



UNIVERSITY OF LEEDS

This is a repository copy of *Massively parallel reporter assays and variant scoring identified functional variants and target genes for melanoma loci and highlighted cell-type specificity*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/194086/>

Version: Accepted Version

Article:

Long, E, Yin, J, Funderburk, KM et al. (20 more authors) (2022) Massively parallel reporter assays and variant scoring identified functional variants and target genes for melanoma loci and highlighted cell-type specificity. *American Journal of Human Genetics*, 109 (12). pp. 2210-2229. ISSN 0002-9297

<https://doi.org/10.1016/j.ajhg.2022.11.006>

© 2022, Elsevier. This is an author produced version of a article published in *American Journal of Human Genetics (AJHG)*. Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

1 **Massively parallel reporter assays with multi-layer annotation identified functional variants and**
2 **susceptibility genes for multiple melanoma loci and highlighted cell-type-specific variants**

3
4 Erping Long^{1,8}, Jinhu Yin^{1,8}, Karen M. Funderburk¹, Mai Xu¹, James Feng¹, Alexander Kane¹, Tongwu Zhang¹,
5 Timothy Myers¹, Alyxandra Golden¹, Rohit Thakur¹, Hyunkyung Kong¹, Lea Jessop¹, Eun Young Kim², Kristine
6 Jones¹, Raj Chari³, Mitchell J. Machiela¹, Kai Yu¹, Melanoma Meta-Analysis Consortium, Mark M. Iles⁴, Maria
7 Teresa Landi¹, Matthew H. Law^{5,6,7}, Stephen J. Chanock¹, Kevin M. Brown¹, Jiyeon Choi^{1*}

8 ¹Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA, ²Department of
9 Internal Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea, ³Genome Modification
10 Core, Frederick National Lab for Cancer Research, National Cancer Institute, Frederick, MD, USA, ⁴Leeds
11 Institute for Data Analytics, School of Medicine, University of Leeds, Leeds, LS9 7TF, UK, ⁵Statistical Genetics,
12 QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4006, Australia, ⁶Faculty of Health,
13 Queensland University of Technology, Queensland, Australia, ⁷School of Biomedical Sciences, University of
14 Queensland, Queensland, Australia, ⁸These authors contributed equally to this work

15 *Correspondence: jiyeon.choi2@nih.gov

16
17 **Abstract**

18 The most recent genome-wide association study (GWAS) of cutaneous melanoma identified 54 risk-associated
19 loci, but functional variants and their target genes for most have not been established. Here, we performed
20 massively parallel reporter assays (MPRA) using malignant melanoma and normal melanocyte cells and further
21 integrated multi-layer annotation to systematically prioritize functional variants and susceptibility genes from
22 these GWAS loci. Of 1,992 risk-associated variants tested in MPRA, we identified 285 from 42 loci (78% of the
23 known loci) displaying significant allelic transcriptional activities in either cell type ($FDR < 1\%$). We further

24 characterized MPRA-significant variants by motif prediction, epigenomic annotation, and statistical/functional
25 fine-mapping to create integrative variant scores, which prioritized one to six plausible candidate variants per
26 locus for the 42 loci and nominated a single variant for 43% of these loci. Overlaying the MPRA-significant
27 variants with genome-wide significant **expression or methylation quantitative trait loci (eQTL/meQTL)** from
28 melanocytes or melanomas identified **candidate susceptibility genes for 60% of variants (172 of 285 variants)**.
29 CRISPRi of top-scoring variants validated their *cis*-regulatory effect on the eQTL target genes, *MAFF* (22q13.1)
30 and *GPRC5A* (12p13.1). Finally, we identified 36 melanoma-specific and 45 melanocyte-specific MPRA-
31 significant variants, a subset of which are linked to cell-type-specific target genes. Analyses of transcription factor
32 availability in MPRA datasets and variant-transcription factor interaction in eQTL datasets highlighted the roles
33 of transcription factors in cell-type-specific variant functionality. In conclusion, MPRA along with variant scoring
34 effectively prioritized plausible candidates for most melanoma GWAS loci and highlighted cellular contexts
35 where the susceptibility variants are functional.

36

37 **INTRODUCTION**

38 Cutaneous melanoma originates from melanocytes and is the deadliest skin cancer ¹, with increasing
39 incidence and burden worldwide². Melanoma has a substantial heritable germline genetic component explained
40 partly by the 54 genome-wide significant risk loci identified through the most recent genome-wide association
41 study (GWAS)³ including 36,760 cases and 375,188 controls. While a subset of these loci are explained by
42 genetic determinants of pigmentation phenotypes – well-known risk factors of melanoma (e.g., *MC1R*⁴, *OCA2*⁵,
43 *SLC45A2*⁶, and *TYR*⁷), molecular mechanisms of most loci have not been characterized, with a few exceptions
44 (e.g., *PARP1*⁸, *MX2*⁹, *TERT*¹⁰, and *AHR*¹¹). Identifying potentially causal variants and their target genes from
45 melanoma GWAS loci is challenging because there are often many co-inherited variants in strong linkage
46 disequilibrium (LD)¹², and these variants display statistically indistinguishable associations with disease risk
47 given the current sample size. Further, most of these risk-associated variants are in non-protein-coding regions¹³
48 and therefore it is difficult to pinpoint the target genes¹⁴.

49 Most non-coding GWAS variants likely function via *cis*-regulatory mechanisms to regulate target gene
50 expression. Classical reporter assays can test this hypothesis by assessing allelic transcriptional activity of
51 individual variants, and Massively Parallel Reporter Assays (MPRA) allow for scaling the reporter assays to test
52 hundreds to thousands of variants, enabling the identification of functional variants among multiple variants
53 indistinguishable as a result of strong LD. Our previous study⁹ using this approach tested 832 variants from 16
54 melanoma loci based on a previous GWAS¹⁵ and prioritized 39 candidate functional variants from 14 loci in the
55 context of a melanoma cell line. While MPRA can functionally test individual variants in a reporter system, this
56 approach does not identify candidate causal genes. Quantitative trait loci (QTL) analysis is a powerful tool to link
57 GWAS variants to candidate susceptibility genes¹⁶. Our previous studies established multi-QTL datasets using
58 cultured melanocytes as well as skin cutaneous melanomas from The Cancer Genome Atlas (TCGA)^{17,18} and
59 demonstrated that a multi-QTL approach of disease-relevant cell/tissue datasets could nominate candidate
60 susceptibility genes for most melanoma GWAS loci. The strategy of combining both MPRA and cell-type
61 specific eQTLs identified the most prominent locus to follow up and led to a discovery of *MX2* as a pleiotropic
62 gene promoting melanoma in a zebrafish model⁹.

63 Despite this progress, a comprehensive understanding of the role of GWAS-identified loci in melanoma
64 susceptibility is still lacking. A more recent melanoma GWAS meta-analysis identified a total of 54 loci reaching
65 genome-wide significance³, increasing the total number of melanoma risk-associated loci by more than three-fold,
66 most of which have not been functionally tested. Moreover, beyond our work focusing on a handful of loci with
67 one prominent candidate variant, most loci tend to have multiple functional variants displaying allelic
68 transcriptional activity, and systematic prioritization schemes are needed to guide further time-consuming
69 functional studies on these more challenging loci. Furthermore, there is growing evidence that the *cis*-regulation
70 of gene expression underlying complex trait susceptibility is cell-type and context specific^{19,20}. Indeed, our
71 previous studies using LD score regression³ and colocalization/TWAS approaches¹⁷ demonstrated that using data
72 from primary human melanocytes, the cell of melanoma origin²¹, is more useful for annotating melanoma GWAS
73 data than any tissue type from GTEx dataset including skin. Still, it is often not clear in studying individual cancer
74 susceptibility variants and genes whether their tumor-promoting potential is more pronounced in the context of

75 early stages (i.e., normal cells) or later stages (i.e., cancer cells) along the evolutionary trajectory of
76 tumorigenesis. Critically, substantial heterogeneity of QTLs between melanocytes and melanomas has been
77 observed in our previous studies¹⁸, highlighting the importance of studying the gene expression regulation in the
78 contexts of both normal and cancer cells. While there have been many approaches and datasets that prioritized
79 functional variants from GWAS loci and linked them to target genes²²⁻²⁴, the relative roles of different trait-
80 relevant cell-types in variant functionality have not been systematically compared and incorporated to prioritizing
81 variants, especially for melanoma and other cancer GWAS.

82 To address these issues and functionally characterize all 54 reported melanoma GWAS loci, we
83 performed MPRA in both malignant melanoma and normal melanocyte cell lines. Multilayered variant functional
84 features, including motif prediction, epigenomic annotation, and statistical/functional fine-mapping were
85 integrated with MPRA data to further prioritize the plausible candidate causal variants by locus. To link
86 functional variants to potential susceptibility genes, expression QTLs (eQTLs) and DNA methylation QTLs
87 (meQTLs) from melanocytes and melanoma were incorporated. Leveraging these approaches, we prioritized
88 plausible candidates from GWAS loci and highlighted significant cell-type specificity of melanoma susceptibility
89 in relevant tumor and normal cell types.

90

91 **MATERIALS AND METHODS**

92 **MPRA variant selection.** For MPRA, we selected candidate variants from each of the 54 genome-wide
93 significant loci (**Table S1**) from the melanoma GWAS meta-analyses by Landi and colleagues³ that meet one of
94 the following three criteria:

- 95 1. Variants with log likelihood ratio (LLR) < 1:1000 relative to the primary lead SNP based on the GWAS *P*
96 values (fixed effect model) from the main meta-analysis; for the locus tagged by rs4731207, LLR < 1:150
97 was applied to test only the strongest candidate variants within an extended/large LD block (~600 variants
98 with LLR < 1:1000)
- 99 2. LD $R^2 > 0.8$ (1000 Genomes, phase 3, EUR populations) with the primary lead SNP for any variant not
100 genotyped or successfully imputed in the GWAS (*P* values not available)

101 3. LD $R^2 > 0.8$ (1000 Genomes, phase 3, EUR populations) with an additional independent lead SNP(s)
102 identified through a conditional analysis³ within 1Mb of a primary lead SNP (regardless of LLR); two
103 additional lead SNPs (rs3212371 and rs73069846) reported in the melanoma GWAS³ were not included in
104 the design.

105 After considering these criteria, 214 variants were dropped due to technical reasons including those that have
106 enzyme digestion sites for either KpnI, XbaI or SfiI within the 145bp encompassing the variant. A total of 1,992
107 melanoma GWAS variants were tested by MPRA. A complete list of variants tested are shown in **Table S2**.

108

109 **MPRA oligo library design.** The oligo library was designed in a similar way to our previous work⁹ with some
110 modifications. For each variant, 145-base sequences encompassing the variant (± 72 bases) with **reference and**
111 **alternative alleles** in both forward and reverse directions were extracted from human genome build GRCh37.
112 **Strand (forward/reverse) was tested in assessing enhancer function of a sequence element, which models the**
113 **relative position of enhancer element to gene promoter.** Each test sequence was randomly associated with 20
114 different randomly generated 12-base sequence tags separated by recognition sequences for restriction enzymes,
115 KpnI (GGTACC) and XbaI (TCTAGA), and flanked by binding sequences for PCR primers and a two-base
116 spacer (204 bases oligo sequences: 5'-ACTGGCCGCTTCACTG-145 bases-GGTACCTCTAGA-12 bases tag-
117 AC (spacer)-AGATCGGAAGAGCGTCG-3'; based on ~maximum length allowed in oligo design). For each
118 variant, a single scrambled sequence of 145-base test sequence was also included and associated with 16 tag
119 sequences (using forward direction and the reference allele) as a background level control for activator inference
120 (see **Transcriptional activator inference** section). **The number of tags is based on down-sampling analysis from**
121 **a previous study²⁵, which suggested that 16 or higher barcode per sequence showed sufficient inter-transfection**
122 **correlation.** When there are additional SNPs other than the test SNP that fall in the 145 bp region, the major allele
123 in the 1000 Genomes EUR populations was used **for both sequences of reference/alternative alleles, ensuring all**
124 **the sequences are fixed except the tested variant.** For indel variants, a 145 total base length was set based on
125 insertion allele, and additional bases were added to each side of the test sequence of the deletion allele to fit 145
126 bases. For the 12-base tag sequence and scrambled sequences, only homopolymers of < 4 bases were used and the

127 enzyme recognition sites for KpnI, XbaI, and SfiI were avoided. A pooled library of 201,792 oligos in a
128 randomized order was synthesized by Agilent Technologies (Santa Clara, CA). A complete list of oligo sequences
129 can be found in **Data S1**.

