CANCER RESEARCH | GENOME AND EPIGENOME

Transcription Factors Involved in Tumorigenesis Are Over-Represented in Mutated Active DNA-Binding Sites in Neuroblastoma [A][C]

Mario Capasso[1,2,3], Vito Alessandro Lasorsa[1,2], Flora Cimmino[1,2], Marianna Avitabile[1,2], Sueva Cantalupo[3], Annalaura Montella[1,2], Biagio De Angelis[4], Martina Morini[5], Carmen de Torres[6,†], Aurora Castellano[4], Franco Locatelli[4,7], and Achille Iolascon[1,2]

ABSTRACT

The contribution of coding mutations to oncogenesis has been largely clarified, whereas little is known about somatic mutations in noncoding DNA and their role in driving tumors remains controversial. Here, we used an alternative approach to interpret the functional significance of noncoding somatic mutations in promoting tumorigenesis. Noncoding somatic mutations of 151 neuroblastomas were integrated with ENCODE data to locate somatic mutations in regulatory elements specifically active in neuroblastoma cells, nonspecifically active in neuroblastoma cells, and nonactive. Within these types of elements, transcription factors (TF) were identified whose binding sites were enriched or depleted in mutations. For these TFs, a gene expression signature was built to assess their implication in neuroblastoma. DNA- and RNA-sequencing data were integrated to assess the effects of those mutations on mRNA levels. The pathogenicity of mutations was significantly higher in transcription factor binding site (TFBS) of regulatory elements specifically active in neuroblastoma cells,

as compared with the others. Within these elements, there were 18 over-represented TFs involved mainly in cell-cycle phase transitions and 15 under-represented TFs primarily regulating cell differentiation. A gene expression signature based on over-represented TFs correlated with poor survival and unfavorable prognostic markers. Moreover, recurrent mutations in TFBS of over-represented TFs such as EZH2 affected MCF2L and ADP-ribosylhydrolase like 1 expression, among the others. We propose a novel approach to study the involvement of regulatory variants in neuroblastoma that could be extended to other cancers and provide further evidence that alterations of gene expression may have relevant effects in neuroblastoma development.

Significance: These findings propose a novel approach to study regulatory variants in neuroblastoma and suggest that noncoding somatic mutations have relevant implications in neuroblastoma development.

Introduction

As most high-throughput sequencing studies of cancer focused mainly on the protein-coding part of the genome, the role of somatic mutations in regulatory regions [i.e., transcription factor binding site (TFBS)] remains underestimated. The most recent literature clearly demonstrates that altered transcriptional regulatory circuits play

[1]Department of Molecular Medicine and Medical Biotechnology, Università degli Studi di Napoli Federico II, Napoli, Italy. [2]CEINGE Biotecnologie Avanzate, Napoli, Italy. [3]IRCCS SDN, Napoli, Italy. [4]Department of Pediatric Haematology and Oncology, IRCCS Ospedale Pediatrico Bambino Gesù, Roma, Italy. [5]Laboratory of Molecular Biology, IRCCS Istituto Giannina Gaslini, Genova, Italy. [6]Developmental Tumor Biology Laboratory, Department of Oncology, Hospital Sant Joan de Déu, Barcelona, Spain. [7]Department of Paediatrics, Sapienza University of Rome, Roma, Italy.

Corresponding Authors: Mario Capasso, University of Naples Federico II, Via G. Salvatore, 486, Napoli 80145, Italy. Phone: 3908-1373-7889; Fax: 3908-1373-7804; E-mail: mario.capasso@unina.it; and Achille Iolascon, achille.iolascon@unina.it

relevant roles in cancer development (1). For instance, a reanalysis of sequencing data from 493 tumors found somatic mutations in TFBSs under positive selection, consistent with the fact that these loci regulate important cancer cell functions (2). Another recent study has demonstrated that noncoding mutations can affect the gene expression of target genes in a large number of tumors (3). Promoter mutations can either create ETS-binding sites activating transcription of oncogenes such as *TERT* (4) or disrupt the same binding sites disabling transcription of tumor suppressors such as *SDHD* (5).

Despite these recent advances, it remains a challenge to establish the pathogenic repercussions of noncoding variants in cancer development. First, noncoding variants can alter a number of functions including transcriptional and posttranscriptional regulation. Second, noncoding regions accumulate mutations at higher rates than coding regions because of a weaker selective pressure (5). As a result, distinguishing driver noncoding mutations from passengers becomes a difficult statistical and computational task. Third, it is challenging to computationally predict whether a noncoding variant affects gene expression or mRNA stability because the logic involved in regulatory element function has not yet been fully elucidated. In this respect, data from ENCODE and Roadmap projects, which map regulatory elements in several tissues and cell lines can facilitate functional interpretation of noncoding somatic mutations.

Neuroblastoma is the most common pediatric, extracranial solid tumor of neural crest origin accounting for the 8%–10% of all pediatric cancers. Despite decades of international efforts to improve the outcome, long-term survivors of high-risk neuroblastoma are about 30% (6). Genomic and transcriptomic biomarkers, including *MYCN* amplification, chromosomal aberrations, and gene expression

AACR American Association for Cancer Research·

signatures are used to stratify patients into different risk groups (6–8). Next-generation sequencing studies of neuroblastoma have documented low somatic mutation rates and few recurrently mutated genes. Indeed, only mutations in *ALK, ATRX*, and *TERT* have been identified as the most frequent genetic abnormalities (9–12). Recent studies have shown that disease-associated genomic variation is commonly located in regulatory elements in the human population (13). Our genome-wide association studies using DNA from nondisease cells (blood) have revealed that many genetic variants that are associated with neuroblastoma susceptibility lie in noncoding regions of the genome (14–16). Functional investigations have shown that the cancer genes *LMO1* (17), *BARD1* (16), *LIN28B* (18), whose expression is affected by risk variants, play a role in neuroblastoma tumorigenesis. Importantly, despite considerable efforts, the molecular events that drive neuroblastoma initiation and progression are still largely unknown. Therefore, a better knowledge of the alterations that shape neuroblastoma genomes and transcriptomes becomes necessary to understand its etiology. Detailed genomic information leading to new drug targets is also the starting point to develop more effective and less toxic treatments (19).

In this work, we propose an alternative method to verify the potential pathogenic consequence of noncoding variants in neuroblastoma using whole-genome sequencing (WGS) data from 151 tumors. First, we tested whether the predicted functional impact of somatic variants in TFBS is a feature of a specific cancer tissue. Second, we tested whether there is an enrichment of somatic mutations in DNA-binding sites for TFs involved in tumorigenesis. Third, we verified whether recurrent somatic mutations in TFBS are associated with changes in mRNA levels.

Overall, our analysis strategy allowed us to find mutated binding sites for TFs that may have key functions in neuroblastoma initiation. Moreover, for some of these mutated sites, we identified candidate target genes whose deregulation may contribute to neuroblastoma development.

Materials and Methods

Samples collection

Neuroblastoma tumor DNA (primary tumors) and matched germline DNA (from peripheral blood) were obtained from the IRCCS Istituto Giannina Gaslini (Genova, Italy), Ospedale Pediatrico Bambino Gesù (Rome, Italy), and Hospital Sant Joan de Déu, (Barcelona, Spain). Primary tumor samples were verified to have >75% viable tumor cell content by histopathology assessment. This study was approved by the Ethics Committee of the Ospedale Bambino Gesù of Rome (protocol no. 20757 of the April 9, 2019). Informed written consent was obtained from the subjects.

DNA extraction from peripheral blood and primary tumor tissues

DNA from peripheral blood (PB) was extracted with QIAamp DNA Mini Kit (Qiagen) according to manufacturer's instructions. DNA from primary tumor tissues was extracted with MasterPure DNA Purification Kit (Epicentre) according to manufacturer's protocol.

DNA quantification and library preparation for sequencing

DNA quality was monitored on 1% agarose gels. Its purity was checked using the NanoPhotometer spectrophotometer (IMPLEN). DNA concentration was measured using Qubit DNA Assay Kit in Qubit 2.0 Flurometer (Life Technologies). A total of 1.0 μg of DNA per sample was used as input material for library preparation. Sequencing

libraries were generated using Truseq Nano DNA HT Sample Preparation Kit (Illumina) following manufacturer's recommendations. Genomic DNA was sonicated to a size of 350 bp, and then fragments were end-polished, A-tailed, and ligated with the full-length adapter for Illumina sequencing with further PCR amplification. At last, PCR products were purified (AMPure XP system) and libraries were analyzed for size distribution using the DNA Nano 6000 Assay Kit of Agilent Bioanalyzer 2100 system (Agilent Technologies) and quantified using real-time PCR.

Somatic mutations detection
In-house WGS data

WGS of 14 normal-primary neuroblastoma sample pairs was performed on an Illumina HiSeq1500 platform. The paired-end sequencing produced 150-bp long reads. Alignment files were obtained by mapping reads versus GRCh37/hg19 reference genome assembly. Somatic SNVs and INDELs were detected with MuTect (20) and Strelka (21), respectively.

