HGG Advances

ARTICLE

CFDP1 is a neuroblastoma susceptibility gene that regulates transcription factors of the noradrenergic cell identity

Daniela Formicola,[1,2,3,8] Vito Alessandro Lasorsa,[2,3,8] Sueva Cantalupo,[2,3] Alessandro Testori,[2,3,4] Antonella Cardinale,[2,7] Marianna Avitabile,[2,3] Sharon Diskin,[5,6] Achille Iolascon,[2,3] and Mario Capasso[2,3,9,*]

Summary

Pleiotropic genetic factors (e.g., DNA polymorphisms) may be involved in the initiation of neuroblastoma (NB) and coronary artery disease (CAD) given their common origin from defects in neural crest development. To discover novel NB susceptibility genes, we conducted a three-stage survey including a meta-analysis of NB and CAD genome-wide association data, prioritization of NB causal variants, and validation in an independent cohort of affected individuals–control subjects. The lead SNP, rs13337397 at the 16q23.1 locus, associated with both diseases in the meta-analysis and with NB in the validation study. All the SNPs in linkage disequilibrium with rs13337397 were annotated using the H3K27ac epigenetic marker of neural crest cells (NCC) and NB cell lines. Indeed, we identified the functional SNP rs13337017, mapping within an enhancer of NCCs and NB cell lines and showing long-range interactions with *CFDP1* by Hi-C analysis. Luciferase assays indicated that the risk allele of rs13337017 increased *CFDP1* expression in NB cell lines. Of note, *CFDP1* high expression associated with unfavorable prognostic markers in an analysis including 498 NB transcriptomes. Moreover, depletion of *CFDP1* markedly decreased viability and migration and increased apoptotic rates in NB cell lines. Finally, transcriptome and qPCR analyses revealed that the depletion of *CFDP1* may affect noradrenergic neuron differentiation by downregulating master regulators of sympathetic noradrenergic identity, including *PHOX2B*, *HAND2,* and *GATA3*. Our data strongly suggest that *CFDP1* acts as oncogene in NB. In addition, we provide evidence that genetic predisposition to NB can be mediated by the alteration of noradrenergic lineage-specific gene expression.

Introduction

Neuroblastoma (NB), the most common extracranial tumor in childhood, is an aggressive metastasis-prone tumor of the sympathetic nervous system.[1] Clinical and biological factors, including age at diagnosis, stage, tumor histopathology, ploidy, and genomic aberrations, define distinct risk strata and determine treatment plans for patients.[1] In young children, *MYCN* amplification and other chromosomal arm-level alterations, such as the deletion of 1p and 11q, gain of 17q[2], and *TERT* rearrangements, have been reported as poor prognostic features.[3] Children older than 6 years present with unique structural variants, with 19p loss and 1q gain being among the most frequent.[4] Somatic point mutations in *ALK* and *ATRX*[5] and NB-specific active regulatory elements[6,7] have been reported in primary tumors as cancer drivers with the potential to predict patient prognosis. In addition, several variants and SNPs involving genes of the base excision repair and of the mRNA post transcriptional modification pathways are known for their association with the occurrence of neuroblastoma.[8,9]

Nevertheless, despite progress in discovering clinically actionable alterations, NB still causes 15% of all deaths due to cancer in children. Furthermore, less than 40% of children survive for more than 5 years after being diagnosed with a high-risk disease.[10]

NB arises from the sympathetic ganglia and adrenal medulla, which originate from the trunk neural crest (NC). During early sympathetic neurogenesis, proper migration and specification of human NC cells (NCC) is regulated by a network of transcription factors, including three master regulators of sympathetic noradrenergic identity of NB: PHOX2B, GATA3, and HAND2.[11] Notably, GATA3 and HAND2 are also required for coronary artery development and physiological formation of the outflow tract.[12–14] In addition, Arima and colleagues[15] reported the involvement of preotic NC in the development of smooth muscles in coronary arteries and adjacent cardiac tissues. Indeed, they observed that ablation of preotic NCs in chick and mouse embryos leads to coronary defects.

Recent genome-wide association studies (GWASs)[16] have demonstrated that common single nucleotide

[1]Dipartimento di Neurobiologia e Medicina Molecolare, IRCSS Fondazione Stella Maris, 56128 Pisa, Italy; [2]CEINGE Biotecnologie Avanzate Franco Salvatore, Via Gaetano Salvatore 486, 80145 Napoli, Italy; [3]Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Università degli Studi di Napoli Federico II, Via Pansini, 5, 80131 Napoli, Italy; [4]Division of Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; [5]Division of Oncology and Center for Childhood Cancer Research, Children's Hospital of Philadelphia, Philadelphia, PA, USA; [6]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; [7]Department of Pediatric Hematology and Oncology, Bambino Gesù Children's Hospital, IRCCS, Roma, Italy
[8]These authors contributed equally
[9]Lead contact
*Correspondence: mario.capasso@unina.it
https://doi.org/10.1016/j.xhgg.2022.100158.

polymorphisms (SNPs) are risk factors for NB and coronary artery disease (CAD). In NB, GWASs have identified several susceptibility loci, including *CASC15, BARD1, LMO1, DUSP12, HSD17B12, DDX4/IL31RA, HACE1, LIN28B, TP53, MLF1, CPZ, CDKN1B, NEFL,* and *SLC16A1*.[17] Other loci, including *PCSK9*, *SORT1*, *MIA3*, *PHACTR1*, *LPA*, *CDKN2A*, *CDKN2B*, *CXCL12,* and *LDLR* have been associated with CAD.[18,19] Some of these loci, such as *LPA*, *CDKN2A*, *CDKN2B, CXCL12*, and *LDLR*, also play a role in NB.[20–24] For example, the genomic region of 9p21 confers susceptibility to CAD. In addition, deletion of 9p21 is a marker of poor prognosis for NB patients.[25] Taken together, given the common origins of NB and CAD, we had strong suggestion that these two diseases might share causal pathways. Furthermore, our hypothesis was also supported by our previous work in which we found genetic risk loci, regulating the expression of relevant genes of NCC development, associated with both congenital heart disease and NB susceptibility.[26]

Based on the above observations, we reasoned that pleiotropic DNA polymorphisms could be modestly associated with both NB and CAD and that, in a single phenotype-based GWAS, these SNPs could remain undetected at the standard threshold for genome-wide significance, mainly because of sample size limitations and multiple testing corrections. Previous studies have shown that new risk variants can be identified by combining GWAS results of two or more disease phenotypes.[27–29] Here, to identify novel susceptibility loci for NB, we performed a cross-phenotype meta-analysis of two large GWASs of NB and CAD, including a total of 36,785 individuals and a replication study of Italian NB affected individuals and control subjects (N = 1,716).

This strategy, coupled with biological investigations, allowed us to identify *CFPD1* as a novel NB susceptibility gene that acts as an oncogene by influencing the transcriptional regulatory circuitry of noradrenergic identity in NB.

Materials and methods

Meta-analysis

To identify genetic loci sharing the GWAS signal between NB and CAD, we performed a cross-phenotype meta-analysis using two large GWAS datasets.

In Stage 1 we analyzed the SNP genotyping data of a large CAD GWAS dataset (n = 15,420 affected individuals and n = 15,062 control subjects) obtained from the CARDIoGRAMplusC4D Consortium[30] and the SNP genotyping data obtained from a recent NB GWAS (n = 2,101 affected individuals and n = 4,202 control subjects) published by McDaniel and colleagues.[31] See details in the supplemental information.

First, for each dataset, we performed whole-genome imputation on a multipopulation reference panel to increase the number of detected SNPs. We then combined the significant SNPs ($p < 0.05$) of each dataset to test for shared genetic variants between these two diseases (cross-phenotype meta-analysis) following a 2-fold approach in which we searched for SNPs with the same direction of association (*same effect*) and for SNPs with inverse direction of association in CAD (*inverse effect*).

In Stage 2, we annotated SNPs emerging from these two meta-analysis studies (n = 434) with epigenomic profiling of NC using H3K27Ac chromatin immunoprecipitation sequencing (ChIP-seq) data (GEO: GSM2664365, GSM2664367, GSE90683) to prioritize SNPs mapping in promoters and/or enhancers (active in NCs) with a significance of association of $p < 1.0 \times 10^{-4}$.

In Stage 3 we selected four SNPs to perform direct genotyping by TaqMan assay (TaqMan Universal PCR Master Mix; Applied Biosystems, Thermo Fisher Scientific) in an independent Italian cohort of 636 NB affected individuals and 1,080 control subjects (details are described in the supplemental information). This study was approved by the Ethics Committee of the University of Naples Federico II.

Hardy-Weinberg equilibrium was evaluated using the goodness-of-fit chi-square test in control subjects. Two-sided chi-square tests were used to evaluate differences in the distributions of allele frequencies between patients and control subjects. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated to assess the relative disease risk conferred by a specific allele.

The statistical power was calculated by using the Genetic Association Study (GAS) Power Calculator (http://csg.sph.umich.edu/abecasis/gas_power_calculator/index.html). The false positive report probability (FPRP) for the significant SNP (rs13337397) was calculated according to the method reported in the published paper by Wacholder et al.[32] We had a power of 82.3% to identify disease-associated SNPs with $p = 0.05$, effect size of 1.35, and a disease allele frequency of 0.10. If we set the disease allele frequency at 0.20 we were able to identify disease-associated SNPs with statistical power of 95.8%.

Analysis of public ChIP-seq data

ChIP-seq reads from the GEO: GSE90683 public dataset were reanalyzed, as reported by Avitabile et al.[33] Briefly, alignment to the human reference genome (assembly GRCh37/hg19) was performed using Bowtie2 (v2.3.4.3). Duplicate and low mapping quality (MQ < 20) reads were removed. Enriched regions (peaks) were detected with HMCan (v1.30) using the data from the input DNA to subtract the background noise from the ChIP signals. To identify enhancers and super-enhancers, we used the LILY package.[11]

DNase-I hypersensitivity level (DHS) signal intensity files of the SK-N-SH NB cell line (GEO: GSE96240) and fetal adrenal gland (FAD) tissues (GEO: GSM530653, GSM817165, GSM1027310, GSM1027311, and GSM817167) were downloaded from the GEO database as processed files. The details are described in the supplemental information.

Identification of causal variants at 16q23.1

To identify potential functional SNPs, we first selected variants in linkage disequilibrium (LD) with the lead SNP rs13337397 ($r^2 \geq 0.8$) (n = 57). LD calculations (r^2 and D′) were performed using the LDlink web tool and allele frequencies of Europeans from the 1000 Genomes Project.

Next, we mapped the selected SNPs onto the predicted regulatory elements (super-enhancers [SEs], enhancers, and promoters) obtained in the previous step (see above). From this analysis, we retained only the SNPs (n = 12) that achieved the highest number of SE annotations (SE = 13). Finally, we selected, for further

consideration, the rs13337017 SNP showing the highest levels of both H3K27ac (in NB cell lines) and DHS signals (in FAD tissues).

To obtain additional functional annotations on the prioritized loci and SNPs, we used the GeneHancer database.[34] GeneHancer assigns confidence scores to gene-regulatory element associations based on a combination of evidence annotations. From the list of *CFDP1* associated regulatory elements, we selected those supported by two or more evidence sources.

HiC data analysis

In-house-generated HiC data of the SK-N-BE NB cell line were processed as reported previously.[35,36] Briefly, sequencing was performed using an Illumina HiSeq platform. Paired-end reads of 150 base pairs (bp) were mapped to the reference genome (build hg19/GRCH37) using Bowtie2. HiCExplorer (v3.7) was used to (1) build the interaction matrix at a resolution of 10 Kb (bin size = 10 Kb), (2) normalize the observed read counts, (3) determine topologically associating domains (TADs, self-interacting genome regions) and their boundaries, and (4) plot the results. Subsequently, starting from the selected SNP coordinates, we extended our region of interest of 0.5 Mb up- and downstream and calculated the statistical significance of the interactions between bins with FitHiC (v2.0.8). p-values were corrected for multiple tests using the Benjamini-Hochberg method (false discovery rate, FDR), and the cutoff was set at 5%. Finally, we annotated these bins using the R-Bioconductor package ChIPseeker (v 1.32.0) to map genomic bins to gene coordinates.

Cell cultures

Human HEK293T, SK-N-BE, SK-N-AS, and SH-EP cell lines were obtained from the American Type Culture Collection (respectively ATCC #CRL-3216, #CRL-2268, #CRL-2137, #CRL-2269). The cell lines used for all experiments were re-authenticated and tested as mycoplasma-free and grown in commercial media, as reported in the supplemental information. Early passage cells were used, and the cumulative culture length was less than 3 months after resuscitation.

Treatment of cells with siRNA *CFDP1*

SK-N-BE, SK-N-AS, and SH-EP cells were plated into a six-well at 70% confluence. Cells were transfected with 10 nM of three unique 27-mer small interfering RNA (siRNA) duplexes for *CFDP1* (human) (Origene, Locus ID 10428) and 10 nM of Trilencer-27 Universal Scrambled Negative Control siRNA Duplex (Origene) using X-tremeGENE (Roche). After 48 h of transfection, the cells were used to assess the silencing of protein and mRNA of *CFDP1*. The experiments were performed in triplicate, and three experimental points were analyzed for each experiment.

In vitro functional assays

We used a luciferase reporter assay to assess the effects of the rs13337017 alleles on gene expression. The functional implications of *CFDP1* in regulating cell behavior and properties were assessed on siRNA silenced (siCFDP1) and control (siScrambled) NB cell lines using viability, wound-healing, and Caspase-3 activity assays. Moreover, we used NB cells to quantify the gene expression of *CFDP1* and regulated TFs by quantitative real-time PCR and to assess CFDP1 protein concentrations by western blotting. Detailed descriptions of the experimental conditions for each experiment, quantitative real-time PCR primers, and antibodies are provided in the supplemental information. In each case, the experiments were replicated, and the statistical significance of the results was assessed using a t test.

Analysis of RNA-sequencing data and differential expression

Total RNA was isolated from two biological replicates of the SK-N-AS NB cell line under treated (siCFDP1) and control (siScrambled) experimental conditions and processed as reported in the supplemental information. The RNA-sequencing (RNA-seq) data (obtained on the Illumina platform) were first checked for quality using FastQC and then analyzed using the Tuxedo suite: TopHat (v2.0.14) and Cufflinks v2.1.0. TopHat was run using the default options by providing the reference genome and its related RefSeq reference transcriptome (assembly GRCh37/hg19). Next, Cufflinks was used to assemble the mapped reads into possible transcripts and generate a final transcriptome assembly. Gene- and transcript-level expression is reported as normalized fragments per kilobase of exon per million fragments mapped (FPKM) counts. Finally, Cuffdiff was used to detect differentially expressed genes and transcripts between the two experimental conditions (siCFDP1 versus siScrambled). The results were analyzed using the R-Bioconductor package CummeRbund (v2.16.0). We deemed differentially expressed genes, those showing FDR below 0.05, and the Log2 transformed fold change greater than +0.5 (up-regulated) or lower than −0.5 (down-regulated). Next, we used the expression data of these genes to perform k-means clustering (using k = 4) to identify groups of genes showing similar expression patterns. The four lists of differentially expressed genes (DEGs) were then used to query the Gene Ontology database of Biological Processes (GO BP) with the R-Bioconductor package ClusterProfileR (v3.18.1). BPs were considered enriched if FDR was less than 0.05.

Results

NB and CAD share risk polymorphisms

Candidate susceptibility loci were selected through a cross-phenotype meta-analysis using two GWAS datasets (Figure 1). The results of this analysis are presented below.

