Functional characterization of a multi-cancer risk locus on chromosome band 2q33.1 near *CASP8*

Hyunkyung Kong[1], Jiyeon Choi[1], Tongwu Zhang[2], Cathrin Gräwe[3], Mai Xu[1], Rohit Thakur[1], Hayley Sowards[1], Rebecca C Hennessey[1], Andrew Vu[1], Jianxin Shi[4], D Timothy Bishop[5], Julia Newton-Bishop[6], Jeremie Nsengimana[7], Mark M. Iles[6,8], Maria Teresa Landi[2], Michiel Vermeulen[3,9], Matthew H. Law[10,11,12], Laufey T. Amundadottir[1], Melanoma Meta-Analysis Consortium[13], Kevin M Brown[1]

[1]Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, 20892, USA

[2]Integrative Tumor Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, 20892, USA

[3]Department of Molecular Biology, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, Nijmegen, 6525 XZ, Netherlands

[4]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, 20892, USA

[5]Leeds Institute of Medical Research, School of Medicine, University of Leeds, Leeds, UK

[6]Leeds Institute for Data Analytics, University of Leeds, Leeds, UK

[7]Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University and NIHR Biomedical Research Centre, Newcastle upon Tyne, UK

[8]NIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals NHS Trust, Leeds, UK

[9]Division of Molecular Genetics, Netherlands Cancer Institute, Amsterdam, Netherlands

[10]Population Health Department, QIMR Berghofer Medical Research Institute, Herston, QLD, Australia

[11]School of Biomedical Sciences, Faculty of Health, Queensland University of Technology, Brisbane, QLD, Australia

[12]School of Biomedical Sciences, University of Queensland, Brisbane, QLD, Australia

Corresponding author: kevin.brown3@nih.gov

Abstract

Genome-wide association studies (GWAS) of melanoma have identified numerous susceptibility loci. However, causal genes and variants underlying risk have yet to be established for most. It is becoming apparent that many functional variants underlying complex traits act via *cis*-regulation that may be context-specific, dependent on availability of specific transcription factors/complexes in specific cell types and cell-states. To characterize a risk locus on chromosome band 2q33.1 associated with melanoma, breast cancer, and keratinocyte cancers, we integrated fine-mapping, cell-type specific expression quantitative trait locus (eQTL) analysis, a massively parallel reporter assay, individual luciferase assays, and SNP-based proteomics. Integrated analysis implicates the presence of multiple functional variants lying primarily within a promoter for *CASP8*. A haplotype containing rs3769823 appeared have the largest effect on expression. Strikingly, both tumor/normal context and this risk-associated haplotype play critical roles in mediating allelic *cis*-regulatory activity. Quantitative mass spectrometry for rs3769823 identified both E4F1, a transcriptional repressor, and IRF2, a transcriptional activator, as binding preferentially to risk-associated rs3969823-A. The binding of these transcription factors was validated via EMSA, supershift, and chromatin immunoprecipitation (ChIP) assays. The relative levels of E4F1 and IRF2 differ by cell-type and play a role in mediating transcriptional activity in a cell-type specific manner. Our results indicate that the top credible causal set variant rs3769823 likely influences expression of *CASP8* and *FLACC1* in a cell-type specific manner and may be a relevant functional variant for multiple cancers associated with this locus.

Keywords

Melanoma; genome-wide association study; GWAS; expression quantitative trait locus; eQTL; fine-mapping; caspase 8; CASP8

Introduction

Over the past decade, genome-wide association studies (GWAS) have identified 54 genome-wide significant susceptibility loci for cutaneous malignant melanoma (CMM) in populations of European ancestry [1-10]. Of these, Barrett and colleagues initially identified a genome-wide significant locus for melanoma on chromosome band 2q33.1 adjacent to the Caspase 8 gene (*CASP8*) [5]. Multiple subsequent melanoma GWAS meta-analyses since have replicated this finding, including the most recent meta-analysis of 36,760 melanoma cases (rs10931936-T, $P = 2.12 \times 10^{-12}$, OR = 1.08)[10]. This region is a multi-cancer risk locus, with highly-correlated risk alleles for melanoma and other cancers sharing a common haplotype, including signals for breast cancer (rs3769821, $P = 3.97 \times 10^{-18}$, r^2 with rs10931936 = 0.63) [11], keratinocyte cancers including both cutaneous basal cell carcinoma (BCC, rs6714430, $P = 5.42 \times 10^{-55}$, r^2 with rs10931936 = 0.99), squamous cell carcinoma (SCC; rs10931936, $P = 1.03 \times 10^{-7}$), or both combined (rs6743068, $P = 1.50 \times 10^{-48}$, r^2 with 10931936 = 0.99)[12,13], and non-small cell lung cancer (rs3769821, $P = 4.45 \times 10^{-8}$, r^2 with rs10931936 = 0.63)[14]. Additionally, the region also harbors a highly-correlated signal for prostate cancer (rs59308963, $P = 2.41 \times 10^{-8}$, r^2 with rs10931936 = 0.93)[15], where the protective allele is highly correlated with risk alleles for melanoma and the other cancers. However, whether causal variants and genes are shared across multiple cancers remains under-explored.

The topologically-associated domain (TAD) at 2q33.1 harbors multiple strong *a priori* candidate genes with well-established roles in apoptosis, including Caspase-8 (*CASP8*), Caspase 10 (*CASP10*), and the CASP8 and FADD-like apoptosis regulator (*CFLAR*). CASP8 is an apical caspase that initiates apoptosis through the extrinsic pathway via a subset of proteins in the tumor necrosis factor superfamily (TNFRSF) that includes TNFR1 (tumor necrosis factor receptor-1), CD95/Fas, and TRAIL-R (TNF-related apoptosis-inducing ligand receptor) [16,17]. CASP8 is composed of a catalytic domain, a protease domain and two death domains comprising the prodomain, which is important for forming the death-inducing signaling complex (DISC), death receptor-FADD-CASP8 [18,19]. Like all caspases, CASP8 is present as a monomeric procaspase zymogen[20]. Binding of CASP8 to the DISC facilitates zymogen maturation by dimerization and proteolytic cleavage of the caspase itself [21,22]. Fully processed active CASP8 is released from DISC and activates downstream effector caspases. Notably, two proteins encoded by adjacent genes at this locus, *CASP10* and *CFLAR*, are highly similar to the N-terminus of *CASP8* and play a different role in the apoptotic pathway by interacting with FADD and CASP8 [23,24]. While CASP10 is an apical caspase like CASP8, CFLAR splice variants

inhibit different processing steps of CASP8 activation through heterodimerization [24-26]. Given that most common trait loci are thought to function via *cis*-gene regulation, often at long distances, other genes within this TAD cannot be ruled out as candidate causal genes.

In this study, we fine-mapped multiple cancer GWAS datasets and integrate tissue-based and cell-type specific quantitative trait locus (QTL) data in order to identify likely causal variants and genes. We identified multiple highly-correlated potentially *cis*-regulatory variants, and establish rs3769823 as a key functional variant, likely acting via altered binding of the IRF2 and E4F1 transcription factors. We establish *cis*-regulation of *CASP8* as a likely causal mechanism underlying melanoma and breast cancer risk at this locus, and given the strong LD between signals, this mechanism may potentially play a role in risk of other associated cancers.

Results

Comparison of the 2q33.1 melanoma risk signal with GWAS of susceptibility to other cancers

The most recent GWAS meta-analyses of melanoma identified rs10931936 as the sentinel variant at the 2q33.1 risk locus (rs10931936-T, $P = 2.12 \times 10^{-12}$, OR = 1.08)[10]. Conditional and joint analysis of summary GWAS meta-analysis data[27,28] identified no additional genome-wide significant association signals at this locus[10]. The next most significant signal within 1 Mb in either direction of rs10931936 is rs563855920 ($P_{conditional} = 3.05 \times 10^{-4}$; **Figure S1**; **Table S1**) more than 300 kb away, suggesting that there are no additional major signals.

This region has been identified as a genome-wide significant risk locus via GWAS of multiple other cancer types including breast cancer (122,977 cases, rs3769821-C, $P = 3.97 \times 10^{-18}$, 1000 Genomes EUR r^2 with rs10931936 = 0.63)[11] and keratinocyte cancers (47,742 cases, rs6743068-A, $P_{MTAG} = 1.50 \times 10^{-48}$, $r^2_{rs10931936} = 0.99$)[13], including cutaneous BCC (31,787 cases, rs6714430-C, $P = 5.42 \times 10^{-55}$, $r^2_{rs10931936} = 0.99$)[13] and SCC (9,674 cases, rs10931936-T, $P = 1.93 \times 10^{-7}$)[13] where the genotypes of these risk alleles are correlated. To formally assess whether these traits may share common causal variants with melanoma risk, we compared melanoma fine-mapping data to other cancer datasets using both HyPrColoc[29] and eCAVIAR[30] and summary statistic data for the set of SNPs +/- 100 kb centered over the melanoma GWAS lead SNP. Colocalization analyses suggest a likely shared causal between melanoma risk and breast cancer (HyPrColoc PP = 0.93; maximum eCAVIAR CLPP = 0.04). Across the larger

region, HyPrColoc finds the strongest evidence for shared causals for rs3769821 (HyPrColoc SNP score = 0.52, r^2 = 0.63 to rs10931936) and rs3769823 (SNP score = 0.11, r^2 = 0.77 to rs10931936) (**Figure S2, Tables S2** and **S3**). Colocalization analyses of keratinocyte cancers (both cutaneous BCC and SCC combined; PP = 0.99, maximum CLPP = 0.09), cutaneous BCC (PP = 0.98; maximum CLPP = 0.02), and SCC (PP = 0.96; maximum CLPP = 0.02) all show evidence for colocalization with melanoma risk, primarily showing the strongest evidence for melanoma lead SNP rs10931936 (SNP score is 0.42, 0.22 and 0.29, respectively for all keratinocyte cancers, BCC, and SCC) (**Tables S2** and **S3**). LocusCompare[31] plots of regional summary data for melanoma compared to keratinocyte cancers or BCC suggest that the signals for these non-melanoma skin cancers may be complex and involve multiple potential signals (**Figures S3**-**S5**). Overall, these results suggest that a risk signal tagged by rs10931936 is common to all tested cancers and nominate several potential shared causal variants.

Fine-mapping of melanoma risk-associated variants at 2q33.1

To comprehensively identify candidate causal variants within the 2q33.1 melanoma risk locus, we performed fine-mapping using multiple complementary approaches. Firstly, using GWAS summary data from Landi and colleagues[10], we identified a set of 24 variants with a log-likelihood ratio (LLR) of up to 1:1000 relative to the lead SNP at the locus (rs10931936; **Figure 1**, **Table S4**). We also performed Bayesian fine-mapping using DAP-G[32,33], identifying two clusters of potential candidate variants. The first cluster consisted of a 95% credible set of 15 variants (set 1; marked by rs10931936, PIP = 0.16; **Figure 1**, **Tables S6** and **S7**). A second cluster consisted of a set of six variants with cumulatively low posterior inclusion probabilities (PIP; set 2; led by rs77292590, PIP = 0.01; **Tables S5** and **S6**) consistent with conditional analysis of the melanoma GWAS that suggested no major second signal at this locus. Considering the possibility of a *cis*-regulatory function for this locus, we also fine-mapped using PAINTOR[34,35], weighting the priors using melanocyte-specific epigenomic annotations from the RoadMap Epigenome Project and locations of a set of genes specifically expressed in melanocytes[36] (**Figure 1**, **Table S7**), identifying a 95% credible set of seven SNPs. Of these, rs3769823, which was also fine-mapped by the other methods, had by far the highest posterior probability (0.789) of any variant. PAINTOR identified only one variant not in the DAP-G credible set 1 (rs74574949, r^2 = 0.02 with rs10931936, DAP-G credible set 3, PIP = 0.0002; **Table S5**). Lastly, we assessed whether other variants not successfully imputed in the meta-analysis, and thus not directly fine-mapped, could also be credible causal variants. Here, we chose an LD-based cutoff of r^2 = 0.625 to the lead variant, identifying an additional nine potential candidate

causal variants. Combining variants from these analyses, we identified a total set of 27 credible causal variants for the melanoma risk signal (**Figure 1**, **Table S8**).

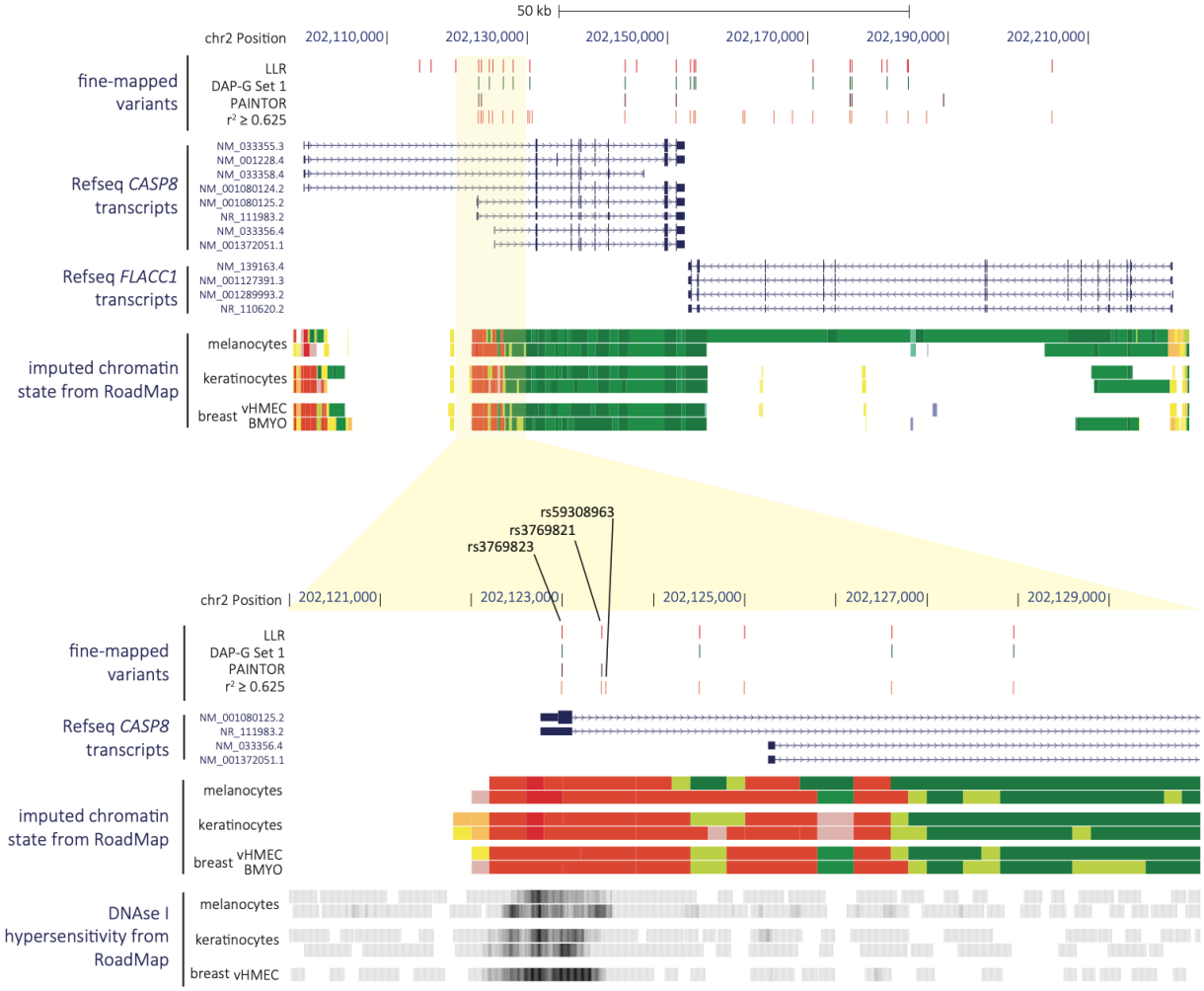

Figure 1. Fine-mapping of multiple melanoma GWAS signals on chromosome band 2q33.1. (Top) A view of the 2q33.1 locus (hg19) including all fine-mapped candidate causal variants nominated by log-likelihood ratio (LLR; red), Bayesian fine-mapping either unweighted (DAP-G, green) or weighted using melanocyte-specific epigenomic annotations (PAINTOR, brown), or LD to the melanoma lead SNP (rs10931936, $r^2 \geq 0.625$, orange). Also shown are tracks representing NCBI RefSeq genes and imputed chromatin state (ChromHMM) from two primary melanocyte cultures, two primary keratinocyte cultures, and two breast mammary cultures generated by the Roadmap Epigenome Project (Red: Active TSS; Orange-Red: Flanking Active TSS; Yellow: Enhancers; Green-Yellow: Genic enhancers; Green: Strong transcription, Dark Green: Weak transcription). (Bottom) A zoomed-in view of the same annotations for a region encompassing select candidate causal variants. Data and rendering was performed using the UCSC Genome Browser.