Publicly available WGS data (Target)

We obtained access to WGS and RNA-seq data of neuroblastoma from the Target project (accession no.: phs000218.v21.p7; project ID: #14831; ref. 12) and included, in our analysis, 137 primary neuroblastoma for which somatic variants were available. The functional annotation of somatic variant calls was performed with ANNOVAR (22) and FunSeq2 (23).

RNA-seq data were available for 161 samples in the form of processed fragments per kilobase of exon model per million reads mapped (FPKM). Eighty-nine neuroblastoma samples had both WGS and gene expression data.
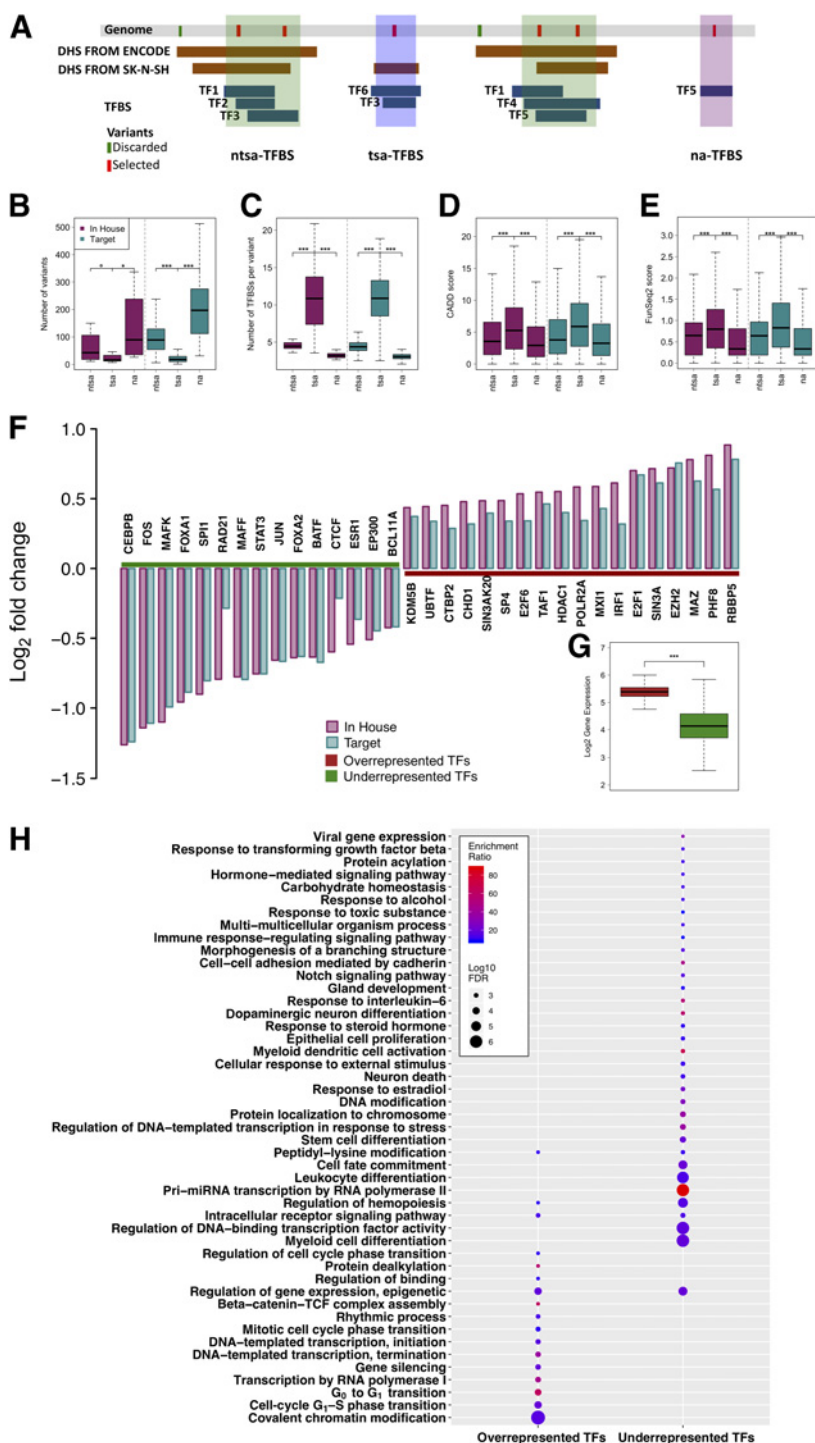
Somatic variants selection criteria

Somatic variants of the in-house and Target data sets were processed as follows. From raw variant calls, we first eliminated those that did not pass MuTect or Strelka quality. As reported in ref. 24, we required that the read count of the variant site in the tumor had to be significantly higher than the normal ($P \leq 0.05$ by two-sided Fisher exact test). Then, we discarded somatic variants falling in genomic duplicated regions. Finally, we filtered out common polymorphisms (minor allele frequency >1%) by using allele frequencies of 1,000 Genomes Project, ExAC and gnomAD databases.

We generated three lists of somatic variants based on DNA regulatory activity. Using ENCODE v3 data, we defined active regulatory elements as genomic regions characterized by the presence of both DNase hypersensitive sites (DHS) and TFBS (**Fig. 1A**).

We built the first somatic variant set in "nontissue-specific and active TFBS" (ntsa-TFBS) by selecting those variants falling in DHS of 125 cell lines and TFBSs of 161 TFs tested in 91 cell lines.

Then, we obtained the second set of somatic variants in "tissue-specific and active TFBS" (tsa-TFBS) by keeping variants falling in DHS of the SK-N-SH neuroblastoma cell line and TFBSs. Here, we excluded the DHS data from the SK-N-BE(2)C neuroblastoma cell line (included in ENCODE v3), and considered only SK-N-SH DHS data for three main reasons. First, SK-N-BE(2)C is characterized by *MYCN* amplification (whereas SK-N-SH does not) that can have effects on chromatin structure (25). Therefore, we discarded SK-N-BE(2)C to reduce the effects of *MYCN* as possible confounder. Second, the DHS data from SK-N-SH cell line were obtained after treatment with all trans-retinoic acid (ATRA) for 48 hours (ENCODE protocol). As reported previously, short-term ATRA treatment promotes growth control and initial induction of differentiation towards neuronal

Figure 1.
Description of somatic variants in TFBS and results from mutational enrichment analysis. **A,** Schematic explanation of the three variant selection strategies. ntsa-TFBS (nontissue-specific and active TFBS) included variants falling in DHS and TFBS ENCODE v3. tsa-TFBS (tissue-specific and active TFBS) comprised variants falling in DHS of the SK-N-SH neuroblastoma cell line and TFBSs of the ENCODE v3. na-TFBS (nonactive TFBS) enclosed variants falling in candidate inactive regulatory elements characterized by TFBSs but not by any DHS of the ENCODE v3. **B,** Box plot showing the median values of noncoding regulatory variants per sample. **C,** Box plot showing the median values of TFBSs per variant. **D** and **E,** Box plots showing the median values of CADD and FunSeq2 pathogenicity scores of somatic variants. **F,** Bar plot showing the $\log_2$-fold change of TFs in tsa-TFBS. **G,** Box plot showing the median expression values of the over-represented TFs (red) and under-represented TFs (green) in 498 tumors (GSE62564). **H,** Dot plot showing the GO biological processes enriched in the lists of over-represented TFs (left) and under-represented TFs (right). Dot color scale is dependent on the enrichment ratio, whereas dot sizes are proportional to the $\log_{10}$ FDR. Statistical significance was calculated with Mann–Whitney test in **B**, **C**, **D**, **E**, and **G**. Fisher exact test and Benjamini–Hotchberg (FDR) correction was used in **F**. Hypergeometric test and Benjamini–Hotchberg (FDR) correction was used in **H**. *, $P < 0.01$; ***, $P < 0.0001$.

phenotype (26). Therefore, the SK-N-SH epigenome, in terms of chromatin accessibility, should resemble that of neurons at very early stages of neuroblastoma development. Third, SK-N-SH shared a lower number of DHSs with other cell lines than SK-N-BE(2)C. The tissue specificity of the SK-N-SH DHSs was assessed as follows. We calculated the number of DHSs regions for each cell line of the ENCODE v3 catalog. Then, we extracted SK-N-SH DHSs and determined the number of these regions that overlapped with the other cell lines.

Finally, for each cell line, we obtained the percentage of DHS regions shared with SK-N-SH by dividing the number of common regions by the number of initial DHSs. The same procedure was applied to assess the tissue specificity of the SK-N-BE(2)C DHSs. As reported in Supplementary Fig. S1A, on average, the other 124 cell lines shared only the 23.64% of DHSs with SK-N-SH, whereas the 37.48% with SK-N-BE(2)C. We also used the Jaccard statistics (27) of the Bedtools software (28) to measure the similarity of DHSs profiles between

SK-N-SH, SK-N-BE(2)C, and each of the other cell lines of the ENCODE catalog. The Jaccard value ranges from 0 (no similarity) to 1 (completely identical). Our results indicated that the SK-N-SH and SK-N-BE(2)C DHSs profiles are quite distant (their Jaccard score was about 0.43). Furthermore, the median Jaccard score in the pairwise comparisons with the other cell lines was about 0.27 and 0.30 for the SK-N-SH and the SK-N-BE(2)C, respectively (Supplementary Fig. S1B). This last result confirmed that SK-N-SH DHSs were more tissue specific than SK-N-BE(2)C DHSs ($P = 7.54 \times 10^{-11}$; Mann–Whitney test).