Stage 1: meta-analysis of NB and CAD GWAS data | We conducted a fixed-effects meta-analysis of two large GWAS studies, 2,101 NB affected individuals and 4,202 control subjects[31] and 15,420 CAD affected individuals and 15,062 control subjects.[30] Whole-genome imputation of the two GWASs was performed using a multipopulation reference panel,[36] which yielded up to 9 million SNPs. As described in the materials and methods section and detailed in the supplemental information, we compared the GWAS data (p value and odds ratio for 9,671,310 SNPs) of NB and CAD using two approaches. First, meta-analysis across NB and CAD returned 218 SNPs with the same direction of association (*same effect*) (Table S1). With the second approach, the meta-analysis across NB and CAD with inverted effects in CAD, we obtained 193 SNPs with an inverse direction of association (*opposite effect*) (Table S2).

Stage 2: Identification of causal variants | To prioritize the top potential functional SNPs, we integrated NB and CAD meta-analysis data with H3K27ac epigenetic markers using three publicly available ChIP-seq data of human

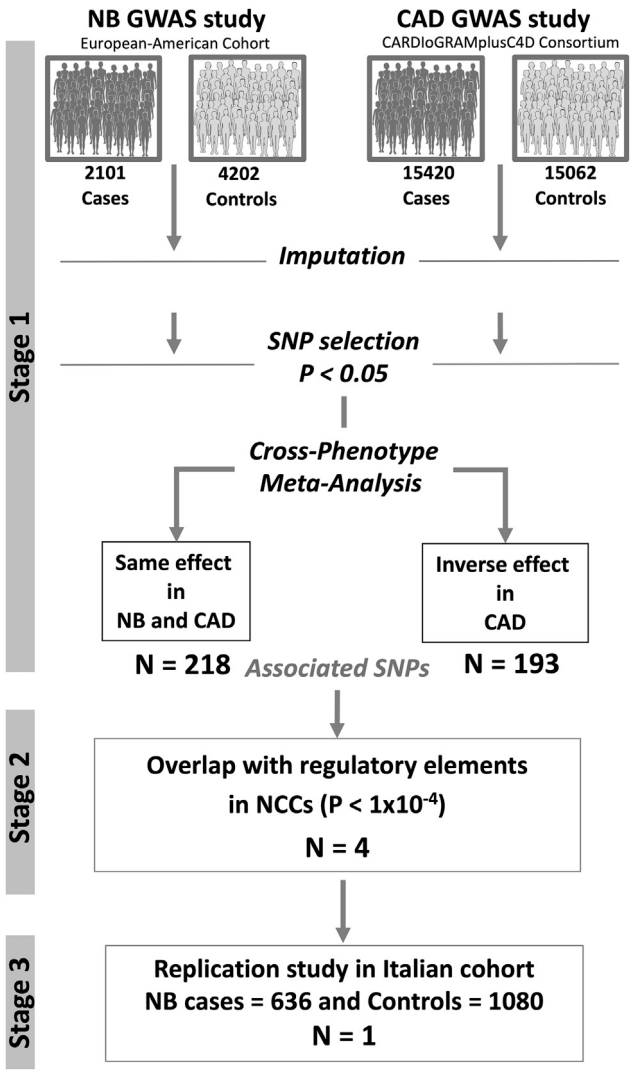

Figure 1. Workflow of the cross-phenotype meta-analysis
Stage 1 | Combination of the summary statistics of two large NB and CAD GWAS data using two approaches: (1) meta-analysis study across NB and CAD and (2) meta-analysis study across NB and CAD with inverted effect in CAD.
Stage 2 | Identification and prioritization of candidate causal SNPs using epigenomic data.
Stage 3 | Replication study in an Italian cohort of 636 NB affected individuals and 1,080 control subjects by direct, PCR-based, genotyping.

NCCs [11,37] to select SNPs mapping in human promoters and/or enhancers (active in NCs) with a p value of association below 1.0×10^{-4} (Figure 1). We identified two SNPs with the same direction of association (green in Table S1) and two SNPs with opposite directions of association (green in Table S2).

Stage 3: Replication in an Italian cohort of NB | We next sought to replicate the results of the association for the four prioritized risk loci in an independent cohort of Italian subjects (n = 636 NB affected individuals and n = 1,080 control subjects) (Figure 1). We identified an association of the minor allele A for rs13337397 SNP at 16q23.1 locus with the NB risk (p = 0.033; OR = 1.339, Table 1).

Functional characterization of NB-CAD cross-associated locus 16q23.1

In order to better understand the functional role of the 16q23.1 locus in NB biology, we further investigated a subset of 57 SNPs in LD ($0.8 < R^2 \leq 1$) with the lead SNP rs13337397 (Table S3). We mapped the selected 57 SNPs on predicted regulatory elements (SEs, enhancers, and promoters) obtained by re-analysis of H3K27ac ChIP-seq peaks of 25 NB cell lines and two human NCCs published in Boeva et al.[11] (GEO: GSE90683). Furthermore, the SNPs were annotated using H3K27ac signal intensity levels from the cell lines listed above. We have also added information regarding the DHS of the SK-N-SH NB cell line (from GEO: GSE96240) and FAD (N = 5) DHS data from GEO: GSE18927 (Figures 2 and S1 and Table S3).

To further filter the list of SNPs, we first selected those SNPs (n = 12) with the highest number of super-enhancer annotations (n = 13) (Table S3). Among the 12 SNPs, we focused on rs13337017, which showed the highest levels of both H3K27ac and DHS (Figures 2 and S1 and Table S3).

Subsequently, we used in-house Hi-C data from the SK-N-BE NB cell line to assess the strength and significance of long-range interactions between rs13337017 SNP and its surrounding genomic regions. We found that rs13337017 and the *CFDP1* locus were mapped within the same topologically associated domain. Furthermore, rs13337017 significantly interacted with the *CFDP1* locus (FDR $\leq 1.45 \times 10^{-16}$) (Figure 2 and Table S4). We also queried the GeneHancer database to retrieve the known regulatory elements associated with *CFDP1* and found that rs13337017 fell within a regulatory element (GeneHancer ID: GH16J075237), with multiple supporting evidence (including Ensembl: ENSR00000542243 and dbSUPER: SE_01027) for its association with *CFDP1* (Figure 2). Our analyses of H3K27ac ChIP-seq data strongly indicated the super-enhancer functions of our prioritized locus at 16q23.1 in NB and NCCs. Nevertheless, we wanted to assess its regulatory activity in the adrenal gland tissues from which NB commonly arises.[38] For this purpose, we used the public database 3DIV (http://3div.kr/). The analysis showed that long-range interactions between the rs13337017 SNP and the promoter of *CFDP1* were maintained in the adrenal gland tissue (Figure S2).

Taken together, these findings provide strong evidence for the regulatory functions of our prioritized locus on chromosome 16q23.1 (including rs13337017) in NB, and are consistent with recent evidence that disease-associated SNPs frequently affect regulatory regions that control tissue-specific and developmental stage-specific gene expression.[39]

The expression profile of *CFDP1* in cells of origin of NB and CAD is comparable

This study was based on the assumption that NB and CAD share the same embryonic origin and regulatory pathways

Table 1. Replication study in Italian cohort

dbSNP identifier	Position (hg19)	Major allele	Minor allele	MAF controls (n = 1,080)	MAF cases (n = 636)	p value	Odds ratio	CI 95% (low-high)
rs11879191	chr19:10,512,911	G	A	0.18	0.18	0.90	1.02	0.803–1.284
rs4387287	chr10:105,677,897	C	A	0.23	0.24	0.63	1.04	0.881–1.233
rs2838330	chr21:45,023,876	A	T	0.44	0.42	0.33	0.93	0.807–1.074
rs13337397*	**chr16:75,295,639**	**C**	**A**	**0.08**	**0.10**	**0.03**	**1.339**	**1.024–1.751**

Significant SNPs obtained in Stage 2 (see Figure 1) were tested by direct genotyping in a cohort of Italian subjects (N = 1,080 healthy control subjects and N = 636 NB patients). MAF, minor allele frequency; CI, confidence interval. Genomic coordinates have been reported for the hg19 assembly. This significant association is in bold. *False Positive Report Probability (FPRP) values equal to 0.11, 0.271, 0.804, 0.976, 0.998, and 1.000 according to the prior probability of 0.25, 0.1, 0.01, 0.001, 0.0001, and 0.00001, respectively.

in the early phases of development. NB arises from trunk-NCC,[40] whereas CAD originates from craniofacial-NCC.[15] To exclude the differential regulation of *CFDP1* in these two branches of NC, we compared the expression profiles of trunk-NCC and craniofacial-NCC in murine embryos (GEO: GSE39191). As expected, we found no statistically significant difference in *CFDP1* expression between the two NC derivatives (Figure 3A).

The rs13337017 acts as an eQTL for *CFDP1*

Expression quantitative trait loci (eQTL) analysis using whole blood tissue from the GTEx database revealed that rs13337017 acts as an eQTL for *CFDP1* ($p = 1.9 \times 10^{-8}$, GTEx Release v7, Figure 3B). Moreover, we observed a significant association of the eQTL in adrenal gland tissues ($p = 0.021$, Figure 3C). To assess the eQTL function of rs13337017 genotypes in NB cell lines, we first obtained *CFDP1* expression profiles in a set of 13 NB cell lines using expression arrays (GEO: GSE78061). We directly genotyped rs13337017 in this set of cell lines (C/C = 11, C/T = 2, and T/T = 0). Finally, we compared *CFDP1* expression according to the rs13337017 genotype and found that the T risk allele was significantly correlated with increased mRNA expression of *CFDP1* ($p = 0.018$, Figure 3D). This association was further confirmed by luciferase reporter gene assays, which showed a higher enhancer activity for the T allele than for the C allele ($p < 0.01$, Figure 3E) in HEK293T and SK-N-AS cell lines.

High *CFDP1* expression associates with worst patient prognosis

The R2 Genomics Analysis and Visualization Platform (http://r2.amc.nl) was used to query the transcriptomic data of 498 NB samples (GEO: GSE62564). Our analysis showed that high *CFDP1* expression was associated with poor overall and event-free survival probabilities ($p \leq 5.2 \times 10^{-7}$) (Figure S3A). Moreover, the expression of *CFDP1* was significantly higher in stage 4 tumors than in lower tumor stages ($p \leq 3.7 \times 10^{-3}$) (Figure S3B). In addition, *CFDP1* expression was significantly higher in four independent NB datasets (including 281 samples) than in normal adrenal gland tissues ($p < 1.0 \times 10^{-4}$, Figure S3C). Finally, by querying gene expression data from the Cancer Cell Line Encyclopedia (GEO: GSE36133) (including 917 samples from 36 cancer types), we found that NB cell lines showed higher expression of *CFDP1* than other cell lines ($p = 4.82 \times 10^{-7}$; ANOVA) (Figure S3D).

In vitro functional analysis of *CFDP1*

To investigate the biological role of *CFDP1* in NB, we transiently knocked down endogenous *CFDP1* using specific siRNA transfection. To conduct this experiment, we first assessed the expression level of CFPD1 protein in NB cell lines by western blotting (Figure S4). From this analysis, we selected the SK-N-BE cell line given its high levels of CFDP1 and the SK-N-AS cell line, given their intermediate noradrenergic identity between NB and NCCs, as reported in Boeva et al. [11] Compared with the control siRNA (siScrambled), three different siRNAs targeting *CFDP1* (siCFPD1) significantly reduced its mRNA (Figure 4A) and protein levels (Figure 4B) in SK-N-AS cells ($p < 0.05$). Moreover, depletion of *CFDP1* markedly decreased cell viability ($p < 0.05$) (Figure 4C), and migration capabilities ($p < 0.05$) (Figures 4D and 4E) and increased apoptotic rates ($p < 0.001$) (Figure 4F) at different time points. The same assays performed on SK-N-BE cells confirmed these results (Figures S5A–S5F). These findings support the involvement of *CFDP1* in the initiation and progression of NB.

The depletion of *CFDP1* affects noradrenergic neuron differentiation

To elucidate the molecular mechanisms involved in *CFDP1*, we sequenced the whole transcriptome (RNA-seq) of the SK-N-AS NB cell lines under siCFDP1 and siScrambled experimental conditions (silenced and control, respectively; see above) and performed differential gene expression analysis. Details of the quality control of RNA-seq analysis are reported in Figures S6A–S6C.

RNA-seq analysis confirmed that *CFDP1* expression was significantly decreased in siCFPD1 compared with that in the control (Log2 FC = −2.36; FDR = 3.6×10^{-3}) (Figure 5A and Table S5). From the differential expression analysis, we identified 278 significantly regulated genes (FDR ≤0.05, Table S5). Subsequently, we performed k-means clustering (k = 4) to group the list of genes based on their expression patterns. In siCFPD1 compared with

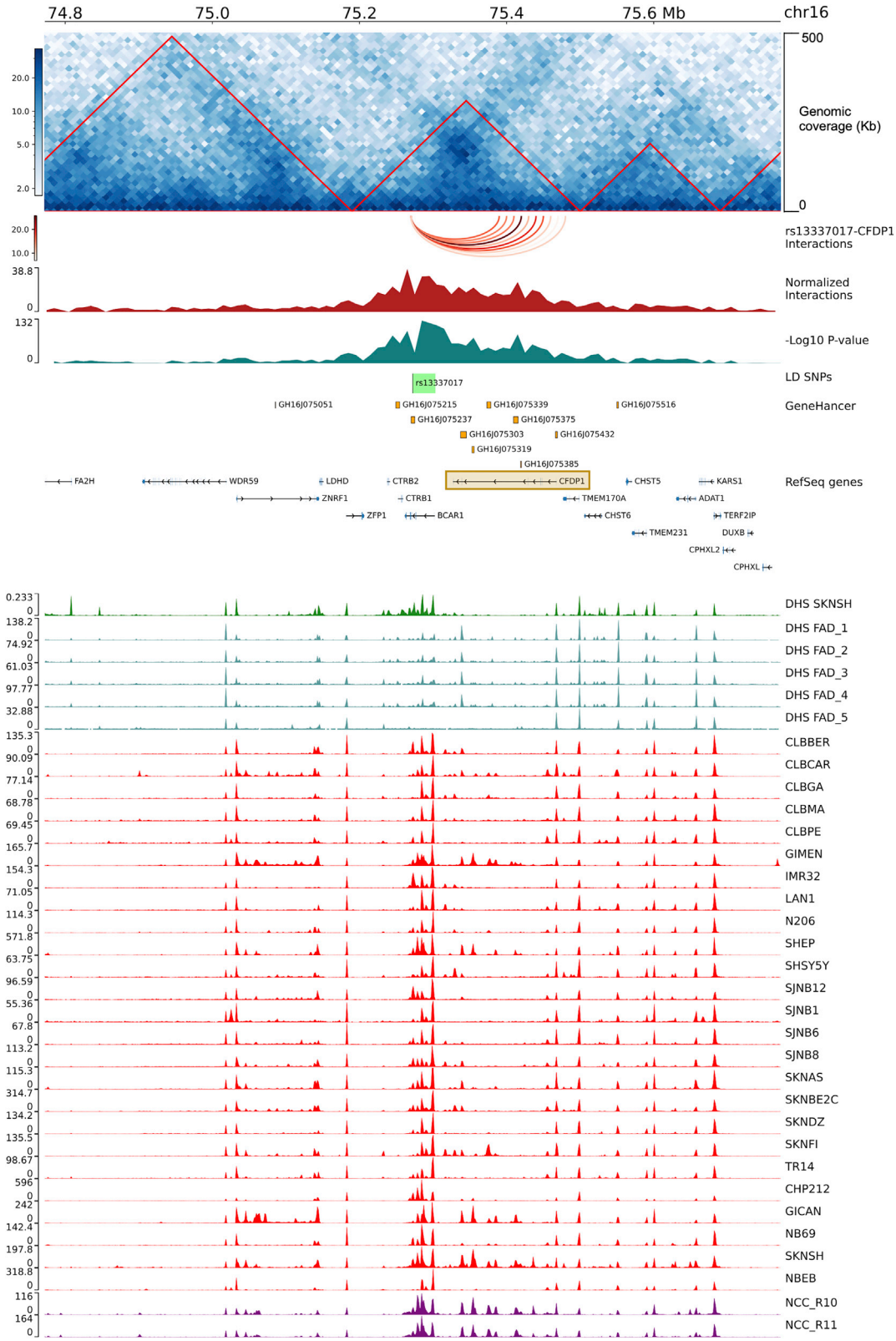

Figure 2. rs13337017 showed long-range significant interactions with *CFPD1* in SK-N-BE NB cell line

