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], Mark M. Iles[4], Stuart MacGregor[5], Laufey T. Amundadottir[1], Matthew H. Law[5], Stacie K. Loftus[6], William J. Pavan[6,8], 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, UK. [5] Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia; [6] Genetic Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

[7] These authors contributed equally to this work

[8] These authors co-supervised this work

[9,10] A complete list of consortium authors appears at the end of this paper

Corresponding authors: bpavan@mail.nih.gov, kevin.brown3@nih.gov

RUNNING TITLE: Cell-type specific melanocyte eQTL dataset

KEY WORDS: eQTL, Melanocyte, Melanoma, cell-type specific

ABSTRACT

Most expression quantitative trait loci (eQTL) studies to date have been performed in heterogenous 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*-eQTLs and 4,997 eGenes (FDR<0.05), which are higher numbers than any GTEx tissue type with a similar sample size. Melanocyte eQTLs differed considerably from those identified in the 44 GTEx tissues, including skin. Over a third of melanocyte eGenes including key genes in melanin synthesis pathways were not observed in two types of GTEx skin tissues or melanoma samples. The melanocyte dataset also identified cell-type specific *trans*-eQTL of a pigmentation-associated SNP for four genes, likely through its *cis*-regulation of *IRF4*, encoding a transcription factor implicated in human pigmentation phenotypes. Melanocyte eQTLs are enriched in *cis*-regulatory signatures found in melanocytes as well as melanoma-associate variants identified through genome-wide association studies (GWAS). Co-localization of melanoma GWAS variants and eQTLs from melanocyte and skin eQTL datasets identified candidate melanoma susceptibility genes for most of the known GWAS loci including unique genes identified by the melanocyte dataset. Further, a transcriptome-wide association study using published melanoma GWAS data uncovered four new 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 highlights 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.

INTRODUCTION

Expression quantitative trait loci (eQTL) 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 datasets derived from normal human tissues have been made publicly available. Perhaps most notable of those is from the Genotype-Tissue Expression (GTEx)(GTEx-Consortium 2015) project representing >44 tissue types of hundreds of post-mortem donors. These studies have collectively emphasized 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; 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) or induced Pluripotent Stem Cells (iPSC) (Kilpinen et al. 2017), eQTL datasets 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 (GTEx-Consortium 2015). These findings highlight the importance of using eQTL datasets 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 datasets (Gamazon et al. 2015; Gusev et al. 2016). Such approaches, usually referred to as transcriptome-wide association studies (TWAS) enable assignments of disease-associated genes 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; Brown et al. 2008; Gudbjartsson et al. 2008; Han et al. 2008; Bishop et al. 2009; Falchi et al. 2009; Nan et al. 2009; Duffy et al. 2010b; Eriksson et al. 2010; Amos et al. 2011; Barrett et al. 2011; Macgregor et al. 2011; Nan et al. 2011; Candille et al. 2012; Zhang et al. 2013; Jacobs et al. 2015; Law et al. 2015; Liu et al. 2015), 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/β-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 datasets are available, the largest ones (GTEx(GTEx-Consortium 2015), MuTHER (Nica et al. 2011)) are derived from heterogenous 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 (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 reflect 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 dataset to study the lineage-specific regulatory function of melanoma- and pigmentation-associated common variants. Our results demonstrate that eQTLs in tissue-specific homogenous cultures are distinct from those of heterogenous tissue samples by identifying melanocyte-specific *cis*-eQTL gene sets as well as a unique *trans*-regulatory network. Our dataset serves as a valuable resource for annotating melanoma risk variants from known GWAS loci as well as identifying new genome-wide significant loci where imputed gene expression levels were significantly associated with melanoma case control status. Our results highlight the value of building cell-type specific eQTL datasets and complementary roles of multiple eQTL datasets in GWAS annotation.

RESULTS

Melanocyte eQTLs are distinct from those in 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 1**). Melanocytes were cultured in lot-matched culture medium in randomized batches to minimize variability that could be introduced by culturing conditions. RNA sequencing and direct SNP genotyping of these samples produced an average

of ~87.9 million reads (paired-end, stranded, 126bps), and ~713,000 SNP genotypes, respectively. With this genotyping data, we further imputed 10,718,646 genotypes using the Haplotype Reference Consortium reference panel and the Michigan Imputation Server (Das et al. 2016). To apply the best practices for eQTL data processing and enable direct comparison to publicly available eQTL data for other tissue types, we followed the analysis methods adopted by the GTEx project (see Methods section), utilizing probabilistic estimation of expression residuals (PEER) to control for hidden batch effects and other confounding effects that could be reflected in the expression data.

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 +/-1Mb of transcription start sites (TSS). After quality control, 19,608 genes (GenCODE Release 19) were expressed above cutoff levels in primary melanocytes, and among them, 4,997 genes exhibited association with genotypes of at least one SNP at a genome-wide significant level (FDR < 0.05, "eGenes"; **Supplemental Table 2**). Of 10,718,646 genotyped or imputed SNPs passing QC, 597,335 unique SNP-gene pairs displayed genome-wide significant correlation (FDR < 0.05, significant eQTLs). The number of identified eGenes and significant eQTLs was approximately three times higher than those from data analyzed without using PEER factors as covariates (**Supplemental Table 2**). Application of PEER factors almost doubled the number of eGenes known to be related to pigmentation phenotypes (0.8% vs 1.5%; Fisher exact test *P*-value = 0.0335). 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 consistent with broad-based pleiotropic cell functions for genes expressed in melanocytes (**Supplemental Table 3**). Further, melanocyte eGenes included 11 out of 24 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 75 out of 379 curated pigmentation genes (based on human and mouse phenotype, OMIM, MGI) including *IRF4*, *TRPM1*, and *MC1R* (**Supplemental Table 4-5**).

When significant melanocyte eQTLs were tested for overlap with 44 GTEx tissue types, none of them displayed a shared eQTL proportion (π_1) higher than 0.75 with melanocytes, including two types of skin samples (π_1 = 0.67 with Skin_Sun_Exposed, and 0.58 with Skin_Not_Sun_Exposed; **Fig. 1**). The largest overlap with melanocyte eQTLs was seen for transformed fibroblasts (π_1 = 0.74). This contrasts with the considerably higher levels of sharing between the two types of skin samples (π_1 = 0.91) or among brain tissues (average π_1=0.87) in GTEx. Notably, while the sample size of our melanocyte collection ranks at the 60th percentile among GTEx tissue types, the number of significant eGenes was much higher than expected (ranked within top 30%) considering a general correlation between sample size and number of eGenes (GTEx-Consortium 2015) (**Supplemental Table 6**).

We further focused the comparison of our melanocyte dataset 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). Collectively, these four eQTL datasets identified 12,136 eGenes, with 382 eGenes shared among all four datasets. Sixty-four percent of melanocyte eGenes were shared with at least one of the three datasets. While 33% and 39% of eGenes from sun-exposed skin and not sun-exposed skin, respectively, overlapped with melanocyte eGenes, 48% of cutaneous melanoma eGenes were shared with melanocyte eGenes. Notably, 1,801 eGenes (36% of melanocyte eGenes) were entirely private to melanocytes, and a total of 6,187 eGenes (51% of eGenes from all four datasets) were specific to only one of four datasets (**Supplemental Fig. S1; Supplemental Table 2**). eGenes from these four datasets collectively accounted for 150 of 379 (40%) curated pigmentation genes,

Figure 1

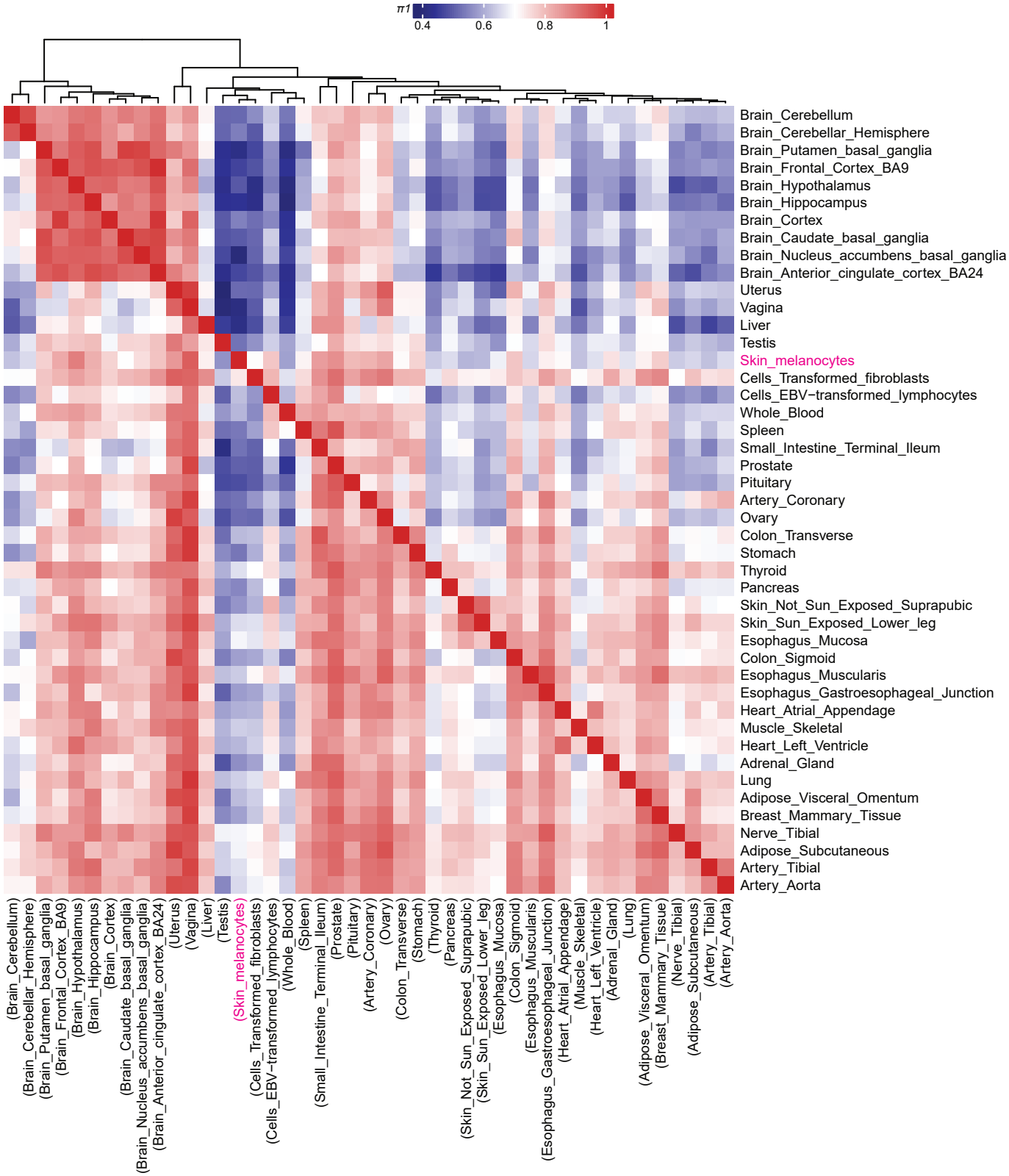
π1
0.4
0.6
0.8
1
Brain_Cerebellum
Brain_Cerebellar_Hemisphere
Brain_Putamen_basal_ganglia
Brain_Frontal_Cortex_BA9
Brain_Hypothalamus
Brain_Hippocampus
Brain_Cortex
Brain_Caudate_basal_ganglia
Brain_Nucleus_accumbens_basal_ganglia
Brain_Anterior_cingulate_cortex_BA24
Uterus
Vagina
Liver
Testis
Skin_melanocytes
Cells_Transformed_fibroblasts
Cells_EBV−transformed_lymphocytes
Whole_Blood
Spleen
Small_Intestine_Terminal_Ileum
Prostate
Pituitary
Artery_Coronary
Ovary
Colon_Transverse
Stomach
Thyroid
Pancreas
Skin_Not_Sun_Exposed_Suprapubic
Skin_Sun_Exposed_Lower_leg
Esophagus_Mucosa
Colon_Sigmoid
Esophagus_Muscularis
Esophagus_Gastroesophageal_Junction
Heart_Atrial_Appendage
Muscle_Skeletal
Heart_Left_Ventricle
Adrenal_Gland
Lung
Adipose_Visceral_Omentum
Breast_Mammary_Tissue
Nerve_Tibial
Adipose_Subcutaneous
Artery_Tibial
Artery_Aorta

Figure 1. Melanocyte eQTLs display a distinct pattern from those of 44 GTEx tissue types. Dendrogram and heatmap 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 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 parenthesis on the bottom. The position of our skin melanocyte eQTL dataset is shown in pink.

with the majority specific to one dataset (**Supplemental Fig. S2**). Melanocyte eGenes were especially enriched with GO terms containing “melanin” compared to those of normal skin tissues and skin cutaneous melanomas (0.22% of melanocyte eGenes vs. 0.04 – 0.08% of three other eQTL datasets; **Supplemental Table 4**).

