The American Journal of Human Genetics, Volume *109*

Supplemental information

Massively parallel reporter assays and variant scoring identified functional variants and target genes for melanoma loci and highlighted cell-type specificity

Erping Long, Jinhu Yin, Karen M. Funderburk, Mai Xu, James Feng, Alexander Kane, Tongwu Zhang, Timothy Myers, Alyxandra Golden, Rohit Thakur, Hyunkyung Kong, Lea Jessop, Eun Young Kim, Kristine Jones, Raj Chari, Mitchell J. Machiela, Kai Yu, Melanoma Meta-Analysis Consortium, Mark M. Iles, Maria Teresa Landi, Matthew H. Law, Stephen J. Chanock, Kevin M. Brown, and Jiyeon Choi

SUPPLEMENTAL FIGURES

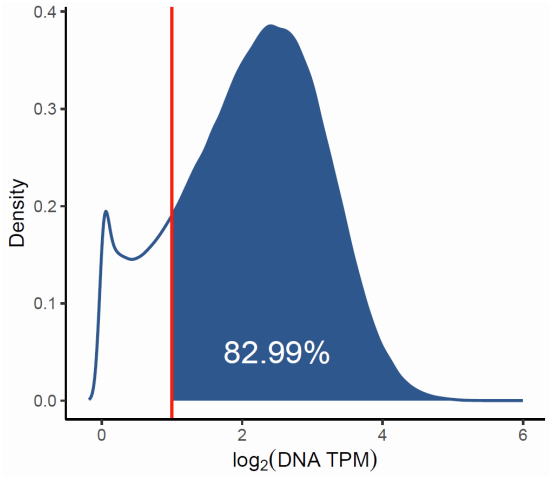

Figure S1. Tag count distribution in the input DNA libraries is shown as log_2(DNA TPM) density. Red line denotes log_2 (DNA TPM) =1 or DNA TPM = 2, which was used as a QC cutoff. Percentage of tags with DNA TPM ≥ 2 are 82.99% of all the detected tags.

