Systematic elucidation of transcriptional network necessary for initiation and maintenance of high-risk neuroblastoma

Presha Rajbhandari

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2016

ABSTRACT

Systematic elucidation of transcriptional network necessary for initiation and maintenance of high-risk neuroblastoma

Presha Rajbhandari

Neuroblastoma is a heterogeneous pediatric malignancy originating from the developing sympathetic nervous system, with poor long-term survival for high-risk patients (~40%). About half of advanced neuroblastomas harbor high-level amplification of the *MYCN* gene, and these tumors show few, if any, additional driver lesions. Despite significant increase in the body of knowledge of genetics in neuroblastoma, all the high-risk patients follow similar therapeutic procedures and little advancement has been made on molecular target based therapies. The major challenge is to dissect the complexity and heterogeneity of these tumors to find driver genes and activated pathways that are essential for the survival of these cancer cells.

We used an integrated systems biology approach to define the core regulatory machinery responsible for maintenance of an aggressive neuroblastoma phenotypic state. In the first part of the thesis, I will discuss our computational approach to decipher the tumor heterogeneity by subtype classification, followed by identification of master regulator protein modules for three distinct molecular subtypes of high-risk neuroblastomas, which were validated in a large independent cohort of cases. We propose that such modules are responsible for integrating the effect of mutations in upstream pathways and for regulating the genetic programs and pathways necessary for tumor state implementation and maintenance.

The second part of the thesis is focused on experimental validation of putative master regulators in the subtype of neuroblastomas associated with *MYCN* amplification. By using RNAi screening followed by experimental and computational analyses to elucidate the interdependencies between the top master regulators, we identified TEAD4-MYCN positive feedback loop as a major tumor maintenance mechanism in this subtype. While MYCN regulates TEAD4 transcriptionally, TEAD4 regulates MYCN through transcriptional and post-translational mechanisms. Jointly, MYCN and TEAD4 regulate 90% of inferred

MR proteins and causally orchestrate 70% of the subtype-specific gene expression signature. TEAD4 gene expression was associated with neuroblastoma patient survival independently of age, tumor stage and MYCN status (P=2.1e-02). In cellular assays, MYCN promoted growth and repressed differentiation, while TEAD4 activated proliferation and DNA damage repair programs, the signature hallmarks of *MYCN*-amplified neuroblastoma cells. Specifically, TEAD4 was shown to induce MYCN-independent proliferation by transactivating key genes implicated in high-risk neuroblastoma pathogenesis, including cyclin-dependent kinases, cyclins, E2Fs, DNA replication factors, checkpoint kinases and ubiquitin ligases. The critical role of the core master regulator module in controlling tumor cell viability, both in vitro and in vivo, and its clinical relevance as a prognostic factor highlights TEAD4 as a novel and highly effective candidate target for therapeutic intervention.

In this thesis, we demonstrate that interrogation of tumor specific regulatory networks with patient-derived gene expression signatures can effectively elucidate molecular subtypes as well as the core transcriptional machinery driving subtype specific hallmarks. This approach enables identification of oncogenic and non-oncogenic dependencies of high-risk neuroblastoma and is applicable to other tumor subtypes.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ARACNe	Algorithm for the Reconstruction of Accurate Cellular Networks
CGH	Comparative genomic hybridization
CIN	Chromosomal instability
CNV	Copy number variation
COG	Children's Oncology Group
DDR	DNA damage response
DGE	Differential gene expression
FDR	False discovery rate
FET	Fisher's exact test
GEP	Gene expression profile
GES	Gene expression signature
GI	Genomic instability
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
INRG	International Neuroblastoma Risk Group
INSS	International Neuroblastoma Staging System
KEGG	Kyoto Encyclopedia of Genes and Genomes
LOH	Loss of heterozygosity
MARINA	Master Regulator Inference algorithm
MES	Mesenchymal
MR	Master Regulator
MYCNA	*MYCN*-amplified
MYCN-NA	*MYCN*-non amplified
NBi	Neuroblastoma interactome
NBL	Neuroblastoma
NRC	Neuroblastoma Research Consortium

RNAi	RNA interference
SA	Sympathoadrenal
shRNA	Short hairpin RNA
siRNA	Short interfering RNA
SNP	Single Nucleotide Polymorphism
SNS	Sympathetic nervous system
TARGET	Therapeutically Applicable Research to Generate Effective Treatments
TF	Transcription factor
VIPER	Virtual Proteomics by Enriched Regulon analysis

Acknowledgements

I am deeply indebted to several people who have supported me directly or indirectly, before and during my graduate studies, and contributed to my scientific and educational journey.

First and foremost, my deepest gratitude to Andrea Califano for his continuous guidance and support since I joined his lab several years ago and throughout my graduate studies. I owe to him my growth as a scientist for allowing me to work independently, while keeping me focused. His positive energy, intellect and broad insight has always been a motivating force, has helped to navigate the difficult problems and to see the projects through the end. He has created a highly collaborative and pleasant lab environment to foster productive scientific endeavors. His own career exemplifies the power of collaborative efforts between the computational and experimental researchers to answer some of the significant challenges in biology. It has been an invaluable, well-rounded learning experience, which I will undoubtedly benefit from as I build my career, and I sincerely thank him for that.

For my thesis project, I got an opportunity to collaborate with and learn from the best people the field and I am indebted to them. John Maris has been a constant source of guidance throughout the project and important decision-makings. He has always been available by providing detailed explanations for every question, big or small. I owe my knowledge about neuroblastoma to him and his amazing contribution to the field. I would also like to thank Anna Lasorella, Antonio Iavarone, Brent Stockwell and Jose Silva for the collaboration and everything that I have learnt from them. Anna Lasorella had been my mentor when I first joined Columbia and I owe to her the experimental expertise that I gained through her rigorous training and guidance; and for helping me navigate through experimental issues.

I would also like to thank my thesis committee members, Brent Stockwell and Darrell Yamashiro for sharing their wisdom and providing valuable guidance at different points, which has tremendously helped with the progress of the project. I would like to extend thanks to my dissertation committee member, Rodney Rothstein for devoting his valuable time and effort.

This work would not be complete without the productive collaborative relationship with several colleagues. I would like to thank Gonzalo Lopez, who has been the computational lead behind this project. This work represents an interdisciplinary effort, where we worked on a highly iterative fashion to integrate the computational and experimental analyses to understand and refine the findings. I would also like to take this opportunity to thank several collaborators including Jiyang Yu at Pfizer; Lori Hart, Daniel Martinez and Mark Yarmarkovich at CHOP; Ruth Rodriguez-Barrueco at Mount Sinai, who have contributed by providing guidance and/or performing experimental/computational analyses for the work presented here.

My colleagues, past and present, have been an important part of the journey. I would like to thank all the members of the Califano lab. A special thanks to Brygida Bisikirska, Mariano Alvarez, Yao Shen, Chuck Karan, Ronald Realubit, Beatrice Salvatori, Pavel Sumazin, Mukesh Bansal, Elena Komissarova, Alexandar Lachman, Federico Giorgi, Forest Ray, Prem Subramanian, Archana Iyer and Hua-Sheung Chiu for your scientific contributions and helpful discussions at different points during the course. I have learnt from each one of you. Additionally, thanks to Mahalaxmi Aburi, William Shin, Adina Grunn, and Carolyn Williams for your support. I would also like to thank the Biological Sciences department and Sarah Kim for the support during my graduate studies. This work was funded by CTD2 and NIH training grant.

Most importantly, my family has been the bedrock, providing unconditional love and support my entire life. I would like to thank my parents, husband and his family, sisters and brother for the support and encouragement throughout my studies. A special thanks to my husband, without his support and sacrifices, I could not have devoted myself to this. Last but not the least, my biggest thanks goes to my wonderful daughter, Anusha, for the love and unending joy she brings to my life... and simply for giving meaning to my life…

To my Dad, Bishnu and Mom, Shanti

Chapter 1

Introduction

Neuroblastoma (NBL) is the most common extracranial solid tumor in children. It remains a challenging malignancy because of its remarkably heterogeneous clinical behavior, ranging from spontaneous maturation and regression to aggressive clinical phenotype (Brodeur and Nakagawara, 1992). Efforts have been made by various international groups for patient risk stratification at diagnosis, to direct appropriate treatment plan (Cohn et al., 2009a). Based on a few clinical and genetic markers with definitive prognostic impact, NBL patients have been categorized into low, intermediate and high-risk groups. Among these, more than 50% of the pediatric NBLs are defined as “high-risk”. Although substantial improvement has been made in treating the low and intermediate risk group of NBL patients, the outcome of high-risk clinical phenotype has improved only modestly. Despite intensive multimodal therapy, the five-year survival rate for these patients is still below 50% and the surviving patients often suffer from chronic side effects. This highlights the importance of finding specific and more effective approaches to treatment of high-risk NBL. Dissection of the molecular profiles of the high-risk NBL patients to understand the complexity and heterogeneity underlying the disease is the crucial first step towards this goal.

With tremendous increase in our understanding of the pathogenesis of cancers, it has become clear that the cancer cells possess the core hallmark capabilities including programs deregulating proliferation, evasion of cell death, cellular metabolism as well as inducing genome instability, metastasis and angiogenesis (Hanahan and Weinberg, 2011a). Decades of studies in various cancers have shown that the cancer cells harbor numerous genetic and epigenetic alterations including gain of function mutation, amplification, translocation or overexpression of oncogenes, and/or loss of function mutation, deletion or epigenetic silencing of tumor suppressors, allowing the cells to undergo malignant transformation and/or maintenance of tumor state.

In an effort to identify these key genes inducing the tumor specific hallmarks, novel technologies have fueled the generation of extensive multi-dimensional datasets characterizing genomic, transcriptomic, epigenetic and proteomic properties of different tissue types, in the past 15 years. Major initiative by several groups like The Cancer Genome Atlas (TCGA) Research Network, International network of cancer genome projects (ICGC) (International Cancer Genome Consortium et al., 2010) and Cancer Genome Characterization Initiative (CGCI) have produced massive comprehensive datasets from large cohort of cancer patients in different tumor types. Similar efforts in the field of NBL has been done by Therapeutically Applicable Research to Generate Effective Treatments (TARGET), Neuroblastoma Research Consortium (NRC) and other international cohorts, to identify the therapeutic targets and prognostic biomarkers in NBL patients — with the ultimate aim to develop more effective and molecularly targeted therapies. Such data not only provides the knowledge base, but also provides an opportunity to interrogate the profiles to understand the molecular basis of cancers. Despite the wealth of information, the major challenge has been to develop strategies to dissect the insurmountable complexity within tumors to uncover the molecular underpinnings of the disease.

With increasing understanding of the complexity of biological systems and the technological capabilities to collect vast amount of information from normal and disease states, systems biology came as a necessary shift from the reductionist approach, which allowed us to view the biological system from a different lens, at different levels and from different perspectives. The landscape of computational approaches to identify driver genes of a phenotype has evolved concurrently with the availability of the high-throughput “omics” measurements. The main approach used in the field is to identify the somatic alterations of genomic information associated with the disease. Besides the already known activating mutations in oncogenes such as *RAS*, *PI3K*, and deactivating mutations in tumor suppressors such as *PTEN*, *BRCA1* and *p53*, these sequencing efforts have shown that sporadic cancers have very low rate of recurrent mutations (Greenman et al., 2007; Lawrence et al., 2013; Tamborero et al., 2013). This stands true for pediatric cancers like NBL, where *ALK* is the only known gene with significant frequency of somatic and familial mutation (Chen et al., 2008; Janoueix-Lerosey et al., 2008a; Mossé et al., 2008; Pugh et al., 2013). On the other hand, many cancers including NBL are marked by chromosomal

instability (CIN), resulting into changes in the copy number of massive number of genes, which could affect the cell by changing the dosage of the genes in the altered regions (Łastowska et al., 2002), imbalance of components of protein complexes affecting the expression or stability of other proteins, to alter the balance of growth and survival signals in the cells (Torres et al., 2007). Identification of genes within these large regions of chromosomal alterations remains a challenge.

While the DNA sequencing efforts will undoubtedly yield new information to identify novel driver mutations, rare mutant subpopulations, biomarkers and improved patient stratification, one of the limitations is that it will not lead to discovery of genes and pathways that are not altered at the genetic level. The cancers containing these multiple genetic and epigenetic abnormalities depend on oncogene (Weinstein and Joe, 2006) or non-oncogene (Luo et al., 2009a) mediated maintenance of tumor phenotype, comprising a vast landscape of tumor dependencies. Furthermore, it is not an easy task to integrate the information from all these regulatory levels to derive a single mechanistic model of the system. However, all these alterations manifest its effect by mediating differential expression of their effector genes, the confluence of which can be captured by its dynamic transcriptome landscape.

Several studies have shown that the classification of patients based on gene expression pattern is a powerful approach to stratify the tumors into molecularly homogeneous subgroups (Alizadeh et al., 2000; Cancer Genome Atlas Network, 2012; Cristescu et al., 2015; Golub et al., 1999; Phillips et al., 2006; van ’t Veer et al., 2002). These subgroups are useful to understand the inherent biology of tumors as it provides us with a list of effector genes that can be interrogated to infer the drivers and the corresponding pathways driving these subgroups. The knowledge of transcriptional regulators residing over such transcriptional states, as read by gene expression signature (GES) would allow us to determine the driving mechanisms of malignant transformation.

Thus, following on recent results from assembly and interrogation of regulatory networks from our laboratory (Califano et al., 2012), we focused on more universal tumor dependencies – that are relatively independent of the specific genetic alteration landscape of tumors, with similar transcriptional

profiles. This was accomplished using the Master Regulator Inference algorithm (MARINa) (Aytes et al., 2014; Carro et al., 2010; Lefebvre et al., 2010) and its extension to the analysis of individual gene expression profiles, which we will call VIPER (Virtual Proteomics by Enriched Regulon analysis) (Alvarez et al., manuscript in press). MARINa and VIPER interrogate regulatory models (reverse-engineered de novo from experimental data) to identify transcriptional regulators called master regulator (MR) proteins that causally implement the transcriptional signature representative of a subtype- or sample-specific tumor state (Aytes et al., 2014; Bisikirska et al., 2015; Carro et al., 2010; Lefebvre et al., 2010; Rodriguez-Barrueco et al., 2015). Critically, our lab has shown that MR proteins are mechanistically responsible for integrating the effect of multiple genetic alterations in upstream pathways to implement the associated tumor phenotype (Chen et al., 2014; Compagno et al., 2009). Interestingly, transcriptome based analysis in NBL are focused more on risk classification (Abel et al., 2011; Asgharzadeh et al., 2006; Oberthuer et al., 2015; Valentijn et al., 2012), and the stratification of high-risk NBL molecular subtypes are lacking. We propose that the MR modules are highly enriched in both essential and synthetic lethal genes, which represent Achilles's heels of cancer that are more generalizable than the diverse repertoire of upstream genetic alterations responsible for their aberrant activity.

Overall, this thesis aims to (1) dissect the heterogeneity of high-risk NBL GEPs to classify them into homogeneous group of tumors (2) predict the MR inducing the GES of each subgroup (3) functionally validate the putative MRs of a subtype. The analysis proceeds in seven steps- Step 1: We analyze large-scale gene expression profiles of high-risk NBL patients and categorized them into clinically and/or biologically distinct/relevant molecular subtypes. Step 2: We analyze gene expression profile data to reverse-engineer NBL transcriptional networks or interactomes (NBi), using the Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNe) (Margolin et al., 2006). Step 3: To infer MR proteins of individual molecular subtypes, we generate subtype specific differential expression signatures by comparing samples representing high-risk molecular subtypes to samples in the low-risk set (control). To infer MR proteins of individual molecular subtypes, we generated TARGET- and NRC-specific differential expression signatures by comparing samples representing high-risk molecular

subtypes to samples in the low-risk set (control). We then used the VIPER suite which incorporates an updated version of MARINa algorithm (msVIPER) to identify the MR proteins that causally regulate these signatures, based on their targets as represented in the cohort-matched interactome. Specifically, msVIPER prioritizes MR proteins based on the enrichment of their transcriptional targets (i.e., their interactome regulons) in genes differentially expressed in the gene expression signature of interest. Step 4: we performed an identical MR analysis, using VIPER, which works at the single sample level, for each of a set of 25 neuroblastoma cell lines suitable for experimental validation. Cell lines were matched to specific subtypes based on overlap of MR proteins. Step 5: We tested the top ranked MRs through a molecular screen, using complementary approaches and RNAi reagents, to mitigate false discovery resulting from off-target effects and technology specific biases Step 6: We experimentally tested the mutual regulatory activity of validated MR proteins to infer the connectivity of the hierarchical regulatory module they comprise. Finally, Step 7: We functionally characterized the regulatory feedback loop and mechanism responsible for maintaining the aberrant activity of the MR module characterized in Step 6 as well as its biological relevance.

The knowledge of the drivers of tumor programs has been a major bottleneck in the molecularly targeted treatment of high-risk NBL. The potential role of the driver genes in promoting high-risk NBL will bring new insights into how these genes are controlling the cancerous state at the mechanistic level and lead to identification of therapeutic targets. Once validated, it can serve as a pipeline that can repeatedly be used to understand other tumor subtypes.

This thesis is divided into seven chapters. Chapter 1 outlines the motivation, objective and significance of the thesis, brief overview of the research design and organization of this document. Chapter 2 gives an overview on current knowledge of neuroblastoma from developmental biology and functional genomics perspective. The current and prospective risk stratification strategies for NBL patients are also described. Chapter 3 introduces the computational approaches used for analysis of GEP data for pathway enrichment analysis, cluster analysis, transcriptional regulatory network inference and master regulator identification algorithms that are relevant for this thesis. Chapter 4 details the molecular

classification, biological characterization and master regulator identification of high-risk NBL subtypes. Chapter 5 details the experimental validation of predicted MRs of MYCNA subtype by in-vitro and in-vivo RNAi screening. It also shows experimental analyses performed to elucidate the modular logic controlling disease state and identification of TEAD4 as a novel MYCNA subtype specific master regulator. Chapter 7 shows the conclusion, contribution of the thesis and future directions.

Chapter 2

Neuroblastoma

Neuroblastoma (NBL) is the most common extracranial solid tumor of childhood, accounting for 8-10% of all childhood cancers and 15% of childhood cancer related deaths (Maris, 2010). It generally occurs in infancy and early childhood, with the average age of diagnosis between one and two years and almost 90% of the cases below five years (Castleberry et al., 1997). NBL remains a challenging malignancy because of its remarkably heterogeneous presentation and has intrigued investigators for decades. For example, children with age below 18 months would undergo spontaneous regression, while older children with similar symptoms would display aggressive clinical phenotype (Brodeur and Nakagawara, 1992).

Based on various genetic and clinical factors, NBL patients have been categorized into low, intermediate and high-risk groups. Among these, more than 50% of the pediatric NBLs are defined as high-risk. Although substantial improvement has been made in treating the low and intermediate risk group of NBL patients, the outcome of high-risk clinical phenotype has improved only modestly, with 5-year survival between 40-50% (cancer.org). The surviving patients often suffer from complications of the intensive treatments that they receive. Despite reports showing improvements in outcome in recent randomized trials (Yu et al., 2010), a significant number of NBL patients still succumb to the disease. Moreover, limited understanding of underlying biology of this disease in children has resulted into therapies using non-specific and toxic agents designed in the first place for adults. A better understanding of the disease from the developmental and cancer biology perspective may offer new insights for targeted therapeutics as genes that control cell growth, differentiation and death during normal development have parallel functions in cancer. While there are a number of excellent reviews on NBL (Brodeur, 2003; Brodeur and Bagatell, 2014; Cheung and Dyer, 2013; Maris, 2010; Maris and Matthay, 1999a; Marshall et al., 2014; Yu et al., 2010), a brief summary of our current understanding of NBL from developmental and cancer genomics perspective as well as current risk-stratification strategies are detailed below.

2.1 Developmental biology of Neuroblastoma

Many pediatric malignancies are embryonal in nature, where the defect in normal embryonic development processes such as terminal differentiation program; or reversion of the cells to pluripotent state could aid the initiation of these tumors (Diede, 2014). The biological phenomenon constituting the normal development is highly orchestrated, resulting into several differentiated cell types from a single pluripotent stem cell. NBL is an embryonic tumor of peripheral nervous system that are believed to originate from immature cells of neural crest origin (Hoehner et al., 1996) – neural crest is a multipotent embryonic structure that are present early during embryogenesis and gives rise to diverse cell types that form various components of sympathetic nervous system (SNS) including sympathetic ganglia, chromaffin cells of the adrenal medulla and paraganglia (Huber, 2006). NBLs originates mainly from adrenal medulla but can also develop in the nerve tissue in the neck, chest, abdomen and pelvis (Maris, 2010). However, the exact mechanism of neuroblastoma formation remains unknown. Gene expression profiling studies indicate that NBLs have expression patterns similar to the SNS cells (De Preter et al., 2006; Hoehner et al., 1996, 1998). It has been suggested that dysregulation of temporally and spatially regulated genes dictating migration, specification, divergence and maturation during normal development could cause defects in the differentiation and cell death mechanisms, hence leading to uncontrolled cell proliferation and resistance to cell death pathways to induce neuroblastoma tumor formation (Marshall et al., 2014).

Several genes that play an important role in SNS development have been shown to be critical for NBL pathogenesis. For example, V-Myc Avian Myelocytomatosis Viral Oncogene Neuroblastoma Derived Homolog (MYCN) expression is transiently high in early post-migratory neural crest to control migration and proliferation of the neural crest cells, and it is gradually reduced in maturing cells (Zimmerman et al., 1986). It has been shown in mice and zebrafish models that sustained overexpression of MYCN blocks development towards chromaffin cell fate and leads to neuroblastoma initiation (Hansford et al., 2004; Weiss et al., 1997; Zhu et al., 2012). In a normal state, the sympathoadrenal cells would differentiate into sympathetic neurons and adrenal chromaffin cells (Huber, 2006). Similarly, Anaplastic Lymphoma Receptor Tyrosine Kinase (ALK) signaling has been shown to be essential for sympathetic neuron proliferation (Reiff et al., 2011). Somatic and germline activating mutation of *ALK* are present in

8-10% of NBL patients and stimulate proliferation of NBL cells (Janoueix-Lerosey et al., 2008b; Mosse et al., 2008). Introduction of *ALK* mutation into neural crest cells of mice is sufficient to induce neuroblastoma and it synergizes with MYCN to promote tumor formation in vivo (Heukamp et al., 2012; Schulte et al., 2013). Similarly, differentiation of SA cells are promoted by a network of transcription factors like mammalian achaete–scute homolog 1 (MASH-1), paired homeodomain protein 2A (PHOX2A) and 2B (PHOX2B), HAND2 and GATA 2/3 (Guillemot et al., 1993; Howard et al., 2000; Lim et al., 2000; Pattyn et al., 1999; Stanke et al., 1999). Heterozygous germline mutation of *PHOX2B* has been shown to be involved in initiation of NBL, especially in patients with disorders of the autonomic nervous system (Mosse et al., 2004).

Similarly, neurotrophins like nerve growth factor (NGF), brain-derived neurotrophic factor (BDNF), neurotrophin-3 and neurotrophin-4 have been shown to be important in SNS development (Fagan et al., 1996; Klein et al., 1993). They signal through two types of receptors: Trk receptor tyrosine kinases and p75 neurotrophin receptor, to regulate neural precursor cell fate, cell growth, survival and repair of SNS (Chao, 2003; Kaplan and Miller, 2000). In NBL tumors, expression of TRK receptor kinases, TRKA and TRKB are strong prognostic indicator, where low level of TRKA expression or high expression of TRKB is indicative of poor prognosis (Brodeur et al., 2009). In fact, tumors with high expression of TRKA expression has been shown to be more prone to cell death or differentiation depending on the presence or absence of NGF in their microenvironment (Brodeur and Bagatell, 2014; Maris and Matthay, 1999a). *MYCN*-amplified tumors, on the other hand, have very low expression of TRKA and high expression of TRKB to block neural differentiation (Brodeur, 2003). Thus, dysregulated expression of a number of key neurodevelopmental regulators could play an important role in neuroblastoma initiation.

2.2 Genetics and molecular biology of neuroblastoma

NBLs have very heterogeneous clinical presentation and efforts have been made to understand the biological basis behind it. Hereditary predisposition of NBL accounts for only 1-2% of NBL patients. 80% of these cases are caused by germline mutations in *ALK* (75%) (Mosse et al., 2008), and of *PHOX2B* (5%) (Bourdeaut et al., 2005; Mosse et al., 2004), to a lesser degree. It has been shown that similar to

other pediatric tumors, high-risk NBLs harbor very low frequency of recurrent somatic mutations (Pugh et al., 2013), making it challenging to classify patients based on DNA sequencing alone. The most frequent mutation is in *ALK* gene (9.2% of cases), followed by *PTPN11* (2.9%), *ATRX* (2.5%), *MYCN* (1.7%) and *NRAS* (0.83%). Further studies have identified less frequent SNPs located at *LMO1* (LIM domain only 1) (Wang et al., 2011), *BARD1* (BRCA1 Associated RING Domain 1) (Capasso et al., 2009), *HACE1* (HECT domain and ankyrin repeat containing E3 ubiquitin protein ligase 1) and *LIN28B* (Lin-28 Homolog B (C. Elegans)) (Diskin et al., 2012). This indicates that NBLs are likely driven by copy number alterations and other epigenetic alterations.

Indeed, genomic instability (GI) is a hallmark of cancers including NBLs (Hanahan and Weinberg, 2011b; Negrini et al., 2010). While most of the hereditary cancers are caused by germline mutations in the genes maintaining genomic integrity such as DNA damage sensing and repair genes, and mitotic checkpoint genes, the molecular basis of sporadic cancers remain unclear. Sequencing efforts in several sporadic cancers have failed to identify genes that are frequently mutated (Cahill et al., 1999; Wood et al., 2007). Consistently, mutation in these genes are a rare event in NBL (Pugh et al., 2013) except for mutations in *BARD1* (BRCA1 Ring Domain 1) gene (Capasso et al., 2009). A form of GI, microsatellite instability (MIN), characterized by repeats of microsatellites (short repeated sequences of DNA), and reported to be caused by loss of function of DNA mismatch repair proteins, are rare in most cancers including NBLs (Hogarty et al., 1998; Martinsson et al., 1995; White et al., 1995).

Instead, GI manifests as chromosomal instability (CIN) in most sporadic cancers including NBL, which involve loss or gain of parts or whole chromosomal fractions. NBL tumors harbor distinct types of chromosomal alteration reflecting the complex selection structure involved in escaping the cellular mechanisms protecting the genomic integrity and hence contributing towards accumulation of genotypes that favor tumor progression. While conventionally, cancer development has been shown to be a progressive phenomenon, with continuous acquisition of genetic alteration, each providing selective advantage and resulting into increasingly malignant phenotype (Stratton et al., 2009); recently, it has been shown that the genomic instability characterized by clustered rearrangements affecting one

or a few chromosomes happen in a single catastrophic event called chromothripsis (Greek: chromo stands for chromosome and thripsis means shattering into pieces) (Stephens et al., 2011). First, multiple rearrangements confined to a few chromosomes or parts of chromosomes are generated because of the DNA double strand breaks (DSB). This is followed by imperfect repair of DSBs by joining of the DNA fragments by the process of nonhomologous end-joining (NHEJ). Chromothripsis has been shown to occur in 18% of advanced stage NBL patients (Molenaar et al., 2012; Valentijn et al., 2015).

Decades of work on these searches have identified several large-scale chromosomal imbalances and gene amplifications, which are associated with NBL pathogenesis. These genomic alterations could aid tumorigenesis through combinatorial effect of multiple genes lying in these regions. The copy number changes of genes mapped to the region of imbalance could alter its gene dosage, such as deletion of tumor suppressor genes or amplification of oncogenes. Furthermore, genes that are involved in activation or repression of genes mapped to other chromosomes, or whose alteration contributes to the imbalance of components of several protein complexes involved in different process could affect NBL pathogenesis. To ascertain the effect of copy number changes in these regions, global gene expression profiling studies have been conducted in parallel. Different groups have retrieved > 15% of the genes that map to these altered regions showing corresponding changes in gene expression (Janoueix-Lerosey et al., 2004; Wang et al., 2006). Even when the altered regions in NBL are known, the challenge, also common to other tumor types is to separate the “driver” genes from the “passenger” genes. A few genes that have been identified in these regions will be discussed below.

The common methods used to detect the chromosomal alterations that have been adapted to the microarray-based technology are Single nucleotide polymorphism (SNP) array and comparative genomic hybridization (CGH) arrays. SNPs are a variation at a single site of DNA, which can provide information on copy number changes as well as loss of heterozygosity (LOH). SNP arrays implement 25 to 50-mer oligonucleotide probes to capture fragments of sample DNA. Hybridization intensities are assessed to infer copy number alterations, as the signal intensity is directly proportional to the amount of DNA in the sample. CGH is a cytogenetic technique based on the principle of competitive DNA

hybridization, where DNA from tumor and normal samples are labeled, hybridized to metaphase spreads and visualized under the microscope. It provides 5-10 megabase resolution. The implementation of CGH array is based on the same principle, but on an array based platform, where the sample and reference DNA are co-hybridized to long oligonucleotide sequence, hence providing higher resolution at the level of 5-10 kilobases of DNA to assess copy number alteration compared to reference (Pinkel et al., 1998).

2.2.1 Genomic alterations in NBL

The full spectrum of genomic alterations that have been observed in NBL has been thoroughly discussed in literature (Brodeur, 2003; Maris and Matthay, 1999a), and the most common aberrations are discussed below.

2.2.1.1 DNA ploidy

Aneuploidy or presence of abnormal number of chromosomes in a cell is a common feature of cancers. It is caused by abnormal mitosis, resulting into duplication or loss of chromosomes in the daughter cells (Griffiths et al., 2000). DNA ploidy is a means of measuring modal chromosomal content of a cell in terms of DNA index or content, with respect to normal diploid cells. In case of NBLs, the tumors have been classified into diploid (DNA index of 1), hyperdiploid (DNA index > 1), hypodiploid (DNA index <1). It has been shown that ploidy has prognostic significance only in younger patients, with it being highly predictive for patients below 1 year, some degree of significance for children between 1 and 2 years of age and not predictive for patients above 2 years old (Kaneko and Knudson, 2000). The patients with hyperdiploid and near triploid tumors have better prognosis and respond better to treatment than patients with diploid or tetraploid tumors (Look et al., 1984, 1991). This is most likely because hyperdiploid and near triploid tumors have whole chromosomal gains, whereas diploid and tetraploid tumors have segmental chromosomal alterations. Studies performed by Children's Oncology Group (COG) showed that hyperdiploid and near triploid tumors were commonly found in infants, while diploid and tetraploid tumors are associated with high-risk disease markers like older age, metastatic disease,

MYCN-amplification, 1q deletion or 17q amplification, indicating that ploidy behaves as a surrogate marker for high-risk associated genetic alterations.

2.2.1.2 Amplification of MYCN and 2p24 locus

MYCN belongs to the MYC family of transcription factors and maps to the distal short arm of chromosome 2 (2p24). It was originally cloned in 1983 (Schwab et al., 1983) and since then several studies have confirmed the oncogenic role in NBL. DNA amplification of this region occurs in ~25% of NBL tumors, and was identified by cytogenetic analysis, indicated by double minute chromatin bodies (DMs) or homogenously staining regions (HSRs), resulting into 50 to 400 copies of genes per cell (Brodeur et al., 1984; Kohl et al., 1983; Schwab et al., 1984). While other genes from the locus could be co-amplified, with the amplicon size of 500-100kb, *MYCN* is the only gene that has shown to be consistently amplified and is believed to be the primary target of the amplification (Reiter and Brodeur, 1996). *MYCN* amplification, indicated by more than 10 copies of *MYCN*, has been shown to be associated with aggressive and advanced disease state with poor prognosis, even in patients younger than one year old and in lower stage, low-risk patients (Brodeur et al., 1984; Seeger et al., 1985a). Hence, it is routinely used as a robust prognostic marker for patient evaluation.

2.2.1.3 Gain of 17q arm

Gain of 17q chromosomal arm is the most common genomic alteration in primary NBL tumors. Comparative genomic hybridization (CGH) assays showed that it occurs in about 50-75% of NBL cases (Bown et al., 1999; Plantaz et al., 1997). Gain of whole chromosome 17 also occurs but 17q gain with a minimal common region of 25 megabases (17q21-qter) is the most frequent event. Even though 17q gain can occur independently, unbalanced 1;17 translocation occurs frequently, resulting into loss of 1p arm and gain of 17q arm (Van Roy et al., 1994). While there could be many potential oncogenes in this region, survivin, an anti-apoptotic gene, has been proposed to be a candidate oncogene that could mediating the pathologic effect of 17q gain (Islam et al., 2000). Even though most of the advanced stage patients have 17q gain and this event is associated with 1p deletion and *MYCN* amplification, this

factor hasn't been added to risk-classification yet. It has been reported that the patients with MYCNA or 11q deletion with 17q gain have worse prognosis than patients without the 17q gain (Carén et al., 2010).

2.2.1.4 Allelic loss of 1p arm

Deletion of short arm of chromosome 1 (1p) in NBL was first reported by Brodeur et al. (Brodeur et al., 1981), and occurs in about 35% of NBL patients (Maris et al., 2001; White et al., 2001). This event has been shown to be associated with advanced stages of disease and *MYCN*-amplification. While not all MYCNA tumors have 1p deletion, all patients with 1p deletion have MYCNA indicating that 1p deletion precedes MYCNA (Brodeur, 2003). While the presence of both 1p deletion and *MYCN* amplification has been shown to be indicative of poor prognosis, its significance as an independent prognostic marker has been controversial (Attiyeh et al., 2005; Caron et al., 1996; Gehring et al., 1995; Maris et al., 2000). Several putative tumor suppressor genes maps to this region including, *CHD5*, *CAMTA1*, *KIF1B*, *TP73*, *microRNA-34a*. *CHD5* (chromodomain helicase DNA binding domain 5) maps to 1p36.3 and is expressed in low level or absent in NBL cells with 1p deletion (Thompson et al., 2003). It is involved in chromatin remodeling and has been shown to have tumor suppressive effects by controlling proliferation, apoptosis and senescence via p19/p53 pathway (Bagchi et al., 2007) and also by controlling neuronal differentiation (Egan et al., 2013). *CAMTA1* (calmodulin binding transcription activator 1) has been mapped to 1p36.3 and has been shown to mediate tumor suppressive effects by controlling cell growth and inducing cellular differentiation (Henrich et al., 2011). *TP73* (tumor protein 73) also maps to this region and is considered to have a tumor suppressive role, where overexpression of p73 promotes cell cycle arrest and apoptosis (Kaghad et al., 1997). microRNA-34a is another tumor suppressor, the overexpression of which leads to dramatic inhibition of cell growth, partly because it represses the expression of BCL2 and MYCN (Cole et al., 2008).

2.2.1.5 Allelic loss of 11q arm

Many studies performing CGH analysis in NBL confirmed allelic loss of 11q in NBLs (Brodeur, 2003). It is found in 43% of NBL patients, is one of the most common deletions in NBL, where it starts between 11q13.3-14.1 and extends to the telomere (Guo et al., 1999). This event has been shown to be

inversely correlated with *MYCN* amplification, and associated with event-free survival in patients lacking *MYCN* amplification. Because of its prognostic value, 11q-LOH has now been added to the current risk-stratification strategy for NBL. Tumor suppressor genes in this region haven't been identified yet.

The pervasive genomic alterations in NBL as discussed above indicate a clear evidence for loss of control of genomic integrity. The recurrence of non-random alterations in specific chromosomal regions indicates that these aberrations provide selective advantage to the cells by overexpressing or repressing the genes that support malignant transformation. Whether these alterations are the consequence of tumor progression or drive tumor formation is still not fully understood.

2.3 Neuroblastoma Staging and risk-classification

NBL remains a challenging malignancy because of its remarkably heterogeneous clinical behavior, ranging from spontaneous maturation and regression to aggressive clinical phenotype (Yamamoto et al., 1998). There is a consensus in the field that the clinical behavior of NBL can be predicted based on different prognostic variables. Hence, major efforts have been made to classify the patients based on multiple clinical variables with prognostic significance to determine best treatment options. This has resulted into remarkably accurate system to perform risk classification, where stage of disease, age of the patient and *MYCN*-amplification status has been shown to be the most important clinical variable predicting outcome.

