

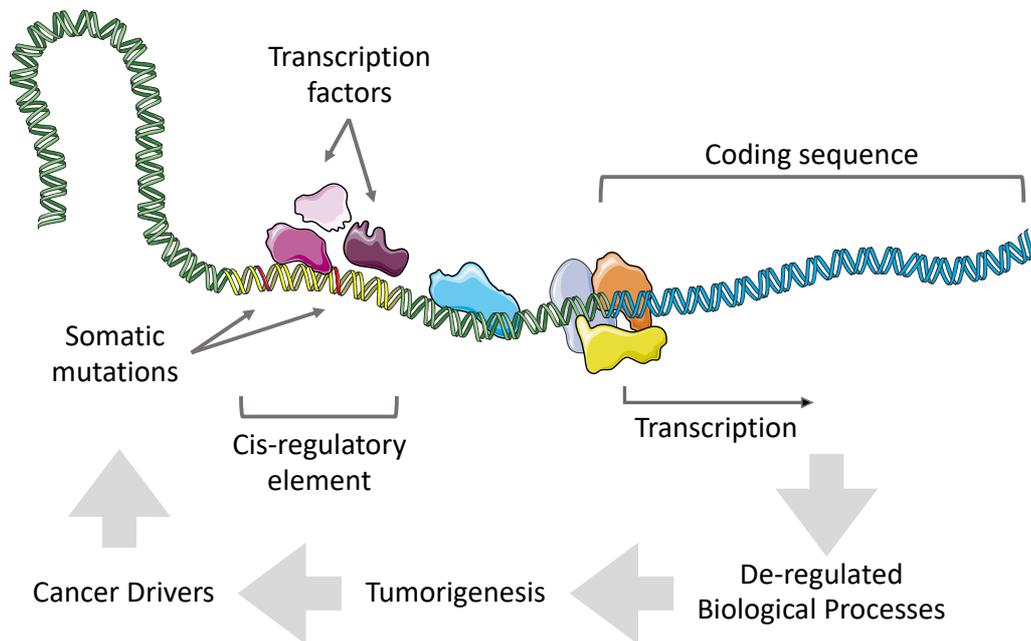
UNIVERSITY OF NAPLES FEDERICO II

DOCTORATE IN MOLECULAR MEDICINE AND MEDICAL BIOTECHNOLOGY

XXXIII CYCLE

Vito Alessandro Lasorsa

Neuroblastoma somatic mutations enriched in cis-regulatory elements collectively affect genes involved in embryonic development and immune system response



Year 2021

UNIVERSITY OF NAPLES FEDERICO II

**DOCTORATE IN
MOLECULAR MEDICINE AND MEDICAL
BIOTECHNOLOGY**

XXXIII CYCLE



**Neuroblastoma somatic mutations enriched in cis-regulatory
elements collectively affect genes involved in embryonic
development and immune system response**

**Tutor
Prof. Mario Capasso**

**PhD Candidate
Vito Alessandro Lasorsa**

Year 2021

Table of contents

List of abbreviations	1
Abstract.....	2
1. Background.....	3
1.1 Regulatory somatic variants as cancer drivers	3
1.2 Somatic variants in cis-regulatory elements	4
1.3 Neuroblastoma	6
1.3.1 Staging and stratification	7
1.3.2 Genetic Predisposition to neuroblastoma.....	9
1.3.3 Genomic features of neuroblastoma.....	10
2. Aim	12
3. Methods.....	13
3.1 Analysis of public ChIP-Seq data and sequencing target selection	13
3.2 Samples collection	13
3.3 DNA extraction from peripheral blood and primary tumor tissues	13
3.4 DNA quantification and library preparation for sequencing.....	14
3.5 Detection of somatic mutations.....	14
3.6 Filtering of somatic mutations	14
3.7 Mutational enrichment in active CREs	15
3.8 Gene expression data analysis.....	16
3.9 DNA extraction, library preparation and HiC sequencing.....	17
3.10 HiC data analysis	17
3.11 Validation data sets of whole genome sequencing (WGS).....	18
3.12 Gene expression analysis, samples clustering and survival analysis	18
3.13 Gene Ontology analysis	18
4. Results	19
4.1 Genome-wide map of CREs	19
4.2 Somatic mutations in active (a)CREs	21
4.3 Mutational enrichment in aCREs.....	25
4.4 Genes interacting with mutated aCREs are involved in neuroblastoma tumorigenesis	31
4.5 Somatic variants in aCREs alter TFs binding motifs.....	34

4.6 Convergence of biological processes of the genes affected by somatic regulatory variants.....	34
5. Discussion.....	38
6. Conclusions	43
8. List of publications.....	44
9. References	47

List of abbreviations

aCRE: active Cis Regulatory Element
BAM: Binary Alignment Map format
ChiP-Seq: Chromatin Immunoprecipitation Sequencing
DHS: DNase Hypersensitive Sites
FDR: False Discovery Rate
GEO: Gene Expression Omnibus
GO: Gene Ontology
GWAS: Genome-Wide Association Study
H3K27ac: Histone H3 acetylated at lysine 27
HiC-seq: all-vs-all Chromosome Conformation Capture Sequencing
INDEL: small insertion/deletion
INRG: International Neuroblastoma Risk Group
INSS: International Neuroblastoma Staging System
NGS: Next Generation Sequencing
PB: Peripheral Blood
PCR: Polymerase Chain Reaction
RNA-Seq: messenger RNA sequencing
SNS: Sympathetic Nervous System
SNV: single nucleotide variants
TAD: Topologically Associating Domain
TD-Seq: Targeted Deep Sequencing
TFBS: Transcription Factor Binding Site
WES: Whole Exome Sequencing
WGS: Whole Genome Sequencing

Abstract

Background | The ever-growing interest towards noncoding cis-regulatory variants as cancer drivers is currently hampered by numerous challenges and limitations of variant prioritization and interpretation methods and tools.

Methods | To overcome these limitations, I focused on active cis regulatory elements (aCREs) to design a customized panel for deep sequencing of 56 neuroblastoma tumor and normal DNA sample pairs. CREs were defined by a reanalysis of H3K27ac ChIP-seq peaks of 25 neuroblastoma cell lines. aCREs were further identified by the presence of open chromatin as defined by DNase Hypersensitivity sites. This provided a small subset of genomic regions with evident regulatory functions in which to search for driver mutations. I tested these regions for an excess of somatic mutations and assessed the statistical significance with a global approach accounting for chromatin accessibility and replication timing. Additional validation was provided by analyzing whole genome sequences of 151 neuroblastomas. For the mutated regions, I determined their candidate target genes through HiC data analysis.

Results | I identified a significant excess of somatic mutations in aCREs of diverse genes including *IPO7*, *HAND2*, and *ARID3A*. A gene expression signature built on basis of these three, and nearby, interacting genes strongly correlated with negative prognostic markers and low survival rates of patients affected by neuroblastoma. Moreover, I observed a convergence of biological functions of the target genes of mutated aCREs and transcription factors with binding motifs altered by mutations towards processes related to embryonic development and immune system response.

Conclusion | My strategy led me to identify somatic mutations in regulatory elements that collectively can drive neuroblastoma onset.

1. Background

1.1 Regulatory somatic variants as cancer drivers

The advent of Next Generation Sequencing (NGS) has revolutionized many areas of research including cancer genomics. To date, hundreds of samples could be rapidly sequenced and thousands of mutations could be quickly analyzed. However, as most high throughput sequencing studies of cancer relied on the Whole Exome Sequencing (WES) and focused mainly on the protein-coding part of the genome (1% to 2%), only the contribution of coding mutations to oncogenesis has been largely clarified. The remaining 98% of the genome is still unexplored. Recently, wide international efforts have shed light on the characteristics of the non-coding genome. Through Chromatin Immunoprecipitation Sequencing (ChIP-Seq) and DNase Hypersensitive Sites (DHS) assays, to date, we have better knowledge of where transcription factors bind to DNA and where chromatin is open (or closed). In this regard, public data from large epigenomic projects such as ENCODE [1], as well as those from smaller projects, have become an essential source of information for researchers.

Nevertheless, the role of somatic mutations in regulatory regions (i.e. TFBSs and cis-regulatory elements) remains underestimated. However, in recent years, there is growing interest in the study of noncoding variants as drivers of carcinogenesis. The most recent literature clearly demonstrates that altered transcriptional regulatory circuits play relevant roles in cancer development [2]. For instance, a re-analysis 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 [3]. Another recent study has demonstrated that noncoding mutations can affect the gene expression of target genes in a large number of tumors [4]. Furthermore, noncoding single nucleotide variants (SNVs) in the promoters of *TERT*,

FOXA1, *PLEKHS1*, *WDR74* and *SDHD* [5-7] and in a cis-regulatory element of *ETVI* [8] have been identified as cancer drivers.

By analyzing the whole genome sequences of 151 neuroblastoma primary tumors, with my group, I have recently demonstrated that somatic SNVs do not occur randomly across noncoding regulatory regions but rather target TFBSs of a specific set of tumorigenic transcription factors (TFs) suggesting that combining mutations within a set of regulatory regions act in network to drive cancer initiation [9].

So far, no other study has investigated the pathogenic implications of noncoding SNVs in neuroblastoma.

1.2 Somatic variants in cis-regulatory elements

Despite the above-mentioned recent advances, it remains a challenge to identify noncoding cancer driver mutations and assess the pathogenic repercussions of these variants in cancer development. A common approach to prioritize somatic noncoding SNVs is to determine genomic regions with high mutation frequency across different cancer samples. However, distinguishing driver from passenger noncoding mutations is challenging since mutation rates can be affected from replication timing [10] and chromatin conformation [11]. Moreover, other factors such as the complexity of the human genome, the need of large cohorts to recruit and the thousands of somatic mutations to analyze, do diminish the chances to obtain significant results after correcting for multiple hypothesis testing. In this regard, cis-regulatory elements (CREs) might represent a highly enriched subset of the regulatory regions of the genome in which to search for such mutations (**Figure 1**). Another advantage of focusing the analyses on CREs is that they are characterized by a peculiar mutation rate that is lower with respect to other genomic regions [7,8]. Thus, analyses of CREs may reduce the risk of obtaining false negative or positive results.

In the last years, the deep sequencing, usually exploited to study the coding DNA, has made a huge contribution to better understand the complex picture of genetic variation in cancer. Indeed, now we know that a compendium of rare subclones of coding variants exist in human tumors that can be responsible for drug resistance, invasion, metastasis, and relapse [12].

However, only few studies have used the targeted deep sequencing (TD-Seq) approach to search for somatic mutations in CREs. For instance, a recent work on breast cancer reported recurrent mutations in the promoter of the cancer driver gene *FOXAI* by deep sequencing of regulatory elements [6].

Therefore, in view of the above observations, I feel that focusing on CREs and using TD-Seq might overcome the limitations of noncoding driver analysis and could lead to the identification of recurrently mutated regions regulating genes involved in tumorigenesis.

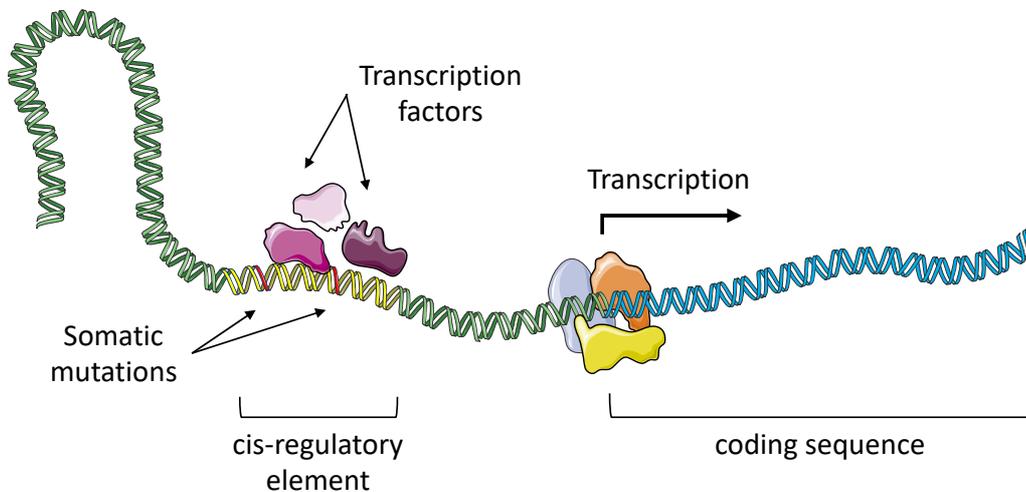


Figure 1 | Cis-regulatory element. Schematic representation of somatic mutations (in red) in cis-regulatory element (in yellow) and its coding sequence (in blue).

1.3 Neuroblastoma

Neuroblastoma is an early childhood tumor. It is the most common malignancy diagnosed in children during their first year of life. It counts 25-50 cases per million individuals [13]. The 90% of neuroblastomas are diagnosed in children less than 10 years old. Its median age of diagnosis is of 18 months [14].

This neuroendocrine tumor arises during the embryonal development of the Sympathetic Nervous System (SNS), which originates from the neural crest. During development, the neural crest gives rise to several cell types including peripheral neurons, enteric neurons, glia, melanocytes, Schwann cells, cells of the craniofacial skeleton and adrenal medulla [15] (**Figure 2**). As a result, neuroblastoma can arise at various anatomical sites along the SNS.

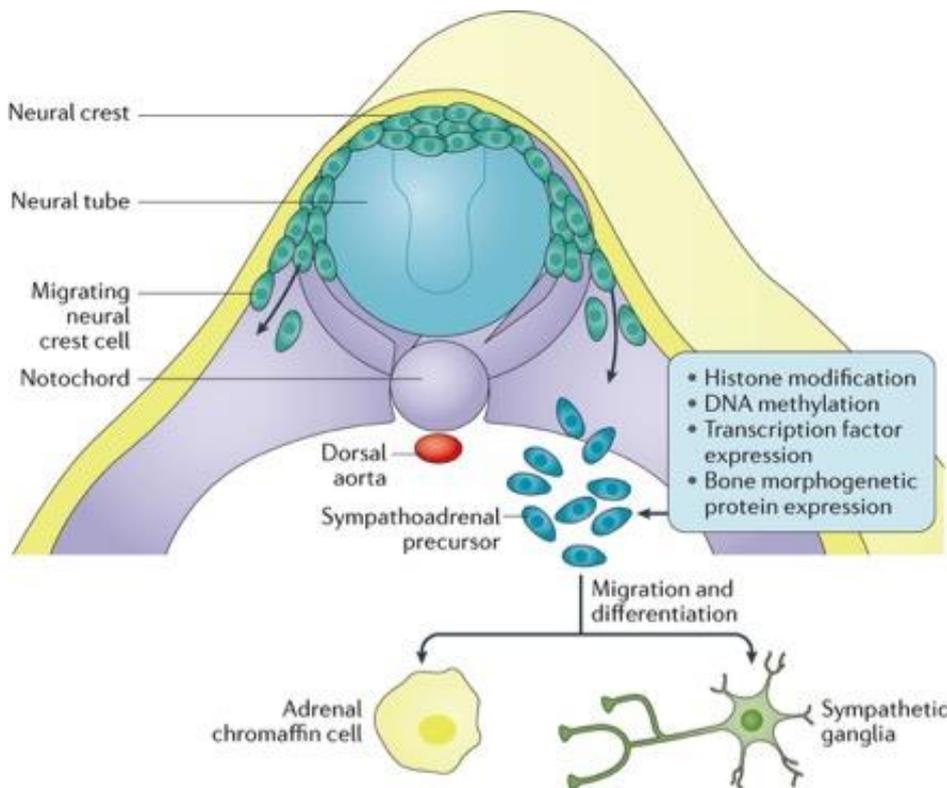


Figure 2 | Cells of origin of neuroblastoma. Adapted from Matthay, KK. et al. (2016) [16].