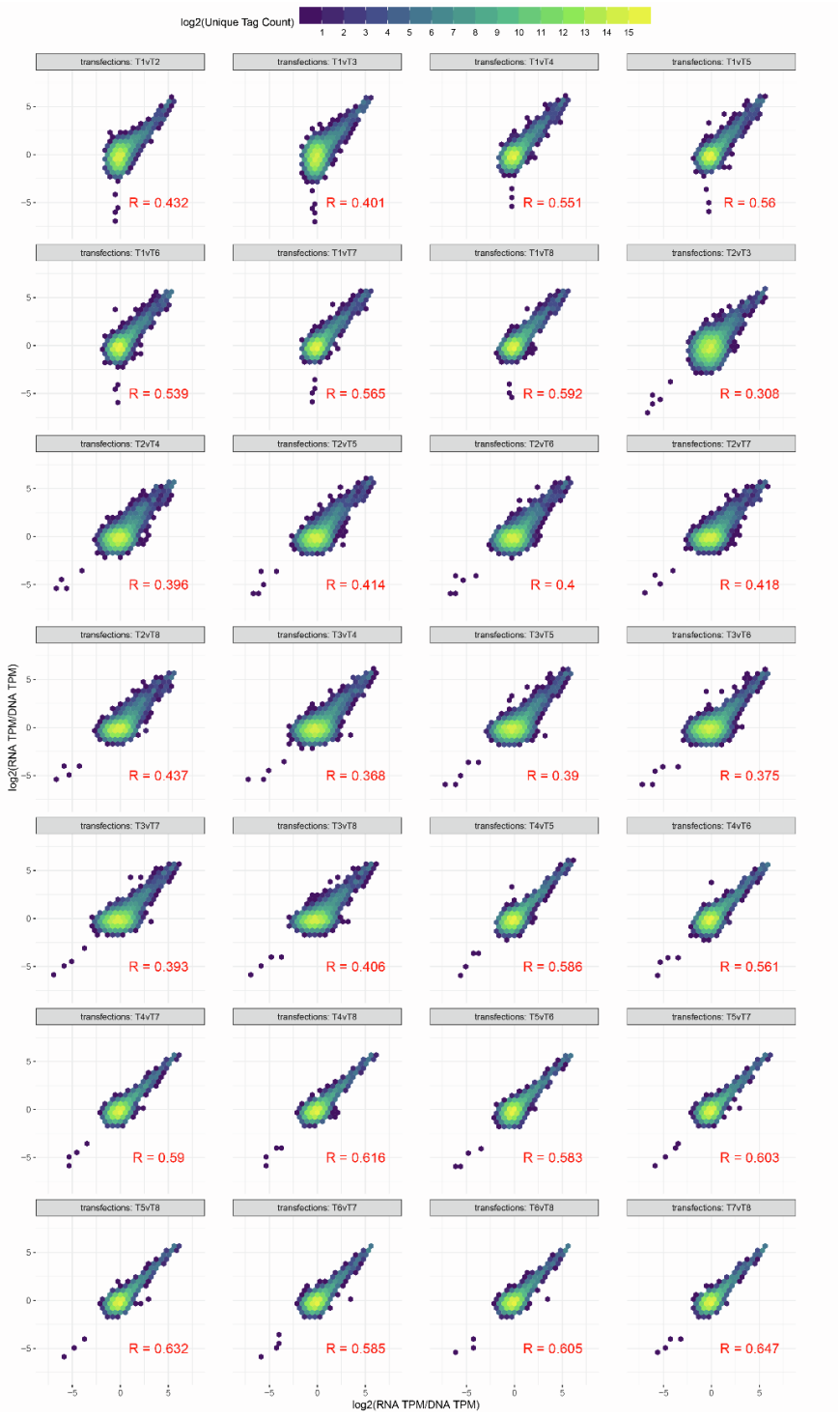

Figure S2. Correlation plots among replicates in melanoma MPRA by tags. Inter-transfection correlation of tag-level normalized tag counts for each tag between transfection replicates are shown for melanoma cells. Log_2(RNA TPM/DNA TPM) value for each tag before QC are plotted with log_2(tag count) shown as color-coded density level. Pair-wise Pearson correlation coefficients are shown in red (R values).

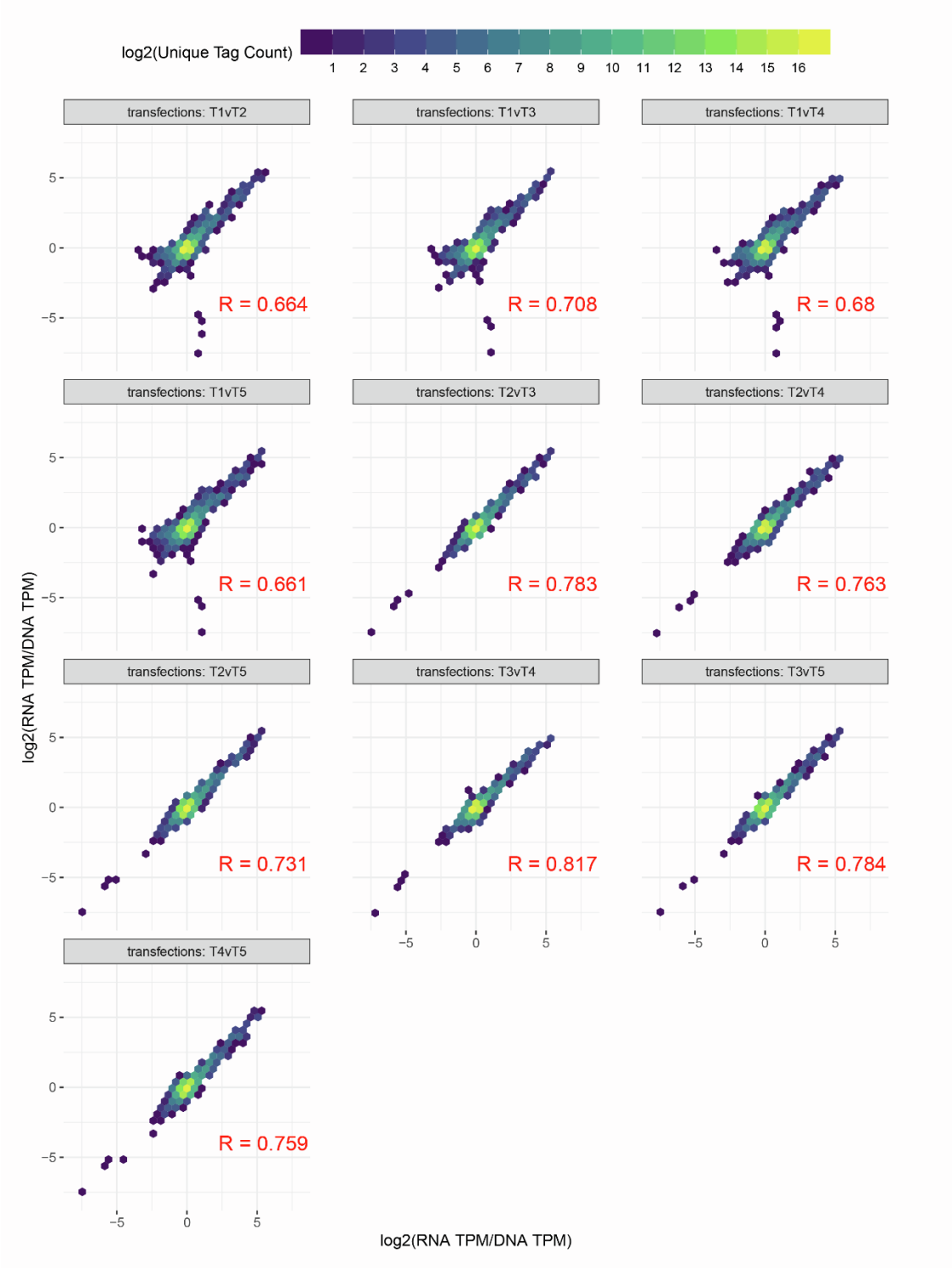

Figure S3. Correlation plots among replicates in melanocyte MPRA by tags. Inter-transfection correlation of tag-level normalized tag counts for each tag between transfection replicates are shown for melanocyte cells. Log_2(RNA TPM/DNA TPM) value for each tag before QC are plotted with log_2(tag count) shown as color-coded density level. Pair-wise Pearson correlation coefficients are shown in red (R values).

