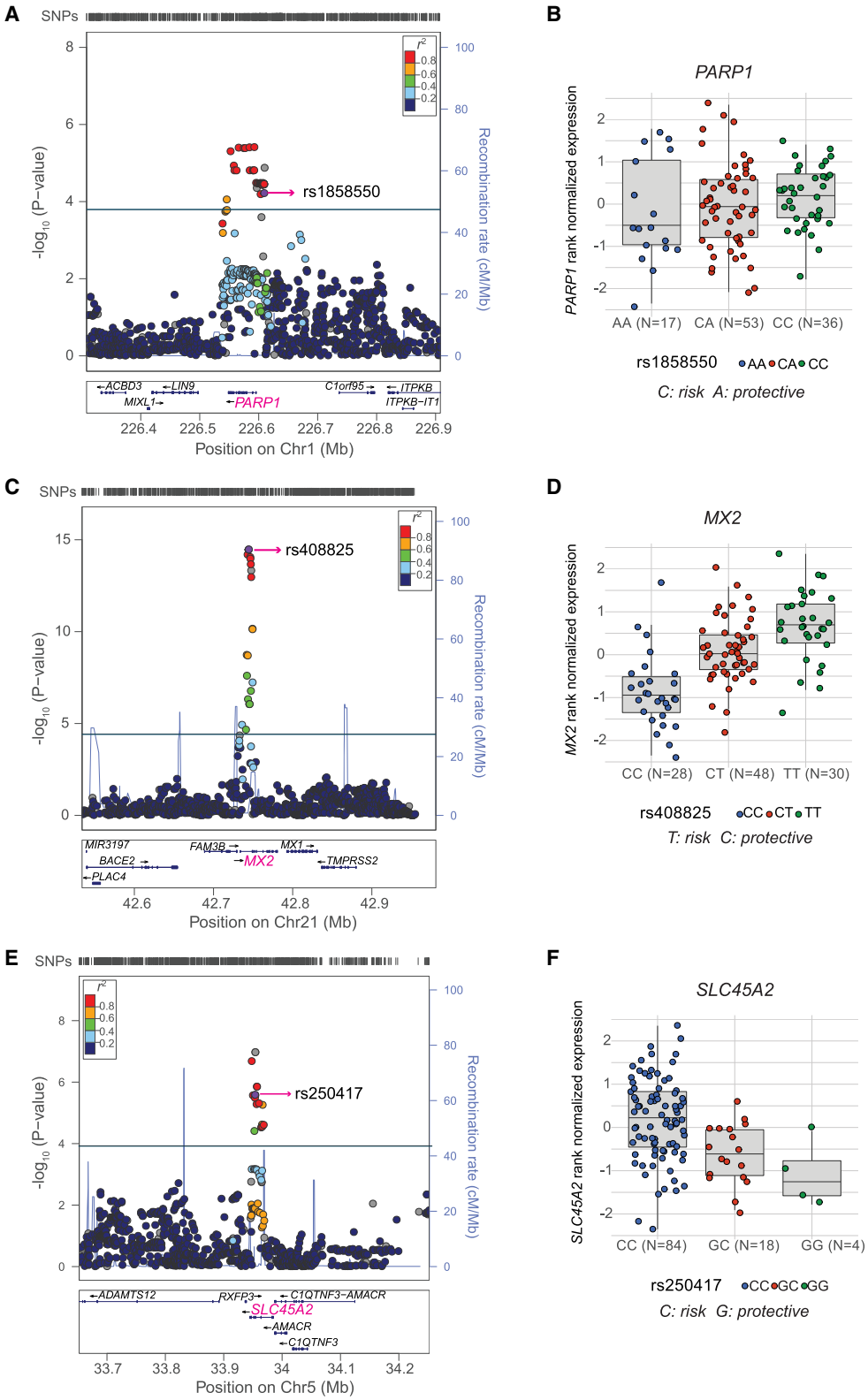

Figure 4. Melanoma GWAS signals colocalizing with melanocyte eQTLs. (*A*,*C*,*E*) LocusZoom plots present the nominal eQTL *P*-values of all tested local SNPs in 300- to 400-kb windows for three significant eQTL genes from three melanoma GWAS loci: (*A*) *PARP1*; (*C*) *MX2*; and (*E*) *SLC45A2*. The gene being measured is highlighted in pink, the index melanoma risk SNP is labeled and highlighted in purple, and r^2 (based on 1000G EUR) of all other SNPs to the index SNP is color-coded. SNPs with missing LD information with the index SNP are shown in gray. Horizontal lines are shown for nominal *P*-value cutoffs for significant eQTLs. Genomic coordinates are based on hg19. (*B*,*D*,*F*) Boxplots present melanocyte expression differences of each gene in relation to the genotypes of the index SNP. Melanoma risk and protective alleles are shown for each locus.

Table 3. Top melanoma TWAS genes using melanocyte eQTL as a reference set

Gene symbol	Chr	GWAS best SNP[a]	GWAS *Z*-score[1 b]	eQTL best SNP[c]	eQTL *Z*-score[d]	GWAS *Z*-score[2 e]	# of SNP[f]	# of weight[g]	Model[h]	TWAS *Z*-score	TWAS *P*-value[i]	FDR passed	GWAS locus[j]
MAFF	22	rs132985	−6.91	rs738322	−5.14	−6.69	340	2	enet	6.67	$\mathbf{2.60 \times 10^{-11}}$	Y	**22q13.1**
CTSS	1	rs12410869	−7.22	rs7521898	3.33	−6.42	315	6	enet	−6.32	$\mathbf{2.65 \times 10^{-10}}$	Y	**1q21.3**
ZFP90	16	rs7184977	5.07	rs11075688	1.06	3.96	319	319	blup	5.20	$\mathbf{1.95 \times 10^{-7}}$	Y	**New**
HEBP1	12	rs2111398	5.08	rs1684387	−5.32	5.08	605	5	lasso	−5.05	$\mathbf{4.48 \times 10^{-7}}$	Y	**New**
CASP8	2	rs10931936	5.50	rs3769823	−3.71	4.45	372	3	lasso	−4.61	$\mathbf{4.06 \times 10^{-6}}$	Y	**2q33-q34**
MSC	8	rs1481853	−4.65	rs6983160	−3.83	−4.60	553	2	lasso	4.60	$\mathbf{4.27 \times 10^{-6}}$	Y	**New**
CBWD1	9	rs661356	4.79	rs2992854	4.99	−1.96	462	1	lasso	−4.54	$\mathbf{5.52 \times 10^{-6}}$	Y	**New**
RP11-383H13.1	8	rs1481853	−4.65	rs6983160	−2.75	−4.60	693	4	lasso	4.50	$\mathbf{6.68 \times 10^{-6}}$	Y	**New**
GPRC5A	12	rs2111398	5.08	rs1684387	−4.02	5.08	584	11	enet	−4.19	2.80×10^{-5}	N	New
RBBP5	1	rs11240466	4.09	rs10900456	1.41	3.98	623	623	blup	3.59	3.37×10^{-4}	N	New
ATP6V1G2-DDX39B	6	rs2239704	4.55	rs2523504	3.51	3.54	278	1	lasso	3.54	3.97×10^{-4}	N	New
CDH1	16	rs7184977	5.07	rs4076177	2.71	4.52	345	345	blup	3.47	5.27×10^{-4}	N	New
CHCHD6	3	rs9851451	3.46	rs9822602	−2.73	3.44	635	3	lasso	−3.44	5.78×10^{-4}	N	New
CTD-2003C8.2	11	rs1554519	4.51	rs7932891	5.66	3.32	746	4	lasso	3.43	6.08×10^{-4}	N	New
UQCC1	20	rs2425025	6.95	rs6060369	−3.14	−3.00	353	353	blup	3.42	6.21×10^{-4}	N	New
PLXNA1	3	rs9851451	3.46	rs4679317	−2.20	3.42	484	2	lasso	−3.42	6.33×10^{-4}	N	New
PARP1	1	rs1865222	−7.12	rs3219090	−1.30	−6.86	404	404	blup	3.41	6.57×10^{-4}	N	1q42.12