Nevertheless, in the majority of cases it affects the adrenal glands and the sympathetic ganglia (**Figure 3**) [16].

The pathologic classification of neuroblastoma is based on the differentiation grade of its cells; the lower the differentiation, the higher the aggressiveness. It shows high biologic (and genetic) heterogeneity and a wide range of clinical behaviors. Despite being among the most lethal childhood cancers, it is associated with one of the highest proportions of spontaneous and complete regression of all human cancers [17,18].

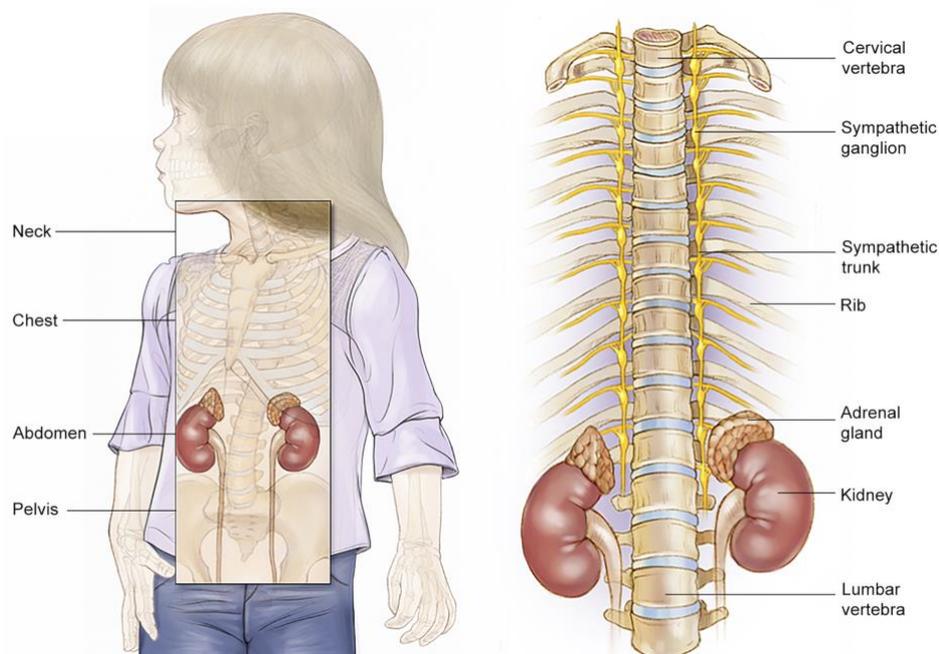


Figure 3 | Anatomical sites of neuroblastoma tumors. Adapted from American Society of Clinical Oncology 2005, cancer.net.

1.3.1 Staging and stratification

Diagnosis and staging criteria are based on the International Neuroblastoma Staging System (INSS), first published in 1988 (**Figure 4**). INSS staging has been used for more than 20 years. It is based on the localization of the primary tumor and on the possibility to eradicate it through surgery and/or therapy

approaches. The INSS system, together with the age at diagnosis, allowed physicians to determine patient’s risk and to plan treatment strategies [19]. Children older than 18 months with Stage 4 (metastatic disease) are at high risk for death from refractory disease. In contrast, infants younger than 12 months with localized tumors (Stages 1-2) are effectively cured, often without cytotoxic therapies.

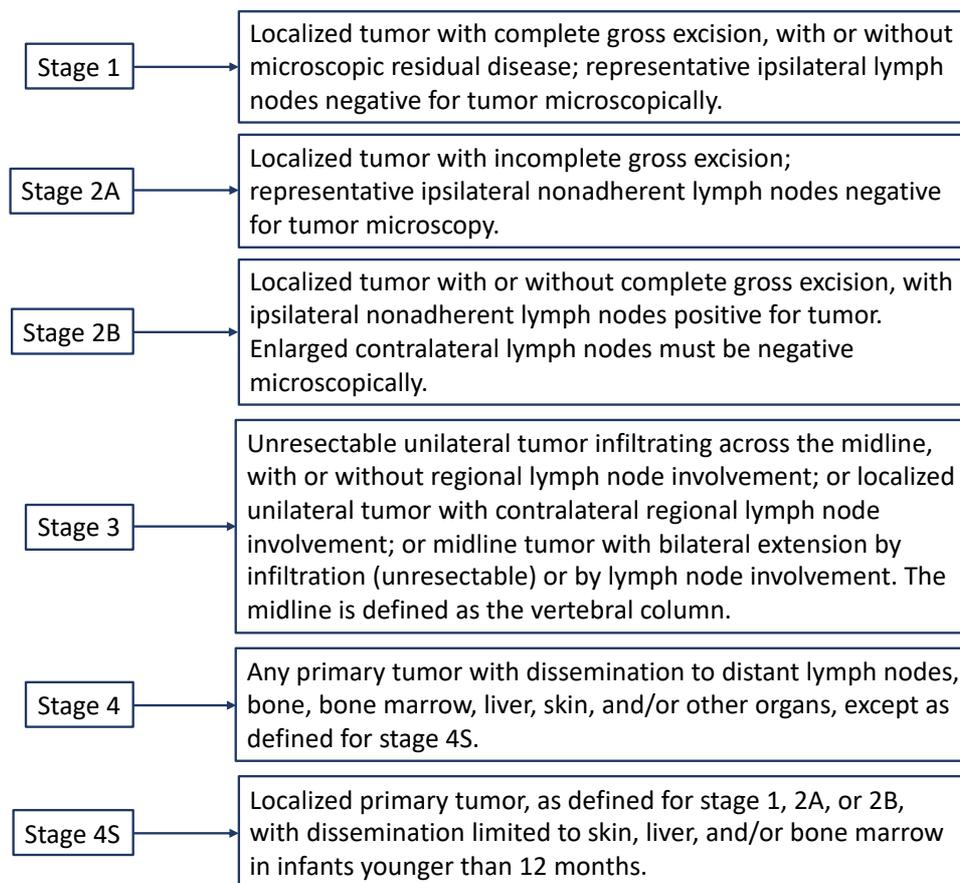


Figure 4 | INSS stages. The international neuroblastoma staging system (INSS).

More recently, in 2009, the International Neuroblastoma Risk Group (INRG) elaborated an updated Staging System including INSS Stage, age at diagnosis, histology, tumor grade, *MYCN* status, 11q alteration status, and DNA ploidy [19]. Based on this system, neuroblastoma patients can be stratified in risk

groups (very low, low, intermediate, and high-risk) on the basis of 5 years survival probabilities (**Figure 5**).

The overall 5-year survival rate in high-risk patients is lower than 40%, despite decades of considerable international efforts to improve the outcome [20].

INRG Stage	Age (months)	Histologic Category	Grade of Tumor Differentiation	MYCN	11q Aberration	Ploidy	Pretreatment Risk Group	
L1/L2		GN maturing; GNB intermixed					A Very low	
L1		Any, except GN maturing or GNB intermixed		NA			B Very low	
				Amp			K High	
L2	< 18	Any, except GN maturing or GNB intermixed		NA	No		D Low	
				NA	Yes		G Intermediate	
	≥ 18		GNB nodular; neuroblastoma	Differentiating	NA	No		E Low
				Poorly differentiated or undifferentiated	NA	Yes		H Intermediate
				Amp		N High		
M	< 18			NA		Hyperdiploid	F Low	
	< 12			NA		Diploid	I Intermediate	
	12 to < 18			NA		Diploid	J Intermediate	
	< 18			Amp			O High	
	≥ 18						P High	
MS	< 18				No		C Very low	
				NA	Yes		Q High	
				Amp			R High	

Figure 5 | INRG classification. International Neuroblastoma Risk Group classification system.

1.3.2 Genetic Predisposition to neuroblastoma

Several germline mutations have been associated with a genetic predisposition to neuroblastoma, including mutations in *ALK* (Anaplastic Lymphoma Kinase) [21] and *PHOX2B* (Paired Like Homeobox 2B; Neuroblastoma Phox) [22]. The primary cause of familial neuroblastoma (about 1% - 2% of patients with neuroblastoma) is a germline mutation in the *ALK* gene (about 75% of familial cases) [21]. Further, also sporadic neuroblastoma shows a germline contribution, either with modest effect sizes for common polymorphic alleles or with greater effect sizes for rare pathogenic variants. As an example of the

latter, rare germline variants of *BARD1* have been identified in children with high-risk neuroblastoma [23].

Recent studies have shown that disease-associated genomic variation is commonly located in regulatory elements in the human population [24]. Our genome-wide association studies (GWASs) on DNA from peripheral blood have revealed that many neuroblastoma susceptibility variants, lie in noncoding regions of the genome [25]. Functional investigations have shown that the cancer genes *LMO1* [26], *BARD1* [27,28], *LIN28B* [29], *SLC16A1* [30] whose expression is affected by such risk variants, play a role in neuroblastoma tumorigenesis.

1.3.3 Genomic features of neuroblastoma

Neuroblastoma genomes are shaped by large chromosomal aberrations, low rates of exonic mutations and genomic alterations promoting telomere maintenance.

In young children, *MYCN* amplification (defined as more than 10 copies per diploid genome) is one of the most common segmental chromosomal aberrations. Moreover, other chromosome arm level alterations such as deletion of 1q (30%) and 11q (45%) and unbalanced gain of 17q (60%) are reported as poor prognostic features [31]. In a recent work, with my research group, I demonstrated that children older than 6 years present unique structural variants with 19p loss and 1q gain among those more recurrent [32].

Recently, we and others highlighted that recurrent somatic coding mutations are infrequent in primary neuroblastoma with activating mutations in *ALK* and inactivating mutations in *ATRX* as the most frequently reported [33-35]. Moreover, rearrangements activating *TERT* locus [36] and mutations affecting genes in RAS and TP53 pathways occur in high-risk neuroblastoma [37].

Together, these studies suggest that an increased understanding of genomic alterations of neuroblastoma could impact patient prognosis and response to therapies. Detailed genomic information leading to new drug targets is also the starting point to develop more effective and less toxic treatments [38].

2. Aim

Decades of international efforts have clarified the contribution of both germline and somatic coding mutations to neuroblastoma genesis. As most high throughput sequencing studies of cancer focused mainly on the protein-coding part of the genome, little is known about the impact of somatic mutations in noncoding DNA. Furthermore, the growing interest towards noncoding cis-regulatory variants as cancer drivers is currently hampered by numerous challenges and limitations of variant prioritization and interpretation methods and tools. So far, no study has investigated the pathogenic implications of noncoding SNVs in neuroblastoma.

I conducted this study with the hypothesis that mutated active regulatory elements could de-regulate genes involved in the tumorigenesis of neuroblastoma. In view of the above observations, I focused on CREs and exploited the TD-Seq approach to overcome the limitations of noncoding driver analysis.

The main aim of this study was to identify somatic mutations in CREs involved in the genesis of neuroblastoma.

With this aim, I set up an alternative method to identify regulatory driver mutations in neuroblastoma. First, I defined the CREs in neuroblastoma by an analysis of H3K27ac ChIP-seq peaks shared among 25 neuroblastoma cell lines. Then, I used this set of CREs as a target to perform TD-Seq of 56 neuroblastomas and normal DNA sample pairs. Then, I tested these regions for an excess of somatic mutations using a global approach accounting for replication timing. Then, I validated the mutated CREs in whole genome sequencing data from 151 neuroblastomas. Finally, I verified if the somatic mutations significantly enriched in CREs can collectively affect specific biological processes.

3. Methods

3.1 Analysis of public ChIP-Seq data and sequencing target selection

H3K27ac Chromatin Immunoprecipitation Sequencing (ChIP-Seq) data of neuroblastoma cell lines, deposited in NCBI Gene Expression Omnibus (GEO) with accession GSE90683 [39], were downloaded and re-analysed as described in [30]. As reported by Boeva et al. [39], the landscape of super-enhancers in neuroblastoma cell lines could delineate 3 groups of cell lines with distinct characteristics. The first group, with a sympathetic noradrenergic identity, counted 18 cell lines. The second, with an NCC-like identity, included 3 cell lines. The third counted 4 cell lines of a mixed type. To select the target of my sequencing panel, I used the BEDTools suite [40] and proceeded as follows. First, I reduced the size of peaks larger than 1100 bp to the 80% of their length (40% up- and down-stream). Second, I obtained the intersections of H3K27ac peaks. Here, if one common region was shorter than 301 bp, I extended it up to 301 bp. Finally, I merged overlapping, book-ended and close regions (up to 100 bp) to get the final set of intervals (N = 13,667; Target length = 9,804,818).

3.2 Samples collection

Neuroblastoma tumor DNA (primary tumors) and matched germline DNA (from peripheral blood) were obtained from Istituto di Ricerca Pediatrica, Padova, 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 University of Naples Federico II.

3.3 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.

3.4 DNA quantification and library preparation for sequencing

DNA quality was monitored on 1% agarose gels. Its purity was checked using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA). DNA concentration was measured using Qubit® DNA Assay Kit in Qubit® 2.0 Fluorometer (Life Technologies, CA, USA). A total of 1.0µg 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 USA) following manufacturer's recommendations. Genomic DNA was sonicated to a size of 350bp, 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, CA, USA) and quantified using real-time PCR.

3.5 Detection of somatic mutations

TD-Seq of 56 normal-primary neuroblastoma sample pairs was performed on an Illumina HiSeq1500 platform. The sequencing produced paired-end reads of 150bp. Mapping BAM files were obtained with BWA and SAMTools [41,42] by aligning the reads versus the GRCh37/hg19 reference genome assembly. Somatic SNVs and small insertions and deletions (INDELS) were detected with MuTect [43] and Strelka [44], respectively. The functional annotation of somatic variants was performed with ANNOVAR [45] and FunSeq2 [46].

3.6 Filtering of somatic mutations

From raw somatic mutation calls, I first discarded those that did not pass the variant callers (MuTect or Strelka) quality filters. To remove possible false

positives, I eliminated somatic mutations falling in genomic duplicated regions. I filtered out common polymorphisms (minor allele frequency > 1%) by using allele frequencies of non-Finnish European populations in 1000 Genomes Project, ExAC and gnomAD databases. I removed off-target mutations. Finally, I retained for downstream analyzes only somatic SNVs falling in DNase Hypersensitive Sites (DHS) of the SK-N-SH neuroblastoma cell line (ENCODE v3) as the DNA regions characterized by H3K27ac and DHS signal peaks are considered as transcriptionally active CREs (aCREs).

3.7 Mutational enrichment in active CREs

After applying stringent quality controls and filtering of somatic variants, the selected SNVs/INDELs were analyzed to distinguish driver ones from passengers using a global approach as previously described [7,8]. In brief, with this approach, I searched for an excess of noncoding mutations in aCREs by assuming that the observed number of tumor samples mutated in any specific region, follows a binomial distribution. This is dependent on the background mutation frequency that, in turn, depends on the effective length of the region and other genomic characteristics, such as the replication timing and the chromatin accessibility. Indeed, the significantly mutated aCREs were identified, by accounting for these factors, as follows.

