Functional annotation and investigation of the 10q24.33 melanoma risk locus identifies a common variant that influences transcriptional regulation of *OBFC1*

Antonella Cardinale[1,2;3], Sueva Cantalupo[1;2], Vito Alessandro Lasorsa[1;2], Annalaura Montella[1;2], Mariangela Succoio[2], Michiel Vermeulen[4], Matteo Esposito[1], Marianna Avitabile[1;2], Daniela Formicola[1,2;9], Massimiliano Scalvenzi[5], Fabrizio Ayla[6], Nicola Zambrano[1;2], Matthew H Law[7], Mark M Iles[8], Mai Xu[10], Kevin M Brown[10], Achille Iolascon[1,2], Mario Capasso[1;2]

[1]Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Università degli Studi di Napoli Federico II, Naples 80136, Italy

[2]CEINGE Biotecnologie Avanzate, Naples 80145, Italy

[3]Department of Pediatric Hematology and Oncology, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy

[4]Department of Molecular Biology, Radboud Institute for Molecular Life Sciences, Oncode Institute, Radboud University, Nijmegen, The Netherlands

[5]Dipartimento di Medicina clinica e Chirurgia, Università degli Studi di Napoli Federico II, Naples 80136, Italy, [6]National Cancer Institute, 'Fondazione G. Pascale'

[7]Statistical Genetics, QIMR Berghofer Medical Research Institute Brisbane, Queensland 4006, Australia

[8]Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK

[9]SOC Genetica Medica, Azienda Ospedaliera Universitaria Meyer, Firenze 50139, Italy

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

Corresponding Author

Mario Capasso
Università degli Studi di Napoli "Federico II", via Gaetano Salvatore 486, 80145 Napoli;
e-mail: mario.capasso@unina.it
Phone: +390813737889

ABSTRACT

The 10q24.33 locus is known to be associated with susceptibility to cutaneous malignant melanoma (CMM), but the mechanisms underlying this association have been not extensively investigated.

Here, we carried out an integrative genomic analysis of 10q24.33 using *in-silico* epigenomic annotations of ChIP-seq data with in vitro reporter gene assays to identify regulatory variants at 10q24 locus. We found two putative functional SNPs enriched in an active enhancer and one in promoter of *OBFC1* in neural crest cells and CMM cells, but only rs2995264 induced enhancer activity. The minor allele G of rs2995264 correlated with lower expression in 470 tumors and was confirmed to increase the risk of CMM development in a cohort of 484 CMM cases and 1801 controls of Italian origin. HiC and 3C experiments validated the interaction between rs2995264 and the promoter region of *OBFC1* and an isogenic model characterized by *CRISPR-Cas9* deletion of the enhancer-SNP region confirmed the potential regulatory activity of rs2995264 on *OBFC1* transcription. We demonstrated that the presence of G-rs2995264 risk allele reduced the binding affinity of homeobox transcriptional factor *MEOX2*. Biologic investigations showed that transient inhibition of *OBFC1* promoted cells viability in CMM cell lines with protective genotype of rs2995264 and high basal expression of *OBFC1* but not in those with risk genotype. Clinically, high levels of *OBFC1* expression associated with histologically favorable tumors. Finally, preliminary results suggested the potential effect of *OBFC1* on the telomerase activity in tumorigenic conditions.

Our results support the hypothesis that decreased expression of *OBFC1* gene through functional hereditable DNA variation can contribute to malignant transformation of normal cutaneous cells.

Introduction

Cutaneous malignant melanoma (CMM) is a cancer of transformed neural crest (NC) derived melanocytes, pigment-producing cells, where both genetic and environmental factors are involved. Major environmental risk factors include a personal and familial history of the disease, cutaneous and pigmentary characteristics, sun exposure and reactions to sun exposure. Phenotypic risk factors are likely to be genetically determined [1]. Besides the rare, deleterious mutations in genes such as *CDKN2A* and *CDK4*, which confer a high CMM risk in their carriers[1; 2], common variants with low effect size are likely to be also involved in melanoma susceptibility.

Genome Wide Association Studies (GWAS) approach have led to insights into the architecture of disease susceptibility through the identification of novel disease-causing genes and mechanisms improving our knowledge of the complex disease etiology.[3] Over the past decade, more than 430 cancer associated common variants at 262 distinct genomic regions have been successfully identified by GWAS[4]**.** Most of these alterations resides in non-coding portion of the human genome and may have regulatory consequences on cancer susceptibility[5]. However, the functional role of the identified risk loci in cancer pathogenesis remains poorly investigated.

Large GWASs have identified several loci associated with CMM risk in the general population: *PARP1*, *SLC45A2*, *TYR*, *MC1R*, *ASIP*, CDKN2A-*MTAP*, *CYP1B1*, *PLA2G6*, *TERT, ATM, ARNT-SETDB1, CDKAL1, OCA2, CCND1, AGR3, CASP8, FTO, CDK10, TMEM38B, OBFC1* and *MX2*[6; 7]. Despite GWAS facilitates the initial identification of a risk locus, this approach presents some limitations mainly due to the difficulty to discern the causal variants. Moreover, common

genetic loci that usually are likely hidden among signals discarded by the multiple testing correction represents a restrictive step of GWAS analytical process. [3]

Post-GWAS strategies are trying to overcome these limitations by leveraging different approaches such as imputation analysis, next generation sequencing (NGS), fine mapping and cross phenotype meta-analysis (CPMA)[8]. Leveraging the concept of pleiotropy, a cross phenotype meta-analysis of CMM and nevus GWAS demonstrated that several risk loci might act through nevus development, in line with clinical evidence[9]. Moreover, another study identified additional risk loci for CMM associated with telomere length or located near prominent telomere maintenance genes, including *POT1, TERC, RTEL1, MPHOSPH6* and *OBFC1.* [10] Recently, we performed a cross-disease meta-analysis of neuroblastoma e CMM GWAS, in which we found diverse neuroblastoma-CMM cross-associated loci. Among these, we further confirmed the association of the previously identified[6] 10q24.33 CMM risk locus (index SNP rs11591710). [11] However, at this risk locus, the identification of the causal genetic variants (i.e., those that actually contribute to the development of CMM) and the detection of the genes whose function is influenced by the same causal variants remain to be established.