Melanocyte eQTLs are enriched with allelic imbalance variants, melanocyte *cis*-regulatory signatures, and melanoma-associated variants

We next sought to determine whether melanocyte eQTLs were corroborated by allelic imbalance variants in heterozygous individuals from the same dataset. To determine genome-wide allele-specific expression (ASE), we performed binomial tests at the single sample level, as well as in the combined set of heterozygous samples for each locus, identifying 48,038 unique allelic imbalance variants at the individual level and 8,635 sample-set wide (FDR < 0.05 or effect size > 0.15; **Supplemental Table 7**). These two analyses identified a total of 48,207 unique ASE variants in 9,710 ASE genes (≥ 1 ASE variant), 30.7% of which are significant melanocyte eGenes, demonstrating an enrichment of ASE in melanocyte eGenes (Fisher’s exact test $P = 6.32 \times 10^{-67}$, Odds Ratio=1.77; **Supplemental Table 7**). When the allelic effects of 48,038 ASE variants (single sample level) were averaged from all the heterozygous individuals and compared between coding regions of melanocyte eGenes and non-eGenes, the allelic effects were significantly larger in the eGene group (Wilcoxon signed rank test $P = 1.67 \times 10^{-34}$; **Supplemental Fig. S3A**). Similarly, when the proportions of heterozygous individuals displaying allelic imbalance at each locus was compared between coding regions of melanocyte eGenes and non-eGenes, the proportions displaying allelic imbalance were likewise significantly higher in the eGene group (Wilcoxon signed rank test $P = 1.27 \times 10^{-81}$; **Supplemental Fig. S3B**). These data indicate that melanocyte eGenes are also supported by allelic imbalance at the individual level as well as population levels.

We then further examined if melanocyte eQTLs were enriched within epigenetic signatures marking melanocyte *cis*-regulatory elements identified by the ENCODE and Epigenome Roadmap projects (www.encodeproject.org; www.roadmapepigenomics.org). We specifically examined regions of open chromatin marked by DNaseI-hypersensitivity sites (DHS), as well as promoter and enhancer histone marks (H3K27ac, H3K4Me1, and H3K4me3) generated from primary cultured human melanocytes from three individuals by the Roadmap project. 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 +/- 1Mb of TSS 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) (**Table 1**). Notably, enrichment of melanocyte eQTL SNPs within these melanocyte epigenetic signatures (with the exception of H3K4me3) was higher than that observed for eQTL SNPs from GTEx skin tissues or melanoma tumors when compared to all the tested SNPs (**Supplemental Fig. 4B**) or to matched controls by 10,000 permutations (**Table 1**). Further, melanocyte eQTL SNPs were also found to be enriched (permutation test $P < 1 \times 10^{-4}$) upstream of genes (within 1-5 kb of the TSS), as well as in gene promoters, 5' UTRs, exonic regions, first exons, first introns, introns, intron-exon boundaries (+/- 200bp encompassing exon splice regions), and 3' UTRs, but not in intergenic regions or annotated lincRNA regions (**Supplemental Fig. S4**). Consistent with this enrichment, most of the melanocyte eQTL SNPs were centered within +/- 250kb of transcription start sites (TSS) (**Supplemental Fig. S5**).

Next, we sought to determine if melanocyte eQTLs were enriched with common risk variants from melanoma GWAS. When melanoma GWAS *P*-values from the most recent meta-analysis (Law et al. 2015) were compared between significant melanocyte eQTL SNPs and non-eQTL SNPs, significant enrichment was observed for eQTL SNPs (**Fig. 2**). Notably, the

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

Epigenetic Mark		DHS	H3K27ac	H3K4Me1	H3K4Me3
Melanocyte eQTLs	Fold enrichment	1.81	5.48	1.99	3.34
	P-value	<0.0001	<0.0001	<0.0001	<0.0001
TCGA SKCM eQTLs	Fold enrichment	1.72	2.69	1.86	3.44
	P-value	<0.0001	<0.0001	<0.0001	<0.0001
Skin Not Sun Exposed eQTLs	Fold enrichment	1.78	2.77	1.92	3.5
	P-value	<0.0001	<0.0001	<0.0001	<0.0001
Skin Sun Exposed eQTLs	Fold enrichment	1.75	2.66	1.88	3.29
	P-value	<0.0001	<0.0001	<0.0001	<0.0001

Epigenetic Mark: DNaseI hypersensitivity sites (DHS) and gene regulatory histone marks (H3K27ac, H3K4me1, and H3K4me3) of primary melanocytes (Epigenome Roadmap database)

Fold enrichment: 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.

P-value: see **Methods** section

Figure 2

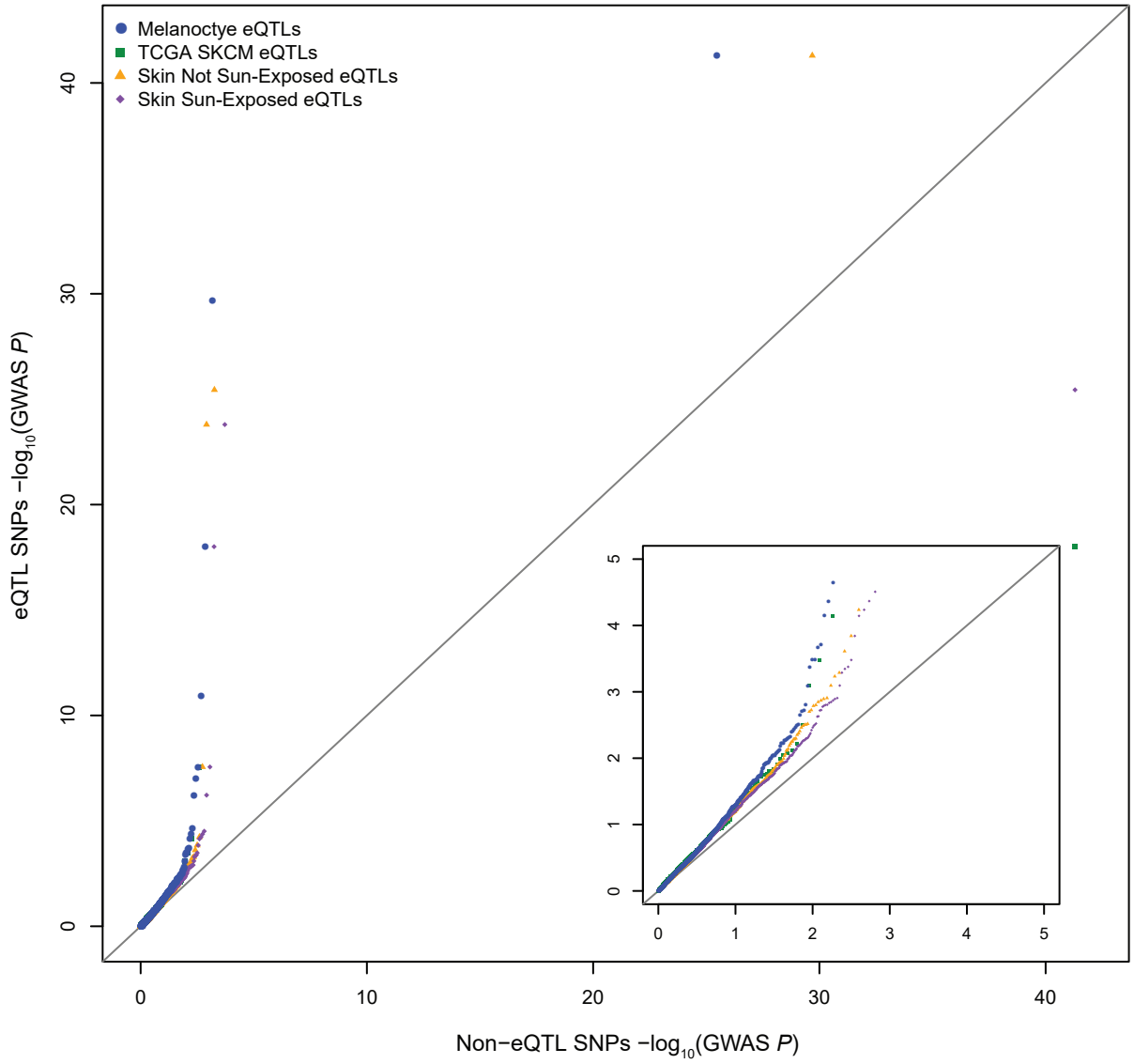
Melanoctye eQTLs
TCGA SKCM eQTLs
Skin Not Sun-Exposed eQTLs
Skin Sun-Exposed eQTLs
eQTL SNPs −log10(GWAS P)
Non−eQTL SNPs −log10(GWAS P)
0
10
20
30
40
0
1
2
3
4
5

Figure 2. Melanocyte eQTLs are enriched for common melanoma risk variants. QQ plot presents melanoma GWAS LD-pruned *P*-values of significant eQTL SNPs versus non-eQTL SNPs for melanocyte dataset compared to three other datasets. SNPs were classified as eQTL SNPs if they were significant eQTLs or is in strong LD (r^2>0.8) with an eQTL SNP (eQTL SNPs threshold: FDR < 0.05) in each dataset. The inset displays a zoomed-in view of a lower –log10 GWAS *P*-value range (0-5 range for X- and Y- axes).

enrichment of melanoma-associated SNPs was more pronounced in melanocyte eQTL SNPs compared to three related tissue types (**Fig. 2**) as well as all the other GTEx tissue types (**Supplemental Fig. S6**), where estimated Lambda value was highest for melanocyte dataset (Lambda=1.51).

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

While the modest size of this dataset significantly limits power, we also performed *trans*-eQTL analyses for the SNPs that are located over 1Mb away from the TSS of each gene or on a different chromosome. In all, we identified 78 genome-wide significant *trans*-eQTL genes, 16 of which had mappability higher than 0.8 (**Supplemental Table 8**). Of these 16, seven *trans*-eQTL SNPs were also *cis*-eQTLs for local genes within 1Mb. Notably, rs12203592 (chr6:396321), among these, is a genome-wide significant *trans*-eQTL SNP for four different genes on four separate chromosomes (*TMEM140*, *MIR3681HG, PLA1A*, and *NEO1*), and also is the strongest *cis*-eQTL SNP for the *IRF4* gene (P = 7.9 x 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 found to be associated with human pigmentation phenotypes (Han et al. 2008). This variant has also been shown to be a functional SNP mediating transcription of *IRF4* in melanocytes via allele-preferential binding of the transcription factor, TFAP2, in turn activating the melanin synthesis enzyme, *TYR* (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*-eQTL (Pearson r = 0.54, 0.65, 0.53, and 0.58; P = 2.67 x 10^{-9}, 5.34 x 10^{-14}, 4.28 x 10^{-9}, and 6.00 x 10^{-11}, respectively; **Supplemental Fig. S7**). Gene ontology analysis of a larger set of 84 genes correlated with rs12203592 *in trans* at a less stringent cutoff (P < 1 x 10^{-5} and mappability > 0.8)

Figure 3

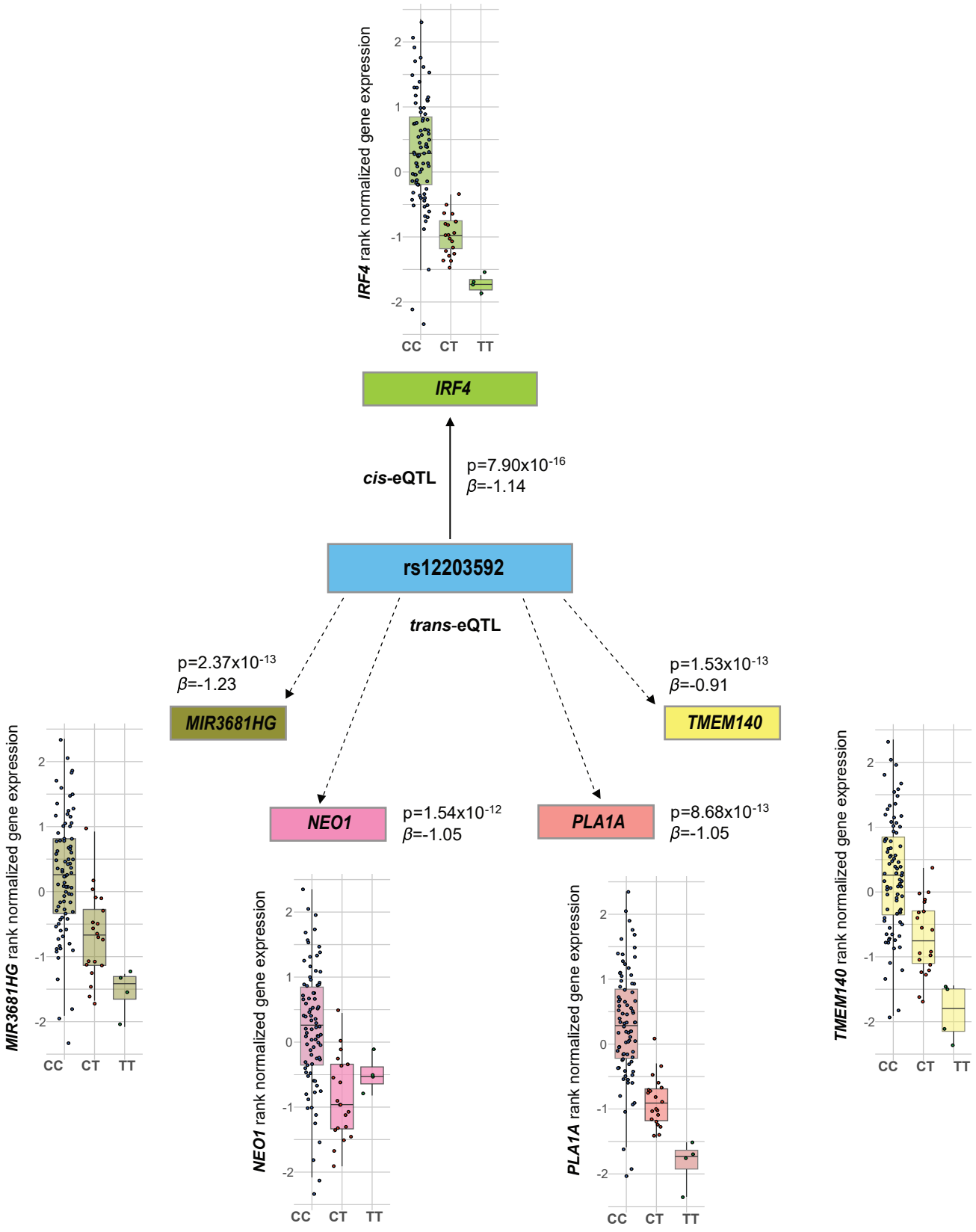
IRF4 rank normalized gene expression
2
1
0
-1
-2
CC
CT
TT
IRF4
cis-eQTL
p=7.90x10-16
β=-1.14
rs12203592
trans-eQTL
p=2.37x10-13
β=-1.23
MIR3681HG
p=1.53x10-13
β=-0.91
TMEM140
NEO1
p=1.54x10-12
β=-1.05
PLA1A
p=8.68x10-13
β=-1.05
MIR3681HG rank normalized gene expression
NEO1 rank normalized gene expression
PLA1A rank normalized gene expression
TMEM140 rank normalized gene expression