130

131 **MPRA library construction, transfection, and sequencing.** MPRA library construction, transfection, and
132 sequencing was performed following published procedures^{9,26} with some modifications. For library cloning, ten
133 femtomoles of gel-purified (10% TBE-Urea polyacrylamide gel) oligo library was amplified by emulsion PCR
134 using 1.5 µl of Herculase II fusion polymerase (Agilent, Santa Clara, CA), 0.5 mg/ml BSA acetylated, 375 µM
135 dNTP, and 3 µM of primers providing SfiI enzyme sites and 25 cycles amplification per 50 µl reaction, then 3X
136 50 µl reactions were combined and cleaned up in column purification step, following the instructions of the
137 Micellula DNA Emulsion & Purification Kit (EURx/CHIMERx, Milwaukee, WI). To verify the oligo sequences,
138 amplicon libraries were prepared using 100 ng of oligos from emulsion PCR using KAPA Hyper Prep Kit (KAPA
139 Biosystems, Wilmington, MA) following the instructions of the manufacturer, and were sequenced with the
140 MiSeq reagent kit v3 (150-cycles). Twelve-base tag sequences plus spacer sequences were used to map each oligo
141 from FASTQ file and count the total read depth. 98% of the designed tag sequences were detected at one or more
142 reads. Sequence-verified oligo library was first cloned into pMPRA1 vector (Addgene, Watertown, MA) using
143 SfiI site followed by electroporation into 10 times higher number of bacterial cells than the number of unique
144 sequences in the oligo library. Cloned pMPRA1 was further digested on KpnI and XbaI sites between the 145 bp
145 test sequence and the 12 bp barcode sequence, where luc2 ORF with a minimal promoter from pMPRAdonor2
146 (Addgene, Watertown, MA) was inserted. The ligation product was transformed by electroporation into 10 times
147 higher number of bacterial cells in the same manner. Cloned final library for transfection was verified on the gel
148 as a single band after KpnI digestion.

149 Three batches of cloned library were used to transfect 8 times into UACC903 melanoma cells and 5 times
150 into an immortalized primary melanocyte cell line (C283T)¹¹, aiming for >100 times higher number of transfected
151 cells than the library complexity in each transfection. The numbers of transfected cells were estimated using
152 transfection efficiency measured by a separate GFP transfection and visualization. For UACC903, cells were

153 transfected using Lipofectamine 3000 and harvested 48 hours after transfection for RNA isolation. For C283T,
154 cells were transfected by electroporation using P2 Primary Cell 4D-Nucleofector X Kit L (Lonza, Basel,
155 Switzerland), following manufacturer's instruction. Nucleofector programs for C283T cell lines were optimized
156 with the P2 Primary Cell 4D-Nucleofector X Kit and GFP visualization. The amount of cloned MPRA library and
157 harvesting time of transfection cells were optimized using qPCR with specific primers (**Data S1**). Electro-
158 transfected C283T cells were harvested at 24 hours for RNA isolation. Total RNA was isolated using Qiagen
159 RNeasy kit (Qiagen, Hilden, Germany), and mRNA was subsequently isolated using PolyA purist MAG kit
160 (Thermo Fisher). cDNA was then synthesized using Superscript III reverse transcriptase, from which short
161 sequences encompassing 12 bp unique tags were amplified using Q5 high-fidelity polymerase (NEB, Ipswich,
162 MA) and primers introducing Illumina TruSeq adapter sequences. Tag sequence libraries were also prepared
163 using input DNA in the same way. Tag sequence libraries were sequenced on NovaSeq 6000 SP flow cells (100
164 bp dual-indexed single end read) to obtain 125-200 million reads per sample for UACC903 transfections and 218-
165 295 million reads per sample for C283T transfections.

166

167 **MPRA data analyses.** Using FASTQ files from input DNA or RNA transcript (cDNA) sequencing, we counted
168 the number of reads (Illumina read 1) completely matching 12 bp barcode sequences (tag counts) plus spacer
169 sequences and the same downstream sequence context including an XbaI recognition site and the 3' of the luc2
170 gene. For each transfection, Tag counts Per Million sequencing reads (TPM) values were calculated by dividing
171 each tag count by the total number of sequence-matching tag counts divided by a million. A pseudo count of 1
172 was added to all TPM values and then TPM ratio was taken as RNA TPM over input DNA TPM and \log_2
173 converted: $\log_2 ((\text{RNA TPM} + 1)/(\text{DNA TPM} + 1))$. We defined this \log_2 transformed TPM ratio as "normalized
174 expression level".

175 From each input DNA library 93.4-94.9% of designed barcode sequences were detected. From RNA
176 samples 90.2-92.8% of barcode sequences were detected in melanoma cells, and 93.7-94.3% in melanocytes
177 (**Table S3**). Median tag counts were 723-810 for DNA input, 412-655 for RNA output from melanoma cells, and
178 754-1018 for RNA output from melanocytes (**Table S3**). In melanoma cells, 96.2-98.6% of unique tags detected

179 in DNA input were recovered in mRNA output, and 98.0-98.9% were recovered in melanocytes (**Table S3**).
180 Reproducibility between transfections were assessed by Pearson correlation of normalized expression level of
181 each barcode between replicates of transfection. To avoid low input DNA counts driving variations in RNA/DNA
182 TPM ratios, we removed tags with < 2 TPM counts (\log_2 DNA TPM < 1) from further analyses. The remaining
183 tags account for 82.99% of all the detected tags (**Fig. S1**).

184 We used the following standard linear regression model to assess the impact of Allele (reference or
185 alternative) on the transcriptional activity (normalized expression level defined as \log_2 ((RNA TPM + 1)/(DNA
186 TPM + 1)), named “Ratio” in following formulas), while adjusting for the effect of Strand (forward or reverse) as
187 a binary covariate and the effect of Transfection replicate as a categorical covariate:

$$188 \quad \text{Ratio} = \text{Allele} + \text{Strand} + \text{Transfection}$$

189 To account for the potential heteroskedasticity in the measurement error, we used the robust sandwich
190 type variance estimate in the Wald test to determine the significance. This analysis was carried out with the R
191 package Sandwich (<https://sandwich.r-forge.r-project.org>). The Wald test P values were corrected for multiple
192 testing using the procedure of Benjamini and Hochberg²⁷. A corrected P (FDR) < 0.01 was used to define
193 “MPRA-significant variants” that display significant allelic transcriptional activity in each cell type.

194

195 **Transcriptional activator inference.** Given that the variants were selected and tested in MPRA regardless of
196 their functional annotation, we assumed that most of the tested sequences are non-functional and therefore the
197 mean normalized expression levels of all the variants were considered as null. Therefore, the putative function of
198 activators (or repressors) was inferred by defining the extreme outliers from the mean expression levels. First, the
199 overall distribution of the normalized expression levels (mean \log_2 ((RNA TPM + 1)/(DNA TPM + 1))) of all tags
200 by variants including reference and alternative alleles and scrambled sequences were calculated. The putative
201 function of activators was inferred by defining the extreme outliers from the mean expression levels (upper limit:
202 $Q3 + 3 \times IQ$), where $Q3$ is 75th percentiles and the interquartile range (IQ) is $Q3 - Q1$ (25th percentiles). Similarly,
203 putative repressor function was inferred based on extreme lower limits ($Q1 - 3 \times IQ$). For each variant,
204 allele/strand-specific normalized expression levels were then calculated using only the tags for reference-forward,

205 reference-reverse, alternative-forward, or alternative-reverse sub-group. The variants with normalized expression
206 levels of one or more of these sub-groups higher than the upper limits were assigned as activators and vice versa
207 for repressors. These assignments were confirmed by the regression analyses comparing normalized expression
208 levels of scrambled sequences with either reference or alternative allele for each strand separately, while still
209 using Transfection as a covariate. The Wald test with robust sandwich type variance estimate were used, and *FDR*
210 < 0.01 was applied.

211

212 **MTSA analyses.** MPRA tag sequence analysis (MTSA) is a sequence-based analysis for estimating tag sequence
213 effects on gene expression in an MPRA experiment using the following steps²⁸. First, tags with low read counts in
214 input DNA (< 200 reads) were removed for the purpose of MTSA analysis. Second, the relative expression (tag
215 expression normalized to mean zero across each set of tags associated with a 145bp sequence) was calculated.
216 Third, a support vector regression (SVR) was trained based on gapped-kmer kernels²⁹ to learn the contribution of
217 each tag sequence to its relative expression level. Fourth, the adjusted expression values (RNA tag counts) were
218 calculated. Finally, the MTSA-corrected *FDR* and \log_2FC (\log_2 -transformed fold difference of mean TPM ratio
219 for alternative allele over mean TPM ratio for reference allele) were the outputs. MTSA-corrected *FDR* is
220 calculated using the approach of linear regression with the robust sandwich type variance estimate in Wald test
221 (see section of **MPRA data analyses**). The MTSA-corrected *FDR*s are compared with original *FDR*s regarding
222 the significance of allelic transcriptional activity ($FDR < 0.01$) and allelic direction.

223

224 **Functional annotations.** The melanocyte open chromatin regions were inferred by the human melanocyte DHS
225 peaks from ENCODE³⁰ ($n = 1$), Epigenome Roadmap database ($n = 2$)³¹, and melanocyte ATAC-seq peaks
226 combined from the cultured melanocytes of 6 individuals that were generated in our laboratory¹¹. The melanoma
227 open chromatin regions were inferred by human melanoma short-term culture FAIRE-Seq peaks from one or
228 more individuals of 11 available from Verfaillie et al³². The enhancer regions were marked if the variant is located
229 within both a human melanocyte H3K27Ac ChIP-Seq peak and a H3K4Me1 ChIP-Seq peak from at least one
230 individual ($n = 2$ available through Epigenome Roadmap database). The promoter regions were marked if the

231 variant is located within both a human melanocyte H3K27Ac ChIP-Seq peak and a H3K4Me3 ChIP-Seq peak
232 from at least one individual (n = 2 available through Epigenome Roadmap database).

233

234 **Motif analysis.** Prediction of variant effects on transcription factor (TF) binding sites was performed using the
235 motifbreakR package and a comprehensive collection of human TF binding sites models (HOCOMOCO, v11).
236 We selected the information content algorithm and used a threshold of 10^{-4} as the maximum P-value for a
237 transcription binding site match in motifbreakR. The strong effect is defined by the difference between alternative
238 allele score and reference allele score larger than 0.7.

