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

Tongwu Zhang,1,7 Jiyeon Choi,1,7 Michael A. Kovacs,1 Jianxin Shi,2 Mai Xu,1 NISC Comparative Sequencing Program,9 Melanoma Meta-Analysis Consortium,10 Alisa M. Goldstein,3 Adam J. Trower,4 D. Timothy Bishop,4 Mark M. Iles,4 David L. Duffy,5 Stuart MacGregor,5 Laufey T. Amundadottir,1 Matthew H. Law,5 Stacie K. Loftus,6 William J. Pavan,6,8 and Kevin M. Brown1,8

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

Most expression quantitative trait locus (eQTL) studies to date have been performed in heterogeneous tissues as opposed to specific cell types. To better understand the cell-type–specific regulatory landscape of human melanocytes, which give rise to melanoma but account for <5% of typical human skin biopsies, we performed an eQTL analysis in primary melanocyte cultures from 106 newborn males. We identified 797,335 cis-eQTL SNPs prior to linkage disequilibrium (LD) pruning and 4997 eGenes (FDR < 0.05). Melanocyte eQTLs differed considerably from those identified in the 44 GTEx tissue types, including skin. Over a third of melanocyte eGenes, including key genes in melanin synthesis pathways, were unique to melanocytes compared to those of GTEx skin tissues or TCGA melanomas. The melanocyte data set also identified trans-eQTLs including those connecting a pigmentation-associated functional SNP with four genes, likely through cis-regulation of IRF4. Melanocyte eQTLs are enriched in cis-regulatory signatures found in melanocytes as well as in melanoma-associated variants identified through genome-wide association studies. Melanocyte eQTLs also colocalized with melanoma GWAS variants in five known loci. Finally, a transcriptome-wide association study using melanocyte eQTLs uncovered four novel susceptibility loci, where imputed expression levels of five genes (ZFP90, HEBPI, MSC, CBWD1, and RP1L1–383H13.I) were associated with melanoma at genome-wide significant P-values. Our data highlight the utility of lineage-specific eQTL resources for annotating GWAS findings, and present a robust database for genomic research of melanoma risk and melanocyte biology.

[Supplemental material is available for this article.]

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

© 2018 Zhang et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
et al. 2017), or smooth muscle cells (Liu et al. 2018), eQTL data sets representing single primary cell types and direct comparison of these to the tissue type of origin have been lacking. eQTLs may be particularly useful for annotating variants associated with complex traits, as such variants are likely enriched for eQTLs (Nicolae et al. 2010). A recent study suggested that two-thirds of candidate common trait susceptibility genes identified as eQTLs are not the nearest genes to the GWAS lead SNPs, highlighting the utility of this approach in annotating GWAS loci (Zhu et al. 2016). Importantly, GWAS variants are enriched in eQTLs in a tissue-specific manner. For instance, whole blood eQTLs are enriched with autoimmune disorder-associated SNPs but not with GWAS SNPs for bipolar disease or type 2 diabetes (The GTEx Consortium 2015). These findings highlight the importance of using eQTL data sets from relevant cell types when following up GWAS loci for a specific disease. In addition to providing functional insights for known GWAS loci, eQTL data may be useful for identification of novel trait-associated loci via imputation of genotype-correlated gene expression levels into GWAS data sets (Gamazon et al. 2015; Gusev et al. 2016). Such approaches, usually referred to as transcriptome-wide association studies (TWASs), enable assignment of potentially disease-associated loci via estimations of their genetically regulated expression.

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

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

Results

Melanocyte eQTLs are distinct from those of other tissue types

In order to create a melanocyte-specific eQTL resource, we obtained primary melanocyte cultures isolated from foreskin of 106 healthy newborn males predominantly of European descent (Supplemental Table S1). We then cultured all 106 lines following a uniform procedure to harvest RNA and DNA, for RNA sequencing and genotyping, respectively (see Methods). Given the relatively small size of our sample set, we initially focused our analysis on local eQTLs (cis-eQTL), where we assessed the association between expression of each gene with common variants within ±1 Mb of transcription start sites (TSSs), following the best practices from the GTEx project (see Methods). In all, we identified 4997 “eGenes” (genes exhibiting association with genotypes of at least one SNP at FDR < 0.05) (Supplemental Table S2) and 597,335 genome-wide “significant eQTLs” (unique SNP-gene pairs showing FDR < 0.05; SNPs were not LD-pruned), which are higher numbers than any GTEx tissue type of similar sample size (Supplemental Table S3). Melanocyte eGenes were enriched with Gene Ontology (GO) terms including metabolic process, mitochondrial translation, biosynthetic process, catalytic activity, and ion-binding, as well as lysosome and metabolic pathways (Supplemental Table S4).

Further, melanocyte eGenes included 46% of genes categorized with GO terms as containing “melanin” (OCA2, TRPC1, CTNS, DCT, MCHR1, SLC45A2, TYR, BCL2, WNT7A, MC1R, and MYOSA) (http://amigo.geneontology.org) and 20% of curated pigment genes (based on human and mouse phenotype, OMIM, MGI) such as IRF4, TRPM1, and MC1R (Supplemental Tables S5, S6), reflecting pigmentation-related biology of melanocytes.

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

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

We next sought to determine whether melanocyte eQTLs were corroborated by allelic imbalance variants in heterozygous
the skin melanocyte eQTLs (estimated Lambda = 1.51) compared to three related tissue types as well as all the other GTEx tissue types (Supplemental Fig. S6).

Enrichment of melanocyte eQTLs was also observed in additional genomic features (Supplemental Figs. S4, S5; Supplemental Material).