Figure 3. Pigmentation trait-associated variant, rs12203592, in *IRF4* is a *trans*-eQTL for six genes in melanocytes. *cis*- or *trans*- eQTL *P*-value and effect size (β) are shown between rs12203592 and *IRF4* or rs12203592 and four genome-wide significant *trans*-eQTL genes (*TMEM140*, *MIR3681HG*, *PLA1A*, and *NEO1*). β value is shown relative to alternative allele, T. Boxplot displays gene expression levels based on rs12203592 genotypes (CC, CT, and TT).

suggested an enrichment in “cell morphogenesis involved in neuron differentiation” (GO: 0061564, $P = 1.46 \times 10^{-5}$) and “cell morphogenesis involved in differentiation” (GO: 0000904, $P = 1.73 \times 10^{-4}$; **Supplemental Table 9**) among others. To assess if *IRF4* expression levels mediate the observed *trans*-eQTL effect for these four genes, we performed mediation analyses in three different ways. Firstly, the residuals of the four *trans*-eQTL gene expression levels were regressed against rs12203592 after accounting for *IRF4* levels. The results indicated considerable increases in association *P*-values of the residuals for all four genes, suggesting that *IRF4* expression accounts for the most of association between the SNP and four *trans*-eQTL genes (**Supplemental Table 10-A**). Secondly, the expression levels of four *trans*-eQTL genes were regressed against rs12203592 with and without *IRF4* level as a covariate. This time, their association effect sizes and *P* - values were not significantly affected by *IRF4* expression (**Supplemental Table 10-B,C**). Finally, we applied a recently published Genomic Mediation analysis with Adaptive Confounding adjustment (GMAC) (Yang et al. 2017) to 1507 eSNP - *cis*-eGene - *trans*-gene trios (*trans*-eQTL cutoff: $P < 1 \times 10^{-5}$), 130 of which include rs12203592. A total of 422 trios displayed a suggestive mediation ($P < 0.05$), and 49 of them were by *IRF4* cis-eQTL including those with *TMEM140* and *NEO1*. (**Supplemental Table 11**). 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 12**), as well as the larger set of 84 *trans*-eQTL genes correlated with rs12203592 ($P < 1 \times 10^{-5}$ and mappability > 0.8; $P = 7.25 \times 10^{-26}$). 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

In order to identify candidate susceptibility genes for melanoma risk loci, we overlaid cutaneous melanoma GWAS SNPs (meta-analysis $P_{fixed} < 5 \times 10^{-8}$) (Law et al. 2015) with significant eQTL SNPs. From the 20 known genome-wide significant melanoma risk loci, 25 genes from 11 of these loci displayed overlapping SNPs (**Table 2**), which is the largest overlap of all 44 GTEx tissue types (**Supplemental Table 6**). Only eight of these twenty-five eQTL genes were the closest gene to a lead GWAS SNP (*PARP1*, *ARNT*, *SLC45A2*, *MTAP*, *TYR*, *OCA2*, *FANCA*, and *MX2*). Three melanoma risk loci are significant eQTLs for known pigmentation genes (*SLC45A2* on chr5p13.2, *TYR* on chr11q14-121, and both *OCA2* and *HERC2* on chr15q13.1; **Supplemental Fig. S8**); eQTLs overlapping significant melanoma GWAS SNPs for *SLC45A2*, *TYR*, and *OCA2* were not observed in any of the 44 GTEx tissue types or TCGA melanoma data. Four loci, among others, strongly overlapped with eQTLs for single genes – *CASP8, MX2, MAFF,* and as previously reported, *PARP1* (Choi et al. 2017) (74-100% of significant GWAS SNPs overlapped eQTLs; **Table 2**; **Fig. 4**).

When a similar approach was applied to the GTEx skin and TCGA melanoma datasets, a total of 33 genes from six melanoma GWAS loci showed significant overlap in any of the three datasets (**Supplemental Tables 13-15**). By including the melanocyte eQTL set, the four melanoma-relevant datasets identified a total of 49 eQTL genes overlapping 11 melanoma GWAS loci. Thirty-two of these eQTL genes (65% of the 49 total genes) were unique to a single dataset, with half of them specific to the melanocyte dataset (**Supplemental Fig. S9; Supplemental Table 16**). While significant eQTLs for pigmentation genes at melanoma risk loci were typically melanocyte-specific, an eQTL for *ASIP* is observed specifically in sun-exposed skin tissue. Among genes displaying consistent overlap across multiple tissue types, *GOLPH3L* was unique in that its levels displayed an opposite direction of effect between normal skin tissues and cutaneous melanomas (**Supplemental Tables 13-15**). We also performed permutation analyses to test for statistically significant enrichment of eQTLs from the four tissue

Table 2. Melanoma GWAS variants that are melanocyte eQTLs

Melanoma GWAS locus	GWAS lead SNP	eQTL gene	eQTL SNP	Chromosome: location	Ref	Alt	eQTL *P*-value	eQTL slope	GWAS *Pfixed* of eQTL SNP	Risk allele	# of GWAS SNPs	# of eQTL SNPs	# of Overlapping SNPs	% Overlapping (of GWAS SNPs)
1q21.3	rs12410869	RP11-316M1.12	rs12410869	1:150856153	G	T	8.06E-08	-0.70	5.21E-13	G	368	380	313	85
1q21.3	rs12410869	ARNT	rs7412746	1:150860471	C	T	5.45E-07	-0.34	2.25E-12	T	368	94	45	12
1q21.3	rs12410869	CTSS	rs6669855	1:150861229	C	T	9.62E-05	0.52	1.66E-12	C	368	292	220	60
1q21.3	rs12410869	CERS2	rs6669855	1:150861229	C	T	1.55E-04	-0.29	1.66E-12	C	368	28	25	7
1q21.3	rs12410869	CTSK	rs12410050	1:150857141	A	G	2.32E-04	0.53	8.23E-12	G	368	11	7	2
1q42.12	rs1858550	PARP1	rs1858550	1:226608104	C	A	6.11E-05	-0.23	1.68E-13	C	68	66	50	**74**
2q33-q34	rs7582362	CASP8	rs7582362	2:202176294	A	G	4.43E-09	0.65	8.88E-09	A	11	132	11	**100**
5p13.2	rs250417	SLC45A2	rs250417	5:33952378	G	C	2.62E-06	0.80	2.30E-12	C	6	21	6	100
9p21	rs7852450	MTAP	rs935055	9:21803183	G	C	1.29E-07	-0.28	2.65E-32	G	394	269	253	64
9p21	rs7852450	RP11-149I2.5	rs935055	9:21803183	G	C	8.94E-05	-0.40	2.65E-32	G	394	77	72	18
11q14-q21	rs1393350	CBX3P7	rs3907665	11:89043946	A	G	1.98E-05	-0.34	2.85E-18	G	558	99	36	6
11q14-q21	rs1393350	TYR	rs1126809	11:89017961	G	A	1.30E-04	0.41	1.67E-26	A	558	172	137	25
15q13.1	rs4778138	OCA2	rs4778138	15:28335820	A	G	8.72E-08	0.62	3.11E-09	A	1	171	1	100
15q13.1	rs4778138	HERC2	rs4778138	15:28335820	A	G	1.17E-04	0.46	3.11E-09	A	1	1	1	100
16q24.3	rs75570604	SPIRE2	rs8049897	16:90024202	G	A	1.62E-06	0.58	1.27E-81	A	1097	655	81	7
16q24.3	rs75570604	CDK10	rs76888998	16:89884310	G	T	3.49E-06	-0.59	1.16E-20	T	1097	65	34	3
16q24.3	rs75570604	DEF8	rs8051733	16:90024206	A	G	1.62E-05	-0.26	6.47E-75	G	1097	55	17	2
16q24.3	rs75570604	TCF25	rs9933627	16:90107353	G	A	5.90E-05	0.20	1.98E-28	A	1097	30	13	1
16q24.3	rs75570604	TUBB3	rs11076628	16:89875246	T	C	1.43E-04	-0.54	1.34E-18	C	1097	136	18	2
16q24.3	rs75570604	FANCA	rs12924011	16:89829539	C	T	1.78E-04	-0.42	1.24E-45	T	1097	219	185	17
16q24.3	rs75570604	TUBB8P7	rs73253776	16:89385582	T	C	2.48E-04	-0.73	1.05E-38	C	1097	7	2	0
20q11.2-q12	rs6059655	MAP1LC3A	rs2144956	20:33145868	G	A	1.68E-04	0.21	2.02E-12	G	285	258	4	1
20q11.2-q12	rs6059655	NCOA6	rs6088579	20:33284624	A	G	2.44E-04	-0.25	1.61E-12	A	285	55	42	15
21q22.3	rs408825	MX2	rs408825	21:42743496	C	T	3.47E-15	0.74	3.21E-15	T	11	21	10	**91**
22q13.1	rs2092180	MAFF	rs132941	22:38545942	T	C	3.20E-20	-0.83	1.61E-12	T	120	244	119	**99**

melanoma GWAS locus: loci that are mainly explained by pigmentation traits are in bold

eQTL SNP: representative eQTL SNP showing the most significant association with melanoma (based on meta-analysis P_{fixed} value). "Chromosome:location", "Ref" and "Alt" alleles, "eQTL *P*-value", "GWAS Pfixed of eQTL SNP", and "Risk allele" are all for this SNP.

eQTL Slope: the effect of the "Alt" allele relative to the "Ref" allele.

# of GWAS SNPs: total number of significant melanoma GWAS SNPs in the locus (Pfixed <5E-8)

# of eQTL SNPs: total number of significant eQTL SNPs for the gene

# of Overlapping SNPs: number of SNPs that are both significant melanoma GWAS and significant eQTL SNPs.

# of Overlapping (of GWAS SNPs): percentage of # of Overlapping SNPs over # of GWAS SNPs. eQTL genes with >20% overlap are in bold.