239

240 **Melanoma GWAS statistical and functional fine-mapping.** Statistical fine-mapping of the 54 genome-wide
241 significant loci from the meta-analysis reported by Landi and colleagues was conducted using FINEMAP v1.4³³.
242 We defined flanking regions as 250 kilobase (kb) on either side of the most significant variant at each locus.
243 Evidence (Z-score) for each variant from the GWAS summary statistics and LD matrix (pre-computed using n =
244 ~337,000 unrelated British-ancestry individuals from the UK Biobank³⁴) were the input for the analysis. For loci
245 with one independent signal identified by the conditional analysis in the original GWAS³, we set the maximum
246 number of causal variants as 2. For loci with multiple conditionally independent signals, we set the maximum
247 number of causal variants equal to the number of independent signals from the GWAS. For an improved fine-
248 mapping efficiency, we also performed a fine-mapping incorporating functional annotation using POLYFUN³⁴ by
249 specifying prior probabilities for FINEMAP analysis. Following the recommended procedure, we incorporated
250 precomputed prior causal probabilities of ~19 million imputed UK Biobank SNPs with MAF > 0.1%, based on a
251 meta-analysis of 15 UK Biobank traits including hair color. The output includes posterior inclusion probability
252 (PIP) for each variant and the index of the credible set that the variant belongs to. A 95% credible set is comprised
253 of variants that cumulatively reach a probability of 95%. The variants with PIP > 0.1% were considered being in
254 the 95% credible set.

255

256 **Integration of MPRA variants with melanoma and melanocyte eQTL and meQTL variants.** Significant
257 eQTLs or meQTLs were defined using the empirical genome-wide significance threshold as described in the
258 previous studies^{17,18}. MPRA-significant variants were linked to target genes if they display significant eQTL or
259 meQTL *P*-values for one of the significant genes (eGenes) or CpG sites (meProbes) in melanocytes or
260 melanomas. Gene assignments to each meProbe are presented based on the Illumina HumanMethylation450
261 BeadChip annotation file, which we define as meGenes. Identification of eQTLs and meQTLs as well as
262 colocalization analyses were previously described^{17,18}. Briefly, melanocyte eQTLs and meQTLs were obtained
263 from a dataset including 106 individuals mainly of European descent. Melanoma eQTLs and meQTLs were based
264 on our previous analyses of 444 skin cutaneous melanoma (SKCM) samples from TCGA with genotype,
265 expression, and methylation data. The colocalization analysis was performed among melanoma GWAS, eQTL,
266 and meQTL datasets using HyPrColoc with detailed parameters described in our previous study¹⁸.

267

268 **Variant prioritization scores.** We established a system to prioritize variants in each locus by assigning an
269 integrative score to each variant based on multi-layer information. Each variant was first assigned scores in the
270 categories listed below (score 0 for no hit, score 1 for a hit, or score 2 for a strong hit), and scores for all the
271 categories were added up to an integrative score. For each locus, the variant(s) with the highest integrative score
272 were assigned as tier-1 variants. Those with the second-highest scores (no less than 70% of the highest score)
273 were assigned as tier-2 variants and the rest as tier-3 variants.

274 1. MPRA scores:

- 275 • Variants displaying significant allelic transcriptional activity ($FDR < 0.01$) in melanoma cells
276 were considered as a hit and those with strong significance $FDR < 10^{-9}$ as a strong hit.
- 277 • Variants displaying significant allelic transcriptional activity ($FDR < 0.01$) in melanocytes were
278 considered as a hit and those with strong significance $FDR < 10^{-9}$ as a strong hit.
- 279 • An assignment as a transcriptional activator function in either melanoma cells or melanocytes
280 was considered as a hit (see Transcriptional activator inference section).

281 2. Chromatin annotation scores:

- 282 • Overlap with an accessible chromatin region (genomic regions defined as peaks from ATAC-seq,
283 DHS-seq, or FAIRE-seq data) reported in at least one dataset was considered as a hit and if in
284 more than one dataset (4 datasets in total; melanocytes in ENCODE and Epigenome Roadmap
285 datasets, melanocytes from in-house data, and melanoma cultures from Verfaillie *et al.*³²) as a
286 strong hit.
- 287 • Overlap with human melanocyte histone modifications consistent with enhancer (marked by both
288 H3K27Ac ChIP-Seq peak and H3K4Me1 ChIP-Seq peak) or promoter region (marked by both
289 H3K27Ac ChIP-Seq peak and H3K4Me3 ChIP-Seq peak) from Epigenome Roadmap database
290 was considered as a hit and overlap with both enhancer and promoter regions as a strong hit.
- 291 3. Fine-mapping scores:
- 292 • Variant included in the 95% credible sets from FINEMAP analyses was considered as a hit and
293 PIP > 0.5 as a strong hit.
- 294 • Variants included in the 95% credible sets from POLYFUN analyses was considered as a hit and
295 PIP > 0.5 as a strong hit.
- 296 4. TF-binding motif scores:
- 297 • Variant displaying a significant match with a TF-binding motif ($P < 10^{-4}$) predicted by
298 motifbreakR analysis was considered as a hit and those displaying strong effects (allelic
299 differences of binding scores > 0.7) as a strong hit.

300

301 **Differentially expressed genes between melanoma and melanocytes.** We profiled differentially expressed
302 genes (DEGs) from RNA-seq data generated for the same melanoma cells (UACC903, n=3) and immortalized
303 primary melanocytes (C283T, n = 3) used for MPRA. Total counts of mappable reads for each annotated gene
304 (hg38) were obtained using featureCounts from the Rsubread package³⁵. The DESeq2 software³⁶ was applied to
305 perform quality control, determine differential expression based on a negative binomial model using count data
306 from both melanoma and melanocytes groups. The Wald test *P* values were corrected for multiple testing using

307 the procedure of Benjamini and Hochberg²⁷. A total of 4,388 DEGs were determined with corrected P (FDR) <
308 0.01 and $|\log_2 \text{fold change}| > 2$.

309

310 **Identification of cell-type specific variants.** To nominate cell-type specific variants we applied the following
311 three criteria and assigned scores for each criterion (score 0 for no hit, score 1 for a hit).

312 1. MPRA allelic effect is exclusively observed in one cell type

- 313 • MPRA allelic effect $FDR < 10^{-9}$ (extreme significance) in one cell type
- 314 • AND MPRA allelic effect $FDR > 0.01$ (non-significance) in the other cell type

315 2. 145bp sequence harboring the variant is an activator in the same cell type where the significant allelic
316 effect is observed

- 317 • MPRA allelic effect $FDR < 0.01$ in one cell type
- 318 • AND 145bp sequence is an activator in the same cell type (see **Transcriptional activator**
319 **inference** section)
- 320 • AND MPRA allelic effect $FDR > 0.01$ in the other cell type

321 3. Predicted TFs binding to the variant display significantly higher abundance in the same cell type where
322 the significant allelic effect is observed

- 323 • MPRA allelic effect $FDR < 0.01$ in one cell type
- 324 • AND the levels of predicted TFs are significantly higher in the same cell type (see DEGs defined
325 in **Differentially expressed genes between melanoma and melanocytes** section)
- 326 • AND MPRA allelic effect $FDR > 0.01$ in the other cell type

327

328 **Cell-type regression analyses.** To directly compare the allelic transcriptional activity of variants between
329 melanoma and melanocyte, we applied a standard linear regression to encode the interaction term between the
330 Cell_Type and Allele, after adjusting the effect of Strand and Transfection:

331 $\text{Ratio} = \text{Allele} + \text{Cell_Type} + \text{Allele} * \text{Cell_Type} + \text{Strand} + \text{Transfection}$

332 The Wald test with robust sandwich type variance estimate on the interaction term was used to determine the
333 significance, which was corrected for multiple testing. The cutoff of corrected cell-type $FDR < 0.01$ was applied.

334

335 **VARIANT-TRANSCRIPTION FACTOR-GENE INTERACTION ANALYSES.** Melanoma or melanocyte-specific candidate Variant-
336 Transcription Factor (TF)-Gene trios were established separately when variants are 1) significant in the MPRA of
337 the corresponding cell-type, 2) predicted to significantly change TF binding by motifbreakR, and 3) linked with
338 genome-wide significant eQTL genes in the corresponding dataset. We identified 38 trios for melanoma and 119
339 trios for melanocyte datasets. For each trio, a Multiple Linear Regression with Interaction model was used for the
340 expression levels of eGene and TF (RNA-Seq by Expectation Maximization or RSEM³⁷), and variant genotype
341 (alternative allele count) (eGene ~ SNP + TF + SNP × TF). A Benjamini-Hochberg²⁷ correction was applied to the
342 corrected P -value (FDR) across each Variant-eGene pair (for testing multiple TFs). The trios with FDR value <
343 5% in SNP × TF are considered as displaying significant variant-TF-eGene interaction.

344

345 **CRISPRi EXPERIMENTS.** CRISPR interference (CRISPRi) was performed in the UACC903 melanoma cell line.
346 Three different guide RNAs (gRNAs) for each variant were designed to target the genomic regions surrounding
347 three tested variants (rs61935859, rs4384, and rs2111398), and the sequences of gRNAs are listed in **Table S4A**
348 Non-targeting gRNA and gRNA targeting the adeno-associated virus site 1 (AAVS1) were used as controls.
349 gRNAs were ligated into the lentiviral vector pRC0608-U6-SpCas9-XPR050-puro-2A-GFP (made by Dr. Raj
350 Chari at Genome Modification Core in the Frederick National Laboratory for Cancer Research). For the
351 generation of lentiviral particles, plasmids encoding gRNA or dCas9-ZIM3 (pRC0528_Lenti-dCas9-ZIM3-Blast
352 from Dr. Raj Chari) were co-transfected into HEK293T cells with psPAX2, pMD2-G and pCAG4-RTR2
353 packaging vectors. Virus particles were collected 2 days after transfection, and titer was measured by Lenti-X
354 GoStix Plus (Takara, CA). UACC903 melanoma cells were infected with dCas9-ZIM3 lentivirus and selected by
355 10 μ g/ml blasticidin to generate UACC903-dCas9-ZIM3 polyclonal stable cell line. UACC903-dCas9-ZIM3 cells
356 were infected with lentivirus harboring gRNA. Twenty-four hours after infection, 2 μ g/ml of puromycin was
357 applied for selection. Surviving cells were harvested 48 h after puromycin selection for RNA and protein

358 isolation. The experiments were performed in at least 3 biological replicates in sets of 6 replicates. Total RNA
359 was isolated using an RNeasy Kit (Qiagen). For optimal synthesis of the relatively large full-length cDNA of
360 *MED13L* (3.2kb), SuperScript III First-Strand Synthesis kit (Thermofisher) was used. The cDNA of *MAFF* and
361 *GPRC5A/HEBP1/EMP1* was generated using High-Capacity cDNA Reverse Transcription kit (Thermofisher).
362 mRNA levels of each gene were measured with a Taqman probe set (**Table S4B**) and normalized to *GAPDH*
363 levels. qPCR triplicates (technical replicates) were averaged to be considered as one data point. Proteins were
364 separated on NuPAGE 3–8% Tris-Acetate Protein Gels (Thermo Fisher) and detected by mouse anti-Cas9 (7A9–
365 3A3, Active Motif) and mouse anti-GAPDH (sc-47724, Santa Cruz) primary antibodies.

366

367 **Statistical analyses.** Cell-based experiments were repeated at least three times with separate cell cultures, and
368 mean values of all the biological replicates are presented. For all plots, individual data points are shown with the
369 median or mean, range (maximum and minimum), and 25th and 75th percentiles (where applicable). The statistical
370 method, number of data points, and number and type of replicates are indicated in each figure legend.