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

Next, we sought to determine if melanocyte eQTLs were enriched with common risk variants from the most recent melanoma GWAS meta-analysis (Law et al. 2015). A quantile-quantile plot demonstrated an enrichment of significant GWAS P-values for eQTL SNPs compared to non-eQTL SNPs (Fig. 2A), which was the most pronounced in melanocyte eQTLs (estimated Lambda = 1.51) compared to three related tissue types as well as all the other GTEx tissue types (Supplemental Fig. S5).

To further assess the enrichment of melanoma heritability in melanocyte-specific expressed genes, we performed LD score regression analysis (Finucane et al. 2015). The results indicated that partitioned melanoma heritability was significantly enriched (2.54-fold; P = 2.45 × 10^{-6}) in melanocyte-specific genes (top 4000 genes) compared to all test SNPs (i.e., cis-SNPs ±1 Mb 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 × 10^{-4}, 10,000 permutations; 1.81- to 5.48-fold) (Table 1) and mostly more pronounced than that observed in GTEx tissue categories (Supplemental Figs. S4B; Table 1).

Figure 1. Melanocyte eQTLs display a distinct pattern from those of 44 GTEx tissue types. Dendrogram and heat map presenting the sharing of eQTLs between human primary melanocytes and 44 other GTEx tissue types. Pairwise π1 statistics were calculated from single-tissue eQTL discoveries in each tissue using all the genome-wide significant eQTL SNP-gene pairs. π1 is only calculated when the gene is expressed and testable both in discovery (columns) and replication (rows) tissues. Higher π1 values indicate an increased replication of eQTLs between two tissue types. π1 values range between 0 and 1 and are color-coded from blue (low sharing) to red (high sharing). Tissues are clustered using the Spearman’s correlation of π1 values. Note that π1 values are not symmetrical, since each entry in row i (replication tissue) and column j (discovery tissue) is an estimate of π1 = Pr(eQTL in tissue i given an eQTL in tissue j).

Discovery tissue names are shown in parentheses on the top of the figure.
Zhang et al.

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

<table>
<thead>
<tr>
<th>Epigenetic marka</th>
<th>DHS</th>
<th>H3K27ac</th>
<th>H3K4Me1</th>
<th>H3K4Me3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melanocyte eQTLs</td>
<td>Fold enrichmentb</td>
<td>1.81</td>
<td>5.48</td>
<td>1.99</td>
</tr>
<tr>
<td></td>
<td>$P$-valuec</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>TCGA SKCM eQTLs</td>
<td>Fold enrichmentb</td>
<td>1.72</td>
<td>2.69</td>
<td>1.86</td>
</tr>
<tr>
<td></td>
<td>$P$-valuec</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Skin not sun-exposed eQTLs</td>
<td>Fold enrichmentb</td>
<td>1.78</td>
<td>2.77</td>
<td>1.92</td>
</tr>
<tr>
<td></td>
<td>$P$-valuec</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Skin sun-exposed eQTLs</td>
<td>Fold enrichmentb</td>
<td>1.75</td>
<td>2.66</td>
<td>1.88</td>
</tr>
<tr>
<td></td>
<td>$P$-valuec</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

aDNase I hypersensitivity sites (DHS) and gene regulatory histone marks (H3K27ac, H3K4me1, and H3K4me3) of primary melanocytes (Epigenome Roadmap database; www roadmapepigenomics.org).
bMean 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.
cSee Methods section.

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

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

While the modest size of this data set limits power, we also performed trans-eQTL analyses for the SNPs that are located over 5 Mb away from the TSS of each gene or on a different chromosome. In all, we identified 15 genome-wide significant trans-eQTL genes (excluding genes of mappability < 0.8 or overlapping low complexity regions) (Supplemental Table S9). Of these, eight trans-eQTL SNPs were also cis-eQTLs for local genes within 1 Mb. Notably, rs12203592 (Chr 6: 396321), among these, is a genome-wide significant trans-eQTL SNP for four different genes on four separate chromosomes (TMEM140, IRF3681HG, PLA1A, and NEO1) and is also the strongest cis-eQTL SNP for the IRF4 gene ($P = 7.9 \times 10^{-16}$, slope = −1.14), which encodes the transcription factor interferon regulatory factor 4. All four genes displayed the same direction of allelic gene expression as IRF4 (Tables 1 and 2; Fig. 3). rs12203592 has previously been associated with human pigmentation phenotypes (Han et al. 2008). This variant was also shown to be a functional SNP mediating transcription of IRF4 in melanocytes via C allele-preferential binding of the transcription factor, TFAP2, by collaborating with melanocyte lineage-specific transcription factor, MITF, in turn activating the melanin synthesis enzyme, TYR. The rs12203592-C allele (prevalent in African populations) is correlated with high IRF4 levels in our melanocyte data set, validating the findings observed in a smaller sample set (Praetorius et al. 2013). Expression correlation analyses in melanocytes indicated that expression levels of TMEM140, IRF3681HG, PLA1A, and NEO1 are significantly correlated with those of IRF4 in the same direction as shown by trans-eQTLs (Pearson $r = 0.34$, 0.65, 0.53, and 0.58; $P = 2.67 \times 10^{-9}$, $5.34 \times 10^{-14}$, $4.28 \times 10^{-9}$, and $6.00 \times 10^{-11}$, respectively) (Supplemental Fig. S8). To assess if IRF4 expression levels mediate the observed trans-eQTL effect for these four genes, we performed mediation analyses using regression-based methods (Supplemental Material; Supplemental Table S10), as well as a recently published Genomic Mediation analysis with Adaptive Confounding adjustment (GMAC) (Yang et al. 2017). We applied the GMAC method to 455 eSNP-cis-eGene trans-eQTL trios (trans-eQTL cutoff: $P < 1 \times 10^{-5}$), 84 of which include rs12203592. A total of 121 trios displayed a suggestive mediation ($P < 0.05$), and 32 of them were by IRF4 cis-eQTL including those with TMEM140 and NEO1 (Supplemental Table S11). In contrast, another cis-eQTL gene, RPS14, sharing two SNPs with three trans-eQTL genes (Supplemental Table S9), did not show suggestive mediation (Supplemental Table S11). These results are consistent with IRF4 expression levels mediating at least part of the observed trans-eQTL effect. We then sought to determine if IRF4 is predicted to bind to the genomic regions encompassing the rs12203592 trans-eQTL genes. Sequence motif enrichment analyses indicated that IRF4 binding motifs were enriched in the genomic regions of TMEM140, IRF3681HG, PLA1A, and NEO1 ($\pm 2$ kb of gene boundary; $P = 1.52 \times 10^{-7}$) (Supplemental Table S12), as well as in the above-mentioned 84 trans-eQTL genes ($P = 7.25 \times 10^{-26}$). Together, our data suggest a melanocyte-specific trans-eQTL network potentially regulated by the transcription factor, IRF4.