Figure4

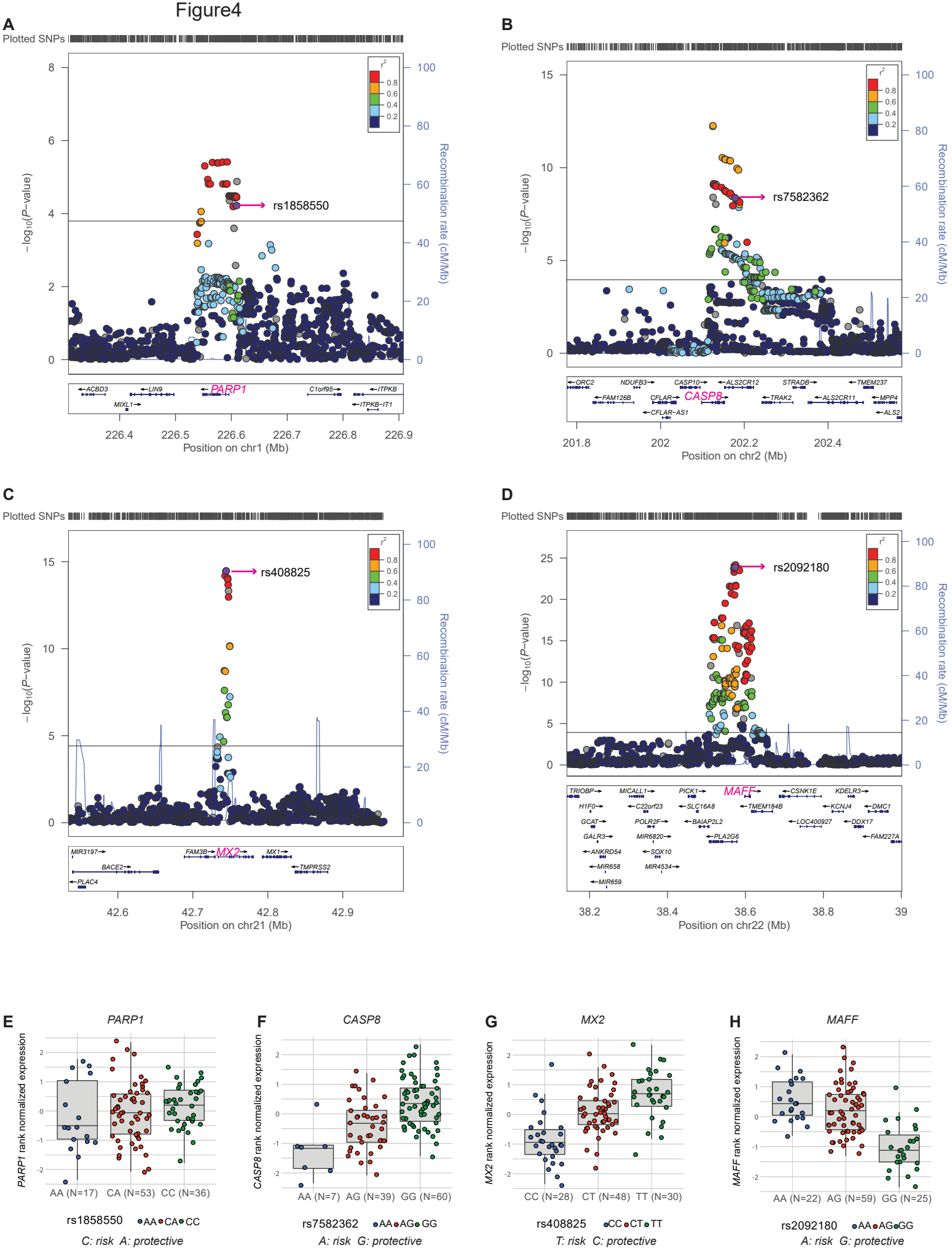
A
Plotted SNPs
rs1858550
PARP1
ACBD3
LIN9
MIXL1
C1orf95
ITPKB
ITPKB-IT1
Position on chr1 (Mb)
B
rs7582362
CASP8
ORC2
NDUFB3
CASP10
ALS2CR12
STRADB
TMEM237
FAM126B
CFLAR
TRAK2
ALS2CR11
MPP4
CFLAR-AS1
ALS2
Position on chr2 (Mb)
C
rs408825
MX2
MIR3197
FAM3B
MX1
BACE2
TMPRSS2
PLAC4
Position on chr21 (Mb)
D
rs2092180
MAFF
TRIOBP
MICALL1
PICK1
CSNK1E
KDELR3
H1F0
C22orf23
SLC16A8
TMEM184B
KCNJ4
DMC1
GCAT
POLR2F
BAIAP2L2
LOC400927
DDX17
GALR3
MIR6820
PLA2G6
FAM227A
ANKRD54
SOX10
MIR658
MIR4534
MIR659
Position on chr22 (Mb)
−log10(P-value)
Recombination rate (cM/Mb)
E
PARP1
PARP1 rank normalized expression
AA (N=17)
CA (N=53)
CC (N=36)
rs1858550
AA CA CC
C: risk A: protective
F
CASP8
CASP8 rank normalized expression
AA (N=7)
AG (N=39)
GG (N=60)
rs7582362
AA AG GG
A: risk G: protective
G
MX2
MX2 rank normalized expression
CC (N=28)
CT (N=48)
TT (N=30)
rs408825
CC CT TT
T: risk C: protective
H
MAFF
MAFF rank normalized expression
AA (N=22)
AG (N=59)
GG (N=25)
rs2092180
AA AG GG
A: risk G: protective

Figure 4. Melanoma GWAS variants that are local eQTLs for a single gene. (*A*-*D*) LocusZoom plots present the nominal eQTL *P*-values of all tested local SNPs in 300-400 kb windows for four significant eQTL genes from four melanoma GWAS loci: (*A*) *PARP1*; (*B*) *CASP8*; (*C*) *MX2*; and (*D*) *MAFF*. 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 grey. Horizontal lines are shown for nominal *P*-value cutoff for significant eQTLs. Genomic coordinates are based on hg19. (*E*-*H*) Box plots 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.

types in melanoma GWAS using three less-stringent GWAS *P*-value thresholds (5×10^{-5}, 5×10^{-6}, and 5×10^{-7}; **Supplemental Table 17**). The results indicated that melanoma-associated SNPs using all four thresholds are significantly enriched (at least 2-fold) in eQTLs. Notably, the number of GWAS loci displaying true overlap was much higher (8-12 loci) for melanocyte eQTLs than for skin tissue or melanoma tumors (2-7 loci). In addition, we applied eCAVIAR (Hormozdiari et al. 2016) to assess colocalization of melanoma GWAS and eQTL signals from melanocyte and two GTEx skin datasets. At a colocalization posterior probability (CLPP) cutoff of 1%, 15 of 20 known melanoma loci displayed colocalization of GWAS and eQTL signals for a total of 86 genes and 55 SNPs across three datasets (**Supplemental Table 18**). While the sun-exposed skin dataset contributed the highest number of genes (49 genes), melanocyte dataset identified 29 genes and added unique colocalizations for pigmentation genes *TYR* and *SLC45A2*, which were absent in other datasets.

We further examined sharing between melanocyte eQTL SNPs and variants associated with melanoma-associated traits, specifically number of melanocytic nevi, as well as pigmentation phenotypes including hair color, eye color, freckling, tanning, and skin darkness. Five lead SNPs from two published GWAS of nevus count (Falchi et al. ; Nan et al.) as well as 84 lead SNPs from 10 GWA studies of pigmentation traits (Stokowski et al. ; Gudbjartsson et al. ; Han et al. ; Nan et al. ; Duffy et al. ; Eriksson et al. ; Candille et al. ; Zhang et al. ; Jacobs et al. ; Liu et al.) together with their LD proxies ($r^2 > 0.8$, EUR) were used for the analysis with melanocyte, skin, and TCGA melanoma eQTLs (**Supplemental Tables 19-21**). Six genes (*MTAP*, *RP11-149I2.5*, *BAIAP2L2*, *MAFF*, *APOBEC3C*, and *PLA2G6*) displayed significant eQTLs for SNPs associated with number of cutaneous nevi, four of which were identified from melanocyte eQTL (**Supplemental Fig. S10**). For pigmentation traits, the GTEx sun-exposed skin tissue dataset identified the most genes (19 genes), followed by both melanocyte and not sun-exposed skin tissue (13 genes; **Supplemental Table 22**). Out of 32 total genes overlapping

pigmentation GWAS variants, 60% were only found in one of four datasets (**Supplemental Fig. S11**). 25% of those 32 genes are known pigmentation genes based on human and mouse phenotype (**Supplemental Table 5**).

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

eQTL data can be utilized for transcriptome-wide association study (TWAS) to impute gene expression levels into GWAS datasets, so that trait-associated differences in gene expression can be inferred from genotype data or summary statistics of GWAS data. Given the enrichment of melanoma risk loci from GWAS for melanocyte eQTLs relative to other trait-relevant tissues, we performed a TWAS using summary statistics (Gusev et al.) from a previously-published meta-analysis of melanoma GWAS data (Law et al. 2015) and the melanocyte eQTL dataset as the reference dataset (see **Methods** section). We used a conservative cutoff for heritability estimates ($P < 0.01$) to minimize false-positive findings; 3,187 out of 4,997 total eGenes including 16 of 25 eQTL genes near known risk loci, passed the threshold and were tested for association (**Supplemental Table 23-24**). Among these, 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 fourth locus being suggestive (*PARP1* on chr1q42.1) (**Table 3**). TWAS further identified novel associations with melanoma at four genomic loci at a genome-wide significant level (*ZFP90* at chr16q22.1, *HEBP1* at chr12p13.1, *MSC* and *RP11-383H13.1* at chr8q13.3, and *CBWD1* at chr9p24.3) (**Table 3** and **Fig. 5**).

We additionally performed TWAS using each of the 44 GTEx tissue types as reference eQTL datasets. 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 dataset (**Supplemental Table 25**). Tibial nerve tissue identified the largest number of genes (11 genes) while

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

Gene Symbol	CHR	BEST.GWAS.ID	BEST.GWAS .Z	EQTL.ID	EQTL.Z	EQTL.GWAS.Z	# of SNP	# of WGT	MODEL	TWAS.Z	TWAS.P	FDR_PASS	gwasID
MAFF	22	rs132985	-6.91	rs738322	-5.14	-6.69	340	2	enet	6.67	**2.60E-11**	Y	22q13.1
CTSS	1	rs12410869	-7.22	rs7521898	3.33	-6.42	315	6	enet	-6.32	**2.65E-10**	Y	1q21.3
ZFP90	16	rs7184977	5.07	rs11075688	1.06	3.96	319	319	blup	5.20	**1.95E-07**	Y	**New**
HEBP1	12	rs2111398	5.08	rs1684387	-5.32	5.08	605	5	lasso	-5.05	**4.48E-07**	Y	**New**
CASP8	2	rs10931936	5.50	rs3769823	-3.71	4.45	372	3	lasso	-4.61	**4.06E-06**	Y	2q33-q34
MSC	8	rs1481853	-4.65	rs6983160	-3.83	-4.60	553	2	lasso	4.60	**4.27E-06**	Y	**New**
CBWD1	9	rs661356	4.79	rs2992854	4.99	-1.96	462	1	lasso	-4.54	**5.52E-06**	Y	**New**
RP11-383H13.1	8	rs1481853	-4.65	rs6983160	-2.75	-4.60	693	4	lasso	4.50	**6.68E-06**	Y	**New**
GPRC5A	12	rs2111398	5.08	rs1684387	-4.02	5.08	584	11	enet	-4.19	2.80E-05	N	New
RBBP5	1	rs11240466	4.09	rs10900456	1.41	3.98	623	623	blup	3.59	3.37E-04	N	New
ATP6V1G2-DDX39E	6	rs2239704	4.55	rs2523504	3.51	3.54	278	1	lasso	3.54	3.97E-04	N	New
CDH1	16	rs7184977	5.07	rs4076177	2.71	4.52	345	345	blup	3.47	5.27E-04	N	New
CHCHD6	3	rs9851451	3.46	rs9822602	-2.73	3.44	635	3	lasso	-3.44	5.78E-04	N	New
CTD-2003C8.2	11	rs1554519	4.51	rs7932891	5.66	3.32	746	4	lasso	3.43	6.08E-04	N	New
UQCC1	20	rs2425025	6.95	rs6060369	-3.14	-3.00	353	353	blup	3.42	6.21E-04	N	New
PLXNA1	3	rs9851451	3.46	rs4679317	-2.20	3.42	484	2	lasso	-3.42	6.33E-04	N	New
PARP1	1	rs1865222	-7.12	rs3219090	-1.30	-6.86	404	404	blup	3.41	6.57E-04	N	1q42.12

CHR: chromosome

BEST.GWAS.ID:rsID of the most significant melanoma GWAS SNP in the locus

BEST.GWAS.Z:GWAS Z-score of the most significant GWAS SNP in the locus

EQTL.ID:rsID of the best eQTL SNP in the locus

EQTL.Z:eQTL Z-score of the best eQTL in the locus

EQTL.GWAS.Z:GWAS Z-score for this eQTL SNP

# of SNP: Number of SNPs in the locus

# of WGT: Weighted number of SNPs in the locus

MODEL: Best performing model

TWAS.Z: TWAS Z-score

TWAS.P: TWAS P-value **(genome-wide significant P-values are in bold)**

gwasID: previously identified GWAS locus (yellow) or newly identified by this study