The figure reports the genomic interactions from the rs13337017 point of view. The genomic tracks are named from top to bottom and described below. The genomic coordinates are from human genome hg19. The interaction matrix is centered on rs13337017 (chr16:75,272,366-75272367) and extended of 0.5 Mb up- and downstream. Genomic coverage is of 500 Kb and the matrix resolution is of 10 Kb (the interactions are calculated between bins of 10 Kb). Red-bordered triangles represent the Topologically Associated Domains (TADs). The arcs track shows the interactions between rs13337017 and *CFPD1* annotated bins. The normalized number of

(legend continued on next page)

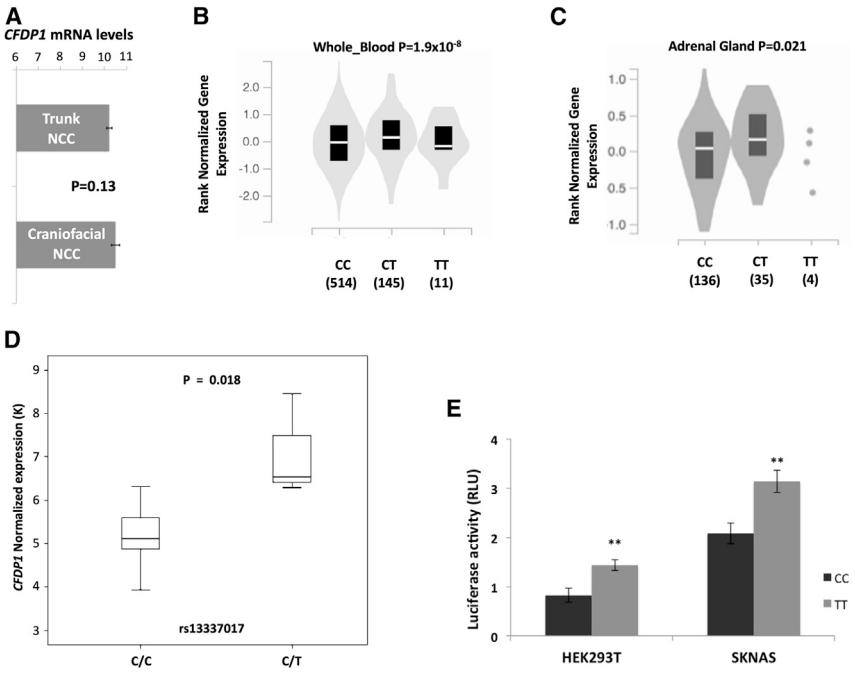

Figure 3. rs13337017 genotypes and *CFDP1* expression association

(A) Comparison of *CFDP1* expression across trunk-NCC and craniofacial-NCC of murine embryos (GEO: GSE39191).

(B) Violin plot reporting the eQTL analysis of rs13337017 in whole blood.

(C) Violin plot showing eQTL analysis of rs13337017 in adrenal gland tissue.

(D) Boxplots reporting *CFDP1* expression grouped by rs13337017 genotypes of NB cell lines in GEO: GSE78061.

(E) Barplots of luciferase reporter gene assays carried out in HEK293T and SK-N-AS. Data shown are the mean ± standard deviation from nine independent transfection experiments, each done in triplicate and compared with promoter-less control. **$p < 0.001$, t test. eQTL analysis performed with GTEx Release v7.

the control, the Clusters 1, 3, and 4 showed increased gene expression, whereas Cluster 2, including *CFDP1,* exhibited decreased expression (Figure 5B). Next, we performed the gene ontology (GO) enrichment analysis of biological processes for the genes included in the four clusters. The results clearly showed that Clusters 1, 3, and 4 (up-regulated genes) contained genes mainly involved in processes related to inflammation, the interferon pathway, and the immune system (Figure 5C and Table S6). The genes in Cluster 2 (down-regulated genes) were mainly involved in processes related to autonomic and sympathetic nervous system development and differentiation of noradrenergic neurons (Figure 5C and Table S6). Notably, the Cluster 2 contained genes known to be involved in conferring noradrenergic cell identities in NB (*PHOX2A*, [41] *INSM1*, [42] *TFAP2B* [43]).

To assess the reliability of RNA-seq data, we measured gene expression by quantitative real-time PCR in our two experimental conditions using seven DEGs that play key roles in conferring noradrenergic and NCC-like identity to NB cells. We confirmed the up-regulation of *GADD45G*, *MAFF*, *FOSL2,* and *CCL20* ($p \leq 0.05$) (Figure 5D) and down-regulation of *PHOX2A*, *TFAP2B,* and *INSM1* ($p \leq 0.05$) (Figure 5D) in siCFDP1 compared with the siScrambled control. Based on these results, we also assessed the expression levels of the transcription factors (TFs) *HAND2*, *GATA3,* and *PHOX2B,* which together participate in a complex network regulating noradrenergic neuron differentiation.[11] For this purpose, we silenced *CFPD1* in three different NB cell lines representative of the three NB cell types: SK-N-BE (noradrenergic cell type), SK-N-AS (intermediate cell type), and SH-EP (NCC type). Quantitative real-time PCR showed down-regulation of *PHOX2B* in SK-N-BE and SK-N-AS cells ($p \leq 0.05$), but not in SH-EP cells (Figure 5E), whereas *HAND2* and *GATA3* were down-regulated in all three cell lines ($p \leq 0.05$) (Figures 5F and 5G). In the RNA-seq data, we observed the same decrease in the expression of these TFs without reaching the statistical threshold of significance, probably because of the low sensitivity of the method (Table S5).

CFDP1 and master regulators of noradrenergic cell identity are co-expressed in NB tumors

Consequently, we assessed whether the observed positive correlation between *CFDP1* and noradrenergic-related TFs (*PHOX2B*, *HAND2*, *GATA3*, *PHOX2A,* and *INSM1*) also subsisted in NB tumors. By querying a public dataset of 498 NB (GEO: GSE62564), we confirmed the direct correlation between *CFDP1* and *PHOX2B*, *GATA3*, and *PHOX2A* ($R > 0.180$; $p \leq 5.24 \times 10^{-5}$), but not with *HAND2* and

interactions (a measure of the strength of interactions). The minus Log10 of the FDR adjusted p value. The region of 30,631 bp containing the list of 58 in Linkage Disequilibrium SNPs (LD SNPs) with rs13337017 is in green. The GeneHancer track reports the regulatory elements, from the GeneHancer database, showing strong associations with *CFDP1* (see materials and methods). The NCBI RefSeq genes (we plotted the longest transcript for each gene). A brown-bordered rectangle highlights the *CFDP1* locus. The ChIP peak-based tracks show, from top to bottom, the DNase-I hypersensitivity levels (DHS) sites of the SK-N-SH NB cell line in dark green (from ENCODE v3); the DHS signal of fetal adrenal gland (n = 5) tissues from Roadmap Epigenomics Consortium (GEO: GSM530653, GSM817165, GSM1027310, GSM1027311, GSM817167) in dark cyan; the H3K27ac data of NB cell lines (from GEO: GSE90683 and GSE65664) in red; and H3K27ac peaks from two human neural crest cells (NCCs) (from GEO: GSE90683) in purple. Data ranges and heatmap color keys are shown on the left of each track, when needed.

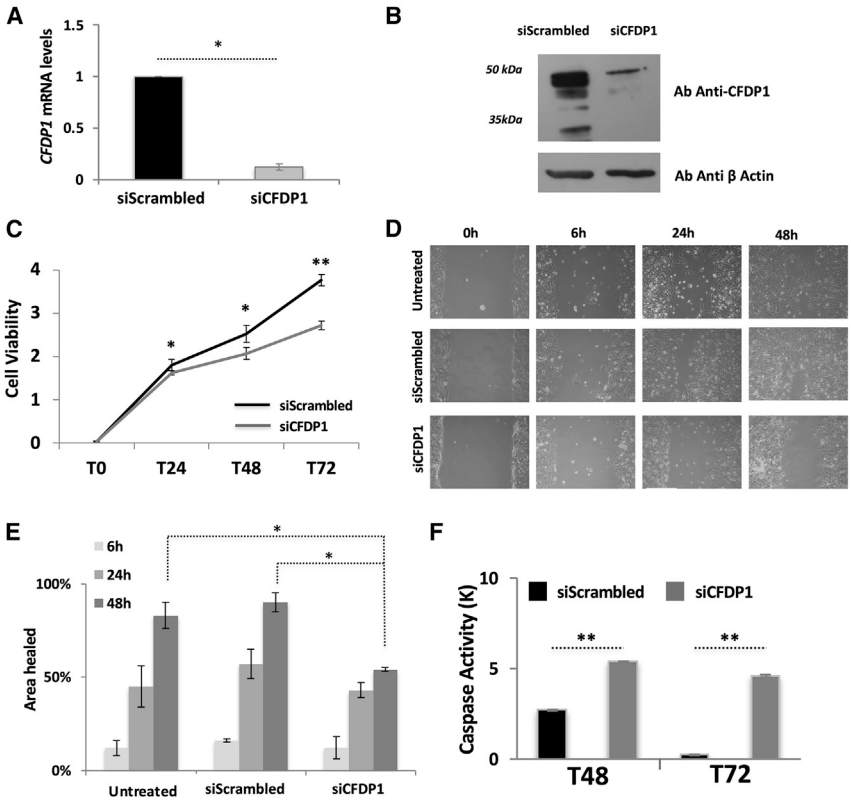

Figure 4. The depletion of *CFDP1* correlates with decreased tumorigenicity in SK-N-AS NB cell line

(A) Barplots showing the results of expression quantification by quantitative real-time PCR of *CFDP1* Control (siScrambled) and *CFDP1* silenced cells (siCFDP1).

(B) CFDP1 protein levels assessed using western blot analysis in Scrambled and siCFDP1 cells.

(C) Line plot reporting cell viability and proliferation measured by MTT assay in Scrambled and siCFDP1 cells. In (A) to (C), the data (mean of three experiments ± standard deviation) are represented as the fold change of siCFDP1 compared with control (siScrambled) cells.

(D) Bright-light microscopy images reporting the results of wound-healing assay observed at different time points (6, 24, and 48 h). The wound distance was measured with ImageJ software.

(E) Barplots reporting the quantifications of distances obtained in (D). The experiments in (D) and (E) were made in triplicate and conducted on untreated, siScrambled, and siCFDP1 SK-N-AS cells.

(F) Barplot of the Caspase-3 activity assay on SK-N-AS cells. The data in siCFDP1 were normalized on the negative (promoter-less) control and reported in thousands (K).

Data are shown as the mean ± standard deviation from three independent transfection experiments, each done in triplicate. In Panel (A), (C), (E), and (F), statistical significance was assessed by t test. *$p < 0.05$; **$p < 0.001$; ***$p < 0.0001$.

INSM1 (Figures S7A–S7E). We can comment on the latter result by speculating that other and more complex regulatory circuitries could involve *HAND2* and *INSM1*. In an additional expression analysis, we observed that *CFDP1* and the noradrenergic-related TFs showed similar expression trends from the undifferentiated NCs to the differentiated adrenal glands ($p \leq 8.23 \times 10^{-3}$) (Figures S8A–S8F). Together, these findings provide strong evidence supporting a direct correlation between *CFDP1* and master regulators of NB, and add a piece to the complex mosaic of the regulatory circuitries of NB.

Discussion

Based on the hypothesis that NB and CAD are NCC-derived diseases that may share the same genetic factors, we conducted a cross-phenotype meta-analysis of GWAS data and replication-based studies to identify novel pleiotropic variants that contribute to NB risk. This strategy exploits the pleiotropic effect of two distinct phenotypes to increase the statistical power necessary for discovering novel disease-predisposing variants discarded by stringent multiple-testing correction in GWAS. We are aware that the different sample size of our two cohorts in the Stage 1 analysis might represent a limitation of this work. However, the reliability of our data was assessed in the replication-based genotyping of an independent cohort of NB affected individuals and control subjects that confirmed the association of the SNPs selected in Stages 1 and 2.

The meta-analysis of NB and CAD GWASs identified the rs13337397 SNP at 16q23.1, in an active enhancer of NCCs, cross-associated with CAD and NB. The same SNP replicated the genetic association with NB risk in an Italian cohort of patients and control subjects. This genetic association supports our hypothesis that the early phases of embryonic development can be altered by pleiotropic risk SNPs. We then focused on functional investigation of the 16q231 locus to unravel its involvement in NB initiation. Fine mapping analyses indicated that the SNP rs13337017 is a potential causative variant; indeed, it resides in a regulatory region of NCCs, the fetal adrenal gland, and NB cell lines. Notably, this genomic region functions as a super-enhancer and strongly interacts with *CFDP1* in NB cells. Moreover, the minor allele of the rs13337017 SNP was associated with increased NB risk and higher *CFDP1* expression in whole blood, adrenal glands, and NB cell lines. Together, these data suggest that NB and CAD are NC-derived diseases that may share the same molecular mechanisms and that our strategy is useful for the identification of novel disease susceptibility genes. Interestingly, *CFDP1* has also been reported as a new CAD susceptibility gene[44]; however, its functional role in CAD development remains unknown.