[a]rsID of the most significant melanoma GWAS SNP for the TWAS gene after QC and IMPG imputation by FUSION program.
[b]GWAS *Z*-score of the most significant GWAS SNP in the locus.
[c]rsID of the best eQTL SNP in the locus.
[d]eQTL *Z*-score of the best eQTL SNP in the locus.
[e]GWAS *Z*-score for the best eQTL SNP.
[f]Number of SNPs in the locus.
[g]Weighted number of SNPs in the locus.
[h]Best performing model.
[i]TWAS *P*-value (genome-wide significant *P*-values are in bold).
[j]Previously identified GWAS locus (shaded) or newly identified by this study.

Consortium 2015) and Mixed population, and using SHAPEIT (Delaneau et al. 2011) for prephasing. Post-imputation genetic variants (single nucleotide variants [SNPs and small insertion-deletion (indel) polymorphisms) with MAF < 0.01 or imputation quality scores (*R*-squared) <0.3 were removed from the final analysis. Overall, ~713,000 genotypes were obtained, and 10,718,646 genotypes were further imputed. Due to the small sample size, we included all samples that passed genotyping QC but histologically carry a range of African and Asian ancestry measured by ADMIXTURE (Alexander et al. 2009) analysis, while accounting for ancestry in the further analyses as covariates. For eQTL analysis, we included the top three genotyping principal components as covariates. The principal components analysis for population substructure was performed using the *struct.pca* module of GLU (Wolpin et al. 2014), which is similar to EIGENSTRAT (Price et al. 2006).

RNA sequencing and data processing

Cells were harvested at log phase, and total RNA was isolated using a miRNeasy Mini kit (217004, Qiagen) in randomized batches. Poly(A) selected stranded mRNA libraries were constructed using Illumina TruSeq Stranded mRNA Sample Prep kits and sequenced on HiSeq 2500 using version 4 chemistry to achieve a minimum of 45 million 126-base paired-end reads (average of ~87.9 million reads). STAR (version 2.5.0b) (Dobin et al. 2013) was used for aligning reads to the human genomic reference (hg19) with the gene annotations from GENCODE Release 19 (https://www.gencodegenes.org/releases/19.html). RSEM (version 1.2.31, http://deweylab.github.io/RSEM/) was used to quantify the gene expression followed by the quantile normalization. Genes were selected based on expression thresholds of >0.5 RSEM and ≥6 reads in at least 10 samples. After processing, 19,608 genes were expressed above cutoff levels in primary melanocytes. For each gene, expression values were further inverse quantile normalized to a standard normal distribution across samples. To control for hidden batch effects and other confounding effects that could be reflected in the expression data, a set of covariates identified using the Probabilistic Estimation of Expression Residuals (PEER) method (Stegle et al. 2010) was calculated for the normalized expression matrices. The top 15 PEER factors were determined based on the sample size and optimizing for the number of eGenes discovered as suggested by the GTEX project (http://www.gtexportal.org) (15 factors for $N < 150$).

Identification of *cis*-eQTLs in primary melanocytes

Cis-eQTL analysis was performed closely following a recent standard procedure adopted by GTEx (The GTEx Consortium 2017; see Supplemental Material for details). In brief, *cis*-eQTL mapping was performed using FastQTL (Ongen et al. 2016), and nominal *P*-values were generated for genetic variants located within ±1 Mb of the TSSs for each gene tested. The beta distribution-adjusted empirical *P*-values from FastQTL were then used to calculate *q*-values (Storey and Tibshirani 2003), and a false discovery rate (FDR) threshold of ≤0.05 was applied to identify genes with a significant eQTL ("eGenes"). The effect size of the eQTLs was defined as the slope of the linear regression and is computed as the effect of the alternative allele (ALT) relative to the reference allele (REF). For each gene, a nominal *P*-value threshold was calculated, and variants with a nominal *P*-value below the gene-level threshold were considered as genome-wide significant *cis*-eQTL variants. The number of identified eGenes and significant eQTLs were approximately three times higher than those from data analyzed

