# Massively parallel reporter assays combined with cell-type specific eQTL informed multiple melanoma loci and identified a pleiotropic function of HIV-1 restriction gene, *MX2,* in melanoma promotion

Jiyeon Choi[1#], Tongwu Zhang[1#], Andrew Vu[1], Julien Ablain[2], Matthew M Makowski[3], Leandro M Colli[1], Mai Xu[1], Harriet Rothschild[2], Cathrin Gräwe[3], Michael A Kovacs[1], Myriam Brossard[4], John Taylor[5], Bogdan Pasaniuc[6], Raj Chari[7], Stephen J Chanock[1], Clive J Hoggart[8], Florence Demenais[4], Jennifer H Barrett[5], Matthew H Law[9], Mark M Iles[5], Kai Yu[1], Michiel Vermeulen[3], Leonard I Zon[2], Kevin M Brown[1*]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, 20892, USA, [2]Stem Cell Program and Division of Hematology/Oncology, Boston Children's Hospital and Dana-Farber Cancer Institute, Boston, MA, 02115, USA, [3]Department of Molecular Biology, Oncode Institute, Radboud University, Nijmegen, Netherlands, [4]Genetic Variation and Human Diseases Unit, Institut National de la Santé et de la Recherche Médicale (INSERM), Université Paris Diderot, Paris, France, [5]Leeds Institute for Data Analytics, School of Medicine, University of Leeds, Leeds, UK, [6]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, 90024, USA, [7]Genome Modification Core, Frederick National Lab for Cancer Research, National Cancer Institute, Frederick, MD, 21701, USA, [8]Department of Medicine, Imperial College London, London, UK, [9]Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia, [#]These authors contributed equally to this work, *Correspondence should be addressed to K.M.B at: Email: kevin.brown3@nih.gov

## Abstract

Genome-wide association studies (GWAS) have identified ~20 melanoma susceptibility loci. To identify susceptibility genes and variants simultaneously from multiple GWAS loci, we integrated

massively-parallel reporter assays (MPRA) with cell type-specific epigenomic data as well as melanocyte-specific expression quantitative trait loci (eQTL) profiling. Starting from 16 melanoma loci, we selected 832 variants overlapping active regions of chromatin in cells of melanocytic lineage and identified 39 candidate functional variants displaying allelic transcriptional activity by MPRA. For four of these loci, we further identified four colocalizing melanocyte *cis*-eQTL genes (*CTSS*, *CASP8*, *MX2*, and *MAFF*) matching the allelic activity of MPRA functional variants. Among these, we further characterized the locus encompassing the HIV-1 restriction gene, *MX2*, on chromosome band Chr21q22.3 and validated a functional variant, rs398206, among multiple high LD variants. rs398206 mediates allelic transcriptional activity via binding of the transcription factor, YY1. This allelic transcriptional regulation is consistent with a significant *cis*-eQTL of *MX2* in primary human melanocytes, where the melanoma risk-associated A allele of rs398206 is correlated with higher *MX2* levels. Melanocyte-specific transgenic expression of human *MX2* in a zebrafish model demonstrated accelerated melanoma formation in a *BRAF*$^{V600E}$ background. Thus, using an efficient scalable approach to streamline GWAS follow-up functional studies, we identified multiple candidate melanoma susceptibility genes and variants, and uncovered a pleiotropic function of *MX2* in melanoma susceptibility.

## Introduction

A series of genome-wide association studies (GWAS) over the past decade have identified about twenty genomic loci associated with cutaneous melanoma[1-10], highlighting the genetic contribution to melanoma susceptibility in the general population. Some of these loci represent genes or regions implicated in melanoma-associated traits e.g., pigmentation phenotypes[11-15] and nevus count[5,16,17]. Other than these loci, however, underlying mechanisms of genetic susceptibility to melanoma in the general population is less well understood. For a small number of these loci, extensive characterization of susceptibility genes and variants under

the GWAS peaks have led to new insights into molecular pathways underlying melanoma susceptibility. *PARP1,* located in the Chr1q42.1 melanoma locus[8], was shown to be a susceptibility gene that has tumor-promoting roles in early events of melanomagenesis through its regulation of melanocyte master transcription factor and oncogene, *MITF*[18], while a functional variant at a multi-cancer locus on Chr5p15.33 was characterized highlighting the role of *TERT* in cancer susceptibility including in melanoma[19]. Still, the molecular mechanisms underlying the majority of common melanoma risk loci remain unexplained.

Recent advances in sequencing technologies have enabled a number of classical molecular assays to be conducted at a large scale. Massively Parallel Reporter Assays (MPRA) scale up conventional luciferase reporter assays for testing transcriptional activities of DNA elements, facilitating evaluation of tens of thousands of different short sequences at the same time in cells, which are then deconvoluted by massively parallel sequencing[20-22]. Incorporation of this approach is particularly attractive for GWAS functional follow-up studies, as 1) linkage disequilibrium (LD) limits statistical fine-mapping and leaves numerous variants as potential functional candidates, and 2) many trait-associated variants are hypothesized to contribute to allelic gene expression through *cis*-regulatory mechanisms that can be tested by reporter assays. Therefore, direct assessment of allelic differences in transcriptional regulation could help prioritize likely functional variants among multiple variants tied by LD. For example, a recent study adopted MPRA to test 2,756 variants from 75 GWAS loci for red blood cell traits and identified 32 functional variants from 23 loci[20].

In addition, expression quantitative trait loci (eQTL) analysis can be a powerful approach for identifying susceptibility genes from GWAS loci, as it informs on genes for which expression levels are correlated with trait-associated variants. While there are a number of publicly available eQTL datasets using tissues representing different human organs including those through the GTEx project[23], most of them are based on bulk tissue samples (e.g., skin tissues)

as opposed to individual cell types. Importantly, melanomas arise from melanocytes, but they account for less than 5% of a typical skin biopsy. To dissect cell-type specific gene expression regulation implicated in melanoma predisposition, a melanocyte eQTL dataset using primary cultures of melanocytes from 106 individuals was established and mapped six melanoma GWAS loci (30% of all the loci) to melanocyte eQTLs[24]. This dataset outperformed eQTLs from bulk skin tissues, other tissue types from GTEx, and melanoma tumors[24], highlighting the utility of cell-type specific eQTL dataset for functional follow-up of GWAS regions.

In this study, we combine MPRA and cell-type specific melanocyte eQTL to scale up the functional annotation process for melanoma GWAS loci and nominate the best candidates for testing in a zebrafish model. Our approach identified a functional risk variant that increases the level of an HIV-1 restriction gene, *MX2,* in cells of melanocytic lineage; subsequent expression of *MX2* in melanocytes of a zebrafish melanoma model accelerated melanoma formation.

## Results

### Massively parallel reporter assays identified melanoma-associated putative functional variants

To identify functional melanoma-associated variants displaying allelic transcriptional function, we used the MPRA approach. Among 20 genome-wide significant melanoma loci from the most recent GWAS meta-analysis[1], we prioritized 16 loci where a potential *cis*-regulatory mechanism could be hypothesized, excluding four pigmentation-associated loci previously explained by functional protein coding variants (*MC1R*, *SLC45A2*, and *TYR*[11-14]) or shown not to be expressed in melanocytes (*ASIP*[15]). To comprehensively analyze genetic signals from these loci, we then performed statistical fine-mapping using the HyperLasso[25] approach. The fine-mapping nominated additional independent signals (**Supplementary Table 1**), from which we selected 30 variants, adding to the 16 lead SNPs from the initial meta-analysis results[1]. To

prioritize melanoma-associated variants to test by MPRA, we first selected 2,748 variants that are in LD ($r^2$ > 0.4) with these 46 primary and secondary lead SNPs (**Methods; Supplementary Fig 1; Supplementary Table 2**). Among them, we further prioritized 832 variants that overlap potentially functional melanoma-relevant genomic signatures, namely, open chromatin regions and promoter/enhancer histone marks in primary melanocytes and/or melanoma short term cultures[26] (**Supplementary Table 3-4**; **Methods**; www.encodeproject.org; www.roadmapepigenomics.org). We then constructed MPRA libraries for these 832 variants using methods adopted from previous studies[20-22,27]. A 145 bp genomic sequence encompassing the risk or protective allele of each variant was tested for their potential as an enhancer or promoter element in luciferase constructs. For each variant, a scrambled sequence for its core 21 bases was also tested as a null (**Supplementary Fig 2**; **Methods**). Transcribed output of tag (barcode) sequences associated with each tested DNA element were then measured by sequencing, after transfections into a melanoma cell line (UACC903) to represent melanoma-specific *trans*-acting factors and the HEK293FT cell line to obtain maximum transfection efficiency. From these data, we initially observed significantly high correlation of transcriptional activities among replicates, and further applied a conservative quality control measure for downstream analyses (**Methods**; **Supplementary Figs 3-7; Supplementary Table 5**).

To nominate variants displaying allelic transcriptional activity, we focused on those displaying significant difference between two alleles (FDR < 0.01), and then further selected those with either allele displaying a significant departure from the null (scrambled core sequence; FDR < 0.01) (**Supplementary Fig 3**). After applying these cutoffs, 39 of the 832 tested variants (~4.7%) qualified as displaying allelic transcriptional activity in the UACC903 melanoma cell dataset alone as well as in the combined total dataset (**Methods**; **Supplementary Fig 8A**; **Supplementary Table 6**). These candidate functional variants are

from 14 melanoma GWAS loci with 1-9 variants per locus (median 1.5 variants), which represents 2-33% of tested variants per locus (**Fig 1**; **Supplementary Table 7; Supplementary Fig 9**). Transcriptional activities of these 39 variants were significantly higher than those of negative controls (8 variants of high LD with the lead SNP but located in non-DHS/non-promoter/enhancer histone mark in melanocytes/melanoma cells; $P < 2.2e-16$, effect size = 0.137; Mann-Whitney U test; **Supplementary Fig 8B**) as well as the rest of the variants (non-significant variants; $P < 2.2e-16$, effect size = 0.109). These 39 variants displayed 1.13 to 3.49-fold difference in transcriptional activity between two alleles (UACC903 cells; **Supplementary Table 6**). We then asked if the observed allelic differences from MPRA are in part due to differential binding of transcription factors. For this, we predicted allelic transcription factor binding affinity of each tested variant using motifbreakR[28]. When the allelic differences were compared, the MPRA-significant variants displayed a higher level of correlation between MPRA allelic activities and predicted allelic motif scores (Pearson $r = 0.24$, $P = 0.149$, $n = 39$; **Supplementary Fig 10A**) compared to non-significant ones (Pearson $r = -0.023$, $P = 0.556$, $n = 793$). We then performed additional statistical fine-mapping of melanoma GWAS data to obtain probability scores for melanoma-associated variants using PAINTOR[29], which integrates association strength with genomic functional annotation. To incorporate melanoma-relevant annotations to this fine-mapping, we included select functional annotations of primary melanocytes (melanocyte-specific expressed genes from our melanocyte dataset, melanocyte enhancers, TF-binding sites, and histone marks from ENCODE and Roadmap database). When overlaid with these probability scores, the 39 significant MPRA variants (FDR < 0.01) displayed the highest median probability score compared to other variant groups with varying FDR cutoffs, which was a 2.12-fold enrichment over all the tested variants with probability scores (**Supplementary Fig 10B**). These data demonstrated that MPRA can quickly narrow down to a small number of plausible functional candidate variants from melanoma GWAS loci using allelic transcriptional activity.

## Integration of MPRA and melanocyte eQTLs identified functional variants and genes from multiple melanoma loci

To prioritize functional variants that contribute to melanoma risk through regulation of nearby gene expression, we turned to cell-type specific melanocyte eQTL data from 106 individuals[24]. 597,335 significant *cis*-eQTL SNPs (+/-1 Mb of TSS, FDR < 0.05, not LD-pruned) were identified in this dataset, with 6 of 20 melanoma GWAS loci displaying significant co-localization/TWAS[24]. As five of these six loci (1q21.3, 1q42.12, 2p22.2, 21q22.3, and 22q13.1) were tested in our MPRA, we overlaid MPRA-significant variants from these loci with genome-wide significant melanocyte eQTLs. Four loci had variants that were significant in both assays, and nine of these variants displayed a consistent direction, in which the direction of allelic expression of local genes matches those of MPRA allelic transcriptional activity (**Supplementary Table 7; Supplementary Fig 10C**). Namely, two MPRA-significant variants (rs2864871 and rs6700022) from the locus on chromosome band 1q21.3 were significant eQTLs for *CTSS* in melanocytes, where lower *CTSS* levels were correlated with melanoma risk. Similarly, two to three variants each (rs2349075, rs529458487, rs398206, rs408825, rs4383, rs4384, and rs6001033) from three other loci (2p22.2, 21q22.3, and 22q13.1) also overlapped with melanocyte eQTLs, where lower *CASP8*, higher *MX2,* and higher *MAFF* levels were correlated with melanoma risk, respectively (**Supplementary Table 8**). Thus, by combining MPRA and cell-type specific melanocyte eQTL, we identified candidate functional variants and susceptibility genes from multiple melanoma GWAS loci.

For the 21q22.3 locus, twenty-two variants were originally tested in MPRA, and three of these variants were significant MPRA variants (**Fig 2A**; **Supplementary Table 9**). Of these, rs398206 in the first intron of *MX2* gene (**Fig 2A**, shown in magenta) displayed a strong transcriptional activator function (1.7 to 4.3-fold above the scrambled sequence) as well as the most significant allelic difference in the MPRA experiment (the lowest P-value of all 832

variants), where the melanoma risk-associated A allele drove significantly higher luciferase expression than protective C allele (3.1-fold in UACC903 cells, FDR = 5.6e-206; **Fig 2B**). Subsequent individual luciferase assays using the same 145bp sequence in two melanoma cell lines validated this finding (2.7 to 5.0-fold allelic difference, P = 1.1e-6 – 5.2e-11; **Fig 2C**; **Supplementary Fig 10D**). rs398206 was also a significant eQTL for levels of *MX2* gene in primary melanocytes, where the melanoma risk-associated A allele is correlated with higher *MX2* expression (Slope = 0.70, P = 6.6e-15; **Fig 2D**). These data demonstrated that integration of MPRA with cell-type specific eQTL efficiently identified functional variants from the 21q22.3 melanoma locus, as well as three additional loci (1q21.3, 2q33-q34, 22q13.1), by uncoupling multiple high-LD variants based on molecular phenotypes. This is a considerable advantage of our integrative approach complementing statistical fine-mapping, where perfect LD variants are impossible to distinguish. Based on the strong evidence for rs398206 on the locus on chromosome band 21q22.3, we focused our efforts of further molecular characterization on this locus.

## Multi-QTL analyses identified *MX2* as a melanoma susceptibility gene in the locus on chromosome band 21q22.3

While melanocyte eQTL consistently identified *MX2* as the best candidate susceptibility gene at the 21q22.3 melanoma locus[24], we further interrogated eQTL data from melanocytes and 44 GTEx tissue types, to comprehensively assess potential melanoma susceptibility gene(s) in this locus. When we inspected eQTL data from 44 GTEx tissue types, rs398206 was a significant eQTL for *MX2* in five other tissue types (testis, transformed skin fibroblasts, ovary, tibial nerve, and whole blood) but no other gene displayed a genome-wide significant eQTL with rs398206 (GTEx portal; https://gtexportal.org).

As the melanocyte *cis*-eQTL analyses used for the above assessments were limited to the genes in +/-1Mb of the tested variants[24], we explored if rs398206 is a marginal eQTL for any gene in the topologically-associated domain (TAD) to account for potential gene regulation mediated by chromatin looping typically occurring within this physical domain. From the genomic interval defined as the TAD encompassing rs398206 (chr21:42,480,000-44,320,000; hg19; retrieved from Hi-C data of SKMEL5 melanoma cell line generated for ENCODE dataset via http://promoter.bx.psu.edu/hi-c/), a total of 21 genes were significantly expressed in melanocytes, for which eQTL analyses were performed. The results demonstrated that *MX2* displayed the most significant eQTL with rs398206 (P = 6.6e-15), while none of the other genes in the TAD displayed even a marginally significant eQTL after adjusting for multiple testing (Bonferroni-corrected cutoff at P < 0.0024 for 21 genes; **Supplementary Table 10**). These data determined that *MX2* is the most likely susceptibility gene at the 21q22.3 melanoma susceptibility locus.

To complement the eQTL data, we also assessed allele-specific expression (ASE) of *MX2* in melanocytes. rs398206 is located in the 5' UTR region of an alternative *MX2* transcript isoform (ENST00000543692; **Supplementary Fig 11A**), the expression levels of which are correlated with the most abundant full-length transcript in melanocytes (ENST00000330714; Pearson *r* = 0.69, P = 1.63e-16; **Supplementary Fig 12**). RNA sequencing data from our previous study did not find genome-wide significant ASE for any melanoma-associated SNP (GWAS P < 5e-8) residing in the transcribed region of *MX2*[24], partly due to low sequence coverage of this transcript that is expressed at a low level. To thoroughly examine allele-specific expression in this region, we genotyped rs398206 in melanocyte cDNA using a Taqman genotyping assay that recognizes both genomic DNA and cDNA. The results demonstrated an over-representation of A allele-bearing transcripts in 27 heterozygous individuals, when the allelic ratio in cDNA was normalized to those in genomic DNA (One-sample Wilcoxon test, P =

223 2.49e-5; **Supplementary Fig 13**). These data are consistent with the eQTL data, where the
224 risk-associated A allele is correlated with higher *MX2* expression.

225 To thoroughly investigate possible mechanisms of allelic *MX2* expression in relation to
226 rs398206, we performed a series of additional QTL analyses in melanocytes addressing
227 alternative modes of gene regulation - splice-QTL (sQTL), DNA methylation QTL (meQTL),
228 microRNA QTL (miQTL), and RNA stability QTL (QTL analysis of estimated mRNA half-life by
229 measuring the differences between exonic and intronic read changes from RNAseq data[30]).
230 Among them, sQTL analyses using LeafCutter[31] suggested that the main effect of the *MX2*
231 eQTL was not driven by alternative isoforms or splicing events (**Supplementary Fig 11B-F;**
232 **Supplementary Material**). Subsequent miQTL and RNA stability QTL analyses did not identify
233 any genome-wide significant QTL for rs398206 in melanocytes (data not shown). meQTL
234 analysis, on the other hand, identified a significant meQTL for rs398026 at a CpG probe near
235 the *MX2* canonical promoter, where the melanoma risk-associated A allele is correlated with
236 lower CpG methylation, which is consistent with higher expression of the full-length isoform
237 (**Supplementary Fig 14**). Two other CpG probes in the first intron of *MX2* (closer to rs398206)
238 also displayed significant meQTLs for rs398206 in melanocytes, where higher CpG methylation
239 is correlated with the risk A allele. These observations are consistent with the previous findings
240 that DNA methylation in promoters is negatively correlated with gene expression, while that of
241 transcribed regions is positively correlated with gene expression[32-36]. Taken together, eQTL,
242 sQTL, and meQTL data are consistent with the hypothesis that *MX2* full-length transcript mainly
243 accounts for the eQTL at rs398206 in melanocytes through a transcriptional mechanism.

244 **rs398206 is a functional variant regulating *MX2* levels via allelic binding of YY1**

245 To identify protein factors mediating the allelic difference observed in MPRA, we
246 performed comparative mass-spectrometry using a 21bp DNA probe encompassing rs398206

with A or C alleles and nuclear extract from the UACC903 melanoma cell line (**Fig 3A**). Among the proteins displaying allelic binding, the most prominent A-allele preferential binding was shown for Yinyang-1 (YY1), a ubiquitous transcription factor having roles in development and cancer[37] as well as in pigmentation pathways of melanocytes[38]. Sequence-based motif prediction was also consistent with this finding, indicating that the sequence around rs398206 forms a consensus binding site for YY1 favoring the A-allele (**Fig 3B**). Subsequent electrophoretic mobility shift assays (EMSAs) validated that this A-allele-preferential binding of nuclear proteins is sequence-specific, as shown by competition with unlabeled probes (**Fig 3C**). Antibody super-shift demonstrated that YY1 is present in this subset of allelic-binding proteins (**Fig 3C**), which was further validated by EMSAs with purified recombinant YY1 protein (**Fig 3C-D**). We subsequently performed chromatin immunoprecipitation (ChIP) using anti-YY1 antibody and demonstrated enrichment of YY1 binding to the genomic DNA region encompassing rs398206 in two melanoma cell lines (**Fig 4A**). Of these two cell lines, UACC647 is heterozygous for rs398206, and thus we performed genotyping of rs398206 using the DNA fragments pulled down by anti-YY1 antibody. DNA fragments pulled down using YY1 antibody displayed a significant enrichment of A allele (Mann-Whitney U test, P = 9.1e-3), while genomic DNA and serial-diluted input DNA displayed equivalent signal from both A and C alleles, indicating clear A-allele preferential binding of YY1 in melanoma cells (**Fig 4B-C**).

Based on this strong allelic YY1 binding, we next asked if YY1 regulates endogenous *MX2* expression levels. siRNA knockdown of YY1 in the UACC903 melanoma cell line demonstrated a weak but consistent reduction of *MX2* levels by four different sets of siRNAs (14-32% decrease, P = 1.5e-3 – 1.9e-5, one-sample Wilcoxon test; **Fig 5A; Supplementary Fig 15D**) indicating a regulation of *MX2* levels by YY1. To further determine if the genomic region encompassing rs398206 regulates endogenous *MX2* levels, we targeted this region by CRISPRi using dCAS9-KRAB-MeCP2[39] in the same melanoma cell line. Four gRNAs targeting the

genomic regions either directly overlapping rs398206 (gRNA 1, 3, and 4) or ~25bp upstream (gRNA 2) resulted in 61-82% reduction in *MX2* expression levels (P = 2.05e-4 – 3.19e-4, one-sample Wilcoxon test; **Fig 5B**), while the same gRNAs do not have effect on nearby *MX1* expression (**Supplementary Fig 15A**). As rs398206 is located in the intronic region of *MX2*, it is formally possible that some of the effect on *MX2* expression could be due to physical blocking of passage of transcriptional machinery by dCAS9-KRAB-MeCP2 system. CRISPRi using dCAS9 without the transcriptional repressor elements, however, displayed little or no effect on *MX2* expression, which is consistent with the CRISPRi effect on *MX2* being mainly transcriptional (**Supplementary Fig 15B, C, E**).

To identify additional support for rs398206 regulating *MX2* via YY1, we examined available chromatin interaction data involving YY1. Notably, YY1 was recently shown to mediate chromatin looping and contribute to interactions between gene promoters and enhancers within TADs[40]. Given this, we examined YY1-mediated chromatin interaction around the genomic region encompassing rs398206 in these published Hi-ChIP data using YY1 antibody. In the human colorectal carcinoma cell line, HCT116, the 5Kb bin harboring rs398206 displayed strong interactions with two adjacent bins encompassing *MX2* promoter area[40] (P = 2.27e-80 and 8.44e-24; **Supplementary Fig 16**), but not with other neighboring gene promoters (at PET count >2). Together these data determined that rs398206 is a functional variant regulating *MX2* expression via differential YY1 binding in the Chr21q22.3 melanoma locus.

**Melanocyte-specific *MX2* expression accelerates melanoma formation in zebrafish**

*MX2* is best known for its function in innate immunity as an HIV-1 restriction gene[41,42]. In GTEx tissue types, the highest *MX2* expression levels are observed in EBV-transformed lymphocytes, whole blood, and spleen, reflecting its main role in innate immune response as an interferon-stimulated gene (GTEx portal; https://gtexportal.org). On the other hand, a previous

study also demonstrated that *MX2* has cell-autonomous function in the proliferation of HeLa cells without IFNα-mediated induction[43]. In our primary melanocyte dataset, *MX2* is expressed at a relatively high level (median expression ranked at top 26.5% of all expressed genes) without IFNα stimulation. To assess co-expressed genes and enriched pathways in melanocytes expressing *MX2* at a higher level, we profiled differentially expressed genes between *MX2*-high (top 25%; n = 28) and *MX2*-low (bottom 25%; n = 28) melanocytes. From 253 differentially expressed genes in *MX2*-high melanocytes (FDR < 0.01 and $|\log_2$ fold difference| > 1; **Supplementary Table 11**), significantly enriched pathways included those relevant to cellular immune response as expected, but also included those affecting cellular growth and cancer (**Fig 6C**; **Supplementary Table 12**) suggesting a possible non-immune function of MX2 in melanocytes. On the other hand, an examination of immune infiltrates in melanomas from TCGA did not provide sufficient evidence for the roles of *MX2* in immune surveillance at least at the time of surgical resections represented in these tumor samples (**Supplementary Material; Supplementary Fig 17**).

Given the possibility of a melanocyte-specific function of *MX2*, we hypothesized that melanocyte-specific *MX2* expression might have roles in early events of melanoma formation. To test this hypothesis, we first asked if *MX2* affects growth of primary melanocytes and melanoma cells in a single culture system. Cell growth assays using the xCELLigence system demonstrated that inducible lentiviral expression of MX2 (2 – 10-fold induction; **Supplementary Fig 18**) resulted in slightly decreased growth of both melanoma cells and primary melanocytes at 100ng/ml of doxycycline treatment, while empty vector transduced cells did not show any difference (**Fig 6A-B**). To begin to understand what genes and pathways might be affected by increased *MX2* expression and could potentially underlie the altered melanoma cells/melanocytes growth, we performed RNA-seq analyses on melanocytes over-expressing MX2 (2 – 10-fold induction; **Supplementary Fig 18**). Differentially expressed genes in MX2-

overexpressing melanocytes compared to controls (158 genes, FDR<10%; melanocytes from 3 individuals, 3 biological replicates each) displayed enrichment of pathways relevant to immune response as well as those involving second messenger mediated kinase signaling and cellular growth, among others (**Supplementary Tables 13-14; Fig 6D**). Since these data did not provide an apparent mechanistic hypothesis linking the effect of increased *MX2* on reduced melanocyte growth in single cultures to its association with melanoma risk, we speculated that the effect of *MX2* on melanocyte growth might change depending on cellular context and microenvironment.