Melanocyte eQTLs identified candidate melanoma susceptibility genes from GWAS loci

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

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

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

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

eQTL data can be utilized for transcriptome-wide association studies to impute gene expression levels into GWAS data sets. We performed a TWAS (Gusev et al. 2016) using summary statistics from the melanoma GWAS meta-analysis (Law et al. 2015) and the melanocyte eQTL data set as the reference data set (see Methods). Using 3187 eGenes passing a conservative cutoff for heritability estimates ($P < 0.01$) (Supplemental Table S18), the TWAS identified genes at three known melanoma loci at a genome-wide significant level (MAFF on Chr22q13.1, CTSS on Chr1q21.3, CASP8 on Chr2q33-q34), with a fourth locus being suggestive (PARP1 on Chr1q42.1) (Table 3). The TWAS further identified novel associations with melanoma at four genomic loci at a genome-wide significant level (ZFPP90 on Chr16q22.1, HEBP1 on Chr12 p13.1, MSC and RP11-38H13.1 at Chr8q13.3, and CBWD1 at Chr9p24.3) (Table 3; Fig. 5).

We additionally performed a TWAS using each of the 44 GTEx tissue types as reference eQTL data sets. Forty-three GTEx tissue types identified one or more melanoma GWAS loci at a genome-wide significant level with a median of three genes per data set. Tibial nerve tissue identified the largest number of genes (11 genes), while melanocytes ranked third (Supplemental Table S19). Collectively, melanocyte and GTEx data sets identified 22 TWAS genes at six previously known melanoma GWAS loci (Chr1q 21.3, Chr1q42.1, Chr2q33-q34, Chr15q 13.1, Chr21q22.3, and Chr22q13.1) as well as nine TWAS genes at eight novel loci. Melanocyte eQTLs alone identified the majority of novel TWAS genes (five of nine), including the genes unique to the melanocyte data set (Supplemental Table S20). In contrast, none of the 44 GTEx tissue data sets produced more than one novel association for melanoma. Four novel melanoma GWAS loci added from 44 GTEx tissue types are ERCC2 on Chr19q13.32, KIF9 on Chr3p21.31, MRAF2 on Chr6q14.2, and ZBTB4 on Chr17p13.1. Finally, we conducted conditional analyses on the TWAS loci displaying marginally significant associations with multiple genes from melanocyte and GTEx tissue data.
sets. The analyses identified 15 jointly significant genes from 14 loci (Supplemental Table S20; Supplemental Material).

Discussion

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

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

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

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

Figure 3. The pigmentation trait-associated variant, rs12203592, in IRF4 is a trans-eQTL for four genes in melanocytes. Cis- or trans-eQTL P-values and effect sizes (\(\beta\)) are shown between rs12203592 and IRF4 or rs12203592 and four genome-wide significant trans-eQTL genes (TMEM140, MIR3681HG, PLA1A, and NEO1). \(\beta\) values are shown relative to alternative alleles (T). Boxplots display gene expression levels based on rs12203592 genotypes (CC, CT, and TT).
blocks displayed variability in target gene prediction across different data sets as well as between eCAVIAR and TWAS approaches. For the melanoma locus at Chr1q21.3, a total of eight genes were colocalized from three eQTL data sets, and TWAS nominated nine genes. Each top CLPP score gene from melanocyte and skin data sets (ARNT, GERS2, and SETD8BI) were also supported by TWASs in more than one tissue type, and TWAS joint/conditional analyses identified CTSS as explaining the most of the effect in this locus (Supplemental Material; Supplemental Fig. S9). These data imply that multiple statistical approaches using diverse tissue types including the cell type of disease origin is beneficial to robust target gene prediction. Collaboration of single-cell-type and whole tissue eQTL was also exemplified in ASIP for hair color GWAS colocalization (Supplemental Material).

TWASs also identified novel melanoma loci by leveraging tissue-specific eQTL data sets and reducing the multiple testing burden associated with GWASs. While identification of trait-associated gene expression differences via TWASs cannot be taken to imply causality for a specific gene, TWASs may nonetheless nominate plausible candidate risk genes at significant loci. Here, we identified five genes at four new melanoma susceptibility loci (ZFP90 on Chr16q22.1, HEBP1 on Chr12p13.1, MSC and RP11-303H3.1 on Chr9q33.3, and CBWD1 on Chr9p24.3) using melanocyte eQTLs as a reference set and four additional new genes/loci (ERCC2, KIP9, MRAP2, and ZBTB4) using 44 GTEx tissue types.