371

372 **RESULTS**

373 **MPRA identified functional variants in 42 melanoma GWAS loci.** We performed MPRA to simultaneously
374 identify functional *cis*-regulatory variants for multiple melanoma GWAS loci. We tested 1,992 variants (median
375 26.5 variants per locus) from 54 genome-wide significant loci (including 11 additional independent signals) based
376 on the recent melanoma GWAS meta-analysis³ (**Table S1**). To select these variants, we primarily considered
377 GWAS statistics (log likelihood ratio < 1:1000 with the primary lead SNPs) and further used linkage
378 disequilibrium (LD) for the variants that are not present in the imputation reference set or poorly imputed in the
379 GWAS data and for the secondary signals ($R^2 > 0.8$ with the lead SNP) (**Fig. 1A; Table S2; Materials and**
380 **Methods**). We assessed 145-bp genomic sequences encompassing the **reference and alternative** alleles of each
381 variant for their potential as a transcriptional enhancer in luciferase constructs in both forward and reverse
382 directions with 20 unique barcodes associated with each tested sequence. A scrambled sequence of the same 145-
383 bp associated with 16 barcodes was also tested as a null for each variant (**Fig. 1B; Materials and Methods**). To

384 test variant function in the cellular contexts representing both tumor and normal states we transfected the MPRA
385 library into a melanoma cell line (UACC903, n=8 transfections) and an immortalized primary melanocyte cell
386 line (C283T, n=5). Each barcode sequence detected in the input DNA or mRNA (cDNA) after transfections was
387 counted by sequencing. Initial quality assessment showed a good correlation of normalized tag counts among
388 transfection replicates by tags (median Pearson $R = 0.553$ and 0.745 for melanoma and melanocyte, respectively;
389 **Fig. S2&S3**) and by variants (median Pearson $R = 0.938$ and 0.947 for melanoma and melanocyte, respectively;
390 **Fig. S4&S5**). High recovery rates of designed tags were observed in the transcribed output (90.2-92.8% for
391 melanoma and 93.7-94.3% for melanocyte; **Table S3**). Details of quality control measure for downstream
392 analyses are shown in **Table S3**.

393 We first focused on the variants displaying allelic transcriptional activity in each cell type, identifying 134
394 (7% of tested variants) in UACC903 melanoma (**Fig. 1C**; **Table S5**) and 208 (10% of tested variants) in C283T
395 melanocyte cell lines (**Fig. 1D**; **Table S6**) that pass an $FDR < 0.01$ cutoff (two-sided Wald test with robust
396 sandwich type variance estimate; multiple testing correction by Benjamini & Hochberg²⁷ method; **Materials and**
397 **Methods**). We defined these 285 unique variants ($FDR < 0.01$ in either cell line; 14% of tested variants) as
398 “MPRA-significant variants”. Seventy-eight percent of the melanoma GWAS loci (42 of 54 loci) displayed at
399 least one MPRA-significant variant. For 83% of these loci (35 of 42 loci) MPRA-significant variants were
400 identified from both cell types, while the rest were from only one cell type (3 loci in melanoma and 4 loci in
401 melanocyte). For 8 loci, a single MPRA-significant variant was identified, while 2-36 MPRA-significant variants
402 were identified for 34 loci (**Fig. 1C-D**).

403 We further inferred a putative enhancer (transcriptional activator) function of the 145-bp around MPRA-
404 significant variants by applying two criteria: 1) the sequence containing either allele displays an extreme outlier
405 expression level (three-time interquartile range above 75th percentiles, **Materials and Methods**) compared to the
406 mean expression level distribution of all the tested tags (assuming that most of the tested variants do not display
407 transcriptional activity), and 2) the same sequence also shows a significantly higher expression level than the
408 matched scrambled sequence ($FDR < 0.01$) (**Materials and Methods**). Among these 285 variants, 57 variants
409 were assigned as activators in the melanoma cells (**Fig. S6A**), 28 variants in melanocytes (**Fig. S6B**), and 15 in

410 both cell lines. Only one variant (rs2911405) was identified as a repressor in melanocytes, which displayed
411 significantly lower expression level than the mean value as well as that of scrambled sequence.

412 Notably, our MPRA design included 206 variants that have been tested in the same UACC903 cell line
413 from our previous study, and 93.2% of them (192 of 206) displayed consistent results between two studies
414 regarding the significance of allelic transcriptional activity ($FDR < 0.01$) and allelic direction (**Table S5**). To
415 detect potential bias from tag sequences in measured *cis*-regulatory activity, we applied a sequence-based
416 correction method, MPRA tag sequence analysis (MTSA)²⁸ (**Materials and Methods**). The regression using
417 MTSA-corrected expression levels demonstrated that 284 of 285 MPRA-significant variants displayed consistent
418 allelic direction before and after the correction. Moreover, 85% (melanoma) and 78% (melanocyte) of the MPRA-
419 significant variants ($FDR < 0.01$) still displayed an allelic difference at a relaxed criteria ($FDR < 0.1$) after
420 correction (**Table S7**). These results supported that the allelic differences detected in this study are robust and
421 reproducible, and we therefore used the normalized expression values before applying MTSA-correction
422 throughout the study.

423
424 **Fine-mapping and motif prediction of functional variants.** To supplement and compare with the variant
425 prioritization based on MPRA, we performed a fine-mapping analysis of the melanoma GWAS data. Statistical
426 fine-mapping of 54 melanoma GWAS loci using FINEMAP³³ nominated 2 to 101 variants per locus (median =
427 32.5) in 95% credible sets. We also performed a fine-mapping using POLYFUN³⁴ incorporating functional
428 annotations (precomputed prior causal probabilities based on a meta-analysis of 15 UK Biobank traits) following
429 the recommended procedure, which further narrowed down the credible set to 2 to 84 variants per locus (median =
430 19) (**Table S8; Materials and Methods**). Complementing and refining these prioritizations, MPRA identified 1
431 to 36 candidate functional variants per locus (median 5 variants) that display significant allelic transcriptional
432 activity from 42 melanoma GWAS loci (**Fig. S7**). MPRA-significant variants displayed slightly higher posterior
433 inclusion probability (PIP) and larger proportion of “high” probability score variants ($PIP > 0.1$) compared to non-
434 significant variants, resulting in higher percentage being included in the 95% credible sets of FINEMAP and
435 POLYFUN, although the enrichments were not statistically significant (**Fig. S8; Table S9**).

436 To assess the roles of transcription factors (TFs) in variant functionality, we predicted allelic TF binding
437 affinity of each MPRA tested variant using motifbreakR³⁸ (**Materials and Methods**). A substantial proportion of
438 MPRA-significant variants (167/285, 58.6%) were predicted to have effects on at least one TF binding site (**Table**
439 **S10**). These predicted allelic binding scores displayed a significant correlation with allelic transcriptional
440 activities measured from MPRA in C283T melanocyte data (Spearman $R = 0.249$, $P = 0.006$), and a non-
441 significant but similar pattern in UACC903 melanoma data ($R = 0.155$, $P = 0.172$) (**Fig. S9**). MPRA-significant
442 variants more frequently overlapped with the genomic regions annotated as open chromatin (32% vs. 28%; Chi-
443 squared $P = 0.0026$) or promoter/enhancer (15% vs. 11%; Chi-squared $P = 0.1998$) in melanoma or melanocyte
444 datasets compared to non-significant variants (**Fig. S10; Materials and Methods**). These results suggested that
445 some of observed allelic differences from MPRA could be attributed to differential binding of TFs and potentially
446 driven by functional *cis*-regulatory elements in melanocyte or melanoma cells.

447

448 **Nominating the most plausible candidate variants using an integrative scoring system.** To further nominate
449 the most plausible variants for in-depth follow-up from each locus, we integrated multi-layer functional
450 annotations and fine-mapping data to the 285 variants prioritized by MPRA. Given that our MPRA system
451 evaluates variants in an episomal setting, we incorporated chromatin features of the genomic regions around these
452 285 variants in melanocyte and melanoma cells. We previously profiled accessible chromatin regions in primary
453 cultures of melanocytes using ATAC-seq¹¹ ($n = 6$ individuals) and compiled other melanocyte and melanoma cell
454 chromatin features (accessible chromatin, promoter, and enhancer histone marks) from public databases and
455 published studies³⁰⁻³² (**Materials and Methods**). We also incorporated the information from the statistical fine-
456 mapping and motif prediction analyses described earlier. To systematically integrate these multi-layer features,
457 we established a scoring system by assigning three-level scores (0 = no hit, 1 = hit, 2 = strong hit) to each of the
458 eight components under four categories (MPRA, chromatin annotation, TF binding, and fine-mapping) for 285
459 MPRA significant variants (**Fig. 2A; Materials and Methods**). Within each locus, the variant(s) displaying the
460 highest integrative score based on these four categories were assigned as tier-1 variants. For some loci, there were

461 variants displaying lower but similar scores to the tier-1 variants (>70% of the highest score), which were
462 assigned as tier-2 variants. The rest were assigned as tier-3 variants (**Materials and Methods; Table S11**).

463 Using this system, we nominated 86 top-score variants including 52 tier-1 and 34 tier-2 variants across
464 the 42 loci (**Fig. 2B; Table S11**), with 1 to 6 top-score variants per locus (median = 2) and a single top-score
465 variant for 18 of 42 loci (43%). Among them were well-characterized functional variants including the top two
466 variants with the highest scores (**Fig. 2C**). For example, rs12913832 in the locus at 15q13.1 (42_15q13.1) is a
467 known functional variant in a melanocyte enhancer element mediating allelic *OCA2* expression⁵. rs398206 in the
468 locus at 21q22.3 (51_21q22.3) was shown to regulate *MX2* expression in melanocytes via allelic binding of YY1,
469 and *MX2* accelerated melanoma formation in a zebrafish model⁹. A third variant (displaying the 4th highest score),
470 rs117132860 in the locus at 7p21.2 (20_7p21.1) is a functional variant driving UVB-responsive allelic expression
471 of *AHR* with a prolonged effect in melanocyte growth and cellular response to UVB exposure¹¹. Re-identification
472 of these known functional variants supported the validity of our scoring system for variant prioritization. For 15
473 other loci with a single top-score variant (**Fig. 2C; Table S11**), prioritized variants include top candidates from
474 our previous study⁹ (e.g., rs3769823 at 8_2q33.1) as well as those from 10 newly discovered loci by the recent
475 GWAS (e.g., rs61935859 at 40_12q24.21, rs4753840 at 35_11q22.3, rs1046793 at 41_13q34, and rs61898347 at
476 36_11q23.3)³. These data demonstrated that most of the melanoma GWAS loci (78%) harbor potential functional
477 variants via *cis*-regulatory mechanism (i.e., allelic transcriptional activity) either with a single prominent
478 candidate (42% of loci) or multiple (up to six) functional candidate variants (58% of the loci) based on the multi-
479 layer functional features.

480

481 **Linking functional variants to target genes using eQTL/meQTL.** To link the candidate functional variants to
482 target susceptibility genes, we used eQTL and meQTL of melanocytes from 106 individuals and of melanoma
483 tissues from 444 individuals with skin cutaneous melanomas from TCGA. We previously identified 597,335
484 significant *cis*-eQTLs and 1,497,502 *cis*-meQTLs (+/-1 Mb of TSS or CpG sites, *FDR* < 0.05, not LD-pruned) in
485 melanocytes, and 209,393 significant *cis*-eQTLs and 3,794,446 *cis*-meQTLs in melanomas^{17,18}. Sixty percent of
486 the MPRA-significant variants (172/285) overlapped genome-wide significant eQTLs or meQTLs in melanocytes

487 or melanomas, nominating 31 candidate eGenes (**Table S12**) and 42 assigned genes for meProbes (that we define
488 as meGenes) in 26 loci (**Table S13**). Among these loci, 9 loci were mapped to a single eGene or meGene (**Fig.**
489 **3A**), 8 loci to 2 eGenes/meGenes (**Fig. 3B**, including those at 5_1q42.12 and 36_11q23.3 to the same gene by
490 both eQTL and meQTL), while 8 loci were mapped to three or more eGenes/meGenes (**Fig. 3C-D**). A total of 23
491 eGenes (23/31, 74.2%) and 25 meGenes (25/42, 59.5%) were further supported by GWAS-QTL colocalization or
492 TWAS/MWAS¹³. Furthermore, a total of 93 MPRA-significant variants from 14 loci displayed a consistent
493 direction between MPRA and eQTL, in which the direction of allelic expression of local genes matches those of
494 MPRA allelic transcriptional levels (**Table S12**). We limited the allelic direction matching analysis to eQTL
495 genes because of the intrinsic complexity of association between DNA methylation levels and target gene
496 expression levels.