In 1998, Children's Oncology Group (COG) established risk-stratification system based on International Neuroblastoma Staging system (INSS) staging system, age at diagnosis (<365 days vs ≥365 days) (Breslow and McCann, 1971), histopathology (Joshi et al., 1992; Shimada et al., 1984), *MYCN*-amplification status (more than 10 copies) (Brodeur et al., 1984; Seeger et al., 1985a) and DNA index (ploidy for patients below 18months) (Look et al., 1984) to classify patients into low-risk, intermediate-risk and high-risk groups (Table 1). This classification system is still in use. While the molecular variables were discussed in the previous section, stratification of patients by stage and histopathology will be discussed below.

Stage of patients, an important variable defining prognosis, was developed by INSS in 1986 and revised in 1988 to classify patients at time of diagnosis based on the origin of tumor, its size and whether or not it has metastasized to other parts of the body, assessment of lymph node involvement and extent of surgical excision (Brodeur et al., 1988, 1993a). Stage I patients have completely resected tumors; Stage 2A and 2B have localized tumors with partial resection and are subcategorized based on the amount of resection and whether it has spread to the nearby lymph nodes and local tissues; Stage 3 patients have partially resected tumors that have infiltrated across the midline; Stage 4 patients have metastatic disease where the tumor has spread to distant organs and lymph nodes; Stage 4S are unique in that even though these patients have metastatic disease, these tumors spontaneously regresses and the patients are below one year of age (Park et al., 2008). It has been postulated that the regression is a result of delayed cell death during the normal process of neural crest development (Brodeur and Bagatell, 2014; Pritchard and Hickman, 1994).

To classify patients based on histology, the International Neuroblastoma Pathology Committee established a prognostically significant classification system, also called Shimada system, based on morphologic features of the neuroblastic tumors (NTs) (Shimada et al., 1999). NTs comprise of neuroblastoma, ganglioneuroblastoma and ganglioneuroma. The tumors are classified as favorable or unfavorable based on the degree of differentiation, mitosis-karyorrhexis index (MKI) and presence or absence of stroma. They are categorized into undifferentiated, poorly differentiated or differentiated tumors based on the degree of differentiated cells. Similarly, tumors are categorized as low, intermediate or high MKI. MKI is a count of mitotic and karyorrhectic cells based on a total count of 5000 cells, where low MKI indicates <2%; intermediate indicates 2-4%; and high indicates >4% of mitotic and karyorrhectic cells. Similarly, based on the Schwannian stromal component of the tumor, they are categorized into neuroblastoma (Schwannian stroma-poor) ganglioneuroblastoma, intermixed (Schwannian stroma-rich), ganglioneuroma (Schwannian stroma-dominant), and ganglioneuroblastoma, nodular (composite, Schwannian stroma-rich/stroma dominant and stroma poor). While ganglioneuroblastoma, intermixed and ganglioneuroma have favorable histology, with 100% 5-year

overall survival; ganglioneuroblastoma, nodular have unfavorable histology with 59% 5-year overall survival (Shimada et al., 2001).

Risk	*MYCN* amplification	Stage*	Age at diagnosis	Overall survival (%)	Current treatment approach
Low risk	No	4S	<12 months	>91 ± 2‡	Supportive care
	No	Locoregional	≤21 years	>95‡	Surgery ± chemotherapy§
Intermediate risk	No	4	<18 months	89 ± 2‡	Surgery and moderate intensity chemotherapy
High risk	Yes	Locoregional	≤21 years	53 ± 4‡	Dose-intensive chemotherapy, surgical resection of residual primary tumour, radiation to primary and resistant metastatic sites, myeloablative therapy with autologous stem cell rescue, anti-GD2 immunotherapy and 13-*cis*-retinoic acid
	Yes	4	<18 months	29 ± 4‡	
	Yes or no	4	≥18 months and ≤21 years	31 ± 1‡	
	No	4	≥12 years	<10‖	

Table 1- 1. Neuroblastoma risk stratification and risk based therapeutic approaches (Cheung and Dyer, 2013)

The analyses of all the mentioned variables are used for risk-classification and appropriate treatment options. For low-risk patients above 12 months of age, surgical resection suffices; infants below 12 months are under observation only; as these tumors almost always regresses. Intermediate-risk patients are treated with surgery and moderate intensity chemotherapy. Although substantial improvement has been made in treating the low and intermediate risk group of NBL patients (overall survival >90%), the outcome of high-risk clinical phenotype, which comprises >50% of NBLs, has improved only modestly, with 5-year survival between 40-50%. Survival rates for different risk groups are derived from the Surveillance, Epidemiology, and End Results databases (seer.cancer.gov). Treatment of high-risk group remains one of the greatest challenges and involves an intensive multimodal therapy. The modest increase in short-term survival is due to intensification of chemotherapy, which results into long-term residual disease and severe complications.

In an effort to define a consensus pre-treatment risk-classification that could be applied internationally, International Neuroblastoma Risk Group (INRG) risk-classification system (Cohn et al., 2009b) was developed after analyzing data available from 8,800 patients in various cohorts worldwide. These

included data collected by COG (North America and Australia), the German Pediatric Oncology and Hematology Group (GPOH), the Japanese Advanced Neuroblastoma Study Group (JANB), the Japanese Infantile Neuroblastoma Co-operative Study Group (JINCS) and International Society of Pediatric Oncology Europe Neuroblastoma Group (SIOPEN) clinical trials. This classification is based on INRG Staging System (INRGSS) (Monclair et al., 2009), which considers radiographic characteristic of tumor unlike the surgical criteria used by INSS to stratify the stage based on clinical and image-defined risk factors (IDRFs). The prognostic value of the IDRFs have been validated by independent studies (Cecchetto et al., 2005; Simon et al., 2008). The patients could be categorized as Stage L1, L2, M or MS where L1 are localized tumors without affecting vital organs; L2 are locoregional tumors with one or more IDRFs; M are metastatic; MS are metastatic in patients younger than 18 months. The COG stages can be translated into INRGSS staging system as follows: INSS 1 → INRGSS L1; INSS 2 and 3 → INRGSS L2; INSS 4 → INRGSS M; and INSS 4S → INRGSS MS (Cohn et al., 2009b). Variables including INRGSS stage, age cutoff (18 months or 547 days) (London et al., 2005), histopathology, grade of tumor differentiation are used for INRG risk classification. Furthermore, the most significant prognostic NBL biomarkers including *MYCN*-amplification status (regardless of age and stage), 11q23 allelic status (Attiyeh et al., 2005; Carén et al., 2010) and DNA index are also incorporated (Ambros et al., 2009). The patients could be categorized by 5-year event-free survival cutoffs into very low-risk (>85%), low-risk (>75 to ≤85%), intermediate-risk (≥50 to ≤75%) and high-risk (<50%) groups. This classification system may replace the COG system in near future as it becomes validated in cooperative clinical trials within United States and Europe (Cohn et al., 2009b).

Even though the INRG risk-classification system is comprehensive by incorporating all the relevant clinical and molecular information available till date, it is widely acknowledged that the categories result from inherent differences in tumor biology and that the classification needs to be further refined as more information becomes available. With the advent of technologies to read the genomic and gene expression patterns, understanding of the underlying tumor biology of NBL is increasing. Several studies have confirmed that information on genome and transcriptome are predictive of patient outcome and could assist risk-stratification (Abel et al., 2011; Asgharzadeh et al., 2006; De Preter et al., 2010;

Garcia et al., 2012; Oberthuer et al., 2006; Ohira et al., 2005; Spitz et al., 2006; Stricker et al., 2014). However, the challenge is to validate these findings in clinical trials before it can be incorporated into the current risk prediction strategies. When verified, it has a potential for more effective molecular diagnostics to guide optimal therapeutic plan. With the increasing knowledge of molecular basis of NBL pathogenesis, these information are slowly being added to the current risk stratification strategies and will keep being updated as new information becomes available, to guide proper therapeutic plan.

Chapter 3

Inference of gene networks from gene expression data

3.1 Introduction

The global pattern of gene expression reflects the biological state of the system, where genes with similar functions and within similar regulatory pathways are highly likely to be co-regulated and hence co-expressed. The availability of microarray and RNA-Sequencing technologies has made it possible to monitor the expression of tens of thousands of genes in parallel. With computational advances in the field of systems biology, understanding of the complexity underlying the biological processes became possible, at different levels of granularity. In this chapter, I will discuss the computational methods used in the field, that are relevant for this thesis. It will mainly focus on analysis of gene expression profiling data in the context of classification of heterogeneous tumor samples into homogeneous groups, pathway analysis approaches to understand the biology of the phenotype, reverse engineering approach to decipher the transcriptional network structure within the data and finally inference of causal transcription factors driving the phenotype.

3.2 Clustering analyses

Analysis of such gene expression data to identify the group of genes that have similar expression patterns is the first step towards classification of heterogeneous tumor samples. Dissection of such transcriptional profiles have proven useful for classification and prognosis of various tumors (Alizadeh et al., 2000; Cancer Genome Atlas Network, 2012; Golub et al., 1999; van ’t Veer et al., 2002). The molecular subclasses sharing intrinsic biological properties are discovered based on global gene expression pattern of each tumor by using various clustering methodologies with the objective of grouping the data based on the similarity of the samples over the whole set of variables (GEP) into a set of disjoint classes or clusters. It is an unsupervised learning approach, which attempts to identify the intrinsic structure within data and does not rely on training sets to classify the data and remains an important tool for class discovery.

A wide variety of clustering algorithms have been proposed to subgroup the GEPs. The GEPs consist of log transformed and normalized expression values that are measured across the full gene set represented in the microarray or RNA-sequencing data. Normalization account for variation between samples that could result from differing RNA amounts, handling and hybridization intensities etc. so that the clustering is performed on corrected values (Cheadle et al., 2003). In order to cluster the samples with similar GEP, these algorithms compute similarity or dissimilarity in gene expression patterns between each pair of samples to partition each sample into subsets (clusters).

The most commonly used statistical analyses for defining measure of similarity between GEPs are correlation coefficient, euclidean distance or mutual information. Correlation is the similarity metric, where higher correlation denotes higher similarity on a set of two-dimensional (bivariate) data. Correlation coefficient (r) is used to measure the strength and direction of relationship between the paired data, where, $-1 \geq r \leq 1$. For example, Pearson's correlation coefficient is a measure of the strength of linear relationship between the paired data. It assumes that the data is normally distributed and has a linear relationship. Spearman's rank correlation is used when the above assumptions are not met and when the data hold monotonic relationship (dependent variable either never decreases or never increases when the independent variable is increasing). It is a measure of the strength of monotonic relationship between the paired data. Mutual information provides statistical dependencies between two variables and is capable of measuring non-linear and non-monotonic relationship. Euclidean distance, on the other hand, is dissimilarity metric, where longer distance denotes less similarity between samples.

Once the proximity measurement is defined, either hierarchical or non-hierarchical clustering methods are applied. Hierarchical clustering attempts to merge smaller clusters into larger ones or split larger clusters into smaller ones in a successive manner. Hierarchical agglomerative clustering (Eisen et al., 1998) is one of the most widely used methods in functional genomics where the merging of similar clusters can be visualized as dendograms (tree) with similar samples in the same subtrees. Agglomerative clustering is a bottom-up approach, where initially each sample is assigned to an

individual cluster and in each successive steps, similar clusters are merged together to form a dendogram, until all the elements belong to the same cluster. Even though it is a powerful method for class discovery, no criteria are defined to establish the number of clusters and cluster boundaries. Other clustering methods like self-organizing map (SOM) (Tamayo et al., 1999), K-means clustering (Kanungo et al., 2000) are partition-based in that the data is separated into a set of disjoint clusters. It provides some advantages by grouping data into a predefined number of non-overlapping clusters, where the samples are assigned to a cluster based on its similarity score. However, the number of clusters needs to be defined a priori.

The limitation of clustering analyses is that there is no means of determining the confidence in the clusters number and cluster assignment obtained. The noise inherent to the gene expression profiling technologies, high dimensionality of the datasets and the biological nature of tumor type results into challenges in cluster analysis. A variety of measures have been suggested for validation of clustering results (Halkidi et al., 2001). Consensus clustering is one of the widely used algorithms, which calculates the frequency with which a pair of samples are clustered together in repeated rounds of clustering with a certain degree of permutation by gene or sample subsampling. It can be based on hierarchical or non-hierarchical clustering methods. The sample-to-sample similarity matrix derived for 2-to-N number of clusters can be visualized to confirm the number of optimal clusters and the stability of the clusters. This algorithm provides an advantage in that a consensus is achieved by repeated rounds of subsampling and clustering rather than performing of single clustering run on the dataset (Bhattacharjee et al., 2001; Monti et al., 2003). The idea is that subsets of samples when repeatedly clustered into a certain number of clusters (2 to N) by sample subsampling or the subset of genes that repeatedly cluster together by gene subsampling provides higher degree of confidence in the results.

One of the major problems in the field has been the lack of reproducibility of these clusters in samples that are independently obtained and analyzed. Thus, whenever possible, the clusters obtained from GEPs from one dataset should be assessed on similar samples in separate datasets to ensure validity of the data obtained.

3.3 Pathway enrichment analysis approaches

As described above, genome-wide clustering of GEPs provides a list of genes that exhibit similar expression pattern and are highly likely to be involved in similar cellular processes. By finding groups of smaller sets of genes from the list that are involved in similar function based on our previous knowledge of biology are more informative to understand the biological system than probing individual genes. Gene set or pathway enrichment analysis allows us to study the statistical significance of gene sets from differential gene expression (DGE) signature to infer the biological processes or pathways inducing a phenotype.

Pathway enrichment analyses help us understand the mechanism of a disease, effect of perturbation with drugs or gene manipulation, to name a few. The availability of a large number of knowledge bases aids this process. There are databases like Gene Ontology (GO) (Ashburner et al., 2000), which provides controlled vocabularies of gene products in terms of biological processes, cellular component and molecular functions. Here, the pathway refers to a list of genes associated with a GO term with no knowledge of the relationship between the genes. Other pathway analysis algorithms incorporate topology-based information to provide a graphical summary of interactions to identify categories including gene regulatory networks, metabolic networks and protein-protein interaction networks. Some of these databases include KEGG (Kanehisa and Goto, 2000), Reactome (Croft et al., 2014) and BioCarta (http://cgap.nci.nih.gov/Pathways/BioCarta_Pathways). All these knowledgebase provide a list of gene sets that could be tested on a dataset of interest to find enrichment for these processes by Overrepresentation analysis (ORA) or Quantitative Enrichment analysis (QEA), described below. Statistical significance of the findings is assessed using various methods, which are also briefly described.

3.3.1 Overrepresentation Analysis

Once a list of genes of interest is generated, the usual next step is to assess whether these genes have significant overlap with the genes mapped to known pathways such as GO biological terms, KEGG and Reactome pathway database. In other words, overrepresentation analysis (ORA) is a means of

analyzing a set of genes to assess whether they are over-represented in genes annotated to a known pathway/biological process compared to what would be expected by random chance (i.e, in a set of randomly collected equal-sized gene list annotated to the same pathway). Computation of P-value is used to assess the statistical significance of the finding, correcting for multiple testing. For ORA, it is a probability of finding x number of genes out of n number of genes annotated to a particular GO term compared to the proportion of the genes annotated to that GO term. It is usually done by Fisher's exact test (FET), which is the most commonly used hypergeometric test. It is used when the data can be divided into 2X2 contingency table to divide the genes that are associated or not associated to a particular GO term, among the gene list of interest (i.e. differentially expressed genes) from all the genes in the reference list, and calculates the probability for non-random association. The closer the value is to zero, the greater the likelihood of the association being non-random. By convention, P-values less than 0.05 is an indication of statistical significance. The uncorrected p-value is often used when we are testing if a particular function is enriched in a set of genes. The corrected p-value, on the other hand, is used when we are trying to find all the significant pathways from a large set of pathways available from the databases.

When performing a large number of tests, it needs to be corrected for multiple testing, because of the increase in Type I error (false positive). The multiple testing corrections adjust the statistical confidence based on the number of tests performed. A simple method is Bonferroni correction also called controlling for Family-wise Error rate (FWER), where, the corrected p-value = α / m, where α = original p-value and m = number of tests performed

However, this is very stringent. Another commonly used method for multiple-testing correction which is less stringent than Bonferroni correction is Benjamini-Hochberg method to control for false discovery rate (FDR) (Benjamini and Hochberg, 1995). This method controls for the false discovery by calculating the threshold p-value that indicates the amount of false positives that can be maximally expected among the significant findings. The threshold p-value corresponds to the false discovery rate (FDR). For e.g., the corrected p-value of 0.05 indicates that while the function is still statistically significant, we are

accepting a maximum FDR of 5%. For m number of tests, the original p-values (α) are ranked from largest to the smallest and desired FDR is defined (e.g, 0.05). The corrected p-value (FDR) = α x (m-rank+1) / m. The original p-value is considered significant if it is less than FDR, and the first p-value to pass this test would be the new p-value cutoff.

A limitation of ORA approach is that it gives equal weight to all the genes and does not consider the magnitude of information contributed by each gene (e.g., the level of differential expression) and assumes that each gene contributes equally to the process.

3.3.2 Quantitative enrichment analysis

The shortcomings of the ORA are overcome by quantitative enrichment analysis (QEA). Gene set enrichment analysis (GSEA) (Subramanian et al., 2005) is one of the most commonly used QEA analysis, which determines whether a pre-specified set of genes (for example, KEGG pathway, GO terms) shows statistically significant enrichment among the top or the bottom (or both) of the ranked list of genes, or is randomly distributed. In most instances, the ranked list of genes contains the differential expression of the genes across two conditions and the gene set contains a specific signature of interest. It uses Kolmogorov-Smirnoff running sum statistic test to determine whether the gene set is overrepresented among the top and/or bottom of the ranked list. In brief, the test begins with the top-ranking gene and if a gene in the gene set is encountered, the running sum increases. If it encounters a gene that is not in the gene set, then the running sum decreases. The final enrichment score (ES) represents the maximum running ES or the degree to which the gene set is overrepresented among the most differentially expressed genes. It is the maximum deviation from zero encountered in the K-S run, where the leading edge subset are the genes that are on the left (positive target enrichment) and right (negative target enrichment) of the gene with the maximum ES.

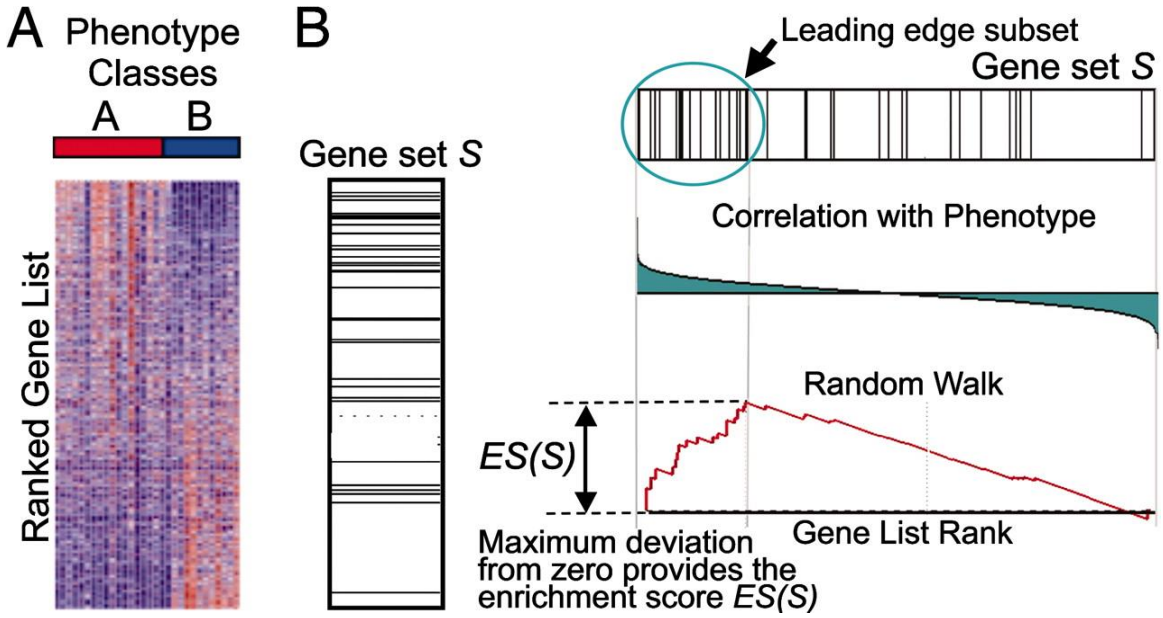


Figure 3- 1

Figure 3- 2. An overview of GSEA.

(A) Heatmap showing the gene list ranked by certain criteria such as the amount of differential expression between two phenotypes (B) Enrichment for the gene set of interest, S, is calculated by its running sum among the ranked gene list; ES denotes the maximum enrichment score and the genes to the left of these are the leading edge gene set (Subramanian et al., 2005).

The statistical significance (p-value) of the ES is determined by permuting sample labels or genes when enough samples are not available and re-computing the ES for the gene set to generate a null distribution and comparing the actual ES score to the distribution of ES scores from the permuted data. Normalized enrichment score (NES) is used to normalize the ES to account for the gene set size. A positive NES means that the gene set is overrepresented at the top of the list (upregulated GES) and negative NES means that the gene set is overrepresented at the bottom of the list (downregulated GES). In both cases, an empirical p-value is calculated. The GSEA enrichment plot provides a highly visual representation of the gene set differential expression by showing whether or not the individual genes within the gene set appear among the most differentially expressed genes. Analogous to ORA approaches, GSEA can be run on collections of gene sets, in which case, FDR is calculated to control

for false positives to adjust for multiple hypothesis testing (Benjamini and Hochberg, 1995). Both ORA and QEA assumes that each gene is independent of each other.

Single sample GSEA (Barbie et al., 2009a) is an extended form of GSEA, where the absolute enrichment of a gene set is evaluated on a single sample basis. The list of genes is ranked by absolute expression in the sample and is subsequently rank transformed. Then, GSEA is performed on any gene set of interest. It provides a way to assess the activity of a certain biological process within a sample.

3.4 Transcriptional regulatory network discovery

Transcriptional regulation plays a vital role in biological processes by specifically controlling the gene expression in response to particular biological signal. A transcriptional regulatory network refers to the regulatory interaction between the regulator or transcription factor (TF) and their target genes. These networks constitute a graph where nodes correspond to genes and the edges correspond to the interaction between the two. Understanding such regulatory architecture is crucial to elucidate cellular phenotypes of interest. These transcriptional networks are highly context dependent and hence these regulatory network need to be built for the specific cell type of interest (Basso et al., 2005).

3.4.1 Reverse engineering algorithm- ARACNe

Various network reconstruction methods have been proposed to retrieve the transcriptional interactions (Emmert-Streib et al., 2012). Direct experimental evidences confirming the targets of the TFs, for example, by chIP-Seq experiments enable genome-wide identification of protein-DNA interaction, which could then be used to build the network. However, to generate a comprehensive set of these experimentally validated protein-DNA interactions is expensive, labor intensive and is further limited by the availability of good antibodies to perform chromatin-immunoprecipitation. On the other hand, reconstruction of regulatory network based on GEP data is called reverse engineering or back engineering, where the information is extracted based on the observation of the behavior of its individual components. Since the abundance of TF protein is often not available, it is approximated by the statistical associations between their mRNA abundance levels. There are several algorithms to infer

these transcriptional interactions based on the statistical dependencies between random variables, which are discussed in detail in literature (Emmert-Streib et al., 2012; Gardner and Faith, 2005), and the ones used for our study are discussed below.

ARACNe (Algorithm for the Reconstruction of Accurate Cellular Networks) is an information theoretic approach that allows reverse engineering of transcriptional network or direct physical transcriptional interaction between the transcription factors and their targets based on GEP datasets (Margolin et al., 2006). It computes mutual information (MI), which is a probabilistic approach to measure the statistical dependency between two variables, in this case, between the TF and each gene available in the GEP. In order to establish statistical significance of the dependency between the two variables, a null distribution is built through gene/sample shuffling and an empirical p-value is calculated. Since ARACNe runs a combinatorial search (TFs by genes), multiple testing correction is needed, and a MI threshold is established according to corrected p-value (p-value <0.05; Bonferroni corrected for the number of tested pairs). The interactions below a MI threshold are excluded. Even though MI does not provide the directionality of the relationship, the non-TF is assumed to be the target of the TF. Additionally, MI does not establish whether an interaction is direct or mediated by intermediate nodes. In order to eliminate indirect interactions, ARACNe implements another information theoretic approach, Data Processing Inequality (DPI), which provides relation between MI values. This is performed on every possible gene-triplets (three genes with mutual information >0), where the edge corresponding to the lowest mutual information is removed, hence resulting into final network representation. Also, in order to adjust for the noise coming from the microarray technology as well as error in MI estimation, it uses bootstrap sampling for network reconstruction. The bootstrap datasets are generated by randomly selecting the samples/GEPs with replacement from the original set. ARACNe is then applied on these datasets to generate separate bootstrap networks and the interactions that exist across significant number of these bootstrap networks are used to build a consensus network. This gives us the “regulon” or the set of genes that are under the control of each TF. These regulatory networks have been validated in several contexts (Basso et al., 2005; Carro et al., 2010). These networks are highly context specific as it depends on the expression and association of the TF and the target genes in the tissues tested.

3.4.2 Master Regulator analysis

Cellular phenotypes are regulated by concerted effect of differentially expressed genes. Even though this list gives us a starting point to understand the biological system, the transcriptome landscape is very unstable as they are downstream of the complex network of regulatory cascades, hence carrying over the noise at each level of regulation. Moreover, the noise from the microarray technology also contributes towards the differential expression. The important question is to identify the TFs that control the GES of the phenotype of interest. While the TF activity sometimes correlates with its differential expression, it does not recapitulate the post-transcriptional, post-translational and cellular compartment localization information that affects its protein activity. Hence, instead of searching for TFs that are differentially expressed, identification of TFs that are causally responsible for inducing the GES are of more value. These functional TFs are called the master regulators (MRs), as they are necessary and sufficient to initiate or maintain a tumor state. MRs represent the regulatory bottlenecks by integrating the upstream signaling from the genetic/epigenetic code and signaling molecules. The pre-requisite for identification of these MRs is the need for reliable context specific genome-wide transcriptional regulatory map, as provided by ARACNe algorithm, described above.

The algorithm developed in the lab, Master Regulator Inference algorithm (MARINA), identifies the transcription factors whose targets (based on ARACNe prediction) are enriched in the GES defined by the differentially expressed genes between two phenotypes of interest (Lefebvre et al., 2010). Firstly, statistical methods such as limma (Smyth, 2005) or student's t-test are used to derive the differential gene expression (DGE) signature. Student's t-test is a simple parametric statistical test used to derive the statistical significance of the differential expression between two groups by comparing the means of the two distributions. It assumes that the population is normally distributed and the variance between the two groups is equal, and the distribution assumption is often true for GEP datasets. Similarly, limma implements a number of statistical principles to first $\log_2$ transform the normalized counts, determine the weight of each gene by estimating the mean-variance relationship to fit the expression values of each gene into a linear model. Benjamini-Hochberg method is used for FDR estimation. Secondly, ARACNe is used to derive genome-wide TF-target interaction map. MARINa tests whether the change in the

activity of the TF leads to enrichment of its regulon in the DGE signature. The enrichment is evaluated by GSEA, for which, the GES is first ranked from the most downregulated to the most upregulated genes. Then GSEA is performed to simultaneously assess whether the positive and negative regulon are respectively enriched in the genes that are overexpressed or underexpressed in the GES, and the MRs are ranked by their p-value.

Similarly, another algorithm developed recently, Virtual Proteomics by Enriched Regulon analysis (VIPER) (Alvarez et al., manuscript in review), exploits the same principle as MARINa and infers protein activity by systematically constructing and analyzing the activity of a TF based on the expression of the target genes (regulon) that are directly regulated by the TF. The algorithm uses a probabilistic framework to account for mode of regulation (i.e, if the target is activated or repressed by the TF), statistical confidence in regulator-target interactions, and target overlap between different TF regulators. It provides an enrichment analysis framework, supporting weighted contribution of each gene in the regulon. To ensure robustness and computational efficiency, mean of ranks of the regulon is used as test statistic and hence this method is called analytic Rank-based Enrichment Analysis (aREA). Therefore, the information from activated, repressed, and undetermined targets are integrated probabilistic weighting of individual genes and provides computationally efficient means of determining the activity of a TF. Differential protein activity is thus quantitated by VIPER as the normalized enrichment score (NES) computed by the aREA algorithm. It can perform TF activity analysis both for multiple samples (msVIPER), and on a sample by sample basis (VIPER), thus allowing analysis of relative activity of each protein in each sample compared to the rest of the samples.

Chapter 4

Molecular classification and master regulator identification of high-risk neuroblastoma

4.1 Background

While significant improvements in outcome have been reported for low and intermediate-risk NBL patients, high-risk group of patients have proven refractory to the current treatment modalities. Understanding the disease at the molecular level will allow us to find targeted therapeutics, thus minimizing the side effects of current therapies for high-risk NBL. The major challenge is to dissect the complexity and heterogeneity of these tumor cells to find the driver genes and activated pathways that are essential for survival of these cells. Unlike other adult tumors, pediatric tumors generally have very low mutational burden, hence challenging the current therapeutic strategies to target tumors with specific oncogenic or tumor suppressive mutations (Maris, 2010; Pugh et al., 2013).

NBLs commonly harbor structural variations affecting large portions of specific chromosomes and accumulation of copies of V-Myc Avian Myelocytomatosis Viral *Oncogene* Neuroblastoma Derived Homolog (MYCN) oncogene. These are highly recurrent events used in risk stratification strategies (Brodeur, 2003; Brodeur and Bagatell, 2014; Mosse et al., 2007; Schleiermacher et al., 2012). Structural variants are the basis of classification of NBLs into three subtypes- subtypes 1, 2 and 3. Subtype 1 lacks segmental chromosomal alterations (SCA), expresses high level of TRKA and is associated with low-risk clinical phenotype. Subtypes 2 and 3 have unbalanced gain of 17q and overexpression of TRKB, and are associated with high-risk and poor prognosis. They harbor mutually exclusive segmental alterations; subtype 2 with segmental loss of chromosome 11q, 3p and 4p; and the more aggressive subtype 3, with deletion of chromosome 1p and *MYCN* amplification (Brodeur and Bagatell, 2014; Maris and Matthay, 1999a; Michels et al., 2007).

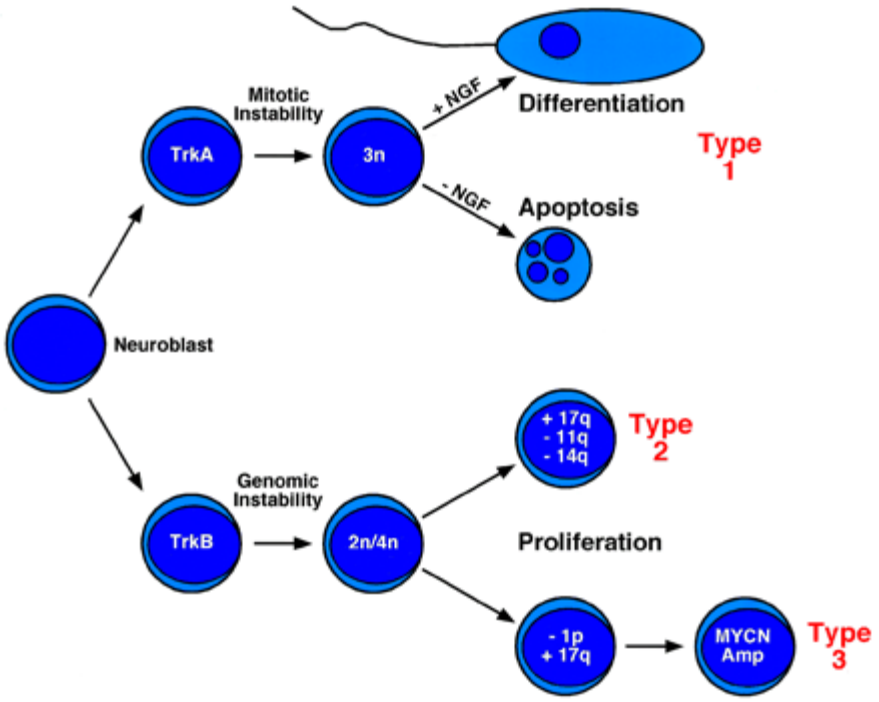


Figure 4- 1. Neuroblastoma subtypes by genomics

The low risk patients (Type I) expressing TRKA gene are more likely to have whole chromosomal gains and can undergo differentiation or apoptosis depending on the presence or absence of NGF in their microenvironment. High-risk patients express TRKB and exhibit chromosomal instability. Most of these tumors have unbalanced gain of 17q chromosome and they can be further stratified into Type 2 tumors with 11q deletion and Type 3 tumors with MYCN amplification and 1p deletion (Maris JM and Matthay KK., 1999).

Despite significant increase in the body of knowledge of NBL genetics, all high-risk patients follow similar therapeutic procedures and little advancement has been made on molecular information based adaptive therapies. The clinical heterogeneity and the resulting subtypes in NBL is most likely the result of the genomic alterations and the associated regulatory bottlenecks that keeps the high-risk NBL cells in highly proliferative and undifferentiated state; the confluence of which can be captured by its dynamic transcriptome landscape. While most of the efforts in transcriptome based analyses are driven towards better risk classification of NBL (Abel et al., 2011; Asgharzadeh et al., 2006; Oberthuer et al., 2015;

Valentijn et al., 2012), systematic investigation of gene expression patterns of high-risk NBL to identify the subtypes and transcriptional regulators inducing the signatures are lacking. Therefore, systems-level dissection of these MRs in high-risk NBL is essential for (a) understanding the mechanisms of tumor malignancy and (b) developing therapeutic agents that are specific for the patient subgroups.

In recent years, several studies have shown that heterogeneous cancers can be clustered into distinct and homogeneous molecular subtypes based on genetic and gene expression profiling (Cancer Genome Atlas Network, 2012; Monti et al., 2005; Lapointe et al., 2004). This can be explained by the existence of transcriptional and post-transcriptional regulatory modules that integrate the aberrant signals from the upstream signaling molecules to drive gene expression of a particular subtype. We hypothesize that the signals transmitted by heterogeneous set of genetic and epigenetic alterations within the cancer cells convene on a few TFs or master regulators (MRs), which drive the gene expression of a particular subtype. Recent advances in systems biology approaches for assembly, interrogation and perturbation of these regulatory pathways has been shown to successfully identify the transcriptional modules driving normal and pathological phenotypes (Aytes et al., 2014; Carro et al., 2010; Lefebvre et al., 2010). In this chapter, I will discuss how we analyzed the gene expression profiles (GEP) of high-risk NBL – to identify the molecular subtypes and the associated master regulator (MR) modules, constituting the natural Achilles' heels for the high-risk subtypes.

The computational analyses described in this chapter were performed by Dr. Gonzalo Lopez.

4.2 Results

In order to recapitulate the heterogeneity of high-risk NBL, we performed systematic dissection of high-risk NBL gene expression profiles (GEPs) to identify the molecular subtypes and master regulators driving each subtype. Furthermore, we performed comprehensive analysis of the co-segregating clinical features as well as pathway enrichment analysis within each subtype.

Datasets

Motivated by the need to understand the basics of high-risk NBL and facilitate discovery of molecular targets of therapeutic value, the NCI Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative (http://target.cancer.gov), gathered a patient cohort enriched in high-risk advanced tumors as defined by COG. These are patients who have clinical stage 4 disease, with late onset (> 18 months) and/or harboring *MYCN* amplification. Relevant for this work, the TARGET dataset comprises 646 samples for which comprehensive clinical information is available, representing 577 high-risk tumors and 65 low-risk tumors (Stage I). It comprises both gene expression profiles using Human Exon 1.0 ST array (n=249 of which n=219 are high risk tumor samples) paired with Illumina SNPs profiles (n=276, of which n=224 are high risk tumor samples). Tumor samples from low-risk, clinical stage 1 patients function as a control group.

In another significant effort, the European group, Neuroblastoma Research Consortium (NRC) collected a large NBL patient cohort from all tumor stages, which were profiled to obtain gene expression profiles using Human Exon 1.0ST array (n= 279) and paired CGH arrays (n=219), also reported with clinical information. In order to maintain consistency for our computational analyses, we filtered the samples to include a total of 97 high-risk patients following TARGET high-risk criteria as determined by COG, with an additional 30 samples matched to TARGET low-risk samples, used as a control. The remaining samples, which couldn't be classified into either high-risk or low-risk groups based on TARGET classification, were still used to create neuroblastoma interactome (NBi) and survival analyses but not for subtype and MR identification.

