Massively parallel reporter assays of melanoma risk variants identify MX2 as a gene promoting melanoma

Jiyeon Choi, Tongwu Zhang, Andrew Yu, Julien Ablain, Matthew M. Makowski, Leandro M. Colli, Mai Xu, Rebecca C. Hennessey, Jinhu Yin, Harriet Rothschild, Cathrin Gräwe, Michael A. Kovacs, Karen M. Funderburk, Myriam Brossard, John Taylor, Bogdan Pasaniuc, Raj Chari, Stephen J. Chanock, Clive J. Hoggart, Florence Demenais, Jennifer H. Barrett, Matthew H. Law, Mark M. Iles, Kai Yu, Michiel Vermeulen, Leonard I. Zon & Kevin M. Brown

Genome-wide association studies (GWAS) have identified ~20 melanoma susceptibility loci, most of which are not functionally characterized. Here we report an approach integrating massively-parallel reporter assays (MPRA) with cell-type-specific epigenome and expression quantitative trait loci (eQTL) to identify susceptibility genes/variants from multiple GWAS loci. From 832 high-LD variants, we identify 39 candidate functional variants from 14 loci displaying allelic transcriptional activity, a subset of which corroborates four colocalizing melanocyte cis-eQTL genes. Among these, we further characterize the locus encompassing the HIV-1 restriction gene, MX2 (Chr21q22.3), and validate a functional intronic variant, rs398206. rs398206 mediates the binding of the transcription factor, YY1, to increase MX2 levels, consistent with the cis-eQTL of MX2 in primary human melanocytes. Melanocyte-specific expression of human MX2 in a zebrafish model demonstrates accelerated melanoma formation in a BRAF^V600E background. Our integrative approach streamlines GWAS follow-up studies and highlights a pleiotropic function of MX2 in melanoma susceptibility.
A series of genome-wide association studies (GWAS) over the past decade has identified about twenty genomic loci associated with cutaneous melanoma1–10, 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 phenotypes11–15 and nevus count5,16,17. 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 insights into molecular pathways underlying melanoma susceptibility. PARP1, located in the Chr1q42.1 melanoma locus8, 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, MIF18, while a functional variant at a multi-cancer locus on Chr3p15.33 was characterized highlighting the role of TERT in cancer susceptibility including in melanoma19. 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 sequencing20–22. 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 2756 variants from 75 GWAS loci for red blood cell traits and identified 32 functional variants from 23 loci20.

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 project23, 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 loci24. This dataset identified more candidate susceptibility genes than using eQTLs from datasets of larger sample size generated from bulk skin tissues, other tissue types from GTEx, and melanoma tumors24, 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. Using our approach we identify 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 accelerates melanoma formation.

Results
MPRA identified melanoma-associated 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-analysis1, 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 (MCIR, SLCA5A2, and TYR11–14) or shown not to be expressed in melanocytes (ASIP15). To comprehensively analyze genetic signals from these loci, we then performed statistical fine-mapping using the HyperLasso25 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 results1. To prioritize melanoma-associated variants to test by MPRA, we first selected 2748 variants that are in LD (r² > 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 cultures26 (Supplementary Data 1; Supplementary Table 3; Methods; www.encodeproject.org; www.roadmapepigenomics.org). We then constructed MPRA libraries for these 832 variants using methods adopted from previous studies20–22,27. 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 (Fig. 1a; 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. 2–6; Supplementary Table 4).

To nominate variants displaying allelic transcriptional activity, we focused on those displaying significant difference between two alleles (FDR < 0.01; two-sided Wald test with robust sandwich type variance estimate; multiple comparisons adjusted using Benjamini & Hochberg method), and then further selected those with either allele displaying a significant difference from the null (scrambled core sequence; FDR < 0.01) (Supplementary Fig. 2). 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. 7a; Supplementary Table 5). These candidate functional variants are from 14 melanoma GWAS loci with 1–9 variants per locus (median 1.5 variants), demonstrating that MPRA narrowed down functional candidate variants to a considerably small number from tens to hundreds of high-LD variants in most of the loci (Fig. 1b; Supplementary Table 6; Supplementary Fig. 8). 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. 7b), as well as the rest of the variants (non-significant variants; P < 2.2e−16, effect size = 0.109). These 39 variants displayed a modest 1.13 to 3.49-fold (median 1.26-fold) difference in transcriptional activity between two alleles consistent with subtle
effects on risk observed for common cancer-associated variants (UACC903 cells; Supplementary Table 5). 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\(^\text{28}\) and subsequently correlated the predicted allelic binding scores with the allelic transcriptional activities measured from MPRA. Notably, the MPRA-significant variants displayed a higher level of correlation compared to that of non-significant ones, shown by a larger Pearson \(r\) (Pearson \(r = 0.24\ vs. \ -0.023\); Supplementary Fig. 9a), while the \(P\)-values for the both tests are not significant (\(P = 0.149\ and 0.556\), respectively). We then asked if MPRA-significant variants are enriched in credible causal variants or those with higher probability scores from a Bayesian fine-mapping approach. For this we performed statistical fine-mapping of melanoma GWAS data using PANTOR\(^\text{29}\) while integrating primary melanocyte-specific functional annotations (Methods). When overlaid with the fine-mapping results, the 39 significant MPRA variants (\(FDR < 0.01\)) indeed 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. 9b). These data demonstrated that MPRA can quickly narrow down to a small number of plausible functional candidate variants from tens to hundreds of high-LD variants from melanoma GWAS loci by measuring allelic transcriptional activity.