Potential molecular mechanisms underlying the 2q33.1 locus

Of these fine-mapped variants, only one, rs3769823, is protein-coding, being a missense variant of codon 14 (K14R) of a single *CASP8* isoform (isoform G; ENST00000358485.4; RefSeq NM_001080125.2; **Figures 1** and **S6**), where the allele encoding K14 is associated with increased risk. We assessed potential functions via Jpred4, which suggested that $CASP8^{K14R}$ may be located within a helical structure (**Figure S7**). To assess potential impacts on protein function, we used five different *in silico* prediction tools via PredictSNP2 [37]. Of these, the majority (4/5) predicted rs3769823 to be benign; FunSeq2 however suggested that this variant may be deleterious (**Table S9**). While we cannot rule out a protein-coding function for rs3769823, these data suggest rs3769823 is benign.

We also assessed splicing QTLs (sQTL) from primary human melanocytes[38]. Notably, a significant risk-associated SNP, rs10804111 ($P_{GWAS} = 6.79 \times 10^{-10}$, LD to rs10931936 $r^2 = 0.62$ and D' = 1; **Table S1**), is a significant sQTL for alternative splicing of the *CASP8* exon 8 to 9 junction ($P_{sQTL} = 7.33 \times 10^{-12}$; **Figure S8**). Specifically, this generates an alternative *CASP8* transcript (isoform H; ENST00000339403.6, RefSeq NR_111983.2; **Figures 1** and **S6**), retaining part of intron 8 and leading to frameshift and premature termination of the *CASP8* transcript, which may act as a dominant-negative inhibitor of the caspase cascade[39]. Here, rs10804111-C is associated with higher melanoma risk and lower levels of the alternative isoform H (**Figure S8B**). To test the potential of rs10804111 to mediate alternative splicing, we generated C- or T- allele-specific mini-genes (**Figure S9A,** see material and methods). Using qRT-PCR in melanoma cells, we confirmed the association between rs10804111-C and lower usage of the exon junction for isoform H (exon 8-9 junction, $P = 7.23 \times 10^{-5}$; exon 8L-9 junction, $P = 1.78 \times 10^{-3}$; **Figure S9B**). However, we note fine-mapping of the melanoma GWAS signal does not identify rs10804111 as a member of any credible causal set at this locus, and formal colocalization between the melanoma risk signal and melanocyte sQTL signal did not provide significant evidence of colocalization.

Given the lack of strong support for protein-coding or splice variants explaining the 2q33.1 signal, as well as a clustering of fine-mapped variants within regions marked as promoter or enhancer in melanocytes (**Figure 1**, **Table S8**), we turned to eQTL analysis to explore the potential of this signal being explained by *cis*-gene regulation. We first investigated previously-published cell-type specific eQTL data generated from primary cultures of human melanocytes

(n=106; **Table S10**)[36]. While multiple genes within the TAD harboring the association signal had nominally significant eQTL, *CASP8* is the only significant eQTL gene for the lead melanoma GWAS SNP (rs10931936; $P_{CASP8\ \mathrm{eQTL}} = 1.2 \times 10^{-9}$, slope = 0.67, where lower *CASP8* expression associated with the T-risk allele; **Figure 2A**, **Table S10**)[36]. Conditional analysis of melanocyte *CASP8* eQTL identified multiple variants comprising only a single eQTL signal within this locus (**Table S11**). Colocalization analysis of the *CASP8* eQTL and melanoma GWAS summary data using both HyPrColoc[29] and eCAVIAR[30] suggest *CASP8* as a strong candidate gene (HyPrColoc PP = 0.97, eCAVIAR colocalization posterior probability/CLPP = 0.09 for rs3769823) with the strongest evidence of shared causal variants for rs3769823 (HyPrColoc SNP score = 0.86, eCAVIAR CLPP = 0.09, r^2 to rs10931936 = 0.77) and rs3769821 (HyPrColoc SNP score = 0.14, eCAVIAR CLPP = 0.03, r^2 to rs10931936 = 0.63) (**Figure 3**, **Table S12** and **S13**). Collectively these results suggest a shared common causal variant between melanoma risk and melanocyte-specific *cis*-regulation of *CASP8*. While no other genes had nominally significant melanocyte eQTLs for the sentinel melanoma risk SNP rs10931936, other strongly risk-associated SNPs at this locus were, notably for the *CASP8*-adjacent gene *FLACC1* (rs3769821, $P = 2.42 \times 10^{-5}$, r^2 to rs10931936 = 0.63; lead *FLACC1* eQTL SNP rs796181752, $P = 9.01 \times 10^{-6}$; **Table S10**), where the direction of effect relative to risk allele is opposite to that for *CASP8*. Colocalization using HyPrColoc[29] and eCAVIAR[30] give discordant results, with eCAVIAR suggesting potential eQTL colocalization (CLPP = 0.04 for rs3769821; **Table S13**) while HyPrColoc does not (HyPrColoc PP < 0.5; **Table S12**). A LocusCompare[31] plot suggests *cis*-regulation of *FLACC1* in melanocytes may be complex and involve multiple potential signals potentially explaining the discordant results (**Figure S10**). Collectively, these data suggest that the melanoma GWAS signal may be explained by *cis*-regulation of *CASP8* and *FLACC1*, with opposite directions of effect for the two genes.

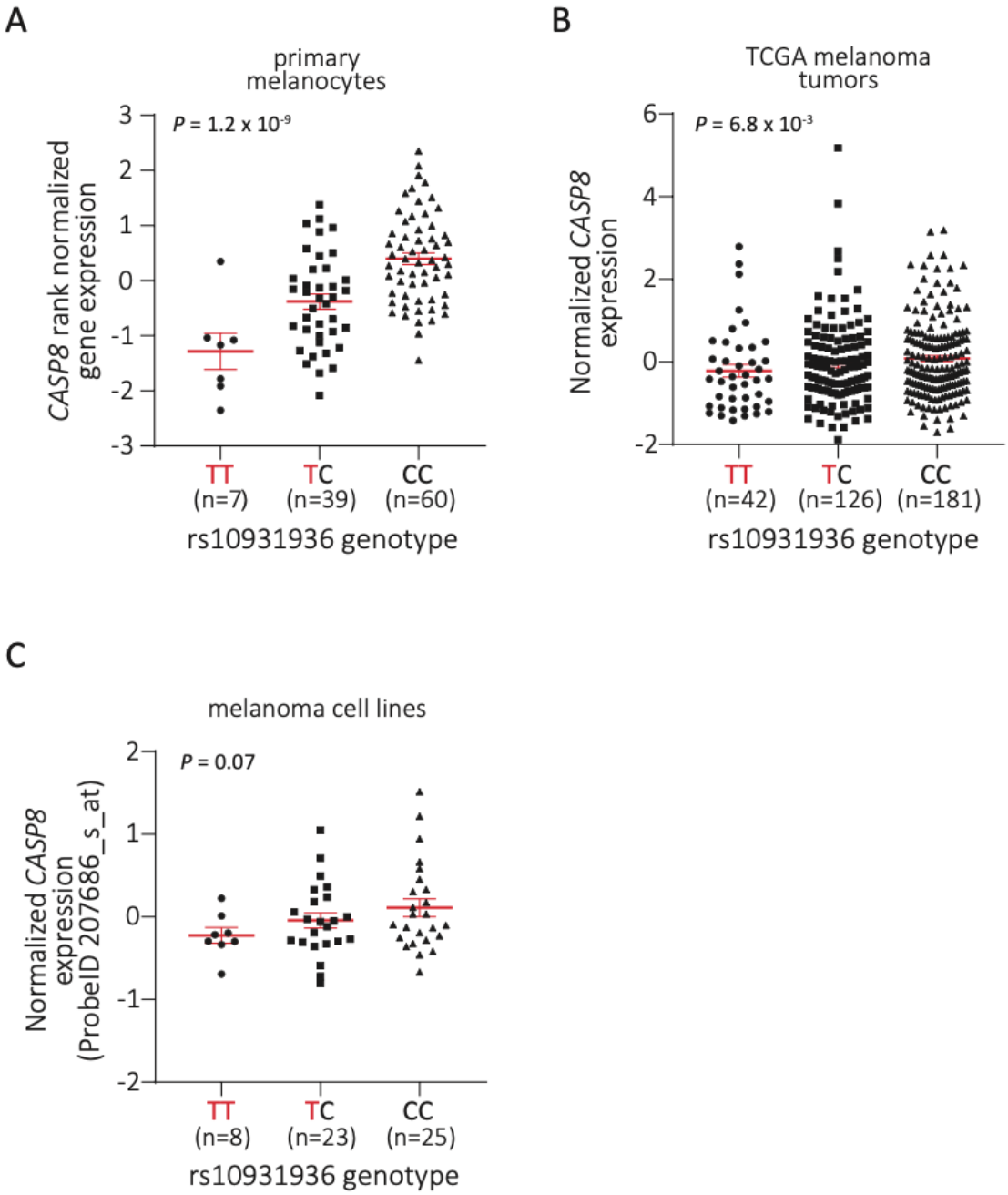

Figure 2. The melanoma risk-associated T allele of rs10931936 is correlated with lower *CASP8* expression. eQTL analyses were performed for rs10931936 combining genotype and expression level data derived from (A) 106 human primary melanocyte cultures (TT, *n*=7; TC, *n*=39; CC, *n*=60), (B) 349 melanoma tumors from TCGA-SKCM (TT, *n*=42; TC, *n*=126; CC, *n*=181), and (C) a panel of 59 early-passage melanoma cell lines (TT, *n*=8; TC, *n*=23; CC, *n*=25), where the risk-T allele is labeled in red. A significant eQTL effect with higher expression driven from the C-protective allele was observed for *CASP8* in both melanocytes and melanoma tumors ($P = 1.2 \times 10^{-9}$ and $P = 6.8 \times 10^{-3}$, respectively), while the result was marginal but in the

same direction in melanoma cell lines (P = 0.07). Significance determined by linear regression; mean with SEM are plotted along with individual data values.

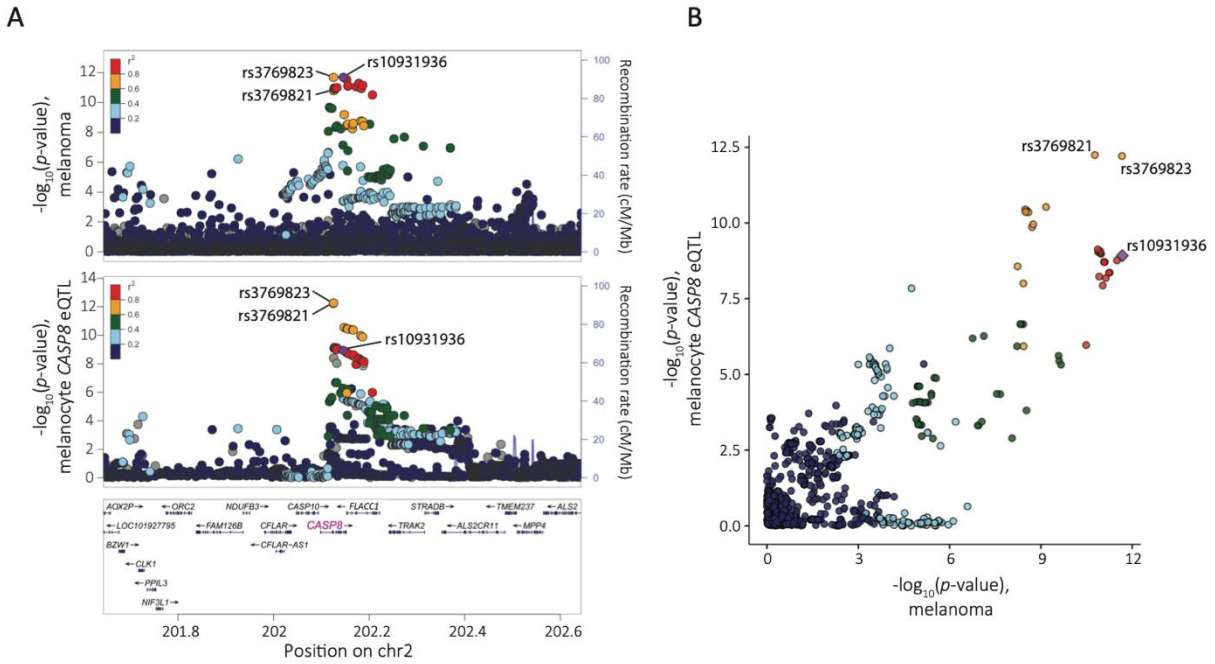

Figure 3. *CASP8* eQTL in human primary melanocyte cultures is colocalized with the melanoma GWAS. (A) LocusZoom plots present –log10 P-values for melanoma GWAS (upper) and melanocyte *CASP8* eQTL (lower) for a 1 Mb region encompassing rs10931936. The melanoma risk lead SNP rs10931936 is labeled and highlighted in purple in both panels, and LD (r^2 based on 1000G EUR) of all other SNPs to the melanoma GWAS lead SNP is color-coded. (B) A LocusCompare plot compares P-values between melanoma GWAS (x-axis) and melanocyte *CASP8* eQTL (y axis) for the same region. Genomic coordinates are based on hg19.

These data are consistent with results from a transcriptome wide association study (TWAS) study we previously performed using the same GWAS summary statistics and melanocyte eQTL data[10], where only *CASP8* is significant and *FLACC1* is marginal (P_{CASP8} = 5.57 x 10^{-12}; P_{FLACC1} = 3.83 x 10^{-5}; **Table S14**). Assessing models trained on each of 47 tissue types in the GTEx v.7; GTEx portal (https://gtexportal.org)[40] identified *CASP8* as a significant TWAS gene in 30 out of 47 GTEx tissues, strikingly with the direction of effect for six tissues being opposite to that observed for melanocytes, including multiple brain tissues and pituitary (**Table S15**)[10].

TWAS across GTEx tissues likewise found a significant positive association between imputed *FLACC1* levels and risk in 37 tissues (**Table S16**). We also assessed eQTLs in both 349 melanoma tumors from TCGA-SKCM[36,41,42] as well as 59 early-passage melanoma cell lines using expression microarray data[43]. In TCGA melanomas, we observed an FDR-significant eQTL for *FLACC1* (P = 6.57 x 10^{-4}), as well marginal eQTLs for *CASP8* (P = 6.82 x 10^{-3}; **Figure 2B**) and *TRAK2* (P = 0.03) (**Table S17**). In melanoma cultures, the rs10931936 risk-T allele is marginally associated with multiple transcripts, including *CASP8* (P = 0.07; **Figure 2C**, **Table S18**). Both *CASP8* eQTLs were in the same direction as observed in melanocytes.

Finally, given potential colocalization of melanoma risk with signals for breast cancer and cutaneous SCC, we also assessed candidate genes specifically in mammary and skin tissues profiled by GTEx v.7[40]. Colocalization analyses with the melanoma risk signal suggests melanoma risk may share common causal variants with eQTLs for *CASP8* (eCAVIAR CLPP = 0.012 for rs3769821) and *FLACC1* (eCAVIAR CLPP = 0.05 for rs3769823) in mammary tissues, and *FLACC1* in skin-not-sun-exposed (HyPrColoc PP = 0.97; SNP score = 0.62 and CLPP = 0.013 for rs3769823) and skin-sun-exposed tissues (HyPrColoc PP = 0.79; SNP score = 0.15 and CLPP = 0.012 for rs3769823) (**Tables S19** and **S20**). TWAS of melanoma risk using mammary and skin tissues found both *CASP8* and *FLACC1* to be significant TWAS genes with the same direction of effect observed in melanocytes (**Tables S15** and **S16**). These results suggest that the 2q33.1 melanoma risk locus may not only share common causal variants with breast and other cancers, but also with *cis*-regulatory variants influencing *CASP8* and *FLACC1* levels in breast and skin tissues, suggesting a potential shared *cis*-regulatory etiology for multiple cancers.

