Università degli Studi di Napoli Federico II

Ph.D. Program in
Computational and Quantitative Biology

XXXVI Cycle

**Thesis for the Degree of Doctor of Philosophy**

# Noncoding regulatory mutations as driving event for the oncogenic core regulatory circuitries of neuroblastoma

by

**Vincenzo Aievola**

Advisor: Prof. Mario Capasso

**Scuola Politecnica e delle Scienze di Base**

**Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione**

# Noncoding regulatory mutations as driving event for the oncogenic core regulatory circuitries of neuroblastoma

Ph.D. Thesis presented
for the fulfillment of the Degree of Doctor of Philosophy
in Computational and Quantitative Biology

by

**Vincenzo Aievola**

**March 2024**

Approved as to style and content by

Prof. Mario Capasso, Advisor

Università degli Studi di Napoli Federico II
Ph.D. Program in Computational and Quantitative Biology
XXXVI cycle - Chairman: Prof. Michele Ceccarelli

https://cqb.dieti.unina.it/

## Candidate's declaration

I hereby declare that this thesis submitted to obtain the academic degree of Philosophiæ Doctor (Ph.D.) in Computational and Quantitative Biology is my own unaided work, that I have not used other than the sources indicated, and that all direct and indirect sources are acknowledged as references.

Parts of this dissertation have been published in international journals and/or conference articles (see list of the author's publications at the end of the thesis).

Napoli, December 15, 2023

Vincenzo Aievola

## Table of contents

## List of abbreviations

ADRN: Adrenergic

ATAC-seq: Assay for transposase-accessible chromatin using sequencing

BAM: Binary alignment map

BWA: Burrows-Wheeler aligner

CHiC: Promoter-capture HiC

CHiCAGO: Capture Hi-C analysis of genomic organisation

ChIP-seq: Chromatin immunoprecipitation sequencing

CRC: Core regulatory circuitry

DHS: DNase hypersensitivity sites

DMEM: Dulbecco's modified eagle medium

DP: Allele depth

DSB: DNA double-strand break

ECM: Extracellular matrix

EFS: Event free survival

EGA: European genome-phenome archive

EMEM: Minimum essential medium eagle

ENCODE: Encyclopedia of DNA elements

EVS: Empirical variant score

FABIAN: Fast binding-site analysis

FBS: Fetal bovine serum

FDR: False discovery rate

FRiP: Fraction of reads in peaks

GEO: Gene expression omnibus

GO: Gene ontology

GWAS: Genome-wide association study

H3K27ac: Histone H3 acetylated at lysine 27

H3K4me1: Histone H3 mono-methylated at lysine 4

H3K4me3: Histone H3 tri-methylated at lysine 4

HiC: all-vs-all chromosome conformation capture

I-type: Intermediate type

ICGC: International cancer genome consortium

INRG: International neuroblastoma risk group

INSS: International neuroblastoma staging system

IP: Immunoprecipitation

IRCCS: Istituti di Ricovero e Cura a Carattere Scientifico

ITH: intratumoral heterogeneity

lincRNA: Long noncoding RNA

MACS: Model-based Analysis of ChIP-Seq

MAF: Minor allele frequency

MES: Mesenchymal

miRNA: micro-RNA

mTF: master TF

N-type: Neuroblastic type

NB: Neuroblastoma

NC: Neural crest

NCBI: National center for biotechnology information

NCC: Neural crest cell

NGS: Next generation sequencing

NSC: Normalized strand cross-correlation

OCR: Open chromatin region

OS: Overall survival

PB: Peripheral blood

PCR: Polymerase chain reaction

PDX: Patient derived xenograft

PWM: Position-weight matrix

RNA-seq: messenger RNA sequencing

RPMI: Roswell Park Memorial Institute

RSC: Relative strand cross-correlation

RT: Room Temperature

S-type: substrate-adherent type

snoRNA: Small nucleolar RNA

SNP: Single nucleotide polymorphism

snRNA: Small nuclear RNA

SNS: Sympathetic nervous system

SNV: Single nucleotide variant

Target: Therapeutically applicable research to generate effective treatments

TCGA: The Cancer Genome Atlas

TF: Transcription factor

TFBS: Transcription factor binding site

TFFM: Transcription factor flexible models

UCSC: University of California Santa Cruz

VAF: Variant Allele Frequency

WES: Whole exome sequencing

WGS: Whole genome sequencing

## List of Figures

## List of Tables

## Abstract

**Background** | Neuroblastoma (NB) is a pediatric tumor composed of mesenchymal (MES) and adrenergic-like (ADRN) cells. These identities derive from the dysregulation of normal cell differentiation, imposed by Core Regulatory Circuitries (CRCs). I hypothesize that noncoding somatic single nucleotide variants (SNVs) in CRCs transcription factor binding sites (TFBSs) may underlie such perturbation, promoting tumorigenesis. **Methods** | MES and ADRN transcriptionally active TFBSs (aTFBSs) were identified integrating 42 ChIP-seq and 12 ATAC-seq experiments in 7 ADRN and 2 MES NB cell lines. Using Fisher test, I tested these regions for an enrichment of somatic SNVs from WGS data of 397 NB patients. This enrichment was correlated with patients clinical and survival data. SNVs in mutated aTFBSs were selected based on their impact on TFBSs through FABIAN-variant. Next, aTFBSs target-genes were identified analyzing promoter capture HiC in 2 ADRN and 2 MES NB cell lines and their expression values were correlated with known NB prognostic markers. **Results** | I found a significant mutation enrichment in aTFBSs of 5 ADRN and 1 MES TFs (FDR $\leq 0.1$), correlated with patients low survival. Then, I selected 689 SNVs affecting the binding of ADRN/MES CRC TFs (Fabian score $\neq 0$). Mutated aTFBSs were found to interact with genes involved in pathways of neuronal differentiation, and MES cells proliferation, thus suggesting potential SNVs impact on NB cell identity definition. aTFBSs carrying SNVs with the highest (cut-off $\geq +0.1$) or the lowest (cut-off $\leq -0.1$) Fabian score, were found to interact with genes whose expression correlated with worse NB outcome. **Conclusions** | These results demonstrated that somatic noncoding SNVs may act synergistically to promote NB tumorigenesis, impacting genes involved in differentiation processes.



**Keywords:** Neuroblastoma, Genomics, somatic mutations, NGS

## Sintesi in italiano

**Contesto** | Il neuroblastoma (NB) è un tumore pediatrico che presenta due identità cellulari, una mesenchimale (MES), l'altra adrenergica (ADRN). Esse emergono dalla deregolazione del differenziamento cellulare e sono definite da 2 circuiti centrali di regolazione (CCR). La mia ipotesi è che varianti somatiche a singolo nucleotide (VSSN) in siti di legame per fattori di trascrizione (FT) dei CCR comportino tale deregolazione, innescando lo sviluppo del NB. **Metodi** | 42 ChIP-seq e 12 ATAC-seq condotte in 9 linee di NB (7 ADRN e 2 MES) sono state integrate per identificare siti di legame attivi per FT dei CCR MES e ADRN. Tramite test di Fisher, è stato valutato il loro arricchimento in VSSN ottenute dal sequenziamento del genoma di 397 pazienti di NB. Tale arricchimento è stato correlato a dati clinici e di sopravvivenza di pazienti con NB. VSSN alteranti il legame a FT CCR-specifici sono state selezionate con Fabian-variant e, tramite promoter capture HiC condotte in 2 linee ADRN e 2 MES di NB, sono stati identificati i geni interagenti con i siti mutati. L'espressione di questi ultimi è stata correlata a noti marker prognostici in NB. **Risultati** | Ho identificato un arricchimento significativo di VSSN nei siti di 5 FT ADRN e 1 MES, associato a scarsa sopravvivenza di pazienti con NB. Inoltre, ho rilevato 689 VSSN influenzanti il legame per FT CCR-specifici (Fabian $\neq 0$). I geni interagenti con i siti mutati sono coinvolti in processi di differenziamento neuronale e proliferazione di cellule MES, suggerendo il coinvolgimento delle VSSN nella definizione dell'identità cellulare del NB. Inoltre, le VSSN con punteggi di Fabian $\geq +0{,}1$ o $\leq -0{,}1$ sono state osservate interagire con geni correlati a prognosi sfavorevole nel NB. **Conclusioni** | Questi risultati dimostrano che VSSN non codificanti possono agire sinergicamente agendo sull'espressione di geni coinvolti nel differenziamento e promuovendo lo sviluppo del NB.

**Parole-chiave:** Neuroblastoma, Genomica, mutazioni somatiche, NGS

# 1. Background

## 1.1. Neuroblastoma, the deadliest solid tumor of infancy

Neuroblastoma (NB) is known as the deadliest solid tumors in early childhood [1]. Indeed, it counts 25-50 cases per million individuals, with 90% NBs diagnosed in children less then 10 years and a median age at diagnosis of 18 months [2]. This neuroendocrine tumor arises from the developing peripheral sympathetic nervous system (SNS), specifically from neural crest (NC)-derived cells undergoing a defective sympathetic neuronal differentiation due to genomic and epigenetic impairments (**Figure 1**).

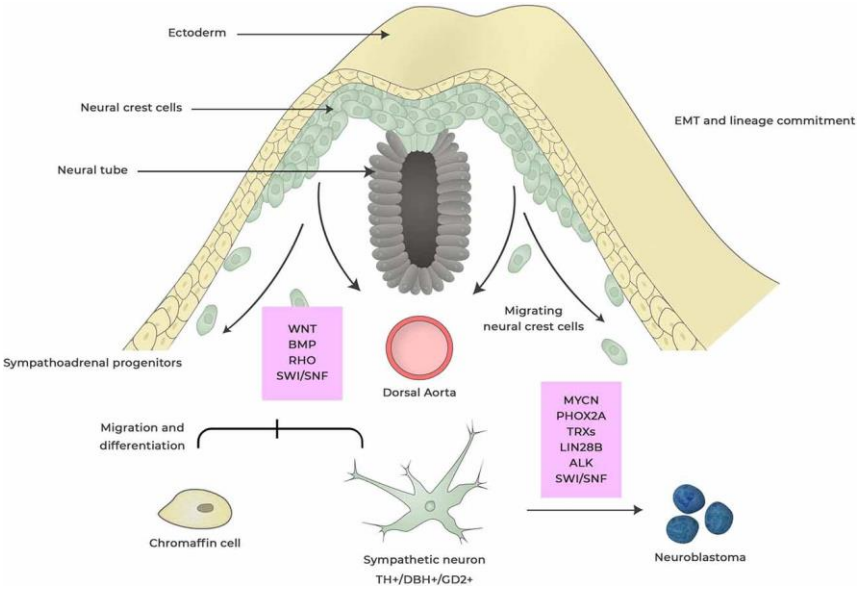


**Figure 1. NB origins from neural crest.** The neural crest cells go through an epithelial-to-mesenchymal transition during embryogenesis, which allows the cells to proliferate, migrate, and differentiate into a variety of cell types that contribute to the anatomical structures of the organism. A complicated network of environmental signals, transcriptional program activation, and epigenetic events play a role in controlling this process. A dysregulation of the components involved in this process can lead to modifications in the specification of cells as well as abnormal migration and differentiation of cells, resulting in hyperplastic lesions that have the potential to develop into NB. *Adapted from Johnsen JI et al. Neuroblastoma-A Neural Crest Derived Embryonal Malignancy. Front Mol Neurosci. 2019 Jan 29;12:9.*

As a result, NB may localize at any final site reached by migrating NC progeny (peripheral neurons, enteric neurons, glia, melanocytes, Schwann cells, cells of the craniofacial skeleton and adrenal medulla), preferentially in the adrenal glands or in the sympathetic ganglia (**Figure 2**) [3].

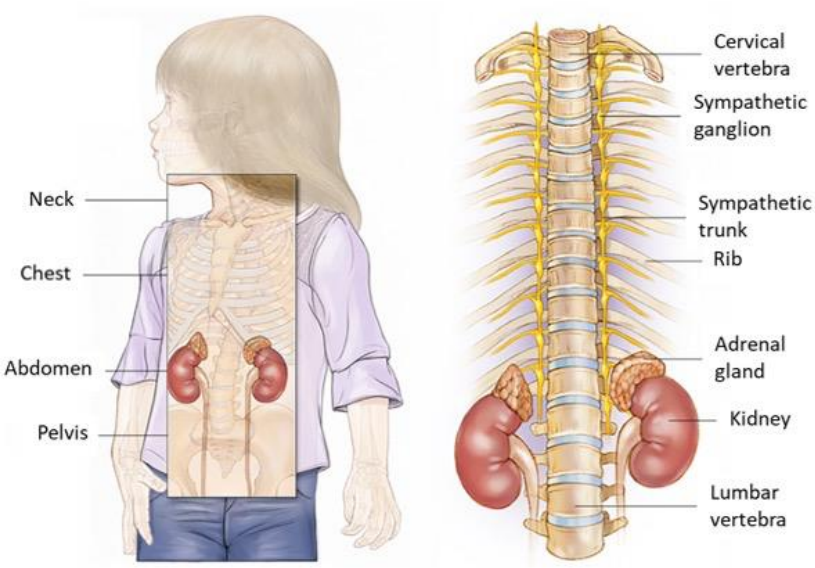


**Figure 2. Clinical Presentations of NB.** *Adapted from American Society of Clinical Oncology 2005*

NB shows high biologic and clinical heterogeneity. Primary tumors localized in the adrenal glands, which represent the vast majority of cases, occur as abdominal masses. On the other hand, tumors arising from sympathetic nerve roots present symptoms that vary depending on the site of origin. For example, NBs arising in the neck can present Horner syndrome (resulting in disrupted sympathetic innervation of one side of the face), while patients with tumors arising in the thorax can be characterized by airways impairment [4].

Metastases occur early in NB with around 50% of patients already presenting bone marrow, bone and lymph nodes metastases at diagnosis next to variable and

severe symptoms such as fever, malaise, pain, and/or cytopenias due to bone and/or bone marrow disease [4].

Despite NB represents one of the most lethal childhood cancers, with 5-year survival rate that can fall below 50% in the most severe cases, it is also associated with one of the highest proportions of spontaneous and complete regression of all human cancers [5].

### 1.2. Neuroblastoma staging and stratification

Multiple staging systems have been used in NB studies, but the most widely accepted within the past 30 years was the International Neuroblastoma Staging System (INSS) [6]. It is a postsurgical staging system developed in 1986 and refined in 1994 which was based on the extent of tumor surgical resection, tumor location at diagnosis and metastases and lymph nodes status (**Figure 3**).

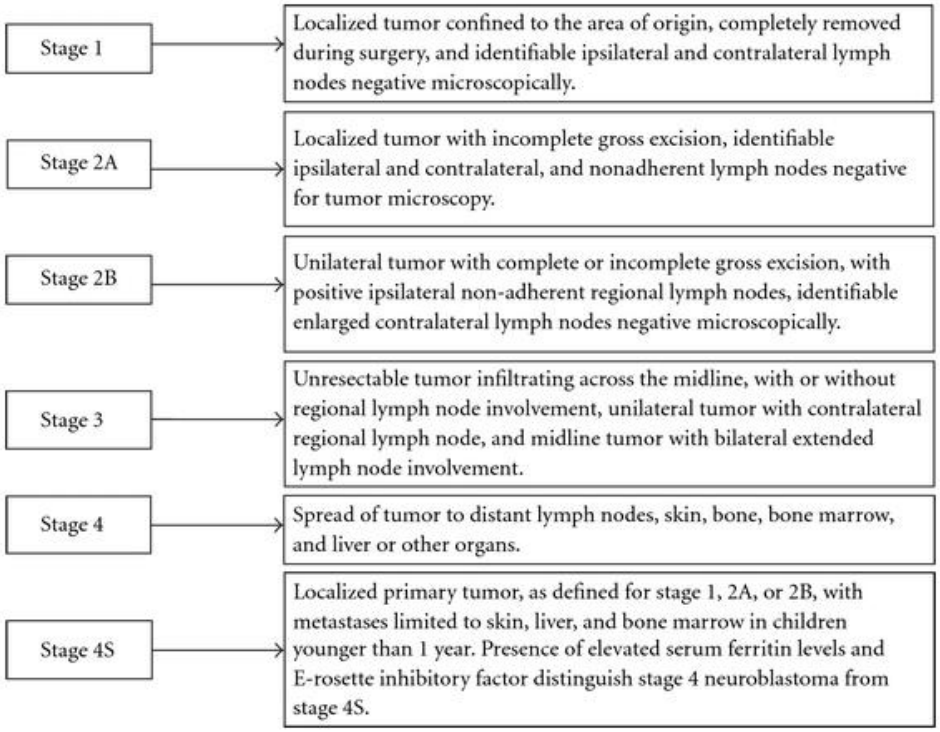


**Figure 3. Staging system for NB according to the INSS.** *Adapted from Roy Choudhury S et al. Targeting angiogenesis for controlling neuroblastoma. J Oncol. 2012; 2012:782020.*

Survival among patients with INSS stage 4 disease was significantly worse than those with stages 1, 2, 3 and 4S stage. The latter represents a peculiar NB category with cases showing disseminated metastases involving distant lymph nodes, bone, bone marrow, liver, and/or other organs (as stage 4), but with more favorable outcome [7]. More recently, in 2004, the International Neuroblastoma Risk Group (INRG) elaborated an updated Staging System based on prognostic factors that are independent of surgical resection extent (age, stage, histology, ploidy, *MYCN* amplification, 11q status and DNA ploidy) (**Figure 4**) [8]. Based on this system, NB patients can be stratified in risk groups (very low, low, intermediate, and high-risk) based on 5 years survival probabilities.

| 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 |
| | | | | | Yes | | G Intermediate |
| | ≥ 18 | GNB nodular; neuroblastoma | Differentiating | NA | No | | E Low |
| | | | | | Yes | | H Intermediate |
| | | | Poorly differentiated or undifferentiated | NA | | | |
| | | | | 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 | | | NA | No | | C Very low |
| | | | | | Yes | | Q High |
| | | | | Amp | | | R High |

**Figure 4. International Neuroblastoma Risk Group (INRG) Consensus Pretreatment Classification schema.** *Adapted from Cohn SL et al. The International Neuroblastoma Risk Group (INRG) classification system: an INRG Task Force report. J Clin Oncol. 2009;27(2):289-297.*

## 1.3. Genetic Predisposition to neuroblastoma

Familial NB accounts for 1-2% of cases, where the main susceptibility genes are *ALK* and *PHOX2B* [9, 10]. *PHOX2B* encodes a transcription factor that drives neural crest development [3]. It represents the first gene found implicated in the arising of familial NBs, with about 10% of cases caused by *PHOX2B* germline mutations [11], such as frameshift and missense variants [12]. Subsequently, the major susceptibility gene identified was *ALK*. In particular, the primary location of *ALK* gain-of-function mutations, which characterize 75% of familial cases, is the kinase domain of the encoded tyrosine kinase receptor [11, 13]. Overall, *ALK* mutations exhibit incomplete penetrance [14].

Additionally, there is evidence of a germline impact in sporadic NB, with smaller effect sizes for common polymorphic alleles and larger effect sizes for

uncommon pathogenic variants. As an example, rare germline variations of *BARD1* have been found in children with high-risk NB [15].

Moreover, Genome-Wide Association Studies (GWAS) on DNA from peripheral blood revealed an increasing number of Single Nucleotide Polymorphisms (SNPs) associated with NB susceptibility both in coding and noncoding regions. Specifically, SNPs found in *BARD1*, *CASC15*, *CASC14*, *DDX4/IL31RA*, *DUSP12*, *HACE1*, *LIN28B*, *LMO1*, and *PT53* showed a strong correlation with NB susceptibility and/or aggressiveness [16-20]. Functional investigations performed for some of these SNPs confirmed their role in controlling the expression of genes implicated in NB carcinogenesis [21-23].

### 1.4. Genomic features of neuroblastoma

NB genome is characterized by large chromosomal aberrations, low exonic mutation rates, and genomic changes that support telomere maintenance. *MYCN* amplification, which is characterized by more than 10 copies per diploid genome, is one of the most prevalent segmental chromosomal abnormalities in young children. In addition, other arm-level chromosomal alterations are reported as poor prognostic features including unbalanced gain of 17q (60%), deletion of 1q (30%) and 11q (45%) [9]. Moreover, children older than 6-years showed 19q loss and 1q gain as unique structural variants [24].

In primary NB, the most commonly somatic coding alterations are represented by activating mutations in *ALK* and inactivating variants in *ATRX* [25-27]. Furthermore, in high-risk NB rearrangements activating the *TERT* locus and mutations in the RAS, TP53 and FGF pathways are reported [26, 28, 29].

Taken together, these studies suggest that a deeper knowledge of the genetic alterations associated with NB may influence the prognosis and treatment of patients. A more detailed picture of genomic alterations can improve the

personalized medicine, contributing to the development of more powerful and less toxic drugs [30].

### 1.5. Neuroblastoma cell identities

As many other tumor types, NB is not uniformly composed by the same cell type, but it represents a more heterogeneous structure with genetically different cell identities reflecting sympathetic nervous system development [31]. The first hint that NB exhibits cellular heterogeneity raised from studies in NB cell lines. In particular, it was reported that NB cells may present with diverse morphological and biochemical properties *in vitro*, defining three types of phenotype called N-type (neuroblast) or ADRN (adrenergic), S-type (substrate-adherent) or MES (mesenchymal) and I-type, which is a biochemical intermediate between N- and S-types [32, 33].

The ADRN population is neuron-like dense cells with small spiny and scant cytoplasm. These cells with delicate neuritic processes grow as aggregates and barely attach to the substrate. On the other hand, the MES cells represent large and flat cells that attach rapidly and strongly to the substrate [32]. ADRN population is characterized by activities of neurotransmitter enzymes such as dopamine-hydroxylase (DBH) and tyrosine hydroxylase (TH), while any catecholamine activities are detected in the MES population, which were reported to exhibit tyrosinase activity (**Figure 5**) [34].