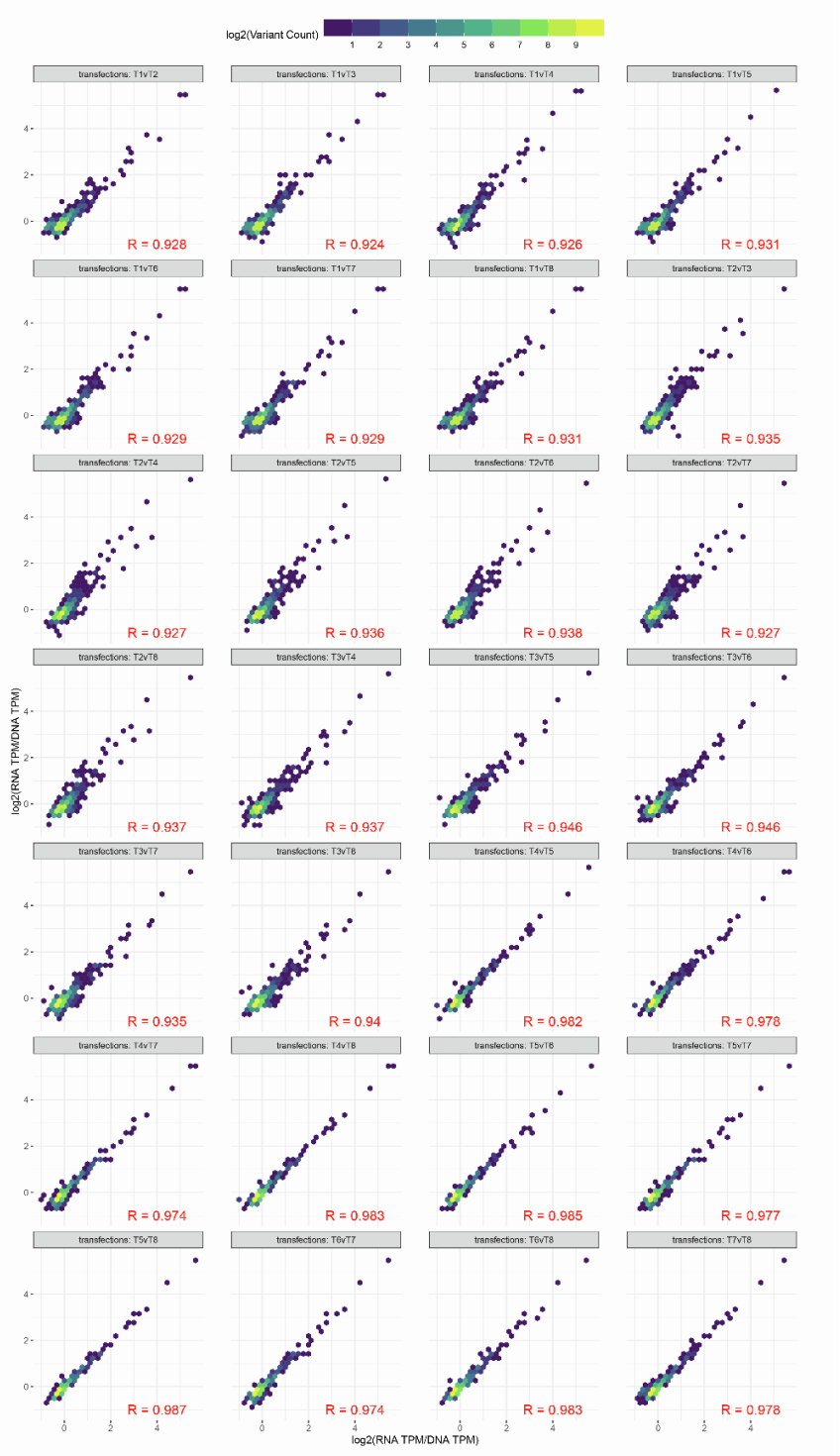

Figure S4. Correlation plots among replicates in melanoma MPRA by variants. Inter-transfection correlation of variant-level normalized tag counts for each variant between transfection replicates are shown for melanoma cells. Log_2(RNA TPM/DNA TPM) value for each variant before QC are plotted with log_2(tag count) shown as color-coded density level. Pair-wise Pearson correlation coefficients are shown in red (R values).