I assumed that the observed number of tumor samples mutated in any specific region followed a binomial distribution, binomial (n, p_i) , where n is the total number of samples with mutation data and p_i is the estimated sample mutation rate for region of interest i under the null hypothesis that the region is not recurrently mutated. I therefore computed the following P value:

$$P(X \geq k) = 1 - P(X < k) = 1 - \sum_{j=0}^{k-1} \binom{n}{j} p_i^j (1 - p_i)^{n-j}$$

I also assumed that p_i depended on the effective length L_i of the region and the estimated nucleotide mutation rate q_i for the region under the null hypothesis as follows:

$$p_i \sim 1 - (1 - q_i)^{L_i}$$

I thus determined the background mutation frequency q_i that was needed to compute a P value using the above equation. Furthermore, to account for genomic factors, such as the replication timing, that affect the nucleotide mutation rates, I corrected the nucleotide mutation rate using replication timing from HeLa, K562, HEPG2, MCF7 and SKNSH cell lines [10]. Briefly, I computed the mean replication times for each cell line across genomic windows (bins) of 100 Kb. Then, I assigned to each aCRE its corresponding mean replication time. Subsequently, for each aCRE, I identified the top 5% of aCREs with the most similar replication times across the cell lines, measured using the Euclidian distance between the vectors of times. The global background nucleotide mutation frequency was then estimated by dividing the total number of observed mutations in the top 5% of the regions of interest by the effective length of these regions. P values were computed using the equation above and adjusted for multiple testing with using the Benjamini-Hochberg method.

3.8 Gene expression data analysis

Gene expression data (RNA-Seq) of neuroblastoma cell lines, deposited in GEO database with accession GSE90683 [39], were downloaded and analyzed as processed files. Gene expression was reported as Log₂ transformed FPKM values.

3.9 DNA extraction, library preparation and HiC sequencing

Cells were prepared as follows. The cell culture of a minimum of 2.5×10^7 cells was verified for its purity. The protocol does not allow DNA contamination. Cells were cross-linked with formaldehyde and lysed. Then, chromatin was digested with restriction enzyme. The resulting sticky ends were filled and marked in with biotinylated nucleotides. Ligation was performed under extremely dilute conditions to create chimeric molecules. The chromatin complexes containing the biotin-labeled ligation products were degraded by proteinase. Then genomic DNA was purified and randomly sheared into short fragments of 350bp by Covaris apparatus. Biotinylated junctions were isolated with streptavidin beads. Then DNA fragments were end polished, A-tailed, and ligated with adapter for Illumina sequencing, further PCR enriched by P5 and indexed P7 oligos. The PCR products were purified and the concentration of the library was determined by Qubit® 2.0 fluorometer. After dilution to 1 ng/ μ L, the resulted libraries were analyzed for size distribution by Agilent® 2100 Bioanalyzer and quantified using real-time PCR. The sequencing of the qualified libraries was performed on an Illumina® HiSeq platform.

3.10 HiC data analysis

The sequencing was performed on an Illumina® HiSeq platform. Paired-end reads with length of 150bp were mapped to the reference genome (build hg19/GRCH37) with Bowtie2 [47]. The alignment BAM file was then filtered to remove duplicates, re-ligation or self-circularization artifacts that can be introduced during Hi-C library preparation. Then I used HiCExplorer tool v2.1.4 [48] to (i) build the interaction matrix at a resolution of 10Kb (bin size=10Kb); [49] normalize the observed interaction matrix; (iii) determine Topologically Associating Domains (TADs, self-interacting genome regions) and their boundaries; and (iv) plot the results. Subsequently, I extended the target CRE regions of interest of 1Mb up- and down-stream and calculated the statistical significance of the interactions between bins with the FitHiC v1.1.3

program [50]. P-values were corrected for multiple tests by Benjamini-Hotchberg method (False Discovery Rate, FDR) and the cutoff was set at 5%. Finally, I annotated those bins with ANNOVAR [45] in order to map genomic bins to gene coordinates.

3.11 Validation data sets of whole genome sequencing (WGS)

In order to validate the presence of somatic mutations in aCREs, I analyzed 14 primary neuroblastoma whole genomes published in [9] and used public data of 137 whole genomes from the TARGET neuroblastoma project for which our lab had authorized access (Accession: phs000218.v21.p7; Project ID: #14831) [35]. Here, I searched for somatic mutations by extending the length of each aCRE of the 50% on both sides.

3.12 Gene expression analysis, samples clustering and survival analysis

I used the R2: Genomics Analysis and Visualization Platform (<http://r2.amc.nl>) to query transcriptomic data of 498 neuroblastoma samples (GSE62564). I used the k-means clustering algorithm, to divide samples in two groups based on expression levels of genes that significantly interacted with mutated aCREs. The overall (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%.

3.13 Gene Ontology analysis

The functional enrichment analysis was performed by using the web tool: WEB-based GENE SeT AnaLysis Toolkit [51]. The Gene Ontology database of non-redundant Biological Processes was used. Significantly enriched GO terms were considered for $FDR \leq 0.05$.

4. Results

4.1 Genome-wide map of CREs

CREs such as promoters and enhancers, in accessible chromatin are commonly bookmarked by the acetylation of lysin 27 on histone 3 (H3K27ac). To define CREs in neuroblastoma, I used the H3K27ac ChIP-seq peaks (GSE90683) shared among 25 neuroblastoma cell lines (see **Methods**) that represent three different tumor identities: sympathetic noradrenergic, NCC-like and mixed type [39]. As outlined in **Figure 6**, I selected 13,437 common peaks (regions overlapping H3K27ac marker) with mean length of about 730 bp (min=300, max=8,265). This represented the target (of about 9.8 Mb) for the TD-Seq of a set of 56 tumor-normal pairs (**Table 1**).

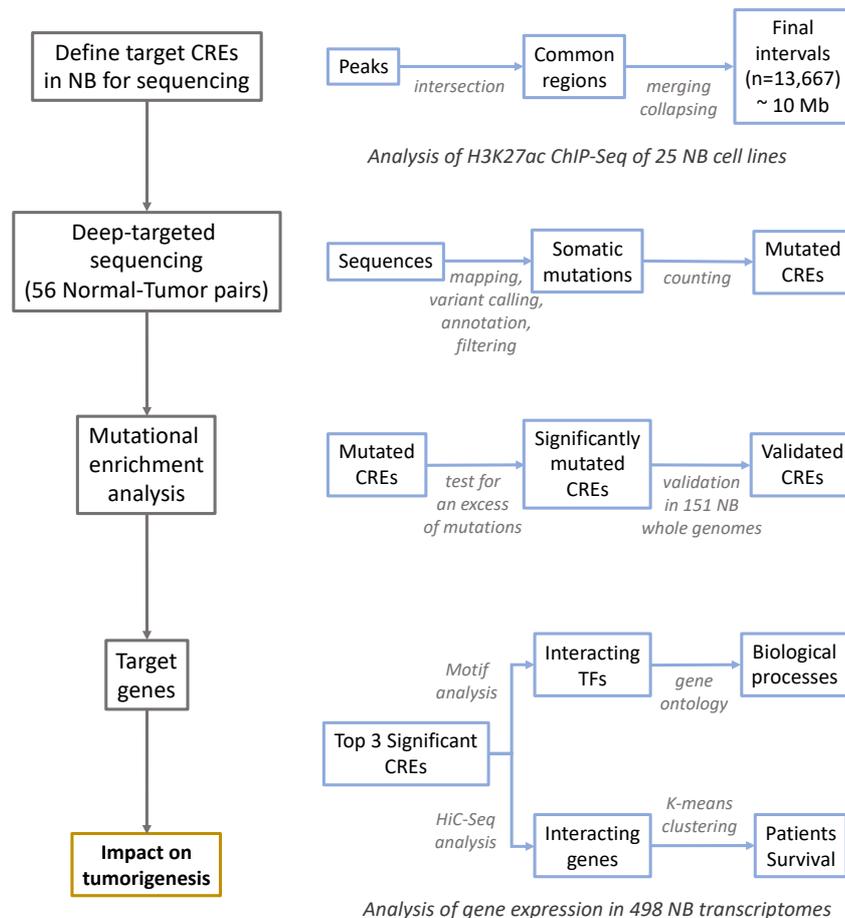


Figure 6 | Schematic workflow describing the analyses steps. First, I defined active cis-regulatory elements in neuroblastoma by a re-analysis of public ChIP-Seq data of H3K27ac. I selected and adjusted the size of the common regions to get the final targets for sequencing of a set of 56 neuroblastoma samples with their matched control tissue. I processed these data and called somatic variants. After filtering the raw somatic variants, I mapped the variants to aCREs and obtained mutation counts for each aCRE. Then, I tested these regions for an excess of somatic mutations and assessed the statistical significance. Further, I searched for additional mutations in aCREs by using the WGS data of 151 neuroblastomas. Then, I identified the candidate target genes of each significantly mutated aCREs by analyzing our HiC-Seq data of the SK-N-BE neuroblastoma cell line. In parallel, I performed a TF binding motif analysis to identify the altered binding motifs of TFs targeting the aCREs. Finally, I focused on the impact on tumor development. On the basis of the expression of target genes I performed a K-means clustering of 498 neuroblastoma samples of a public data set. In brief, I divided these samples in two groups and evaluated survival rates and clinical prognostic markers by k-means group. In parallel, based on the TFs with altered binding motifs, I performed a GO enrichment analysis to find out which biological processes they were involved in.

Table 1 | Clinical features of collected neuroblastoma samples.

Clinical Parameters	Samples (n=56)
INSS stage 1	8/56 (14%)
INSS stage 2	4/56 (7%)
INSS stage 3	12/56 (21%)
INSS stage 4	29/56 (52%)
INSS stage 4s	3/56 (6%)
Age >18	25/56 (45%)
Age ≤18	31/56 (55%)
MYCN amplified	10/56 (18%)
MYCN not amplified	46/56 (82%)
High risk	27/56 (48%)
Non High risk	29/56 (52%)

4.2 Somatic mutations in active (a)CREs

I collected high quality sequencing data with, on average, the 1.94% of reads discarded for low quality (**Figure 7A**).

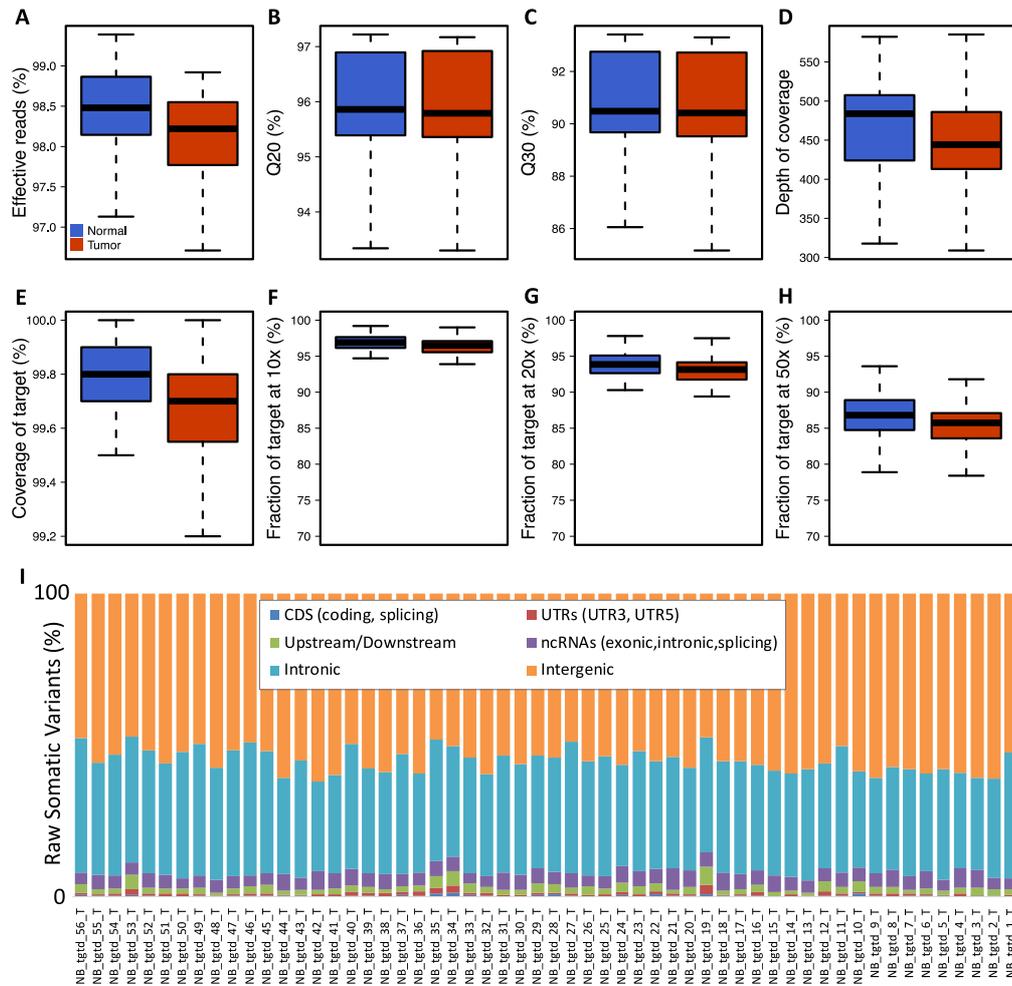


Figure 7 | Descriptive statistics of the Targeted Sequencing of 56 normal-primary NB sample pairs. (A) Box plot reporting the percentage of clean reads used for the alignment. (B) The percentage of bases with quality score above 20 (Q20). (C) The percentage of bases with quality score above 30 (Q30). (D) Box plot reporting the average Depth of coverage obtained after the alignment. (E) Box plot reporting the average coverage of target regions and the percentage of target regions covered with at least 10 (F), 20 (G) and 50 reads (H). (I) The bar plot shows the functional categories of the annotated raw somatic variant calls. The functional annotation was performed with ANNOVAR by using the NCBI-RefSeq database.

The percentage of bases with quality scores above 20 (Q20) and above 30 (Q30) was 95.91% and 90.79%, respectively (**Figure 7B-C**). After the alignment I obtained a depth of coverage of 459x (**Figure 7D**) and I covered the 99.63% of the target regions (**Figure 7E**). Furthermore, the fraction of the target covered with at least 10, 20 and 50 reads, was, on average, the 96.24%, 92.89% and the 85.31%, respectively (**Figure 7F-H**). This allowed me to obtain a highly reliable set of raw somatic variant calls. Indeed, I identified a total of 83,962 somatic SNVs/INDELs mainly mapping to intronic (36.80%) and intergenic (55.14%) regions (**Figure 7I**). I applied stringent criteria to filter raw somatic variant calls. As detailed in **Methods**, I discarded poor quality variants, off targets, common polymorphisms, and possible false positives mapping to problematic genomic regions. Here, I wanted to focus on somatic variants falling in transcriptionally active CREs (aCREs) characterized by the presence of DNase I hypersensitive sites (DHSs). Therefore, I further removed variants that did not map to these sites that indicate regions of open chromatin.

After this multistep filtering, I retained a total of 1,296 SNVs/INDELs in aCREs (median per sample = 10). As reported in **Figure 8A**, I obtained 1.0183 variants per Mb. The filtered somatic variants were mainly located in regions closer to genes (such as exonic, ncRNA exonic, UTR5, upstream) if compared to the raw somatic variants (**Figure 8B**). Indeed, whereas intergenic mutations drastically decreased from 52% (raw variants) to 7.10% (filtered variants), the percentage of intronic mutations remained comparable (42.50% and 43.60%, respectively). Furthermore, among the filtered variants, I observed an increased proportion of UTR (1.40% of raw and 15.70% of filtered variants, respectively), of exonic (0.84% of raw and 11.60% of filtered variants, respectively), and of upstream/downstream variants (3.30% of raw and 21.90% of filtered variants, respectively) (**Figure 8B**). Together, these data support the hypothesis that the selected somatic variants can affect DNA regions (aCREs) that regulate the transcription of neighboring genes.