**eQTL prioritized functional candidates in four 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\(^\text{24}\), 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 to TWAS\(^\text{24}\). For the purpose of nominating the most plausible candidates for functional follow-up, we mainly focused on the five of these six loci with melanocyte eQTL support (1q21.3, 1q42.12, 2q33-q34, 21q22.3, and 22q13.1) that were also tested in our MPRA. While there were initially 17 to 116 eQTL variants per locus at a genome-wide significant level, MPRA further narrowed them down to 3 to 9 functional variants in four loci based on transcriptional activity (Supplementary Table 6). Among them, a total of nine variants from four loci displayed a consistent direction between MPRA and eQTL, in which the direction of allelic expression of local genes matches those of MPRA allelic difference in the UACC903 set. The most significant variant from each locus is labeled. Putative function of 39 signficant MPRA variants were identified from the loci on Chr6p22.3 and Chr7p21.1. Source data are provided as a Source Data file.

**Fig. 1 MPRA identified 39 functional variants from 16 melanoma GWAS loci.**

- **a** MPRA workflow. Oligo libraries were synthesized using 145 bp 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 10 bp 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. Luc: luciferase gene, AAAAA: poly-A tail.

- **b** 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. A two-sided Wald test with robust sandwich type variance estimate was used. Multiple comparisons were adjusted using the Benjamini and Hochberg method. Dashed horizontal lines indicate the FDR 1% cutoff for allelic difference in the UACC903 set. The most significant variant from each locus is labeled. Putative function of 39 significant MPRA variants are shown as activator (red circle), repressor (blue circle), or both (purple circle) (expression levels of either allele is higher, lower, or higher and lower than those of scrambled sequence, respectively). Gray variants above the FDR 1% cutoff are those that failed additional criteria (allelic difference in the combined data or significant departure from the scrambled control). No significant variants were identified from the loci on Chr6p22.3 and Chr7p21.1. Source data are provided as a Source Data file.
Fig. 9c). 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 (2q33-q34, 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 7). For 1q42.12, none of the eQTL significant variants passed the MPRA significance cutoff, and therefore were not prioritized. Thus, by combining MPRA and cell-type specific melanocyte eQTL, we efficiently prioritized candidate functional variants and susceptibility genes from multiple melanoma GWAS loci.

MX2 is a susceptibility gene in the 21q22.3 melanoma locus. To validate the efficiency of our prioritization scheme, we performed a thorough validation and functional characterization for one of the four loci with combined MPRA and melanocyte eQTL support. We chose the locus on 21q22.3, where MPRA identified the most significant variant. In this locus, twenty-two variants were originally tested in MPRA (Supplementary Table 8), nineteen of which are located in the first intron of the MX2 gene (Fig. 2a), and the remaining three upstream. Of these, three variants were significant in MPRA, and rs398206, in particular, (Fig. 2a, shown in magenta) displayed the lowest P-value of all 832 tested variants. rs398206 displayed a strong transcriptional activation function (1.7 to 4.3-fold above the scrambled sequence), as well as significant allelic difference, 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−26; Fig. 2b). Subsequent individual luciferase assays using the same 145 bp sequence in two melanoma cell lines validated this finding (2.7 to 5.0-fold allelic difference, P = 1.1e−6−5.2e−11; two-tailed, unpaired t-test assuming unequal variance; Fig. 2c; Supplementary Fig. 9d). rs398206 was also a significant eQTL for levels of MX2 gene expression in primary melanocytes, where the melanoma risk-associated A allele is correlated with higher MX2 expression (Slope = 0.70, P = 6.6e−15; linear regression; Fig. 2d).

While melanocyte eQTL consistently identified MX2 as the best candidate susceptibility gene at the 21q22.3 melanoma locus24, 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, tubal nerve, and whole blood) but no other gene displayed a genome-wide significant eQTL with rs398206 (using a nominal P-value threshold set for each gene based on a genome-wide empirical P-value as defined by GTEx V6p; 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 variants24, 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/hc-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; linear regression), 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 9). 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. 10a), the expression levels of which are correlated with the most abundant full-length transcript in melanocytes (ENST00000330714; Pearson r = 0.69, P = 1.6e−16; Supplementary Fig. 11). 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 MX224, 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 that SNP 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 = 2.49e−5; Supplementary Fig. 12). These data are consistent with the eQTL data, where the risk-associated A allele is correlated with higher MX2 expression.

To thoroughly investigate possible mechanisms of allelic MX2 expression in relation to rs398206, we performed additional QTL analyses in melanocytes addressing alternative modes of gene regulation, including splice-QTL (sQTL) and DNA methylation QTL (meQTL). sQTL analyses using LeafCutter30 suggested that the main effect of the MX2 eQTL was not driven by alternative isoforms or splicing events (Supplementary Fig. 10b; Supplementary Note). meQTL analysis, on the other hand, identified a significant meQTL for rs398206 at a CpG probe near the MX2 canonical promoter, where the melanoma risk-associated A allele is correlated with lower CpG methylation, which is consistent with higher expression of the full-length isoform (Supplementary Fig. 13). Two other CpG probes in the first intron of MX2 (closer to rs398206) also displayed significant meQTLs for rs398206 in melanocytes, where higher CpG methylation is correlated with the risk A allele. These observations are consistent with the previous findings that DNA methylation in promoters is negatively correlated with gene expression, while that of transcribed regions is positively correlated with gene expression31–35. Taken together, eQTL, sQTL, and meQTL data are consistent with the hypothesis that MX2 full-length transcript mainly accounts for the eQTL at rs398206 in melanocytes through a transcriptional mechanism.

rs398206 regulates MX2 levels via allelic binding of YY1. To identify protein factors mediating the allelic difference observed in MPRA, we performed comparative mass-spectrometry using a 21 bp 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 cancer36, as well as in pigmentation pathways of melanocytes37. 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
**Fig. 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 19 variants located in the first intron of MX2 coming 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. Ensembl predicted transcripts from archive 75 are shown.