Figure 5

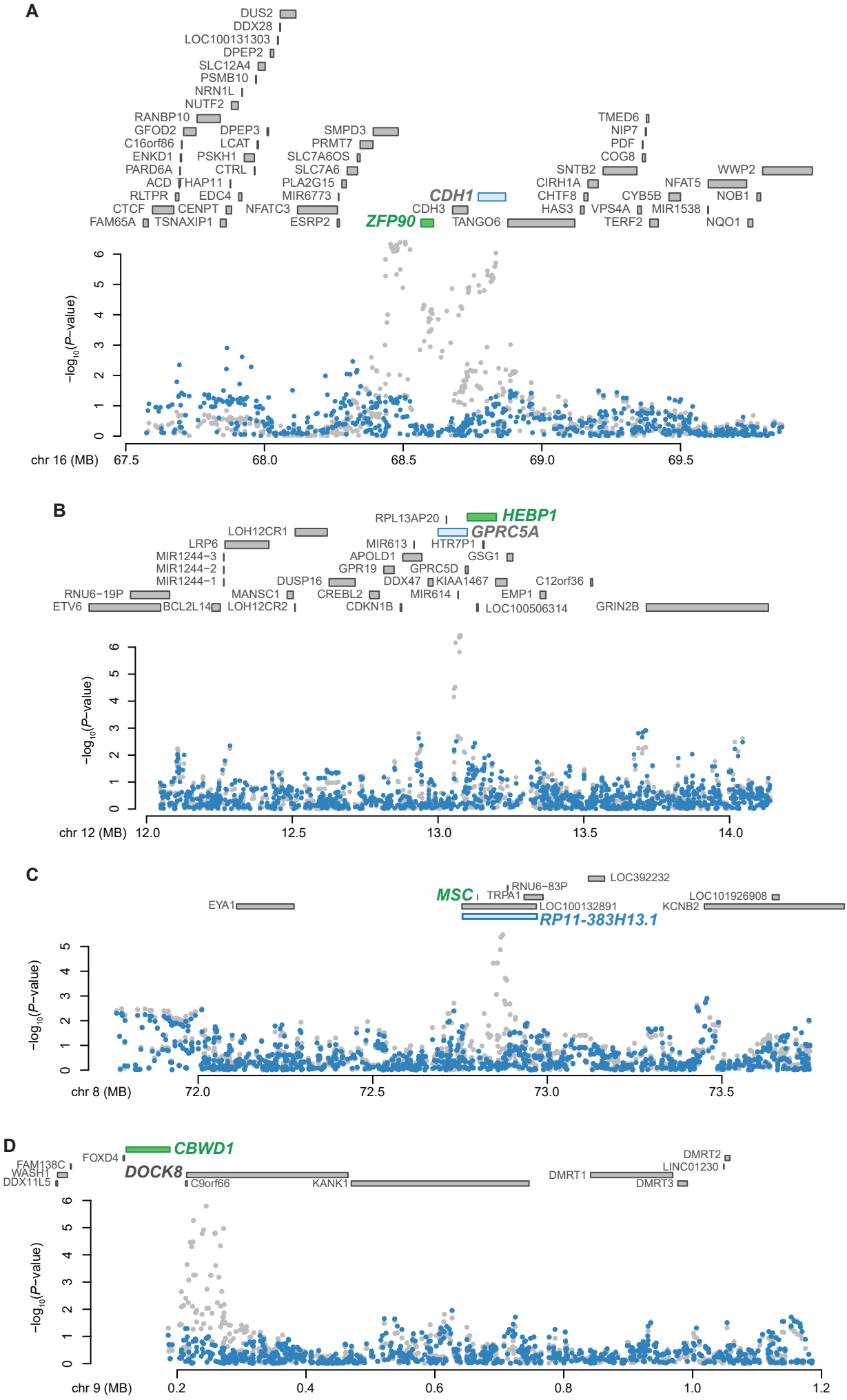
A
DUS2
DDX28
LOC100131303
DPEP2
SLC12A4
PSMB10
NRN1L
NUTF2
RANBP10
GFOD2
DPEP3
C16orf86
LCAT
ENKD1
PSKH1
PARD6A
CTRL
ACD
THAP11
RLTPR
EDC4
CTCF
CENPT
FAM65A
TSNAXIP1
SMPD3
PRMT7
SLC7A6OS
SLC7A6
PLA2G15
MIR6773
NFATC3
ESRP2
CDH1
CDH3
ZFP90
TANGO6
TMED6
NIP7
PDF
COG8
SNTB2
WWP2
CIRH1A
NFAT5
CHTF8
CYB5B
NOB1
HAS3
VPS4A
MIR1538
TERF2
NQO1
−log10(P−value)
chr 16 (MB)
67.5
68.0
68.5
69.0
69.5
B
RPL13AP20
HEBP1
LOH12CR1
GPRC5A
LRP6
MIR613
HTR7P1
MIR1244−3
APOLD1
GSG1
MIR1244−2
GPR19
GPRC5D
MIR1244−1
DUSP16
DDX47
KIAA1467
C12orf36
RNU6−19P
MANSC1
CREBL2
MIR614
EMP1
ETV6
BCL2L14
LOH12CR2
CDKN1B
LOC100506314
GRIN2B
−log10(P−value)
chr 12 (MB)
12.0
12.5
13.0
13.5
14.0
C
LOC392232
RNU6−83P
MSC
TRPA1
LOC101926908
EYA1
LOC100132891
KCNB2
RP11-383H13.1
−log10(P−value)
chr 8 (MB)
72.0
72.5
73.0
73.5
D
CBWD1
FOXD4
FAM138C
WASH1
DDX11L5
DOCK8
C9orf66
KANK1
DMRT2
LINC01230
DMRT1
DMRT3
−log10(P−value)
chr 9 (MB)
0.2
0.4
0.6
0.8
1.0
1.2

Figure 5. TWAS using melanocyte eQTL 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 x 10^{-7}, TWAS Z = 5.2) is shown in green, along with a second marginally significant gene, *CDH1* (*P* = 5.27 x10^{-4}, Z = 3.47) 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 of for melanoma TWAS gene *HEBP1* (TWAS *P* = 4.65 x 10^{-7}, TWAS Z=-5.04) and a second marginally significant gene, *GPRC5A* (TWAS *P*= 2.8 x 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 x 10^{-6}, Z = 4.6) and *RP11-383H13.1* (*P* = 6.68 x 10^{-6}, Z = 4.5) on chromosome band 8q13.3. The Manhattan plot show 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, *CBWD* (*P* = 5.52 x 10^{-6}, Z = -4.54) and a marginally significant gene, *DOCK8* (*P* = 2.7 x 10^{-3}, Z = 2.99) on chromosome band 9p24.3.

melanocytes ranked third (**Supplemental Table 24**). Collectively, melanocyte and GTEx datasets identified 22 TWAS genes at six previously known melanoma GWAS loci (chr1q21.3, chr1q42.1, chr2q33-q34, chr15q13.1, chr21q22.3, and chr22q13.1) as well as nine TWAS genes at eight novel loci. Melanocyte eQTLs alone identified five of the nine novel TWAS genes, three of which (*HEBP1*, *MSC* and *RP11-383H13.1*) were only genome-wide significant when using the melanocyte eQTL dataset (**Supplemental Table 25**). In contrast, none of the 44 GTEx tissue datasets produced more than one novel association for melanoma. Five 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 datasets. Using melanocyte data, we noted two significant genes on chr16q22.1: *ZFP90* and *CDH1.* Upon conditioning the analysis on predicted expression of *ZFP90,* little GWAS signal remained, and *ZFP90* was jointly significant in this locus, while considerable GWAS signal still remained when the analysis was conditioned on *CHD1* (joint *P*-values are 1.9×10^{-7} for *ZFP90* and 0.43 for *CDH1* when conditioned on *ZFP90*; **Fig. 5A**). Similarly, conditional analyses indicated that *MSC* but not *RP11-383H13.1* explains the most of melanoma GWAS signal at chr8q13.3 (Joint *P* values are 4.3×10^{-6} for *MSC* and 0.9 for *RP11-383H13.1* when conditioned on *MSC*; **Fig.5C**). The same approach was then applied to the loci with marginally significant TWAS signals for multiple genes or a single gene coming from multiple tissue reference sets including melanocytes and GTEx tissues. Chr1q21.3 among them was the most complicated melanoma locus, where nine genes from 41 tissue types (**Supplemental Table 25**) exhibited genome-wide significant TWAS signal. Among these genes, *CTSS* was significant in a joint model, and when conditioned on *CTSS*, most of the GWAS signal disappeared (**Supplemental Fig. S12**). This analysis further identified significant TWAS genes that explain the most of the GWAS signal for

two other complex loci (*CASP8* on chr2q33-q34, and *MAFF and DMC1* on chr22q13.1; **Supplemental Table 25**).

DISCUSSION

In this study, we established a cell-type specific eQTL dataset using primary cultures of human melanocytes. Our dataset 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 dataset, we further identified novel melanoma TWAS loci. Our data highlight the utility of building even a modestly sized cell-type specific dataset.

Through *trans*-eQTL analysis, this melanocyte dataset identified IRF4 as a potential regulator of melanocytic-lineage specific gene expression for a set of downstream genes. IRF4, or Interferon regulatory factor 4, was initially known as a transcription factor expressed in cells of immune system with roles in immune responses (Huber and Lohoff 2014), but variation near this gene has also been found to be associated with human pigmentation phenotypes (Han et al. 2008) as well as number of melanocytic nevi (Duffy et al. 2010a). A recent functional study identified rs12203592 as a functional variant regulating *IRF4* expression through allele-specific binding of TFAP2A and mediating pigmentation phenotype via regulation of the melanin synthesis enzyme, TYR (Praetorius et al. 2013). rs12203592 is a *cis*-eQTL for *IRF4* in several other tissues in GTEx dataset including sun-exposed skin and whole blood. Notably, skin eQTL displays the same directional changes relative to the SNP as in melanocytes albeit with much smaller effect, while eQTLs of all the other tissues including blood show the opposite directional change of *IRF4* levels. In a blood cell *trans*-eQTL meta-dataset (Westra et al. 2013) (Blood eQTL browser; http://genenetwork.nl/bloodeqtlbrowser/), rs12203592 is a *trans*-eQTL for a

different set of genes than those in melanocytes. These observations suggest cell-type specific nature of *trans*-eQTLs as well as cell-type specific regulation of transcription factors. Our data suggests that 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, are potential 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.

Our melanocyte eQTL dataset nominated several candidate genes near known melanoma risk loci that may potentially play a role in mediating melanoma risk. The majority (68%) of these candidate susceptibility genes are not the closest gene to the lead GWAS SNP or larger association signal. Among them, *MAFF* on chr22q13.1 is one of the most significant melanocyte eQTL genes, with the eQTL clearly colocalizing with the melanoma GWAS locus (supported by eCAVIAR results). At this locus, the GWAS SNP is located in the first intron of *PLA2G6*, ~25 kb 5' of *MAFF*. Further, *MAFF* was also the most significant melanoma TWAS gene when using melanocyte expression data as a reference. *MAFF* is a small Maf protein regulated by EGF signaling (Amit et al. 2007) and plays a role in the oxidative stress response (Katsuoka et al. 2005). These findings warrant further functional study in the context of melanoma risk, given the vulnerability of melanocytes to oxidative stress attributable to melanin production (Denat et al. 2014). For the risk locus at chromosome band 1q21.3, five melanocyte eQTL genes overlapped the GWAS signal (*RP11-316M1.12, ARNT*, *CTSS*, *CERS2*, and *CTSK*). Adding to this complexity, eQTLs for four additional genes (*HORMAD1*, *GOLPH3L*, *ANXA9*, and *SETDB1*) in GTEx skin tissues and TCGA melanoma datasets also overlapped this locus. While it is possible that different subsets of these genes play a role in different spatio-temporal contexts to contribute to melanoma susceptibility, our TWAS conditional analyses using melanocyte expression data (as well as data from 44 GTEx datasets) indicated that

expression of *CTSS* may contribute to the most of the melanoma GWAS signal at this locus. At chr16q24.3, another complex locus, eQTLs for seven genes overlapped the melanoma GWAS signal in melanocytes and thirteen more collectively from two GTEx skin types and TCGA melanoma datasets. While protein-coding variants of *MC1R,* determinants of "red hair color" phenotype (Beaumont et al. 2007) are thought to best explain the melanoma GWAS signal at this locus (Bishop et al. 2009), our data suggests that expression levels of additional genes including *CDK10* (cell cycle regulating gene) (Malumbres 2014) and *FANCA* (involved in genome stability) (Palovcak et al. 2017) may be under control of melanoma-associated variants in melanocytes and related tissues.

While melanocyte expression data suggest the majority of melanoma loci (11 or 20) show overlap with genes whose expression levels in actively growing primary melanocytes are correlated with common risk-associated genetic variants, we also noted examples of potential context-dependent or cell non-autonomous gene expression contributing to melanoma susceptibility. For example, eQTLs for *ASIP,* an important paracrine regulator of melanogenesis (Suzuki et al. 1997), overlapping the chr20q11.2-q12 melanoma locus were only observed in sun-exposed skin and tibial nerve tissue but not in melanocytes. This is consistent with the findings that *ASIP* is not expressed in melanocytes of epidermis but in dermal components of skin (Liu et al. 2015). *ASIP* encodes an antagonist to melanocortin 1 receptor, MC1R, on the melanocytic cell membrane (Wolf Horrell et al. 2016), and is not expressed at a detectable levels in our melanocytes (median FPKM = 0). Analyses including both melanocyte and skin tissue data are thus complementary and allow for a more comprehensive understanding of context-dependent eQTLs underlying melanoma GWAS loci.