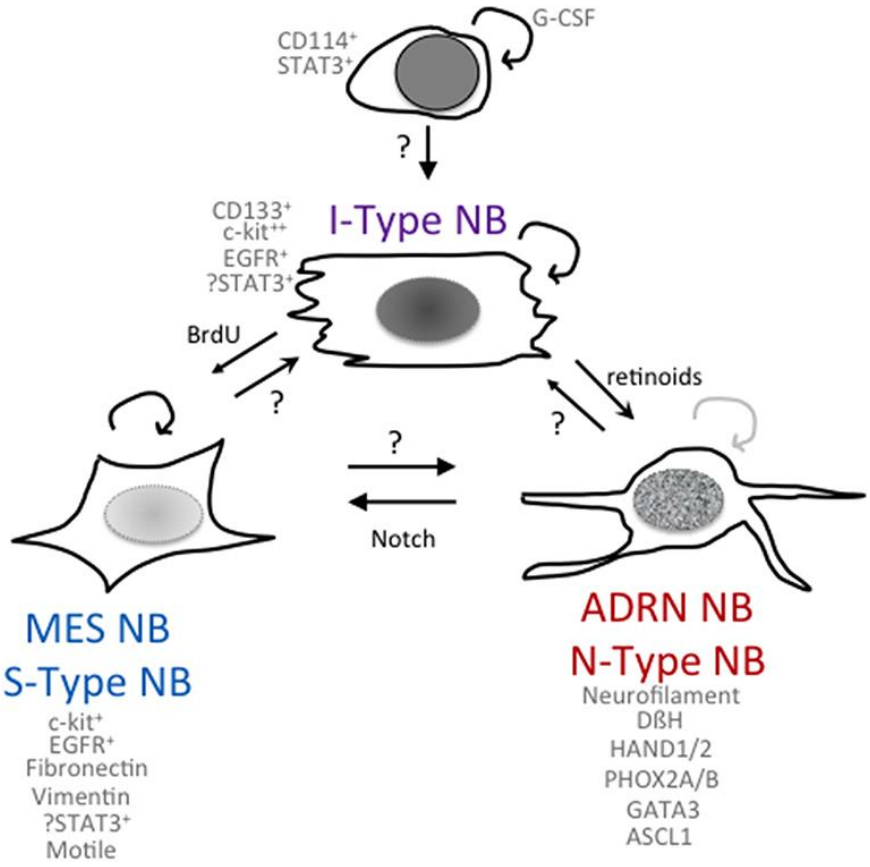


**Figure 5. Schematic of potential relationships among NB cells with different phenotypes.** *Adapted from Veschi V et al. Cancer Stem Cells and Neuroblastoma: Characteristics and Therapeutic Targeting Options. Front Endocrinol (Lausanne). 2019; 10:782*.

Further characterization of ADRN and MES cells was performed considering several extracellular matrix (ECM) proteins and intermediate filaments, known to be cell- and tissue-specific [35]. Whereas ADRN cells express *NEFL*, *NEFM and NF200,* S-type cells express *VIM* and *FNI* genes resulting in the production of intermediate filaments vimentin and fibronectin, responsible for flatness and adhesiveness [33]. Furthermore, previous studies have also highlighted extensive production of stromal collagens of isotypes I and III in MES cells, which is linked to the expression of *COL1A1* and *COL3A1* genes [36, 37].

## 1.6. Core Regulatory Circuitries underlie adrenergic and mesenchymal neuroblastoma identities

Phenotypical differences between cell states can be attributed to differences in expression and activity of master TFs (mTFs) [38]. This concept was conclusively demonstrated by Yamanaka and colleagues, which highlighted the role of Oct3/4, Sox2, c-Myc, and Klf4 in establishing embryonic stem cell identity in mouse adult fibroblasts, thus highlighting that cell transcriptome determining cell state is flexible and can be rewired in the context of forced expression of small, selected sets of mTFs [39]. Distinct cell states and phenotypes are produced by distinct cohorts of mTFs inducing distinct transcriptional programs [39]. In addition, beside the regulation of lineage-specific genes, these mTFs regulates its own expression forming autoregulatory loops by the physical binding of each mTF protein to enhancer regulatory elements distal to the gene body [40]. This network organization model is defined the ‘‘Core Regulatory Circuitry’’ (CRC) [41, 42].

In the context of NB, ADRN and MES specific clusters of super-enhancers (SEs), large genomic domains densely occupied by master TFs and mediators, were identified using H3K27ac ChiP-seq data in human NB cell lines, patient-derived xenografts (PDXs), and primary NCC lines [43]. This highlighted that at the basis of ADRN and MES phenotype there are specific CRCs [41]. In particular, what emerged from this study was that the majority of NB cell lines and PDXs showed an ADRN identity controlled by super-enhancer-associated mTFs, which was highly co-expressed and formed a conserved CRC in these samples (ADRN-CRC). Interestingly, PHOX2B, HAND2 and GATA3 were shown to be co-localized in the regulatory regions of their respective loci, in addition to numerous other notable driver genes such as *MYCN* and *ALK*. On the other hand, another group of NB cell lines exhibited a more MES-like identity driven by a

CRC module containing mTFs, such as FOSL1 and FOSL2 (MES-CRC). Moreover, a third group of NB cell lines was characterized by a mixture of both ADRN and MES identities. A mixed identity preferentially shifted towards the ADRN one was highlighted by transcriptomic and epigenetic analysis of NB tumors [44, 45].

In recent years, several studies were also focused on a more deeper investigation of the main mTFs involved in both NB cell identity specification [46, 47].

To date, the identified MES-specific CRC consists of 20 TFs, including PRRX1, FOSL1, FOSL2, JUN, and NOTCH3, while the ADRN CRC includes 18 TFs, among which PHOX2B, GATA3, HAND2, TBX2, LMO1, ASCL1, ISL1, all involved in the specification of the sympathetic nervous system (**Figure 6**) [3, 46-49].

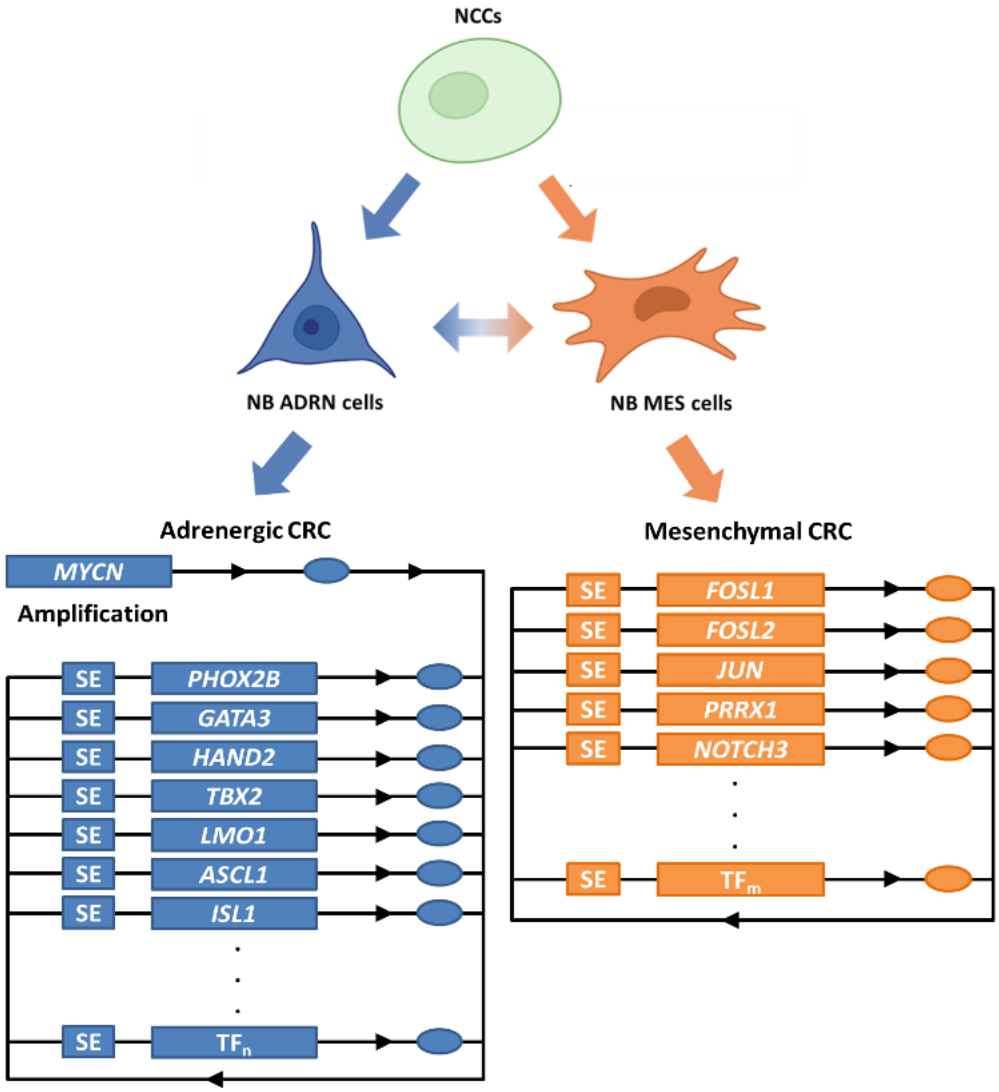


**Figure 6. Schematization of ADRN and MES CRCs.** ADRN CRC is depicted in blue, while the MES one is depicted in orange. In boxes the name of the main CRC mTFs are reported. SE: super-enhancer.

## 1.7. Neuroblastoma plasticity

Epigenetic changes underlie the interconversion between ADRN and MES cell states. Indeed, several studies have shown that targeting lineage-specific CRCs can induce ADRN/MES trans-differentiation *in vitro*. The transitional and

epigenetic landscape of ADRN cells can be reprogrammed towards a MES state by over-expressing the Paired related homeobox protein 1 (PRRX1) gene, a MES-specific CRC transcription factor (TF) [46]. Later, the Neurogenic locus notch homolog protein 3 (NOTCH3) was identified as a master regulator of ADRN-to-MES reprogramming [50].

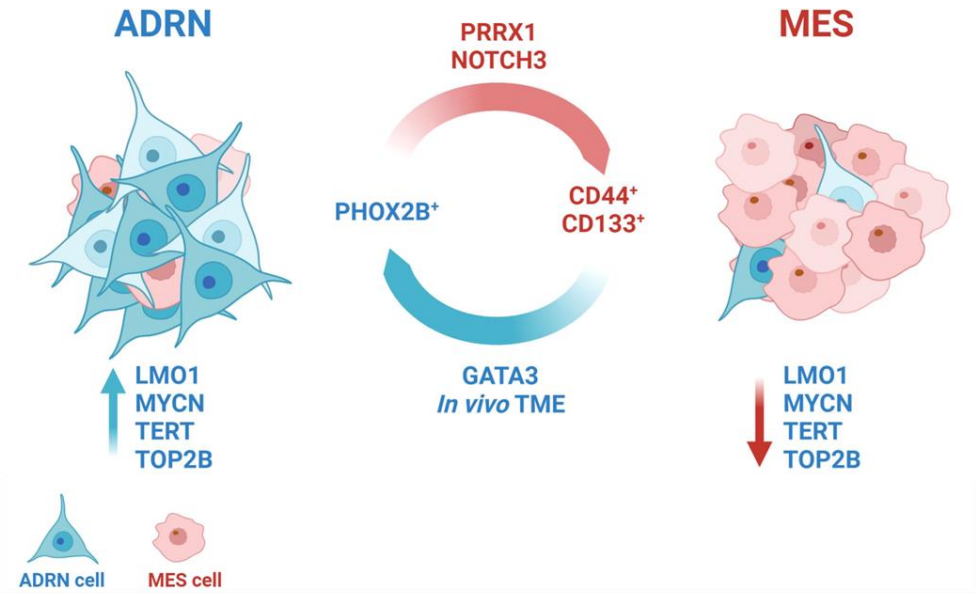


**Figure 7. Neuroblastoma plasticity.** The interconversion from the ADRN to the MES occurs spontaneously and bidirectionally *in vivo*. The transition can be induced through the activation or overexpression of specific transcription factors (TFs), such as GATA3, PRRX1, NOTCH3. *Adapted from D'Amico, Silvia et al. Two bullets in the gun: combining immunotherapy with chemotherapy to defeat neuroblastoma by targeting adrenergic-mesenchymal plasticity. Frontiers in immunology vol. 14 1268645. 2 Oct. 2023.*

Numerous additional genes have been found to support the maintaining of a specific transcriptional program. As an instance, *TERT* was found to be sufficient to induce a reversible switch from ADRN and MES subclones which exhibit different levels of telomerase activity and telomeric protein expression [51]. Likewise, it was discovered that *TOP2B* was necessary to suppress the alternative MES-like epigenetic state and preserve the ADRN-like transcriptional signature in SHSY5Y cells [52]. In fact, *TOP2B* silencing in SHSY5Y cells causes a

downregulation of 47% of the genes in the ADRN signature and upregulation of 38% of the genes in the MES [52].

Despite is currently unclear whether NB tumors can adopt a pure MES phenotype, there are several lines of evidence suggesting the interconversion between ADRN and MES phenotypes *in vivo*. In 2017, ADRN-type and MES-type cells were separated from primary patients' samples using CD133 cell surface expression as a marker for MES-type cells. Tracing of CD133 expression in prolonged culture of purified MES and ADRN NB cells highlighted that they were able to interconvert between states, thus confirming plasticity of NB [46]. The same group performed a transcriptomic characterization on 33 NB cell lines (8 MES-like and 25 ADRN-like) obtaining two identity-specific gene expression signatures which they compared with the transcriptional profiling of 122 NB patients of all stages, highlighting the presence of both ADRN and MES signature [46]. More recently, a single-cell RNA sequencing (scRNA-seq) highlighted the presence of ADRN tumor cells with a few MES-like characteristics in NB tumor samples [45]. All together, these results underlined that NB samples are a mixture of both cell types, with tumors mainly shifted toward ADRN state and confirmed the intra-tumoral heterogeneity of NB.

Dynamic and heterogeneous interconversion between ADRN and MES cell subtypes has been associated with responses to therapy in NB [53]. In this scenario, MES cells, which represent a more proliferative and chemo-resistant phenotype, were found enriched in post-therapy and relapsed tumors, suggesting that the conversion of cell identity and plasticity of NB is involved in response to therapy [46]. Remarkably, the link between NB plasticity and resistance to treatment is highlighted in our recent work in which etoposide and cisplatin resistant NB and their parental cells were investigated by means of single-cells RNA-seq analysis. In particular, we found that drug resistance was associated

with an alteration in the expressions of drug targets and genes involved in DNA double-strand break (DSB) repair, such as *BARD1*, *BRCA1* and *PARP1* [54]. Interestingly, ADRN signature was enriched in cisplatin-resistant cells compared to parental cells which were characterized by MES ones, while the etoposide-resistant cells were equally enriched in MES and ADRN genes, suggesting a higher plasticity potential [54]. All together, these results suggest that drug treatment induces the formation of cell subpopulations with distinct transcriptome profiles.

### 1.8. Noncoding somatic mutations in neuroblastoma tumorigenesis

The advent of NGS has revolutionized many areas of research including cancer genomics. To date, thousands of samples could be quickly sequenced and analyzed revealing thousands of potentially pathogenic mutations. Several studies focused mainly on the protein-coding part of the genome (1% to 2%) to identify cancer treatment targets by using Whole Exome Sequencing (WES). Despite it resulted in the discovery of new therapeutic targets, Whole Genome Sequencing (WGS) studies performed by international consortia, such as The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC), have shown that the majority of somatic mutations occur in the noncoding genome which is still poorly investigated [55, 56].

Since NB appears to be characterized by a lack of somatic coding variants, noncoding regions of the genome are currently the subject of interest for the identification of alterations contributing to NB initiation, progression, maintenance, and metastasis [27, 57].

Among the alterations in the nonconding regions of the genome, single nucleotide variants (SNVs) can influence transcriptional and post-transcriptional gene regulation playing a role in the development of cancer. Falling in cis-

regulatory elements, such as promoters and enhancers, the noncoding SNVs can alter the binding of TFs, which in turn causes a dysregulation of gene expression [58]. However, despite recent evidence regarding the role of somatic noncoding SNVs in the context of cancer, the identification and functional characterization of driver SNVs remain difficult tasks. Moreover, the high number of passenger mutations, due to the weak selective pressure in the noncoding portion of the genome, complicates the identification of putative driver tumorigenic mutations. Therefore, the application of computational approaches to prioritize somatic driver SNVs is required.

One of the most well-established approaches to identify candidate cancer driver SNVs is to investigate DNA elements with putative regulatory function [59]. Generally, regulatory regions can be identified by integrating several assays followed by massive parallel sequencing: DNase hypersensitivity sites (DHSs) and Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) which detect open chromatin and active regions of the genome; Chromatin Immunoprecipitation Sequencing (ChIP-Seq) experiments which can identify the region bound by TFs and/or characterized by specific histone modifications (such as H3K27ac, H3K4me3 and H3K4me1), which give information about the type (promoter or enhancer) and the state (active or non-active) of a regulatory element [59]. In this regard, public data from large epigenomic projects such as ENCODE, as well as those from smaller projects, have become an essential source of information for researchers [60].

By applying computational approaches to prioritize somatic noncoding variants, recent studies have demonstrated that SNVs falling in regulatory regions can affect the expression of genes implicated in cancer onset and development [61, 62]. In this regard, my research team has recently identified the enrichment of somatic noncoding mutations in NB-specific cis-active regulatory regions that

affect the expression of target-genes involved in tumor formation and immune escape [61]. Moreover, we further highlighted that NB tissue-specific active binding sites of tumorigenic transcription factors (TFs) are significantly enriched in somatic SNVs, demonstrating the presence of direct links between TFs, active regulatory elements and somatic mutations [62].

Taken together, these results suggest that regulatory elements can represent new resource of oncogenic somatic variants that can synergistically act to drive cancer initiation.

# 2. Aim

NB represents a neural crest derived tumor characterized by intra-tumoral heterogeneity reflected by two major cell identities (ADRN and MES) defined by specific CRCs. Years of international efforts have highlighted the role of somatic coding mutations in NB tumorigenesis, leading to the identification of several therapeutic targets. Despite aggressive therapies, NB still causes death in 50% of high-risk patients, to date. Since recent evidence highlighted noncoding genome as a new source of driver mutations, the research of alterations contributing to NB tumorigenesis is now shifted to noncoding regions. However, limitations of variant prioritization and interpretation techniques are currently hinder the research of cancer drivers in noncoding portion of the genome.

In this scenario, I hypothesize that somatic noncoding mutations affecting the ADRN and MES CRCs can contribute to NB tumorigenesis.

With this in mind, I set up a computational pipeline to identify noncoding SNVs in active binding sites of CRC master TFs which can affect the binding for these TFs, altering the regulation of genes involved in pathways correlated to developmental and differentiation processes and the definition of ADRN and MES identities.

Results from this project can lead to the identification of potential somatic noncoding drivers which can synergically contribute to the perturbation of still uninvestigated genes that can be targeted by novel therapies.

# 3. Methods

## 3.1. Cell culture

The human SHSY5Y (ACC 209, DSMZ), SKNBE2C (CRL-2268, ATCC), SHEP (CRL-2269, ATCC) and GIMEN (ACC654, DSMZ) cell lines were grown in Dulbecco's Modified Eagle Medium (DMEM, Sigma), 1:1 mixture of Minimum essential Medium Eagle (EMEM, Lonza) and Nutrient Mixture F12 (Sigma), Roswell Park Memorial Institute (RPMI) 1640 Medium, respectively. In all cases medium was supplemented with 10% heat-inactivated FBS (Sigma), 1vmmol/L L- glutamine, penicillin (100 U/mL), and streptomycin (100mg/mL; Invitrogen). Cells were cultured at 37°C, 5% CO2 in a humidified atmosphere and only early-passage cells were employed. All cell lines were re-authenticated and tested as mycoplasma-free.

## 3.2. Chromatin Immunoprecipitation Sequencing (ChIP-seq) in neuroblastoma cell lines

### 3.2.1. ChIP–seq experiments

13 ChIP experiments were performed by using iDeal ChIP-seq kit for Transcription Factors (C01010055, Diagenode) according to manufacturer instructions. Briefly, 4 million cells per IP reaction were fixed by cross-linking solution (Fixation Buffer, 11% formaldehyde) for 15 minutes at room temperature (RT) with gentle shaking. After stopping fixation with glycine, cells were washed in phosphate-buffered saline (PBS) and lysed in Lysis buffer iL1b at 4°C for 20 minutes. Cells were centrifugated at 500 x g and 4°C for 5 minutes and then resuspend in Lysis buffer iL2 and incubated for 10 minutes at 4°C. After a centrifugation for 5 minutes at 500 x g and 4°C, Shearing buffer iS1b with Proteinase Inhibitor Cocktail (Thermofisher) was added to cell pellet and

incubated on ice for 10 minutes. The lysates were subject to 30 cycles of sonication (30 s ON 30 s OFF, high frequency) using Diagenode Bioruptor Plus (Diagenode). In parallel, 30 µl of beads per IP were washed in ChIP buffer iC1b 3 times and were mixed to 5 µg of ChIP-grade antibodies and Protein A-coated magnetic beads for 4 hours at 4°C. ChIP-grade antibodies used were: anti-GATA3 (ab199428; Abcam), anti MYCN (ab227822; Abcam), anti-JUN (ab32137; Abcam), anti-FOSL2 (19967; Cell Signaling) and anti-PRRX1 (SAB1412737; Sigma-Aldrich). After collecting 2.5 µL of sonicated lysate as input, the samples were incubated with antibody beads conjugates overnight at 4 °C with rotation. Next day, the beads were washed and decrosslinking was performed for 4 h at 65 °C and samples were then purified and eluted. After DNA quantification, samples were prepared to a final concentration ≥ 2 ng/µL and sent to Novogene for sequencing. Samples were sequenced on an Illumina sequencing platform, thus obtaining data of 12 ChIP-seq related to the binding profiles of 2 ADRN CRC TFs (GATA3 and MYCN) in 2 ADRN NB cell lines (SKNBE2C and SHSY5Y) and 3 MES CRC TFs (JUN, FOSL2 and PRRX1) in 2 MES NB cell lines (SHEP and GIMEN).

### 3.2.2. Publicly available ChIP-seq data

Raw data of 29 ChIP-seq related to the binding profiles of 8 ADRN CRC TFs (GATA3, HAND2, ISL1, TBX2, MYCN, PHOX2B, ASCL1, LMO1) in 6 ADRN NB cell lines (SKNBE2C, Kelly, NGP, COGN415, LAN5 and NB1643) deposited in NCBI Gene Expression Omnibus (GEO) with accessions GSE94824 [47], GSE169616 [63], GSE80151 [49], GSE94782 [64], GSE120074 [65], GSE65664 [66], GSE138315 [64] were downloaded and re-analysed as described in the next section.

### 3.2.3. ChIP-seq data analysis

Fastq files for ChIP-seq were processed using Nextflow together with the nf-core ChIP-seq v1.2.2 pipelines [67]. In brief, reads adapters were trimmed using Trim Galore [68]. Alignment was performed using BWA [69], duplicates were removed, and peaks were called using MACS2 [70] in narrow peak mode. Peaks with FDR $\leq 0.01$ were retained. Quality control parameters, such as the fraction of reads in peaks (FRiP), normalized strand cross-correlation coefficient (NSC) and relative strand cross-correlation coefficient (RSC), were determined according to the Encyclopedia of DNA Elements (ENCODE) guidelines [71]. Sets of peaks related to the binding of the same TF in the same cell line were considered as replicates and integrated using the procedure described in [72]. Briefly, the lengths of overlapping peaks within replicates were averaged and centered at the average of their peak summits.