497 For example, rs61935859, a single tier-1 top-score variant in the locus at 12q24.21 (40_12q24.21), is
498 linked to a single eGene, *MED13L* (**Fig. 2C, Fig. 3A**), with a matched direction of allelic expression. Namely, the
499 melanoma risk-associated G allele displayed 1.6- and 1.05-fold higher transcriptional activity in MPRA ($FDR =$
500 $5.48e-92$ and $4.93e-4$ in UACC903 and C283T, respectively) and is correlated with higher *MED13L* levels in the
501 melanocyte dataset (slope 0.48 relative to G allele and eQTL $P = 5.49e-9$). In the locus at 22q13.1 (52_22q13.1),
502 the tier-1 variant, rs4384 (**Fig. 3B**), was the only tier-1 top-score variant and also with a matched direction of
503 allelic expression with eGene *MAFF*, where the melanoma risk-associated G allele increased transcription by 1.3-
504 fold in MPRA ($FDR = 5.05e-41$ in UACC903) and is also correlated with higher *MAFF* levels in the melanocyte
505 dataset (slope 0.89 relative to G allele and eQTL $P = 7.11e-25$). *MAFF* encodes a basic leucine zipper (bZIP) TF
506 and has been reported to be involved in multiple cancers. In the locus at 16q22.1 (44_16q22.1), two top-score
507 variants, rs9928796 and rs7199991 (**Fig. 3C**), are linked to *CDHI* (increased with risk) and *FTLPI4* (increased
508 with risk) with matched directions of allelic expression in the melanocyte eQTL dataset and MPRA in UACC903.
509 While *FTLPI4* is a pseudogene, *CDHI* encodes E-cadherin. E-cadherin is a cell adhesion molecule responsible
510 for the adhesion of melanocytes to keratinocytes³⁹, and loss of E-cadherin was observed in melanoma
511 progression⁴⁰, in line with its roles in epithelial-to-mesenchymal transitions⁴¹. In the locus at 12p13.1
512 (37_12p13.1), rs2111398 and rs850934 (**Fig. 3C**) are the top-score variants linked to 4 eGenes (*GPRC5A*,

513 *HTR7P1*, *HEBP1*, and *EMP1*; decreased gene expression associated with risk for all four genes) with matched
514 directions of allelic expression in the melanocyte eQTL dataset and MPRA in UACC903. We note that some of
515 these variants displaying strong allelic function in UACC903 including those at 22q13.1, 16q22.1, and 12p13.1
516 were only significant eQTLs/meQTLs in melanocyte dataset potentially due to relatively lesser statistical power
517 in the heterogeneous TCGA tumor tissue dataset. Significantly enriched pathways in these 31 eGenes and 42
518 meGenes (67 unique genes) consistently highlighted those relevant to cellular immune response and apoptosis
519 signaling (**Table S14**). Thus, by combining MPRA and molecular QTLs in melanomas and melanocytes, we
520 nominated candidate susceptibility genes linked to one or more plausible functional variants from 48% of the
521 known melanoma GWAS loci.

522

523 **Validation of functional variants and target genes by CRISPRi.** To further determine whether the
524 genomic regions encompassing the prioritized functional variants regulate expression levels of target genes, we
525 performed CRISPRi of three representative top-tier variants using dCas9-ZIM3 system in the UACC903
526 melanoma cell line (**Fig. 4A, Materials and Methods**). We focused on loci 1) that have not been previously
527 characterized, 2) with eGenes identified in GWAS-eQTL colocalization or TWAS, 3) with eGenes and tier-1
528 variants displaying a matching allelic direction between eQTL and MPRA, and 4) with the variants located in
529 annotated enhancers/promoters in melanomas or melanocytes. Using these criteria, we selected 5 variant-eGene
530 pairs from 3 loci (rs61935859-*MEDI3L* at 12q24.21, rs4384-*MAFF* at 22q13.1, and rs2111398-
531 *GPRC5A/HEBP1/EMP1* at 12p13.1) and targeted each SNP using three different guide-RNAs (gRNAs).
532 CRISPRi followed by qPCR demonstrated a 31-60% reduction of *MAFF* levels upon targeting the region
533 encompassing rs4384 for all three gRNAs ($P = 1.56e-6$, 0.031 and $8.17e-7$, two-tailed t-test, $n = 24$, combined
534 from 4 biological replicates; **Fig. 4B**). We also observed a 27-30% reduction of *GPRC5A* levels for all three
535 gRNAs targeting rs2111398 ($P = 0.005$, $P = 0.002$ and $P = 0.002$, two-tailed t-test, $n = 24$, combined from 4
536 biological replicates; **Fig. 4C**). No significant changes of *HEBP1* or *EMP1* levels were observed for all three
537 gRNAs in this locus (at $P < 0.017$ cutoff for testing three genes). For *MEDI3L*, we did not observe significant
538 changes in three biological replicates (**Fig. S11**). These data identified *MAFF* and *GPRC5A* as plausible

539 melanoma susceptibility genes regulated by functional *cis*-regulatory variants and demonstrated that our scoring
540 strategy could nominate the most plausible loci, functional variants, and candidate susceptibility genes for further
541 in-depth characterization.

542

543 **Cell-type specificity of melanoma-associated functional variants.** Given that MPRA-significant variants
544 displayed cell-type dependent allelic activity, we further inspected the cell-type specific functionality of these
545 variants. Namely, 57 variants displayed significant allelic transcriptional activity in both melanoma and
546 melanocyte cell lines, while 77 variants were only significant for melanoma and 151 variants only for
547 melanocytes (**Fig. 5A**). Notably, 1.6-times more variants were identified in melanocytes, even though the total
548 transfected cells were more for UACC903 (transfection events = 8 in UACC903 and 5 in C283T), potentially due
549 to higher transfection efficiency of C283T cells. On the other hand, allelic differences in transcriptional activities
550 were significantly larger for 134 variants significant in melanoma (median 1.14-fold, range 1.07 to 2.88-fold,
551 **Table S5**) than for 208 variants significant in melanocyte (median 1.06-fold; range 1.03 to 2.47-fold, **Table S6**)
552 ($P = 5e-15$, two-tailed unpaired t-test), which is consistent with elevated global transcription levels observed in
553 cancer cells^{42,43}. For the 57 variants that are significant in both cell types, allelic differences displayed a similar
554 pattern with a larger effect size in melanoma (median 1.19-fold in melanoma vs. 1.08-fold in melanocyte) and a
555 significant difference between two cell types in a paired test ($P = 0.00037$, two-tailed paired t-test). For example,
556 the variant rs398206 in the locus at 21q22.3 (51_21q22.3) showed significant allelic transcriptional activity both
557 in melanoma (MPRA $FDR = 0$) and melanocyte (MPRA $FDR = 3.83e-18$), but its allelic effect was stronger in
558 melanoma (allelic difference of 2.88-fold in melanoma vs. 1.16-fold in melanocyte; **Fig. S12A**). We further
559 inspected cell-type dependent “activator” (**Materials and Methods**) function of the DNA sequences harboring
560 MPRA-significant variants. Among 285 MPRA-significant variants, ~2-fold more variants also displayed
561 activator function in melanoma (57 variants) compared to melanocytes (28 variants). Moreover, 32% of 77
562 melanoma-only allelic variants were also located in melanoma-only activators, while 5% of 151 melanocyte-only
563 variants were in melanocyte-only activators. These observations suggested substantial cell-type specificity of

564 melanoma-associated functional variants between melanoma and melanocyte and their potentially larger allelic
565 effect sizes accompanied by stronger transcriptional activity in melanoma cells in our system.

566 To formally nominate the cell-type specific variants, we further assessed 77 melanoma-only and 151
567 melanocyte-only variants in MPRA ($FDR < 1\%$, **Fig. 5A**). We applied three criteria for these variants as follows:
568 1) a variant shows strong allelic transcriptional activity (MPRA $FDR < 10^{-9}$) in the same cell type, 2) 145bp
569 encompassing the variant is a *cis*-activator from the MPRA of the same cell type, or 3) the level of TF predicted
570 to show allelic binding is significantly higher in the same cell type based on the differentially expressed gene
571 analysis between UACC903 melanoma and C283T melanocyte cells (**Fig. 5A**). We reasoned that an extreme
572 allelic significance cutoff (MPRA $FDR < 10^{-9}$) could help reduce potential false positives coming from technical
573 differences (e.g., transfection efficiency, potential tag sequence effect). Further, we hypothesized that potential
574 drivers of cell-type dependency in allelic transcriptional activity could be enhancer strength and/or differential
575 availability of allele-preferential binding TFs between two cell types used in MPRA. To test this hypothesis, we
576 performed a transcriptome analysis of UACC903 and C283T cells by sequencing the same mRNA samples from
577 MPRA transfections ($n = 3$ from each cell type). A total of 4,388 differentially expressed genes (DEGs; $P < 0.01$
578 and $|\log_2 \text{fold change}| > 2$) were identified using DESeq2. After applying the three criteria, a total of 36 of 77
579 variants met at least one criterion in melanoma (**Table S15**) and 45 of 151 variants in melanocytes (**Table S16**),
580 which we define as melanoma-specific and melanocyte-specific variants, respectively (**Fig. 5A**). One example is
581 rs4384 in the locus at 22q13.1 (52_22q13.1), which only showed significant allelic transcriptional activity in
582 melanoma (MPRA $FDR = 5.05e-41$, **Fig. S12B**) and was nominated by all three criteria. To confirm the cell-type
583 specificity, we applied a linear regression to encode the interaction between cell type and allelic effect (**Materials**
584 **and Methods**). Notably, all 5 variants nominated by all three criteria displayed significant interaction between
585 allelic effect and cell-type ($FDR < 0.01$). Moreover, 82% (melanoma-specific) and 75% (melanocyte-specific) of
586 variants nominated by at least two criteria displayed significant interaction between allelic effect and cell-type
587 ($FDR < 0.01$) (**Tables S15-16**). These results further validated the cell-type specific variants nominated using
588 three criteria.

589 Notably, we observed 10 loci with only melanoma-specific variants (examples in **Fig. 5B**), 12 loci with
590 both melanoma- and melanocyte-specific variants (an example in **Fig. 5C**), and 11 loci with only melanocyte-
591 specific variants (examples in **Fig. 5D**). We further looked into the QTL-based target genes assigned to these cell-
592 type specific variants in the matching cell type. As shown in **Fig. 5E**, a total of 5 eGenes and 12 meGenes from
593 melanomas are linked with melanoma-specific variants, while 12 eGenes and 11 meGenes from melanocytes are
594 linked with melanocyte-specific variants (**Fig. 5F**). In the locus at 5p15.33 (11_5p15.33), two of the three MPRA-
595 significant variants are melanoma-specific, are meQTLs for *TERT* only in melanomas but are meQTLs for
596 *CLPTMIL* in both melanomas and melanocytes. Notably, *TERT* expression is re-activated in transformed
597 melanoma cells but not in differentiated melanocytes. In the locus at 1q42.12 (5_1q42.12), a single variant
598 rs1865220 is melanocyte-specific and an eQTL for *PARP1*, consistent with its role of mediating melanocyte
599 growth⁸. Many other loci displayed both melanoma- and melanocyte-specific variants that are linked with target
600 genes. Two variants in the locus at 6p21.32 (18_6p21.32) are melanoma-specific and eQTLs for two *HLA* genes,
601 *HLA-DQA1* and *HLA-DQB1*, in melanomas, while two variants in the same locus are melanocyte-specific and
602 meQTLs for an immunoproteasome gene, *PSMB9*, in melanocytes. The locus at 1q21.3 (2_1q21.3) presents 2
603 each of melanocyte-specific and melanoma-specific variants, where *CTSS* in melanocytes and *HORMAD1* in
604 melanomas are representative target genes. The locus at 16q22.1 (44_16q22.1) presents 5 melanocyte-specific and
605 7 melanoma-specific variants with the common target gene, *CDH1*.