Prioritization of potential *cis*-regulatory variants using epigenomic data

Given the strong evidence for the 2q33.1 GWAS signals for melanoma as well as perhaps other cancers being explained by *cis*-regulation of *CASP8* or potentially *FLACC1*, we next prioritized credible set variants for assessment of allelic *cis*-regulatory activity. We prioritized for further study eight variants with active imputed ChromHMM chromatin states consistent with melanocyte gene regulatory regions as annotated by Regulome DB and HaploReg (**Figure 1**, **Table S8**; individual melanocyte regulatory marks are shown in **Figure S11**)[44]. Most of these eight variants also overlapped H3K27ac histone marks and or open chromatin (FAIRE-seq) from a panel 11 melanoma cultures[45] (**Figure S12**). Given the shared GWAS signal between melanoma, breast cancer, and keratinocyte cancers, we additionally prioritized fine-mapped

variants that overlap regulatory imputed ChromHMM states in primary keratinocytes and mammary epithelial tissues (**Table S8**); all such variants were already prioritized from melanocyte ChromHMM data. Thus, we proceeded to move forward with eight fine-mapped variants for functional assessment. Of these eight, three variants in close proximity (rs3769823, rs3769821, rs59308963, within 641 bp), are located in a single continuous region of active transcription start site (TSS)-proximal promoter state (TssA, **Table S8**)[44] in melanocytes, keratinocytes, and mammary epithelial cells, overlapping the first exon and first intron of two annotated *CASP8* isoforms, isoform G and H (**Figures S11** and **S12**), potentially representing a *cis*-regulatory haplotype.

Identification of functional *cis*-regulatory fine-mapped variants

To identify potentially functional *cis*-regulatory variants from these eight candidates, we performed individual reporter assays for each in the context of ~140 bp of DNA sequence surrounding the variant cloned in both forward and reverse orientations and tested in two different melanoma cell lines (summarized in **Table S21**). Only one variant, rs3769823, showed a significant allelic difference in both forward and reverse orientations in both cell lines (**Figures 4A** and **S13**, **Table S21**), albeit with the opposite direction of effect relative to the melanocyte *CASP8* eQTL. rs3769821 also had a significant albeit weaker allelic effect in the reverse direction in both cell lines with the direction of effect matching the *CASP8* eQTL, while in one cell line in the forward direction there we noted a reproducible allelic difference in the opposite direction (**Figure S14**, **Table S21**). Lastly, two other variants, rs59308963 and rs1861270 (**Table S21**), showed significant but weak allelic difference across cell lines only in the reverse orientation (**Figure S15**, **Table S21**), with the effect direction for only rs59308963 matching the *CASP8* eQTL. We also examined two previously published melanoma MPRA studies which assayed five of these variants in human melanoma cells as well as HEK293T kidney cells (MPRA v.1[46]; **Figure S16**), or six of these variants in both melanoma cells and human primary melanocytes (MPRA v.2)[47] with both studies testing a similar insert size (**Table S21**). Of the eight variants tested in either study, only rs3769823 showed an FDR-significant (FDR < 0.01) allelic effect in melanoma cells ($P_{\text{MPRA v.1}} = 1.86 \times 10^{-12}$, $P_{\text{MPRA v.2}} = 1.09 \times 10^{-52}$; **Table S21**), where the direction of effect was consistent with data from individual reporter assays and where the allelic effect is opposite to the *CASP8* eQTL. Likewise, in the second MPRA study, of the six variants tested, only rs3769823 was FDR-significant in primary melanocytes ($P_{\text{MPRA v.2}} = 3.96 \times 10^{-9}$), albeit with a smaller effect size (beta = -0.11) as compared to that observed in melanoma cells ($\text{beta}_{\text{MPRA v.1}} = -0.65$; $\text{beta}_{\text{MPRA v.2}} = -0.45$) (**Table S21**). Of the remaining variants,

rs3769821 was significant only in a joint analysis of data from both melanoma and HEK293T kidney cells ($P_{\text{MPRA v.1}} = 3.11 \times 10^{-4}$; **Table S21**) with direction of effect matching that observed in individual luciferase assays (reverse orientation) and consistent with the *CASP8* eQTL. Given the potential of shared signal in this region with breast cancer, we also tested three variants, rs3769823, rs3769821, rs59308963, in multiple breast cancer cell lines replicating the significant allelic effect and direction observed in melanoma for rs3769823, but with inconsistent results for the other two variants (**Figures S13**-**S15**). Collectively, these data suggest the potential for multiple *cis*-regulatory variants, with risk-associated allelic effects in differing directions and in some cases possibly dependent on cellular context, with the strongest evidence for rs3769823.

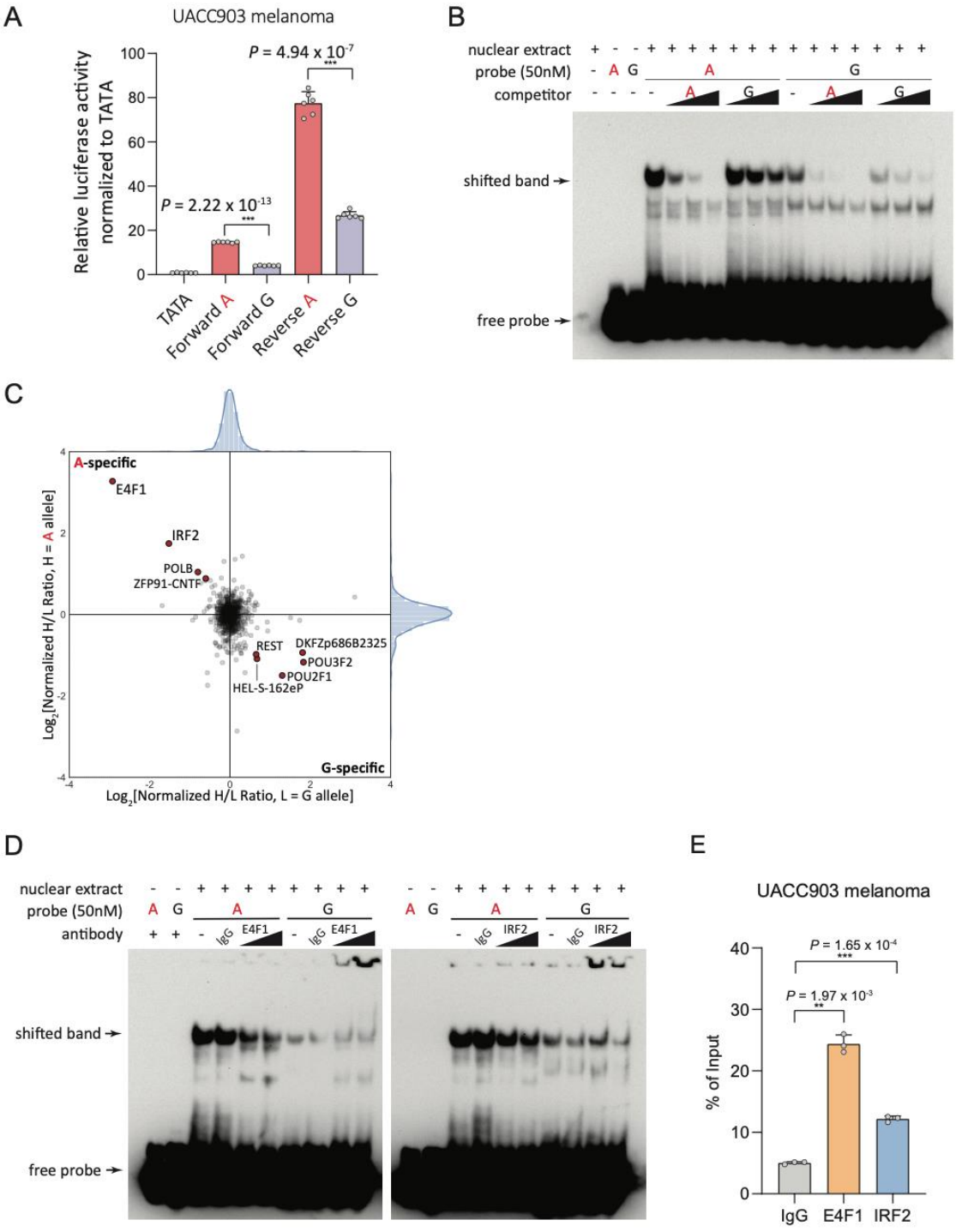

Figure 4. The melanoma-associated rs3769823 is a functional *cis*-regulatory variant and displays allele-preferential binding by E4F1 and IRF2. (A) Individual luciferase reporter activity assays for rs3769823 were conducted using the melanoma cell line UACC903. 138 bp sequences encompassing rs3769823 construct were cloned 5' of the pGL4.23 minimal TATA promoter and transfected. Luciferase activity was measured 24 h after transfection and was

normalized against *Renilla* luciferase activity. One representative set from three biological replicate experiments is shown. Mean with SEM are plotted. Individual *P* values are shown for A-risk allele versus G-protective allele for the replicate shown (two-tailed, unpaired t-test assuming unequal variances). TATA, minimal promoter control; A, risk allele construct marked in red; G, protective allele construct. (B) EMSAs were performed using 21 bp biotin-labeled double-stranded oligonucleotides for the A-risk (red) or G-protective (black) alleles of rs3769823 and nuclear extract from UACC903 melanoma cells. 2x, 5x, or 10x molar excess of unlabeled competitor was added in specified lanes. One representative set from three replicate experiments is shown. (C) Allele-specific rs3769823 binding proteins were identified by quantitative mass spectrometry using UACC903 melanoma nuclear extracts and 21 bp biotinylated double-stranded DNA probed with A-risk or G-protective alleles. The dimethyl-labeling ratios of proteins bound to A/G or G/A probes are plotted on the *x* and *y* axes for label-swapping experiments. Red-filled circles highlight proteins enriched above the background in both experiments. (D) Antibody super-shift assay of rs3769823 EMSA in UACC903 melanoma cells using anti-E4F1 or anti-IRF2 antibody is shown, where the A-risk specific shifted band (arrows) is diminished. One representative set from three independent experiments is shown. (E) ChIP was performed using antibody to anti-E4F1, anti-IRF2 or anti-IgG and chromatin prepared from UACC903 melanoma cells, followed by qPCR. DNA quantity was normalized to input DNA for each immunoprecipitation. Mean of qPCR triplicates with SEM are plotted and *P*-values shown for one representative experiment from three biological replicates.

Three of these variants are in close proximity and located within a common melanocyte enhancer element, and individual luciferase assays for all three reflect enhancer activity well above empty vector controls. We speculated that these may influence enhancer function in the context of a haplotype and thus performed reporter assays for a larger 641 bp region harboring all three variants in the context of the four most common haplotypes observed in the 1000 Genomes EUR sample set. Here, the haplotype harboring risk alleles of all three represent 27% of alleles, which that containing protective alleles for each is 64% (**Table S22**). Comparing the all-risk (A/C/ATTCTGTC for rs3769823, rs3769821, and rs59308963) to all-protective (G/T/-) haplotypes in melanocytes, we observed significantly lower reporter activity for the risk haplotype in the forward direction (cloned relative to the direction of *CASP8* transcription (**Figure S17A).** This direction matches the eQTL direction of effect for rs3769823 and *CASP8* in melanocytes, melanomas, skin, and mammary tissues (**Figure S18A-F**). In contrast, reporter assays in melanoma cells and breast cancer cells showed the opposite direction of effect, with higher reporter expression driven from the all-risk haplotype (**Figures S17B, C**). These data suggest a complex *cis*-regulatory architecture overlapping annotated melanocyte promoter for multiple *CASP8* isoforms, that allelic patterns of regulation may be highly dependent on both

adjacent functional variants as well as cellular context, and where the observed direction of effect in melanocytes *in vitro* is consistent with *CASP8* as plausible causal gene.

The genomic region containing rs3769823, rs3769821 and rs59308963 physically interacts with *FLACC1*

Given that rs3769823, rs3769821, rs59308963 are all correlated with not only *CASP8* expression but also *FLACC1* in melanocytes (rs3769823, $P = 7.14 \times 10^{-4}$, slope = -0.23, **Figure S18A**; rs3769821, $P = 2.42 \times 10^{-5}$, slope = -0.28; rs59308963, $P = 0.01$, slope = -0.19; **Table S10**), we investigated whether these or other credible causal variants physically interact with *FLACC1*. We first assessed interactions using a region-specific Capture-C assay performed in five independent human primary melanocyte cultures, where capture baits were tiled across the entire region of association [48,49] at 2q33.1. Here, we observed a direct physical contact between the restriction fragments containing these three variants and the *FLACC1* promoter (**Figure S19A**, **Table S23** and **S24**). We did not observe physical interaction between these variants and any other gene within the TAD containing the association signal. We further assessed this interaction using chromatin conformation capture (3C) in two independent melanocyte cultures as well as the UACC903 melanoma cell line, using the region harboring these three variants as bait and using primers spanning *CASP8* and *FLACC1*. Here, we confirmed a physical interaction between these variants and the *FLACC1* promoter in both melanocyte cultures, with little evidence of a strong association in melanoma cells (**Figure S19B**). These data suggest that the transcriptional regulatory region harboring rs3769823, rs3769821, and rs59308963 could potentially play a role in regulation not only of *CASP8*, but also *FLACC1*.

Multiple transcription factors bind to rs3769823 in an allele-preferential manner

Given potential *cis*-regulatory roles for rs3769823, rs3769821, rs59308963, we further assessed whether these variants influence the binding of nuclear proteins to the sequences encompassing them. We first performed electromobility shift assays (EMSAs) using nuclear extracts from melanoma cell lines, primary melanocytes, or breast cancer cell lines. We observed preferential binding of nuclear protein to the A-risk allele compared to G-protective allele of rs3769823 across all cell lines tested, where competition with unlabeled probe containing the A-risk allele is stronger than that containing the G allele, demonstrating specificity (**Figures 4B** and **S20**). For rs3769821, we observed suggestive evidence of preferential binding of nuclear protein to the T-protective allele compared to C-risk allele in melanocytes, melanoma, and breast cancer cell lines, albeit with weak specificity (**Figure S21A**). Finally, for rs59308963,

EMSA results show preferential binding to the insertion-risk allele compared to deletion allele with weak specificity in melanocytes, melanoma, and breast cancer cells (**Figure S22A**). These data suggest a strong and consistent pattern of allelic protein binding to rs3769823, consistent with the large effect observed in reporter assays, and weaker evidence for allelic binding for the other two variants.

We also performed quantitative mass spectrometry with oligonucleotides corresponding to the risk or the protective allele of each of these variants incubated with nuclear extracts from melanoma or breast cancer cell lines to directly identify allele preferential binding proteins. For rs3769823, we noted shared allele-specific binding proteins between both cell lines; specifically, IRF2 and E4F1 were observed to preferentially bind the A-risk allele, while REST and POU2F1 bound the G-protective allele (**Figures 4C** and **S23A**). The data for A-specific binding proteins are consistent with the patterns observed from the EMSA data, suggesting E4F1 or IRF2 as potential risk-allele binding transcription factors. Further, sequence-based motif prediction was also consistent with these mass spectrometry data, indicating that the sequence around rs3769823 forms a consensus binding site for both E4F1 and IRF2, favoring binding to the A-risk allele (**Figure S24**). For rs3769821, most of the potential preferential binding was found to be the T-protective allele via EMSA, however mass spectrometry found few T-specific binders and no overlapping T-binding proteins between cells; PRDM14 and ZBTB9 were both found to preferentially bind the C-allele (**Figure S21B**). We also identified BNC2 as a strong risk-insertion binding protein to the risk-associated allele of rs59308963 in melanoma cells (**Figure S22B**).

Based on the clear results and strong allelic pattern of *cis*-regulatory activity for rs3769823, we sought to verify whether E4F1 or IRF2 proteins bound the A-allele by using antibodies against these proteins in conjunction with EMSAs. Antibodies against either E4F1 or IRF2 consistently resulted in loss of A allele-specific protein binding in both melanoma and breast cancer cell lines (**Figures 4D** and **S23B**). To further establish the binding of E4F1 or IRF2 to rs3769823, we performed chromatin immunoprecipitation (ChIP) for E4F1 or IRF2 followed by quantitative PCR, noting an enrichment of binding at rs3769823 in melanoma cell lines with homozygous for rs3769823-A as well as primary melanocytes heterozygous for this variant (**Figures 4E** and **S25B**). We assessed enrichment at other locations across the *CASP8* gene that had previously been shown via ChIP-seq to be bound by IRF2 or E4F1, and DHS data from ENCODE (**Figure S25A**), however ChIP only showed enrichment of both in the region over rs3769823

(UACC903(A/A), P_{E4F1} = 1.46 x 10^{-3}, P_{IRF2} = 0.016; C87(A/G), P_{E4F1} = 3.25 x 10^{-3}, P_{IRF2} = 9.47 x 10^{-3}; **Figure S25B**). Using the heterozygous melanocyte culture, we also assessed allelic enrichment in the immunoprecipitants using quantitative real-time PCR genotyping assay and noted a significant enrichment of the A allele as compared to the G allele at rs3769823 in primary melanocytes for both transcription factors (**Figure S25C**). Together, these data clearly suggest that E4F1 and IRF2 may act through rs3769823 at this locus.