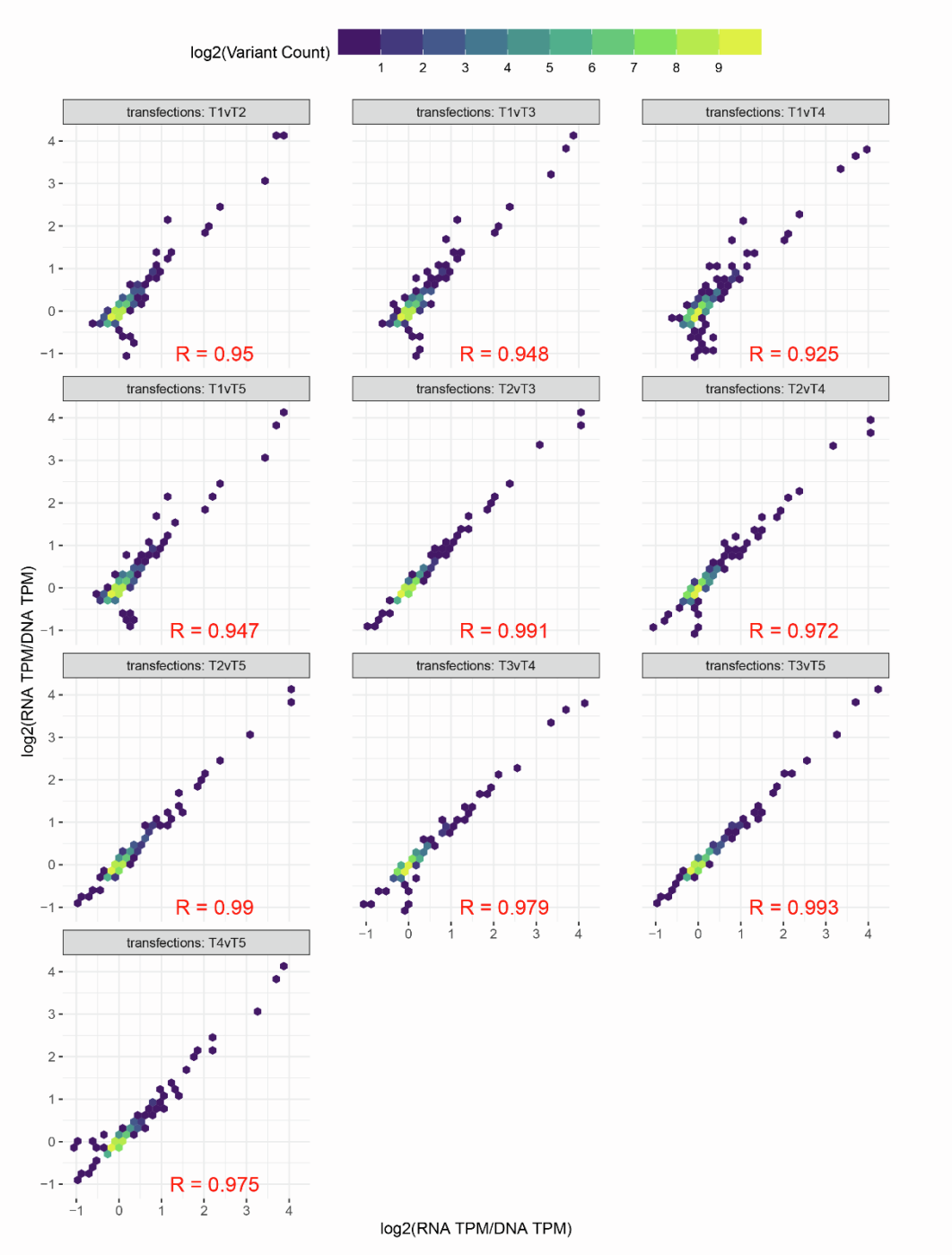

Figure S5. Correlation plots among replicates in melanocyte MPRA by variants. Inter-transfection correlation of variant-level normalized tag counts for each variant between transfection replicates are shown for melanoma cells. Log_2(RNA TPM/DNA TPM) value for each variant before QC are plotted with log_2(tag count) shown as color-coded density level. Pair-wise Pearson correlation coefficients are shown in red (R values).