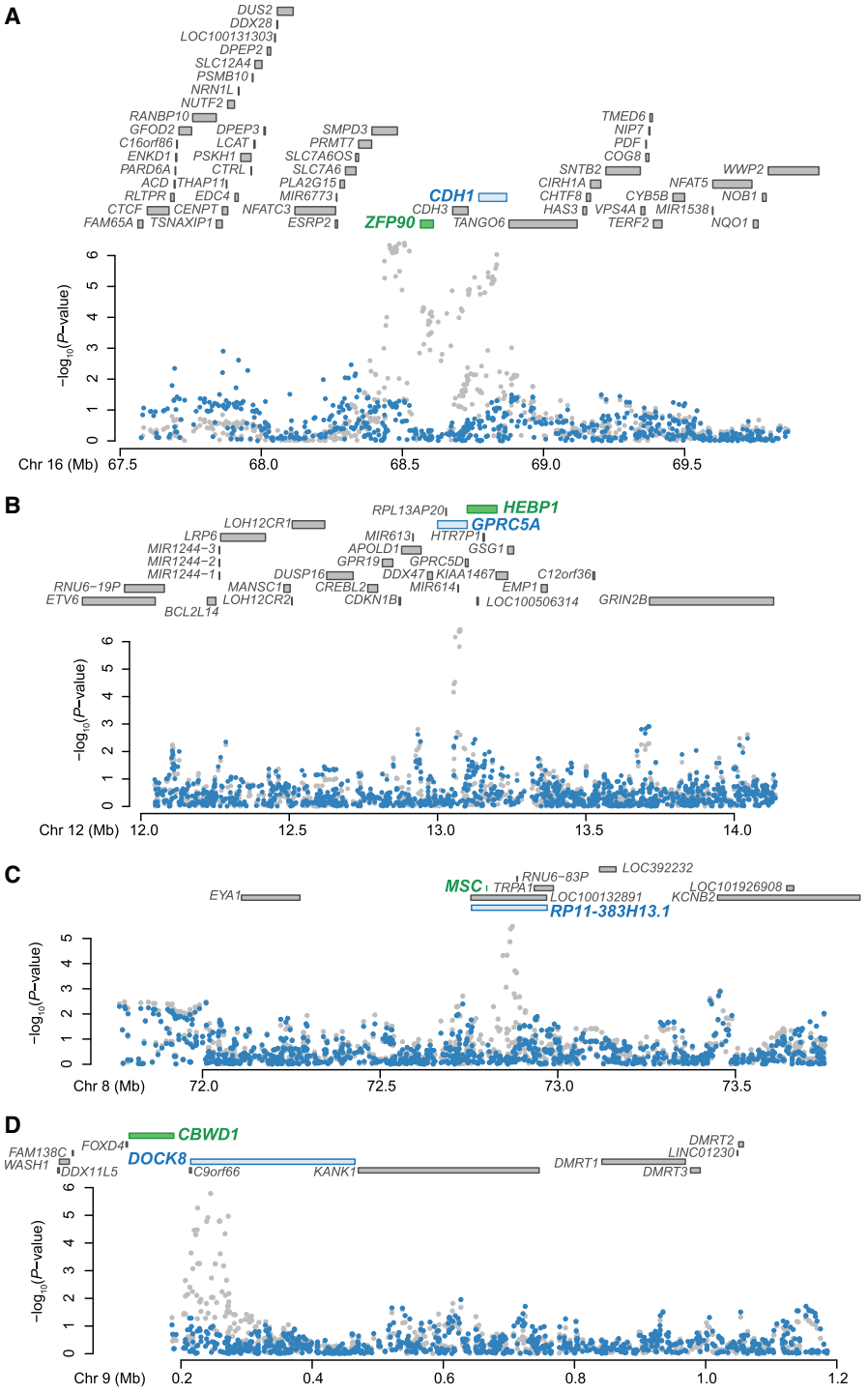

Figure 5. TWAS using melanocyte eQTL data as a reference set identified five new melanoma-associated genes in four new loci. (*A*) The new melanoma TWAS gene, *ZFP90* on Chromosome 16 (TWAS $P = 1.95 \times 10^{-7}$, TWAS $Z = 5.2$) is shown in green, along with a second marginally significant gene, *CDH1* ($P = 5.27 \times 10^{-4}$, $Z = 3.47$) in blue, and other annotated genes at the locus (coordinates are hg19). The Manhattan plot presents the melanoma GWAS *P*-values before (gray) and after (blue) conditioning on imputed melanocyte-specific gene expression of the gene in green (*ZFP90* in this locus). (*B*) A similar plot for the melanoma TWAS gene *HEBP1* (TWAS $P = 4.65 \times 10^{-7}$, TWAS $Z = -5.04$) and a second marginally significant gene, *GPRC5A* (TWAS $P = 2.8 \times 10^{-5}$, TWAS $Z = -4.19$) on chromosome band 12p13.1. (*C*) A similar plot for two new melanoma TWAS genes, *MSC* ($P = 4.27 \times 10^{-6}$, $Z = 4.6$) and *RP11-383H13.1* ($P = 6.68 \times 10^{-6}$, $Z = 4.5$) on chromosome band 8q13.3. The Manhattan plot shows the melanoma GWAS *P*-values before (gray) and after (blue) conditioning on imputed melanocyte-specific gene expression of *MSC*. (*D*) A similar plot of new melanoma TWAS gene, *CBWD1* ($P = 5.52 \times 10^{-6}$, $Z = -4.54$) and a marginally significant gene, *DOCK8* ($P = 2.7 \times 10^{-3}$, $Z = 2.99$) on chromosome band 9p24.3.

without using PEER factors as covariates (Supplemental Table S2). Application of PEER factors almost doubled the number of eGenes known to be related to pigmentation phenotypes (0.8% vs. 1.5%; Fisher's exact test *P*-value = 0.0335).

Pairwise eQTL sharing between primary melanocytes and 44 GTEx tissues

To test the sharing of all significant SNP-gene pairs of our melanocyte eQTL study with those identified in 44 tissue types by GTEx (FDR < 0.05) (The GTEx Consortium 2013; The GTEx Consortium 2017), we calculated pairwise π_1 statistics (indicating the proportion of true positives) using Storey's QVALUE software (https://github.com/StoreyLab/qvalue) (Storey and Tibshirani 2003). A heat map was drawn based on the pairwise π_1 values, where higher π_1 values indicate an increased replication of eQTLs. Tissues are grouped using hierarchical clustering on rows and columns separately with a distance metric of $1 - \rho$, where ρ is the Spearman's correlation of π_1 values. π_1 is only calculated when the gene is expressed and testable in both the discovery and the replication tissues.

Identification of *trans*-eQTLs in primary melanocytes

Trans-eQTL analysis was performed for SNPs that are located over 5 Mb away from the TSS of each gene or on a different chromosome. Genes of mappability <0.8 or overlapping low complexity regions defined by the RepeatMasker (Smit et al. 2013–2015) library were excluded from the analysis. The nominal *P*-values for gene-SNP pairs in *trans*-eQTL analysis were calculated using Matrix-eQTL (Shabalin 2012). We performed a permutation test followed by the Benjamini–Hochberg procedure to identify significant *trans*-eQTLs following our previous approach (Shi et al. 2014). For each nominal *P*-value threshold p, we calculated the number of genes (denoted as $N_1(p)$) that have at least one SNP in its *trans* region with nominal *P*-value less than the threshold p. Here, $N_1(p)$ denotes the number of *trans*-eQTL genes at *P*-value threshold p. Next, we performed 100 permutations to estimate the number of genes (denoted as $N_0(p)$) detected to have *trans*-eQTL signals at nominal *P*-value p under the global null hypothesis. By definition, one can calculate the FDR as $\text{FDR} = N_0(p)/N_1(p)$. We chose $p = 3.25 \times 10^{-11}$ to control FDR at a desired level of 0.1.

Identifying *cis*-mediators for *trans*-eQTLs in primary melanocytes

We applied the Genomic Mediation analysis with Adaptive Confounding adjustment (Yang et al. 2017) algorithm to identify *cis*-mediators for *trans*-eQTLs in primary melanocyte eQTL data. Only the trios with evidence of both *cis* and *trans* association were kept. The *cis*-eQTL SNP with the smallest *P*-value for each gene (eQTL *FDR* < 0.05) and *trans*-association *P*-value $<10^{-5}$ was selected as one trio. Up to five PEER factors and other covariates (top 10 genotype PCs) were adjusted. One hundred thousand permutations for testing mediation were performed, and trios with suggestive mediation were reported using a mediation *P*-value threshold <0.05.

Allele-specific expression

ASE analysis was performed based on the GATK best practices pipeline in allelic expression analysis published by the Broad Institute (Castel et al. 2015; see Supplemental Material for details). Following quality control, we evaluated the significance of allelic imbalance using a binomial test in each individual level, comparing the observed to the subject- and genotype-specific expected allele ratios (Ongen et al. 2014), while accounting for base-specific mapping bias (Lappalainen et al. 2013; Zhang et al. 2018). In addition, the effect size of allelic expression (AE, defined as |0.5-Reference ratio|) was calculated. We defined significant ASE genes as genes with at least one genetic variant exhibiting a minimum effect size of 0.15 or a significant difference from the expected allele ratio of 0.5 at FDR < 0.05 (calculated using the Benjamini and Hochberg approach) (Benjamini and Hochberg 1995) in one or more individuals. Significant ASE genes were then grouped into melanocyte eGenes and non-eGenes, and |Mean AE| values as well as percentage of individuals displaying allelic imbalance were compared between two groups (Wilcoxon rank-sum and signed ranked tests).