CFDP1 belongs to the evolutionarily conserved Bucentaur (BCNT) protein superfamily and is expressed during

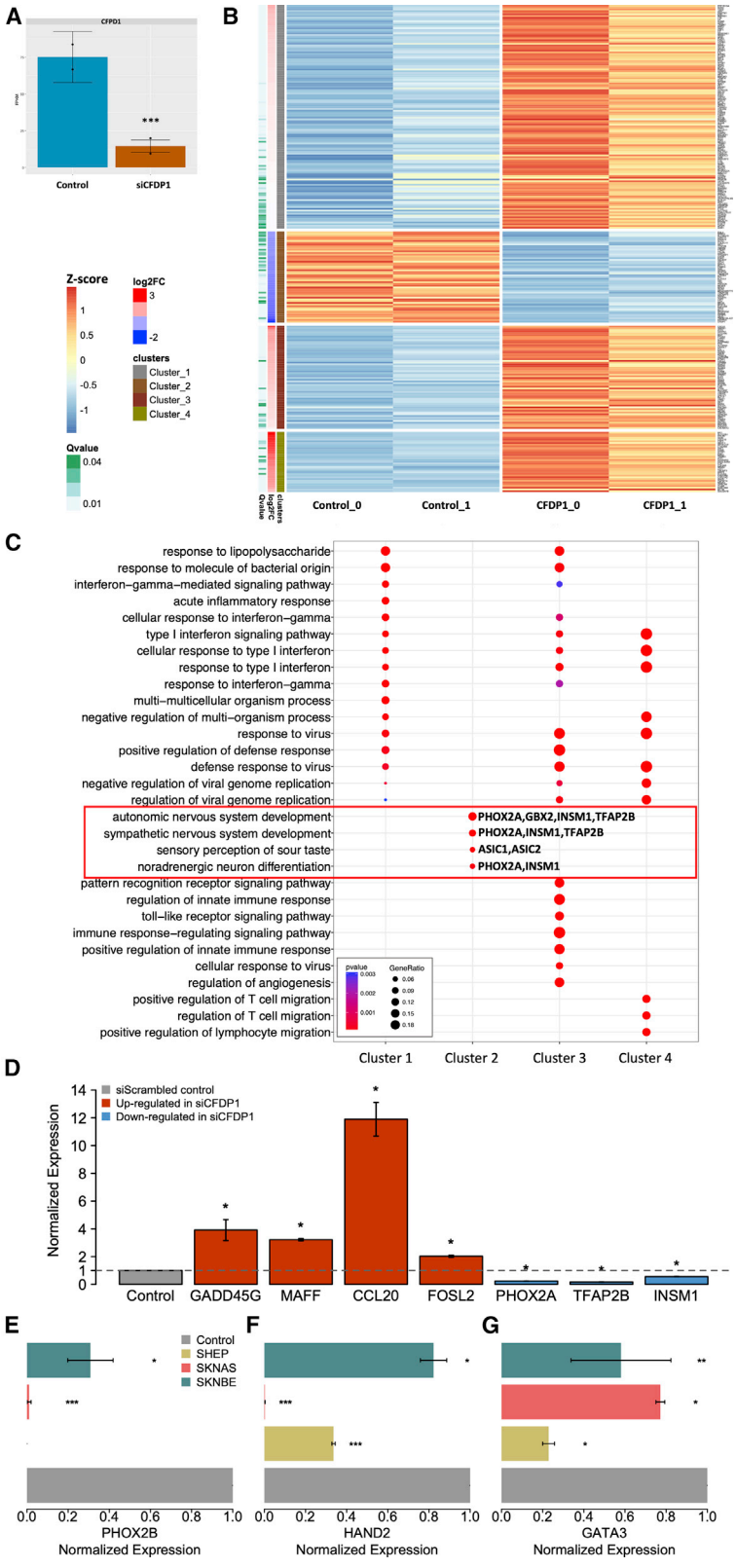

Figure 5. The depletion of *CFDP1* affects the expression of gene regulators of NB cell identity

(A) Barplots showing the expression levels of *CFDP1* in control (siScrambled) and *CFDP1* silenced cells (siCFDP1) as measured by RNA-seq in SK-N-AS.

(B) Heatmap of the differentially expressed genes grouped according to the k-means clustering algorithm. Color key on the left.

(C) Dotplots reporting the Gene Ontology biological process enrichment results of clustered gene lists.

(D) Barplot showing the results of expression quantification by quantitative real-time PCR for four up-regulated genes (red) and three down-regulated genes (blue) in SK-N-AS.

(E) Expression quantification by quantitative real-time PCR of *PHOX2B* in SK-N-AS and SK-N-BE NB cell lines.

(F) Expression quantification by quantitative real-time PCR of *HAND2* in SK-N-AS, SK-N-BE, and SH-EP NB cell lines.

(G) Expression quantification by quantitative real-time PCR of *GATA3* in SK-N-AS, SK-N-BE, and SH-EP NB cell lines. The data in (D) to (G) are reported as fold-changes of induction in siCFDP1 compared with control cells. In Panel (A), (D), (E), (F), and (G) statistical significance was assessed using a t test. *$p < 0.05$; **$p < 0.001$; ***$p < 0.0001$.

Williams-Beuren syndrome, which is characterized by developmental disorders affecting many body sites, including the heart, blood vessels, bone, and eye.[46] Craniofacial birth defects involving bones can be caused by abnormal development of NCC.[47,48] Based on these observations, we propose that risk variants can induce *CFDP1* overexpression and contribute to cellular malignant transformation in the early phases of embryogenesis by altering the normal program of NCC differentiation into sympathetic neurons. Our hypothesis was supported by literature reporting predisposing variants with *cis-effects* on *LIN28B* in NB. [49] *LIN28B* is involved in reprogramming of human and mouse fibroblasts into pluripotent stem cells during the embryogenesis.[50,51]

Our work demonstrated that the minor allele of rs13337017 was correlated with increased *CFDP1* mRNA levels and conferred a risk of NB development, suggesting that *CFPD1* acts as an oncogene. Accordingly, *CFDP1* expression was significantly higher in NB than in other types of tumors and normal tissues, and was correlated with poor clinical outcomes in patients with NB. Transient knockdown of *CFDP1* results in significant growth inhibition and of apoptosis induction in NB cells. This result is consistent with previous reports indicating that *CFDP1* stimulates cell proliferation and prevents apoptosis in embryonic fibroblasts. [52]

early embryogenesis. In mice, *CFDP1* is highly expressed during neural tube development[45] and its ablation causes early embryonic lethality.[45] Recent studies have proposed *CFDP1* involvement in craniofacial development and osteogenesis in vertebrates and linked this gene with

Evidence of high expression levels of *CFDP1* has been reported in a wide range of human tissues, including cancer tissues.[16,53] Furthermore, recent studies have associated the 16q23.1 locus, which contains *CFDP1*, with different cancer types, such as multiple myeloma,[54] acute

lymphoblastic leukemia,[55] pancreatic cancer,[56,57] and breast cancer.[58] Taken together, these results provide evidence that *CFDP1* may play an oncogenic role in NB and potentially in other cancer diseases. Nevertheless, further studies are necessary to assess the implications of *CFDP1* in NB.

In addition, using RNA-seq, we investigated the possible molecular mechanisms involving *CFDP1* in NB cell lines. We compared the transcriptome of *CFDP1* depleted cells with that of untreated control cells. Analysis of the DEGs revealed that sympathetic nervous system development and noradrenergic neuron differentiation were the most significantly enriched GO terms among the down-regulated genes. Interestingly, depletion of *CFDP1* also reduced the expression of three master regulators (*PHOX2B*, *HAND2,* and *GATA3*) of noradrenergic cell identity in NB.[11] Since *CFPD1* seems to play a role in chromatin organization[59] and in promoting neural differentiation,[48] we speculated that its alteration could affect normal differentiation programs controlled by the master regulators *PHOX2B*, *HAND2,* and *GATA3* and thus promote NB initiation and/or progression.

In light of our results, additional studies, involving mouse models, become necessary to better understand the oncogenic role of *CFPD1*. Furthermore, they could help in deepening the biochemical and physiological processes behind the occurrence and development of NB. In the future, such levels of information could also be useful for alternative and personalized therapy strategies.

In conclusion, our results indicate that a polymorphism within a super-enhancer element influences NB susceptibility by modulating *CFDP1* expression leading to oncogenic dependency in tumor cells, likely by altering the control of noradrenergic differentiation. Furthermore, this study demonstrated that using GWAS data through cross-phenotype meta-analysis and integrating functional genomics data could be a useful strategy to identify novel disease susceptibility loci and genes.

Data and code availability

The RNA-sequencing data generated and analyzed in this study are available at the NCBI GEO database under the accession number: GEO: GSE216674. Public data and data repositories are referenced within the manuscript. This study did not generate code.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2022.100158.

Acknowledgments

This study was supported by grants from Associazione Italiana per la Ricerca sul Cancro (Grant no. 25796 to M.C. and Grant no. 20757 to A.I.); Fondazione Italiana per la Lotta al Neuroblastoma (to M.C.); Associazione Oncologia Pediatrica e Neuroblastoma (to M.C.); Regione Campania "SATIN" grant 2018-2020 (to M.C.); and Associazione Giulio Adelfio onlus (to M.C. and A.I.).

Author contributions

M.C., D.F., and V.A.L. drafted the manuscript. D.F., A.C., and M.A. performed *in vitro* functional assays. V.A.L. performed bioinformatic analyses and curated data collection, analysis, and visualization. A.T. performed bioinformatic analyses of the GWAS data. S.C. supported in DNA sample collecting and genotyping. S.D. provided GWAS data for neuroblastoma. A.C., S.C., and M.A. critically reviewed the manuscript. S.D. and A.I. critically reviewed the study design and the manuscript. M.C. designed and supervised the study. All the authors have read and approved the final version of the manuscript.

Declaration of interests

The authors declare no competing interests.

Received: August 30, 2022
Accepted: October 31, 2022

Web resources

GEO DataSets: https://www.ncbi.nlm.nih.gov/gds

GAS Power Calculator: http://csg.sph.umich.edu/abecasis/gas_power_calculator/index.html

GTEx: https://gtexportal.org/home/

R2 Genomics Analysis and Visualization Platform: http://r2.amc.nl

3DIV: http://3div.kr/

References

1. Matthay, K.K., Maris, J.M., Schleiermacher, G., Nakagawara, A., Mackall, C.L., Diller, L., and Weiss, W.A. (2016). Neuroblastoma. Nat. Rev. Dis. Primers *2*, 16078. https://doi.org/10.1038/nrdp.2016.78.
2. Capasso, M., and Diskin, S.J. (2010). Genetics and genomics of neuroblastoma. Cancer Treat Res. *155*, 65–84. https://doi.org/10.1007/978-1-4419-6033-7_4.
3. Peifer, M., Hertwig, F., Roels, F., Dreidax, D., Gartlgruber, M., Menon, R., Kramer, A., Roncaioli, J.L., Sand, F., Heuckmann, J.M., et al. (2015). Telomerase activation by genomic rearrangements in high-risk neuroblastoma. Nature *526*, 700–704. https://doi.org/10.1038/nature14980.
4. Lasorsa, V.A., Cimmino, F., Ognibene, M., Mazzocco, K., Erminio, G., Morini, M., Conte, M., Iolascon, A., Pezzolo, A., and Capasso, M. (2020). 19p loss is significantly enriched in older age neuroblastoma patients and correlates with poor prognosis. NPJ Genom Med. *5*, 18. https://doi.org/10.1038/s41525-020-0125-4.
5. Pugh, T.J., Morozova, O., Attiyeh, E.F., Asgharzadeh, S., Wei, J.S., Auclair, D., Carter, S.L., Cibulskis, K., Hanna, M., Kiezun, A., et al. (2013). The genetic landscape of high-risk neuroblastoma. Nat. Genet. *45*, 279–284. https://doi.org/10.1038/ng.2529.
6. Lasorsa, V.A., Montella, A., Cantalupo, S., Tirelli, M., de Torres, C., Aveic, S., Tonini, G.P., Iolascon, A., and Capasso, M. (2022). Somatic mutations enriched in cis-regulatory elements affect

genes involved in embryonic development and immune system response in neuroblastoma. Cancer Res. CAN-20-3788. https://doi.org/10.1158/0008-5472.
7. Capasso, M., Lasorsa, V.A., Cimmino, F., Avitabile, M., Cantalupo, S., Montella, A., De Angelis, B., Morini, M., de Torres, C., Castellano, A., et al. (2020). Transcription factors involved in tumorigenesis are over-represented in mutated active DNA-binding sites in neuroblastoma. Cancer Res. *80*, 382–393. https://doi.org/10.1158/0008-5472.CAN-19-2883.
8. Zhuo, Z., Lu, H., Zhu, J., Hua, R.X., Li, Y., Yang, Z., Zhang, J., Cheng, J., Zhou, H., Li, S., et al. (2020). METTL14 gene polymorphisms confer neuroblastoma susceptibility: an eight-center case-control study. Mol. Ther. Nucleic Acids *22*, 17–26. https://doi.org/10.1016/j.omtn.2020.08.009.
9. Zhuo, Z., Zhou, C., Fang, Y., Zhu, J., Lu, H., Zhou, H., Wu, H., Wang, Y., and He, J. (2020). Correlation between the genetic variants of base excision repair (BER) pathway genes and neuroblastoma susceptibility in eastern Chinese children. Cancer Commun. *40*, 641–646. https://doi.org/10.1002/cac2.12088.
10. Park, J.R., Bagatell, R., Cohn, S.L., Pearson, A.D., Villablanca, J.G., Berthold, F., Burchill, S., Boubaker, A., McHugh, K., Nuchtern, J.G., et al. (2017). Revisions to the international neuroblastoma response criteria: a consensus statement from the national cancer institute clinical trials planning meeting. J. Clin. Oncol. *35*, 2580–2587. https://doi.org/10.1200/JCO.2016.72.0177.
11. Boeva, V., Louis-Brennetot, C., Peltier, A., Durand, S., Pierre-Eugene, C., Raynal, V., Etchevers, H.C., Thomas, S., Lermine, A., Daudigeos-Dubus, E., et al. (2017). Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. Nat. Genet. *49*, 1408–1413. https://doi.org/10.1038/ng.3921.
12. Raid, R., Krinka, D., Bakhoff, L., Abdelwahid, E., Jokinen, E., Karner, M., Malva, M., Meier, R., Pelliniemi, L.J., Ploom, M., et al. (2009). Lack of Gata3 results in conotruncal heart anomalies in mouse. Mech. Dev. *126*, 80–89. https://doi.org/10.1016/j.mod.2008.10.001.
13. Song, H., Suehiro, J., Kanki, Y., Kawai, Y., Inoue, K., Daida, H., Yano, K., Ohhashi, T., Oettgen, P., Aird, W.C., et al. (2009). Critical role for GATA3 in mediating Tie2 expression and function in large vessel endothelial cells. J. Biol. Chem. *284*, 29109–29124. https://doi.org/10.1074/jbc.M109.041145.
14. Barnes, R.M., Firulli, B.A., VanDusen, N.J., Morikawa, Y., Conway, S.J., Cserjesi, P., Vincentz, J.W., and Firulli, A.B. (2011). Hand2 loss-of-function in Hand1-expressing cells reveals distinct roles in epicardial and coronary vessel development. Circ. Res. *108*, 940–949. https://doi.org/10.1161/CIRCRESAHA.110.233171.
15. Arima, Y., Miyagawa-Tomita, S., Maeda, K., Asai, R., Seya, D., Minoux, M., Rijli, F.M., Nishiyama, K., Kim, K.S., Uchijima, Y., et al. (2012). Preotic neural crest cells contribute to coronary artery smooth muscle involving endothelin signalling. Nat. Commun. *3*, 1267. https://doi.org/10.1038/ncomms2258.
16. Orian-Rousseau, V., Mink, S., Mengwasser, J., HogenEsch, H., Guo, F., Thies, W.G., Hofmann, M., Herrlich, P., and Ponta, H. (2005). Genes upregulated in a metastasizing human colon carcinoma cell line. Int. J. Cancer *113*, 699–705. https://doi.org/10.1002/ijc.20644.
17. Tonini, G.P., and Capasso, M. (2020). Genetic predisposition and chromosome instability in neuroblastoma. Cancer Metastasis Rev. *39*, 275–285. https://doi.org/10.1007/s10555-020-09843-4.
18. Mehta, N.N. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Circulation. Cardiovascular Genet. *4*, 327–329. https://doi.org/10.1161/CIRCGENETICS.111.960443.
19. Chen, Z., and Schunkert, H. (2021). Genetics of coronary artery disease in the post-GWAS era. J. Intern. Med. *290*, 980–992. https://doi.org/10.1111/joim.13362.
20. Tan, S.L., Nishi, M., Ohtsuka, T., Matsui, T., Takemoto, K., Kamio-Miura, A., Aburatani, H., Shinkai, Y., and Kageyama, R. (2012). Essential roles of the histone methyltransferase ESET in the epigenetic control of neural progenitor cells during development. Development *139*, 3806–3816. https://doi.org/10.1242/dev.082198.
21. Pan, L., Ma, X., Wen, B., Su, Z., Zheng, X., Liu, Y., Li, H., Chen, Y., Wang, J., Lu, F., et al. (2015). Microphthalmia-associated transcription factor/T-box factor-2 axis acts through Cyclin D1 to regulate melanocyte proliferation. Cell Prolif *48*, 631–642. https://doi.org/10.1111/cpr.12227.
22. Williams, A.L., and Bohnsack, B.L. (2015). Neural crest derivatives in ocular development: discerning the eye of the storm. Birth defects research. Part C, Embryo today : reviews *105*, 87–95. https://doi.org/10.1002/bdrc.21095.
23. Yeh, C.W., Kao, S.H., Cheng, Y.C., and Hsu, L.S. (2013). Knockdown of cyclin-dependent kinase 10 (cdk10) gene impairs neural progenitor survival via modulation of raf1a gene expression. J. Biol. Chem. *288*, 27927–27939. https://doi.org/10.1074/jbc.M112.420265.
24. Dong, D., Jiang, M., Xu, X., Guan, M., Wu, J., Chen, Q., and Xiang, L. (2012). The effects of NB-UVB on the hair follicle-derived neural crest stem cells differentiating into melanocyte lineage in vitro. J. Dermatol. Sci. *66*, 20–28. https://doi.org/10.1016/j.jdermsci.2012.01.012.
25. Takita, J., Hayashi, Y., Nakajima, T., Adachi, J., Tanaka, T., Yamaguchi, N., Ogawa, Y., Hanada, R., Yamamoto, K., and Yokota, J. (1998). The p16 (CDKN2A) gene is involved in the growth of neuroblastoma cells and its expression is associated with prognosis of neuroblastoma patients. Oncogene *17*, 3137–3143. https://doi.org/10.1038/sj.onc.1202232.
26. Testori, A., Lasorsa, V.A., Cimmino, F., Cantalupo, S., Cardinale, A., Avitabile, M., Limongelli, G., Russo, M.G., Diskin, S., Maris, J., et al. (2019). Exploring shared susceptibility between two neural crest cells originating conditions: neuroblastoma and congenital heart disease. Genes *10*. https://doi.org/10.3390/genes10090663.
27. Kar, S.P., Beesley, J., Amin Al Olama, A., Michailidou, K., Tyrer, J., Kote-Jarai, Z., Lawrenson, K., Lindstrom, S., Ramus, S.J., Thompson, D.J., et al. (2016). Genome-wide meta-analyses of breast, ovarian, and prostate cancer association studies identify multiple new susceptibility loci shared by at least two cancer types. Cancer Discov. *6*, 1052–1067. https://doi.org/10.1158/2159-8290.CD-15-1227.
28. Fehringer, G., Kraft, P., Pharoah, P.D., Eeles, R.A., Chatterjee, N., Schumacher, F.R., Schildkraut, J.M., Lindstrom, S., Brennan, P., Bickeboller, H., et al. (2016). Cross-cancer genome-wide analysis of lung, ovary, breast, prostate, and colorectal cancer reveals novel pleiotropic associations. Cancer Res. *76*, 5103–5114. https://doi.org/10.1158/0008-5472.CAN-15-2980.
29. Hung, R.J., Ulrich, C.M., Goode, E.L., Brhane, Y., Muir, K., Chan, A.T., Marchand, L.L., Schildkraut, J., Witte, J.S., Eeles, R., et al. (2015). Cross cancer genomic investigation of