Here, we performed an integrative genomic analysis of 10q24.33 CMM risk locus that led us to identify a gene-regulatory variant (rs2995264) with enhancer features within intronic region of *OBFC1* gene. *In silico* and *in vitro* studies demonstrated that the G risk allele of rs2995264 SNP correlated with a decreased expression of *OBFC1* gene suggesting its role as tumor-suppressor in CMM, which has been experimentally confirmed in melanoma cell lines. Despite the need for follow-up functional studies, a preliminary model of pathways potentially important for the CMM development is emerging through this approach.

Methods

Identification of causal variant at 10q24.33

To identify potential functional SNPs we used multiple sources of *in silico* functional annotation from public databases, as detailed in **Supplementary Information**. We obtained the genome binding/occupancy profiling by high throughput sequencing the epigenetic marker H3K27ac in 7 human derived-melanoma (GSE75352) and 2 human neural crest cell (hNCC) lines (GSE90683) through the National Center for Biotechnology Information (NCBI). We first selected the variants in Linkage Disequilibrium (LD) with the rs11591710 lead SNP ($0.5 < r2 \leq 1$) (total including the lead SNP, n= 33) in European population using LDlink (analysistools.cancer.gov/LDlink)_and occurring in regulatory genomic regions in CMM and hNCC (n = 9).

CMM replication in an Italian cohort

The genomic DNA of CMM patients was extracted from peripheral blood using a Maxwell® RSC Blood DNA Kit (Promega, Madison, WI, USA), and DNA concentration and purity were evaluated using a NanoDrop™ 8000 Spectrophotometer. The rs2488001 in LD (r^2=0.98) with rs2995264 SNP was typed by TaqMan® SNP Genotyping Assay (Applied Biosystems by Thermo Fisher Scientific, Waltham, MA, USA) in an Italian cohort of 484 CMM cases and 1801 controls. To monitor quality control, three DNA samples per genotype were genotyped by Sanger sequencing (3730 DNA analyzer, Applied Biosystems) and included in each 384-well reaction plate; genotype concordance was 100%. This study was approved by the Ethics Committee of the Medical University of Naples

Statistical analysis

A comparison of the genotypic and allelic frequencies between the groups was performed using the chi square test. Statistical significance was established at $P < 0.05$. Hardy–Weinberg

equilibrium was evaluated using the goodness-of-fit chi-square test in control and case subjects ($P > 0.05$). Conditional analysis was performed with GCTA software[12] using the summary statistics of melanoma GWAS[6] including 12874 cases and 23203 controls. The association plot was generated using LocusZoom. [13] All LD calculations (r2 and D') were performed using the LDlink suite (https://ldlink.nci.nih.gov/?tab=home) and data from the 1000 Genomes Project European ancestry populations.

HiC data analysis

As detailed in **Supplementary Information**, the sequencing was performed on an Illumina® HiSeq platform. Paired-end reads with length of 150bp were mapped to the reference genome (build hg19/GRCH37) with Bowtie2[14]. The alignment BAM file was then filtered to remove duplicates, re-ligation or self-circularization artifacts that can be introduced during Hi-C library preparation. Then we used HiCExplorer tool v3.5.1[14] to (i) build the interaction matrix at a resolution of 10Kb (bin size=10Kb); (ii) normalize the observed interaction matrix; (iii) determine Topologically Associating Domains (TADs, self-interacting genome regions) and their boundaries; and (iv) plot the results. Subsequently, we extended our region of interest containing LD SNPSs of 1Mb up- and down-stream and calculated the statistical significance of the interactions between bins with the FitHiC v2.0.7 program[15]. P-values were corrected for multiple tests by Benjamini-Hotchberg method (False Discovery Rate, FDR) and the cutoff was set at 1%. Finally, we annotated those bins with ANNOVAR[16] in order to map genomic bins to gene coordinates.

In vitro functional study

A detailed description of the Luciferase reporter assays and the experiments performed to evaluate the *OBFC1* effect on CMM cell line phenotype is reported in **Supplementary Information**.

Chromosome Conformation Capture

3C procedure is described more in details in **Supplementary Information**. CMM cells were used, A375 and CJM respectively. In order to obtain a negative control for the 3C analysis, in addition to the non-crosslinked sample, we selected a fragment chr10:105693245-105701218 that did not show any characteristic that can be associated with a regulatory region and for which is not expected an interaction with the promoter region. To guarantee the correct setting of the PCR experimental conditions it has been also necessary to produce, through overlapping PCR, a PCR product consisting of the restriction fragments corresponding to those of the intronic region and promoter. PCR products were resolved on 2% agarose gels. In order to normalize 3C-PCR signals, we used a loading control (internal primers located in the *GAPDH* gene[17]). The amount of DNA input was first titrated, and bands analyzed semi-quantitatively using ImageJ software; the background was subtracted, and data normalized to an internal region unaffected by the restriction digest (LC region) [18; 19]. Two biological replicates were prepared and analyzed in three technical repeats.