606 For 10 of 36 melanoma-specific variants, the expression levels of TFs predicted to show allelic binding
607 were higher in UACC903 melanoma cells compared to C283T melanocytes (DEGs with $FDR < 0.01$ and $|\log_2$
608 fold change| > 2 , $n = 3$, **Materials and Methods; Table S15**). Notably, HES1 and HEY2, which are known
609 targets of NOTCH signaling pathway⁴⁴ and induced in cancers, were linked to four variants from distinct loci
610 (52_22q13.1, 11_5p15.33, 16_6p22.3, and 44_16q22.1), and three of these variants are in melanoma-specific
611 activators. For 31 of 45 melanocyte-specific variants, the levels of predicted allelic TFs were higher in C283T
612 melanocytes compared to UACC903 melanomas (**Table S16**). For 22 variants among them, differentially
613 expressed TFs (EGR4, HIC1, TBX5, TCF4, THRB, ARID3A, FOSL2, JUN, JUNB, FOXF2, KLF8, MEIS1,
614 IRF1, IRF7, and IRF9) were linked to melanocyte-specific variants from more than one locus. These data

615 suggested that melanoma risk-associated variants within and across multiple GWAS loci could be functional in
616 different cellular contexts representing normal/primary melanocytes and transformed/melanoma cells, and TF
617 levels could potentially contribute to the context dependency.

618

619 **Effect of transcription factors on allelic expression of susceptibility genes.** Given the suggested roles of TFs in
620 the allelic transcriptional activity of melanoma-associated variants including cell type-specific ones, we further
621 investigated the interaction of MPRA-significant variants and the levels of allelic binding TFs on target eGene
622 expression in large-scale eQTL datasets. For this, we included 38 variant-TF-eGene trios from melanoma data, by
623 selecting MPRA-significant variants in UACC903 ($FDR < 0.01$), significant allelic binding of a TF to the variant
624 predicted by motifBreakR, and genome-wide significant eQTL target gene for the variant in TCGA melanomas.
625 We included 119 trios from melanocyte data, similarly selecting MPRA-significant variants in C283T, their
626 predicted TFs, and target genes in melanocyte eQTL dataset.

627 Using a multiple linear regression interaction model⁴⁵ (**Materials and Methods**), significant variant-TF-
628 eGene interactions were identified for 7 melanoma trios and 7 melanocyte trios at FDR 5% (**Table S17**). In the
629 melanoma analysis, 4 variants from the locus at 1q21.3 (2_1q21.3) significantly interacted with 7 different TFs
630 (ATF6, E4F1, REST, ESRRG, ZNF143, ATF5, and FOXJ3) and all were linked to an eGene, *HORMADI* (**Table**
631 **S17**). Notably, this locus has a large LD block with multiple functional variants including 2 tier-1, 2 tier-2, 2
632 melanoma-specific, and 2 melanocyte-specific variants with 9 potential target genes (**Fig. 3D, 5E-F**). *HORMADI*
633 is a melanoma-specific eQTL gene for multiple MPRA-significant variants, and rs10305673 (melanoma-specific
634 variant), among them, showed a significant interaction with a TF, REST, in the TCGA melanoma dataset ($FDR =$
635 0.000488; **Table S17**). Further, one of the tier-1 variants at this locus, rs2864871, showed a significant interaction
636 with three TFs in the TCGA melanoma dataset (ATF6, E4F1, and ESRRG; $FDR = 0.000423, 0.000423,$ and
637 0.00493, respectively; **Table S17**). These data suggested that these TF-interacting MPRA functional variants
638 potentially mediate *HORMADI* expression regulation that might contribute to melanoma susceptibility at this
639 locus. In melanocyte analysis, 3 variants from 3 loci (including 2 melanocyte-specific variants) significantly
640 interacted with 6 different TFs (FLI1, THRB, ETV4, ELF1, ETS1, and POU3F1) and 4 eGenes (*GPRC5A*,

641 *CDH1*, *HEBP1*, and *CASP8*) (**Table S17**). Notably, the variant rs850936 (melanocyte cell type score = 1) showed
642 an interaction with 4 ETS-domain TFs (FLI1, ETV4, ELF1, and ETS1) on the expression of *GPRC5A* and/or
643 *HEBP1*. Among them, *FLI1* was a DEG displaying higher levels in C283T compared to UACC903, suggesting
644 that FLI1 might mediate cell-type specific allelic function of this variant in melanocytes (**Fig. S13A**). Moreover,
645 the variant rs4783674 (melanocyte cell type score = 2) showed an interaction with a TF, *THRB*, on *CDH1* levels
646 in melanocytes. Notably, the level of *THRB* was significantly higher in C283T melanocytes compared to
647 UACC903 melanoma cells, which further supported the hypothesis that *THRB* mediates melanocyte-specific
648 variant functionality altering *CDH1* expression to contribute to melanoma risk in this locus (**Fig. S13B**). Together
649 these data suggested that for a subset of MPRA-significant variants including cell-type specific variants, also
650 interact with TFs to affect the target eQTL gene levels, and TF availability might play an important role in variant
651 functionality including their cell-type specificity.

652

653 **DISCUSSION**

654 In this study, we performed MPRA of 1,992 variants selected from 54 melanoma GWAS loci to narrow down to a
655 small set (285, 14%) of functional variants displaying allelic transcriptional activity. To further reduce this set, we
656 constructed a score that leveraged multi-layer genetic and functional features including epigenomic annotation
657 from relevant cell types, GWAS fine-mapping scores, and motif-prediction, in addition to allelic functionality
658 measured by MPRA. This score nominated a small number of top-score variants for 42 of 54 known melanoma
659 GWAS loci, most of which had not been functionally tested before. The validity of the MPRA-significant variants
660 and the scoring system was demonstrated by re-identification of the well-characterized variants from three
661 melanoma loci as the top two variants^{5,9} and another high-ranking variant¹¹ among all 285 variants. By integrating
662 this variant scoring system with expression and methylation QTLs from disease-relevant cell types (melanoma
663 and melanocyte), we linked the functional variants to their potential target eGenes or meGenes. Some of these
664 variant-gene connections were validated using a CRISPRi system in a relevant cell type. Given that *in vitro* and *in*
665 *vivo* characterization of candidate susceptibility genes is laborious and time-consuming, a tiered nomination of
666 loci, variants, and genes for 48% of melanoma GWAS loci by our study will inform future functional follow-up

667 studies. Compared to our previous study⁹, the current study presents significant advances regarding the number of
668 tested loci (over two thirds of 54 loci tested for the first time), cellular context (primary melanocytes melanoma
669 cells were formally compared), further variant prioritization via scoring system, and variant-to-gene linkage via
670 both eQTL and meQTL.

671 Our systematic profiling of melanoma GWAS loci provided a few general observations regarding genetic
672 susceptibility to melanoma. Unbiased testing of all the known melanoma GWAS loci identified at least one
673 functional variant for 78% of these loci, adding support to the body of knowledge that transcriptional regulation is
674 a main mechanism that GWAS variants exert their function. As expected, the loci that are mainly explained by
675 coding variants of pigmentation genes (e.g., 5p13.2, 11q14.3) did not present strong functional variants based on
676 MPRA. Our integrative variant scoring system indicated that in 42% of the cases melanoma GWAS loci presented
677 a single prominent variant based on the overlap of variant transcriptional activity and multiple functional
678 annotation features. On the other hand, a larger proportion of the loci (58%) exhibited more than one equally
679 plausible functional variants, suggesting that multiple functional variants could potentially contribute to one or
680 more target genes in each locus. This observation is somewhat consistent with the recent study that identified
681 multiple causal regulatory variants that are in high-LD for a subset of lymphoblastoid cell eQTL loci using
682 MPRA⁴⁶.

683 We provided further support to a few melanoma susceptibility genes that have not been studied before by
684 validating the connections between the top-score variants and their target eGenes using CRISPRi system. For the
685 locus at 22q13.1, *MAFF* was identified as a target of the top-score variant, rs4384. Higher levels of *MAFF* are
686 correlated with the melanoma risk-associated allele in melanocytes, matching the allelic activity of rs4384 in
687 MPRA. *MAFF* encodes a bZIP TF that lacks a transactivation domain that forms heterodimers with several
688 regulators of antioxidant responses (e.g., NRF2⁴⁷ and BACH1⁴⁸), regulating genes in stress response and
689 detoxification pathways⁴⁹. *MAFF* has been shown to act as an oncogene that plays a vital role in tumor invasion
690 and metastasis⁴⁸. The variant rs4384 is also a melanoma-specific variant predicted to bind HES1 in melanoma
691 context. Although the interaction of rs4384 and HES1 on *MAFF* expression could not be tested because *MAFF*
692 was not a significant eGene in the TCGA melanoma dataset, HES1-mediated *MAFF* regulation in melanoma can

693 be investigated as a potential mechanism of melanoma susceptibility in this locus. For the locus at 12p13.1,
694 *GPRC5A* was validated as a target of the region harboring rs2111398, the top-score variant of the locus, using
695 CRISPRi to target this region and assessing multiple eQTL target. *GPRC5A* is an orphan G protein-coupled
696 receptor that has an important role in growth and survival of cancer cells⁵⁰ and sustaining cell adhesion⁵¹. The
697 melanoma risk-associated allele is correlated with lower expression of *GPRC5A* in melanocytes, which is
698 consistent with the allelic activity of rs2111398 in MPRA. We did not observe significant effect of CRISPRi on
699 *MED13L* levels in our system. Given that *MED13L* plays an essential role in general transcription regulation as
700 well as embryonic development⁵², it is possible that multiple layers of redundant regulatory mechanism⁵³ hindered
701 the detection of relatively small effect of a single enhancer. It is also possible that there are other target gene(s)
702 that were not detected in our QTL datasets.

703 Our study highlighted the cell-type specific functionality of cancer-associated variants in the contexts of
704 tumor and cell of tumor origin. We identified a subset of MPRA-significant variants as melanoma- (13%) or
705 melanocyte-specific (16%) variants, while most of the variants are functional in both. Notably, these cell-
706 type/context-specific variants were distributed evenly across melanoma GWAS loci, suggesting that both tumor
707 and cell-of-origin contexts may play a role across melanoma loci. For example, two top melanoma-specific
708 variants (rs452384 and rs31487) in the locus at 5p15.33 were identified based on their strong allelic transcription
709 and enhancer activity restricted to melanoma cell line. Notably, these variants are also significant meQTLs for
710 CpG probes within *TERT* in the TCGA melanoma dataset but not in the melanocyte dataset. Given that *TERT*
711 expression is reactivated in most cancers including melanoma⁵⁴, two of three MPRA-significant variants at this
712 locus being melanoma-specific is consistent with them contributing to target gene expression in tumor context
713 rather than normal melanocyte context. Moreover, rs452384 in the locus at 5p15.33 and another top melanoma-
714 specific variant, rs4384 in the locus at 22q13.1, are both predicted to modulate the binding of a NOTCH1 target
715 gene *HES1*, which displayed elevated expression in the UACC903 melanoma cell line compared to C283T
716 melanocytes and has previously been shown to promote tumorigenesis⁵⁵. This observation and identification of
717 two other melanoma-specific variants (rs6914598 at 6p22.3 and rs57688464 at 16q22.1) potentially recruiting
718 another NOTCH1 target, *HEY2*, suggested that tumor-specific activation of TFs could mediate the activity of

719 melanoma-specific variants across multiple loci. Consistent with this observation, our previous TWAS analysis
720 demonstrated that increased *NOTCH2* levels (located in 1p12) in melanocytes are associated with melanoma
721 risk⁴⁶. NOTCH signaling is involved in maintaining melanocyte stem cells and melanoblasts⁵⁶, and Notch1 was
722 shown to reprogram mature melanocytes into stem-like cells⁵⁷. *NOTCH1* was also shown to be elevated in
723 melanomas and promote growth and survival of melanoma cells⁵⁸. Interaction analysis of the functional variants
724 and their TF partners further validated a melanoma-specific variant (REST-rs10305673-*HORMAD1*) and two
725 melanocyte-specific variants (FLI1-rs850936-*GPRC5A* and THRB-rs4783674-*CDHI*) identified through MPRA
726 in the large-scale expression datasets. These data further supported the roles of TFs in mediating cell-type specific
727 variant function contributing to melanoma susceptibility. Future studies exploring the effects of these TFs on
728 target gene expression in relevant cell types using CRISPR knock-out/knock-in of TF motifs or direct modulation
729 of TF levels will be informative. Although we identified more melanocyte-specific functional variants than
730 melanoma-specific ones through MPRA, we observed larger allelic effect sizes and stronger enhancer activities of
731 MPRA-significant variants in the UACC903 melanoma cell line in general. This could be due to increased global
732 transcription levels in cancer cells by oncogene-induced activation and amplification of general transcription that
733 have been observed before^{42,43}.