inflammation pathway for five common cancers: lung, ovary, prostate, breast, and colorectal cancer. J. of the Natl Cancer Inst. *107*. https://doi.org/10.1093/jnci/djv246.
30. Coronary Artery Disease Genetics, C. (2011). A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nat. Genet. *43*, 339–344. https://doi.org/10.1038/ng.782.
31. McDaniel, L.D., Conkrite, K.L., Chang, X., Capasso, M., Vaksman, Z., Oldridge, D.A., Zachariou, A., Horn, M., Diamond, M., Hou, C., et al. (2017). Common variants upstream of MLF1 at 3q25 and within CPZ at 4p16 associated with neuroblastoma. PLoS Genet. *13*, e1006787. https://doi.org/10.1371/journal.pgen.1006787.
32. Wacholder, S., Chanock, S., Garcia-Closas, M., El Ghormli, L., and Rothman, N. (2004). Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. Journal of the National Cancer Institute *96*, 434–442. https://doi.org/10.1093/jnci/djh075.
33. Avitabile, M., Lasorsa, V.A., Cantalupo, S., Cardinale, A., Cimmino, F., Montella, A., Capasso, D., Haupt, R., Amoroso, L., Garaventa, A., et al. (2020). Association of PARP1 polymorphisms with response to chemotherapy in patients with high-risk neuroblastoma. J. Cell Mol. Med. *24*, 4072–4081. https://doi.org/10.1111/jcmm.15058.
34. Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M., et al. (2017). GeneHancer: Genome-wide Integration of Enhancers and Target Genes in GeneCards (Database). https://doi.org/10.1093/database/bax028.
35. Cardinale, A., Cantalupo, S., Lasorsa, V.A., Montella, A., Cimmino, F., Succoio, M., Vermeulen, M., Baltissen, M.P., Esposito, M., Avitabile, M., et al. (2022). Functional annotation and investigation of the 10q24.33 melanoma risk locus identifies a common variant that influences transcriptional regulation of OBFC1. Hum. Mol. Genet. *31*, 863–874. https://doi.org/10.1093/hmg/ddab293.
36. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. *5*, e1000529. https://doi.org/10.1371/journal.pgen.1000529.
37. Rada-Iglesias, A., Prescott, S.L., and Wysocka, J. (2013). Human genetic variation within neural crest enhancers: molecular and phenotypic implications. Philosophical transactions of the Royal Society of London. Series B, Biological sciences *368*, 20120360. https://doi.org/10.1098/rstb.2012.0360.
38. Tolbert, V.P., Coggins, G.E., and Maris, J.M. (2017). Genetic susceptibility to neuroblastoma. Curr. Opin. Genet. Dev. *42*, 81–90. https://doi.org/10.1016/j.gde.2017.03.008.
39. Parker, S.C., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., Program, N.C.S., et al. (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Prod. of the Natl. Acad. of Sci. of the USA *110*, 17921–17926. https://doi.org/10.1073/pnas.1317023110.
40. Brodeur, G.M. (2003). Neuroblastoma: biological insights into a clinical enigma. Nat. Rev. Cancer *3*, 203–216. https://doi.org/10.1038/nrc1014.
41. Seo, H., Hong, S.J., Guo, S., Kim, H.S., Kim, C.H., Hwang, D.Y., Isacson, O., Rosenthal, A., and Kim, K.S. (2002). A direct role of the homeodomain proteins Phox2a/2b in noradrenaline neurotransmitter identity determination. J. Neurochem. *80*, 905–916. https://doi.org/10.1046/j.0022-3042.2002.00782.x.
42. Wildner, H., Gierl, M.S., Strehle, M., Pla, P., and Birchmeier, C. (2008). Insm1 (IA-1) is a crucial component of the transcriptional network that controls differentiation of the sympatho-adrenal lineage. Development *135*, 473–481. https://doi.org/10.1242/dev.011783.
43. Ikram, F., Ackermann, S., Kahlert, Y., Volland, R., Roels, F., Engesser, A., Hertwig, F., Kocak, H., Hero, B., Dreidax, D., et al. (2016). Transcription factor activating protein 2 beta (TFAP2B) mediates noradrenergic neuronal differentiation in neuroblastoma. Mol. Oncol. *10*, 344–359. https://doi.org/10.1016/j.molonc.2015.10.020.
44. Klarin, D., Zhu, Q.M., Emdin, C.A., Chaffin, M., Horner, S., McMillan, B.J., Leed, A., Weale, M.E., Spencer, C.C.A., Aguet, F., et al. (2017). Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. Nat. Genet. *49*, 1392–1397. https://doi.org/10.1038/ng.3914.
45. Diekwisch, T.G., Marches, F., Williams, A., and Luan, X. (1999). Cloning, gene expression, and characterization of CP27, a novel gene in mouse embryogenesis. Gene *235*, 19–30.
46. Makeyev, A.V., and Bayarsaihan, D. (2011). Molecular basis of Williams-Beuren syndrome: TFII-I regulated targets involved in craniofacial development. Cleft Palate Craniofac J *48*, 109–116. https://doi.org/10.1597/09-093.
47. Trainor, P.A. (2010). Craniofacial birth defects: the role of neural crest cells in the etiology and pathogenesis of Treacher Collins syndrome and the potential for prevention. Am. J. Med. Genet. *152A*, 2984–2994. https://doi.org/10.1002/ajmg.a.33454.
48. Snider, T.N., and Mishina, Y. (2014). Cranial neural crest cell contribution to craniofacial formation, pathology, and future directions in tissue engineering. Birth defects research. Part C, Embryo today : reviews *102*, 324–332. https://doi.org/10.1002/bdrc.21075.
49. Diskin, S.J., Capasso, M., Schnepp, R.W., Cole, K.A., Attiyeh, E.F., Hou, C., Diamond, M., Carpenter, E.L., Winter, C., Lee, H., et al. (2012). Common variation at 6q16 within HACE1 and LIN28B influences susceptibility to neuroblastoma. Nat. Genet. *44*, 1126–1130. https://doi.org/10.1038/ng.2387.
50. Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917–1920. https://doi.org/10.1126/science.1151526.
51. Melton, C., Judson, R.L., and Blelloch, R. (2010). Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. Nature *463*, 621–626. https://doi.org/10.1038/nature08725.
52. Luan, X., and Diekwisch, T.G. (2002). CP27 affects viability, proliferation, attachment and gene expression in embryonic fibroblasts. Cell Prolif *35*, 207–219.
53. Messina, G., Celauro, E., Atterrato, M.T., Giordano, E., Iwashita, S., and Dimitri, P. (2015). The Bucentaur (BCNT) protein family: a long-neglected class of essential proteins required for chromatin/chromosome organization and function. Chromosoma *124*, 153–162. https://doi.org/10.1007/s00412-014-0503-8.
54. Mitchell, J.S., Li, N., Weinhold, N., Forsti, A., Ali, M., van Duin, M., Thorleifsson, G., Johnson, D.C., Chen, B., Halvarsson, B.M., et al. (2016). Genome-wide association study

identifies multiple susceptibility loci for multiple myeloma. Nat. Commun. *7*, 12050. https://doi.org/10.1038/ncomms12050.

55. Shi, Y., Du, M., Fang, Y., Tong, N., Zhai, X., Sheng, X., Li, Z., Xue, Y., Li, J., Chu, H., et al. (2016). Identification of a novel susceptibility locus at 16q23.1 associated with childhood acute lymphoblastic leukemia in Han Chinese. Hum. Mol. Genet. *25*, 2873–2880. https://doi.org/10.1093/hmg/ddw112.
56. Childs, E.J., Mocci, E., Campa, D., Bracci, P.M., Gallinger, S., Goggins, M., Li, D., Neale, R.E., Olson, S.H., Scelo, G., et al. (2015). Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. Nat. Genet. *47*, 911–916. https://doi.org/10.1038/ng.3341.
57. Wolpin, B.M., Rizzato, C., Kraft, P., Kooperberg, C., Petersen, G.M., Wang, Z., Arslan, A.A., Beane-Freeman, L., Bracci, P.M., Buring, J., et al. (2014). Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. Nat. Genet. *46*, 994–1000. https://doi.org/10.1038/ng.3052.
58. Higginbotham, K.S., Breyer, J.P., McReynolds, K.M., Bradley, K.M., Schuyler, P.A., Plummer, W.D., Freudenthal, M.E., Trentham-Dietz, A., Newcomb, P.A., Parl, F.F., et al. (2012). A multistage genetic association study identifies breast cancer risk loci at 10q25 and 16q24. Cancer Epidemiol. Biomarkers Prev. *21*, 1565–1573. https://doi.org/10.1158/1055-9965.EPI-12-0386.
59. Messina, G., Atterrato, M.T., Prozzillo, Y., Piacentini, L., Losada, A., and Dimitri, P. (2017). The human Cranio Facial Development Protein 1 (Cfdp1) gene encodes a protein required for the maintenance of higher-order chromatin organization. Sci. Rep. *7*, 45022. https://doi.org/10.1038/srep45022.

HGGA, Volume *4*

Supplemental information

CFDP1 is a neuroblastoma susceptibility gene that regulates transcription factors of the noradrenergic cell identity

Daniela Formicola, Vito Alessandro Lasorsa, Sueva Cantalupo, Alessandro Testori, Antonella Cardinale, Marianna Avitabile, Sharon Diskin, Achille Iolascon, and Mario Capasso

Supplemental material

Supplemental Figures (pages 2-10)

Figure S1. The rs13337017 mapped within a regulatory region of open and active chromatin.
Figure S2. rs13337017 showed long-range significant interactions with *CFPD1* in Adrenal Gland.
Figure S3. High expression of *CFDP1* is associated with poor prognosis and high tumor stages in NB patients.
Figure S4. Levels of CFDP1 protein in NB cell lines.
Figure S5. The depletion of *CFDP1* affects gene expression and tumorigenicity in SK-N-BE NB cell line.
Figure S6. Main features and quality control of RNA-seq data.
Figure S7. *CFDP1* expression was correlated with that of noradrenergic transcription factors.
Figure S8. *CFDP1* and noradrenergic transcription factors showed similar expression patterns.

Supplemental information (pages 11-18)

Supplemental references (page 19)

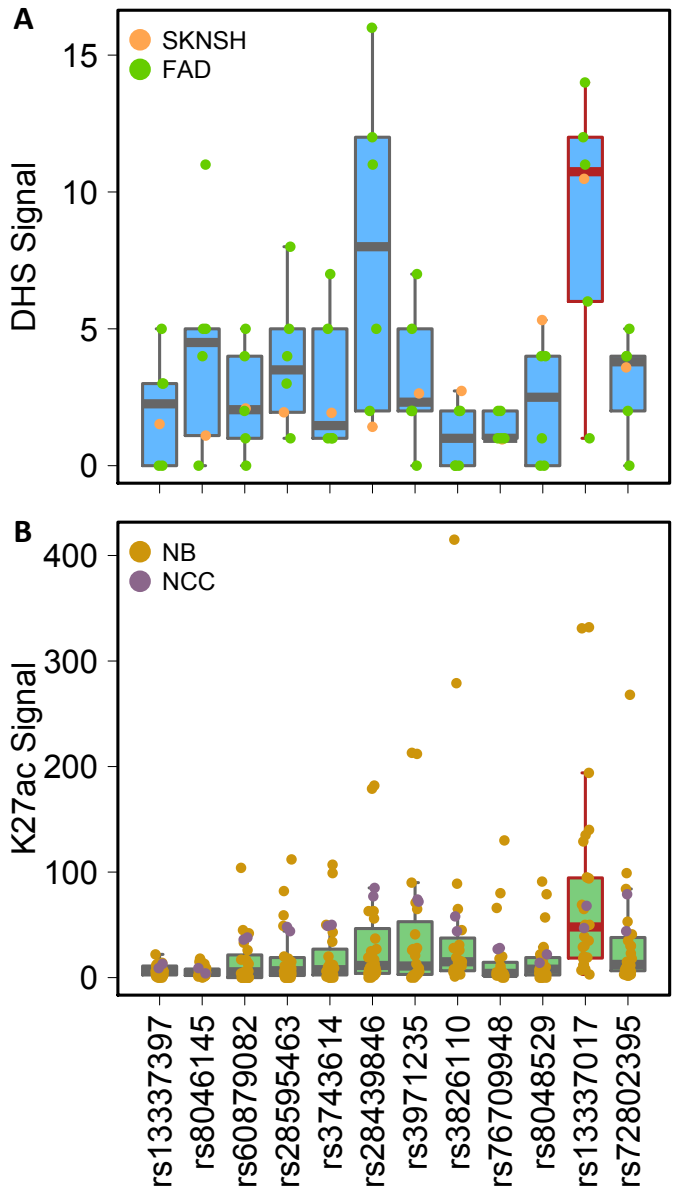

Figure S1. **The rs13337017 mapped within a regulatory region of open and active chromatin.** The figure shows boxplots reporting (**A**) the DHS and (**B**) the H3K27ac signal intensities obtained on the list of 12 prioritized SNPs. SK-N-SH: the SK-N-SH NB cell lines. FAD: Fetal Adrenal Gland. NB: the set of NB cell lines. NCC: Neural Crest Cells. The darked-bordered box indicates the SNP with the highest median value of both DHS and H3K27ac.