Finally, we obtained the third set of somatic variants in "nonactive TFBS" (na-TFBS) by including those variants falling in inactive regulatory elements characterized by only TFBSs.

From ntsa-TFBS variants, we subtracted those exclusive of SK-N-SH regulatory elements. From all three sets, were further excluded somatic variants in exons, UTRs, and splice junctions of protein-coding and ncRNA genes.

Mutational enrichment

We compared the ratio of the number of mutations in active binding sites of single TF to the total number mutations in all active binding sites of all TFs and that of the total of active binding sites of the same TF to all active binding sites of all TFs using a Fisher exact test. We performed this test for all TFs and corrected the resulting P values for multiple testing (FDR) using the Benjamini–Hochberg procedure. The threshold for the $\log_2$ of fold enrichment was set at ± 0.2. The TFs that showed a positive or negative fold enrichment were indicated as either over-represented (or)TF or under-represented (ur)TF, respectively.

We performed this enrichment analysis separately for the three variant selections (tsa-TFBS, ntsa-TFBS, and na-TFBS) in the two datasets under study (in-house and Target). Then, for each variant selection, we merged the two lists (in-house and Target) to get common significantly orTF and urTFs. Finally, we used these lists of TFs to perform a Gene Ontology (GO) and search for enriched Biological Processes by using the WebGestalt tool (29).

Gene expression analysis, samples clustering, and survival analysis

R2: Genomics Analysis and Visualization Platform (http://r2.amc.nl) was used to query transcriptomic data of 498 neuroblastoma samples (GSE62564). We used the k-means clustering algorithm to divide samples in two groups based on significantly enriched orTFs and urTFs expression levels. Then, multivariate Cox proportional regression analysis was performed to assess the independence of K-means grouping from the other risk factors as reported in ref. 30. HRs and 95% confidence interval (CI) for survival were calculated. The overall survival (OS) and the event-free survival (EFS) probabilities were calculated by using the Kaplan–Meier method. The log-rank test statistical significance was set at 5%.

Variant clusterization, gene expression correlation, and chromatin interactions

We used the Target samples with somatic variants and gene expression data ($n = 89$) to search for clusters of somatic variants within the enriched tsa-TFBSs. With variants clustered within 300bp to each other, we identified the nearest protein-coding genes excluding pseudogenes, hypervariable genes (Immunoglobulins and T-cell receptors), and ncRNA genes. We also searched for candidate targets by examining public HiC sequencing data obtained on the neuroblastoma cell line SK-N-DZ and hosted at the 3D-genome Interaction Viewer (31). In brief, we used the genomic coordinates of mutated tsa-TFBSs with clustered variants to obtain candidate interacting genes

showing the bias-removed interaction frequency greater than 2-fold of background signal. Subsequently, we compared the expression levels of these genes between mutated and non-mutated samples. t test was used to determine the statistical differences between the groups. The statistical significance threshold was set at $P < 0.05$ after FDR correction.

All the analyses were performed within the R environment for statistical computing (www.r-project.org).

Availability of data and materials

The WGS datasets generated and analyzed during this study are available from the corresponding author on reasonable request. Public data and data repositories are referenced within the manuscript.

Results

WGS data

To detect somatically mutated, active TFBSs and study the functions of their respective TFs, we analyzed 14 primary neuroblastoma whole genomes (Supplementary Table S1) and used public data from the Target neuroblastoma project encompassing 137 whole genomes. The sequencing and the bioinformatics analysis pipeline produced high quality and reliable variant calls (Supplementary Fig. S2A–S2C). We applied stringent filtering criteria to discard false positives, common polymorphisms, genomic duplications, and coding regions from somatic variant calls of both datasets. After filtering, we obtained 19,750 (median = 1,410.7 per sample) and 282,312 (median = 2,060.7 per sample) somatic variants for the in-house and target datasets, respectively (Supplementary Fig. S3A). In agreement with previous reports, on average, we selected 0.47 and 0.68 somatic variants per megabase for the in-house and target datasets, respectively (Supplementary Fig. S3B; ref. 24).

Noncoding somatic variants show high pathogenic scores in neuroblastoma-active TFBS

We selected three different types of variants (ntsa-TFBS, tsa-TFBS, and na-TFBS) based on their position in regulatory DNA elements (see Materials and Methods and **Fig. 1A**). The median number of variants in tsa-TFBS category was significantly lower as compared with the other two categories in both datasets ($P < 0.05$; Mann–Whitney test; **Fig. 1B**). As depicted in **Fig. 1C**, somatic variants in tsa-TFBS altered, on average, a significantly higher number of binding sites compared with ntsa-TFBS and na-TFBS ($P < 0.0001$; Mann–Whitney test). Moreover, the median of pathogenic scores, estimated by CADD (32) and FunSeq2 (23) (**Fig. 1D** and **E**), was significantly higher in tsa-TFBS than the others ($P < 0.0001$; Mann–Whitney test).

Enrichment analysis identifies orTFs and urTFs in neuroblastoma-active TFBS

To determine whether particular TFs harbored more somatic variants than the expected in their binding sites, we performed an enrichment analysis in which the ENCODE catalogue of TFBS was our reference (see Materials and Methods; Supplementary Table S2). When we analyzed somatic mutations belonging to tsa-TFBS category, we found (**Fig. 1F**) 18 orTFs and 15 urTFs whereas the analysis based on somatic mutations belonging to ntsa-TFBS and na-TFBS categories returned a lower number of significantly enriched TFs (Supplementary Figs. S4A and S4B). Furthermore, we used the RNA-seq data from 498 neuroblastoma samples (GSE62564) to assess the expression levels orTFs and urTFs. We found that orTFs expression was significantly higher than urTFs as reported in **Fig. 1G** ($P = 2.278 \times 10^{-136}$; Mann–Whitney test). Together, these data indicate that somatic variants of tsa-TFBS fell in TF-binding sites specifically active in neuroblastoma.

orTFs in mutated TFBS active in neuroblastoma are enriched in cell cycle and gene regulation processes

To assess biological processes related to orTFs and urTFs, we performed a GO enrichment analysis. The results showed two distinct groups of biological processes characterizing orTFs and urTFs, respectively. Indeed, we found that orTFs were highly significantly enriched in biological processes belonging to cell-cycle phase transitions, covalent chromatin modification, gene silencing, and others (**Fig. 1H**). In contrast, urTFs were enriched in processes involved in cell differentiation, cell fate commitment, stem cell differentiation, among the others (**Fig. 1H**). Of note, we did not find any significant GO terms for orTFs and urTFs of the other variant categories ntsa-TFBS and na-TFBS.

orTFs- and urTFs-based gene signatures show opposite ability to predict unfavorable outcome

We also asked whether the expression of orTFs and urTFs could correlate with neuroblastoma patient's survival. We used a well-annotated RNA-seq dataset (GSE62564, $n = 498$). We performed k-means clustering of samples based on expression levels of orTFs and urTFs, separately. For orTFs, the clustering algorithm distinguished two groups of samples characterized by low (group 0) and high (group 1) expression levels as reported in **Fig. 2A** (see also Supplementary Fig. S5A; $P = 5.495 \times 10^{-35}$; Mann–Whitney test). The group 1 strongly correlated with markers of neuroblastoma aggressiveness (**Fig. 2B**) such as the *MYCN* amplification ($P = 2.4 \times 10^{-29}$; χ^2 test), the INSS stage 4 ($P = 1.27 \times 10^{-29}$; χ^2 test), and the high-risk tumors ($P = 1.5 \times 10^{-51}$; χ^2 test). Furthermore, group 1 patients showed inferior OS and EFS as compared with the group 0 (**Fig. 2C**; $P < 2.0 \times 10^{-16}$; log-rank test). We confirmed the prognostic significance of k-means grouping based on orTFs by multivariate Cox proportional regression analysis (Supplementary Table S3). Indeed, the orTF-based gene signature resulted to be a prognostic factor independently from other known clinical markers (*MYCN* amplification, INSS stage, and age at diagnosis). In contrast, the k-means clustering of samples based on the urTFs gene expression returned opposite results. Two markedly separated groups characterized by low (group 0) and high (group 1) expression levels were identified (**Fig. 2D**; Supplementary Fig. S5B; $P = 4.471 \times 10^{-38}$; Mann–Whitney test). Here, the group 1, characterized by high expression levels of urTFs, correlated with favorable markers as depicted in **Fig. 2E** (*MYCN*-amplified vs. *MYCN*-nonamplified, $P = 3.4 \times 10^{-04}$; stage 4 vs. non-stage 4 tumors, $P = 0.035$; high-risk vs. non-high-risk tumors, $P = 6.5 \times 10^{-03}$; χ^2 test) and better survival probabilities (**Fig. 2F**).