### 3.2.4. TF binding site analysis

TF motif enrichment analysis was performed using Homer v4.11 findMotifsGenome.pl tool [73] on the ChIP-seq peaks specific to each NB cell line. Results were filtered for highly enriched ($p < 1 \cdot 10^{-10}$ and Target sequences > 5%) TF binding motifs.

## 3.3. Assay for transposase-accessible chromatin using sequencing (ATAC–seq) in neuroblastoma cell lines

### 3.3.1. ATAC-seq experiments

For ATAC-seq experiments performed by AZENTA life sciences, SHSY5Y, SKNBE2C, SHEP and GIMEN frozen cell suspension were prepared. According to GENEWIZ-protocol, a total amount of 1,500,000 cells in 1500 µL were cryopreserved in cell culture media supplemented with 10% DMSO and sent for

sequencing. Data of 4 ATAC-seq related to the chromatin accessibility profiles in 2 ADRN (SKNBE2C and SHSY5Y) and 2 MES NB cell lines (SHEP and GIMEN) were obtained.

### 3.3.2. Publicly available ATAC-seq data

11 ATAC-seq data related to the chromatin accessibility profiles in 6 ADRN NB cell lines (SKNBE2C, Kelly, NGP, COGN415, LAN5 and NB1643) deposited in NCBI Gene Expression Omnibus (GEO) with accessions GSE94824 [47], GSE80152 [49], GSE138315 [64] were downloaded and re-analysed as described in the next section.

### 3.3.3. ATAC-seq data analysis

Fastq files for ATAC-seq were processed using Nextflow together with nfcore-core ATAC-seq v1.2.2 pipelines [67]. In brief, reads adapters were trimmed using Trim Galore [68]. Alignment was performed using BWA [69], duplicates were removed, and peaks were called using MACS2 [70] in narrow peak mode. Peaks with FDR $\leq$ 0.01 were retained. Similarity between different ATAC-seq experiments was evaluated using Jaccard's coefficient (the ratio of the number of intersecting base pairs between two sets of regions to the number of base pairs in the union of the two sets), which was calculated using BEDTools suite. Sets of peaks in the same cell line with Jaccard's coefficient $\geq$ 0.4 were considered as replicates and integrated using the procedure described in [72].

## 3.4. Definition of CRC active Transcription Factor Binding Sites (aTFBSs)

ATAC-seq and ChIP-seq in the same cell line were integrated to select active TFBSs (aTFBSs). Briefly, I selected only TFBSs bound by the same TF overlapping with open chromatin regions (OCRs). Subsequently, TF-specific aTFBSs found in different NB cell lines were integrated to define aTFBSs sets

representative of the active regions bound by each CRC TF. The latter step was accomplished using the strategy described in [72].

### 3.5. Tumors collection

80 NB 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.

### 3.6. 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.

### 3.7. 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.

### 3.8. Somatic mutations detection

#### 3.8.1. In-house WGS data

WGS of 80 normal-primary NB 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 were detected with Strelka [74].

#### 3.8.2. Publicly available WGS data (EGA and Target)

We obtained access to WGS alignment files from the European Genome-phenome Archive (EGA) (accession no.: EGAD00001001687, EGAD00001006739 and EGAD00001006626) including in our analysis 180 primary NBs. Somatic SNVs were detected with Strelka [74]. In addition we extended our analysis to WGS data of 137 NBs from the Target project (accession no.: phs000218.v21.p7; project ID: #14831; ref. 12) for which somatic variants were already available.

### 3.9. Somatic variants selection criteria

Somatic variants of the in-house and EGA datasets were processed as follows. From raw variant calls, we first eliminated those that did not pass Strelka Empirical Variant Score (EVS). As reported in [74], this is done with the EVS model, a supervised random forest classifier trained on labeled data from sequencing runs carried out under a variety of conditions (different sequencers, sample preparation, and coverage). The EVS model provides an aggregate

quality score for each variant, which takes in account features like (1) the genotype probability computed by the core variant probability model, (2) root-mean-square mapping quality, (3) strand bias, (4) the fraction of reads consistent with locus haplotype model, and (5) the complexity of the reference context as measured by metrics such as homopolymer length and compressibility. Subsequently, I removed SNVs with Variant Allele Frequency (VAF) < 5% and Allele Depth (DP) < 8 and I filtered out common polymorphisms with minor allele frequency (MAF) >1% by using gnomAD databases. Then, considering also the Target datasets, I retained only SNVs mapping on autosomal chromosomes. Finally, SNVs from the three cohorts were merged together.

### 3.10. Somatic mutation enrichment analysis in active CRC TFBSs

I extended aTFBSs ± 1kb from their summits and I split them in 200 bp core regions (containing peak summits) and 900 bp flanking regions (> 200 bp and ≤ 1 kb). Then, I removed regions that could introduce any analysis bias (those overlapping with coding sequences, UCSC blacklisted regions [75] and regions with low unique mappability scores [76]). Moreover, I excluded from the analysis all regions overlapping other TFBSs (stored in the ENCODE database [60]) in their flanking regions. I mapped SNVs on the resulting set of regions and aligned them taking peak summits as reference. Subsequently, the enrichment for mutations in core regions respect to flanking ones was tested. To do this, I compared the ratio of the total number of mutated versus the total number of non-mutated nucleotides within the core regions and flanking regions using Fisher test. p-values were corrected for multiple testing using the Benjamini-Hochberg procedure. Region sets with Fold enrichment > 1 and FDR ≤ 0.1 were considered significantly enriched in somatic mutations in core regions respect to flanking ones. The mutation enrichment analysis workflow is schematized in **Figure 8**.

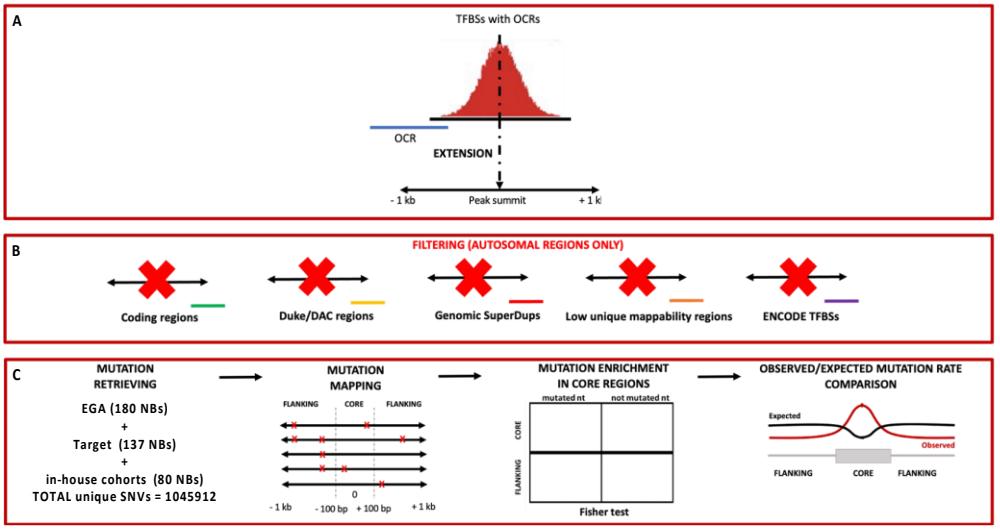


**Figure 8. Schematic work of mutation enrichment analysis.** aTFBSs are defined based on the overlap with OCRs and extended ± 1kb from their summits (**A**). Regions overlapping with coding sequences, UCSC blacklisted regions and regions with low unique mappability scores were removed together with regions overlapping other TFBSs (stored in the ENCODE database) in their flankings (**B**). SNVs were mapped on filtered regions, which were aligned taking peak summits as reference and splitted in 200 bp core regions (containing peak summits) and 900 bp flanking regions (>200 bp and ≤ 1 kb). The ratio of the total number of mutations versus the total number of non-mutated nucleotides was compared within the core regions and flanking ones. The mutation enrichment was tested using Fisher test and p-values were corrected using the Benjamini-Hochberg procedure. Observed mutation rate was compared to the expected one due to local sequence context (**C**).

## 3.11. Comparison with background mutation rate

To check if the observed mutation enrichment was expected due to the local sequence context (a biological factor not related to tumor onset [77, 78]), I randomly introduced the same number of mutations observed at each position in the stacked regions following the probability of occurrence of each mutation according to its tri-nucleotide context [78]. This procedure was repeated 1,000 times obtaining 1,000 simulated mutated sets of regions. Subsequently, I computed an empirical one-sided p-value as the fraction of the simulations with more expected mutations than observed in cores. p-values ≤ 0.1 highlighted any significant association between enrichment of somatic SNVs in core regions and core sequence composition.

### 3.12. Somatic SNVs drivers prioritization

SNVs falling in significantly mutated aTFBSs cores were selected and their impact on CRC TF binding affinity was evaluated with FABIAN-variant tool [79]. Multiple databases of positional weighted matrices (PWM) and transcription factor flexible models (TFFMs) were used as reference (JASPAR, CIS-BP, HOCOMOCO, hPDI, Jolma, SwissRegulon, UniPROBE). FABIAN-variant generates a joint score S between −1 (likely TFBS loss) and 1 (likely TFBS gain). I prioritized all SNVs affecting ADRN and MES CRC TF binding with a non-zero FABIAN score. The workflow of the prioritization strategy previously described was reported in **Figure 9**.

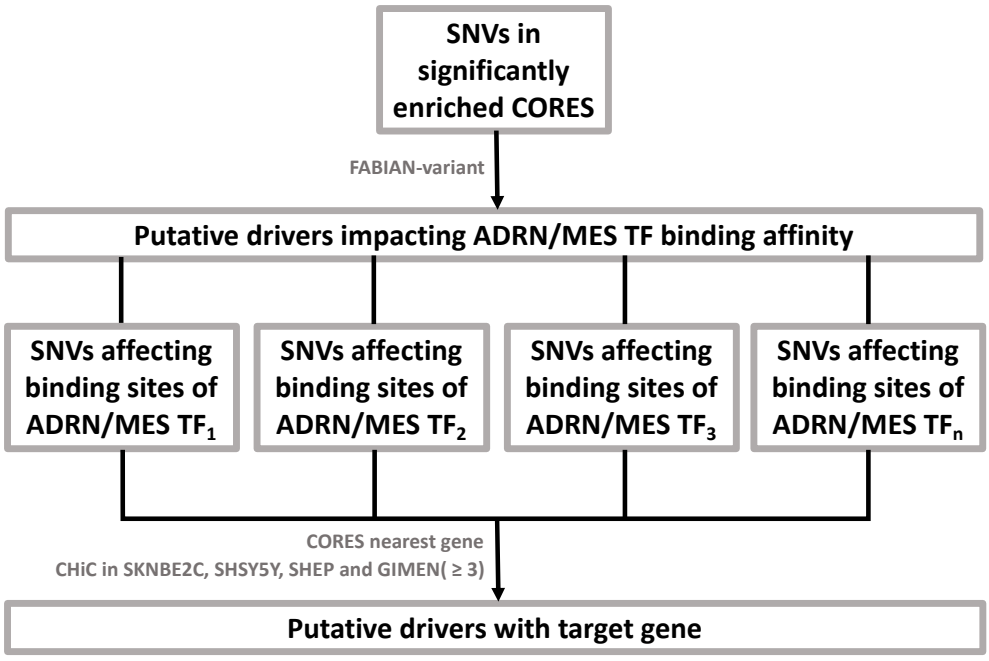


**Figure 9. Schematic workflow of Prioritization strategy.** SNVs falling in significantly mutated aTFBSs cores were selected and their impact on CRC TF binding affinity was evaluated with FABIAN-variant tool. I assigned mutated aTFBSs to their target genes if significant physical interactions were present between the aTFBSs and a gene promoter. In case no significant interactions could be determined, the closest gene was assigned to the aTFBSs.

### 3.13. Promoter Capture Hi-C (CHiC) in neuroblastoma cell lines

#### 3.13.1. Promoter CHiC experiments

5-10 million of crosslinked SHSY5Y, SKNBE2C, SHEP and GIMEN cells by 2% formaldehyde were sent to ARIMA Genomics to perform Promoter CHiC, according to manufacturer protocol. DNA libraries were prepared using the Arima-HiC kit. Target enrichment was performed by using Arima Human Promoter Panel designed to the restriction fragment of the promoters of 23 711 genes from the human GRCh37 Ensemble database, version 95, including: 18,741 protein-coding genes, 84 antisense RNAs, 170 lincRNA's, 1,878 miRNA's, 938 snoRNA's, and 1,898 snRNA's. The probes were manufactured using 1x tiling with repeat masking and balance boosting. Arima Capture-HiC libraries were sequenced via Illumina sequencers in 'paired-end' mode.

#### 3.13.2. Promoter CHiC analysis

Briefly, 150 bp paired-end reads were trimmed and mapped to the reference genome (build hg19/GRCH37). The alignment BAM file was then filtered to remove duplicates, re-ligation or self-circularization artifacts that can be introduced during CHi-C library preparation. These steps were performed using HiCUP [80]. Interaction calling and significance thresholding was performed using CHiCAGO tool [81]. Interactions were called with default parameters at 5 kb resolution and significant ones were defined as those with a CHiCAGO score $\geq 3$. I assigned mutated aTFBSs to their target genes if significant physical interactions were present between the aTFBSs and a gene promoter. In case no significant interactions could be determined, the closest gene was assigned to the aTFBSs.

### 3.14. Gene Ontology analysis

The functional enrichment analysis was performed by using the web tool: WEB-based GEne SeT AnaLysis Toolkit [82]. The Gene Ontology database of non-redundant Biological Processes was used. Significantly enriched GO terms were considered for FDR ≤ 0.05.

### 3.15. Gene expression analysis, survival analysis and correlation with clinical data

R2: Genomics Analysis and Visualization Platform (http://r2.amc.nl) was used to query transcriptomic data of 498 neuroblastoma samples (GSE62564). The expression of genes interacting with aTFBSs carrying SNVs with Fabian score ≥ +0.1 or ≤ - 0.1 were correlated with the expression of CRC TF whose binding was affected by noncoding SNVs. I compared the expression levels of these genes between *MYCN* amplified and *MYCN* non-amplified patients, stage 4 and stage1,2,3,4s patients, high risk and low risk patients using one way ANOVA test. Gene expression values were also tested in relation to overall (OS) and event free survival (EFS) data in order to identify genes associated with worse outcomes. The OS and the EFS probabilities were calculated by using the Kaplan–Meier method. The log-rank test statistical significance was set at 5%. Multivariate Cox proportional regression analysis was performed to evaluate the prognostic significance of the mutational burden in CRC aTFBSs and currently used prognostic factors, such as age at diagnosis (>18 months vs. <18 months), International Neuroblastoma Staging System (INSS) stage (stages 4 vs. stages 1, 2, 3, and 4s), MYCN status (amplified vs. not amplified), and risk group stratification. Hazard Ratios and 95% confidence interval for survival rates were calculated.

# 4. Results

## 4.1. Identification of ADRN and MES CRC aTFBSs in NB cell lines

To define transcriptionally active TF binding sites (aTFBSs), defined as TFBSs overlapping OCRs in the same cell line, I processed and integrated in-house and public ChIP-seq and ATAC-seq raw data in ADRN and MES NB cell lines (**Figure 10)**.

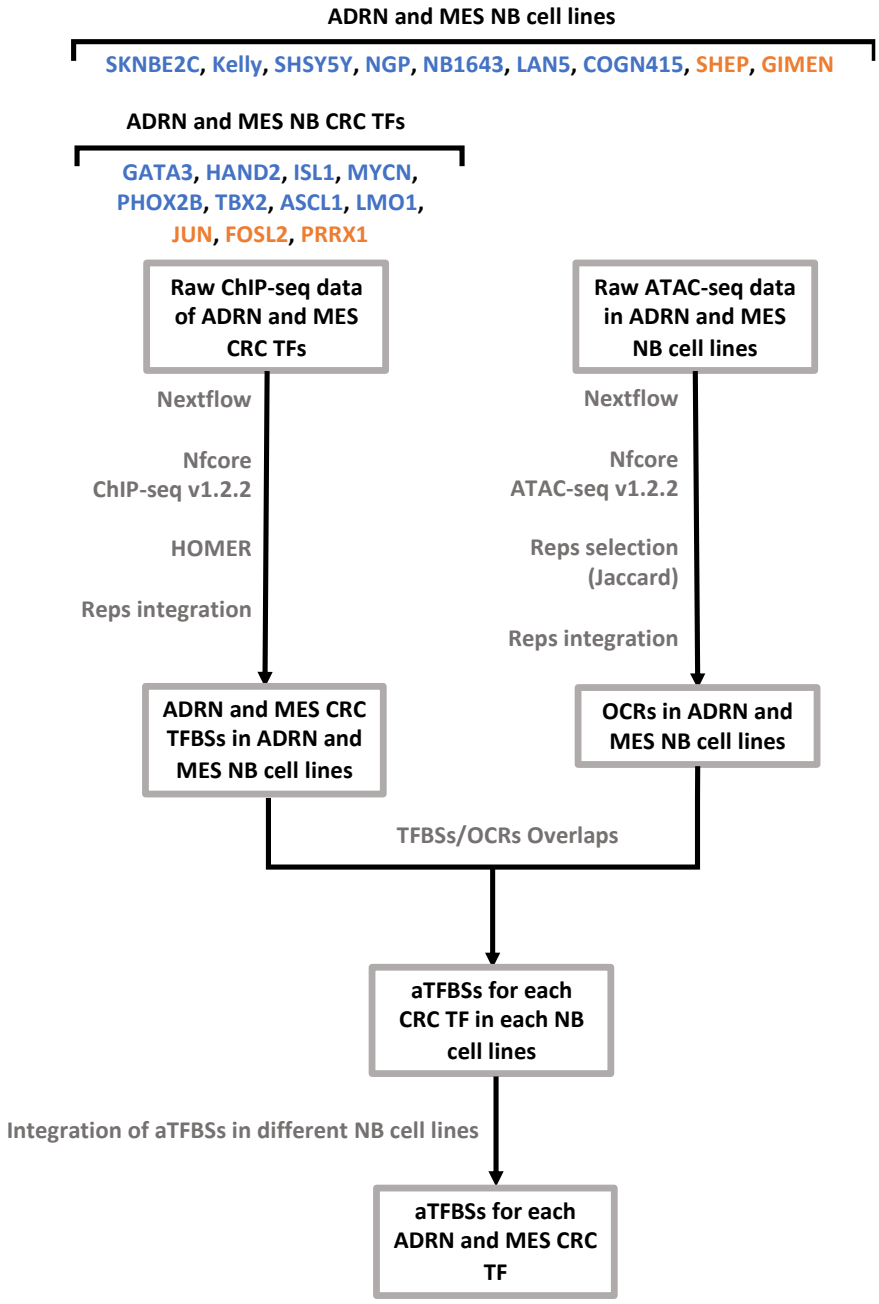


**Figure 10. Schematic workflow of ADRN and MES CRC aTFBS identification.** ADRN NB cell lines are reported in blue, MES ones are highlighted in orange.

To select specific TFBSs of ADRN CRC TFs, I processed and integrated 29 raw public and 7 in-house ChIP-seq of ADRN CRC TFs (GATA3, HAND2, ISL1, MYCN, PHOX2B, TBX2, ASCL1 and LMO1) in 7 ADRN NB cell lines (SKNBE2C, Kelly, SHSY5Y, NGP, NB1643, LAN5 and COGN415). Regarding MES CRC TFBSs, I analyzed data from 6 in-house ChIP-seq experiments of MES CRC TFs (JUN, FOSL2 and PRRX1) in 2 MES NB cell lines (SHEP and GIMEN).

After peak calling and replicates integration for the same TF in the same cell line, I obtained 44,438 (min=1,822; max=287,448) specific-TFBSs on average with a mean length of about 432 bp (min=229; max=1001). Every identified TFBSs set was found significantly enriched in motifs bound by the respective ADRN or MES CRC TF, confirming the reliability of data (p-value $\leq 1 \cdot 10^{-10}$ and proportion of sequences with motif > 5%) (**Table 1**).