TWAS can be a powerful approach to leverage tissue-specific eQTL datasets and reduce multiple testing burden associated with GWAS. While identification of trait-associated gene expression differences via TWAS cannot be taken to imply causality for a specific gene,

TWAS may nonetheless nominate plausible candidate risk genes at significant loci. Here, we applied our cell-type specific reference dataset to TWAS of melanoma risk, identifying 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). Higher *ZFP90* mRNA levels (TWAS $P = 1.95 \times 10^{-7}$) were associated with melanoma risk. This gene encodes a zinc-finger protein implicated in susceptibility to obesity (Schadt et al. 2005) as well as systemic lupus erythematosus (Morris et al. 2016) with strong eQTL evidence in mouse liver and human blood cells, respectively, suggesting that this gene is under strong genetic control by heritable variants. At chr12p13.1, lower expression levels of *HEBP1*, a ubiquitously expressed heme-binding protein found in vesicles and in the nucleus which influences the MAP kinase pathway via proteolysis (Devosse et al. 2011), were found to be associated with melanoma (TWAS $P = 4.48 \times 10^{-7}$). A suggestive association was also observed for a second gene at this locus, *GPRC5A* (TWAS $P = 2.8 \times 10^{-5}$, lower levels were associated with melanoma), encoding a G-protein coupled receptor implicated in oncogenes suppression in lung cancer (Wang et al. 2016). Increased *MSC* mRNA levels (chr8q13.3;TWAS $P = 4.27 \times 10^{-6}$) were associated with melanoma; this gene encodes an E-protein-binding transcription factor, musculin, with known roles in myogenesis (Lu et al. 1999) and induced regulatory T cell development (Wu et al. 2017). Lastly, *CBWD1* at the chr9p24.3 locus (TWAS $P = 5.52 \times 10^{-6}$, lower levels were associated with melanoma) encodes cobalamin synthetase W domain-containing protein 1. While expression levels of *CBWD1* explain most of the melanoma GWAS signal in this locus by conditional analysis, an adjacent gene, *DOCK8*, also showed a suggestive TWAS association ($P = 6.6 \times 10^{-3}$, higher levels were associated with melanoma) and is a strong candidate gene given its role as a GTP exchange factor and activator of the oncogene, *RAC1* (Wang et al. 2015), a recurrently mutated somatic driver gene in cutaneous melanoma (Halaban 2015).

Combining melanoma TWAS data generated from each of 44 GTEx tissue types identified four additional new genes/loci (*ERCC2*, *KIF9*, *MRAP2*, and *ZBTB4*). *ERCC2* on chr19q13.32 is a nucleotide excision repair gene targeting UV-induced DNA damage and implicated in Xeroderma Pigmentosum (Taylor et al. 1997). *KIF9* on chr3p21.31 encodes kinesin-like protein KIF9 which is required in spindle length control and mitotic progression (Andrieu et al. 2012). *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). *ZBTB4* gene product is a methyl-CpG-binding protein, loss of which promotes tumorigenesis by increasing genome instability (Roussel-Gervais et al.). Seemingly relevant functions of these new candidate genes warrant further studies on their roles in melanomagenesis. In all, our primary melanocyte eQTL dataset 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 datasets.

METHODS

Melanocyte culture

We obtained frozen aliquots of melanocytes isolated from foreskin of 106 healthy newborn males who are mainly of European descent following an established protocol (Halaban et al. 2000) from SPORE in Skin Cancer Specimen Resource Core at Yale University. Cells were grown in Dermal Cell Basal Medium (ATCC® PCS-200-030™) supplemented with 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, and trypsinized with Trypsin/EDTA Solution (R-001-100, Cascade Biologics) as well as Trypsin Neutralizer Solution (R-002-100, Cascade

Biologics). Media was changed every 2-3 days when necessary. Throughout the whole process two specific lot numbers of medium and supplement were used for consistency, and for the final passage of at least 2 days before harvesting cells for RNA and DNA, a single lot number of medium and supplement was used for the whole panel. Every step of cell culture, DNA/RNA isolation, and sequencing/genotyping processes were performed in randomized batches. Melanocytes were first thawed and cultured for two passages to be expanded and re-frozen. A fraction of samples failed to grow back from frozen stocks or harbored irrepressible contamination, and was excluded from the study. The final number of samples included in the eQTL analysis was 106. For RNA and DNA isolation, a vial of expanded stock for each sample was thawed and cultured for two passages and harvested at log phase to capture gene expression profiles of actively proliferating cells. First round and second round of cultures were performed in randomized batches to minimize variability that could be introduced by culture condition over time. During both rounds of cultures, cell morphology, pigmentation status, and growth speed were closely monitored and recorded. Before harvesting the cell, 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

Cells were harvested at log phase by trypsinization and kept at -80C°. Genomic DNA was isolated from frozen pellets in randomized batches using the Gentra Puregene Cell Kit (158745, Qiagen). DNA quantity and quality was assessed by NanoDrop8000 spectrophotometer, Quant-iT PicoGreen dsDNA Assay kit (Thermo Fisher Scientific), and agarose gel running. All samples were profiled using Applied Biosystems Identifiler STR panel prior to genotyping to exclude unexpected duplicate samples or contaminated samples from the analyses. DNA samples were then genotyped on the Illumina OmniExpress arrays (HumanOmniExpress-24-v1-1-a) in randomized batches of 24 samples per chip at the Cancer Genomics Research Laboratory of

the Division of Cancer Epidemiology and Genetics (NCI/NIH). Genome Studio (Illumina, Inc.) was used to cluster and normalize the raw genotyping data and the genotype calls and SNP statistics were generated and exported to PLINK format (https://www.cog-genomics.org/plink2). After genotype quality assessment, only the samples passing the following cutoffs were included in the further analyses: minor allele frequency (MAF>0.01), genotype call rate (both sample and SNP call rate>0.95), Hardy Weinberg Equilibrium (HWE>10^{-7}), and sample heterozygosity (>0.3 & <0.38). Included samples also passed gender check (all male) and cryptic relatedness test (Pi_HAT < 0.05), and computed principal components (PCs), assessment of ancestry and population structure were also performed by using PLINK2 and GLU (https://github.com/bioinformed/glu-genetics). After genotype quality control, genotypes were imputed using Michigan Imputation Server (Das et al. 2016) based on 1000 Genomes (Phase 3, v5) reference panel and Mixed population, and using SHAPEIT for pre-phasing. Post-imputation genetic variants (single nucleotide variants (SNP) and small insertion-deletion (indel) polymorphisms) with MAF<0.01 or imputation quality scores (R-squared) <0.3 were removed from the final analysis. For the duplicate samples, the sample with the highest SNP call rate was selected. Due to the small sample size, we included all samples that pass genotyping QC but histologically carry a range of African and Asian ancestry while accounting for ancestry in the further analyses as covariates. Based on estimation of population admixture using ADMIXTURE analysis, three individuals displayed mainly African ancestry (YRI score > 0.8) and another 12 admixture of European and African ancestries (both CEU and YRI scores > 0.2 and < 0.8). Three other individuals displayed mainly Asian ancestry (CHB score > 0.8) and four more individuals admixture of European and Asian ancestries (both CEU and CHB scores > 0.2 and < 0.8). The remaining 84 samples displayed mainly European ancestry (CEU score > 0.8 or CEU score > 0.7 with secondary ancestry scores < 0.2). For eQTL analysis, we included the top 3 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

Cell were harvested at log phase by washing with cold PBS on ice followed by lysis with QIAzol lysis reagent and stored at -80°C. Total RNA was isolated using miRNeasy Mini Kit (217004, Qiagen) in randomized batches with following changes to the protocol. To reduce melanin carryover, cell lysate in QIAzol reagent was heated at 65°C for 5 minutes before the addition of chloroform and phase separation. To maximize the yield, pass-through after binding was re-loaded on the miRNeasy column. After elution and before storage, RNase inhibitor (RNaseOUT Recombinant Ribonuclease Inhibitor, 10777019, Thermo Fisher Scientific) was added to purified RNA. RNA quantity and quality was assessed using NanoDrop8000 spectrophotometer and Bioanalyzer, which yielded RIN>9 for all 106 samples and RIN=10 for >75% of the samples. Poly-A selected stranded mRNA libraries were constructed from 1 µg total RNA using the Illumina TruSeq Stranded mRNA Sample Prep Kits according to manufacturer's instructions except where noted. Amplification was performed using 10 cycles to minimize the risk of over-amplification. Unique barcode adapters were applied to each library, and libraries were pooled in randomized batches for sequencing. The pooled libraries were sequenced on multiple lanes of a HiSeq2500 using version 4 chemistry to achieve a minimum of 45 million 126 base paired reads. The data was processed using RTA version 1.18.64 and CASAVA 1.8.2. The updated pipeline originally published by University of North Carolina for processing TCGA RNA-sequencing data (https://cghub.ucsc.edu/docs/tcga/UNC_mRNAseq_summary.pdf) was used to analyze the RNA- sequence data generated from our samples. In summary, starting from raw Illumina FASTQ files, FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), RSeQC (Wang et al. 2012) and QuickRNAseq (Zhao et al. 2016) were used to do the quality check and data interactive visualization. STAR (version 2.5.0b) (Dobin et al. 2013) was used for

aligning reads to the human genomic reference (hg19) with the gene annotation 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 to transcripts per million (TPM), and then quantile normalization was applied to the TPM in all samples to obtain the final RSEM value. Using the RNA-Seq BAM file and after QC genotyping data, we further checked whether the reads are contaminated as a mixture of two samples by VerifyBamID and no contamination was found (Jun et al. 2012). For eQTL analysis, we used the same method of post-processing gene expression data to GETx project (http://www.gtexportal.org). Genes were selected based on expression thresholds of >0.5 RSEM in at least 10 samples and ≥6 reads in at least 10 samples. Expression values were quantile normalized to the average empirical distribution observed across samples. For each gene, expression values were inverse quantile normalized to a standard normal distribution across samples. A set of covariates identified using the Probabilistic Estimation of Expression Residuals (PEER) method (Stegle et al. 2010) were 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 (15 factors for N<150).

Identification of *cis*-eQTLs in primary melanocytes

Cis-eQTLs analysis was performed closely following a recent standard procedure adopted by GTEx (Aguet et al. 2016) (https://www.gtexportal.org/home/documentationPage#staticTextAnalysisMethods). In summary, *cis*-eQTL mapping was performed using FastQTL (Ongen et al. 2016), using the expression data, imputed genotype data and covariates described above. First, nominal *P*-values were generated for each variant-gene pair by testing the alternative hypothesis that the slope of a linear regression model between genotype and expression deviates from 0. Genetic

variants located within +/- 1Mb of the TSSs for each gene were tested for *cis*-eQTL effects of the corresponding gene. Variants in imputed VCF were selected based on the following thresholds: the minor allele was observed in at least 10 samples (the minor allele frequency was ≥ 0.05). Second, the beta distribution-adjusted empirical *P*-values from FastQTL were 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 adaptive permutations mode was used with the setting "--permute 1000 10000". 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). Last, to identify the list of all significant variant-gene pairs associated with eGenes, a genome-wide empirical *P*-value threshold, p_t, was defined as the empirical p-value of the gene closest to the 0.05 FDR threshold. p_t was then used to calculate a nominal p-value threshold for each gene based on the beta distribution model (from FastQTL) of the minimum *P*-value distribution $f(p_{min})$ obtained from the permutations for the gene. Specifically, the nominal threshold was calculated as $F^{-1}(p_t)$, where F^{-1} is the inverse cumulative distribution. For each gene, variants with a nominal p-value below the gene-level threshold were considered significant and included in the final list of variant-gene pairs.

Identification *cis*-eQTLs in TCGA SKCM dataset

Level 2 TCGA SKCM Genotyping data were collected from NCI Genomic Data Commons Data Portal (GDC Legacy Archive) (https://portal.gdc.cancer.gov). Only Blood Derived Normal sample genotyping data were kept (TCGA sample type code 10). Genotype calls for the 906,600 SNP probes assayed using Affymetrix Genome-Wide Human SNP Array 6.0 platform were processed using Birdseed. Genotype calls were coded as 0, 1 or 2 according to the number of variant alleles and filtered according to a Birdseed confidence threshold of 0.05. The genotyped data were subject to additional quality control filtration criteria same as previous

melanocyte genotyping data processing. SNPs with call rates <0.9 or minor allele frequencies (MAF) <0.05 were excluded, as were SNPs out of Hardy Weinberg equilibrium with $p<10^{-6}$. All samples with a call rate below 0.95 were excluded. Cryptic Relatedness Check (Pi_HAT<0.05), computed PCs, assessment of ancestry and population structure were also performed using PLINK2 and GLU. In total, 443 samples and 736,502 SNPs passing all the criteria were included in the analysis. Genotypes passing the QC were also imputed using the Michigan Imputation Server (Das et al. 2016) based on 1000 Genomes (Phase 3, v5) reference panel and EUR population and using SHAPEIT for pre-phasing. Similar post-imputation filters were applied as follows. MAF<0.01 or imputation quality scores (R-squared) <0.3 were removed from the final analysis. Level 3 RNA-Seq data of TCGA SKCM were also collected from NCI Genomic Data Commons Data Portal (GDC Legacy Archive). The expression quantification was quantile normalized as "RSEM" (Li and Dewey 2011) with the NCBI gene annotation reference ("TCGA.hg19.June2011.gaf"). Genes were selected based on expression thresholds of RSEM >=3 and SD >0.2. In addition, Level 3 SKCM Copy number data were collected from cBioPortal (http://www.cbioportal.org) (Cerami et al. 2012; Gao et al. 2013). Both relative linear DNA copy-number values and putative copy-number calls were determined using GISTIC 2.0 (Mermel et al. 2011) and obtained for each gene. We performed a similar *cis*-eQTL analysis for TCGA SKCM data by using FastQTL (Ongen et al. 2016). For the covariates, both the top 3 genotyping principal components and regional DNA copy numbers were included. Two *cis*-eQTL datasets were generated based on the copy number covariates. First, we performed eQTL analysis using the relative linear copy-number values as a covariate for 349 TCGA SKCM samples with expression and genotyping data available. Second, based on the putative copy-number calls determined using GISTIC 2.0, we performed the same *cis*-eQTL analysis using FastQTL exclusively for each gene and only included the sample with copy neutral status for the same tested gene. The first method was used for comparisons with other datasets throughout the manuscript.