While most of these genes have known functions that might have relevance in melanomagenesis (Supplemental Material), ERCC2 on Chr1q13.32, among them, is a nucleotide excision repair gene targeting UV-induced DNA damage and implicated in Xeroderma pigmentosum (Taylor et al. 1997). MRAP2 on Chr6q14.2 encodes melanocortin-2-receptor accessory protein 2, which interacts with all melanocortin receptor proteins (MCRs) together with MRP1 to regulate cell surface expression of MCRs (Ramachandrapa et al. 2013). Seemingly relevant functions of these new candidate genes warrant further studies on their roles in melanomagenesis. In all, our primary melanocyte eQTL data set considerably advanced identification of candidate melanoma susceptibility genes from known and new melanoma loci through multiple approaches, which highlights the unique value of cell-type-specific eQTL data sets.

Methods

Melanocyte culture

We obtained frozen aliquots of melanocytes isolated from foreskin of 106 healthy newborn males, mainly of European descent, following an established protocol (Halaban et al. 1000) from the SPORE in Skin Cancer Specimen Resource Core at Yale University. Cells were grown in lot-matched Dermal Cell Basal Medium (ATCC PCS-200-030) supplemented with lot-matched Melanocyte Growth Kit (ATCC PCS-200-041) and 1% amphotericin B/penicillin/streptomycin (120-096-711, Quality Biological) at 37°C with 5% CO2. Every step of cell culture, DNA/RNA isolation, and sequencing/genotyping processes was performed in randomized batches. Before harvesting the cells, media was taken and tested for mycoplasma contamination using MycoAlert PLUS mycoplasma detection kit (LT07-710, Lonza). All 106 samples were negative for mycoplasma contamination.

Genotyping and imputation

Genotyping DNA was isolated from melanocytes following a standard procedure while minimizing melanin carry-over and genotyped on the Illumina OmniExpress arrays (HumanOmniExpress-24-v1-1-a) at the Cancer Genomics Research Laboratory of the Division of Cancer Epidemiology and Genetics (NCI/NIH). Following quality control, genotypes were imputed using Michigan Imputation Server (Das et al. 2016) based on 1000 Genomes (Phase 3, v5) reference panel (The 1000 Genomes Project).
Figure 4. Melanoma GWAS signals colocalizing with melanocyte eQTLs. (A,C,E) LocusZoom plots present the nominal eQTL P-values of all tested local SNPs in 300- to 400-kb windows for three significant eQTL genes from three melanoma GWAS loci: (A) PARP1; (C) MX2; and (E) SLC45A2. The gene being measured is highlighted in pink, the index melanoma risk SNP is labeled and highlighted in purple, and $r^2$ (based on 1000G EUR) of all other SNPs to the index SNP is color-coded. SNPs with missing LD information with the index SNP are shown in gray. Horizontal lines are shown for nominal P-value cutoffs for significant eQTLs. Genomic coordinates are based on hg19. (B,D,F) Boxplots present melanocyte expression differences of each gene in relation to the genotypes of the index SNP. Melanoma risk and protective alleles are shown for each locus.
for population substructure was performed using the eQTL analysis, we included the top three genotyping principal counting for ancestry in the further analyses as covariates. For ensured by ADMIXTURE (Alexander et al. 2009) analysis, while ac-
but histologically carry a range of African and Asian ancestry mea-
sample size, we included all samples that passed genotyping QC
10,718,646 genotypes were further imputed. Due to the small
EIGENSTRAT (Price et al. 2006).