Recurrent mutations within orTF-binding sites can affect the expression of target genes

Consequently, we wanted to detect possible mutational hotspots in orTF binding sites and evaluate whether these recurrent regulatory variants could affect the expression of Target genes. To address this issue, we selected from the Target neuroblastoma cohort, those samples with both somatic mutations and gene expression data available ($n = 89$). We found three variants (in two samples) nearby the telomeric end of chromosome 13q. The variants clustered within a binding site of EZH2 located 18,076 bp downstream the gene ADP-Ribosylhydrolase Like 1 (*ADPRHL1*; **Fig. 3A**). HiC data of the SK-N-DZ neuroblastoma cell line confirmed the interactions between the mutated regulatory region and *ADPRHL1* and indicated significant interactions also with MCF.2 Cell Line Derived Transforming Sequence Like (*MCF2L*) located 435,000 bp apart (see Supplementary Table S4; Supplementary Figs. S6A–S6D). Of note, the presence of the variants significantly

reduced the expression of both genes (**Fig. 3B** and **C**) when we compared mutated and wild-type samples ($P = 3.35 \times 10^{-03}$ for *ADPRHL1*, $P = 3.97 \times 10^{-18}$ for *MCF2L*; t test). Consistent with the repressive role of EZH2, we observed an inverse, although weak, correlation between the gene expression of *EZH2* and its targets in the gene expression dataset (GSE62564; **Fig. 3D** and **E**). Nevertheless, the low expression of *ADPRHL1* and *MCF2L* correlated with poor OS and EFS ($P < 1.00 \times 10^{-07}$; log-rank test, **Fig. 3F** and **G**). Furthermore, the expression of both genes inversely correlated with the unfavorable neuroblastoma prognostic markers such as stage 4, high-risk, and *MYCN* amplification ($P < 1.00 \times 10^{-03}$; Mann–Whitney test, **Fig. 3H** and **I**).

We found a second cluster of somatic variants (three variants in three samples) nearby the transcription starting site of *CTTNBP2* gene (cortactin binding protein 2) on chromosome 7q. In particular, two variants were immediately upstream the transcription starting site, whereas the other one was within the first intron. As shown in **Fig. 4A**, these variants fell in the binding sites of six orTFs (RBBP5, EZH2, SIN3A, POLR2A, TAF1, E2F1). Here, HiC data of the SK-N-DZ neuroblastoma cell line indicated long-range interactions between the regulatory region and *ASZ1* and *WNT2* (Supplementary Table S4) that did not show significant changes in gene expression between mutated and wild-type samples. A motif analysis with the R-Bioconductor package "motifbreakeR" showed that two of these variants weakly altered the binding motif of SIN3A and TAF1 (**Fig. 4B**). Also, in this case, the *CTTNBP2* gene expression in the mutated samples was significantly reduced compared with wild-type (**Fig. 4C**; $P = 8.86 \times 10^{-07}$; t-test). Pairwise Pearson correlations between CTTNBP2 and each of the six TFs (summarized in **Fig. 4D**) indicated that *CTTNBP2* expression was directly dependent on *POLR2A*, and inversely correlated with *E2F1* and *RBBP5* expression levels. Low mRNA levels of *CTTNBP2* correlated with a poor survival (**Fig. 4E**) and unfavorable clinical parameters such as stage 4, high-risk, and *MYCN* amplification (**Fig. 4F**; $P < 0.05$; Mann–Whitney test).

The third cluster of somatic variants (two variants in two samples, Supplementary Table S4) was within the first intron of lymphocyte-specific protein 1 (*LSP1*, F-actin binding and cytoskeleton associated protein) on chromosome 11p. As shown in **Fig. 5A**, these variants could alter the binding site of POLR2A. We assessed that the gene expression in the mutated samples was significantly reduced compared with wild-type (**Fig. 5B**; $P = 2.48 \times 10^{-03}$; t-test). Pearson correlation (**Fig. 5C**) indicated that *LSP1* expression was directly dependent on *POLR2A*. Survival analysis based on *LSP1* expression showed that both OS and EFS were significantly influenced when the gene expression was lower than the median (**Fig. 5D**). Furthermore, the expression of *LSP1* inversely correlated with the markers of neuroblastoma aggressiveness such as stage 4, high-risk, and *MYCN* amplification (**Fig. 5E**; $P < 0.01$; Mann–Whitney test). Interestingly, we found an inverse correlation between *ALK* and *LSP1* expression levels ($r = -0.188$, $P = 2.35 \times 10^{-05}$) in the cohort of 498 neuroblastoma samples as previously observed in white blood cells and thymocytes (33).

This third cluster of variants also showed significant long-range interactions (Supplementary Table S4) with the tyrosine hydroxylase (*TH*) gene. In this case, its expression in the mutated samples was significantly increased compared with wild-type ($P = 9.14 \times 10^{-06}$; T test; Supplementary Fig. S7A). Pearson correlation (Supplementary Fig. S7B) indicated that *TH* expression was directly dependent on *POLR2A*. Tumors with high *TH* expression exhibited a poor survival (Supplementary Fig. S7C). Furthermore, high *TH* expression inversely

Figure 2.
Results of the *k*-means clustering of neuroblastoma samples ($n = 498$, GSE62564). The clustering was based on the expression levels of over-represented (**A**–**C**) and under-represented TFs (**D**–**F**). **A,** Box plot showing the median expression levels of group 0 ($n = 269$) and group 1 ($n = 229$). **B,** Clinical features of neuroblastoma samples in group 0 and group 1 based on orTFs. **C,** EFS (top) and OS (bottom) probabilities of neuroblastoma samples in group 0 and group 1. **D,** Box plot showing the median expression levels of group 0 ($n = 378$) and group 1 ($n = 120$). **E,** Clinical features of neuroblastoma samples in group 0 and group 1 based on urTFs. **F,** EFS (top) and OS (bottom) probabilities of neuroblastoma samples in group 0 and group 1. **A** and **D**, Mann–Whitney test; **B** and **E**, chi-square test; **C** and **F**, log-rank test. ***, $P < 0.0001$.
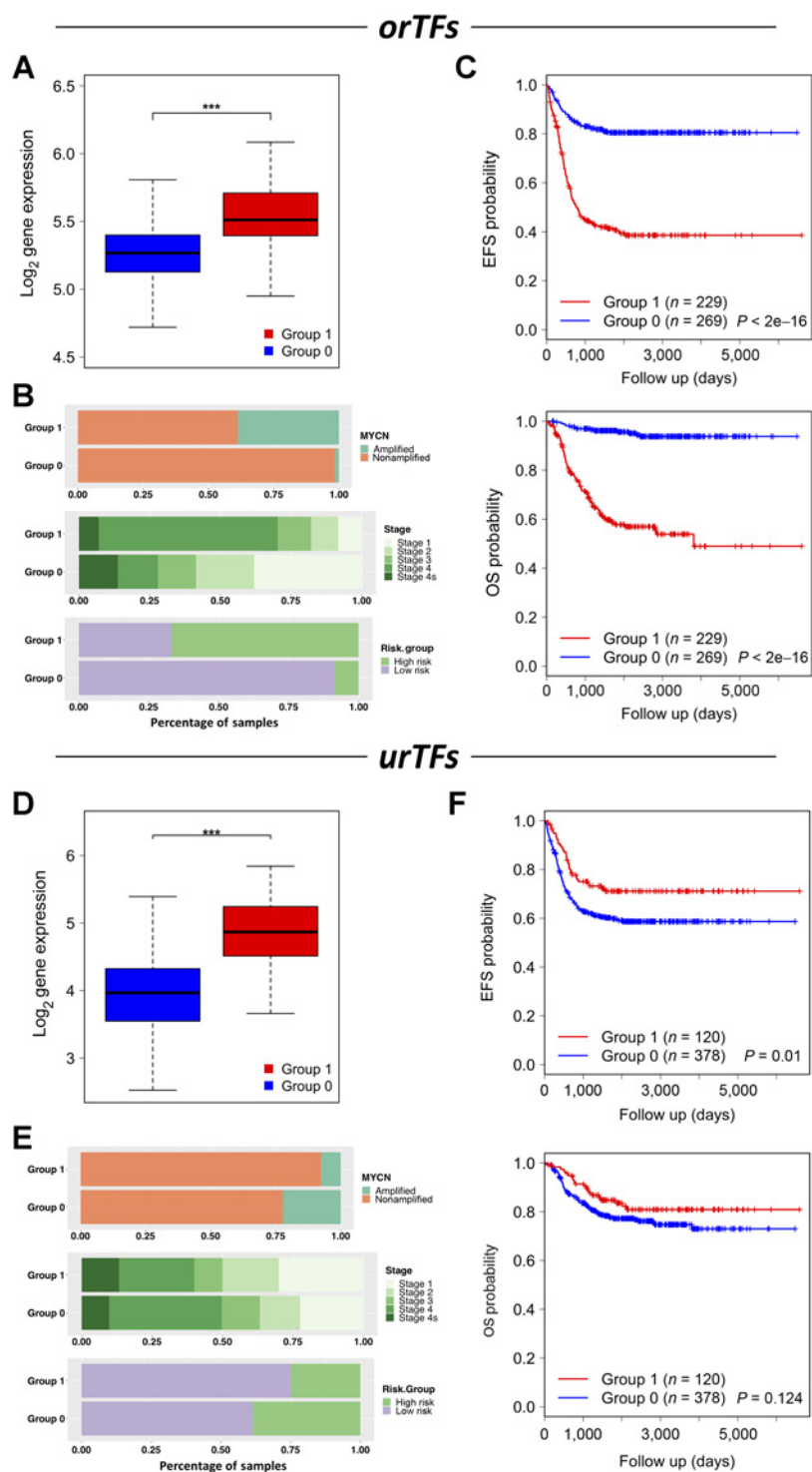
correlated with the markers of neuroblastoma aggressiveness (Supplementary Fig. S7D; $P < 0.05$; Mann–Whitney test).