To test this idea and establish a melanocyte-specific role for *MX2* expression in the development or progression of melanoma, we examined transgenic expression of human *MX2* in a zebrafish melanoma model, in conjunction with the most recurrent somatic driver event of melanoma, $BRAF^{V600E}$. Using the previously developed miniCoopR transgene system[44], we over-expressed human *MX2* exclusively in the melanocytic-lineage using an *MITF* promoter in the background of $BRAF^{V600E}$ and $p53^{-/-}$. The results demonstrated that zebrafish with transgenic human *MX2* expression presented an accelerated melanoma formation (46% of fish developed melanoma by 19 weeks; n = 184) compared to those with GFP controls (33% of fish by 19 weeks; n = 194) in this genetic background (P = 0.003; log-rank test; **Fig 6E**). These data are consistent with *MX2* expression contributing to an increased melanoma risk in part by a melanocyte-specific mechanism.

## Discussion

In this study, we adopted an integrative approach combining MPRA with cell-type specific epigenomic and eQTL data to efficiently nominate functional variants and susceptibility genes from 20 known melanoma GWAS loci. Molecular characterization of functional variants and susceptibility genes from a GWAS locus can represent a significant commitment of time

and effort as the functions of these genes and variants could be obscure unless the relevant cell types and molecular contexts are considered. By using cell-type specific eQTL to prioritize candidate variants from MPRA, we were able to maximize the probability of finding the most plausible candidates for intense characterization in a time-efficient way. In the future, incorporation of MPRA and cell-type specific eQTL with additional genome-scale datasets, including cell-type specific chromatin interaction data as well as chromatin features of different cellular contexts will further identify strong leads for additional loci with candidate melanoma susceptibility genes and variants.

Our integrative approach efficiently identified the most plausible susceptibility genes and functional variants from four melanoma GWAS loci. For the melanoma locus on chromosome band 22q13.1, increased *MAFF* levels were correlated with risk. MAFF is a small Maf protein regulated by EGF signaling[45] and plays a role in the oxidative stress response[46], which is relevant to melanomagenesis, given the vulnerability of melanocytes to oxidative stress attributable to melanin production[47]. For the locus on chromosome band 1q21.3 and 2q33-q34, lower *CTSS* and *CASP8* levels were correlated with the risk, respectively. CTSS is a member of cathepsin proteases, initially known as lysosomal enzymes[48]. Increased expression of *CTSS* is correlated with poor prognosis in the context of some cancers (breast and colorectal cancer) but also correlated with better outcome in others (lung cancer). CASP8 is mainly known for its function in apoptosis[49], and GWAS also implicated the *CASP8* locus for breast cancer[50] and basal cell carcinoma[51]. Our results provide strong support for these three genes and warrant further in-depth characterization.

Through molecular interrogation, we demonstrated that a melanoma-associated intronic variant, rs398206, contributes to allelic expression of *MX2* via modifying an enhancer element recruiting the transcription factor, YY1. Our multi-QTL analyses of primary melanocytes further supported a transcriptional mechanism, while ruling out alternative mechanisms (splicing, RNA

stability, or microRNA-related). Thorough investigation of marginal eQTLs in the TAD further validated that *MX2* is the best target of this *cis*-regulation.

Our zebrafish model provided further support for *MX2* as a melanoma susceptibility gene accelerating melanoma formation when expressed in the cells of melanocytic-lineage. *MX2* has been mainly known as an effector of innate-immunity, conferring restriction to HIV-1 infection[41,42], and its roles in melanomagenesis have not been studied. Our findings suggest a cell-autonomous role of *MX2* in promoting melanoma formation when exclusively expressed in cells of melanocytic-lineage, in the presence of $BRAF^{V600E}$, a frequent somatic driver mutation. Our single cell-type growth assays for *MX2* also support the zebrafish data by showing growth effects for MX2 in melanocytic-lineage without the presence of neighboring cell types, albeit in the opposite direction. Further interrogation of how MX2 promotes melanocyte growth will enhance our understanding of precise molecular pathways involved. Nevertheless, our findings established *MX2* as a new gene displaying pleiotropic roles in melanoma susceptibility and immune response, building on to the established roles of telomere biology (*TERT*, Chr5p15.33)[19] and oncogene-induced senescence (*PARP1*, Chr1q42.1)[18] in genetic susceptibility to melanoma in the general population.

## Methods

### Melanoma GWAS fine-mapping

Fine-mapping of the 20 genome-wide significant loci from the meta-analysis reported by Law and colleagues[1] was conducted following a very similar approach to that of Barrett, *et al*[52]. Using the results from Law, *et al*[1] a window was defined as 1Mb on either side of the most significant variant at each locus. The only exception to this was the region that included the *ASIP* gene (20q11.2-q12), where a 6Mb region was instead defined, as this region demonstrated a long-range linkage disequilibrium. Melanoma case/control status was regressed on each genotyped

and imputed variant in turn across these regions, with the first four principal components as covariates to account for stratification on 12,419 cases and 14,242 controls from the meta-analysis (only the Harvard GWAS samples and the endometriosis controls from the Q-MEGA_610k study were unavailable). Each region was further narrowed down to the interval covering 500kb on either side of the most extreme SNPs with p-value < $10^{-6}$ in the initial single SNP analysis and any variants with an imputation INFO score < 0.5 (for variants with MAF >= 0.03) or INFO score < 0.8 (for variants with MAF < 0.03) was removed. A Bayesian-inspired penalized maximum likelihood approach implemented in HyperLasso[25] was applied to these regions. 100 iterations of HyperLasso were then conducted, using all variants in each region and a Normal Exponential Gamma prior distribution for SNP effects with a shape parameter 1.0[53,54] and scale parameter such that type 1 error is $10^{-4}$. Both the study (as a categorical variable) and the first four components were included as covariates. Because of the stochastic nature of the order in which variables are tested for inclusion, this produced a number of potential models, including some that can be considered to 'correspond' to one another, because they differ only by substituting genetic variants that are in very strong LD ($r^2 > 0.9$). By dropping equivalent models, a reduced set of models was produced and was then further reduced by dropping any model whose likelihood was inferior to that of the best model by a factor >= 10. For each remaining model, a logistic regression was conducted using the SNPs in the model to generate adjusted odds ratios. For SNPs retained in any of the models, LD blocks were defined (based on both the HyperLasso results and strength of LD) and the most significant SNP (in a multivariable analysis) from each block was selected. rs36115365 in the region near *TERT* gene (5p15.33) was not identified in the fine-mapping but included for variant selection as it was identified previously based on functional evidence[19]. Subsequent analysis showed that the risk-associated alleles at rs36115365 and at rs2447853 (the most significant SNP in the region at the time) are in negative LD and when adjusted for the latter SNP, rs36115365 has a P-value of $10^{-4}$ (**Supplementary Table 1**). Similarly, for the locus on

chromosome band 2p22.2, the optimal model was a 2-SNP, but the secondary signal at rs163094 displayed low INFO scores in some studies rendering imputation less optimal and hence was not used for variant selection. Instead, the best SNP identified by 1-SNP model (rs1056837; a missense variant of *CYP1B1*) was included as an alternative (**Supplementary Table 1**). Since the effect of the region around *MC1R* gene (16q24.3) on melanoma risk is mainly explained by several well-established coding variants[12], we did not include this region in our fine-mapping data.

**MPRA workflow**

Luciferase reporter libraries were constructed by taking 145bp genomic sequence encompassing the risk or protective allele of each variant (**Supplementary Fig 2**). We also included a scrambled sequence for each variant, where 21 bases encompassing the variant were scrambled to serve as a pseudo-baseline. Each 145bp sequence was tested in both forward and reverse directions and was assigned 10 different unique 10bp barcode sequences to minimize aberrant effects of a specific barcode sequence. Resulting sequences were tested in luciferase constructs harboring TATA minimal promoter (for potential enhancer function) or no promoter sequence (for potential promoter function). Libraries were then transfected into a melanoma cell line (UACC903) to represent melanoma-specific *trans*-acting factors and the HEK293FT cell line to obtain maximum transfection efficiency. Resulting transcribed output as well as DNA input were then quantified by sequencing. Transcriptional activity of each sequence was determined by measuring the ratio of transcribed Tag counts Per Million sequencing reads (TPM) compared to those of DNA input (**Supplementary Fig 3**). We observed high inter-transfection as well as inter-library correlations of these transcriptional activities given the same input sequences ($log_2$-transformed TPM ratios between replicates; median inter-transfection Pearson $r$ = 0.984 for transfection replicates, and 0.854 for inter-library replicates; **Supplementary Figs 4-6; Supplementary Table 5**). After removal of tags that were poorly

445 represented at the DNA level (TPM < 6), 77.47% of the tags from the input sequences were
446 retained for the further analyses (**Supplementary Fig 7**).

447 **MPRA variant selection**

448 Among 20 genome-wide significant loci from the melanoma meta-analyses by Law and
449 colleagues[1], we prioritized 16 loci where potential *cis*-regulatory mechanism could be applied.
450 We excluded the other 4 loci containing genes that are implicated in melanoma-associated
451 pigmentation phenotypes (*SLC45A2*, *TYR*, *MC1R,* and *ASIP* loci), as for many of these genes,
452 coding variants were shown to alter the protein functions. To select high-LD proxy variants for
453 16 melanoma GWAS loci (Law, *et al*., 2015[1]), we used the following criteria:

454 1. Primary lead SNPs were taken from Law, *et al*.[1] meta-analysis paper and supplemented
455 by those from additional HyperLasso analysis when there are alternative best SNPs
456 available.
457 2. For 8 loci, HyperLasso analysis nominated independent multiple secondary signals and
458 these lead SNPs were also added.
459 3. SNPs of $r^2 > 0.4$ with the primary or secondary lead SNPs using 1000 Genomes phase3
460 EUR or CEU populations were selected as "high-LD variants" (n = 2,748)

461 To prioritize high-LD variants overlapping melanocyte/melanoma open chromatin regions and/or
462 active promoter/enhancer histone marks, we used one or more of the following criteria:

463 1. Variant is located within a human melanocyte DHS peak from one or more individuals of
464 three available through ENCODE and Epigenome Roadmap database.
465 2. Variant is located within a human melanocyte H3K27Ac ChIP-Seq peak from one or
466 more individuals **AND** a H3K4Me1 ChIP-Seq peak from one or more individuals of two
467 and three available through Epigenome Roadmap database, respectively.

3. Variant is located within a human melanocyte H3K27Ac ChIP-Seq peak from one or more individuals **AND** a H3K4Me3 ChIP-Seq peak from one or more individuals of two and three available through Epigenome Roadmap database, respectively.
4. Variant is located within a human melanoma short-term culture FAIRE-Seq peak from one or more individuals of 11 available from Verfaillie et al[26].

Based on the above criteria, **832** melanoma GWAS variants were selected to be tested by MPRA. We also included 8 additional variants from Chr1q21.3 that were of $r^2 > 0.8$ with the lead SNP but did not overlap with any functional signature listed above and assigned them as negative controls. Of 832 variants, 306 as well as 8 negative controls were repeated in two libraries to ensure cross-library consistency (see **MPRA oligo library design**). These 306 variants are also $r^2 > 0.6$ with their lead SNPs and supported by both open chromatin and histone mark annotation from melanocyte or melanoma data. A complete list of variants tested are listed in **Supplementary Table 3**.

## MPRA oligo library design

Oligo libraries were designed mainly following the guidelines from published works[21,27] with some modifications. Two libraries containing 32,580 (library 1) and 36,660 (library 2) unique sequence of 200-mer oligos (total of 50,400 unique sequences across two libraries with 18,840 repeated in both) were synthesized by Agilent Technologies (Santa Clara, CA). Composition of each library by GWAS locus and repeated variants are listed in **Supplementary Tables 3-4**. For each variant, 145 bases encompassing the variant with either risk or protective allele in both forward and reverse directions were synthesized together with 10 different 10 base random barcode sequences. These two parts of sequences were separated by recognition sequences for restriction enzymes KpnI (GGTACC) and XbaI (TCTAGA), and flanked by binding sequences for PCR primers (200 bases oligo sequences: 5'-ACTGGCCGCTTCACTG-145 bases-GGTACCTCTAGA-10 bases tag-AGATCGGAAGAGCGTCG-3'). For each variant, a

493 scrambled sequence (core 21 bases encompassing the SNP with the reference allele were
494 shuffled) was also tested in forward and reverse directions in the same manner. This is
495 equivalent to a total of 60 unique sequences designed per variant. When there are additional
496 SNPs other than the test SNP that fall in the 145bp region, major allele in EUR population was
497 used. For indels, 145 bases length was set based on insertion allele and the deletion allele was
498 left shorter than 145 bases. Random 10 base tag sequences were generated once so that each
499 library has up to 36,660 unique tag sequences (the same 36,660 tag sequences were used for
500 each library). For the 10 base tag sequence and scrambled 21 base core sequence, only
501 homopolymers of <4 bases were used and the enzyme recognition sites for KpnI, XbaI, and SfiI
502 were avoided. A complete list of oligo sequences can be found in **Supplementary File 1** as an
503 R object.

504 **MPRA library construction and sequencing**

505 MPRA library construction and sequencing was performed following published protocols with
506 some modifications[21,27].

507 *Cloning of the libraries*

508 Ten femtomole each of gel-purified (10% TBE-Urea polyacrylamide gel) oligo libraries was
509 amplified by emulsion PCR using Herculase II fusion polymerase and 2 μM of primers providing
510 SfiI enzyme sites (**Supplementary Table 15**), following the instructions of the Micellula DNA
511 Emulsion & Purification Kit (EURx/CHIMERx). Amplified oligos were quantified using KAPA
512 qPCR assay and verified by DNA sequencing on Ion PGM. Amplicon libraries were prepared
513 using 30ng of oligos from emulsion PCR using Ion Plus Fragment Library Kit and were
514 sequenced on Ion PGM for an average 203bp and 175bp read length at 6.7 million and 5.6
515 million reads per sample for library 1 and library 2, respectively. To verify oligo library design,
516 21bp sequences within oligos including variant site and +/- 10bp were used to map to each

517 sequencing read. Linux command “fgrep” was used and only 100% sequence match was kept.
518 We then counted the total read depth for each variant represented by the matched sequences,
519 and then calculated the proportion of variant sequences that were verified. For both library 1
520 and library 2, more than 97% of unique sequences representing the variants in the library were
521 detected with at least 10 sequencing reads. In addition, we found similar proportion and read
522 depth for sequences representing both forward and reverse directions in both libraries. If we use
523 the actual tag sequences as a bait, 82% of tags could be verified, with a caveat that some tags
524 were amplified but not detected because of relatively poor sequencing quality in this position of
525 the amplicon. Sequence-verified oligo libraries were first cloned into pMPRA1 vector (Addgene)
526 using SfiI site by electroporation into 10 times higher number of bacterial cells than the number
527 of unique sequences in the oligo library. Cloned pMPRA1 was further digested on KpnI and
528 XbaI sites between 145bp test sequence and 10bp barcode sequence, where luc2 ORF with or
529 without a minimal promoter was ligated from pMPRAdonor2 and pMPRAdonor1 (Addgene),
530 respectively. The ligation product was transformed by electroporation into 10 times higher
531 number of bacterial cells in the same manner. Cloned final library for transfection was verified
532 on the gel as a single band after KpnI digestion.

533 *Transfections and sequencing library preparation*

534 Each library was transfected at least four times (two transfections for each promoter type) into
535 HEK293FT or UACC903 melanoma cells aiming > 100 times higher number of transfected cells
536 than the library complexity considering transfection efficiency estimated by a separate GFP
537 transfection and visualization. A summary of transfections is listed in **Supplementary Table 5**.
538 Cells were transfected using Lipofectamine 3000 and harvested at 24 hours after transfection
539 for RNA isolation. Total RNA was isolated using Qiagen RNeasy kit, and mRNA was
540 subsequently isolated using PolyA purist MAG kit. cDNA was then synthesized using
541 Superscript III, from which only short sequences encompassing 10 bp unique barcodes were

amplified using Q5 high-fidelity polymerase and primers introducing Illumina TruSeq adapter sequences (**Supplementary Table 15**). Tag sequence libraries were also prepared using input DNA in the same way. Each tag sequence library was sequenced on a single lane of HiSeq2500 (125 bp paired end read).

## MPRA data analyses

### *Obtaining normalized tag counts*

Using FASTQ files from input DNA or RNA transcript sequencing, we counted the number of reads (Illumina read 1) completely matching 10bp barcode sequences (tag counts) and the same downstream sequence context ("TCTAGAATTATTACACGGCG") including an XbaI recognition site and the 3' of the *luc2* gene. For each transfection (equivalent to one sequencing run), Tag counts Per Million sequencing reads (TPM) values were calculated by dividing each tag count by the total number of sequence-matching tag counts divided by a million. TPM ratio was then taken as RNA TPM over input DNA TPM and $\log_2$ converted.

### *Quality control*

From each input DNA library at least 92.1% of barcode sequences were detected, and > 89.3% were covered at 10 reads or higher. From RNA samples 87.4 – 93.3% of barcode sequences were detected, and 84.8 – 90.8% were covered at 10 reads or higher (**Supplementary Table 5**). Barcodes showing 10 tag counts or lower were excluded from the further analyses. Median tag counts for the barcodes that were included in the analyses were 48,973-49,903 for DNA input and 46,471-49,758 for RNA output. Reproducibility between transfections were assessed by Pearson correlation of $\log_2$-transformed TPM ratio of each barcode between replicates of transfection. We observed correlation coefficient of 0.944 or higher for each library transfected to HEK293FT cells and 0.935 or higher for UACC903 cells (**Supplementary Fig 4**). Correlation test was also performed between repeated sequences across libraries. We observed correlation

coefficient of 0.821 or higher for HEK293FT cells (**Supplementary Fig 5**) and 0.815 or higher for UACC903 cells (**Supplementary Fig 6**). To avoid low input DNA counts driving variations in RNA/DNA TPM ratios, we removed tags with < 6 TPM counts from further analyses. The remaining tags account for 77.47% of all the detected tags (**Supplementary Fig 7**).

*Identification of functional GWAS variants*

We analyzed the normalized MPRA measurement ($log_2$ transformed TPM ratio) using a standard linear regression model. We used the Wald test to test the impact of "allele" on MPRA level, after adjusting the effect of "Strand" (forward or reverse direction) as a binary covariate, the effect of "Transfection" as a categorical covariate with 18 levels (accounting for different promoter status and cell types as well as cross-transfection variations). To account for the potential heteroskedasticity in the measurement error, we used the robust sandwich type variance estimate in the Wald test as recommended by Long and Ervin[55], and used the R package "Sandwich" to conduct the analysis. To assess overall transcriptional activity of the 145bp DNA element including the variant, we used variant-specific scrambled sequences as a null. $Log_2$ transformed TPM ratios of scrambled sequences were regressed against those of either reference or alternative allele while using the same covariates ("Strand" and "Transfection"). $Log_2$ TPM ratio for each tag in each transfection was considered as an experimental replicate for regression. The same set of analyses was done only using data from UACC903 melanoma cells and further dropping data for repeated variants from one of two libraries (library 1) to allow subsequent enrichment analyses. Variants showing FDR < 0.01 for both allelic difference and departure from null (for either allele) in both UACC903 only and combined set were called as significant MPRA variants. Complete processed MPRA data can be found in **Supplementary File 1** as an R object.

**Motif analysis**

Prediction of variant effects on transcription factor binding sites was performed using the motifbreakR package[28] and a comprehensive collection of human transcription factor binding sites models (HOCOMOCO)[56]. We selected the information content algorithm and used a threshold of 0.001 as the maximum P-value for a transcription binding site match in motifbreakR. $Log_2$ fold change between alternative allele score and reference allele score were used to predict the transcription factor motif effect for each variant.

**Statistical fine-mapping using PAINTOR**

PAINTOR 3.0 (http://bogdan.bioinformatics.ucla.edu/software/paintor) was used to estimate the posterior probability of any SNP within a melanoma locus to be causal. We used default parameters in PAINTOR (window size of 100Kb, max causals 2) and filtered out all the SNPs with P-value > 0.5 for computational efficiency. The pairwise LD between all SNPs in each window was computed using the 1000 Genomes EUR data. Functional annotations were provided as part of PAINTOR software which was complimented with a melanocyte specific gene set annotation[24]. In order to determine which annotations are relevant to the phenotype being considered, we ran PAINTOR on each annotation independently and then selected 4 annotations specific to primary melanocytes with high sum of log Bayes factors for the final model to compute trait-specific posterior probabilities for causality. These 4 annotations include melanocyte-specific expressed genes from our melanocyte dataset[24], melanocyte enhancers, TF-binding sites, and histone marks from ENCODE and Roadmap. Aside from the variants not meeting our analysis parameters, 462 out of 832 MPRA-tested variants were assigned a posterior probability by PAINTOR and were used for enrichment analyses.

**Melanocyte eQTL, sQTL, meQTL, and RNA stability QTL**

Primary melanocyte eQTL data was obtained from our previous study[24], where 106 individuals mainly of European decent were analyzed. For the marginal eQTL analysis of the genes located

in the TAD including rs398206, 21 genes were selected based on expression thresholds of >0.5 RSEM and ≥6 reads in at least 10 samples. Using FastQTL[57], nominal P-value was generated between each gene and all the SNPs +/-2 Mb of rs398206 to test the alternative hypothesis that the slope of a linear regression model between the genotypes and expression levels deviates from 0. The same set of covariates as that used for the eQTL analyses was applied (three top genotype PCs and 10 top PEER factors). A Bonferroni-corrected cutoff of $P < 0.0024$ for 21 genes was then applied to select the genes showing marginal eQTL with rs398206. For sQTL, meQTL, and RNA stability QTL, we performed similar QTL analyses as our previous eQTL study using the same genotype data, population structure covariates, and statistical approaches. We replaced normalized gene expression levels with normalized splice junction events (sQTL), normalized methylation values (meQTL), and normalized mRNA stability measures (RNA stability QTL). We also re-calculated the top 15 PEER factors according to these phenotype values. For sQTL analysis, STAR[58] was used to map the RNA-Seq reads onto the genome (hg19) and then LeafCutter[31] was applied to quantify the splice junctions following the procedures described by the authors (http://davidaknowles.github.io/leafcutter/articles/sQTL.html). For meQTL analysis, we performed genome-wide DNA methylation profiling on Illumina Infinium Human Methylation 450K BeadChip. Methylation levels of all 106 primary melanocyte samples was measured according to the manufacturer's instruction at Cancer Genomics Research Laboratory at NCI. Measurement of raw methylation densities and quality control were conducted using the RnBeads pipeline[59] and the minfi package[60] (http://bioconductor.org/packages/minfi/). In total, we retained 635,022 probes for the downstream meQTL analysis. No batch effects were identified and there were no plating issues. To obtain the final methylation levels (beta value) for meQTL anlaysis, normalization was performed using the preprocessFunnorm algorithm implemented in minfi R package[60]. For RNA stability QTL, we calculated mRNA half-life by measuring the differences between exonic and intronic read changes from 106 melanocyte

RNAseq data using REMBRANDTS package (https://github.com/csglab/REMBRANDTS). The unbiased estimates of differential mRNA stability (Δexon–Δintron–bias), relative to the average of all samples, was obtained from the output file "the stability.filtered.mx.txt".

### *MX2* isoform analysis

Taqman assays targeting unique junctions of *MX2* transcript isoforms were obtained from Thermo Fisher (full-length transcript: Hs01550809_m1 and AP323EZ; ENST00000543692: APYMKKU; ENST00000418103: AP2W9U3). Custom assay design was based on Ensembl75 GRCh37 annotation. RNA was isolated from primary cultures of melanocyte from 106 individuals mainly of European decent[24], and cDNA was synthesized using iScript Advanced cDNA Synthesis Kit (Bio-Rad). Taqman assays were performed in triplicates (technical replicates) to be averaged to single data points and normalized to *TBP* levels. *TBP* was selected among 16 conventional human control genes as being one of the least variable genes in melanocyte dataset based on RNAseq data.

### Cell culture

Melanoma cell lines were grown in the medium containing RPMI1640, 10% FBS, 20 mM HEPES, and Amphotericin B/penicillin/streptomycin. All cell lines were tested negative for mycoplasma contamination.

### Luciferase assays

For each tested SNP, the exact same 145bp sequence encompassing rs398206 as tested in MPRA was amplified from genomic DNA of HapMap CEU panel samples carrying either risk or protective allele. Primers were designed to carry 15 base 5' overhangs recognizing either side of pGL4.23 vector after KpnI single cut in both forward and reverse direction to facilitate recombination (**Supplementary Table 15**). Amplified fragments containing 145 bp sequence were then cloned into pGL4.23 vector using In-Fusion HD Cloning kit (Clontech). The resulting

constructs were co-transfected with renilla luciferase into melanoma cell lines (UACC903 and UACC502) using Lipofetamine 2000 reagent following the manufacturer's instructions (Thermo Fisher) in 24-well format. Cells were harvested at 24hrs after transfection for luciferase activity assays. All the experiments were performed in at least three biological replicates in sets of 6 replicates.

**EMSA and super-shifts**

Forward and reverse strand of 21-mer DNA oligos encompassing rs398206 were synthesized with 5' biotin labeling (Life Technologies; **Supplementary Table 15**) and were annealed to make double stranded probes. Nuclear extracts were prepared from actively growing melanoma cells (UACC2331) using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Scientific). Probes were bound to 2 µg nuclear extracts pre-incubated with 1 µg poly d(I-C) (Roche) or 100-750ng YY1 full-length recombinant protein (31332, Active Motif) in binding buffer containing 10 mM Tris (pH 7.5), 50 mM KCl, 1 mM DTT, and 10 mM MgCl2 at 4ºC for 30 min. For competition assay, unlabeled competitor oligos were added to the reaction mixture 5 min prior to the addition of probes. Completed reactions were run on 5% or 4-20% native acrylamide gel and transferred blots were developed using LightShift Chemiluminescent EMSA kit (Thermo Scientific) and imaged on Chemidoc Touch (Bio-Rad). For antibody-supershifts, 0.6-1.2 µg of antibody against YY1 (sc-1703X, Santa Cruz) or rabbit normal IgG (sc-2027, Santa Cruz) were bound to nuclear extract prior to poly d(I-C) (Roche) incubation at 4ºC for 1 hr.