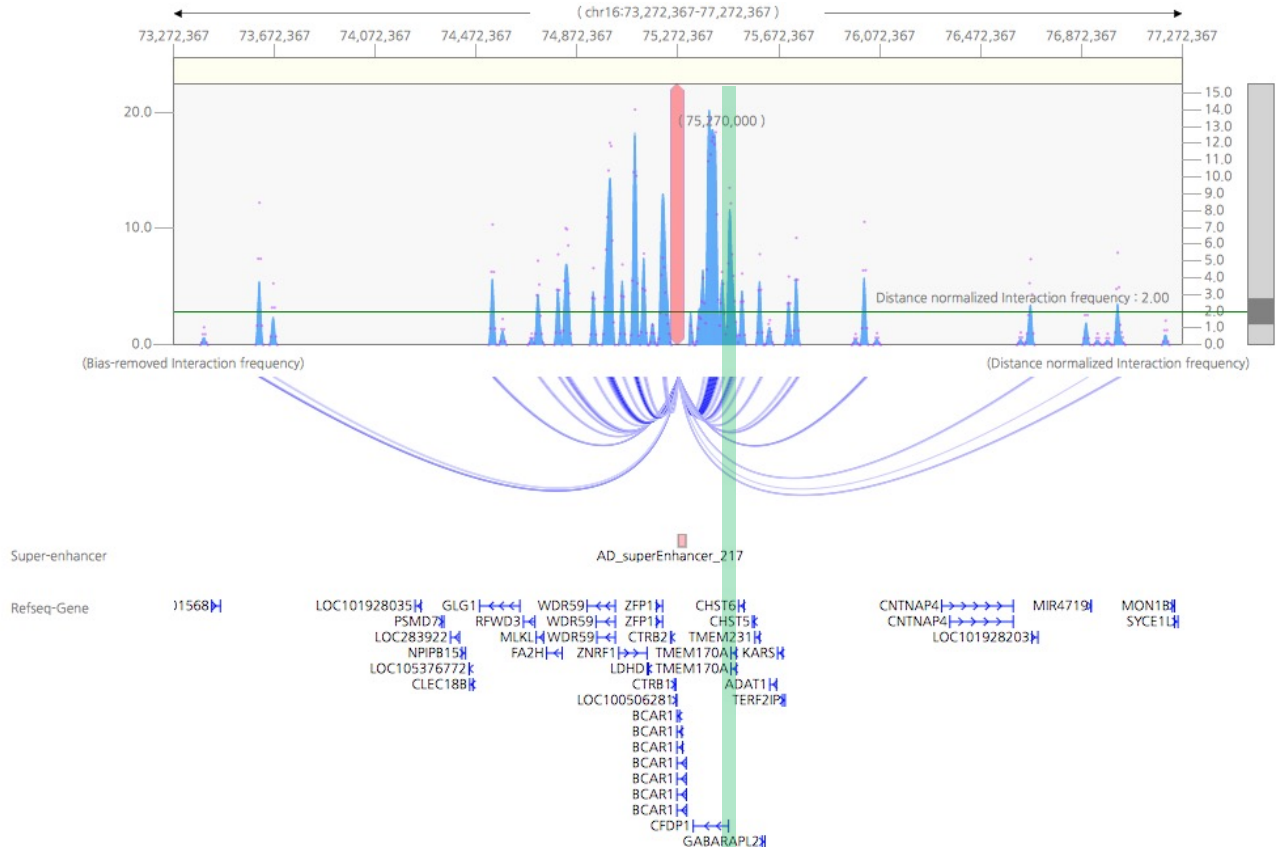

Figure S2. **rs13337017 showed long-range significant interactions with *CFPD1* in Adrenal Gland**. One-to-all interaction plot of Adrenal Gland HiC data (3DIV database) is shown from the rs13337017 point of view. Y-axes on the left and on the right indicate bias-removed interaction frequency (blue peaks) and distance-normalized interaction frequency (magenta dots), respectively. The green line indicates the cut-off for distance-normalized interaction frequency. The arcs represent significant interactions for the rs13337017. The green highlight indicates the *CFDP1* promoter.

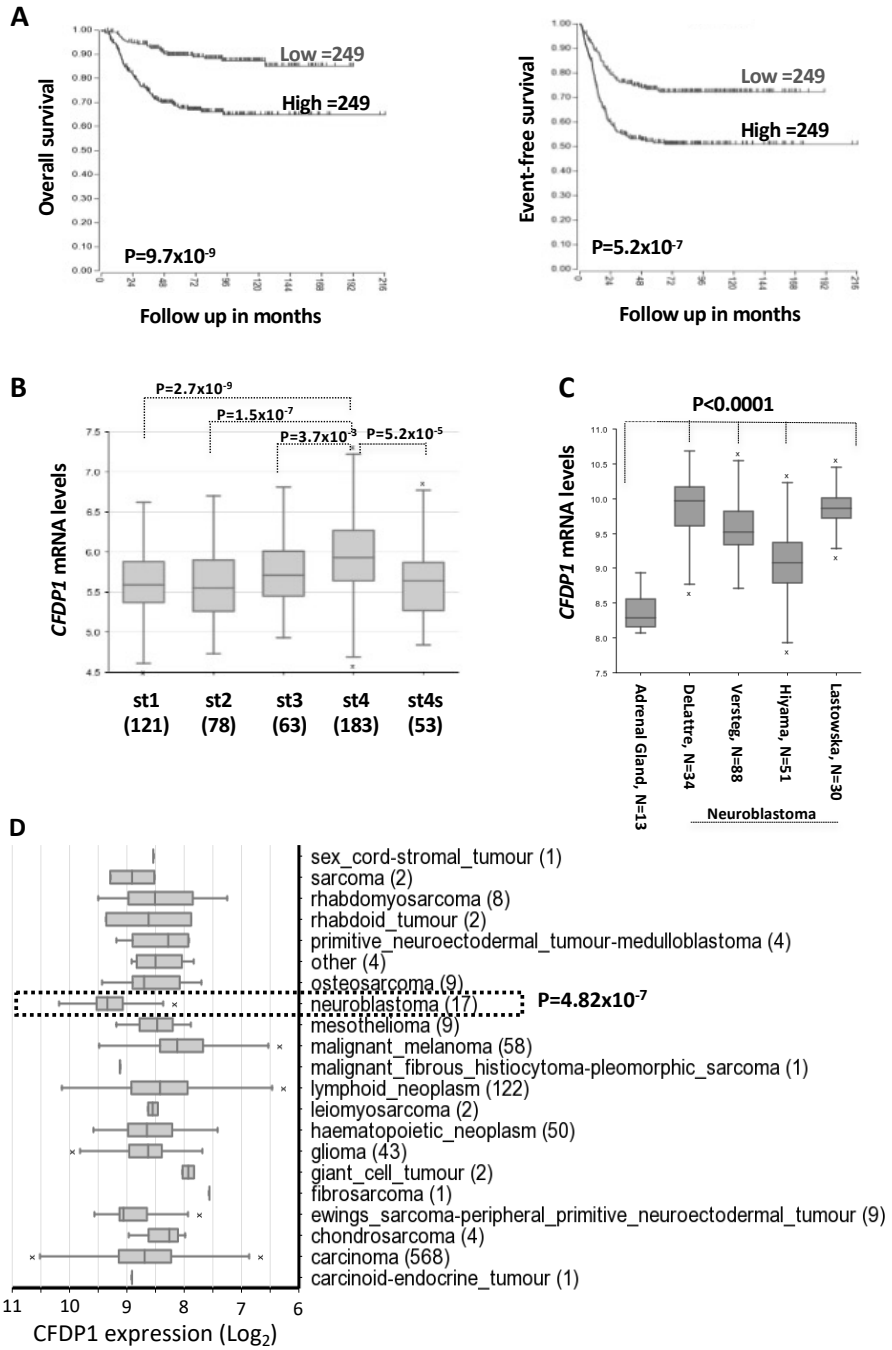

Figure S3. **High expression of *CFDP1* is associated with poor prognosis and high tumor stages in NB patients**. **A** | Kaplan-Meier curves showing the Overall and the Event-free survival probabilities based on *CFDP1* median expression in 498 NB patients (GSE62564). **B** | Boxplot reporting *CFDP1* expression in 498 NB patients (GSE62564) stratified by tumor stage. **C** | Boxplot showing *CFDP1* normalized expression in four independent sets of NB patients (a total of 281 samples) compared to normal Adrenal Gland. Delattre, N=34 (GSE14880); Versteeg, N=88 (GSE16476); Hiyama, N=51 (GSE16237); Lastowska, N=30 (GSE13136); Adrenal Gland, N=13

(GSE3526, GSE7307, GSE8514). **D** | Boxplots of *CFDP1* expression in the catalogue of the Cancer Cell Line Encyclopedia – CCLE (GSE36133). Cell lines are grouped by tumor type.

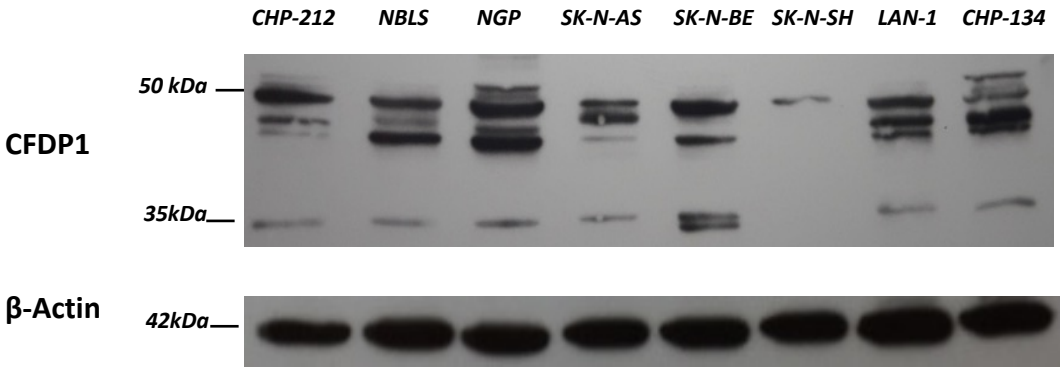

Figure S4. **Levels of CFDP1 protein in NB cell lines**. The figure shows the Western blot analysis of protein extracts from NB cells. The bands at 50 kDa and 35 kDa report the recognition of CFDP1 by a rabbit polyclonal antibody.

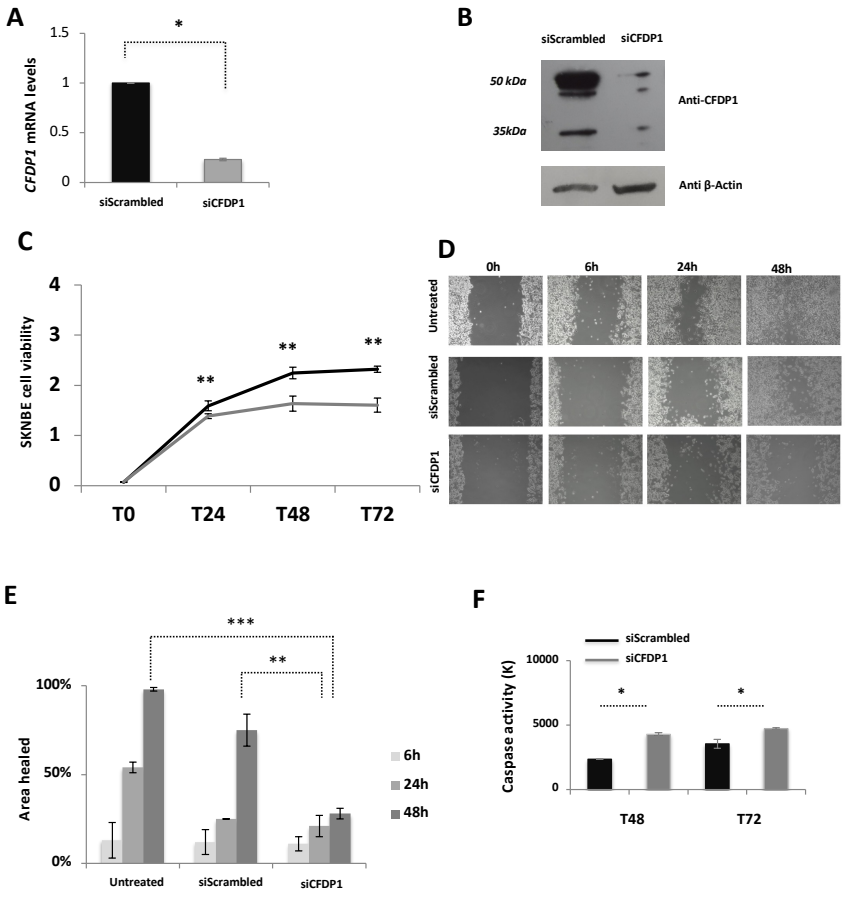

Figure S5. **The depletion of *CFDP1* affects gene expression and tumorigenicity in SK-N-BE NB cell lines. A** | Barplots showing the results of expression quantification by qRT-PCR of *CFDP1* Control (siScrambled) and *CFDP1* silenced cells (siCFDP1). **B** | CFDP1 protein levels assessed using Western Blot analysis in Scrambled and siCFDP1 cells. **C** | Line plot reporting cell viability and proliferation measured by MTT assay in Scrambled and siCFDP1 cells. In panels **A** to **C**, the data (mean of three experiments), are represented as the fold change of siCFDP1 compared to control (siScrambled) cells. **D** | Bright-light microscopy images reporting the results of wound-healing assay observed at different time points (6, 24 and 48 hours). The wound distance was measured with *ImageJ* software. **E** | Barplots reporting the quantifications of distances obtained in **D**. The experiments in panels **D** and **E** were made in triplicates and conducted on untreated, siScrambled and siCFDP1 SK-N-BE cells. **F** | Barplot of the Caspase-3 activity assay on SK-N- BE cells. The data in siCFDP1 were normalized on the negative (promoter-less) control and reported in thousands (K). Data are shown as the mean ± standard deviation from three independent transfection experiments, each done in triplicate. Statistical significance was assessed by T-test. *: P<0.05; **: P<0.001; ***: P<0.0001.