**Table 1. ADRN and MES TFBSs selected.**

| Cell line | Cell Identity[1] | TF[2] | N replicates[3] | N peaks[4] | M width[5] | p-value[6] |
|---|---|---|---|---|---|---|
| SKNBE2C | ADRN | GATA3 | 4 | 26260 | 367 | $1 \cdot 10^{-2379}$ |
| SKNBE2C | ADRN | MYCN | 3 | 48971 | 331 | $1 \cdot 10^{-2225}$ |
| SKNBE2C | ADRN | PHOX2B | 2 | 58480 | 341 | $1 \cdot 10^{-8124}$ |
| SKNBE2C | ADRN | ISL1 | 1 | 46730 | 393 | $1 \cdot 10^{-1793}$ |
| SKNBE2C | ADRN | HAND2 | 1 | 61801 | 352 | $1 \cdot 10^{-1925}$ |
| SKNBE2C | ADRN | TBX2 | 1 | 17562 | 341 | $6.62 \cdot 10^{-27}$ |
| Kelly | ADRN | MYCN | 3 | 68111 | 301 | $1 \cdot 10^{-1444}$ |
| Kelly | ADRN | LMO1 | 2 | 1822 | 497 | $6.3 \cdot 10^{-43}$ |
| Kelly | ADRN | GATA3 | 1 | 20298 | 367 | $1 \cdot 10^{-640}$ |
| Kelly | ADRN | HAND2 | 1 | 49550 | 378 | $1 \cdot 10^{-1444}$ |
| Kelly | ADRN | ISL1 | 1 | 9570 | 353 | $1 \cdot 10^{-464}$ |
| Kelly | ADRN | PHOX2B | 1 | 120974 | 304 | $1 \cdot 10^{-16175}$ |
| Kelly | ADRN | TBX2 | 1 | 29888 | 302 | $9.2 \cdot 10^{-82}$ |
| Kelly | ADRN | ASCL1 | 1 | 4564 | 229 | $1 \cdot 10^{-1140}$ |
| SHSY5Y | ADRN | GATA3 | 4 | 21164 | 374 | $1 \cdot 10^{-2768}$ |
| SHSY5Y | ADRN | MYCN | 1 | 5846 | 616 | $1 \cdot 10^{-180}$ |
| NGP | ADRN | GATA3 | 1 | 4921 | 231 | $1 \cdot 10^{-988}$ |
| NGP | ADRN | MYCN | 2 | 287448 | 298 | $1 \cdot 10^{-1760}$ |
| NB1643 | ADRN | MYCN | 3 | 41117 | 581 | $1 \cdot 10^{-231}$ |
| LAN5 | ADRN | MYCN | 1 | 55957 | 636 | $1 \cdot 10^{-504}$ |
| COGN415 | ADRN | MYCN | 1 | 44799 | 670 | $1 \cdot 10^{-347}$ |
| SHEP | MES | JUN | 1 | 49379 | 1001 | $1 \cdot 10^{-6985}$ |
| SHEP | MES | FOSL2 | 1 | 68532 | 704 | $1 \cdot 10^{-16319}$ |
| SHEP | MES | PRRX1 | 1 | 4441.00 | 444 | $5.3 \cdot 10^{-305}$ |
| GIMEN | MES | JUN | 2 | 5210.00 | 469 | $1 \cdot 10^{-2292}$ |
| GIMEN | MES | FOSL2 | 1 | 2016.00 | 356 | $1 \cdot 10^{-1562}$ |

[1] Cell identity: Adrenergic (ADRN) or Mesenchymal (MES) NB cell identity classification
[2] TF: Transcription Factor
[3] N replicates: Number of ChIP-seq replicates per transcription factor in the same NB cell line
[4] N peaks: Number of peaks obtained after peak calling and replicates integration
[5] M width: Average of peaks widths after peak calling and replicates integration
[6] p-value: p-value of the motif enrichment analysis performed with HOMER

Subsequently, I identified open chromatin regions (OCRs) processing and integrating 11 raw public and 2 in-house ATAC-seq data, in the same 7 ADRN NB cell lines (SKNBE2C, Kelly, SHSY5Y, NGP, NB1643 LAN5, COGN415). The same analysis was performed by using 2 in-house ATAC-seq experiments obtained in 2 MES NB cell lines (SHEP and GIMEN) to define MES-specific OCRs. After peak calling, I compared ATAC-seq experiments by using Jaccard similarity score to select ATAC-seq replicates characterized by the most similar chromatin accessibility profiles. Interestingly, I observed that ADRN NB cell lines resulted more similar and grouped together forming a cluster distinct from the MES NB cells one (**Figure 11**). This finding further validates our experimental approach in distinguishing ADRN and MES NB cells using epigenetic markers.

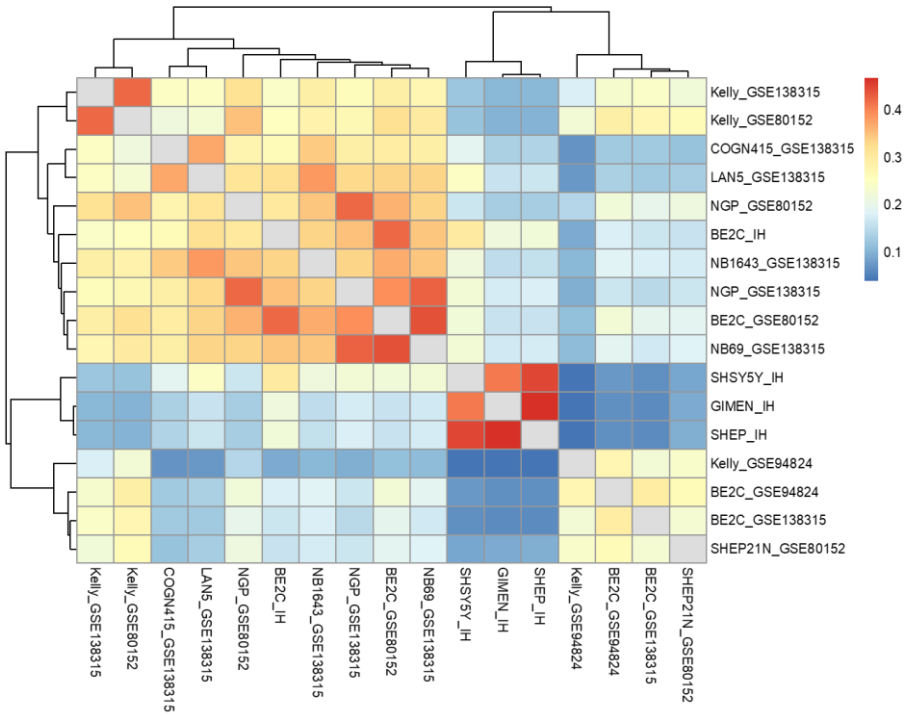


**Figure 11. ADRN chromatin accessibility profiles are distant from MES ones.** Matrix showing the Jaccard scores of pairwise comparisons of ATAC-seq experiments in 7 ADRN and 2 MES NB cell lines. IH: in-house.

For NB cell lines with multiple replicates, ATAC-seq experiments with the highest Jaccard similarity score (cut-off ≥ 0.4) were selected. In this way, I obtained ADRN and MES-specific sets of 97,351 (min= 60,196; max=116,644) OCRs on average with a mean length of about 574 bp (min=347; max=880) (**Table 2**).

**Table 2. ADRN and MES OCRs selected.**

| Cell line | Cell identity[7] | N replicates[8] | N peaks[9] | M width[10] |
|---|---|---|---|---|
| SKNBE2C | ADRN | 2 | 107029 | 381 |
| Kelly | ADRN | 2 | 60196 | 366 |
| SHSY5Y | ADRN | 1 | 111022 | 880 |
| NGP | ADRN | 2 | 116644 | 347 |
| NB1643 | ADRN | 1 | 71835 | 493 |
| LAN5 | ADRN | 1 | 109452 | 527 |
| COGN415 | ADRN | 1 | 99359 | 466 |
| SHEP | MES | 1 | 103602 | 843 |
| GIMEN | MES | 1 | 97020 | 863 |

To identify ADRN and MES CRC aTFBSs, I integrated TFBSs (**Table 1**) and OCRs (**Table 2**) sets previously obtained. Afterwards, since ADRN NB cell lines showed highly similar OCRs profiles compared to the MES ones (**Figure 11**) and to avoid cell line specific biases that could result in analyses artifacts, aTFBSs associated to the same TF in ADRN NB cell lines were merged together. The same procedure was adopted for MES-specific aTFBSs. In this way, I obtained ADRN and MES-specific sets of 29,637 (min= 835; max=72,157) aTFBSs on average with a mean length of about 492 bp (min=235; max=996) (**Table 3**).

[7] Cell identity: Adrenergic (ADRN) or Mesenchymal (MES) NB cell identity classification
[8] N replicates: Number of ATAC-seq replicates per NB cell line
[9] N peaks: Number of peaks obtained after peak calling and replicates integration
[10] M width: Average of peaks widths after peak calling and replicates integration

**Table 3. ADRN and MES CRC aTFBSs identified in ADRN and MES NB cell lines.**

| Cell line | Cell identity[11] | TF[12] | aTFBSs[13] | M width[14] |
|---|---|---|---|---|
| SKNBE2C, Kelly, NGP, SHSY5Y | ADRN | GATA3 | 30475 | 378 |
| SKNBE2C, Kelly | ADRN | HAND2 | 46112 | 414 |
| SKNBE2C, Kelly | ADRN | ISL1 | 27011 | 445 |
| SKNBE2C, Kelly, NGP, COGN415, LAN5, NB1643, SHSY5Y | ADRN | MYCN | 72157 | 438 |
| SKNBE2C, Kelly | ADRN | PHOX2B | 25699 | 441 |
| SKNBE2C, Kelly | ADRN | TBX2 | 26688 | 331 |
| Kelly | ADRN | ASCL1 | 3026 | 235 |
| Kelly | ADRN | LMO1 | 1508 | 535 |
| SHEP, GIMEN | MES | JUN | 40409 | 996 |
| SHEP, GIMEN | MES | FOSL2 | 52087 | 768 |
| SHEP | MES | PRRX1 | 835 | 431 |

These sets of aTFBSs represented the target for the somatic mutation enrichment.

## 4.2. Somatic SNVs from EGA, Target and in-house neuroblastoma cohorts

To detect significantly mutated ADRN and MES CRC aTFBSs, I analyzed tumor-normal matched WGS data of three cohorts of NB patients: 2 public available cohorts of 180 (EGA) and 137 (Target) NBs, and 1 in-house cohort encompassing 80 primary NBs. For Target cohort, I collected already called publicly available somatic variants. Instead, for both the EGA and in-house cohorts, I collected high quality sequencing data as confirmed by QC parameters obtained after reads alignment. Indeed, I observed 98.3% mapped reads and a mean coverage of 68.4 both for EGA normal and tumor samples (**Figure 12**).

[11] Cell identity: Adrenergic (ADRN) or Mesenchymal (MES) NB cell identity classification
[12] TF: Transcription Factor
[13] aTFBSs: active Transcription Factor Binding Sites
[14] M width: Average of aTFBSs widths after integration of aTFBSs sets of the same TF in different NB cell lines

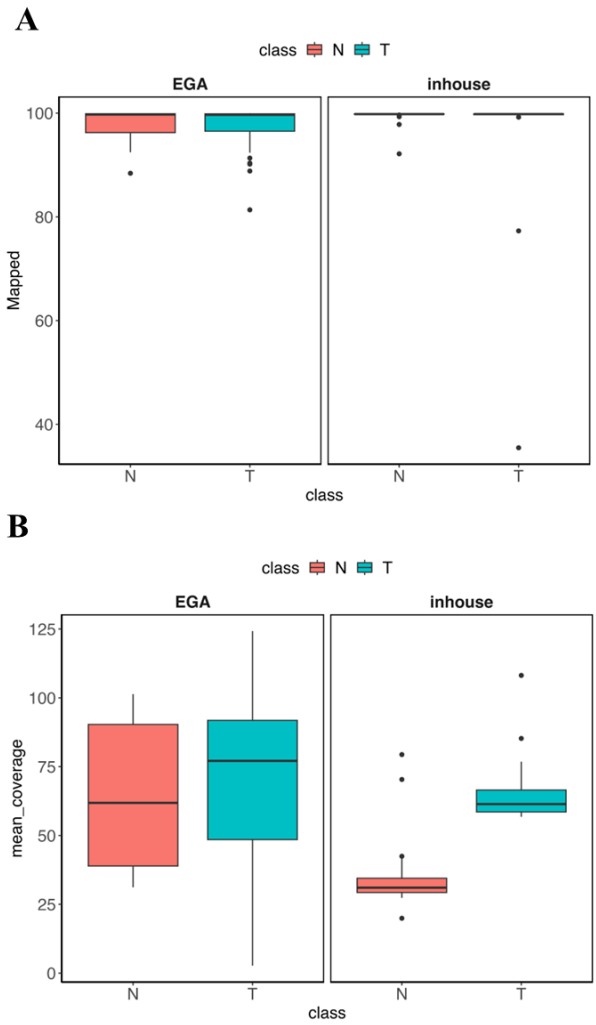


**Figure 12. Descriptive statistics of the EGA (180 NBs) and in-house (80 NBs) tumor-normal matched WGS data.** Distributions of statistics related to normal samples (in salmon) and tumor ones (in green) for EGA (on left) and in-house (on the right) cohorts. Box plots reporting **A,** mean coverage after reads alignment and **B,** percentage of mapped reads.

Both in tumor and normal samples, the proportions of bases covered with at least 10, 20 and 50 reads were 91%, 87% and 55%, respectively (**Figure 13**).

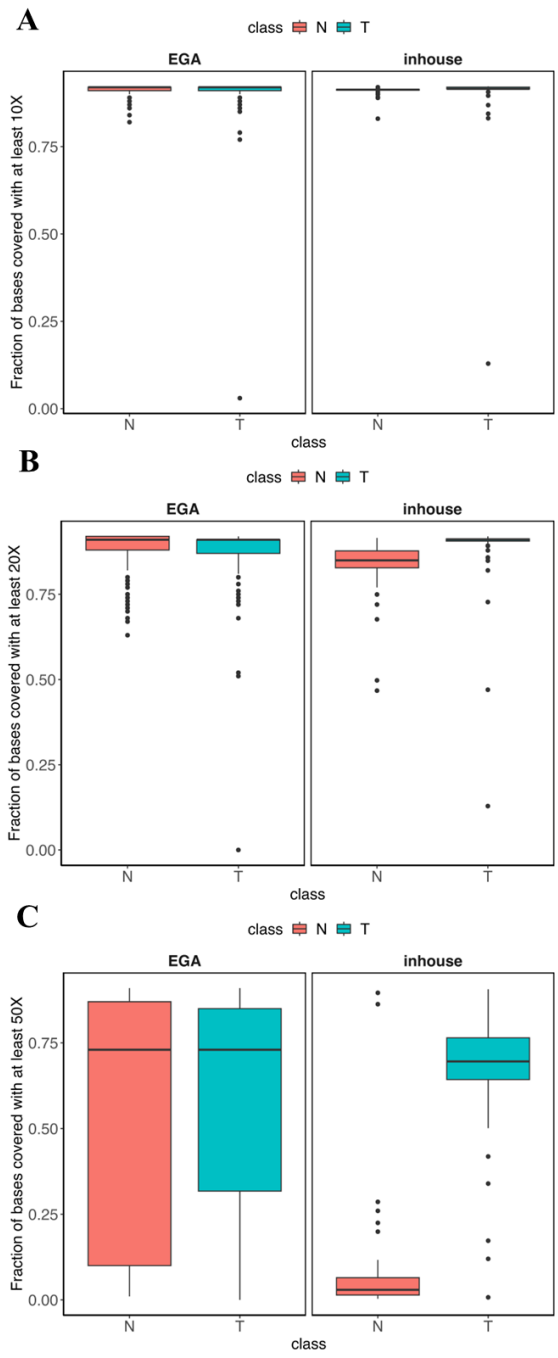


**Figure 13. Descriptive statistics of the EGA (180 NBs) and in-house (80 NBs) tumor-normal matched WGS data.** Distributions of statistics related to normal samples (in salmon) and tumor ones (in green) for EGA (on left) and in-house (on the right) cohorts. Box plots reporting the percentage of bases covered with at least (**A)** 10, (**B**) 20, (**C**) 50 reads.

Regarding in-house cohort, I observed a proportion of mapped reads of 99% for both normal and tumor samples with a mean coverage of 48% and 57%, respectively (**Figure 12**). In normal samples, the proportions of bases covered with at least 10, 20 and 50 reads were 91%, 86% and 37% respectively (**Figure 13**). In tumor ones, the proportions of bases covered with at least 10, 20 and 50 reads were 90%, 86% and 70%, respectively (**Figure 13**). For both EGA and in-house datasets, I called somatic SNVs using Strelka and retained only somatic mutations passing EVS score. Subsequently, I applied stringent filtering criteria to discard false positives (VAF ≤ 5%, Allele mutated Depth ≤ 8) and common polymorphisms (MAF ≥ 1%). After filtering, I obtained 532,886 (median = 2,340 per sample), 1,672,393 (median = 2,796 per sample) and 290,392 (median = 1,990 per sample) somatic SNVs for EGA, in-house and Target datasets, respectively (**Figure 14**).

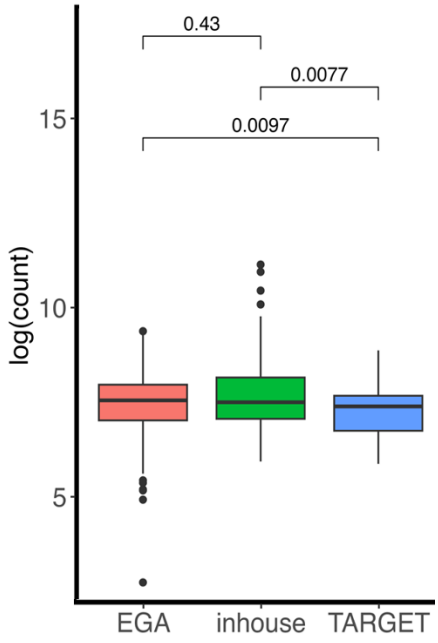


**Figure 14. Somatic SNVs filtering summary.** Box plot showing the count in logarithmic scale of filtered somatic SNVs per sample after removing false positives (VAF ≤ 5%, Allele mutated Depth ≤ 8) and common polymorphisms (MAF ≥ 1%).

After filtering steps, I merged all somatic mutations from EGA, in-house and Target cohorts, obtaining one dataset of 1,045,912 unique SNVs.

### 4.3. Mutational enrichment analysis in ADRN and MES CRC aTFBSs

As detailed in Methods section (**Figure 8**), to assess if CRC aTFBSs were characterized by significant enrichment of somatic mutations, I used a global approach that accounts for trinucleotide sequence context and correct for any potential bias due to the overlap of aTFBSs with coding sequences, regions with mappability problems and binding sites of TF which are not involved in NB CRCs. I evaluated if aTFBS core regions were significantly enriched in somatic SNVs respect to flanking regions. Starting from a total of 19,794 filtered aTFBSs of ADRN CRC TFs (GATA3, HAND2, ISL1, MYCN, TBX2, PHOX2B, LMO1 and ASCL1) (**Table 3**), I observed an enrichment of somatic SNVs for 2811, 4010, 3013, 4573, 2301 aTFBSs of GATA3, HAND2, ISL1, MYCN and TBX2, respectively (Fold Enrichment > 1, FDR ≤ 0.1) (**Figure 15**).

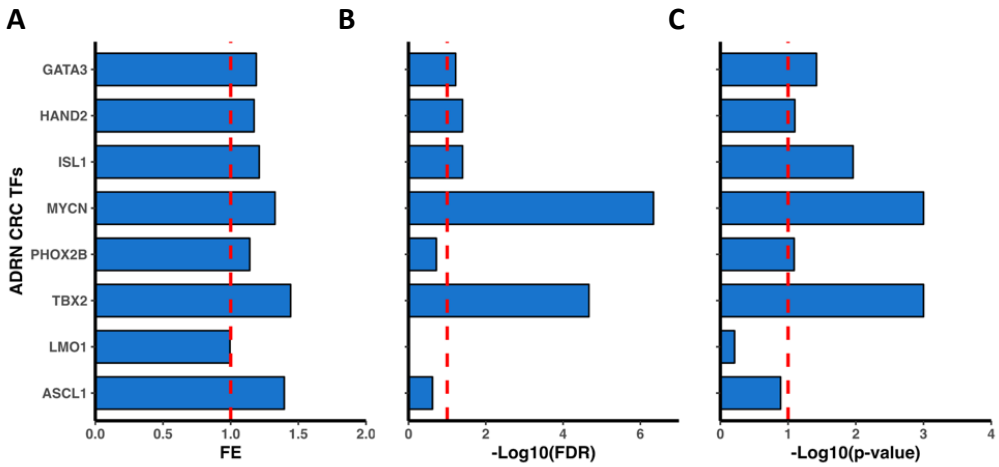


**Figure 15. Mutation enrichment analysis results for ADRN aTFBSs.** aTFBSs sets for CRC ADRN TFs in NB cell lines were investigated. Bar plots in blue show Fold Enrichments (FE) of somatic mutations in aTFBSs cores respect to flanking regions (**A**), -$Log_{10}$(FDR) associated to Fisher Test (**B**) and -$Log_{10}$(p-value) related to permutation tests (**C**). These values are reported for each set of GATA3, HAND2, ISL1, MYCN, PHOX2B, TBX2, LMO1 and ASCL1 aTFBSs. Red line indicates thresholds for FE, -$Log_{10}$(FDR) and -$Log_{10}$(p-value).

Furthermore, I observed that the mutation accumulation is not related to sequence composition (p-value ≤ 0.1).

The same approach was exploited to investigate significantly mutated MES aTFBSs. In brief, to test for a somatic SNVs enrichment in MES aTFBSs cores respect to flanking regions, I started from a total of 4546 filtered aTFBSs of MES CRC TFs (JUN, FOSL2 and PRRX1) (**Table 3**) and I observed an enrichment of somatic SNVs for 2476 aTFBSs of FOSL2 (Fold Enrichment > 1, FDR ≤ 0.1) (**Figure 16**).

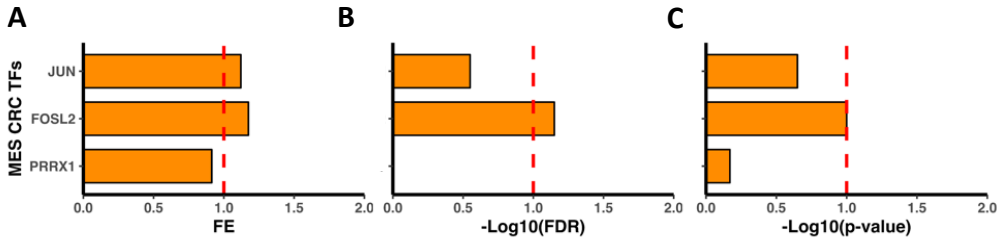


**Figure 16. Mutation enrichment analysis results for MES aTFBSs.** aTFBSs sets for CRC MES TFs in NB cell lines were investigated. Bar plots in orange show Fold Enrichments of somatic mutations in aTFBSs cores respect to flanking regions (FE) (**A**), -$Log_{10}$(FDR) associated to Fisher Test (**B**) and -$Log_{10}$(p-value) related to permutation tests (**C**). These values are reported for each set of JUN, FOSL2 and PRRX1 aTFBSs. Red line indicates thresholds for FE, -$Log_{10}$(FDR) an -$Log_{10}$(p-value).

Again, in significantly mutated MES aTFBSs, I observed somatic mutations accumulation not due to local sequence context (p-value ≤ 0.1). Taken together, these results seem to suggest a somatic mutation burden affecting subsets of ADRN and MES aTFBSs.

### 4.4. The enrichment of somatic mutations in CRC aTFBSs is associated to NB patients low survival probabilities

To assess if the enrichment of somatic mutations in CRC aTFBSs was linked to any relevant feature of NB patients, I categorized 282 out of 397 patients with available clinical data into two groups: the first with at least one noncoding somatic SNVs in significantly enriched CRC aTFBSs sets, the second without

noncoding somatic mutations in these regions. In this way, I identified 182 patients belonging to the first group and 100 patients belonging to the second one. For both groups, I tested the overall survival (OS) probability, highlighting that patients with SNVs in enriched CRC aTFBSs had a worse OS (p-value = 0.00017) (**Figure 17A**). On multivariable analysis, the noncoding mutational burden still significantly influenced the patients' risk of death (p-value = 0.015). However, the patients' worst outcome was not independent from other prognostic factors such as the amplification of *MYCN* (p-value < 0.001) (**Figure 17B**).