E4F1 and IRF2 as potential repressor and activator of *CASP8* transcription, respectively

Given the allelic binding observed for E4F1 and IRF2 to the A-risk allele of *cis*-regulatory SNP rs3769823, we next explored relationships between levels of these transcription factors and expression of *CASP8* and *FLACC1* in various cell and tissue types. In primary melanocytes, we observe clear negative correlation between both *E4F1* mRNA levels with *CASP8* (*E4F1*, *P* = 3.80 x 10^{-6}, Pearson r = -0.43; **Figure S26**, **Table S25**) while *IRF2* is not significant (*IRF2, P* = 0.39, r = 0.08; **Figure S26**, **Table S25**); multiple linear regression shows both rs3769823 and *E4F1* to be significant (*E4F1 P* = 1.70 x 10^{-15}; rs3769823 *P* = 5.39 x 10^{-7}; *IRF2 P* = 0.43; **Table S26**). These data are consistent with a known role for E4F1 as a transcriptional repressor[50] and suggests a role in regulation of *CASP8*. This direction of effect is consistent with transcriptional activity observed in reporter assays in melanocytes potentially being attributable in part of a repressive effect for E4F1 when considering the longer three-variant risk haplotype, but not the short (138 bp) fragment assays considering rs3769823 alone. In contrast, in TCGA melanoma tumors, while significant negative correlation with *E4F1* was observed (*E4F1*, *P* = 2.26 x 10^{-4}, r = -0.20; **Figure S26**, **Table S25**), a much stronger positive with *CASP8* was observed for *IRF2* (*IRF2*, *P* = 1.69 x 10^{-27}, r = 0.54; **Figure S26, Table S25**); both transcription factors and rs3769823 are significant via multiple linear regression (*E4F1 P* = 0.0018; *IRF2 P* = 2 x 10^{-16}; rs3769823 *P* = 0.035; **Table S27**), again with the strongest effect for *IRF2.* We did not find significant evidence for an interaction between rs3769823 and either of the transcription factors (**Table S27**). We were able to replicate this positive correlation specifically between *IRF2* and *CASP8* in independent panels of melanoma cell lines (*P* = 4.23 x 10^{-4}, r = 0.38; **Table S25**) and the Leeds Melanoma Cohort (*P* = 8.74 x 10^{-4}, r = 0.30; **Table S25**). This direction of effect is consistent with a potential role for *IRF2* as a potential activator in both individual and haplotype-based reporter assays, but inconsistent with the direction of the relatively weak eQTL for CASP8 in melanoma tumors. We formally tested for potential interactions between *E4F1* and rs3769823 in melanocyte expression data as well as between *E4F1* and/or *IRF2* and rs3769823

in TCGA melanoma tumors but did not find significant supporting evidence (**Tables S26** and **S27**).

We also assessed transcripts encoding two T-protective allele binding proteins for correlations with potential target genes, observing a positive correlation between the transcriptional repressor *REST,* as well as *POU2F1,* and *CASP8* in melanocytes (*REST*, $P = 7.21 \times 10^{-11}$; r = 0.58; *POU2F1*, $P = 8.40 \times 10^{-3}$, r = 0.25), a weak negative correlation for *POU2F1* in TCGA melanomas ($P = 0.04$, r = -0.11), and a positive correlation for *REST* in the Leeds Melanoma Cohort ($P = 1.75 \times 10^{-14}$, r = 0.30) (**Table S25**). This positive correlation with *CASP8* in melanocytes and some melanoma tumors is inconsistent with the established role for REST as a transcriptional repressor. Still, this direction of effect is consistent with a potential role for REST repressing reporter activity in both individual (short fragment) reporter assays for rs3769823 as well as both previously-reported MPRA assays. Multiple linear regression considering all three individually correlated risk-A and protective-G binding alleles and genotype for rs3769823 in melanocytes suggests all four are significant (*REST* P = 6.71×10^{-7}; rs3769823 $P = 4.29 \times 10^{-6}$; *POU2F1* $P = 0.0019$; *E4F1* $P = 0.010$; **Table S28**). We note similarly complex patterns of correlation between both risk- and protective- allele rs3769823 binding proteins in non-melanocytic tissues (GTEx skin, GTEx breast mammary, and TCGA breast invasive carcinoma; **Table S25**) and expression of *CASP8* as well as *FLACC1* (**Table S25**). These data are consistent with the activity of the regulatory region encompassing rs3769823 being modulated by a complex interplay of multiple allele-specific transcriptional regulators interacting with risk and protective alleles in a tissue-specific manner.

Finally, we also investigated correlations between *CASP8* and allele-specific binding proteins for rs3769821 (**Table S29**) and rs59308963 (**Table S30**), observing negative correlations with the transcripts for rs3769821 risk-allele binding protein ZBTB9 and the rs59308963 risk-allele binding protein BNC2 (*ZBTB9*, TCGA $P = 2.15 \times 10^{-11}$, TCGA r = -0.35, Leeds $P = 3.40 \times 10^{-3}$, Leeds r = -0.12; *BNC2*, TCGA $P = 2.70 \times 10^{-4}$, TCGA r = -0.19, Leeds $P = 1.43 \times 10^{-13}$, Leeds r = -0.29; **Tables S29** and **S30**). For *FLACC1*, we note similar correlations in melanocytes to those observed for *CASP8*, with negative correlation observed for *E4F1* ($P = 3.40 \times 10^{-6}$, r = -0.43), and positive correlations observed for *REST* and *POU2F1* (*REST*, $P = 1.72 \times 10^{-10}$, r = 0.57; *POU2F1*, $P = 4.71 \times 10^{-12}$, r = 0.61), but do not observe consistent correlations for *IRF2* or *REST* across melanoma datasets (**Table S25**). Similar to what is observed for rs3769823, these correlations vary considerably between melanocytic and non-melanocytic tissues (**Tables S29**

and **S30**) observing similarly complex patterns of correlation. Collectively, these data suggest a potentially complex gene regulatory interplay, prominently of multiple allele-specific rs3769823 binding proteins, as well as other transcriptional regulators, some of which may bind this region differentially in risk and protective haplotypes.

To determine the effects of depleting the two rs3769823 A-risk allele-specific binding proteins on *CASP8* transcription in melanoma, primary melanocyte cultures, and breast cancer cell lines, we knocked these genes down using siRNAs. Knockdown of *E4F1* resulted in an increase in *CASP8* levels while *IRF2* knockdown results in reduction of *CASP8* RNA levels in melanoma (**Figure 5A** and **5B**, **S27A** and **S27B**), breast cancer (**Figure S28A** and **S28B**), and melanocytes (**Figure S29A**), consistent with respective roles for E4F1 and IRF2 as a potential repressor and activator in cells of melanocytic lineage. Short fragment (138 bp) reporter assays for rs3769823 in conjunction with knockdown of A-risk allele binding protein IRF2 resulted in a significant decrease of expression driven primarily from the A-risk allele compared to the protective G-allele in both forward and reverse directions in melanoma (UACC903 cells, $P_{forward} = 1.22 \times 10^{-7}$, $P_{reverse} = 2.84 \times 10^{-8}$, **Figure 5C**; UACC1113, $P_{forward} = 1.65 \times 10^{-5}$, $P_{reverse} = 2.20 \times 10^{-3}$, **Figure S27C**) and breast cancer cell lines (T47D, $P_{forward} = 2.92 \times 10^{-7}$, $P_{reverse} = 3.18 \times 10^{-7}$, **Figure S28C**), consistent with a potentially prominent role for IRF2 in *CASP8* regulation in melanoma cells. Consistent with the weaker correlations between *E4F1* and *CASP8* in melanoma cells, *E4F1* knockdown only modestly increased reporter activity for the A-risk allele, and only in the forward direction, suggesting a potentially weak repressor role for E4F1 in the context of melanoma cells (UACC903, $P_{forward} = 8.05 \times 10^{-4}$, $P_{reverse} = 0.74$, **Figure 5C**; UACC1113, $P_{forward} = 3.47 \times 10^{-3}$, $P_{reverse} = 0.02$, **Figure S27C**). In primary melanocytes, reporter assays for rs3769823 with knockdown of A-risk-allele binding E4F1 resulted in a significant increase of promoter activity in the A-risk allele compared to the protective G-allele in both direction (C23 melanocytes, $P_{forward} = 1.27 \times 10^{-9}$, $P_{reverse} = 0.013$, **Figure S29B**), suggesting a potentially repressor role for E4F1 in regulation of *CASP8* in melanocytes.

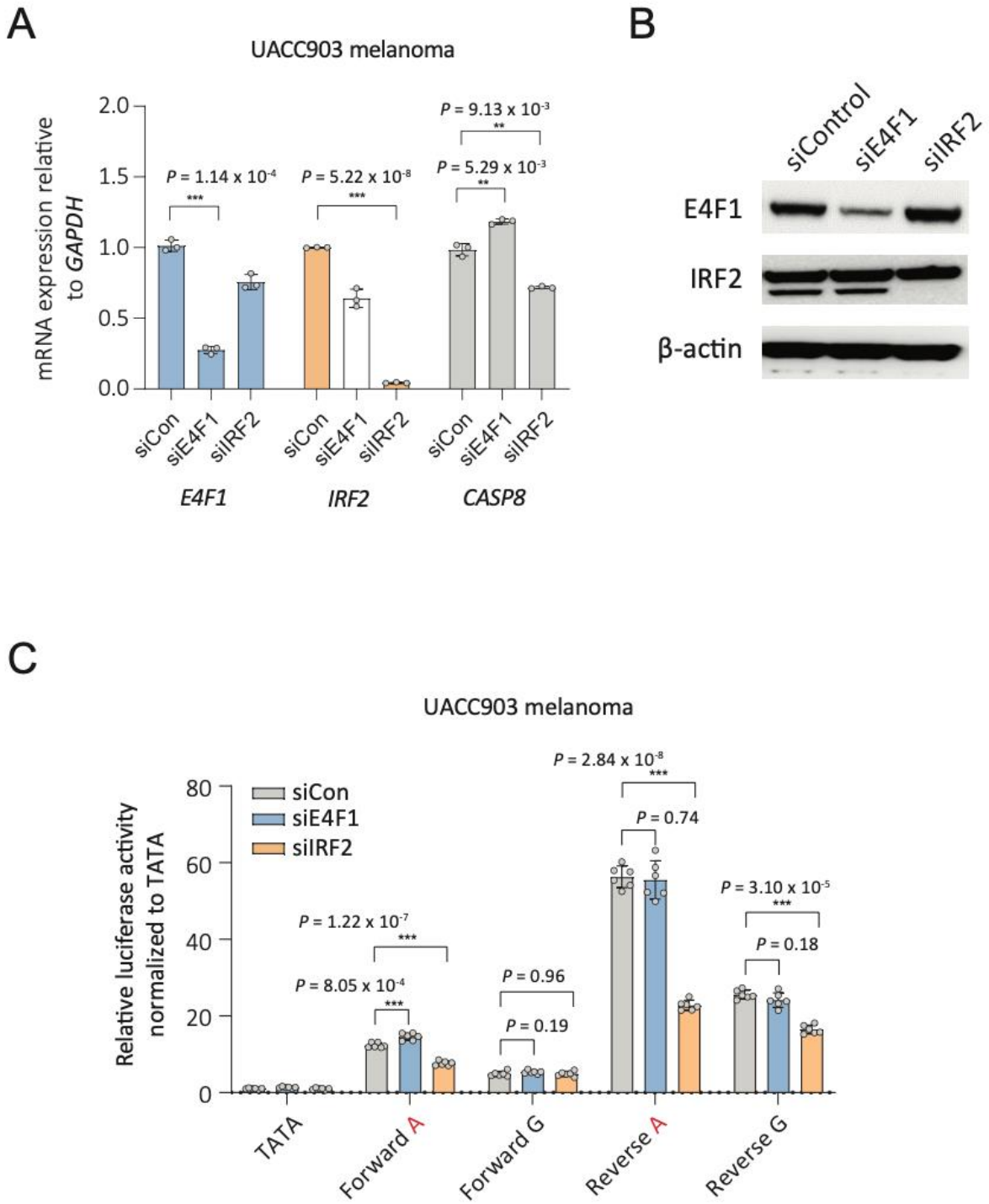

Figure 5. *E4F1* or *IRF2* and rs3769823 regulate *CASP8* expression in UACC903 melanoma cells. (A) *E4F1* or *IRF2* were knocked down using respective pools of four different siRNAs in UACC903 melanoma cells, and *E4F1*, *IRF2* and *CASP8* levels were measured. *GAPDH-*

normalized *E4F1*, *IRF2* or *CASP8* mRNA levels are shown as fold change over those from non-targeting siRNA. A representative experiment from three biological replicates is shown (individual datapoints, mean, and SEM are plotted). *P*-values are shown from one representative set. (B) Protein levels were examined using anti-E4F1, anti-IRF2, or anti-GAPDH antibody with UACC903 cell lysates of siRNAs transfected with either non-targeting control, *E4F1*, or *IRF2* siRNAs. GAPDH was used as a loading control. One representative set of three replicate experiments is shown. (C) Individual luciferase assays were performed by co-transfecting rs3769823 luciferase constructs with *E4F1*, *IRF2,* or non-targeting control siRNAs into UACC903 melanoma cells. *Renilla*-normalized relative luciferase activities are shown relative to an empty construct containing only a minimal promoter (TATA). One representative set is shown from three biological replicates. Mean with SEM is plotted, n = 6 technical replicates. A two-tailed t-test assuming unequal variances was used to calculate all P values shown against control siRNA.

Discussion

In this study, we focused on a melanoma susceptibility locus over the *CASP8* gene on 2q33.1 that was originally identified by Barrett and colleagues[5] and has since been replicated in a large melanoma GWAS meta-analysis (rs10931936-T, $P_{fixed} = 2.12 \times 10^{-12}$, $P_{random} = 2.17 \times 10^{-8}$, OR = 1.08)[10]. In order to comprehensively identify additional independent signals as well as nominate one or more sets of credible causal variants associated with such signals, we used multiple complementary approaches. Conditional and joint analyses previously found no evidence for a second genome-wide significant signal at 2q33.1[10]; here, we observe no evidence for a prominent independent melanoma risk signal of marginal significance. Fine-mapping using Bayesian approaches similarly identified one prominent signal at this locus with little evidence for additional signals.

Germline variants is this region have been identified as associated with multiple cancers, including breast cancer (rs3769821) [11], keratinocyte cancers (cutaneous BCC and SCC combined, rs6743068) [13], cutaneous BCC (rs6714430) [13], cutaneous SCC (rs10931936) [13], and non-small cell lung cancer (rs3769821) [51] where the direction of effect is the same and the risk alleles share a common haplotype with the melanoma risk allele of rs10931936. In addition, a highly correlated risk signal was previously identified for prostate cancer (rs59308963) [52] where the direction of effect is opposite to that of melanoma and the other cancers. This locus was also identified via GWAS for chronic lymphocytic leukemia (CLL; rs3769825) [53], however this signal appears to be considerably less correlated with the melanoma risk signal (r^2 with

rs10931936 = 0.26). For several of these cancers with summary data readily available, we evaluated evidence for colocalization, with the results suggesting sharing of one or more causal variants between melanoma and signals for breast cancer, cutaneous BCC, and cutaneous SCC and further suggesting that our work to fine-map identify functional, potentially causal variants for melanoma risk may be relevant for these other cancers.

Using multiple complementary fine-mapping approaches, we identified a set of 27 potential causal variants. Collectively, these variants present multiple potential alternative or complementary functional mechanisms for influencing gene function and cancer risk. Notably, rs3769823 is a missense protein-coding variant in codon 14 of an alternative first exon of *CASP8* utilized by multiple isoforms and expressed in human primary melanocytes (isoforms G and H; **Figure S6**). However, a majority of algorithms designed to predict variant function suggest this variant is likely benign and K14R is not located within conserved domains or known or predicted sites of post-translational modification. Despite this, we did not directly test for an effect of the two alleles of this variant on CASP8 protein function and cannot rule out a role for altered protein sequence as a mechanism underlying risk at this locus. Zhang and colleagues studied the functionality of rs3769823 by expressing the two alleles in lung cancer cell lines and noted differences in proliferation and cell migration *in vitro* and *in vivo* [54]. Notably, expression from the K14R (protective) allele generated more CASP8 protein, which the authors posited may be due to altered protein stability. Downstream phenotypic assays did not distinguish between the effects of protein abundance and protein sequence-specific effects, and potential transcriptional differences for the two constructs were not controlled for, so a protein-coding function for this variant on CASP8 remains possible but still unclear. Further work is needed to assess potential effects of this protein-coding change on CASP8 protein function.