**b** Transcriptional activity of 145 bp sequences encompassing rs398206 from MPRA are shown as normalized tag counts (log2(RNA TPM/DNA TPM)). Results from UACC903 melanoma cells are shown for both alleles in forward (For) and reverse (Rev) 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 145 bp 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. Center lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles. \( P \)-value and slope were derived from linear regression with no multiple-testing correction applied. Source data are provided as a Source Data file.
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 scanned a ~5 Mb genomic region encompassing rs398206 in three melanoma cell lines representing each genotype of rs398206 (AA, AC, and CC). We observed prominent enrichment of YY1 binding on top of rs398206 in the AA cell line, a weaker but clear enrichment in the AC line, and even weaker binding enrichment over rs398206 in the CC line (Fig. 4a; Supplementary Fig. 14). Given that differences between cell lines (e.g., DNA copy number differences, accessibility of chromatin in the region encompassing rs398206, YY1 levels, and variability in formaldehyde fixing and chromatin shearing efficiency) may also contribute to differential YY1 binding, we also assessed allele-specific YY1 binding in the heterozygous AC cell line (UACC647). 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, $P = 9.1 \times 10^{-3}$), 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, $P = 1.5 \times 10^{-3}–1.9 \times 10^{-5}$, one-sample Wilcoxon test; Fig. 4a; Supplementary Fig. 15a) 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 in the same melanoma cell line. Four gRNAs

---

**Fig. 3 rs398206 displays allele-preferential binding to YY1.**

**a** Quantitative mass-spectrometry of rs398206 using nuclear extract of UACC903 and 21 bp 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(H)-mass or low(L)-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. Proteins passing the inter-quartile >3 cutoff for both axes are color-coded for consistent (red circle) and inconsistent (yellow circle) direction from label-swapping. Names of consistently identified proteins 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 21 bp 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. Representative sets from three replicates (c) and one replicate (d) are shown. Source data are provided as a Source Data file.
targeting the genomic regions either directly overlapping rs398206 (gRNA 1, 3, and 4) or ~25 bp 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. 15b). 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 the 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. 15c–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 TADs39. 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 a strong interaction with the adjacent bin encompassing MX2 promoter area39 (PET count = 18, \( P = 2.27 \times 10^{-8} \), hypergeometric test; Supplementary Fig. 16). Together these data determined that rs398206 is a functional variant regulating MX2 expression via differential YY1 binding in the Chr21q22.3 melanoma locus.

**MX2 accelerates melanoma formation in zebrafish.** MX2 is best known for its function in innate immunity as an HIV-1 restriction gene40,41. 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 induction42. 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 coexpressed 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 106 individuals. We identified 253 differentially expressed genes in MX2-high melanocytes (FDR < 0.01 and \( \log_{2} \) fold difference > 1; Supplementary Data 2), which include many of the top correlated immune-response genes based on pairwise comparisons of all expressed genes in the full melanocyte set (Supplementary Data 3; Supplementary Table 10). Significantly enriched pathways in these 253 differentially expressed genes included those relevant to cellular immune response as might be expected, but also included those affecting cellular growth and cancer (Fig. 6a; Supplementary Table 11) suggesting a possible melanocyte-specific function of MX2 not limited to immune function. On the other hand, an examination of immune infiltrates in melanomas from TCGA did not provide sufficient evidence for the roles of MX2 in tumor immune surveillance other than weak correlations with infiltration of a few cell types (Supplementary Note; 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 (~20-fold induction at 72 h; Supplementary Fig. 18a, b) resulted in slightly decreased growth of both melanoma cells and primary melanocytes at 100 ng/ml of doxycycline treatment, while empty vector transduced cells did not show any difference (Fig. 6b, c). 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 at 72 h; Supplementary Fig. 18a, c). Differentially expressed genes in MX2-overexpressing melanocytes compared to controls (158 genes, FDR < 10%; melanocytes from 3 individuals, 3 biological replicates each; Supplementary Data 4) displayed enrichment of pathways relevant to immune response, as well as those involving second messenger mediated kinase signaling and cellular growth, among others (Supplementary Table 12; Fig. 6d). We subsequently examined pathways enriched at a shorter time point after MX2 induction (6 h, melanocytes from one individual, 3 biological replicates) and observed consistent results highlighting intracellular second messenger
signaling, cancer, cell growth, neurotransmitters, cardiovascular signaling, in addition to cellular immune response (Supplementary Note; Supplementary Data 5 and 6; Supplementary Tables 13 and 14). Since these data did not provide strong support for a specific 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\textsuperscript{V600E}. Using the previously developed miniCoopR transgene system\textsuperscript{43}, we over-expressed human MX2 exclusively in the melanocytic-lineage using an MITF promoter in the background of BRAF\textsuperscript{V600E} and p53\textsuperscript{−/−}. 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, f). While we did not see significant difference in tumor count and pigmentation between the two groups, we observed larger tumor size in the MX2 group compared to GFP control group at median onset (P = 0.028; chi-square test), which is consistent with accelerated tumor formation demonstrated through survival analysis (Supplementary Fig. 19). We further performed histopathology analyses of zebrafish melanomas and did not observe gross difference in tumor morphology and invasion between MX2 and GFP groups (Supplementary Fig. 20a–d). Immunohistochemistry for phospho-Histone H3 also did not display a significant difference in cell proliferation between two groups (P = 0.47, two-tailed t-test, n = 5; Supplementary Fig. 20e–g). Together these data are consistent with MX2 expression contributing to an increased melanoma risk, however consistent with in vitro data, likely through a mechanism other than increased cell proliferation.