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Chr</th>
<th>GWA best SNP*</th>
<th>GWA Z-score</th>
<th>GWA best SNP*</th>
<th>eQTL best SNP*</th>
<th>eQTL Z-score</th>
<th>GWA best SNP*</th>
<th>eQTL best SNP*</th>
<th>TWAS Z-score</th>
<th>TWAS P-value</th>
<th>FDR passed</th>
<th>GWAS locus</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAFF</td>
<td>22</td>
<td>rs132985</td>
<td>−6.91</td>
<td>rs738322</td>
<td>−5.14</td>
<td>−6.69</td>
<td>340</td>
<td>2</td>
<td>enet</td>
<td>6.67</td>
<td>2.60 × 10−11</td>
<td>Y</td>
</tr>
<tr>
<td>CTSS</td>
<td>1</td>
<td>rs12410869</td>
<td>−7.22</td>
<td>rs7521898</td>
<td>3.33</td>
<td>−6.34</td>
<td>315</td>
<td>6</td>
<td>enet</td>
<td>−6.32</td>
<td>2.65 × 10−10</td>
<td>Y</td>
</tr>
<tr>
<td>ZFP90</td>
<td>16</td>
<td>rs11075688</td>
<td>1.06</td>
<td>rs11075688</td>
<td>3.96</td>
<td>319</td>
<td>319</td>
<td>blup</td>
<td>5.02</td>
<td>1.95 × 10−7</td>
<td>Y</td>
<td>New</td>
</tr>
<tr>
<td>HEBP1</td>
<td>12</td>
<td>rs1684387</td>
<td>−5.32</td>
<td>rs1684387</td>
<td>5.08</td>
<td>605</td>
<td>5</td>
<td>lasso</td>
<td>−5.05</td>
<td>4.48 × 10−7</td>
<td>Y</td>
<td>New</td>
</tr>
<tr>
<td>CASP8</td>
<td>2</td>
<td>rs12936922</td>
<td>−3.71</td>
<td>rs12936922</td>
<td>4.45</td>
<td>372</td>
<td>3</td>
<td>lasso</td>
<td>−4.61</td>
<td>4.06 × 10−6</td>
<td>Y</td>
<td>New</td>
</tr>
<tr>
<td>MSCP</td>
<td>8</td>
<td>rs1481853</td>
<td>−3.83</td>
<td>rs1481853</td>
<td>−6.60</td>
<td>553</td>
<td>3</td>
<td>lasso</td>
<td>4.60</td>
<td>4.27 × 10−6</td>
<td>Y</td>
<td>New</td>
</tr>
<tr>
<td>CWB11D1</td>
<td>9</td>
<td>rs661356</td>
<td>4.99</td>
<td>rs661356</td>
<td>−1.96</td>
<td>462</td>
<td>4</td>
<td>lasso</td>
<td>−4.54</td>
<td>5.52 × 10−6</td>
<td>Y</td>
<td>New</td>
</tr>
<tr>
<td>RP11-383H13.1</td>
<td>8</td>
<td>rs1481853</td>
<td>−2.75</td>
<td>rs1481853</td>
<td>−4.60</td>
<td>693</td>
<td>4</td>
<td>lasso</td>
<td>4.50</td>
<td>6.68 × 10−6</td>
<td>Y</td>
<td>New</td>
</tr>
<tr>
<td>GPRCSA</td>
<td>12</td>
<td>rs1211398</td>
<td>−4.02</td>
<td>rs1211398</td>
<td>5.08</td>
<td>584</td>
<td>11</td>
<td>enet</td>
<td>−4.19</td>
<td>2.80 × 10−5</td>
<td>N</td>
<td>New</td>
</tr>
<tr>
<td>RBBPS</td>
<td>1</td>
<td>rs11240466</td>
<td>4.09</td>
<td>rs10900456</td>
<td>3.98</td>
<td>623</td>
<td>623</td>
<td>blup</td>
<td>3.59</td>
<td>3.37 × 10−3</td>
<td>N</td>
<td>New</td>
</tr>
<tr>
<td>ATP6V1G2-</td>
<td>6</td>
<td>rs2239704</td>
<td>4.35</td>
<td>rs2239704</td>
<td>3.31</td>
<td>354</td>
<td>278</td>
<td>lasso</td>
<td>3.54</td>
<td>3.97 × 10−3</td>
<td>N</td>
<td>New</td>
</tr>
<tr>
<td>DDX39B</td>
<td>16</td>
<td>rs7814977</td>
<td>5.07</td>
<td>rs7814977</td>
<td>4.52</td>
<td>345</td>
<td>345</td>
<td>blup</td>
<td>3.47</td>
<td>5.27 × 10−4</td>
<td>N</td>
<td>New</td>
</tr>
<tr>
<td>CHCHD6</td>
<td>3</td>
<td>rs9851451</td>
<td>3.46</td>
<td>rs9851451</td>
<td>3.44</td>
<td>633</td>
<td>3</td>
<td>lasso</td>
<td>−3.44</td>
<td>5.78 × 10−7</td>
<td>N</td>
<td>New</td>
</tr>
<tr>
<td>CTD-2003C8.2</td>
<td>11</td>
<td>rs1554519</td>
<td>4.51</td>
<td>rs7932891</td>
<td>3.66</td>
<td>746</td>
<td>4</td>
<td>lasso</td>
<td>3.43</td>
<td>6.08 × 10−5</td>
<td>N</td>
<td>New</td>
</tr>
<tr>
<td>UQCC1</td>
<td>20</td>
<td>rs2425025</td>
<td>6.95</td>
<td>rs6060369</td>
<td>−3.14</td>
<td>353</td>
<td>353</td>
<td>blup</td>
<td>3.42</td>
<td>6.21 × 10−3</td>
<td>N</td>
<td>New</td>
</tr>
<tr>
<td>PLXNA1</td>
<td>3</td>
<td>rs9851451</td>
<td>3.46</td>
<td>rs4679317</td>
<td>−2.20</td>
<td>438</td>
<td>2</td>
<td>lasso</td>
<td>−3.42</td>
<td>6.33 × 10−4</td>
<td>N</td>
<td>New</td>
</tr>
<tr>
<td>PARP1</td>
<td>1</td>
<td>rs1865222</td>
<td>−7.12</td>
<td>rs3210900</td>
<td>−1.30</td>
<td>404</td>
<td>404</td>
<td>blup</td>
<td>3.41</td>
<td>6.57 × 10−8</td>
<td>N</td>
<td>1q42.12</td>
</tr>
</tbody>
</table>

*rsID of the most significant melanoma GWAS SNP for the TWAS gene after QC and IMPG imputation by FUSION program.

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

GWAS Z-score of the best eQTL SNP in the locus.

GWAS Z-score of the best eQTL SNP.

GWAS Z-score of the best eQTL SNP.

GWAS Z-score of the best eQTL SNP.

GWAS Z-score of the best eQTL SNP.

Weighted number of SNPs in the locus.

Best performing model.

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

Previously identified GWAS locus (shaded) or newly identified by this study.

 Consortium (2015) and Mixed population, and using SHAPEIT (Delaneau et al. 2011) for prephasing. Post-imputation genetic variants (single nucleotide variants [SNPs] and small insertion-deletion [indel] polymorphisms) with MAF<0.01 or imputation quality scores (R-squared) <0.3 were removed from the final analysis. Overall, ~713,000 genotypes were obtained, and 10,716,646 genotypes were further imputed. Due to the small sample size, we included all samples that passed genotyping QC but histologically carry a range of African and Asian ancestry measured by ADMIXTURE (Alexander et al. 2009) analysis, while accounting for ancestry in the further analyses as covariates. For eQTL analysis, we included the top three genotyping principal components as covariates. The principal components analysis for population substructure was performed using the struct.pca module of GLU (Wolpin et al. 2014), which is similar to EIGENSTRAT (Price et al. 2006).