734 We acknowledge several limitations of the current study. First, MPRA-significant variants were not
735 identified for 22% (12/54) of the melanoma GWAS loci. While these loci might have alternative mechanisms that
736 could not be tested by MPRA, incorporating additional cell types (e.g., immune cells) and relevant exposures or
737 contexts (e.g., exposure to ultraviolet radiation⁵⁹) as well as adopting a lentiviral system to reflect genomic context
738 in MPRA approaches could potentially identify additional functional variants. Second, 38% (16/42) of the loci
739 with MPRA-significant variants are not supported by any genome-wide significant QTLs in melanoma or
740 melanocyte datasets. This could be attributed to limited statistical power for lower-frequency variants and
741 heterogeneity in melanoma tumor samples further limiting the eQTL detection¹⁷ as well as cellular contexts of
742 eQTL detection that were not incorporated in these datasets. The power issue in the tumor eQTL dataset as well as
743 potential differences between episomal enhancer activity tested in MPRA and endogenous expression measured in
744 QTL datasets also limited our variant-TF interaction analyses as many melanoma-specific variants (e.g., rs4384

745 and rs2111398) are showing stronger allelic activities in UACC903 cell line but are linked to melanocyte eQTLs.
746 To complement eQTL-based approaches, adopting chromatin interaction methods (e.g., capture-Hi-C⁶⁰) will be
747 beneficial for better sensitivity in variant-gene linkage. For example, Activity-By-Contact (ABC) model utilizes
748 epigenomic features and Hi-C data to predict the enhancer-gene connections⁶¹. An initial query of ABC model
749 based on skin fibroblasts data (foreskin_fibroblast-Roadmap, ABC scores no less than 0.015) nominated
750 candidate genes for 31 variants among 285 MPRA-significant variants, which includes 12 variants that are not
751 linked to any gene based on eQTL/meQTL (**Table S18**).

752 In conclusion, we provide a strategy to profile multiple cancer GWAS loci using high-throughput variant
753 screening and prioritization while incorporating the contexts of tumor and cell of tumor origin, which could be
754 applied to other cancer GWAS follow-up studies.

755

756 **Declaration of interests**

757 The authors declare no competing interests.

758

759 **Acknowledgements**

760 This work has been supported by the Intramural Research Program (IRP) of the Division of Cancer Epidemiology
761 and Genetics, National Cancer Institute, US National Institutes of Health. This work utilized the Biowulf cluster
762 computing system at the NIH. The results appearing here are in part based on data generated by the TCGA
763 Research Network. We would like to thank members at the National Cancer Institute Cancer Genomics Research
764 Laboratory (CGR) for help with sequencing efforts. We would like to thank the research participants and
765 employees of 23andMe, Inc. for making this work possible. The content of this publication does not necessarily
766 reflect the views or policies of the US Department of Health and Human Services, nor does the mention of trade
767 names, commercial products, or organizations imply endorsement by the US Government.

768

769 **Web resources**

770 POLYFUN, <https://github.com/omerwe/polyfun>

771 FINEMAP, <http://www.christianbenner.com/>
772 motifbreakR, <https://github.com/Simon-Coetzee/motifBreakR>
773 MTSA, <https://github.com/Dongwon-Lee/mtsa>
774 Rsubread, <https://bioconductor.org/packages/release/bioc/html/Rsubread.html>
775 DESeq2, <https://github.com/mikelove/DESeq2>
776 NIH Biowulf Cluster, <http://hpc.nih.gov>
777 The Cancer Genome Atlas (TCGA) Research Network, <http://cancergenome.nih.gov/>

778

779 **Data and code availability**

780 The sequencing data generated during this study (MPRA sequencing and RNA-seq data) are being submitted to
781 Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>). A complete list of MPRA oligo sequences
782 can be found in **Data S1**. The raw Illumina HumanMethylation450 BeadChips data are accessible through GEO
783 under the accession GEO: GSE166069; melanocyte genotype data, RNA-seq expression data, and all
784 eQTL/meQTL association results are accessible through Genotypes and Phenotypes (dbGaP) under accession
785 dbGaP: phs001500.v2.p1. Data from the 2020 melanoma GWAS meta-analysis performed by Landi and
786 colleagues were obtained from dbGaP (dbGaP: phs001868.v1.p1), with the exclusion of self-reported data from
787 23andMe, Inc. and UK Biobank. The full GWAS summary statistics for the 23andMe discovery dataset will be
788 made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the
789 privacy of the 23andMe participants. Please visit <https://research.23andme.com/collaborate/#dataset-access/> for
790 more information and to apply to access the data. Summary data from the remaining self-reported cases are
791 available from the corresponding authors of that manuscript³ (Matthew Law, matthew.
792 law@qimrberghofer.edu.au; Mark Iles, m.m.iles@leeds.ac.uk; and Maria Teresa Landi, landim@mail.nih.gov).
793

794 **REFERENCES**

- 795 1. Shain, A. H. & Bastian, B. C. From melanocytes to melanomas. *Nat. Rev. Cancer* **16**, 345–358 (2016).
- 796 2. Karimkhani, C. *et al.* The global burden of melanoma: results from the Global Burden of Disease Study
797 2015. *Br. J. Dermatol.* **177**, 134–140 (2017).
- 798 3. Landi, M. T. *et al.* Genome-wide association meta-analyses combining multiple risk phenotypes provide
799 insights into the genetic architecture of cutaneous melanoma susceptibility. *Nat. Genet.* **52**, 494–504 (2020).
- 800 4. Beaumont, K. A. *et al.* Receptor function, dominant negative activity and phenotype correlations for MC1R
801 variant alleles. *Hum. Mol. Genet.* **16**, 2249–2260 (2007).
- 802 5. Visser, M., Kayser, M. & Palstra, R.-J. HERC2 rs12913832 modulates human pigmentation by attenuating
803 chromatin-loop formation between a long-range enhancer and the OCA2 promoter. *Genome Res.* **22**, 446–
804 455 (2012).
- 805 6. Tsetskhladze, Z. R. *et al.* Functional assessment of human coding mutations affecting skin pigmentation
806 using zebrafish. *PLoS One* **7**, e47398 (2012).
- 807 7. Halaban, R. *et al.* Aberrant retention of tyrosinase in the endoplasmic reticulum mediates accelerated
808 degradation of the enzyme and contributes to the dedifferentiated phenotype of amelanotic melanoma cells.
809 *Proc. Natl. Acad. Sci. U. S. A.* **94**, 6210–6215 (1997).
- 810 8. Choi, J. *et al.* A common intronic variant of PARP1 confers melanoma risk and mediates melanocyte growth
811 via regulation of MITF. *Nat. Genet.* **49**, 1326–1335 (2017).
- 812 9. Choi, J. *et al.* Massively parallel reporter assays of melanoma risk variants identify MX2 as a gene
813 promoting melanoma. *Nat. Commun.* **11**, 2718 (2020).
- 814 10. Fang, J. *et al.* Functional characterization of a multi-cancer risk locus on chr5p15.33 reveals regulation of
815 TERT by ZNF148. *Nat. Commun.* **8**, 15034 (2017).
- 816 11. Xu, M. *et al.* A UVB-responsive common variant at chromosome band 7p21.1 confers tanning response and
817 melanoma risk via regulation of the aryl hydrocarbon receptor, AHR. *Am. J. Hum. Genet.* **108**, 1611–1630
818 (2021).
- 819 12. Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).

- 820 13. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA.
821 *Science* **337**, 1190–1195 (2012).
- 822 14. Gallagher, M. D. & Chen-Plotkin, A. S. The post-GWAS era: From association to function. *Am. J. Hum.*
823 *Genet.* **102**, 717–730 (2018).
- 824 15. Law, M. H. *et al.* Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant
825 melanoma. *Nat. Genet.* **47**, 987–995 (2015).
- 826 16. Liu, Y., Ye, Y., Gong, J. & Han, L. Expression quantitative trait loci (eQTL) analysis in cancer. *Methods*
827 *Mol. Biol.* **2082**, 189–199 (2020).
- 828 17. Zhang, T. *et al.* Cell-type-specific eQTL of primary melanocytes facilitates identification of melanoma
829 susceptibility genes. *Genome Res.* **28**, 1621–1635 (2018).
- 830 18. Zhang, T. *et al.* Cell-type-specific meQTLs extend melanoma GWAS annotation beyond eQTLs and inform
831 melanocyte gene-regulatory mechanisms. *Am. J. Hum. Genet.* **108**, 1631–1646 (2021).
- 832 19. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues.
833 *Science* **369**, 1318–1330 (2020).
- 834 20. Kim-Hellmuth, S. *et al.* Cell type-specific genetic regulation of gene expression across human tissues.
835 *Science* **369**, (2020).
- 836 21. Liu, J., Fukunaga-Kalabis, M., Li, L. & Herlyn, M. Developmental pathways activated in melanocytes and
837 melanoma. *Arch. Biochem. Biophys.* **563**, 13–21 (2014).
- 838 22. Mountjoy, E. *et al.* An open approach to systematically prioritize causal variants and genes at all published
839 human GWAS trait-associated loci. *Nat. Genet.* **53**, 1527–1533 (2021).
- 840 23. Boer, C. G. *et al.* Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. *Cell*
841 **184**, 4784-4818.e17 (2021).
- 842 24. van Arensbergen, J. *et al.* High-throughput identification of human SNPs affecting regulatory element
843 activity. *Nat. Genet.* **51**, 1160–1169 (2019).
- 844 25. Ulirsch, J. C. *et al.* Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood
845 Cell Traits. *Cell* **165**, 1530–1545 (2016).

- 846 26. Melnikov, A. *et al.* Systematic dissection and optimization of inducible enhancers in human cells using a
847 massively parallel reporter assay. *Nat. Biotechnol.* **30**, 271–277 (2012).
- 848 27. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to
849 multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).
- 850 28. Lee, D. *et al.* Sequence-based correction of barcode bias in massively parallel reporter assays. *Genome Res.*
851 **31**, 1638–1645 (2021).
- 852 29. Lee, D. LS-GKM: a new gkm-SVM for large-scale datasets. *Bioinformatics* **32**, 2196–2198 (2016).
- 853 30. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome.
854 *Nature* **489**, 57–74 (2012).
- 855 31. Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28**, 1045–
856 1048 (2010).
- 857 32. Verfaillie, A. *et al.* Decoding the regulatory landscape of melanoma reveals TEADS as regulators of the
858 invasive cell state. *Nat. Commun.* **6**, 6683 (2015).
- 859 33. Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-wide association
860 studies. *Bioinformatics* **32**, 1493–1501 (2016).
- 861 34. Weissbrod, O. *et al.* Functionally informed fine-mapping and polygenic localization of complex trait
862 heritability. *Nat. Genet.* **52**, 1355–1363 (2020).
- 863 35. Liao, Y., Smyth, G. K. & Shi, W. The R package Rsubread is easier, faster, cheaper and better for alignment
864 and quantification of RNA sequencing reads. *Nucleic Acids Res.* **47**, e47 (2019).
- 865 36. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data
866 with DESeq2. *Genome Biol.* **15**, 550 (2014).
- 867 37. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a
868 reference genome. *BMC Bioinformatics* **12**, 323 (2011).
- 869 38. Coetzee, S. G., Coetzee, G. A. & Hazelett, D. J. motifbreakR: an R/Bioconductor package for predicting
870 variant effects at transcription factor binding sites. *Bioinformatics* **31**, 3847–3849 (2015).