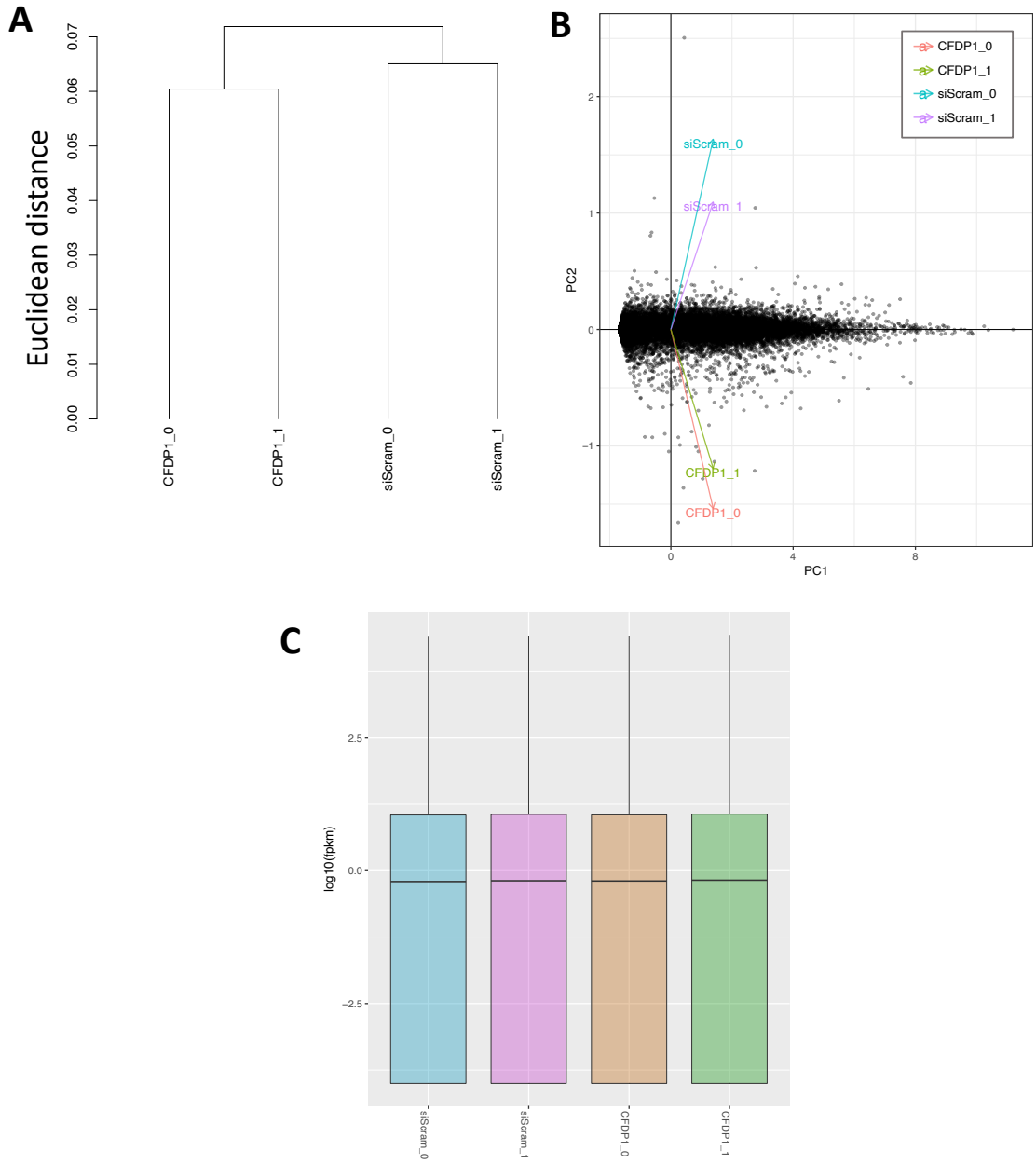

Figure S6. **Main features and quality control of RNA-seq data**. **A** | Hierarchical clustering of samples. **B** | Principal Component Analysis (PCA) **C** | Gene expression overview.

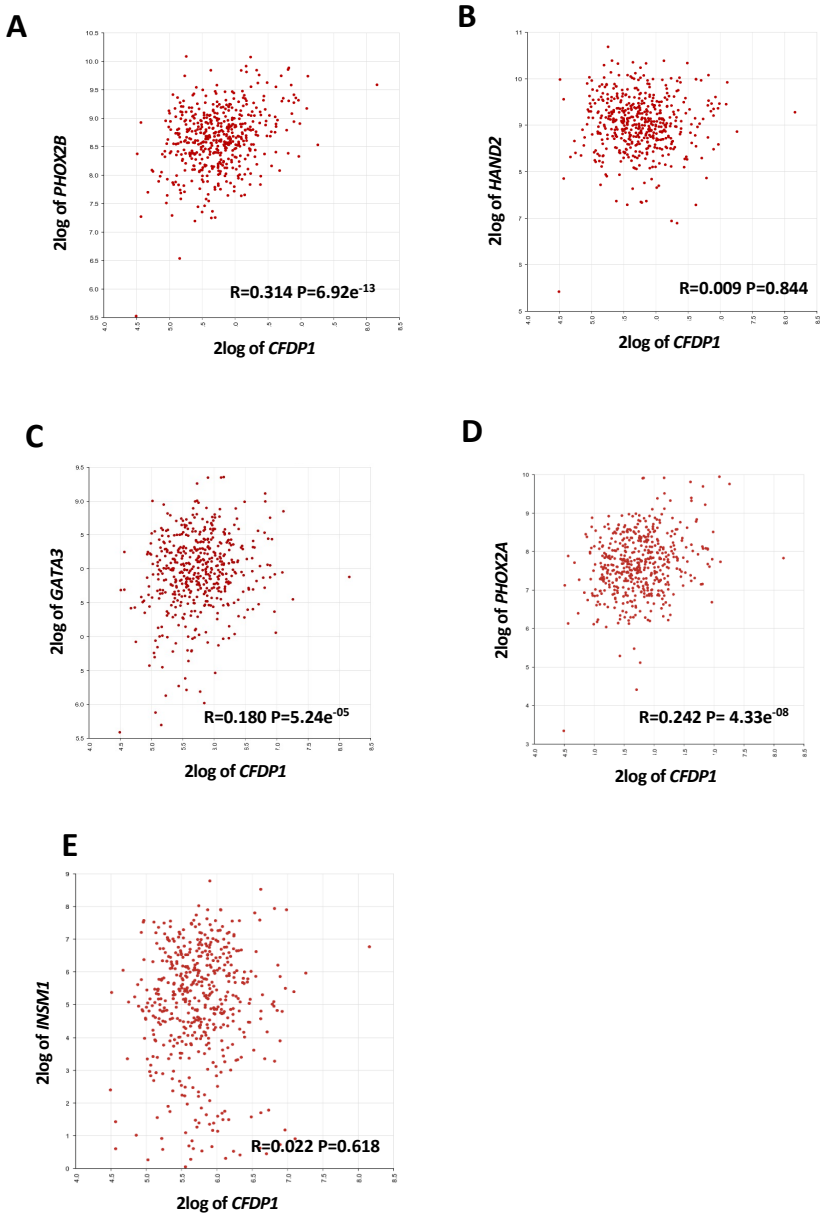

Figure S7. ***CFDP1*** **expression was correlated with that of noradrenergic transcription factors**. The figure reports scatterplots of pairwise co-expression correlations between *CFDP1* and noradrenergic-related transcription factors in 498 NB patients (GSE62564). Gene expression was compared as Log_2 normalized values. **A** | *CFDP1* versus *PHOX2B*. **B** | *CFDP1* versus *HAND2*. **C** | *CFDP1* versus *GATA3*. **D** | *CFDP1* versus *PHOX2A*. **E** | *CFDP1* versus *INSM1*. R: Pearson correlation coefficient.

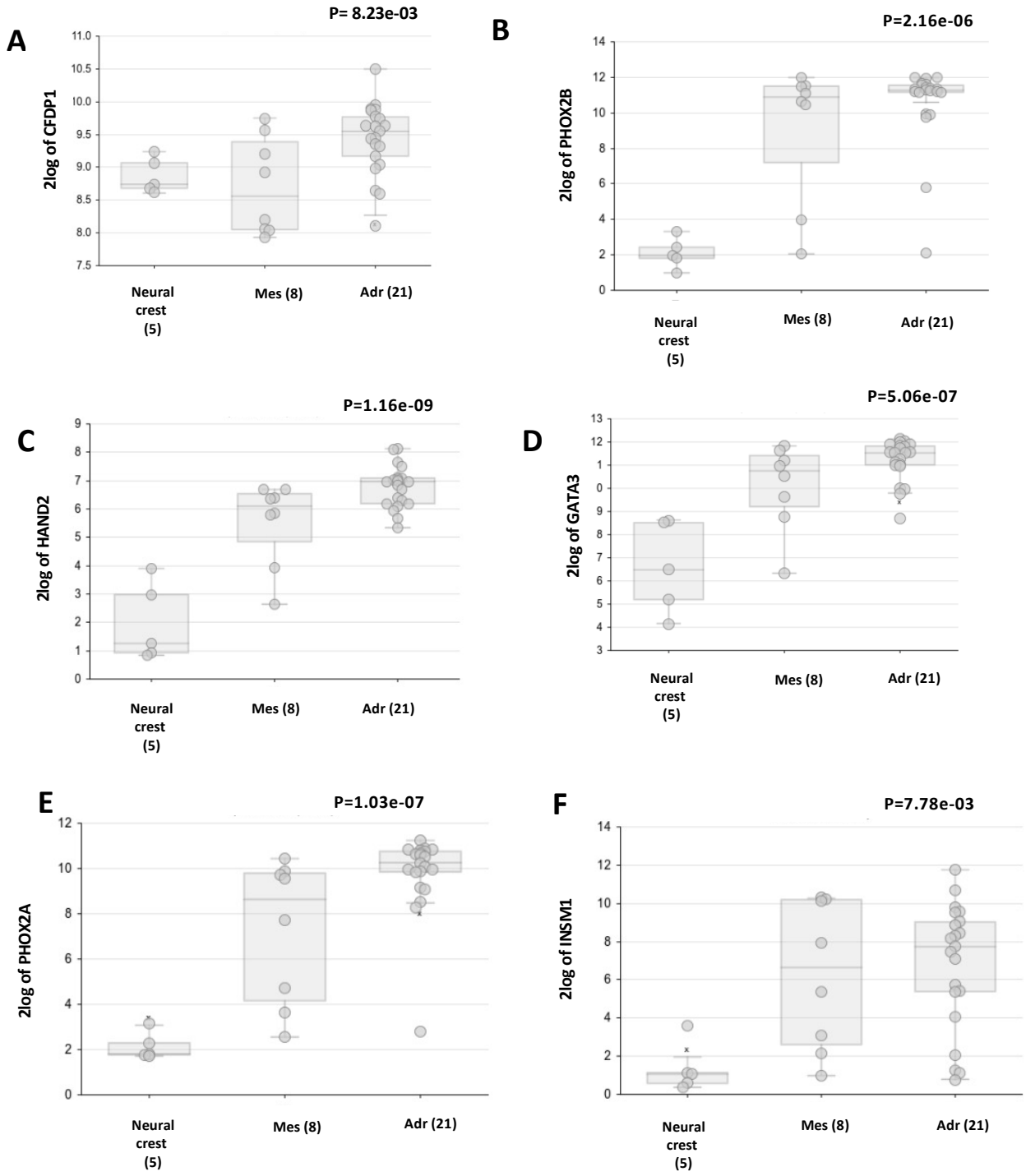

Figure S8. ***CFDP1* and noradrenergic transcription factors showed similar expression patterns**. The figure shows boxplots reporting gene expression of *CFDP1* and noradrenergic-related transcription factors in Neural Crests (n=5), Mesenchymal (n=8) and Adrenal cell lines (n=21). **A** | Expression of *CFDP1*. **B** | Expression of *PHOX2B*. **C** | Expression of *HAND2*. **D** | Expression of *GATA3*. **E** | Expression of *PHOX2A*. **F** | Expression of *INSM1*.

Supplemental Information

Genome-wide SNP genotyping data of CAD and NB

Genotype data from the GWAS of CAD, were obtained from the CARDIoGRAMplusC4D Consortium (http://www.cardiogramplusc4d.org) and were published in [1]. In detail, this data set included 8,424 cases of European ancestry, recruited by the Precocious Coronary Artery Disease Study (PROCARDIS) and Heart Protection Study (HPS), and 6,996 cases of South Asian ancestry (chiefly from Pakistan and India), recruited by the Pakistan Risk of Myocardial Infarction Study (PROMIS) and London Life Sciences Prospective Population (LOLIPOP) study. Each of these studies recruited controls, from within the same self-reported ethnic. The paper reporting the CAD GWAS data set also describes the methods trans-ethnic meta-regression of genome-wide association studies between European and South Asian individuals.

Genotyping data for 2101 NB cases and 4202 controls were published in [2].

Genotype imputation and association testing

Genotypes were phased using SHAPEIT (v2.r837) [3] and data from 1000 Genomes Phase 3 (NCBI build 37, haplotype release date October 2014). Subsequently, imputation was performed using IMPUTE2 (v2.3.2) [4] for SNPs and indel variants annotated in 1000 Genomes Phase 3. Variants with MAF<5% and/or IMPUTE2-info quality score <0.8 were removed. For population stratification, a measure of African admixture as estimated by the ADMIXTURE (v1.3.0) software [5] was used in both cohorts.

Definition of independently associated loci

To find the shared genetic risk variants between NB and CAD, we compared the whole genome results (p-value and OR for 9,671,310 SNPs) of NB and CAD GWAS data using two approaches:

a) *a meta-analysis study across NB and CAD*

b) *meta-analysis study across NB and CAD with inverted effect in CAD* by using the method of inverse variance weighted, fixed-effect meta-analysis with the METAL program [6, 7]. This step becomes necessary when, for a given SNP, the risk allele for the first phenotype is the protective allele for the other phenotype under consideration. Indeed, in order to select those SNPs, in the CAD GWAS, we conducted a fixed-effect meta-analysis with inverted Odds Ratios. We identified the 411 independent NB-CAD cross-associated SNPs according to the following algorithm: $xi \geq H+2\times(1.5xIQR)$ where H is the third quartile (Q^3) of wP values defined according to Tukey's hinge, IQR is interquartile range (Q3-Q1).

Genotyping by TaqMan assay

DNA was extracted and purified from peripheral blood according to the manufacturer's protocol by using a QIAamp tissue kit (Qiagen, Hilden, Germany). DNA concentration was determined by means of NanoDrop® (Thermo Scientific, Wilmington, DE) ND-1000 spectrophotometer, and samples were diluted to 10 ng/μL. SNP genotyping was carried out according to the TaqMan® genotyping protocol (TaqMan Universal PCR Master Mix; appliedbiosystems by Thermo Fisher Scientific) with 20-ng DNA template.

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%.

Analysis of public ChIP-Seq data

To identify potentially functional SNPs, we used multiple sources of *in silico* functional annotation from public databases. To identify SNPs that are located in human promoter genes and/or enhancer elements, we integrated NB and CAD meta-analysis data with epigenomic profiling by high throughput sequencing of the epigenetic marker H3K27ac in hNCCs GSE28874-GSM2664365-GSM2664367 [8].

Analysis of public ChIP-Seq data

H3K27ac Chromatin Immunoprecipitation Sequencing (ChIP-Seq) data published in [8, 9] and deposited in NCBI GEO (GSE90683) were downloaded and processed, starting from sequence files, following the analysis pipeline described below.

In brief, ChIP-Seq reads were mapped to the human reference genome version hg19/GRCh37 using Bowtie2 (version 2.3.4) [10]. Reads of low mapping quality ($Q \leq 20$) were discarded; duplicate reads were removed to avoid the detection of signal from genomic amplification regions. Enriched regions (peaks) were called using HMCan (version v1.30) [11] with default parameters except for large bin length, 10 Kb instead of 25 Kb. Histone Modification in Cancer (HMCan) is a Hidden Markov Model (HMM) based tool developed to detect histone modification in cancer ChIP-Seq data. It applies three correction steps to the data: copy number correction, GC bias correction and noise level correction. HMCan has two main stages: data profiling stage, and peak calling stage. In data profiling stage, HMCan constructs the density profile from the aligned reads, and then normalizes it. Normalization includes: Library size normalization, copy number normalization, GC bais normalization and noise level correction. For peak calling, HMCan estimates initial HMMs parameters using right side exact

Poisson's test. Then, it applies iterative HMMs to fine tune parameters and perform final peak calling. HMCan output included density profiles corrected for the GC content and copy number bias and peak files. To classify H3K27ac peak regions and call super-enhancers, enhancers and promoters, we used LILY [8], a modified version of ROSE [12, 13]. Briefly, the LILY tool stitches together large H3K27ac peaks using a default distance of 12.5 Kb to call a super-enhancer. If peaks are farther, it calls them as enhancers. Promoters are called if peaks are ±2.5 Kb from transcription start sites. Each region receives a score corresponding to the sum of normalized H3K27ac density values. We used those H3K27ac peak files to annotate our list of SNPs for the presence of active regulatory elements.