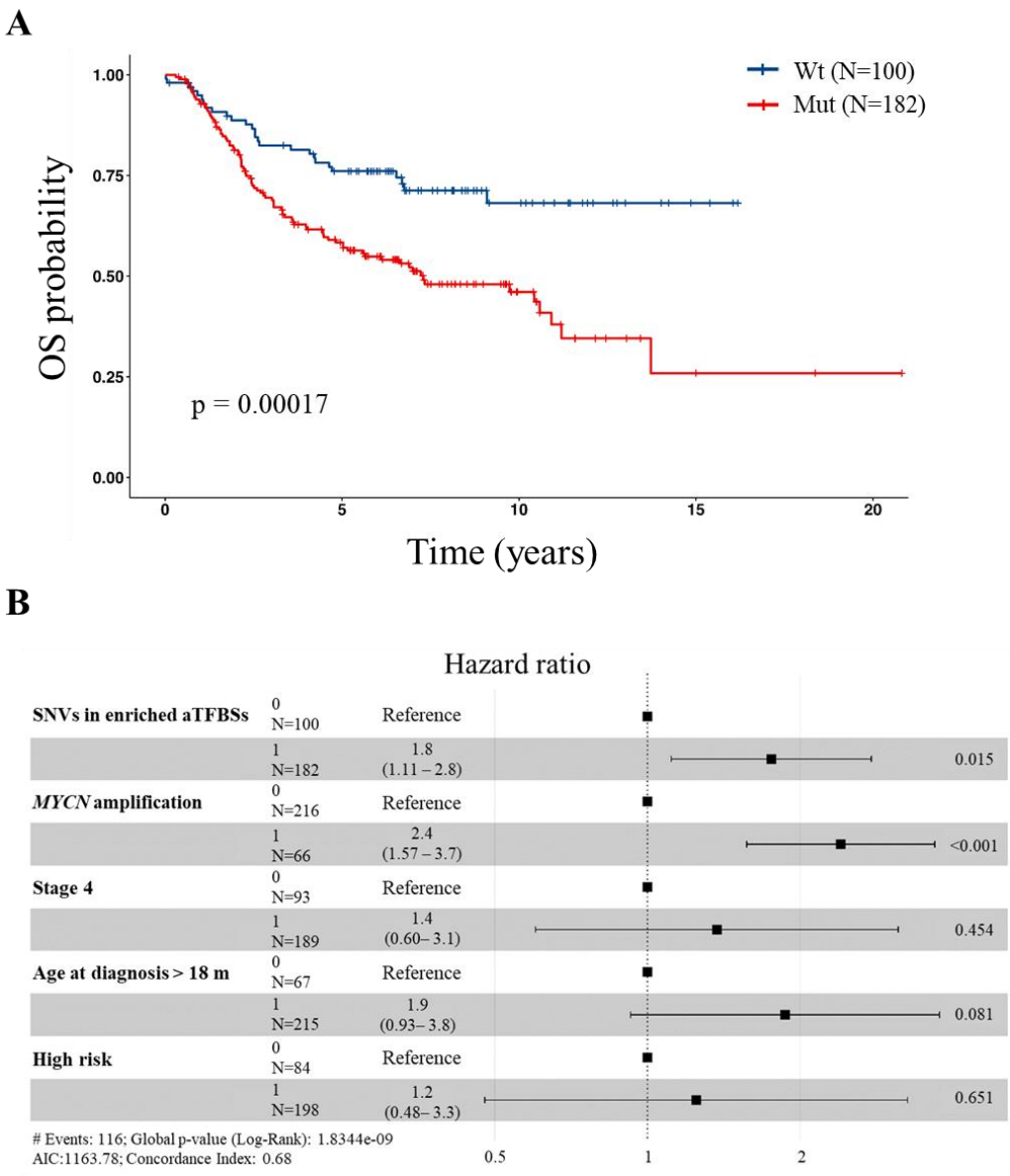


**Figure 17. Somatic variants at significantly mutated CRC aTFBSs affect NB patients overall survival.** **A,** OS of patients carrying (Mut; N = 182) and not carrying (Wt; N = 100) noncoding SNVs in significantly mutated CRC aTFBSs. Statistical significance was calculated with log-rank test. **B**, Forest plot of Multivariate Cox proportional regression analysis evaluating the prognostic significance of the mutational burden in CRC aTFBSs (SNVs in enriched aTFBSs) using known NB prognostic factors related to bad outcome (MYCN amplification, Stage 4 classification, Age at diagnosis > 18 months and High risk classification) as covariates. Hazard Ratios and 95% confidence interval for survival rates were calculated.

Similarly, I tested event-free survival (EFS) probability of patients carrying at least one mutation in enriched CRC aTFBSs and patients with non-mutated aTFBSs. I found that patients of the first group had a significant worse EFS (p-value = 0.00015) (**Figure 18A**). Taking into account other clinical covariates, the noncoding mutation enrichment still significantly influences the patients' risk of unfavorable outcome (p-value = 0.009) together with the age at diagnosis > 18 months (p-value = 0.023) (**Figure 18B**).

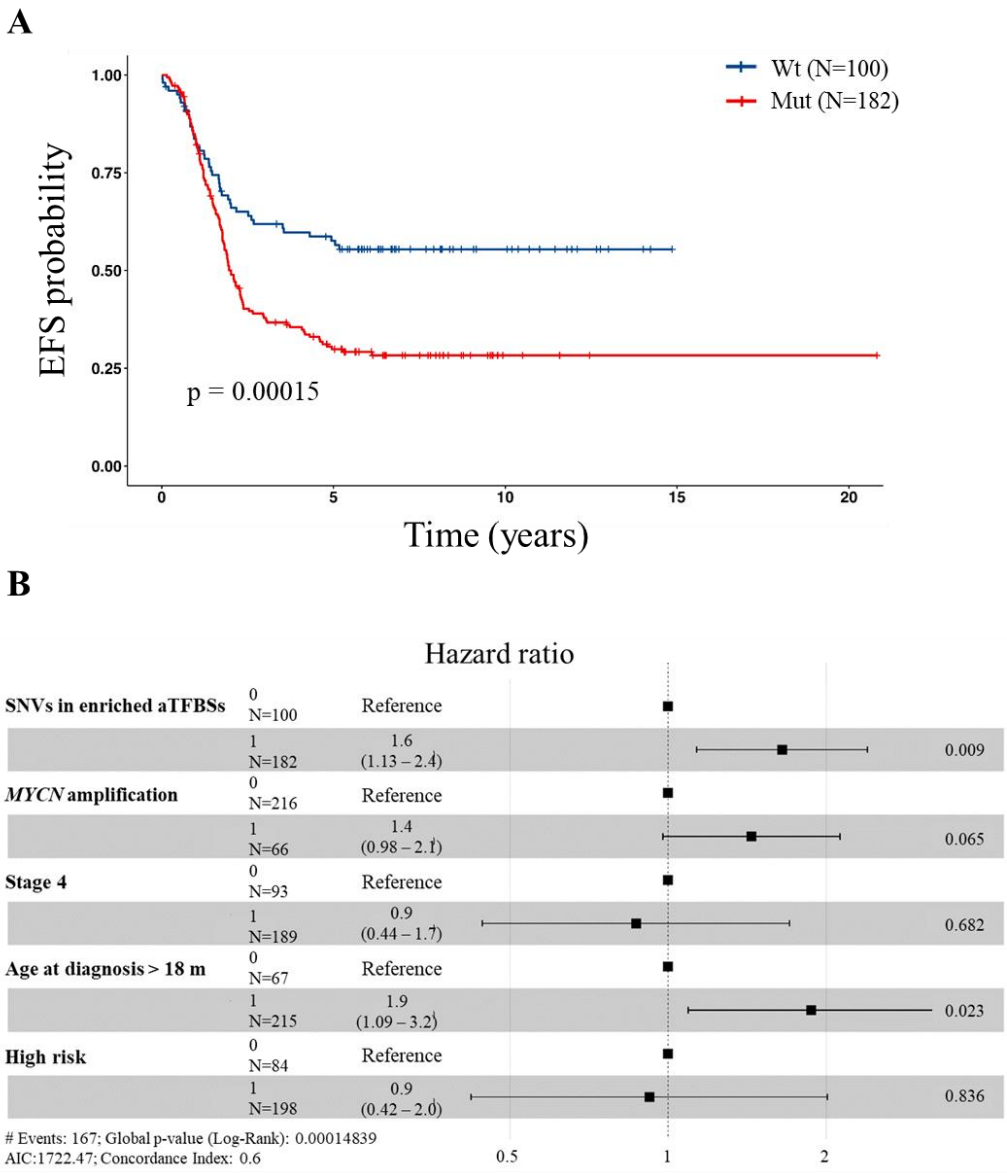


**Figure 18. Somatic variants at significantly mutated CRC aTFBSs affect NB patients event free survival. A**, EFS of patients carrying (Mut; N = 182) and not carrying (Wt; N = 100) noncoding SNVs in significantly mutated CRC aTFBSs. Statistical significance was calculated with log-rank test. **B**, Forest plot of Multivariate Cox proportional regression analysis evaluating the prognostic significance of the mutational burden in CRC aTFBSs (SNVs in enriched aTFBSs) using currently used prognostic factors related to bad outcome (MYCN amplification, Stage 4 classification, Age at diagnosis > 18 months and High risk classification) as covariates. Hazard Ratios and 95% confidence interval for survival rates were calculated.

## 4.5. NB tumorigenesis is triggered by SNVs impacting the binding of ADRN and MES CRC TFs which perturb genes involved in neuronal development and differentiation

To discover somatic noncoding SNVs that perturb ADRN and MES CRC, I identified SNVs which strongly altered the binding for CRC TFs and detected mutated aTFBSs target genes whose regulation could be altered by selected mutations.

Firstly, I selected the SNVs located in aTFBSs cores characterized by a FABIAN-variant score $\neq 0$. Afterwards, I identified predicted target genes of mutated aTFBSs by analyzing promoter Capture HiC data in 2 adrenergic-like (SKNBE2C and SHSY5Y) and 2 mesenchymal-like (SHEP and GIMEN) NB cell lines. For mutated binding sites with any significant interaction detected by CHiC, I reported their closest gene.

As result, I identified a total of 689 somatic SNVs affecting ADRN and MES TF motifs. In particular, 159, 157, 153, 318, 205 and 35 SNVs were predicted to impact the binding of GATA3, HAND2, ISL1, MYCN, TBX2 and FOSL2 respectively. These SNVs mapped on aTFBSs interacting with a total of 1,056 putative target genes, some of which were already known to be involved in NB tumorigenesis (i.e. *PHOX2B, CTTNBP2, MAML2, FGFR1*).

Subsequently, to identify pathways in which mutated aTFBSs target genes were mainly involved, I performed a GO analysis that highlighted significant enrichment (FDR $\leq 0.05$) of 23 biological processes involved in neuronal development and differentiation (e.g., “gland development”, “pattern specification process”, “synapse organization”, “axon development” and “neuron projection guidance”) formation of mature structures originating from neural crest progenitor cells (e.g., “ear development”, “embryonic skeletal system

development" and "sensory organ morphogenesis") and specification of neural crest-like and mesenchymal cell identities (e.g., "stem cell differentiation", "negative regulation of cell development, "mesenchymal cell proliferation" and " mesenchyme development")(**Figure 19**).

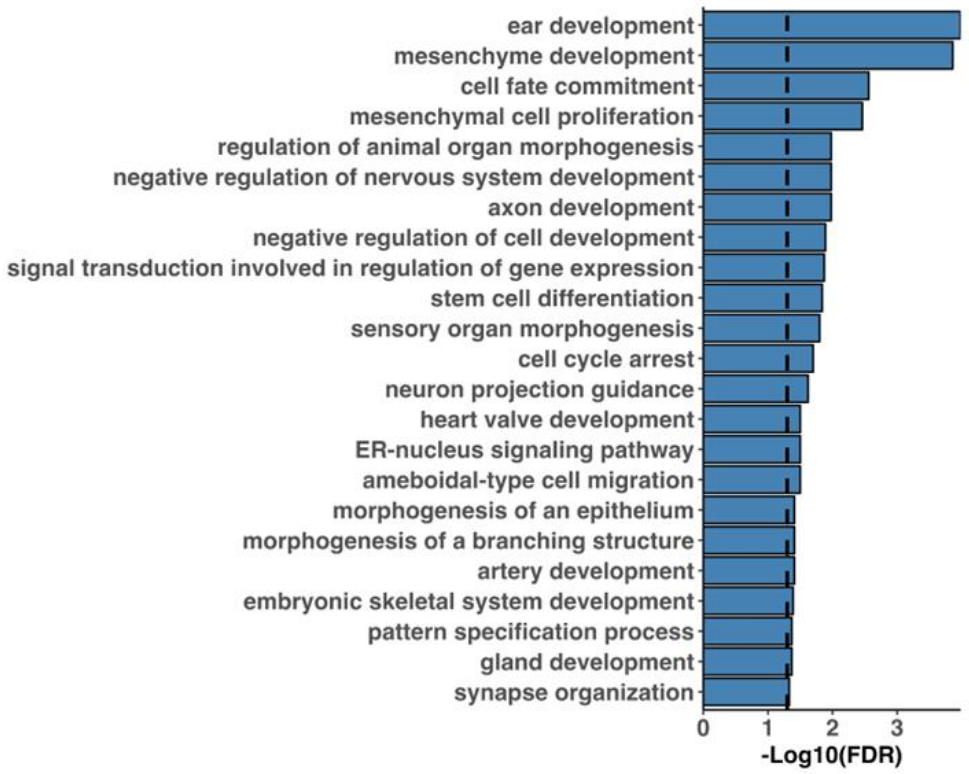


**Figure 19. Enriched Biological Processes of genes interacting with mutated ADRN and MES aTFBSs.** The Gene Ontology enrichment was performed with WebGestalt tool (www.webgestalt.org) using the non-redundant set of Biological Processes. The bar plot shows the -$Log_{10}$ of the enrichment adjusted p-value (FDR). The black dashed line represents the cut-off for statistical significance (set at 0.05).

Considering that NB derives from perturbations of cell differentiation and development, I focused only on genes implicated in such type of biological processes (18 out of 23 significantly enriched pathways). Moreover, to select genes interacting with aTFBSs carrying SNVs with strongly impact on CRC TFs binding affinity, I restricted the analysis on variants with Fabian score $\geq +0.1$ or $\leq -0.1$ (**Table 4**).

**Table 4. Target genes selected on the basis of somatic SNVs Fabian score (≥ +0.1 or ≤ -0.1) and involved in differentiation processes. Variants were stratified for CRC TFs.**

| TF[15] | SNV ID[16] | Fabian Score | Target Gene | Methodology[17] | GO term[18] |
|---|---|---|---|---|---|
| GATA3 | chr13:79713153:C>A | 0.59 | *POU4F1* | CHiC SHSY5Y | Synapse organization |
| | chr1:58610003:G>T | 0.50 | *TACSTD2* | CHiC SHSY5Y | Morphogenesis of an epithelium |
| | chr11:95755776:G>A | 0.31 | *MTMR2* | nearest | Synapse organization |
| | chr2:171027362:G>T | 0.17 | *MYO3B* | nearest | Sensory organ morphogenesis |
| | chr17:54757219:G>T | 0.14 | *NOG* | CHiC SHSY5Y | Morphogenesis of an epithelium |
| | chr9:120149880:T>A | -0.90 | *ASTN2* | nearest | Morphogenesis of an epithelium |
| HAND2 | chr20:57893545:C>G | 0.37 | *EDN3* | nearest | Mesenchyme development |
| | | | *GNAS* | CHiC SHSY5Y | Embryonic skeletal system development |
| | chr10:131086996:A>T | 0.29 | *MGMT* | CHiC SHSY5Y | Gland development |
| | chr3:181416180:G>A | 0.28 | *SOX2* | nearest | Gland development |
| | chr6:107867836:T>G | 0.19 | *SOBP* | nearest | Sensory organ morphogenesis |
| | chr3:77115218:G>T | -0.12 | *ROBO2* | nearest | Synapse organization |
| | chr11:63845903:T>A | -0.21 | *FLRT1* | nearest | Synapse organization |
| | chr2:192916500:A>T | -0.27 | *TMEFF2* | nearest | Morphogenesis of an epithelium |
| | chr17:36906003:T>C | -0.38 | *CACNB1* | CHiC SHSY5Y | Synapse organization |
| | chr5:67845133:C>T | -0.40 | *PIK3R1* | nearest | Axon development |
| MYCN | chr6:37592900:T>G | 0.78 | *MDGA1* | nearest | Synapse organization |
| | chr14:37125587:A>C | 0.69 | *PAX9* | CHiC SKNBE2C and SHSY5Y | Regulation of animal organ morphogenesis |
| | chr4:94755933:C>G | 0.66 | *ATOH1* | nearest | Axon development |

[15] TF: Transcription Factor interacting with mutated aTFBSs
[16] SNV ID: Single Nucleotide Variant Identifier
[17] Methodology: Approach used to identify target genes (CHiC or nearest gene method)
[18] GO term: The most represented Gene Ontology terms in which target genes are implicated

| | | | | | |
|---|---|---|---|---|---|
| | chr13:110263103:C>T | 0.22 | *IRS2* | CHiC SKNBE2C | Gland development |
| | chr10:131086996:A>T | 0.22 | *MGMT* | CHiC SHSY5Y | Gland development |
| | chr2:239067632:A>C | 0.18 | *HES6* | CHiC SKNBE2C | Pattern specification process |
| | chr18:13222542:T>C | -0.11 | *LDLRAD4* | CHiC SHSY5Y | Mesenchyme development |
| | chr2:74347571:T>G | -0.11 | *DGUOK* | CHiC SKNBE2C and SHSY5Y | Negative regulation of cell development |
| | chr17:79367670:T>G | -0.11 | *FSCN2* | CHiC SKNBE2C | Sensory organ morphogenesis |
| | chr11:63845903:T>A | -0.24 | *FLRT1* | nearest | Synapse organization |
| | chr5:3996569:C>A | -1.00 | *IRX1* | nearest | Morphogenesis of an epithelium |
| ISL1 | chr2:192952351:G>T | 0.17 | *TMEFF2* | nearest | Morphogenesis of an epithelium |
| | chr4:111708394:G>A | 0.11 | *PITX2* | nearest | Gland development |
| | chr6:137753475:G>T | -0.24 | *OLIG3* | CHiC SHSY5Y | Cell fate commitment |
| | chr18:40427987:C>T | -0.59 | *RIT2* | nearest | Negative regulation of cell development |
| TBX2 | chr9:20684229:T>G | 0.14 | *MLLT3* | CHiC SKNBE2C | Morphogenesis of an epithelium |
| | chr10:74020551:T>G | -0.11 | *SPOCK2* | CHiC SHSY5Y | Synapse organization |
| | chr4:112217746:T>A | -0.14 | *PITX2* | nearest | Gland development |
| | chr17:27918839:T>G | -0.14 | *SSH2* | CHiC SKNBE2C and SHSY5Y | Axon development |
| FOSL2 | chr5:5203712:C>A | 0.43 | *ADAMTS16* | CHiC SHEP | Morphogenesis of an epithelium |

To predict the regulation of specific-target gene expression by its TFs, I evaluated the correlation between target gene and TF expression values by analysing a well-annotated RNA-seq dataset (GSE62564, n = 498) ($R \neq 0$ and p-value $\leq 0.05$). Finally, to retain genes potentially involved in NB tumorigenesis, I selected those whose expression was significantly associated with markers of bad prognosis in NB such as poor OS and EFS probabilities, *MYCN* amplification, Stage 4 and High-risk classifications (p-value $\leq 0.05$).

In this way, I found 9 target genes (*ROBO2, CACNB1, PIK3R1, MDGA1, HES6, LDLRAD4, DGUOK, IRX1* and *SPOCK2*) of HAND2, MYCN and TBX2

mutated aTFBSs whose expression was correlated with CRC TF expression, OS, EFS, *MYCN* status, INSS Stage and Risk categories in NB patients (**Table 5**).

**Table 5. Expression of target genes of HAND2, MYCN and TBX2 correlates with NB prognosis markers.** p-values (p) of genes whose downregulation is significantly associated with NB bad prognosis markers (*MYCN* amplification, Stage 4 and High risk stratification) and lower OS and EFS are highlighted in red, while those of genes whose upregulation is significantly associated with NB bad prognosis markers and lower OS and EFS are highlighted in blue.

| TF[19] | Target Gene | TF corr[20] | OS[21] | EFS[22] | *MYCN*$_a$ vs *MYCN*$_{na}$[23] | St4 vs St1/2/3/4s[24] | High R vs Low R[25] |
|---|---|---|---|---|---|---|---|
| HAND2 | *ROBO2* | R:0.36 p:$9.24\cdot10^{-17}$ | p:$4.4\cdot10^{-7}$ | p:$2.9\cdot10^{-12}$ | p:$1.0\cdot10^{-9}$ | p:$1.4\cdot10^{-6}$ | p:$1.3\cdot10^{-10}$ |
| | *CACNB1* | R:0.32 p:$1.92\cdot10^{-13}$ | p:$3.5\cdot10^{-15}$ | p:$2.8\cdot10^{-12}$ | p:$5.0\cdot10^{-32}$ | p:$2.4\cdot10^{-17}$ | p:$2.3\cdot10^{-25}$ |
| | *PIK3R1* | R:0.12 p:$6.6\cdot10^{-3}$ | p:$9.6\cdot10^{-21}$ | p:$4.9\cdot10^{-11}$ | p:$1.2\cdot10^{-14}$ | p:$5.5\cdot10^{-37}$ | p:$3.6\cdot10^{-33}$ |
| MYCN | *MDGA1* | R:-0.51 p:$1.1\cdot10^{-34}$ | p:$3.0\cdot10^{-14}$ | p:$1.0\cdot10^{-06}$ | p:$5.0\cdot10^{-33}$ | p:$5.2\cdot10^{-23}$ | p:$7.8\cdot10^{-38}$ |
| | *HES6* | R:0.31 p:$1.6\cdot10^{-12}$ | p:$1.7\cdot10^{-6}$ | p:$6.6\cdot10^{-9}$ | p:$8.2\cdot10^{-15}$ | p:$3.1\cdot10^{-6}$ | p:$1.4\cdot10^{-12}$ |
| | *LDLRAD4* | R:-0.18 p:$3.1\cdot10^{-5}$ | p:$1.2\cdot10^{-5}$ | p:$1.2\cdot10^{-4}$ | p:$2.1\cdot10^{-4}$ | p=$5.2\cdot10^{-7}$ | p:$7.1\cdot10^{-8}$ |
| | *DGUOK* | R:0.45 p:$3.4\cdot10^{-26}$ | p:$2.5\cdot10^{-8}$ | p:$2.3\cdot10^{-4}$ | p:$2.2\cdot10^{-19}$ | p:$6.5\cdot10^{-7}$ | p:$3.0\cdot10^{-14}$ |
| | *IRX1* | R:0.18 p:$4.5\cdot10^{-5}$ | p:0.01 | p:0.02 | p:$2.8\cdot10^{-8}$ | p:$9.2\cdot10^{-3}$ | p:$1.0\cdot10^{-7}$ |
| TBX2 | *SPOCK2* | R:-0.29 p:$6.8\cdot10^{-11}$ | p:$1.7\cdot10^{-9}$ | p:$9.1\cdot10^{-9}$ | p:$3.9\cdot10^{-20}$ | p:$2.4\cdot10^{-19}$ | p:$8.2\cdot10^{-20}$ |

Among mutated aTFBSs target genes involved in neuronal development and differentiation, the Calcium Voltage-Gated Channel Auxiliary Subunit Beta 1 (*CACNB1*) gene was predicted to interact with a HAND2 aTFBS carrying a

[19] TF: Transcription Factor interacting with mutated aTFBSs
[20] TF corr: Correlation between target gene and specific CRC TF expression.
[21] OS: Overall Survival
[22] EFS: Event Free Survival
[23] *MYCN*$_a$ vs *MYCN*$_{na}$: Comparison of target genes expression levels between *MYCN* amplified and *MYCN* nonamplified patients
[24] St4 vs St1/2/3/4s: Comparison of target genes expression levels between Stage 4 and Stage 1, 2, 3 and 4s patients
[25] High R vs Low R: Comparison of target genes expression levels between High risk and Low risk patients

somatic SNV which decreased the affinity for this TF (FABIAN score: -0.38) (**Figure 20A**).