We also observed a significant *CASP8* sQTL in melanocyte and other GTEx eQTL datasets relevant to cancers with associations for this locus (marked by rs10804111; r^2 to rs10931936 = 0.62; D’ = 1) reflecting alternative splicing events at the exon 8 to exon 9 junction. The alternative splicing event is more strongly associated with the protective allele generates an alternative *CASP8* transcript, isoform H. This transcript retains only part of canonical intron 8 and introduces a frameshift in the transcript, leading to premature termination. The resultant protein retains the *CASP8* death effector domain but lacks the catalytic domain, and notably, Himeji and colleagues demonstrated that this protein could inhibit apoptosis mediated by wild-type CASP8 in a dominant-negative manner[39]. Thus, the transcript preferentially driven by the

protective allele could potentially lead to loss of *CASP8* function; this stands in functional contrast to the melanocyte eQTL observed where the protective allele of rs10931936 is associated with higher rather than lower levels of *CASP8* transcript. Nonetheless, while LD of rs10804111 with rs10931936 is comparable to that of other fine-mapped credible causal variants for melanoma at this locus, none of the fine-mapping approaches we applied nominates rs10804111 as a credible causal variant itself. Still, much as for the protein-coding change caused by rs3769823, we cannot rule out a role for altered *CASP8* splicing contributing to melanoma risk.

In contrast, eQTL and TWAS results in human primary melanocytes[10,36] provide strong evidence suggesting *cis*-gene regulation as a likely mechanism underlying melanoma risk attributable to this locus, where melanoma risk-associated alleles were significantly correlated with lower levels of *CASP8*. Both TWAS and colocalization approaches suggest that melanoma risk and melanocyte eQTL for *CASP8* likely share one or more common causal variant, with colocalization strongly nominating rs3769823. Both also provide some support for *FLACC1*, where in contrast to the *CASP8* eQTL, higher *FLACC1* levels are associated with the melanoma risk allele. Several fine-mapped melanoma risk variants at this locus are clearly located within melanocyte gene-regulatory regions, including regions annotated as active promoters, consistent with a potential role for this signal regulating *CASP8* levels. Consistent with observed *FLACC1* QTLs, melanocyte-specific capture-C and 3C assays showed a physical interaction between fine-mapped melanoma risk variants and the *FLACC1* promoter region (**Figure S19**) suggesting risk variants indeed may play a role in regulating both genes. Notably, none of these analyses suggest a prominent role for other nearby genes, including the apoptosis regulatory genes *CASP10* and *CFLAR*.

Based on these data, we comprehensively assessed fine-mapped risk variants for potential *cis*-regulatory function, first identifying eight variants overlapping with *cis*-regulatory histone marks in disease relevant tissue (melanocytes, keratinocytes, human mammary epithelial cells, and breast myoepithelial primary cells). We assessed each for allelic *cis*-regulatory potential both in two previously published massively parallel reporter assays performed in melanocytes and/or melanoma cells, identifying rs3769823 as the only significant variant, with allelic effects in both melanoma cells and to a lesser degree melanocytes. We likewise tested each in individual luciferase reporter assays in two melanoma cell lines, again identifying rs3769823 as the only variant with consistently significant allelic *cis*-regulatory activities in both cell lines and in both

forward and reverse orientations where the direction of reporter effect was consistent with the melanocyte *FLACC1* eQTL but opposite that of *CASP8* (e.g., the risk allele is associated with higher reporter expression). Nonetheless, in individual assays, several variants displayed allelic activity in one orientation, including two additional variants rs3769821 and rs59308963, located nearby rs3769823 in the same annotated gene promoter region, suggesting potential functional contributions from more than one variant. Testing commonly occurring rs3769823/rs3769821/rs59308963 haplotypes in a reporter assay revealed cell-type dependent effects; the risk-associated A/C/ATTCTGTC haplotype was associated with lower expression relative to the protective G/T/- haplotype in melanocytes but higher expression in melanoma cells in the forward orientation relative to *CASP8* transcription (**Figure S17A, B**). This stands in contrast to the results from Camp and colleagues, who tested risk and protective haplotypes of rs3769823/rs3769821 and observed lower reporter gene expression driven from the risk haplotype in both normal breast and breast cancer cell lines [55]. The differing direction of effect observed in melanocytes for rs3769823 in short-fragment versus longer-fragment haplotype analysis suggests that while rs3769823 is strongly functional, the allelic effects are influenced by other nearby and potentially functional variants.

All three variants showed allelic patterns of protein binding via EMSA. Previously, in the context of lung cancer susceptibility, Long and colleagues observed a *cis*-regulatory element shared across many lung cell types overlapping rs3769823 and noted that this variant is predicted to alter IRF8 DNA binding [56]. Here, we applied a quantitative mass spectrometry workflow and identified multiple allele-specific nuclear binding proteins from both melanoma and breast cancer cells that bound these variants in an allele specific manner. In particular, we identified proteins that typically act either as a transcriptional repressor (E4F1) or activator (IRF2) as both preferentially binding the A-risk allele of rs3769823, and verified binding of both factors via ChIP in melanoma cells and primary melanocytes. While we did not observe significantly different IRF8 binding in this context, the DNA-binding motifs of IRF2 and IRF8 are quite similar. We observed a similar situation for G-protective allele specific binding proteins, where both an activator (POU2F1) and repressor (REST) were found to bind to this allele in an allele preferential manner. Thus, allelic transcriptional activity may be highly dependent on relative availability of these factors as well as competition for binding under different cellular contexts. Indeed, gene expression correlation analyses suggest, for example, that E4F1 may play a prominent role for *CASP8* regulation in melanocytes, while the activator IRF2 may play a larger role in regulation of *CASP8* in melanomas. Indeed, reporter assay data for rs3769823 in

conjunction with *E4F1* or *IRF2* knockdown are consistent with this, where in melanoma cells knockdown of IRF2 resulted in a significant decrease of expression driven from the A-risk allele, while knockdown of E4F1 resulted in a weaker increase of reporter activity driven from the A-allele. Given the similarity of DNA-binding motifs for IRF proteins, it may be possible that in addition, tissue-specific availability of other IRF family members could also drive or repress gene expression via rs3769823 and possibly explain the pleiotropy at this signal for prostate cancer where the direction of effect is opposite that of melanoma and the other cancers [15]. In contrast knockdown of E4F1 in melanocytes resulted in a significant increase of activity driven from the A-risk allele. We further observed additional allele specific binding proteins for potentially functional adjacent variants rs3769821 and rs59308963, which could well further modulate transcriptional output themselves as well as potentially contribute to preference of transcription factor binding to rs3769823.

Our work provides strong evidence for two potential causal genes at this locus, *CASP8* and *FLACC1*. Of these, *CASP8* represents a very strong *a priori* candidate cancer risk gene, given its long-established role as an apoptotic initiator caspase and the fact that lower *CASP8* levels are associated with cancer risk. Little is known about the function of *FLACC1* (formerly *ALS2CR12*). It has been found to be component of sperm tail flagellum [57], and functional studies of *ALS2CR12*-like genes in *C. elegans* suggest that similar genes play a role in gene and or repetitive DNA silencing, a function appearing to be dependent on the biogenesis of small RNAs [58,59]. Notably this region harbors multiple genes involved in apoptosis, including *CASP10*, encoding another apoptotic initiator caspase [60-62] sharing similar substrate specificity to CASP8 [63] but possibly acting as a negative regulator of CASP8-mediated cell death [64]. Also located within this region is *CFLAR* (also known as c-FLIP), encoding a protein structurally-similar to CASP8 but lacking caspase activity and acting to prevent CASP8 dimerization and inhibiting apoptosis [19,65-69]. Despite the interacting roles of these genes, we observe little evidence that the risk signal drives differences in expression of these antiapoptotic genes. Nonetheless, we cannot rule out potential regulation of these genes from this signal under specific contexts. We do note, however, that at least in human melanocytes we observe no evidence of physical interaction between fine-mapped cancer risk variants (**Figure S19**) and the promoter regions of either gene.

In summary, our study functionally interrogated the multi-cancer risk locus on chromosome band 2q33.1, providing strong evidence for *cis*-regulation of *CASP8* via multiple functional variants as a primary causal mechanism influencing cancer risk.

Material and methods

Melanoma GWAS summary data

For all analyses using melanoma GWAS data, we used summary statistics from both histologically confirmed, as well as self-reported melanoma cases (from 23andMe and UK Biobank), and controls as previously described[10]. In total, summary data are derived from 36,760 melanoma cases and 375,188 controls; all participants provided informed consent and participation was IRB approved. Participants from 23andMe provided informed consent and participated in the research online under a protocol approved by the external AAHRPP-accredited IRB, Ethical and Independent Review Services (E and I Review).

Conditional and joint analysis of melanoma GWAS

Conditional and joint analysis was performed with summary melanoma GWAS meta-analysis data by using Genome-wide Complex Trait Analysis (GCTA, v.1.26.0)[28] to identify independent associated variants. We set the collinearity threshold (-cojo-collinear) to $R^2 = 0.05$ to detect only completely independent SNPs. Linkage disequilibrium (LD) between SNPs was calculated using 5,000 randomly selected individuals of European ancestry from UK Biobank.

Colocalization of the 2q33.1 melanoma association signal with other GWAS and QTL data

For colocalization analyses, we used GWAS summary data from breast cancer (122,977 cases and 105,974 controls; dbGaP accession number phs001265.v1.p1; downloaded from https://bcac.ccge.medschl.cam.ac.uk/bcacdata/oncoarray/oncoarray-and-combined-summary-result/gwas-summary-results-breast-cancer-risk-2017/)[11] and keratinocyte cancers (47,742 cases and 634,414 controls; dbGaP accession number phs000360.v3.p1)[13], including cutaneous BCC (31,787 cases and 619,351 controls)[13] and cutaneous squamous cell carcinoma (SCC; 9,674 cases and 625,657 controls)[13]. Colocalization analyses for the 2q33.1 locus were performed using both Hypothesis Prioritization in multi-trait Colocalization (HyPrColoc)[29] and eQTL and GWAS Causal Variant Identification in Associated Regions (eCAVIAR)[30] as implemented in the ezQTL web-based tool

(https://analysistools.cancer.gov/ezqtl/)[70]. For both, we used the set of SNPs in +/- 100 kb window size from the melanoma GWAS lead SNP rs10931936[36,38]. For HyPrColoc, a posterior probability (PP) above 50% (0.5) was considered to display a positive colocalization and SNP score above 5% (0.05) was considered to share a common causal variant. For eCAVIAR, the colocalization posterior probability (CLPP) score was calculated with a maximum number of two causal SNPs in the locus; a CLPP score above 1% (0.01) was considered to display a positive colocalization.

GWAS fine-mapping

Fine-mapping was performed using several approaches. Firstly, in order to inclusively nominate candidate causal variants for assessment, we used the log-likelihood ratio (LLR) of each SNP in the region relative to lead variant rs10931936, and considered all variants within a ratio of >1:1000. Variants not successfully imputed from the GWAS that thus could not be fine-mapped in this manner nor via other methods requiring summary data, hence we used an LD-based threshold. Specifically, we chose this threshold to be consistent with the observed LD (r^2) between LLR-mapped credible set variants with, where the minimum LD from this set with rs10931936 was $r^2 = 0.625$ (1000 Genomes Projects Phase 3 EUR data; 1000G EUR). We thus considered variants not present in the summary dataset where 1000G EUR $r^2 \geq 0.625$. We also performed Bayesian fine-mapping of melanoma GWAS summary data[10] for the 2q33.1 risk region using both Deterministic Approximation of Posteriors (DAP-G)[32,33] and Probabilistic Annotation INtegraTOR (PAINTOR)[34,35]. DAP-G (version 1.0) analysis used a 500 Kb window size centered over the association signal (median position of chr2:202,123,966) and allowing for a maximum number of five causals. A UK Biobank (UKBB) LD reference panel (from 337K UKBB participants, downloaded from https://alkesgroup.broadinstitute.org/UKBB_LD/)[71] was aligned with the summary statistics from the melanoma GWAS meta-analysis[10] using R (version 4.1.0) for input into DAP-G. PAINTOR analyses were performed as described previously[46] using PAINTOR 3.0 with a window size of 100 Kb, maximum causals set to 4, 4 melanocyte-specific epigenomic annotations and pairwise LD derived from 1000 Genomes phase 3 EUR (1000G EUR) data computed using PLINK version 1.9 and R version 4.1.0. Functional annotations include a set of 2,000 melanocyte-specific expressed genes from our melanocyte dataset[36], melanocyte enhancers, transcribed regions, and a histone mark (H3T11ph) from ENCODE and Roadmap.

Prediction of variant effect on protein function and structure

We predicted the functional impact of missense variant rs3769823 on function of CASP8 protein through the PredictSNP2 [37] web-based tool (https://loschmidt.chemi.muni.cz/predictsnp2/), which integrates five functional prediction methods including CADD (Combined Annotation Dependent Depletion) [72], DANN (Deleterious Annotation of genetic variants using Neural Network) [73], FATHMM (Functional Analysis Through Hidden Markov Model) [74], FunSeq2 [75], and GWAVA (Genome-Wide Annotation of VAriants)[76]. We also predicted potential effects of CASP8^{K14R} on secondary structure of the protein through JPred4 [77] web server (https://www.compbio.dundee.ac.uk/jpred4/), which also predicts solvent accessibility and coiled-coil regions. Predictions from AlphaMissense[78] for rs3769823 are not available.

Quantitative trait locus (QTL) data and other expression data

QTL data from 106 primary human melanocyte cultures primarily of European ancestry were generated as described previously for both gene expression (eQTL)[36] as well as for splicing (sQTL) and CpG methylation (meQTL)[38] (dbGaP accession: phs001500.v1.p1). Conditional analysis of the *CASP8* melanocyte eQTL was performed using QTLTools (https://qtltools.github.io/qtltools/)[79], combining genotype and expression level data, along with a matrix of covariates including the top three genotype principal components and the top 15 PEER factors. For other tissues, we used pre-analyzed *cis*-eQTL data from the Genotype-Tissue Expression (GTEx) project (v.7), which were downloaded from the GTEx portal (https://gtexportal.org/home/index.html)[40]. eQTL analysis was performed previously[36] using relative linear copy-number values as a covariate for 349 melanoma tumors from The Cancer Genome Atlas Project (TCGA) skin cutaneous melanoma (SKCM) dataset[41,42]. eQTL data for 59 early passage melanoma cell lines was performed as described previously[43] using expression and genotyping data from NCBI GEO (accessions GSE78995 for expression data; GSE99193 for genotyping data). We specifically assessed genes within the TAD harboring the lead SNP at 2q33.1 (rs10931936; chr2:201,760,000-203,080,000; hg19) based on Hi-C results from SKMEL5 melanoma cells as visualized by the 3D Genome Browser (http://3dgenome.fsm.northwestern.edu/view.php).

***CASP8* isoform analysis**

To assess splicing QTLs (sQTLs), we used the same genotype data, population structure covariates, and statistical approaches as used for melanocyte eQTL analysis[36], replacing normalized gene expression levels with normalized splice junction events. STAR[80] was used to map the RNA-Seq reads onto the genome (hg19) and then LeafCutter[81] was applied to quantify

the splice junctions following the procedures described by the authors (http://davidaknowles.github.io/leafcutter/articles/sQTL.html). Taqman real-time PCR assays targeting unique junctions of *CASP8* transcript isoforms were obtained from Thermo Fisher Scientific (Waltham, MA; all isoform transcripts: Hs01018151_m1; exon 8 to 9 junction: Hs01018160_m1; exon 8L (alternative spliced) to 9 junction: Hs04405665_m1). RNA was isolated from primary melanocyte cultures derived from 106 individuals mainly of European decent[36], and cDNA was synthesized using iScript™ Advanced cDNA Synthesis Kit (Bio-Rad, Hercules, CA). Taqman assays were performed in triplicates (technical replicates) and normalized to *TBP* or *PPIA* levels.

Exon-trap analysis of alternatively spliced exons of *CASP8* genes

rs10804111 is an intronic variant; no variants strongly linked to the *CASP8* melanocyte sQTL were located in consensus splice or branch point sequences, and none could create a cryptic splice site responsible for the observed allelic splicing pattern. We therefore assessed potential effects of the lead sQTL SNP rs10804111 in a minigene assay. The intron from the exon 8L to 9 junction is >5,500 bp and not amenable to cloning in a simple minigene vector. We therefore constructed minigene vectors to contain *CASP8* exons 8/8L and surrounding intron (100 bp upstream of exon 8 and 100 bp downstream intronic sequence from the exon 8L junction) and *CASP*8 exon 9 and surrounding intronic sequence (100 bp upstream and 100 bp downstream intronic sequence), linked together with 201 bp of intronic sequence surrounding each allele of rs10804111 (**Figure S8**). *CASP8* exons 8 to 9 (including intron 8; 1,181 bp), along with the 5' flanking of exon 8 and 3' flanking of 9 sequences (100 bp for each) were custom-synthesized from Thermo Fisher Scientific and cloned in sense orientation using XhoI and BamHI restriction sites of the Exontrap vector pET01 (MoBiTec, Germany) to generate the pET01-rs10804111(C/T) mini-gene constructs containing each allele of rs10804111. Sequence-verified pET01 constructs were transfected into UACC903 melanoma cells using Lipofectamine™ 2000 (Invitrogen, Waltham, MA) in 6-well format. Cells were harvested 48 h after transfection, and total RNA was extracted with the QIAGEN QIAcube using RNeasy kit with on-column DNase I treatment (Qiagen, Germany). 1 µg of total RNA was converted into cDNA with SuperScript™ III reverse transcriptase (Invitrogen) and a vector-specific cDNA primer listed in **Table S31**. Taqman assays targeting unique junctions of pET01-rs10804111 were obtained from Thermo Fisher Scientific (vector exon 2 (VE2) as a control: APH6CGV; VE1 to VE2 junction: APPRMTJ; *CASP8* exon 8 to exon 9 junction: APKA42T; *CASP8* exon 8L to exon 9 junction: APNKT7M; listed in **Table S1**). Taqman assays were performed in triplicate and normalized to VE2 levels.