CRISPR-based enhancer deletion

We used the CRISPR/Cas9 system to generate HEK 293 isogenic cell line with a deletion of the enhancer region (hg19/chr10:105,668,100-105,669,000), as confirmed by the peak in H3K27ac Chip-Seq of HEK293T cell line (Encode project track ENCR000FCH). To delete this enhancer element, following Bauer et al and Ran, F. A. et al guidelines[20; 21], two pSpCas9(BB)-2A-GFP (PX458) vectors expressing Cas9 (Addgene) and the desired single guide RNAs

(sgRNAs) (designed using CRISPOR, CHOPCHOP and CRISPR Design tool), were co-transfected into HEK293T cell line using Transfectin Lipid Reagent (Bio-Rad). sgRNA2 (5'-ACAGGCCTGCGGTGAGTCAG-3') and sgRNA3 (5'- CGGGATGAGTCAGTGCGAGC-3') were paired with sgRNA6 (5'- CAGCTATGGGCAGTACACTG-3') to make clones with different genomic deletions (720 bp and 756 bp, respectively). After 48h of transfection, GFP-positive single cells were FACS-sorted by size into 96-well plates. To identify and distinguish both mono-allelic and bi-allelic deletions, a PCR using two primer pairs flanking the single guide RNA (sgRNA) cleavage sites (Primer-OUT-Forward: 5'-TGCGAGGTCATTCTGGTCTTG-3'; Primer-OUT-Reverse: 5'-AACTTTGTGACCAAGAGCGT-3') was performed with KAPA HiFi HotStart PCR Kit (Roche), following manufacturer's instructions. Other two primers falling into the deleted sequence (Primer-IN-Forward: 5'- GTGAGTCAGGGGAAGCAGAA-3'; Primer-OUT-Reverse: 5'-TCCAGCTATGGGCAGTACAC-3') were used to confirm deletion occurred.

These two sets of primer were used to screen single cell-deriving clones to evaluate which ones was edited correctly.

Affinity Purification Mass Spectrometry (AP-MS)

Quantitative AP-MS following SNP DNA pulldown was performed on the basis of procedures described by Choi et al[22-24]. For DNA pulldown, 500 pmol of annealed, forward-strand 5′-biotinylated oligonucleotide probe was coupled to Streptavidin Sepharose beads (GE Healthcare). rs2995264-A and 2995264-G allele probe sequences are: 5'-TGTACTTTCT***G***TTTCAAAAGA-3' ; 5'-TGTACTTTCT***A***TTTCAAAAGA-3'.

Chromatin immunoprecipitation (ChiP)

A375 and UACC1816, both in basal condition and after transfection with MEOX2 origene expressing plasmid pCMV6-ENTRY MEOX2 were used and a detailed description of the procedure is shown in **Supplementary Information**. In order to validate MEOX2-ChIP reaction, MEOX2 binding sites in individual genes were identified using JASPAR Web Tools

(http://jaspar.genereg.net/). The promoter of p21 gene known for MEOX2 binding[25] was used as positive control and analyzed using (qRT)- PCR. NFT3 genomic region (chr12:5542580-5542721) with the highest H3K27me3 peak and lowest H3K27ac Chip-seq peak in skin keratinocytes experiments, available in Encode (ENCSR621FNM; ENCSR736ZEG; ENCSR709ABP; ENCSR793NQA), was used as negative control (Neg Ctrl).

Cell culture

The human A375 cell line and the human CJM cell line were donated from Professor Nick Hayward (QIMR Berghofer Medical Research Institute, Australia). The human UACC1816 cell line was donated from Dr. Kevin Brown (Translational Genomics Research Institute TGen, Arizona USA). The human HEK293T cell lines were obtained from the American Type Culture Collection (CRL-3216). HEK293T and A375 cell lines were grown in Dulbecco's Modified Eagle Medium (DMEM; Sigma); CJM cell line was grown in RPMI-1640 Media (Sigma); UACC1816 cell line was grown in RPMI-1640 Media (Sigma) supplemented with 25mM HEPES (Sigma). The mediums were supplemented with 10% heat-inactivated FBS (Sigma), 1 mmol/L L-glutamine, penicillin (100 U/mL), and streptomycin (100mg/mL; Invitrogen). The cells were cultured at 37° C, 5% CO2 in a humidified atmosphere. The cell lines used for all the experiments were re-authenticated and tested as mycoplasma-free. Early-passage cells were used and cumulative culture length was less than 3 months after resuscitation. Total cellular RNA extraction, reverse transcription and Quantitative Real Time (qRT)-PCR were performed as previously described[26].

Results

Integrative genomic analysis and functional annotation of candidate SNPs.

We previously found the risk locus 10q24.33, with the rs11591710 index SNP in an intronic region of *OBFC1,* reaching genome-wide significance of association in a cross-disease meta-analysis of neuroblastoma e CMM GWAS [11]. Particularly, the minor allele of rs11591710 resulted to be associated with increased risk of developing CMM[11]. Therefore, we decided to further functionally analyze the 10q24.33 locus. To highlight potentially functional variants, we annotated 32 SNPs in LD ($0.5<r^2<1$) with the lead SNP rs11591710 with the regulatory elements super-enhancer, enhancer and promoter histone obtained by an analysis of H3K27ac ChIP-Seq data (**Supplementary Information**) derived from 7 human CMM cell lines and 2 hNC cell lines (GSE90683) deposited in GEO database. To prioritize causal functional variants in CMM, we first selected those SNPs overlapping at least 1 histone marker in CMM and hNCC. The analysis pipeline allowed us to classify acetylation peaks as super-enhancers (SE), enhancers or promoters **(Supplementary Table 1; see Supplementary Information**). The three top SNPs rs34685262 (r^2=0.77 with the index SNP), rs2995264 (r^2=0.68 with the index SNP) and rs35176048 (r^2=0.68 with the index SNP) fell in the same H3K27ac peak, located in the intron 3 of *OBFC1*, called as SE or enhancer in 9 or 8 cell lines whereas rs4387287 (r^2= 0.65 respect to the index SNP) was annotated in the promoter region of the same gene in 8 cell lines (**Fig. 1 A and Supplementary Table 1)**. Among the three top SNPs, the rs35176048 SNP resulted to be located in H3K27ac but not in H3K4me1 peak in melanoma and other ENCODE cell lines (**Supplementary Table 1**). The two missense variants rs10786775 and rs2487999 were predicted to be benign (**Supplementary Table 1**). Based on these observations rs35176048, rs10786775, rs2487999 SNPs were not considered for further analysis. Instead, the SNPs rs34685262, rs2995264 and rs4387287, located in a highly predicted functional regions (**Fig. 1 A**), were tested for the induction of enhancer activity of these three variants through luciferase reporter assay in HEK293T and A375 cells in order to validate their regulatory properties (**Supplementary Fig. 1**). Only the construct

containing rs2995264-G risk allele induced a significant decrease of enhancer activity compared to the construct containing rs2995264-A reference allele (**Supplementary Fig. 1**).