The addition of stage 1 tumors to the analysis is fundamental, provided that normal tissue is not available. Nonetheless, we reason that this group of tumors conform an ideal reference subtype, presenting a differentiated phenotype, lack of segmental chromosomal alterations and overall good prognosis. This ultimately allowed us to obtain differential gene expression (DGE) profiles for each high-risk subtype, representing individual prognostic/progression signatures.

4.2.1 Molecular classification of high-risk NBL

We used the GEPs of high-risk NBLs from TARGET and NRC dataset for subtype identification. The relative expression for a gene in a sample is derived by z-score transformation. We performed unsupervised clustering by selecting ~2,500 genes with coefficient of variation >0.1, which represent genes with high variability in expression across samples, relative to the average expression for that gene. These genes were used to perform consensus clustering (Monti et al., 2003), separately on the two datasets (Figure 4-2).

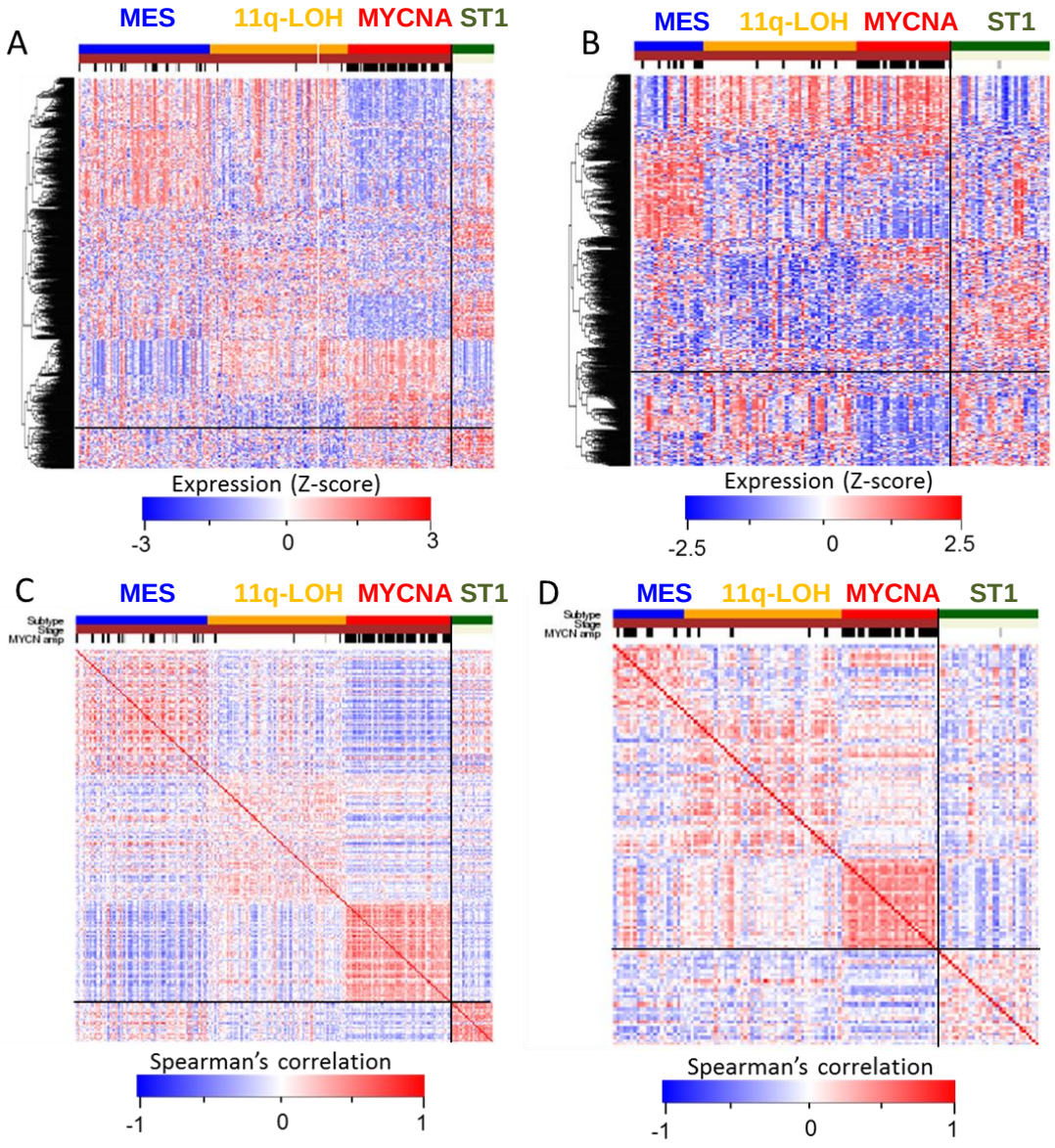


Figure 4- 2. Clustering of high-risk NBL primary tumor GEPs from two patient cohort

Unsupervised consensus clustering was applied on high-risk NBL affymetrix HuEx 1.0 ST microarray sample profiles. Optimal division with reproducibility across cohorts was assessed in three group clustering; Stage 1 samples were not included in the clustering analysis, also shown in the heatmaps. (A)(B) Gene level hierarchical clustering of expression Z scores for highly variable genes filtered by inter-quantile ratio (IQR 4th quantile) comprising 4331 genes from (A) TARGET and (B) NRC high risk cohort respectively. (C)(D) Spearman correlation matrices of (C) TARGET and (D) NRC cohorts respectively.

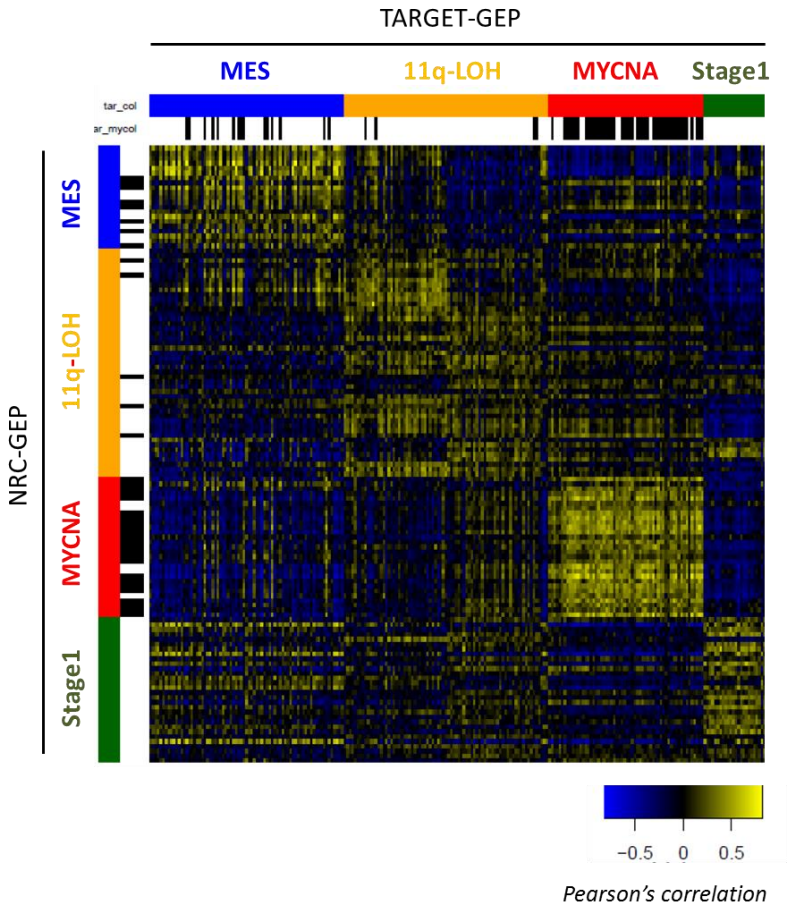


Figure 4- 3. Implementation of unsupervised consensus clustering of high-risk NBL GEPs to establish molecular subtypes.

Three subgroups were identified according to robustness of clustering and consistency between two cohorts (TARGET and NRC). Cluster validity was measured by correlation between the GEPs from TARGET (x-axis) and NRC (y-axis) cohort on a sample-by-sample basis.

Consensus matrices were derived for 2-6 clusters and assessed for optimal cluster number and cluster stability. Our results showed that both datasets optimally clustered into three distinct subgroups (Figure 4-2). In order to assess cluster reproducibility, we studied across-dataset similarities of their GES by performing 1:1 correspondence analysis between TARGET and NRC GEPs. There was a striking overlap in the subtypes in both datasets (Figure 4-3).

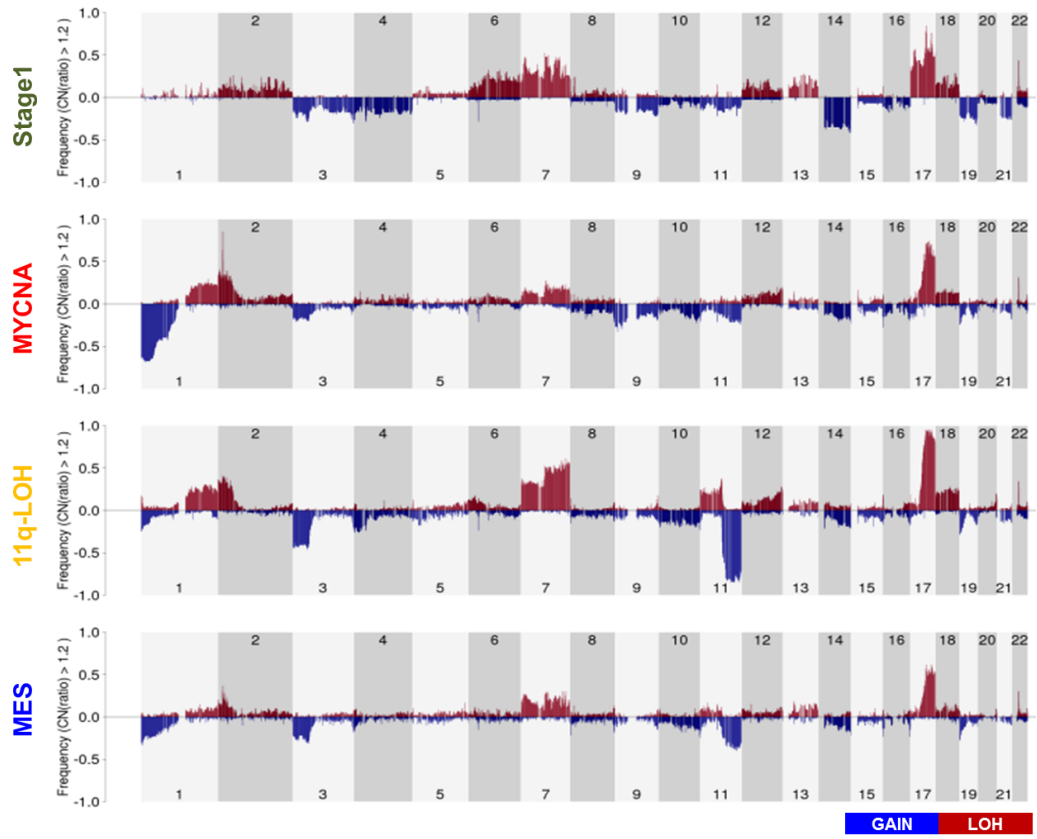


Figure 4- 4. Segmental chromosomal alterations in high-risk NBL subtypes.

Copy number frequency per genomic location (integrated from both TARGET and NRC data) of individual molecular subtypes showing segregated pattern of 11q, 3p and 1p loss. The data represents the log ratio of the quantity of DNA in the sample versus the normal control. The normal copy number where the sample has copy number equal to that of normal control equals 0. The measurements of copy number greater than 0 represent gain of copy number and less than 0 represents loss of copy number with equal number. Copy number gains are considered when the log2 ratio between tumor and blood > 1.1 while losses are considered for log2 ratios < 0.9. In order to compute copy number frequencies for large segments (Figure S1G, S1H), we obtained the average log2 ratio for genes mapping to each segment and 0.9 (loss) and 1.1 (gain) cutoffs apply.

As segmental chromosomal alterations are characteristic of NBL tumors, we assessed if the clusters co-classify with the overall frequencies of copy number alterations in these patients. For this, we analyzed the overlap of copy number profiles derived from SNP arrays (TARGET) and CGH arrays (NRC) with their respective GEPs. Copy number gain and losses are first assessed in the individual samples and the frequency of copy number alteration of each gene across the samples within each subtype (Figure 4-3), is averaged to derive the relative frequency of the alterations across the genome (Figure 4-4). We observed unbalanced loss of 1p, 3p and 11q and unbalanced gains of 1p, 2p and 17q chromosomal arms, as previously reported (Brodeur and Bagatell, 2014; Maris and Matthay, 1999b). The copy number variations differed significantly between the subtypes and clustered into specific regions, hence resulting into recurrent chromosomal imbalances. As previously described, all high-risk NBLs displayed unbalanced gain of 17q arm, with 11q-LOH subtype showing higher dosage compared to the other subtypes.

One cluster co-segregated with *MYCN* amplification status and hence will be referred as *MYCN*-amplified (MYCNA) subtype (Figure 4-3, 4-4). It comprised 78% and 83% of the subtype specific tumors in TARGET and NRC, respectively. As previously reported, these samples exhibit partial amplification of 2p arm (where *MYCN* gene resides) and 1p deletion (Brodeur, 2003). This subtype comprises mostly of *MYCN*-amplified tumors. However, MYCN and MYC overexpression have been reported to transcriptionally regulate similar set of target genes by binding to E-box sequence (CACGTG) in the target gene promoter (Valentijn et al., 2012). It has been reported that in NBLs, expression of MYCN and MYC expression are mutually exclusive by binding to the promoter of each other and repressing it (Breit and Schwab, 1989). A few samples with overexpression of MYC co-classifies into MYCNA subtype. Similarly, the activity of MYC/N is highest in MYCNA subtype tumors (Figure 4-5).

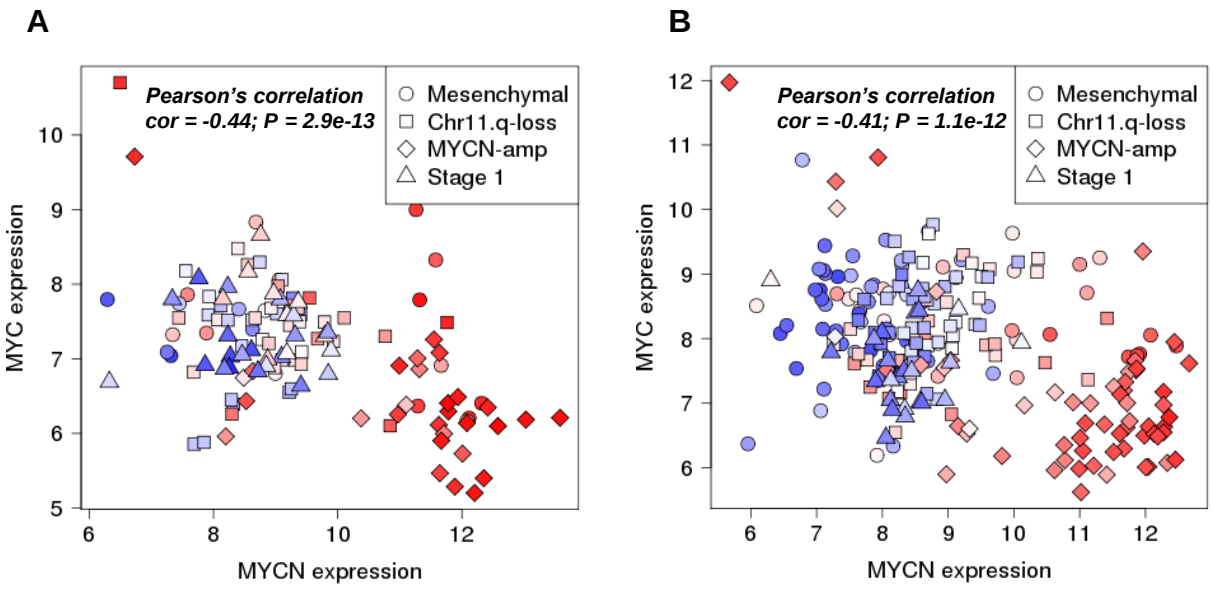


Figure 4- 5. MYCN and MYC expression and MYC/MYCN activity in NBL subtypes.

(A) TARGET and (B) NRC samples show negative Pearson's correlation (showed in text in the upper left plot area) of MYCN versus MYC expression. MYC/N activity is expressed in blue to red colors; activity was measured as the single sample normalized enrichment score (NES) of a MYCN functional signature (Valentijn et al., PNAS, 2012) using GSEA after rank transformation of the gene expression signatures.

Similarly, the second cluster with 90% prevalence of loss of Chr. 11 long arm (10% LOH cutoff), could be segregated by Chr11q loss of heterozygosity (LOH) and hence is referred as Chr11q-LOH (11q-LOH) subtype (Figure 4-3, 4-4). As reported previously, this alteration co-exist with 3p deletion (Brodeur and Bagatell, 2014).

Interestingly, we also recovered a third subtype that is not defined by any marked genomic alteration. We refer this subgroup as Mesenchymal (MES) subtype, the details of which will be presented below.

4.2.1.1 Activated and deactivated pathways in distinct molecular subtypes

The differential expression of genes in the subtypes and the genomic alterations associated with it suggests differential activity of biological processes in these subtypes. The availability of a large number of knowledge base driven pathway analysis methods facilitate this process. We evaluated the DGE signature between each subtype compared to Stage I subgroup by two methods: Gene Ontology (GO) biological processes terms and REACTOME biological pathway enrichment analysis. This was performed by GSEA analysis on the gene sets for all the GO and REACTOME categories within the subtype specific GES, and corrected for multiple testing via FDR estimation (Figure 4-6).

MYCN-amplified subtype: Both REACTOME and GO analyses revealed that this subtype was characterized by activation of cell cycle programs and DNA replication program indicating that these cells must have high proliferation rate to increase cell numbers. Similarly, RNA and protein metabolism programs were also activated to support the high cell growth rate of these cells (Figure 4-6). These two processes have been shown to be complementary and independent of each other (Dowling et al., 2010; Su and O'Farrell, 1998). This subtype also showed activation of DNA repair program and the interplay between DNA replication and repair programs has a significant impact on maintenance of genomic integrity (Lowndes and Murguia, 2000).

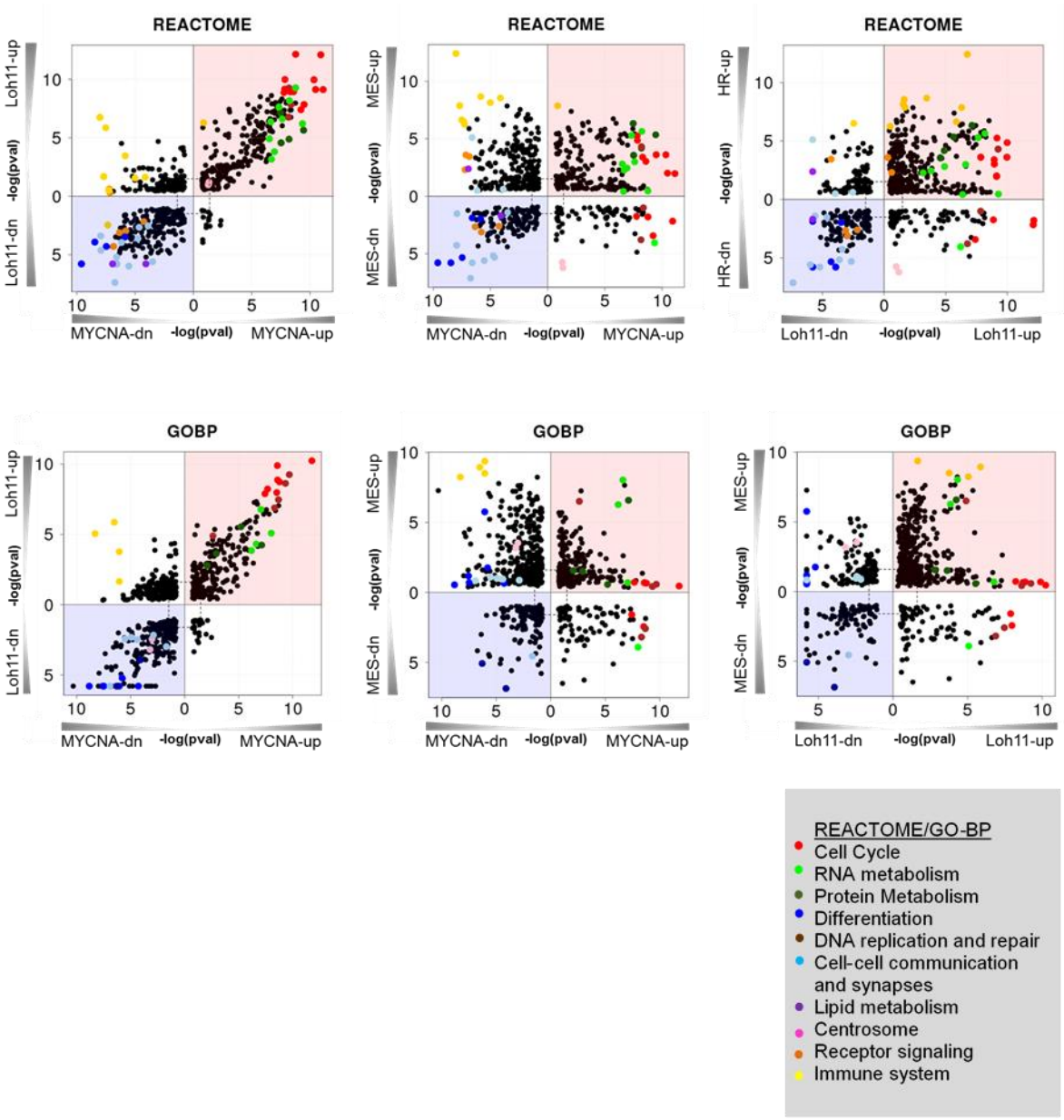


Figure 4- 6. Activated and repressed biological programs in the high-risk NBL subtypes. Gene set enrichment analysis was performed on REACTOME pathway and GO biological processes gene sets to find enriched pathways for the DGE signature of MYCNA, 11q-LOH and MES subtype. Differentially activated or repressed pathways are compared between subtypes. Axis represents $-\log_{10}$ of the p-value while retaining the directionality of the normalized enrichment score. The dashed line represent FDR cutoff of <0.05.

Similarly, this subtype was marked by repression of differentiation pathways, cell to cell communication and synapses, lipid metabolism and immune system pathways (Figure 4-6). The association of formation of synapses and neuronal differentiation is well known where the differentiating cells have neurite outgrowth to reach out to the neighboring cells. Also, lipids like cholesterol has been shown to be less abundant in undifferentiated compared to differentiated neuroblastoma cells (Gulaya et al., 1989). And cholesterol has been shown to enhance the differentiation process in neuroblastoma cells (Sarkanen et al., 2007). Hence, all these repressed processes in MYCNA subtype seemed to be intertwined and cross-regulatory.

Chr11q-LOH subtype: Similar to MYCNA subtype, this subtype showed activation of cell proliferation and cell growth programs and repression of differentiation programs. Overall, the pathway analysis pattern is very similar to that of MYCNA subtype (Spearman correlation = 0.79) (Figure 4-6), only differing in a stronger activation of immune related pathways possibly due to a more infiltrated histology, as discussed below. In order to assess whether the immune signal is represented by the entire samples within the subtype or is contributed by a subset of samples, we assessed the expression of the gene sets previously reported for markers of immune cells (Yoshihara et al., 2013). These gene sets were derived by identifying the genes that were upregulated in various types of immune cells when compared to their expression in normal tissues. This gene set was then crossed with the leukocyte methylation signature previously reported to correlate with leukocyte infiltration in ovarian cancer to find the overlapping genes to finally derive the immune signature (Carter et al., 2012). Indeed, we observed that subsets of the samples in this subtype were highly infiltrated by immune cells. In addition, these samples were also assessed for stromal infiltration. The stromal gene set were derived by comparing the paired tumor samples with its stromal component in breast, ovarian and colorectal cancer (Tothill et al., 2008). The tumor samples with immune infiltration also displayed stromal infiltration.

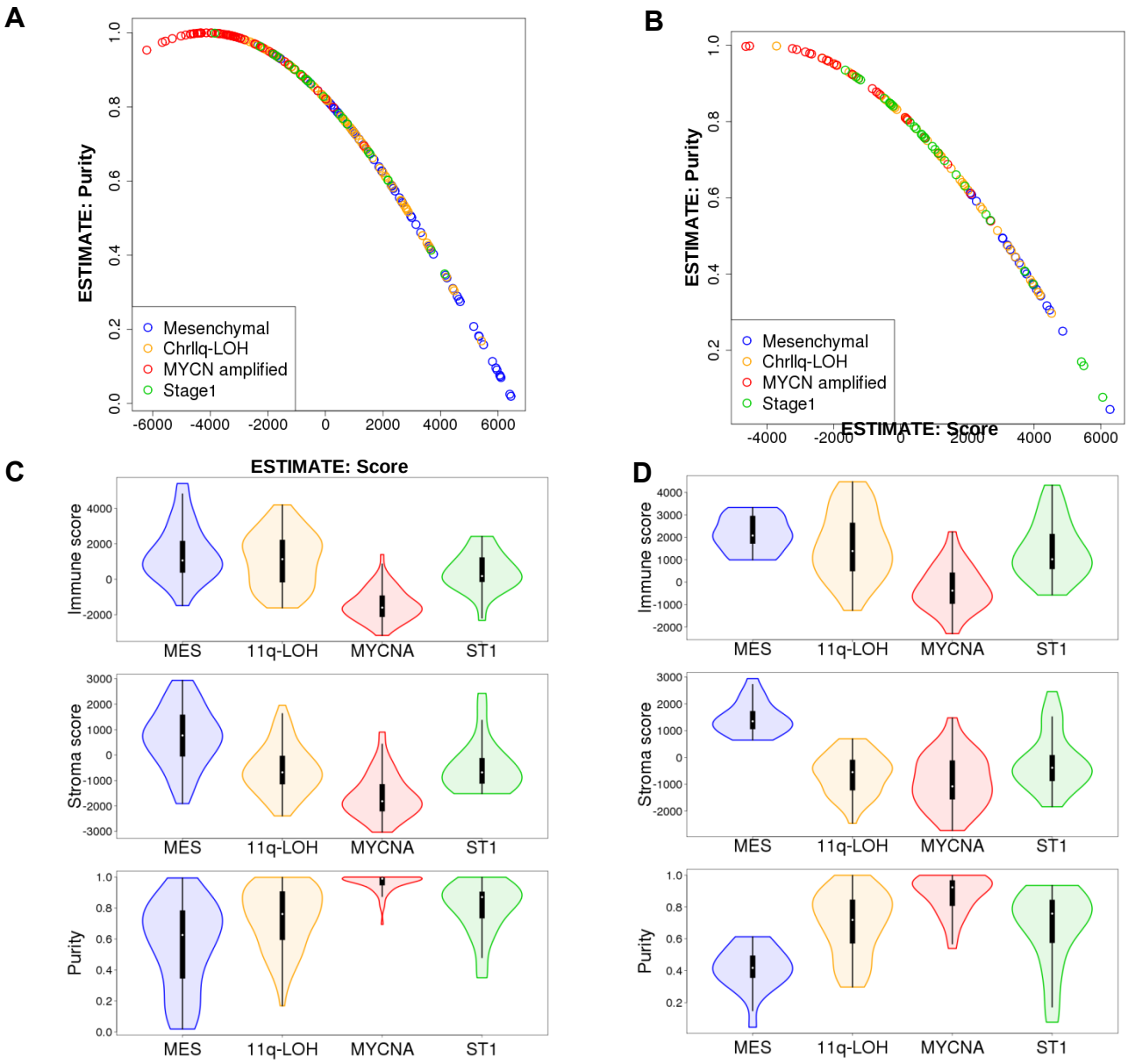


Figure 4- 7. Measure of immune and stromal infiltration in NBL subtypes.

Estimation of tumor purity in terms of immune and stromal cell infiltration by using the algorithm ESTIMATE, where score of 1 corresponds to highest purity in (A) TARGET) and (B) NRC patient samples divided by subtypes. (C, D) Determination of immune score, stromal score and corresponding purity of the patient samples from (C) TARGET) and (D) NRC cohort.

The sample purity was further assessed by using ESTIMATE algorithm, which estimates the tumor purity based on enrichment of the immune and stromal signature in the individual samples (Figure 4-7-A, B) (Yoshihara et al., 2013). We observed that 11q-LOH and MES subtype sample were highly infiltrated while, MYCNA subtype displayed high degree of purity (Figure 4-7-C, D). MYCN has also been shown to repress expression of CCL2, a chemokine required for natural killer cell chemoattraction (Song et al., 2007).

Mesenchymal subtype: Unlike the other two subtypes, this subtype didn't display hyper activation of proliferative and cell growth programs. However, this subtype also showed repression of differentiation programs indicating that all high-risk NBL tumors have undifferentiated programs. This subtype displayed activation of immune pathways more strongly than 11q-LOH subtype. Evaluation of tumor purity by immune and stromal signature as well as ESTIMATE score indicated that all the samples in this subtype were highly infiltrated (Figure 4-7). Recent studies showed that the MES subtype of gliomas have more infiltrating lymphocytes compared to the other subtypes and were more responsive to dendritic cell immunotherapy (Prins et al., 2011). Hence, we further evaluated whether this subtype exhibit mesenchymal GES. Indeed, this subtype presented a strong MES GES reported previously for glioma (Phillips et al., 2006) (Figure 4-8 A). In order to disentangle tumor specific signatures from those of tumor infiltrating components to confirm that the signal is intrinsic to neuroblastoma cells, we performed similar analysis in a panel of NBL cell lines. We studied whether the relevant signatures can be recapitulated in the NBL cell lines using co-expression analysis. Remarkably, while the expression of immune and stromal related signatures disappears in cell lines, the mesenchymal component remains intact, confirming that the MES signature is tumor cell autonomous (Figure 4-8 B).

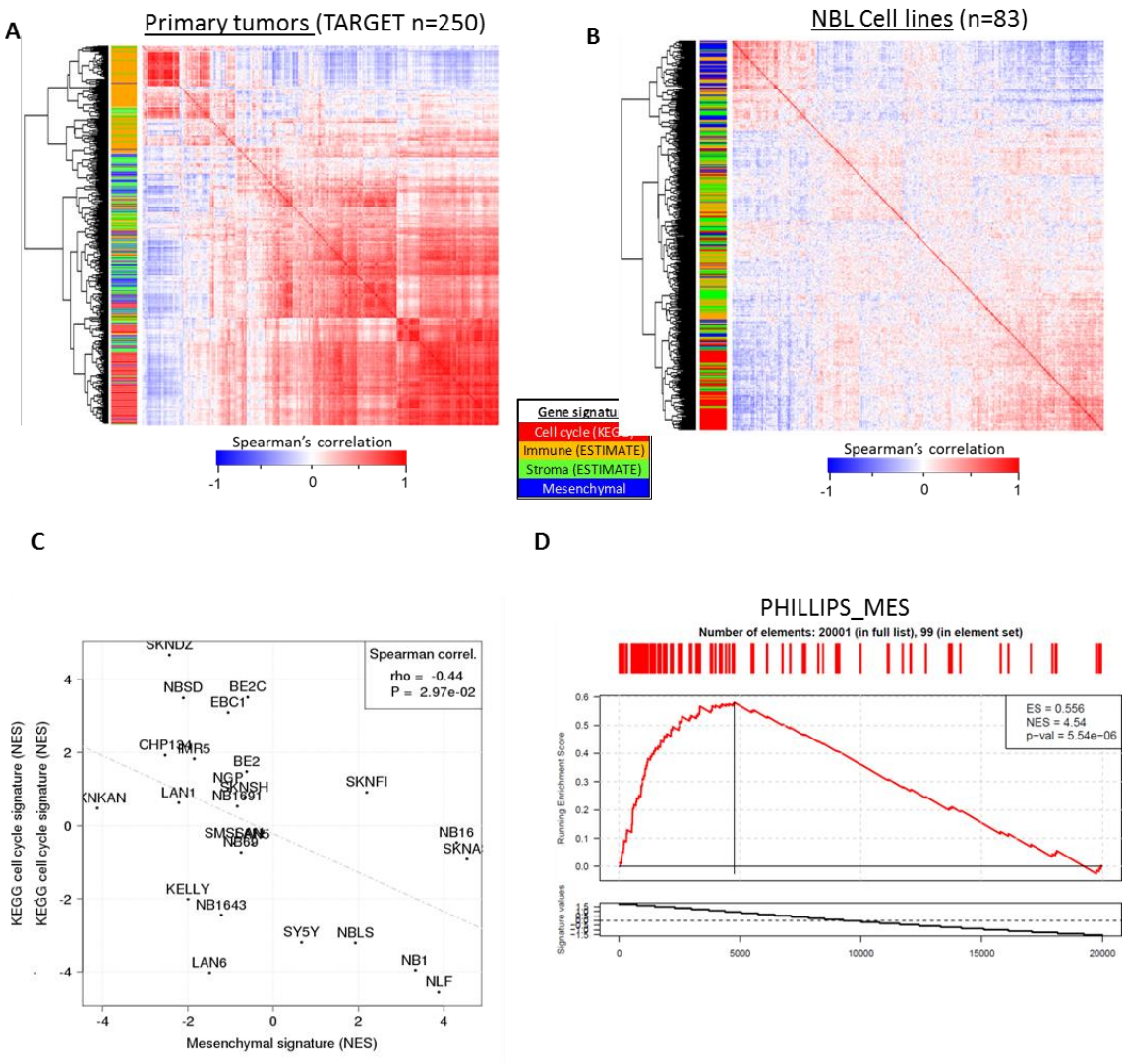


Figure 4- 8. Mesenchymal gene expression signature in patient samples and cell lines.

(A) Co-expression analysis of molecular subtype segregating signatures in tumor patient samples show independent patterns of expression for cell cycle, mesenchymal and infiltrating signatures (immune and stroma) (B) Co-expression analysis in NB cell lines show orthogonal expression of cell cycle and mesenchymal genes and lack of co-expression of infiltrating signatures (C) single cell line GSEA analysis representing NES of mesenchymal and KEGG cell cycle signatures in the CHOP panel of 25 NB cell lines indicate negative correlation between mesenchymal and proliferative activities in cell lines (D) GSEA plot showing enrichment of MES-GES in a cell line representing MES subtype, SK-N-AS.

Similarly, single sample GSEA analysis performed on the cell lines indicated that the MES and proliferative signature were negatively correlated (Figure 4-8 C). The analysis identified SK-N-AS as the cell line with strongest enrichment for MES signature (NES = 4.54, p-value = 5.5e-6), compared to patient derived samples (Figure 4-8 D). With all these evidences, we refer this subgroup as mesenchymal (MES) subtype (Figure 4-3, 4-7, 4-8).

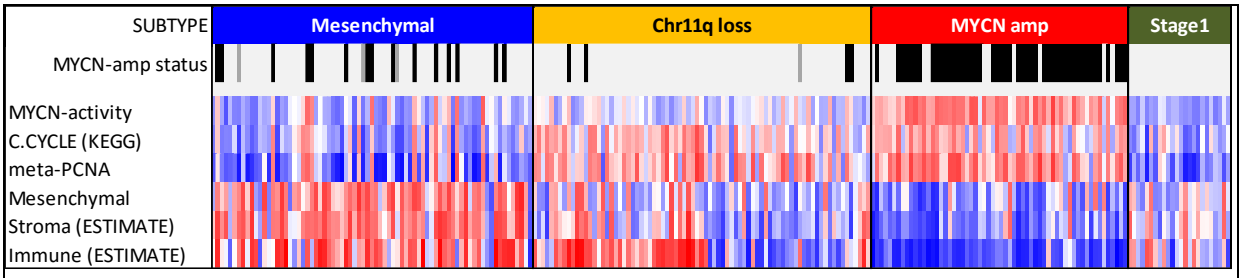


Figure 4- 9. Summary of activated biological programs in the high-risk NBL subtypes.
GSEA enrichment analysis for gene sets defining proliferative, mesenchymal, immune and stromal infiltration programs were performed in the GES of MYCNA, 11q-LOH, MES and Stage I patient samples. Red color denotes activation and blue denotes repression of the respective programs.

In summary, MYCNA and 11q-LOH subtype exhibited strong enrichment for proliferative programs based on GSEA analysis of cell cycle gene set from KEGG database, meta-PCNA signature, comprising a gene list set that is most positively correlated with the proliferation marker, PCNA (Venet et al., 2011). 11q-LOH and MES subtype displayed strong enrichment for stromal and immune signature (Yoshihara et al., 2013); and MES subtype displayed strong enrichment for mesenchymal signature defined for high-grade gliomas (Phillips et al., 2006) (Figure 4-9).