Capture-C data

Capture-C data were generated and analyzed as previously described [48,49]. Briefly, regions for bait design were determined for each locus via identification of the set of SNPs with a LLR < 100 relative to the leading SNP of a given GWAS signal. For the 2q33.1 locus, capture baits were tiled across the entire region of association (chr2:202,114,359 to 202,205,025); baited fragments are listed in **Table S32**. Capture-C libraries were made using the Arima HiC Kit (Arima Genomics, Carlsbad, CA) and the KAPA HyperPrep Kit (KAPA Biosystems, Wilmington, MA) following the manufacturer's protocol. Data are presented for region-specific Capture-C runs with five independent human primary melanocyte cultures (C56, C140, C205, C24, and C27) with three biological replicates for each. Paired-end sequencing reads from biological replicates were pre-processed with the HiCUP pipeline[82] and aligned to human genome version 19 via bowtie2[83,84]. Chromatin interaction loops were detected at one- and four-fragment resolutions via Capture Hi-C Analysis of Genomic Organization (CHiCAGO) pipeline v.1.16.0[85]. Because we observed no distant interactions between fine-mapped variants and gene promoter at one-fragment resolution, data presented are from four-fragment analysis. Visualization was performed using the WashU Epigenome Browser[86].

Chromatin conformation capture (3C)

Chromatin conformation capture (3C) assays were done based on the protocol from the Dekker Lab[87]. RP11-1078D1 bacterial artificial chromosome (BAC) clones were purchased from BACPAC Resources Center and purified with the QIAGEN large-construct Maxi Kit to cover the 2q33.1 genomic region (chr2:202,045,328 to 202,245,875). BAC libraries were made by HindIII digestion of BAC plasmids, followed by ligation and DNA purification. To generate melanoma cell and melanocyte 3C libraries, we fixed and lysed 1 x 10^8 cells, followed by HindIII digestion and ligation. Both the BAC libraries, and melanoma and melanocytes 3C libraries were amplified with a Taqman assay containing a primer localized to various regions of chr2:202,045,328 to 202,245,875 and a fixed primer harboring rs3769823, rs3769821, and rs59308963 plus a FAM-labeled probe annealing to 3' of the fixed primer. Primers are listed in **Table S31**. PCR cycle conditions were as follows: 95°C, 5min; [95°C, 30s; 62°C, 30s; 72°C, 30s]$_{x34}$; 72°C, 10min. Amplification for different primer pairs from the melanoma and melanocyte 3C libraries was normalized to that of BAC libraries.

Epigenomic annotations

Chromatin Primary Core Marks Segmentation by HMM (ChromHMM), DNase Hypersensitivity (DHS), and histone modification data for primary human melanocyte cultures, primary human keratinocyte culture, and human breast mammary cultures "breast variant human mammary epithelial cells" and "breast myoepithelial primary cells" (shown as vHMEC and BMYO, respectively, in **Figure 1**) were obtained from the Roadmap Epigenomics Project (downloaded through the UCSC Genome Browser; http://genome.ucsc.edu/). Peaks from FAIRE-seq and H3K27ac ChIP for 11 melanoma cultures samples were obtained from the Gene Expression Omnibus (GEO; accession GSE60666)[45].

Cell culture

We originally obtained melanoma cell lines (UACC903 and UACC1113) from the University of Arizona Cancer Center (UACC), which were maintained in RPMI-1640 medium with L-glutamine (Quality Biological, 112-025-101) supplemented with 10% FBS (GenClone, 25-514H), 20 mM HEPES (pH 7.9; Gibco, 15630), and 100 U/ml penicillin-streptomycin (Gibco, 15140). Primary melanocytes from foreskin healthy newborn males, mainly of European descent, were grown in M254 media (Invitrogen, M254500) with HMGS-2 supplement (Invitrogen, S0165). MCF7 and T47D breast cancer cell lines were purchased from ATCC. MCF7 cells were grown in DMEM with L-glutamine (Quality Biological, 112-014-101) supplemented with 10% FBS, and 100 U/ml penicillin-streptomycin. T47D cells were maintained in RPMI 1640 supplemented with 10% FBS, and 100 U/ml penicillin-streptomycin. All cell lines were grown in a 5% CO_2 humidified incubator at 37 °C and were tested for mycoplasma contamination every 3-6 months.

Luciferase reporter assays

Sequences encompassing each variant were PCR-amplified (primers listed in **Table S31**) from genomic DNA of HapMap CEU panel samples with the appropriate genotypes to obtain clones with each genotype, and cloned into the HindIII and XhoI sites of the pGL4.23[*luc2*/minP] (Promega, Madison, MI) luciferase vector in the 5'-to-3' or 3'-to'5' orientation using In-Fusion HD Cloning Kit (Takara Bio). Sequence-verified pGL4.23 constructs were co-transfected with the pGL4.73[*hRluc*/SV40] Renilla luciferase control vector (Promega) into melanoma cell lines (UACC903 and UACC1113) using Lipofectamine™ 2000 (Invitrogen) in 24-well format. Luciferase activity was measured 24 h after transfection with the Dual Luciferase Reporter Assay System (Promega). Luciferase activity was normalized to the Renilla luciferase activity. All the experiments were performed in six replicate wells and repeated for at least three

biological replicates. Significance between alleles was assessed using a two-tailed, unpaired t-test assuming unequal variance.

Massively parallel reporter assay data

We used data from two massively parallel reporter assay (MPRA) datasets assessing allelic *cis*-regulatory potential for candidate causal variants from this locus in melanoma cells and melanocytes[46,47]. Firstly, in a prior MPRA analysis (MPRA v.1)[46], 823 variants from 20 genome-wide significant loci from a prior melanoma meta-analysis[8] were tested in a melanoma cell line, UACC903, as well as HEK293FT cells. For this study, an LD-based selection criteria was used ($r^2 > 0.4$ with the sentinel SNP using 1,000 Genomes phase 3 EUR data), further filtering candidate causal variants to test only those located within annotated open chromatin regions and promoter/enhancer histone marks found in primary melanocytes and/or short-term cultures. A second MPRA study (MPRA v.2)[47] investigated 54 loci from a more recent melanoma GWAS meta-analysis[10] in both melanoma cells and melanocytes. Here, 1,992 candidate causal variants were tested from 54 genome-wide significant loci. Candidate causal variants were fine-mapped using log-likelihood ratios (LLR < 1:1,000 with the lead SNP) or where variants were not successfully imputed or tested in the GWAS using an LD-based criteria ($r^2 > 0.8$ with the lead SNP, 1,000 Genomes EUR).

EMSAs and supershift assays

Nuclear extracts from melanoma cell lines (UACC903 and UACC1113) or human melanocytes were prepared using the NE-PER Nuclear and Cytoplasmic Extraction Kit (Thermo Fisher Scientific). Nuclear extract lysates for MCF7 and T47D breast cancer cell lines were purchased from Abcam (ab14860 and ab14896, respectively). Oligonucleotides for each SNP were synthesized with or without biotin labeling at the 5' end and HPLC purified (21-28nt, Life Technologies, listed in **Table S31**). Forward and reverse strands were then annealed to generate double stranded 5'-end labelled probes or unlabeled competitors. 50 nM probes were added to 2 µg of nuclear extract preincubated with 1 µg of poly d(I-C) (Roche) in LightShift™ 1 × binding buffer (Thermo Fisher Scientific) for 30 min on ice. Competition experiments were performed by adding 1-100 fold more unlabeled competitor oligonucleotides to the reaction mixture 5 min before the addition of labelled probes. Anti-E4F1 (Abcam, ab70615) or anti-IRF2 (Abcam, ab245658) antibodies for supershift assays were incubated with nuclear extract for 1 h at 4 °C prior to adding poly d(I-C). The reactions were run on 5 – 10% TBE gels (Bio-Rad Criterion) on ice at 120 V, transferred to Biodyne B membranes (VWR), transferred blots were

crosslinked (Stratagene UV Stratalinker 1800) and detected using the LightShift Chemiluminescent EMSA Kit (Thermo Fisher Scientific) and imaged on Chemidoc Touch (Bio-Rad). One representative data is shown from three biological replicates.

Quantitative mass spectrometry

MCF7 breast cancer cells were grown in RPMI 1640 (Gibco) supplemented with 10% FBS, 100 U/ml penicillin and 100 µg/ml streptomycin (Gibco). The same culture conditions were used for UACC903 melanoma cells with additional 20 mM HEPES (pH 7.9). Nuclear lysates for DNA affinity purifications were collected as described previously[88]. Oligonucleotide probes with 5'-biotinylation of the forward strand were ordered from Integrated DNA Technologies (**Table S31**). DNA pulldowns and on-bead trypsin digestion were performed in Eppendorf tubes as described previously[89]. Briefly, forward and reverse oligos were annealed with a 1.5X molar excess of the reverse strand. For each pulldown, annealed DNA oligos (500 pmol) were immobilized on 10 µl (20 µl slurry) Streptavidin-Sepharose beads (GE Healthcare, Chicago, IL). Immobilized DNA oligos were incubated with nuclear extracts (500 µg) from UACC903 and MCF7, and 10 µg of non-specific competitor DNA (5 µg polydIdC, 5 µg polydAdT). After washing away unbound proteins, beads were resuspended in elution buffer (2 M Urea, 100 mM TRIS (pH 8), 10 mM DTT), peptides were alkylated with 50 mM iodoacetamide, and on-bead digested with 0.25 µg trypsin. Tryptic peptides were labeled by stable isotope dimethyl labeling on StageTips, as described previously[89]. Samples were eluted from the StageTips and matching light and medium labeled samples were combined. Samples were loaded onto a 30 cm column (heated at 40 °C) packed in-house with 1.8 µm Reprosil-Pur C18-AQ (Dr Maisch) and eluted using a gradient from 9 to 32% Buffer B (80% acetonitrile, 0.1% formic acid) over 114 min at a flow rate of 250 nL/min using an Easy-nLC 1000 (Thermo Fisher Scientific). Peptides were sprayed directly into either a Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific) or a Q Exactive mass spectrometer (Thermo Fisher Scientific). The mass spectrometers were operated as described previously[46,89]. 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[90]. 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. rs3769823 and rs3769821 were assayed in both UACC903 and MCF7 cells, while rs59308963 was assayed only in UACC903.

Chromatin immunoprecipitation and genotyping

Following the manufacturer's instructions of the Active Motif ChIP-IT High-Sensitivity kit, actively growing melanoma cells (UACC903 and UACC1649) or C87 primary melanocytes were fixed with 1% formaldehyde when 80-90% confluent. Nuclei of 1.0 – 1.5 × 10^7 cells were prepared and sheared with ME220 Focused-ultrasonicator (Covaris, Woburn, MA) for 20 min following the instructions. 10 – 30 µg sheared chromatin from 1.5 – 4.5 × 10^6 cells were used for each immunoprecipitation reaction with anti-E4F1, anti-IRF2 or anti-IgG (Abcam, ab37415) following the instructions. Purified immunoprecipitated DNA or input DNA was analyzed by SYBR Green qPCR for enrichment of target sites using the primers listed in **Table S31**. ChIP DNA or input DNA from UACC1649 or C87 melanocyte cell lines (heterozygous for rs3769823) was assayed for genotyping rs3769823 by Taqman genotyping assay (Assay ID: C__25808407_20). All experiments were performed in triplicate (technical replicates) and repeated for at least three biological replicates.

siRNA-mediated knockdown of E4F1 and IRF2

SMARTpool ON-TARGETplus Human siRNAs to *E4F1* (L-011847-00-0005), *IRF2* (L-011705-02-0005) and ON-TARGETplus Non-targeting controls (D-001810-01-05) were purchased from Dharmacon (Lafayette, CO). Transfections were carried out using Lipofectamine™ RNAiMAX (Invitrogen) according to the manufacturer's instructions. Cells were collected after 48h for RNA and protein extractions.

qPCR

RNA was isolated using RNeasy Mini Column (Qiagen), which was always complemented by DNase treatment. cDNA was synthesized from the total RNA using iScript™ Advanced cDNA Synthesis Kit (Bio-Rad). Gene expression levels were quantified by qPCR using Taqman assays for *E4F1* (Hs00231773_m1), *IRF2* (Hs01082884_m1), *CASP8* (Hs01018151_m1, Hs01018149, Hs01018160_m1, Hs04405665_m1 and Hs01022432_m1), and *GAPDH* (Hs99999905_m1) from Thermo Fisher Scientific. Gene expression levels of *E4F1*, *IRF2* and *CASP8* were normalized to *GAPDH*. Each experiment was performed in triplicate and replicated three times. Significance was assessed using a Student's two-tailed T-test assuming unequal variances.

Western blotting

Whole-cell extracts prepared in RIPA buffer with protease inhibitors (Complete Protease Inhibitor, Roche) or phosphatase inhibitors (PhosSTOP, Roche). Samples were quantified using

DC Protein Assay (Bio-Rad) and electrophoresed on 4–12% Bis Tris Plus Bolt gels in MES buffer (Invitrogen). Proteins were transferred to nitrocellulose using an iBlot2 (Invitrogen). Blots were blocked with 5% non-fat dry milk in Tris-buffered saline with 0.1% Tween-20 (TBST). Primary and secondary antibodies were diluted in 1-2% milk in TBST, and all washes were performed with TBST. Blots were rinsed briefly with PBS before the addition of ECL Prime Western Blotting Detection Reagent (Amersham, UK). Images were captured on ChemiDoc™ Gel Imaging System (Bio-Rad). The antibodies of anti-E4F1 (Santa Cruz, sc-514718), anti-IRF2 (Santa Cruz, sc-101069,) and anti-β-actin (Santa Cruz, sc-47778) were used as primary antibodies, and goat anti-mouse HRP (Santa Cruz, sc-2005) for secondary antibodies.

Analysis of melanocyte and melanoma expression data

Transcriptomic data were generated as described previously from human primary melanocytes[36], TCGA-SKCM[36,41,42], early passage melanoma cultures[43], the Leeds Melanoma Cohort[91], and GTEx tissues (v.7; https://gtexportal.org/home/index.html)[40]. In brief, RNA Sequencing by Expectation maximization (RSEM, version 1.2.31, http://deweylab.github.io/RSEM/) was used to quantify the gene expression followed by the quantile normalization to obtain the final RSEM for primary melanocytes, TCGA-SKCM, and GTEx tissues as previously described[36]. For background correction and quantile normalization, Robust Multi-array Average (RMA) algorithm with default settings (Affymetrix) was performed for melanoma cell lines[43]. A primary melanoma transcriptomic dataset from the Leeds Melanoma Cohort study on 703 patients was generated using the Illumina DASL Human HT12 v4 array platform, as previously described[91] (European Genome-phenome Archive accession EGAS00001002922; https://ega-archive/org/studies/EGAS00001002922). 16 samples were removed in quality control (expression data from only one extraction per patient) leaving a cohort of 687 patients[91]. LUMI[92] was used for background correction and quantile normalization as previously reported[92]. For GTEx (v.7) transcriptomic data, genes were selected based on expression thresholds of > 0.5 RSEM and ≥ 6 reads in at least 10 samples, and expression values for each gene were inverse quantile normalized to a standard normal distribution across samples and described previously[36]. We used R (version 4.1.0) for Pearson correlation to analyze the correlations between candidate target genes and allelic-specific binding transcription factors. The significance level was set at $P < 0.05$ and [slope] > 0.4 for all these tests. Using transcriptomic and genotype datasets from multiple tissues, we also ran multiple linear regression on R (version 4.1.0) to evaluate the influence of allelic-specific binding transcription factor or genetic variable of rs3769823 on the *CASP8*.