**Chromatin immunoprecipitation and genotyping**

Melanoma cells (UACC903 and UACC647) were fixed with 1% formaldehyde when ~85% confluent, following the instructions of Active Motif ChIP-IT high sensitivity kit. 7.5 x $10^6$ cells were then homogenized and sheared by sonication using a Bioruptor (Diagenode) at high setting for 15 min, with 30 sec on and 30 sec off cycles. Sheared chromatin from 2 x $10^6$ cells

were used for each immunoprecipitation with antibodies against YY1 (sc-1703X, Santa Cruz), or normal rabbit IgG (sc-2027; Santa Cruz) following the manufacturer's instructions. Purified pulled-down DNA or input DNA was assayed by SYBR Green qPCR for enrichment of target sites using primers listed in **Supplementary Table 15**. Relative quantity of each sample was driven from standard curve of each primer set and normalized to 1/100 input DNA. For genotyping rs398206, pulled down DNA, input DNA, or genomic DNA from UACC647 cell line (heterozygous for rs398206) was used as template DNA for Taqman genotyping assay (Assay ID: C___2265405_20). All experiments were performed in at least three biological replicates in sets of triplicates.

**Mass spectrometry**

Nuclear lysates for mass spectrometry analysis were collected from UACC903 cells grown in RPMI 1640 media (Gibco) supplemented with 10% FBS, 20 mM HEPES (pH 7.9), 100 U/ml penicillin and 100 µg/ml streptomycin (Gibco)[61]. 21bp oligonucleotide probes encompassing rs398206 were ordered via custom synthesis from Integrated DNA Technologies with 5'-biotinylation of the forward strand (**Supplementary Table 15**). Forward and reverse DNA oligos were annealed using a 1.5X molar excess of the reverse strand. DNA pulldowns and on-bead digestion were performed on a 96-well filterplate system as described previously[62]. In short, 500 pmol of annealed DNA oligos were immobilized on 10 µl (20 µl slurry) Streptavidin-Sepharose beads (GE Healthcare) for each pulldown. Immobilized DNA oligos were incubated with 500 µg of UACC903 nuclear extract and 10 µg of non-specific competitor DNA (5 µg polydAdT, 5 µg polydIdC). After washing away unbound proteins, beads were resuspended in elution buffer (2 M Urea, 100 mM TRIS (pH 8), 10 mM DTT), alkylated with 55 mM iodoacetamide, and on-bead digested with 0.25 µg trypsin. After desalting using Stage tips, peptides were labelled by stable isotope dimethyl labeling, as described previously[62]. Each pulldown was performed in duplicate and label swapping was performed between replicates to eliminate labeling bias. Matching light

and heavy peptides were combined and loaded onto a 30cm column (heated at 40°C) packed in-house with 1.8 um Reprosil-Pur C18-AQ (Dr. Maisch, GmbH). The peptides were eluted from the column using a gradient from 9 to 32% Buffer B (80% acetonitrile, 0.1% formic acid) in 114 minutes at a flow rate of 250 nL/min using an Easy-nLC 1000 (Thermo Fisher Scientific). Samples were sprayed directly into a Thermo Fisher Orbitrap Fusion Tribrid mass spectrometer. Target values for full MS were set to 3e5 AGC target and a maximum injection time of 50 ms. Full MS were recorded at a resolution of 120000 at a scan range of 400-1500 m/z. The most intense precursors with a charge state between 2 and 7 were selected for MS/MS analysis, with an intensity threshold of 10000 and dynamic exclusion for 60s. Target values for MS/MS were set at 2e4 AGC target with a maximum injection time of 35ms. Ion trap scan rate was set to 'rapid' with an isolation width of 1.6 m/z and collision energy of 35%. Scans were collected in data-dependent top-speed mode in cycles of 3 seconds. Thermo RAW files were analyzed with MaxQuant 1.6.0.1 by searching against the UniProt curated human proteome (released June 2017) with standard settings[63]. Protein ratios were normalized by median ratio shifting and used for outlier calling. An outlier cutoff of 1.5 inter-quartile ranges in two out of two biological replicates was used.

**siRNA knockdown of YY1**

siRNA knockdown of YY1 was performed in the UACC647 melanoma cell line using ON-TARGETplus YY1 siRNAs (J-011796-08, J-011796-09, J-011796-10, and J-011796-11; Dharmacon). Non-targeting siRNA and siRNA targeting *GAPDH* were used for negative and positive control, respectively. Six picomole of siRNA was transfected into 5 x $10^4$ cells using Lipofectamine RNAiMax (Thermo Fisher) following the reverse transfection procedure in 24-well format. Cells were harvested at 72 hours after transfection for RNA isolation. The experiments were performed in 4 biological replicates in sets of 6 replicates. Total RNA was isolated using RNeasy kit (Qiagen) and cDNA was generated using iScript Advanced cDNA Synthesis Kit (Bio-

Rad). *MX2* levels were measured using Taqman probe set (Assay ID: Hs01550809_m1) specifically detecting the full-length isoform and normalized to *GADPH* levels. qPCR triplicates (technical replicates) were averaged to be considered as one data point. Cells were also harvested for protein isolation from each biological replicate to assess knockdown efficiency by Western blot analysis. Total cell lysates were generated with RIPA buffer (Thermo Scientific, Pittsburgh, PA) and subjected to water bath sonication. Samples were resolved by 4-12% Bis-Tris ready gel (Invitrogen, Carlsbad, CA) electrophoresis. The primary antibodies used were rabbit anti-YY1 (sc-1703X, Santa Cruz), and mouse anti-GAPDH (sc-51907, Santa Cruz).

**CRISPRi of rs398206**

CRISPRi was performed in UACC903 melanoma cell line (AA genotype for rs398206) using four different gRNAs targeting the genomic region on or near rs398206 (gRNA sequences are listed in **Supplementary Table 15**). Guide RNA target sites were identified using the sgRNA Scorer 2.0 algorithm[64]. Non-targeting gRNA and gRNA targeting the adeno-associated virus site 1 (AAVS1) were used as controls. For each sgRNA, forward and reverse oligonucleotides were annealed and cloned into vector carrying the sgRNA scaffold using the BsmBI restriction enzyme (NEB). For CRISPRi, 400ng of the vectors containing gRNAs, 500ng of dCas9-KRAB-MeCP2 (Addgene: 110821) or dCAS9 (Addgene: 47316), and 100 ng of pCMV6-entry vector (carrying neomycin resistance marker) were co-transfected into 2 x $10^5$ cells using Lipofectamine 2000 (Thermo Fisher) following a reverse transfection procedure scaled to 12-well format. Half the amount of DNA, lipofectamine, and cells were used when conducting 24-well format of culture. Cells were treated with 1mg/ml Geneticin (Gibco) 24 hours after transfection. Cells were harvested 48 hours after drug selection for RNA and protein isolation. The experiments were performed in at least 3 biological replicates in sets of 5-6 replicates. Total RNA was isolated using RNeasy kit (Qiagen) and cDNA was generated using iScript Advanced cDNA Synthesis Kit (Bio-Rad). *MX2* levels were measured using Taqman probe set (Assay ID:

763 Hs01550809_m1) specifically detecting the full-length isoform and normalized to *GADPH* levels.
764 qPCR triplicates (technical replicates) were averaged to be considered as one data point.
765 UACC903 cells tested negative for mycoplasma. Cells were concomitantly transfected and
766 harvested for protein isolation from one representative set of dCAS9 vs. dCas9-KRAB-MeCP2
767 experiments (**Supplementary Fig 15B-C**) for western blotting following the same procedure
768 described before. Proteins were separated on NuPAGE 3-8% Tris-Acetate Protein Gels
769 (Thermo Fisher). The primary antibodies used were mouse anti-CAS9 (7A9-3A3, Active Motif),
770 and mouse anti-GAPDH (sc-51907, Santa Cruz).

771 ***MX2* allele-specific expression**

772 Melanocyte cells were grown in Dermal Cell Basal Medium (ATCC PCS-200-030)
773 supplemented with Melanocyte Growth Kit (ATCC PCS-200-041) and 1% amphotericin
774 B/penicillin/streptomycin (120-096-711, Quality Biological) as described before[24]. Total RNA
775 was isolated using a miRNeasy Mini kit (217004, Qiagen) further treated with CTAB-Urea
776 following a previously described method[65] to remove excess melanin pigmentation. cDNA was
777 synthesized from total RNA using iScript Advanced cDNA Synthesis Kit (Bio-Rad). Genomic
778 DNA and cDNA were then genotyped for rs398206 using custom Taqman genotyping probe set
779 (ANRWEYM) recognizing both genomic DNA and cDNA (ENST00000543692) with a 5bp 5'
780 overhang on the left primer for cDNA based on Ensembl archive 75 annotation. From a total of
781 44 samples heterozygous for rs398206, 27 samples passing QC (Ct values lower than 38 for
782 both alleles in cDNA and genomic DNA) were used to calculate A/C allelic ratio based on dRn
783 values.

784 **MX2 over-expression and growth assays**

785 Melanoma cells and melanocyte growth assays were conducted using lentiviral transduction of
786 *MX2* cDNA under the control of tetracycline-inducible promoter using pINDUCER20 vector

(Addgene). The *MX2* cDNA clone (RC206437) in the pCMV6-entry backbone was purchased from Origene and full-length *MX2* cDNA sequence was sub-cloned to pENTR-1A vector by introducing stop codons and removing 3' Myc-DDK tag before being transferred to pINDUCER20 vector (adapter sequence is listed in **Supplementary Table 15**). BamHI and MluI sites on pCMV6-entry vector and BamHI and XhoI sites on pENTR-1A were used for sub-cloning. Primary human melanocytes were obtained from Invitrogen and/or the Yale SPORE in Skin Cancer Specimen Resource Core and grown under standard culture conditions using Medium M254 (Invitrogen) with Human Melanocyte Growth Supplement-2 (Invitrogen). For lentivirus production, lentiviral vectors were co-transfected into HEK293FT cells with packaging vectors psPAX2, pMD2-G, and pCAG4-RTR2. Virus was collected two days after transfection and concentrated by Vivaspin20. Cells were incubated with virus for 24 hours, followed by drug selection (1 mg/ml Geneticin, Gibco), before being subjected to experimental treatments and assays. For xCELLigence assays, optimized number of cells for each cell type were seeded to RTCA E-plate 16 and grown until the Cell Index stabilized. Varying amounts of doxycycline were then added, and the Cell Index was monitored for 72 hours. All experiments were performed in 3 biological replicates in sets of triplicates. For each round, cells were concomitantly infected and harvested for protein isolation at 72 hours of doxycycline treatment to assess MX2 levels by western blotting. The primary antibodies used were rabbit anti-MX2 (NBP1-81018, Novus Biologicals), and mouse anti-GAPDH (sc-51907, Santa Cruz).

**Differentially expressed genes in *MX2*-high melanocytes**

From the RNA-seq data of primary melanocytes (n = 106), we profiled differentially expressed genes (DEGs) between *MX2*-high (top 25%; n = 28) and *MX2*-low (bottom 25%; n = 28) samples. Total counts of mappable reads for each annotated gene (GENCODE v19) was obtained using featureCounts from Rsubread package[66]. The SARTools[67] workflow was used to perform quality control, apply differential analysis and generate reports based on the count data

from both *MX2*-high and *MX2*-low groups. edgeR[68] was selected as the statistical methodology to determine differential expression based on the negative binomial distributions. The final DEG list with criteria FDR < 0.01 and $|\log_2$ fold difference$| > 1$ was applied to Ingenuity Pathway Analysis (IPA).

**RNA-seq of melanocytes over-expressing MX2**

For RNA-seq analyses of MX2 over-expressing melanocytes, primary cultures of melanocytes from three individuals (C23, C29, and C53) were selected based on their low basal *MX2* expression levels. Cells were grown and infected with the lentiviral system using *MX2* cDNA cloned into pINDUCER20 or empty vector as described above. Following drug selection, cells were treated with 0 or 100 ng/ml doxycycline (total of three conditions for each cell line: 0 or 100 ng/ml doxycycline for pINDUCER20-*MX2* infected cells, and 100 ng/ml doxycycline treatment for empty vector infected cells) for 72 hours before being harvested for RNA and protein isolation. For each cell line, three separate infections (biological replicates) were performed and sequenced for transcriptome analysis (total of 27 samples sequenced: 3 conditions, 3 cell lines, and 3 biological replicates). Western blotting was performed for each cell line to estimate the level of MX2 induction. Total RNA was isolated in the same way as previously described[24]. Sequencing library was constructed following Illumina TruSeq Standard mRNA Library protocol. 150 bp paired-end sequencing was performed on NovaSeq 6000 to achieve at least 50 million reads per sample (range 53.0-82.4 M). FASTQ raw data was received and quality control was performed by the MultiQC RNA-Seq module[69] (https://multiqc.info). Quasi-mapping algorithm Salmon[70] was used to provide fast and bias-aware quantification of transcript expression using GENCODE human transcripts database (release 29). A principal component analysis was performed based on the expression qualification, and based on the results, differentially expressed genes (DEGs) were calculated with DESeq2[71] adjusting for cell line, biological replicate, and library construction batch. The expression threshold FDR < 0.1 was recognized

as DEGs after *MX2* over expression by comparing pINDUCER20-*MX2*-infected cells with (100 ng/ml) or without doxycycline treatment. The list of significant DEGs was analyzed using IPA for pathway enrichment analysis. Threshold of $P < 0.05$ and non-zero z-scores were used for identifying significantly enriched pathways. DEG analysis of cells infected with empty vector followed by 100 ng/ml doxycycline treatment vs. those infected with pINDUCER20-*MX2* with no treatment was performed as a control. IPA analysis using DEGs from this control analysis (1838 genes at FDR < 0.01 cutoff) did not overlap in the same direction of change with the main pathways enriched by MX2 overexpression except for "Apelin Endothelial Signaling Pathway" (**Supplementary Table 14 and 16**).

**Zebrafish melanoma model**

The *MX2* open reading frame was cloned under the control of the melanocyte-specific mitfa promoter into the miniCoopR expression vector[44]. Tg(*mitfa:BRAF*$^{V600E}$), *p53*-/-, *mitfa*-/- embryos were injected at the one cell stage with either miniCoopR *mitfa:MX2* or miniCoopR *mitfa:EGFP* (as a negative control). Embryos were sorted for melanocyte rescue at 5 days post fertilization and raised to adulthood. Tumor formation was monitored weekly between weeks 10 and 19 post-injection. There were no observable differences between the negative control and *MX2* group in melanocyte rescue efficiency, overall pigmentation of fish, and morphology or pigmentation of melanomas. Three independent experiments of different sample sizes were performed by independent injections of DNA constructs replicating similar results. Kaplan-Meier survival curve was plotted using the combined data from these three sets, and P-value was calculated using log-rank test. Zebrafish were handled humanely according to our vertebrate animal protocol that implements the principles of replacement, reduction, and refinement ('three Rs'), has been approved by Boston Children's Hospital Animal Care Committee, and includes detailed experimental procedures for all *in vivo* experiments described in this paper.

## Statistical analyses

All cell-based experiments were repeated at least three times with separate cell cultures. When a representative set is shown, replicate experiments displayed similar patterns. For all plots, individual data points are shown with the median or mean, range (maximum and minimum), and 25th and 75th percentiles (where applicable). The statistical method, number of data points, and number and type of replicates are indicated in each figure legend.

## Data availability

The data generated during the current study are deposited in Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) as a SuperSeries under the accession number GSE129250. A complete list of oligo sequences for MPRA libraries and complete processed MPRA data can be found in **Supplementary File 1** as an R object. Melanocyte eQTL data and expression data from 106 individuals are available through the database of Genotypes and Phenotypes (dbGAP, https://www.ncbi.nlm.nih.gov/gap) under accession number phs001500.v1.p1.

## URLs

ENCODE Project, https://www.encodeproject.org/; Roadmap Epigenomics Project, http://www.roadmapepigenomics.org/; UCSC Genome Browser, http://genome.ucsc.edu/; GTEx Portal, http://www.gtexportal.org/home/testyourown; TCGA, https://cancergenome.nih.gov/; motifbreakR, http://bioconductor.org/packages/motifbreakR/; PAINTOR, https://github.com/gkichaev/PAINTOR_V3.0; 3D Genome Browser, http://promoter.bx.psu.edu/hi-c/; LeafCutter, https://davidaknowles.github.io/leafcutter/articles/sQTL.html; TIMER, https://cistrome.shinyapps.io/timer/; CIBERSORT, https://cibersort.stanford.edu/; minfi: http://bioconductor.org/packages/minfi/; REMBRANDTS: https://github.com/csglab/REMBRANDTS

## Acknowledgements

The results appearing here are in part based on data generated by the TCGA Research Network (http://cancergenome.nih.gov/). Data were also obtained from the GTEx Portal (V6p) on 2 December 2015 or dbGaP accession phs000424.v6.p1 on 17 December 2015. We would like to thank K. Jones, S. Bass, K. Teshome, S. Brodie and other members at the National Cancer Institute Cancer Genomics Research Laboratory (CGR) for help with sequencing efforts and xCELLigence assays, B. Hennessey, H. Kong, and L. Mehl from the National Cancer Institute, Laboratory of Translational Genomics for proofreading the manuscript, and A. Jermusyk, O. Onabajo, L. Jessop, and L. Amundadottir from National Cancer Institute, Laboratory of Translational Genomics for helpful discussions, and S. Loftus and W. Pavan from National Human Genome Research Institute for the help with melanocyte eQTL study. This work has been supported by the Intramural Research Program (IRP) of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health. The content of this publication does not necessarily reflect the views or policies of the US Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government. The Vermeulen lab is part of the Oncode Institute, which is partly funded by the Dutch Cancer Society (KWF).

## Author contributions

J.C., T.Z., and K.M.B. conceived and planned the study. J.C., T.Z., L.M.C., M.A.K., K.Y., and B.P. designed and analyzed MPRA assays. J.C., A.V. and M.X. conducted experiments for MPRA and MX2 molecular characterization. R.C. designed CRISPR assays. T.Z. performed all the data analyses. M.M.M., C.G., and M.V. conducted proteomics analyses. M.B., J.T., B.P., C.H., F.D., J.H.B., M.H.L., and M.M.I. performed fine-mapping of melanoma GWAS data. J.A.,

H.R., and L.I.Z. designed and performed zebrafish experiments. J.C., T.Z., and K.M.B. wrote the manuscript. K.M.B. and S.J.C. helped supervise the project.

## Competing interests

The authors declare no competing financial interests.

## References


1. Law, M.H. *et al.* Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. *Nat Genet* **47**, 987-995 (2015).
2. Gudbjartsson, D.F. *et al.* ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. *Nat Genet* **40**, 886-91 (2008).
3. Brown, K.M. *et al.* Common sequence variants on 20q11.22 confer melanoma susceptibility. *Nat Genet* **40**, 838-40 (2008).
4. Bishop, D.T. *et al.* Genome-wide association study identifies three loci associated with melanoma risk. *Nat Genet* **41**, 920-5 (2009).
5. Falchi, M. *et al.* Genome-wide association study identifies variants at 9p21 and 22q13 associated with development of cutaneous nevi. *Nat Genet* **41**, 915-9 (2009).
6. Iles, M.M. *et al.* A variant in FTO shows association with melanoma risk not due to BMI. *Nat Genet* **45**, 428-32, 432e1 (2013).
7. Rafnar, T. *et al.* Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat Genet* **41**, 221-7 (2009).
8. Macgregor, S. *et al.* Genome-wide association study identifies a new melanoma susceptibility locus at 1q21.3. *Nat Genet* **43**, 1114-8 (2011).
9. Barrett, J.H. *et al.* Genome-wide association study identifies three new melanoma susceptibility loci. *Nat Genet* **43**, 1108-13 (2011).
10. Ransohoff, K.J. *et al.* Two-stage genome-wide association study identifies a novel susceptibility locus associated with melanoma. *Oncotarget* **8**, 17586-17592 (2017).
11. Duffy, D.L. *et al.* Multiple pigmentation gene polymorphisms account for a substantial proportion of risk of cutaneous malignant melanoma. *J Invest Dermatol* **130**, 520-8 (2010).
12. Beaumont, K.A. *et al.* Receptor function, dominant negative activity and phenotype correlations for MC1R variant alleles. *Hum Mol Genet* **16**, 2249-60 (2007).
13. Tsetskhladze, Z.R. *et al.* Functional assessment of human coding mutations affecting skin pigmentation using zebrafish. *PLoS One* **7**, e47398 (2012).
14. Halaban, R. *et al.* Endoplasmic reticulum retention is a common defect associated with tyrosinase-negative albinism. *Proc Natl Acad Sci U S A* **97**, 5889-94 (2000).
15. Liu, F. *et al.* Genetics of skin color variation in Europeans: genome-wide association studies with functional follow-up. *Hum Genet* **134**, 823-35 (2015).
16. Duffy, D.L. *et al.* Novel pleiotropic risk loci for melanoma and nevus density implicate multiple biological pathways. *Nat Commun* **9**, 4774 (2018).
17. Duffy, D.L. *et al.* IRF4 variants have age-specific effects on nevus count and predispose to melanoma. *Am J Hum Genet* **87**, 6-16 (2010).

18. Choi, J. *et al.* A common intronic variant of PARP1 confers melanoma risk and mediates melanocyte growth via regulation of MITF. *Nat Genet* **49**, 1326-1335 (2017).
19. Fang, J. *et al.* Functional characterization of a multi-cancer risk locus on chr5p15.33 reveals regulation of TERT by ZNF148. *Nat Commun* **8**, 15034 (2017).
20. Ulirsch, J.C. *et al.* Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits. *Cell* **165**, 1530-1545 (2016).
21. Kheradpour, P. *et al.* Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res* **23**, 800-11 (2013).
22. Melnikov, A. *et al.* Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol* **30**, 271-7 (2012).
23. GTEx Consortium *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).
24. Zhang, T. *et al.* Cell-type-specific eQTL of primary melanocytes facilitates identification of melanoma susceptibility genes. *Genome Res* **28**, 1621-1635 (2018).
25. Hoggart, C.J., Whittaker, J.C., De Iorio, M. & Balding, D.J. Simultaneous analysis of all SNPs in genome-wide and re-sequencing association studies. *PLoS Genet* **4**, e1000130 (2008).
26. Verfaillie, A. *et al.* Decoding the regulatory landscape of melanoma reveals TEADS as regulators of the invasive cell state. *Nat Commun* **6**, 6683 (2015).
27. Melnikov, A., Zhang, X., Rogov, P., Wang, L. & Mikkelsen, T.S. Massively parallel reporter assays in cultured mammalian cells. *J Vis Exp* (2014).
28. Coetzee, S.G., Coetzee, G.A. & Hazelett, D.J. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics* **31**, 3847-9 (2015).
29. Kichaev, G. *et al.* Improved methods for multi-trait fine mapping of pleiotropic risk loci. *Bioinformatics* **33**, 248-255 (2017).
30. Alkallas, R., Fish, L., Goodarzi, H. & Najafabadi, H.S. Inference of RNA decay rate from transcriptional profiling highlights the regulatory programs of Alzheimer's disease. *Nat Commun* **8**, 909 (2017).
31. Li, Y.I. *et al.* Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet* **50**, 151-158 (2018).
32. Baylin, S.B. & Jones, P.A. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer* **11**, 726-34 (2011).
33. Kulis, M. *et al.* Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat Genet* **44**, 1236-42 (2012).
34. Maunakea, A.K. *et al.* Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**, 253-7 (2010).
35. Varley, K.E. *et al.* Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res* **23**, 555-67 (2013).
36. Wolff, E.M. *et al.* Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer. *PLoS Genet* **6**, e1000917 (2010).
37. Gordon, S., Akopyan, G., Garban, H. & Bonavida, B. Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene* **25**, 1125-42 (2006).
38. Li, J. *et al.* YY1 regulates melanocyte development and function by cooperating with MITF. *PLoS Genet* **8**, e1002688 (2012).
39. Yeo, N.C. *et al.* An enhanced CRISPR repressor for targeted mammalian gene regulation. *Nat Methods* **15**, 611-616 (2018).
40. Weintraub, A.S. *et al.* YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171**, 1573-1588 e28 (2017).