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. We demonstrate that MPRA is a high-throughput variant prioritization tool complementing statistical fine-mapping. While Bayesian fine-mapping methods could nominate a small number of credible causal variants for functional testing, these methods are nonetheless limited by their dependence on imputability, imputation quality, population LD reference, causal assumptions, and choice of functional annotation datasets. Further, variants that are tightly linked by LD are often still difficult to distinguish based solely on the genetic data and require individual functional testing. To this end, MPRA provides an agnostic approach to quickly screen a large number of variants linked by LD without relying on assumptions about LD structure or number of causals. By applying a conservative cutoff, we identified 39 variants displaying allelic transcriptional activity from 14 melanoma GWAS loci and showed that they are more likely to change transcription factor binding preference and more likely to be causal compared to the rest of the tested variants. Starting from an average of 52 high-LD melanoma-associated variants per locus that overlap active chromatin regions, MPRA narrowed the candidate variants down to an average of 2.78 per locus. These results highlight the utility of our approach in complementing statistical fine-mapping and breaking up LD structure using a functional assay. To further prioritize variants functioning through cis-regulation of local gene expression, we integrated the 39 MPRA significant variants with melanocyte eQTL data which best represents the cell type of origin for melanoma. By choosing the variants displaying the same allelic directions as those from eQTL, we nominated a short list of candidate genes and variants from four loci. In particular, from the locus on chromosome band 21q22.3, 5 of 22 total tested variants are in near perfect LD (r² > 0.99) with the GWAS lead SNP (rs4088285), including the lead SNP from the original study identifying this locus (rs45430) (Supplementary Table 8). While both Bayesian fine-mapping using PAINTOR in this study and eQTL colocalization using eCAVIAR from our previous study24 assigned higher probability scores for some of these perfect LD SNPs as causal variants for melanoma risk and/or eQTL, MPRA results for this set of 22 SNPs determined rs398206 (r² = 0.94 with the lead SNP) as the most significant functional variant. Integrating this result with melanocyte eQTL nominated MX2 as the best candidate susceptibility gene, and we performed thorough functional characterization of this locus including generating an animal model as a proof of principle.

In addition, our integrative approach efficiently identified the most plausible susceptibility genes and functional variants from three other 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 and plays a role in the oxidative stress response, which is relevant to melanomagenesis, given the vulnerability of melanocytes to oxidative stress attributable to melanin production. 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. 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, and GWAS also implicated the CASP8 locus for breast cancer and basal cell carcinoma. Our results provide strong support for these three genes and warrant further in-depth characterization.

Our approach also successfully identified a small number of candidate functional variants (median 1 variant per locus) in nine
melanocyte loci not currently supported by cell-type specific melanocyte eQTLs (Supplementary Table 6). These variants may represent plausible functional candidates for loci where we are underpowered to detect melanocyte cis-eQTLs from only 106 cultures. Alternatively, as our eQTL dataset is based on gene expression patterns captured in cultured primary melanocytes, it is possible that these variants may function within specific cellular contexts not captured in this eQTL dataset. Given that our MPRA assay was conducted in melanoma cells, it may be possible that such variants, for example, may only be functioning in the context of oncogenic signaling (e.g., oncogenic BRAF), tumor suppressor loss (e.g., CDKN2A), or immortalizing events (e.g., mutation of the TERT promoter). Future evaluation of these variants in the context of additional genome-scale datasets, including cell-type specific chromatin interaction data, as well as chromatin features of different cellular contexts, may strengthen or clarify evidence for functionality of leads for these loci.

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 an alternative mechanism through splicing. While our data demonstrated the YY1 transcription factor plays a major role in MX2 regulation via the functional SNP, it is also possible that other trans-acting factors, perhaps among other proteins identified through mass-spec with less pronounced allelic difference (e.g., KLF9, ZBTB7A, PAX3), could also play a role in the observed allelic transcriptional 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 infection40,41, and its roles in melanoma-genesis have not been studied well. Our findings suggest a role of MX2 in promoting melanoma formation when exclusively expressed in cells of melanocytic-lineage, in the presence of BRAFV600E, a frequent somatic driver mutation. Unlike the relatively high level of MX2 overexpression in the zebrafish model, increased melanoma risk and MX2 levels associated with this common functional variant (rs398206 risk allele frequency = 0.61, EUR) in human populations are rather modest, which is nearly always the case for cancer-associated common variants. It is possible that the effect of increased MX2 levels have roles only in specific contexts during the process of melanoma development. Our single cell-type growth assays for MX2 also show growth effects for MX2 in melanocytic-lineage in the absence of neighboring cell types from the skin, albeit in the opposite direction. A recent study also reported that MX2 over-expression resulted in decreased growth in primary melanocytes and melanoma cells, while the effect could be the opposite in a subset of melanoma cells.53. Our zebrafish tumor proliferation assay also suggested that accelerated melanoma formation by MX2 might not be due to increased proliferation per se. Given this apparent complexity, we speculate that MX2 function in the growth of cells of melanocytic lineage might depend on cellular context (e.g., somatic driver events) and or interaction with microenvironment (e.g., immune cells). Further studies will be required to tease out the discrepancy between single cell-type assay and zebrafish model, as well as the molecular mechanisms of how MX2 contributes to melanoma promotion. Nevertheless, our findings established MX2 as a 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, Chr1q42.2)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 et al.1 was conducted following a very similar approach to that of Barrett et al.32. Using the results from Law et al.32 a window was defined as 1 Mb 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 6 Mb 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 available). Each region was further narrowed down to the interval covering 500 kb on either side of the most extreme SNPs with p-value < 10−8 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 HyperLasso33 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 1.0035.7 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 (r2 > 0.9). By dropping equivalent models, a reduced set of models was produced and was then further reduced by dropping each 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 evidence19. 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 MCIR gene (16q24.3) on melanoma risk is mainly explained by several well-established coding variants12, we did not include this region in our fine-mapping data.

**MPRA variant selection.** Among 20 genome-wide significant loci from the melanoma meta-analyses by Law and colleagues1, we prioritized 16 loci where potential leads could be applied. We selected 4 loci containing genes that are implicated in melanoma-associated pigmentation phenotypes (SLC45A2, TYR, MCIR, and ASIP loci), as for many of these genes, coding variants were shown to alter the protein functions. To select high-LD proxy variants for 16 melanoma GWAS loci (Law et al., 2015)1, we used the following criteria:

1. Primary lead SNPs were taken from Law et al.1 meta-analysis paper and supplemented by those from additional HyperLasso analysis when there are alternative best SNPs available.
2. For 8 loci, HyperLasso analysis nominated independent multiple secondary signals and these lead SNPs were also added.
3. SNPs of r2 > 0.4 with the primary or secondary lead SNPs using 1000 Genomes phase3 EUR or CEU populations were selected as high-LD variants (n = 2748).