RNA sequencing and data processing

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

Identification of cis-eQTLs in primary melanocytes

Cis-eQTL analysis was performed closely following a recent standard procedure adopted by GTEx (The GTEx Consortium 2017; see Supplemental Material for details). In brief, cis-eQTL mapping was performed using FastQTL (Ongen et al. 2016), and nominal P-values were generated for genetic variants located within ±1 Mb of the TSSs for each gene tested. The beta distribution-adj STEral empirical P-values from FastQTL were then used to calculate q-values (Storey and Tibshirani 2003), and a false discovery rate (FDR) threshold of ≤0.05 was applied to identify genes with a significant eQTL ("eGenes"). The effect size of the eQTLs was defined as the slope of the linear regression and is computed as the effect of the alternative allele (ALT) relative to the reference allele (REF). For each gene, a nominal P-value threshold was calculated, and variants with a nominal P-value below the gene-level threshold were considered as genome-wide significant cis-eQTL variants. The number of identified eGenes and significant eQTLs were approximately three times higher than those from data analyzed.
and a marginally significant gene, 383H13.1 (10 Genome Research). TWAS using melanocyte eQTL data as a reference set identified five new melanoma-associated genes in four new loci. (A) A similar plot for two new melanoma TWAS genes, MSC (P = 4.27 × 10^{-6}, Z = 4.6) and RP11-383H13.1 (P = 6.68 × 10^{-6}, Z = 4.5) on chromosome band 12p13.1. The Manhattan plot shows the melanoma GWAS P-values before (gray) and after (blue) conditioning on imputed melanocyte-specific gene expression of MSC. (B) A similar plot of new melanoma TWAS gene, CBWD1 (P = 5.52 × 10^{-6}, Z = 4.54) and a marginally significant gene, DOCK8 (P = 2.7 × 10^{-4}, Z = 2.99) on chromosome band 9p24.3 without using PEER factors as covariates (Supplemental Table S2). Application of PEER factors almost doubled the number of eGenes known to be related to pigmentation phenotypes (0.8% vs. 1.5%; Fisher’s exact test P-value = 0.0335). Pairwise eQTL sharing between primary melanocytes and 44 GTEx tissues

To test the sharing of all significant SNP-gene pairs of our melanocyte eQTL study with those identified in 44 tissue types by GTEx (FDR < 0.05) (The GTEx Consortium 2013; The GTEx Consortium 2017), we calculated pairwise π1 statistics (indicating the proportion of true positives) using Storey’s QVALUE software (https://github.com/StoreyLab/qvalue) (Storey and Tibshirani 2003). A heat map was drawn based on the pairwise π1 values, where higher π1 values indicate an increased replication of eQTLs. Tissues are grouped using hierarchical clustering on rows and columns separately with a distance metric of 1 − ρ, where ρ is the Spearman’s correlation of π1 values. π1 is only calculated when the gene is expressed and testable in both the discovery and the replication tissues.

**Identification of trans-eQTLs in primary melanocytes**

Trans-eQTL analysis was performed for SNPs that are located over 5 Mb away from the TSS of each gene or on a different chromosome. Genes of mappability <0.8 or overlapping low complexity regions defined by the RepeatMasker (Smit et al. 2013–2015) library were excluded from the analysis. The nominal P-values for gene-SNP pairs in trans-eQTL analysis were calculated using Matrix-eQTL (Shabalin 2012). We performed a permutation test followed by the Benjamini–Hochberg procedure to identify significant trans-eQTLs following our previous approach (Shi et al. 2014). For each nominal P-value threshold, we calculated the number of genes (denoted as N(1)(p)) that have at least one SNP in its trans region with nominal P-value less than the threshold p. Here, N(1)(p) denotes the number of trans-eQTL genes at P-value threshold p. Next, we performed 100 permutations to estimate the number of genes (denoted as N(0)(p)) without using PEER factors as covariates to have trans-eQTL signals at nominal P-value under the global null hypothesis. By definition, one can calculate the FDR as FDR = N(0)(p)/N(1)(p). We chose p = 3.25 × 10^{-11} to control FDR at a desired level of 0.1.
Identifying cis-mediators for trans-eQTLs in primary melanocytes

We applied the Genomic Mediation analysis with Adaptive Confounding adjustment (Yang et al. 2017) algorithm to identify cis-mediators for trans-eQTLs in primary melanocyte eQTL data. Only the trios with evidence of both cis and trans association were kept. The cis-eQTL SNP with the smallest P-value for each gene (eQTL FDR < 0.05) and trans-association P-value < 10^-4 was selected as one trio. Up to five PEER factors and other covariates (top 10 genotype PCs) were adjusted. One hundred thousand permutations for testing mediation were performed, and trios with suggestive mediation were reported using a mediation P-value threshold < 0.05.