Discussion

Emerging evidence suggests that noncoding somatic mutations can promote cancer development by affecting the functions of DNA regulatory elements and, as a consequence, the normal transcriptional program (1–3). Despite these recent advances, it remains a challenge to establish the pathogenic role of noncoding variants in cancer development. In this work, we have demonstrated that mutated tissue-specific and active regulatory elements are enriched in binding sites of TFs that might have a role in the development and progression of neuroblastoma.
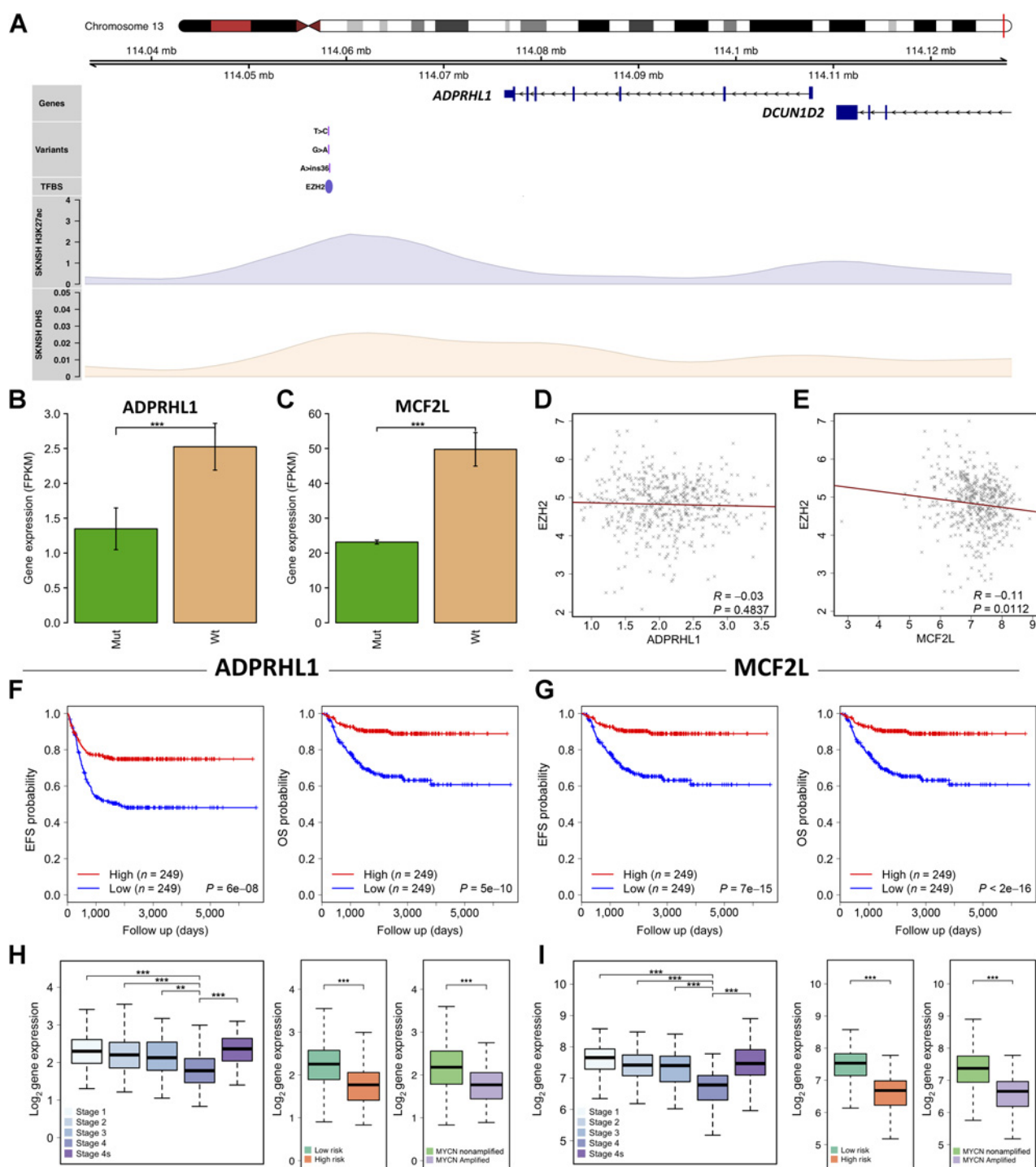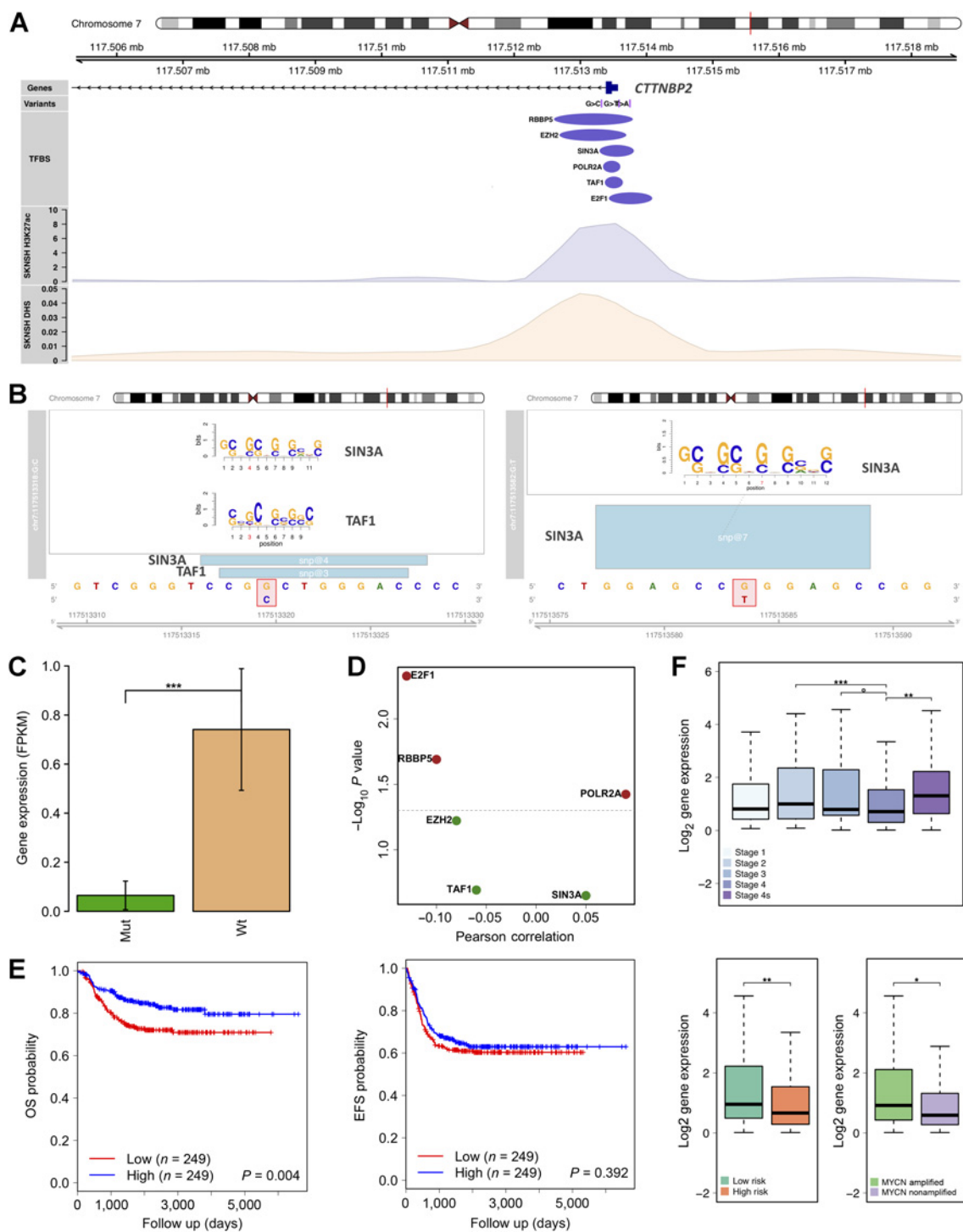
Capasso et al.

Figure 3.