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

1. Variant is located within a human melanocyte DHS peak from one or more individuals of available through ENCODE and Epigenome Roadmap database.
2. Variant is located within a human melanocyte H3K27Ac ChIP-seq peak from one or more individuals and a H3K4Me1 ChIP-seq peak from one or more individuals of two 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 EPIC chip-seq.

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-020-16590-1 | www.nature.com/naturecommunications
more individuals of two and three available through Epigenome Roadmap database, respectively.

4. Higher number of transfected cells than the library complexity. The number of promoter type) into HEK293FT or UACC903 melanoma cells aiming >100 times using Qiagen RNeasy kit (Qiagen, Hilden, Germany), and mRNA was subsequently reverse transcription-PCR. The ligation product was transformed by electroporation into 10 times more individuals of two and three available through Epigenome Roadmap database, respectively.

MPRA oligo library design. Oligo libraries were designed mainly following the guidelines from published work21,27 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, 18,840 repeated in both) were synthesized by Agilent Technologies (Santa Clara, CA). Composition of each library by GWAS loci and repeated variants are listed in Supplementary Data 1 and Supplementary Table 3. 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 amplification (Supplementary Table 2). A total of 100 oligo sequences were used for each barcode. The 10 base sequences and scrambled 21 base core sequence, only homopolymers of ≤4 bases were used and the enzyme recognition sites for KpnI, XbaI, and SfiI were avoided. A complete list of oligo sequences can be found in Supplementary Data 7.

MPRA library construction and sequencing. MPRA library construction and sequencing was performed following published protocols with some modifications21,27,28. Random 10 base tag sequences were generated once so that each library has up to 36,660 unique tag sequences (the same 36,660 tag sequences were used for each library). The 10 base sequence and scrambled 21 base core sequence, only homopolymers of ≤4 bases were used and the enzyme recognition sites for KpnI, XbaI, and SfiI were avoided. A complete list of oligo sequences can be found in Supplementary Data 7.

MPRA data analyses. Using FASTQ files from input DNA or RNA transcript sequencing, we counted the number of reads ( Illumina run 1) completely matching 10 bp barcode sequences (tag counts) and the same downstream sequence context (TCTAGATTATACCCGCC) including an XbaI recognition site and the 5’ 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, converted.

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 78.4%–93.3% of barcode sequences were detected, and 84.8%–90.8% of barcode sequences were detected. Two libraries combined (10% TBE-Urea gel-puri- fied oligo libraries were amplified by emulsion PCR using Herculase II fusion polymerase (Agilent) and 2 µM of primers providing SfiI enzyme sites (Supplementary Table 15), following the instructions of the Micellula DNA Emulsion & Purification Kit (EUA/CHIMERX, Milwaukee, WI). Amplified oligos were sequenced to estimate the PCR assay and verified by DNA sequence analysis in PGM. Amplicon libraries were prepared using 30 ng of oligos from emulsion PCR using Ion Plus Fragment Library Kit (Thermo Fisher, Waltham, MA) and were sequenced using Ion Plus Fragment Library Kit (Thermo Fisher, Waltham, MA) and were sequenced on Ion PGM for an average 203 bp and 175 bp read length at 6.7 million and 5.6 million reads per sample for library 1 and library 2, respectively. To verify oligo binding, 21 base sequences within including variant sites and ≤10 bp were used to map to each sequencing read. Linux command fgrep was used and only 100% sequence match was kept. We then counted the total read depth for each variant represented by the matched sequences, and then calculated the proportion of variant sequences that were verified. For both library 1 and library 2, more than 90% of unique sequences representing the variants in the library were detected with at least 10 sequencing reads. In addition, we found similar proportion and read depth for sequences representing both forward and reverse directions in both libraries. If we use the actual tag sequences as a bait, 82% of tags could be verified, with a caveat that some tags were amplified but not detected because of relatively poor sequencing quality in this position of the amplicon. Sequence-verified oligo libraries were first cloned into pMPRA1 vector (Addgene, Water- town, MA) using SfiI site by electroporation into 10 times higher number of bacterial cells than the number of unique sequences in the oligo library. Cloned pMPRA1 was further digested on KpnI and XbaI sites between 145 bp test sequence and 10 bp barcode sequence, where luciferase ORF was with or without minimal promoter was ligated from pMPRADonor2 and pMPRADonor1 (Addgene), respectively. The ligation product was transformed by electroporation into 10 times higher number of bacterial cells in the same manner. Cloned final library for transfection was verified on the gel as a single band after KpnI digestion. Each library was transfected at least four times (two transfections for each promoter type) into HEK293FT or UACC903 melanoma cells aiming >100 times higher number of transfected cells than the library complexity. The number of transfected cells were estimated using transfection efficiency measured by a separate GFP transfection and visualization. A summary of transfections is listed in Supplementary Table 15. Transfections were transfected with Lipofectamine 2000 and harvested at 24 h after transfection for RNA isolation. Total RNA was isolated using Qiagen RNeasy kit (Qiagen, Hilden, Germany), and mRNA was subsequently harvested at 24 h after transfection for RNA isolation. Total RNA was isolated using Qiagen RNeasy kit (Qiagen, Hilden, Germany), and mRNA was subsequently isolated using PolyA purist MAG kit (Thermo Fisher). cDNA was then synthesized using Superscript III, from which only short sequences encompassing 10 bp unique barcodes were amplified using Q5 high-fidelity polymerase (NEB, Ipswich, MA) 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 variants with melanocyte eQTL variants. Significant eQTL genes were defined as candidate genes significantly identified by eCAVIAR (CLP > 0.01) or TWAS (P-value < 0.05/number of genes tested) in melanocytes24. Among all 832 variants, variants were selected if they display genome-wide sig- nificant melanocyte eQTL P-values (nominal P-value < gene-level genome-wide threshold as defined in Zhang et al.24) for one of the significant eQTL genes. Six genes (ARNT, CTSS, PARP1, CASP8, MX2, and MAFF) from five loci had signi- ficant eQTL SNPs included in MPRA, and the final MPRA-significant variants overlapped eQTL SNPs in four loci, where a subset of SNPs also displayed the same allelic direction 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 Data 7.