Based on this hypothesis, I asked if gene expression could be influenced by neighboring aCREs. Overall, I found that the presence of the aCREs, identified herein, correlated with increased expression of proximal (up to 0.25 Mb away, mean=2.95) and distal genes (up to 0.5 Mb away, mean=2.82) when I analyzed RNA-Seq data from the same set of 25 neuroblastoma cell lines (GSE90683) ($P < 1.0 \times 10^{-04}$) (**Figure 8C**). As reported in **Figure 8D-E**, the selected somatic variants showed significantly higher pathogenicity, assessed with CADD [52] and FunSeq2 [46] scores, when compared to the raw set of somatic variants; ($P < 2.2 \times 10^{-16}$). I found 60 missense SNVs (83.33%) and 12 truncating SNV/INDELS (16.66%) (**Figure 8F**) but only three of these hit known cancer genes (reported in COSMIC catalog): *ARID1A* (1 inframe deletion) and *ARID1B* (2 inframe deletions in two different patients), *GATA3* (1 missense SNV). Somatic deleterious mutations of the chromatin remodelers *ARID1A* and *ARID1B* have been already reported in neuroblastoma tumors [53]. This data further supports the reliability of the results obtained with the sequencing strategy.

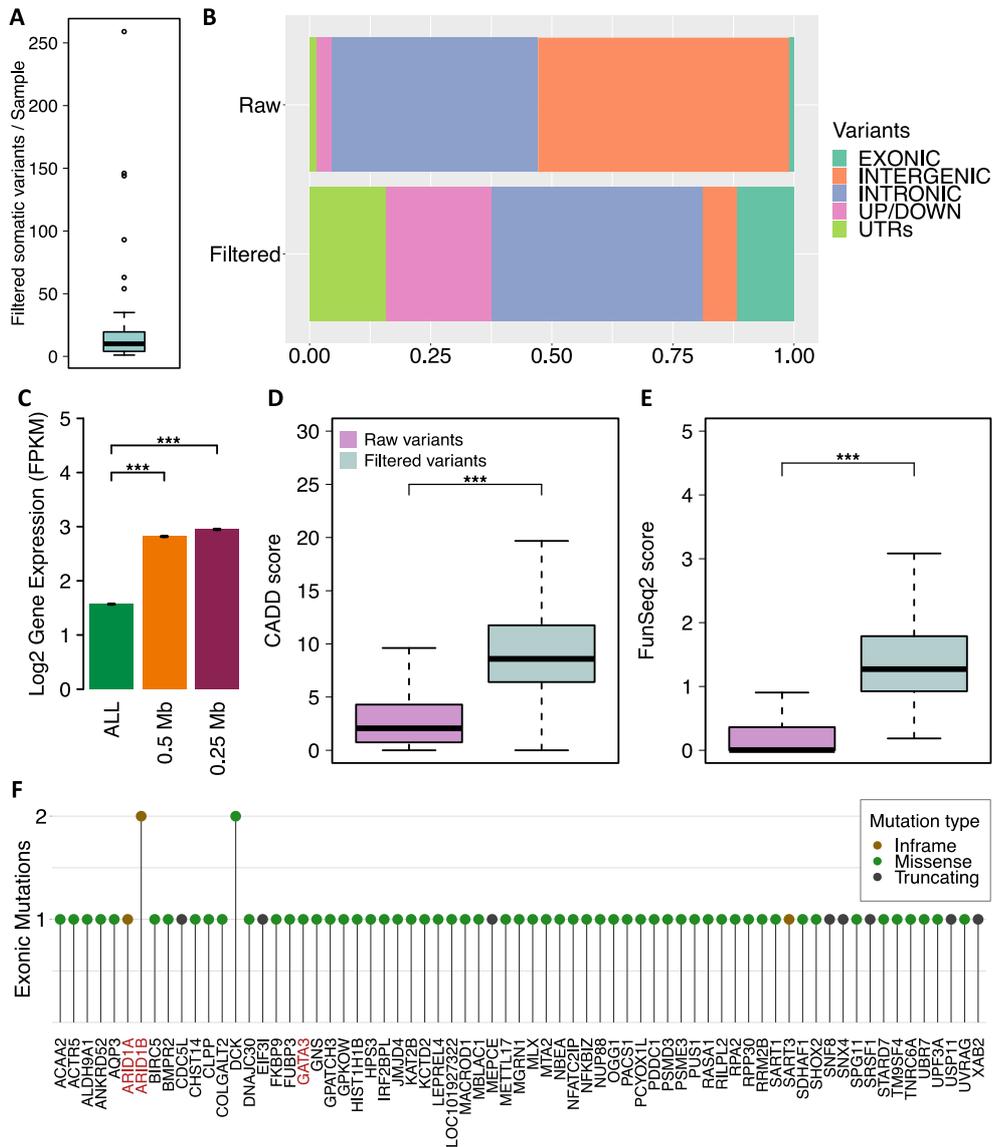


Figure 8 | Characteristics of the selected somatic variants.

(A) Box plot showing the median number of filtered somatic variants per Megabase (Mb). (B) Stacked bar plot reporting the location of Raw and Filtered somatic variants. (C) Bar plot with the median expression of "ALL" genes in the RNA-Seq dataset of neuroblastoma cell lines (GSE90683) compared to the median expression of aCREs proximal (up to 0.25 Mb away) and distal genes (up to 0.5 Mb away). (D) Box plot showing the median levels of CADD pathogenicity scores for Raw and Filtered somatic variants. (E) Box plot showing the median levels of FunSeq2 pathogenicity scores for Raw and Filtered somatic variants. (F) Lollipop plot reporting the mutated genes with number and type of somatic exonic variants. Gene names written in red indicate known cancer genes. $P < 0.05$ (°), $P < 0.01$ (*), $P < 0.001$ (**), $P < 0.0001$ (***) . P value was calculated by Mann-Whitney Test.

4.3 Mutational enrichment in aCREs

To assess if aCREs could harbor more somatic variants than the expected, I tested these regions for an excess of noncoding mutations using a global approach that corrects the mutation rate for replication time as previously described [7,8] and detailed in **Methods** section. In total, 1151 individual aCREs contained at least one mutation. One aCRE was significantly enriched for mutations with FDR<0.05 and 8 with FDR<0.15 (**Table 2**). The validation performed on a set of 151 tumor normal pairs analyzed by WGS showed that two out of the 9 aCREs were respectively mutated in additional two patients (**Table 2**).

Table 2 | Recurrently mutated CREs in 56 neuroblastoma samples.

CRE region	Closest genes	Pathway	Region	Target (bp)	Mutations TD-Seq (N)	Pvalue	FDR	Mutations WGS (N)
chr11:9385430-9386340	TMEM41B(dist=49770) IPO7(dist=20084)	Nuclear Import	intergenic	910	4	3.70E-05	0.043	0
chr4:174440675-174448940	SCRG1(dist=125307) HAND2(dist=1728)	Heart Development	intergenic	8265	3	3.65E-04	0.114	2
chr6:157099410-157101080	ARID1B	Chromatin organization	exonic	1670	3	5.79E-04	0.114	0
chr2:198379730-198381565	HSPE1-MOB4	-	intronic	1835	3	6.69E-04	0.114	0
chr19:925145-927105	ARID3A	Regulation of TP53 Activity	intronic	1960	3	6.69E-04	0.114	2
chr2:70370175-70370560	LINC01816(dist=17809) C2orf42(dist=6760)	-	intergenic	385	3	7.68E-04	0.114	0
chr11:66085170-66085700	CD248	-	upstream	530	3	7.81E-04	0.114	0
chr16:28874330-28875755	SH2B1	Neurotrophin signaling	upstream	1425	3	7.94E-04	0.114	0
chr12:110905620-110907360	GPN3, FAM216A	-	intronic	1740	3	9.42E-04	0.120	0

In bold: the top significant CRE and the validated CREs with somatic mutations found in WGS data set. Chromosomal locations and closest genes (as upstream, downstream) are also provided.

By using our in-house generated HiC data on SK-N-BE cell line, I mapped noncoding regulatory mutations of the most significant aCREs to their candidate target genes. Then, I calculated the intensity and the significance of genomic interactions within a window of 1 Mb around the aCREs (see **Methods**).

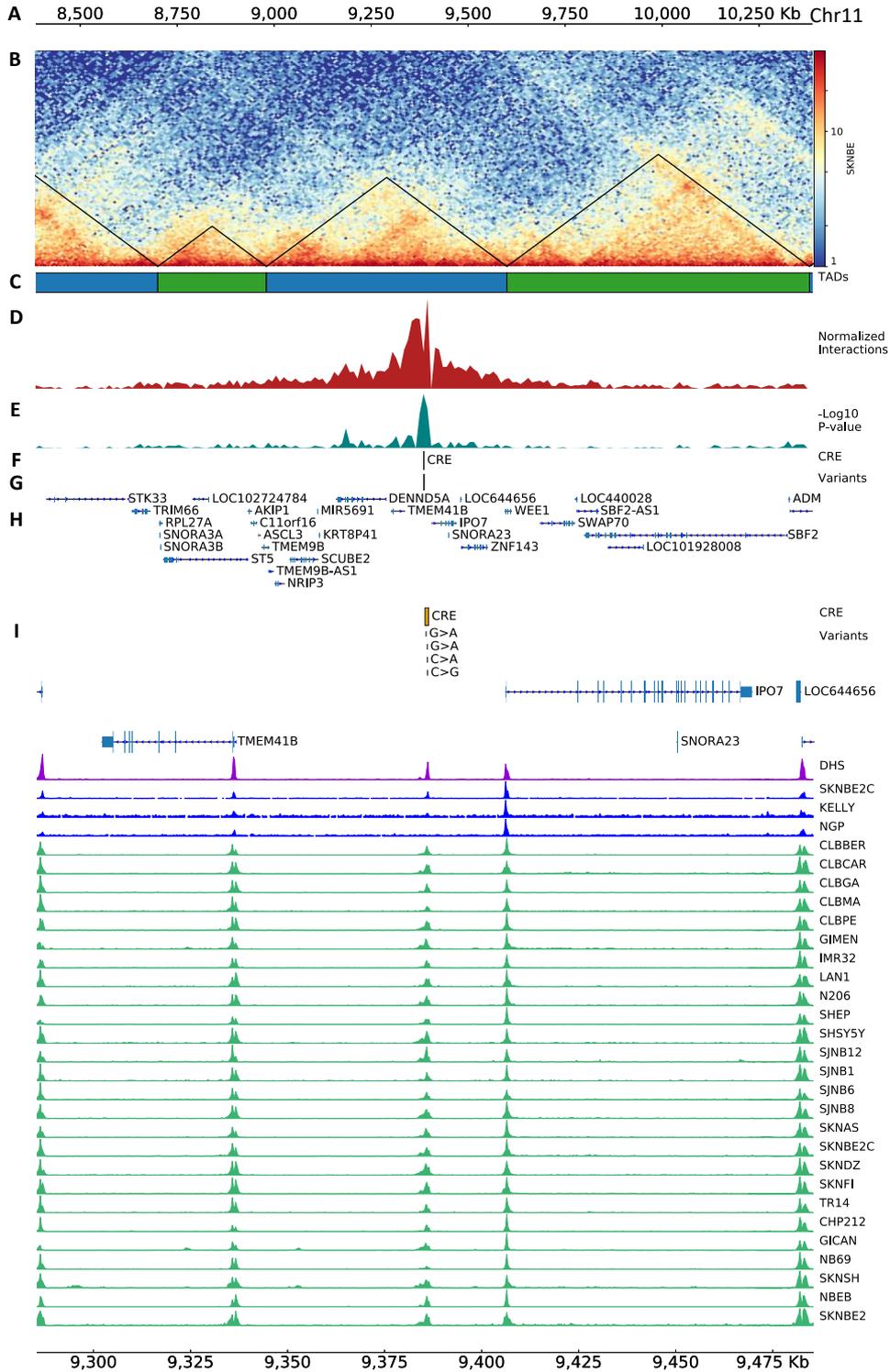


Figure 9 | Genomic interactions for the IPO7-aCRE.

The figure reports tracks named from top to bottom. **(A)** Genomic coordinates (hg19). **(B)** The interaction matrix for the chr11:9385430-9386340 region extended of 1Mb up- and down-stream. Black bordered triangles represent

TADs. (C) TADs boundaries. (D) Normalized number of interactions. (E) Minus Log10 transformed P value. (F) The aCRE region of 1960 bp. (G) The cluster of somatic mutations falling in the aCRE and its extended ends. (H) RefSeq genes (I plotted the longest isoform for each gene). (I) Zoom-in showing 100 Kbs up- and down-stream the aCRE. Here, in the listed order, I show the aCRE, the somatic mutations, the *IPO7* gene, the DHS sites of the SKNSH neuroblastoma cell line in purple (ENCODE v3), the *MYCN* ChIP-Seq data of neuroblastoma cell lines in GSE80151 (in blue), and the H3K27ac data of neuroblastoma cell lines (GSE90683 and GSE65664).

The most significantly mutated aCRE (Table 2 and Figure 9) mapped at chr11:9385430-9386340 (910 bp, carrying four mutations) between *TMEM41B* (distance = 49,770 bp) and *IPO7* (distance = 20,084 bp). I named this aCRE as “*IPO7*-aCRE”. I found significant interactions with *TMEM41B* and *IPO7* (FDR=1.75x10⁻¹³). As reported in literature, *IPO7* is regulated positively by c-Myc and negatively by p53 [54]. Furthermore, I found that *IPO7* is a target of *MYCN* (blue tracks of ChIP-Seq data in Figure 9) and its high expression correlates with *MYCN* amplification (P=1.60x10⁻²⁶) as shown in Figure 10. Of note, one significantly (*DENND5A*) and two not significantly interacting genes (*DENND2B* and *DENND7B*), all located in the same TAD of the *IPO7*-aCRE, belong to DENN domain protein family.

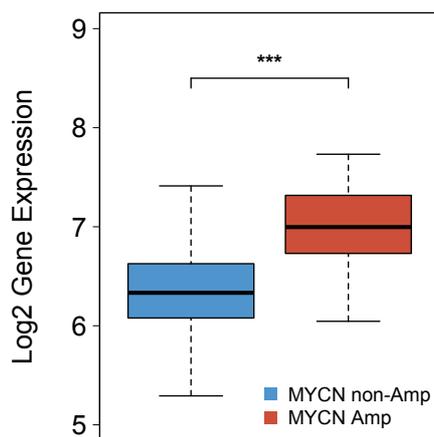


Figure 10 | *IPO7* expression correlates with *MYCN* amplification status. The boxplot shows the expression of *IPO7* stratified by *MYCN* amplification status in neuroblastoma samples of GSE62564 data set. *MYCN* Amp: n=406. *MYCN* non-Amp: n=97. P=1.6x10⁻²⁶; Mann-Whitney Test.