Assessing enrichment in putative functional elements

To assess the enrichment of *cis*-eQTL in putative functional elements of primary melanocytes, we collected the DNase-seq and ChIP-seq data from the Epigenome Roadmap Project (http://www.roadmapepigenomics.org) (Roadmap Epigenomics Consortium et al. 2015). For each putative functional element, we merged peak callings from all samples into one, and all the significant melanocyte eQTL SNP-Gene pairs were used for the enrichment analyses using a similar method to a recent publication (Zhang et al. 2018). Briefly, we performed randomizations for testing whether an eQTL SNP set is enriched for given histone mark regions. Note the following procedure controls for the distribution of minor allele frequencies of a given eQTL SNP set: (1) For K eQTL SNPs, we determined the number (denoted as X_0) of eQTL SNPs functionally related with the histone mark. (2) We randomly sampled 10,000 SNP sets. Each SNP set had K SNPs in linkage equilibrium, with minor allele frequency distribution similar to the original K eQTL SNPs. For the n^{th} sampled SNP set, we calculated the number (denoted as x_n) of SNPs functionally related with the histone mark. We had $\{x_1, \ldots, x_{10,000}\}$, corresponding to the sampled 10,000 SNP sets. (3) Enrichment fold change was calculated as $\text{FC} = (X_0/(\sum_{n=1}^{10,000} x_n/10,000))$, where the denominator represented the average number of SNPs functionally related with the histone mark under the null hypothesis. The *P*-value for enrichment was calculated as $P = \{n: x_n \geq X_0\}/10,000$, i.e., the proportion of SNP sets functionally more related with the histone mark than the given eQTL SNP set. If $x_n < X_0$ for all sample SNP sets, we reported the *P* value as $P < 10^{-4}$. In addition, we also assessed enrichment of *cis*-eQTLs in different genomic regions including 5′/3′ UTR, promoter, exon, intron, and intergenic and lncRNA regions as described in the R annotatr package (https://github.com/hhabra/annotatr) (Cavalcante and Sartor 2017).

Enrichment of melanoma GWAS variants in eQTLs

Two methods were used to evaluate if the melanoma GWAS variants were enriched in eQTLs of different data sets. First, QQ plots were used to show the differences in melanoma association *P*-values from the most recent meta-analysis (Law et al. 2015) between the significant eQTL SNPs and non-eQTL SNPs. For all GWAS variants, we first performed LD pruning using PLINK ($r^2 = 0.1$ and window size 500 kb) (Purcell et al. 2007). If a pruned SNP is an eQTL or is in LD ($r^2 > 0.8$) with an eQTL SNP, the SNP is classified as an eQTL SNP. Otherwise, it is classified as a non-eQTL SNP. The lambda values were estimated using the "estlambda2" function in R package "QQperm" (https://github.com/cran/QQperm). For the second method, a simulation procedure was applied to identify overlap and test for enrichment of eQTLs in melanoma GWAS SNPs closely following a previously published method (Hannon et al. 2016; see Supplemental Material for details).

Melanoma heritability enrichment of tissue-specific genes

We used stratified LD score regression implemented in the LDSC program (https://github.com/bulik/ldsc) to estimate the enrichment of melanoma heritability for SNPs around tissue and cell-type–specific genes as described previously (Finucane et al. 2018). In brief, to reduce batch effects, RNA-seq data for both GTEx tissues and our melanocytes were quantified as RPKM using RNA-SeQC (v1.18) (DeLuca et al. 2012) followed by quantile normalization. To define the tissue-specific genes, we calculated the *t*-statistic of each gene for a given tissue, excluding all samples from the same tissue category (we treated the tissue category for melanocytes as "Skin") (see Supplemental Table S8 and Supplemental Material for details). We selected the top 1000, 2000, and 4000 tissue-specific genes by *t*-statistic, added a 100-kb window around their transcribed regions to define tissue-specific genome annotation, and applied stratified LD score regression on a joint SNP annotation to estimate the heritability enrichment against the melanoma GWAS meta-analysis (Law et al. 2015). The results using the top 4000 tissue-specific genes showed significant enrichment (FDR < 0.05) for melanocyte and all three tissue types in the "Skin" category. The overall pattern was consistently observed in results using 2000 and 1000 genes, while melanocyte was significant in results from 2000 but not in those from 1000 genes (Supplemental Table S8; Supplemental Fig. S7). Importantly, some of the top enriched tissues outside of the "Skin" category (e.g., Colon_Transverse) displayed high median expression level correlation with melanocytes (Pearson's $r = 0.95$ between melanocyte and Colon_Transverse) (Supplemental Fig. S10).

Colocalization analysis of GWAS and eQTL data

We performed colocalization analysis for 20 GWAS loci from the most recent GWAS meta-analysis using CAusal Variants Identification in Associated Regions (eCAVIAR, http://genetics.cs.ucla.edu/caviar/index.html) (Hormozdiari et al. 2016). For each locus, both GWAS and eQTL summary statistics (from our melanocyte data set and two GTEx skin tissues) of selected variants in that locus were extracted as the input for eCAVIAR. We selected 50 SNPs both upstream of and downstream from the GWAS lead SNP for each GWAS locus. We computed the CLPP score with a maximum number of two causal SNPs in each locus. We used a CLPP > 1% (0.01) cutoff for colocalization. Thus, for a given GWAS variant (either the lead SNP itself or the SNPs in near perfect

LD with the lead SNP using the cutoff $r^2 > 0.99$), an eGene with a CLPP score above the colocalization cutoff is considered a target gene. We also highlight the eGenes with CLPP > 0.05 as they were more robust across minor changes in analyses criteria compared to those on the borderline (between 0.01 and 0.05) in our analyses.

Performing TWASs with GWAS summary statistics

We performed 45 transcriptome-wide association studies by predicting the function/molecular phenotypes into GWAS using melanoma GWAS summary statistics and both GTEx and melanocyte RNA-seq expression data. TWAS/FUSION (http://gusevlab.org/projects/fusion/) was used to perform the TWAS analysis, allowing for multiple prediction models, independent reference LD, additional feature statistics, and cross-validation results (Gusev et al. 2016). In brief, we collected the summary statistics including no significance thresholding from the most recently published cutaneous melanoma meta-analysis (Law et al. 2015). The precomputed expression reference weights for GTEx gene expression (V6) RNA-seq across 44 tissue types were downloaded from the TWAS/FUSION website (http://gusevlab.org/projects/fusion/). We computed functional weights from our melanocyte RNA-seq data one gene at a time. Genes that failed quality control during a heritability check (using minimum heritability *P*-value of 0.01) were excluded from the further analyses. We restricted the *cis*-locus to 500 kb on either side of the gene boundary. A genome-wide significance cutoff (TWAS *P*-value < 0.05/number of genes tested) was applied to the final TWAS result. Multiple associated features in a locus were observed, and thus we performed the joint/conditional analysis to identify which are conditionally independent for each melanoma susceptibility locus using a permutation test with a maximum of 100,000 permutations and initiate permutation *P*-value threshold of 0.05 for each feature. We also checked how much GWAS signal remained after conditioning on imputed expression levels of each associated feature by using "FUSION.post_process.R" script.