Integration of MPRA variants with melanocyte eQTL variants. Significant eQTL genes were defined as candidate genes significantly identified by eCAVIAR (CLP > 0.01) or TWAS (P-value < 0.05/number of genes tested) in melanocytes24. Among all 832 variants, variants were selected if they display genome-wide sig- nificant melanocyte eQTL P-values (nominal P-value < gene-level genome-wide threshold as defined in Zhang et al.24) for one of the significant eQTL genes. Six genes (ARNT, CTSS, PARP1, CASP8, MX2, and MAFF) from five loci had signi- ficant eQTL SNPs included in MPRA, and the final MPRA-significant variants overlapped eQTL SNPs in four loci, where a subset of SNPs also displayed the same allelic directions between MPRA and melanocyte eQTL for four genes (CTSS, CASP8, MX2, and MAFF; Supplementary Table 6).

Motif analysis. Prediction of variant effects on transcription factor binding sites was performed using the motifbreakR package29 and a comprehensive collection of human transcription factor binding sites models (HOCOMOCO)30. 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 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.uc.edu/software/painter) 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 complemented with a melanocyte specific gene set annotation24. 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 true specific posterior probabilities. These 4 annotations include melanocyte-specific expressed genes from our melanocyte dataset24, melanocyte enhancers, transcribed regions, and a histone mark (H3T11ph) 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 and meQTL for the MX2 locus. Primary melanocyte eQTL data was obtained from our previous study24, where 106 individuals mainly of European decent were analyzed. For the marginal eQTL analysis of the genes located in the TAD including rs938206, 21 genes were selected based on expression thresholds of >0.5 by RSEM quantification (RNA-Seq by Expectation Maximization29) and 26 reads in at least 10 samples. Using FastQTL30, nominal P-value was generated between each gene and all the SNPs +− 2 Mb of rs938206 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 rs938206. For sQTL and meQTL, we performed similar QTL analyses as our previous eQTL study using the same genotype data, population structure covariates, and statistical approaches. We replaced normalized expression levels with normalized splice junction events (sQTL), and normalized methylation values (meQTL). We also re-calculated the top 15 PEER factors according to these phenotype values. For sQTL analysis, STAR29 was used to map the RNA-Seq reads onto the genome (hg19) and then LeafCutter30 was applied to quantify the splice junctions following the procedures described in http://bioconductor.org/packages/minfiR (http://bioconductor.org/packages/minfiR.html). For meQTL analysis, we performed genome-wide DNA methylation profiling on Illumina Infinium Human Methylation 450 K 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 pipeline62 and the miniprocessor system as described previously65. 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 analysis, normalization was performed using the preprocessFunFunnal algorithm implemented in minfi R package63.

MX2 isoform analysis. TagMan assays targeting unique junctions of MX2 transcript isoforms were obtained from Thermo Fisher (Waltham, MA, full-length transcript: Hs01550809_m1 and AP323EZ; ENST00000543692: APYMKKU; MX2_fusion: Hs01550809_m1 and AP323EZ; ENST00000543692: APYMKKU; MX2_cDNA: Hs01550809_m1 and AP323EZ; ENST00000543692: APYMKKU). Custom assay design was based on Ensembl7/GRCh37. RNA was isolated from primary cultures of melanocytes from 106 individuals mainly of European decent24, and CDNA was synthesized using iScript Advanced cDNA Synthesis Kit (Bio-Rad). TagMan 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 RNA-seq data.

Cell culture. Melanoma cell lines (UACC903, UACC647, UACC2331, UACC502, UACC2545, and UACC612) 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 145 bp sequence encompassing rs938206 as tested in MPRA was amplified from genomic DNA of HapMap CEU panel samples carrying either risk or protective allele. Primers were designed from standard curve of each primer set and normalized to 1/100 input DNA. For genotyping rs938206, input DNA or genomic DNA from each cell line, and pulled-down DNA from UACC647 cell line (heterozygous for rs938206) was used as input template DNA for TaqMan genotyping assay (Assay ID: C_2262545_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). 21 bp oligonucleotide probes encompassing rs938206 were ordered via exact same 145 bp sequence from Integrated DNA Technologies targeting a known YY1 binding site near 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 filter plate as described previously46. In short, 500 pmol of annealed DNA oligos were immobilized on 10 μl (20 μl after 2×) streptavidin-Sepharose beads (GE Healthcare, Chicago, IL) 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 overnight digested with 200 ng of calf kidney pancreatic RNase A and 200 ng of calf kidney pancreatic DNase I. The peptides were eluted from the column using a gradient from 9 to 32% Buffer B (80% acetonitrile, 0.1% formic acid) in 114 min at a flow rate of 250 μL/min using an Easy-NLC 1000 (Thermo Fisher). Samples were sprayed directly into a Thermo Fisher Orbitrap Fusion Trubrid mass spectrometer. Target values for full MS were set to 3±5 AGC target and a maximum injection time of 59 ms. Full MS were recorded at a resolution of 120,000 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 60 s. Target values for MS/MS were set at 2±4 AGC target with a maximum injection time of 155 ms set to ‘rapids’ with a dynamic exclusion of 1.6 m/z and collision energy of 35%. Scans were collected in data-dependent top-speed mode in cycles of 3 s. Thermo RAW files were analyzed with MaxQuant 1.6.0.1 by searching against the UniProt curated human protein database (released June 2017) with standard settings64. Protein identifications were normalized by median ratio shifting and using mixture calling. An oq value 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 cell line UACC647 melanoma cell line using ON-TARGETplus 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 × 105 cells using Lipofectamine 2000 (ThermoFisher Scientific) following the reverse transfection procedure in 24-well format. Cells were harvested at 72 h 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 by using Taqman probe set (Assay ID: Hs01550809_m1) specifically detecting the full-length isoform and normalized to GAPDH 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 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-MX2 [NP_81018, Novus Biologicals], and mouse anti-GAPDH (sc-51907, Santa Cruz). Uncropped gel images and molecular weight marker images for western blot analysis are provided for the corresponding figure in Source Data file.