41. Goujon, C. *et al.* Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. *Nature* **502**, 559-62 (2013).
42. Kane, M. *et al.* MX2 is an interferon-induced inhibitor of HIV-1 infection. *Nature* **502**, 563-6 (2013).
43. King, M.C., Raposo, G. & Lemmon, M.A. Inhibition of nuclear import and cell-cycle progression by mutated forms of the dynamin-like GTPase MxB. *Proc Natl Acad Sci U S A* **101**, 8957-62 (2004).
44. Ceol, C.J. *et al.* The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. *Nature* **471**, 513-7 (2011).
45. Amit, I. *et al.* A module of negative feedback regulators defines growth factor signaling. *Nat Genet* **39**, 503-12 (2007).
46. Katsuoka, F. & Yamamoto, M. Small Maf proteins (MafF, MafG, MafK): History, structure and function. *Gene* **586**, 197-205 (2016).
47. Denat, L., Kadekaro, A.L., Marrot, L., Leachman, S.A. & Abdel-Malek, Z.A. Melanocytes as instigators and victims of oxidative stress. *J Invest Dermatol* **134**, 1512-1518 (2014).
48. Olson, O.C. & Joyce, J.A. Cysteine cathepsin proteases: regulators of cancer progression and therapeutic response. *Nat Rev Cancer* **15**, 712-29 (2015).
49. McIlwain, D.R., Berger, T. & Mak, T.W. Caspase functions in cell death and disease. *Cold Spring Harb Perspect Biol* **5**, a008656 (2013).
50. Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92-94 (2017).
51. Stacey, S.N. *et al.* New basal cell carcinoma susceptibility loci. *Nat Commun* **6**, 6825 (2015).
52. Barrett, J.H. *et al.* Fine mapping of genetic susceptibility loci for melanoma reveals a mixture of single variant and multiple variant regions. *Int J Cancer* **136**, 1351-60 (2015).
53. Stephens, M. & Balding, D.J. Bayesian statistical methods for genetic association studies. *Nat Rev Genet* **10**, 681-90 (2009).
54. Vignal, C.M., Bansal, A.T. & Balding, D.J. Using penalised logistic regression to fine map HLA variants for rheumatoid arthritis. *Ann Hum Genet* **75**, 655-64 (2011).
55. Long, J.S. & Ervin, L.H. Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *The American Statistician* **54**, 217-224 (2000).
56. Kulakovskiy, I.V. *et al.* HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res* **46**, D252-D259 (2018).
57. Ongen, H., Buil, A., Brown, A.A., Dermitzakis, E.T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479-85 (2016).
58. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
59. Assenov, Y. *et al.* Comprehensive analysis of DNA methylation data with RnBeads. *Nat Methods* **11**, 1138-1140 (2014).
60. Aryee, M.J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-9 (2014).
61. Spruijt, C.G., Baymaz, H.I. & Vermeulen, M. Identifying specific protein-DNA interactions using SILAC-based quantitative proteomics. *Methods Mol Biol* **977**, 137-57 (2013).
62. Hubner, N.C., Nguyen, L.N., Hornig, N.C. & Stunnenberg, H.G. A quantitative proteomics tool to identify DNA-protein interactions in primary cells or blood. *J Proteome Res* **14**, 1315-29 (2015).
63. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-72 (2008).

64. Chari, R., Yeo, N.C., Chavez, A. & Church, G.M. sgRNA Scorer 2.0: A Species-Independent Model To Predict CRISPR/Cas9 Activity. *ACS Synth Biol* **6**, 902-904 (2017).
65. Lagonigro, M.S. *et al.* CTAB-urea method purifies RNA from melanin for cDNA microarray analysis. *Pigment Cell Res* **17**, 312-5 (2004).
66. Liao, Y., Smyth, G.K. & Shi, W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* (2019).
67. Varet, H., Brillet-Gueguen, L., Coppee, J.Y. & Dillies, M.A. SARTools: A DESeq2- and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. *PLoS One* **11**, e0157022 (2016).
68. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-40 (2010).
69. Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-8 (2016).
70. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417-419 (2017).
71. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

## Figure Legends

**Figure 1. MPRA identified 39 functional variants from 16 melanoma GWAS loci** Volcano plots of MPRA results for each melanoma GWAS locus. Inverse P-values and effect sizes of allelic difference from UACC903 transfections are shown for each of the 16 loci tested. Dashed horizontal lines indicate the FDR 1% cutoff. The most significant SNP from each locus is labeled. Putative function of significant MPRA variants are shown as activator, repressor, or both (expression levels of either allele is higher, lower, or higher and lower than those of scrambled sequence).

**Figure 2. rs398206 is a functional *cis*-regulatory variant and a significant *cis*-eQTL for *MX2* levels in melanocytes** (A) Variants that were tested in MPRA from the Chr21q22.3 melanoma locus are shown relative to the genomic position of *MX2*. Only the variants from the primary GWAS signal are shown (the other three from a secondary signal are located upstream of the *MX2* genic region). ChromHMM annotation (Primary Core Marks segmentation) of Penis Foreskin Melanocyte Primary Cells from Roadmap Epigenomics Project is shown

(Red/OrangeRed: Active_TSS/Flanking_Active_TSS, Yellow/GreenYellow: Enhancers/Genic_enhancers, Green/DarkGreen: Strong_transcription/Weak_transcription). Melanoma FAIRE-seq track of 11 samples is from a study by Verfaillie and colleagues[26]. Ensembl predicted transcripts from archive 75 are shown. (B) Transcriptional activity of 145bp sequences encompassing rs398206 from MPRA are shown as normalized tag counts ($log_2$(RNA TPM/DNA TPM)). Results from UACC903 melanoma cells are shown for both alleles in forward and reverse directions, where results from promoter and enhancer constructs were combined. Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots. Density is reflected in the width of the shape. (C) Individual luciferase activity assays of 145bp sequences encompassing rs398206 is shown for UACC903. pGL4.23 construct including minimal TATA promoter was used. One representative set is shown from three biological replicates. Mean with SEM, n = 6. All constructs are significantly higher than pGL4.23 (TATA) control ($P < 0.0001$, two-tailed, unpaired t-test assuming unequal variance). (D) eQTL plot of *MX2* levels in primary melanocytes in relation to rs398206 is shown for three genotype groups.

**Figure 3. rs398206 displays allele-preferential binding to YY1** (A) Quantitative mass-spectrometry of rs398206 using nuclear extract of UACC903 and 21bp double-stranded DNA probes with A (risk) or C (protective) alleles. A-allele specific interacting proteins are shown in the bottom right quadrant, and C-allele specific interactors in the top left quadrant. Using label-swapping of high- or low-mass label, the A-bound/C-bound ratio is shown on the x-axis, and the C-bound/A-bound ratio on the y-axis. Only the protein names of inter-quartile > 3 are shown. (B) YY1 binding motif is shown as a position weight matrix at the top (motif obtained from HOCOMOCO database and plotted using weblogo3). The genomic sequence surrounding rs398206 is shown at the bottom with the risk-associated A allele matching the consensus YY1-

binding motif. Genomic positions are hg19. (C-D) EMSAs using 21bp double-stranded DNA probes with A or C alleles of rs398206 and nuclear extract from UACC2331 melanoma cells (C) or purified recombinant YY1 protein (D). Antibody super-shift using anti-YY1 antibody is shown at the last lanes of (C), where the A-specific band (arrows) is diminished.

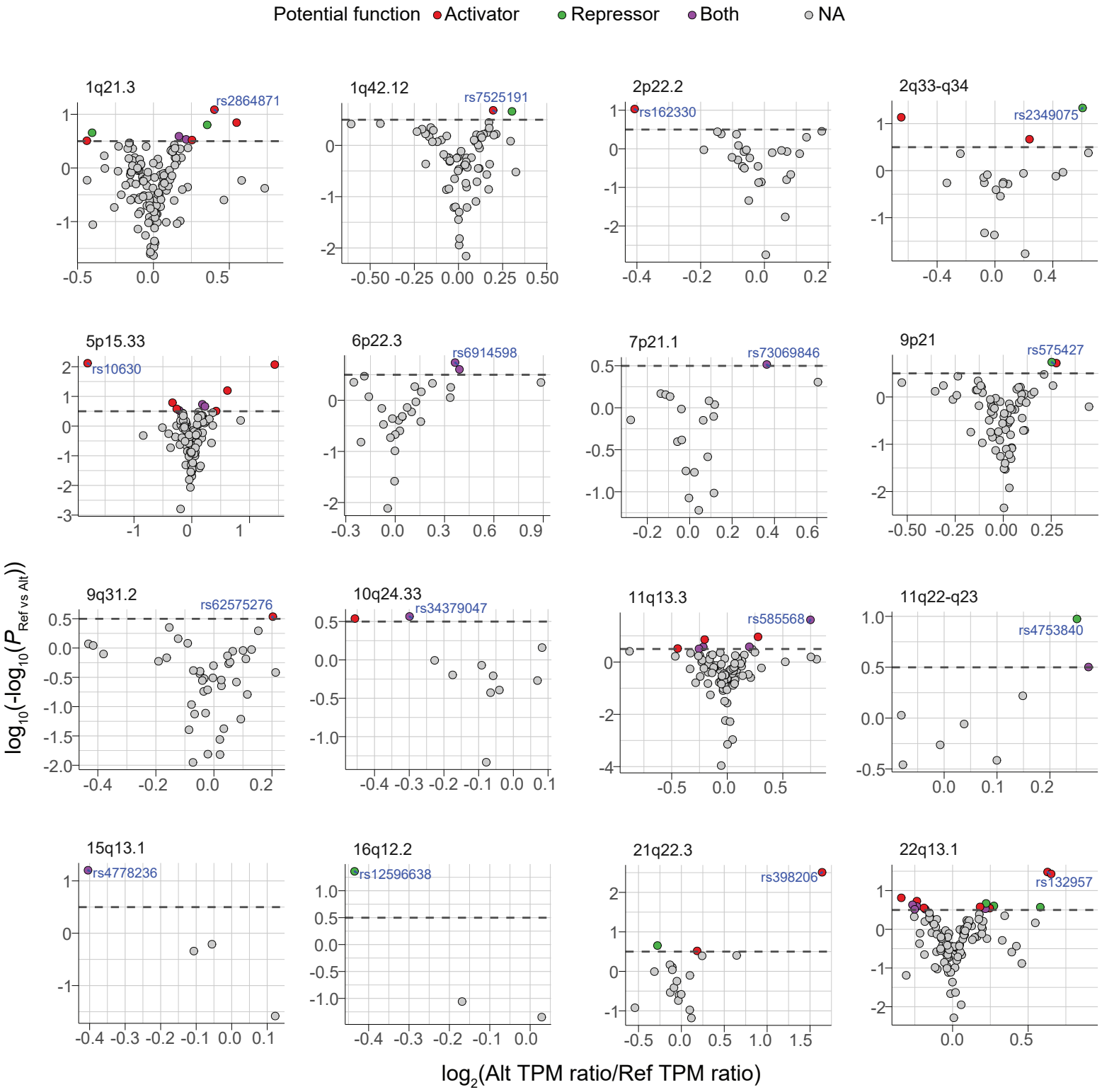
**Figure 4. YY1 preferentially binds to the rs398206-A allele in melanoma cells** (A) Chromatin immunoprecipitation using anti-YY1 antibody or normal IgG followed by qPCR. On the top panel, genomic positions of the amplicons using five qPCR primer sets are shown relative to rs398206 (red dashed line) in the first intron of *MX2*. qPCR results from two melanoma cell lines are shown on the bottom panel with genotype for rs398206 indicated in the X-axis label. Relative quantities are shown as fold over input DNA. Mean of PCR triplicates with SEM are plotted. One representative set of three biological replicates for each cell line is shown. (B) Taqman genotyping of rs398206 using YY1 ChIP DNA in UACC647 melanoma cells (heterozygous for rs398206). The blue shaded areas mark HapMap CEU DNA controls showing separation of three genotype clusters. qPCR triplicates were plotted separately. One representative set of three biological replicates is shown. Normalized A and C allele intensity is shown as dRn values on x and y-axis, respectively. (C) Combined plot for A/C allelic ratio from three rounds of biological replicates of UACC647. Each dot represents A/C ratio calculated from $2^{-(\text{average dCt(alleleA-alleleC)})}$ from qPCR triplicates. Input includes three 10-fold serial dilutions from each round. Mean with SEM are plotted. Two-tailed Mann-Whitney U test was used.

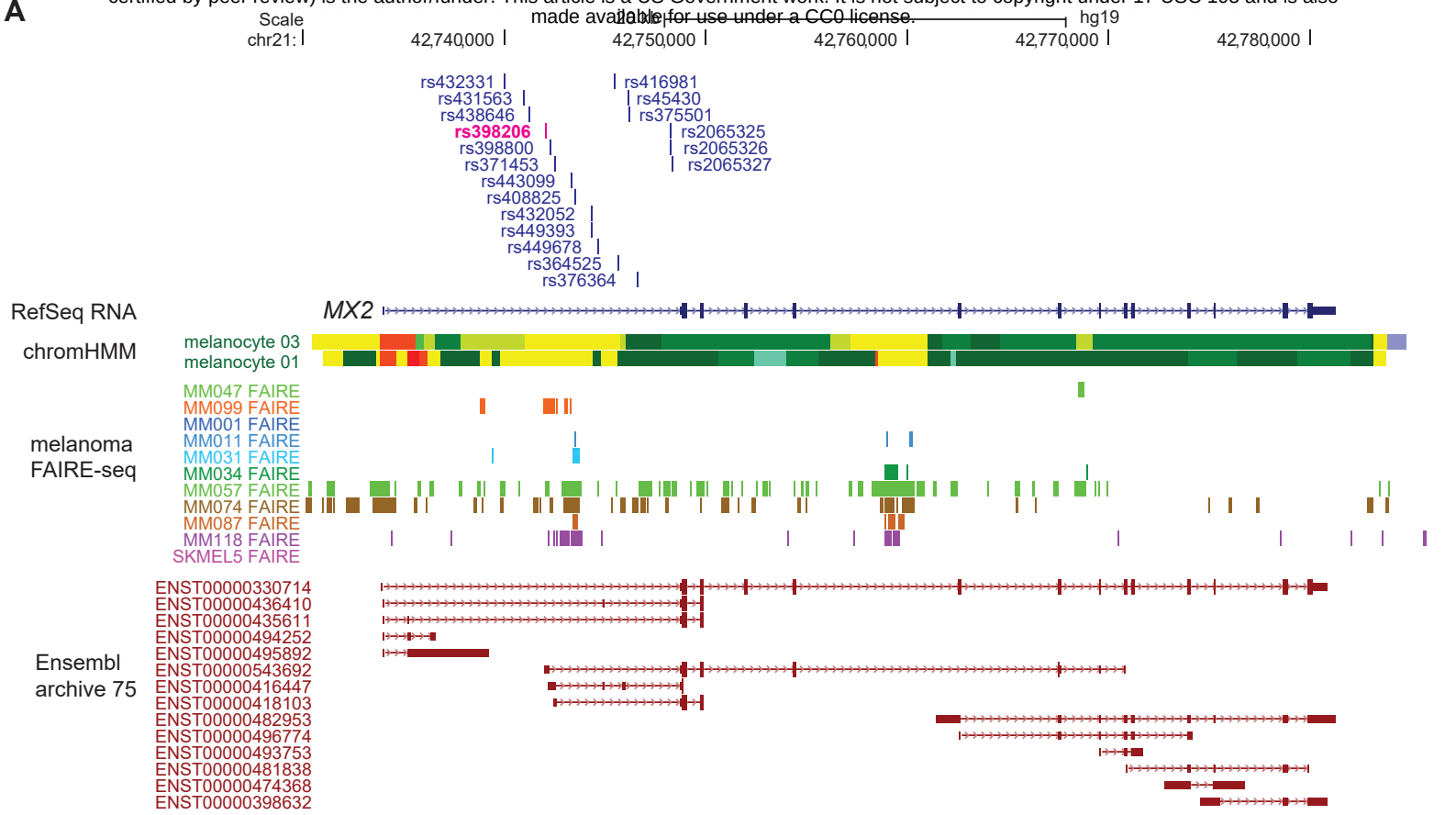
**Figure 5. YY1 and rs398206 affect *MX2* expression in melanoma cells** (A) YY1 was knocked down using four different siRNAs in UACC647 cells, and *MX2* levels were measured. *GAPDH*-normalized *MX2* mRNA levels are shown as fold change over those from non-targeting siRNA. Four biological replicates of n = 6 were combined (total n = 24). (B) CRISPRi using dCAS9-KRAB-MeCP2 and four different gRNAs targeting rs398206 in UACC903 cells. *MX2* mRNA levels (GAPDH-normalized) are shown as fold change over those from non-targeting gRNA.

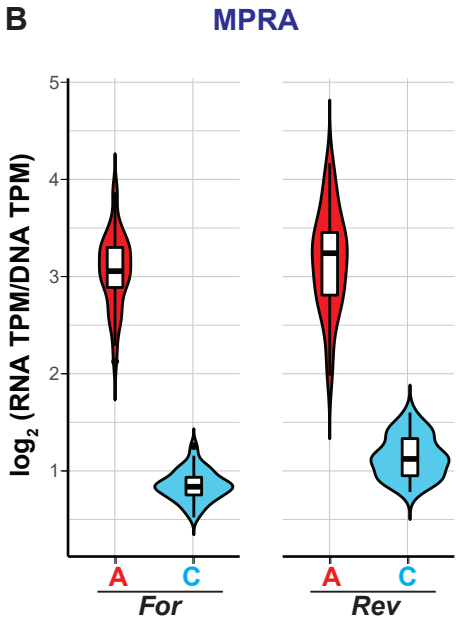
Three biological replicates of n = 6 were combined (total n = 18, except gRNA-3, n=17). gRNA 1, 3, and 4 directly overlap rs398206. gRNA 2 targets ~25bp upstream of rs398206. AAVS1 (gRNA targeting adeno-associated virus integration site on Chr19). Box: Median and $25^{th}$ to $75^{th}$ percentile. Whisker: $10^{th}$ to $90^{th}$ percentile. P - values are shown from one-sample Wilcoxon test (two-sided) for difference from non-targeting siRNA/gRNA. Dotted line denotes the *MX2* levels in non-targeting siRNA/gRNA control.

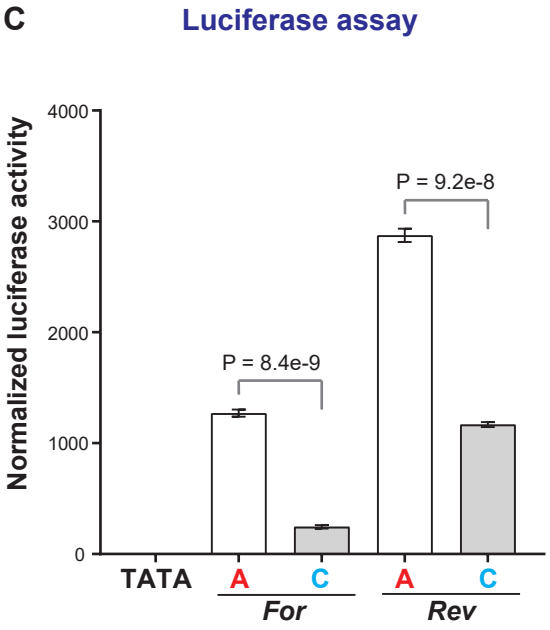
**Figure 6. MX2 accelerates melanoma formation** (A-B) Cell growth and movement of human primary melanocytes (A) or melanoma cell line UACC2545 (B) infected with an inducible lentiviral construct of *MX2* cDNA or Empty pINDUCER20 vector were measured on xCELLigence system. Cell Index values were normalized relative to those at the time of doxycycline addition (dotted vertical line: Dox). The amount of doxycycline is shown in ng/ml and color-coded. Mean Normalized Cell Index (colored dots) and SD (gray vertical lines) are plotted (n = 3). A representative set of three biological replicates is shown. (C-D) Ingenuity Pathway Analysis of differentially expressed genes from *MX2*-high versus *MX2*-low melanocytes from 106 individuals (C) or RNA sequencing of *MX2*-overexpressed versus control melanocytes from 3 individuals (D). (C) 252 differentially expressed genes (FDR < 1% and >2-fold change) between *MX2*-high and *MX2*-low melanocytes (top and bottom quantile based on *MX2* levels; n = 28 each) from 106 individuals were used as input for the analysis. (D) 158 differentially expressed genes (FDR < 10%) between *MX2*-overexpressing (100 ng/ml doxycycline) versus control (no doxycycline) melanocytes using 3 biological replicates for 3 individuals were used as input for the analysis. Significantly enriched canonical pathways (P < 0.05 and |Z-score| >1) are color-coded for the direction of effect relative to *MX2*-high melanocytes (C) or *MX2*-overexpressing melanocytes (D). A weaker to stronger shade of each color represent the relative magnitude of Z-scores: Positive Z-score between 1 and 2.646 and negative Z-score between -1 and -3.464 (C) or positive between 1 and 1.134 and negative

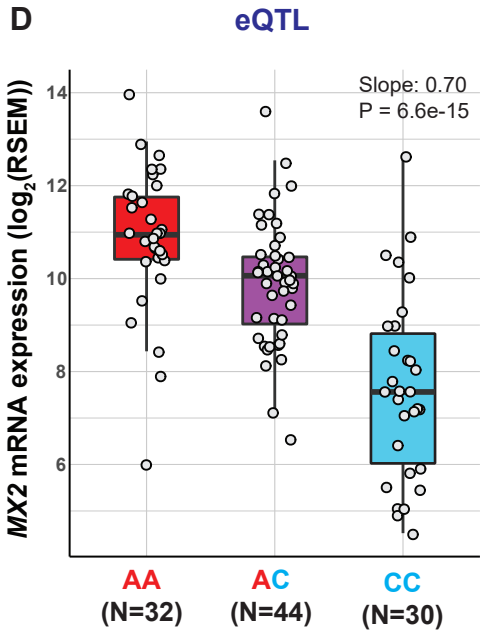
between -1.342 and -2.236 (D), where lightest red is closer to 1 and lightest blue is closer to -1. (E) Melanoma-free survival curves of a zebrafish melanoma model[44] (Tg(*mitfa:BRAF*$^{V600E}$), *p53*-/-, *mitfa*-/-). The fish were injected at the one cell stage with either miniCoopR *mitfa:MX2* or miniCoopR *mitfa:EGFP* and monitored weekly for melanoma formation. The percentage of melanoma-free fish was combined from three independent experiments and plotted. Log-rank test was used.