ADPRHL1 and *MCFL1* deregulation by somatic regulatory variants. **A,** From top to bottom, the plot tracks show the *ADPRHL1* gene, the clustered somatic variants, the binding site of EZH2, the track for H3K27ac histone modifications in SK-N-SH neuroblastoma cell line (GSE90683), and the ENCODE track of DHS signal in the SK-N-SH neuroblastoma cell line. As reported in Supplementary Table S4, although the nearby gene *DCUN1D2* showed significant interactions with the mutated locus, it did not show significant changes in gene expression. **B,** *ADPRHL1* expression in the mutated ($n = 2$) and wild-type samples ($n = 96$) of Target project. **C,** *MCF2L* expression in the mutated ($n = 2$) and wild-type samples ($n = 87$) of Target project. **D,** *ADPRHL1* versus *EZH2* expression correlation. **E,** *MCF2L* versus *EZH2* expression correlation. **F,** EFS and OS based on *ADPRHL1* median expression in the GSE62564 data set. **G,** EFS and OS based on *MCF2L* median expression in the GSE62564 data set. **H,** *ADPRHL1* expression by clinical features. From left to right, INSS stage, risk group, and MYCN amplification. **I,** *MCF2L* expression by clinical features. From left to right, INSS stage, risk group, and *MYCN* amplification. Two genes correlation coefficients in **D** and **E** were calculated with Pearson method. Statistical significance was calculated with Mann–Whitney test in **B**, **C**, **D**, **E**, **H**, and **I**. Log-rank test was used in **F** and **G**. **, $P < 0.001$; ***, $P < 0.0001$.

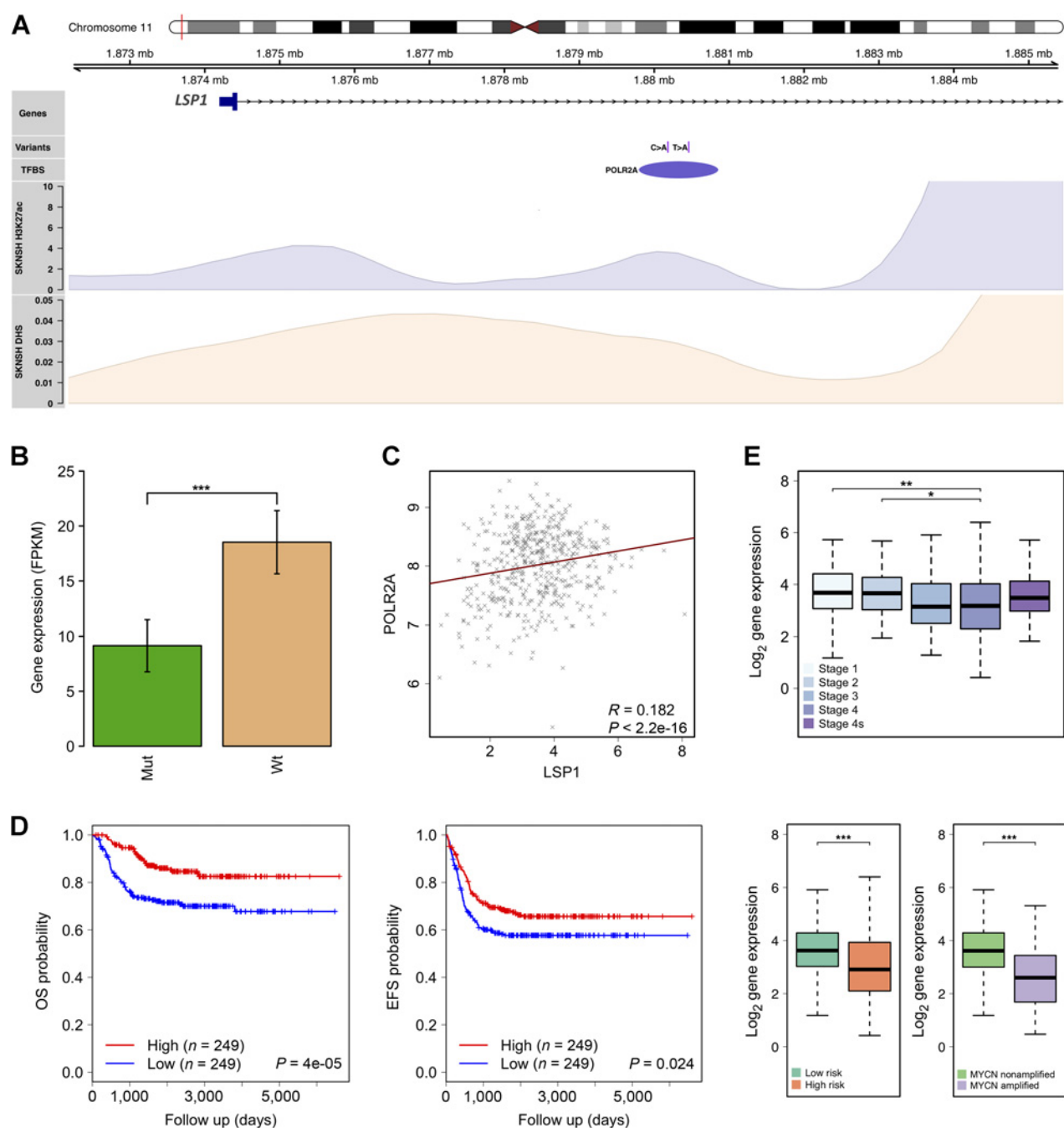**Figure 4.**
CTTNBP2 deregulation by somatic regulatory variants. **A,** From top to bottom, the plot tracks show the *CTTNBP2* gene, the clustered somatic variants, the binding sites of six orTFs, the track for H3K27ac histone modifications in SK-N-SH neuroblastoma cell line (GSE90683), and the ENCODE track of DHS signal in the SK-N-SH neuroblastoma cell line. **B,** SIN3A and TAF1 binding motifs alteration. **C,** *CTTNBP2* expression in the mutated ($n = 3$) and wild-type samples ($n = 86$) of Target project. **D,** Volcano plot summarizing the gene expression Pearson correlations between *CTTNBP2* and each of the six TFs reported in **A**. The gray dashed line marks the cutoff for statistical significance. **E,** EFS and OS results according to *CTTNBP2* median expression. **F,** *CTTNBP2* expression by clinical features. From top to bottom-right, INSS stage, risk group, and *MYCN* amplification. Statistical significance was calculated with Mann–Whitney test in **C**, **D**, and **F**. Log-rank test was used in **E**. °, $P < 0.05$; *, $P < 0.01$; **, $P < 0.001$; ***, $P < 0.0001$.

Capasso et al.

Figure 5.

LSP1 deregulation by somatic regulatory variants. **A,** From top to bottom, the plot tracks show the *LSP1* gene, the clustered somatic variants, the binding site of EZH2, the track for H3K27ac histone modifications in SK-N-SH neuroblastoma cell line (GSE90683), and the ENCODE track of DHS signal in the SK-N-SH neuroblastoma cell line. **B,** *LSP1* expression in the mutated ($n = 2$) and wild-type samples ($n = 87$) of Target project. **C,** Pearson correlation of *LSP1* versus *POLR2A* gene expressions. **D,** OS and EFS results according to *LSP1* median expression. **E,** *LSP1* expression by clinical features. From top to bottom right, INSS stage, risk group, and *MYCN* amplification. Statistical significance was calculated with Mann–Whitney test in **B, C,** and **E**. Log-rank test was used in **D**. *, $P < 0.01$; **, $P < 0.001$; ***, $P < 0.0001$.

By using available chromatin characteristics and epigenomic data, we set up three different approaches to determine active TFBSs in neuroblastoma tumors and select the variants that fell within. Our results showed that, despite their limited number, somatic variants occurring in TFBSs active in neuroblastoma were characterized by higher pathogenicity than those occurring in TFBSs localized outside or within open chromatin regions defined by ENCODE project in 125 cell lines. Furthermore, the median number of binding sites altered by each variant was higher in tsa-TFBS than the others. In addition, the enrichment analysis of

mutated binding sites in ntsa-TFBSs or in na-TFBSs returned small sets of TFs that were not significantly enriched in any GO terms. These results provide evidence that each tumor may have a specific profile of pathogenic mutations, which may change according to the tissue specificity of the DNA regulatory elements.