4.2.1.2 Clinical features co-segregating with the subtypes

The distinct molecular subtypes of high-risk NBL complemented with differential pathway activation indicate that this classification discriminates the underlying biology and has functional significance. In order to evaluate whether the pathways relevant for each subtype results into changes in phenotype, we evaluated the clinical profile of the patient samples in each of the subtypes we discovered.

We observed that the patients in all the three subtypes had poor outcome in both TARGET and NRC dataset based on Kaplan Meier curve analysis (Figure 4-10 A, F). This is consistent with poor prognosis reported for all high-risk group of NBL. All the subtypes had unfavorable histology based on Shimada index, which is a histopathological system used to classify neuroblastic tumors based on degree of differentiation of the tumor cells, presence of stromal cells and Mitosis karyorrhexis Index (MKI) index (Figure 4-10 B). All the subtypes showed undifferentiated phenotype (Figure 4-10 C), consistent with the pathway analysis showing that all the subtypes had downregulation of differentiation programs (Figure 4-7). MYCNA, 11q-LOH and MES subtype, displayed highest frequency for high, intermediate and low MKI respectively (Figure 4-10 D). While the frequency of mitotic vs. karyorrhectic cells are not known, the MKI index has been shown to be highly correlated with mitotic index and hence is used to estimate the mitotic index that is used by Shimada classification (Joshi et al., 1996). Again, this profile correlates with the activation of proliferative pathways in these subtypes, with MYCNA and 11q-LOH subtype being more proliferative than MES subtype.

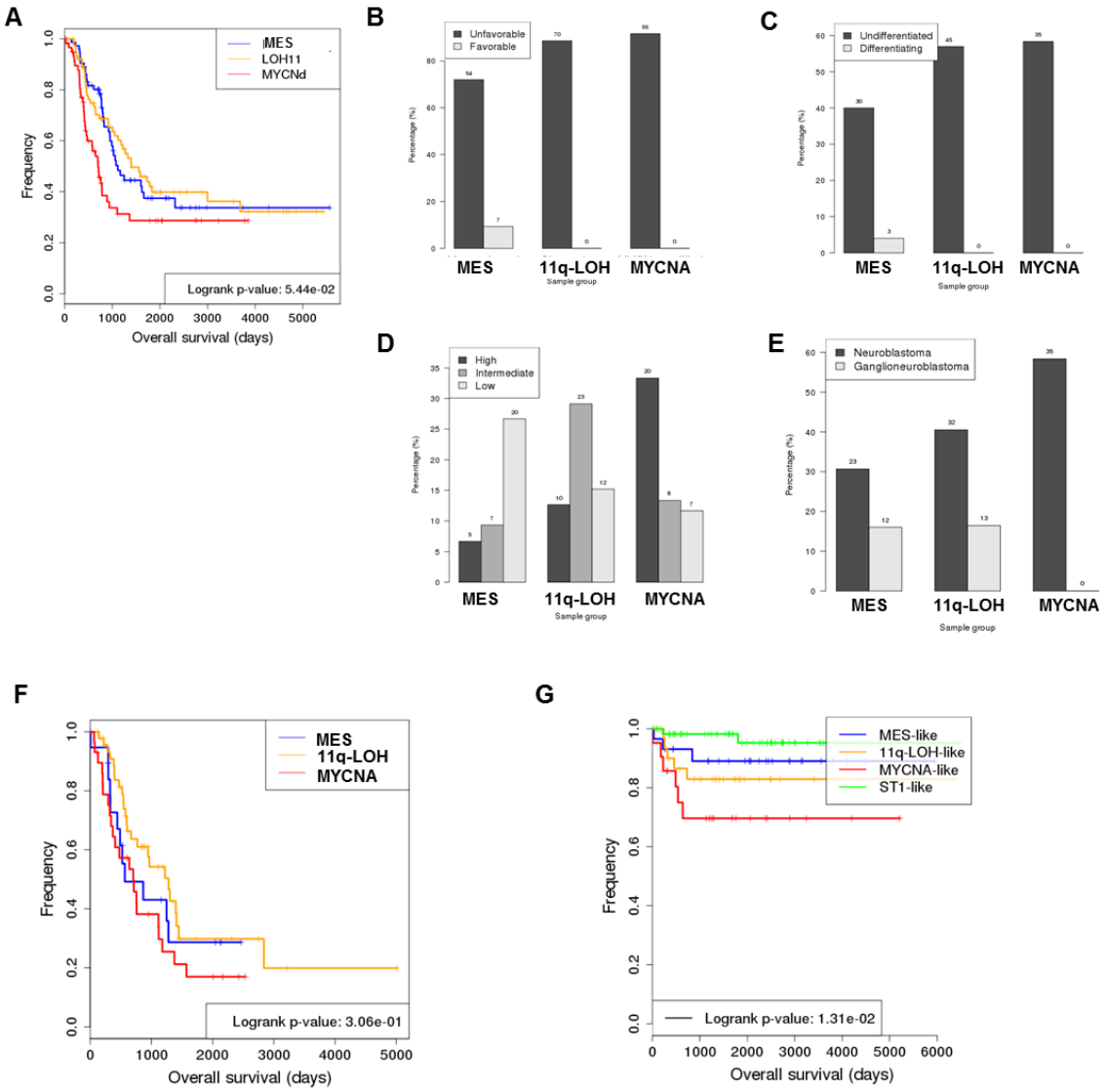


Figure 4- 10. Clinical significance of high risk molecular subtypes

(A) Kaplan-Meier and survival analysis across high-risk subtypes shows borderline separation in TARGET dataset (logrank test, P~0.05). Co-segregation of histological variables assessed by frequency of (B) favorable and unfavorable histology (C) differentiation (D) and mitosis-karyorrhexis index (MKI) (E) ganglioneuroblastoma samples (F) Kaplan-Meier and survival analysis across high-risk subtypes shows no separation in NRC dataset (G) Kmeans clusters of intermediate risk samples derived from high risk average signature centroids show separation in a Kaplan-Meier survival analysis (logrank test, P=0.013)

Lastly, neuroblastic tumors can be discriminated as neuroblastoma or ganglioneuroblastoma based on their histology (degree of neuroblastic maturation and presence of schwannian stromal component), which are derived from identical neoplastic cells that develop along different lineages. Evaluation of this parameter showed that MYCNA subtype comprised of only neuroblastoma tumors compared to higher frequency of ganglioneuroblastoma in Chr11q-LOH and MES subtype (Figure S2B). This partly explains the high rate of stromal infiltration in these two subtypes as ganglioneuroblastoma has been shown to be stroma-rich tumors comprising more than 50% schwannian stroma (Ambros et al., 2002).

4.2.2 Master Regulator identification of high-risk NBL subtypes

The gene expression profile and the associated biological pathways are driven by transcriptional regulatory module directly and indirectly controlling the expression of the genes driving the specific subtypes. Our approach follows a well-established pipeline for inference of regulatory modules involved in tumorigenesis and tumor progression from specific tumor types (Aytes et al., 2014; Carro et al., 2010; Lefebvre et al., 2010), using GEPs from tumor samples as an input. We run ARACNe and VIPER analysis independently on the TARGET and NRC cohorts, using cohort-specific subtype signatures and interactomes, followed by reproducibility assessment and result integration.

To assemble NBL specific transcriptional interactome (NBi), we first ran ARACNe on primary NBL tumor GEPs independently on TARGET and NRC cohort samples. ARACNe is an information theoretic approach for identifying transcriptional interactions among genes based on their mutual information. We built NBi from TARGET and NRC dataset and retrieved 205,271 and 359,846 interactions respectively, with 81,035 common interactions (p-value=0, odds ratio = 65.02) (Figure 4-11 A). We then applied the novel algorithm msVIPER (Alvarez et al, in press) in order to retrieve candidate MRs for each high-risk subtypes using a group of stage 1 tumors as control; as a result, we produced two independent MR protein lists for each subtype.

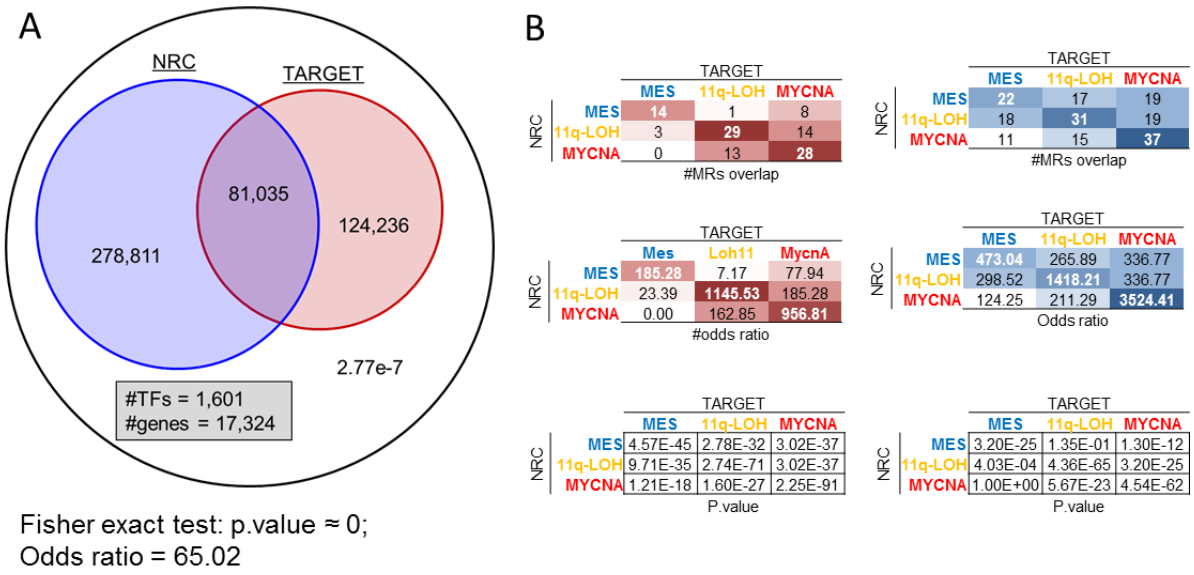


Figure 4- 11. NBi ARACNe network and master regulator analysis reproducibility

(A) Venn diagram representing overlap and the number of interactions inferred by ARACNe transcriptional networks produced from TARGET and NRC GEPs. The total number of putative interactions is estimated using the number of genes annotated as transcription factor and/or DNA binding from Gene ontology database (REF) multiplied by the total number of human genes (filtered by availability in the HumanExon mycroarray plattform). Fisher's exact test to assess overlap between TARGET and NRC NBi transcriptional interactions represented as text. (B) Fisher's exact test for the overlap of the top 50 predicted positive (red) and negative (blue) MRs across subtypes from TARGET and NRC datasets.

This algorithm identifies the MRs based on the overlap of ARACNe predicted targets of a TF and DGE signature of each subtype using Stage I samples as a reference group. msVIPER inferred subtype specific MRs and computed enrichment scores for downstream targets of each transcriptional regulator in the NBi. Hence, each MR can be ranked by their p-value. msVIPER analysis of the two interactomes produced two independent MR lists for each subtype, showing striking cross-cohort reproducibility (Figure 4-11 B).

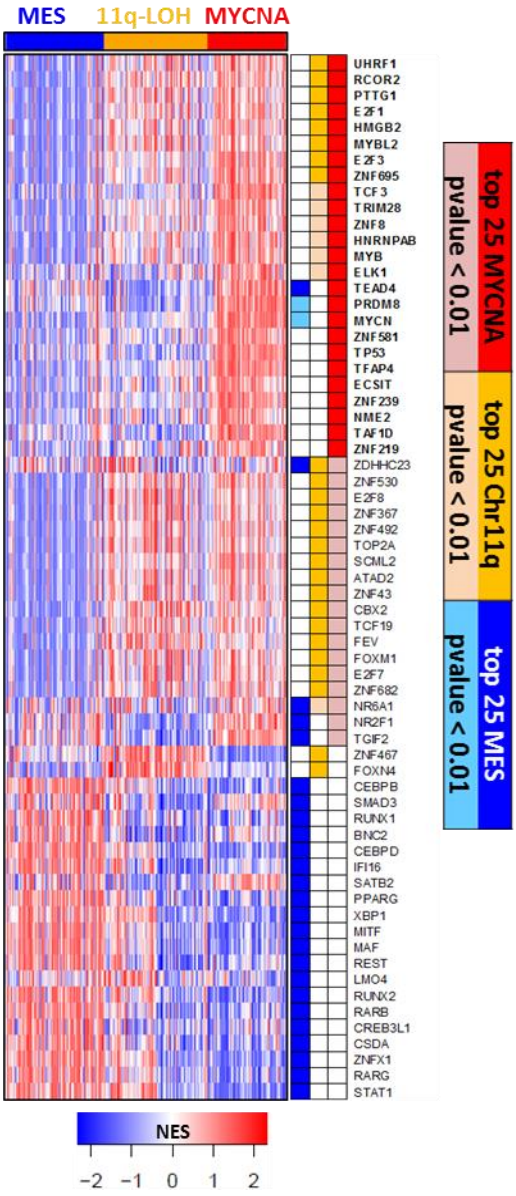


Figure 4- 12. Master regulators representing the three subtypes of high-risk NBL.

Heatmap of top activated MRs (red) of each subtype are represented using VIPER inference of TF activity using stage 1 samples as a control group.

Since the cohorts produced highly congruent results in this analysis, we generated a final MR list by integrating the p-values from each independent analysis using FET (Fisher, 1922) to obtain the final rankings (Figure 4-12). Consistent with the subtype specific programs, we retrieved several MRs that were known to be positively associated with the phenotype. For MYCNA subtype, the MRs included MYCN, which is a well-known driver of this subtype, and also recovered many cell cycle related TFs like E2Fs, MYBL2 and HMGB2. Similarly, 11q-LOH subtype also shared some cell cycle related MRs with MYCNA subtype. Even though both subtypes have proliferative phenotype, they have common as well as distinct MRs driving the specific subtypes. As expected, the MES MRs were entirely different from the other two subtypes and we recovered the previously validated MRs of MES subtype of glioma including CEBPD, CEBPB, RUNX2, FOSL2 (Carro et al., 2010).

4.4 Materials and Methods

Patient Samples

All samples profiles used in this analysis belong to primary tumors of NBL patients collected and organized by the TARGET. Data were generated using clinically annotated, matched tumor and normal samples obtained from patients enrolled in various Children's Oncology Group (COG) biology and clinical trials. The TARGET NBL cohort is composed mostly by high risk patients as defined by the Children's Oncology Group (COG). High risk tumors are those which present stage 4 based on the INSS (International NBL Staging System) in patients diagnosed after 18 months age. Also tumors harboring *MYCN* amplification diagnosed after 12 months are considered high risk. The TARGET cohort also integrates a group of low risk, INSS stage 1 tumors from patients diagnosed before 12 months age.

NBL gene expression profiles

The NBL primary tumor gene expression profile datasets used to define the subtypes and to assemble the transcriptional interactomes are organized by the TARGET (Therapeutically Applicable Research to Generate Effective Treatments) consortium and the NRC (Neuroblastoma Research Council) consortium. Profiles were obtained in both datasets using Affymetrix arrays Human Exon 1.0ST platform. Gene expression intensities were normalized using Robust Multi-array. Average implemented in the Affymetrix Power Tools suit using the “rma-sketch” option over the core probeset annotation.TARGET dataset comprises 249 GEPs from which 214 high risk and 24 are low risk. NRC dataset comprises 278 GEPs from 96 high risk tumors and 30 low risk tumors.

Clustering analysis of high risk primary tumor samples

We used ConsensusClusterPlus (Monti et al., 2003) R bioconductor package implementation of Consensus Clustering. This method provides quantitative evidences for determining the number and membership of possible clusters within a dataset, such as microarray gene expression. The algorithm implements several clustering algorithms and generates consensus for multiple resampling. First we collected genes with higher coefficient of variation (CV > 0.1), which represent genes with greater sample variability in each dataset (#TARGET = 3401; #NRC=2853) we selected those genes which

present in both lists (#common=2435). Consensus clustering was based on hierarchical clustering algorithm implemented in R “hclust” function where "ward.D" method was selected to segregate samples based on “Pearson” distance measures. Consensus clustering requires a predefined number of final clusters k; we ran the algorithm using k values 2 to 6.

Analysis of segmental DNA copy number frequencies.

SNP arrays where initially analyzed as described previously (Attiyeh et al., 2009) in order to obtain segmentation profiles corrected by aneuploidy and gene level copy number.

Differential expression analysis and null distributions for tumor subtypes

In order to calculate differential expressed genes in high risk tumor subtypes we calculated two sided Student's t-test comparisons for each high risk subtype versus a pool of stage 1 tumor GEPs. Additionally, we generated null models for each signature by sample permutation (i=1000) and t-test. These null models are required for the Master Regulator analysis msViper algorithm described below.

Assembly of the NBL specific transcriptional interactome (NBLi)

NBLi was assembled from NBL GEPs using ARACNe (Algorithm for the Reconstruction of Accurate Cellular Networks) an information theoretic algorithm (Margolin et al., 2006). ARACNe allows reverse engineering of transcriptional networks (i.e. TF-target interactions) from large GEP datasets. Inference of TF-target interactions is inferred using the significance of Mutual Information between TF genes (1800 transcriptional regulators defined by Gene Ontology annotation) and candidate target genes where the p-value cutoff is established empirically. In order to remove indirect interactions, ARACNe implements a Data Process Inequality (dpi) tolerance pruning system. ARACNe software is freely available for academic purposes at http://wiki.c2b2.columbia.edu/califanolab/index.php/Software. We ran 100 bootstrapped networks by resampling 80% samples every time and integrated the resulting networks into a consensus network. The probability of every edge in the network was corrected using Bonferroni multiple testing correction method and only the edges with $p<0.5$ after correction are

included in the final network. This process was used independently to obtain TARGET and NRC interactomes.

Identification of Master regulators of high risk NBL subtypes

msVIPER was used to interrogate the TARGET and NRC interactomes to obtain Master Regulators from each high risk subtype differential expression signature in both TARGET and NRC datasets independently. The results from both dataset were driven independently and return empirical p-values; therefore we can integrate the results using the Fisher's method for the integration of probabilities which states that $X^2 \sim -2 \sum_{i=0}^{k} ln(Pi)$. This approach returns an X2 statistic from which an analytical p-value can be obtained. Master Regulator Rank for both TARGET and NRC datasets as well as the integrated ranking from which the top 25 master regulators were selected for further experimental validation (Table S4).

Gene set and pathway enrichment analysis

We used gene set enrichment analysis (GSEA) in order to calculate differentially active pathways and gene sets across the relevant GES. We derived the gene sets for the biological processes from REACTOME and GO database. P-values were corrected by multiple testing using Benjamini and Hochberg method (Benjamini and Hochberg, 1995).

4.3 Discussion

Recent advances in high-throughput technologies have resulted into emergence of multi-dimensional dataset characterizing the genetic, epigenetic and functional properties of tumor phenotypes. Dissection of these molecular profiles to decipher the heterogeneous tumor samples and identify the causal drivers represents a considerable challenge. Current classification of cancer subtypes suffer from lack of reproducibility and hence cannot be relied upon for clinical practice (Lavasani and Moinfar, 2012). The results presented in this chapter show that we could reproduce the subtypes and master regulatory modules when applied on high-risk NBL samples obtained from two independent datasets; TARGET and NRC, hence increasing the confidence in the subtype classification and master regulator prediction.

Furthermore, our transcriptome based classification largely overlapped with the reported classification based on genomic alterations (Michels et al., 2007) showing that these alterations could be recapitulated by the transcriptome landscape of the tumor. Our analysis classified the subtypes with mutually exclusive structural variants (a) 1 p deletion and MYCNA, and (b) 11q deletion as Moreover, there are several genomic alterations like 3p and 4p deletion, chr.7 gain co-exist with 11q-LOH subtype, suggesting a complex interaction between these alterations to create a subtype specific signature. Similarly, we observed higher dosage for 17q gain in this subtype compared to MYCNA or MES subtype suggesting that dosage of 17q might also play a role in inducing 11q-LOH subtype GES. While some genes mapping to the altered regions have been reported to play a role in NBL pathogenesis — such as tumor suppressor, *CHD5*, in 1p deletion region (Fujita et al., 2008); oncogenes such as *BIRC5* (Lamers et al., 2011) and *PPMD1* (Saito-Ohara et al., 2003) in 17q amplification region, exhaustive search for identification of oncogenic drivers and tumor suppressor genes within these regions is a daunting task. This is demonstrated by one such study showing that more than 1000 genes have been identified within the eight recurrent genomic alterations in NBL (loss of 1p, 3p, 4p, 11q and 10q; gain of 2p, 17q and *MYCN* amplicon), which shows correlation with its corresponding mRNA expression level, indicating that they are likely to have a functional role in NBL (Łastowska et al., 2007).

Rather than searching for driver genes in the genomically altered regions, our approach provided an efficient way to identify the MRs that are causal to driving the subtype specific programs. This is because the confluence of the spectrum of genetic and epigenetic factors, integrated by the master regulatory genes within the regulatory networks and driving the phenotype can be better captured by gene expression of their target genes. Our results show that the core regulatory machinery responsible for implementation and stability, including canalization and integration of mutational events are recapitulated by their distinct transcriptional states. For example, even though *MYCN* is amplified only in a subset of MYCNA subtype, the MRs that account for the dependencies of the entire subtype with high MYC/N activity would be missed by solely genetic based analyses. Following the same principle, other TFs within the genomically altered regions or the TFs acting as a key mediator of the genes in these

altered regions can also be captured by their transcriptional activity. While both MYCNA and 11q-LOH subtype are highly proliferative, these programs are likely driven by different genomic alterations converging on a subset of common as well as distinct MRs to drive the phenotype as shown by our MR predictions (Figure 4-12). Besides *MYCN*-amplification and chr.11q-deletion specific for the subtypes, higher dosage of 17q gain might also contribute towards higher activation of some of the proliferative MRs.

Furthermore, our approach allowed identification of MES subtype without genomic alterations, which were missed by genomic alteration based classification, suggesting that this subtype maybe sustained by the MRs (CEBPβ and CEBPδ), that drive mesenchymal phenotype, thus recapitulating the mechanism responsible for mesenchymal transformation of high-grade glioma (Figure 4-8) (Carro et al., 2010). The high rate of stromal and immune cell infiltration in this subtype and a subset of Chr11q-LOH samples indicate that the interplay with tumor microenvironment might be important for pathogenesis of these tumors. While there has been several reports depicting infiltration of *MYCN*-non-amplified tumors (Asgharzadeh et al., 2012; Metelitsa et al., 2004), the interaction between immune system and this subtype hasn't been studied. Further studies are needed to fully understand the biology behind it. We do realize that the infiltration affects the gene expression as well as copy number variation data (Carter et al., 2012; de Ridder et al., 2005) hence obscuring the interpretation of some of our analyses. Single cell analysis of these patient samples would be crucial to provide insight into tumor cell-autonomous as well as non-cell-autonomous processes driving these tumors (Alizadeh et al., 2015). This is critical as the heterogeneity among tumor samples as well as within an individual tumor is starting to be realized and will be central to personalized medicine. It is becoming clearer that sequencing on single tumor-biopsy samples from the patient's primary tumor is not the representative of the genomic properties of the tumor (Gerlinger et al., 2012) as it doesn't capture the evolution of tumor. Unlike other tumors where immune infiltration has been predictive of patient outcome (Fridman et al., 2012; Pagès et al., 2009), all the high-risk NBL subtypes have equally bad prognosis. Whether the interaction of these NBL cells with its microenvironment plays a role in response to therapy remains to be determined. It is an attractive

hypothesis, particularly because it has already been shown in other cancer types (Gajewski et al., 2013; Pagès et al., 2009).

In this chapter, we showed that high-risk neuroblastoma tumors can be classified based on gene expression signature, each recapitulating its unique genomic and genetic driven transcriptional programs. We further defined the predominant mechanistic pathways in each subtype resulting into subtype specific clinical phenotype, where all the high-risk NBL tumors were undifferentiated with MYCNA and 11q-LOH subtype showing strong activation of proliferative program and MES subtype showing strong activation of mesenchymal program. Finally, we showed that the retrieved subtype specific MRs recapitulated the subtype specific programs, where proliferative programs seem to be driving MYCNA and Chr11q-LOH subtype, and MES-GEP and highly infiltrated microenvironment seems to be driving the MES subtype (Figure 4-9). Our data suggests that a more personalized approach to identify the target genes that are required for sustaining each subtype could facilitate development of targeted therapeutics and better patient outcome.

Chapter 5

Identification of TEAD4 as a master regulator of *MYCN*-amplified subtype of high-risk neuroblastoma

5.1 Background

Amplification of *MYCN* occurs in about 20% of all NBL and 45% of high-risk cases (Brodeur et al., 1984). *MYCN* amplification as well as overexpression has been shown to be associated with aggressive disease and poor prognosis (Bordow et al., 1998; Chan et al., 1997). This event is strongly associated with advanced stage of disease and poor outcome, regardless of the clinical factors that are the most predictive of outcome such as age and stage of the disease. Moreover, among the high-risk group, patients with *MYCN* amplification have been shown to have more aggressive phenotype (Brodeur and Bagatell, 2014). Because of its clinical significance, integration of *MYCN* copy number status into NBL risk stratification and treatment planning has been applied.

Considering the importance of MYCN oncogene in pathogenesis of MYCNA subtype of high-risk NBL, significant efforts have been expended into finding therapeutic targets for *MYCN*-amplified NBL tumors. Knockdown of MYCN in *MYCN*-amplified neuroblastoma has been shown to induce differentiation and apoptosis in human and mice models (Kang et al., 2006). However, direct targeting of MYCN is challenging because of the lack of binding pockets for the small molecules. A common strategy devoted to that end is the identification of synthetic lethal target genes through genome-wide RNA interference (RNAi) screens. The synthetic lethal relationship offers an attractive hypothesis for development of specific therapeutic agents to identify genes that are essential only in the context of specific genetic background, usually performed on a cell line with inducible MYC/N ON/OFF system. RNAi reagents like small interfering RNA (siRNA) and short hairpin RNA (shRNA) are consistently used in large-scale screens and combined with phenotypic assays to evaluate the consequences of eliminating specific genes. Promising results have been obtained in *MYCN* amplified specific tumor synthetic lethal screens leading to identification of AURKA (Otto et al., 2009), CDK1 (Goga et al., 2007; Sjostrom et al., 2005), CDK2 (Molenaar et al., 2009), and CHK1 kinases (Cole et al., 2011). Instead of focusing on genome-

wide or kinase specific screens, our RNAi screen is focused on top ranked putative MRs, thus allowing us to (a) identify universal tumor dependencies that canalize the effects of the upstream genomic variants and signaling molecules, and (a) reduce the noise inherent to the genome-wide screens.

As shown in the previous chapter, MYCNA subtype is characterized by high activation of cell proliferation, cell growth and DNA repair programs. A brief overview of DNA replication and repair as well as a brief description of MYCN and TEAD4 transcription factors are described below.

5.1.1 DNA replication and damage repair control

Cell cycle regulation is an intricate process strictly controlled by activities of cyclins and the associated CDKs, coupled with DNA replication to ensure high fidelity in genome duplication. The critical steps during replication are origin licensing and origin firing; where licensing refers to commitment to S-phase and origin firing marks the beginning of the S-phase. Replication is initiated at several hundred sites on the chromosome to ensure proper duplication of the entire genome and that it occurs only once during a cell cycle. The cell prepares for DNA replication in G1 phase, by assembling the multiprotein complexes at the origins of DNA replication. Firstly, origin recognition complex (ORC) is recruited, followed by Cdc6 and Cdt1 dependent loading of MCM2-7 complex possessing DNA helicase function (Remus and Diffley, 2009). Once the pre-replicative complexes are assembled on DNA origins, initiation of DNA synthesis is triggered by S-phase specific CDKs and Cyclin A and E and CDC7, which promotes binding and activation of MCM complex by Cdc45 and GINS. This marks commitment to S-phase. DNA is consequently unwound at each replication fork, and DNA polymerase α is loaded to initiate DNA synthesis (Kelly and Brown, 2000). This entire process is tightly regulated to ensure that the DNA replication happens only once in a cell cycle to ensure genomic integrity. This is performed by degradation of Cdt1 by Cyclin A/CDK2 complex, which requires the degradation complex to tether to PCNA on the replicative fork and inactivation of Cdt1 and Cdc6 by Geminin. Consistently, deregulated expression of Cdt1 and Cdc6 or depletion of geminin has been shown to drive re-replication (Blow and Gillespie, 2008).

Another layer of regulation to recognize and repair DNA damage is necessary to maintain genomic integrity. During cell cycle, S phase is the most prone to DNA damage events, as DNA replication has to occur with high fidelity. Various mechanisms can induce replication stress in the cells, which leads to halting the cell cycle progression to allow the cell to repair the damage. One of the main mechanisms leading to replication stress in precancerous and cancerous lesions is induced by oncogenes through aberrant DNA replication (Gaillard et al., 2015) and deregulation of cell cycle progression by activating cyclin-dependent kinases (CDKs) that functions in G1 and S phase (Hartwell and Kastan, 1994). Oncogenes such as MYC/N, Cyclin E, Cyclin D, Ras and E2Fs have been shown to induce replication stress and genomic instability (Halazonetis et al., 2008a; Hills and Diffley, 2014). In general, these oncogenes induce replication stress in the cells, leading to fork stalling, i.e, the DNA replication initiation complex are uncoupled leading to long stretches of ssDNA. If the stalled forks are unable to restart, they collapse, thereby leading to DNA double strand breaks (DSB), which occurs at specific chromosomal loci called common fragile sites (Arlt et al., 2006). This leads to activation of ATM and ATR mediated DNA damage response (DDR) pathway to induce both DNA repair and checkpoint signaling (Bartkova et al., 2005; Gorgoulis et al., 2005). While ATM responds to DSBs, ATR primarily responds to ssDNA coated with Replication protein A (RPA), present at processed DSB ends and stalled replication forks. Both ATR and ATM mediates its effect through CHK1 and CHK2 kinases respectively to induce G2/M arrest by sequestration and degradation of Cdc25A and Cdc25C, which are primarily involved in activating the CDK1 to allow G2/M transition. In addition, Chk1 has also been shown to respond to stalled replication forks by inducing intra-S phase checkpoint to block the activation of late replication origins when DNA synthesis from early origins are inhibited (Feijoo et al., 2001). Activation of the CHK1/2 kinases act as a double-edged sword, as tumors exhibiting genomic instability have activated DDR pathways in an effort to restrain the replicative stress to inhibit cancer development in early stages. This is because cells undergoing high load of replicative stress and hence activated DDR pathways, when aided with cell cycle checkpoint defect such as p53 inactivation induces DSBs which leads to cytotoxicity by mitotic catastrophe and apoptosis (Malumbres and Barbacid, 2009; Toledo et al., 2011; Vitale et al., 2011). Hence, hyperactivation of DDR pathways in cells with CIN serve to restrain the oncogene mediated replication stress inherent to the cells, thereby allowing proliferation of the cells in

the presence of damaged DNA. Consistently, NBLs and other tumors with activated MYC/N, indicative of replicative stress, has been shown to be sensitive to CHK1 and WEE1 inhibition (Cole et al., 2011; Murga et al., 2011; Russell et al., 2013).

5.1.1 MYCN transcription factor

MYCN belongs to the MYC proto-oncogene family of transcription factors, comprised of MYC, MYCN and MYCL. It was first discovered in neuroblastoma cell lines as a distinct homolog of avian myelocytomatosis viral oncogene (v-myc) that is different from c-MYC (Kohl et al., 1983; Schwab et al., 1983). In addition to neuroblastoma, MYCN has been implicated in other cancers of embryonic or neuroectodermal origin such as retinoblastoma, medulloblastoma, Wilm's tumor, glioblastoma, rhabdomyosarcoma and small cell lung cancer (Schwab, 2004). Early studies showed that both MYC and MYCN are capable of cellular transformation in the presence of other oncogenes such as RAS (Land et al., 1983; Schwab et al., 1985; Yancopoulos et al., 1985). Similarly, transgenic expression of MYCN in peripheral neural crest of mice has been shown to induce neuroblastoma formation, where these tumors showed prolonged latency and recurrent chromosomal aberrations (Weiss et al., 1997).

MYC proteins are activated by different mechanisms like amplification of *MYCN* in neuroblastoma (Seeger et al., 1985b), translocation of *C-MYC* in Burkitt's lymphoma (Dalla-Favera et al., 1982), or their increased protein stability by acting downstream of the signaling pathways (Nilsson and Cleveland, 2003). It has been accepted that the genetic alteration is not a required event and that overexpression or increased activity of MYC/N proteins by various mechanisms above a certain threshold level is enough to induce cellular transformation. Turnover of MYC/N are regulated by sequential phosphorylation at S62 via MAPK and CDK1, followed by phosphorylation at T58 via GSK-3B (Sears et al., 2000; Sjostrom et al., 2005). When S62 is dephosphorylated by PP2A, it binds FBW7 or other E3 ligases leading to ubiquitination and degradation (Welcker et al., 2004; Yada et al., 2004a). Even though the biochemical properties of MYC and MYCN are similar, the expression pattern of MYC and MYCN differ significantly. While MYC is expressed ubiquitously and has high expression in rapidly proliferating cells during development as well as in adult tissue, MYCN expression is restricted during

embryogenesis and is downregulated in differentiated adult tissues (Meyer and Penn, 2008). The expression of MYCN and MYC has been shown to be inversely correlated as they repress each other (Breit and Schwab, 1989). In NBLs, *MYC* amplification is not common and MYC mRNA is higher in *MYCN*-non amplified tumors (Westermann et al., 2008).

Both MYCN and MYC, to a greater extent, are very well studied in the context of their role in normal development as well as tumorigenesis. Decades of studies have shown that it plays an important physiological and pathophysiological role and have been coined as bona fide oncogenes, capable of inducing cellular transformation. MYCN and MYC are highly homologous both structurally and functionally. They are members of basic helix-loop-helix-leurine zipper (bHLH-LZ) transcription factor family and share similar DNA binding, protein-protein interaction and transactivating domains. The carboxy-terminal domain of MYC/N is necessary for dimerization with other bHLH domain containing transcription factors such as MYC Associated Factor X (MAX) MAX and sequence specific DNA binding. While the leucine zipper (LZ) domain is required for protein-protein interaction, bHLH domain binds the consensus E-box sequences (CANNTG) in the target gene promoter to transactivate them. This accounts for their functional redundancy by sharing common targets (Blackwell et al., 1993; Blackwood and Eisenman, 1991). MYC/N has also been shown to heterodimerize with MIZ1 and SP1 to repress its targets (Iraci et al., 2011; Wanzel et al., 2003). Not surprisingly, there is a large overlap in the targets of MYC and MYCN (Valentijn et al., 2012; Westermann et al., 2008; Zeller et al., 2003), and replacement of *c-myc* by *mycn* in mouse model can rescue the embryonic lethal phenotype, otherwise induced upon *c-myc* or *mycn* knockout (Charron et al., 1992; Davis et al., 1993; Malynn et al., 2000). MYC/N has been shown to regulate several key cellular processes (Meyer and Penn, 2008) including cell proliferation by transactivating several cell cycle and DNA replication genes as well as non-transcriptional control of DNA replication (Bell et al., 2006; Dominguez-Sola et al., 2007), repressing differentiation by regulating genes like p27Kip1, SKP2, p21, GADD45 (Evans et al., 2015; Kang et al., 2006), ribosome biogenesis and cell growth (van Riggelen et al., 2010a), cellular metabolism (Dang, 2013), and chromosomal instability induced by DNA-replication stress through transcriptional control of cell cycle genes as well as by directly increasing replication origin activity (Dominguez-Sola et al., 2007; Felsher and Bishop, 1999;

Vafa et al., 2002). The mechanism of regulation of apoptosis is not fully elucidated and has been thought to be partly mediated by blocking p53 mediated response (Van Maerken et al., 2009).

5.1.2 TEAD4 transcription factor

TEA Domain Family Member 4 (TEAD4) belongs to highly conserved transcriptional enhancer factor (TEF) family of transcription factor (Drosophila: Scalloped), characterized by a highly conserved TEA DNA binding domain that binds a consensus sequence (CATTCCA/T) in the enhancers and promoters of the target genes to activate transcription (Kaneko and DePamphilis, 1998). Furthermore, all the four TEAD proteins have more than 80% amino acid sequence conservation in the protein-protein interaction domain at the c-terminus of the protein (Mahoney et al., 2005; Vassilev et al., 2001a). Most of the late embryonic and adult tissues express at least one kind of TEAD proteins, suggesting that it mediates common cellular functions such as cell proliferation in different tissue types (Kaneko and DePamphilis, 1998). Among them, TEAD2 and TEAD4 are among the earliest transcription factor that are expressed during mammalian development and may serve to activate the genes critical for development. While inactivation of TEAD2 led to neural tube defects, inactivation of TEAD4 resulted into preimplantation lethal phenotype (Yagi et al., 2007). Similarly, it has been shown to contribute to pathogenesis of several cancers (Lim et al., 2013; Liu et al., 2015; Wang et al., 2015).