Other analyses

IRF4 motif enrichment analyses were performed using the AME module in The MEME Suite (http://meme-suite.org) (Bailey et al. 2009) and inputted shuffled sequences as control. *IRF4* motifs were downloaded from HOCOMOCO v10 database (http://hocomoco.autosome.ru/motif/IRF4_HUMAN.H10MO.C) (Kulakovskiy et al. 2013). All the statistical analyses were performed in R (R Core Team 2018).

Reference genome build statement

Our data were mapped to GRCh37/hg19 to allow maximum comparability with the GTEx and other public data sets we used in the manuscript. Mapping the reads of our data to the most current GRCh38 would not significantly affect the global eQTL analyses and conclusions of the current paper, and only minor differences are expected.

Data access

Genotyping and RNA sequencing data of 106 primary human melanocytes as well as processed eQTL data from this study have been submitted to the database of Genotypes and Phenotypes (dbGAP, https://www.ncbi.nlm.nih.gov/gap) under accession number phs001500.v1.p1.

Members of the Melanoma Meta-Analysis Consortium

Matthew H. Law,[11] D. Timothy Bishop,[12] Jeffrey E. Lee,[13] Myriam Brossard,[14,15] Nicholas G. Martin,[16] Eric K. Moses,[17] Fengju Song,[18] Jennifer H. Barrett,[12] Rajiv Kumar,[19] Douglas F. Easton,[20] Paul D.P. Pharoah,[21] Anthony J. Swerdlow,[22,23] Katerina P. Kypreou,[24] John C. Taylor,[12] Mark Harland,[12] Juliette Randerson-Moor,[12] Lars A. Akslen,[25,26] Per A. Andresen,[27] Marie-Françoise Avril,[28] Esther Azizi,[29,30] Giovanna Bianchi Scarrà,[31,32] Kevin M. Brown,[33] Tadeusz Dębniak,[34] David L. Duffy,[16] David E. Elder,[35] Shenying Fang,[13] Eitan Friedman,[30] Pilar Galan,[36] Paola Ghiorzo,[31,32] Elizabeth M. Gillanders,[37] Alisa

[11]Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4006, Australia
[12]Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, LS9 7TF, UK
[13]Department of Surgical Oncology, University of Texas MD Anderson Cancer Center, Houston, TX, 77030, USA
[14]INSERM, UMR 946, Genetic Variation and Human Diseases Unit, 75013 Paris, France
[15]Institut Universitaire d'Hématologie, Université Paris Diderot, Sorbonne Paris Cité, 75013 Paris, France
[16]Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4006, Australia
[17]Centre for Genetic Origins of Health and Disease, Faculty of Medicine, Dentistry and Health Sciences, University of Western Australia, Perth, Western Australia, 6009, Australia
[18]Department 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 300060, China
[19]Division of Molecular Genetic Epidemiology, German Cancer Research Center, 69120 Heidelberg, Germany
[20]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB2 1TN, UK
[21]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge CB2 1TN, UK
[22]Division of Genetics and Epidemiology, The Institute of Cancer Research, London SW3 6JB, UK
[23]Division of Breast Cancer Research, The Institute of Cancer Research, London SW3 6JB, UK
[24]Department of Dermatology, University of Athens School of Medicine, Andreas Sygros Hospital, Athens 161 21, Greece
[25]Centre for Cancer Biomarkers (CCBIO), Department of Clinical Medicine, University of Bergen, 5007 Bergen, Norway
[26]Department of Pathology, Haukeland University Hospital, 5021 Bergen, Norway
[27]Department of Pathology, Molecular Pathology, Oslo University Hospital, Rikshospitalet, 0372 Oslo, Norway
[28]Assistance Publique-Hôpitaux de Paris, Hôpital Cochin, Service de Dermatologie, Université Paris Descartes, 75006 Paris, France
[29]Department of Dermatology, Sheba Medical Center, Tel Hashomer, Sackler Faculty of Medicine, Tel Aviv 6997801, Israel
[30]Oncogenetics Unit, Sheba Medical Center, Tel Hashomer, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 6997801, Israel
[31]Department of Internal Medicine and Medical Specialties, University of Genoa, 16126 Genova GE, Italy
[32]Laboratory of Genetics of Rare Cancers, Istituto di Ricovero e Cura a Carattere Scientifico Azienda Ospedaliera Universitaria (IRCCS AOU) San Martino l'Istituto Scientifico Tumori Istituto Nazionale per la Ricerca sul Cancro, 16132 Genova GE, Italy
[33]Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health, Bethesda, MD 20892, USA
[34]International Hereditary Cancer Center, Pomeranian Medical University, 70-204 Szczecin, Poland
[35]Department of Pathology and Laboratory Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104, USA
[36]Université Paris 13, Equipe de Recherche en Epidémiologie Nutritionnelle (EREN), Centre de Recherche en Epidémiologie et Statistiques, INSERM U1153, Institut National de la Recherche Agronomique (INRA) U1125, Conservatoire National des Arts et Métiers, Communauté d'Université Sorbonne Paris Cité, 93000 Bobigny, France

M. Goldstein,[33] Nelleke A. Gruis,[38] Johan Hansson,[39] Per Helsing,[40] Marko Hočevar,[41] Veronica Höiom,[39] Christian Ingvar,[42] Peter A. Kanetsky,[43] Wei V. Chen,[44] Maria Teresa Landi,[33] Julie Lang,[45] G. Mark Lathrop,[46] Jan Lubiński,[34] Rona M. Mackie,[45,47] Graham J. Mann,[48] Anders Molven,[26,49] Grant W. Montgomery,[50] Srdjan Novaković,[51] Håkan Olsson,[52,53] Susana Puig,[54,55] Joan Anton Puig-Butille,[54,55] Wenting Wu,[56,57] Abrar A. Qureshi,[58] Graham L. Radford-Smith,[59,60,61] Nienke van der Stoep,[62] Remco van Doorn,[38] David C. Whiteman,[63] Jamie E. Craig,[64] Dirk Schadendorf,[65,66] Lisa A. Simms,[57] Kathryn P. Burdon,[67] Dale R. Nyholt,[50,68] Karen A. Pooley,[20] Nick Orr,[69] Alexander J. Stratigos,[24] Anne E. Cust,[70] Sarah V. Ward,[17] Nicholas K. Hayward,[71] Jiali Han,[56,57] Hans-Joachim Schulze,[72] Alison M. Dunning,[21] Julia A. Newton Bishop,[12] Florence Demenais,[14,15] Christopher I. Amos,[73] Stuart MacGregor,[11] and Mark M. Iles[12]

Members of the NISC Comparative Sequencing Program

Beatrice B. Barnabas,[74] Gerard G. Bouffard,[74] Shelis Y. Brooks,[74] Holly Coleman,[74] Lyudmila Dekhtyar,[74] Xiaobin Guan,[74] Joel Han,[74] Shi-ling Ho,[74] Richelle Legaspi,[74] Quino L. Maduro,[74] Catherine A. Masiello,[74] Jennifer C. McDowell,[74] Casandra Montemayor,[74] James C. Mullikin,[74] Morgan Park,[74] Nancy L. Riebow,[74] Karen Schandler,[74] Brian Schmidt,[74] Christina Sison,[74] Raymond Smith,[74] Sirintorn Stantripop,[74] James W. Thomas,[74] Pamela J. Thomas,[74] Meghana Vemulapalli,[74] and Alice C. Young[74]