Pairwise eQTL sharing between primary melanocytes and 44 GTEx tissues

To assess replication of *cis*-eQTL and eGenes in the publicly available Genotype Tissue Expression (GTEx) project (Consortium 2013; Aguet et al. 2016), we collected eQTL results from 44 tissue types with >= 70 samples (The GTEx Analysis V6p, https://gtexportal.org/). The eGene and significant SNP-gene associations based on permutations were collected (GTEx_Analysis_v6p_eQTL.tar) and every SNP-gene association test (including non-significant tests) were download from GTEx website. To test the sharing of all significant SNP-gene pairs of our melanocytes eQTL study with the ones identified in 44 tissue types by GTEx, we used the threshold of *FDR* <0.05 and calculated the pairwise π_1 statistics. We used Storey's QVALUE software (Storey and Tibshirani 2003) (https://github.com/StoreyLab/qvalue) to calculate the π_1, which indicates the proportion of true positives. A heat map was drawn based on the pairwise π_1 values. The pairwise π_1 statistics are reported for single-tissue eQTL discoveries in each tissue. 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 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*-eQTL in primary melanocytes

The nominal p-values for gene-SNP pairs in *trans*-eQTL analysis were calculated using the Matrix-eQTL program (Shabalin 2012). We performed multiple testing to identify significant *trans*-eQTLs following our previous approach (Shi et al. 2014). SNPs with call rates <0.9 or minor allele frequencies (MAF) <0.05 were excluded, as were SNPs out of Hardy Weinberg equilibrium with $p<10^{-6}$. For each nominal *P*-value threshold p, we calculated the number of genes (denoted as $N_1(p)$) that has 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 FDR as $FDR = N_0(p)/N_1(p)$. We chose $p = 3.98 \times 10^{-11}$ to control FDR as a desired level 0.1. Among the genes that pass the genome-wide significance cutoff, only the *trans*-genes with mappability > 0.8 were reported following the procedures described below.

Mappability of transcripts for *trans*-eQTL quality control

Mappability of transcripts for *trans*-gene was measured following a recently published work by GTEx consortium (GTEx-Consortium et al. 2017). Mappability of every *k*-mer of the reference human genome (hg19) computed by the ENCODE project was downloaded from the UCSC genome browser under accession; wgEncodeEH000318 and wgEncodeEH000317. Exon- and untranslated region (UTR)- mappability of a gene were computed as the average mappability of all *k*-mers in exonic regions and UTRs, respectively. We chose *k* = 100 for exonic regions, as it was the closest to our RNA-Seq read length among all available values of *k*. However, as UTRs are generally small regions, and 36 is the smallest among all possible values of *k*, we chose *k* = 36 for UTRs. We computed the mappability of a gene using the weighted average of its exon-mappability and UTR mappability, weights being proportional to the total length of exonic regions and UTRs, respectively.

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

We applied the Genomic Mediation analysis with Adaptive Confounding adjustment (GMAC) (Yang et al. 2017) algorithm to identify *cis*-mediators for *trans*-eQTLs in primary melanocytes eQTL data. We considered only the trios with evidence of both *cis* and *trans* association. For genes with multiple *cis*-eSNPs as eQTLs, only one *cis*-eSNP for each gene (i.e., the high-quality

SNP with the smallest *P*-value) was selected and was included in the subsequent mediation analyses. When a *trans*-association *P*-value is less than 10^{-5}, the eSNP, its corresponding *cis* transcript and the trans transcript were treated as a candidate trio and were retained for mediation analysis. Up to 5 PEER factors and other covariates (top 10 genotype PCs) were adjusted. 100,000 permutations for testing mediation were performed and trios with suggestive mediation were reported using mediation *P*-value threshold <0.05.

Allele specific Expression (ASE)

ASE analysis was performed based on the GATK best practices pipeline in allelic expression analysis published by the Broad Institute (Castel et al. 2015). We included heterozygous loci in exonic regions when the imputation quality was R^2>0.9 and probability of heterozygosity >95%. Specifically, RNA-Seq data BAM files were post-processed according to the pipeline: duplicate marking using Picard (https://broadinstitute.github.io/picard/), and indel realignment and base recalibration using GATK (https://software.broadinstitute.org/gatk/). The ASEReadCounter implemented in GATK was used to calculate read counts per allele per sample (Castel et al. 2015).

We included only expressed heterozygous variant sites with more than 30 reads combined from both alleles and performed the significance testing after down-sampling to 30 reads per site. We excluded loci with >5% mapping bias kindly provided by Dr. Halit Ongen at the university of Geneva, Switzerland(Ongen et al. 2014) and loci in regions with low mappability (ENCODE 125bp mappability score < 1) from the final analysis. Two methods were used to generate the significant allelic expression sites. First, 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). To minimize false ASE events resulting from mapping bias specific to different DNA bases, we compared the observed allelic ratio for each coding heterozygous variant to the overall ratio for that specific allele pair in each sample (i.e.

for each of the following pairs: AC, AG, AT, CA, CG, CT, GA, GC, GT, TA, TC, TG heterozygotes)(Lappalainen et al. 2013), so that we were able to take into account the expected allele imbalance in the ASE analysis. In addition, the effect size of allelic expression (AE, defined as |0.5-Reference ratio|) were 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 Singed Ranked Test). Second, we applied the similar binomial test to each heterozygous locus but combined down-sampled 30 reads from all heterozygous individuals. The same criteria (FDR<0.05 or effect size>0.15) was applied to generate the significant ASE genes.

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 Roadmap Epigenomics project (http://www.roadmapepigenomics.org). Promoter and enhancer histone marks as well as DHS information of penis foreskin primary melanocytes are available for different subsets of three individuals: ChIP-seq BED files for H3K4me1 (skin01: GSM941728; skin02: GSM941730; skin03: GSM958152) and H3K4me3 from 3 individuals (skin01: GSM941719; skin02: GSM941731; skin03: GSM958151), H3K27AC for 2 individuals (skin01: GSM1127072; skin03: GSM1127073 and GSM958157), and DHS data from 2 individuals (skin01: GSM774243 and GSM774244; skin02: GSM1027307and GSM1027312) were downloaded. MACS2 (http://liulab.dfci.harvard.edu/MACS/) (Zhang et al. 2008) was used to generated histone H3K4me1/Histone H3K4me3/Histone H3K27ac/DHS peaks using the default settings and FDR

1% cutoff. 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. 2017). Briefly, we performed randomizations for testing whether an eQTL SNP set is enriched for a given histone mark regions. Note that 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 nth sampled SNP set, we calculated the number (denoted as x_n) of SNPs functionally related with the histone mark. We had $\{x_1, \cdots, x_{10000}\}$, corresponding to the sampled 10000 SNP sets, and 3.) Enrichment fold change was calculated as $FC = \frac{X_0}{\frac{\sum_{n=1}^{10000} x_n}{10000}}$, 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 $P = \{n: x_n \geq X_0\}/10000$, 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 set, we reported 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, intergenic and lncRNA region as described in R annotatr package (https://github.com/hhabra/annotatr).

Enrichment of melanoma GWAS variants in eQTLs

Two methods were used to evaluate if the melanoma GWAS variants are enriched in eQTLs of different datasets. First, QQ plots were used to show the differences in melanoma association *P*-values between the significant eQTL SNPs and non-eQTL SNPs. For all the GWAS variants, we first performed LD pruning using PLINK ($r^2 = 0.1$ and window size 500 kb), so that the

remaining SNPs are independent. Then, based on eQTL data, these pruned SNPs were classified into two groups. If a SNP is an eQTL or is in LD ($r^2 > 0.8$) with an eQTL SNP, the SNP is classified as eQTL SNPs. Otherwise, it is classified as non-eQTL SNPs. QQ plots were generated using the melanoma GWAS *P*-values from the most recent meta-analysis (Law et al. 2015) for eQTL SNPs versus non-eQTL SNPs. Deviation from the 45-degree line indicates that melanoma GWAS SNPs are strongly enriched in eQTL SNPs. The lambda values were estimated using the "estlambda2" function in R package "QQperm". For the second method, a similar simulation procedure was applied to identify overlap and test for enrichment of eQTLs in melanoma GWAS SNPs (Hannon et al. 2016). All significant eQTLs were 'clumped' based on the best eQTL *p* value using PLINK to create a list of quasi-independent SNPs ($r^2<0.25$ for all pairs of SNPs within 250kb) and to prevent LD between SNPs in the set biasing the results. A more stringent clumping procedure was used for SNPs located in chr5:25000000-35000000, where the window for pairwise SNP comparisons was extended to 10,000kb. 1,000,000 simulated sets matched for allele frequency were drawn to calculate the expected overlap between the eQTL SNP and melanoma GWAS variants at four GWAS significance thresholds (*P*<5E-5,5E-6,5E-7,5E-8) and generate empirical *P*-values. Empirical significance for enrichment of eQTLs in GWAS variants was ascertained by counting the number of simulations with at least as many SNPs set and dividing by the number of simulations performed. Fold change statistics were calculated as the true overlap divided by the mean overlap of these simulations.

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). eCAVIAR is a statistical framework

that quantified the probability of the variant to be causal both in GWAS and eQTL studies, while allowing an arbitrary number of causal variants. For each locus, both GWAS and eQTL (from human melanocytes cultures in our study and two GTEx skin tissues) summary statistics of selected variants in that locus were extracted as the input for eCAVIAR. We selected 50 SNPs in both upstream and downstream of GWAS lead SNP for each GWAS locus. We computed the CLPP score with maximum number of two causal SNPs in each locus. We used CLPP >1% (0.01) cutoff for co-localization. Thus, for a given GWAS variant, an eGene with a CLPP score above the colocalization cutoff is considered a target gene.

Performing TWAS with GWAS summary statistics

We performed 45 transcriptome-wide association studies (TWAS) by predicting the function/molecular phenotypes into GWAS using melanoma GWAS summary statistics and both GTEx and melanocyte RNA-Seq expression data. The new framework 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 data including no significance thresholding in LD-score format (https://github.com/bulik/ldsc/wiki/Summary-Statistics-File-Format) 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 post mortal tissue were downloaded (http://gusevlab.org/projects/fusion/). We computed our functional weights from our melanocyte RNA-Seq data one gene at a time. Genes that failed quality control during heritability check (using minimum heritability *P*-value 0.01) were excluded from the further analyses. We restricted the *cis*-locus to 500kb 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 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 analysis

IRF4 motif enrichment analysis were performed using the AME module in The MEME Suite (http://meme-suite.org) and inputted shuffled sequences as control. *IRF4* motif were download from HOCOMOCO v10 database (http://hocomoco.autosome.ru/motif/IRF4_HUMAN.H10MO.C). All the statistical analyses were performed in R (https://cran.r-project.org/).

DATA ACCESS

Both genotyping data and RNA sequencing data of our 106 primary human melanocytes from this study are in the process of being submitted to the database of Genotypes and Phenotypes (dbGAP).