The TEAD proteins, and TEAD4 to a greater degree, has been studied mainly in the context of hippo signaling pathway, as it has been shown to be required for of transcriptional activity of the oncogenes Yes-Associated Protein 1 (YAP) and WW domain containing transcriptional regulator 1 (TAZ). In brief, hippo signaling pathway is a tumor suppressive pathway, the activation of which phosphorylates YAP and TAZ to retain it in the cytoplasm. Upon translocation to the nucleus, it binds with TEAD factors to coactivate target gene expression to promote cell proliferation and pro survival signals (Harvey et al., 2013). Specifically, TEAD proteins have been shown to be required for YAP-induced cell growth, cellular transformation and epithelial to mesenchymal transition (EMT); and TAZ-induced EMT (Zhang et al., 2009; Zhao et al., 2008a). Recently, it has been shown that YAP/TAZ/TEAD complexes bind to

the distal enhancer elements to regulate the expression of the target genes to induce oncogenic growth (Zanconato et al., 2015).

Computational analyses in this chapter were performed by Dr. Gonzalo Lopez.

5.2 Results

While we performed MR analysis on each subtype, we focused our efforts on experimental validation of MYCNA subtype MRs. We concentrated on systematic validation of MRs of the MYCNA subtype for several reasons: (a) it is strongly associated with the most recurrent NBL genetic alteration, affecting ~30% of primary NBL (Seeger et al., 1985a); (b) MYCN overexpression is associated with aggressive disease and poor prognosis (Bordow et al., 1998; Chan et al., 1997); (c) Analysis of high-risk NBL samples indicated that the MYCNA subtype has highest purity in terms of immune and stromal infiltration (Figure 4-7), thus contributing to high reliability on GEP based MR predictions; finally (d) it represents the subtype with the largest number of established, MR- and genetic-matched cell lines for validation purposes, thus making the analysis robust and non-controversial.

5.2.1 Master Regulators of *MYCN*-amplified subtype

For experimental validation, we chose the top 25 putative MRs of MYCNA subtype, that were integrated from the two cohorts: TARGET and NRC (Figure 5-1 A). The MRs were first predicted separately based on TARGET or NRC NBi, followed by integration of these two predictions, both of which showed significant overlap. This was assessed by performing GSEA enrichment analysis for the top 50 positive and negative MRs derived from each dataset on the ranked list of TFs on the other ($NES_{\text{TARGET activated MRs on NRC}}$ = 4.13, p-value = 3.7 e -05; $NES_{\text{TARGET deactivated MRs on NRC}}$ = -4.14, p-value = 3.53 e -05; $NES_{\text{NRC activated MRs on TARGET}}$ = 4.4, p-value = 1.08 e -05; $NES_{\text{NRC deactivated MRs on TARGET}}$ = -4.58, p-value = 4.7 e -06) (Figure 5-1 B and C).

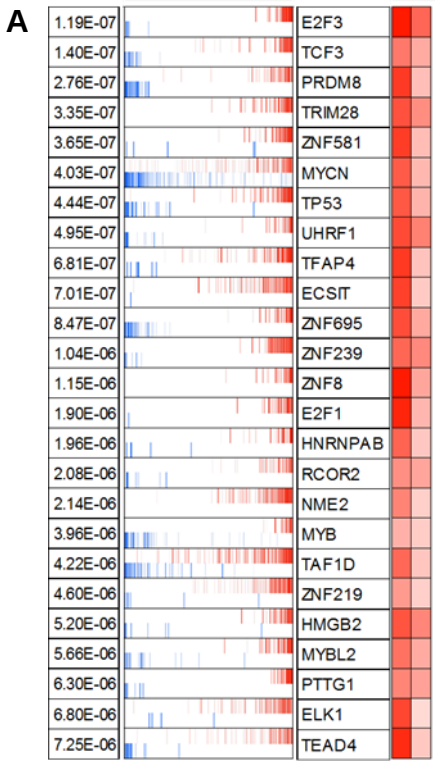
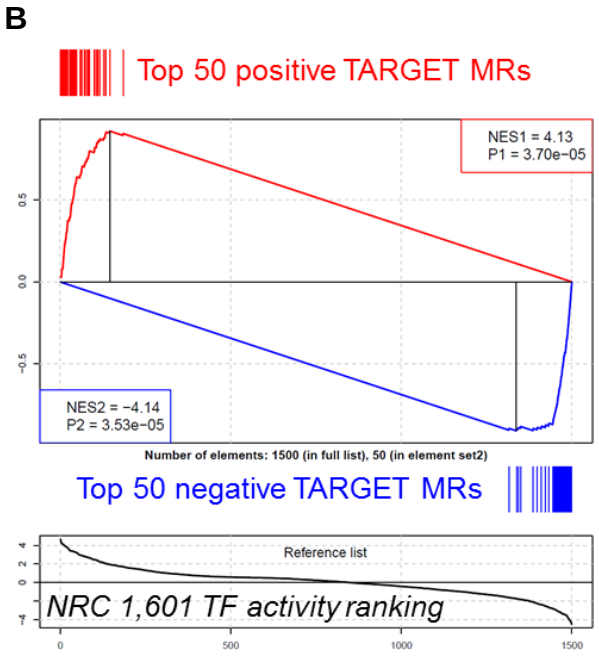
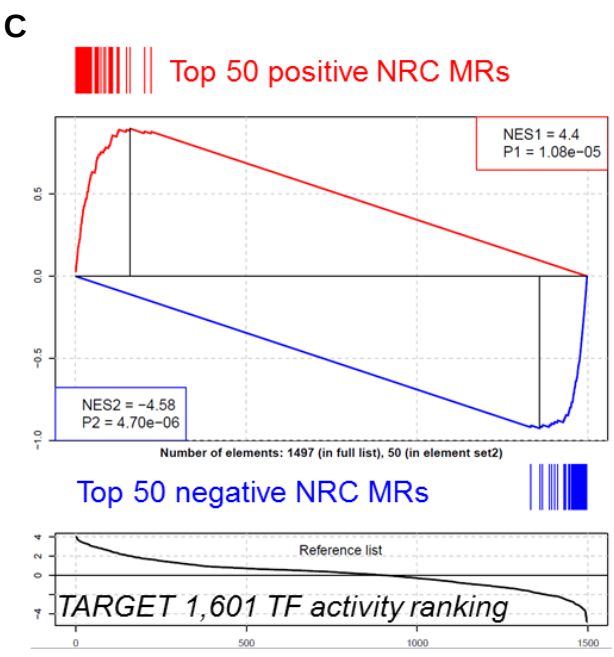


Figure 5- 1. Master regulators of MYCNA subtype. (A) The top 25 combined MRs of MYCNA subtype chosen for validation. The map shows distribution of positively (red) and negatively (blue) regulated targets of each MR ranked by differential expression between MYCNA subtype vs. stage I patient samples. (B, C) Reproducibility of MYCNA positive MRs derived from TARGET and NRC dataset assessed by GSEA enrichment analysis for the top 50 positive and negative (B) TARGET derived MRs in the NRC TF activity ranking and (C) *vice versa*.





5.2.2 Identification of cell lines representing MYCNA subtype

For experimental validation, it is necessary to establish the cell lines representing the subtype as well as appropriate control. In order to select multiple case/control systems for experimental validation, we analyzed GEPs of a panel of 28 NBL cell lines available from Children's hospital of Philadelphia (CHOP). These cell lines have gene expression as well as copy number variation data available. Due to unavailability of Stage 1 patient sample derived cell lines, we used a few measures to find appropriate controls from Stage 4 derived cell lines. We used VIPER algorithm, which allows single sample level evaluation of a TF activity by measuring the enrichment of its ARACNe inferred targets in genes differentially expressed between the cell line vs. stage I samples. In order to consider a cell line representative of MYCNA subtype, cells must harbor *MYCN* amplification, high expression/activity of MYCN and overall high activity of the 25 MRs chosen for validation. Suitable controls, on the contrary, would have low expression and activity of MYCN and also low VIPER-inferred activity of the top 25 MRs.

Most of the MYCNA cell lines showed higher activity of the top 25 MRs, with SK-N-Be2 being the best model for the subtype (p-value = 3.7e-24) (Figure 5-2). We selected several cell lines and chose a few that represent MYCNA subtype or the control group based on the efficiency of lentiviral transduction. The cell lines chosen showed near 100% lentiviral transduction efficiency tested by expression of GFP upon transduction of cells with the lentiviral particles expressing GFP. This allowed us to perform experiments without the need for antibiotic selection, hence reducing the noise level for the experimental results. We chose several cell lines representing MYCNA subtype including SK-N-Be2, IMR-5, IMR-32, LAN1, SK-N-DZ and NB-1. Similarly, we used SK-N-AS, SY-5Y and SK-N-FI as controls. As an exception, we included NLF as a control cell line, which despite harboring amplification of *MYCN*, showed low level of MYCN protein expression and activity as well as low MYCNA activity of the MRs (Figure 5-2, 5-20 A).

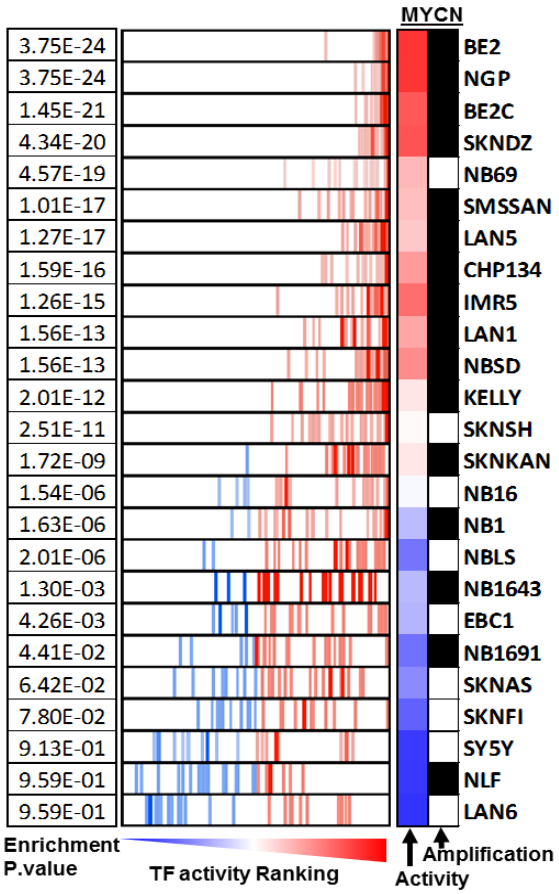


Figure 5- 2. Cell lines with high MR activity show segregation between MYCN amplified and non-amplified subtypes.

VIPER activity (Alvarez, M., 2013, bioconductor package) of the top 25 MRs of MYCNA subtype included in validation screens are computed in the CHOP panel for 28 NB cell lines. Overall activity is represented as the enrichment P value (left column), ranked positions of the top 25 MRs in each cell line (center). MYCN activity from average signature and MYCN amplification status occupy the columns on the right.

5.2.2 RNAi screening: pooled shRNA, individual shRNA and siRNA

While RNAi technology offers a way to measure the functional consequences of depletion of a gene in the system of interest, there are certain limitations inherent to the technology. We used multiple approaches and RNAi reagents to mitigate false discovery resulting from off-target effects and technology specific biases: (a) in-vitro and in-vivo pooled shRNA screen (b) individual shRNA screen (c) siRNA screen. To avoid cell line specific bias, we used multiple cell lines in our study. The strategy we used for experimental validation is to use the appropriate phenotypic assay defining the subtype. We assessed cellular viability by resazurin assay as our primary readout since MYCNA subtype is marked by highly proliferative signature and clinical phenotype with high mitotic index (Figure 4-6, 4-10).

(a) Pooled shRNA screening

To identify the MRs essential for both *in-vitro* and *in-vivo* tumorigenesis of MYCNA cells, we performed loss-of-function screen in MYCNA and control cell lines using TRC shRNA library comprising 4-5 hairpins per MR, bought from Sigma. Performing the assay on a pooled format allowed us to collect data with many hairpins per gene at a time. The goal of this screening was to identify the genes that are essential selectively for MYCNA cells compared to the control cells, evaluated by the depletion of shRNAs from the population in the respective cells. The cell lines were transduced with MR specific hairpins in separate wells in a 96-well plate such that each cell contains only a single type of integrated virus. The idea is that if a gene is essential for the cell, the cells expressing the specific hairpin against that gene will be depleted from the population over time. The shRNA sequence can act as a barcode for fate of the cell.

In vitro study was done on MYCNA cell lines (SK-N-Be2 and IMR-5) and control cell lines (NLF and SK-N-AS). The cells were transduced with MR specific and control hairpins; and depletion of hairpin representation was quantitated by deep sequencing, by comparing time 0 with time-28 days. Similarly, we also performed *in-vivo* study to ensure more physiologic environment for MR validation. In-vivo study was done on SK-N-Be2 and SK-N-AS as these cells presented more tumorigenic potential *in-vivo*. We assessed the shRNA representation in the resulting tumor, 4 weeks after implantation of the pool of shRNA transduced cells into nude mice. Hairpin level data was first integrated into a gene-level representation and then used to compute a MYCNA and control group level score using the ScreenBeam algorithm (Yu et al., 2015) (Table 5-1). Overall, a gene is identified as a candidate if (a) the shRNA is depleted more than 2-fold with a p-value of <0.05 (Z score -1.96) in the MYCNA cells, (b) there is more than 2-fold hairpin depletion in MYCNA group compared to the control group, and (c) the candidates are common to both *in-vitro* and *in-vivo* screen. Using this criterion, we identified 7 MRs that were common in both in-vitro and in-vivo screens: MYCN, TEAD4, HNRNPAB, HMGB2, PRDM8, E2F3, ECSIT (Figure 5-3). Furthermore, several of the predicted MRs induced loss of cell viability in both MYCNA and control cells (Table 5-1), suggesting a potential role across high-risk NBL. It is not

surprising as the MRs predicted for each subtype used Stage I samples as control and hence recapitulated subtype specific as well as high-risk NBL specific MRs.

Gene Symbol	z-score MYCN-amplified (in-vitro)	log2FC-MYCN non-amplified (in-vitro)	log2FC-MYCN amplified (in-vitro)	log2FC.MYCNAmp vs non-amplified (in-vitro)	z-score MYCN-amplified (in-vivo)	log2FC-MYCN non-amplified (in-vivo)	log2FC-MYCN amplified (in-vivo)	log2FC.MYCNAmp vs non-amplified (in-vivo)
E2F1	-2.13	-0.99	-2.33	-1.35	-1.06	-2.03	-1.38	0.65
E2F3	-1.99	-0.41	-2.20	-1.79	-2.41	0.01	-2.87	-2.88
ECSIT	-4.11	-1.56	-4.47	-2.90	-3.21	-3.61	-5.17	-1.55
ELK1	-1.65	-0.60	-2.04	-1.44	-3.24	-1.31	-4.36	-3.05
HMGB2	-2.10	-0.56	-3.11	-2.56	-2.98	-1.09	-3.85	-2.76
HNRNPAB	-3.35	-1.16	-2.55	-1.40	-4.10	-1.31	-5.32	-4.00
MYB	-2.49	-1.18	-2.20	-1.02	-1.07	-2.29	-2.54	-0.25
MYBL2	-2.71	-1.74	-2.17	-0.42	-2.10	-4.12	-2.44	1.68
MYCN	-3.85	-2.32	-4.22	-1.90	-2.99	-3.06	-4.60	-1.54
NME2	-3.60	-1.21	-1.58	-0.37	-1.37	-2.10	-2.13	-0.03
PRDM8	-2.71	-1.91	-5.10	-3.20	-2.85	-3.67	-6.15	-2.47
PTTG1	-3.28	-3.30	-2.57	0.74	-2.51	-5.25	-3.91	1.34
RCOR2	-3.22	-1.00	-2.51	-1.51	-2.18	-2.01	-2.68	-0.67
TAF1D	-2.83	-1.24	-3.64	-2.40	-1.83	-1.95	-2.24	-0.29
TCF3	-1.48	-0.71	-1.54	-0.83	-0.70	-0.91	-1.26	-0.35
TEAD4	-2.89	-1.55	-3.95	-2.40	-2.53	-1.45	-3.75	-2.30
TFAP4	-4.30	-1.67	-4.51	-2.84	-4.23	-4.27	-5.09	-0.82
TP53	-2.90	-1.64	-2.50	-0.85	-3.57	-2.48	-5.96	-3.48
TRIM28	-1.05	-1.68	-1.22	0.47	-1.35	-2.14	-2.04	0.10
UHRF1	-3.03	-0.77	-2.40	-1.63	-0.80	-1.65	-1.32	0.33
ZNF219	-1.59	-2.10	-1.58	0.52	-2.45	-3.05	-3.83	-0.78
ZNF239	-2.01	-1.23	-1.72	-0.49	-1.35	-1.53	-2.65	-1.13
ZNF581	-1.50	-0.18	-2.32	-2.14	-2.68	-0.68	-2.66	-1.98
ZNF695	-2.86	-1.01	-1.59	-0.59	-1.74	-2.41	-3.01	-0.60
ZNF8	-2.16	-1.70	-2.45	-0.75	-3.75	-2.72	-5.10	-2.38
NegControl	0.55	0.71	0.50	-0.21	-1.57	-1.87	-1.87	0.00

Table 5- 1. In-vitro and in-vivo pooled shRNA screening to identify MYCNA subtype specific MRs

Pooled in-vivo and in-vitro shRNA screening was performed on *MYCN*-amplified and non-amplified cell lines. Integrated z-score and fold change of shRNA dropout for each gene was derived for each cell line from individual shRNAs and the values for *MYCN*-amplified cell lines vs non-amplified cell lines were further integrated. Log_2Fold change of *MYCN* amp cell lines vs *MYCN* non-amp cell lines is derived for both *in-vivo* and *in-vitro* shRNA screening.

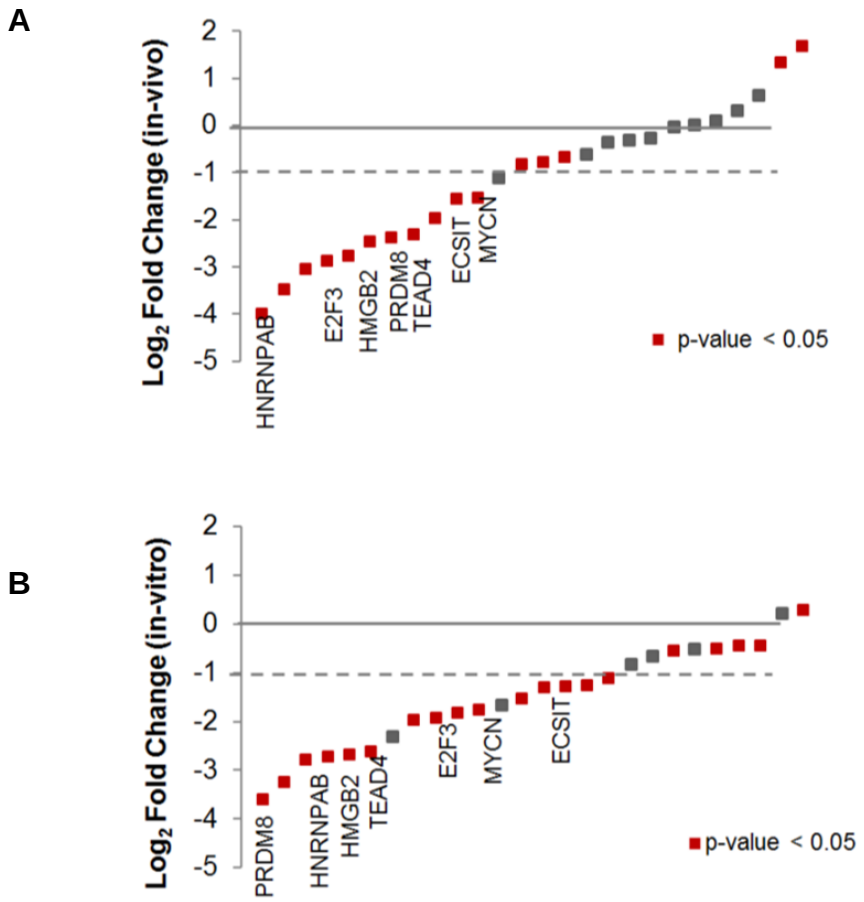


Figure 5- 3. In-vivo and in-vitro pooled shRNA screening of top 25 MRs of MYCNA subtype

(A) In-vivo pooled shRNA screening in MYCNA cells (SK-N-Be2) vs. control cells (SK-N-AS) and (B) In-vitro pooled shRNA screening in MYCNA cells (SK-N-Be2, IMR-5) vs. control cells (NLF, SK-N-AS), depicting changes in hairpin representation in MYCNA cells divided by control cells. For both (A, B), tumor-enriched shRNAs were amplified, sequenced and counted to identify enrichment and dropouts. shRNA abundance for a gene was integrated into a score and calculated as a ratio of T (final day) to T(0 day). The MRs were first screened to include only the ones with $p < 0.05$ in MYCNA group (red). MRs specific for MYCNA subtype were identified by the average fold change between MYCNA cells vs control cells. The grey dashed line shows the cutoff for -2.0 fold change.

In-vitro pooled shRNA experiments were performed by Ruth Rodriguez-Barrueco in the lab of Dr. Jose Silva, and in-vivo pooled shRNA experiments were performed by Dr. Antonio Iavarone and Dr. Anna Lasorella. Computational analyses of pooled shRNA screening results were performed by Dr. Jiyang Yu.

(b) Individual shRNA screening

Furthermore, we performed extensive individual shRNA validation in multiple NBL cell lines, where we selected two out of four to five shRNAs per MR that exhibited highest silencing efficiency at mRNA or protein level. This allowed us to perform further detailed and focused screen in multiple cell lines to ensure that the candidate MRs don't emerge because of cell-line specific dependencies. We used concentrated lenti-virus to transduce the selected shRNAs in five MYCNA cell lines (SK-N-Be2, IMR-5, IMR-32, NB-1 and LAN-1) and three *MYCN*-non-amplified (MYCN-NA) cell lines (SY-5Y, SK-N-AS and SK-N-FI). We measured relative cell viability normalized to control shRNA for each MR specific hairpin in each cell line, 96 hrs post transduction. In order to identify MRs representing MYCNA specific dependencies, we compared MR depletion effect (average over the two hairpins) in the MYCNA group compared to the control group (Table 5-2).

Cell lines	MYCN amplified					Control		
shRNA	SKNBE2 T96	IMR5 T72	IMR32 T96	NB1 T96	LAN1 T96	SY5Y T72	SKNAS T96	SKNF1 T96
TAF1D sh1	0.7	0.3	0.2	1.3	0.8	0.9	0.9	0.9
TAF1D sh2	0.6	0.3	0.3	0.6	1.0	1.0	0.8	1.0
HNRNPAB sh1	0.7	0.4	0.9	0.9	0.8	0.8	0.9	0.9
HNRNPAB sh2	0.9	0.4	0.6	1.0	1.0	1.9	0.9	0.8
ECSIT sh1	0.9	0.3	0.4	1.0	1.1	1.4	0.9	0.8
ECSIT sh2	0.6	0.4	0.2	0.2	0.7	0.6	0.6	0.7
TEAD4 sh1	0.9	0.5	0.5	0.6	0.9	0.9	0.9	0.8
TEAD4 sh2	0.4	0.3	0.1	0.1	0.5	0.4	0.5	0.6
PRDM8 sh1	0.4	0.3	0.2	0.5	0.8	0.2	0.7	0.8
PRDM8 sh2	0.5	0.2	0.1	0.1	0.8	0.2	0.6	0.8
PTTG1 sh1	0.3	0.1	0.2	0.3	0.5	0.7	0.2	0.5
PTTG1 sh2	0.5	0.2	0.1	0.5	0.5	0.5	0.4	0.6
RCOR2 sh1	0.8	0.5	0.2	0.5	0.8	0.8	0.8	1.0
RCOR2 sh2	0.5	0.2	0.6	0.7	0.7	0.4	0.6	0.7
TCF3 sh1	0.7	0.5	0.3	0.1	0.9	0.4	0.6	0.7
TCF3 sh2	0.9	0.7	0.6	0.8	1.0	1.0	0.7	0.9
E2F1 sh1	0.9	0.6	0.6	0.7	0.9	0.4	0.6	0.8
E2F1 sh2	0.7	0.8	0.2	0.5	0.8	0.9	0.5	0.7
E2F3 sh1	0.9	0.8	0.4	1.2	1.0	0.9	0.8	0.9
E2F3 sh2	0.8	0.2	0.2	0.4	0.9	0.4	0.9	0.7
TRIM28 sh1	0.7	0.6	1.0	0.8	0.9	0.7	0.5	0.8
TRIM28 sh2	0.8	0.2	0.2	0.8	0.9	0.5	0.8	0.8
ZNF219 sh1	0.9	0.2	1.1	1.2	1.1	0.7	0.8	1.0
ZNF219 sh2	0.6	0.3	0.3	0.7	0.6	0.5	0.6	0.7
HMGB2 sh1	0.9	0.6	0.3	0.4	0.6	0.7	0.8	0.7
HMGB2 sh2	0.9	0.7	0.7	1.1	1.0	0.6	0.7	1.0
ZNF8 sh1	0.7	0.4	0.3	0.4	1.0	0.6	0.7	0.8
ZNF8 sh2	1.0	0.9	0.7	1.0	1.1	1.0	1.0	0.6
ZNF239 sh1	0.9	0.6	0.8	0.4	1.0	0.6	0.9	0.9
ZNF239 sh2	0.8	0.9	0.3	1.0	0.7	1.0	0.9	0.8
NME1 sh1	1.2	0.7	1.0	1.2	1.0	1.1	1.1	1.0
NME1 sh2	0.6	0.2	0.3	0.7	0.9	0.5	0.8	0.7
UHRF1 sh1	1.1	0.3	0.6	0.8	1.0	0.6	0.8	1.0
UHRF1 sh2	0.8	0.9	0.5	1.0	1.0	0.9	1.0	1.0
TFAP4 sh1	0.7	0.7	1.0	0.8	0.9	0.8	0.7	0.8
TFAP4 sh2	0.7	1.2	1.2	0.9	0.7	1.1	1.0	0.9
MYCN sh1	0.9	1.0	0.6	1.2	1.0	0.7	0.8	1.0
MYCN sh2	0.8	0.8	0.6	1.2	1.0	1.5	0.6	1.1
ELK1 sh1	1.1	0.4	0.9	1.3	1.2	1.1	1.1	0.8
ELK1 sh2	1.0	1.2	0.7	1.0	1.1	1.0	1.0	0.7
ZNF581 sh1	1.2	0.7	1.4	1.4	1.3	0.9	1.0	1.1
ZNF581 sh2	0.8	1.1	0.7	0.9	0.7	1.1	1.1	0.8
ZNF695 sh1	1.1	0.9	1.3	0.9	1.0	1.2	0.8	1.0
ZNF695 sh2	0.6	0.4	0.2	0.2	0.7	0.3	0.6	0.6
TP53 sh1	0.5	0.9	1.6	1.4	1.1	1.7	0.8	0.9
TP53 sh2	0.6	1.2	1.5	1.1	1.0	1.5	0.7	0.9
MYB sh1	1.1	0.9	1.2	1.5	1.0	1.4	1.1	0.9
MYB sh2	1.2	1.1	1.3	1.4	1.1	1.2	1.1	0.9
MYBL2 sh1	0.8	0.3	0.3	1.3	0.8	0.8	0.6	0.9
MYBL2 sh2	0.9	0.5	0.4	0.5	0.9	0.6	0.7	0.7
PLK1	0.5	0.3	0.2	0.3	0.7	0.6	0.6	0.7

Table 5- 2. Individual shRNA screening to identify MRs of MYCNA subtype

A panel of MYCN-amplified and non-amplified cell lines were transduced and cell viability was measured 72 to 96hrs post-transduction. The experiment was done in triplicate and relative viability was measured relative to control shRNA. Representative experiments are shown.

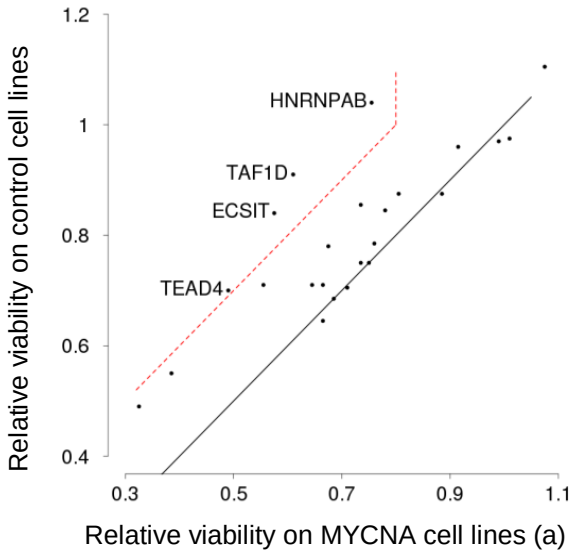


Figure 5- 4. Comparison of MR depletion in MYCNA vs control cells by individual shRNA

Scatter plot of average relative cell viability of MYCNA cells (SK-N-Be2, IMR-5, IMR-32, NB-1, and LAN-1) vs. control cells (SY-5Y, SK-N-AS, SK-N-FI) upon transduction with 2 shRNAs per MR, normalized to control shRNA, measured 72 to 96hrs post transduction. Relative cell viability for MYCN-amplified vs non-amplified groups were derived by taking average effect of 2 hairpins across the cell lines for each group. The red dashed lines shows the cutoff of a<0.8 and (b-a)>0.2.

These shorter-term assays show that MYCNA cells were significantly more sensitive to TEAD4, TAF1D, HNRNPAB and ECSIT silencing, compared to MYCN-NA cell lines (Figure 5-4). We did not detect MYCN as a hit in this screen at the time points we chose, as the viral load was not adjusted for the higher copy number of *MYCN* in MYCNA cells. However, we confirmed that depletion of MYCN in MYCNA cells, SK-N-Be2, induce differentiation and decrease cell viability seven days post transduction (Figure 5-5). Clearly, short-time assays are better suited to detecting MRs producing direct effect on proliferation, while longer-term assays are optimally suited at elucidating multifunctional dependencies. Consistently, additional MYCNA specific MR dependencies were detected in the long-term assays.

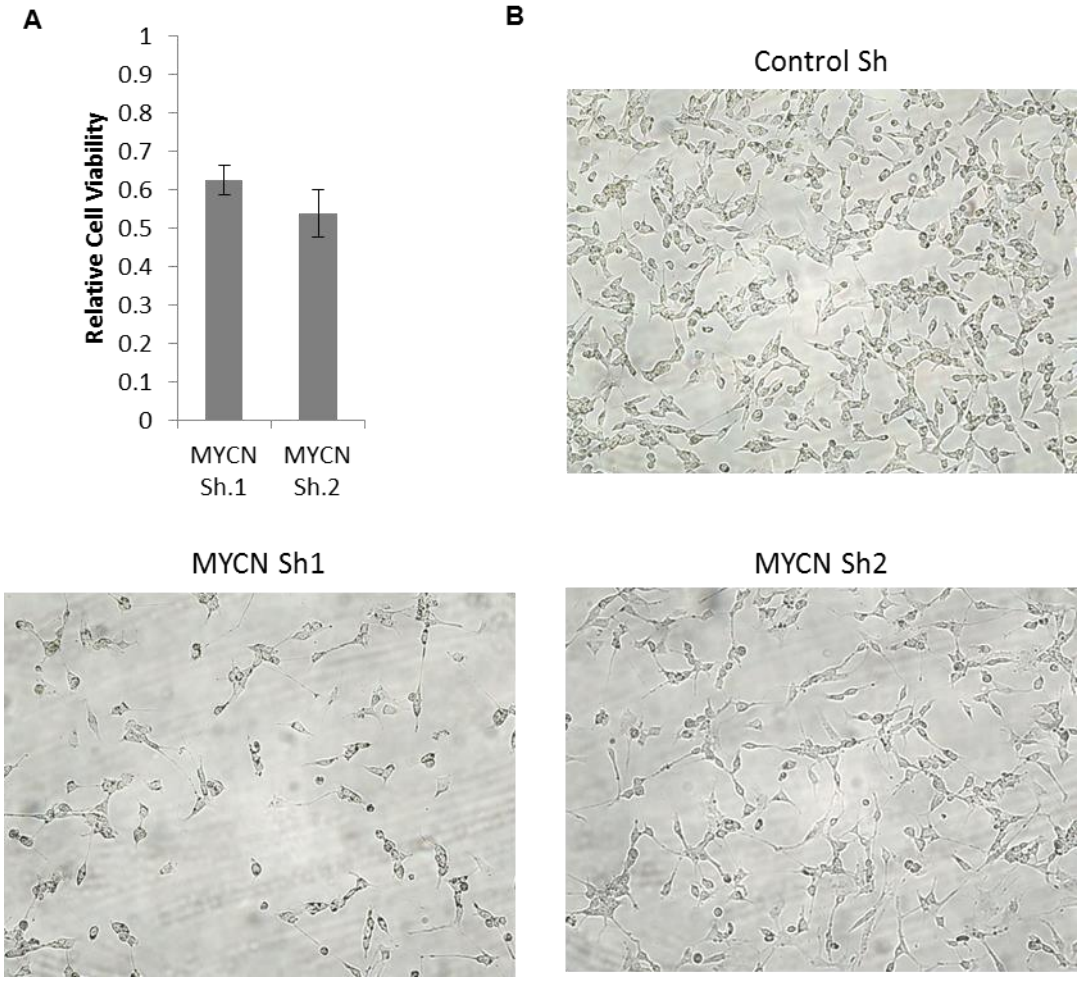


Figure 5- 5. Effect on cell viability and cell morphology upon MYCN knockdown.

(A) Relative cell viability and (B) Microscopy images showing morphological changes of SK-N-Be2 cells transduced with MYCN shRNA compared to Control shRNA, measured 7 days post-transduction.

(c) siRNA screening

Cell lines	MYCN amp			Control	
siRNA	IMR5 T96	SKNBE2 T96	SKNDZ T96	SY5Y T96	SKNAS T96
si ZNF219	0.7	0.6	0.8	0.9	1.3
si TFAP4	0.6	0.5	0.6	0.9	1
si TEAD4	0.7	0.7	0.9	1.1	1.2
si HMGB2	0.9	0.9	0.9	1.2	1.3
si PRDM8	0.8	0.7	0.9	1.1	1.2
si HNRNPA	0.7	0.5	0.9	1	0.8
si TRIM28	0.9	0.8	0.9	1	1.1
si MYB	1	0.8	0.9	1.2	1.1
si MYBL2	0.5	0.7	1	0.9	1
si ZNF581	0.9	0.7	1	1.1	1.1
si UHRF1	1	0.9	0.9	1.1	1.1
si TAF1D	1.2	0.8	1	1.2	1.1
si ZNF239	1	0.9	0.9	1	1.2
si E2F3	1	1.1	0.8	1.1	1.1
si ECSIT	1	0.7	0.9	1.1	0.8
si ZNF695	0.8	0.9	0.9	1.1	0.8
si E2F1	0.9	0.8	0.9	1.1	0.8
si CBX2	0.6	0.8	0.9	0.9	0.8
si NME2	1.3	0.7	0.8	1	1
si MYCN	1.3	0.7	1	1.1	1
si ELK1	0.8	0.8	0.8	1	0.7
si PTTG1	1.2	0.7	1	1.1	0.9
si TCF3	1.1	0.7	0.9	1.1	0.8
si ZNF8	0.9	0.9	1	1	0.8
si RCOR2	1.4	0.7	1	1.1	0.9
si TP53	1.3	0.8	0.9	1.1	0.6
si PLK1	0.1	0.3	0.5	0.6	0.1
si AURKA	0.4	0.5	0.9	0.9	0.7
si siDeath	0.2	0.2	0.4	0.6	0.1

For siRNA screen, we used ON-TARGETplus siRNA pool against the 25 MRs. By using a pool of four siRNAs, we dramatically dilute the off-target effect known to be caused by partial complementary anti-sense sequence acting as a micro-RNA to repress the expression of non-specific genes. First, we confirmed that the MRs were efficiently knocked down by their respective siRNA pool by qRT-PCR. We then transfected MYCNA cell lines (SK-N-Be2, IMR-5 and SK-N-DZ) and MYCN-NA cell lines (SY-5Y and SK-N-AS) with MR specific siRNAs. Similar to shRNA screen, we derived relative cell viability of MR-siRNA vs control-siRNA treated cells and derived the average effect of gene depletion in MYCNA group vs MYCN-NA group (Table 5-3). Our results confirmed TFAP4, HNRNPAB, MYBL2, TEAD4 and ZNF219 as MYCNA specific dependencies (Figure 5-6).