In GWAS, including 12874 cases and 23203 controls[6], the allele G of SNP rs2995264 resulted to protect against CMM development ($P=8.5x10^{-7}$, OR=0.87) (**Supplementary Table 2**). We sought to replicate this genetic association (using the SNP rs2488001 in high LD, r2=0.98, with rs2995264) in an independent cohort of 484 CMM cases and 1801 controls of Italian origin performing PCR-based genotyping. The minor allele G confirmed to increase the risk of CMM onset in the Italian population (**Supplementary Table 3**). Moreover, to investigate whether more than one association signal may exist at 10q24.33, we conditioned our analysis of locus 10q24.33 on the SNP rs2995264 using summary statistics of CMM GWAS[6]. We confirmed that no evidence for a separate association signal was observed at 10q24.33 locus (**Supplementary Fig. 2**).

Based on the above-reported results, we decided to further functionally characterize the SNP rs2995264, which resulted to be the most significant functional SNP at 10q24.33 and located in an enhancer region of CMM and hNCC cell lines.

Evaluation of rs2995264 allele-specific enhancer activity toward *OBFC1* promoter.

To evaluate functional role of rs2995264, we verified if the candidate variant affected gene expression, by performing cis-expression Quantitative Trait Loci (eQTL) analysis. The analysis of gene expression variation using 470 genome-wide expression and SNP arrays of CMM tumors demonstrated that the SNP rs2995264 altered expression of *OBFC1* gene. Particularly, the presence of the G risk allele significantly correlated with decreased *OBFC1* mRNA expression (**Fig. 1 B**). These results were further confirmed in skin tissue and cultured

fibroblasts (GTEx portal data, **Supplementary Fig. 3**) and by luciferase assay (**Supplementary Fig. 1**). Since rs2995264 affected *OBFC1* expression and based on the previous evidences, we hypothesized that this SNP is located in an enhancer that physical interacts with *OBFC1* promoter. To demonstrate this assumption, we interrogated the public database Enhancer Atlas[27] that provides a list of long-range chromatin interaction partners for the queried locus obtained from 105 different human cell/tissue types. According to the analysis of Enhancer Atlas database, in the human foreskin tissue, *OBFC1* promoter interacts with the enhancer where the SNP rs2995264 is located (**Supplementary Fig. 4**). Subsequently, we confirmed this interaction using HiC sequencing data obtained on the COLO-829 CMM cell line to (**Fig. 1 C**). Our analysis showed that the genomic bin (10 Kb) containing rs2995264 significantly interacted with a total of 11 genes. In particular, the SNP rs2995264 strongly interacted with both the up-stream (distance = 2058 bp; FDR = 1.76×10^{-38}) and the down-stream (distance = 7327 bp; FDR = 1.26×10^{-16}) regions of *OBFC1* (**Fig. 1 C and Supplementary Table 4**).

To further validate physical interactions between polymorphic enhancer containing rs2995264 and *OBFC1* promoter, we performed Chromosome Conformation Capture (3C) analysis in A375 and CJM melanoma cells. Sequences that are held nearby by genic regulation factors in chromosomal structure, but which might be far distant from one another on the linear chromosome, can be ligated and subsequently detected by PCR. A schematic representation of our 3C experiment is given in **Fig. 2 A**. In addition to the enhancer/SNP and the promoter region of *OBFC1* gene, we examined a genomic region physically opposite to the enhancer containing rs2995264 and lacking typical characteristics of a regulatory element as a negative control (mock region). An artificial template consisting of promoter region linked to mock region and obtained by overlap extension PCR was used as a positive control to

validate technical set up (**Supplementary Fig. 5**). Specific products were amplified in both cell lines with primers targeting the restriction fragments of enhancer/SNP and *OBFC1* promoter, in samples that had been cross-linked, but not in samples that were not cross-linked (**Fig. 2 B**). Ultimately, the results of 3C in the analyzed CMM cell lines, confirmed the interaction between the *OBFC1* promoter and the regulatory element associated to the genetic variant rs2995264.

To further illustrate the importance of this regulatory element in inducing expression of *OBFC1* gene, we deleted the enhancer region containing rs2995264 by CRISPR/Cas9 system (**Fig. 2 C**) in HEK293T cell line that is frequently used for genome editing due to its high efficacy. The regulatory element was targeted by two single guides RNA (sgRNAs) pairs that efficiently deleted the region overlapping the enhancer (**Supplementary Fig. 6**). We confirmed that the homozygous deletion of the enhancer region decreased *OBFC1* expression levels compared to wild type HEK293T cells (**Fig. 2 D**). Collectively, these results provide strong evidence for the role of the enhancer containing rs2995264 in regulating *OBFC1* gene expression.

To identify proteins that bind the SNP rs2995264 in an allele-preferential manner, we used affinity purification mass spectrometry (AP-MS)[22; 28]: DNA pull down using nuclear A375 and UACC1816 extracts identified rs2995264-A and rs2995264-G preferential interactors, and MEOX2 resulted as most significant interactor (**Fig. 2 E**). In view of this, ChiP experiments were performed to determine whether the sequence containing rs2995264 actually binds MEOX2 nuclear protein in allele preferential manner in A375 (A/A) and UACC1816 (G/G) cell lines in basal conditions and after 48h of transfection with MEOX2 expressing plasmid (**Supplementary Fig. 7**). **Figure 2 F** indicates a MEOX2 binding enrichment in the presence of

rs2995264-A allele. Accordingly, MEOX2 transcription factor recognizes and binds specifically to the rs2995264-A allele, with greater affinity, compared with the rs2995264-G allele. Altogether, these data suggest that the rs2995264 polymorphism alters the binding of MEOX2 transcription factor, possibly leading to alteration of the *OBFC1* transcriptional machinery.