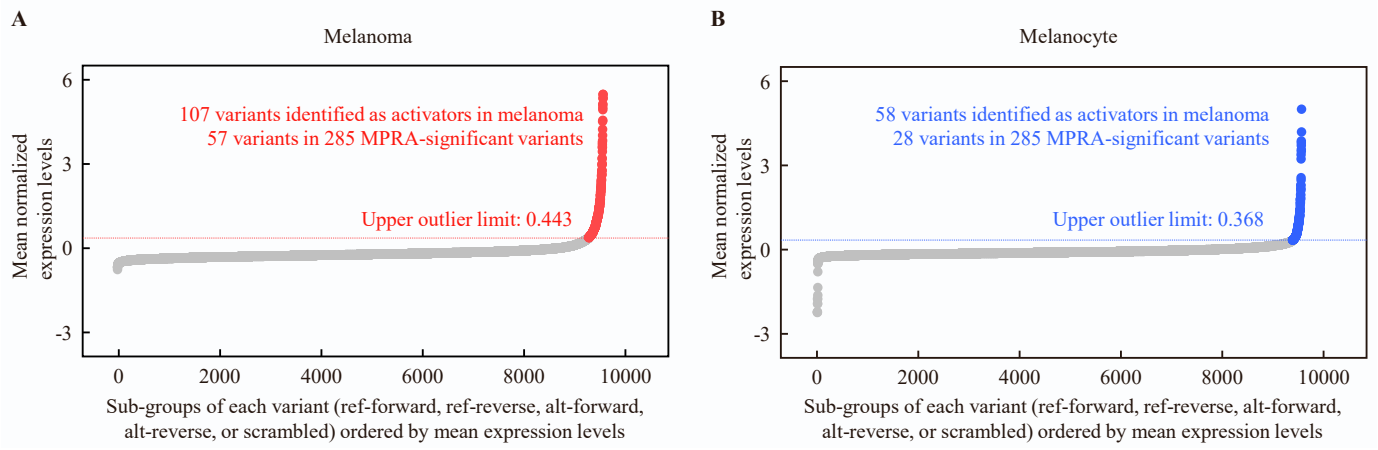

Figure S6. Putative function of activators in UACC903 melanoma (A) and C283T melanocyte (B). The overall distribution of the normalized expression levels of all tags by variants including reference or alternative alleles and scrambled sequences are presented. Each dot represents the mean normalized expression level of all the tags (up to 20) in one of five subgroups of each variant (ref-forward, ref-reverse, alt-forward, alt-reverse, or scrambled). The upper outlier limit is defined by Q3 + 3 x IQ), where Q3 is 75th percentiles and the interquartile range (IQ) is Q3 - 25th percentiles. The variants with normalized expression levels of one or more of these sub-groups higher than the upper limits were assigned as activators.

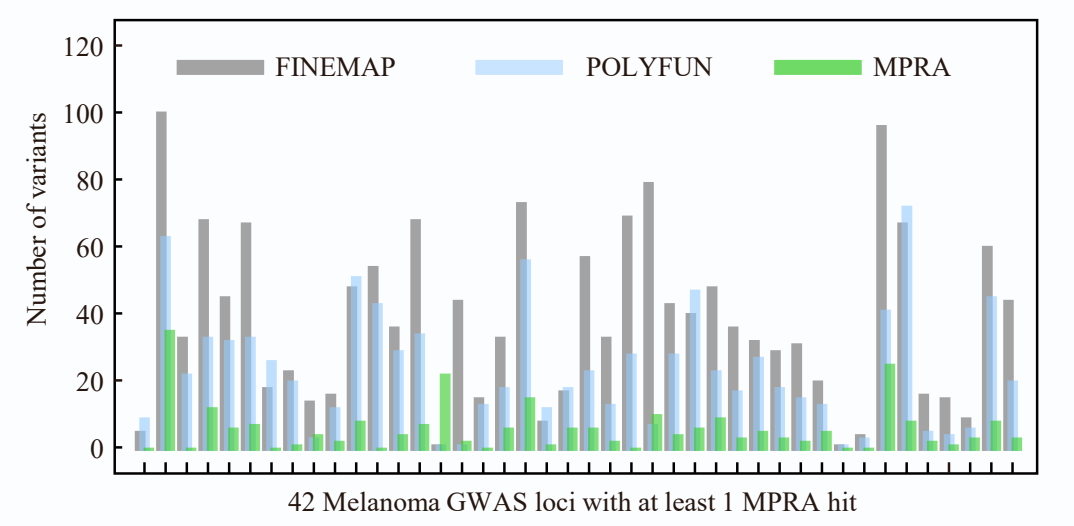

Figure S7. MPRA-significant variants (green) represented a considerable prioritization compared to the 95% credible sets from FINEMAP (grey) and POLYFUN (blue). Melanoma GWAS loci were ordered by chromosomes (defined in **Table S1**).

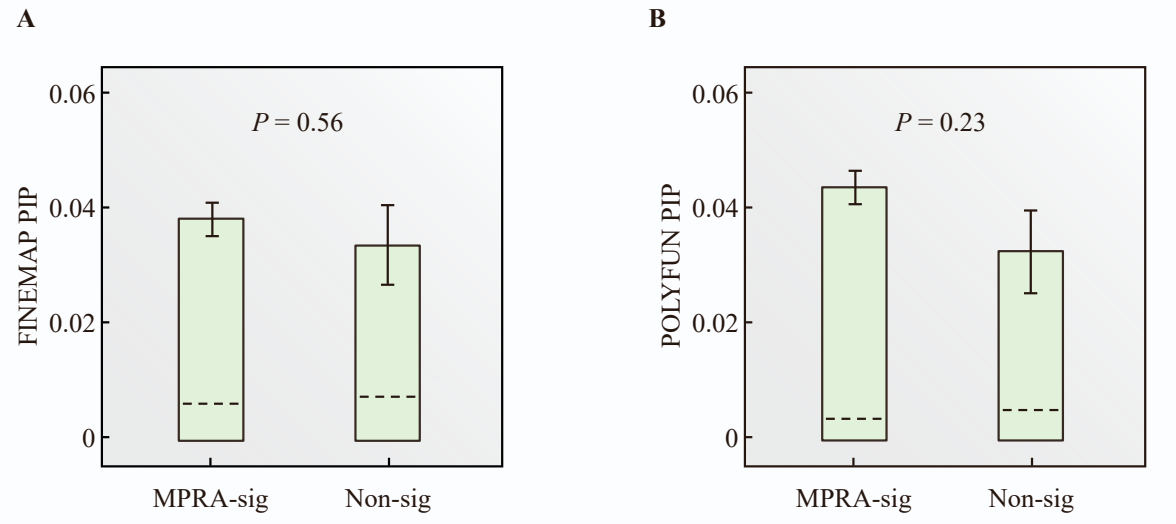

Figure S8. Probability score (PIP) of MPRA-significant variants compared to non-significant variants in FINEMAP (A) and POLYFUN (B). Error bar indicates the standard error. *P* values are based on two-tailed unequal variance t-test.