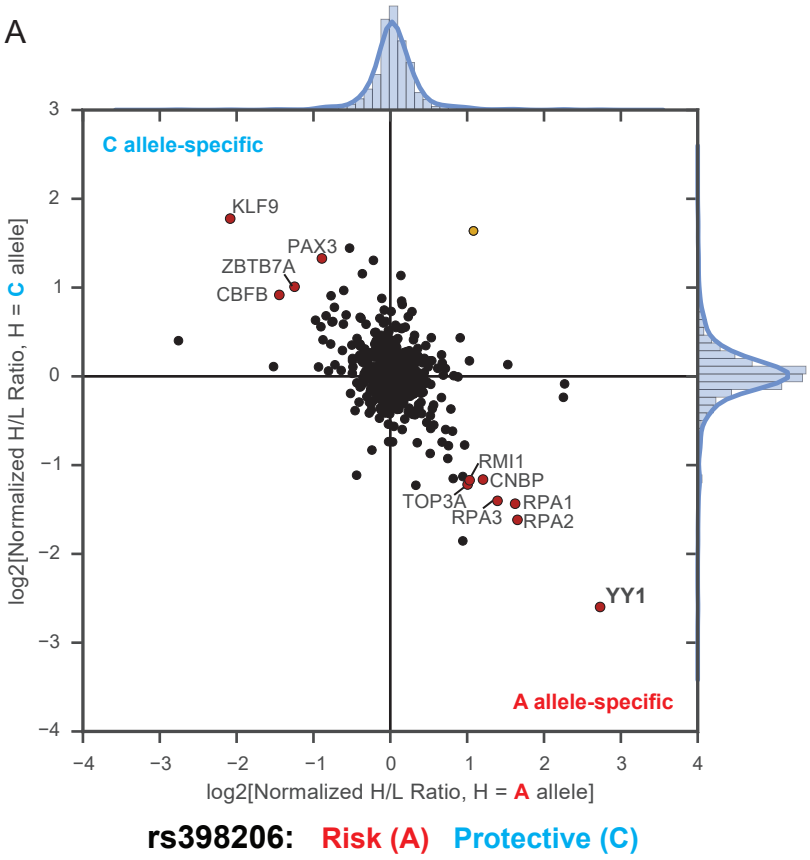


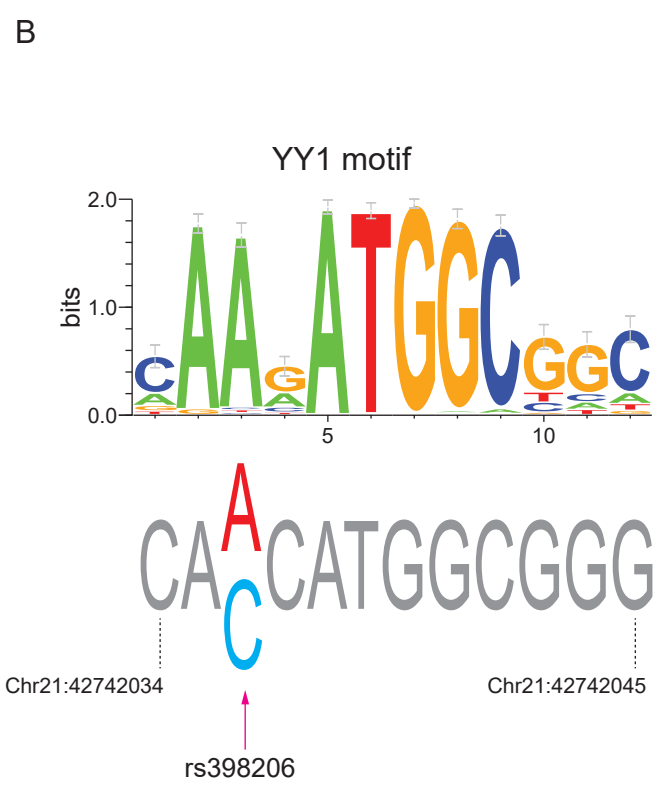



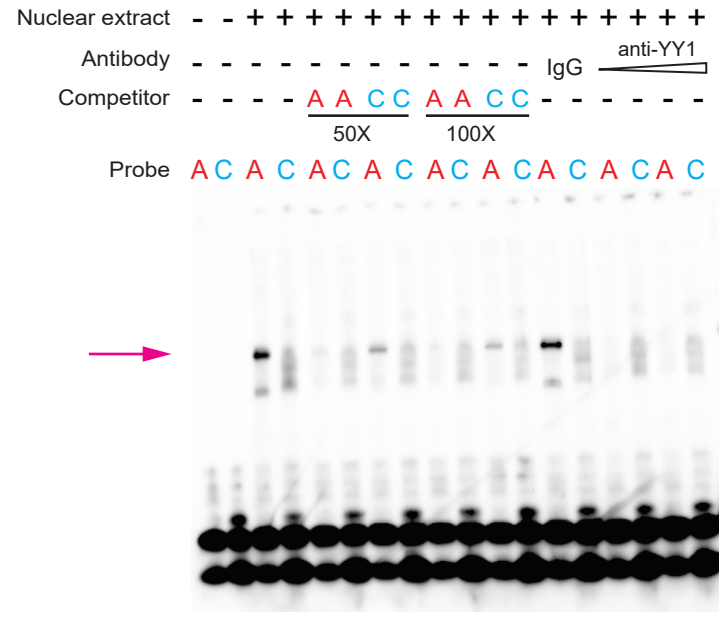



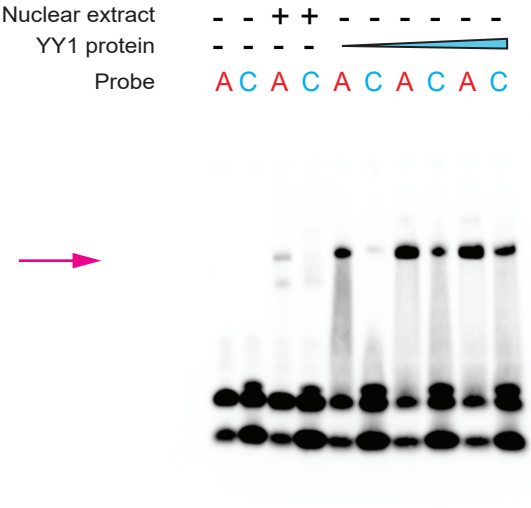



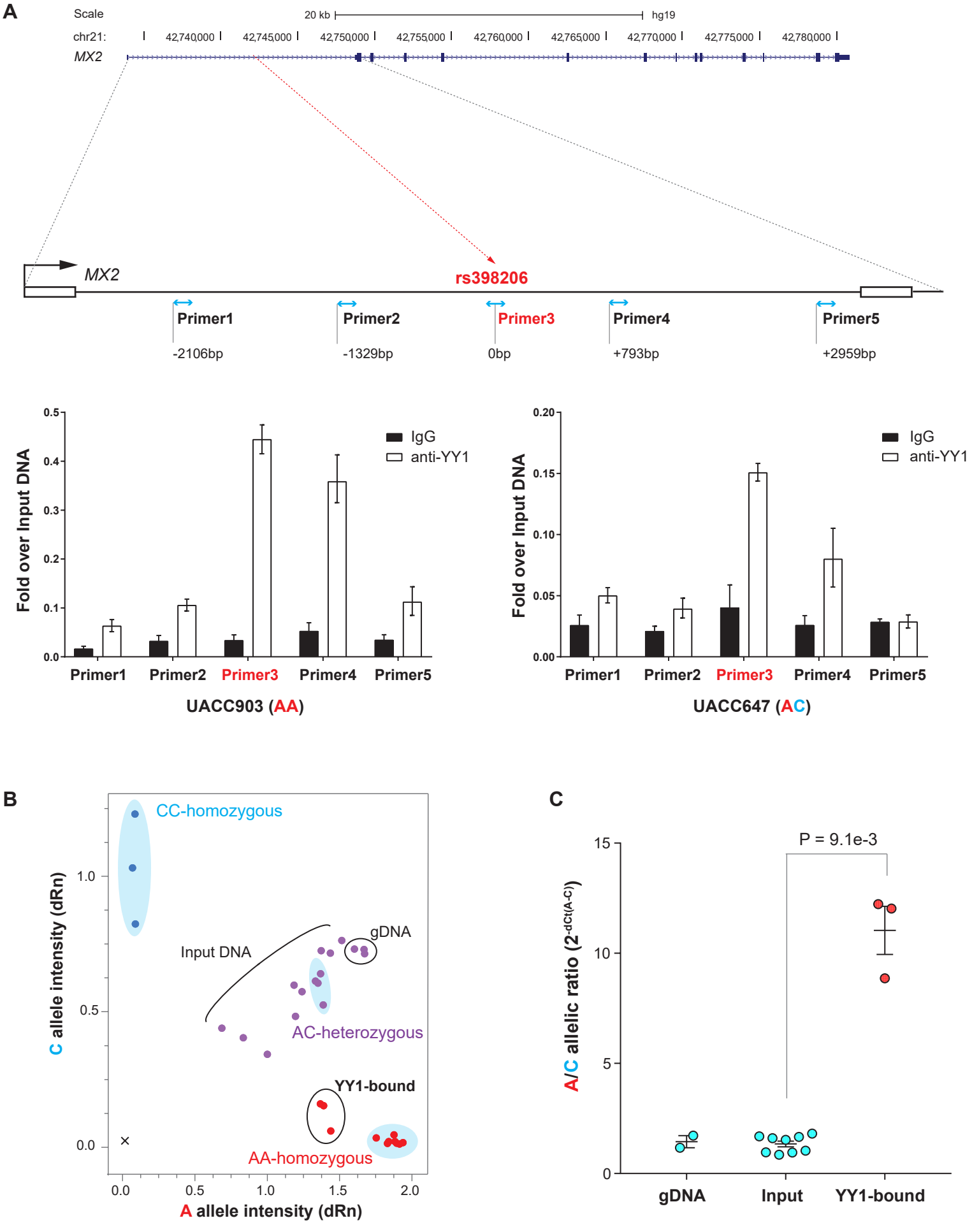

**rs398206: Risk (A) Protective (C)**



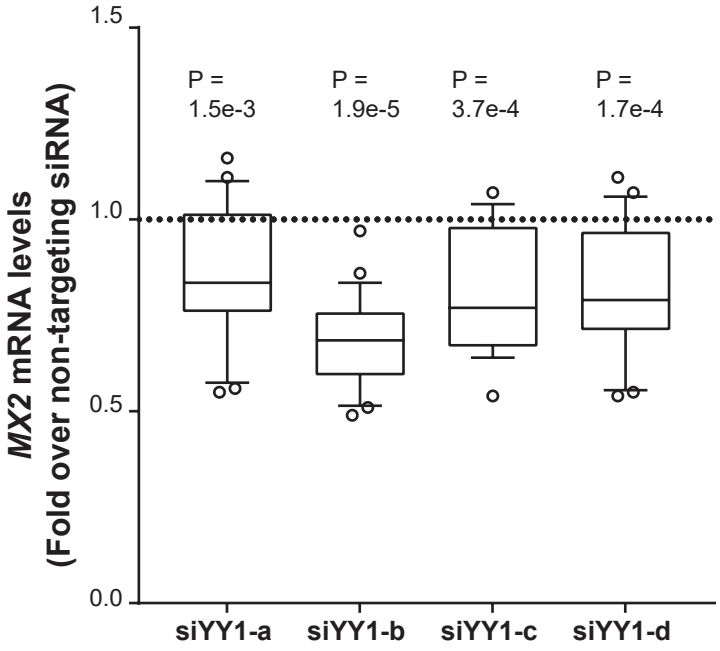



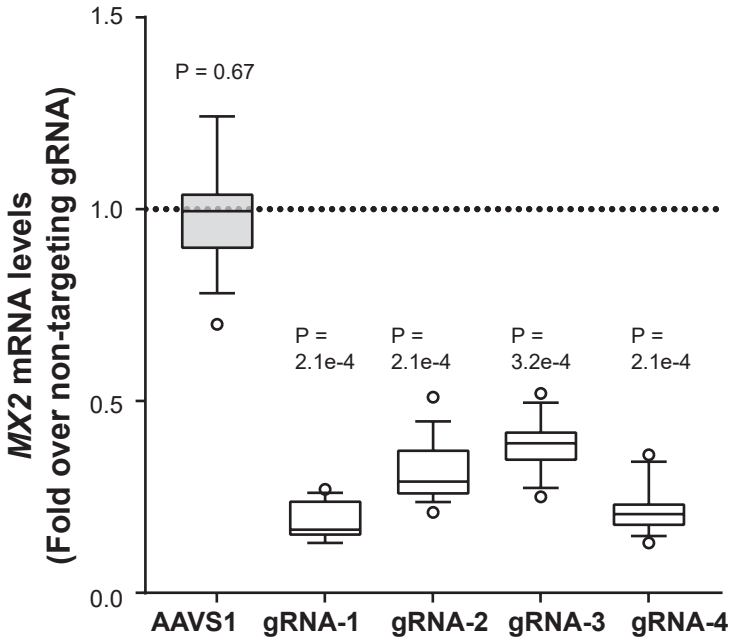





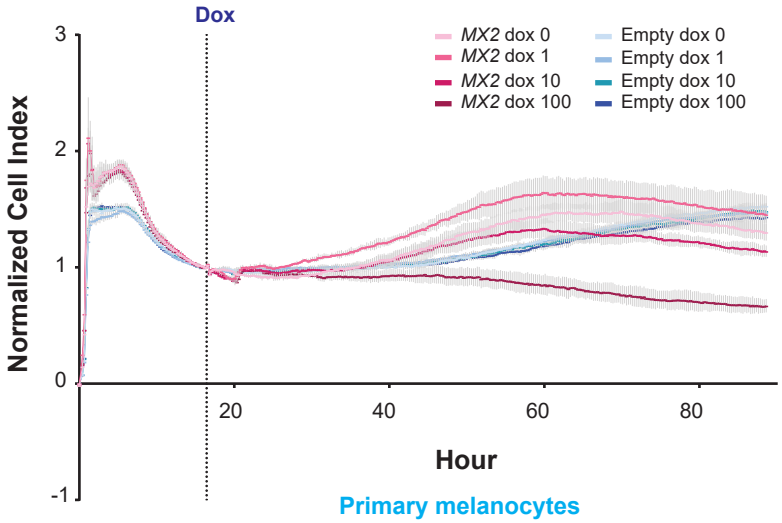
A



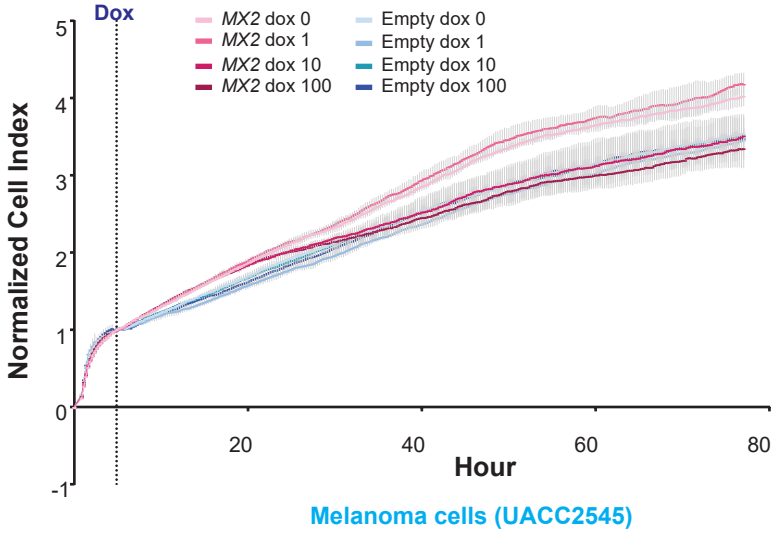

B



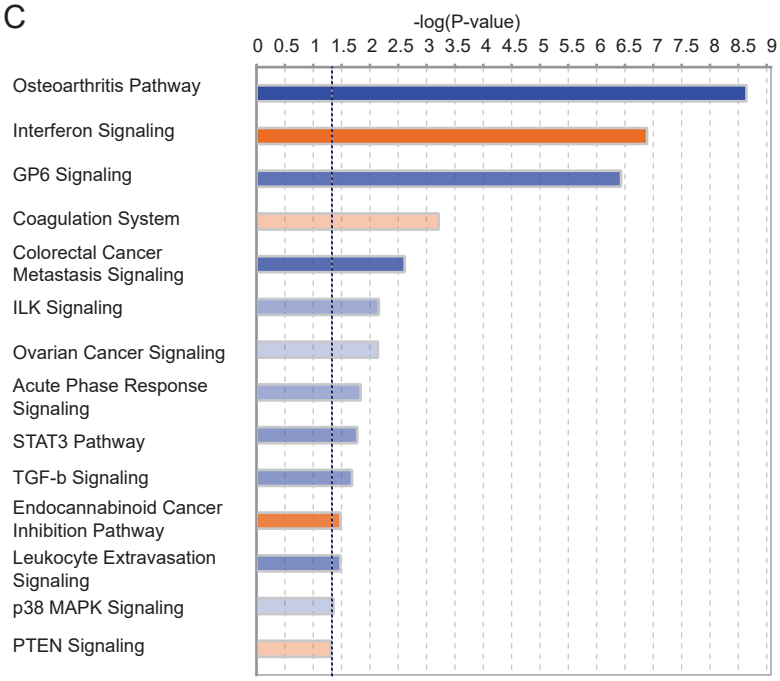
A



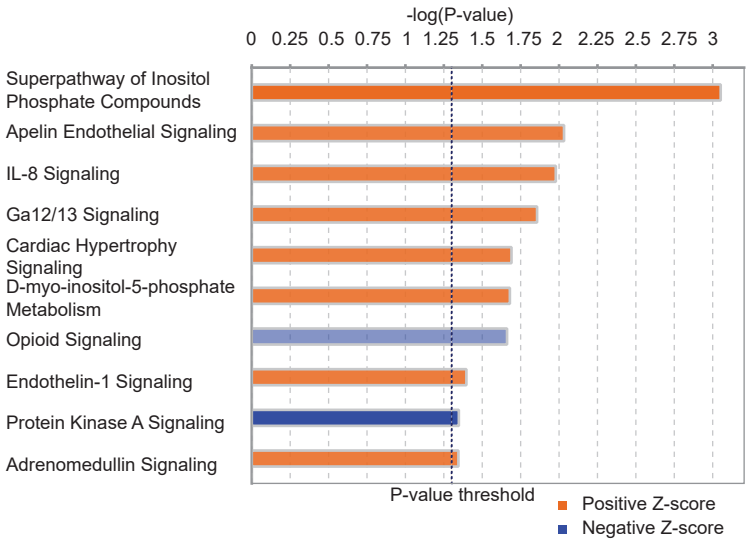

B



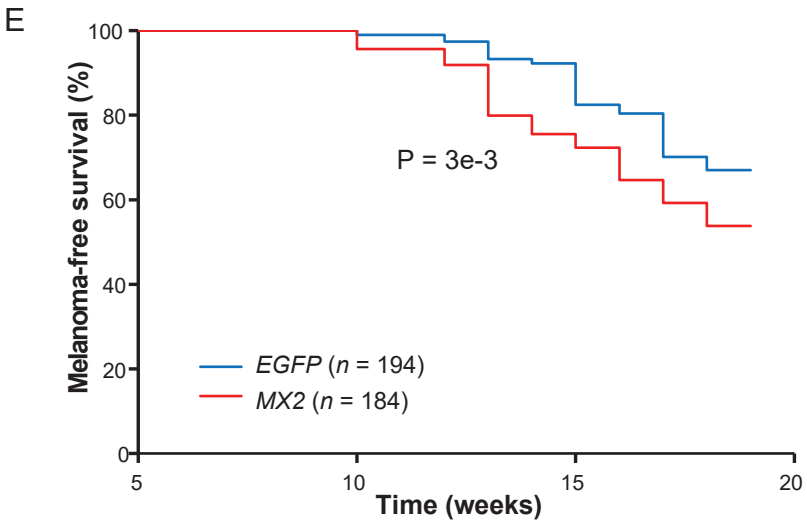

C




D




E



## Supplementary Material

### Splice-QTL analyses of *MX2*

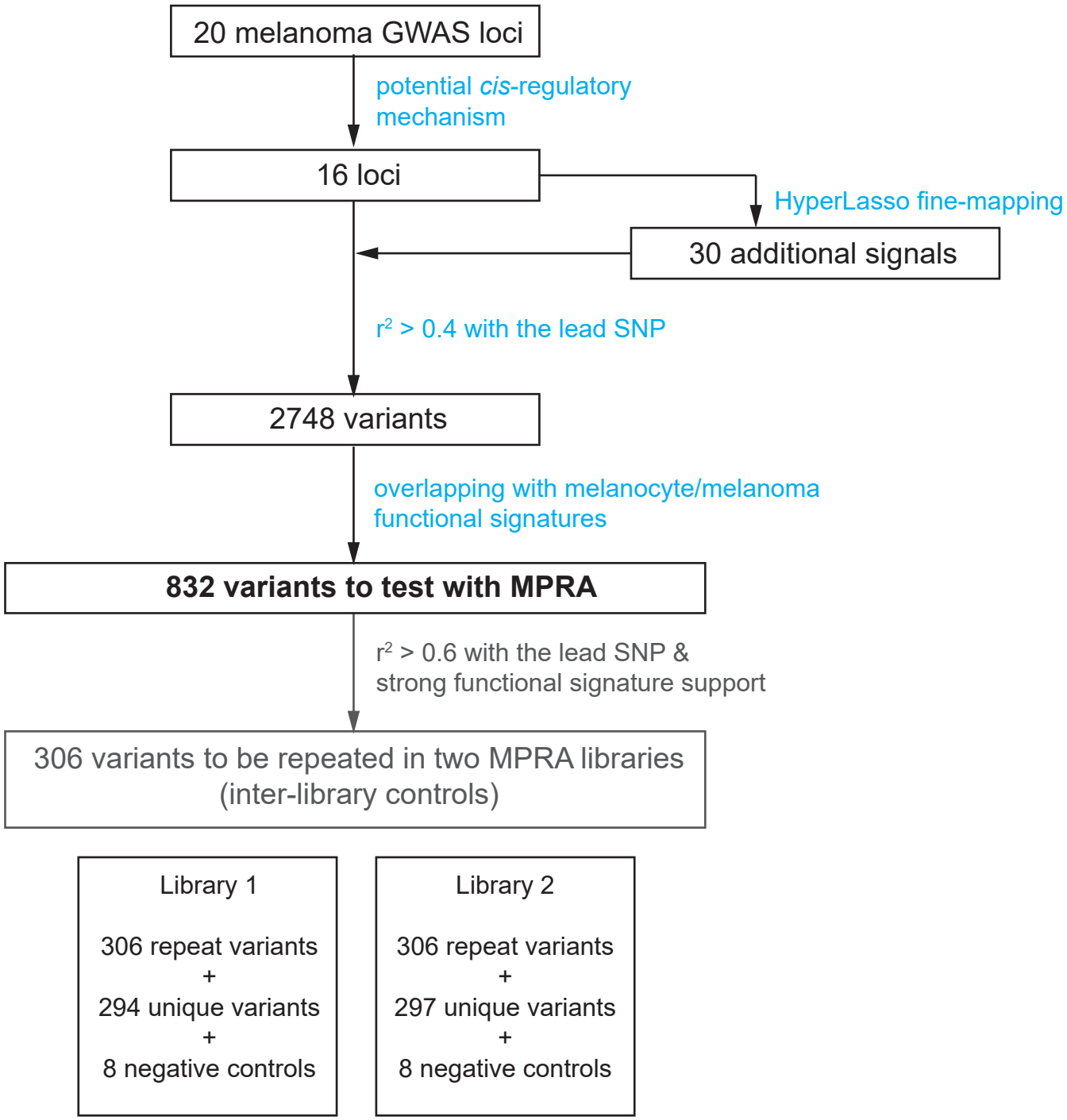
For sQTL we used LeafCutter[1] to assess splice junction-level QTL focusing on alternative intron exclusion and exon joining within a shared cluster as opposed to estimated isoform-levels. sQTL initially indicated that rs398206 was associated with an alternative intron excision of *MX2* in primary melanocytes in the opposite direction of *MX2* eQTL (*P* = 7.79e-7, slope = -0.53; **Supplementary Fig 11A-B**). To seek additional support of this observation in other tissue types, we performed sQTL analysis in 44 GTEx tissue types. In 21 tissue types including blood, testis, ovary, and fibroblasts, the same pattern of significant sQTLs were observed for rs398206 or correlated SNPs (lowest D' = 0.87, EUR), where the protective, C allele favors an alternative junction usage producing an alternative *MX2* transcript (ENST00000418103). Moreover, in 10 tissue types, this sQTL was reciprocated by risk, A allele favoring the junction usage producing the full-length transcript (ENST00000330714), raising a potential hypothesis of alternative promoter usage of *MX2* in these tissue types. Thorough inspection of raw data in melanocytes, however, indicated that the finding was driven by the junction reads of low coverage (i.e. < 4% of the samples showed three or more reads mapped to the junction spanning Chr21:42742322:42748763) (**Supplementary Fig 11C-F**). Since this junction in melanocytes was not mapped to the reference genome (Ensembl75), nor was it detected in PacBio long-read sequencing (data not shown), we performed isoform-specific qPCR of two other *MX2* alternative transcript isoforms (ENST00000543692 and ENST00000418103), which are predicted to use similar junctions. The results for both isoforms displayed similar expression patterns to that of the full-length isoform (ENST00000330714) relative to rs398206 genotypes, suggesting that sQTL finding was false-positive (**Supplementary Fig 11C-F**). Together these data suggest the main effect of *MX2* eQTL in melanocytes was not driven by alternative isoforms or splicing events.

### *MX2* and immune infiltrates in melanomas

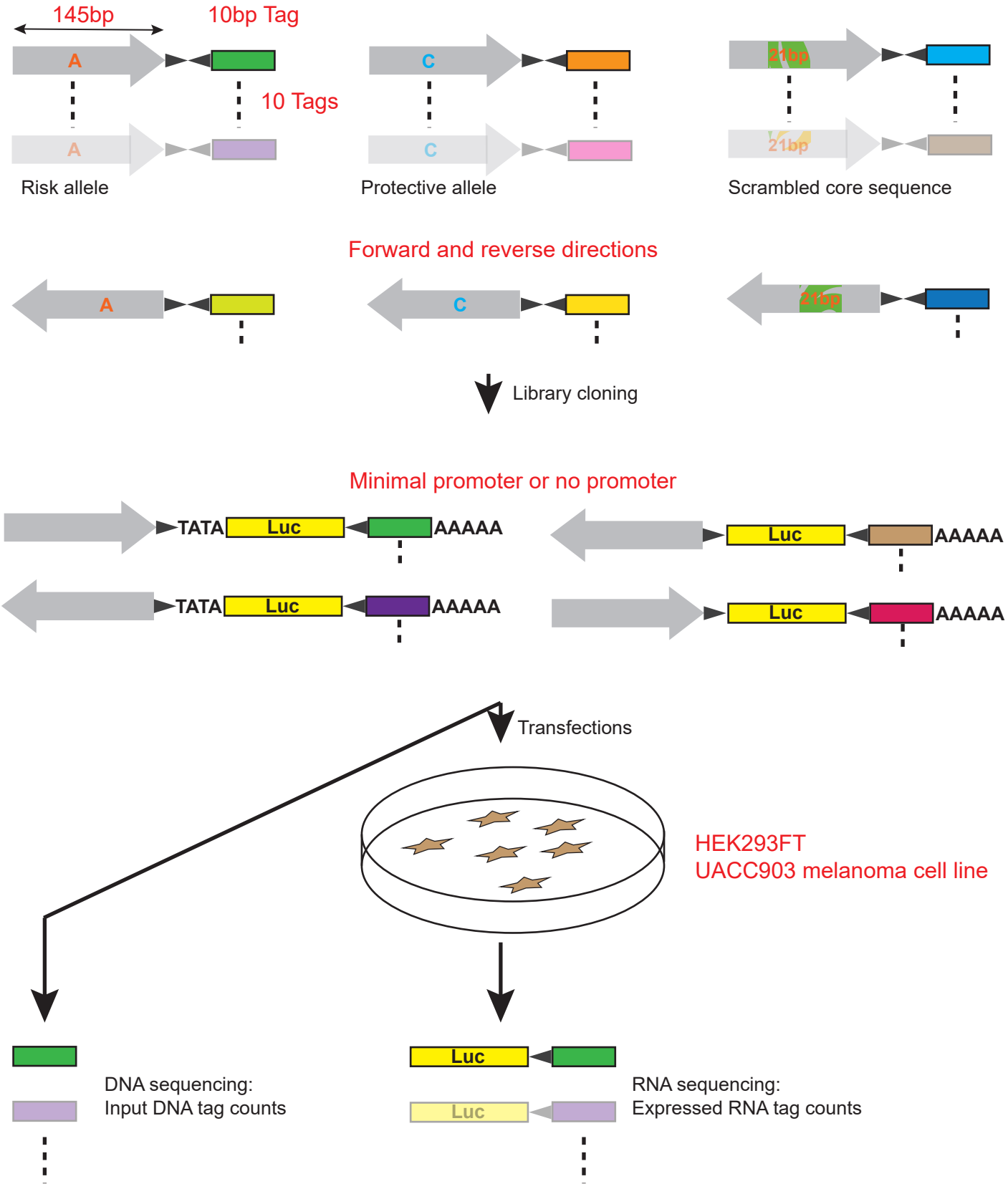
To explore the possibility that *MX2* plays its roles mainly through immune response during melanomagenesis, we also asked if *MX2* levels are correlated with immune cell infiltration in TCGA melanoma samples. Using cell type deconvolution programs, TIMER[2] and CIBERSORT[3], we observed weak correlations between *MX2* levels and infiltration of CD4+ T cells, neutrophils and dendritic cells among 6 cell type models (TIMER; purity-corrected partial Pearson correlation *r* = 0.221, 0.279, 0.273, and *P* = 2.36e-6, 1.68e-9, 4.31e-9, respectively; **Supplementary Fig 17A**). When we examined correlations with proportions of 22 types of immune cells established by CIBERSORT, we did not observe a significant correlation with *MX2* levels (data not shown). Instead, weak correlations between melanoma-associated rs398206 A allele count and fractions of Macrophage M1 and M2 were observed (Pearson correlation *r* = 0.204 and 0.211, and *P* = 0.01 and 0.013, respectively; n = 147 samples with deconvolution *P* < 0.05; **Supplementary Fig 17B**). Together these data did not provide sufficient evidence that *MX2* roles in melanomagenesis are mainly through its roles in immune cell infiltration to tumor.

