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)

Supplemental information (pages 11-18)

Supplemental references (page 19)

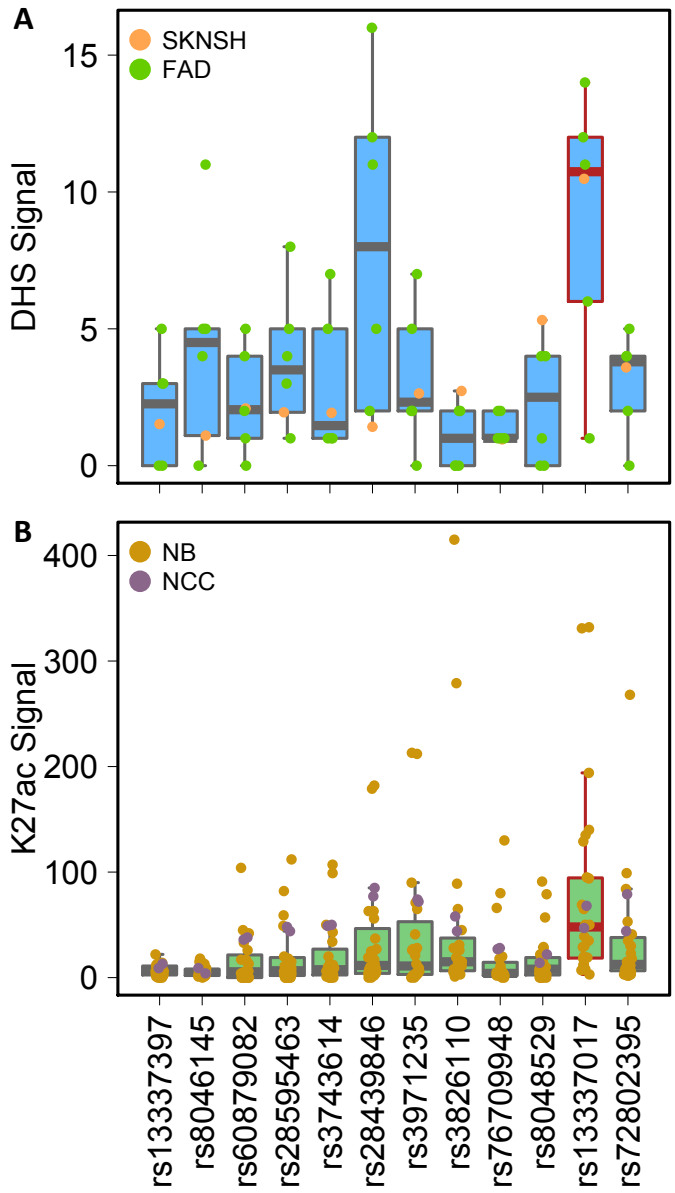

Figure S1. **The rs13337017 mapped within a regulatory region of open and active chromatin.** The figure shows boxplots reporting (**A**) the DHS and (**B**) the H3K27ac signal intensities obtained on the list of 12 prioritized SNPs. SK-N-SH: the SK-N-SH NB cell lines. FAD: Fetal Adrenal Gland. NB: the 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.