Acknowledgments

This work has been supported by the Intramural Research Program (IRP) of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health (NIH). The content of this publication does not necessarily reflect the views or policies of the US Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government. M.M.I. was supported by Cancer Research UK (CRUK) C588/A19167 and NIH CA083115. This research has been conducted using the UK Biobank Resource under Application Number 3071. This work utilized the computational resources of the NIH high-performance computational capabilities Biowulf cluster (http://hpc.nih.gov). We acknowledge contributions to human melanocyte genotyping from The National Cancer Institute Cancer Genomics Research Laboratory (CGR) and to RNA sequencing from The NIH Intramural Sequencing Center (NISC), and NCI Center for Cancer Research Sequencing Facility (CCR-SF) and the Yale University Skin SPORE Specimen Resource Core. S.M. is supported by an Australian Research Council Fellowship. We thank A. Vu, L. Mehl, and H. Kong for proofreading the manuscript. The previously published melanoma meta-analysis (Law et al. 2015) makes use of data from two dbGap data sets (accession numbers phs000519.v1.p1 and phs000187.v1.p1). The Leeds component of the meta-analysis was funded by the European Commission under the 6th Framework Programme, contract no. LSHC-CT-2006-018702; by Cancer Research UK Programme Awards, C588/A4994 and C588/A10589, and Cancer Research UK Project Grant C8216/A6129; and by US National Institutes of Health R01 ROI CA83115. The Cambridge component of the meta-analysis was supported by Cancer Research UK grants

[37]Inherited Disease Research Branch, National Human Genome Research Institute, US National Institutes of Health, Baltimore, MD 21224, USA
[38]Department of Dermatology, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands
[39]Department of Oncology-Pathology, Karolinska Institutet, Karolinska University Hospital, Stockholm, 171 76 Solna, Sweden
[40]Department of Dermatology, Oslo University Hospital, Rikshospitalet, 0372 Oslo, Norway
[41]Department of Surgical Oncology, Institute of Oncology Ljubljana, 1000 Ljubljana, Slovenia
[42]Department of Surgery, Clinical Sciences, Lund University, P663+Q9 Lund, Sweden
[43]Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL 33612 USA
[44]Department of Genetics, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[45]Department of Medical Genetics, University of Glasgow, Glasgow G12 8QQ, UK
[46]McGill University and Génome Québec Innovation Centre, Montreal, QC H3A 0G1, Canada
[47]Department of Public Health, University of Glasgow, Glasgow G12 8QQ, UK
[48]Centre for Cancer Research, University of Sydney at Westmead, Millennium Institute for Medical Research and Melanoma Institute Australia, Sydney, NSW 2145, Australia
[49]Gade Laboratory for Pathology, Department of Clinical Medicine, University of Bergen, 5007 Bergen, Norway
[50]Molecular Biology, the University of Queensland, Brisbane, QLD 4072, Australia
[51]Department of Molecular Diagnostics, Institute of Oncology Ljubljana, 1000 Ljubljana, Slovenia
[52]Department of Oncology/Pathology, Clinical Sciences, Lund University, P663+Q9 Lund, Sweden
[53]Department of Cancer Epidemiology, Clinical Sciences, Lund University, P663+Q9 Lund, Sweden
[54]Melanoma Unit, Departments of Dermatology, Biochemistry and Molecular Genetics, Hospital Clinic, Institut d'Investigacions Biomèdica August Pi Suñe, Universitat de Barcelona, 08007 Barcelona, Spain
[55]Centro de Investigación Biomédica en Red (CIBER) de Enfermedades Raras, Instituto de Salud Carlos III, Planta 0 28029 Madrid, Spain
[56]Department of Epidemiology, Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, IN 46202, USA
[57]Melvin and Bren Simon Cancer Center, Indiana University, Indianapolis, IN 46202, USA
[58]Department of Dermatology, Warren Alpert Medical School of Brown University, Providence, RI 02903, USA
[59]Inflammatory Bowel Diseases, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4006, Australia
[60]Department of Gastroenterology and Hepatology, Royal Brisbane and Women's Hospital, Brisbane, QLD 4029, Australia
[61]University of Queensland School of Medicine, Herston Campus, Brisbane, QLD 4072, Australia
[62]Department of Clinical Genetics, Center of Human and Clinical Genetics, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands
[63]Cancer Control Group, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4006, Australia
[64]Department of Ophthalmology, Flinders University, Adelaide, SA 5042, Australia
[65]Department of Dermatology, University Hospital Essen, 45147 Essen, Germany
[66]German Consortium for Translational Cancer Research (DKTK), 69120 Heidelberg, Germany
[67]Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS 7005, Australia
[68]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD 4000, Australia
[69]Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London SM2 5NG, UK
[70]Cancer Epidemiology and Services Research, Sydney School of Public Health, University of Sydney, Sydney, NSW 2006, Australia
[71]Oncogenomics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4006, Australia
[72]Department of Dermatology, Fachklinik Hornheide, Institute for Tumors of the Skin at the University of Münster, 48149 Münster, Germany
[73]Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Hanover, NH 03755, USA
[74]NIH Intramural Sequencing Center, Bethesda, MD 20892, USA

C490/A10124 (SEARCH) and C1287/A10118 (MAPLES). The WAMHS data collection for the meta-analysis was supported by funding from the Scott Kirkbride Melanoma Research Centre and The University of Western Australia; we thank the Western Australian DNA Bank for assistance and support with biospecimens, and the Ark for informatics support. We also thank The Western Australian Cancer Registry for their assistance. For the University of Texas MD Anderson Cancer Center data used in the meta-analysis, research support to collect data and develop an application to support this project was provided by 3P50CA093459, 5P50CA097007, 5R01ES011740, and 5R01CA133996.

References

The 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature* **526:** 68–74.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19:** 1655–1664.

Amos CI, Wang LE, Lee JE, Gershenwald JE, Chen WV, Fang S, Kosoy R, Zhang M, Qureshi AA, Vattathil S, et al. 2011. Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. *Hum Mol Genet* **20:** 5012–5023.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37:** W202–W208.

Barrett JH, Iles MM, Harland M, Taylor JC, Aitken JF, Andresen PA, Akslen LA, Armstrong BK, Avril MF, Azizi E, et al. 2011. Genome-wide association study identifies three new melanoma susceptibility loci. *Nat Genet* **43:** 1108–1113.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol* **57:** 289–300.

Bishop DT, Demenais F, Iles MM, Harland M, Taylor JC, Corda E, Randerson-Moor J, Aitken JF, Avril MF, Azizi E, et al. 2009. Genome-wide association study identifies three loci associated with melanoma risk. *Nat Genet* **41:** 920–925.

Brown KM, Macgregor S, Montgomery GW, Craig DW, Zhao ZZ, Iyadurai K, Henders AK, Homer N, Campbell MJ, Stark M, et al. 2008. Common sequence variants on 20q11.22 confer melanoma susceptibility. *Nat Genet* **40:** 838–840.