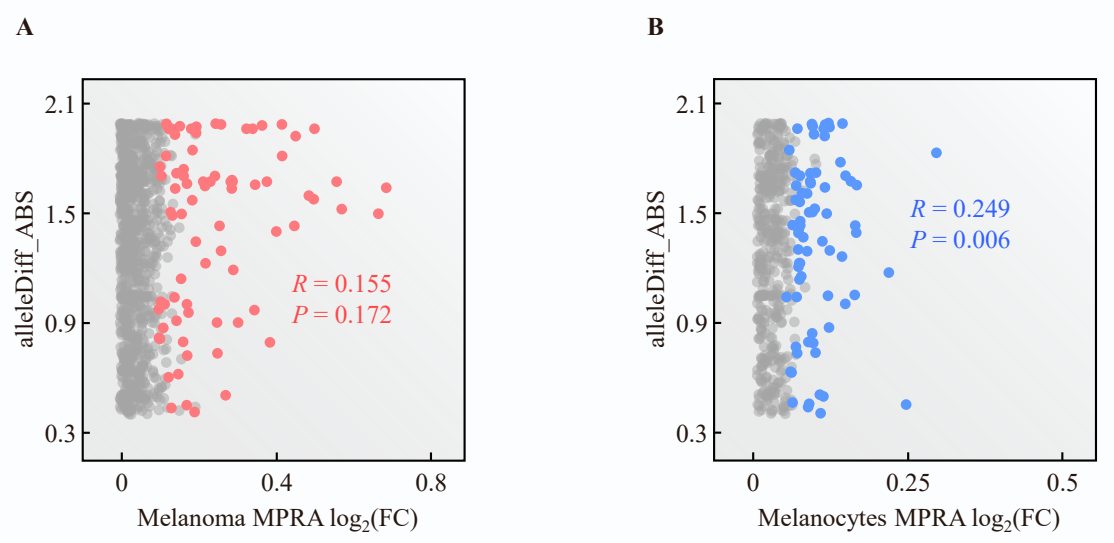

Figure S9. Correlation between predicted allelic binding scores and allelic transcriptional activities among MPRA-significant variants in melanoma (A, red dots) and melanocyte (B, blue dots). The Y-axis displays the absolute values of allelic binding scores between reference and alternative alleles (highest scores for each variant). The X-axis displays the absolute values of log2-transformed fold change (FC) between MPRA mean TPM ratios of reference and alternative alleles. Considering the skewed distribution of log2(FC), the *R* scores and *P* values were calculated by the Spearman correlation.