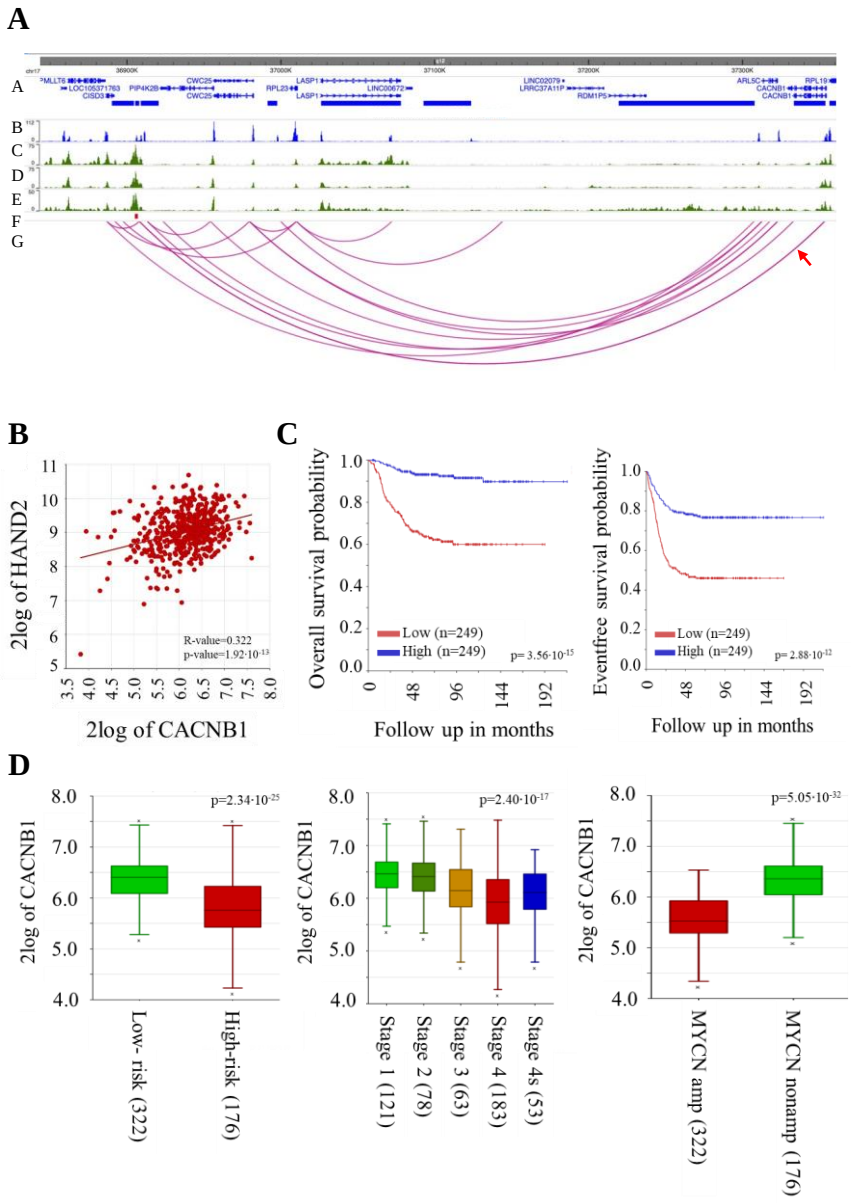


**Figure 20.** ***CACNB1*** **represents a putative target of a mutated HAND2 aTFBS and its downregulation is correlated with NB worse outcome. A,** Capture HiC results highlighting the interaction between mutated active HAND2 binding site and *CACNB1* promoter. From top to bottom, the plot tracks show the Refseq genes (A), interacting promoters in SHSY5Y (B), H3K27ac profiles in SKNBE2C (C), Kelly (D) and NGP (E), mutated aTFBSs (F) and the interaction arcs in SHSY5Y (G). In red and blue are highlighted interacting *CACNB1* promoter and HAND2 aTFBS, respectively. A red arrow indicates the arc connecting regions of interest. **B,** *CACNB1* versus *HAND2* expression correlation in the GSE62564 data set. **C,** EFS and OS based on *CACNB1* median expression in the GSE62564 data set. **D,** *CACNB1* expression in NB patients stratified by clinical features. From left to right, risk group, INSS stage and *MYCN* amplification investigating the GSE62564 data set.

Consistent with the role of HAND2 as transcriptional activator, I observed positive correlation between *CACNB1* and *HAND2* expression (GSE62564, R = 0.32, p-value = $1.9 \cdot 10^{-13}$) (**Figure 20B)**. Notably, the decreased expression of *CACNB1* was significantly associated with poor OS (p-value = $3.56 \cdot 10^{-15}$) and EFS (p-value = $2.88 \cdot 10^{-12}$) probabilities (**Figure 20C**). In addition, *CACNB1* low expression was correlated with bad NB prognosis markers, such as high-risk category (p-value = $2.34 \cdot 10^{-25}$), INSS stage (p-value = $2.40 \cdot 10^{-17}$) and MYCN amplification (p-value = $5.05 \cdot 10^{-32}$) (**Figure 20D**).

Another selected gene involved in neuronal differentiation and development was *HES6*. The Hes Family BHLH Transcription Factor 6 (*HES6*) gene was predicted to interact with a MYCN aTFBS, carrying a somatic SNV which increased the affinity for MYCN (FABIAN score: 0.18) (**Figure 21A**)

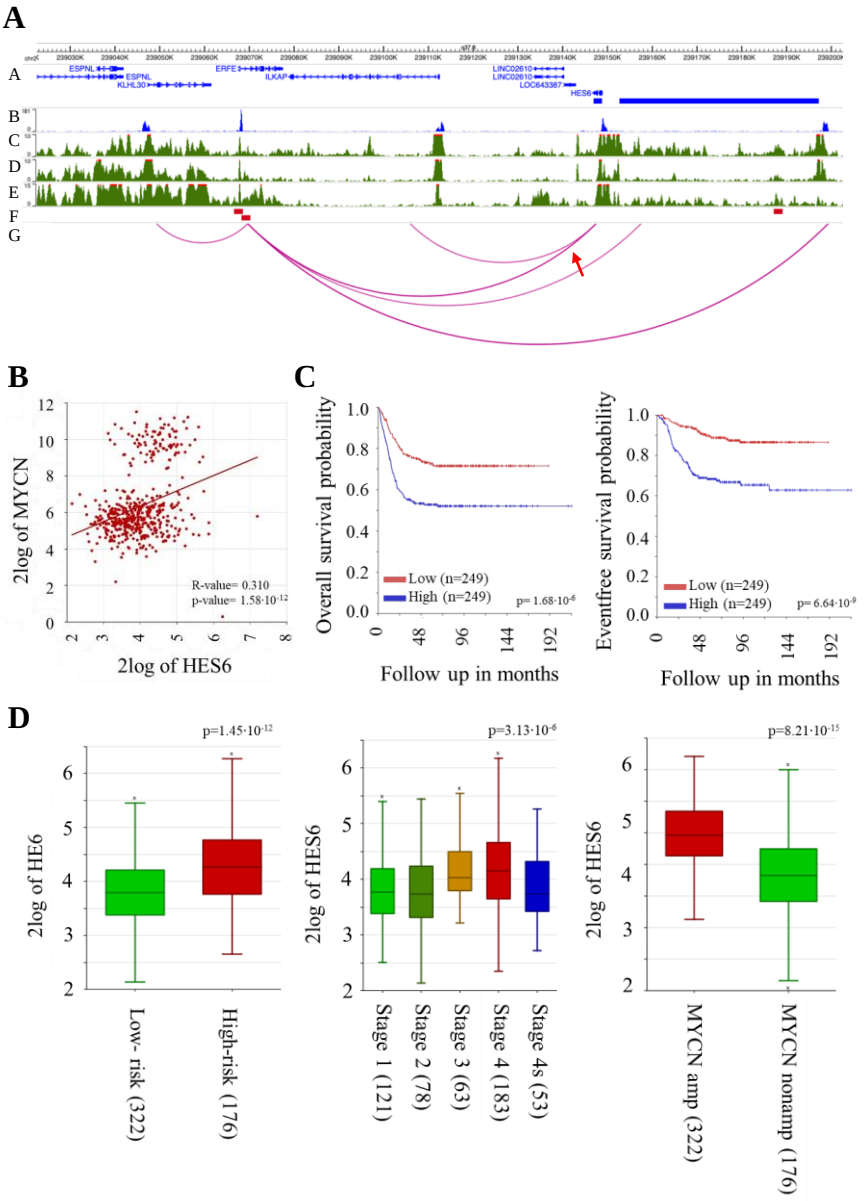


**Figure 21.** ***HES6*** **represents a putative target of a mutated MYCN aTFBS and its downregulation is correlated with NB worse outcome. A,** Capture HiC results highlighting the interaction between mutated active MYCN binding site and *HES6* promoter. From top to bottom, the plot tracks show the Refseq genes (A), interacting promoters in SKNBE2C (B), H3K27ac profiles in SKNBE2C (C), Kelly (D) and NGP (E), mutated aTFBSs (F) and the interaction arcs in SKNBE2C (G). In red and blue are highlighted interacting *HES6* promoter and HAND2 aTFBS, respectively. A red arrow indicates the arc connecting regions of interest. **B,** *HES6* versus *MYCN* expression correlation in the GSE62564 data set. **C,** EFS and OS based on *HES6* median expression in the GSE62564 data set. **D,** *HES6* expression in NB patients stratified by clinical features. From left to right, risk group, INSS stage and MYCN amplification investigating the GSE62564 data set.

Consistent with the role of MYCN as transcriptional activator, I observed a positive correlation between *HES6* and *MYCN* expression (GSE62564, R = 0.31, p-value = $1.6 \cdot 10^{-12}$) (**Figure 21B)**. Highlighting a putative role of *HES6* in NB tumorigenesis, the increased expression of this gene was significantly associated with poor OS (p-value = $1.68 \cdot 10^{-6}$) and EFS (p-value = $6.64 \cdot 10^{-9}$) probabilities (**Figure 21C**) and poor outcome (high-risk vs. non-high-risk tumors, p-value = $1.45 \cdot 10^{-12}$; Stage 4 vs. non-Stage 4 tumors, p-value = $3.13 \cdot 10^{-6}$; and MYCN-amplified vs. MYCN-non-amplified, p-value = $8.21 \cdot 10^{-15}$) (**Figure 21D**). Altogether, these results suggest that NB tumorigenesis may be triggered by perturbing the binding of ADRN and MES CRCs TFs and, in turn, the regulation of genes involved in processes of cell development and differentiation and in pathways associated to the definition of neural-crest and mesenchymal cell identities.

# 5. Discussion

Challenges in treating high-risk NBs emerge from the development of therapy-resistances. This phenomenon is associated to NB intratumoral heterogeneity (ITH) that occurs in the shape of ADRN and MES NB cell identities imposed by ADRN and MES CRCs activity [43, 44, 46]. In this scenario, the investigation of biological and genetic mechanisms behind the definition of these two identities has become of primary importance.

In the context of NB, which is characterized by a paucity of coding somatic mutations, the noncoding genome can represent a novel reservoir of oncogenic drivers [55, 56, 83]. In particular, SNVs can potentially affect gene regulation, contributing to cancer onset. However, distinguishing noncoding driver SNVs from the huge number of passenger mutations (due to the lower selective pressure of the noncoding genome) represents a challenging task. However, this limitation can be overcome by applying ad hoc computational strategies to prioritize somatic noncoding variants located in regulatory regions.

Based on what reported so far, the aim of my thesis was the application of a computational analysis to identify putative driver somatic SNVs which impact active binding sites of ADRN and MES CRC master TFs affecting the expression of genes involved in differentiation and developmental processes.

To detect putative cancer driver SNVs perturbing the function of CRC master TFs, I firstly identified aTFBSs in ADRN an MES NB cell lines, starting from ChIP-seq and ATAC-seq data. By comparing OCRs profiles between ADRN and MES NB cell lines, I observed the formation of identity-specific clusters, which is in line with the current knowledge about the differences in epigenomic and transcriptomic profiles between ADRN and MES identities [43, 46]. Of note, this

is the first time that differences between ADRN and MES phenotypes are highlighted in terms of chromatin accessibility profiles.

After the definition of ADRN and MES specific-aTFBSs sets, I performed a mutation enrichment analysis exploiting a reliable set of somatic SNVs obtained from WGS data of 397 NB patients belonging to three independent cohorts: EGA (accession no.: EGAD00001001687, EGAD00001006739 and EGAD00001006626), Target (accession no.: phs000218.v21.p7; project ID: #14831; ref. 12) and an in-house cohort. The mutation enrichment analysis highlighted that aTFBSs of 6 ADRN and MES CRC master TFs (GATA3, HAND2, MYCN, ISL1, TBX2 and FOSL2) were characterized by a significant high rate of mutations.

Interestingly, patients carrying SNVs enriched in CRC aTFBSs had a significant worse OS and EFS respect to patients without noncoding SNVs in these regulatory regions. In particular, as highlighted by Multivariate Cox proportional regression, a significant increase of patients' risk of death is correlated with both the mutational burden in CRC aTFBSs and *MYCN* amplification. Considering that MYCN is a stabilizer of ADRN CRC [3] and since the majority of the significantly mutated aTFBSs identified are involved in ADRN CRC, it is reasonable that *MYCN* amplification and the mutational burden can contribute together to the worse OS probability of NB patients.

Regarding the EFS analysis, although the increase patients' risk of relapses is also significantly correlated with the age at the diagnosis greater than 18 months, the association with the mutational burden in CRC aTFBSs remains the most significant. Overall, these results strongly support that noncoding mutations in CRC aTFBSs may have a potential prognostic value. However, future studies on independent sets of patients are required to confirm this hypothesis.

The application of ad hoc filtering criteria based on the predicted effects of SNVs on TF binding showed that 689 SNVs caused gains or losses of CRC TFs binding. Starting from the hypothesis that these SNVs, affecting the binding of CRC master TFs, can alter the expression of genes potentially involved in NB tumorigenesis, I identified putative target genes of mutated CRC aTFBSs, analysing CHiC data in 2 ADRN and 2 MES NB cell lines.

Among the putative targets of mutated aTFBSs, I found genes already known to be involved in NB tumorigenesis, such as *FGFR1*, *CTTNBP2*, *PHOX2B* and *MAML2*. *FGFR1* represents a potential therapeutic target in NB, as highlighted in one recent work of my research group [29].

In another our recent study, a cluster of somatic SNVs in the promoter region of *CTTNBP2* was predicted to alter the binding of NB tissue-specific TFs, reducing the expression of this gene [62], a phenomenon that was correlated with bad NB prognosis markers [62].

In the context of NB cell identity definition, *PHOX2B* and *MAML2* constitute two master TFs belonging to ADRN and MES CRCs, respectively. *PHOX2B* is regarded as a master TF implicated in neuronal development. Indeed, its absence results in the lack of the expression of other sympathetic nervous system-specific TFs, which causes the total loss of peripheral autonomic neurons [3]. On the other hand, *MAML2* represents a transcriptional activator of the NOTCH pathway and regulates target genes, including lineage-specific TFs, involved in cell fate decisions in the context of MES cell identity [50].

Overall, the identification of genes already known to be implicated in NB tumorigenesis and cell identity among the targets of mutated aTFBSs of master CRC TFs demonstrated the reliability of my bioinformatic approach to distinguish potential driver somatic noncoding SNVs from passenger ones.

GO analysis of mutated aTFBSs target genes highlighted enrichment of pathways related to neuronal development and differentiation, mirroring the ADRN phenotype. On the other hand, I also observed a significant enrichment of pathways related to specification of neural crest-like and mesenchymal proliferation, reflecting the MES phenotype.

Starting from the assumption that NB derives from the impairment of cell differentiation and developmental processes [31], I focused only on genes involved in cell differentiation and development-related pathways. In addition, I restricted further analysis only on 35 targets of aTFBSs carrying mutations which strongly impact CRC master TFs bindings. To identify novel genes potentially involved in NB tumorigenesis, I evaluated the expression of the 35 selected targets in correlation with NB prognosis markers, overall and event free survival exploiting RNA-seq data of a cohort of well annotated NB patients. I found 3 genes (*ROBO2, CACNB1, PIK3R1)* interacting with HAND2 aTFBSs, 5 MYCN aTFBSs targets *(MDGA1, HES6, LDLRAD4, DGUOK, IRX1)* and 1 target gene interacting with TBX2 aTFBS (*SPOCK2*) whose high or low expression significantly correlated with worse OS and EFS probabilities, MYCN amplification, Stage 4 and high-risk category in NB patients.

3 SNVs in HAND2 active binding sites were predicted to decrease the affinity for this TF and were found to interact with genes implicated in synapse organization, axon development, neuron projection guidance and mesenchyme development. *ROBO2* encoded a transmembrane receptor that exerts its activity by binding to the secreted ligand Slit. Activation of this receptor plays multiple roles during nervous system development, from axon targeting and synaptogenesis to cell migration [84-87]. Of note, *ROBO2* expression was reported to increase during treatment with 9-cis retinoic acid in NB cells, strongly supporting its involvement in NB cell differentiation [88]. This is in line with the

observed correlation between bad NB prognosis markers and *ROBO2* downregulation in NB patients.

*CACNB1* encoded the β1-subunit of human neuronal voltage-dependent calcium channels involved in processes of synaptic transmission and plasticity [89, 90]. Regarding its role in tumorigenesis, *CACNB1* upregulation was correlated with low overall survival and progression in ovarian and colorectal cancer [91, 92]. Conversely, I observed a significant association between *CACNB1* downregulation and the worst cases of NB, suggesting a novel role of this gene as potential tumor suppressor.

Finally, *PIK3R1* encodes the regulatory subunits of phosphoinositide 3-kinases (PI3Ks) enzymes which are essential during synapse and dendritic spine development [93-95]. Of note, dysregulated PI3K activity was associated to several mental disorders. Furthermore, in NB it was found highly expressed in favourable cases compared to aggressive ones [96]. Similar results emerged from the correlation between *PIK3R1* expression and markers of bad prognosis in 498 NB cases, confirming its role as tumor suppressor.

Regarding MYCN, aTFBSs interacting with *MDGA1* and *HES6* carried 2 SNVs predicted to increase affinity for this TF. On the other hand, 3 SNVs in the aTFBSs of this TF interacted with *LDLRAD4*, *DGUOK* and *IRX1* and were predicted to decrease MYCN affinity. These genes were implicated in different biological processes, such as synapse organization, pattern specification, neuron projection guidance, negative regulation of cell development and negative regulation of nervous system development. *MDGA1* plays a key role in development of neurons and encodes a synaptic adhesion molecule (the so-called MAM domain-containing glycosylphosphatidylinositol anchors) essential for driving the formation, maturation, and plasticity of synaptic connections [97].

Interestingly, *MDGA1* was identified as high susceptibility gene for several neurological disorders, such as Autism Spectrum Disorders, Bipolar Disorders and Schizophrenia [98-100]. Regarding its role in cancer, *MDGA1* knockdown was found to decrease renal cancer cell migration and proliferation [101]. In NB, I observed a lower *MDGA1* expression in the most severe cases, suggesting a dual role for this gene which may act as oncogene and tumor suppressor in a cancer type specific manner.

*HES6* encodes for a member of a subfamily of basic helix-loop-helix transcription repressors governing cell fate in neural development. Indeed, it was shown that *HES6* expression in undifferentiated cortical progenitors was sufficient to induce neuronal differentiation [102]. Moreover, numerous lines of evidences have already linked *HES6* increased expression with poor survival of advanced astrocytoma, glioblastoma, prostate cancer, and kidney cancer patients [103-105]. The same was observed in NB, where high *HES6* expression significantly correlated with bad NB prognostic markers, overall and event free survival of the patients.

*LDLRAD4* is known to negatively regulate TGF-β signaling thus playing a pivotal role in multiple cell processes such as proliferation, motility and differentiation [106]. Consistently with its implication in migration and proliferation processes, *LDLRAD4* high expression was previously implicated in the progression of hepatic cancer [107]. On the other hand, I observed that high-risk NB patients are characterized by *LDLRAD4* decreased mRNA levels, which could be linked to the role of this gene in cellular differentiation processes.

*DGUOK* encodes deoxyguanosine kinase, which is implicated in the phosphorylation of purine deoxyribonucleosides in the mitochondrial matrix, which is essential for mitochondrial DNA preservation. Recently, *DGUOK*

activity was linked to self-renewal of lung cancer cells, which is supported by *DGUOK* increased expression in patients with the worst outcome [108]. As emerged from GO analysis, *DGUOK* results also implicated in negative regulation of nervous system development, suggesting that its upregulation can perturb formation of neuronal cells. This is consistent with the correlation between *DGUOK* high expression and the worst cases of NB, which is known to be triggered by impaired neuronal differentiation.

*IRX1* is a TF belonging to the Iroquois homeobox TFs family, which govern development and cell differentiation in several tissues during embryonic development [109]. Several studies highlighted a dual tumorigenic role of this gene, which can act either as tumor suppressor or oncogene in a cancer type specific manner. Indeed, *IRX1* overexpression was found both to inhibit cancer cell growth, invasion, and tumorigenesis in gastric cancer [110] and drive tumor metastasis in osteosarcoma patients [111]. In NB, it may act as oncogene, as suggested by the high *IRX1* expression in high-risk NB cases.