Buil A, Brown AA, Lappalainen T, Vinuela A, Davies MN, Zheng HF, Richards JB, Glass D, Small KS, Durbin R, et al. 2015. Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat Genet* **47:** 88–91.

Candille SI, Absher DM, Beleza S, Bauchet M, McEvoy B, Garrison NA, Li JZ, Myers RM, Barsh GS, Tang H, et al. 2012. Genome-wide association studies of quantitatively measured skin, hair, and eye pigmentation in four European populations. *PLoS One* **7:** e48294.

Castel SE, Levy-Moonshine A, Mohammadi P, Banks E, Lappalainen T. 2015. Tools and best practices for data processing in allelic expression analysis. *Genome Biol* **16:** 195.

Cavalcante RG, Sartor MA. 2017. annotatr: genomic regions in context. *Bioinformatics* **33:** 2381–2383.

Choi J, Xu M, Makowski MM, Zhang T, Law MH, Kovacs MA, Granzhan A, Kim WJ, Parikh H, Gartside M, et al. 2017. A common intronic variant of *PARP1* confers melanoma risk and mediates melanocyte growth via regulation of *MITF*. *Nat Genet* **49:** 1326–1335.

Costin GE, Hearing VJ. 2007. Human skin pigmentation: melanocytes modulate skin color in response to stress. *FASEB J* **21:** 976–994.

Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, et al. 2016. Next-generation genotype imputation service and methods. *Nat Genet* **48:** 1284–1287.

Delaneau O, Marchini J, Zagury JF. 2011. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9:** 179–181.

DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, Reich M, Winckler W, Getz G. 2012. RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* **28:** 1530–1532.

Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M, et al. 2009. Common regulatory variation impacts gene expression in a cell type–dependent manner. *Science* **325:** 1246–1250.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29:** 15–21.

Duffy DL, Zhao ZZ, Sturm RA, Hayward NK, Martin NG, Montgomery GW. 2010. Multiple pigmentation gene polymorphisms account for a substantial proportion of risk of cutaneous malignant melanoma. *J Invest Dermatol* **130:** 520–528.

The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489:** 57–74.

Eriksson N, Macpherson JM, Tung JY, Hon LS, Naughton B, Saxonov S, Avey L, Wojcicki A, Pe'er I, Mountain J. 2010. Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet* **6:** e1000993.

Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, Ellis P, Langford C, Vannberg FO, Knight JC. 2012. Genetics of gene expression in primary immune cells identifies cell type–specific master regulators and roles of HLA alleles. *Nat Genet* **44:** 502–510.

Falchi M, Bataille V, Hayward NK, Duffy DL, Bishop JA, Pastinen T, Cervino A, Zhao ZZ, Deloukas P, Soranzo N, et al. 2009. Genome-wide association study identifies variants at 9p21 and 22q13 associated with development of cutaneous nevi. *Nat Genet* **41:** 915–919.

Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, Anttila V, Xu H, Zang C, Farh K, et al. 2015. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47:** 1228–1235.

Finucane HK, Reshef YA, Anttila V, Slowikowski K, Gusev A, Byrnes A, Gazal S, Loh PR, Lareau C, Shoresh N, et al. 2018. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet* **50:** 621–629.

Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, Eyler AE, Denny JC, GTEx Consortium, Nicolae DL, et al. 2015. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* **47:** 1091–1098.

The GTEx Consortium. 2013. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45:** 580–585.

The GTEx Consortium. 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348:** 648–660.

The GTEx Consortium. 2017. Genetic effects on gene expression across human tissues. *Nature* **550:** 204–213.

Gudbjartsson DF, Sulem P, Stacey SN, Goldstein AM, Rafnar T, Sigurgeirsson B, Benediktsdottir KR, Thorisdottir K, Ragnarsson R, Sveinsdottir SG, et al. 2008. *ASIP* and *TYR* pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. *Nat Genet* **40:** 886–891.

Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, Jansen R, de Geus EJ, Boomsma DI, Wright FA, et al. 2016. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48:** 245–252.

Halaban R, Cheng E, Smicun Y, Germino J. 2000. Deregulated E2F transcriptional activity in autonomously growing melanoma cells. *J Exp Med* **191:** 1005–1016.

Han J, Kraft P, Nan H, Guo Q, Chen C, Qureshi A, Hankinson SE, Hu FB, Duffy DL, Zhao ZZ, et al. 2008. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet* **4:** e1000074.

Hannon E, Spiers H, Viana J, Pidsley R, Burrage J, Murphy TM, Troakes C, Turecki G, O'Donovan MC, Schalkwyk LC, et al. 2016. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat Neurosci* **19:** 48–54.

Hormozdiari F, van de Bunt M, Segre AV, Li X, Joo JWJ, Bilow M, Sul JH, Sankararaman S, Pasaniuc B, Eskin E. 2016. Colocalization of GWAS and eQTL signals detects target genes. *Am J Hum Genet* **99:** 1245–1260.

Huber M, Lohoff M. 2014. IRF4 at the crossroads of effector T-cell fate decision. *Eur J Immunol* **44:** 1886–1895.

Hysi PG, Valdes AM, Liu F, Furlotte NA, Evans DM, Bataille V, Visconti A, Hemani G, McMahon G, Ring SM, et al. 2018. Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability. *Nat Genet* **50:** 652–656.

Jacobs LC, Hamer MA, Gunn DA, Deelen J, Lall JS, van Heemst D, Uh HW, Hofman A, Uitterlinden AG, Griffiths CE, et al. 2015. A genome-wide association study identifies the skin color genes *IRF4*, *MC1R*, *ASIP*, and *BNC2* influencing facial pigmented spots. *J Invest Dermatol* **135:** 1735–1742.

Kane M, Zang TM, Rihn SJ, Zhang F, Kueck T, Alim M, Schoggins J, Rice CM, Wilson SJ, Bieniasz PD. 2016. Identification of interferon-stimulated genes with antiretroviral activity. *Cell Host Microbe* **20:** 392–405.

Kilpinen H, Goncalves A, Leha A, Afzal V, Alasoo K, Ashford S, Bala S, Bensaddek D, Casale FP, Culley OJ, et al. 2017. Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* **546:** 370–375.

Kim-Hellmuth S, Bechheim M, Putz B, Mohammadi P, Nedelec Y, Giangreco N, Becker J, Kaiser V, Fricker N, Beier E, et al. 2017. Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nat Commun* **8:** 266.