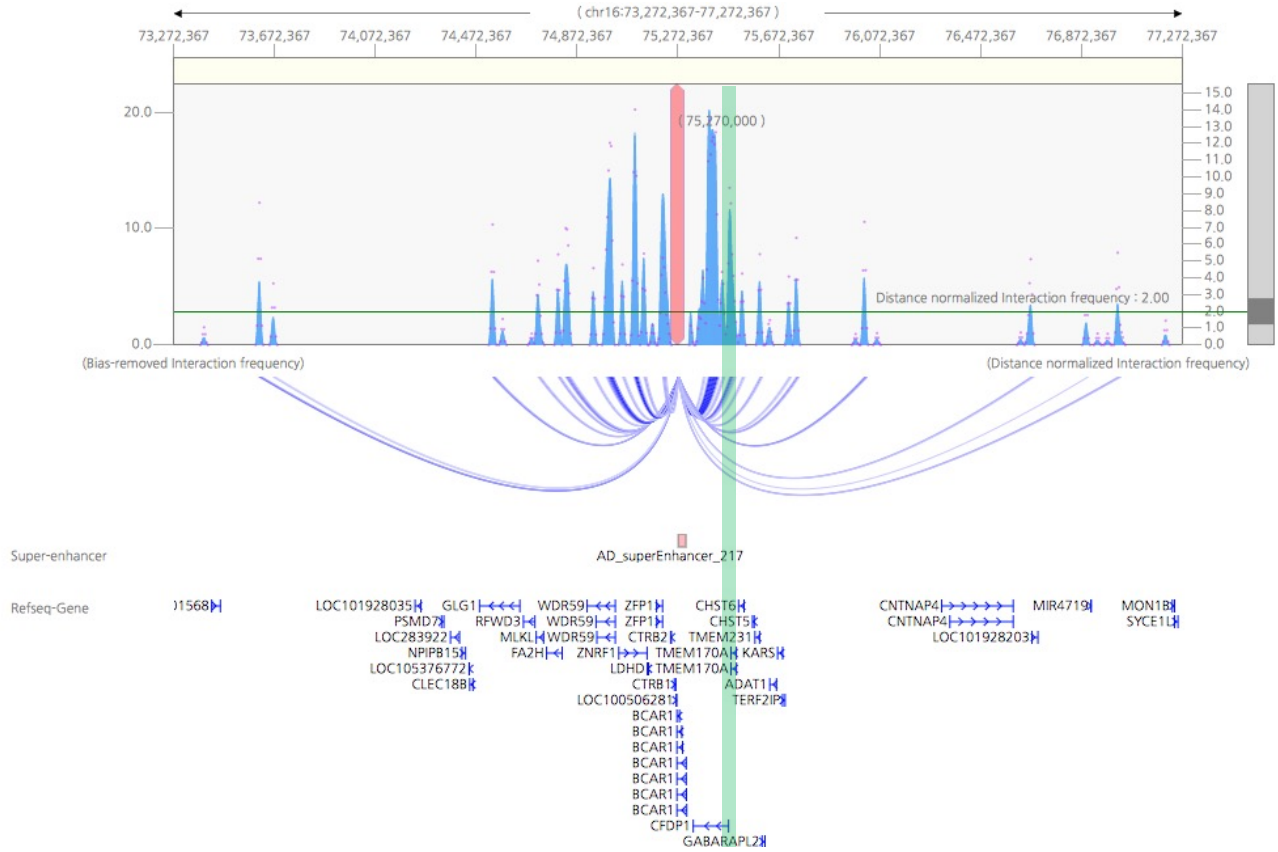

Figure S2. **rs13337017 showed long-range significant interactions with *CFPD1* in Adrenal Gland**. One-to-all interaction plot of Adrenal Gland HiC data (3DIV database) is shown from the rs13337017 point of view. Y-axes on the left and on the right indicate bias-removed interaction frequency (blue peaks) 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.