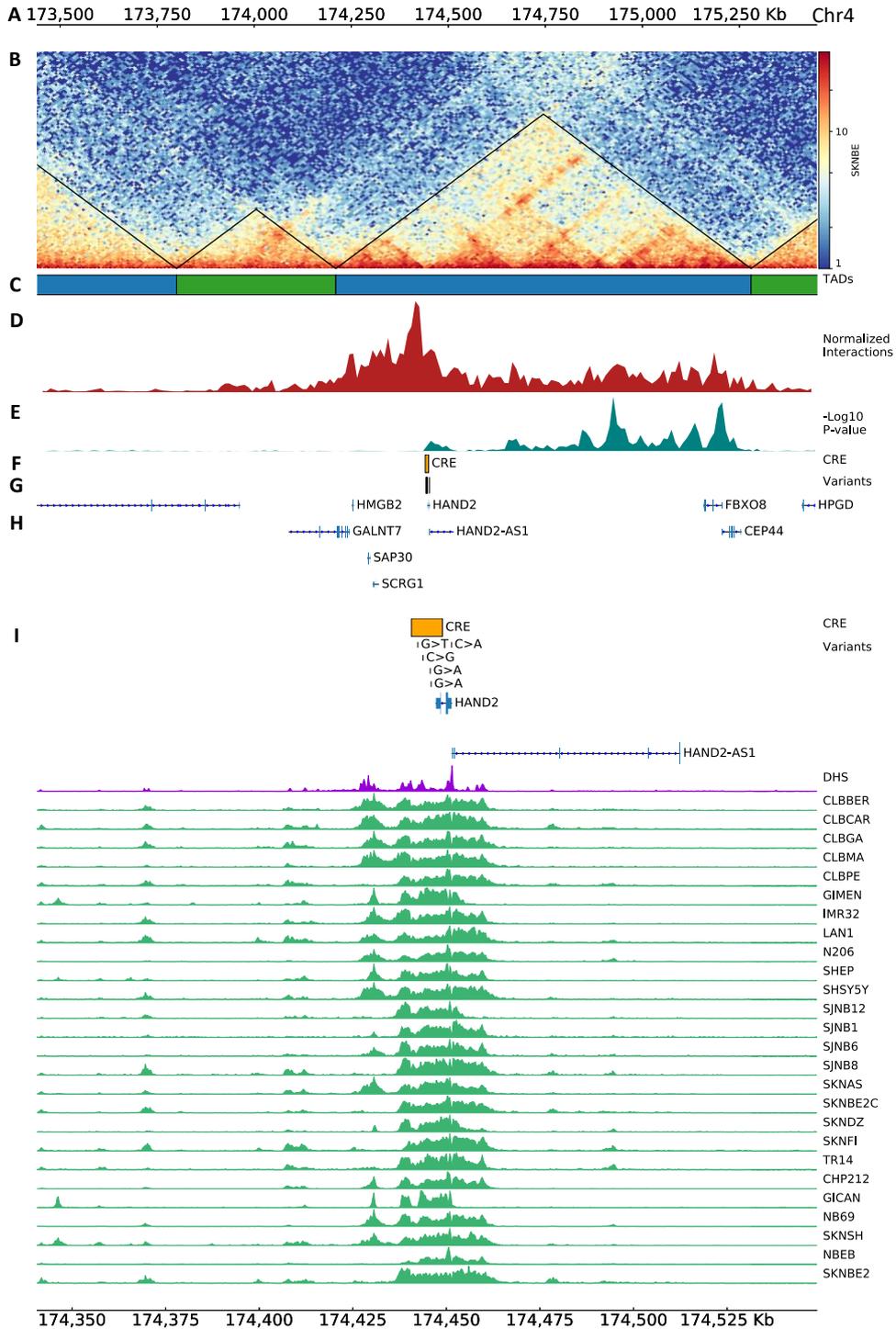


Figure 11 | Genomic interactions for the HAND2-aCRE.

The figure reports tracks named from top to bottom. (A) Genomic coordinates (hg19). (B) The interaction matrix for the chr4:174440675-174448940 region extended of 1Mb up- and down-stream. Black bordered triangles represent TADs. (C) TADs boundaries. (D) Normalized number of interactions. (E)

Log10 transformed P value. (F) The aCRE region of 8,265 bp. (G) The cluster of somatic mutations falling in the aCRE and its extended ends. (H) RefSeq genes (I plotted the longest isoform for each gene). (I) Zoom-in showing 100 Kbs up- and down-stream the aCRE. Here, in the listed order, I show the aCRE, the somatic mutations, the *HAND2* gene and its antisense, the DHS sites of the SKNSH neuroblastoma cell line in purple (ENCODE v3) and the H3K27ac data of neuroblastoma cell lines (GSE90683 and GSE65664).

I found that the second aCRE (**Table 2** and **Figure 11**), harboring 5 mutations, at chr4:174440675-174448940 (an intergenic region of 8,265 bp) significantly interacted with *HAND2* (FDR=3.09x10⁻⁰⁷) and its antisense RNA *HAND2-ASI* (FDR=1.17x10⁻⁰⁴). I named this aCRE as “HAND2-aCRE”. I also found strong and significant interactions with two ncRNAs: *LINC02269* (FDR=2.16x10⁻¹⁴) and *LINC02268* (FDR=1.15x10⁻¹³) and with two protein coding genes: *FBXO8* (FDR=1.15x10⁻¹³) and *CEP44* (FDR=1.15x10⁻¹³). Of note, I observed that all of the significant interactions were on the right of the aCRE and involved the majority of the *HAND2* topologically associating domain (TAD). Notably, the 5 mutations identified, are located in a super-enhancer region that regulates *HAND2* gene [39].

The third aCRE (**Table 2** and **Figure 12**), harboring 5 mutations, at chr19:925145-927105 (an *ARID3A* intronic region of 1,960 bp) mapped in a genomic region near the chromosome 19 telomere characterized by high gene density. This aCRE, that I named as “ARID3A-aCRE”, significantly interacted with *ARID3A* (FDR=4.68x10⁻⁰⁹), a direct TP53 effector, and with other 24 genes (mean FDR=4.38x10⁻⁰³) mainly involved in the immune response as determined by GO functional enrichment analysis (**Figure 13**).

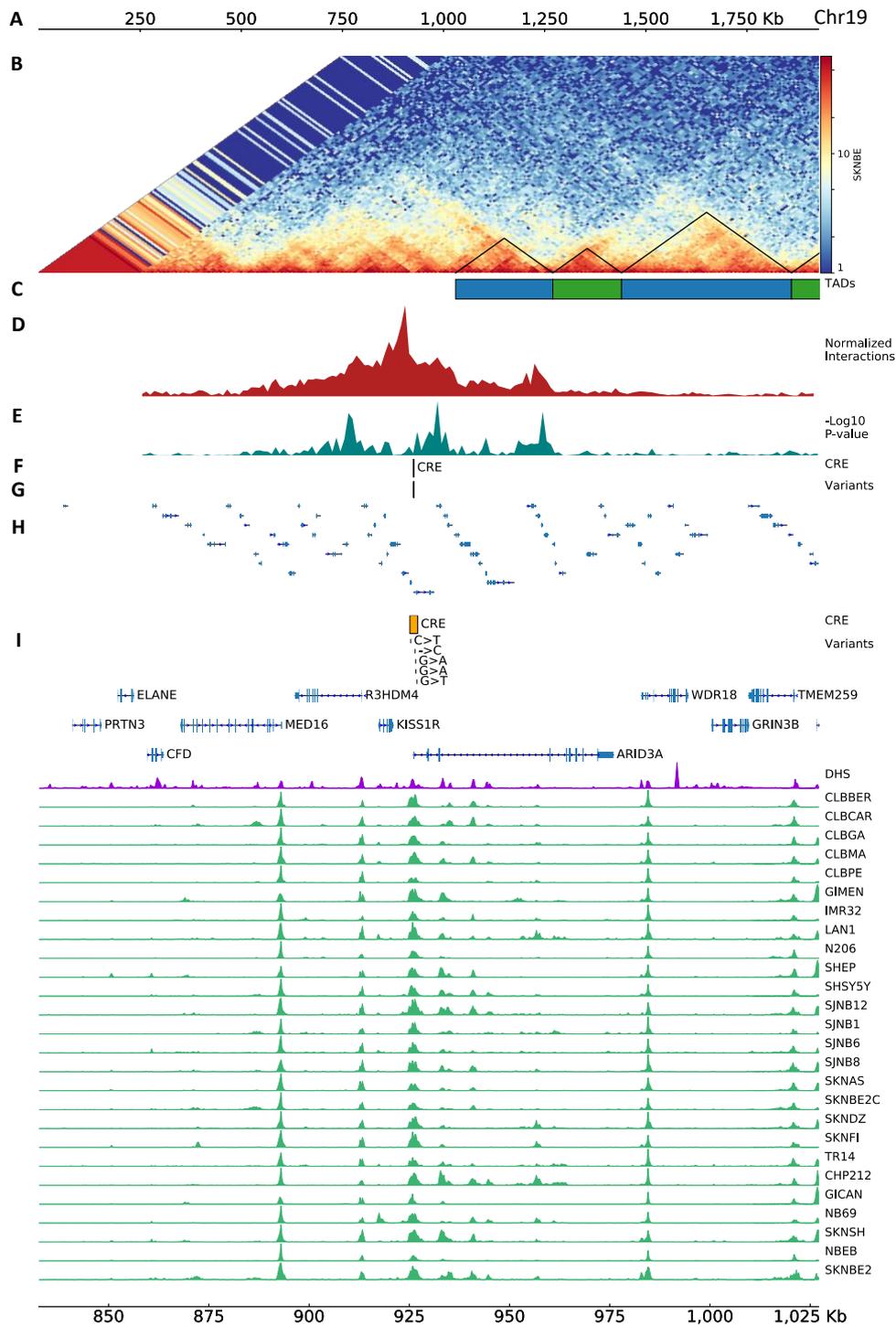


Figure 12 | Genomic interactions for the ARID3A-aCRE.

The figure reports tracks named from top to bottom. (A) Genomic coordinates (hg19). (B) The interaction matrix for the chr19:925145-927105 region extended of 1Mb up- and down-stream. Black bordered triangles represent TADs. (C) TADs boundaries. (D) Normalized number of interactions. (E)

Log10 transformed P value. (F) The aCRE region of 1960 bp. (G) The cluster of somatic mutations falling in the aCRE and its extended ends. (H) RefSeq genes (I plotted the longest isoform for each gene). (I) Zoom-in showing 100 Kbs up- and down-stream the aCRE. Here, in the listed order, I show the aCRE, the somatic mutations, the ARID3A gene, the DHS sites of the SKNSH neuroblastoma cell line in purple (ENCODE v3) and the H3K27ac data of neuroblastoma cell lines (GSE90683 and GSE65664).

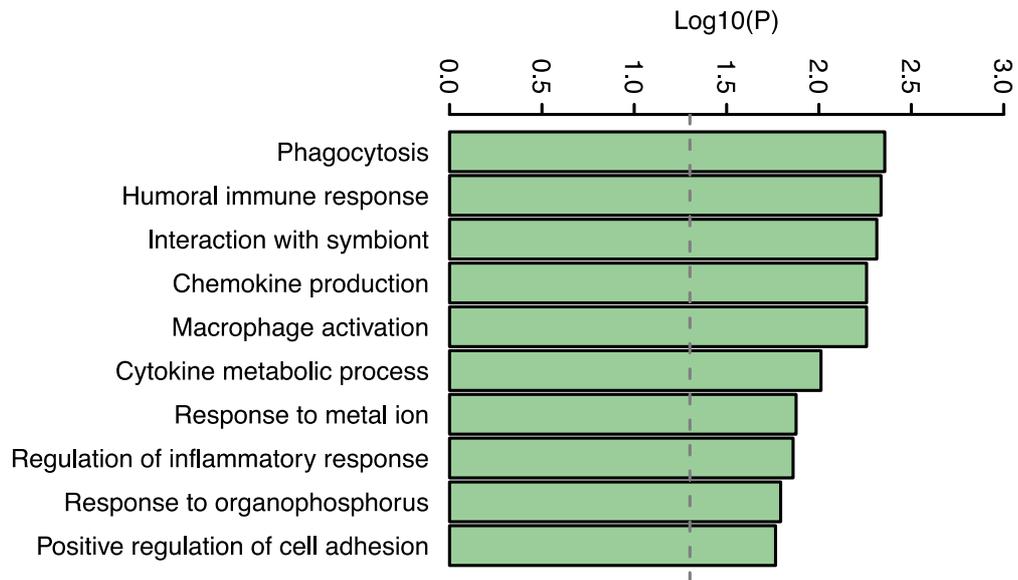


Figure 13 | Enriched Biological Processes of genes interacting with the ARID3A-aCRE.

The Gene Ontology enrichment was performed with WebGestalt tool (www.webgestalt.org) using the non-redundant set of Biological Processes. The bar plot shows the $-\text{Log}_{10}$ of the enrichment P value. The grey dashed line represents the cutoff for statistical significance (set at 0.05).

4.4 Genes interacting with mutated aCREs are involved in neuroblastoma tumorigenesis

The analysis of HiC data on SK-N-BE neuroblastoma cell line for each of the significantly mutated aCREs, returned their interacting neighboring genes (see above). Here, I wanted to assess the importance of significantly mutated aCREs in regulating nearby genes expression, and if the expression of these genes could be linked with the characteristics of the disease. With this aim, for

each of the 3 selected aCREs, I performed a k-means clustering of 498 neuroblastoma samples (GSE62564), based on expression levels of significantly interacting genes.

For the genes interacting with the IPO7-aCRE (in **Figure 9**), I found that the k-means Group 1 (n=216) showed higher expression as compared to Group 0 (n=282) (**Figure 14-A**; $P=5.96 \times 10^{-15}$). As reported in **Figure 14-B**, the Group 1 showed strong correlations with markers of poor outcome (*MYCN*-amplified vs. *MYCN*-non-amplified, $P=7.79 \times 10^{-10}$; Stage 4 vs. non-Stage 4 tumors, $P=5.25 \times 10^{-14}$; high-risk vs. non-high-risk tumors, $P=6.92 \times 10^{-15}$) and worse EFS (Event-Free Survival) and OS (Overall Survival) probabilities (**Figure 14-CD**) ($P=3.0 \times 10^{-06}$ and $P=4.0 \times 10^{-09}$, respectively).

For HAND2-aCRE (see **Figure 11**), I identified two groups characterized by low (Group 0; n=479) and high (Group 1; n=19) expression levels as shown in **Figure 14-E** ($P=1.77 \times 10^{-13}$). The Group 1 strongly correlated with markers of neuroblastoma aggressiveness (**Figure 14-F**) such as the *MYCN* amplification ($P=2.26 \times 10^{-09}$), the INSS stage 4 ($P=2.64 \times 10^{-04}$) and the high-risk patients ($P=1.72 \times 10^{-05}$). Furthermore, Group 1 patients showed inferior EFS and OS as compared to the Group 0 (**Figure 14-GH**) ($P=4.0 \times 10^{-08}$ and $P=2.0 \times 10^{-16}$, respectively).

For the ARID3A-aCRE, interacting with *ARID3A* other 24 genes (in **Figure 12**), the k-means clustering identified two groups characterized by low (Group 1; n=241) and high (Group 0; n=257) expression levels as shown in **Figure 14-I** ($P=6.99 \times 10^{-13}$). The Group 0 strongly correlated with markers of neuroblastoma aggressiveness (**Figure 14-J**) such as the *MYCN* amplification ($P=2.74 \times 10^{-04}$), the INSS stage 4 ($P=1.25 \times 10^{-06}$) and the high-risk patients ($P=7.51 \times 10^{-08}$). Furthermore, Group 1 patients showed inferior EFS and OS as compared to the Group 0 (**Figure 14-KL**) ($P=1.0 \times 10^{-09}$ and $P=2.0 \times 10^{-09}$, respectively).