In this first part of our work, we observed that certain TFs (orTF) of active regulatory regions (tsa-TFBS) had their binding sites mutated more often than the expected. GO analysis suggested their involvement in promoting cell-cycle phase transitions, covalent chromatin modification, and gene silencing. Some of them are known to play key roles in tumor development. For example, retinoblastoma-binding protein 5 (RBBP5), part of the MLL1/MLL complex, is involved in H3 "Lys-4" methylation at key developmental loci. In embryonic stem cells, it plays a crucial role in the differentiation potential, particularly along the neural lineage. RBBP5 takes part in large complexes involved in nearly all steps of tumor development including transcriptional activation of key oncogenes, cell proliferation, survival, initiation and progression, invasion, and metastasis (34). Several small molecules and drugs have been tested to target MLL1 complexes in cancer (35, 36). Furthermore, *PHF8* is a histone lysine demethylase acting as a transcription activator. It is involved in cell-cycle progression where it controls G_1–S transition (37). It promotes oncogenesis and tumor progression in a number of cancer types including prostate cancer (38), breast cancer (39), and others. MYC-Associated Zinc Finger Protein 9 (*MAZ*) has dual roles in transcription initiation and termination (40). It transactivates numerous oncogenes (*MYC, HRAS, PPARG, TSG101, VEGF, CAV1,* and *PTHR1*) and transrepresses others (*NOS3, MYB,* and *TERT*). Clinical evidence has revealed that *MAZ* is overexpressed in prostate (41) and breast cancer (42). *In vitro* experiments have shown that *MAZ* knockdown suppresses prostate and breast cancer cell proliferation, implying that *MAZ* functions as a promoter in cancer development and cell proliferation. EZH2 is a member of the polycomb repressive complex 2 involved in maintaining the transcriptional repressive programs over cell generations. Recent studies have shown that the suppression of *EZH2* inhibited neuroblastoma growth *in vitro* and *in vivo* and repressed neuronal differentiation (43, 44). Several EZH2 inhibitors have been developed for cancer therapy (45). The TF E2F1, a key cell-cycle regulator, targets genes encoding proteins that regulate cell-cycle progression through the G_1–S transition as well as proteins important in DNA repair and apoptosis (46–48). Additional experimental studies are needed to determine the role of these TFs in neuroblastoma development and progression.

In contrast, urTFs (TFs with binding sites mutated more rarely than the expected) were involved in cell differentiation and cell fate commitment. Among these factors we found key regulators of cell proliferation and differentiation. FOS and JUN form the AP-1 complex, which is implicated in regulation of cell proliferation, differentiation, and transformation (49). In neuroblastoma, the AP-1 complex is involved in conferring NCC-like identity to tumor cells (50). *FOXA1* and *FOXA2* are transcription factors involved in embryonic development, establishment of tissue-specific gene expression, and regulation of gene expression in differentiated tissues. They are thought to act as pioneer factors opening the compacted chromatin for other proteins (51).

Having established the importance of enriched TFs, we asked whether their expression levels (over- and under-represented TFs, separately) could influence neuroblastoma patient's prognosis. We used RNA-seq and clinical annotation data of 498 neuroblastoma tumors to answer this question. First, we found that the expression value of orTFs was significantly higher than that of urTFs. This result pointed to key biological implications for orTFs in neuroblastoma

development. Then, we clustered the 498 neuroblastomas based on orTFs and urTFs expression, separately. In each case, the *K*-means clustering algorithm identified two distinct groups. For orTFs, the group 1 included samples characterized by high gene expression and low survival probabilities. In contrast, group 0 included samples with low expression of these TFs and with higher survival rates. Moreover, group 1 significantly associated with stage 4, *MYCN* amplified and high-risk clinical features. Conversely, the clustering of samples based on urTFs expression, returned opposite, but less significant, results. Indeed, in this case, patients included in group 0 were characterized by worse prognosis and unfavorable clinical markers. We interpret these results by assuming that TFs enriched in active mutated binding sites might form a signature reflecting the activity of key pathways in neuroblastoma. In addition, further clinical studies could assess the prognostic value of this signature for more refined patient stratification in risk groups.

In a subsequent step, we searched for possible mutational hot spots that could influence the expression of candidate target genes. Overall, we identified three clusters of variants with significant changes of target gene expression between mutated and wild-type samples. Here, it should be pointed out that the small set of initial variants in tsa-TFBSs limited our search for clusters that contained no more than three variants in three different samples. However, we found that the expression of four candidate target genes (*ADPRHL1, MCF2L, CTTNBP2,* and *LSP1)* was significantly reduced in mutated samples compared with wild-type. We observed that the reduced expression of *ADPRHL1, MCF2L,* and *LSP1* negatively influenced both OS and EFS of patients with neuroblastoma and significantly correlated with clinical and biological markers of aggressive disease.

ADPRHL1 is a protein ADP-ribosylhydrolase that reverses the action of ADP-ribosyltransferases. So far, this gene has been little studied: it has been indicated as potential prognostic marker of uveal melanoma (52) and seems to play a relevant role during cardiogenesis (53).

MCF2L encodes a guanine nucleotide exchange factor that interacts specifically with the GTP-bound Rac1 and plays a role in the Rho/Rac signaling pathways (54). Beyond its role in regulating breast cancer cell migration, MCFL2 is suggested to participate in axonal transport in neurons and is highly expressed in brain (55).

LSP1, an F-actin binding protein regulating transendothelial migration of white blood cells, is involved in ERK/MAPK pathways (56). Downregulation of *LSP1* has been reported to be a risk factor for liver tumor and to promote its development and metastasis (57, 58). Of interest, in white blood cells and thymocytes, *LSP1* expression was inversely correlated with anaplastic lymphoma kinase (*ALK*; ref. 33), which is the most common, somatically mutated gene in sporadic neuroblastoma and is the major susceptibility gene to familial neuroblastoma (59). Accordingly, we observed this inverse correlation between *LSP1* and *ALK* also in 498 neuroblastoma samples.

TH encodes an enzyme with important roles in the physiology of adrenergic neurons. High levels of *TH* mRNA in peripheral blood and bone marrow of patients with neuroblastoma at diagnosis can predict poor prognosis (60). Our survival analysis using *TH* expression in neuroblastoma tumors showed opposite results.

Further functional investigations are needed to elucidate the functional effect of this cluster of noncoding variants.

So far, no study has investigated the role of the above-reported genes in neuroblastoma. Thus, for the importance of their interactors and the pathways they are involved in, they are noteworthy and could be

considered for further studies to better understand and deconvolute molecular events of neuroblastoma initiation.

Conclusions

To conclude, our study proposes a novel approach to verify the potential pathogenic consequence of regulatory variants in neuroblastoma that could be extended to other cancers. Overall, we demonstrated the existence of a direct link between TFBS within active regulatory elements and somatic mutations. Moreover, we provide further evidence that alteration of gene expression regulatory circuits can have relevant implications in neuroblastoma development.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: M. Capasso, A. Iolascon
Development of methodology: M. Capasso, V.A. Lasorsa
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): V.A. Lasorsa, F. Cimmino, M. Avitabile, S. Cantalupo, A. Montella, B. De Angelis, M. Morini, C. de Torres, A. Castellano, F. Locatelli
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): V.A. Lasorsa

References

1. Bradner JE, Hnisz D, Young RA. Transcriptional addiction in cancer. Cell 2017; 168:629–43.
2. Melton C, Reuter JA, Spacek DV, Snyder M. Recurrent somatic mutations in regulatory regions of human cancer genomes. Nat Genet 2015;47:710–6.
3. Zhang W, Bojorquez-Gomez A, Velez DO, Xu G, Sanchez KS, Shen JP, et al. A global transcriptional network connecting noncoding mutations to changes in tumor gene expression. Nat Genet 2018;50:613–20.
4. Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. Highly recurrent TERT promoter mutations in human melanoma. Science 2013;339:957–9.
5. Weinhold N, Jacobsen A, Schultz N, Sander C, Lee W. Genome-wide analysis of noncoding regulatory mutations in cancer. Nat Genet 2014;46:1160–5.
6. Capasso M, Diskin SJ. Genetics and genomics of neuroblastoma. Cancer Treat Res 2010;155:65–84.
7. Barbieri E, De Preter K, Capasso M, Johansson P, Man TK, Chen Z, et al. A p53 drug response signature identifies prognostic genes in high-risk neuroblastoma. PLoS ONE 2013;8:e79843.
8. Formicola D, Petrosino G, Lasorsa VA, Pignataro P, Cimmino F, Vetrella S, et al. An 18 gene expression-based score classifier predicts the clinical outcome in stage 4 neuroblastoma. J Transl Med 2016;14:142.
9. Esposito MR, Binatti A, Pantile M, Coppe A, Mazzocco K, Longo L, et al. Somatic mutations in specific and connected subpathways are associated with short neuroblastoma patients' survival and indicate proteins targetable at onset of disease. Int J Cancer 2018;143:2525–36.
10. Lasorsa VA, Formicola D, Pignataro P, Cimmino F, Calabrese FM, Mora J, et al. Exome and deep sequencing of clinically aggressive neuroblastoma reveal somatic mutations that affect key pathways involved in cancer progression. Oncotarget 2016;7:21840–52.
11. Peifer M, Hertwig F, Roels F, Dreidax D, Gartlgruber M, Menon R, et al. Telomerase activation by genomic rearrangements in high-risk neuroblastoma. Nature 2015;526:700–4.
12. Pugh TJ, Morozova O, Attiyeh EF, Asgharzadeh S, Wei JS, Auclair D, et al. The genetic landscape of high-risk neuroblastoma. Nat Genet 2013;45:279–84.
13. Khurana E, Fu Y, Colonna V, Mu XJ, Kang HM, Lappalainen T, et al. Integrative annotation of variants from 1092 humans: application to cancer genomics. Science 2013;342:1235587.
14. Capasso M, Devoto M, Hou C, Asgharzadeh S, Glessner JT, Attiyeh EF, et al. Common variations in BARD1 influence susceptibility to high-risk neuroblastoma. Nat Genet 2009;41:718–23.
15. Capasso M, Diskin SJ, Totaro F, Longo L, De Mariano M, Russo R, et al. Replication of GWAS-identified neuroblastoma risk loci strengthens the role of BARD1 and affirms the cumulative effect of genetic variations on disease susceptibility. Carcinogenesis 2013;34:605–11.
16. Cimmino F, Avitabile M, Diskin SJ, Vaksman Z, Pignataro P, Formicola D, et al. Fine mapping of 2q35 high-risk neuroblastoma locus reveals independent functional risk variants and suggests full-length BARD1 as tumor-suppressor. Int J Cancer 2018;143:2828–37.
17. Wang K, Diskin SJ, Zhang H, Attiyeh EF, Winter C, Hou C, et al. Integrative genomics identifies LMO1 as a neuroblastoma oncogene. Nature 2011;469:216–20.
18. Diskin SJ, Capasso M, Schnepp RW, Cole KA, Attiyeh EF, Hou C, et al. Common variation at 6q16 within HACE1 and LIN28B influences susceptibility to neuroblastoma. Nat Genet 2012;44:1126–30.
19. Esposito MR, Aveic S, Seydel A, Tonini GP. Neuroblastoma treatment in the post-genomic era. J Biomed Sci 2017;24:14.
20. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013;31:213–9.
21. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 2012;28:1811–7.
22. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
23. Fu Y, Liu Z, Lou S, Bedford J, Mu XJ, Yip KY, et al. FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. Genome Biol 2014;15:480.
24. Ma X, Liu Y, Liu Y, Alexandrov LB, Edmonson MN, Gawad C, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature 2018;555:371–6.
25. He S, Liu Z, Oh DY, Thiele CJ. MYCN and the epigenome. Front Oncol 2013;3:1.
26. Sidell N, Altman A, Haussler MR, Seeger RC. Effects of retinoic acid (RA) on the growth and phenotypic expression of several human neuroblastoma cell lines. Exp Cell Res 1983;148:21–30.
27. Favorov A, Mularoni L, Cope LM, Medvedeva Y, Mironov AA, Makeev VJ, et al. Exploring massive, genome scale datasets with the GenometriCorr package. PLoS Comput Biol 2012;8:e1002529.
28. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841–2.

29. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res 2019;47:W199205.

30. Russo R, Cimmino F, Pezone L, Manna F, Avitabile M, Langella C, et al. Kinome expression profiling of human neuroblastoma tumors identifies potential drug targets for ultra high-risk patients. Carcinogenesis 2017;38:1011–20.

31. Yang D, Jang I, Choi J, Kim MS, Lee AJ, Kim H, et al. 3DIV: a 3D-genome interaction viewer and database. Nucleic Acids Res 2018;46:D52–D7.

32. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 2019;47:D886–D94.

33. Pulford K, Jones M, Banham AH, Haralambieva E, Mason DY. Lymphocyte-specific protein 1: a specific marker of human leucocytes. Immunology 1999;96:262–71.

34. Lu K, Tao H, Si X, Chen Q. The histone H3 lysine 4 presenter WDR5 as an oncogenic protein and novel epigenetic target in cancer. Front Oncol 2018;8:502.

35. Zhou H, Bao J, Zhu X, Dai G, Jiang X, Jiao X, et al. Retinoblastoma binding protein 5 correlates with the progression in hepatocellular carcinoma. Biomed Res Int 2018;2018:1073432.

36. Vedadi M, Blazer L, Eram MS, Barsyte-Lovejoy D, Arrowsmith CH, Hajian T. Targeting human SET1/MLL family of proteins. Protein Sci 2017;26:662–76.

37. Liu W, Tanasa B, Tyurina OV, Zhou TY, Gassmann R, Liu WT, et al. PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. Nature 2010;466:508–12.

38. Tong D, Liu Q, Liu G, Yuan W, Wang L, Guo Y, et al. The HIF/PHF8/AR axis promotes prostate cancer progression. Oncogenesis 2016;5:e283.

39. Wang Q, Ma S, Song N, Li X, Liu L, Yang S, et al. Stabilization of histone demethylase PHF8 by USP7 promotes breast carcinogenesis. J Clin Invest 2016;126:2205–20.

40. Bossone SA, Asselin C, Patel AJ, Marcu KB. MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating transcriptional initiation and termination. Proc Natl Acad Sci U S A 1992;89:7452–6.

41. Jiao L, Li Y, Shen D, Xu C, Wang L, Huang G, et al. The prostate cancer-up-regulated Myc-associated zinc-finger protein (MAZ) modulates proliferation and metastasis through reciprocal regulation of androgen receptor. Med Oncol 2013;30:570.

42. Ray A, Dhar S, Ray BK. Control of VEGF expression in triple-negative breast carcinoma cells by suppression of SAF-1 transcription factor activity. Mol Cancer Res 2011;9:1030–41.

43. Chen L, Alexe G, Dharia NV, Ross L, Iniguez AB, Conway AS, et al. CRISPR-Cas9 screen reveals a MYCN-amplified neuroblastoma dependency on EZH2. J Clin Invest 2018;128:446–62.

44. Li Z, Takenobu H, Setyawati AN, Akita N, Haruta M, Satoh S, et al. EZH2 regulates neuroblastoma cell differentiation via NTRK1 promoter epigenetic modifications. Oncogene 2018;37:2714–27.

45. Kim KH, Roberts CW. Targeting EZH2 in cancer. Nat Med 2016;22:128–34.

46. Dyson N. The regulation of E2F by pRB-family proteins. Genes Dev 1998;12:2245–62.

47. Muller H, Bracken AP, Vernell R, Moroni MC, Christians F, Grassilli E, et al. E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. Genes Dev 2001;15:267–85.

48. Poppy Roworth A, Ghari F, La Thangue NB. To live or let die - complexity within the E2F1 pathway. Mol Cell Oncol 2015;2:e970480.

49. Bejjani F, Evanno E, Zibara K, Piechaczyk M, Jariel-Encontre I. The AP-1 transcriptional complex: Local switch or remote command? Biochim Biophys Acta Rev Cancer 2019;1872:11–23.

50. Boeva V, Louis-Brennetot C, Peltier A, Durand S, Pierre-Eugene C, Raynal V, et al. Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. Nat Genet 2017;49:1408–13.

51. Katoh M, Igarashi M, Fukuda H, Nakagama H, Katoh M. Cancer genetics and genomics of human FOX family genes. Cancer Lett 2013;328:198–206.

52. Wan Q, Tang J, Han Y, Wang D. Co-expression modules construction by WGCNA and identify potential prognostic markers of uveal melanoma. Exp Eye Res 2018;166:13–20.

53. Smith SJ, Towers N, Saldanha JW, Shang CA, Mahmood SR, Taylor WR, et al. The cardiac-restricted protein ADP-ribosylhydrolase-like 1 is essential for heart chamber outgrowth and acts on muscle actin filament assembly. Dev Biol 2016;416:373–88.

54. Horii Y, Beeler JF, Sakaguchi K, Tachibana M, Miki T. A novel oncogene, ost, encodes a guanine nucleotide exchange factor that potentially links Rho and Rac signaling pathways. EMBO J 1994;13:4776–86.

55. Liu Z, Adams HC 3rd, Whitehead IP. The rho-specific guanine nucleotide exchange factor Dbs regulates breast cancer cell migration. J Biol Chem 2009;284:15771–80.

56. Harrison RE, Sikorski BA, Jongstra J. Leukocyte-specific protein 1 targets the ERK/MAP kinase scaffold protein KSR and MEK1 and ERK2 to the actin cytoskeleton. J Cell Sci 2004;117:2151–7.

57. Zhang H, Wang Y, Liu Z, Yao B, Dou C, Xu M, et al. Lymphocyte-specific protein 1 inhibits the growth of hepatocellular carcinoma by suppressing ERK1/2 phosphorylation. FEBS Open Bio 2016;6:1227–37.

58. Koral K, Paranjpe S, Bowen WC, Mars W, Luo J, Michalopoulos GK. Leukocyte-specific protein 1: a novel regulator of hepatocellular proliferation and migration deleted in human hepatocellular carcinoma. Hepatology 2015;61:537–47.

59. Trigg RM, Turner SD. ALK in neuroblastoma: biological and therapeutic implications. Cancer 2018;10. pii: E113.

60. Viprey VF, Gregory WM, Corrias MV, Tchirkov A, Swerts K, Vicha A, et al. Neuroblastoma mRNAs predict outcome in children with stage 4 neuroblastoma: a European HR-NBL1/SIOPEN study. J Clin Oncol 2014;32:1074–83.

Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

AACR American Association for Cancer Research

Transcription Factors Involved in Tumorigenesis Are Over-Represented in Mutated Active DNA-Binding Sites in Neuroblastoma

Mario Capasso, Vito Alessandro Lasorsa, Flora Cimmino, et al.

Updated version	Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-19-2883
Supplementary Material	Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2019/11/28/0008-5472.CAN-19-2883.DC2

Cited articles	This article cites 59 articles, 9 of which you can access for free at: http://cancerres.aacrjournals.org/content/80/3/382.full#ref-list-1

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/80/3/382. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.