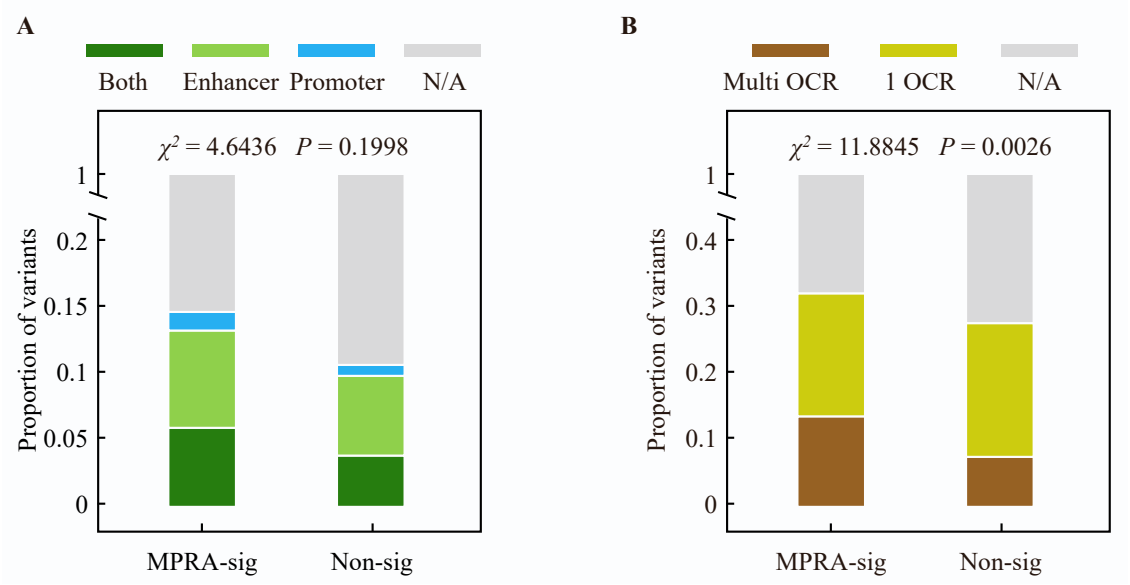

Figure S10. Enrichment of MPRA-significant variants in chromatin features compared to non-significant variants. (A) The proportion of variants located in genomic regions annotated in melanocyte or melanoma datasets as enhancer, promoter, both enhancer and promoter, or none (N/A) are shown. **(B).** The proportion of variants located in open chromatin regions (OCR) assessed in a total of 4 melanocyte or melanoma datasets (multiple datasets, one dataset, or none). The types and sources of genomic annotations (promoter, enhancer, and open chromatin) used are described in "Functional annotations" section of the **Methods**. Enrichment *P*-values are based on Chi-squared test.

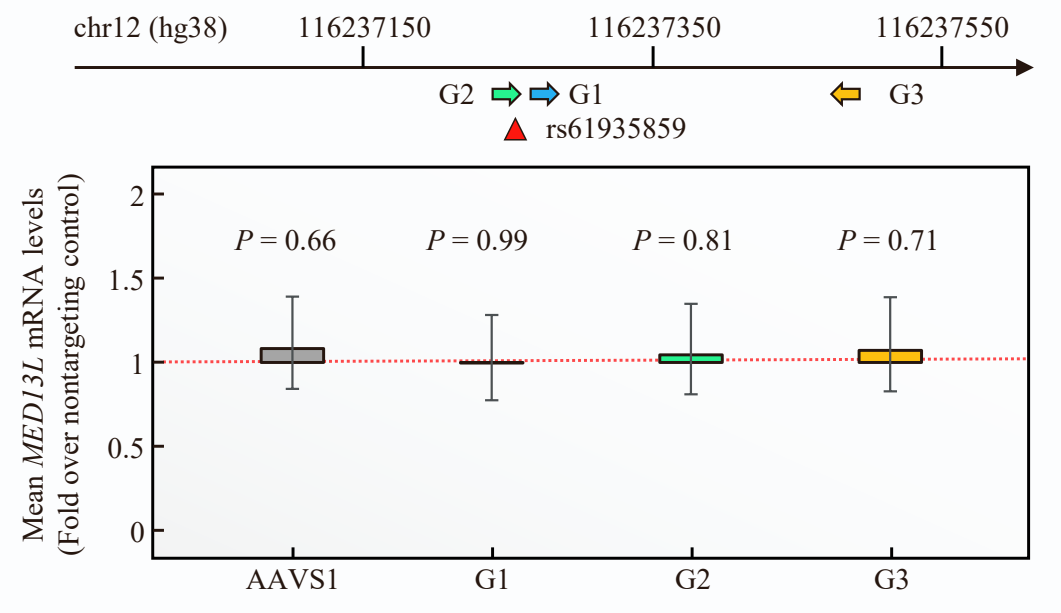

Figure S11. CRISPRi using gRNAs targeting rs61935859 in UACC903 cells. CRISPRi using three gRNAs (G1, G2, and G3) targeting the region (genomic coordinates in hg38) surrounding rs61935859. The levels of *MED13L* transcript (*GAPDH*-normalized) are shown as fold change over those from nontargeting gRNA. Four biological replicates of n = 3-6 were combined (total n = 21). Error bars refer to the standard error. *P*-values are calculated by two-sample t-test (two-sided) with unequal variance from non-targeting controls (dotted red lines).