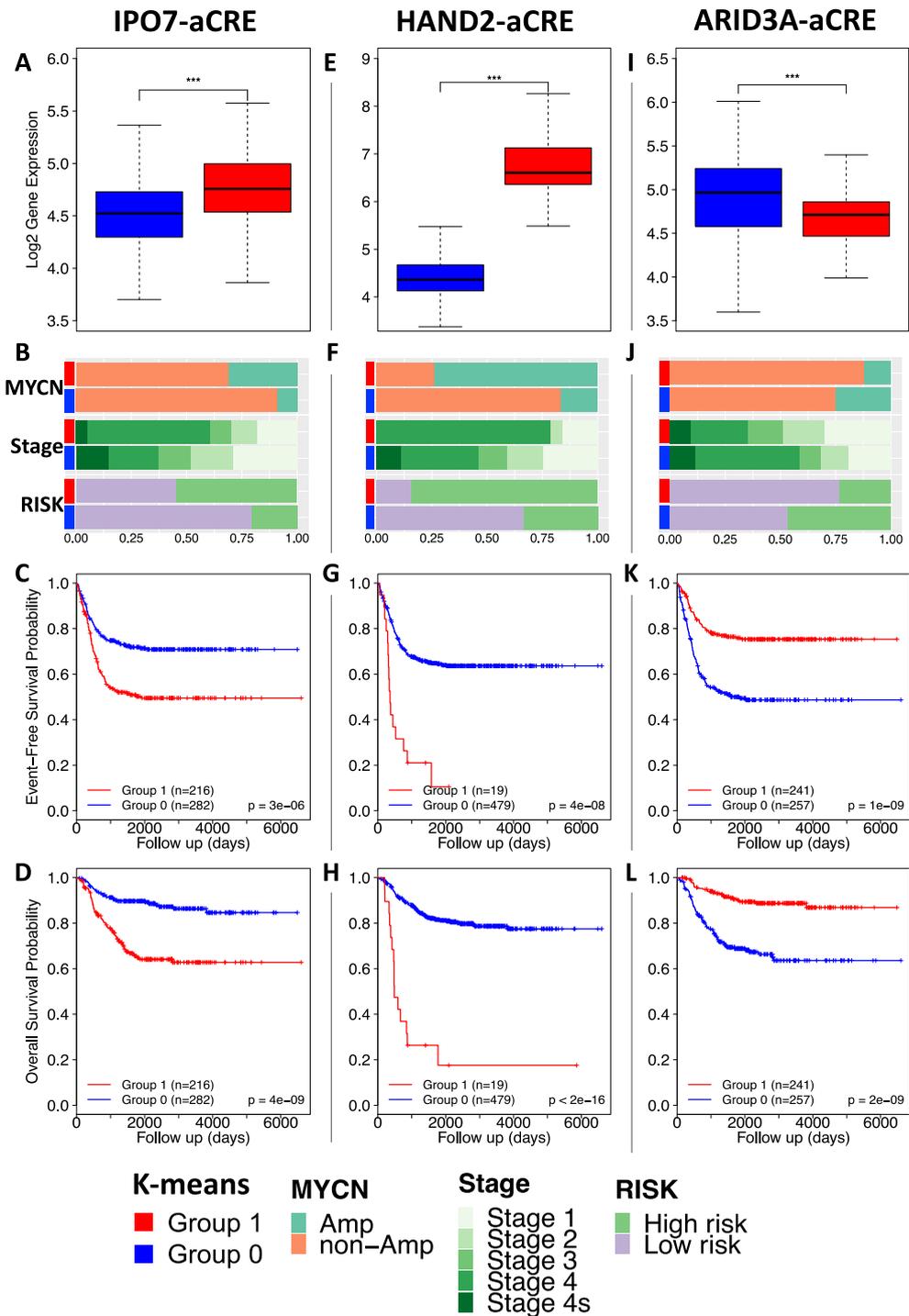


Figure 14 | Results of the k-means clustering of neuroblastoma samples (n=498, GSE62564) based on aCREs interacting genes.

The clustering was based on the expression levels of genes significantly interacting with the mutated aCREs as determined by analysis of HiC data. (A-D) IPO7-aCRE. (E-H) HAND2-aCRE. (I-L) ARID3A-aCRE. Panels A, E, I:

box plots showing the median expression levels of K-means Group 0 and Group 1. Statistical significance was calculated with Mann-Whitney test. Panels **B, F, J**: Clinical features of neuroblastoma samples in Group 0 and Group 1. From top to bottom: *MYCN* amplification status; INSS stages; Risk groups. Statistical significance was calculated with Chi-Square test. Panels **C, G, K**: Event free survival of neuroblastoma samples in Group 0 and Group 1. Statistical significance was calculated with Log rank test. Panels **D, H, L**: Overall survival probabilities of neuroblastoma samples in Group 0 and Group 1. Statistical significance was calculated with Log rank test. $P < 0.05$ ($^{\circ}$), $P < 0.01$ ($*$), $P < 0.001$ ($**$), $P < 0.0001$ ($***$).

4.5 Somatic variants in aCREs alter TFs binding motifs

I conducted a motif analysis with the R-Bioconductor package “motifbreakR” to assess if the somatic variants within the selected aCREs could disrupt or create TF binding motifs. The results showed that the selected variants ($n=14$) altered the binding motif of 118 TFs: 41, 27, and 50 motifs for IPO7-aCRE, HAND2-aCRE and ARID3A-aCRE, respectively.

4.6 Convergence of biological processes of the genes affected by somatic regulatory variants

I assessed the biological processes in which the TFs with altered motifs were involved in, by Gene Ontology (GO) enrichment analysis (**Figure 15**). I observed that, for each mutated aCRE, the closest gene exerts the biological functions similar to those that I found to be enriched among the TFs whose motifs are altered by SNVs (**Figure 15**). In particular, regarding the IPO7-aCRE, the GO terms enriched in the list of TFs “protein localization to nucleus” and “embryonic organ development” coincided with the functions of *IPO7*. Indeed, it mediates the nuclear import of proteins [55], and it is involved in the late stage of neural ectoderm differentiation [56]. Regarding HAND2-aCRE, among the list of TFs there was an over-representation of functions related to the development of diverse tissues that reflects the *HAND2* functions. *HAND2* is a transcription factor involved in the development of heart [57], limb [58], and neural crest derivatives [59]. Finally, the biological functions of the TFs targeting ARID3A-aCRE, were related to p53 signaling

pathway, embryonic development and differentiation of immune system cells. Accordingly, ARID3A is a DNA binding protein that is essential during early embryonic developmental stages [60,61]. Furthermore, it is also required in late stages of embryogenesis for normal erythroid lineage differentiation and hematopoietic stem cell production [62,63]. Moreover, *ARID3A* is a transcriptional target of p53 and cooperates with it to activate p21(WAF1) transcription [64].



Figure 15 | Gene Ontology of TFs with altered binding motifs.

Dot plot reporting the Gene Ontology enrichment analysis of the Biological Processes involving the TFs harboring somatic mutations in their binding motifs. Dot color scale is dependent on the Enrichment Ratio whereas dot sizes are proportional to the Log_{10} FDR. Statistical significance ($\text{FDR} \leq 0.05$) was calculated with Hypergeometric Test and Benjamini-Hotchberg correction. The lower part of the figure reports the biological functions of the aCRE target genes. Coloured arrows highlight the convergence of biological processes between the aCRE target genes and TFs, with mutated binding motifs, regulating the same aCRE.

Of note, as represented in **Figure 15**, the GO term “myeloid cell differentiation”, a biological process that is commonly altered in cancer [65], was shared among the three lists of TFs.

The comparison between the biological functions over-represented in the three lists of TFs with those of all target genes of mutated aCREs showed that most of the biological terms were related to processes of immune response and embryonic development (**Table 3**) suggesting that the combined effect of noncoding cancer driver mutations alters gene sets involved in specific molecular mechanisms underlying the tumorigenesis of neuroblastoma.

Table 3 | Biological functions of genes affected by regulatory somatic mutations.

	*Transcription factors with motifs altered by SNVs		°Target genes of mutated aCREs	
	Embryonic development	Immune response	Embryonic development	Immune response
IPO7-aCRE	embryonic organ development, endocrine system development, animal organ formation	type 2 immune response, myeloid cell differentiation	IPO7 [56] DENND5A [66]	DENND5A, ^DENND2B, ^DENND7B [67]
HAND2-aCRE	mesenchyme development, pericardium development, cell fate commitment, neural tube development, respiratory tube development, in utero embryonic development	myeloid cell differentiation, leukocyte differentiation	HAND2 [57-59]	-
ARID3A-aCRE	endocrine system development, animal organ formation, in utero embryonic development, muscle tissue development, coronary vasculature development	myeloid cell differentiation, interaction with symbiont, response to antibiotic	ARID3A [60-62], STK11 [68]	ARID3A [62,63] Functional enrichment in 24 genes (Additional file 2: Figure S4 and Additional file 3: Table S6)

* The two columns report the enriched biological functions in the TF sets.

° The two columns report the name of target genes involved in embryonic development and immune response.

^ The DENND2B and DENND7B genes did not show significant interactions with the aCREs but localized in the same TAD and all belong to the DENN domain protein family.

5. Discussion

To verify the role of noncoding variants in driving tumorigenesis, I performed a deep sequencing of a set of transcriptionally active CREs in 56 neuroblastomas and normal sample pairs. The bioinformatic analyses and stringent filtering steps, led me to obtain a high reliable and potentially pathogenic set of somatic SNVs falling in aCREs. I observed that most of these mutations were mainly located in noncoding regions (UTRs and Introns) in proximity of genes. These data suggest that noncoding somatic variants can affect aCREs that regulate the transcription of neighboring genes.

The mutational enrichment analysis showed that nine aCREs had a significant high rate of mutations. For two of these, I found additional somatic mutations in a set of 151 neuroblastoma WGS. Based on these data, I decided to perform additional investigation on the IPO7-aCRE showing the adjusted P-value less than 0.05 and on HAND2-aCRE and ARID3A-aCRE also found to be mutated in the independent set of 151 tumors profiled by WGS.

The most significantly mutated aCRE, IPO7-aCRE, was of 910 bp on chromosome 11. It showed significant interactions with *IPO7*, *TMEM41B* and *DENND5A*. *IPO7* is involved in nuclear import and export and it is upregulated in diverse tumors [39,40]. It is regulated positively by c-Myc and negatively by p53, and, under control of these two genes it mediates nuclear import of ribosomal proteins and export of ribosomal subunits, both required for ribosome biogenesis. *IPO7* is a target of MYCN and its high expression correlates with *MYCN* amplification. One significantly interacting gene (*DENND5A*) and two (*DENND2B* and *DENND7B*) not significantly interacting genes but in the same TAD of the aCRE can regulate Rab GTPase pathways [69] that when impaired can promote diseases including immunodeficiencies, neurological disorders, and cancer [66]. Interestingly, upregulation of Trk

receptors (TrkA and TrkB), which play a critical role in development of neuroblastoma [70], due to *DENND5A* deficiency can lead to striking alterations in neuronal development [67]. *TMEM41B* (Transmembrane Protein 41B) is required for autophagosome formation [71] and have been found mutated in pulmonary carcinoid tumors [72].

The second aCRE (HAND2-aCRE), was in an intergenic region of 8,265 bp on chromosome 4. It showed significant interactions with *HAND2*, its antisense long noncoding RNA (lncRNA) *HAND2-AS1*, *FBXO8* and *CEP44*. *HAND2*, Heart and Neural Crest Derivatives Expressed 2, is highly expressed in neural crest-derived cells and encodes a transcription factor of particular importance during neuronal development [73,74]. It also has key roles in neuroblastoma, taking part in defining the core regulatory circuitries of the sympathetic noradrenergic cell identity [39]. *HAND2-AS1* is found downregulated in numerous cancer types. It acts as a tumor suppressor by inhibiting cancer cells proliferation, migration and invasiveness [75-77].

The third, ARID3A-aCRE, was an *ARID3A* intronic region of 1,960 bp on chromosome 19. The chromosomal region is characterized by high gene density. This aCRE significantly interacted with *ARID3A* and other 24 genes that I found to be mainly involved in the immune response. ARID3A (AT-Rich Interaction Domain 3A) is a direct TP53 effector and is required for trophoderm, hematopoietic and B cell lineage development [60,61,63]. Among the genes that showed most strong interactions with the ARID3A-aCRE, I found the enzyme *STK11*, a tumor suppressor required by for normal embryonic development [68].

Based on the above reported results, I could comment that the identified genes regulated by the mutated aCREs show key roles in tumor development.

Additional experimental studies could highlight the functional roles of these genes in neuroblastoma.

Having identified the significantly mutated aCREs and their interactions with surrounding genes, I wanted to assess the importance of these CREs in regulating surrounding genes expression, and if those genes could have a role in biology of neuroblastoma genesis. To this scope, I generated a gene expression signature based on the interacting genes of each aCRE by using a well-annotated RNA-Seq data set consisting of 498 neuroblastoma samples. Each of these gene signatures identified two groups of samples characterized by low and high expression levels. For IPO7-aCRE and for HAND2-aCRE, tumors were classified by the group of genes at high expression strongly correlated with markers of neuroblastoma aggressiveness and with poor survival rates of the patients. Conversely, for ARID3A-aCRE the low expression of the interacting genes correlated with aggressive disease and poor prognosis. I interpret these results by assuming that genes regulated by mutated aCREs 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.

Afterwards, I asked if the somatic variants within the selected aCREs could disrupt or create TF binding motifs. Overall, the results showed that the selected variants altered the binding motif of 118 TFs. Next I assessed the biological processes in which TFs with altered motifs were involved in. The variants in the IPO7-aCRE modified 41 motifs for TFs controlling embryonic organ development and immune response that are biological processes also linked to the functions of the interacting genes *IPO7*, *DENND5A*, *DENND2B* and *DENND7B*. Indeed, *IPO7* is a target of *MYCN*, required for normal embryonic development, and has been reported to play an essential role in mouse embryonic stem cells differentiation into neural ectoderm cells [56].

DENND5A, *DENND2B* and *DENND7B* belong to DENN domain protein family that regulates Rab GTPase pathways [69] whose dysregulation can promote immunodeficiencies [66]. Moreover, *DENND5A* play a key role in neuronal development [67]. Furthermore, I found that the 27 binding motifs, altered in the HAND2-aCRE, were for TFs mainly involved in embryonic development that is the main biological role of *HAND2* [60,61]. In the end, the most significant biological process that characterized the 50 motifs altered in the ARID3A-aCRE was myeloid cell differentiation and many biological terms linked with embryonic development were found. Again, I observed a convergence of biological functions of the target genes of a specific mutated aCREs and of TFs targeting the same aCRE. Indeed, as mentioned above, the target genes of ARID3A-aCRE were enriched for GO terms related to the innate immune system and response to infection and the same *ARID3A* is known to regulate the early embryonic developmental stages and contribute to immune system formation [63]. Another interesting significant GO term shared between *ARID3A* and TFs was the signal transduction by p53 class mediator. Taken together, these data suggest that somatic mutations enriched in diverse aCREs can collectively impact on the functions of the TFs and the aCRE regulated genes, which are involved in the same molecular mechanisms.

Of note, myeloid cell differentiation, a biological process that is commonly altered in cancer [65], was shared between the three groups of TFs. The contribution of myeloid cells to tumor pathogenesis has been largely recognized and during the last years their crucial role in promoting tumor angiogenesis, cell invasion and metastasis has been appreciated. Mounting of evidence indicates that in cancer, myeloid cell differentiation is diverted. This hijacking, leads myeloid precursors to differentiate into potent immunosuppressive cells named myeloid-derived suppressor cells [78,79]. Indeed, in cancer, these cells negatively regulate anti-tumor immunity thus promoting tumor growth, metastasis and angiogenesis [80]. The activation of

abnormal myelopoiesis and recruitment of immature myeloid cells into cancer tissues is governed by diverse soluble factors and depends on the upregulation of STAT3 and other key TFs. Here, I found that aCREs in neuroblastoma tumors are enriched in multiple somatic SNVs altering the binding sites of TFs (including STAT3 and other STAT family members) that play a role in myeloid cell differentiation. Immunosuppressive tumor microenvironment, mediated by myeloid-derived suppressor cells, is one of the immune escaping pathways adopted by neuroblastoma cells. Targeting of myeloid-derived suppressor cells can potentiate the effect of checkpoint inhibitors for immunotherapy in human cancers, including high-risk childhood neuroblastoma [81].