Acknowledgements

This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH). The contributions of the NIH author(s) are considered Works of the United States Government. The findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services. We would like to thank members at the National Cancer Institute's Cancer Genomics Research Laboratory (CGR) for help with sequencing efforts and Rouf Banday from the National Cancer Institute for assistance with minigene assay design and data interpretation. This work has been supported by the Intramural Research Program (IRP) of the National Cancer Institute, US National Institutes of Health. We thank Jeremy Bravo Narula for assistance with the manuscript. We thank all the cohorts, funders, and investigators who contributed to the melanoma GWAS, as originally acknowledged by Landi, 23andMe, and colleagues[10]; data from this GWAS were used toward fine-mapping. JN receives funding from Horizon Europe 101136622. The Leeds Melanoma Cohort recruitment, follow up and data collection was supported by CRUK C588/A19167, C8216/A6129, C588/A10721 and NIH CA83115. Mark Iles is supported in part by the National Institute for Health and Care Research (NIHR) Leeds Biomedical Reearch Centre (BRC) (NIHR203331). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the department of Health and Social Care.

Declaration of interest

The authors declare no competing interests.

Data and code availability

Conditional fine-mapping data are available in **Tables S1** and **S11**; Bayesian fine-mapping results are available in **Tables S5**-**S7**. Colocalization analyses are available are in **Tables S2, S3**, **S12**, **S13**, **S19**, and **S20**. Data from the 2020 melanoma GWAS meta-analysis performed by Landi and colleagues was obtained from dbGap (dbGap: phs001868.v1.p1), with the exclusion

of self-reported data from 23andMe and UK Biobank. The full GWAS summary statistics for the 23andMe discovery dataset will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please visit the 23andMe website for more information and to apply to access the data. Summary data from the remaining self-reported cases are available from the corresponding authors of that manuscript (Matthew H Law, matthew.law@qimrberghofer.edu.au; Mark M Iles, m.m.iles@leeds.ac.uk; and Maria Teresa Landi, landim@mail.nih.gov). Melanocyte eQTL data and RNA-seq expression data from 106 individuals are available from dbGap (dbGap: phs001500.v1.p1), and pre-analyzed cis-eQTL data from GTEx (v.7) were downloaded from the GTEx portal. eQTL data from 59 early melanoma cell lines was available from NCBI Gene Expression Omnibus (accessions GSE78995 for expression data; GSE99193 for genotyping data). Leeds Melanoma Cohort data from 703 patients was available in the European Genome-phenome Archive accession EGAS00001002922. MPRA data are available in the NCBI Gene Expression Omnibus as a SuperSeries under the accession number GEO: GSE129250 and GSE210356. Capture-C data have been previously published[48,49]; interaction data including both called interactions as well as raw sequencing data are available through ArrayExpress (accession: E-MTAB-15079). Specifically, for the 2q33.1 locus restriction fragments baited for region-specific Capture-C assays are provided in **Table S32**, and loops called by ChICAGO are in **Tables S23** and **S24**. Pearson correlations are available in **Table S25**, **S29** and **S30**. Multiple linear regression results are available in **Tables S26**-**S28**. Luciferase assay fragments, other primers and Taqman assays are listed in **Table S31**.

Web resources

23andMe, https://research.23andme.com/collaborate/#dataset-access/

GTEx portal, https://www.gtexportal.org/home/

https://ldlink.nci.nih.gov/?tab=home

NCBI GEO, https://www.ncbi.nlm.nih.gov/geo/

RegulomeDB, https://regulomedb.org/regulome-search/

The Cancer Genome Atlas (TCGA) Research Network,

https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga

References

1. Brown, K.M., Macgregor, S., Montgomery, G.W., Craig, D.W., Zhao, Z.Z., Iyadurai, K., Henders, A.K., Homer, N., Campbell, M.J., Stark, M., et al. (2008). Common sequence variants on 20q11.22 confer melanoma susceptibility. Nat Genet *40*, 838-840. 10.1038/ng.163.
2. Gudbjartsson, D.F., Sulem, P., Stacey, S.N., Goldstein, A.M., Rafnar, T., Sigurgeirsson, B., Benediktsdottir, K.R., Thorisdottir, K., Ragnarsson, R., Sveinsdottir, S.G., et al. (2008). ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. Nat Genet *40*, 886-891. 10.1038/ng.161.
3. Falchi, M., Bataille, V., Hayward, N.K., Duffy, D.L., Bishop, J.A., Pastinen, T., Cervino, A., Zhao, Z.Z., Deloukas, P., Soranzo, N., et al. (2009). Genome-wide association study identifies variants at 9p21 and 22q13 associated with development of cutaneous nevi. Nat Genet *41*, 915-919. 10.1038/ng.410.
4. Bishop, D.T., Demenais, F., Iles, M.M., Harland, M., Taylor, J.C., Corda, E., Randerson-Moor, J., Aitken, J.F., Avril, M.F., Azizi, E., et al. (2009). Genome-wide association study identifies three loci associated with melanoma risk. Nat Genet *41*, 920-925. 10.1038/ng.411.
5. Barrett, J.H., Iles, M.M., Harland, M., Taylor, J.C., Aitken, J.F., Andresen, P.A., Akslen, L.A., Armstrong, B.K., Avril, M.F., Azizi, E., et al. (2011). Genome-wide association study identifies three new melanoma susceptibility loci. Nat Genet *43*, 1108-1113. 10.1038/ng.959.
6. Macgregor, S., Montgomery, G.W., Liu, J.Z., Zhao, Z.Z., Henders, A.K., Stark, M., Schmid, H., Holland, E.A., Duffy, D.L., Zhang, M., et al. (2011). Genome-wide association study identifies a new melanoma susceptibility locus at 1q21.3. Nat Genet *43*, 1114-1118. 10.1038/ng.958.
7. Iles, M.M., Law, M.H., Stacey, S.N., Han, J., Fang, S., Pfeiffer, R., Harland, M., Macgregor, S., Taylor, J.C., Aben, K.K., et al. (2013). A variant in FTO shows association with melanoma risk not due to BMI. Nat Genet *45*, 428-432, 432e421. 10.1038/ng.2571.
8. Law, M.H., Bishop, D.T., Lee, J.E., Brossard, M., Martin, N.G., Moses, E.K., Song, F., Barrett, J.H., Kumar, R., Easton, D.F., et al. (2015). Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. Nat Genet *47*, 987-995. 10.1038/ng.3373.
9. Ransohoff, K.J., Wu, W., Cho, H.G., Chahal, H.C., Lin, Y., Dai, H.J., Amos, C.I., Lee, J.E., Tang, J.Y., Hinds, D.A., et al. (2017). Two-stage genome-wide association study identifies a novel susceptibility locus associated with melanoma. Oncotarget *8*, 17586-17592. 10.18632/oncotarget.15230.
10. Landi, M.T., Bishop, D.T., MacGregor, S., Machiela, M.J., Stratigos, A.J., Ghiorzo, P., Brossard, M., Calista, D., Choi, J., Fargnoli, M.C., et al. (2020). Genome-wide association meta-analyses combining multiple risk phenotypes provide insights into the genetic architecture of cutaneous melanoma susceptibility. Nat Genet *52*, 494-504. 10.1038/s41588-020-0611-8.
11. Michailidou, K., Lindstrom, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemacon, A., Soucy, P., Glubb, D., Rostamianfar, A., et al. (2017). Association analysis identifies 65 new breast cancer risk loci. Nature *551*, 92-94. 10.1038/nature24284.
12. Stacey, S.N., Helgason, H., Gudjonsson, S.A., Thorleifsson, G., Zink, F., Sigurdsson, A., Kehr, B., Gudmundsson, J., Sulem, P., Sigurgeirsson, B., et al. (2015). New basal cell carcinoma susceptibility loci. Nat Commun *6*, 6825. 10.1038/ncomms7825.
13. Liyanage, U.E., Law, M.H., Han, X., An, J., Ong, J.S., Gharahkhani, P., Gordon, S., Neale, R.E., Olsen, C.M., andMe Research, T., et al. (2019). Combined analysis of keratinocyte cancers identifies novel genome-wide loci. Hum Mol Genet *28*, 3148-3160. 10.1093/hmg/ddz121.

1112 14. Hart, K., Landvik, N.E., Lind, H., Skaug, V., Haugen, A., and Zienolddiny, S. (2011). A combination
1113 of functional polymorphisms in the CASP8, MMP1, IL10 and SEPS1 genes affects risk of non-
1114 small cell lung cancer. Lung Cancer *71*, 123-129. 10.1016/j.lungcan.2010.04.016.
1115 15. Lubahn, J., Berndt, S.I., Jin, C.H., Klim, A., Luly, J., Wu, W.S., Isaacs, S., Wiley, K., Isaacs, W.B.,
1116 Suarez, B.K., et al. (2010). Association of CASP8 D302H polymorphism with reduced risk of
1117 aggressive prostate carcinoma. Prostate *70*, 646-653. 10.1002/pros.21098.
1118 16. Shi, Y. (2002). Mechanisms of caspase activation and inhibition during apoptosis. Mol Cell *9*,
1119 459-470. 10.1016/s1097-2765(02)00482-3.
1120 17. Boldin, M.P., Goncharov, T.M., Goltsev, Y.V., and Wallach, D. (1996). Involvement of MACH, a
1121 novel MORT1/FADD-interacting protease, in Fas/APO-1- and TNF receptor-induced cell death.
1122 Cell *85*, 803-815. 10.1016/s0092-8674(00)81265-9.
1123 18. Muzio, M., Chinnaiyan, A.M., Kischkel, F.C., O'Rourke, K., Shevchenko, A., Ni, J., Scaffidi, C., Bretz,
1124 J.D., Zhang, M., Gentz, R., et al. (1996). FLICE, a novel FADD-homologous ICE/CED-3-like protease,
1125 is recruited to the CD95 (Fas/APO-1) death--inducing signaling complex. Cell *85*, 817-827.
1126 10.1016/s0092-8674(00)81266-0.
1127 19. Kischkel, F.C., Hellbardt, S., Behrmann, I., Germer, M., Pawlita, M., Krammer, P.H., and Peter,
1128 M.E. (1995). Cytotoxicity-dependent APO-1 (Fas/CD95)-associated proteins form a death-
1129 inducing signaling complex (DISC) with the receptor. EMBO J *14*, 5579-5588.
1130 20. Boatright, K.M., Renatus, M., Scott, F.L., Sperandio, S., Shin, H., Pedersen, I.M., Ricci, J.E., Edris,
1131 W.A., Sutherlin, D.P., Green, D.R., and Salvesen, G.S. (2003). A unified model for apical caspase
1132 activation. Mol Cell *11*, 529-541. 10.1016/s1097-2765(03)00051-0.
1133 21. Pop, C., Fitzgerald, P., Green, D.R., and Salvesen, G.S. (2007). Role of proteolysis in caspase-8
1134 activation and stabilization. Biochemistry *46*, 4398-4407. 10.1021/bi602623b.
1135 22. Oberst, A., Pop, C., Tremblay, A.G., Blais, V., Denault, J.B., Salvesen, G.S., and Green, D.R. (2010).
1136 Inducible dimerization and inducible cleavage reveal a requirement for both processes in
1137 caspase-8 activation. J Biol Chem *285*, 16632-16642. 10.1074/jbc.M109.095083.
1138 23. Catts, V.S., and Weickert, C.S. (2012). Gene expression analysis implicates a death receptor
1139 pathway in schizophrenia pathology. PLoS One *7*, e35511. 10.1371/journal.pone.0035511.
1140 24. Thome, M., Schneider, P., Hofmann, K., Fickenscher, H., Meinl, E., Neipel, F., Mattmann, C.,
1141 Burns, K., Bodmer, J.L., Schroter, M., et al. (1997). Viral FLICE-inhibitory proteins (FLIPs) prevent
1142 apoptosis induced by death receptors. Nature *386*, 517-521. 10.1038/386517a0.
1143 25. Budihardjo, I., Oliver, H., Lutter, M., Luo, X., and Wang, X. (1999). Biochemical pathways of
1144 caspase activation during apoptosis. Annu Rev Cell Dev Biol *15*, 269-290.
1145 10.1146/annurev.cellbio.15.1.269.
1146 26. Krueger, A., Schmitz, I., Baumann, S., Krammer, P.H., and Kirchhoff, S. (2001). Cellular FLICE-
1147 inhibitory protein splice variants inhibit different steps of caspase-8 activation at the CD95
1148 death-inducing signaling complex. J Biol Chem *276*, 20633-20640. 10.1074/jbc.M101780200.
1149 27. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide
1150 complex trait analysis. Am J Hum Genet *88*, 76-82. 10.1016/j.ajhg.2010.11.011.
1151 28. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Genetic Investigation of, A.T.C., Replication,
1152 D.I.G., Meta-analysis, C., Madden, P.A., Heath, A.C., Martin, N.G., et al. (2012). Conditional and
1153 joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing
1154 complex traits. Nat Genet *44*, 369-375, S361-363. 10.1038/ng.2213.
1155 29. Foley, C.N., Staley, J.R., Breen, P.G., Sun, B.B., Kirk, P.D.W., Burgess, S., and Howson, J.M.M.
1156 (2021). A fast and efficient colocalization algorithm for identifying shared genetic risk factors
1157 across multiple traits. Nat Commun *12*, 764. 10.1038/s41467-020-20885-8.

1158 30. Hormozdiari, F., van de Bunt, M., Segre, A.V., Li, X., Joo, J.W.J., Bilow, M., Sul, J.H., Sankararaman,
1159 S., Pasaniuc, B., and Eskin, E. (2016). Colocalization of GWAS and eQTL Signals Detects Target
1160 Genes. Am J Hum Genet *99*, 1245-1260. 10.1016/j.ajhg.2016.10.003.
1161 31. Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations.
1162 Bioinformatics *28*, 1353-1358. 10.1093/bioinformatics/bts163.
1163 32. Wen, X., Lee, Y., Luca, F., and Pique-Regi, R. (2016). Efficient Integrative Multi-SNP Association
1164 Analysis via Deterministic Approximation of Posteriors. Am J Hum Genet *98*, 1114-1129.
1165 10.1016/j.ajhg.2016.03.029.
1166 33. Lee, Y., Luca, F., Pique-Regi, R., and Wen, X. (2018). Bayesian Multi-SNP Genetic Association
1167 Analysis: Control of FDR and Use of Summary Statistics. bioRxiv.
1168 34. Kichaev, G., Yang, W.Y., Lindstrom, S., Hormozdiari, F., Eskin, E., Price, A.L., Kraft, P., and
1169 Pasaniuc, B. (2014). Integrating functional data to prioritize causal variants in statistical fine-
1170 mapping studies. PLoS Genet *10*, e1004722. 10.1371/journal.pgen.1004722.
1171 35. Kichaev, G., and Pasaniuc, B. (2015). Leveraging Functional-Annotation Data in Trans-ethnic
1172 Fine-Mapping Studies. Am J Hum Genet *97*, 260-271. 10.1016/j.ajhg.2015.06.007.
1173 36. Zhang, T., Choi, J., Kovacs, M.A., Shi, J., Xu, M., Program, N.C.S., Melanoma Meta-Analysis, C.,
1174 Goldstein, A.M., Trower, A.J., Bishop, D.T., et al. (2018). Cell-type-specific eQTL of primary
1175 melanocytes facilitates identification of melanoma susceptibility genes. Genome Res *28*, 1621-
1176 1635. 10.1101/gr.233304.117.
1177 37. Bendl, J., Musil, M., Stourac, J., Zendulka, J., Damborsky, J., and Brezovsky, J. (2016).
1178 PredictSNP2: A Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different
1179 Characteristics of Variants in Distinct Genomic Regions. PLoS Comput Biol *12*, e1004962.
1180 10.1371/journal.pcbi.1004962.
1181 38. Zhang, T., Choi, J., Dilshat, R., Einarsdottir, B.O., Kovacs, M.A., Xu, M., Malasky, M., Chowdhury,
1182 S., Jones, K., Bishop, D.T., et al. (2021). Cell-type-specific meQTLs extend melanoma GWAS
1183 annotation beyond eQTLs and inform melanocyte gene-regulatory mechanisms. Am J Hum
1184 Genet *108*, 1631-1646. 10.1016/j.ajhg.2021.06.018.
1185 39. Himeji, D., Horiuchi, T., Tsukamoto, H., Hayashi, K., Watanabe, T., and Harada, M. (2002).
1186 Characterization of caspase-8L: a novel isoform of caspase-8 that behaves as an inhibitor of the
1187 caspase cascade. Blood *99*, 4070-4078. 10.1182/blood.v99.11.4070.
1188 40. Consortium, G.T., Laboratory, D.A., Coordinating Center -Analysis Working, G., Statistical
1189 Methods groups-Analysis Working, G., Enhancing, G.g., Fund, N.I.H.C., Nih/Nci, Nih/Nhgri,
1190 Nih/Nimh, Nih/Nida, et al. (2017). Genetic effects on gene expression across human tissues.
1191 Nature *550*, 204-213. 10.1038/nature24277.
1192 41. Cancer Genome Atlas, N. (2015). Genomic Classification of Cutaneous Melanoma. Cell *161*,
1193 1681-1696. 10.1016/j.cell.2015.05.044.
1194 42. Hoadley, K.A., Yau, C., Hinoue, T., Wolf, D.M., Lazar, A.J., Drill, E., Shen, R., Taylor, A.M.,
1195 Cherniack, A.D., Thorsson, V., et al. (2018). Cell-of-Origin Patterns Dominate the Molecular
1196 Classification of 10,000 Tumors from 33 Types of Cancer. Cell *173*, 291-304 e296.
1197 10.1016/j.cell.2018.03.022.
1198 43. Choi, J., Xu, M., Makowski, M.M., Zhang, T., Law, M.H., Kovacs, M.A., Granzhan, A., Kim, W.J.,
1199 Parikh, H., Gartside, M., et al. (2017). A common intronic variant of PARP1 confers melanoma
1200 risk and mediates melanocyte growth via regulation of MITF. Nat Genet *49*, 1326-1335.
1201 10.1038/ng.3927.
1202 44. Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-
1203 Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111
1204 reference human epigenomes. Nature *518*, 317-330. 10.1038/nature14248.