The last SNV was predicted to decrease the affinity of TBX2 to its aTFBS interacting with *SPOCK2*, which is involved in synapse organization. It encodes a complex type of secreted proteoglycan that can inhibit neuronal cell attachment and neurite extension [112]. As the case of *IRX1*, also *SPOCK2* is reported to behave both as oncogene and tumor suppressor in different types of cancer. Indeed, *SPOCK2* high expression has been correlated both with tumor growth in ovarian cancer [113] and longer overall survival of pancreatic ductal adenocarcinoma patients [114]. I observed that *SPOCK2* decreased expression exacerbate NB outcome, thus suggesting a potential role as tumor suppressor of this gene in NB tumorigenesis.

All together these findings suggested that somatic noncoding SNVs impacting on binding of master CRC TFs, may potentially affect the expression of genes involved in neuronal cell development and differentiation processes, promoting tumorigenesis. However, computational approaches are not sufficient to truly demonstrate the connections between SNVs impact on CRC master TFs binding and deregulation of genes involved in such type of biological processes. For this reason, functional experiments will be conducted such as luciferase assays to compare differential report activity between the mutated and the wild-type allele of each selected CRC aTFBS, ChIP-qPCR with specific antibodies for CRC master TFs to validate the allele specific binding of CRC TFs and CRISPR-Cas9 gene-editing methodology to evaluate the direct effect of noncoding somatic mutations on the expression of target genes.

# 6. Conclusion

In this work, I used an alternative approach to detect and study regulatory cancer driver mutations, overcoming limitations in distinguishing putative somatic drivers from noncoding passengers and leading to identification of mutated regulatory regions that could have relevant implications in NB development.

What emerges is that SNVs impacting the binding for master CRC TFs can act synergically impairing biological processes of development and differentiation, promoting NB tumorigenesis. Prospectively, functional validation to ascertain the biological roles of this noncoding somatic mutations and their impact on gene regulation represents an important definitive task to address.

In conclusion, considering the heterogeneity of NB tumor environment, the identification of novel driver mutations and genes implicated in the definition of specific tumor cell phenotypes can represent a breakthrough in personalized therapy. This can lead to extend the therapeutic interventions targeting the whole spectrum of cellular identities that drive NB, thus finding successful treatments for tumors that are still resistant to differentiation therapy.

## 7. Aknowledgements

This study was supported by grants from Associazione Italiana per la Ricerca sul Cancro (AIRC) grant no. 25796 to M. Capasso, from Fondazione Italiana per la Lotta al Neuroblastoma to M. Capasso and from OPEN Associazione Oncologia Pediatrica e Neuroblastoma ONLUS to M. Capasso.

## 8. List of publications

1) Avitabile M, Bonfiglio F, **Aievola V**, Cantalupo S, Maiorino T, Lasorsa VA, Domenicotti C, Marengo B, Zbyněk H, Vojtěch A, Iolascon A, Capasso M. Single-cell transcriptomics of neuroblastoma identifies chemoresistance-associated genes and pathways. Comput Struct Biotechnol J. 2022 Aug 18;20:4437-4445.

2) Bonfiglio F, Lasorsa VA, Cantalupo S, D'Alterio G, **Aievola V**, Boccia A, Ardito M, Furini S, Renieri A, Morini M, Stainczyk S, Westermann F, Paolella G, Eva A, Iolascon A, Capasso M. Inherited rare variants in homologous recombination and neurodevelopmental genes are associated with increased risk of neuroblastoma. EBioMedicine. 2023 Jan;87:104395.

## 9. References

1 Johnsen, J.I., Dyberg, C., and Wickstrom, M.: ‘Neuroblastoma-A Neural Crest Derived Embryonal Malignancy’, Front Mol Neurosci, 2019, 12, pp. 9

2 Matthay, K.K., Maris, J.M., Schleiermacher, G., Nakagawara, A., Mackall, C.L., Diller, L., and Weiss, W.A.: ‘Neuroblastoma’, Nat Rev Dis Primers, 2016, 2, pp. 16078

3 Ponzoni, M., Bachetti, T., Corrias, M.V., Brignole, C., Pastorino, F., Calarco, E., Bensa, V., Giusto, E., Ceccherini, I., and Perri, P.: ‘Recent advances in the developmental origin of neuroblastoma: an overview’, J Exp Clin Cancer Res, 2022, 41, (1), pp. 92

4 Qiu, B., and Matthay, K.K.: ‘Advancing therapy for neuroblastoma’, Nat Rev Clin Oncol, 2022, 19, (8), pp. 515-533

5 Ratner, N., Brodeur, G.M., Dale, R.C., and Schor, N.F.: ‘The "neuro" of neuroblastoma: Neuroblastoma as a neurodevelopmental disorder’, Ann Neurol, 2016, 80, (1), pp. 13-23

6 Roy Choudhury, S., Karmakar, S., Banik, N.L., and Ray, S.K.: ‘Targeting angiogenesis for controlling neuroblastoma’, J Oncol, 2012, 2012, pp. 782020

7 Voute, P.A., van der Kleij, A.J., De Kraker, J., Hoefnagel, C.A., Tiel-van Buul, M.M., and Van Gennip, H.: ‘Clinical experience with radiation enhancement by hyperbaric oxygen in children with recurrent neuroblastoma stage IV’, Eur J Cancer, 1995, 31A, (4), pp. 596-600

8 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., Mosseri, V., Simon, T., Garaventa, A., Castel, V., Matthay, K.K., and Force, I.T.: ‘The International

Neuroblastoma Risk Group (INRG) classification system: an INRG Task Force report', J Clin Oncol, 2009, 27, (2), pp. 289-297

9 Capasso, M., and Diskin, S.J.: 'Genetics and genomics of neuroblastoma', Cancer Treat Res, 2010, 155, pp. 65-84

10 Capasso, M., Montella, A., Tirelli, M., Maiorino, T., Cantalupo, S., and Iolascon, A.: 'Genetic Predisposition to Solid Pediatric Cancers', Front Oncol, 2020, 10, pp. 590033

11 Ritenour, L.E., Randall, M.P., Bosse, K.R., and Diskin, S.J.: 'Genetic susceptibility to neuroblastoma: current knowledge and future directions', Cell Tissue Res, 2018, 372, (2), pp. 287-307

12 Bachetti, T., and Ceccherini, I.: 'Causative and common PHOX2B variants define a broad phenotypic spectrum', Clin Genet, 2020, 97, (1), pp. 103-113

13 Mosse, Y.P., Laudenslager, M., Longo, L., Cole, K.A., Wood, A., Attiyeh, E.F., Laquaglia, M.J., Sennett, R., Lynch, J.E., Perri, P., Laureys, G., Speleman, F., Kim, C., Hou, C., Hakonarson, H., Torkamani, A., Schork, N.J., Brodeur, G.M., Tonini, G.P., Rappaport, E., Devoto, M., and Maris, J.M.: 'Identification of ALK as a major familial neuroblastoma predisposition gene', Nature, 2008, 455, (7215), pp. 930-935

14 Bresler, S.C., Weiser, D.A., Huwe, P.J., Park, J.H., Krytska, K., Ryles, H., Laudenslager, M., Rappaport, E.F., Wood, A.C., McGrady, P.W., Hogarty, M.D., London, W.B., Radhakrishnan, R., Lemmon, M.A., and Mosse, Y.P.: 'ALK mutations confer differential oncogenic activation and sensitivity to ALK inhibition therapy in neuroblastoma', Cancer Cell, 2014, 26, (5), pp. 682-694

15 Bosse, K.R., Diskin, S.J., Cole, K.A., Wood, A.C., Schnepp, R.W., Norris, G., Nguyen le, B., Jagannathan, J., Laquaglia, M., Winter, C., Diamond, M., Hou, C., Attiyeh, E.F., Mosse, Y.P., Pineros, V., Dizin, E., Zhang, Y., Asgharzadeh, S., Seeger, R.C., Capasso, M., Pawel, B.R., Devoto, M., Hakonarson, H., Rappaport, E.F., Irminger-Finger, I., and Maris, J.M.: 'Common variation at BARD1 results in the expression of an oncogenic isoform that influences neuroblastoma susceptibility and oncogenicity', Cancer Res, 2012, 72, (8), pp. 2068-2078

16 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., Bosse, K., Diamond, M., Laudenslager, M., Winter, C., Bradfield, J.P., Scott, R.H., Jagannathan, J., Garris, M., McConville, C., London, W.B., Seeger, R.C., Grant, S.F., Li, H., Rahman, N., Rappaport, E., Hakonarson, H., and Maris, J.M.: 'Common variations in BARD1 influence susceptibility to high-risk neuroblastoma', Nat Genet, 2009, 41, (6), pp. 718-723

17 Capasso, M., Diskin, S.J., Totaro, F., Longo, L., De Mariano, M., Russo, R., Cimmino, F., Hakonarson, H., Tonini, G.P., Devoto, M., Maris, J.M., and Iolascon, A.: '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, (3), pp. 605-611

18 Wang, K., Diskin, S.J., Zhang, H., Attiyeh, E.F., Winter, C., Hou, C., Schnepp, R.W., Diamond, M., Bosse, K., Mayes, P.A., Glessner, J., Kim, C., Frackelton, E., Garris, M., Wang, Q., Glaberson, W., Chiavacci, R., Nguyen, L., Jagannathan, J., Saeki, N., Sasaki, H., Grant, S.F., Iolascon, A., Mosse, Y.P., Cole, K.A., Li, H., Devoto, M., McGrady, P.W., London, W.B., Capasso, M.,

Rahman, N., Hakonarson, H., and Maris, J.M.: ‘Integrative genomics identifies LMO1 as a neuroblastoma oncogene’, Nature, 2011, 469, (7329), pp. 216-220

19 Diskin, S.J., Capasso, M., Schnepp, R.W., Cole, K.A., Attiyeh, E.F., Hou, C., Diamond, M., Carpenter, E.L., Winter, C., Lee, H., Jagannathan, J., Latorre, V., Iolascon, A., Hakonarson, H., Devoto, M., and Maris, J.M.: ‘Common variation at 6q16 within HACE1 and LIN28B influences susceptibility to neuroblastoma’, Nat Genet, 2012, 44, (10), pp. 1126-1130

20 Tonini, G.P., and Capasso, M.: ‘Genetic predisposition and chromosome instability in neuroblastoma’, Cancer Metastasis Rev, 2020, 39, (1), pp. 275-285

21 Cimmino, F., Avitabile, M., Diskin, S.J., Vaksman, Z., Pignataro, P., Formicola, D., Cardinale, A., Testori, A., Koster, J., de Torres, C., Devoto, M., Maris, J.M., Iolascon, A., and Capasso, M.: ‘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, (11), pp. 2828-2837

22 Avitabile, M., Lasorsa, V.A., Cantalupo, S., Cardinale, A., Cimmino, F., Montella, A., Capasso, D., Haupt, R., Amoroso, L., Garaventa, A., Quattrone, A., Corrias, M.V., Iolascon, A., and Capasso, M.: ‘Association of PARP1 polymorphisms with response to chemotherapy in patients with high-risk neuroblastoma’, J Cell Mol Med, 2020, 24, (7), pp. 4072-4081

23 Avitabile, M., Succoio, M., Testori, A., Cardinale, A., Vaksman, Z., Lasorsa, V.A., Cantalupo, S., Esposito, M., Cimmino, F., Montella, A., Formicola, D., Koster, J., Andreotti, V., Ghiorzo, P., Romano, M.F., Staibano, S., Scalvenzi, M., Ayala, F., Hakonarson, H., Corrias, M.V., Devoto, M., Law, M.H., Iles, M.M., Brown, K., Diskin, S., Zambrano, N., Iolascon, A., and 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, 41, (3), pp. 284-295

24 Lasorsa, V.A., Cimmino, F., Ognibene, M., Mazzocco, K., Erminio, G., Morini, M., Conte, M., Iolascon, A., Pezzolo, A., and Capasso, M.: '19p loss is significantly enriched in older age neuroblastoma patients and correlates with poor prognosis', NPJ Genom Med, 2020, 5, pp. 18

25 Esposito, M.R., Binatti, A., Pantile, M., Coppe, A., Mazzocco, K., Longo, L., Capasso, M., Lasorsa, V.A., Luksch, R., Bortoluzzi, S., and Tonini, G.P.: '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, (10), pp. 2525-2536

26 Lasorsa, V.A., Formicola, D., Pignataro, P., Cimmino, F., Calabrese, F.M., Mora, J., Esposito, M.R., Pantile, M., Zanon, C., De Mariano, M., Longo, L., Hogarty, M.D., de Torres, C., Tonini, G.P., Iolascon, A., and Capasso, M.: 'Exome and deep sequencing of clinically aggressive neuroblastoma reveal somatic mutations that affect key pathways involved in cancer progression', Oncotarget, 2016, 7, (16), pp. 21840-21852

27 Pugh, T.J., Morozova, O., Attiyeh, E.F., Asgharzadeh, S., Wei, J.S., Auclair, D., Carter, S.L., Cibulskis, K., Hanna, M., Kiezun, A., Kim, J., Lawrence, M.S., Lichenstein, L., McKenna, A., Pedamallu, C.S., Ramos, A.H., Shefler, E., Sivachenko, A., Sougnez, C., Stewart, C., Ally, A., Birol, I., Chiu, R., Corbett, R.D., Hirst, M., Jackman, S.D., Kamoh, B., Khodabakshi, A.H., Krzywinski, M., Lo, A., Moore, R.A., Mungall, K.L., Qian, J., Tam, A., Thiessen, N., Zhao, Y., Cole, K.A., Diamond, M., Diskin, S.J., Mosse, Y.P., Wood, A.C., Ji, L., Sposto, R., Badgett, T., London, W.B., Moyer, Y., Gastier-Foster, J.M., Smith, M.A., Guidry Auvil, J.M., Gerhard, D.S., Hogarty, M.D., Jones, S.J.,

Lander, E.S., Gabriel, S.B., Getz, G., Seeger, R.C., Khan, J., Marra, M.A., Meyerson, M., and Maris, J.M.: 'The genetic landscape of high-risk neuroblastoma', Nat Genet, 2013, 45, (3), pp. 279-284

28 Peifer, M., Hertwig, F., Roels, F., Dreidax, D., Gartlgruber, M., Menon, R., Kramer, A., Roncaioli, J.L., Sand, F., Heuckmann, J.M., Ikram, F., Schmidt, R., Ackermann, S., Engesser, A., Kahlert, Y., Vogel, W., Altmuller, J., Nurnberg, P., Thierry-Mieg, J., Thierry-Mieg, D., Mariappan, A., Heynck, S., Mariotti, E., Henrich, K.O., Gloeckner, C., Bosco, G., Leuschner, I., Schweiger, M.R., Savelyeva, L., Watkins, S.C., Shao, C., Bell, E., Hofer, T., Achter, V., Lang, U., Theissen, J., Volland, R., Saadati, M., Eggert, A., de Wilde, B., Berthold, F., Peng, Z., Zhao, C., Shi, L., Ortmann, M., Buttner, R., Perner, S., Hero, B., Schramm, A., Schulte, J.H., Herrmann, C., O'Sullivan, R.J., Westermann, F., Thomas, R.K., and Fischer, M.: 'Telomerase activation by genomic rearrangements in high-risk neuroblastoma', Nature, 2015, 526, (7575), pp. 700-704

29 Cimmino, F., Montella, A., Tirelli, M., Avitabile, M., Lasorsa, V.A., Visconte, F., Cantalupo, S., Maiorino, T., De Angelis, B., Morini, M., Castellano, A., Locatelli, F., Capasso, M., and Iolascon, A.: 'FGFR1 is a potential therapeutic target in neuroblastoma', Cancer Cell Int, 2022, 22, (1), pp. 174

30 Esposito, M.R., Aveic, S., Seydel, A., and Tonini, G.P.: 'Neuroblastoma treatment in the post-genomic era', J Biomed Sci, 2017, 24, (1), pp. 14

31 Zeineldin, M., Patel, A.G., and Dyer, M.A.: 'Neuroblastoma: When differentiation goes awry', Neuron, 2022, 110, (18), pp. 2916-2928

32 Biedler, J.L., Helson, L., and Spengler, B.A.: 'Morphology and growth, tumorigenicity, and cytogenetics of human neuroblastoma cells in continuous culture', Cancer Res, 1973, 33, (11), pp. 2643-2652

33 Ciccarone, V., Spengler, B.A., Meyers, M.B., Biedler, J.L., and Ross, R.A.: 'Phenotypic diversification in human neuroblastoma cells: expression of distinct neural crest lineages', Cancer Res, 1989, 49, (1), pp. 219-225

34 Ross, R.A., and Biedler, J.L.: 'Presence and regulation of tyrosinase activity in human neuroblastoma cell variants in vitro', Cancer Res, 1985, 45, (4), pp. 1628-1632

35 Gautier, M., Thirant, C., Delattre, O., and Janoueix-Lerosey, I.: 'Plasticity in Neuroblastoma Cell Identity Defines a Noradrenergic-to-Mesenchymal Transition (NMT)', Cancers (Basel), 2021, 13, (12)

36 DeClerck, Y.A., and Lee, C.: 'Collagen biosynthesis in human neuroblastoma cell lines: evidence for expression of glial cell properties', J Natl Cancer Inst, 1985, 75, (3), pp. 431-439

37 DeClerck, Y.A., Bomann, E.T., Spengler, B.A., and Biedler, J.L.: 'Differential collagen biosynthesis by human neuroblastoma cell variants', Cancer Res, 1987, 47, (24 Pt 1), pp. 6505-6510

38 Shendy, N.A.M., Zimmerman, M.W., Abraham, B.J., and Durbin, A.D.: 'Intrinsic transcriptional heterogeneity in neuroblastoma guides mechanistic and therapeutic insights', Cell Rep Med, 2022, 3, (5), pp. 100632

39 Takahashi, K., and Yamanaka, S.: 'Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors', Cell, 2006, 126, (4), pp. 663-676

40 Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., Gifford, D.K., Melton, D.A., Jaenisch, R., and Young, R.A.: 'Core transcriptional regulatory circuitry in human embryonic stem cells', Cell, 2005, 122, (6), pp. 947-956

41 Saint-Andre, V., Federation, A.J., Lin, C.Y., Abraham, B.J., Reddy, J., Lee, T.I., Bradner, J.E., and Young, R.A.: 'Models of human core transcriptional regulatory circuitries', Genome Res, 2016, 26, (3), pp. 385-396

42 Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke, H.A., and Young, R.A.: 'Super-enhancers in the control of cell identity and disease', Cell, 2013, 155, (4), pp. 934-947

43 Boeva, V., Louis-Brennetot, C., Peltier, A., Durand, S., Pierre-Eugene, C., Raynal, V., Etchevers, H.C., Thomas, S., Lermine, A., Daudigeos-Dubus, E., Geoerger, B., Orth, M.F., Grunewald, T.G.P., Diaz, E., Ducos, B., Surdez, D., Carcaboso, A.M., Medvedeva, I., Deller, T., Combaret, V., Lapouble, E., Pierron, G., Grossetete-Lalami, S., Baulande, S., Schleiermacher, G., Barillot, E., Rohrer, H., Delattre, O., and Janoueix-Lerosey, I.: 'Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries', Nat Genet, 2017, 49, (9), pp. 1408-1413

44 Gartlgruber, M., Sharma, A.K., Quintero, A., Dreidax, D., Jansky, S., Park, Y.G., Kreth, S., Meder, J., Doncevic, D., Saary, P., Toprak, U.H., Ishaque, N., Afanasyeva, E., Wecht, E., Koster, J., Versteeg, R., Grunewald, T.G.P., Jones, D.T.W., Pfister, S.M., Henrich, K.O., van Nes, J., Herrmann, C., and Westermann, F.: 'Super enhancers define regulatory subtypes and cell identity in neuroblastoma', Nat Cancer, 2021, 2, (1), pp. 114-128

45 Jansky, S., Sharma, A.K., Korber, V., Quintero, A., Toprak, U.H., Wecht, E.M., Gartlgruber, M., Greco, A., Chomsky, E., Grunewald, T.G.P., Henrich, K.O., Tanay, A., Herrmann, C., Hofer, T., and Westermann, F.: 'Single-cell transcriptomic analyses provide insights into the developmental origins of neuroblastoma', Nat Genet, 2021, 53, (5), pp. 683-693

46 van Groningen, T., Koster, J., Valentijn, L.J., Zwijnenburg, D.A., Akogul, N., Hasselt, N.E., Broekmans, M., Haneveld, F., Nowakowska, N.E., Bras, J., van Noesel, C.J.M., Jongejan, A., van Kampen, A.H., Koster, L., Baas, F., van Dijk-Kerkhoven, L., Huizer-Smit, M., Lecca, M.C., Chan, A., Lakeman, A., Molenaar, P., Volckmann, R., Westerhout, E.M., Hamdi, M., van Sluis, P.G., Ebus, M.E., Molenaar, J.J., Tytgat, G.A., Westerman, B.A., van Nes, J., and Versteeg, R.: 'Neuroblastoma is composed of two super-enhancer-associated differentiation states', Nat Genet, 2017, 49, (8), pp. 1261-1266

47 Durbin, A.D., Zimmerman, M.W., Dharia, N.V., Abraham, B.J., Iniguez, A.B., Weichert-Leahey, N., He, S., Krill-Burger, J.M., Root, D.E., Vazquez, F., Tsherniak, A., Hahn, W.C., Golub, T.R., Young, R.A., Look, A.T., and Stegmaier, K.: 'Selective gene dependencies in MYCN-amplified neuroblastoma include the core transcriptional regulatory circuitry', Nat Genet, 2018, 50, (9), pp. 1240-1246

48 Decaesteker, B., Denecker, G., Van Neste, C., Dolman, E.M., Van Loocke, W., Gartlgruber, M., Nunes, C., De Vloed, F., Depuydt, P., Verboom, K., Rombaut, D., Loontiens, S., De Wyn, J., Kholosy, W.M., Koopmans, B., Essing, A.H.W., Herrmann, C., Dreidax, D., Durinck, K., Deforce, D., Van Nieuwerburgh, F., Henssen, A., Versteeg, R., Boeva, V., Schleiermacher, G., van Nes, J., Mestdagh, P., Vanhauwaert, S., Schulte, J.H., Westermann, F., Molenaar, J.J., De Preter, K., and Speleman, F.: 'TBX2 is a neuroblastoma core regulatory circuitry component enhancing MYCN/FOXM1 reactivation of DREAM targets', Nat Commun, 2018, 9, (1), pp. 4866