***OBFC1* has tumor suppressor effect in cutaneous malignant melanoma.**

To unravel the potential *OBFC1* contribution to CMM development, we tested the gene expression in three independent mRNA expression array data sets. Expression profile of *OBFC1* was significantly lower in CMM unfavorable histology when compared to histologically benign tumors (nevi) (**Fig. 3 A)** and in metastatic melanoma when compared to primary melanomas (**Fig. 3 B and C**). We found no significant correlation between *OBFC1* expression and patient survival (**Supplementary Fig. 8**). Together, these data provide evidence that *OBFC1* might play a biological role in CMM initiation rather than progression. So, we planned to test whether decreased *OBFC1* levels could lead to cellular transformation required in tumor onset. We selected A375 and CJM (A/A rs2995264 genotype) cell lines with high *OBFC1* expression and UACC1816 (G/G rs2995264 genotype) with low *OBFC1* expression, validated by real time PCR and Western Blot analysis (**Supplementary Fig. 9 A-B**). We thus examined the consequence of *OBFC1* knocked down by using short interfering RNA (siRNA) against *OBFC1*. Compared with scrambled siRNA used as a control, three different siRNA (siRNA_A, siRNA_B, siRNA_C) specific for *OBFC1*, significantly reduced *OBFC1* mRNA and protein levels at 48h post transfection (**Fig. 3 D-E-F**). Knockdown of OBFC1 markedly increased cell viability in A375 and CJM cell lines, carrying the AA-rs2995264 protective genotype and high *OBFC1* expression, compared to the control cells (siScrambled) (A375: T24 $P \leq 1X10^{-3}$; T48 $P \leq 2.1X10^{-6}$; T72 $P \leq 1.2X10^{-7}$; CJM T24 $P \leq 6X10^{-3}$; T48 $P \leq 7X10^{-3}$; T72 $P \leq 5.5X10^{-7}$), whereas in UACC1816, with GG-rs2995264 risk genotype and low basal OBFC1 expression, we observed

non-substantial differences in cell viability after OBFC1 silencing (**Fig. 3 G-H-I**). Importantly, these findings indicate that *OBFC1* has potential tumor suppressor effect in CMM and the lack of its expression due to disease-predisposing alleles may contribute to CMM progression by promoting cells proliferation.

***OBFC1* functions in telomeres maintenance by regulating telomerase activity.**

The 10q24.33 locus has been associated with telomere length and cutaneous malignant melanoma traits.[6; 29] As OBFC1 belongs to CST complex which turns off telomerase activity by inhibiting its binding to telomeric DNA[30], we hypothesized that decreased *OBFC1* expression could predispose to CMM risk allowing telomere maintenance. To demonstrate our assumptions, we measured telomerase activity after *OBFC1* transient silencing in melanoma cell lines (**Supplementary Fig. 10 A-B**) with a PCR-based assay that permitted quantitation of telomerase enzymatic activity. Compared to control cells (siRNA Scrambled), *OBFC1* knockdown markedly increased telomerase activity in melanoma cell lines, confirming loss of CST complex capability to suppress telomerase access to lengthening telomeres (**Fig. 4 A**). Literature data showed that endogenous telomerase action at telomeres is restricted to the cell cycle S phase[31-33]. Consistent with these findings, we expected that increased telomerase activity after *OBFC1* depletion could coincide with a higher percentage of cells in S phase. Here, we examined the proportions of A375 and CJM, cells with high *OBFC1* expression, at each stage of the cell cycle by flow cytometry after siRNA *OBFC1* treatment (**Supplementary Fig. 10 C-F**) and we found a direct proportionality between levels of telomerase activity and the percentage of cells in S phase (**Fig. 4 B-C**). It has been known that the cyclin gene with the highest transcription rate during S phase is *CCNA2*[34]. So, we decided to reinforce the previous evidence of S phase cells accumulation with evaluation of *CCNA2* mRNA levels after *OBFC1* depletion. The mRNA of *CCNA2* was higher in *OBFC1* silenced CMM cells, as expected (**Fig. 4 D**).

Discussion

OBFC1 locus (10q24.33) has been previously identified as CMM susceptibility locus[6; 11]. However, at this locus, most of the functional variant(s) responsible of biological mechanisms accounting for the risk and genes involved in CMM pathogenesis have to be better elucidated.

To functional characterize the CMM risk variant(s) at 10q24.33 and to determine the genes affected by the same variants, we have carried out an integrative genomic analysis of 10q24.33 locus that reached significance in our neuroblastoma-CMM GWAS meta-analysis performed previously[11].

We developed a specific strategy based on epigenomic annotations of large number of CMM and NCC cell lines to identify regulatory variants at 10q24.33 locus, which could affect transcriptional machinery. We found three putative functional SNPs (rs2995264, rs34685262, and rs4387287) enriched in active enhancers of NCC and CMM cells, but only rs2995264 induced enhancer activity. We also confirmed that the minor allele G of rs2995264 associated with CMM development in an independent cohort of Italian origin. HiC and 3C experiments confirmed the interaction between the SNP rs2995264 and the promoter region of *OBFC1.* Moreover, the isogenic model characterized by *CRISPR-Cas9* deletion of the enhancer region containing the SNP rs2995264 confirmed the enhancer regulatory potential on *OBFC1* transcription.

In line with the established rs2995264 functional properties, we have also predicted and validated, *in-vitro,* that G-rs2995264 risk allele decreased the binding affinity of *MEOX2*, a homeobox transcriptional factor that seems to mediate carcinogenesis by altering the normal

mechanisms of angiogenesis and cell proliferation[35-37]. These results provide evidence that a functional DNA variant in the enhancer region of *OBFC1* influences CMM susceptibility.