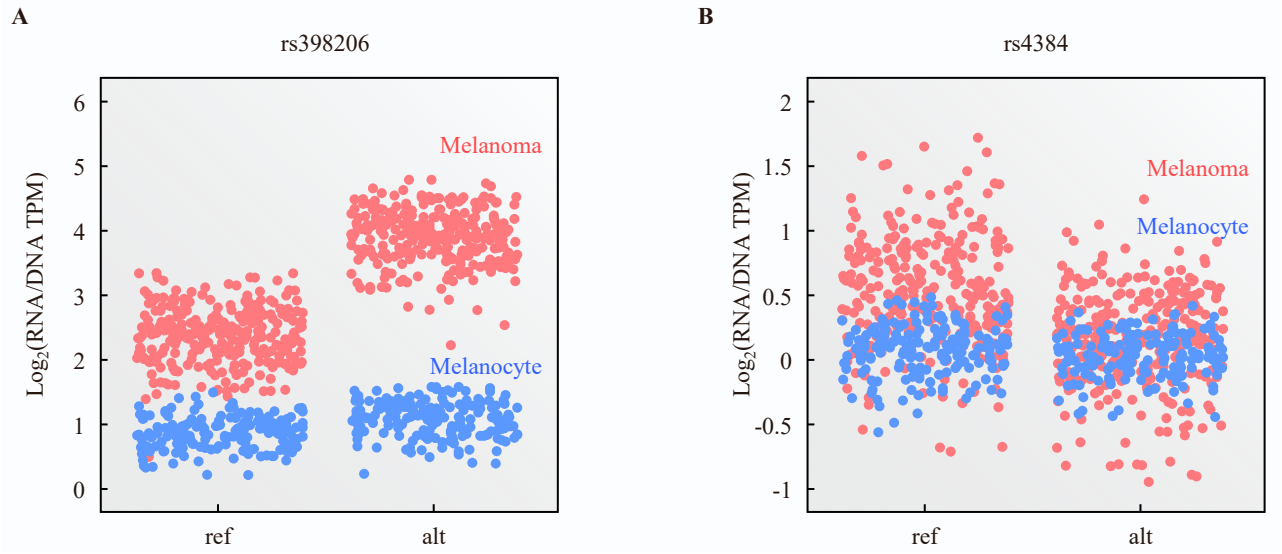

Figure S12. MPRA-significant variants, (A) rs398206 (B) rs4384, showing significant cell-type specificity. Each dot represents a unique tag in the context of UACC903 melanoma (red) and C283T melanocyte (blue).

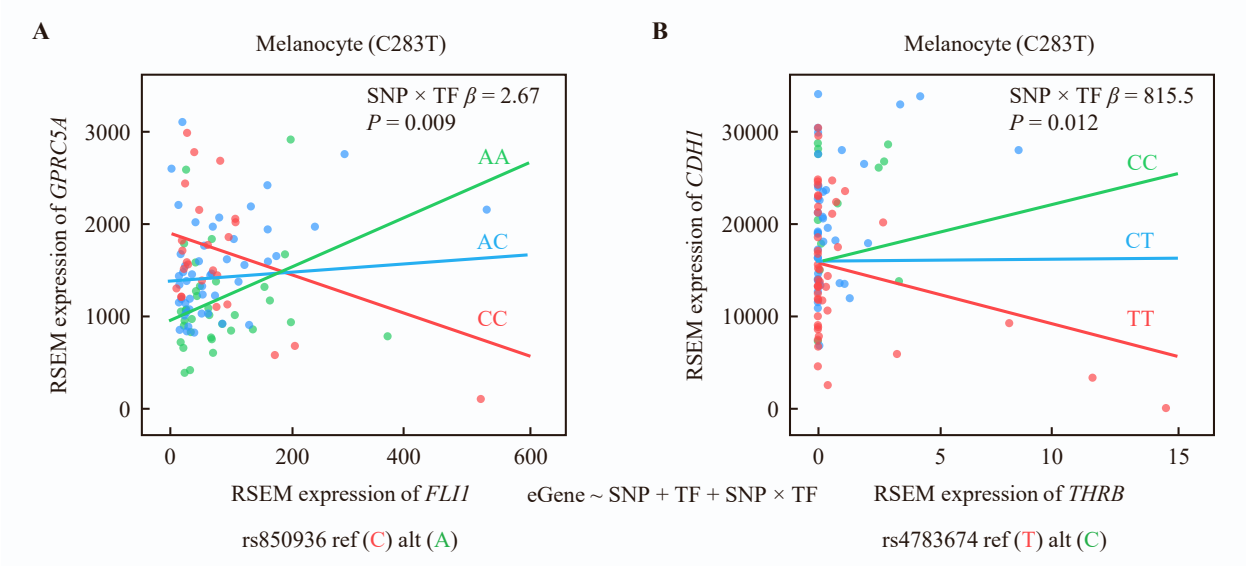

Figure S13. Two examples of variant-TF-eGene interactions. Variant-TF-eGene interactions are identified using a multiple linear regression interaction model of RSEM expression of eGene, TF, and variant genotype. Using 'sjPlot' package 'plot_model' function, linear lines are predicted and plotted separately for each genotype, with SNP × TF interaction beta (refers to alternative allele) and *FDR* value (corrected by the Benjamini-Hochberg method) are displayed on the chart.

SUPPLEMENTAL REFERENCES

1. Landi, M.T., Bishop, D.T., MacGregor, S., Machiela, M.J., Stratigos, A.J., Ghiorzo, P., Brossard, M., Calista, D., Choi, J., Fargnoli, M.C., et al. (2020). Genome-wide association meta-analyses combining multiple risk phenotypes provide insights into the genetic architecture of cutaneous melanoma susceptibility. Nat. Genet. *52*, 494–504.
2. Choi, J., Zhang, T., Vu, A., Ablain, J., Makowski, M.M., Colli, L.M., Xu, M., Hennessey, R.C., Yin, J., Rothschild, H., et al. (2020). Massively parallel reporter assays of melanoma risk variants identify MX2 as a gene promoting melanoma. Nat. Commun. *11*, 2718.
3. Zhang, T., Choi, J., Dilshat, R., Einarsdóttir, B.Ó., Kovacs, M.A., Xu, M., Malasky, M., Chowdhury, S., Jones, K., Bishop, D.T., et al. (2021). Cell-type-specific meQTLs extend melanoma GWAS annotation beyond eQTLs and inform melanocyte gene-regulatory mechanisms. Am. J. Hum. Genet. *108*, 1631–1646.