Allele-specific expression

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

Assessing enrichment in putative functional elements

To assess the enrichment of cis-eQTL in putative functional elements of primary melanocytes, we collected the DNase-seq and ChIP-seq data from the Epigenome Roadmap Project (http://www.roadmapepigenomics.org) (Roadmap Epigenomics Consortium et al. 2015). For each putative functional element, we merged peak callings from all samples into one, and all the significant melanocyte eQTL SNP-Gene pairs were used for the enrichment analyses using a similar method to a recent publication (Zhang et al. 2018). Briefly, we performed randomizations for testing whether an eQTL SNP set is enriched for given histone mark regions. Note the following procedure controls for the distribution of minor allele frequencies of a given eQTL SNP set: (1) For K eQTL SNPs, we determined the number (denoted as \( n_0 \)) of eQTL SNPs functionally related with the histone mark. (2) We randomly sampled 10,000 SNP sets. Each SNP set had K SNPs in linkage equilibrium, with minor allele frequency distribution similar to the original K eQTL SNPs. For the \( n^{th} \) sampled SNP set, we calculated the number (denoted as \( n_x \)) of SNPs functionally related with the histone mark. We had \[ \{ x_1, \ldots, x_{10,000} \} \], corresponding to the sampled 10,000 SNP sets. (3) Enrichment fold change was calculated as \( FC = (X_0 / (\sum_{i=1}^{10,000} x_i / 10,000)) \), where the denominator represented the average number of SNPs functionally related with the histone mark under the null hypothesis. The \( P \)-value for enrichment was calculated as \( P = [n_x > n_0] / 10,000 \), i.e., the proportion of SNP set functionally more related with the histone mark than the given eQTL SNP set. If \( x_i < X_0 \) for all sample SNP sets, we reported the \( P \)-value as \( P < 10^{-4} \). In addition, we also assessed enrichment of cis-eQTLs in different genomic regions including 5′/3′ UTR, promoter, exon, intron, and intergenic and lncRNA regions as described in the Rajannotatr package (https://github.com/hhabra/annotatr) (Cavalcante and Sartor 2017).

Enrichment of melanoma GWAS variants in eQTLs

Two methods were used to evaluate if the melanoma GWAS variants were enriched in eQTLs of different data sets. First, QQ plots were used to show the differences in melanoma association \( P \)-values from the most recent meta-analysis (Law et al. 2015) between the significant eQTL SNPs and non-eQTL SNPs. For all GWAS variants, we first performed LD pruning using PLINK (r² > 0.1 and window size 500 kb) (Purcell et al. 2007). If a pruned SNP is an eQTL or is in LD (r² > 0.8) with an eQTL SNP, the SNP is classified as an eQTL SNP. Otherwise, it is classified as a non-eQTL SNP. The lambda values were estimated using the “estlambda2” function in R package “QQperm” (https://github.com/cran/QQperm). For the second method, a simulation procedure was applied to identify overlap and test for enrichment of eQTLs in melanoma GWAS SNPs closely following a previously published method (Hannon et al. 2016; see Supplemental Material for details).

Melanoma heritability enrichment of tissue-specific genes

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

Colocalization analysis of GWAS and eQTL data

We performed colocalization analysis for 20 GWAS loci from the most recent GWAS meta-analysis using Causal Variants Identification in Associated Regions (eCAVIAR, http://genetics.cs.ucla.edu/caviar/index.html) (Hormozdiari et al. 2016). For each locus, both GWAS and eQTL summary statistics (from our melanocyte data set and two GTeX skin tissues) of selected variants in that locus were extracted as the input for eCAVIAR. We selected 50 SNPs both upstream of and downstream from the GWAS lead SNP for each GWAS locus. We computed the CLPP score with a maximum number of two causal SNPs in each locus. We used a CLPP > 1% (0.01) cutoff for colocalization. Thus, for a given GWAS variant (either the lead SNP itself or the SNPs in near perfect
LD with the lead SNP using the cutoff $r^2 > 0.99$), an eGene with a CLPP score above the colocalization cutoff is considered a target gene. We also highlight the eGenes with CLPP > 0.05 as they were more robust across minor changes in analyses criteria compared to those on the borderline (between 0.01 and 0.05) in our analyses.

Performing TWASs with GWAS summary statistics

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

Other analyses

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

Reference genome build statement

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

Data access

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

Members of the Melanoma Meta-Analysis Consortium

Cell-type-specific melanocyte eQTL data set


37Inherited Disease Research Branch, National Human Genome Research Institute, US National Institutes of Health, Baltimore, MD 21224, USA
38Department of Dermatology, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands
39Department of Oncology-Pathology, Karolinska Institutet, Karolinska University Hospital, Stockholm, 171 76 Solna, Sweden
40Department of Dermatology, Oslo University Hospital, Rikshospitalet, 0372 Oslo, Norway
41Department of Surgical Oncology, Institute of Oncology Ljubljana, 1000 Ljubljana, Slovenia
42Department of Surgery, Clinical Sciences, Lund University, P663+Q9 Lund, Sweden
43Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL 33612 USA
44Department of Genetics, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
45Department of Medical Genetics, University of Glasgow, Glasgow G12 9QQ, UK
46McGill University and Génome Québec Innovation Centre, Montreal, QC H3A 0G1, Canada
47Department of Public Health, University of Glasgow, Glasgow G12 9QQ, UK
48Centre for Cancer Research, University of Sydney at Westmead, Millennium Institute for Medical Research and Melanoma Institute Australia, Sydney, NSW 2145, Australia
49Gade Laboratory for Pathology, Department of Clinical Medicine, University of Bergen, 5007 Bergen, Norway
50Molecular Biology, the University of Queensland, Brisbane, QLD 4072, Australia
51Department of Molecular Diagnostics, Institute of Oncology Ljubljana, 1000 Ljubljana, Slovenia
52Department of Oncology/Pathology, Clinical Sciences, Lund University, P663+Q9 Lund, Sweden
53Department of Cancer Epidemiology, Clinical Sciences, Lund University, P663+Q9 Lund, Sweden
54Melanoma Unit, Departments of Dermatology, Biochemistry and Molecular Genetics, Hospital Clinic, Institut d’Investigacions Biomédiques August Pi Sunyer, Universitat de Barcelona, 08007 Barcelona, Spain
55Centro de Investigación Biomédica en Red (CIBER) de Enfermedades Raras, Instituto de Salud Carlos III, Planta 0 28029 Madrid, Spain
56Department of Epidemiology, Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, IN 46202, USA
57Melvin and Bren Simon Cancer Center, Indiana University, Indianapolis, IN 46202, USA
58Department of Dermatology, Warren Alpert Medical School of Brown University, Providence, RI 02903, USA
59Inflammatory Bowel Diseases, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4006, Australia
60Department of Gastroenterology and Hepatology, Royal Brisbane and Women’s Hospital, Brisbane, QLD 4029, Australia
61University of Queensland School of Medicine, Herston Campus, Brisbane, QLD 4072, Australia
62Department of Clinical Genetics, Center of Human and Clinical Genetics, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands
63Cancer Control Group, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4006, Australia
64Department of Ophthalmology, Flinders University, Adelaide, SA 5042, Australia
65Department of Dermatology, University Hospital Essen, 45147 Essen, Germany
66German Consortium for Translational Cancer Research (DKTK), 69120 Heidelberg, Germany
67Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS 7005, Australia
68Department of Cell Biology, National Institutes of Health, Bethesda, MD 20892, USA