Our data are strengthened by NHGRI-EBI GWAS catalog, which mentioned both the association of rs2995264 with cutaneous malignant melanoma and the association of four SNPs (rs2487999, rs4387287, rs9420907 and rs9419958) located in *OBFC1* genomic region with telomere length[38], beforehand suggesting an involvement of *OBFC1* gene in tumorigenic transformation.

Furthermore, we demonstrated that the G risk allele correlated with low *OBFC1* expression in CMM tumors, and the decreased *OBFC1* expression correlated with histologically unfavorable melanoma tumors, thus indicating a potential tumor suppressor effect of *OBFC1* in melanoma. Our findings are in accord with the work of Phelan et al. showing that the minor allele of an epithelial ovarian cancer (EOC) predisposing SNP, in complete LD with rs2995264, correlated with low *OBFC1* expression in ovarian cancer tissues[39].

OBFC1 (OB Fold-containing Protein 1), a human homolog of yeast *STN1*, is a subunit of an alpha accessory factor that stimulates the activity of DNA polymerase α primase, the enzyme that initiates DNA replication[40]. *OBFC1* is also known to be a key component of telomere-associated CST complex that binds telomeric single-stranded DNA in vitro and localizes at telomeres in vivo[41]. The contribution of *OBFC1* to cancer susceptibility firstly emerged from the large genetic association analyses of patients with different histotypes of EOC that identified 10q24.33 as a risk locus associated with borderline serous EOC.[39] Our work confirms the involvement of *OBFC1* gene in cancer predisposition demonstrating that transient knockdown of *OBFC1* resulted in significant increase of cells viability in CMM cells

with AA-rs2995264 protective genotype and basal *OBFC1* expression, but had little effect on cells with GG-rs2995264 risk genotype and lower basal *OBFC1* expression.

It is known that in physiological conditions the human CST (*CTC1, OBFC1* and *TEN1*) complex inhibits telomerase activity[42], as demonstrated by the excessive telomerase activity resulted after CST depletion[30]. Additional experimental evidences showed that ectopic over-expression of an *OBFC1* truncation mutation also led to telomere length increase over time[41].

Here we present preliminary data supporting a role of *OBFC1* in telomerase homeostasis, with an observed enhancement of telomerase activity in CMM cells after *OBFC1* depletion, and provide evidence for a putative molecular mechanism that confers genetic susceptibility to melanoma. Additional research efforts and *in-vitro* validation experiments of our data are needed to be performed in future studies in order to elucidate the molecular mechanisms of *OBFC1* in promoting telomere maintenance required to CMM malignant transformation.

In conclusion, our results support the hypothesis that decreased expression of *OBFC1* gene through functional hereditable DNA variation can contribute to malignant transformation of normal cutaneous cells. Moreover, we provide preliminary data suggesting the potential effect of *OBFC1* on the telomerase activity in tumorigenic conditions. This study has demonstrated that post-GWAS strategies are a useful step for the identification of causal functional variants at previous identified cancer risk loci and for the elucidation of the key roles of genes involved in tumor biology.

References

1. Bressac-de-Paillerets, B., Avril, M.F., Chompret, A., and Demenais, F. (2002). Genetic and environmental factors in cutaneous malignant melanoma. Biochimie 84, 67-74.
2. Zuo, L., Weger, J., Yang, Q., Goldstein, A.M., Tucker, M.A., Walker, G.J., Hayward, N., and Dracopoli, N.C. (1996). Germline mutations in the p16INK4a binding domain of CDK4 in familial melanoma. Nature genetics 12, 97-99.
3. Tam, V., Patel, N., Turcotte, M., Bosse, Y., Pare, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. Nat Rev Genet 20, 467-484.
4. Sud, A., Kinnersley, B., and Houlston, R.S. (2017). Genome-wide association studies of cancer: current insights and future perspectives. Nature reviews Cancer 17, 692-704.
5. Liu, N.Q., Ter Huurne, M., Nguyen, L.N., Peng, T., Wang, S.Y., Studd, J.B., Joshi, O., Ongen, H., Bramsen, J.B., Yan, J., et al. (2017). The non-coding variant rs1800734 enhances DCLK3 expression through long-range interaction and promotes colorectal cancer progression. Nature communications 8, 14418.
6. 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. Nature genetics 47, 987-995.
7. Antonopoulou, K., Stefanaki, I., Lill, C.M., Chatzinasiou, F., Kypreou, K.P., Karagianni, F., Athanasiadis, E., Spyrou, G.M., Ioannidis, J.P.A., Bertram, L., et al. (2015). Updated field synopsis and systematic meta-analyses of genetic association studies in cutaneous melanoma: the MelGene database. The Journal of investigative dermatology 135, 1074-1079.
8. Solovieff, N., Cotsapas, C., Lee, P.H., Purcell, S.M., and Smoller, J.W. (2013). Pleiotropy in complex traits: challenges and strategies. Nat Rev Genet 14, 483-495.
9. Duffy, D.L., Zhu, G., Li, X., Sanna, M., Iles, M.M., Jacobs, L.C., Evans, D.M., Yazar, S., Beesley, J., Law, M.H., et al. (2018). Novel pleiotropic risk loci for melanoma and nevus density implicate multiple biological pathways. Nat Commun 9, 4774.
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.
11. Avitabile, M., Succoio, M., Testori, A., Cardinale, A., Vaksman, Z., Lasorsa, V.A., Cantalupo, S., Esposito, M., Cimmino, F., Montella, A., et al. (2020). Neural crest-derived tumor neuroblastoma and melanoma share 1p13.2 as susceptibility locus that shows a long-range interaction with the SLC16A1 gene. Carcinogenesis 41, 284-295.
12. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Genetic Investigation of, A.T.C., Replication, D.I.G., Meta-analysis, C., Madden, P.A., Heath, A.C., Martin, N.G., et al. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet 44, 369-375, S361-363.
13. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336-2337.
14. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357-359.
15. Ay, F., Bailey, T.L., and Noble, W.S. (2014). Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. Genome Res 24, 999-1011.
16. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38, e164.