## References

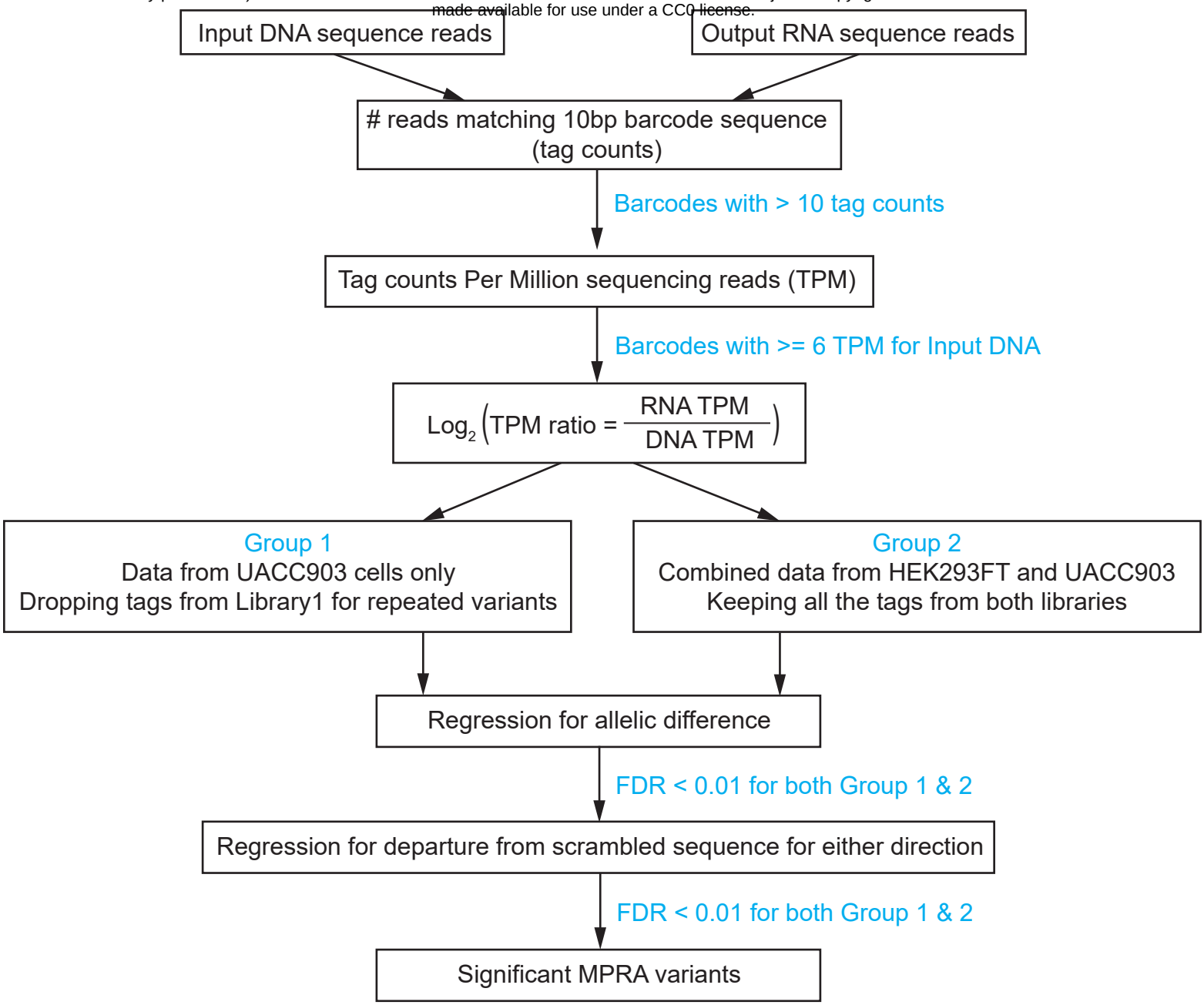
1. Li, Y.I. *et al.* Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet* **50**, 151-158 (2018).
2. Li, T. *et al.* TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res* **77**, e108-e110 (2017).
3. Newman, A.M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12**, 453-7 (2015).
4. Li, B. & Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
5. Zhang, C., Zhang, B., Lin, L.-L. & Zhao, S. Evaluation and comparison of computational tools for RNA-seq isoform quantification. *BMC Genomics* **18**, 583 (2017).
6. Weintraub, A.S. *et al.* YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171**, 1573-1588 e28 (2017).



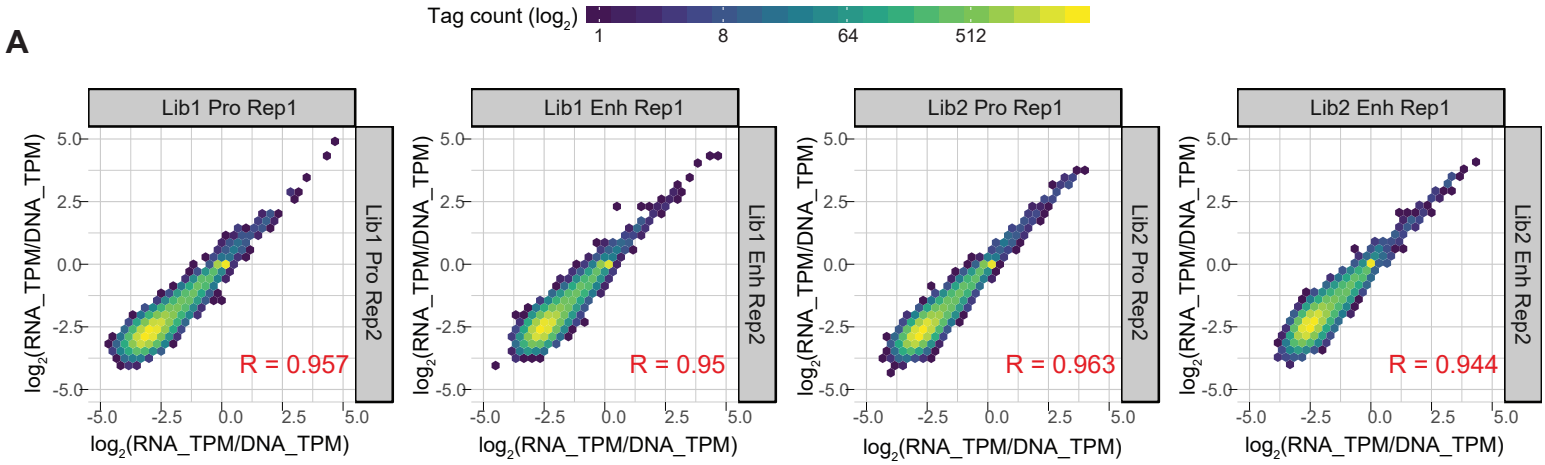
**Supplementary Figure 1** Variant selection from melanoma GWAS loci for MPRA.

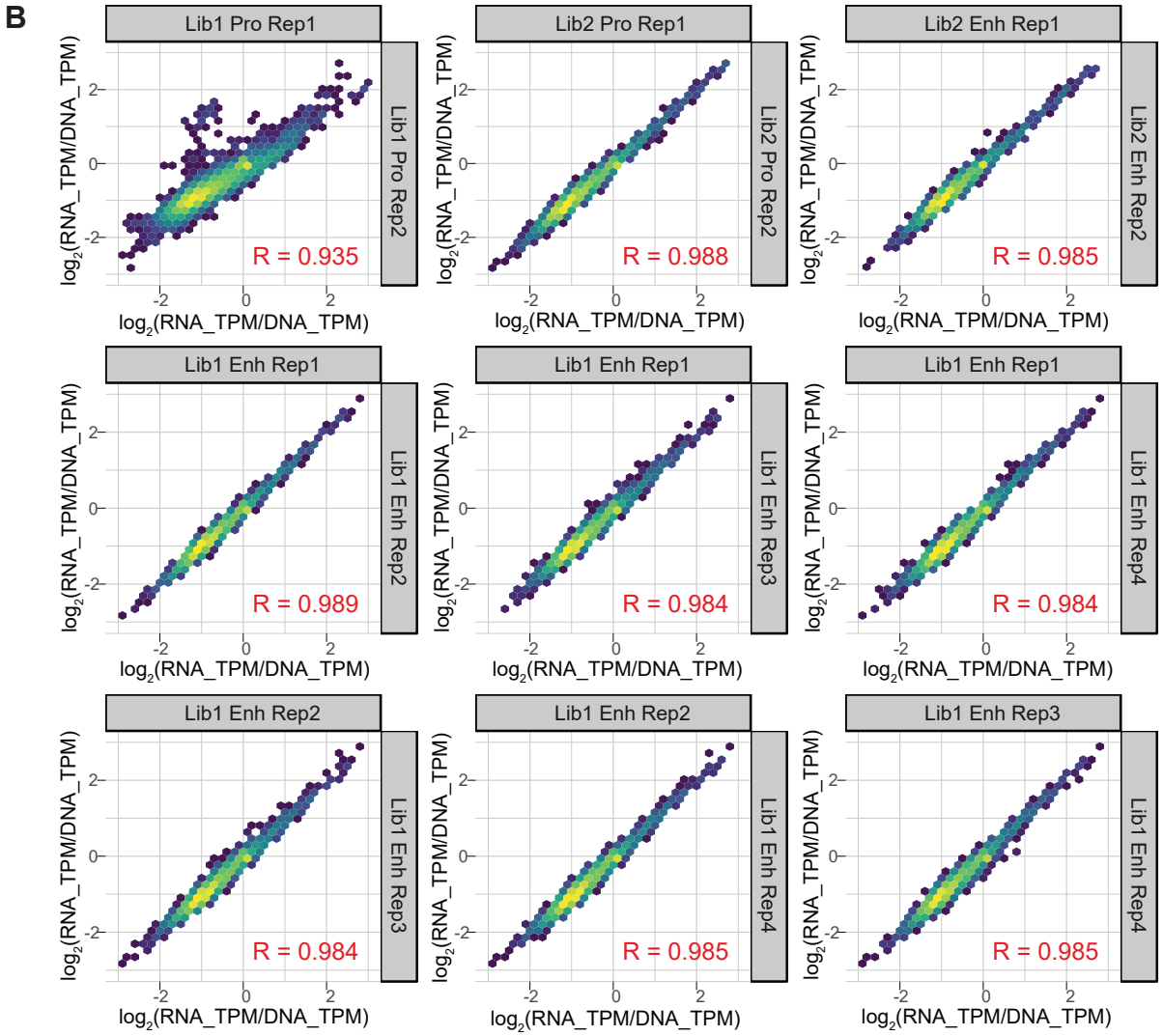


**Supplementary Figure 2** MPRA workflow. Oligo libraries were synthesized using 145bp of sequence encompassing each variant with risk or protective alleles or a scrambled sequence for core 21 bases in both forward and reverse directions, that was flanked by enzyme recognition sites and sequencing primer sequences as well as 10bp barcodes (10 tags per unique sequence). Libraries were cloned into luciferase constructs with or without a minimal TATA promoter. Cloned libraries were then transfected into HEK293FT cells or UACC903 melanoma cells to generate expressed RNA tag libraries. Both input DNA and RNA libraries were sequenced to assess the tag counts associated with the test sequences.

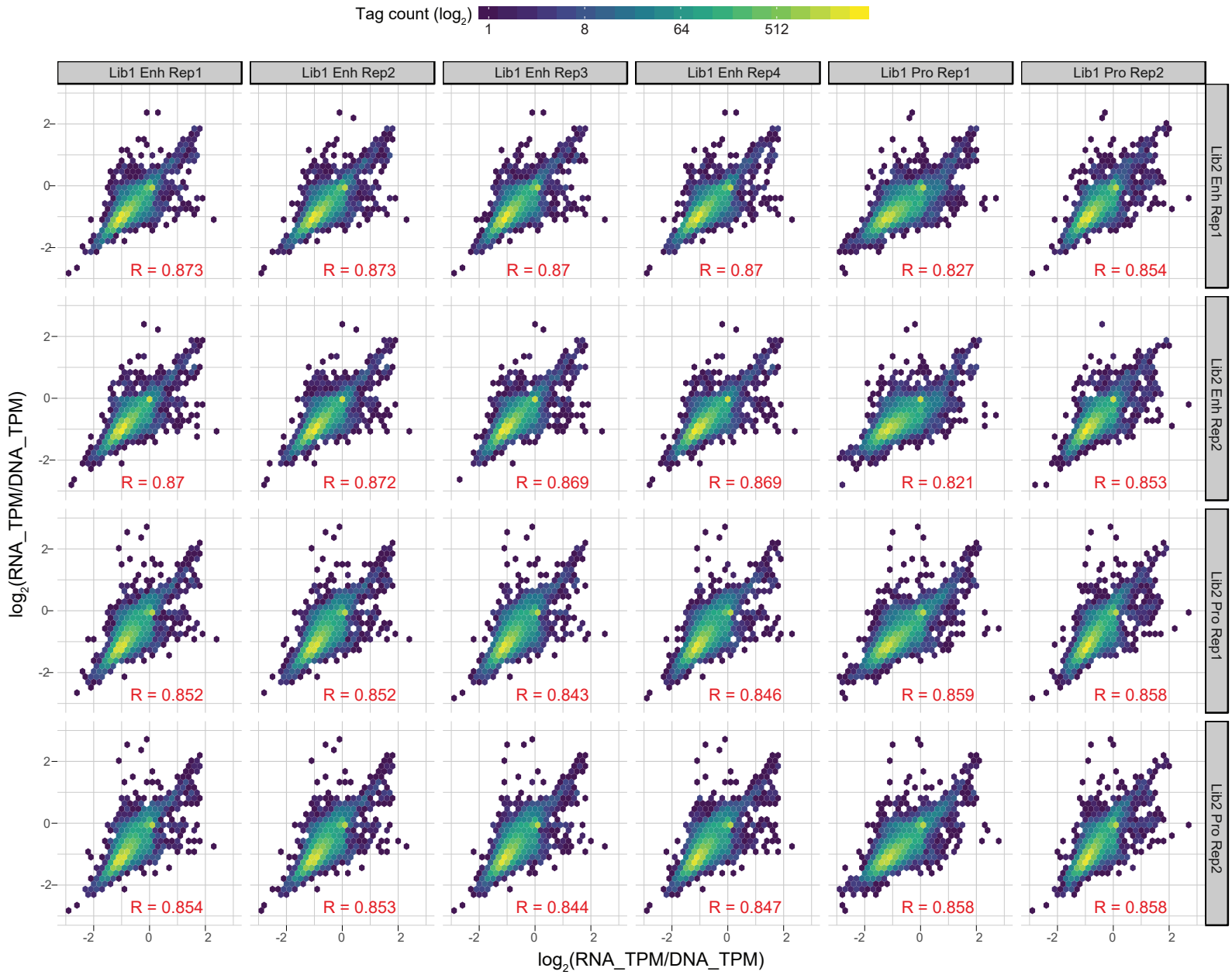


**Supplementary Figure 3** MPRA data analyses workflow.

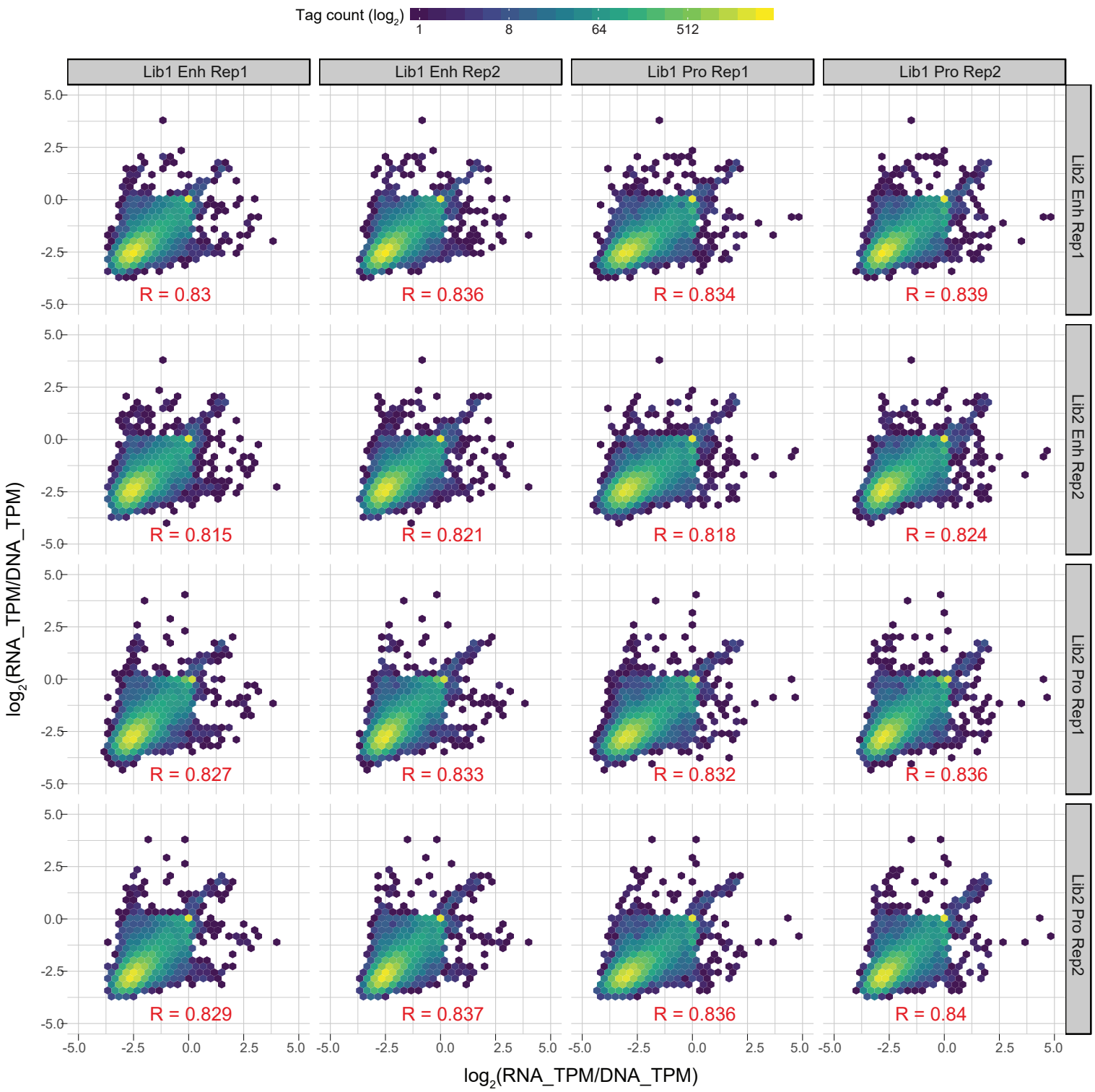


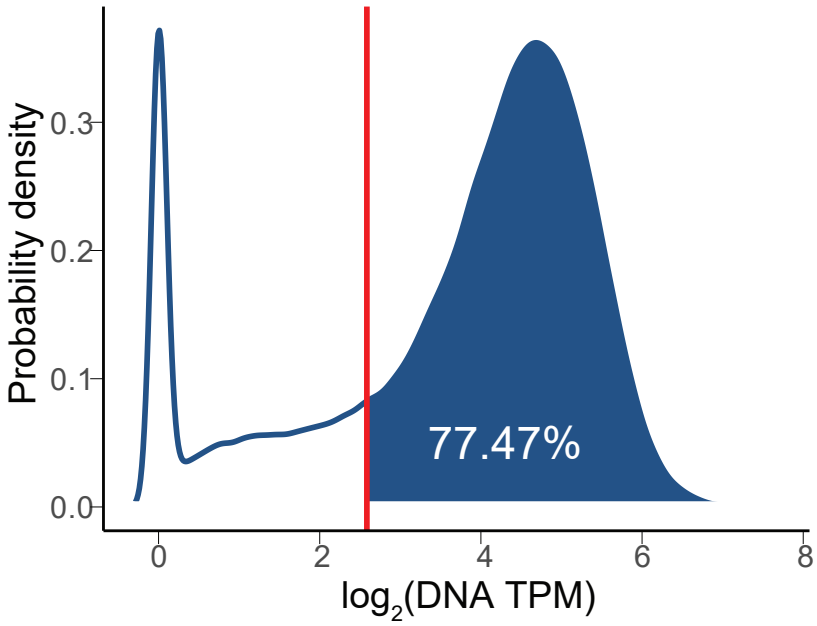
**Supplementary Figure 4** Inter-transfection correlation of normalized tag counts for each tag between transfection replicates are shown for transfections of UACC903 cells (A) and HEK293FT cells (B). $Log_2$(RNA TPM/DNA TPM) value for each tag before QC are plotted with the normalized tag count shown as color-coded density level. Pair-wise Pearson correlation coefficients are shown in red (R values). Lib1 and 2: Library 1 and 2, Pro: MPRA construct testing promoter function containing no promoter element, Enh: MPRA construct testing enhancer function containing minimal TATA promoter, Rep1 through 4: transfection replicates 1 through 4.



**Supplementary Figure 5** Inter-library correlation of normalized tag counts for tags that were repeated in Library 1 and Library 2 are shown for transfections of HEK293FT cells. $Log_2$(RNA TPM/DNA TPM) value for each tag before QC are plotted with the normalized tag count shown as color-coded density level. Pair-wise Pearson correlation coefficients are shown in red (R values). Lib1 and 2: Library 1 and 2, Pro: MPRA construct testing promoter function containing no promoter element, Enh: MPRA construct testing enhancer function containing minimal TATA promoter, Rep1 and 2: transfection replicates 1 and 2.

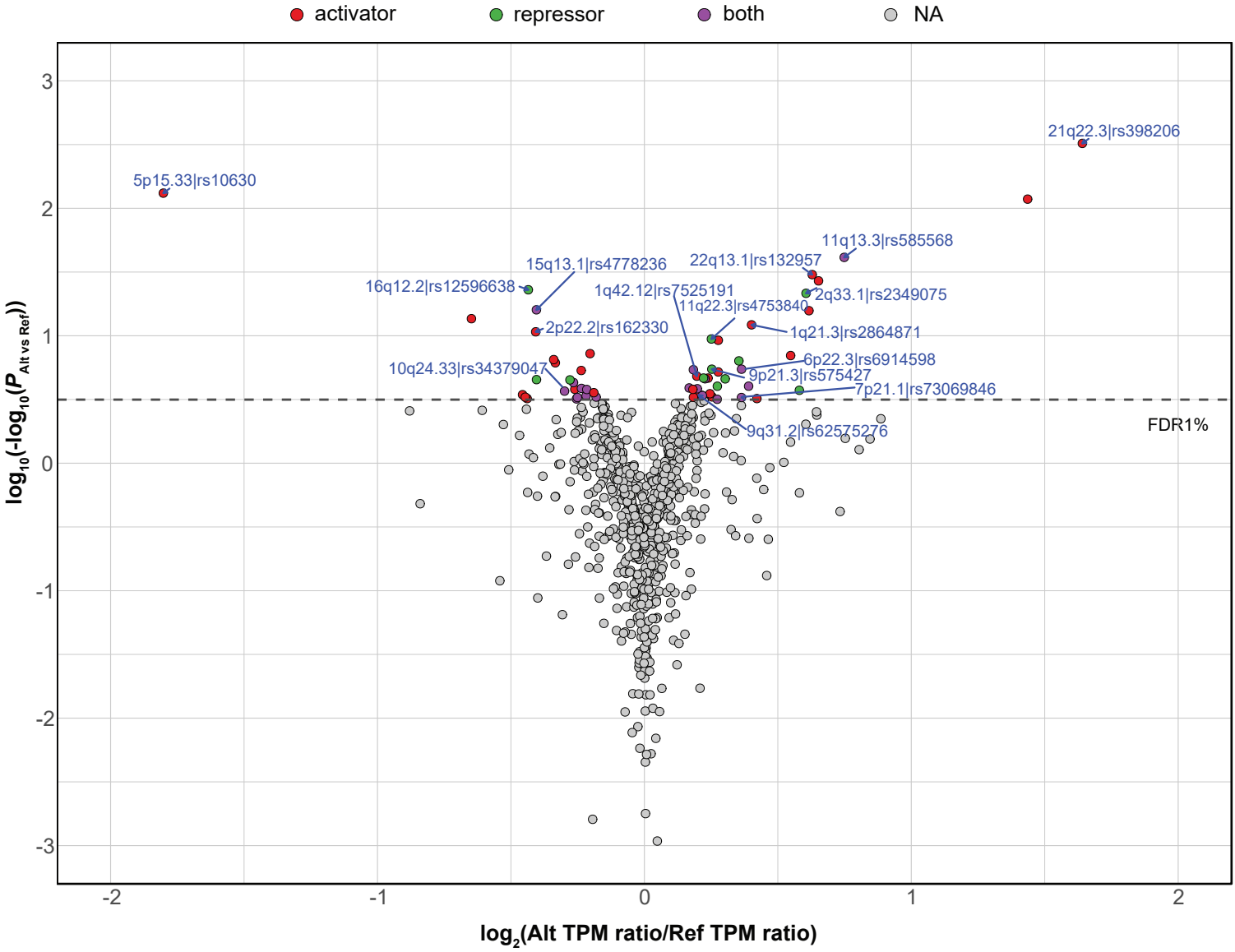


**Supplementary Figure 6** Inter-library correlation of normalized tag counts for tags that were repeated in Library 1 and Library 2 are shown for transfections of UACC903 cells. $Log_2$(RNA TPM/DNA TPM) value for each tag before QC are plotted with the normalized tag count shown as color-coded density level. Pair-wise Pearson correlation coefficient are shown in red (R values). Lib1 and 2: Library 1 and 2, Pro: MPRA construct testing promoter function containing no promoter element, Enh: MPRA construct testing enhancer function containing minimal TATA promoter, Rep1 and 2: transfection replicates 1 and 2.