1205 45. Verfaillie, A., Imrichova, H., Atak, Z.K., Dewaele, M., Rambow, F., Hulselmans, G., Christiaens, V.,
1206 Svetlichnyy, D., Luciani, F., Van den Mooter, L., et al. (2015). Decoding the regulatory landscape
1207 of melanoma reveals TEADS as regulators of the invasive cell state. Nat Commun *6*, 6683.
1208 10.1038/ncomms7683.
1209 46. Choi, J., Zhang, T., Vu, A., Ablain, J., Makowski, M.M., Colli, L.M., Xu, M., Hennessey, R.C., Yin, J.,
1210 Rothschild, H., et al. (2020). Massively parallel reporter assays of melanoma risk variants identify
1211 MX2 as a gene promoting melanoma. Nat Commun *11*, 2718. 10.1038/s41467-020-16590-1.
1212 47. Long, E., Yin, J., Funderburk, K.M., Xu, M., Feng, J., Kane, A., Zhang, T., Myers, T., Golden, A.,
1213 Thakur, R., et al. (2022). Massively parallel reporter assays and variant scoring identified
1214 functional variants and target genes for melanoma loci and highlighted cell-type specificity. Am J
1215 Hum Genet *109*, 2210-2229. 10.1016/j.ajhg.2022.11.006.
1216 48. Xu, M., Mehl, L., Zhang, T., Thakur, R., Sowards, H., Myers, T., Jessop, L., Chesi, A., Johnson, M.E.,
1217 Wells, A.D., et al. (2021). A UVB-responsive common variant at chromosome band 7p21.1
1218 confers tanning response and melanoma risk via regulation of the aryl hydrocarbon receptor,
1219 AHR. Am J Hum Genet *108*, 1611-1630. 10.1016/j.ajhg.2021.07.002.
1220 49. Thakur, R., Xu, M., Sowards, H., Yon, J., Jessop, L., Myers, T., Zhang, T., Chari, R., Long, E., Rehling,
1221 T., et al. (2025). Mapping chromatin interactions at melanoma susceptibility loci uncovers
1222 distant cis-regulatory gene targets. Am J Hum Genet *112*, 1625-1648.
1223 10.1016/j.ajhg.2025.04.015.
1224 50. Rizos, H., Diefenbach, E., Badhwar, P., Woodruff, S., Becker, T.M., Rooney, R.J., and Kefford, R.F.
1225 (2003). Association of p14ARF with the p120E4F transcriptional repressor enhances cell cycle
1226 inhibition. J Biol Chem *278*, 4981-4989. 10.1074/jbc.M210978200.
1227 51. Dai, J., Lv, J., Zhu, M., Wang, Y., Qin, N., Ma, H., He, Y.Q., Zhang, R., Tan, W., Fan, J., et al. (2019).
1228 Identification of risk loci and a polygenic risk score for lung cancer: a large-scale prospective
1229 cohort study in Chinese populations. Lancet Respir Med *7*, 881-891. 10.1016/S2213-
1230 2600(19)30144-4.
1231 52. Schumacher, F.R., Al Olama, A.A., Berndt, S.I., Benlloch, S., Ahmed, M., Saunders, E.J., Dadaev, T.,
1232 Leongamornlert, D., Anokian, E., Cieza-Borrella, C., et al. (2018). Association analyses of more
1233 than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet *50*, 928-936.
1234 10.1038/s41588-018-0142-8.
1235 53. Berndt, S.I., Skibola, C.F., Joseph, V., Camp, N.J., Nieters, A., Wang, Z., Cozen, W., Monnereau, A.,
1236 Wang, S.S., Kelly, R.S., et al. (2013). Genome-wide association study identifies multiple risk loci
1237 for chronic lymphocytic leukemia. Nat Genet *45*, 868-876. 10.1038/ng.2652.
1238 54. Zhang, X., Qin, N., Fan, J., Zhang, C., Sun, Q., Gu, Y., Zhu, M., Zhang, E., Dai, J., Jin, G., et al. (2022).
1239 A causal variant rs3769823 in 2q33.1 involved in apoptosis pathway leading to a decreased risk
1240 of non-small cell lung cancer. Cancer Biol Med *19*, 1385-1396. 10.20892/j.issn.2095-
1241 3941.2022.0068.
1242 55. Camp, N.J., Lin, W.Y., Bigelow, A., Burghel, G.J., Mosbruger, T.L., Parry, M.A., Waller, R.G., Rigas,
1243 S.H., Tai, P.Y., Berrett, K., et al. (2016). Discordant Haplotype Sequencing Identifies Functional
1244 Variants at the 2q33 Breast Cancer Risk Locus. Cancer Res *76*, 1916-1925. 10.1158/0008-
1245 5472.CAN-15-1629.
1246 56. Long, E., Yin, J., Shin, J.H., Li, Y., Li, B., Kane, A., Patel, H., Sun, X., Wang, C., Luong, T., et al.
1247 (2024). Context-aware single-cell multiomics approach identifies cell-type-specific lung cancer
1248 susceptibility genes. Nat Commun *15*, 7995. 10.1038/s41467-024-52356-9.
1249 57. Choi, E., and Cho, C. (2011). Expression of a sperm flagellum component encoded by the
1250 Als2cr12 gene. Gene Expr Patterns *11*, 327-333. 10.1016/j.gep.2011.03.003.

1251 58. Leyva-Diaz, E., Stefanakis, N., Carrera, I., Glenwinkel, L., Wang, G., Driscoll, M., and Hobert, O.
1252 (2017). Silencing of Repetitive DNA Is Controlled by a Member of an Unusual Caenorhabditis
1253 elegans Gene Family. Genetics *207*, 529-545. 10.1534/genetics.117.300134.
1254 59. Reddy, K.C., Dror, T., Sowa, J.N., Panek, J., Chen, K., Lim, E.S., Wang, D., and Troemel, E.R. (2017).
1255 An Intracellular Pathogen Response Pathway Promotes Proteostasis in C. elegans. Curr Biol *27*,
1256 3544-3553 e3545. 10.1016/j.cub.2017.10.009.
1257 60. Kischkel, F.C., Lawrence, D.A., Tinel, A., LeBlanc, H., Virmani, A., Schow, P., Gazdar, A., Blenis, J.,
1258 Arnott, D., and Ashkenazi, A. (2001). Death receptor recruitment of endogenous caspase-10 and
1259 apoptosis initiation in the absence of caspase-8. J Biol Chem *276*, 46639-46646.
1260 10.1074/jbc.M105102200.
1261 61. Sprick, M.R., Rieser, E., Stahl, H., Grosse-Wilde, A., Weigand, M.A., and Walczak, H. (2002).
1262 Caspase-10 is recruited to and activated at the native TRAIL and CD95 death-inducing signalling
1263 complexes in a FADD-dependent manner but can not functionally substitute caspase-8. EMBO J
1264 *21*, 4520-4530. 10.1093/emboj/cdf441.
1265 62. Wang, J., Chun, H.J., Wong, W., Spencer, D.M., and Lenardo, M.J. (2001). Caspase-10 is an
1266 initiator caspase in death receptor signaling. Proc Natl Acad Sci U S A *98*, 13884-13888.
1267 10.1073/pnas.241358198.
1268 63. Fischer, U., Stroh, C., and Schulze-Osthoff, K. (2006). Unique and overlapping substrate
1269 specificities of caspase-8 and caspase-10. Oncogene *25*, 152-159. 10.1038/sj.onc.1209015.
1270 64. Horn, S., Hughes, M.A., Schilling, R., Sticht, C., Tenev, T., Ploesser, M., Meier, P., Sprick, M.R.,
1271 MacFarlane, M., and Leverkus, M. (2017). Caspase-10 Negatively Regulates Caspase-8-Mediated
1272 Cell Death, Switching the Response to CD95L in Favor of NF-kappaB Activation and Cell Survival.
1273 Cell Rep *19*, 785-797. 10.1016/j.celrep.2017.04.010.
1274 65. Irmler, M., Thome, M., Hahne, M., Schneider, P., Hofmann, K., Steiner, V., Bodmer, J.L., Schroter,
1275 M., Burns, K., Mattmann, C., et al. (1997). Inhibition of death receptor signals by cellular FLIP.
1276 Nature *388*, 190-195. 10.1038/40657.
1277 66. Humphreys, L., Espona-Fiedler, M., and Longley, D.B. (2018). FLIP as a therapeutic target in
1278 cancer. FEBS J *285*, 4104-4123. 10.1111/febs.14523.
1279 67. Leverkus, M., Neumann, M., Mengling, T., Rauch, C.T., Brocker, E.B., Krammer, P.H., and
1280 Walczak, H. (2000). Regulation of tumor necrosis factor-related apoptosis-inducing ligand
1281 sensitivity in primary and transformed human keratinocytes. Cancer Res *60*, 553-559.
1282 68. Wachter, T., Sprick, M., Hausmann, D., Kerstan, A., McPherson, K., Stassi, G., Brocker, E.B.,
1283 Walczak, H., and Leverkus, M. (2004). cFLIPL inhibits tumor necrosis factor-related apoptosis-
1284 inducing ligand-mediated NF-kappaB activation at the death-inducing signaling complex in
1285 human keratinocytes. J Biol Chem *279*, 52824-52834. 10.1074/jbc.M409554200.
1286 69. Siegmund, D., Hadwiger, P., Pfizenmaier, K., Vornlocher, H.P., and Wajant, H. (2002). Selective
1287 inhibition of FLICE-like inhibitory protein expression with small interfering RNA oligonucleotides
1288 is sufficient to sensitize tumor cells for TRAIL-induced apoptosis. Mol Med *8*, 725-732.
1289 70. Zhang, T., Klein, A., Sang, J., Choi, J., and Brown, K.M. (2022). ezQTL: A Web Platform for
1290 Interactive Visualization and Colocalization of Quantitative Trait Loci and GWAS. bioRxiv.
1291 71. Weissbrod, O., Hormozdiari, F., Benner, C., Cui, R., Ulirsch, J., Gazal, S., Schoech, A.P., van de
1292 Geijn, B., Reshef, Y., Marquez-Luna, C., et al. (2020). Functionally informed fine-mapping and
1293 polygenic localization of complex trait heritability. Nat Genet *52*, 1355-1363. 10.1038/s41588-
1294 020-00735-5.
1295 72. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general
1296 framework for estimating the relative pathogenicity of human genetic variants. Nat Genet *46*,
1297 310-315. 10.1038/ng.2892.

1298 73. Quang, D., Chen, Y., and Xie, X. (2015). DANN: a deep learning approach for annotating the
1299 pathogenicity of genetic variants. Bioinformatics *31*, 761-763. 10.1093/bioinformatics/btu703.
1300 74. Shihab, H.A., Rogers, M.F., Gough, J., Mort, M., Cooper, D.N., Day, I.N., Gaunt, T.R., and
1301 Campbell, C. (2015). An integrative approach to predicting the functional effects of non-coding
1302 and coding sequence variation. Bioinformatics *31*, 1536-1543. 10.1093/bioinformatics/btv009.
1303 75. Fu, Y., Liu, Z., Lou, S., Bedford, J., Mu, X.J., Yip, K.Y., Khurana, E., and Gerstein, M. (2014).
1304 FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. Genome Biol *15*,
1305 480. 10.1186/s13059-014-0480-5.
1306 76. Ritchie, G.R., Dunham, I., Zeggini, E., and Flicek, P. (2014). Functional annotation of noncoding
1307 sequence variants. Nat Methods *11*, 294-296. 10.1038/nmeth.2832.
1308 77. Drozdetskiy, A., Cole, C., Procter, J., and Barton, G.J. (2015). JPred4: a protein secondary
1309 structure prediction server. Nucleic Acids Res *43*, W389-394. 10.1093/nar/gkv332.
1310 78. Cheng, J., Novati, G., Pan, J., Bycroft, C., Zemgulyte, A., Applebaum, T., Pritzel, A., Wong, L.H.,
1311 Zielinski, M., Sargeant, T., et al. (2023). Accurate proteome-wide missense variant effect
1312 prediction with AlphaMissense. Science *381*, eadg7492. 10.1126/science.adg7492.
1313 79. Delaneau, O., Ongen, H., Brown, A.A., Fort, A., Panousis, N.I., and Dermitzakis, E.T. (2017). A
1314 complete tool set for molecular QTL discovery and analysis. Nat Commun *8*, 15452.
1315 10.1038/ncomms15452.
1316 80. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and
1317 Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.
1318 10.1093/bioinformatics/bts635.
1319 81. Li, Y.I., Knowles, D.A., Humphrey, J., Barbeira, A.N., Dickinson, S.P., Im, H.K., and Pritchard, J.K.
1320 (2018). Annotation-free quantification of RNA splicing using LeafCutter. Nat Genet *50*, 151-158.
1321 10.1038/s41588-017-0004-9.
1322 82. Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., and Andrews,
1323 S. (2015). HiCUP: pipeline for mapping and processing Hi-C data. F1000Res *4*, 1310.
1324 10.12688/f1000research.7334.1.
1325 83. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat
1326 Methods *9*, 357-359. 10.1038/nmeth.1923.
1327 84. Langmead, B., Wilks, C., Antonescu, V., and Charles, R. (2019). Scaling read aligners to hundreds
1328 of threads on general-purpose processors. Bioinformatics *35*, 421-432.
1329 10.1093/bioinformatics/bty648.
1330 85. Cairns, J., Freire-Pritchett, P., Wingett, S.W., Varnai, C., Dimond, A., Plagnol, V., Zerbino, D.,
1331 Schoenfelder, S., Javierre, B.M., Osborne, C., et al. (2016). CHiCAGO: robust detection of DNA
1332 looping interactions in Capture Hi-C data. Genome Biol *17*, 127. 10.1186/s13059-016-0992-2.
1333 86. Li, D., Hsu, S., Purushotham, D., Sears, R.L., and Wang, T. (2019). WashU Epigenome Browser
1334 update 2019. Nucleic Acids Res *47*, W158-W165. 10.1093/nar/gkz348.
1335 87. Naumova, N., Smith, E.M., Zhan, Y., and Dekker, J. (2012). Analysis of long-range chromatin
1336 interactions using Chromosome Conformation Capture. Methods *58*, 192-203.
1337 10.1016/j.ymeth.2012.07.022.
1338 88. Spruijt, C.G., Baymaz, H.I., and Vermeulen, M. (2013). Identifying specific protein-DNA
1339 interactions using SILAC-based quantitative proteomics. Methods Mol Biol *977*, 137-157.
1340 10.1007/978-1-62703-284-1_11.
1341 89. Makowski, M.M., Willems, E., Fang, J., Choi, J., Zhang, T., Jansen, P.W., Brown, K.M., and
1342 Vermeulen, M. (2016). An interaction proteomics survey of transcription factor binding at
1343 recurrent TERT promoter mutations. Proteomics *16*, 417-426. 10.1002/pmic.201500327.

1344 90. Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized
1345 p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*,
1346 1367-1372. 10.1038/nbt.1511.
1347 91. Nsengimana, J., Laye, J., Filia, A., O'Shea, S., Muralidhar, S., Pozniak, J., Droop, A., Chan, M.,
1348 Walker, C., Parkinson, L., et al. (2018). beta-Catenin-mediated immune evasion pathway
1349 frequently operates in primary cutaneous melanomas. J Clin Invest *128*, 2048-2063.
1350 10.1172/JCI95351.
1351 92. Du, P., Kibbe, W.A., and Lin, S.M. (2008). lumi: a pipeline for processing Illumina microarray.
1352 Bioinformatics *24*, 1547-1548. 10.1093/bioinformatics/btn224.

1353