17. Hagege, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., de Laat, W., and Forne, T. (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat Protoc 2, 1722-1733.
18. Palmer, M.B., Majumder, P., Green, M.R., Wade, P.A., and Boss, J.M. (2007). A 3' enhancer controls snail expression in melanoma cells. Cancer Res 67, 6113-6120.
19. Marasek, P., Dzijak, R., Studenyak, I., Fiserova, J., Ulicna, L., Novak, P., and Hozak, P. (2015). Paxillin-dependent regulation of IGF2 and H19 gene cluster expression. J Cell Sci 128, 3106-3116.
20. Bauer, D.E., Canver, M.C., and Orkin, S.H. (2015). Generation of genomic deletions in mammalian cell lines via CRISPR/Cas9. J Vis Exp, e52118.
21. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8, 2281-2308.
22. Choi, J., Xu, M., Makowski, M.M., Zhang, T., Law, M.H., Kovacs, M.A., Granzhan, A., Kim, W.J., Parikh, H., Gartside, M., et al. (2017). A common intronic variant of PARP1 confers melanoma risk and mediates melanocyte growth via regulation of MITF. Nat Genet 49, 1326-1335.
23. Butter, F., Davison, L., Viturawong, T., Scheibe, M., Vermeulen, M., Todd, J.A., and Mann, M. (2012). Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding. PLoS genetics 8, e1002982.
24. Hubner, N.C., Nguyen, L.N., Hornig, N.C., and Stunnenberg, H.G. (2015). A quantitative proteomics tool to identify DNA-protein interactions in primary cells or blood. J Proteome Res 14, 1315-1329.
25. Chen, Y., Leal, A.D., Patel, S., and Gorski, D.H. (2007). The homeobox gene GAX activates p21WAF1/CIP1 expression in vascular endothelial cells through direct interaction with upstream AT-rich sequences. J Biol Chem 282, 507-517.
26. Capasso, M., Ayala, F., Russo, R., Avvisati, R.A., Asci, R., and Iolascon, A. (2009). A predicted functional single-nucleotide polymorphism of bone morphogenetic protein-4 gene affects mRNA expression and shows a significant association with cutaneous melanoma in Southern Italian population. J Cancer Res Clin Oncol 135, 1799-1807.
27. Gao, T., He, B., Liu, S., Zhu, H., Tan, K., and Qian, J. (2016). EnhancerAtlas: a resource for enhancer annotation and analysis in 105 human cell/tissue types. Bioinformatics 32, 3543-3551.
28. Makowski, M.M., Willems, E., Fang, J., Choi, J., Zhang, T., Jansen, P.W., Brown, K.M., and Vermeulen, M. (2016). An interaction proteomics survey of transcription factor binding at recurrent TERT promoter mutations. Proteomics 16, 417-426.
29. Codd, V., Nelson, C.P., Albrecht, E., Mangino, M., Deelen, J., Buxton, J.L., Hottenga, J.J., Fischer, K., Esko, T., Surakka, I., et al. (2013). Identification of seven loci affecting mean telomere length and their association with disease. Nat Genet 45, 422-427, 427e421-422.
30. Chen, L.Y., Redon, S., and Lingner, J. (2012). The human CST complex is a terminator of telomerase activity. Nature 488, 540-544.
31. Hockemeyer, D., and Collins, K. (2015). Control of telomerase action at human telomeres. Nature structural & molecular biology 22, 848-852.
32. Greider, C.W. (2016). Regulating telomere length from the inside out: the replication fork model. Genes & development 30, 1483-1491.
33. Gallardo, F., Laterreur, N., Wellinger, R.J., and Chartrand, P. (2012). Telomerase caught in the act: united we stand, divided we fall. RNA biology 9, 1139-1143.
34. Yam, C.H., Fung, T.K., and Poon, R.Y. (2002). Cyclin A in cell cycle control and cancer. Cellular and molecular life sciences : CMLS 59, 1317-1326.

35. Tian, L., Tao, Z.Z., Ye, H.P., Li, G.Y., Zhan, Z.F., and Tuo, H.W. (2018). Over-expression of MEOX2 promotes apoptosis through inhibiting the PI3K/Akt pathway in laryngeal cancer cells. Neoplasma 65, 745-752.
36. Ohshima, J., Haruta, M., Arai, Y., Kasai, F., Fujiwara, Y., Ariga, T., Okita, H., Fukuzawa, M., Hata, J., Horie, H., et al. (2009). Two candidate tumor suppressor genes, MEOX2 and SOSTDC1, identified in a 7p21 homozygous deletion region in a Wilms tumor. Genes, chromosomes & cancer 48, 1037-1050.
37. Zhou, P., Chen, Z., Chang, R.M., Jiang, W., Wu, L.L., and Wang, Z.M. (2012). Growth arrest-specific homeobox is associated with poor survival in patients with hepatocellular carcinoma. Med Oncol 29, 3063-3069.
38. Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., et al. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic acids research 42, D1001-1006.
39. Phelan, C.M., Kuchenbaecker, K.B., Tyrer, J.P., Kar, S.P., Lawrenson, K., Winham, S.J., Dennis, J., Pirie, A., Riggan, M.J., Chornokur, G., et al. (2017). Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. Nature genetics 49, 680-691.
40. Casteel, D.E., Zhuang, S., Zeng, Y., Perrino, F.W., Boss, G.R., Goulian, M., and Pilz, R.B. (2009). A DNA polymerase-{alpha}{middle dot}primase cofactor with homology to replication protein A-32 regulates DNA replication in mammalian cells. The Journal of biological chemistry 284, 5807-5818.
41. Wan, M., Qin, J., Songyang, Z., and Liu, D. (2009). OB fold-containing protein 1 (OBFC1), a human homolog of yeast Stn1, associates with TPP1 and is implicated in telomere length regulation. The Journal of biological chemistry 284, 26725-26731.
42. Wang, F., Podell, E.R., Zaug, A.J., Yang, Y., Baciu, P., Cech, T.R., and Lei, M. (2007). The POT1-TPP1 telomere complex is a telomerase processivity factor. Nature 445, 506-510.