6. Conclusions

In this work, I used an alternative approach to detect and study regulatory cancer driver mutations. My strategy led me to identify mutated regulatory regions that could have relevant implications in tumor development and immune escape. Further studies will be necessary to ascertain the functional roles in neuroblastoma of genes and regulatory regions that I identified.

To conclude, my findings provide evidence that noncoding somatic regulatory variants can collectively fuel tumor initiation and progression by the alteration of genes involved in the immune response and embryonic development.

8. List of publications

>> Russo R, Andolfo I, **Lasorsa VA**, Cantalupo S, Marra R, Frisso G, Abete P, Cassese GM, Servillo G, Esposito G, Gentile I, Piscopo C, Della Monica M, Fiorentino G, Russo G, Cerino R, Limone A, Zollo M, Iolascon A, Capasso M. The TNFRSF13C H159Y variant is associated with severe COVID-19: a retrospective study of 500 patients from Southern Italy. *Eur J Hum Genet*. Submitted.

>> **Lasorsa VA**, Cantalupo S, de Torres C, Aveic S, Tonini GP, Iolascon A, Capasso M. Neuroblastoma somatic mutations enriched in cis-regulatory elements collectively affect genes involved in embryonic development and immune system response. *Cancer Res*. Under revision.

>> Cimmino F, **Lasorsa VA**, Vertella S, Iolascon A, Capasso M. A targeted gene panel for circulating tumor DNA sequencing in neuroblastoma. *Front Oncol*. Co-first authorsip.

> Jemaà M, Sime W, Abassi Y, **Lasorsa VA**, Bonne Køhler J, Michaelis M, Cinatl J Jr, Capasso M, Massoumi R. Gene Expression Signature of Acquired Chemoresistance in Neuroblastoma Cells. *Int J Mol Sci*. 2020 Sep 16;21(18):6811. PMID: 32948088.

> Sime W, Jemaà M, Abassi Y, **Lasorsa VA**, Bonne Køhler J, Hansson K, Bexell D, Michaelis M, Cinatl J Jr, Strand D, Capasso M, Massoumi R. Discovery of *epi*-Enprioline as a Novel Drug for the Treatment of Vincristine Resistant Neuroblastoma. *Int J Mol Sci*. 2020 Sep 8;21(18):6577. PMID: 32911859.

> Russo R, Andolfo I, **Lasorsa VA**, Iolascon A, Capasso M. Genetic Analysis of the Coronavirus SARS-CoV-2 Host Protease *TMPRSS2* in Different Populations. *Front Genet*. 2020 Aug 4;11:872. PMID: 32849840.

> Orlandella FM, Mariniello RM, Mirabelli P, De Stefano AE, Iervolino PLC, **Lasorsa VA**, Capasso M, Giannatiempo R, Rongo M, Incoronato M, Messina F, Salvatore M, Soricelli A, Salvatore G. miR-622 is a novel potential biomarker of breast carcinoma and impairs motility of breast cancer cells through targeting NUA1 kinase. *Br J Cancer*. 2020 Aug;123(3):426-437. PMID: 32418991.

> **Lasorsa VA**, Cimmino F, Ognibene M, Mazzocco K, Erminio G, Morini M, Conte M, Iolascon A, Pezzolo A, Capasso M. 19p loss is significantly enriched in older age neuroblastoma patients and correlates with poor prognosis. *NPJ Genom Med*. 2020 Apr 15;5:18. PMID: 32337068.

- > Andolfo I, **Lasorsa VA**, Manna F, Rosato BE, Formicola D, Iolascon A, Capasso M. Kinome multigenic panel identified novel druggable EPHB4-V871I somatic variant in high-risk neuroblastoma. *J Cell Mol Med*. 2020 Jun;24(11):6459-6471. PMID: 32336043.
- > Avitabile M, **Lasorsa VA**, Cantalupo S, Cardinale A, Cimmino F, Montella A, Capasso D, Haupt R, Amoroso L, Garaventa A, Quattrone A, Corrias MV, Iolascon A, Capasso M. Association of PARP1 polymorphisms with response to chemotherapy in patients with high-risk neuroblastoma. *J Cell Mol Med*. 2020 Apr;24(7):4072-4081. PMID: 32103589.
- > Cimmino F, Avitabile M, **Lasorsa VA**, Pezone L, Cardinale A, Montella A, Cantalupo S, Iolascon A, Capasso M. Functional characterization of full-length BARD1 strengthens its role as a tumor suppressor in neuroblastoma. *J Cancer*. 2020 Jan 14;11(6):1495-1504. PMID: 32047556.
- > Capasso M, **Lasorsa VA**, Cimmino F, Avitabile M, Cantalupo S, Montella A, De Angelis B, Morini M, de Torres C, Castellano A, Locatelli F, Iolascon A. Transcription Factors Involved in Tumorigenesis Are Over-Represented in Mutated Active DNA-Binding Sites in Neuroblastoma. *Cancer Res*. 2020 Feb 1;80(3):382-393. PMID: 31784426. Co-first authorsip.
- > Amoroso L, Ognibene M, Morini M, Conte M, Di Cataldo A, Tondo A, D'Angelo P, Castellano A, Garaventa A, **Lasorsa VA**, Podestà M, Capasso M, Pezzolo A. Genomic coamplification of CDK4/MDM2/FRS2 is associated with very poor prognosis and atypical clinical features in neuroblastoma patients. *Genes Chromosomes Cancer*. 2020 May;59(5):277-285. PMID: 31756773.
- > Avitabile M, Succio M, Testori A, Cardinale A, Vaksman Z, **Lasorsa VA**, Cantalupo S, Esposito M, Cimmino F, Montella A, Formicola D, Koster J, Andreotti V, Ghiorzo P, Romano MF, Staibano S, Scalvenzi M, Ayala F, Hakonarson H, Corrias MV, Devoto M, Law MH, Iles MM, Brown K, Diskin S, Zambrano N, Iolascon A, Capasso M. Neural crest-derived tumor neuroblastoma and melanoma share 1p13.2 as susceptibility locus that shows a long-range interaction with the SLC16A1 gene. *Carcinogenesis*. 2020 May 14;41(3):284-295. PMID: 31605138.
- > Testori A, **Lasorsa VA**, Cimmino F, Cantalupo S, Cardinale A, Avitabile M, Limongelli G, Russo MG, Diskin S, Maris J, Devoto M, Keavney B, Cordell HJ, Iolascon A, Capasso M. Exploring Shared Susceptibility between Two Neural Crest Cells Originating Conditions: Neuroblastoma and Congenital Heart Disease. *Genes (Basel)*. 2019 Aug 30;10(9):663. PMID: 31480262.

- > Cimmino F, Avitabile M, **Lasorsa VA**, Montella A, Pezone L, Cantalupo S, Visconte F, Corrias MV, Iolascon A, Capasso M. HIF-1 transcription activity: HIF1A driven response in normoxia and in hypoxia. *BMC Med Genet*. 2019 Feb 26;20(1):37. PMID: 30808328.
- > Esposito MR, Binatti A, Pantile M, Coppe A, Mazzocco K, Longo L, Capasso M, **Lasorsa VA**, Luksch R, Bortoluzzi S, Tonini GP. 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 Nov 15;143(10):2525-2536 PMID: 29992558.
- > Sime W, Niu Q, Abassi Y, Masoumi KC, Zarrizi R, K hler JB, Kjellstr m S, **Lasorsa VA**, Capasso M, Fu H, Massoumi R. BAP1 induces cell death via interaction with 14-3-3 in neuroblastoma. *Cell Death Dis*. 2018 May 1;9(5):458. PMID: 29686263.
- > Romano S, Simeone E, D'Angelillo A, D'Arrigo P, Russo M, Capasso M, **Lasorsa VA**, Zambrano N, Ascierio PA, Romano MF. FKBP51s signature in peripheral blood mononuclear cells of melanoma patients as a possible predictive factor for immunotherapy. *Cancer Immunol Immunother*. 2017 Sep;66(9):1143-1151. PMID: 28434031.
- > Calabrese FM, Clima R, Pignataro P, **Lasorsa VA**, Hogarty MD, Castellano A, Conte M, Tonini GP, Iolascon A, Gasparre G, Capasso M. A comprehensive characterization of rare mitochondrial DNA variants in neuroblastoma. *Oncotarget*. 2016 Aug 2;7(31):49246-49258. PMID: 27351283.
- > Formicola D, Petrosino G, **Lasorsa VA**, Pignataro P, Cimmino F, Vetrella S, Longo L, Tonini GP, Oberthuer A, Iolascon A, Fischer M, Capasso M. An 18 gene expression-based score classifier predicts the clinical outcome in stage 4 neuroblastoma. *J Transl Med*. 2016 May 17;14(1):142. PMID: 27188717.
- > **Lasorsa VA**, Formicola D, Pignataro P, Cimmino F, Calabrese FM, Mora J, Esposito MR, Pantile M, Zanon C, De Mariano M, Longo L, Hogarty MD, de Torres C, Tonini GP, Iolascon A, Capasso M. Exome and deep sequencing of clinically aggressive neuroblastoma reveal somatic mutations that affect key pathways involved in cancer progression. *Oncotarget*. 2016 Apr 19;7(16):21840-52. PMID: 27009842.

9. References

1. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57-74.
2. Bradner, J.E.; Hnisz, D.; Young, R.A., Transcriptional addiction in cancer. *Cell* **2017**, *168*, 629-643.
3. Melton, C.; Reuter, J.A.; Spacek, D.V.; Snyder, M., Recurrent somatic mutations in regulatory regions of human cancer genomes. *Nature genetics* **2015**, *47*, 710-716.
4. Zhang, W.; Bojorquez-Gomez, A.; Velez, D.O.; Xu, G.; Sanchez, K.S.; Shen, J.P.; Chen, K.; Licon, K.; Melton, C.; Olson, K.M., *et al.*, A global transcriptional network connecting noncoding mutations to changes in tumor gene expression. *Nature genetics* **2018**, *50*, 613-620.
5. Huang, F.W.; Hodis, E.; Xu, M.J.; Kryukov, G.V.; Chin, L.; Garraway, L.A., Highly recurrent tert promoter mutations in human melanoma. *Science* **2013**, *339*, 957-959.
6. Rheinbay, E.; Parasuraman, P.; Grimsby, J.; Tiao, G.; Engreitz, J.M.; Kim, J.; Lawrence, M.S.; Taylor-Weiner, A.; Rodriguez-Cuevas, S.; Rosenberg, M., *et al.*, Recurrent and functional regulatory mutations in breast cancer. *Nature* **2017**, *547*, 55-60.
7. Weinhold, N.; Jacobsen, A.; Schultz, N.; Sander, C.; Lee, W., Genome-wide analysis of noncoding regulatory mutations in cancer. *Nature genetics* **2014**, *46*, 1160-1165.
8. Orlando, G.; Law, P.J.; Cornish, A.J.; Dobbins, S.E.; Chubb, D.; Broderick, P.; Litchfield, K.; Hariri, F.; Pastinen, T.; Osborne, C.S., *et al.*, Promoter capture hi-c-based identification of recurrent noncoding mutations in colorectal cancer. *Nature genetics* **2018**, *50*, 1375-1380.
9. 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.*, Transcription factors involved in tumorigenesis are over-represented in mutated active DNA-binding sites in neuroblastoma. *Cancer research* **2020**, *80*, 382-393.
10. Hansen, R.S.; Thomas, S.; Sandstrom, R.; Canfield, T.K.; Thurman, R.E.; Weaver, M.; Dorschner, M.O.; Gartler, S.M.; Stamatoyannopoulos, J.A., Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 139-144.
11. Schuster-Bockler, B.; Lehner, B., Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **2012**, *488*, 504-507.
12. Salk, J.J.; Schmitt, M.W.; Loeb, L.A., Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nature reviews. Genetics* **2018**, *19*, 269-285.

13. Stiller, C.A.; Parkin, D.M., International variations in the incidence of neuroblastoma. *International journal of cancer* **1992**, *52*, 538-543.
14. London, W.B.; Castleberry, R.P.; Matthay, K.K.; Look, A.T.; Seeger, R.C.; Shimada, H.; Thorner, P.; Brodeur, G.; Maris, J.M.; Reynolds, C.P., *et al.*, Evidence for an age cutoff greater than 365 days for neuroblastoma risk group stratification in the children's oncology group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **2005**, *23*, 6459-6465.
15. Anderson, D.J.; Carnahan, J.F.; Michelsohn, A.; Patterson, P.H., Antibody markers identify a common progenitor to sympathetic neurons and chromaffin cells in vivo and reveal the timing of commitment to neuronal differentiation in the sympathoadrenal lineage. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **1991**, *11*, 3507-3519.
16. Matthay, K.K.; Maris, J.M.; Schleiermacher, G.; Nakagawara, A.; Mackall, C.L.; Diller, L.; Weiss, W.A., Neuroblastoma. *Nature reviews. Disease primers* **2016**, *2*, 16078.
17. Brodeur, G.M.; Pritchard, J.; Berthold, F.; Carlsen, N.L.; Castel, V.; Castleberry, R.P.; De Bernardi, B.; Evans, A.E.; Favrot, M.; Hedborg, F., *et al.*, Revisions of the international criteria for neuroblastoma diagnosis, staging, and response to treatment. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **1993**, *11*, 1466-1477.
18. Cecchetto, G.; Mosseri, V.; De Bernardi, B.; Helardot, P.; Monclair, T.; Costa, E.; Horcher, E.; Neuenschwander, S.; Tomà, P.; Rizzo, A., *et al.*, Surgical risk factors in primary surgery for localized neuroblastoma: The Inesgl study of the european international society of pediatric oncology neuroblastoma group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **2005**, *23*, 8483-8489.
19. Cohn, S.L.; Pearson, A.D.; London, W.B.; Monclair, T.; Ambros, P.F.; Brodeur, G.M.; Faldum, A.; Hero, B.; Iehara, T.; Machin, D., *et al.*, The international neuroblastoma risk group (inrg) classification system: An inrg task force report. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **2009**, *27*, 289-297.
20. Luksch, R.; Castellani, M.R.; Collini, P.; De Bernardi, B.; Conte, M.; Gambini, C.; Gandola, L.; Garaventa, A.; BIASONI, D.; Podda, M., *et al.*, Neuroblastoma (peripheral neuroblastic tumours). *Critical reviews in oncology/hematology* **2016**, *107*, 163-181.
21. Mossé, Y.P.; Laudenslager, M.; Longo, L.; Cole, K.A.; Wood, A.; Attiyeh, E.F.; Laquaglia, M.J.; Sennett, R.; Lynch, J.E.; Perri, P., *et al.*, Identification of alk as a major familial neuroblastoma predisposition gene. *Nature* **2008**, *455*, 930-935.