Table 5- 3. siRNA screening to identify MRs of MYCNA subtype

A panel of *MYCN*-amplified and non-amplified cell lines were transduced with ON-Target smartpool siRNA and cell viability was measured 96hrs post-transfection by Presto Blue reagent. The experiment was done in triplicate and relative viability was measured relative to control siRNA. Representative experiments are shown.

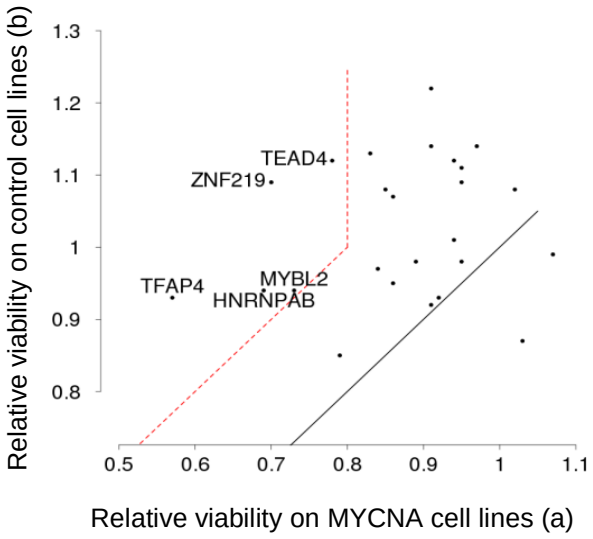


Figure 5- 6. Comparison of MR depletion in MYCNA vs control cells by pooled siRNA

Scatter plot of average cell viability of MYCNA cells (SK-N-Be2, IMR-5, SK-N-DZ) vs control cells (SY-5Y, SK-N-AS) upon transfection with ON-Target smartpool siRNA against each MR normalized to control siRNA, measured 96hrs post transfection. Relative cell viability for *MYCN*-amplified vs non-amplified groups were derived by taking average effect of siRNAs across the cell lines for each group. The red dashed lines shows the cutoff of a<0.8 and (b-a)>0.2.

Comparative analysis of the three screens identified eleven potential MYCNA-specific MR dependencies, while TEAD4 and HNRNPAB confirmed across all three screens (Figure 5-7). We thus further assessed the ability of these genes to regulate each other and the MYCNA signature.

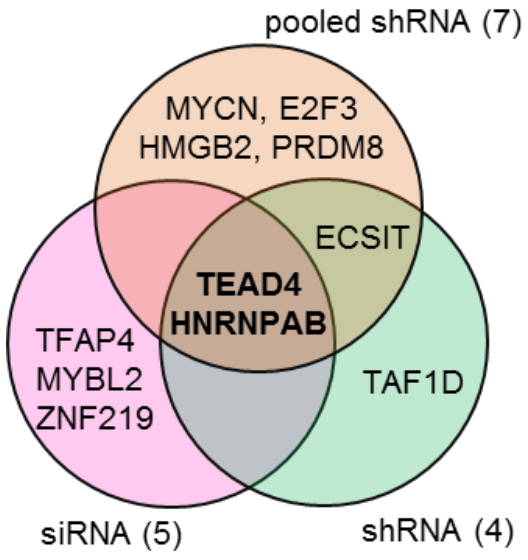


Figure 5- 7. Compilation of RNAi screening results

Venn diagram depicting potential MYCNA subtype specific MRs from (Figure 5-3) MRs common to both in-vitro and in-vivo negative selection pooled shRNA screening (Figure 5-4) individual shRNA screening (Figure 5-6) and siRNA screenings.

5.2.4 MR subnetwork are controlled by MYCN and TEAD4

Cellular phenotype is often controlled by a complement of several MRs acting cooperatively. We have shown that the transcriptional state of physiologic (Kushwaha et al., 2015; Lefebvre et al., 2010), tumor-related (Aytes et al., 2014; Carro et al., 2010; Chudnovsky et al., 2014; Piovan et al., 2013; Rodriguez-Barrueco et al., 2015) and other disease-related (Brichta et al., 2015; Chen et al., 2014; Ikiz et al., 2015; Repunte-Canonigo et al., 2015) phenotypes is controlled by a small number of proteins acting cooperatively within regulatory modules (checkpoints). Within this context, a “master regulator” is a protein whose activity is required (either individually or synergistically), to mechanistically induce normal/aberrant activity of the module, thus maintaining the phenotype “state”.

Thus, elucidating MR proteins in a specific phenotype context requires understanding their ability to regulate each other and the phenotype. To elucidate the transcriptional circuitry connecting the candidate MYCNA-specific MRs identified in the previous section (Figure 5-7), we performed extensive qRT-PCR, at 48h following their lentiviral-mediated silencing to assess their mutual regulatory potential (Figure 5-8 A). This early time point was chosen to minimize confounding effects due to indirect interactions. These data revealed a highly modular structure with hierarchical organization (Figure 5-8 B). Indeed, ten out of eleven MRs were interconnected by 20 intra-module regulatory interactions, 13 of which were also predicted by ARACNe. In addition, based on previous reports, MYCN (Breit and Schwab, 1989) and TEAD4 (Home et al., 2012; Lim et al., 2013) may control their expression through a positive auto-regulatory loop. Such feedback loops represent the hallmark of regulatory structure controlling phenotype stability, for instance by implementing bistable switches (Xiong and Ferrell, 2003). Such control structures are required to buffer against fluctuations in the abundance/activity of individual proteins (Albert et al., 2000; Hartwell et al., 1999). In the MR network, TEAD4 is directly downstream of MYCN and represents a critical regulatory bottleneck that controls the majority of other MYCNA specific MRs identified by our analysis. Indeed, positive TEAD4 regulation by MYCN was further verified at the protein level (Figure 5-9). Based on our qRT-PCR data, MYCN and TEAD4 jointly regulate (10/11) 90% of the MYCNA-specific MR dependencies. Hence, our analysis identifies the TEAD4/MYCN regulatory

loop and the large number of feed-forward loops they implement in regulating other MYCNA MRs as the critical mechanism controlling the stability of this subtype.

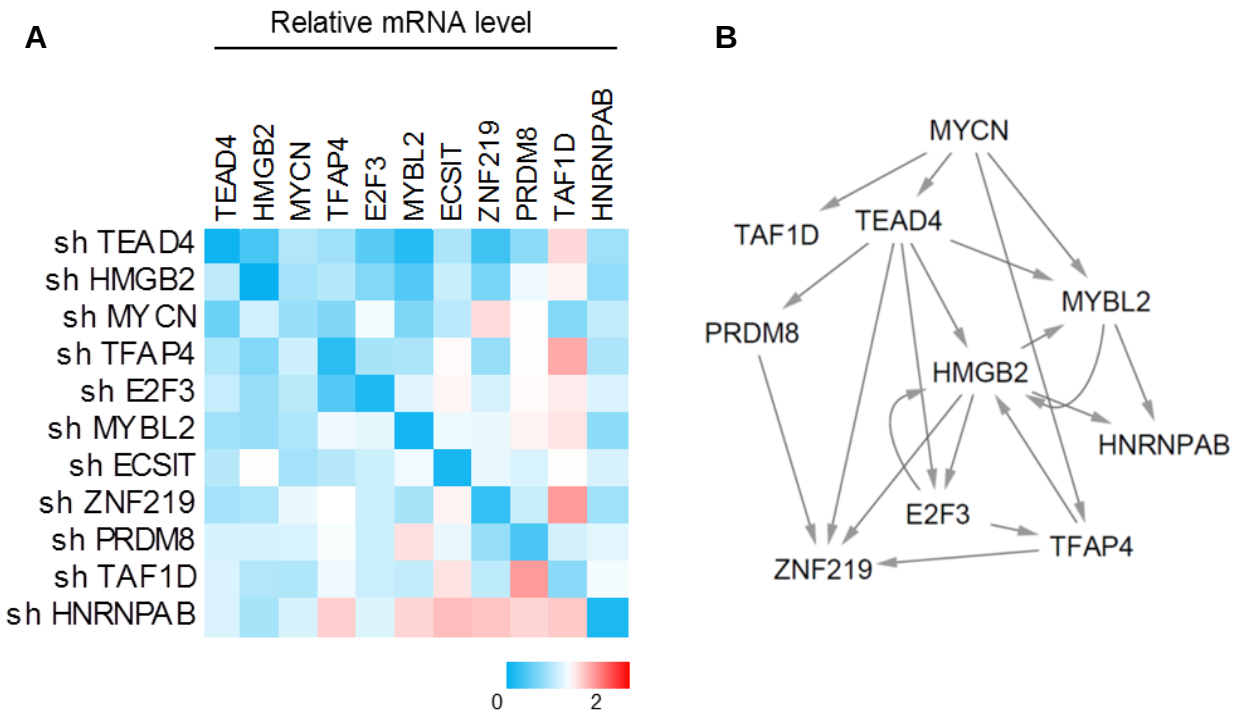


Figure 5- 8. MR inter-regulatory transcriptional network.

(A) Heatmap representing gene expression changes of MYCNA subtype specific potential MRs (Figure 3G), upon transduction of SK-N-Be2 cells with control or respective shRNAs against the MRs. The regulations were evaluated by qRT-PCR 48hrs post-transduction. Samples were run in triplicate and representative experiments are shown. (B) The inter-regulatory network derived from the results in (A) where genes showing >1.5 fold downregulation of transcript upon treatment with the shMR was considered to be a target.

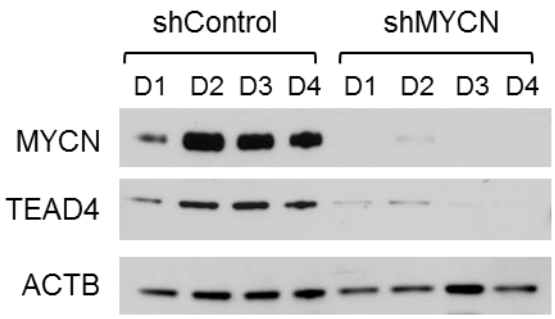


Figure 5- 9. MYCN regulates TEAD4 expression

Immunoblot showing downregulation of TEAD4 protein upon knockdown of MYCN 24, 48, 72 and 96hr post-transduction.

5.2.4 MYCNA signature and biological programs are controlled by TEAD4 and MYCN

To further validate the role of MYCN and TEAD4 as MRs of MYCNA subtype, we examined their global effect on MYCNA-GES. We performed lentivirus mediated silencing of MYCN and TEAD4 in SK-N-Be2 cells, followed by gene expression profiling using RNA-Seq. Again, we analyzed the results 48hrs post transduction to enrich the analysis for direct targets of the respective MR (Figure 5-10 A). Depletion of both MYCN and TEAD4 significantly reversed the MYCNA-GES toward a Stage I signature ($p < 10\text{-}16$) (Figure 5-10 B). GSEA analysis of the MYCNA-GES in up-regulated and down-regulated genes following MYCN and TEAD4 silencing showed highly significant enrichment, with MYCN silencing enriching more for downregulated genes (Figure 5-10 C) than TEAD4 (Figure 5-10 D). Indeed, MYCN and TEAD4 implement the MYCNA subtype signature by regulating both shared and complementary targets, accounting for 29% and 21% of MYCNA overexpressed genes; and 25% and 11% of MYCNA underexpressed genes, respectively (Figure 5-10 E). Taken together, MYCN and TEAD4 regulate ~70% of the genes differentially expressed in the MYCNA-GES, based on differential gene expression signature (FDR <0.01) (Figure 5-10 E).

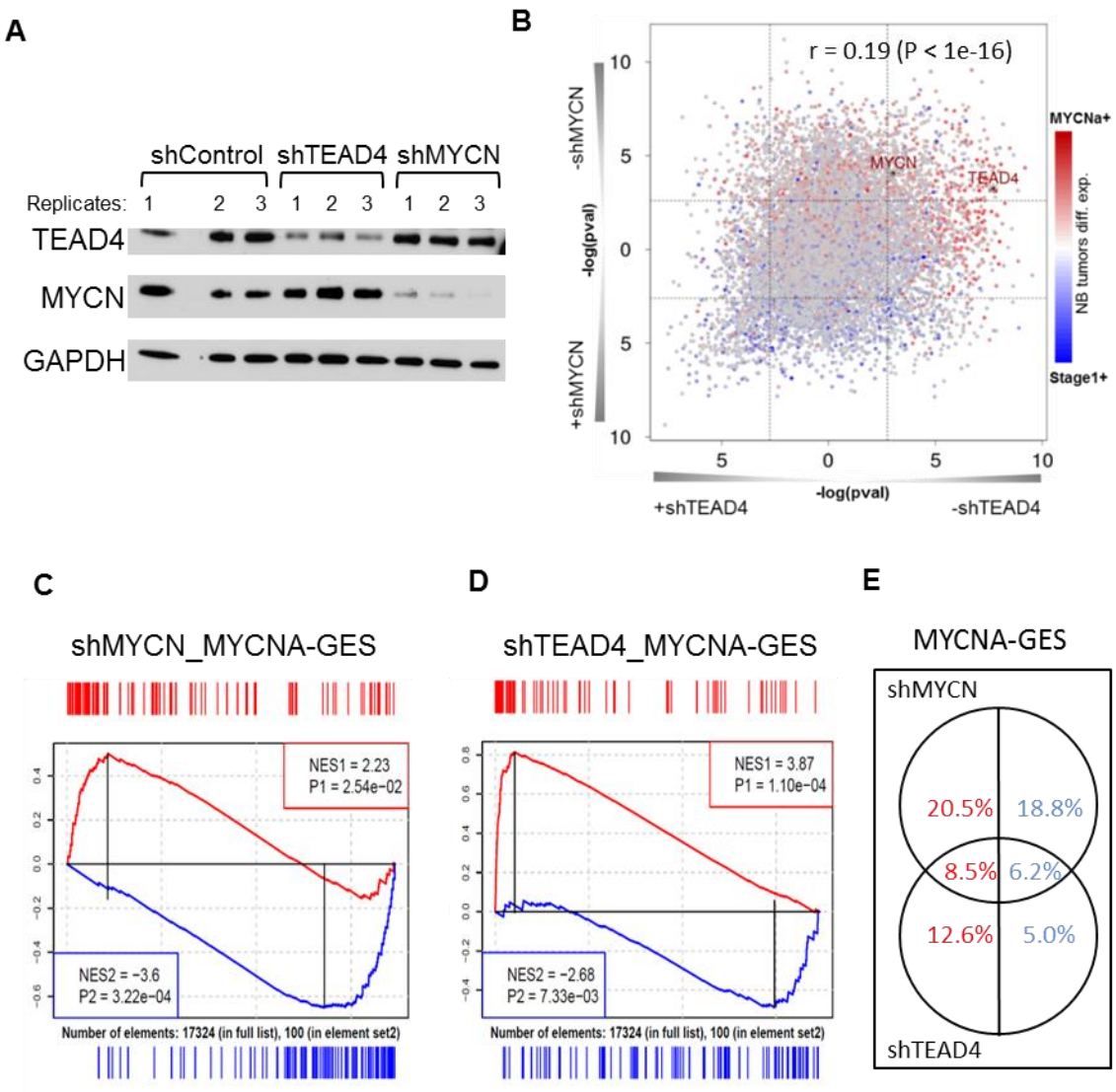


Figure 5- 10. Regulation of MYCNA subtype signature by TEAD4 and MYCN.

Differential GES of TEAD4 (x-axis) and MYCN (y-axis) knock-down compared with MYCNA vs stage1 signature (red-blue heat colors). GSEA analysis of differentially expressed genes after TEAD4 knock-down (D) and MYCN knock-down (E) in the signature of MYCNA versus Stg1 tumors. (F) Venn-diagram showing proportion of MYCNA-upGES (red) and MYCNA-downGES (blue) regulated by MYCN, TEAD4 or both.

ARACNe inferred targets not only supported this conclusion but were also highly enriched in genes differentially expressed following MYCN and TEAD4 silencing (Figure 5-11), thus confirming the overall validity of the regulatory model.

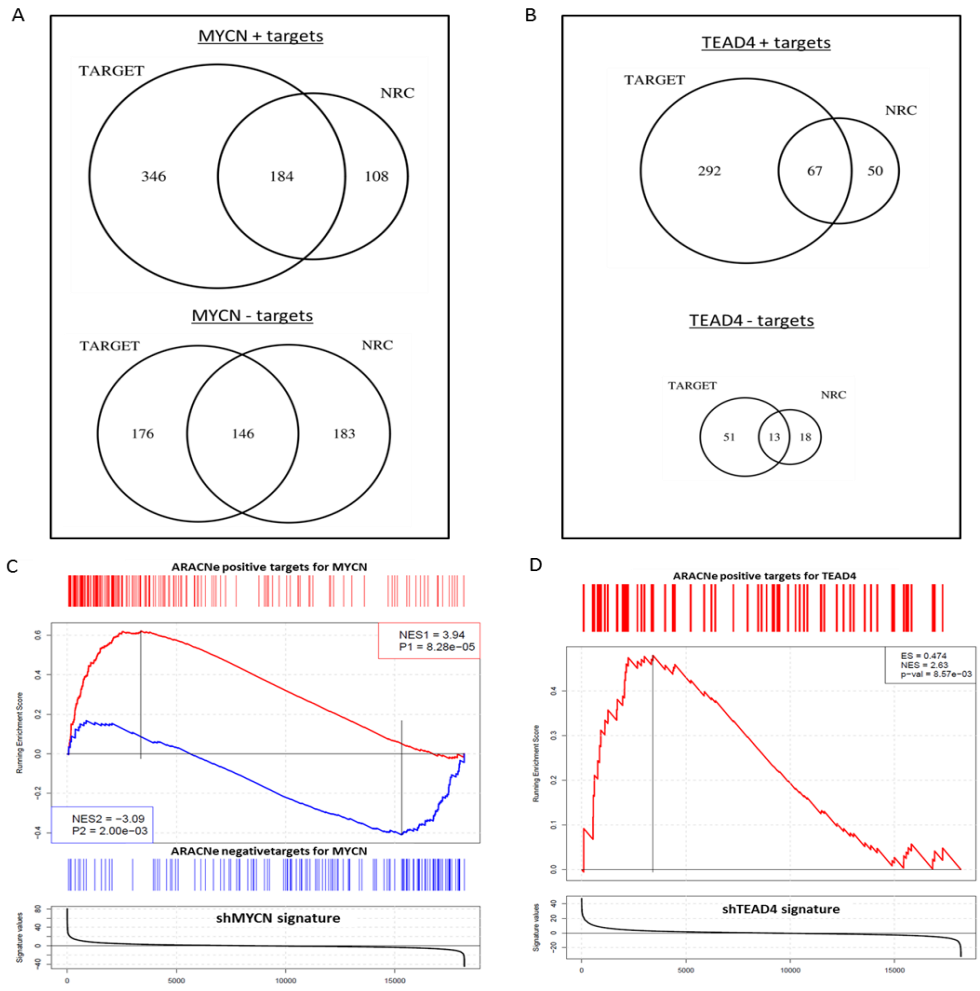


Figure 5- 11. ARACNe inferred regulon overlap and size for TEAD4 and MYCN.

ARACNe inferred interactions by Mutual Information are not directional therefore we present them divided in positive and negative according to the sign of Spearman's correlation between each TF and target. Venn diagram shows the overlap across cohorts of positive and negative targets of (A) MYCN and (B) TEAD4. (C) GSEA analysis of NBi overlapping predicted positive and negative targets of MYCN in the shMYCN RNA-seq signature. (F) GSEA analysis of ARACNe overlapping predicted positive targets of TEAD4 in the shTEAD4 RNA-seq signature.

To gain insight into the MYCNA subtype specific biological programs controlled by MYCN and TEAD4, we performed REACTOME pathway and Gene ontology (GO) biological processes enrichment analysis on their differentially regulated genes overlapping with MYCNA-GES. For this, we took all the gene sets in the REACTOME and GO database and performed GSEA on the ranked list of differentially regulated genes overlapping with MYCNA-GES, to specifically understand its role in the context of MYCNA subtype specific programs. Indeed, we found that depletion of MYCN significantly reversed multiple cellular processes that are activated and repressed in MYCNA cells (Figure 5-12 A and B). In contrast, depletion of TEAD4 significantly reversed the activated programs in MYCNA subtype (Figure 5-12 A and B). Consistent with previous findings, our data indicates that MYCN acts as both activator and repressor (Gartel et al., 2001; Kretzner et al., 1992), while TEAD4 acts mainly as a transcriptional activator (Vassilev et al., 2001b).

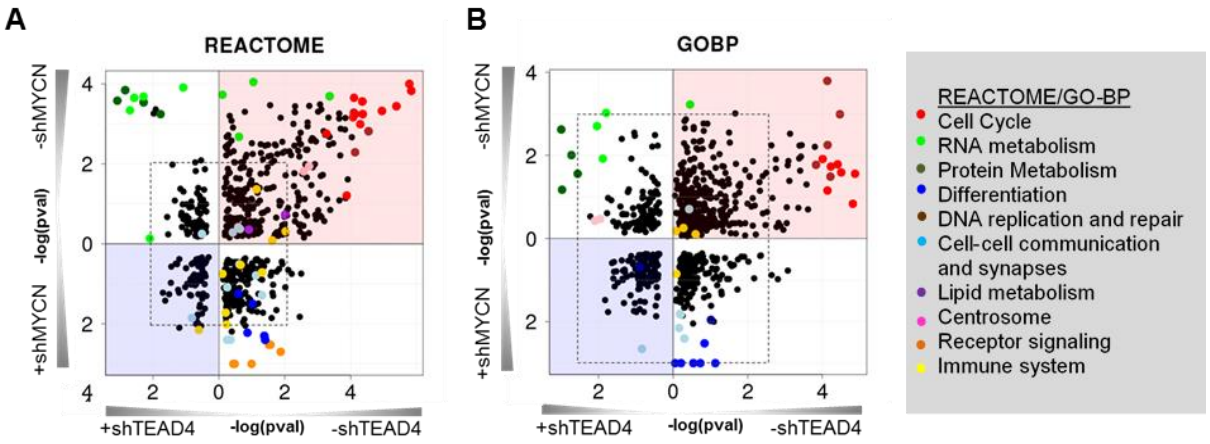


Figure 5- 12. Biological processes controlled by TEAD4 and MYCN

REACTOME pathway enrichment analysis performed on shTEAD4 (x-axis) and shMYCN (y-axis) signatures that overlap with MYCNA GES. Axis represents $-\log_{10}$ of the p-value while retaining the directionality of the normalized enrichment score. The dashed line represent FDR cutoff of <0.05.

Consistent with its known role, MYCN-specific activated genes were highly enriched in cell-growth/metabolism programs, including protein biosynthesis, ribosomal biogenesis, rRNA processing,

RNA processing and splicing (Coller et al., 2000; van Riggelen et al., 2010b). In contrast, the programs that were repressed by MYCN included neuronal differentiation, actin cytoskeleton organization, axon guidance and cell adhesion molecules, suggesting active prevention of cell differentiation. Indeed, actin dependent mechanisms play an established role in neuritogenesis and neuronal differentiation (da Silva and Dotti, 2002) as well as axon guidance (Dent and Gertler, 2003) Repression of Cdc42 and RAC1 (Govek et al., 2005; Yuan et al., 2003) by MYCN is likely a key factors in controlling actin based structure and thus neuronal development. Consistent with these findings, we observe activation of neurite outgrowth and neuronal differentiation programs upon MYCN silencing in these cells (Figure 5-5). On the other hand, TEAD4 induced dramatic activation of proliferative programs in cells where MYCN was also active, suggesting cooperative control not mediated by either MYCN or TEAD4 alone. Both TFs showed enrichment for cell cycle and DNA damage response programs. Finally, the residual MYCNA-GES program not regulated by either MYCN or TEAD4 showed enrichment for RNA and mRNA processing programs.

Sensitivity and subtype specificity of MYCN and TEAD4 silencing depends on the expression level of downstream targets of these MRs. Indeed, we observed that MYCN and TEAD4 induced targets are more expressed in the MYCNA subtype compared to other high-risk subtypes, confirming MYCN/TEAD4 addiction of the MYCNA subtype (Figure 5-13). Overall, our data suggests that MYCN and TEAD4 represent the key regulatory bottlenecks responsible for implementation and maintenance of the MYCNA subtype transcriptional signature.

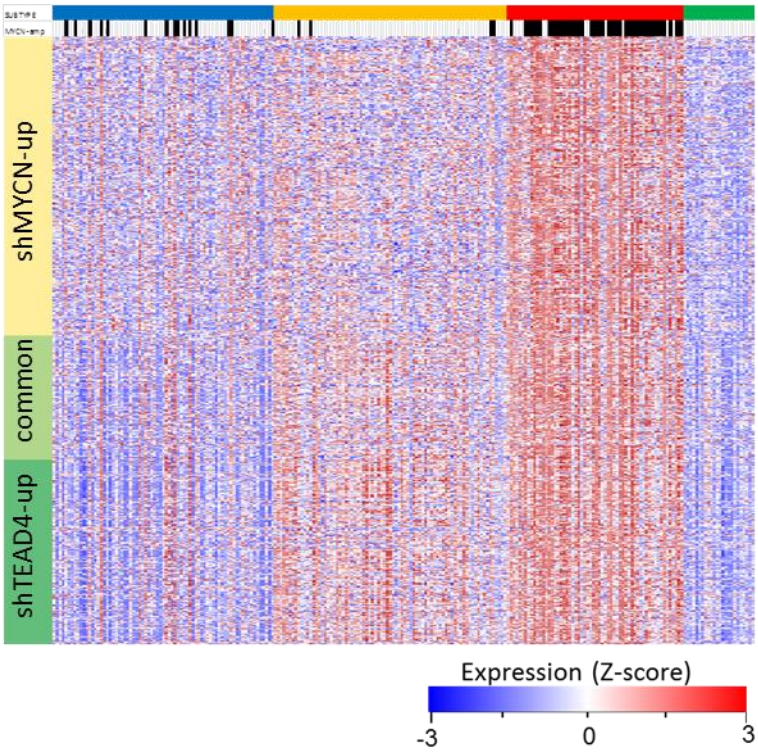


Figure 5- 13.Expression level of MYCN and TEAD4 functional signatures.

The heatmap shows the expression represented as Z-score of genes down-regulated (FDR < 0.01) after knockdown of TEAD4 and MYCN.

5.2.5 TEAD4 drives proliferative program in MYCNA neuroblastoma cells

MYCNA tumors are characterized by high-proliferative capacity and TEAD4 depletion mediated transcriptional changes in MYCNA cells, SK-N-Be2, demonstrated that it controls various cell-cycle dependent programs (Figure 5-14). GSEA analysis of RNA-Seq data upon TEAD4 depletion indicated significant enrichment for cell cycle specific GES of TEAD4 activated genes (p-value: 6.7E-05) (Figure 5-14 A). The genes that are most downregulated in response to TEAD4 depletion contained 49 out of 128 genes in the cell cycle gene set. We observed dramatic transcriptional activation of several critical cell cycle components involved in DNA replication including origin licensing, origin firing and G1/S and G2/M cell cycle checkpoint genes. Several of these genes have been implicated in high-risk NBL and some with positive association with MYCNA cells (underlined), by TEAD4. These include cyclin-

dependent kinases (CDK2, CDK1, CDC25B) (Chen et al., 2013; Sato et al., 2001; Sjostrom et al., 2005), Cyclins (Cyclin D1) (Molenaar et al., 2008), E2Fs (E2F1, E2F2, E2F3) (Strieder and Lutz, 2003), DNA replication (PCNA, MCM7, CDC6) (Feng et al., 2008; Keim et al., 1993; Shohet et al., 2002; Tsai et al., 2004), checkpoint kinases (CHEK1, CHEK2, WEE1) (Cole et al., 2011; Pugh et al., 2013; Russell et al., 2013) and ubiquitin-proteasome system (SKP2) (Evans et al., 2015; Muth et al., 2010).

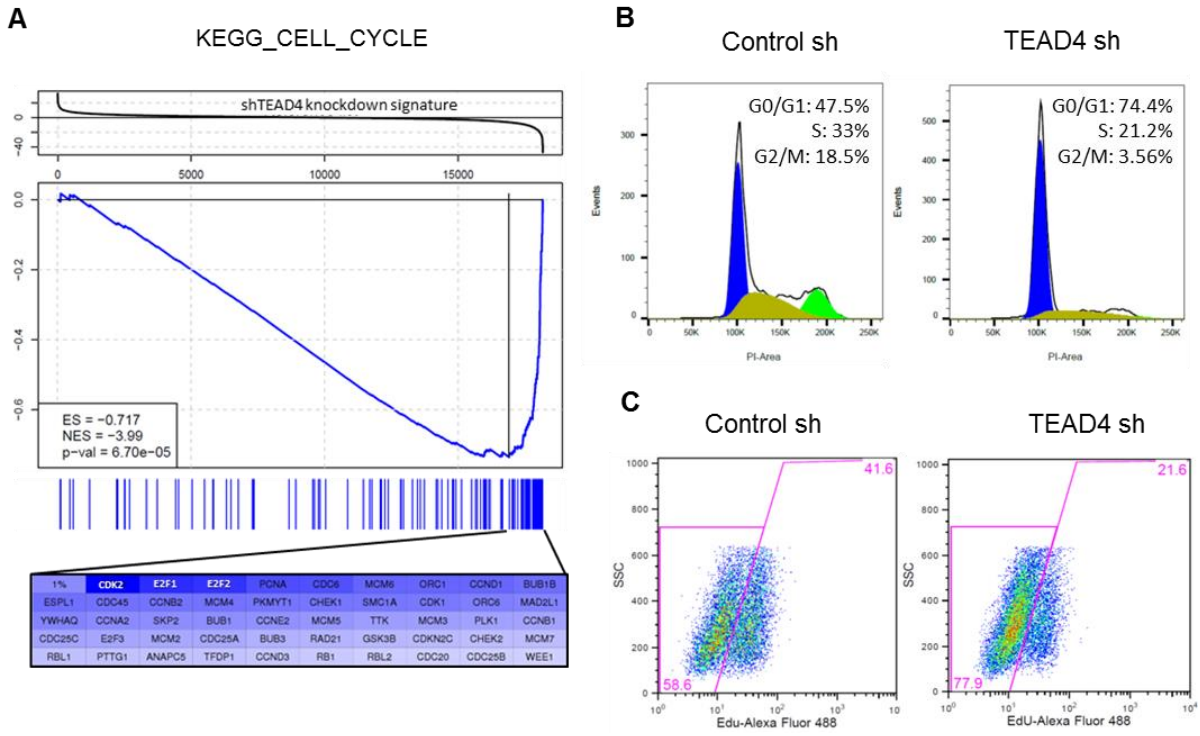


Figure 5- 14. TEAD4 is required for proliferation of *MYCN*-amplified cells

(A) GSEA plot evaluating enrichment for cell cycle gene set in KEGG database among the genes in shTEAD4 signature. (B) Cell cycle profile and (C) cellular proliferation, assessed upon treatment of SK-N-Be2 cells with control or TEAD4 shRNA, 48hrs post transduction. Representative experiments are shown.

We further investigated the phenotypic influence of TEAD4 on cell cycle and proliferation of MYCNA cells. Fluorescence-activated cell sorting (FACS) analysis upon knockdown of TEAD4 in SK-N-Be2 showed that TEAD4 depletion induced significant accumulation of cells in G0/G1 with concomitant decrease of cells in S phase (Figure 5-14 B). Consistent with this, we observed decrease in proliferating

cells by EdU staining (Figure 5-14 C). Our data showed strong concordance between gene expression changes and the phenotypes observed upon TEAD4 knockdown. Collectively, these findings suggest TEAD4 as a critical component driving cellular proliferation of MYCNA cells.

5.2.5 TEAD4 drives DNA damage response program in MYCNA neuroblastoma cells

Furthermore, MYCNA tumors are characterized by activated DNA damage response (DDR) (Figure 4-6), which has also been reported in the literature (Dominguez-Sola et al., 2007; Murga et al., 2011). GSEA analysis of RNA-Seq data upon TEAD4 depletion indicated significant enrichment of several DDR pathways that appeared to be activated by TEAD4 (Figure 5-15 A). A closer look into the genes showed activation of several critical genes such as Fanconi anemia gene family (FANCA, FANCI, FANCB, FANCD2), CHEK1, Rad51, BLM. Of particular interest was CHK1, the inhibition of which has been shown to be sensitive in MYC/N activated NBLs. We observed that upon silencing of TEAD4, CHK1 protein was depleted in a time dependent manner, with a concomitant increase in DNA damage marker, γ-H2Ax in MYCNA cell lines, SK-N-Be2 and LAN-1 (Figure 5-15 B, C).

A

REACTOME Pathways	NES	P-value	FDR	Gene no	Gene set size
G2_M_CHECKPOINTS	-4.049	0.000	0.004	29	41
CELL_CYCLE_CHECKPOINTS	-3.964	0.000	0.004	43	110
DNA_REPAIR	-3.942	0.000	0.004	44	104
ACTIVATION_OF_ATR_IN_RESPONSE_TO_REPLICATION_STRESS	-3.798	0.000	0.005	24	35
NUCLEOTIDE_EXCISION_REPAIR	-3.581	0.000	0.009	15	49
INHIBITION_OF_REPLICATION_INITIATION_OF_DAMAGED_DNA_BY_RB1_E2F1	-3.528	0.000	0.010	9	11
HOMOLOGOUS_RECOMBINATION_REPAIR_OF_REPLICATION_INDEPENDENT_DOUBLE_STRAND_BREAKS	-3.325	0.001	0.013	10	16
DOUBLE_STRAND_BREAK_REPAIR	-3.299	0.001	0.014	10	22
G2_M_DNA_DAMAGE_CHECKPOINT	-2.996	0.003	0.024	6	9
P53_INDEPENDENT_G1_S_DNA_DAMAGE_CHECKPOINT	-2.813	0.005	0.034	23	47
BASE_EXCISION_REPAIR	-2.972	0.003	0.025	9	18

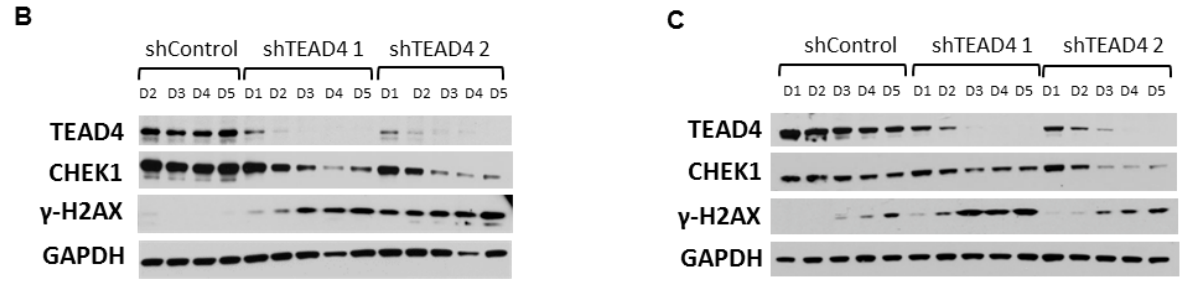


Figure 5- 15. TEAD4 activates DNA damage response programs

(A) Enrichment for REACTOME pathways involved in DNA damage response programs in TEAD4 knockdown signature. (B, C) Immunoblot showing protein levels of TEAD4, CHEK1 and γ-H2AX in (B) SK-N-Be2 and (C) LAN-1 cells transduced with control or two different TEAD4 shRNAs in a time course experiment.