Figure Legend

Figure 1. H3K27ac activity at 10q24.33 and SNP genomic interactions. A) From the top to the bottom of the figure, it is showed the list of SNPs in LD with the lead SNP rs11591710 (zoom-in showing 83,154 bp), the H3K27ac data of NCC cell lines (GSE90683) and of CMM cell

lines (GSE75352 and GSE82332). Functionally relevant SNPs that were further validated are highlighted by colored rectangles. *rs2995264* (red rectangle) showed significant results. **B)** Microarray-based expression profiling on primary tumors demonstrates that lower *OBFC1* expression correlates with GG-rs2995264 risk genotypes (P value= 0.002). **C)** Genomic interactions within the interval chr10:105111147-106194301 (genomic coordinates, hg19) obtained by HiC of the COLO-829 cell line are showed on the right panel. Vertical dashed lines limit our region of interest containing LD SNPs (83,154 bp). Black bordered triangles represent TADs. Bolded triangles highlight interactions between the genomic bin containing the *rs2995264* (chr10:105660000-105670000) and the regions up-stream (distance=2058bp) or down-stream (distance=7327bp) of *OBFC1* (see **Supplementary Table 4**). Significantly interacting genes are underlined in red.

Figure 2. The enhancer containing rs2995264 interacts with the *OBFC1* promoter in CMM cells. A) Schematic representation of Chromosome Conformation Capture (3C) experiment displays approximate positions of analyzed regions, direction of transcription of *OBFC1* and EcoRI cutting sites within the area. Primers P1 and P2 were designed to amplify a novel ligation product formed between the restriction fragments that encode the promoter region and enhancer DNA, respectively. Also the DNA fragment (P2 and P3) for which is not expected an interaction with the promoter region is shown. **B)** The interaction between the enhancer/SNP region and the promoter region (P1-P2) and between the promoter region (P2) and a distal element (P3) in A375 and CJM cells was assessed. The interaction frequency corresponds with the intensity of amplified PCR products analyzed gels are shown in **Supplementary Fig. 5** and **Supplementary Information**. Data are shown as mean ± SD. **C)** Design of CRISPR-mediated enhancer deletion in HEK293T cells showing the sites of targeted deletion in the intron 3 of *OBFC1*. The target regions are indicated by the dashed lines

(deletion A of 720 bp and deletion B of 756 bp) flanked by the pairs of single guides RNA (sgRNA2 yellow, sgRNA3 green and sgRNA6 orange). Agarose gel image with validation PCR results of heterozygous and homozygous deletions are shown in **Supplementary Fig. 6**. **D)** mRNA and protein are collected from H293T selected isogenic lines. (qRT)-PCR and Western Blot analysis are performed to verify the effect of homozygous enhancer deletion on *OBFC1* mRNA and protein levels. Data shown are the mean ± standard deviation from two independent (qRT)-PCR experiments, each done in triplicate. Data are shown as mean ± SD (* P value < 0.01). **E)** Allele-specific binding proteins were identified by mass spectrometry using CMM cell nuclear extract and biotinylated double-stranded oligonucleotides. The dimethyl-labeling ratios of proteins bound to A protective allele (orange) or G risk allele (blu) probes are plotted on the *x* and *y* axes. **F)** MEOX2 preferentially bound to A protective allele of rs2995264 both in basal conditions (-) and after MEOX2-overexpression (+), as determined by ChiP-assay. Data shown are the mean ± SD from two independent (qRT)-PCR experiments, each done in triplicate.

Figure 3. Low *OBFC1* expression is associated with unfavorable histology in CMM and transient knockdown of *OBFC1* influences CMM cell viability in a genotype-specific manner. (A-B-C) Box-plots showing the mRNA expression of *OBFC1* in GSE3189 (P value= 0.025), GSE112509 (P value= 0.0007) and TCGA datasets (P value= 0.08). P value obtained by T-test. **(D-E-F)** *OBFC1* siRNA knockdown as measured by (QRT)-PCR and Western Blot 48h post transfections for experiments. Data shown are the mean ± standard deviation from two independent (QRT)-PCR experiments, each done in triplicate. *P<0.01; P value obtained by T-test. **(G-H)** In cells homozygous for *rs2995264* protective A allele and with higher *OBFC1* expression levels (**Supplementary Fig. 9**), *OBFC1* transient knockdown leads to significant growth increase. **I)** In cells homozygous for *rs2995264* risk G allele and with low *OBFC1*

expression, *OBFC1* transient knockdown does not affect cell growth. **(G-H-I)** Data shown are the mean ± standard deviation from two independent MTT experiments, each done in six-duplicate; P value obtained by T-test.

Figure 4. *OBFC1* transient knockdown enhances telomerase activity. A) *OBFC1* transient knockdown in CMM cells increases significantly telomerase activity. Data shown are the mean ± standard deviation from two independent experiments, each done in triplicate; P value obtained by T-test. **(B-C)** Percentage of cell cycle distribution of CMM cells after transient *OBFC1* knockdown. Data shown are the mean ± standard deviation from three independent transfections experiments, each done in duplicate; P value obtained by T-test (* P= 0.03). **D)** Evaluation of CyclinA2 mRNA expression levels, marker of S phase, in CMM cells analyzed by flow cytometry for cell cycle distribution. Data shown are the mean ± standard deviation from two independent real time PCR experiments, each done in triplicate; P value obtained by T-test (* $P < 0.01$)