Other public epigenomics data

DNase-I hypersensitivity signal data of SH-N-SH cell lines were downloaded as processed files from the ENCODE catalogue (ENCSR046XHQ).

The DHS signal of Fetal Adrenal Gland (n=5) tissues from Roadmap Epigenomics Consortium were obtained (GSM530653, GSM817165, GSM1027310, GSM1027311, GSM817167) from the GEO database as processed files.

ChIP-Seq peak files of H3K27ac of hNCCs (GSM714807, GSM2664365, GSM2664367) were downloaded as processed files from the GEO database.

Cell cultures

The human HEK293T, SK-N-BE, SK-N-AS, SH-EP cell lines were obtained from the American Type Culture Collection (respectively ATCC #CRL-3216, #CRL-2268, #CRL-2137, #CRL-2269). HEK293T cell lines were grown in Dulbecco's Modified Eagle Medium (DMEM; Sigma); SK-N-BE cell line was grown in Dulbecco’s Modified Eagle Medium (DMEM; Sigma) / Ham’s F12 medium (Sigma) and SK-N-AS cell line was grown in Eagle’s Minimum Essential Medium (EMEM, Lonza) / Ham's F-12 medium (Sigma). SHEP cell line was grown in Dulbecco’s Modified Eagle Medium (DMEM; Sigma) / Ham’s F12 medium (Sigma). The mediums were supplemented with 10% heat-inactivated FBS (Sigma), 1mM L-glutamine, penicillin (100 U/ml) and streptomycin (100 µg/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.

Construction of luciferase reporter gene plasmids

The enhancer region of 963bp expanding from 827bp upstream to 135 bp downstream the variant rs13337017 was cloned upstream of the firefly luciferase gene. PCR primers contained recognition sites for KpnI in the forward (5'-AAAAAAGGTACCAGTGACTGTGCCTCTGTTCT-3') and BglII in the reverse primer (5'-AAAAAAAGATCTAGATCTGCTCTCTCCACTGAAAGACC-3'), were designed to amplify 963 bp from the genomic DNA of cell lines heterozygous for the rs13337017 C/T allele. After cutting the fragment with KpnI and BglII restriction enzyme (Biolabs) we cloned it into the pGL3-Promoter-Vector (Promega). The resulting plasmids containing the rs13337017-C allele and rs13337017-T were obtained. The sequence of each construct was confirmed by direct sequencing.

In vitro functional analysis

HEK293T cells were transfected using Trasfectin (Biorad), and SK-N-AS cells using X-tremeGENE (Roche) with 1 ug of pGL3-Promoter-Vector rs13337017-C and rs13337017-T constructs. Cells were subsequently starved in serum-free medium for 8 hrs. Fifteen nanograms pRL-TK Vector (Promega) was co-transfected as a normalizing control. Cells were induced to re-enter the cell cycle by the addition of fresh medium supplemented with 10% FBS for 12 and 24 hrs. At these time-points, the cells were harvested, lysed and analyzed for luciferase activity using the Dual-Luciferase Reporter Assay System (Promega) on a TD20/20 Luminometer (Turner Designs). Results are reported as relative luciferase activities, which are obtained by dividing firefly luciferase activity with Renilla luciferase activity. Data represent the means ± S.D. of three independent transfections.

Treatment of cells with siRNA *CFDP1*

Gene-specific siRNA-27 duplexes (pooled siRNA A-B-C) Trilencer-27 for *CFDP1* (Human) (origene, Locus ID 10428) and Universal Scrambled Negative Control siRNA Duplex (origene) were used at a concentration of 10 nM after a reconstitution step by using a resuspension Duplex Buffer (Origene).

SK-N-BE, SK-N-AS and SH-EP cells were plated into a 6well at 70% confluent. Oligos were transfected with X- tremeGENE HP DNA Transfection Reagent (Roche). The ratio selected to achieve >90% transfection efficiency between siRNAs (10nM) and X-tremeGENE reagent was 1:3, respectively.

After 48 h of transfection cells were harvested to assess the silencing of protein and mRNA of *CFDP1*. The experiments were performed in triplicate and for each experiment three experimental points were analyzed.

Cell viability assay

SK-N-BE, SK-N-AS and SH-EP cells transfected with siRNA-*CFDP1* and siRNA-Scrambled were seeded as six replicates into 96-well plates at a density of 104 cells per well. The cells were then incubated for a further time course (0 - 24h - 48h -72h). At each correspondent time point the 3-(4,5 dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay was added to a final concentration of 0.5 µg/ml according to the manufacture protocol (Promega, Milan, Italy). Plates were analyzed with EnVision multimode plate reader (Perkinelmer). The experiments have been repeated twice.

Wound healing assay

The appropriate number of SK-N-BE and SK-N-AS cells were placed in a 12-well plate and were transfected with the previously described siRNA in order to reach 100% confluence in 24 hours. After 24 hours from transfection, in a sterile environment, a 200-µl pipette tip was used to press firmly against the top of the tissue culture plate and swiftly to make a vertical wound down through the cell monolayer. Following the generation and inspection of the wound initial pictures of each single well, in triplicate, were taken (time 0). Images were taken using bright-light microscopy also in different time points (6 h, 24 h, 48 h). To analyse the results of snapshot pictures, was measured the distance of one side of the wound to the other using a scale bar with *ImageJ* software. All experiment was made in triplicate.

Caspase-3 activity assay

Caspase-3 activity was assayed in SK-N-BE and SK-N-AS cell lines previously transfected with siRNA-*CFDP1* and siRNA-Scrambled. Then, cells were lysed by 200 ul of freeze-thawing lysis buffer (50mM HEPES pH 7,4; 0,1% CHAPS; 1mM DTT; 0,1 mM EDTA; 0,1% NP40). The cell lysate was centrifuged at 13.000 g for 5 minutes at 4°C and the supernatant obtained was assayed for caspase-3 activity using the ENZCHEK caspase-3 assay kit with a DEVD-AMC substrate (Molecular Probes, Eugene, OR, USA) according to the manufacturer's instructions.

Real-time RT-PCR

The expression levels of *CFDP1* and *GAPDH* were analysed using real-time, quantitative PCR. RNA concentration and purity were evaluated by measuring the optical density at 280 and 260 nM using the Nanodrop ND-8000. One microgram of total RNA was reverse transcribed into cDNA using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). The cDNA samples were diluted to 20 ng/µl. Gene-specific primers were designed by using OLIGO 7 primer analysis software

[14] as: *CFDP1* Forward (F) 5'-GCACCCTTGAGAAGTCCAAA-3' and *CFDP1* Reverse (R) 5'-TTCCGTTCAATGTACCCCTC-3'; *GAPDH* (F) 5'- CCACATCGCTCAGACACCAT-3', (R) 5'-AGTTAAAAGCAGCCCTGGTGAC-3'. The amplicon generated for *CFDP1* mRNA was 109 bp, while the amplicon generated for *GAPDH* was 91bp. Real-time PCR was performed using SYBR Green PCR Master Mix (AppliedBiosystems). All real-time PCR reactions were performed using the 7900HT Fast Real-Time PCR System (Applied Biosystems). The experiments were carried out in triplicate for each data point. The housekeeping gene *GAPDH* was used as the internal control. Relative gene expression was calculated using the $2-\Delta CT$ method, where the ΔCT was calculated using the differences in the mean CT between the selected gene and the internal control (*GAPDH*). The mean fold change of $2-(\text{average } \Delta\Delta CT)$ was determined using the mean difference in the ΔCT between the gene of interest and the internal control.

Western blotting

Cell pellets were resuspended and lised in a RIPA buffer (10 mM Tris-Cl pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS and 140 mM NaCl) in the presence of a protease inhibitors cocktail (Roche). The protein concentrations were determined by Bradford assays (Bio-Rad). Thirty micrograms of protein were loaded and separated using 10% polyacrylamide gels, and transferred onto polyvinylidene difluoride membranes (Bio-Rad). The membranes were blocked with 5% BSA (Sigma) in phosphate-buffered saline (PBS) with 0.1 % Tween (PBS-T) for 1 h, and then probed with Anti-Cfdp1 Antibody (TA340096; Origene) antibody, diluted 1:500 with 5% nonfat dried milk in 1% PBS buffer containing 0.2 % Tween 20 and 0,02 % Sodium Azid. After a wash in PBS-T, the membranes were incubated with horseradish-peroxidase-conjugated anti-rabbit secondary antibody (1:2,000 dilution; ImmunoReagent), and then the positive bands were visualized using the ECL kit SuperSignal West Pico Chemiluminescent Substrate (Pierce). A mouse anti β -actin antibody (1:10000 dilution; A2228; Sigma) was used as the control for equal loading. The digital image was analysed using the public domain *ImageJ* software. Briefly, the digital image was opened in *ImageJ* software, the bands were outlined using a squared region tool and added to ROI (Region of Interest) manager. The background was subtracted and the analysis tool was used to measure the integrated optical density (IOD) of the individual bands.

cDNA library construction for RNA sequencing

Transcriptome high-throughput sequencing was performed in the control group (SK-N-AS cells transfected with siScrambled alone, Samples-ID: siScrambled SK-N-AS) and the treatment group (SK-N-AS cells transfected with CFDP1 siRNA, Samples-ID; siCFDP1 SK-N-AS). Total RNA was

isolated from SK-N-AS cells using TRIzol (ThermoFisher Scientific, Waltham, MA, USA) according to manufacturer's instructions. RNA purity was checked using the NanoPhotometer spectrophotometer (Implen, Inc., Westlake Village, CA, USA). RNA concentration was measured using the NanoDrop ND1000 spectrophotometer (Nano-Drop, Wilmington, DE, USA). RNA with an OD260/280 between 1.8 and 2.2 and an OD260/230 ≥ 1.8 was used for the construction of cDNA libraries. RNA integrity was assessed using Agilent 2100 Bioanalyzer RNA Nano chip device (Agilent, Santa Clara, CA, USA).

In total, one μg RNA per sample was used as input material for RNA sample preparations. This study included two groups of two biological replicates.

The mRNA content was concentrated from total RNA using RNase-free DNase I (TaKaRa) and magnetic oligo (dT) beads. The mRNA was mixed with the fragmentation buffer and broken into short fragments (~200 bp long). Then, the first strand of cDNA was synthesized with random hexamer primers. The second strand was synthesized using the SuperScript Double-Stranded cDNA Synthesis kit (Invitrogen, Camarillo, CA) and was purified via magnetic beads. The ends were repaired and tailed with a single 3' adenosine. Subsequently, the cDNA fragments were ligated to sequencing adapters. The RNA-Seq was performed on an Illumina platform.

Brefly, we obtained about 25 milions of high-quality reads (Q20 = 96.51%) per sample and the alignment pipeline returned an overall mapping rate of 87.48%. Based on whole transcriptome normalized FPKM counts, we first assessed the reliability of our data in terms of distance between the two experimental conditions. The hierarchical clustering of samples clearly showed two main branches of the dendrogram separating silenced and control cells (**Supplemental Figure 6A**). The Principal Component Analysis (PCA) confirmed the diversity between experimental conditions and the similarity between replicates (**Supplemental Figure 6B),** as expected, we did not observe significant differences in gene expression from a global point of view (**Supplemental Figure 6C**).

Quantitative real-time PCR assay

For validation of the transcriptome result, we subjected seven significantly differential expressed unigenes on related pathways to qRT-PCR analysis. Redundant RNA from the cDNA library preparation was used to perform reverse transcription according to the Invitrogen protocol. qRT-PCR were performed as described above. The primers used in qRT-PCR assays are listed in **Supplemental information table 1**.

Supplemental information table 1. Primer sequences for qRT-PCR

Gene	PRIMER FORWARD	PRIMER REVERSE
GADD45G	CCGGAAAGCACAGCCAGGAT	ACAGAAGGTCACATTGTCGG
MAFF	GCTCTAAAGATCAAGCGAGA	CACGGTTTTTGAGTGTGCGG
FOSL2	AGATGAGCAGCTGTCTCCTG	TCTGTCTCCGCCTGCAGCTT
CCL20	AGTTGTCTGTGTGCGCAAAT	TTCTGTTCTTGGGCTATGTC
PHOX2A	TACTCGGCAGTGCCCTACAA	CTCACGCGTGTAAATGTCGG
TFAP2B	GTGTTTTGCTCCGTCCCAGG	TTGGCTCTTCTGAGGACTCC
INSM1	ACTTCGGCAACCCCGAGGCT	ACTTGTGCTTCTCGTGCTCG
PHOX2B	GAGTCCAGGTGTGGTTCCAG	GTCAGTGCTCTTGGCCTCTT
HAND2	CCAGCTACATCGCCTACCTC	TTCAAGATTTCGTTCAGCTC
GATA3	TACGGAAACTCGGTCAGGG	GAAGGGGCTGAGATTCCAGG

Supplemental References

1. Coronary Artery Disease Genetics, C., *A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease.* Nat Genet, 2011. **43**(4): p. 339-44.
2. McDaniel, L.D., et al., *Common variants upstream of MLF1 at 3q25 and within CPZ at 4p16 associated with neuroblastoma.* PLoS Genet, 2017. **13**(5): p. e1006787.
3. Delaneau, O., J. Marchini, and J.F. Zagury, *A linear complexity phasing method for thousands of genomes.* Nat Methods, 2011. **9**(2): p. 179-81.
4. Howie, B., et al., *Fast and accurate genotype imputation in genome-wide association studies through pre-phasing.* Nat Genet, 2012. **44**(8): p. 955-9.
5. Alexander, D.H., J. Novembre, and K. Lange, *Fast model-based estimation of ancestry in unrelated individuals.* Genome Res, 2009. **19**(9): p. 1655-64.
6. Chung, D., et al., *GPA: a statistical approach to prioritizing GWAS results by integrating pleiotropy and annotation.* PLoS Genet, 2014. **10**(11): p. e1004787.
7. Sivakumaran, S., et al., *Abundant pleiotropy in human complex diseases and traits.* Am J Hum Genet, 2011. **89**(5): p. 607-18.
8. Boeva, V., et al., *Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries.* Nat Genet, 2017. **49**(9): p. 1408-1413.
9. Rada-Iglesias, A., S.L. Prescott, and J. Wysocka, *Human genetic variation within neural crest enhancers: molecular and phenotypic implications.* Philos Trans R Soc Lond B Biol Sci, 2013. **368**(1620): p. 20120360.
10. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.
11. Ashoor, H., et al., *HMCan: a method for detecting chromatin modifications in cancer samples using ChIP-seq data.* Bioinformatics, 2013. **29**(23): p. 2979-86.
12. Whyte, W.A., et al., *Master transcription factors and mediator establish super-enhancers at key cell identity genes.* Cell, 2013. **153**(2): p. 307-19.
13. Loven, J., et al., *Selective inhibition of tumor oncogenes by disruption of super-enhancers.* Cell, 2013. **153**(2): p. 320-34.
14. Rychlik, W., *OLIGO 7 primer analysis software.* Methods Mol Biol, 2007. **402**: p. 35-60.