Members of the NISC Comparative Sequencing Program

Beatrice B. Bamahas,74 Gerard G. Boulfard,74 Shelis Y. Brooks,74 Holly Coleman,74 Lyudmila Dekhtyar,74 Xiaobin Guan,74 Joel Han,74 Shi-ling Ho,74 Richelle Legasi,74 Quino L. Maduro,74 Catherine A. Masiello,74 Jennifer C. McDowell,75 Casandra Montemayor,74 James C. Mullikin,74 Morgan Park,74 Nancy L. Biebow,74 Karen Schandler,74 Brian Schmidt,74 Christina Sison,74 Raymond Smith,74 Sirintorn Stantonriop,74 James W. Thomas,74 Pamela J. Thomas,74 Meghana Vemulapalli,74 and Alice C. Young74

Acknowledgments

This work has been supported by the Intramural Research Program (IRP) of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health (NIH). The content of this publication does not necessarily reflect the views or policies of the US Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government. M.M.I. was supported by Cancer Research UK (CRUK) CS88/ A19167 and NIH CA083115. This research has been conducted using the UK Biobank Resource under Application Number 3071. This work utilized the computational resources of the NIH high-performance computational capabilities Biowulf cluster (http://hpc.nihs.gov). We acknowledge contributions to human melanocyte genotyping from The National Cancer Institute Cancer Genomics Research Laboratory (CGR) and to RNA sequencing from The NIH Intramural Sequencing Center (NISC), and NCI Center for Cancer Research Sequencing Facility (CCR-SF) and the Yale University Skin SPORe Specimen Resource Core. S.M. is supported by an Australian Research Council Fellowship. We thank A. Vu, L. Mehl, and H. Kong for proofreading the manuscript. The previously published melanoma meta-analysis (Law et al. 2015) makes use of data from two dbGap data sets (accession numbers phs000519.v1.p1 and phs000187.v1.p1). The Leids component of the meta-analysis was funded by the European Commission under the 6th Framework Programme, contract no. LSHC-CT-2006-018702; by Cancer Research UK Programme Awards, C588/A19494 and C588/A10589, and Cancer Research UK Project Grant C8216/A6129; and by US National Institutes of Health R01 ROI CA83115. The Cambridge component of the meta-analysis was supported by Cancer Research UK grants

69Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD 4000, Australia
68Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London SM2 5NG, UK
67Cancer Epidemiology and Services Research, Sydney School of Public Health, University of Sydney, Sydney, NSW 2006, Australia
65Genome Informatics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4006, Australia
66Department of Dermatology, Fachklinik Hornheide, Institute for Tumors of the Skin at the University of Münster, 48149 Münster, Germany
64Department of Gastroenterology and Hepatology, Royal Brisbane and Women’s Hospital, Brisbane, QLD 4029, Australia
63Department of Clinical Genetics, Center of Human and Clinical Genetics, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands
62Cancer Control Group, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4006, Australia
61Department of Ophthalmology, Flinders University, Adelaide, SA 5042, Australia
60Department of Gastroenterology and Hepatology, Royal Brisbane and Women’s Hospital, Brisbane, QLD 4029, Australia
59Department of Dermatology, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands
58Department of Cell Biology, National Institutes of Health, Bethesda, MD 20892, USA
57Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Hanover, NH 03755, USA
56NIH Intramural Sequencing Center, Bethesda, MD 20892, USA
References


Nicola DL, Gamaon 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.


Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet 81*: 559–575.


Received December 7, 2017; accepted in revised form September 21, 2018.
Cell-type–specific eQTL of primary melanocytes facilitates identification of melanoma susceptibility genes

Tongwu Zhang, Jiyeon Choi, Michael A. Kovacs, et al.

Genome Res. published online October 17, 2018
Access the most recent version at doi:10.1101/gr.233304.117

Supplemental Material

http://genome.cshlp.org/content/suppl/2018/10/17/gr.233304.117.DC1

Creative Commons License

This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to Genome Research go to:
http://genome.cshlp.org/subscriptions

© 2018 Zhang et al.; Published by Cold Spring Harbor Laboratory Press