5.2.6 TEAD4 positively regulates MYCN expression in MYCNA cells

To further elucidate the association between TEAD4 and MYCN, we performed a time course experiment upon knockdown of TEAD4 in SK-N-Be2 cells and assessed its effect on MYCN. We observed potent downregulation of MYCN protein upon depletion of TEAD4 by two independent hairpins, in a time dependent manner (Figure 5-16 A). Meanwhile, regulation of MYCN mRNA was not as robust (Figure 5-16 B), indicating indirect post-transcriptional and post-translation regulation of MYCN by TEAD4. There was no evidence of direct protein-protein interaction between TEAD4 and

MYCN by co-immunoprecipitation assay (data not shown). In order to search for genes that could be mediating MYCN regulation, we assessed the transcript level changes of the known modulators of MYCN protein turnover such as AURKA, FBXW7, HUWE1, TRPC4AP and CDK1 complex (Choi et al., 2010; Otto et al., 2009; Sjostrom et al., 2005; Welcker et al., 2004; Zhao et al., 2008c), upon depletion of TEAD4 in our RNA-Seq data. We found that only AURKA and CDK1 were significantly downregulated (AURKA: 2 fold; p-value 2.48E-07; CDK1: 2 fold; p-value 1.38E-07) upon TEAD4 depletion.

We further confirmed that TEAD4 regulates AURKA and CDK1 at protein and mRNA level (Figure 5-16 A and 5-14 B). In addition, MYCN regulation by TEAD4 was validated in another MYCNA cell line, LAN-1, albeit to a lesser degree (Figure 5-16 C and D).

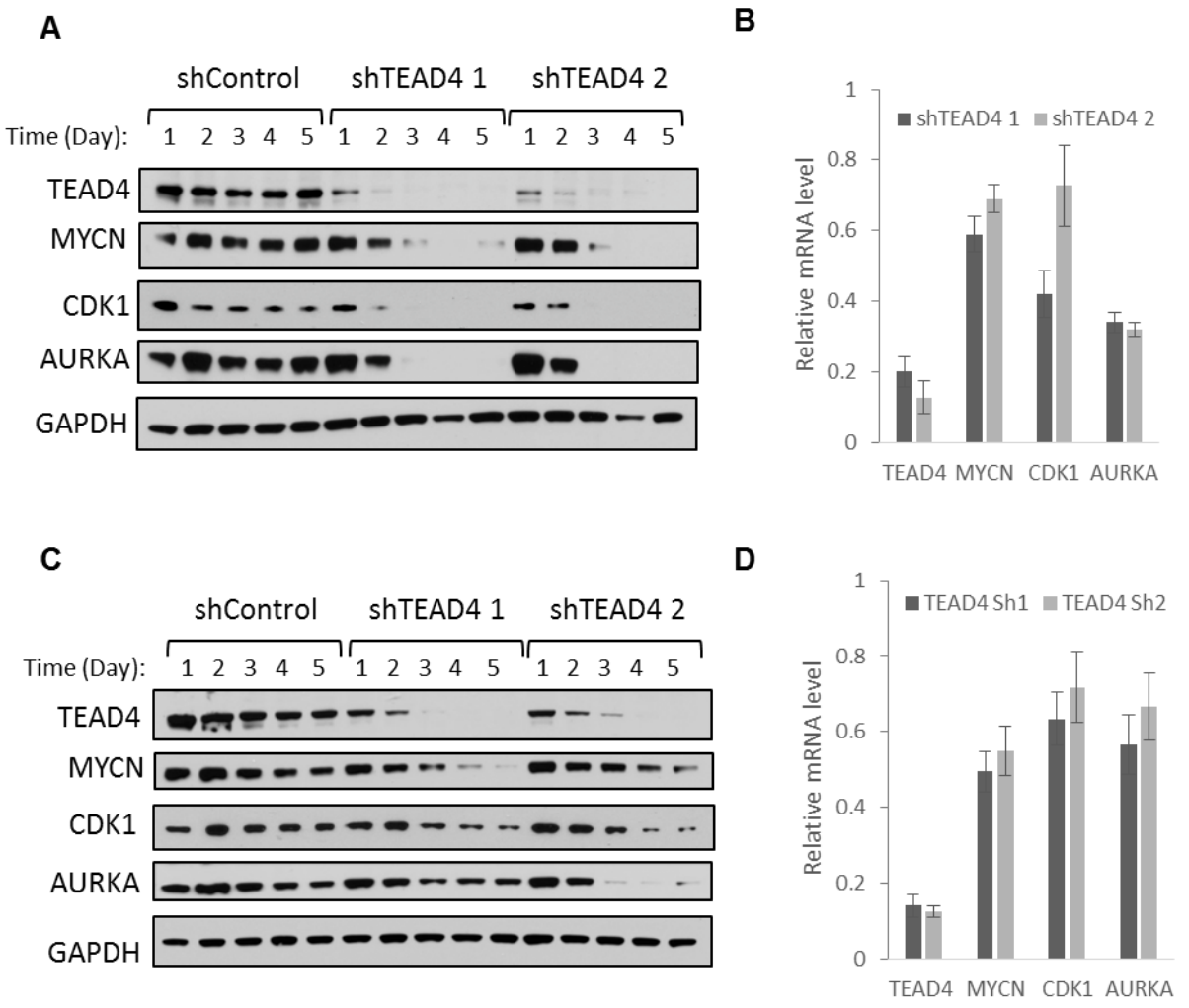


Figure 5- 16. TEAD4 positively regulates MYCN expression

(A, C) Immunoblot analysis of TEAD4, MYCN, CDK1 and AURKA proteins in (A) SK-N-Be2 and (B) LAN-1 cells transduced with control or two different TEAD4 shRNAs in a time course experiment (B, D) qPCR analysis showing TEAD4, MYCN, CDK1 and AURKA transcript level in the corresponding samples in (B) SK-N-Be2 and (D) LAN-1 cells, 2 days post-transduction. Error bars represent standard error calculated from samples run in triplicate. Similar results were obtained at 1, 3, 4 and 5 days' time points (data not shown).

Accordingly, cycloheximide experiments revealed that MYCN turnover was increased by 2-fold upon knockdown of TEAD4, compared to control cells (Figure 5-17 A and B). As cycloheximide inhibits protein synthesis, the decay of the protein of interest can be determined by measuring the protein level over a course of time, by immunoblotting. Consistent with these findings, the degradation of MYCN could be efficiently blocked upon addition of proteasome inhibitor, MG-132 (Figure 5-17 C and D). MG-132 reduces the degradation of ubiquitin-conjugated proteins and hence allows us to assess whether the regulation of MYCN by TEAD4 is post-translational. The degradation of MYCN requires phosphorylation of Serine 62 site by CDK1-Cyclin B complex, followed by phosphorylation of Threonine 58 by GSK3 (Sjostrom et al., 2005). Similarly, AURKA stabilizes MYCN protein by competing with FBXW7 ubiquiin ligase, which degrades MYCN when S62 and T58 are phosphorylated.

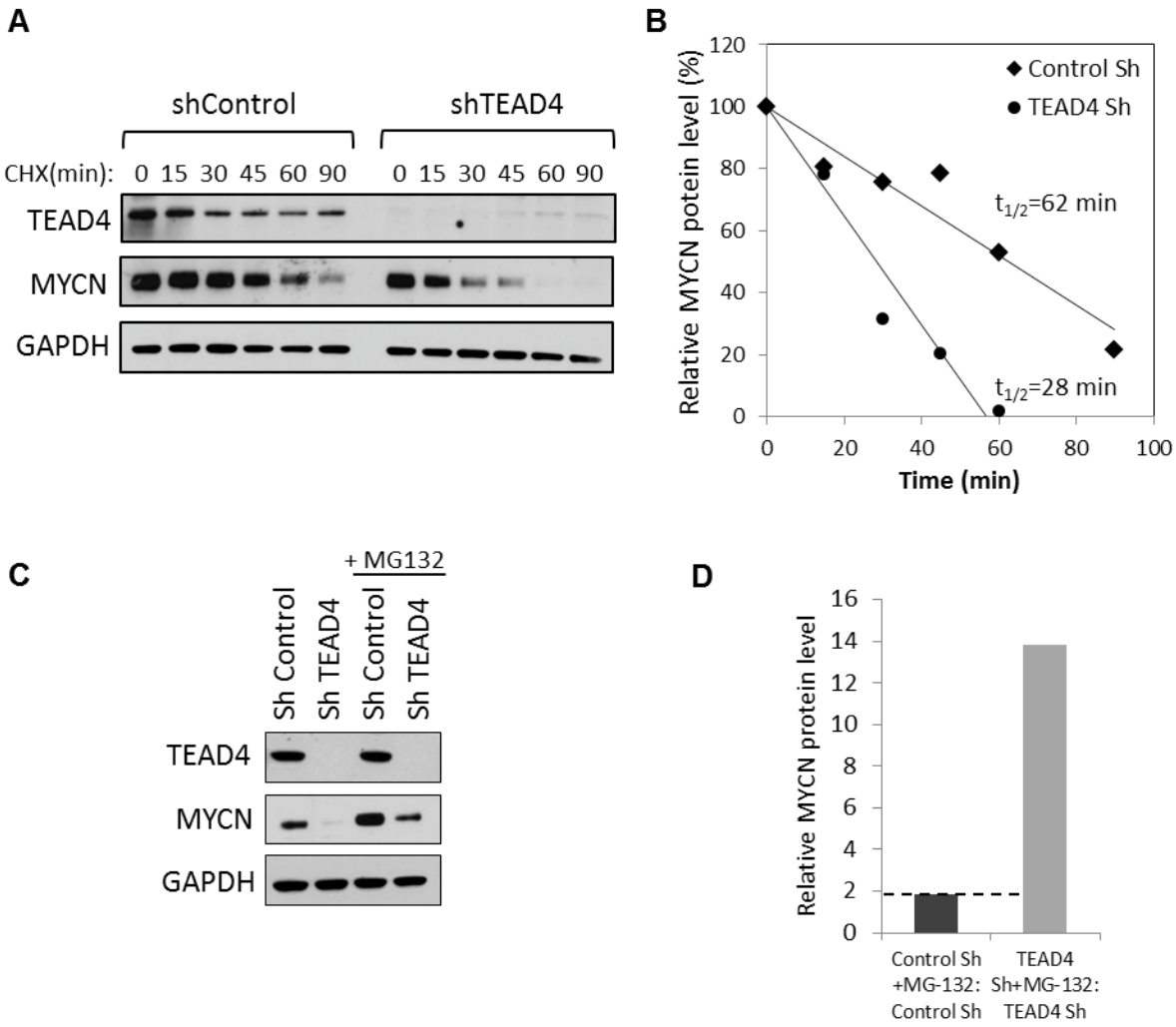


Figure 5- 17. TEAD4 mediates protein stability of MYCN

(A) Immunoblot of TEAD4 and MYCN proteins in SK-N-Be2 cells transduced with control and TEAD4 shRNA 72hrs post transduction, and treated with CHX for indicated times (B) Quantification of MYCN protein stability from results shown in (A) where MYCN level is normalized to GAPDH (C) Immunoblot of TEAD4 and MYCN proteins 72hrs post-induction. To inhibit proteosomal degradation, cells were treated with DMSO or MG-132 4hrs before harvesting. (D) Densitometry analysis of MYCN proteins from results shown in (C), where MYCN level is normalized to GAPDH.

5.2.7 TEAD4 positively regulates MYC protein

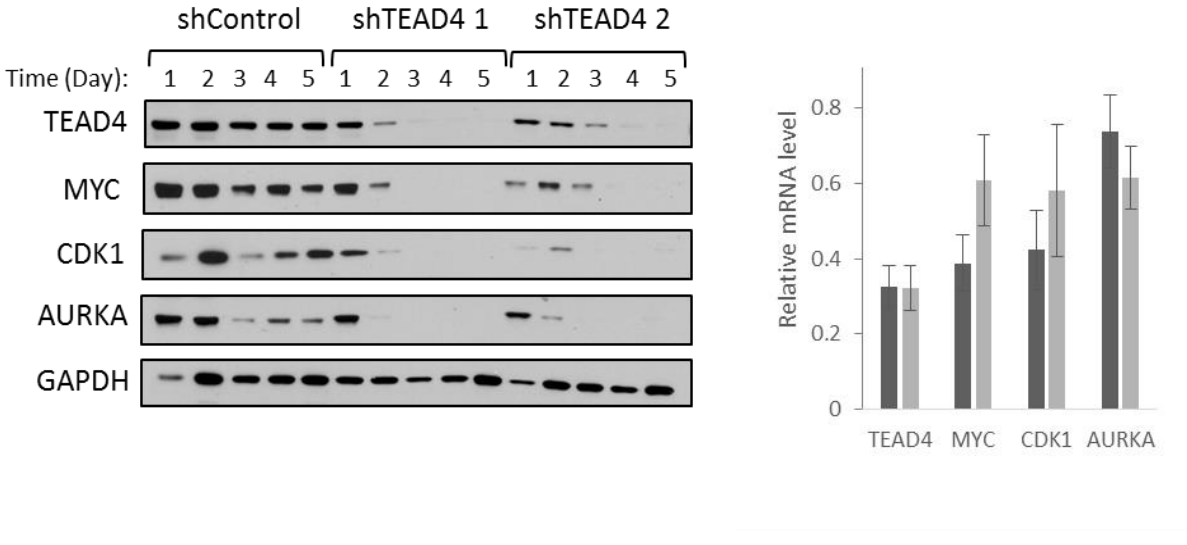


Figure 5- 18. TEAD4 mediates regulation of MYC in *MYCN*-non amplified NBL cell lines

(A, C) Immunoblot analysis of TEAD4, MYC, CDK1 and AURKA proteins in SY-5Y cells transduced with control or two different TEAD4 shRNAs in a time course experiment (B) qPCR analysis showing TEAD4, MYCN, CDK1 and AURKA transcript level in the corresponding samples 2 days post-transduction. Error bars represent standard error calculated from samples run in triplicate.

Stabilization and degradation of MYCN and MYC proteins requires sequential phosphorylation at Serine 62 and Threonine 58 and the sequence around this region is conserved in both (Sjostrom et al., 2005; Yada et al., 2004b). While there are conflicting data in the literature regarding regulation of MYC by AURKA (Otto et al., 2009; Yang et al., 2004), CDK1 has been shown to regulate both MYCN and MYC (Seth et al., 1991; Sjostrom et al., 2005). We examined the regulation of MYC by TEAD4 in MYCN-NA cells, SY-5Y and SK-N-AS. Indeed, we observed that TEAD4 regulates MYC as well (Figure 5-18). Similar to MYCNA cell lines, we observed potent downregulation of MYC, CDK1 and AURKA upon TEAD4 silencing in SY-5Y cell line. On the other hand, while MYC is potently downregulated upon TEAD4 silencing in SK-N-AS cell line, AURKA and CDK1 were not, indicating an alternative mechanism

of MYC regulation by TEAD4 in these cells. Regulation of both MYCN and MYC by TEAD4 suggests that it could play an important role in NBL cells overexpressing MYC/N.

5.2.8 TEAD4 expression/activity is positively correlated with MYCN and MYC

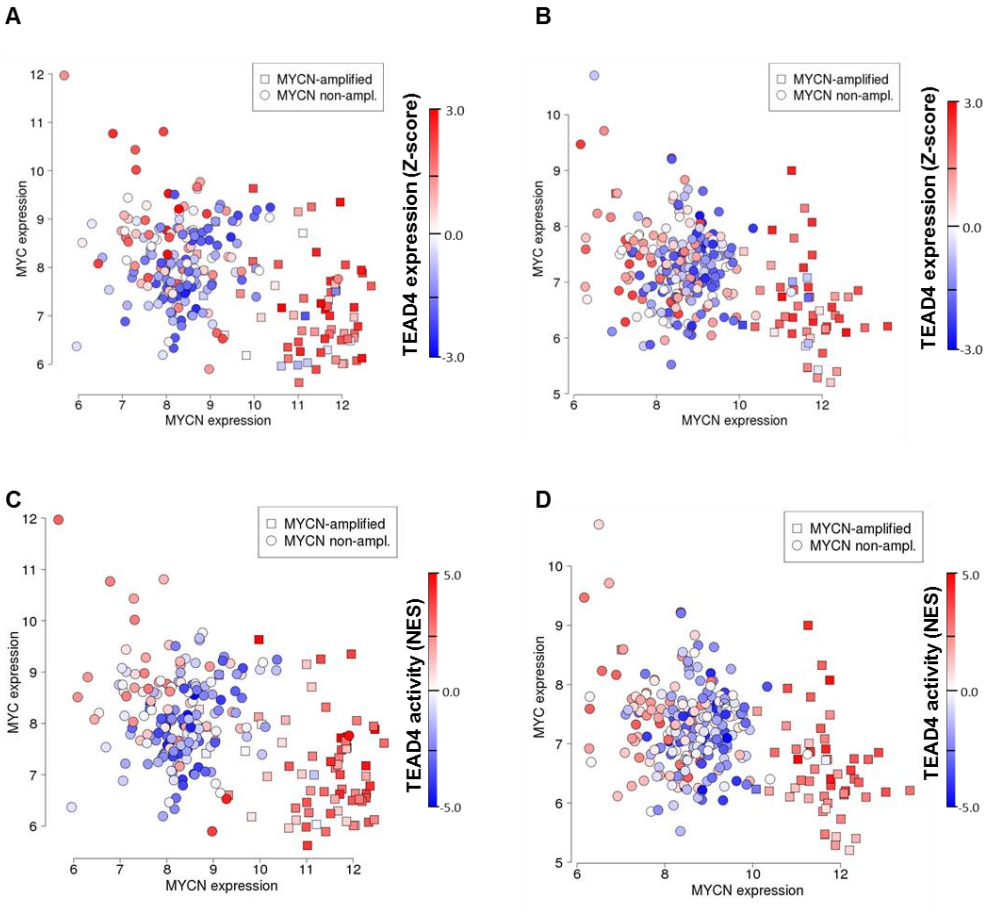


Figure 5- 19. TEAD4 expression and activity is increased in both MYCN and MYC over-expression tumors

Scatter plot represent MYCN and MYC expression in x and y axis respectively. Samples are separated in MYCN amplified and non-amplified. Color code indicates the expression level (A)(B) and the single sample activity represented as normalized enrichment score (NES) (C)(D) of TEAD4. The analysis is shown replicated in TARGET (A)(C) and NRC (B)(D) cohorts.

To gain further insight into the interplay between TEAD4 and MYCN/MYC, we assessed their correlation in expression in NBL patient samples. It has been reported that the expression of MYCN and MYC occurs in a mutually exclusive fashion by repressing each other at specific promoter regions in NBL cells (Breit and Schwab, 1989). We sought to determine whether TEAD4 expression and activity correlate with MYCN amplification/expression and MYC expression. Indeed, both TEAD4 expression and activity, to a greater degree, are positively correlated with *MYCN*-amplification, MYCN expression as well as MYC expression (Figure 5-19).

Taken together, our data indicates that TEAD4 could play an important role in driving MYCN mediated oncogenesis through positive feedback loop, thus providing means to inhibit their activity in high-risk NBL.

5.2.9 TEAD4 is required for growth of NBL cells

The observation that TEAD4 regulates MYCN/MYC and cell cycle related programs implicates a context specific role of TEAD4 in high-risk NBL. To evaluate the phenotypic consequences of TEAD4 depletion in MYCNA and MYCN-NA cell lines, we chose cell lines with varying degree of expression of MYCN and MYC. These cell lines express similar levels of TEAD4 protein expression (Figure 5-20 A). We performed soft agar colony forming assay to determine the effect of TEAD4 depletion in tumorigenecity of these cells. Knockdown of TEAD4 led to dramatic reduction in colony formation in MYCNA cell lines, SK-N-Be2 and LAN-1. In contrast, there was no change in colony formation in SK-N-FI and 40% decrease in SK-N-AS cell lines (Figure 5-20 B). The decrease in colony formation of SK-N-AS compared to SK-N-FI cells might also be attributed to higher expression of MYC in SK-N-AS. Knockdown of TEAD4 in these cell lines were confirmed at a protein level (Figure 5-20 C).

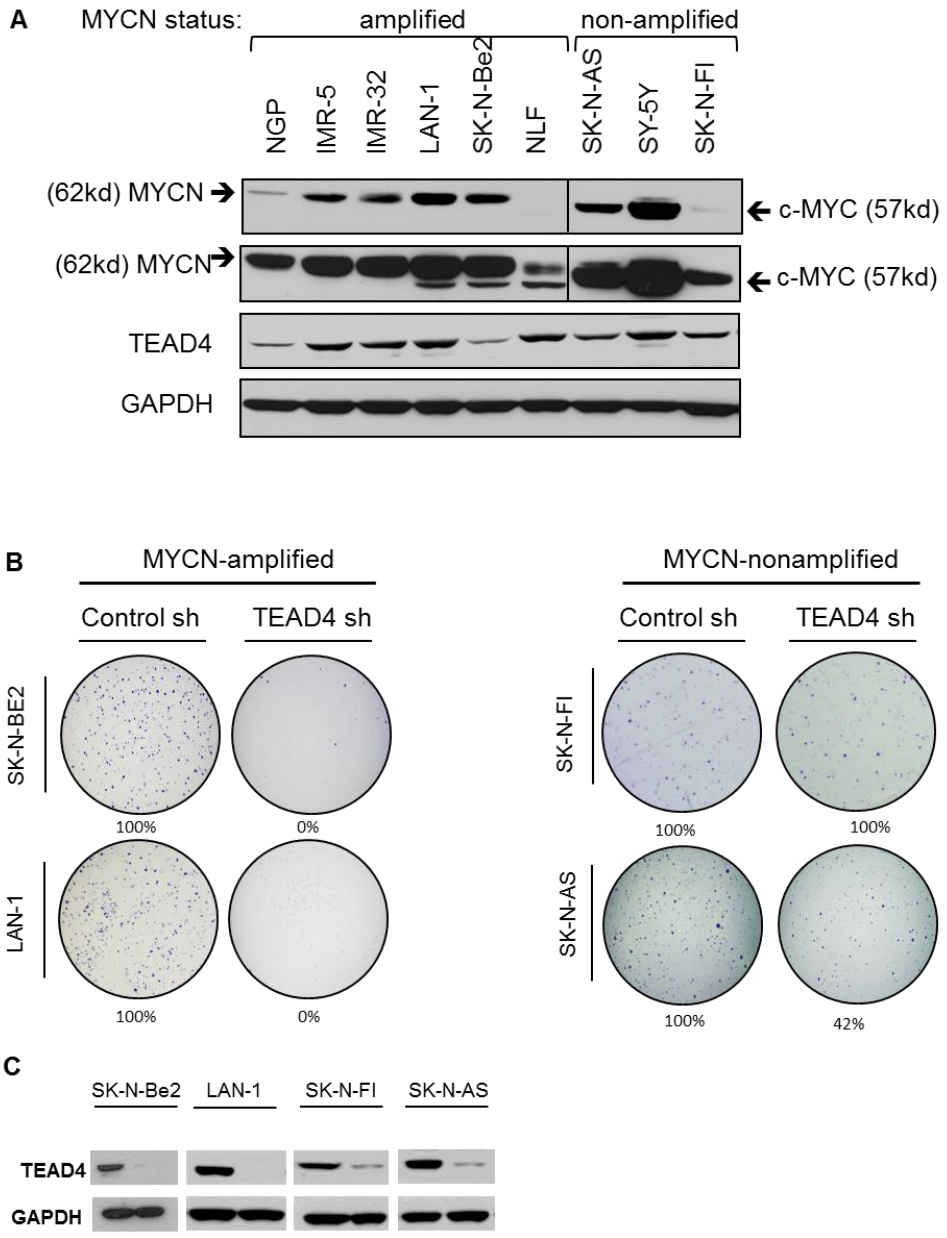


Figure 5- 20. TEAD4 is required for cell growth of cells of NBL cells with high MYC/N expression

A) Immunoblot analysis of TEAD4, MYCN and MYC proteins in a panel of MYCN-amplified and non-amplified cell lines. (A) The effect of TEAD4 on anchorage-independent growth in MYCNA and control cell lines was evaluated by soft agar colony assays, 21 days post transduction (B) Immunoblot analysis confirming silencing of TEAD4 in the corresponding cell lines.

5.2.10 TEAD4 protein is differentially expressed in NBL patient subgroups

Analysis of TEAD4 mRNA expression level and protein activity in NBL patient samples revealed significantly higher level of TEAD4 expression and activity in MYCNA subtype compared to other high risk NBL subtypes, with Stage I samples exhibiting lowest level of TEAD4 (Kruskal-Wallis test, p-value = 3.05e-11). (Figure 5-21 A and B). We further performed tissue microarrays to confirm differential level of protein expression of TEAD4 in various NBL tissue samples. We analyzed TMA containing primary tumors from Children's hospital of Philadelphia, comprising 116 neuroblastoma samples at different stages, risk-levels and *MYCN*-amplification status. Our results showed that high-risk group expressed higher level of TEAD4 protein compared to intermediate and low-risk group, with it being least expressed in normal tissues (adrenal, brain, tonsil) (Figure 5-21 C). We derived a final score (intensity of staining (0-3) X percentage of cells stained (0-100)) to and observed that high risk group has the highest score/TEAD4 expression (high-risk: 89.82; intermediate risk: 62.81; low-risk: 62.57; normal tissue: 53.33). Among the high-risk group, *MYCN* amplified cells had higher expression of TEAD4 compared to MYCN-NA cells by intensity and final score (MYCNA: 139.44; MYCN-NA: 66.31) (Figure 5-21 D). Representative images of TEAD4 protein expression analyzed in primary NBL samples by immunohistochemistry on a tissue microarray chip, segregated by stage of patients are shown (Figure 5-21 E).

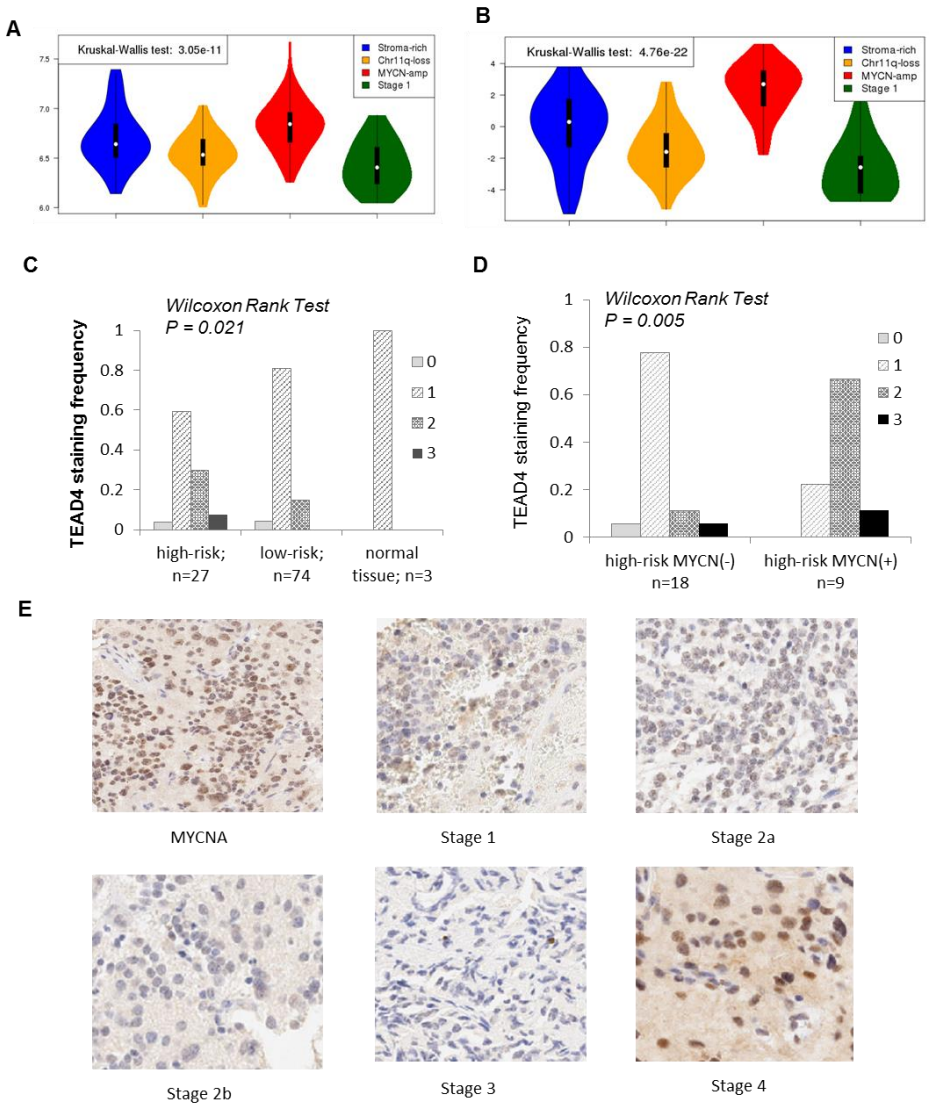


Figure 5- 21. TEAD4 expression and activity in NBL patient samples

(A) TEAD4 mRNA expression and (B) VIPER activity derived from NBL patient GEPs across NBL subtypes (C) Histogram of primary NBL samples stained for TEAD4 by immunohistochemistry on a TMA, segregated by risk level and (D) MYCNA status in high-risk NBL, showing differential pattern of TEAD4 protein staining intensity (where 0=no staining; 1=low staining; 2=moderate staining; 3=high staining) (E) Representative immunohistochemistry staining of NBL tissue microarrays, separated by Stage of patients and MYCN amplification status, with TEAD4 antibody. Immunohistochemistry staining and analysis were performed by Daniel Martinez and Mark Yarmarkovich in the lab of Dr. John Maris at Children's hospital of Philadelphia, PA.

5.2.11 TEAD4 expression is an independent predictor of prognosis in high-risk NBL

To understand the clinical relevance of TEAD4 expression in NBL patient samples, we assessed the association of TEAD4 expression with NBL subtypes as well as various clinical and pathological features. We performed multivariate analysis to study the prognostic value of TEAD4 expression in NBL tumors using cox regression (Cox, 1972). We focused our analysis on NRC cohort since TARGET cohort is mainly composed of high-risk tumor patients, therefore is not valid for an unbiased survival analysis. Interestingly, TEAD4 expression is an exceptional predictor of patient survival (p-value: 5.75e-8, HR=7.8) (Figure 5-22).

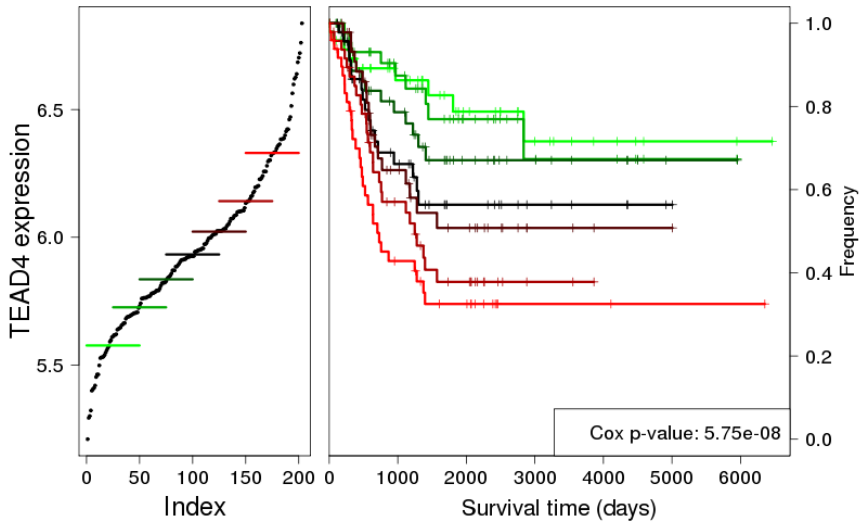


Figure 5- 22. High TEAD4 expression is associated with NBL patient outcome.

TEAD4 expression is Kaplan-Meier curve depicting corresponding increase in poor outcome with increasing expression of TEAD4. P-value was calculated using a cox proportional hazards model after removing stage I patient samples.

A remarkable feature of NBLs is that there exist fairly good systems to predict outcome at diagnosis using clinical variables and genetic markers (Brodeur and Bagatell, 2014; Maris, 2010). The most important clinical variables predicting patient outcome are stage of disease and age of the patient

(Brodeur et al., 1993b). Moreover, *MYCN*-amplification has been shown to be associated with poor prognosis regardless of age of patient and stage of disease (Brodeur et al., 1984; Maris and Matthay, 1999b; Seeger et al., 1985a). To determine whether TEAD4 provides predictive power over these most relevant clinical covariates, we performed multivariate cox regression analysis of TEAD4 expression with these clinical groups. TEAD4 expression predicts survival independently of the key prognostic variable including stage, age group stratified at 18 months, *MYCN* amplification status and a combination of the three variables (Table 5-4).

	HR	HR 95% CI		P.value
TEAD4	7.84	3.73	16.49	**5.75E-08**
TEAD4+Stage	4.98	2.36	10.49	**2.39E-05**
TEAD4+MYCNA	3.99	1.66	9.61	**2.04E-03**
TEAD4+Age	7.14	3.25	15.67	**9.53E-07**
TEAD4+Stage+MYCNA+Age	3.69	1.40	9.73	**8.39E-03**

Table 5- 4. Tead4 Multivariate Cox proportional hazards regression analysis with clinical covariates

As TEAD4 and MYCN regulate each other by positive feedback loop, their expressions are highly correlated with *MYCN* amplified cells expressing significantly higher level of TEAD4, we assessed whether TEAD4 has prognostic information independent of MYCN. We observed that TEAD4 predicts survival independent of MYCN expression (Table 5-5). However, in order to account for the effect of MYCN protein stabilization and MYC functional redundancy, we measured the total MYC/N transcriptional activity (Valentijn et al., 2012) by single sample GSEA (Barbie et al., 2009b) and observed that TEAD4 predicts survival independent of MYC/N protein activity (Table 5-5).

	TEAD4				Biomarker				Additive effect		
Expression	**HR**	**HR 95% CI**		**P.value**	**HR**	**HR 95% CI**		**P.value**	**Conc.**	**Wald.test**	**Log.test**
MYCN	5.09	2.23	11.61	**1.11E-04**	1.23	1.06	1.42	**5.26E-03**	0.69	2.41E-09	8.46E-09
NTRK1	2.68	1.14	6.33	**2.43E-02**	0.63	0.52	0.77	**7.21E-06**	0.73	1.17E-10	1.93E-11
Signature											
MYCN-act (ref)	4.28	1.96	9.35	**2.70E-04**	4.27	2.23	8.18	**1.21E-05**	0.73	1.25E-11	1.62E-11
meta-PCNA	6.01	2.82	12.79	**3.28E-06**	4.24	2.11	8.51	**4.90E-05**	0.73	1.43E-10	3.49E-11

CI, confidence interval; HR, hazard ratio;

Table 5- 5. Tead4 Multivariate Cox proportional hazards regression analysis with other biomarkers and relevant signatures

Similarly, NTRK1 expression is one of the most reliable biomarker of good prognosis of NBL ((Brodeur and Bagatell, 2014). TEAD4 again predicted survival independent of NTRK1 expression (Table 5-5). Additionally, we measured whether TEAD4 association with survival can be explained only by its impact in cell cycle and the proliferative level of the tumor. Most biomarkers associated with cancer survival have been deemed to be correlated with proliferation (Venet et al., 2011), we measured the activity of a meta-PCNA signature and again observed TEAD4 independency as a survival predictor (Table 5-5). Hence, the survival data implies that TEAD4 expression could be used as an important biomarker of patient outcome in NBL patients.

5.2.12 TEAD4 expression is not correlated with YAP/TAZ expression and activity in NBL cells

As YAP/TAZ/TEAD protein complex has been shown to be an effector of hippo signaling pathway to control the transcriptional activation of their target genes, we assessed whether there was differential activity of YAP/TAZ in NBL cells. As activity of YAP/TAZ is defined by their nuclear/cytoplasmic localization, we tried several measures to study their differential activity in high-risk NBL subtypes. Surprisingly, YAP and TAZ activity, as defined by the expression of their target genes retrieved from msigDB, were not correlated with the expression of TEAD4 (Figure 5-23 A, B).

Similarly, we observed that there was no clear discrimination in nuclear localization of YAP/TAZ in MYCNA vs. MYCN-NA cell lines (Figure 5-23 C). Furthermore, expression and activity of YAP/TAZ did not show any prognostic value (Figure 5- 24).