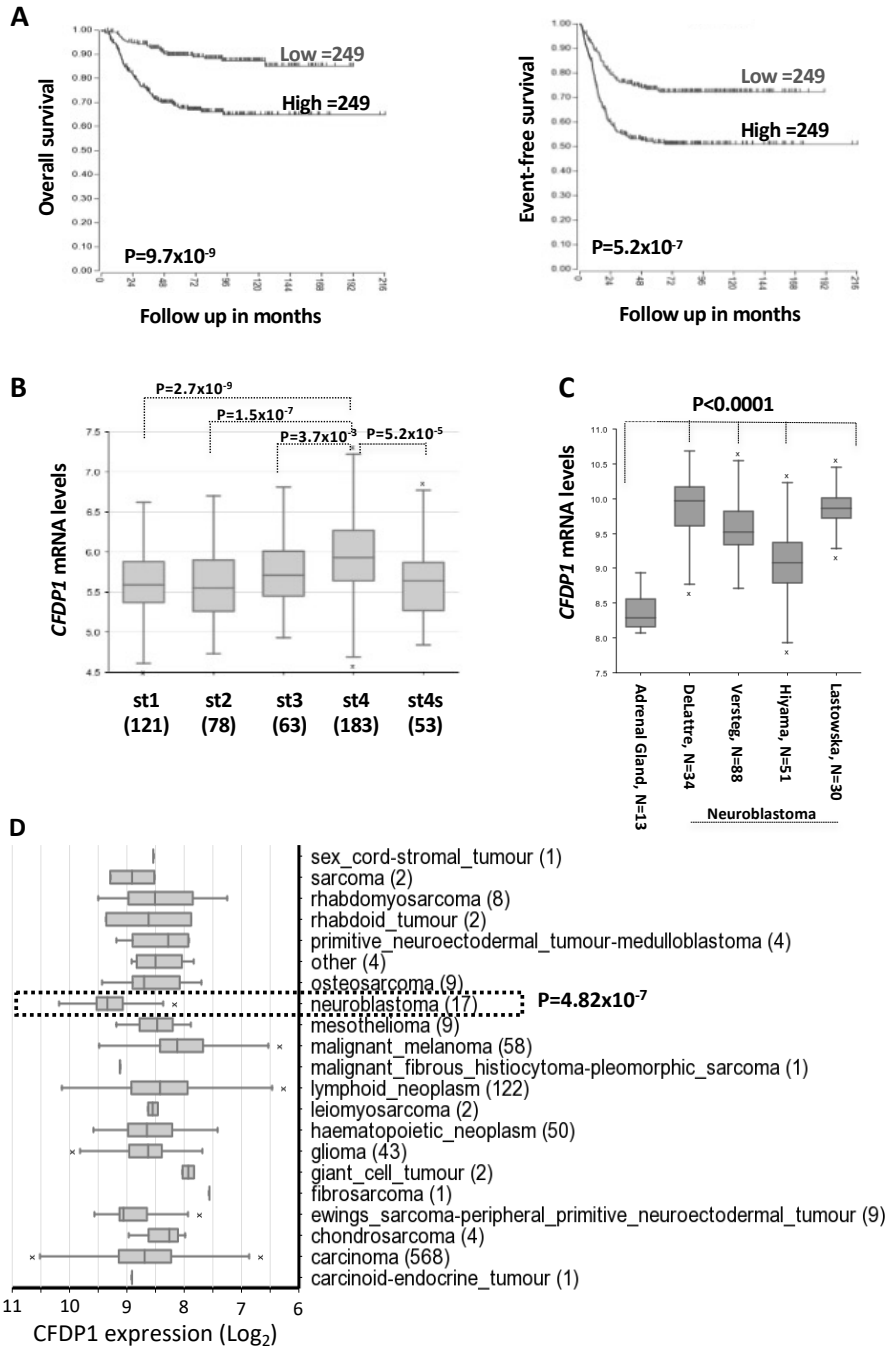

Figure S3. **High expression of *CFDP1* is associated with poor prognosis and high tumor stages in NB patients**. **A** | Kaplan-Meier curves showing the Overall and the Event-free survival probabilities based on *CFDP1* median expression in 498 NB patients (GSE62564). **B** | Boxplot reporting *CFDP1* 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.