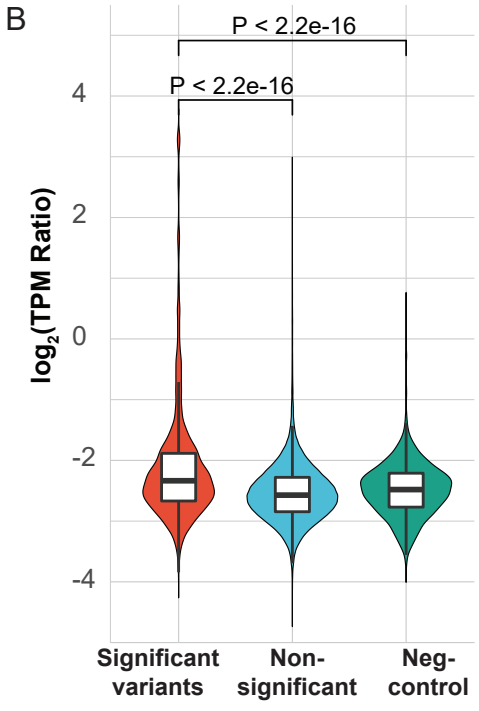


**Supplementary Figure 7** Tag count distribution in the input DNA libraries are shown as $\log_2$ (DNA TPM) density. Red line denotes $\log_2$ (DNA TPM) ~ 2.58 or DNA TPM = 6, which was used as a QC cutoff. Percentage of tags with DNA TPM >= 6 are 77.47% of all the detected tags.

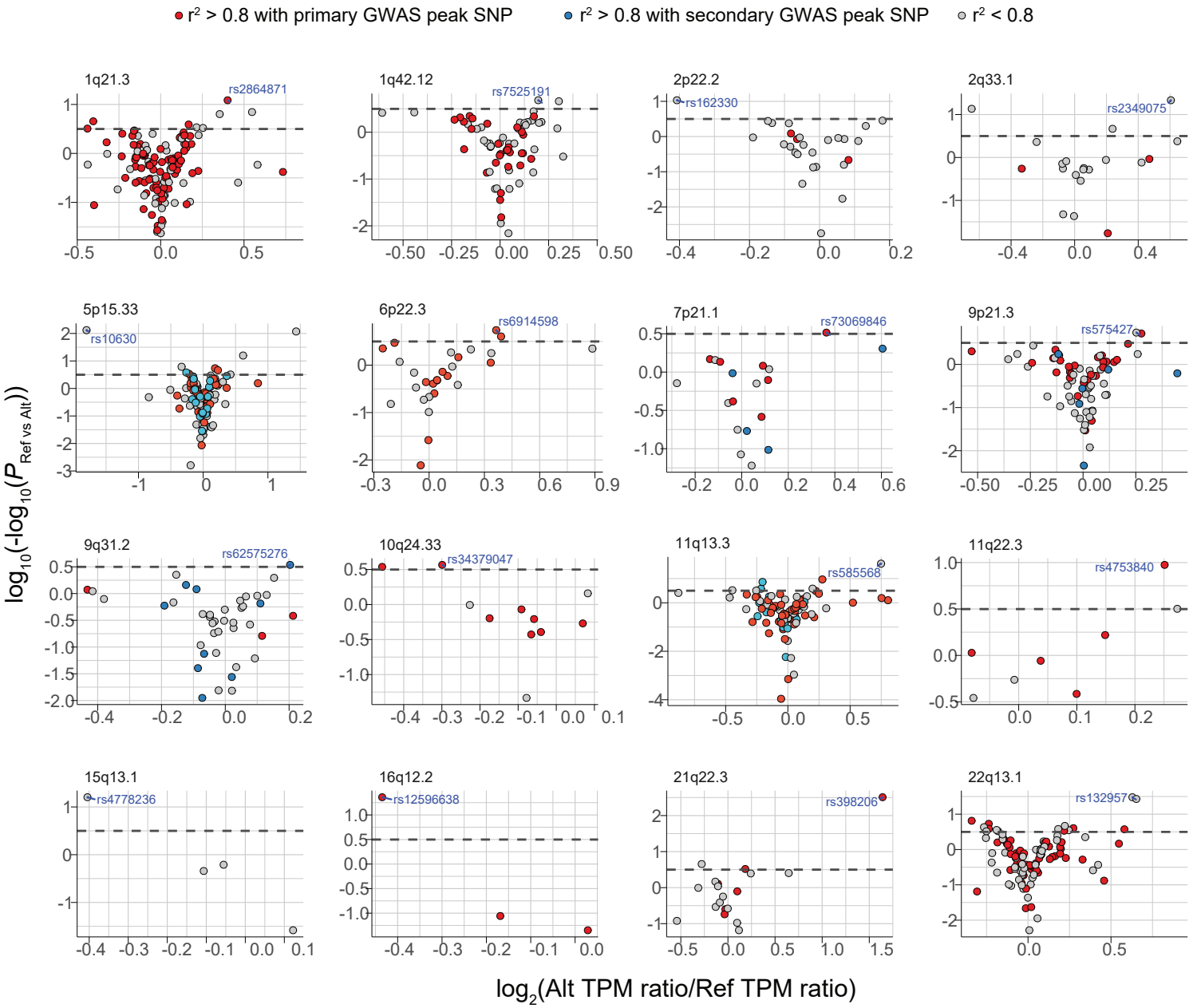
A




B

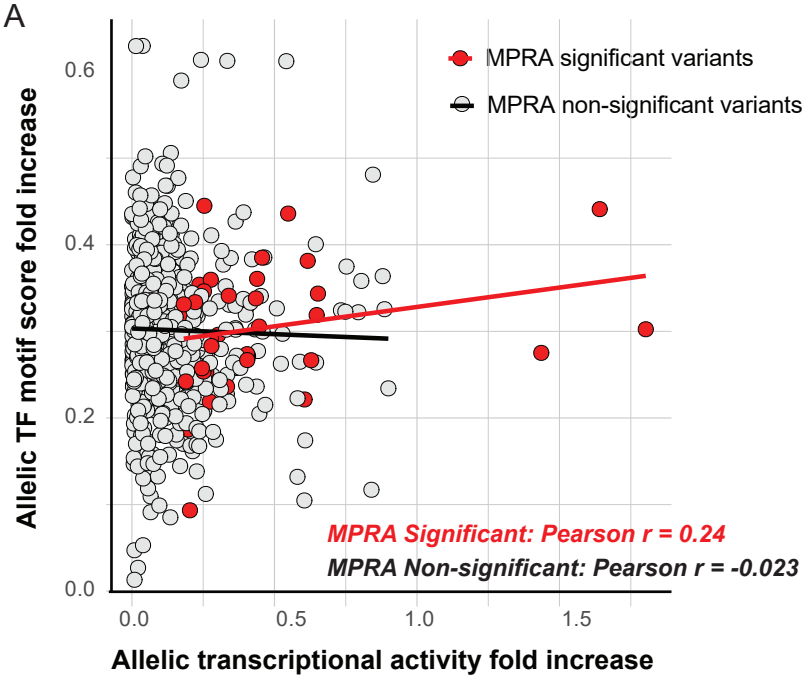


**Supplementary Figure 8** MPRA results from all 16 melanoma loci (A) Melanoma GWAS variants were plotted for their inverted regression P-values of allelic transcriptional difference in UACC903 melanoma cells. Allelic effect size is shown on the X-axis as log-transformed allelic fold difference using the ratio of RNA TPM over DNA TPM (TPM ratio). Putative function of significant MPRA variants are shown as activator, repressor, or both (expression levels of either allele is higher, lower, or higher and lower than those of scrambled sequence). Chromosome band and SNP ID are shown for the variant displaying the lowest FDR from each locus. (B) Transcriptional activity was shown as log-transformed RNA TPM over DNA TPM (TPM ratio) for 39 significant variants (FDR < 0.01), non-significant (the rest of tested variants), and negative controls (8 variants) using data from UACC903 cells. Mann-Whitney U test.

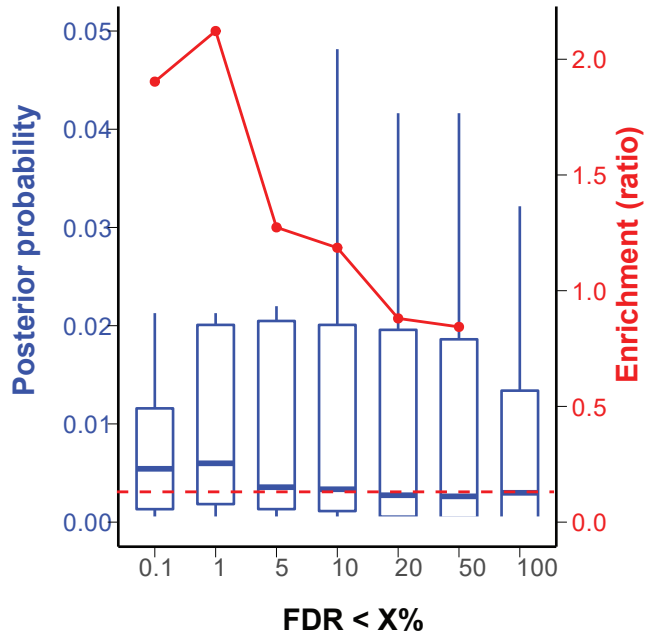


**Supplementary Figure 9** Volcano plots of MPRA results for each melanoma GWAS locus presented based on $r^2$ with the primary or secondary GWAS peak SNP of each locus.

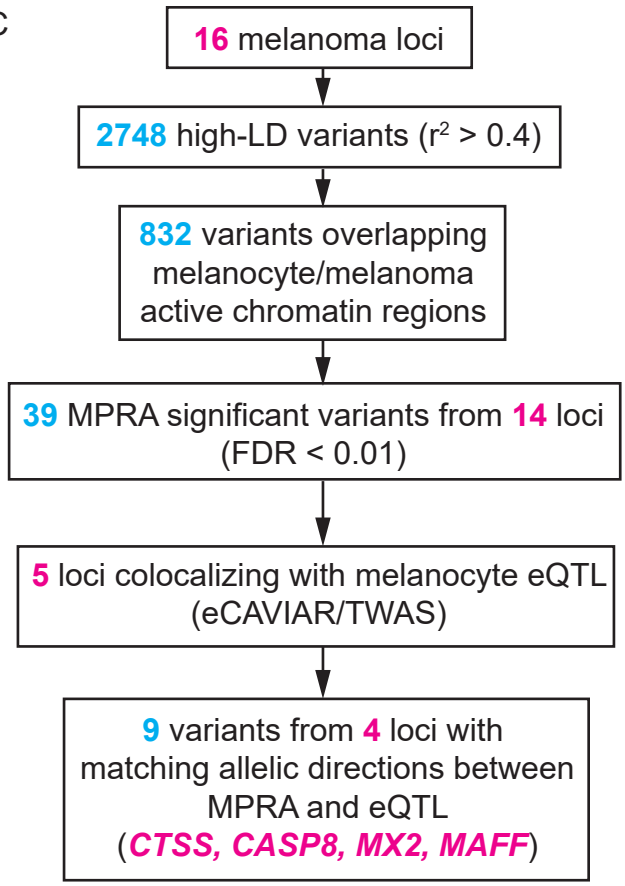
A



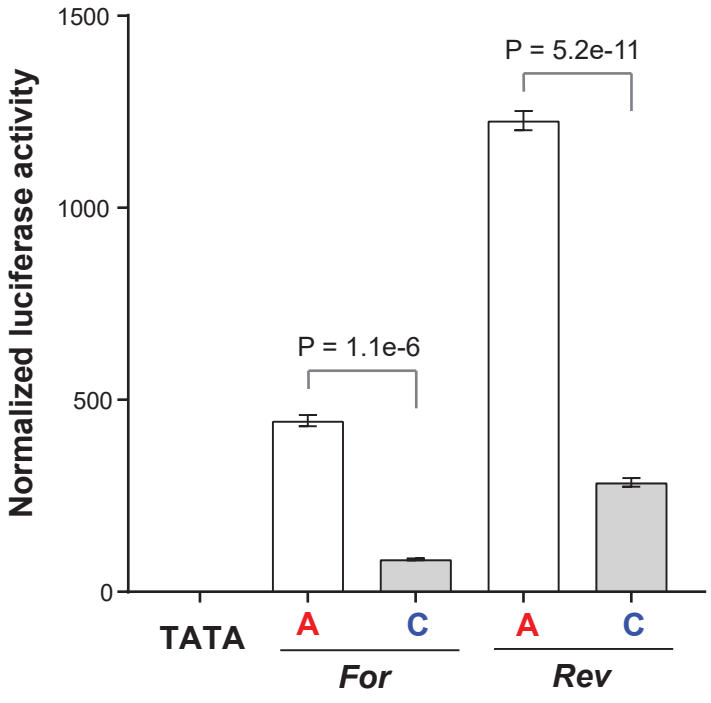

B



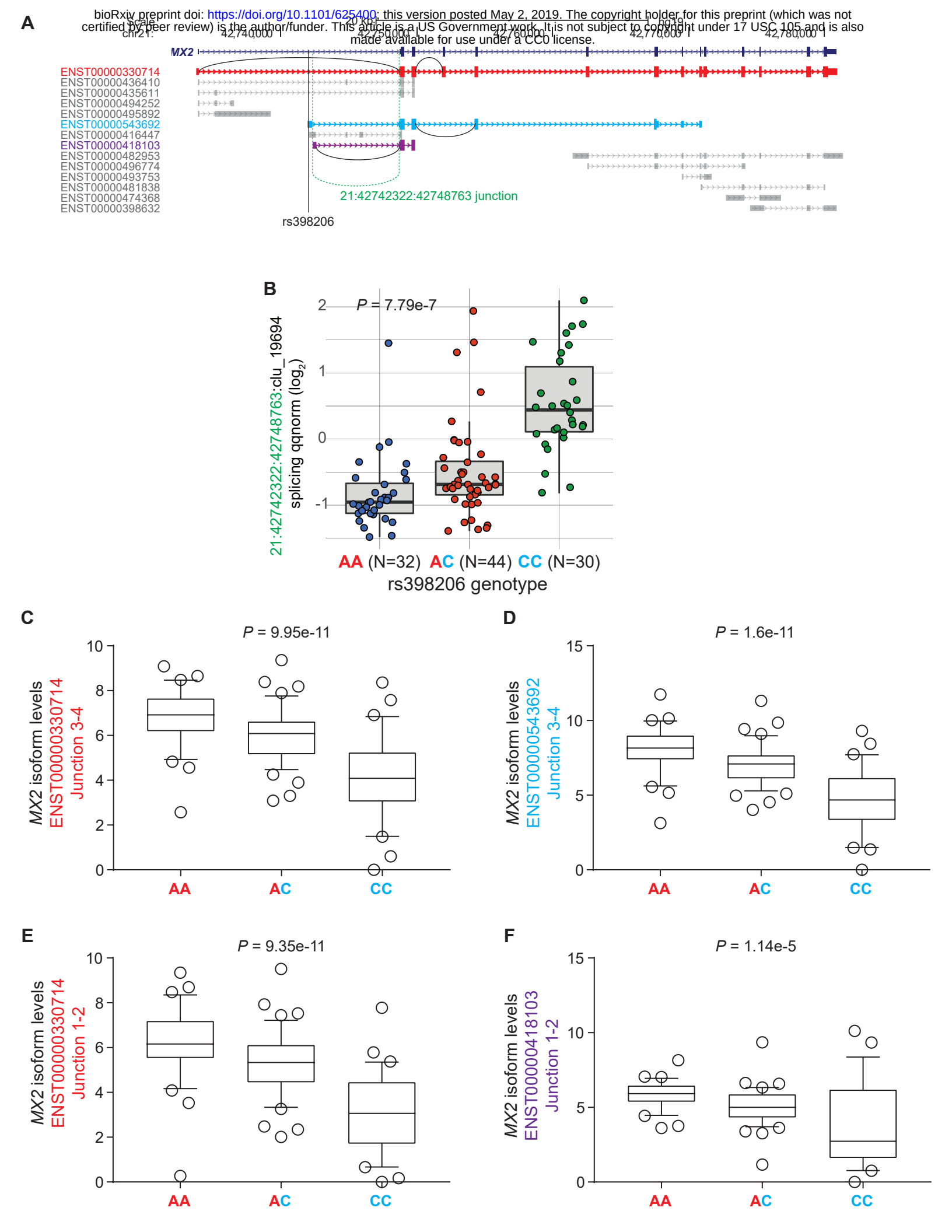

C




D

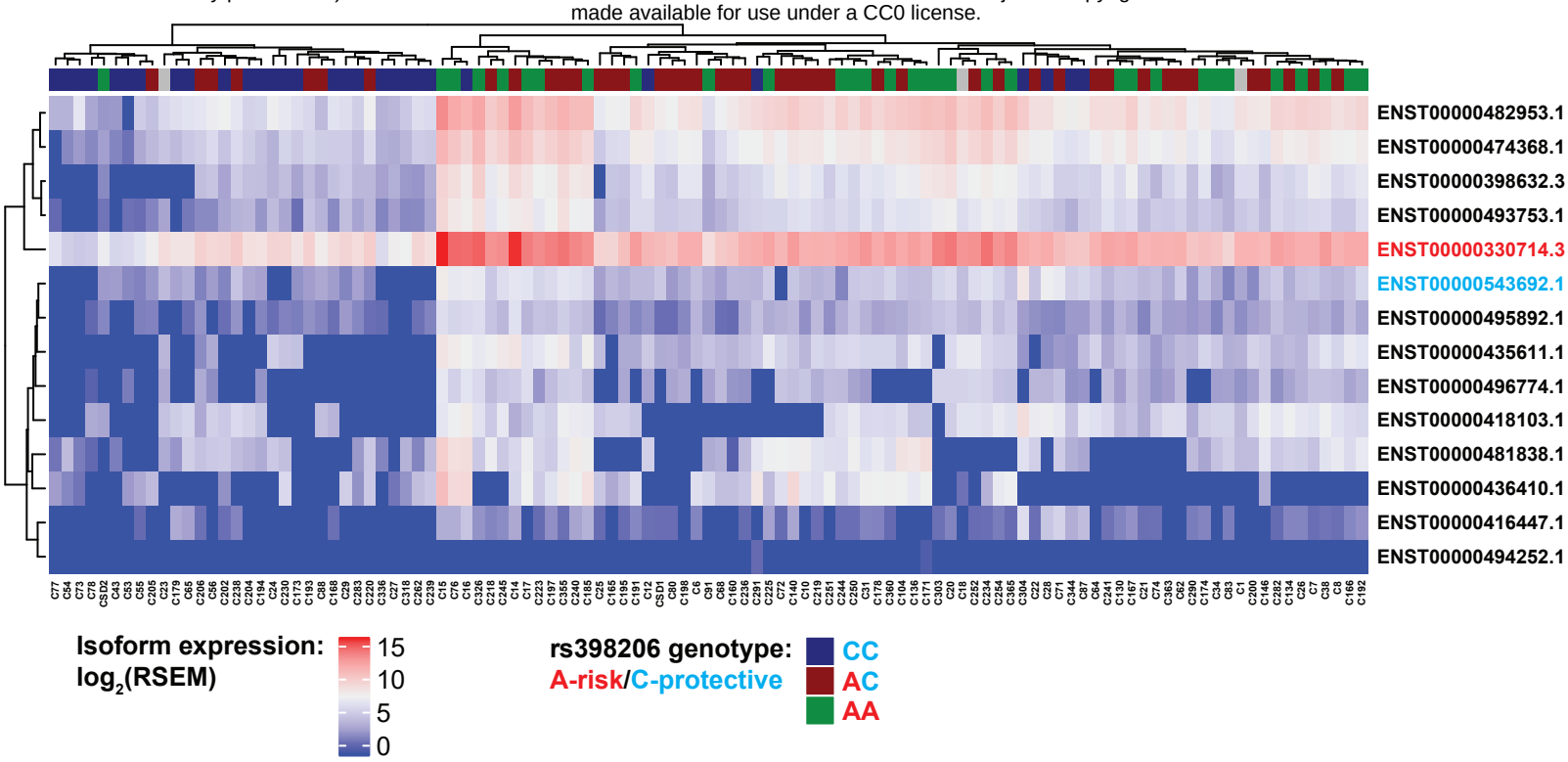


**Supplementary Figure 10** (A) Allelic transcriptional activity fold increase (relative to the allele with lower activity) from MPRA was plotted against allelic fold increase of transcription factor binding motif score (relative to the allele with lower score) for 39 MPRA-significant variants (FDR < 0.01) and non-significant variants. Fold changes in both axes are $\log_2$-transformed. The scores for the most significant TF for each variant was used (TF motif prediction $P < 0.001$). Pearson correlation *r* scores for each group are shown. $P = 0.149$ for significant variants, and $P = 0.556$ for non-significant variants. (B) Posterior probabilities (left Y-axis in blue) of MPRA variants are plotted for subsets of variants with increasing FDR cutoffs (median, 25$^{th}$ and 75$^{th}$ percentile). Red dashed line indicates the median probability score when including up to FDR = 100%. Enrichment ratios of median posterior probability score from each subset over that of FDR = 100% group are shown as dots and a trend line on the right Y-axis in red. (C) A flowchart of variant prioritization using melanocyte eQTL. (D) Individual luciferase activity assays of 145bp sequences encompassing rs398206 is shown for UACC502. pGL4.23 construct including minimal TATA promoter was used. One representative set is shown from three biological replicates. Mean with SEM, n = 6. All constructs are significantly higher than pGL4.23 (TATA) control ($P < 0.0001$). Two-tailed, unpaired t-test assuming unequal variance.

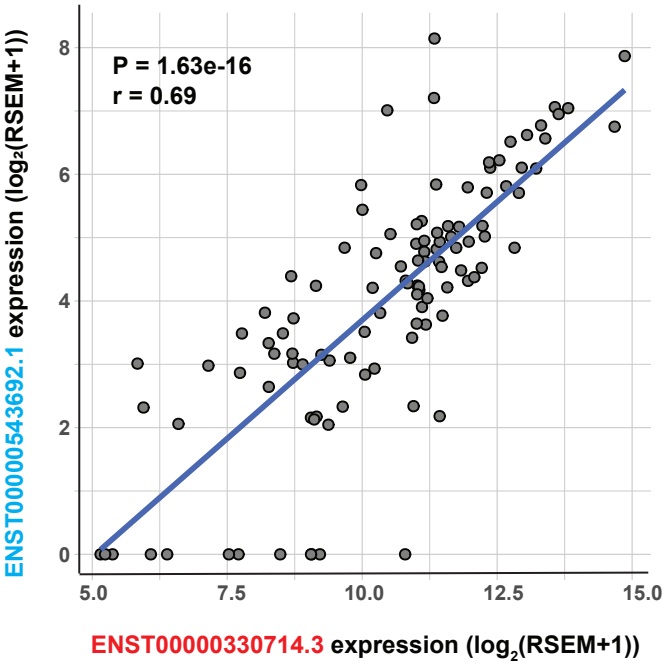


**Supplementary Figure 11** (A) Genomic map of *MX2* transcript isoforms based on Ensembl75 (GRCh37). Half-circle lines denote the isoform-specific splice junctions tested with Taqman probes. Dashed lines indicate the chr21:42742322:42748763 junction identified by melanocyte sQTL analysis. (B) melanocyte sQTL plot of an *MX2* splice junction relative to rs398206 genotypes using LeafCutter[1]. The Y-axis displays standardized and normalized "percent spliced in" (ΔPSI) of the junction, Chr21:42742322:42748763, within the cluster 19694 (clu_19694) of introns sharing splice sites. sQTL nominal *P*-value is shown. The genome-wide significance threshold for cluster 19694 is *P* = 9.12e-6. (C-F) qPCR validation of splice-junction specific *MX2* isoform levels in melanocytes relative to rs398206 genotypes using Taqman probes targeting unique junctions of ENST00000330714, junction 3-4 (C), ENST00000543692, junction 3-4 (D), ENST00000330714, junction 1-2 (E), and ENST00000418103, junction 1-2 (F). Linear regression using average dCt values of PCR triplicates normalized over *TBP* levels against rs398206 A allele count. For plotting, dCt values were converted to *MX2* isoform levels by using $\log_2$ (fold-change over the sample showing the highest dCt).