49 Zeid, R., Lawlor, M.A., Poon, E., Reyes, J.M., Fulciniti, M., Lopez, M.A., Scott, T.G., Nabet, B., Erb, M.A., Winter, G.E., Jacobson, Z., Polaski, D.R., Karlin, K.L., Hirsch, R.A., Munshi, N.P., Westbrook, T.F., Chesler, L., Lin, C.Y.,

and Bradner, J.E.: ‘Enhancer invasion shapes MYCN-dependent transcriptional amplification in neuroblastoma’, Nat Genet, 2018, 50, (4), pp. 515-523

50 van Groningen, T., Akogul, N., Westerhout, E.M., Chan, A., Hasselt, N.E., Zwijnenburg, D.A., Broekmans, M., Stroeken, P., Haneveld, F., Hooijer, G.K.J., Savci-Heijink, C.D., Lakeman, A., Volckmann, R., van Sluis, P., Valentijn, L.J., Koster, J., Versteeg, R., and van Nes, J.: ‘A NOTCH feed-forward loop drives reprogramming from adrenergic to mesenchymal state in neuroblastoma’, Nat Commun, 2019, 10, (1), pp. 1530

51 Yu, E.Y., Zahid, S.S., Aloe, S., Falck-Pedersen, E., Zhou, X.K., Cheung, N.V., and Lue, N.F.: ‘Reciprocal impacts of telomerase activity and ADRN/MES differentiation state in neuroblastoma tumor biology’, Commun Biol, 2021, 4, (1), pp. 1315

52 Khazeem, M.M., Casement, J.W., Schlossmacher, G., Kenneth, N.S., Sumbung, N.K., Chan, J.Y.T., McGow, J.F., Cowell, I.G., and Austin, C.A.: ‘TOP2B Is Required to Maintain the Adrenergic Neural Phenotype and for ATRA-Induced Differentiation of SH-SY5Y Neuroblastoma Cells’, Mol Neurobiol, 2022, 59, (10), pp. 5987-6008

53 D'Amico, S., Tempora, P., Gragera, P., Krol, K., Melaiu, O., De Ioris, M.A., Locatelli, F., and Fruci, D.: ‘Two bullets in the gun: combining immunotherapy with chemotherapy to defeat neuroblastoma by targeting adrenergic-mesenchymal plasticity’, Front Immunol, 2023, 14, pp. 1268645

54 Avitabile, M., Bonfiglio, F., Aievola, V., Cantalupo, S., Maiorino, T., Lasorsa, V.A., Domenicotti, C., Marengo, B., Zbynek, H., Vojtech, A., Iolascon, A., and Capasso, M.: ‘Single-cell transcriptomics of neuroblastoma identifies chemoresistance-associated genes and pathways’, Comput Struct Biotechnol J, 2022, 20, pp. 4437-4445

55 Tomczak, K., Czerwinska, P., and Wiznerowicz, M.: 'The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge', Contemp Oncol (Pozn), 2015, 19, (1A), pp. A68-77

56 Consortium, I.T.P.-C.A.o.W.G.: 'Pan-cancer analysis of whole genomes', Nature, 2020, 578, (7793), pp. 82-93

57 Sweet-Cordero, E.A., and Biegel, J.A.: 'The genomic landscape of pediatric cancers: Implications for diagnosis and treatment', Science, 2019, 363, (6432), pp. 1170-1175

58 Spielmann, M., and Mundlos, S.: 'Looking beyond the genes: the role of non-coding variants in human disease', Hum Mol Genet, 2016, 25, (R2), pp. R157-R165

59 Gan, K.A., Carrasco Pro, S., Sewell, J.A., and Fuxman Bass, J.I.: 'Identification of Single Nucleotide Non-coding Driver Mutations in Cancer', Front Genet, 2018, 9, pp. 16

60 Consortium, E.P.: 'An integrated encyclopedia of DNA elements in the human genome', Nature, 2012, 489, (7414), pp. 57-74

61 Lasorsa, V.A., Montella, A., Cantalupo, S., Tirelli, M., de Torres, C., Aveic, S., Tonini, G.P., Iolascon, A., and Capasso, M.: 'Somatic Mutations Enriched in Cis-Regulatory Elements Affect Genes Involved in Embryonic Development and Immune System Response in Neuroblastoma', Cancer Res, 2022, 82, (7), pp. 1193-1207

62 Capasso, M., Lasorsa, V.A., Cimmino, F., Avitabile, M., Cantalupo, S., Montella, A., De Angelis, B., Morini, M., de Torres, C., Castellano, A., Locatelli, F., and Iolascon, A.: 'Transcription Factors Involved in Tumorigenesis Are Over-

Represented in Mutated Active DNA-Binding Sites in Neuroblastoma', Cancer Res, 2020, 80, (3), pp. 382-393

63 Zimmerman, M.W., Durbin, A.D., He, S., Oppel, F., Shi, H., Tao, T., Li, Z., Berezovskaya, A., Liu, Y., Zhang, J., Young, R.A., Abraham, B.J., and Look, A.T.: 'Retinoic acid rewires the adrenergic core regulatory circuitry of childhood neuroblastoma', Sci Adv, 2021, 7, (43), pp. eabe0834

64 Upton, K., Modi, A., Patel, K., Kendsersky, N.M., Conkrite, K.L., Sussman, R.T., Way, G.P., Adams, R.N., Sacks, G.I., Fortina, P., Diskin, S.J., Maris, J.M., and Rokita, J.L.: 'Epigenomic profiling of neuroblastoma cell lines', Sci Data, 2020, 7, (1), pp. 116

65 Wang, L., Tan, T.K., Durbin, A.D., Zimmerman, M.W., Abraham, B.J., Tan, S.H., Ngoc, P.C.T., Weichert-Leahey, N., Akahane, K., Lawton, L.N., Rokita, J.L., Maris, J.M., Young, R.A., Look, A.T., and Sanda, T.: 'ASCL1 is a MYCN- and LMO1-dependent member of the adrenergic neuroblastoma core regulatory circuitry', Nat Commun, 2019, 10, (1), pp. 5622

66 Oldridge, D.A., Wood, A.C., Weichert-Leahey, N., Crimmins, I., Sussman, R., Winter, C., McDaniel, L.D., Diamond, M., Hart, L.S., Zhu, S., Durbin, A.D., Abraham, B.J., Anders, L., Tian, L., Zhang, S., Wei, J.S., Khan, J., Bramlett, K., Rahman, N., Capasso, M., Iolascon, A., Gerhard, D.S., Guidry Auvil, J.M., Young, R.A., Hakonarson, H., Diskin, S.J., Look, A.T., and Maris, J.M.: 'Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism', Nature, 2015, 528, (7582), pp. 418-421

67 Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso, P., and Nahnsen, S.: 'The nf-core framework for community-curated bioinformatics pipelines', Nat Biotechnol, 2020, 38, (3), pp. 276-278

68 Hulselmans, F.K.F.J.P.E.E.A.M.W.B.S.-B.G.: 'FelixKrueger/TrimGalore: v0.6.10', 2023

69 Li, H., and Durbin, R.: 'Fast and accurate short read alignment with Burrows-Wheeler transform', Bioinformatics, 2009, 25, (14), pp. 1754-1760

70 Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S.: 'Model-based analysis of ChIP-Seq (MACS)', Genome Biol, 2008, 9, (9), pp. R137

71 Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., Chen, Y., DeSalvo, G., Epstein, C., Fisher-Aylor, K.I., Euskirchen, G., Gerstein, M., Gertz, J., Hartemink, A.J., Hoffman, M.M., Iyer, V.R., Jung, Y.L., Karmakar, S., Kellis, M., Kharchenko, P.V., Li, Q., Liu, T., Liu, X.S., Ma, L., Milosavljevic, A., Myers, R.M., Park, P.J., Pazin, M.J., Perry, M.D., Raha, D., Reddy, T.E., Rozowsky, J., Shoresh, N., Sidow, A., Slattery, M., Stamatoyannopoulos, J.A., Tolstorukov, M.Y., White, K.P., Xi, S., Farnham, P.J., Lieb, J.D., Wold, B.J., and Snyder, M.: 'ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia', Genome Res, 2012, 22, (9), pp. 1813-1831

72 Yang, Y., Fear, J., Hu, J., Haecker, I., Zhou, L., Renne, R., Bloom, D., and McIntyre, L.M.: 'Leveraging biological replicates to improve analysis in ChIP-seq experiments', Comput Struct Biotechnol J, 2014, 9, pp. e201401002

73 Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K.: 'Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities', Mol Cell, 2010, 38, (4), pp. 576-589

74 Kim, S., Scheffler, K., Halpern, A.L., Bekritsky, M.A., Noh, E., Kallberg, M., Chen, X., Kim, Y., Beyter, D., Krusche, P., and Saunders, C.T.: 'Strelka2: fast and accurate calling of germline and somatic variants', Nat Methods, 2018, 15, (8), pp. 591-594

75 Amemiya, H.M., Kundaje, A., and Boyle, A.P.: 'The ENCODE Blacklist: Identification of Problematic Regions of the Genome', Sci Rep, 2019, 9, (1), pp. 9354

76 Derrien, T., Estelle, J., Marco Sola, S., Knowles, D.G., Raineri, E., Guigo, R., and Ribeca, P.: 'Fast computation and applications of genome mappability', PLoS One, 2012, 7, (1), pp. e30377

77 Elliott, K., and Larsson, E.: 'Non-coding driver mutations in human cancer', Nat Rev Cancer, 2021, 21, (8), pp. 500-509

78 Liang, Y., Gronbaek, C., Fariselli, P., and Krogh, A.: 'Correction to: Context dependency of nucleotide probabilities and variants in human DNA', BMC Genomics, 2022, 23, (1), pp. 356

79 Steinhaus, R., Robinson, P.N., and Seelow, D.: 'FABIAN-variant: predicting the effects of DNA variants on transcription factor binding', Nucleic Acids Res, 2022, 50, (W1), pp. W322-W329

80 Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., and Andrews, S.: 'HiCUP: pipeline for mapping and processing Hi-C data', F1000Res, 2015, 4, pp. 1310

81 Freire-Pritchett, P., Ray-Jones, H., Della Rosa, M., Eijsbouts, C.Q., Orchard, W.R., Wingett, S.W., Wallace, C., Cairns, J., Spivakov, M., and Malysheva, V.: 'Detecting chromosomal interactions in Capture Hi-C data with CHiCAGO and companion tools', Nat Protoc, 2021, 16, (9), pp. 4144-4176

82 Zhang, B., Kirov, S., and Snoddy, J.: 'WebGestalt: an integrated system for exploring gene sets in various biological contexts', Nucleic Acids Res, 2005, 33, (Web Server issue), pp. W741-748

83 Rheinbay, E., Nielsen, M.M., Abascal, F., Wala, J.A., Shapira, O., Tiao, G., Hornshoj, H., Hess, J.M., Juul, R.I., Lin, Z., Feuerbach, L., Sabarinathan, R., Madsen, T., Kim, J., Mularoni, L., Shuai, S., Lanzos, A., Herrmann, C., Maruvka, Y.E., Shen, C., Amin, S.B., Bandopadhayay, P., Bertl, J., Boroevich, K.A., Busanovich, J., Carlevaro-Fita, J., Chakravarty, D., Chan, C.W.Y., Craft, D., Dhingra, P., Diamanti, K., Fonseca, N.A., Gonzalez-Perez, A., Guo, Q., Hamilton, M.P., Haradhvala, N.J., Hong, C., Isaev, K., Johnson, T.A., Juul, M., Kahles, A., Kahraman, A., Kim, Y., Komorowski, J., Kumar, K., Kumar, S., Lee, D., Lehmann, K.V., Li, Y., Liu, E.M., Lochovsky, L., Park, K., Pich, O., Roberts, N.D., Saksena, G., Schumacher, S.E., Sidiropoulos, N., Sieverling, L., Sinnott-Armstrong, N., Stewart, C., Tamborero, D., Tubio, J.M.C., Umer, H.M., Uuskula-Reimand, L., Wadelius, C., Wadi, L., Yao, X., Zhang, C.Z., Zhang, J., Haber, J.E., Hobolth, A., Imielinski, M., Kellis, M., Lawrence, M.S., von Mering, C., Nakagawa, H., Raphael, B.J., Rubin, M.A., Sander, C., Stein, L.D., Stuart, J.M., Tsunoda, T., Wheeler, D.A., Johnson, R., Reimand, J., Gerstein, M., Khurana, E., Campbell, P.J., Lopez-Bigas, N., Drivers, P., Functional Interpretation Working, G., Group, P.S.V.W., Weischenfeldt, J., Beroukhim, R., Martincorena, I., Pedersen, J.S., Getz, G., and Consortium, P.: 'Analyses of non-coding somatic drivers in 2,658 cancer whole genomes', Nature, 2020, 578, (7793), pp. 102-111

84 Campbell, D.S., Stringham, S.A., Timm, A., Xiao, T., Law, M.Y., Baier, H., Nonet, M.L., and Chien, C.B.: 'Slit1a inhibits retinal ganglion cell arborization and synaptogenesis via Robo2-dependent and -independent pathways', Neuron, 2007, 55, (2), pp. 231-245

85 Cho, J.H., Lepine, M., Andrews, W., Parnavelas, J., and Cloutier, J.F.: 'Requirement for Slit-1 and Robo-2 in zonal segregation of olfactory sensory neuron axons in the main olfactory bulb', J Neurosci, 2007, 27, (34), pp. 9094-9104

86 Cho, J.H., Prince, J.E., Cutforth, T., and Cloutier, J.F.: 'The pattern of glomerular map formation defines responsiveness to aversive odorants in mice', J Neurosci, 2011, 31, (21), pp. 7920-7926

87 Xiao, T., Staub, W., Robles, E., Gosse, N.J., Cole, G.J., and Baier, H.: 'Assembly of lamina-specific neuronal connections by slit bound to type IV collagen', Cell, 2011, 146, (1), pp. 164-176

88 Sung, P.J., Boulos, N., Tilby, M.J., Andrews, W.D., Newbold, R.F., Tweddle, D.A., and Lunec, J.: 'Identification and characterisation of STMN4 and ROBO2 gene involvement in neuroblastoma cell differentiation', Cancer Lett, 2013, 328, (1), pp. 168-175

89 Brust, P.F., Simerson, S., McCue, A.F., Deal, C.R., Schoonmaker, S., Williams, M.E., Velicelebi, G., Johnson, E.C., Harpold, M.M., and Ellis, S.B.: 'Human neuronal voltage-dependent calcium channels: studies on subunit structure and role in channel assembly', Neuropharmacology, 1993, 32, (11), pp. 1089-1102

90 Zhang, W., Jiang, H.H., and Luo, F.: 'Diverse organization of voltage-gated calcium channels at presynaptic active zones', Front Synaptic Neurosci, 2022, 14, pp. 1023256

91 Gao, P., He, M., Zhang, C., and Geng, C.: 'Integrated analysis of gene expression signatures associated with colon cancer from three datasets', Gene, 2018, 654, pp. 95-102

92 Cai, J., Qiu, J., Wang, H., Sun, J., and Ji, Y.: 'Identification of potential biomarkers in ovarian carcinoma and an evaluation of their prognostic value', Ann Transl Med, 2021, 9, (18), pp. 1472

93 Jaworski, J., Spangler, S., Seeburg, D.P., Hoogenraad, C.C., and Sheng, M.: 'Control of dendritic arborization by the phosphoinositide-3'-kinase-Akt-mammalian target of rapamycin pathway', J Neurosci, 2005, 25, (49), pp. 11300-11312

94 Chan, C.B., Liu, X., Ensslin, M.A., Dillehay, D.L., Ormandy, C.J., Sohn, P., Serra, R., and Ye, K.: 'PIKE-A is required for prolactin-mediated STAT5a activation in mammary gland development', EMBO J, 2010, 29, (5), pp. 956-968

95 Lee, C.C., Huang, C.C., and Hsu, K.S.: 'Insulin promotes dendritic spine and synapse formation by the PI3K/Akt/mTOR and Rac1 signaling pathways', Neuropharmacology, 2011, 61, (4), pp. 867-879

96 Fransson, S., Abel, F., Kogner, P., Martinsson, T., and Ejeskar, K.: 'Stage-dependent expression of PI3K/Akt-pathway genes in neuroblastoma', Int J Oncol, 2013, 42, (2), pp. 609-616

97 Zhou, Q.: 'Organizing the synaptic junctions', J Biol Chem, 2023, 299, (5), pp. 104716

98 Perez-Garcia, C.G., and O'Leary, D.D.M.: 'Formation of the Cortical Subventricular Zone Requires MDGA1-Mediated Aggregation of Basal Progenitors', Cell Rep, 2016, 14, (3), pp. 560-571

99 Bucan, M., Abrahams, B.S., Wang, K., Glessner, J.T., Herman, E.I., Sonnenblick, L.I., Alvarez Retuerto, A.I., Imielinski, M., Hadley, D., Bradfield, J.P., Kim, C., Gidaya, N.B., Lindquist, I., Hutman, T., Sigman, M., Kustanovich, V., Lajonchere, C.M., Singleton, A., Kim, J., Wassink, T.H., McMahon, W.M.,

Owley, T., Sweeney, J.A., Coon, H., Nurnberger, J.I., Li, M., Cantor, R.M., Minshew, N.J., Sutcliffe, J.S., Cook, E.H., Dawson, G., Buxbaum, J.D., Grant, S.F., Schellenberg, G.D., Geschwind, D.H., and Hakonarson, H.: 'Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes', PLoS Genet, 2009, 5, (6), pp. e1000536

100 Li, J., Liu, J., Feng, G., Li, T., Zhao, Q., Li, Y., Hu, Z., Zheng, L., Zeng, Z., He, L., Wang, T., and Shi, Y.: 'The MDGA1 gene confers risk to schizophrenia and bipolar disorder', Schizophr Res, 2011, 125, (2-3), pp. 194-200

101 Butz, H., Ding, Q., Nofech-Mozes, R., Lichner, Z., Ni, H., and Yousef, G.M.: 'Elucidating mechanisms of sunitinib resistance in renal cancer: an integrated pathological-molecular analysis', Oncotarget, 2018, 9, (4), pp. 4661-4674

102 Gratton, M.O., Torban, E., Jasmin, S.B., Theriault, F.M., German, M.S., and Stifani, S.: 'Hes6 promotes cortical neurogenesis and inhibits Hes1 transcription repression activity by multiple mechanisms', Mol Cell Biol, 2003, 23, (19), pp. 6922-6935

103 Vias, M., Massie, C.E., East, P., Scott, H., Warren, A., Zhou, Z., Nikitin, A.Y., Neal, D.E., and Mills, I.G.: 'Pro-neural transcription factors as cancer markers', BMC Med Genomics, 2008, 1, pp. 17

104 Ramos-Montoya, A., Lamb, A.D., Russell, R., Carroll, T., Jurmeister, S., Galeano-Dalmau, N., Massie, C.E., Boren, J., Bon, H., Theodorou, V., Vias, M., Shaw, G.L., Sharma, N.L., Ross-Adams, H., Scott, H.E., Vowler, S.L., Howat, W.J., Warren, A.Y., Wooster, R.F., Mills, I.G., and Neal, D.E.: 'HES6 drives a critical AR transcriptional programme to induce castration-resistant prostate

cancer through activation of an E2F1-mediated cell cycle network', EMBO Mol Med, 2014, 6, (5), pp. 651-661

105 Hartman, J., Lam, E.W., Gustafsson, J.A., and Strom, A.: 'Hes-6, an inhibitor of Hes-1, is regulated by 17beta-estradiol and promotes breast cancer cell proliferation', Breast Cancer Res, 2009, 11, (6), pp. R79

106 Nakano, N., Maeyama, K., Sakata, N., Itoh, F., Akatsu, R., Nakata, M., Katsu, Y., Ikeno, S., Togawa, Y., Vo Nguyen, T.T., Watanabe, Y., Kato, M., and Itoh, S.: 'C18 ORF1, a novel negative regulator of transforming growth factor-beta signaling', J Biol Chem, 2014, 289, (18), pp. 12680-12692

107 Liu, Z., Huo, X., Zhao, S., Yang, J., Shi, W., Jing, L., Li, W., Li, Y., Ma, L., Gao, Y., and Diao, A.: 'Low density lipoprotein receptor class A domain containing 4 (LDLRAD4) promotes tumorigenesis of hepatic cancer cells', Exp Cell Res, 2017, 360, (2), pp. 189-198

108 Lin, S., Huang, C., Sun, J., Bollt, O., Wang, X., Martine, E., Kang, J., Taylor, M.D., Fang, B., Singh, P.K., Koomen, J., Hao, J., and Yang, S.: 'The mitochondrial deoxyguanosine kinase is required for cancer cell stemness in lung adenocarcinoma', EMBO Mol Med, 2019, 11, (12), pp. e10849

109 Sekiya, A., Takasawa, K., Arai, Y., Horike, S.I., Akutsu, H., Umezawa, A., and Nishino, K.: 'Variation of DNA methylation on the IRX1/2 genes is responsible for the neural differentiation propensity in human induced pluripotent stem cells', Regen Ther, 2022, 21, pp. 620-630

110 Guo, X., Liu, W., Pan, Y., Ni, P., Ji, J., Guo, L., Zhang, J., Wu, J., Jiang, J., Chen, X., Cai, Q., Li, J., Zhang, J., Gu, Q., Liu, B., Zhu, Z., and Yu, Y.: 'Homeobox gene IRX1 is a tumor suppressor gene in gastric carcinoma', Oncogene, 2010, 29, (27), pp. 3908-3920

111 Lu, J., Song, G., Tang, Q., Zou, C., Han, F., Zhao, Z., Yong, B., Yin, J., Xu, H., Xie, X., Kang, T., Lam, Y., Yang, H., Shen, J., and Wang, J.: 'IRX1 hypomethylation promotes osteosarcoma metastasis via induction of CXCL14/NF-kappaB signaling', J Clin Invest, 2015, 125, (5), pp. 1839-1856

112 Iozzo, R.V., and Schaefer, L.: 'Proteoglycan form and function: A comprehensive nomenclature of proteoglycans', Matrix Biol, 2015, 42, pp. 11-55

113 Jiao, M., Sun, W., Li, L., Li, C., Zhou, J., Li, Q., and Duan, L.: 'Clinical significance of SPOCK2 expression signature for high-grade serous ovarian cancer patients', Front Genet, 2022, 13, pp. 878123

114 Aghamaliyev, U., Su, K., Weniger, M., Koch, D., D'Haese, J.G., Werner, J., and Bazhin, A.V.: 'SPOCK2 gene expression is downregulated in pancreatic ductal adenocarcinoma cells and correlates with prognosis of patients with pancreatic cancer', J Cancer Res Clin Oncol, 2023, 149, (11), pp. 9191-9200