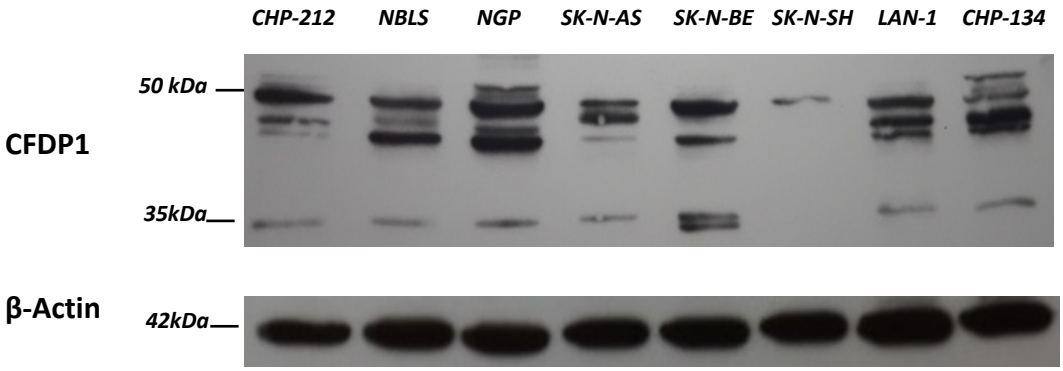

Figure S4. **Levels of CFDP1 protein in NB cell lines**. The figure shows the Western blot analysis of protein extracts from NB cells. The bands at 50 kDa and 35 kDa report the recognition of CFDP1 by a rabbit polyclonal antibody.

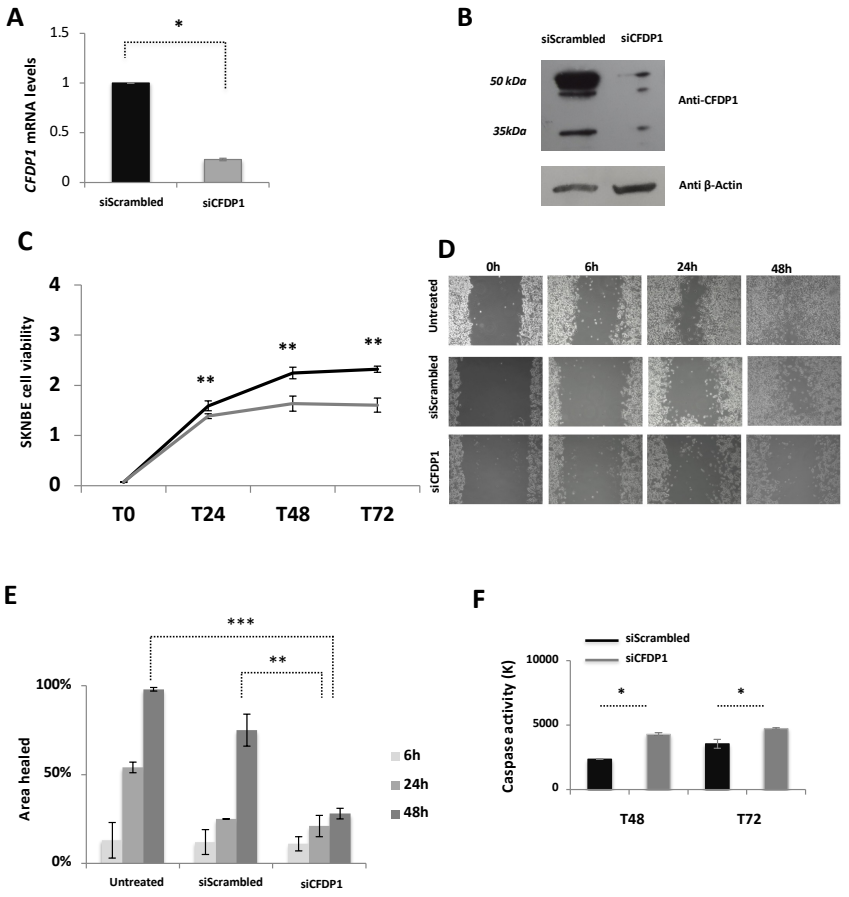

Figure S5. **The depletion of *CFDP1* affects gene expression and tumorigenicity in SK-N-BE NB cell lines. A** | Barplots showing the results of expression quantification by qRT-PCR of *CFDP1* Control (siScrambled) and *CFDP1* silenced cells (siCFDP1). **B** | CFDP1 protein levels assessed using Western 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.