Members of the Melanoma Meta-Analysis Consortium

Law MH[1*], Bishop DT[2*], Lee JE[3#], Brossard M[4,5#], Martin NG[6], Moses EK[7], Song F[8], Barrett JH[2], Kumar R[9], Easton DF[10], Pharoah PD[11], Swerdlow AJ[12,13], Kypreou KP[14], Taylor JC[2], Harland M[2], Randerson-Moor J[2], Akslen LA[15,16], Andresen PA[17], Avril MF[18], Azizi E[19,20], Scarrà GB[21,22], Brown KM[23], Dębniak T[24], Duffy DL[6], Elder DE[25], Fang S[3], Friedman E[20], Galan P[26], Ghiorzo P[21,22], Gillanders EM[27], Goldstein AM[23], Gruis NA[28], Hansson J[29], Helsing P[30], Hočevar M[31], Höiom V[29], Ingvar C[32], Kanetsky PA[33], Chen WV[34]; GenoMEL Consortium; Essen-Heidelberg Investigators; SDH Study Group; Q-MEGA and QTWIN Investigators; AMFS Investigators; ATHENS Melanoma Study Group, Landi MT[23], Lang J[35], Lathrop GM[36], Lubiński J[24], Mackie RM[35,37], Mann GJ[38], Molven A[16,39], Montgomery GW[40], Novaković S[41], Olsson H[42,43], Puig S[44,45], Puig-Butille

JA[44,45], Wu W[46,47], Qureshi AA[48], Radford-Smith GL[49,50,51], van der Stoep N[52], van Doorn R[28], Whiteman DC[53], Craig JE[54], Schadendorf D[55,56], Simms LA[47], Burdon KP[57], Nyholt DR[40,58], Pooley KA[10], Orr N[59], Stratigos AJ[14], Cust AE[60], Ward SV[7], Hayward NK[61], Han J[46,47], Schulze HJ[62], Dunning AM[11], Bishop JA[2], Demenais F[4,5#], Amos CI[63#], MacGregor S[1*], Iles MM[2*].

[1]Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [2]Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK. [3]Department of Surgical Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas, USA. [4]INSERM, UMR 946, Genetic Variation and Human Diseases Unit, Paris, France. [5]Institut Universitaire d'Hématologie, Université Paris Diderot, Sorbonne Paris Cité, Paris, France. [6]Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [7]Centre for Genetic Origins of Health and Disease, Faculty of Medicine, Dentistry and Health Sciences, University of Western Australia, Perth, Western Australia, Australia. [8]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, China. [9]Division of Molecular Genetic Epidemiology, German Cancer Research Center, Heidelberg, Germany. [10]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [11]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK. [12]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK. [13]Division of Breast Cancer Research, The Institute of Cancer Research, London, UK. [14]Department of Dermatology, University of Athens School of Medicine, Andreas Sygros Hospital, Athens, Greece. [15]Centre for Cancer Biomarkers (CCBIO), Department of Clinical Medicine, University of Bergen, Bergen, Norway. [16]Department of Pathology, Haukeland University Hospital, Bergen, Norway. [17]Department 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. [19]Department of Dermatology, Sheba Medical Center, Tel Hashomer, Sackler Faculty of Medicine, Tel Aviv, Israel. [20]Oncogenetics Unit, Sheba Medical Center, Tel Hashomer, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel. [21]Department of Internal Medicine and Medical Specialties, University of Genoa, Genoa, Italy. [22]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, Genoa, Italy. [23]Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health, Bethesda, Maryland, USA. [24]International Hereditary Cancer Center, Pomeranian Medical University, Szczecin, Poland. [25]Department of Pathology and Laboratory Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. [26]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, Bobigny, France. [27]Inherited Disease Research Branch, National Human Genome Research Institute, US National Institutes of Health, Baltimore, Maryland, USA. [28]Department of Dermatology, Leiden University Medical Center, Leiden, the Netherlands. [29]Department of Oncology-Pathology, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden. [30]Department of Dermatology, Oslo University Hospital, Rikshospitalet, Oslo, Norway. [31]Department of Surgical Oncology, Institute of Oncology Ljubljana, Ljubljana, Slovenia. [32]Department of Surgery, Clinical Sciences, Lund University, Lund, Sweden. [33]Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA. [34]Department of Genetics, University of Texas MD Anderson Cancer Center, Houston, Texas, USA. [35]Department of Medical Genetics, University of Glasgow, Glasgow, UK. [36]McGill University and Génome Québec Innovation Centre, Montreal, Quebec, Canada. [37]Department of Public Health, University of Glasgow, Glasgow, UK. [38]Centre for Cancer Research, University of Sydney at Westmead, Millennium Institute for Medical Research and Melanoma Institute Australia, Sydney, New South Wales, Australia. [39]Gade Laboratory for Pathology, Department of Clinical Medicine, University of Bergen, Bergen, Norway.

[40]Molecular Biology, the University of Queensland, Brisbane, Australia. [41]Department of Molecular Diagnostics, Institute of Oncology Ljubljana, Ljubljana, Slovenia. [42]Department of Oncology/Pathology, Clinical Sciences, Lund University, Lund, Sweden. [43]Department of Cancer Epidemiology, Clinical

Sciences, Lund University, Lund, Sweden. [44]Melanoma Unit, Departments of Dermatology, Biochemistry and Molecular Genetics, Hospital Clinic, Institut d'Investigacions Biomèdica August Pi Suñe, Universitat de Barcelona, Barcelona, Spain. [45]Centro de Investigación Biomédica en Red (CIBER) de Enfermedades Raras, Instituto de Salud Carlos III, Barcelona, Spain. [46]Department of Epidemiology, Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, Indiana, USA. [47]Melvin and Bren Simon Cancer Center, Indiana University, Indianapolis, Indiana, USA. [48]Department of Dermatology, Warren Alpert Medical School of Brown University, Providence, Rhode Island, USA. [49]Inflammatory Bowel Diseases, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [50]Department of Gastroenterology and Hepatology, Royal Brisbane and Women's Hospital, Brisbane, Queensland, Australia. [51]University of Queensland School of Medicine, Herston Campus, Brisbane, Queensland, Australia. [52]Department of Clinical Genetics, Center of Human and Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands. [53]Cancer Control Group, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [54]Department of Ophthalmology, Flinders University, Adelaide, South Australia, Australia. [55]Department of Dermatology, University Hospital Essen, Essen, Germany. [56]German Consortium for Translational Cancer Research (DKTK), Heidelberg, Germany. [57]Menzies Institute for Medical Research, University of Tasmania, Hobart, Tasmania, Australia. [58]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia. [59]Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, UK. [60]Cancer Epidemiology and Services Research, Sydney School of Public Health, University of Sydney, Sydney, New South Wales, Australia. [61]Oncogenomics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [62]Department of Dermatology, Fachklinik Hornheide, Institute for Tumors of the Skin at the University of Münster, Münster, Germany. [63]Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Hanover, New Hampshire, USA.

Members of the NISC Comparative Sequencing Program

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

NIH Intramural Sequencing Center, Bethesda, Maryland, 20892

ACKNOWLEDGMENTS

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. SM is supported by an Australian Research Council Fellowship.

DISCLOSURE DECLARATION (including any conflicts of interest)

REFERENCES

Aguet F, Brown AA, Castel S, Davis JR, Mohammadi P, Segre AV, Zappala Z, Abell NS, Fresard L, Gamazon ER et al. 2016. Local genetic effects on gene expression across 44 human tissues. *bioRxiv*.

Amit I, Citri A, Shay T, Lu Y, Katz M, Zhang F, Tarcic G, Siwak D, Lahad J, Jacob-Hirsch J et al. 2007. A module of negative feedback regulators defines growth factor signaling. *Nat Genet* **39**: 503-512.

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.

Andrieu G, Quaranta M, Leprince C, Hatzoglou A. 2012. The GTPase Gem and its partner Kif9 are required for chromosome alignment, spindle length control, and mitotic progression. *FASEB J* **26**: 5025-5034.

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. *Nature genetics* **43**: 1108-1113.

Beaumont KA, Shekar SN, Newton RA, James MR, Stow JL, Duffy DL, Sturm RA. 2007. Receptor function, dominant negative activity and phenotype correlations for MC1R variant alleles. *Hum Mol Genet* **16**: 2249-2260.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* **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.

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.

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E et al. 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**: 401-404.

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.

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

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.

Denat L, Kadekaro AL, Marrot L, Leachman SA, Abdel-Malek ZA. 2014. Melanocytes as instigators and victims of oxidative stress. *J Invest Dermatol* **134**: 1512-1518.

Devosse T, Dutoit R, Migeotte I, De Nadai P, Imbault V, Communi D, Salmon I, Parmentier M. 2011. Processing of HEBP1 by cathepsin D gives rise to F2L, the agonist of formyl peptide receptor 3. *J Immunol* **187**: 1475-1485.

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, Iles MM, Glass D, Zhu G, Barrett JH, Hoiom V, Zhao ZZ, Sturm RA, Soranzo N, Hammond C et al. 2010a. IRF4 variants have age-specific effects on nevus count and predispose to melanoma. *Am J Hum Genet* **87**: 6-16.

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

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.

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

Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E et al. 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**: pl1.

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

GTEx-Consortium, Laboratory DA, Coordinating Center -Analysis Working G, Statistical Methods groups-Analysis Working G, Enhancing Gg, Fund NIHC, Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida et al. 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. 2015. RAC1 and melanoma. *Clin Ther* **37**: 682-685.

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.

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.

Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR, Boehnke M, Kang HM. 2012. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am J Hum Genet* **91**: 839-848.

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.

Katsuoka F, Motohashi H, Ishii T, Aburatani H, Engel JD, Yamamoto M. 2005. Genetic evidence that small maf proteins are essential for the activation of antioxidant response element-dependent genes. *Mol Cell Biol* **25**: 8044-8051.

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.

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.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**: 323.

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 J, Fukunaga-Kalabis M, Li L, Herlyn M. 2014. Developmental pathways activated in melanocytes and melanoma. *Arch Biochem Biophys* **563**: 13-21.

Lu J, Webb R, Richardson JA, Olson EN. 1999. MyoR: a muscle-restricted basic helix-loop-helix transcription factor that antagonizes the actions of MyoD. *Proc Natl Acad Sci U S A* **96**: 552-557.

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. *Nature genetics* **43**: 1114-1118.

Malumbres M. 2014. Cyclin-dependent kinases. *Genome Biol* **15**: 122.

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.

Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. 2011. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**: R41.

Morris DL, Sheng Y, Zhang Y, Wang YF, Zhu Z, Tombleson P, Chen L, Cunninghame Graham DS, Bentham J, Roberts AL et al. 2016. Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. *Nat Genet* **48**: 940-946.

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.

Palovcak A, Liu W, Yuan F, Zhang Y. 2017. Maintenance of genome stability by Fanconi anemia proteins. *Cell Biosci* **7**: 8.

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.

Ramachandrappa S, Gorrigan RJ, Clark AJ, Chan LF. 2013. The melanocortin receptors and their accessory proteins. *Front Endocrinol (Lausanne)* **4**: 9.

Roussel-Gervais A, Naciri I, Kirsh O, Kasprzyk L, Velasco G, Grillo G, Dubus P, Defossez PA. 2017. Loss of the Methyl-CpG-Binding Protein ZBTB4 Alters Mitotic Checkpoint, Increases Aneuploidy, and Promotes Tumorigenesis. *Cancer Res* **77**: 62-73.

Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, Sieberts SK, Monks S, Reitman M, Zhang C et al. 2005. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet* **37**: 710-717.

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. *Nature communications* **5**: 3365.

Sitaram A, Marks MS. 2012. Mechanisms of protein delivery to melanosomes in pigment cells. *Physiology (Bethesda)* **27**: 85-99.

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 U S A* **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.

Suzuki I, Tada A, Ollmann MM, Barsh GS, Im S, Lamoreux ML, Hearing VJ, Nordlund JJ, Abdel-Malek ZA. 1997. Agouti signaling protein inhibits melanogenesis and the response of human melanocytes to alpha-melanotropin. *J Invest Dermatol* **108**: 838-842.

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 U S A* **94**: 8658-8663.

Wang J, Farris AB, Xu K, Wang P, Zhang X, Duong DM, Yi H, Shu HK, Sun SY, Wang Y. 2016. GPRC5A suppresses protein synthesis at the endoplasmic reticulum to prevent radiation-induced lung tumorigenesis. *Nat Commun* **7**: 11795.

Wang L, Wang S, Li W. 2012. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**: 2184-2185.

Wang SJ, Cui HY, Liu YM, Zhao P, Zhang Y, Fu ZG, Chen ZN, Jiang JL. 2015. CD147 promotes Src-dependent activation of Rac1 signaling through STAT3/DOCK8 during the motility of hepatocellular carcinoma cells. *Oncotarget* **6**: 243-257.

Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE et al. 2013. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**: 1238-1243.

Wolf Horrell EM, Boulanger MC, D'Orazio JA. 2016. Melanocortin 1 Receptor: Structure, Function, and Regulation. *Front Genet* **7**: 95.

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.

Wu C, Chen Z, Dardalhon V, Xiao S, Thalhamer T, Liao M, Madi A, Franca RF, Han T, Oukka M et al. 2017. The transcription factor musculin promotes the unidirectional development of peripheral Treg cells by suppressing the TH2 transcriptional program. *Nat Immunol* **18**: 344-353.

Yang F, Wang J, Consortium GT, 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, Lykke-Andersen S, Zhu B, Xiao W, Hoskins JW, Zhang X, Rost LM, Collins I, Bunt MV, Jia J et al. 2017. Characterising cis-regulatory variation in the transcriptome of histologically normal and tumour-derived pancreatic tissues. *Gut* doi:10.1136/gutjnl-2016-313146.

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 Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.

Zhao S, Xi L, Quan J, Xi H, Zhang Y, von Schack D, Vincent M, Zhang B. 2016. QuickRNASeq lifts large-scale RNA-seq data analyses to the next level of automation and interactive visualization. *BMC Genomics* **17**: 39.

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.