- 871 39. Tang, A. *et al.* E-cadherin is the major mediator of human melanocyte adhesion to keratinocytes in vitro. *J.*
872 *Cell Sci.* **107 (Pt 4)**, 983–992 (1994).
- 873 40. Hsu, M. Y., Wheelock, M. J., Johnson, K. R. & Herlyn, M. Shifts in cadherin profiles between human
874 normal melanocytes and melanomas. *J. Investig. Dermatol. Symp. Proc.* **1**, 188–194 (1996).
- 875 41. Onder, T. T. *et al.* Loss of E-cadherin promotes metastasis via multiple downstream transcriptional
876 pathways. *Cancer Res.* **68**, 3645–3654 (2008).
- 877 42. Lin, C. Y. *et al.* Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **151**, 56–67 (2012).
- 878 43. Kotsantis, P. *et al.* Increased global transcription activity as a mechanism of replication stress in cancer. *Nat.*
879 *Commun.* **7**, 13087 (2016).
- 880 44. Aster, J. C., Pear, W. S. & Blacklow, S. C. The Varied Roles of Notch in Cancer. *Annu. Rev. Pathol.* **12**,
881 245–275 (2017).
- 882 45. Flynn, E. D. *et al.* Transcription factor regulation of eQTL activity across individuals and tissues. *PLoS*
883 *Genet.* **18**, e1009719 (2022).
- 884 46. Abell, N. S. *et al.* Multiple causal variants underlie genetic associations in humans. *Science* **375**, 1247–1254
885 (2022).
- 886 47. Moon, E. J. & Giaccia, A. Dual roles of NRF2 in tumor prevention and progression: possible implications in
887 cancer treatment. *Free Radic. Biol. Med.* **79**, 292–299 (2015).
- 888 48. Moon, E. J. *et al.* The HIF target MAFF promotes tumor invasion and metastasis through IL11 and STAT3
889 signaling. *Nat. Commun.* **12**, 4308 (2021).
- 890 49. Kannan, M. B., Solovieva, V. & Blank, V. The small MAF transcription factors MAFF, MAFG and MAFK:
891 current knowledge and perspectives. *Biochim. Biophys. Acta* **1823**, 1841–1846 (2012).
- 892 50. Hirano, M. *et al.* Novel reciprocal regulation of cAMP signaling and apoptosis by orphan G-protein-coupled
893 receptor GPRC5A gene expression. *Biochem. Biophys. Res. Commun.* **351**, 185–191 (2006).
- 894 51. Bulanova, D. R. *et al.* Orphan G protein-coupled receptor GPRC5A modulates integrin β 1-mediated
895 epithelial cell adhesion. *Cell Adh. Migr.* **11**, 434–446 (2017).

- 896 52. Utami, K. H. *et al.* Impaired development of neural-crest cell-derived organs and intellectual disability
897 caused by MED13L haploinsufficiency. *Hum. Mutat.* **35**, 1311–1320 (2014).
- 898 53. Payne, J. L. & Wagner, A. Mechanisms of mutational robustness in transcriptional regulation. *Front. Genet.*
899 **6**, 322 (2015).
- 900 54. Akincilar, S. C., Unal, B. & Tergaonkar, V. Reactivation of telomerase in cancer. *Cell. Mol. Life Sci.* **73**,
901 1659–1670 (2016).
- 902 55. Gao, F. *et al.* Hes1 is involved in the self-renewal and tumorigenicity of stem-like cancer cells in colon
903 cancer. *Sci. Rep.* **4**, 3963 (2014).
- 904 56. Moriyama, M. *et al.* Notch signaling via Hes1 transcription factor maintains survival of melanoblasts and
905 melanocyte stem cells. *J. Cell Biol.* **173**, 333–339 (2006).
- 906 57. Zabierowski, S. E. *et al.* Direct reprogramming of melanocytes to neural crest stem-like cells by one defined
907 factor. *Stem Cells* **29**, 1752–1762 (2011).
- 908 58. Bedogni, B. Notch signaling in melanoma: interacting pathways and stromal influences that enhance Notch
909 targeting. *Pigment Cell Melanoma Res.* **27**, 162–168 (2014).
- 910 59. Watson, M., Holman, D. M. & Maguire-Eisen, M. Ultraviolet radiation exposure and its impact on skin
911 cancer risk. *Semin. Oncol. Nurs.* **32**, 241–254 (2016).
- 912 60. Baxter, J. S. *et al.* Capture Hi-C identifies putative target genes at 33 breast cancer risk loci. *Nat. Commun.* **9**,
913 (2018).
- 914 61. Fulco, C. P. *et al.* Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR
915 perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).

916

917 **Figure titles and legends**

918 **Figure 1. MPRA analysis of candidate variants and risk loci identified in melanoma GWAS.** (A) Overall
919 workflow from melanoma GWAS summary statistics³ to candidate variants for MPRA analysis. (B) MPRA
920 design. Oligo libraries were synthesized using 145 bp of sequence encompassing each variant with reference or
921 alternative allele in both forward and reverse (F & R) directions, which are associated with 12-bp barcodes (20
922 tags per unique sequence). For each variant, a scrambled sequence of 145 bp test sequence was also included and
923 associated with 16 tag sequences (using forward direction and reference allele) as a null. Libraries were cloned
924 into luciferase constructs and then transfected into UACC903 melanoma cells or melanocyte cells to generate
925 expressed RNA tag libraries. Both input DNA and RNA libraries were sequenced to assess the tag counts
926 associated with the test sequences. A summary of MPRA results in UACC903 melanoma cells (C) and
927 melanocyte cells (D). *FDR* values for allelic transcriptional activity of each variant measured by MPRA are
928 displayed in Manhattan plots. Horizontal lines represent an *FDR* cutoff of 0.01 ($-\log_{10}(FDR) = 2$), and variants
929 displaying significant allelic transcriptional activity are shown separately for melanoma (red) and melanocyte
930 (blue) experiments. Bar graphs under the Manhattan plots show the percentage of variants displaying significant
931 allelic transcriptional activity ($FDR < 0.01$, red for melanoma and blue for melanocyte; ≥ 0.01 , grey) by
932 melanoma GWAS loci ordered by chromosomes (defined in **Table S1**). Bar graphs on the right present the
933 summarized statistics as to the numbers of tested vs. MPRA-significant variants in total or by locus for each cell
934 type. **Notes:** LLR = log likelihood ratio; *FDR* = false discovery rate; ref = reference allele; alt = alternative
935 alleles.

936

937 **Figure 2. Integrative scores for prioritizing plausible candidate variants.** (A) The functional (MPRA, motif
938 prediction, and chromatin annotations) and fine-mapping features (credible sets and posterior possibility, PIP)
939 were incorporated to evaluate the candidate variants. For each locus, the variant(s) with the highest combined
940 score were assigned as Tier-1 variants (green) and those with the second-highest scores (no less than 70% of the
941 highest score) were assigned as Tier-2 variants (yellow). (B) The overall prioritization from MPRA-significant
942 variants to Tier-1 (green) and Tier-2 (yellow) variants are shown. Each bar represents a melanoma GWAS locus.
943 (C) Examples of melanoma GWAS loci with known functional variants or substantial prioritization performance.
944 Hits are given 1 score for the variants (MPRA, blue dots; chromatin annotation, light green dots; motif, yellow
945 dots; and fine-mapping, light red dots). Strong hits are given 2 scores (MPRA, purple dots; chromatin annotation,
946 dark green dots; motif, orange dots; and fine-mapping, dark red dots). No hits are shown with gray dots.
947 Definition of hits and strong hits are presented in **Methods**. No dots (gray lines) are presented if functional/fine-
948 mapping features are unavailable for the given variant.

949

950 **Figure 3. Linking the candidate variants to the target genes via QTLs.** MPRA-significant variants in
951 melanoma or melanocyte datasets are presented by locus if they display genome-wide significant eQTL or
952 meQTL *P*-values for one of the significant genes (eGenes with gene names in black, dark green square for
953 matched direction in eQTL and MPRA and light green square for not matched direction) or the nearest gene based
954 on methylation sites (meGenes with gene names in blue, blue square) in melanocyte or melanoma datasets. The
955 loci were presented based on the total number of eGenes and meProbes, with 1 in **(A)**, 2 (including those with the
956 same assigned gene name for eGene and meGene) in **(B)**, 3 or more in **(C)** and **(D)**. Variants are ordered by
957 integrative scores for each locus with tier-1 variants shown in green, tier-2 variants in yellow, and tier-3 variants
958 in black. Asterisks next to the gene names indicate the genes identified in GWAS-QTL colocalization, TWAS, or
959 MWAS.
960

961 **Figure 4. CRISPRi using gRNAs targeting prioritized variants in UACC903 cells.** (A) gRNA plasmids were
962 packed into lentiviral particles in HEK293T cells and then transduced into dCas9-Zim3 expressing UACC903
963 cells. Twenty-four hours after infection, transduced UACC903-dCas9-ZIM3 cells were selected with 2 µg/ml of
964 puromycin. Survived cells were harvested 48 hours after puromycin selection for RNA isolation. (B) CRISPRi
965 using three gRNAs (G1, G2, and G3) targeting the region (genomic coordinates in hg38) surrounding rs4384. The
966 levels of *MAFF* transcript (*GAPDH*-normalized) are shown as fold change over those from nontargeting gRNA.
967 Four biological replicates of n = 6 were combined (total n =24). Error bars refer to the standard error. *P*-values are
968 calculated by two-sample t-test (two-sided) with unequal variance from non-targeting controls (dotted red lines).
969 (C) CRISPRi using three gRNAs targeting the region surrounding rs2111398. The levels of
970 *GPR5CA/HEBP1/EMP1* transcripts (*GAPDH*-normalized) are shown as fold change over those from nontargeting
971 gRNA.

972

973 **Figure 5. Cell-type specificity at the level of variants and genes.** (A) overall analysis from 285 MPRA-
974 significant variants to 77 variants only significant in melanoma (red dots represent melanoma MPRA $FDR < 0.01$)
975 and 151 variants only significant in melanocyte (blue dots represent melanocyte MPRA $FDR < 0.01$). Three
976 criteria were applied to further prioritize variants with cell-type specificity, including MPRA $FDR < 10^{-9}$, putative
977 role of activator, and TF identified as the high-expressed DEGs in the specific cell type. Variants meeting at least
978 one criterion were further prioritized. Representative loci with variants showing cell-type specificity in melanoma
979 (B), both (C), or melanocyte (D) are shown. Asterisks next to variant IDs represent the number of criteria that are
980 met for that variant. (E) Variants meeting at least one criterion in the MPRA of melanoma cells are presented if
981 they are also a genome-wide significant eQTL (eGenes, green) or meQTL (meGenes, blue) in the TCGA
982 melanoma QTL dataset. The variants are grouped and ordered by GWAS loci with locus IDs shown at the top of
983 each group of variants. (F) Variants meeting at least one criterion in the MPRA of melanocyte cells are presented
984 if they are also a genome-wide significant eQTL (eGenes, green) or meQTL (meGenes, blue) in the melanocyte
985 QTL dataset. The variants are grouped and ordered by GWAS loci with locus IDs shown at the top of each group
986 of variants.
987