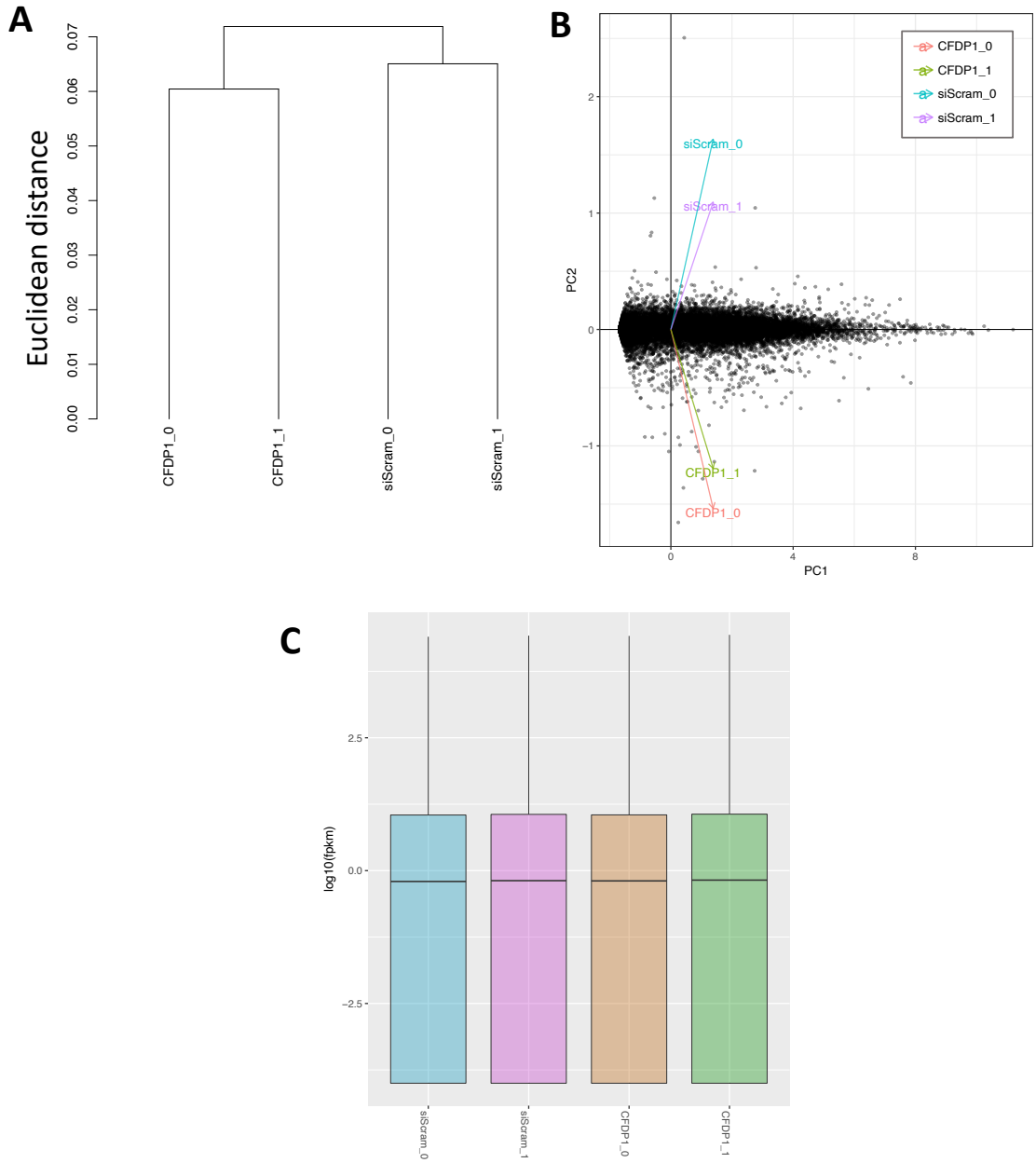

Figure S6. **Main features and quality control of RNA-seq data**. **A** | Hierarchical clustering of samples. **B** | Principal Component Analysis (PCA) **C** | Gene expression overview.