Kulakovskiy IV, Medvedeva YA, Schaefer U, Kasianov AS, Vorontsov IE, Bajic VB, Makeev VJ. 2013. HOCOMOCO: a comprehensive collection

of human transcription factor binding sites models. *Nucleic Acids Res* **41:** D195–D202.
Lappalainen T, Sammeth M, Friedlander MR, 't Hoen PA, Monlong J, Rivas MA, Gonzalez-Porta M, Kurbatova N, Griebel T, Ferreira PG, et al. 2013. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501:** 506–511.
Law MH, Bishop DT, Lee JE, Brossard M, Martin NG, Moses EK, Song F, Barrett JH, Kumar R, Easton DF, et al. 2015. Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. *Nat Genet* **47:** 987–995.
Liu J, Fukunaga-Kalabis M, Li L, Herlyn M. 2014. Developmental pathways activated in melanocytes and melanoma. *Arch Biochem Biophys* **563:** 13–21.
Liu F, Visser M, Duffy DL, Hysi PG, Jacobs LC, Lao O, Zhong K, Walsh S, Chaitanya L, Wollstein A, et al. 2015. Genetics of skin color variation in Europeans: genome-wide association studies with functional follow-up. *Hum Genet* **134:** 823–835.
Liu B, Pjanic M, Wang T, Nguyen T, Gloudemans M, Rao A, Castano VG, Nurnberg S, Rader DJ, Elwyn S, et al. 2018. Genetic regulatory mechanisms of smooth muscle cells map to coronary artery disease risk loci. *Am J Hum Genet* **103:** 377–388.
Macgregor S, Montgomery GW, Liu JZ, Zhao ZZ, Henders AK, Stark M, Schmid H, Holland EA, Duffy DL, Zhang M, et al. 2011. Genome-wide association study identifies a new melanoma susceptibility locus at 1q21.3. *Nat Genet* **43:** 1114–1118.
McCall MN, Illei PB, Halushka MK. 2016. Complex sources of variation in tissue expression data: analysis of the GTEx lung transcriptome. *Am J Hum Genet* **99:** 624–635.
Nan H, Kraft P, Qureshi AA, Guo Q, Chen C, Hankinson SE, Hu FB, Thomas G, Hoover RN, Chanock S, et al. 2009. Genome-wide association study of tanning phenotype in a population of European ancestry. *J Invest Dermatol* **129:** 2250–2257.
Nan H, Xu M, Zhang J, Zhang M, Kraft P, Qureshi AA, Chen C, Guo Q, Hu FB, Rimm EB, et al. 2011. Genome-wide association study identifies nidogen 1 (*NID1*) as a susceptibility locus to cutaneous nevi and melanoma risk. *Hum Mol Genet* **20:** 2673–2679.
Nica AC, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M, Travers M, Potter S, Grundberg E, Small K, et al. 2011. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet* **7:** e1002003.
Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. 2010. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet* **6:** e1000888.
Ongen H, Andersen CL, Bramsen JB, Oster B, Rasmussen MH, Ferreira PG, Sandoval J, Vidal E, Whiffin N, Planchon A, et al. 2014. Putative *cis*-regulatory drivers in colorectal cancer. *Nature* **512:** 87–90.
Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. 2016. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32:** 1479–1485.
Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK. 2010. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464:** 768–772.
Praetorius C, Grill C, Stacey SN, Metcalf AM, Gorkin DU, Robinson KC, Van Otterloo E, Kim RS, Bergsteinsdottir K, Ogmundsdottir MH, et al. 2013. A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. *Cell* **155:** 1022–1033.
Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38:** 904–909.
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81:** 559–575.
R Core Team. 2018. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Ramachandrappa S, Gorrigan RJ, Clark AJ, Chan LF. 2013. The melanocortin receptors and their accessory proteins. *Front Endocrinol (Lausanne)* **4:** 9.
Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518:** 317–330.
Scoggins CR, Ross MI, Reintgen DS, Noyes RD, Goydos JS, Beitsch PD, Urist MM, Ariyan S, Sussman JJ, Edwards MJ, et al. 2006. Gender-related differences in outcome for melanoma patients. *Ann Surg* **243:** 693–698; discussion 698–700.
Shabalin AA. 2012. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28:** 1353–1358.
Shi J, Marconett CN, Duan J, Hyland PL, Li P, Wang Z, Wheeler W, Zhou B, Campan M, Lee DS, et al. 2014. Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue. *Nat Commun* **5:** 3365.
Sitaram A, Marks MS. 2012. Mechanisms of protein delivery to melanosomes in pigment cells. *Physiology (Bethesda)* **27:** 85–99.
Smit A, Hubley R, Green P. 2013–2015. RepeatMasker Open-4.0. http://www.repeatmasker.org.
Stegle O, Parts L, Durbin R, Winn J. 2010. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput Biol* **6:** e1000770.
Stokowski RP, Pant PV, Dadd T, Fereday A, Hinds DA, Jarman C, Filsell W, Ginger RS, Green MR, van der Ouderaa FJ, et al. 2007. A genomewide association study of skin pigmentation in a South Asian population. *Am J Hum Genet* **81:** 1119–1132.
Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci* **100:** 9440–9445.
Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D, et al. 2007. Population genomics of human gene expression. *Nat Genet* **39:** 1217–1224.
Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, Manolescu A, Karason A, Palsson A, Thorleifsson G, et al. 2007. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet* **39:** 1443–1452.
Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Jakobsdottir M, Steinberg S, Gudjonsson SA, Palsson A, Thorleifsson G, et al. 2008. Two newly identified genetic determinants of pigmentation in Europeans. *Nat Genet* **40:** 835–837.
Taylor EM, Broughton BC, Botta E, Stefanini M, Sarasin A, Jaspers NG, Fawcett H, Harcourt SA, Arlett CF, Lehmann AR. 1997. Xeroderma pigmentosum and trichothiodystrophy are associated with different mutations in the *XPD* (*ERCC2*) repair/transcription gene. *Proc Natl Acad Sci* **94:** 8658–8663.
Visconti A, Duffy DL, Liu F, Zhu G, Wu W, Chen Y, Hysi PG, Zeng C, Sanna M, Iles MM, et al. 2018. Genome-wide association study in 176,678 Europeans reveals genetic loci for tanning response to sun exposure. *Nat Commun* **9:** 1684.
Wendt J, Mueller C, Rauscher S, Fae I, Fischer G, Okamoto I. 2018. Contributions by *MC1R* variants to melanoma risk in males and females. *JAMA Dermatol* **154:** 789–795.
Westra HJ, Franke L. 2014. From genome to function by studying eQTLs. *Biochim Biophys Acta* **1842:** 1896–1902.
Wolpin BM, Rizzato C, Kraft P, Kooperberg C, Petersen GM, Wang Z, Arslan AA, Beane-Freeman L, Bracci PM, Buring J, et al. 2014. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. *Nat Genet* **46:** 994–1000.
Yang F, Wang J, GTEx Consortium, Pierce BL, Chen LS. 2017. Identifying *cis*-mediators for *trans*-eQTLs across many human tissues using genomic mediation analysis. *Genome Res* **27:** 1859–1871.
Zhang M, Song F, Liang L, Nan H, Zhang J, Liu H, Wang LE, Wei Q, Lee JE, Amos CI, et al. 2013. Genome-wide association studies identify several new loci associated with pigmentation traits and skin cancer risk in European Americans. *Hum Mol Genet* **22:** 2948–2959.
Zhang M, Lykke-Andersen S, Zhu B, Xiao W, Hoskins JW, Zhang X, Rost LM, Collins I, Bunt MV, Jia J, et al. 2018. Characterising *cis*-regulatory variation in the transcriptome of histologically normal and tumour-derived pancreatic tissues. *Gut* **67:** 521–533.
Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, Montgomery GW, Goddard ME, Wray NR, Visscher PM, et al. 2016. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48:** 481–487.

Received December 7, 2017; accepted in revised form September 21, 2018.