A

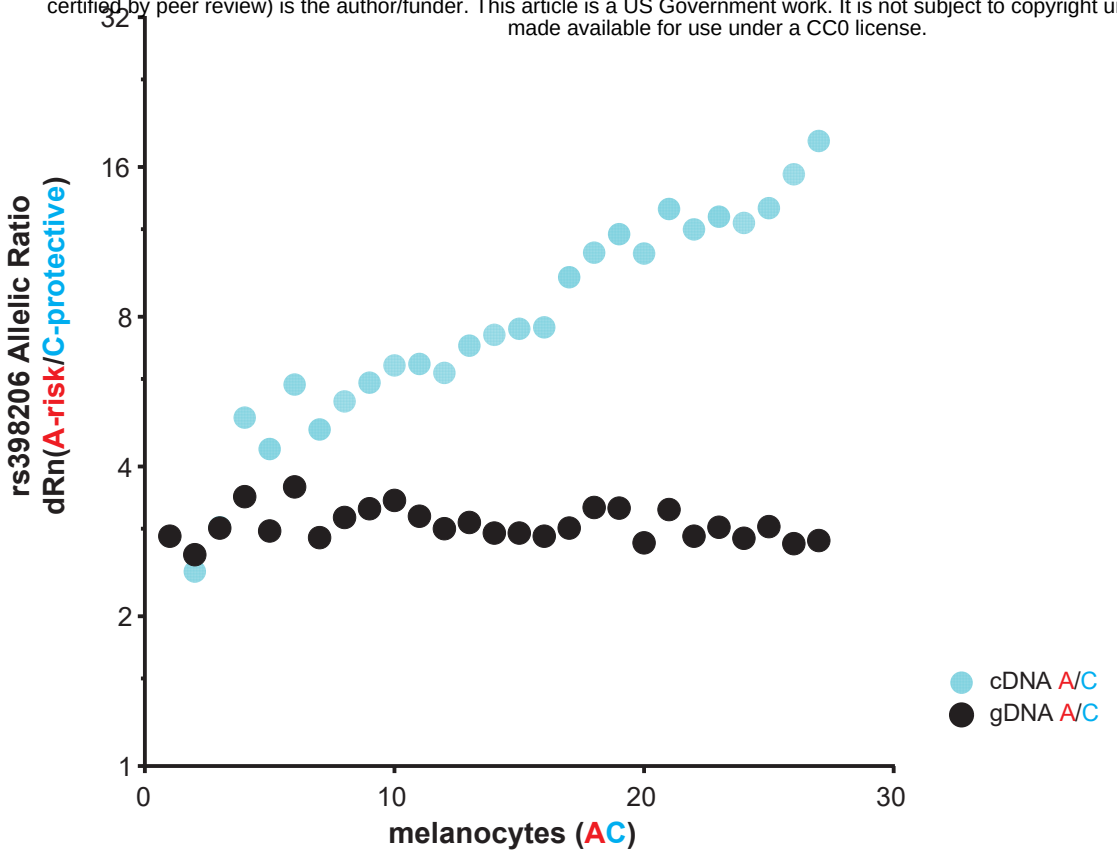



B

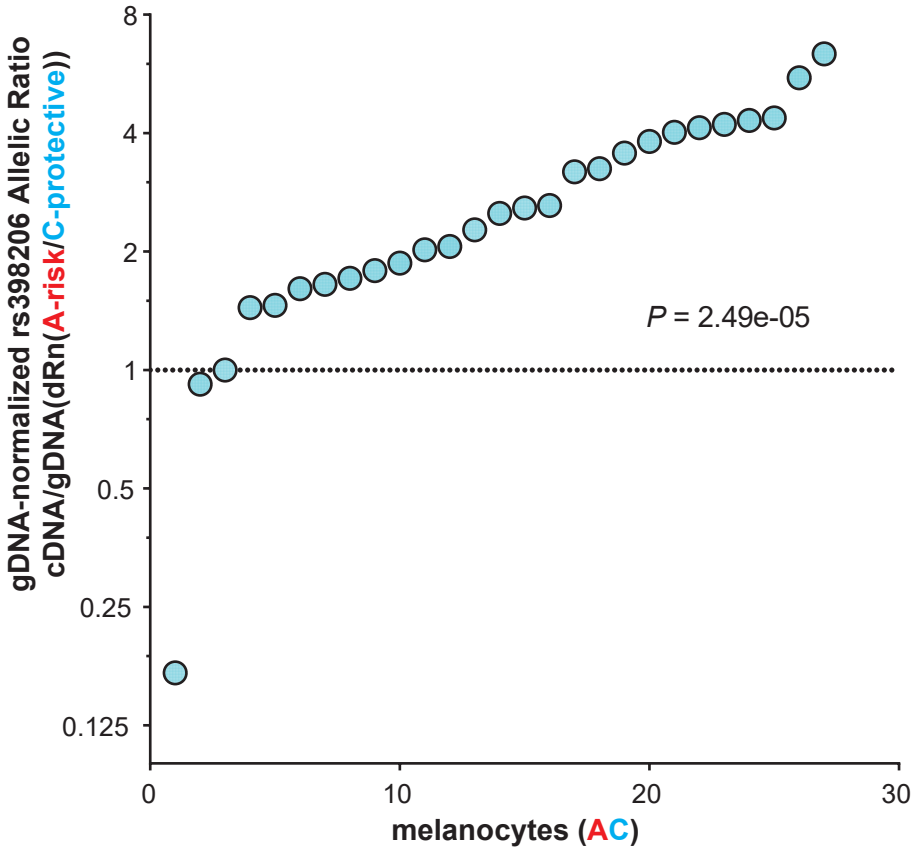


**Supplementary Figure 12** (A) Heatmap of clustered *MX2* transcript isoform expression levels are shown for 106 primary cultures of melanocytes. Isoform quantification from melanocyte RNA-Seq data was performed using RSEM expression quantification package (iterations of Expectation-Maximization algorithms to assign reads to the isoforms from which they originate)[4,5]. rs398206 genotypes are shown (A: risk, C: protective). (B) Expression correlation between isoform ENST00000330714.3 (full-length, most abundant) and ENST00000543692.1 (rs398206 is located at the 5' UTR). Pearson correlation *r* and P-value are shown.

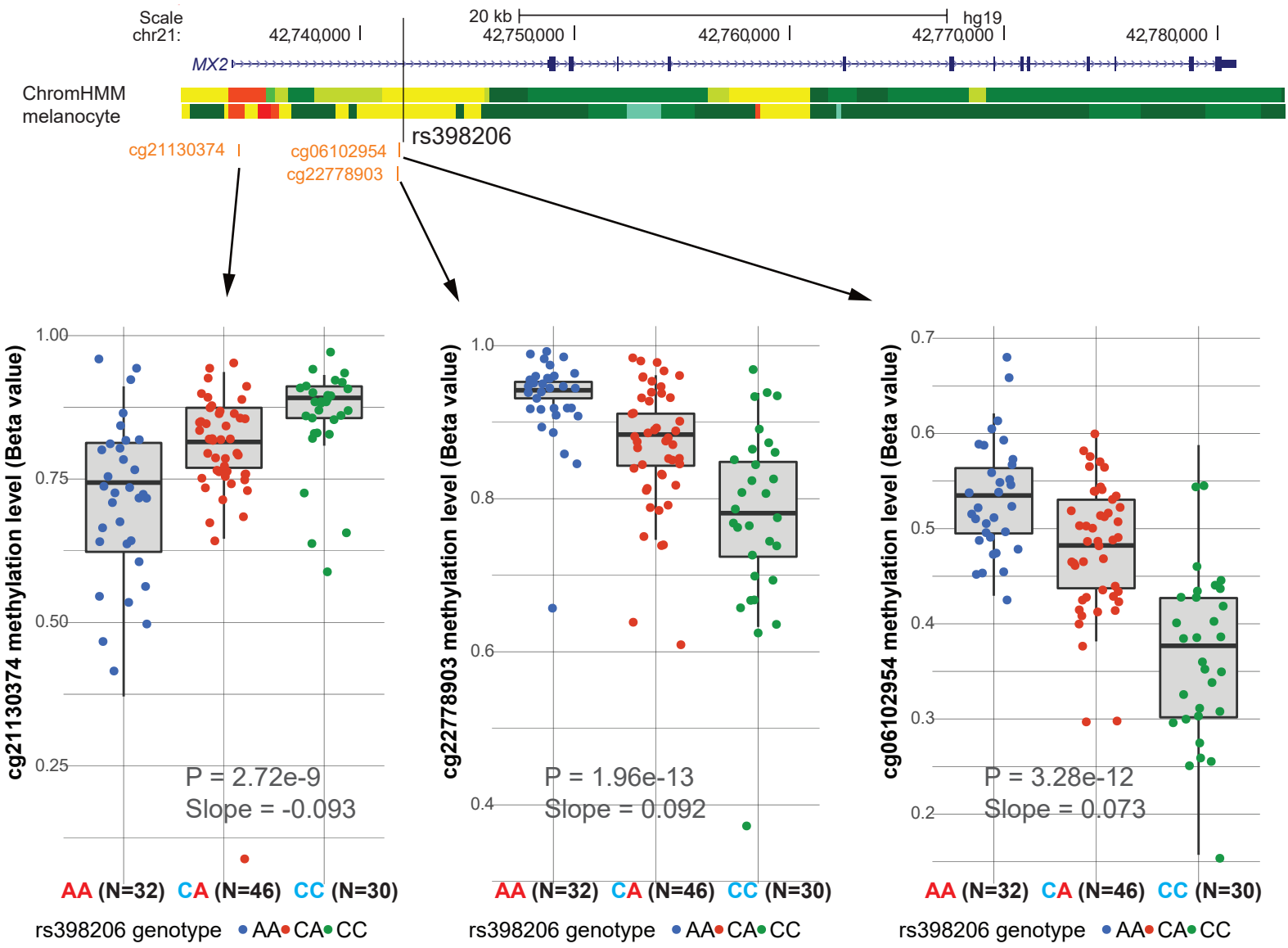
A



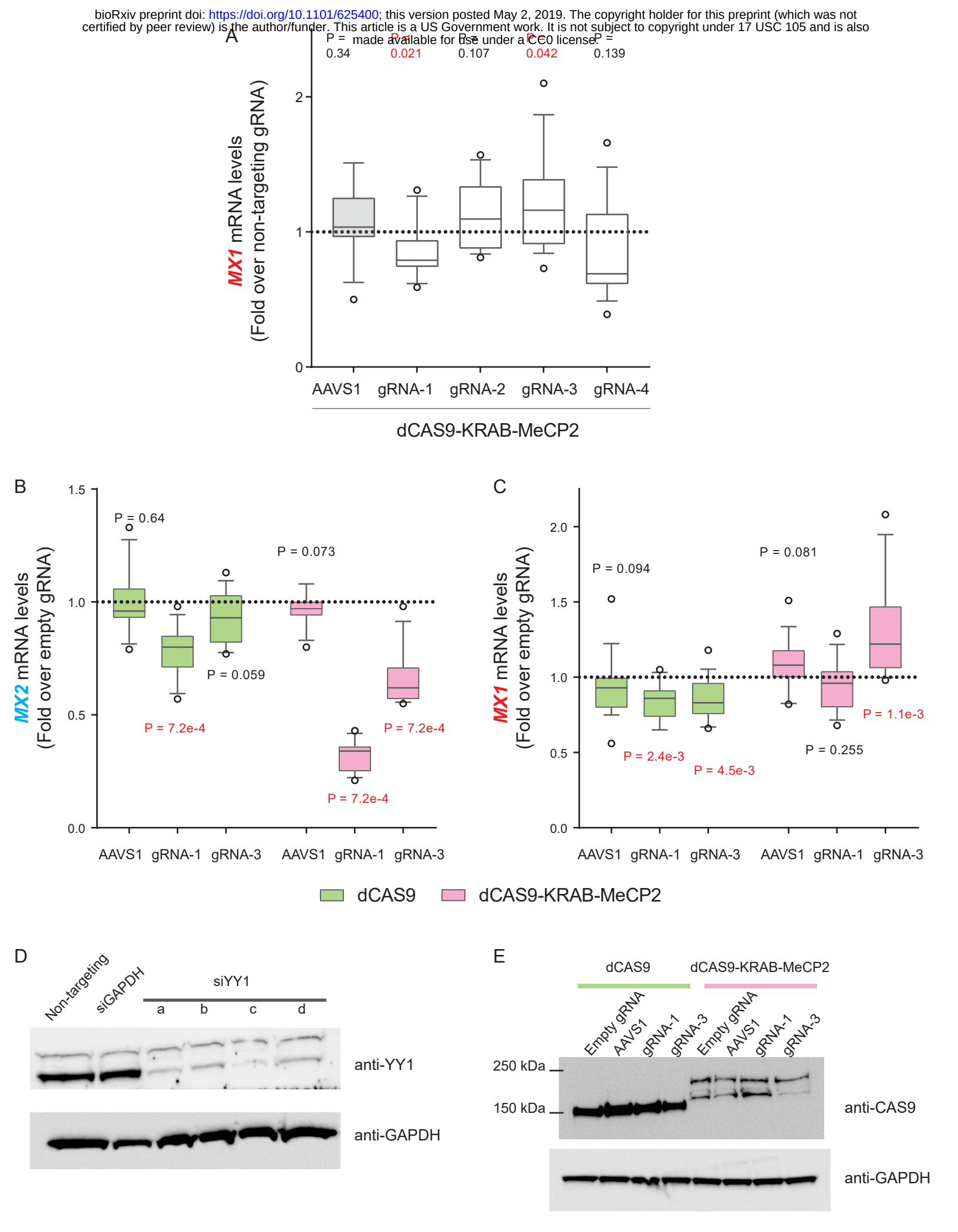

B



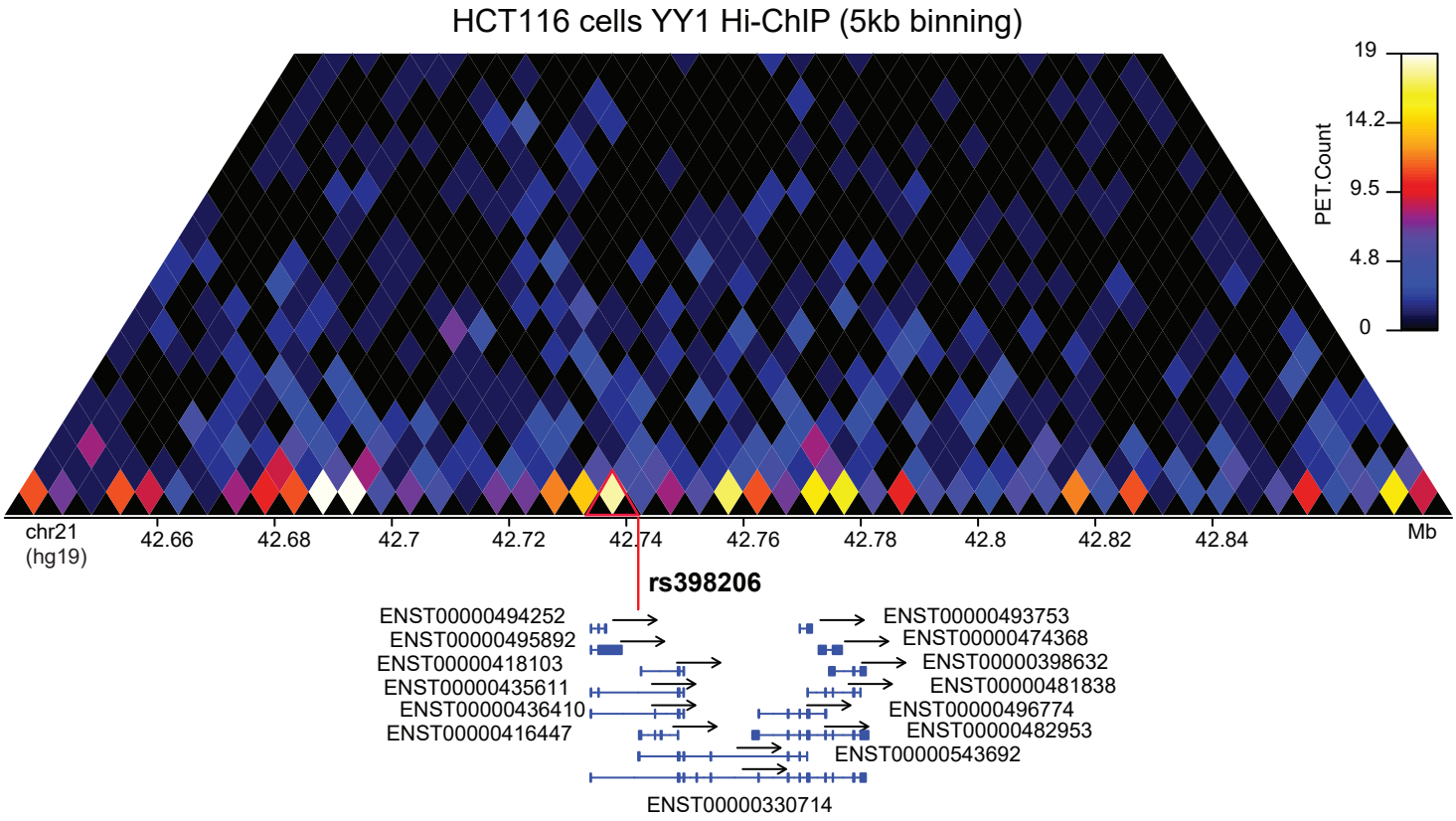
**Supplementary Figure 13** (A) Allele-specific expression of *MX2* transcripts harboring rs398206 in a subset of heterozygous human primary melanocytes. Melanocyte cDNA and genomic DNA (gDNA) from 27 heterozygous individuals were genotyped using a Taqman probe set for rs398206. A/C allelic ratio of dRn values (average of PCR duplicates) are plotted for gDNA and cDNA. (B) cDNA A/C allelic ratio normalized over gDNA A/C ratio from the same sample is plotted for each sample. Dotted line denotes normalized A/C ratio of 1. One-sample Wilcoxon signed rank test was used. Y-axes are displayed in $\log_2$-scale.



**Supplementary Figure 14** Genomic map of *MX2* gene (RefSeq transcript; hg19) with the positions of three CpG probes from the Illumina Human Methylation 450K platform. ChromHMM tracks of primary melanocytes from two individuals are shown. Box plots of meQTL from primary melanocytes (n = 106) between each CpG probe and rs398206 are shown at the bottom (median and 25$^{th}$ and 75$^{th}$ percentile). meQTL P-values and slopes (relative to A-allele) are shown on the plot.

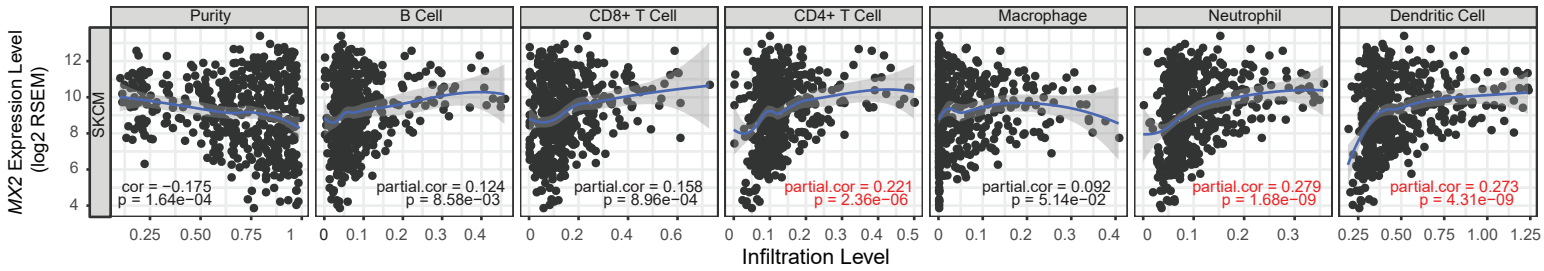


**Supplementary Figure 15** CRISPRi using gRNAs targeting rs398206 in UACC903 cells. (A) CRISPRi using dCAS9-KRAB-MeCP2 and four gRNAs targeting the area immediately surrounding rs398206. Levels of *MX1* transcript (*GAPDH*-normalized) are shown as fold change over those from non-targeting gRNA. Three biological replicates of n = 6 were combined (total n=18, except gRNA-3, n=17). gRNAs 1, 3, and 4 directly overlap rs398206, while gRNA 2 targets ~25bp upstream of rs398206. (B,C) CRISPRi using dCAS9 or dCAS9-KRAB-MeCP2 and gRNAs 1 and 3. Levels of *MX2* (B) or *MX1* (C) transcript (*GAPDH*-normalized) are shown as fold change over those from empty gRNA vector, pRC0215, within each set using the same type of dCAS9 constructs. Three biological replicates of n = 5 were combined (total n = 15). AAVS (gRNA targeting adeno virous integration site on Chr19). Box: Median and $25^{th}$ to $75^{th}$ percentile. Whisker: $10^{th}$ to $90^{th}$ percentile. *P* - values are shown from one-sample Wilcoxson test (two-sided) for difference from non-targeting siRNA/gRNA ($P < 0.05$ are shown in red). Dotted line denotes the *MX2* levels in non-targeting siRNA/gRNA control. (D) Western blotting using anti-YY1 and anti-GAPDH antibodies and cell lysates of UACC903 transfected with four different siRNAs targeting YY1. Proteins were isolated at 72hrs following transfection. Non-targeting: non-targeting siRNA, siGAPDH: positive control siRNA targeting *GAPDH*. (E) Western blotting was performed using anti-CAS9 (recognizing both ~160 kDa dCAS9 and ~200 kDa dCAS9-KRAB-MeCP2) and anti-GAPDH antibodies, and cell lysates of UACC903 co-transfected with dCAS9 or dCAS9-KRAB-MeCP2 and indicated gRNAs. Relative positions of protein ladders are shown on the left side for the anti-CAS9 blot. Proteins were collected from one representative set of three sets of total transfections.

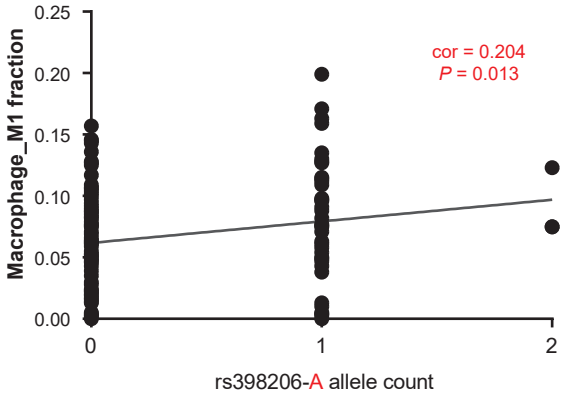


**Supplementary Figure 16** YY1-mediated chromatin interaction map of the genomic region encompassing *MX2* was plotted using YY1 Hi-ChIP data in HCT116 cell line reported by Weintraub and colleagues[6]. Black triangles at the bottom row of the heatmap represent 5kb bins for measuring physical interactions. Each diamond of the heatmap displays the paired-end tag counts (PET.Count) between two bins connected by the diamond. The genomic position of rs398206 is shown with red vertical line, and red triangle highlights the interaction between the bin harboring rs398206 and the neighboring bin encompassing the *MX2* promoter region.

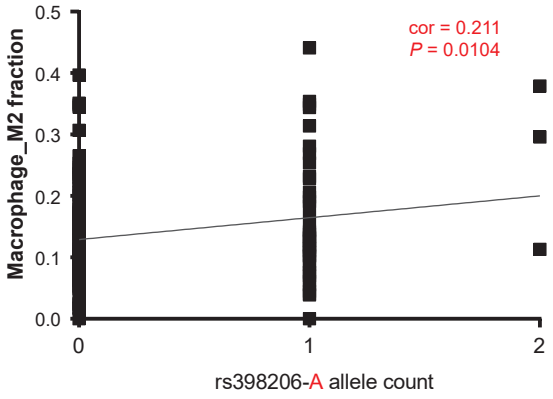
A




B

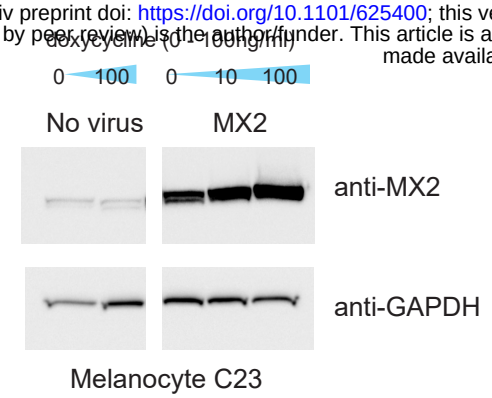



C

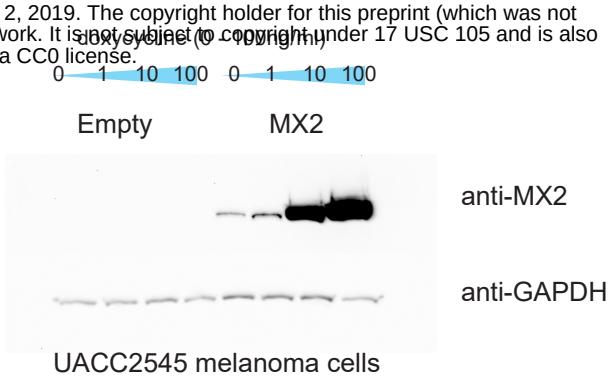


**Supplementary Figure 17** (A) Correlation of *MX2* levels in TCGA SKCM samples with tumor purity and six types of immune cell infiltration levels using TIMER[2] program. Purity is defined as % malignant cells in a tumor tissue inferred from CNA data. Purity-corrected partial Spearman's correlation and statistical significance are shown. Correlation values with partial correlation coefficient > |0.2| and $P < 0.05$ are shown in red. (B) CIBERSORT[3] analyses of TCGA SKCM samples. Pearson correlation coefficients are shown between rs398206 genotype (A allele count) and estimated fractions of Macrophage M1 (B) and M2 (C) using 147 samples displaying significant deconvolution ($P < 0.05$). Only these two cell types among 22 types of leukocytes showed correlation coefficient > |0.2| and $P < 0.05$.

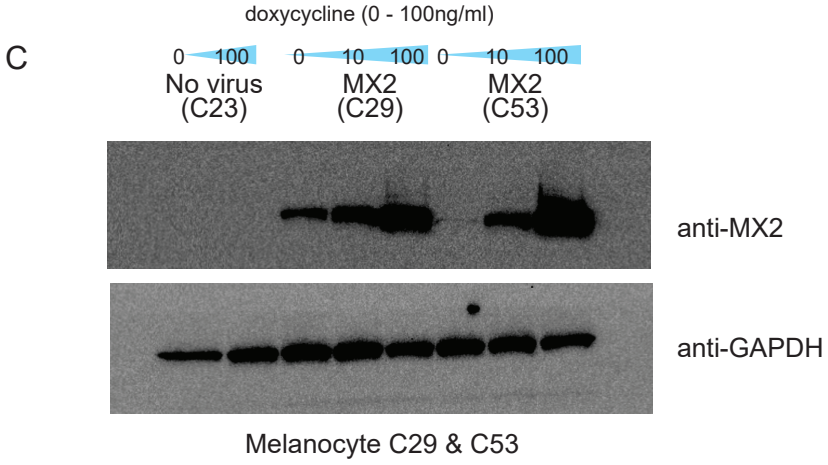
A




B




C



**Supplementary Figure 18.** (A-B) Western blotting using anti-MX2 and anti-GAPDH antibodies and cell lysates from primary human melanocytes (A) or UACC2545 melanoma cells (B) infected with lentivirus containing *MX2* cDNA, empty pINDUCER20, or no virus with a varying amount of doxycycline treatment. Cells were infected concomitantly with each round of xCELLigence growth assay and harvested at 72hrs of doxycycline treatment. One representative set from each cell type are shown from three biological replicates. (C) Western blotting using anti-MX2 and anti-GAPDH antibodies and cell lysates from two additional melanocytes cultures from different individuals infected with lentivirus containing MX2 cDNA with a varying amount of doxycycline treatment. Cells were infected concomitantly with one representative round of RNAseq experiment out of three biological replicates and harvested at 72hrs of doxycycline treatment.