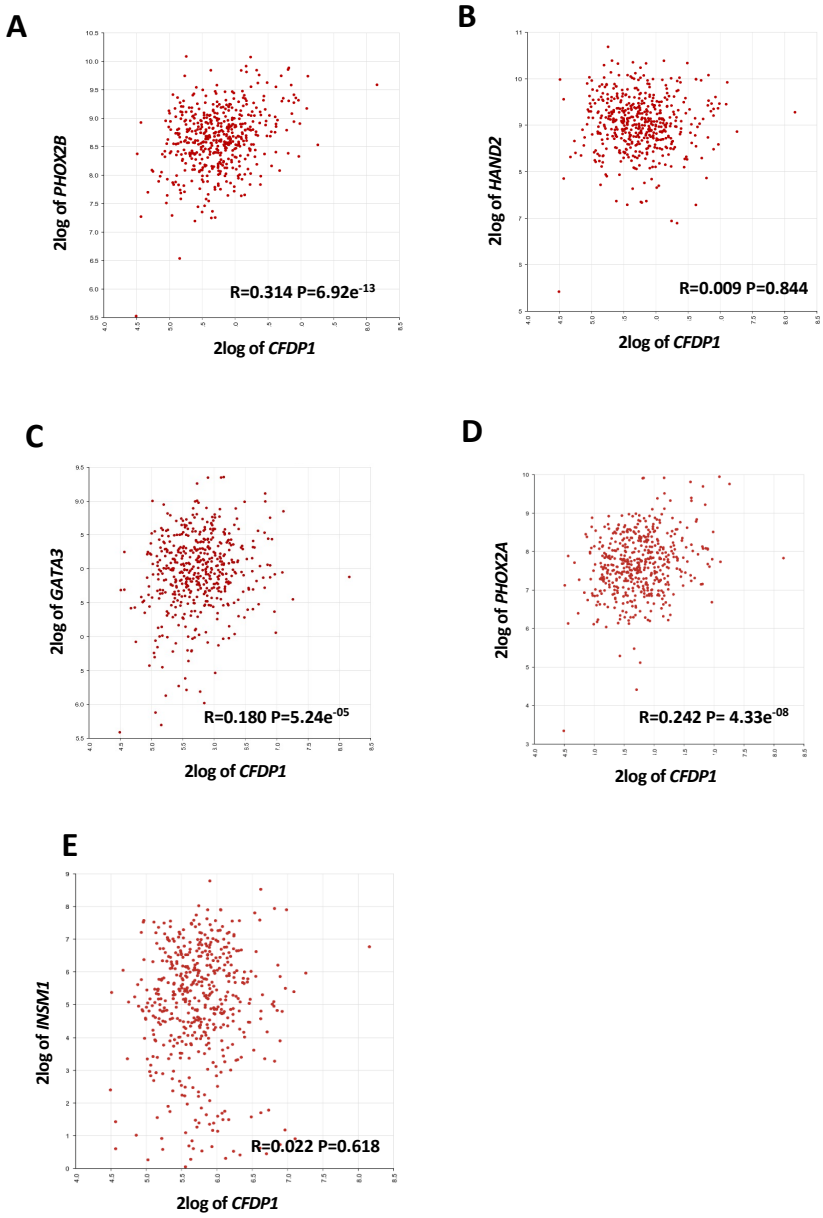

Figure S7. ***CFDP1*** **expression was correlated with that of noradrenergic transcription factors**. The figure reports scatterplots of pairwise co-expression correlations between *CFDP1* and noradrenergic-related transcription factors in 498 NB patients (GSE62564). Gene expression was compared as 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.