**DIFFERENTIALLY EXPRESSED GENES IN MX2-HIGH MELANOGENES.** 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. Totals counts of mappable reads for each annotated gene (GENCODE v19) was obtained using featureCounts from Rsubread package69. The final DEG list with criteria FDR < 0.01 and |log2 ratio| > 1.32e-05 was applied to Ingenuity Pathway Analysis (IPA). We also performed pairwise gene expression correlation between MX2 and the rest of 37,854 genes expressed in these 106 primary melanocytes. Bonferroni-correction was applied to P-values of Pearson correlations FDR < 0.05. Negative correlation levels significantly correlated (FDR < 0.05, N = 377) with those of MX2 were selected for IPA analysis.

**MX2-allele specific expression.** Melanocytes grown in Dermal Cell Basal Medium (ATCC PCS-200-030) supplemented with Melanocyte Growth Kit (ATCC PCS-200-041) and 1% amphotericin B/penicillin/streptomycin (120-096-711, Quality Biological) as described before23. Total RNA was isolated using a mirNeasy Mini kit (217004, Qiagen) and treated with CTAB-Urea following a previously described method88 to remove excess melanin pigmentation. cDNA was synthesized from total RNA using iScript Advanced cDNA Synthesis Kit (Bio-Rad). Genomic DNA and cDNA were then genotyped for rs398206 using custom Taqman genotyping probe set (ANRWETYM) recognizing both genomic DNA and cDNA (ENST0000534692) with a 5 bp homologous sequence at the 3’ UTR end for cDNA and 5’ end for genomic DNA. Genotyping analysis 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 module72 (https://multiqc.info). Quasi-mapping algorithm Salmon27 was used to provide fast and bias-aware quantification of transcript expression levels. For MX2 induction experiment, genotyping and sequencing were performed in at least 3 biological replicates. The final component principal component analysis was performed based on the expression quantification, and based on the results, differentially expressed genes (DEGs) were calculated with DESeq222 adjusting for cell line, biological replicate, and library construction batch. The expression threshold FDR < 0.1 was recognized as DEGs after adjusting for multiple testing. 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. infected with pNDUCellR20-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 12 and 16).

**MX2 over-expression and growth assays.** Melanoma cells and melanocyte growth assays were conducted using lentiviral transduction of MX2 cdNA under the control of tetracycline-inducible promoter using pNDUCellR20 vector (Addgene; 15780, TRC-204637) infected with pCMV6-entrant vector (Addgene; 20222, TRC-204637). Cells were 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 pNDUCellR20 vector (adapter sequence is listed in Supplementary Table 15). BamHI and MluI sites on pCMV6-entrant 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 Stem Cell 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 pPAX2, pMD2-G, and PGC4-A-RT2. Virus was collected two days after transfection and concentrated by Vivasin20. Cells were incubated with virus for 24 h, 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 10,000 cells/cm2 in 24-well plates using log-rank test. Pigmentation, size, and number of melanomas per fish were qualitatively assessed at median onset in the two groups. Data from the three independent experimental sets were combined and P-values were calculated using
chi-square test. For histology analyses, zebrafish were euthanized at 20 weeks of age, and fixed in 4% paraformaldehyde overnight at 4°C. Paraffin embedding, sectioning, hematoxylin and eosin (H&E) staining were performed according to standard techniques by the Brigham & Women’s Hospital Pathology Core. Immunohistochemistry was performed on the Leica Biosystems Bond III automated staining platform. Phospho-Histone H3 (Ser10) Antibody (#9701, Cell Signaling Technologies, Danvers, MA) was run at 1:500 dilution using the Leica Biosystems Refine Detection Kit with citrate antigen retrieval, after boiling of the melanin pigment. Histologic sections were imaged using an Olympus BX41 microscope with an Olympus DP70 camera. p-H3 foci were counted on 7 distinct tumor areas per fish for a total surface area of 1 mm². Zebrafish were handled humanely according to our vertebrate animal protocol that implements the principles of replacement, reduction, and refinement (“three R’s”), 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.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The sequencing data generated during this study (MPRA sequencing and RNA-seq data) are deposited in Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) as a SuperSeries under the accession number GSE129250 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE129250]. A complete list of oligo sequences for MPRA libraries and complete processed MPRA data can be found in Supplementary Data 7, as well as in the Source Data. Melanocyte eQTL data and RNA-seq expression data from 106 individuals are available through the database of Genotypes and Phenotypes (dbGAP, [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= gönderilen GSE99519]). Data from the 2015 melanoma GWAS meta-analysis performed on chromosome 22q13 associated with development of cutaneous nevi. Nat. Genet. 41, 840 (2008).