22. Mosse, Y.P.; Laudenslager, M.; Khazi, D.; Carlisle, A.J.; Winter, C.L.; Rappaport, E.; Maris, J.M., Germline *phox2b* mutation in hereditary neuroblastoma. *American journal of human genetics* **2004**, *75*, 727-730.
23. Bosse, K.R.; Diskin, S.J.; Cole, K.A.; Wood, A.C.; Schnepf, R.W.; Norris, G.; Nguyen le, B.; Jagannathan, J.; Laquaglia, M.; Winter, C., *et al.*, Common variation at *bard1* results in the expression of an oncogenic isoform that influences neuroblastoma susceptibility and oncogenicity. *Cancer research* **2012**, *72*, 2068-2078.
24. Khurana, E.; Fu, Y.; Colonna, V.; Mu, X.J.; Kang, H.M.; Lappalainen, T.; Sboner, A.; Lochovsky, L.; Chen, J.; Harmanci, A., *et al.*, Integrative annotation of variants from 1092 humans: Application to cancer genomics. *Science* **2013**, *342*, 1235587.
25. Tonini, G.P.; Capasso, M., Genetic predisposition and chromosome instability in neuroblastoma. *Cancer metastasis reviews* **2020**, *39*, 275-285.
26. Wang, K.; Diskin, S.J.; Zhang, H.; Attiyeh, E.F.; Winter, C.; Hou, C.; Schnepf, R.W.; Diamond, M.; Bosse, K.; Mayes, P.A., *et al.*, Integrative genomics identifies *lmo1* as a neuroblastoma oncogene. *Nature* **2011**, *469*, 216-220.
27. Cimmino, F.; Avitabile, M.; Diskin, S.J.; Vaksman, Z.; Pignataro, P.; Formicola, D.; Cardinale, A.; Testori, A.; Koster, J.; de Torres, C., *et al.*, Fine mapping of 2q35 high-risk neuroblastoma locus reveals independent functional risk variants and suggests full-length *bard1* as tumor-suppressor. *International journal of cancer* **2018**, *143*, 2828-2837.
28. Capasso, M.; Devoto, M.; Hou, C.; Asgharzadeh, S.; Glessner, J.T.; Attiyeh, E.F.; Mosse, Y.P.; Kim, C.; Diskin, S.J.; Cole, K.A., *et al.*, Common variations in *bard1* influence susceptibility to high-risk neuroblastoma. *Nature genetics* **2009**, *41*, 718-723.
29. Diskin, S.J.; Capasso, M.; Schnepf, R.W.; Cole, K.A.; Attiyeh, E.F.; Hou, C.; Diamond, M.; Carpenter, E.L.; Winter, C.; Lee, H., *et al.*, Common variation at 6q16 within *hac1* and *lin28b* influences susceptibility to neuroblastoma. *Nature genetics* **2012**, *44*, 1126-1130.
30. Avitabile, M.; Succio, M.; Testori, A.; Cardinale, A.; Vaksman, Z.; Lasorsa, V.A.; Cantalupo, S.; Esposito, M.; Cimmino, F.; Montella, A., *et al.*, Neural crest-derived tumor neuroblastoma and melanoma share 1p13.2 as susceptibility locus that shows a long-range interaction with the *slc16a1* gene. *Carcinogenesis* **2020**, *41*, 284-295.
31. Capasso, M.; Diskin, S.J., Genetics and genomics of neuroblastoma. *Cancer treatment and research* **2010**, *155*, 65-84.
32. Lasorsa, V.A.; Cimmino, F.; Ognibene, M.; Mazzocco, K.; Erminio, G.; Morini, M.; Conte, M.; Iolascon, A.; Pezzolo, A.; Capasso, M., 19p loss is significantly enriched in older age neuroblastoma patients and correlates with poor prognosis. *NPJ genomic medicine* **2020**, *5*, 18.

33. Esposito, M.R.; Binatti, A.; Pantile, M.; Coppe, A.; Mazzocco, K.; Longo, L.; Capasso, M.; Lasorsa, V.A.; Luksch, R.; Bortoluzzi, S., *et al.*, Somatic mutations in specific and connected subpathways are associated with short neuroblastoma patients' survival and indicate proteins targetable at onset of disease. *International journal of cancer* **2018**, *143*, 2525-2536.
34. Lasorsa, V.A.; Formicola, D.; Pignataro, P.; Cimmino, F.; Calabrese, F.M.; Mora, J.; Esposito, M.R.; Pantile, M.; Zanon, C.; De Mariano, M., *et al.*, Exome and deep sequencing of clinically aggressive neuroblastoma reveal somatic mutations that affect key pathways involved in cancer progression. *Oncotarget* **2016**, *7*, 21840-21852.
35. 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.*, The genetic landscape of high-risk neuroblastoma. *Nature genetics* **2013**, *45*, 279-284.
36. Peifer, M.; Hertwig, F.; Roels, F.; Dreidax, D.; Gartlgruber, M.; Menon, R.; Kramer, A.; Roncaioli, J.L.; Sand, F.; Heuckmann, J.M., *et al.*, Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* **2015**, *526*, 700-704.
37. Ackermann, S.; Cartolano, M.; Hero, B.; Welte, A.; Kahlert, Y.; Roderwieser, A.; Bartenhagen, C.; Walter, E.; Gecht, J.; Kerschke, L., *et al.*, A mechanistic classification of clinical phenotypes in neuroblastoma. *Science* **2018**, *362*, 1165-1170.
38. Esposito, M.R.; Aveic, S.; Seydel, A.; Tonini, G.P., Neuroblastoma treatment in the post-genomic era. *Journal of biomedical science* **2017**, *24*, 14.
39. 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.*, Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nature genetics* **2017**, *49*, 1408-1413.
40. Quinlan, A.R.; Hall, I.M., Bedtools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841-842.
41. Li, H.; Durbin, R., Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* **2009**, *25*, 1754-1760.
42. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; Genome Project Data Processing, S., The sequence alignment/map format and samtools. *Bioinformatics* **2009**, *25*, 2078-2079.
43. Cibulskis, K.; Lawrence, M.S.; Carter, S.L.; Sivachenko, A.; Jaffe, D.; Sougnez, C.; Gabriel, S.; Meyerson, M.; Lander, E.S.; Getz, G., Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature biotechnology* **2013**, *31*, 213-219.
44. Saunders, C.T.; Wong, W.S.; Swamy, S.; Becq, J.; Murray, L.J.; Cheetham, R.K., Strelka: Accurate somatic small-variant calling from

- sequenced tumor-normal sample pairs. *Bioinformatics* **2012**, *28*, 1811-1817.
45. Wang, K.; Li, M.; Hakonarson, H., Annovar: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* **2010**, *38*, e164.
 46. Fu, Y.; Liu, Z.; Lou, S.; Bedford, J.; Mu, X.J.; Yip, K.Y.; Khurana, E.; Gerstein, M., Funseq2: A framework for prioritizing noncoding regulatory variants in cancer. *Genome biology* **2014**, *15*, 480.
 47. Langmead, B.; Salzberg, S.L., Fast gapped-read alignment with bowtie 2. *Nature methods* **2012**, *9*, 357-359.
 48. Ramirez, F.; Bhardwaj, V.; Arrigoni, L.; Lam, K.C.; Gruning, B.A.; Villaveces, J.; Habermann, B.; Akhtar, A.; Manke, T., High-resolution tads reveal DNA sequences underlying genome organization in flies. *Nature communications* **2018**, *9*, 189.
 49. Maris, J.M.; Weiss, M.J.; Mosse, Y.; Hii, G.; Guo, C.; White, P.S.; Hogarty, M.D.; Mirensky, T.; Brodeur, G.M.; Rebbeck, T.R., *et al.*, Evidence for a hereditary neuroblastoma predisposition locus at chromosome 16p12-13. *Cancer Res* **2002**, *62*, 6651-6658.
 50. Ay, F.; Bailey, T.L.; Noble, W.S., Statistical confidence estimation for hi-c data reveals regulatory chromatin contacts. *Genome research* **2014**, *24*, 999-1011.
 51. Liao, Y.; Wang, J.; Jaehnig, E.J.; Shi, Z.; Zhang, B., Webgestalt 2019: Gene set analysis toolkit with revamped uis and apis. *Nucleic acids research* **2019**, *47*, W199-W205.
 52. Rentzsch, P.; Witten, D.; Cooper, G.M.; Shendure, J.; Kircher, M., Cadd: Predicting the deleteriousness of variants throughout the human genome. *Nucleic acids research* **2019**, *47*, D886-D894.
 53. Sausen, M.; Leary, R.J.; Jones, S.; Wu, J.; Reynolds, C.P.; Liu, X.; Blackford, A.; Parmigiani, G.; Diaz, L.A., Jr.; Papadopoulos, N., *et al.*, Integrated genomic analyses identify *arid1a* and *arid1b* alterations in the childhood cancer neuroblastoma. *Nature genetics* **2013**, *45*, 12-17.
 54. Golomb, L.; Bublik, D.R.; Wilder, S.; Nevo, R.; Kiss, V.; Grabusic, K.; Volarevic, S.; Oren, M., Importin 7 and exportin 1 link c-myc and p53 to regulation of ribosomal biogenesis. *Molecular cell* **2012**, *45*, 222-232.
 55. Jäkel, S.; Görlich, D., Importin beta, transportin, ranbp5 and ranbp7 mediate nuclear import of ribosomal proteins in mammalian cells. *The EMBO journal* **1998**, *17*, 4491-4502.
 56. Sangel, P.; Oka, M.; Yoneda, Y., The role of importin- β s in the maintenance and lineage commitment of mouse embryonic stem cells. *FEBS open bio* **2014**, *4*, 112-120.
 57. Srivastava, D.; Thomas, T.; Lin, Q.; Kirby, M.L.; Brown, D.; Olson, E.N., Regulation of cardiac mesodermal and neural crest development by the bhlh transcription factor, dhand. *Nature genetics* **1997**, *16*, 154-160.

58. Charité, J.; McFadden, D.G.; Olson, E.N., The bhlh transcription factor dhand controls sonic hedgehog expression and establishment of the zone of polarizing activity during limb development. *Development (Cambridge, England)* **2000**, *127*, 2461-2470.
59. Howard, M.; Foster, D.N.; Cserjesi, P., Expression of hand gene products may be sufficient for the differentiation of avian neural crest-derived cells into catecholaminergic neurons in culture. *Developmental biology* **1999**, *215*, 62-77.
60. Rhee, C.; Edwards, M.; Dang, C.; Harris, J.; Brown, M.; Kim, J.; Tucker, H.O., Arid3a is required for mammalian placenta development. *Developmental biology* **2017**, *422*, 83-91.
61. Rhee, C.; Lee, B.K.; Beck, S.; Anjum, A.; Cook, K.R.; Popowski, M.; Tucker, H.O.; Kim, J., Arid3a is essential to execution of the first cell fate decision via direct embryonic and extraembryonic transcriptional regulation. *Genes & development* **2014**, *28*, 2219-2232.
62. Kim, P.G.; Canver, M.C.; Rhee, C.; Ross, S.J.; Harriss, J.V.; Tu, H.C.; Orkin, S.H.; Tucker, H.O.; Daley, G.Q., Interferon- α signaling promotes embryonic hsc maturation. *Blood* **2016**, *128*, 204-216.
63. Webb, C.F.; Bryant, J.; Popowski, M.; Allred, L.; Kim, D.; Harriss, J.; Schmidt, C.; Miner, C.A.; Rose, K.; Cheng, H.L., *et al.*, The arid family transcription factor bright is required for both hematopoietic stem cell and b lineage development. *Molecular and cellular biology* **2011**, *31*, 1041-1053.
64. Lestari, W.; Ichwan, S.J.; Otsu, M.; Yamada, S.; Iseki, S.; Shimizu, S.; Ikeda, M.A., Cooperation between arid3a and p53 in the transcriptional activation of p21waf1 in response to DNA damage. *Biochemical and biophysical research communications* **2012**, *417*, 710-716.
65. Wynn, T.A., Myeloid-cell differentiation redefined in cancer. *Nature immunology* **2013**, *14*, 197-199.
66. Stenmark, H., Rab gtpases as coordinators of vesicle traffic. *Nature reviews. Molecular cell biology* **2009**, *10*, 513-525.
67. Han, C.; Alkhatir, R.; Froukh, T.; Minassian, A.G.; Galati, M.; Liu, R.H.; Fotouhi, M.; Sommerfeld, J.; Alfrook, A.J.; Marshall, C., *et al.*, Epileptic encephalopathy caused by mutations in the guanine nucleotide exchange factor denn5a. *American journal of human genetics* **2016**, *99*, 1359-1367.
68. Zhao, R.X.; Xu, Z.X., Targeting the lkb1 tumor suppressor. *Current drug targets* **2014**, *15*, 32-52.
69. Marat, A.L.; Dokainish, H.; McPherson, P.S., Denn domain proteins: Regulators of rab gtpases. *The Journal of biological chemistry* **2011**, *286*, 13791-13800.
70. Brodeur, G.M.; Minturn, J.E.; Ho, R.; Simpson, A.M.; Iyer, R.; Varela, C.R.; Light, J.E.; Kolla, V.; Evans, A.E., Trk receptor expression and inhibition in neuroblastomas. *Clinical cancer research : an official*

- journal of the American Association for Cancer Research* **2009**, *15*, 3244-3250.
71. Morita, K.; Hama, Y.; Izume, T.; Tamura, N.; Ueno, T.; Yamashita, Y.; Sakamaki, Y.; Mimura, K.; Morishita, H.; Shihoya, W., *et al.*, Genome-wide crispr screen identifies tmem41b as a gene required for autophagosome formation. *The Journal of cell biology* **2018**, *217*, 3817-3828.
 72. Asiedu, M.K.; Thomas, C.F., Jr.; Dong, J.; Schulte, S.C.; Khadka, P.; Sun, Z.; Kosari, F.; Jen, J.; Molina, J.; Vasmatazis, G., *et al.*, Pathways impacted by genomic alterations in pulmonary carcinoid tumors. *Clinical cancer research : an official journal of the American Association for Cancer Research* **2018**, *24*, 1691-1704.
 73. Dai, Y.S.; Cserjesi, P., The basic helix-loop-helix factor, hand2, functions as a transcriptional activator by binding to e-boxes as a heterodimer. *The Journal of biological chemistry* **2002**, *277*, 12604-12612.
 74. Hendershot, T.J.; Liu, H.; Clouthier, D.E.; Shepherd, I.T.; Coppola, E.; Studer, M.; Firulli, A.B.; Pittman, D.L.; Howard, M.J., Conditional deletion of hand2 reveals critical functions in neurogenesis and cell type-specific gene expression for development of neural crest-derived noradrenergic sympathetic ganglion neurons. *Developmental biology* **2008**, *319*, 179-191.
 75. Gokulnath, P.; de Cristofaro, T.; Manipur, I.; Di Palma, T.; Soriano, A.A.; Guarracino, M.R.; Zannini, M., Long non-coding rna hand2-as1 acts as a tumor suppressor in high-grade serous ovarian carcinoma. *International journal of molecular sciences* **2020**, *21*.
 76. Gong, J.; Fan, H.; Deng, J.; Zhang, Q., Lncrna hand2-as1 represses cervical cancer progression by interaction with transcription factor e2f4 at the promoter of c16orf74. *Journal of cellular and molecular medicine* **2020**, *24*, 6015-6027.
 77. Wei, P.; Yang, J.; Zhang, D.; Cui, M.; Li, L., Lncrna hand2-as1 regulates prostate cancer cell growth through targeting the mir-106a-5p/rbm24 axis. *OncoTargets and therapy* **2020**, *13*, 4523-4531.
 78. Gabilovich, D.I.; Ostrand-Rosenberg, S.; Bronte, V., Coordinated regulation of myeloid cells by tumours. *Nature reviews. Immunology* **2012**, *12*, 253-268.
 79. Joyce, J.A.; Pollard, J.W., Microenvironmental regulation of metastasis. *Nature reviews. Cancer* **2009**, *9*, 239-252.
 80. Sica, A.; Bronte, V., Altered macrophage differentiation and immune dysfunction in tumor development. *The Journal of clinical investigation* **2007**, *117*, 1155-1166.
 81. Mao, Y.; Eissler, N.; Blanc, K.L.; Johnsen, J.I.; Kogner, P.; Kiessling, R., Targeting suppressive myeloid cells potentiates checkpoint inhibitors to control spontaneous neuroblastoma. *Clinical cancer*

*research : an official journal of the American Association for Cancer
Research* **2016**, 22, 3849-3859.