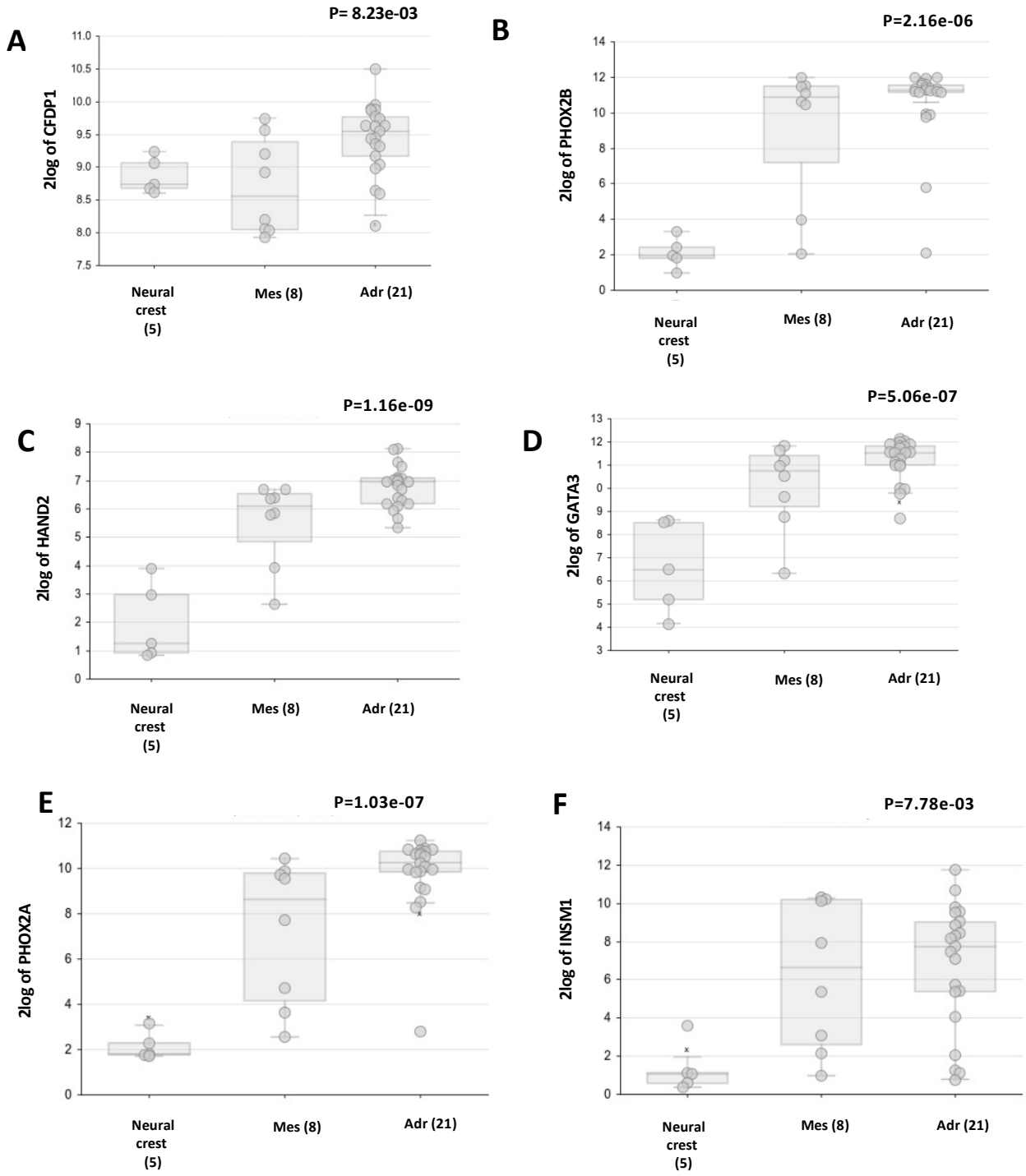

Figure S8. ***CFDP1* and noradrenergic transcription factors showed similar expression patterns**. The figure shows boxplots reporting gene expression of *CFDP1* and noradrenergic-related transcription factors in Neural Crests (n=5), Mesenchymal (n=8) and Adrenal cell lines (n=21). **A** | Expression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