Received: 6 June 2019; Accepted: 12 May 2020;
Published online: 01 June 2020

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, 886–891 (2008).
3. Brown, K. M. et al. Common sequence variants on 20q11.22 confer melanoma susceptibility. Nat. Genet. 40, 838–840 (2008).
4. Bishop, D. T. et al. Genome-wide association study identifies three loci associated with melanoma risk. Nat. Genet. 41, 920–925 (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, 915–919 (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 melanoma due to BMI. Nat. Genet. 41, 1114–1118 (2011).
8. Barrett, J. H. et al. Genome-wide association study identifies three new melanoma susceptibility loci. Nat. Genet. 43, 1108–1113 (2011).
9. Ransohoff, K. J. et al. Two-stage genome-wide association study identifies a novel susceptibility locus associated with melanoma. Oncotarget 8, 15786–15792 (2017).
10. Duffy, D. L. et al. Multiple pigmentation gene polymorphisms account for a substantial proportion of risk of cutaneous malignant melanoma. J. Invest. Dermatol. 130, 520–528 (2010).
11. Beaumont, K. A. et al. Receptor function, dominant negative activity and phenotype correlations for MCIR variant alleles. Hum. Mol. Genet. 16, 2259–2260 (2007).
12. Tetschkladze, Z. R. et al. Functional assessment of human coding mutations affecting skin pigmentation using zebrafish. PLoS ONE 7, e47398 (2012).
13. Halaban, R. et al. Endoplasmic reticulum retention is a common defect associated with tyrosinase-negative albinism. Proc. Natl Acad. Sci. USA 97, 5559–5562 (2000).
14. Liu, F. et al. Genetics of skin color variation in Europeans: genome-wide association studies with functional follow-up. Hum. Genom. 9, 833–835 (2015).
15. Duffy, D. L. et al. Novel pleiotropic risk loci for melanoma and nevus density implicate multiple biological pathways. Nat. Commun. 9, 4774 (2018).
16. Cho, J. et al. Common single nucleotide polymorphisms are associated with susceptibility to melanoma. Am. J. Hum. Genet. 87, 6–16 (2010).
17. Gue, C. et al. The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. Nature 471, 513–517 (2011).
44. Chundra, V. K. et al. Examining the impact of imputation errors on fine-mapping using DNA methylation QTL as a model trait. *Genetics* **212**, 577–586 (2019).

45. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **19**, 491–504 (2018).

46. Amit, I. et al. A module of negative feedback regulators de... (2014).

47. Katsuoka, F. & Yamamoto, M. Small Maf proteins (MafF, MafG, MafK): history, structure and function. *Gene* **586**, 197–205 (2016).

48. Denat, L., Kadekaro, A. L., Leachman, S. A. & Abdel-Malek, Z. A. Melanocyt... (2015).

49. McIlwain, D. R., Berger, T. & Mak, T. W. Caspase functions in cell death and signaling. *Nat. Rev. Cancer* **15**, 712–729 (2015).

50. McIlwain, D. R., Berger, T. & Mak, T. W. Caspase functions in cell death and signaling. *Cold Spring Harb. Perspect. Biol.* **5**, a018656 (2013).

51. Michailidou, K. et al. Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92–94 (2017).

52. Stacey, S. N. et al. New basal cell carcinoma susceptibility loci. *Nat. Commun.* **6**, 6825 (2015).

53. Juravlevic-K, M. et al. MX 2 is a novel regulator of cell cycle in melanoma cells. *Pigment Cell Melanoma Res.* **33**, 446–457 (2019).

54. Barrett, J. H. et al. Fine mapping of genetic susceptibility loci for melanoma reveals a mix of single variant and multiple variant regions. *Int. J. Cancer* **136**, 1351–1360 (2015).

55. Stephens, M. & Balding, D. J. Bayesian statistical methods for genetic association studies. *Nat. Rev. Genet.* **10**, 681–690 (2009).

56. 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**, 655–664 (2011).

57. Long, J. S. & Ervin, L. H. Using heteroscedasticity consistent standard errors in the linear regression model. *Am. Statistician* **54**, 217–224 (2000).

58. 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).

59. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

60. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479–1485 (2016).

61. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

62. Assenov, Y. et al. Comprehensive analysis of DNA methylation data with RnBeads. *Nat. Methods* **11**, 1138–1140 (2014).

63. Arvey, M. J. et al. Minfi: a flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1390–1392 (2014).

64. Spruijt, C. G., Baymaz, H. I. & Vermeulen, M. Identifying specific protein-DNA interactions using SILAC-based quantitative proteomics. *Proteomics* **16**, 4177–4192 (2016).

65. 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**, 1367–1372 (2008).

66. 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).

67. Lagonigro, M. S. et al. CTAB-urea method purifies RNA from melanin for cDNA microarray analysis. *Pigment Cell Res.* **17**, 312–315 (2002).

68. Liao, Y., Smyth, G. K. & Shi, W. The DESeq2 and EdgeR-Based R pipeline for comprehensive differential analysis of RNA-seq data. *PLoS ONE* **11**, e0157022 (2016).

69. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

70. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

71. 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).

72. 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).

**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, H. Kong, and L. Mehl from the National Cancer Institute, Laboratory of Translational Genomics for proofreading the manuscript, and A. Jermakyc, O. Otubajo, L. Jessop, and L. Arnundadottir from National Cancer Institute, Laboratory of Translational Genomics for helpful discussions, and S. Lotus 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. This work was also supported by grants from the National Institutes of Health (RO1 CA110384, P101 CA163222, RO1 CA1083115), Cancer Research UK (C588/A19167), Institut National du Cancer, France (INCa_5982), Programme Hospitalier de Recherche Clinique (AOM-07-195), Fondation pour la Recherche Médicale (FDT2010928343), and the Howard Hughes Medical Institute Investigator Program (L.I.Z.). The Vermeulen lab is part of the Oncode Institute, which is partly funded by the Dutch Cancer Society (KWF). We also thank all the cohorts, funders, and investigators who contributed to the melanoma GWAS, acknowledged by Law and colleagues, from which data was used towards fine-mapping. 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.

**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 R.C. designed and analyzed MPRA assays. J.C., A.V., R.C.H., J.Y., and M.X. conducted experiments for MPRA and MX2 molecular characterization. R.C. designed CRISPR assays. T.Z. and K.M.F. performed data analyses. M.M.G., C.C., 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. A.V. and J.A. contributed equally to this work.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-16590-1.

Correspondence and requests for materials should be addressed to K.M.B.

Peer review information *Nature Communications* thanks Bernhard Weber and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2020.