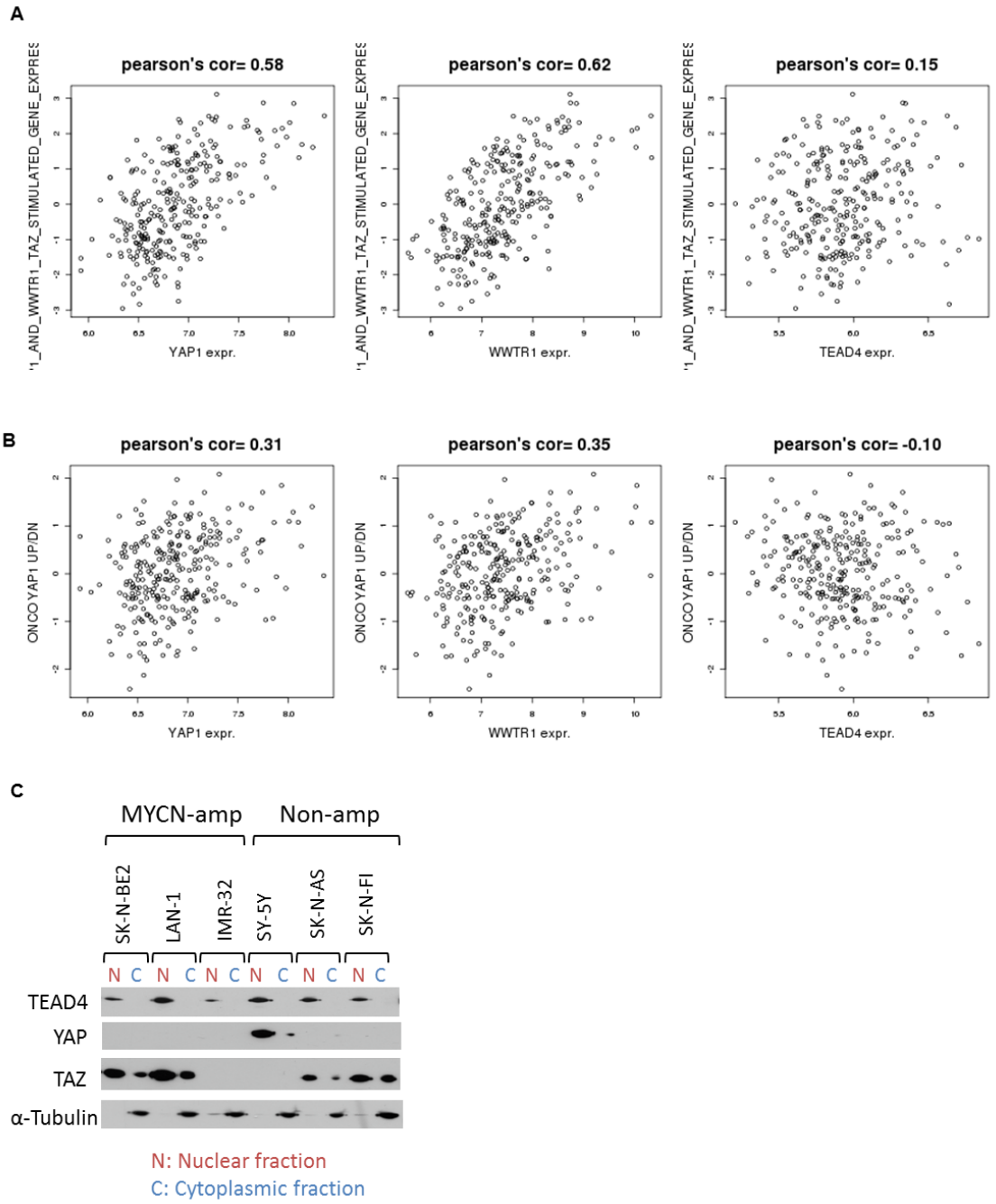


Figure 5- 23. TEAD4 expression does not correlate with YAP/TAZ or activity

(A) Scatterplot showing correlation between YAP/TAZ activities based on gene sets retrieved from REACTOME and msigDB databases (C) Immunoblot analysis of nuclear and cytoplasmic localization of TEAD4, YAP and TAZ in *MYCN* amplified and non-amplified cell lines.

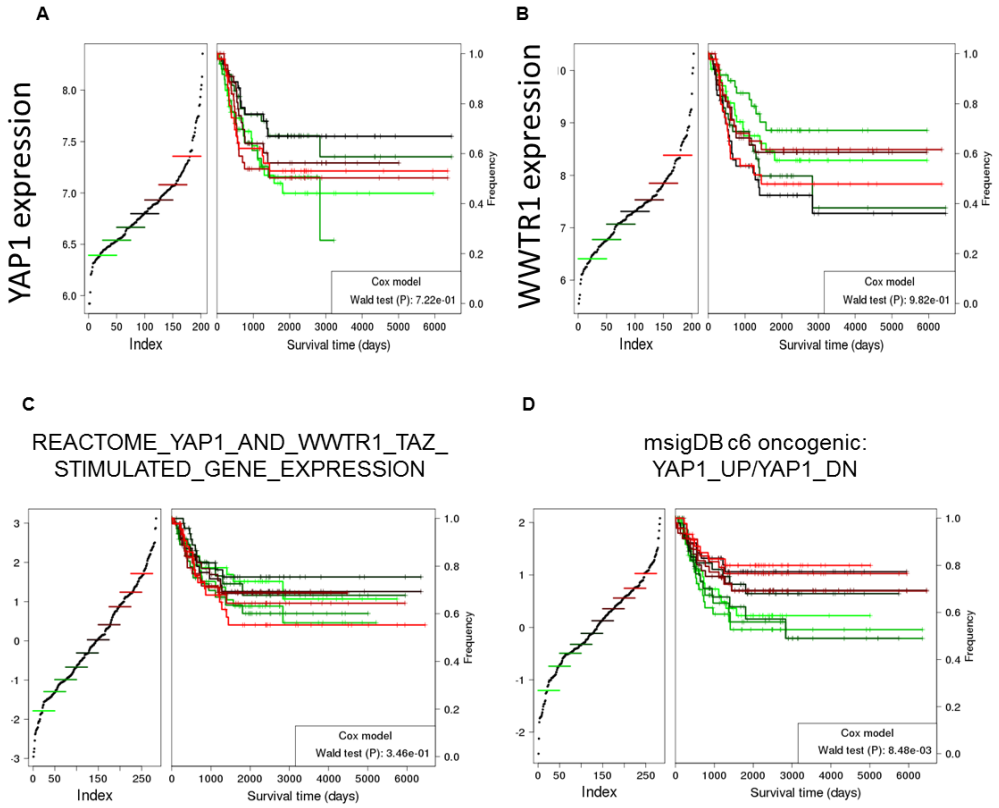


Figure 5- 24. YAP/TAZ expression and activity has no prognostic value in NBL

Kaplan-Meier curve depicting outcome corresponding to increasing expression of (A) YAP and (B) TAZ as well as activity of (C) YAP and TAZ and (D) YAP derived from msigDB based gene set expression. P-value was calculated using a cox proportional hazards model after removing stage I patient samples.

5.3 Materials and methods

Cell lines and cell culture

All neuroblastoma cell lines were maintained in DMEM or RPMI-1640 supplemented with 10% FBS, 20 mM L-glutamine and antibiotics. 293FT cells were maintained in DMEM supplemented with 10% FBS and antibiotics. Treatment of SK-N-Be2 cells with cycloheximide (100µg/ml) and MG-132 (20µM) were carried out as indicated.

Pooled shRNA screening

Lentiviral particles encoding Control shRNA (SHC002) or 4-5 hairpins against each MR were purchased from Sigma. 5ul of the lentivirus were infected into SK-N-BE2, IMR-5, SK-N-AS and NLF in the presence of 2-8ug/ml polybrene. The cells are transduced with unique shRNAs in a 96 well plate, followed by pooling of the cells and puromycin selection with 0.5µg/ml puromycin. For in-vitro screening, he cells were collected after puromycin selection (T_0= day 4) and (T_f= day 28). For in-vivo screening, input cells were collected (T_0= 10 days) and 3 X 106 cells were implanted into nude mice and tumors were collected (T_f= 4 weeks) after implantation. Genomic DNA was isolated by shRNA representation measured at each time points; shRNA cassettes were amplified from the genomic DNA using primers with Illumina adapters and barcode sequences. The samples were pooled together and the amplicons were sequenced using Illumina MiSeq® instrument. The raw sequencing data of shRNA screens was deconvoluted and normalized to count per million in each sample (Yu et al., 2013). Then we analyzed the normalized shRNA abundance data by ScreenBEAM method (Yu et al., 2015) to identify essential genes, whose hairpins were depleted over time, in each cell line, with in vitro and in vivo being separated. To integrate gene-level dropout scores of multiple cell lines in one group, for example, MYC-amplified, we performed the meta-analysis using Stouffer's z score method (Stouffer et al., 1949).

Lentiviral transduction

Lentiviral particles encoding control shRNA (SHC002, SHC202) and 2 hairpins against each MR were produced as described previously (Lefebvre et al., 2010). We used pLKO.1-puro lentiviral constructs (Sigma) (Table S1). The lentiviruses were concentrated using Amicon Ultra Centrifugal filter unit-100K (Millipore) for 40 min or ultracentrifugation at 25,000rpm in SW28 rotor for 1.5 hrs (Beckman). Neuroblastoma cell lines were transduced with the viral particles in the presence of 2-8µg/ml polybrene. The transductions were performed with a mix of 2 hairpins per gene to control for off-target effects, unless otherwise stated.

siRNA transfection

ON-TARGET plus siRNA smartpool against the MRs and non-targeting control (Dharmacon) (Table S2) were used to transfect the neuroblastoma cell lines at a concentration of 20nM, using Dharmafect reagent (Dharmacon).

Immunoblot analysis

For immunoblotting, whole-cell lysates were prepared by using RIPA buffer (Teknova) with protease inhibitor cocktail (Roche). Proteins extracts were resolved by SDS–PAGE and analyzed by standard procedures using the following antibodies: TEAD4 (ab58310; abcam), MYCN (9405; Cell Signaling), MYC (5605; Cell Signaling) GAPDH (sc-32233; Santa Cruz), α-tubulin (sc-8035; Santa Cruz), AURKA (14475; Cell Signaling), CDK1 (9116; Cell Signaling), CHEK1 (2360; Cell Signaling), YAP/TAZ (14475; Cell Signaling). Densitometry analyses were performed using ImageJ software.

RNA isolation, RNA-Seq and RT-PCR analyses

Total RNA were extracted using Direct-zol RNA Miniprep kit (Zymo Research). For RT-PCR, cDNA was prepared using qScript Flex cDNA Synthesis kit (Quanta Biosciences), and the amount of transcripts were quantified using Power SYBR Green master mix (Thermo Fisher Scientific), with the respective oligonucleotides (Table S3) in Applied Biosystems 7300 RT-PCR system. The number of copies of each gene was normalized to the housekeeping gene, GAPDH and HPRT1. For RNA-Seq, the integrity of RNA was analyzed using Bioanalyzer (Agilent Technologies). RNA-Seq libraries were made with Illumina TruSeq RNA Sample Prep protocol by Columbia Genome Center and sequenced on Illumina NextSeq 500. All libraries have >30 milllion reads sequenced. The libraries were aligned to hg19 human genome using TopHat. Raw counts of each gene were generated using GenomicFeatures R-system package (Bioconductor). The data was then normalized using voom (Law et al., 2014) and significantly differentially expressed genes were retrieved using limma R-system package (Bioconductor).

Tissue microarray analyses

TEAD4 (ab58308; abcam) antibody was used to stain formalin fixed paraffin embedded tissue slides. Staining was performed on a Bond Max automated staining system (Leica Microsystems). The Bond Refine polymer staining kit (Leica Microsystems) was used. The standard protocol was followed with the exception of the primary antibody incubation which was extended to 1 hour at room temperature. TEAD4 antibody was used at 1:1K dilution and antigen retrieval was performed with E1 (Leica Microsystems) retrieval solution for 20min. Slides were rinsed, dehydrated through a series of ascending concentrations of ethanol and xylene, then coverslipped. Stained slides were then digitally scanned at 20x magnification on an Aperio CS-O slide scanner (Aperio, Vista CA).

Cell viability assays

Cell viability was measured by resazurin based assay using Presto blue cell viability reagent (Thermo Fisher Scientific) by following manufacturer's protocol.

Cell cycle, proliferation and apoptotic assay by Flow Cytometry

For cell cycle analysis, SK-N-Be2 cells were fixed in 70% ice-cold ethanol, stained with propidium iodide (Sigma). Proliferation assay was performed by evaluation of EdU incorporation by using Click-It EdU Alexa Fluor 488 kit (Thermo Fisher Scientific) using manufacturer's protocol. Similarly, apoptotic assay was performed using Annexin V staining kit (BD Pharmingen), according to manufacturer's protocol). All analyses were performed on BD LSRII Cell Analyzer and FlowJo software (Tree Star).

Colony forming assays

Soft agar colony forming assays were performed as described previously (Franken et al., 2006). In brief, 5000-10000 cells were seeded in Sea Plaque Low Melt Agarose (Lonza) in growth media for 3 weeks. Cells are then fixed, stained with crystal violet and imaged under dissecting microscope. The colonies were quantitated using ImageJ software.

Bioinformatic analyses

Refer to Chapter 4 materials and methods.

5.4 Discussion

With the emergence of high-throughput data characterizing genetic, epigenetic and functional properties of normal and tumor related samples, the list of putative causal drivers of malignant phenotypes are continuously expanding. The validity of these putative drivers is often lost in the heterogeneity of tumor samples and awaits functional validation. In the previous chapter, we have shown that classification of tumors into distinct molecular subtypes followed by interrogation of context-specific regulatory network can be used to identify the transcription factors driving subtype specific signature crucial for deciphering the heterogeneity of high-risk NBL. In this chapter, we have shown that such computational approaches when integrated with experimental evidence aids elucidation of key mechanisms underlying subtype specificity and decipher the control points in tumors. In the case of NBL tumors with high activity of MYC/N, our results demonstrated that TEAD4 is the critical MR of this subtype. TEAD4 is neither mutated nor is the most differentially expressed gene in MYCNA cells and hence has escaped the detection process that relies on identification of classical oncogenes by these measures.

Even though genome-wide RNAi screenings have resulted in important discoveries in different fields and contexts, reproducibility and specificity has been an issue (Jackson and Linsley, 2010; Mohr et al., 2010). The noise and false discovery inherent to the high-throughput approaches are exacerbated by the off-target effects (Mohr et al., 2010, 2014). These shortcomings of the RNAi screening strategies were offset by our evidence-based focused screening by multiple RNAi platforms. Moreover, it allowed maximal representation of individual shRNAs for in-vivo pooled screening and hence reduced the clonal dominance inducing hairpin representation bias, which is a recurrent problem in such screens (Nolan-Stevaux et al., 2013). Hence, we could identify TEAD4 as a master regulator of MYCNA subtype with high degree of confidence.

Indeed, further experimentation demonstrated that TEAD4 is necessary for sustaining the proliferative state of MYCNA cells by robustly activating several inter-regulatory components of cell cycle machinery involved in intricate control of the system. Most likely, the ability of TEAD4 to promote proliferation probably stems from its ability to regulate several genes implicated in S-phase entry and progression, by activating several components of origin licensing and assembly of pre-replicative complexes, to ensure robustness of DNA replication. In addition, our results also indicated that TEAD4 is critical in stabilizing MYCN as well as MYC proteins. This is significant in the light of recent findings suggesting that malignant neuroblastoma progression is driven by protein overexpression of MYCN or MYC (Huang et al., 2011; Wang et al., 2013; Westermann et al., 2008), and that these MYC-driven NBLs comprises the highest-risk NBL cases (Wang et al., 2015b). Furthermore, our results indicate that TEAD4 could potentially act upstream of the kinases that has been shown to be synthetic lethal to MYC/N addicted cells like AURKA (Otto et al., 2009), CDK1 (Goga et al., 2007; Sjostrom et al., 2005), CDK2 (Molenaar et al., 2009), and CHK1 kinases (Cole et al., 2011), some of which were validated at the protein level (Figure 5-16, 5-17).

As observed with MYC/N and other oncogenes, it is plausible that activation of TEAD4 in concert with MYC/N allows the cells to undergo rapid proliferation and replication stress, and simultaneously activates the DDR pathways including ATR-Chk1 signaling molecules to support the replication machinery and restrain the replicative stress – to allow continued proliferation and survival of the cells. Recent studies have shown that the enhanced oncogene induced replication stress and DNA damage leads to dependency on stress support mechanisms that could be the Achilles's heel of tumor cells with high replication stress (Bartek et al., 2012). It has been postulated that the oncogene mediated DNA DSBs in precancerous lesions contribute to genomic instability and when combined with p53 inactivation allows the cells to be dependent on DDR response pathways for tumor formation and maintenance (Halazonetis et al., 2008b). Consistently, inhibition of Chk1 and Wee1 has been shown to exhibit strong synthetic lethality in MYC driven tumors including NBLs with high MYC/N activity by inducing potent genotoxic response (Cole et al., 2011; Murga et al., 2011; Russell et al., 2013). As MYC/N are capable of inducing oncogene-induced replication stress and genomic instability

(Dominguez-Sola et al., 2007; Felsher and Bishop, 1999; Vafa et al., 2002), the positive feedback loop between TEAD4 and MYC/N could potentially support this mechanism for tumor initiation and maintenance. The landscape of oncogene and non-oncogene addiction is vast and potentially provide crucial cancer drug targets (Luo et al., 2009a). TEAD4 could potentially contribute towards this phenomenon, where the MYC/N activated tumor cells depend on these genes to maintain balance between cellular proliferation and stress pathways to continue survival and expansion of the cells. Together, TEAD4 and MYCN controls the hallmarks of MYCNA subtype such as high proliferation and cell growth, activated DNA damage response pathways, and undifferentiated phenotype.

TEAD4 has been shown to be critical effector of hippo signaling pathway by mediating YAP and TAZ dependent cellular growth, oncogenesis and mesenchymal transformation (Harvey et al., 2013; Zhao et al., 2008b). However, we observed that TEAD4 expression is not correlated with active form of YAP or TAZ and that YAP and TAZ do not exhibit prognostic value like TEAD4. Recent studies showed that TEAD4 could function independently of Hippo pathway (Liu et al., 2015), and that it does indeed have a transcriptional activation domain which allows it to induce transcriptional activation independent of the YAP/TAZ binding domain (Qiao et al., 2015) — in contrast to the long believed view that TEAD4 is unable to induce transcriptional activation by itself (Xiao et al., 1991). Additional experiments are required to assess whether TEAD4 acts in concert with YAP/TAZ or independent of the binding partners, in the context of NBL. TEAD4 has been shown to be required for early stages of embryonal development (Yagi et al., 2007) and pathogenesis of several cancers (Lim et al., 2013; Liu et al., 2015; Wang et al., 2015). Whether TEAD4 has a role in neural crest development or only during tumorigenesis remains to be established.

Further studies are required for in-depth understanding of the implication of TEAD4 activity in NBL biology particularly in relevance to clinical significance. Unlike the NBL cell lines, which expressed TEAD4 at a similar level, disregard of MYC or MYCN expression level or *MYCN*-amplification status, we observed that TEAD4 was overexpressed in MYCNA tumors among the high-risk group of patients, hence providing an opportunity for more specificity in inhibiting the growth of MYCNA cells by inhibiting

TEAD4 activity in these cells. Given that TEAD4 expression predicts patient outcome better than MYCNA status, it could potentially be used as a biomarker to facilitate better clinical evaluation.

Chapter 6

Conclusions and Future directions

The emergence of multidimensional datasets characterizing genome-wide transcriptional, genetic, epigenetic, and functional properties of large normal and tumor-related samples are aimed at understanding the biological basis of cancer – to provide more effective targeted therapeutic approaches. The computational approaches to extract biological insights from these datasets, and experimental approaches to ensure validity of the computational predictions are essential to identify bona fide targets of therapeutic value. This thesis is a contribution towards this objective — providing systematic dissection of the transcriptional, genetic and clinical profile of the neuroblastoma patients, followed by interrogation of context-specific transcriptional regulatory pathways using computational and experimental approaches.

The contribution comes both in terms of developing a pipeline that could be applied to dissect the heterogeneity and identify the molecular drivers of other tumor types; and also by providing a molecular target for MYCNA subtype of high-risk NBL. Firstly, the pipeline yields subtype specific targets or master regulator modules by assembly and interrogation of the transcriptional networks in tumor specific context. In general, cancers including NBL have a complex and heterogeneous genomic landscape, which channels through the functional bottlenecks or master regulators that are responsible for maintaining the tumor state. With increasing amount of studies revealing diverse mutations and low paucity of high-frequency mutations in most tumors (Cancer Genome Atlas Network, 2012; Jones et al., 2008; Parsons et al., 2008; Wood et al., 2007), it is becoming increasingly evident that the we should shift our focus to identify the critical nodes within the complex regulatory network that the tumorigenic cells rely on. Rather than focusing on the initiating genetic events, we focused on core regulatory machinery responsible for implementation and maintenance of tumor state. Our results demonstrate the validity of our hypothesis; which led to elucidation of three distinct subtypes of high-risk NBLs as well as core regulatory machinery responsible for their implementation and stability, including canalization and integration of mutational events and the structural variants that represent the hallmark of this disease.

Hence, the strategy shown here for computational prediction and systematic experimental validation could effectively be used to process other tumor phenotypes to yield validated targets that constitute either oncogene or non-oncogene dependencies of the tumor (Luo et al., 2009b).

The major challenge in NBL remains discovery and implementation of targeted therapeutics — as the children with high-risk NBL continue to be treated with conventional chemotherapy. While the patients respond to therapy in the beginning, eventually, they relapse and succumb to the disease. Even though synthetic lethal screens have identified some potential drug targets for *MYCN*-amplified tumors (Cole et al., 2011; Molenaar et al., 2009; Otto et al., 2009; Russell et al., 2013; Sjostrom et al., 2005); similar to other tumors, the complexity and heterogeneity of NBL's genomic landscape suggests that such an approach for single target may lead to relapse following expansion of drug resistant subclones harboring either bypass mechanisms or alternative genetic variants. Our approach provides an advantage by finding the MRs that incorporate signaling from the genetic information as well as signaling cascade, and hence provide more universal dependency, as evidenced by our results. Thus, following on recent results from assembly and interrogation of regulatory networks, we focused on more universal and core tumor dependencies – that are relatively independent of the specific genetic alteration landscape of tumors with similar transcriptional profiles.

By performing exhaustive experimental analyses of MYCNA subtype MRs, we have identified TEAD4 as the key transcriptional node, which in concert with MYCN drives the hallmarks of MYCNA subtype of NBL including hyper proliferation, activation of DDR pathways, activated cellular growth programs and repression of differentiation programs. We have demonstrated that TEAD4 is necessary for proliferation of MYCNA cells by transcriptional regulation of several genes implicated in cell cycle progression and DNA replication. Furthermore, it also activates DDR pathways including ATR-Chk1 signaling molecules, which has recently been shown to be a necessary mechanism taken by the tumor cells with high replication stress and genomic instability to cope with the stress for sustained cell proliferation (Bartek et al., 2012). In addition, our results also indicated that TEAD4 is critical in stabilizing MYCN as well as MYC proteins, which is significant in the light of recent findings that NBLs with high MYC/N comprises

the highest risk cases (Wang et al., 2015b). Furthermore, our results indicate that TEAD4 indeed represents the universal dependency of MYC/N activated tumors as demonstrated by its ability to regulate several kinases that has been shown to be synthetic lethal to MYC/N addicted cells like AURKA (Otto et al., 2009), CDK1 (Goga et al., 2007; Sjostrom et al., 2005), CDK2 (Molenaar et al., 2009), and CHK1 kinases (Cole et al., 2011). Given that MYCNA tumors express higher level of TEAD4 mRNA and protein, and the biological relevance of the TEAD4 regulated programs, it is plausible that TEAD4 acts in concert with MYCN to induce rapid proliferation and replication stress in NBL cells, and by activating the DDR programs allows the cells to survive the inherent stress. Hence, inhibition of TEAD4 activity in these cells offers an attractive therapeutic strategy. In addition, our results indicate that TEAD4 expression predicts patient outcome better than MYCNA status, and hence could potentially be used as a biomarker to facilitate better clinical evaluation.

Even though our results indicate that TEAD4 could potentially play an important role in NBL biology, particularly in cells with high MYC/N activity, further investigation is required to ensure the validity of TEAD4 as a potential target in patient tumors as well as to understand the mechanism in-depth, by which TEAD4 drives the subset of high-risk NBL. The first step towards ensuring its validity in patient tumors would be to test whether TEAD4 is required for initiating and/ or maintaining NBL tumors in-vivo. This could be tested on xenograft mice models where NBL cells stably infected with inducible TEAD4 shRNA would be tested for tumor growth compared to control shRNA. Another important question that could be looked into is whether knockdown of TEAD4 in tumors that are already formed would lead to regression of tumors. This can be performed by injecting the mice with NBL cells stably infected with inducible TEAD4 shRNA or control shRNA and let the tumor develop for a certain time period before inducing TEAD4 shRNA expression.

Upon validation of the role of TEAD4 in-vivo, the following step would be to assess the chemical tractability of TEAD4 activity. As with other TFs, TEAD4 being a TF, lacks the small molecule binding pockets for inhibition of its activity. Therefore, we could screen for small molecule compounds that could potentially inhibit the activity of TEAD4 protein. This could be performed by treating a panel of MYCNA

and MYCN-NA cell lines with the available drug libraries to measure the sensitivity of the drugs in MYCNA cells compared to the control cells. The compounds that are selective for MYCNA cells could further be analyzed by gene expression profiling upon drug perturbation to observe whether the compound inhibits the activity of TEAD4. This would be done by assessing the overlap of the target genes induced upon drug treatment to the ARACNe predicted targets of TEAD4 and/or the functional targets of TEAD4 retrieved by performing RNA-Seq of MYCNA cells upon knockdown of TEAD4. In order to further evaluate the effect of drug in-vivo, we could test the drugs in the patient derived xenograft (PDX) tumors, which are NBL patient primary tumor that are implanted directly into immunodeficient mice. By performing drug sensitivity test on a panel of PDX models, we could evaluate the anti-tumor activity and specificity of the compound to prioritize the appropriate compounds for clinical development. We have already initiated an effort to identify the drugs specific for the individual subtypes in collaboration with the lab of Dr. Brent Stockwell (Columbia University, NY) and Dr. John Maris (Children's hospital of Philadelphia, PA).

With regard to further probing into mechanism of action of TEAD4 in the context of NBL, many thought provoking questions have arisen, with our data serving as a preliminary starting point. Firstly, as TEAD4 seemingly controls the programs that are very similar to the ones invoked upon replication stress, it would be very interesting to follow up on this idea to find out if TEAD4 indeed plays a part in inducing replication stress in MYC/N activated cells. A simplified version of the experiment would be to induce retroviral mediated overexpression of TEAD4 in wild-type and p53 or p21 null mouse embryonic fibroblasts (MEFs) and observe karyotypic analysis on metaphase spreads to detect if any chromosomal alterations are induced. The rationale behind using p53 or p21 null cells is that NBL tumors are assumed to have defective G1/S arrest system because of inactivation of p53 by various mechanisms (Tweddle et al., 2003). We could also evaluate if overexpression of TEAD4 induces transformation of MEFs by colony forming assays to confirm its oncogenic potential. As TEAD4 seemed to regulate several genes involved in origin licensing and origin firing, we could test to evaluate if overexpression of TEAD4 could induce aberrant origin firing and show signs of aberrant proliferation.

Another question as to whether or not TEAD4 acts independently or with other coactivators such as YAP and TAZ, in the context of NBL is also of interest. Increasing number of studies in several tumor contexts have started to show the importance of Hippo signaling pathway in tumor cell proliferation, cell growth and metastasis, with TEAD family of proteins acting as an essential TF for YAP/TAZ mediated oncogenic action (Harvey et al., 2013). The experiments presented in this thesis, particularly the ones where TEAD4 was depleted were performed in cells with intact YAP/TAZ. Therefore, it would be interesting to assess whether TEAD4 mediated function is dependent or independent of YAP/TAZ. This could be done by studying whether the phenotype induced by TEAD4 is also present when YAP/TAZ are also depleted in the cells and if YAP/TAZ depletion phenocopies TEAD4.

In summary, we have elucidated the master regulators and tumor checkpoints of high-risk NBL and identified TEAD4 as a critical transcriptional node driving a subset of these tumors. The ultimate and ideal goal for NBL patients would be to find targeted therapeutics and provide personalized therapy for improved efficacy — this thesis is a small contribution towards this overarching goal — by providing understanding of the molecular circuitry of high-risk NBL, where our findings have indicated that TEAD4 could provide an oncogenic vulnerability, inhibition of the activity of which could be beneficial for a subset of NBL patients. Our findings also provide a preliminary ground to understand further – the role of TEAD4 in NBL biology.

References

Abel, F., Dalevi, D., Nethander, M., Jörnsten, R., De Preter, K., Vermeulen, J., Stallings, R., Kogner, P., Maris, J., and Nilsson, S. (2011). A 6-gene signature identifies four molecular subgroups of neuroblastoma. Cancer Cell Int. *11*, 9.

Albert, null, Jeong, null, and Barabasi, null (2000). Error and attack tolerance of complex networks. Nature *406*, 378–382.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature *403*, 503–511.

Alizadeh, A.A., Aranda, V., Bardelli, A., Blanpain, C., Bock, C., Borowski, C., Caldas, C., Califano, A., Doherty, M., Elsner, M., et al. (2015). Toward understanding and exploiting tumor heterogeneity. Nat Med *21*, 846–853.

Ambros, I.M., Hata, J., Joshi, V.V., Roald, B., Dehner, L.P., Tüchler, H., Pötschger M.Sc., U., and Shimada, H. (2002). Morphologic features of neuroblastoma (Schwannian stroma-poor tumors) in clinically favorable and unfavorable groups. Cancer *94*, 1574–1583.

Ambros, P.F., Ambros, I.M., Brodeur, G.M., Haber, M., Khan, J., Nakagawara, A., Schleiermacher, G., Speleman, F., Spitz, R., London, W.B., et al. (2009). International consensus for neuroblastoma molecular diagnostics: report from the International Neuroblastoma Risk Group (INRG) Biology Committee. Br J Cancer *100*, 1471–1482.

Arlt, M.F., Durkin, S.G., Ragland, R.L., and Glover, T.W. (2006). Common fragile sites as targets for chromosome rearrangements. DNA Repair *5*, 1126–1135.

Asgharzadeh, S., Pique-Regi, R., Sposto, R., Wang, H., Yang, Y., Shimada, H., Matthay, K., Buckley, J., Ortega, A., and Seeger, R.C. (2006). Prognostic Significance of Gene Expression Profiles of Metastatic Neuroblastomas Lacking MYCN Gene Amplification. JNCI J Natl Cancer Inst *98*, 1193–1203.

Asgharzadeh, S., Salo, J.A., Ji, L., Oberthuer, A., Fischer, M., Berthold, F., Hadjidaniel, M., Liu, C.W.-Y., Metelitsa, L.S., Pique-Regi, R., et al. (2012). Clinical significance of tumor-associated inflammatory cells in metastatic neuroblastoma. J. Clin. Oncol. *30*, 3525–3532.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. Nat Genet *25*, 25–29.

Attiyeh, E.F., London, W.B., Mossé, Y.P., Wang, Q., Winter, C., Khazi, D., McGrady, P.W., Seeger, R.C., Look, A.T., Shimada, H., et al. (2005). Chromosome 1p and 11q deletions and outcome in neuroblastoma. N. Engl. J. Med. *353*, 2243–2253.

Aytes, A., Mitrofanova, A., Lefebvre, C., Alvarez, M.J., Castillo-Martin, M., Zheng, T., Eastham, J.A., Gopalan, A., Pienta, K.J., Shen, M.M., et al. (2014). Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. Cancer Cell *25*, 638–651.

Bagchi, A., Papazoglu, C., Wu, Y., Capurso, D., Brodt, M., Francis, D., Bredel, M., Vogel, H., and Mills, A.A. (2007). CHD5 is a tumor suppressor at human 1p36. Cell *128*, 459–475.

Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009a). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108–112.

Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009b). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108–112.

Bartek, J., Mistrik, M., and Bartkova, J. (2012). Thresholds of replication stress signaling in cancer development and treatment. Nat Struct Mol Biol *19*, 5–7.

Bartkova, J., Horejsí, Z., Koed, K., Krämer, A., Tort, F., Zieger, K., Guldberg, P., Sehested, M., Nesland, J.M., Lukas, C., et al. (2005). DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. Nature *434*, 864–870.

Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. Nat Genet *37*, 382–390.

Bell, E., Premkumar, R., Carr, J., Lu, X., Lovat, P.E., Kees, U.R., Lunec, J., and Tweddle, D.A. (2006). The role of MYCN in the failure of MYCN amplified neuroblastoma cell lines to G1 arrest after DNA damage. Cell Cycle *5*, 2639–2647.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) *57*, 289–300.

Bhattacharjee, A., Richards, W.G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., et al. (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. PNAS *98*, 13790–13795.

Bisikirska, B., Bansal, M., Shen, Y., Teruya-Feldstein, J., Chaganti, R., and Califano, A. (2015). Elucidation and pharmacological targeting of novel molecular drivers of follicular lymphoma progression. Cancer Res.

Blackwell, T.K., Huang, J., Ma, A., Kretzner, L., Alt, F.W., Eisenman, R.N., and Weintraub, H. (1993). Binding of myc proteins to canonical and noncanonical DNA sequences. Mol. Cell. Biol. *13*, 5216–5224.

Blackwood, E.M., and Eisenman, R.N. (1991). Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. Science *251*, 1211–1217.

Blow, J.J., and Gillespie, P.J. (2008). Replication Licensing and Cancer - a Fatal Entanglement? Nat Rev Cancer *8*, 799–806.

Bordow, S.B., Norris, M.D., Haber, P.S., Marshall, G.M., and Haber, M. (1998). Prognostic significance of MYCN oncogene expression in childhood neuroblastoma. J. Clin. Oncol. *16*, 3286–3294.

Bourdeaut, F., Trochet, D., Janoueix-Lerosey, I., Ribeiro, A., Deville, A., Coz, C., Michiels, J.-F., Lyonnet, S., Amiel, J., and Delattre, O. (2005). Germline mutations of the paired-like homeobox 2B (PHOX2B) gene in neuroblastoma. Cancer Lett. *228*, 51–58.

Bown, N., Cotterill, S., Lastowska, M., O'Neill, S., Pearson, A.D., Plantaz, D., Meddeb, M., Danglot, G., Brinkschmidt, C., Christiansen, H., et al. (1999). Gain of chromosome arm 17q and adverse outcome in patients with neuroblastoma. N. Engl. J. Med. *340*, 1954–1961.

Breit, S., and Schwab, M. (1989). Suppression of MYC by high expression of NMYC in human neuroblastoma cells. J. Neurosci. Res. *24*, 21–28.

Breslow, N., and McCann, B. (1971). Statistical Estimation of Prognosis for Children with Neuroblastoma. Cancer Res *31*, 2098–2103.

Brichta, L., Shin, W., Jackson-Lewis, V., Blesa, J., Yap, E.-L., Walker, Z., Zhang, J., Roussarie, J.-P., Alvarez, M.J., Califano, A., et al. (2015). Identification of neurodegenerative factors using translatome-regulatory network analysis. Nat. Neurosci. *18*, 1325–1333.

Brodeur, G.M. (2003). Neuroblastoma: biological insights into a clinical enigma. Nat. Rev. Cancer *3*, 203–216.

Brodeur, G.M., and Bagatell, R. (2014). Mechanisms of neuroblastoma regression. Nat Rev Clin Oncol *11*, 704–713.

Brodeur, G.M., and Nakagawara, A. (1992). Molecular basis of clinical heterogeneity in neuroblastoma. Am J Pediatr Hematol Oncol *14*, 111–116.

Brodeur, G.M., Green, A.A., Hayes, F.A., Williams, K.J., Williams, D.L., and Tsiatis, A.A. (1981). Cytogenetic features of human neuroblastomas and cell lines. Cancer Res. *41*, 4678–4686.

Brodeur, G.M., Seeger, R.C., Schwab, M., Varmus, H.E., and Bishop, J.M. (1984). Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. Science *224*, 1121–1124.

Brodeur, G.M., Seeger, R.C., Barrett, A., Berthold, F., Castleberry, R.P., D'Angio, G., De Bernardi, B., Evans, A.E., Favrot, M., and Freeman, A.I. (1988). International criteria for diagnosis, staging, and response to treatment in patients with neuroblastoma. J. Clin. Oncol. *6*, 1874–1881.

Brodeur, G.M., Pritchard, J., Berthold, F., Carlsen, N.L., Castel, V., Castelberry, R.P., Bernardi, B.D., Evans, A.E., Favrot, M., and Hedborg, F. (1993a). Revisions of the international criteria for neuroblastoma diagnosis, staging, and response to treatment. JCO *11*, 1466–1477.

Brodeur, G.M., Pritchard, J., Berthold, F., Carlsen, N.L., Castel, V., Castelberry, R.P., De Bernardi, B., Evans, A.E., Favrot, M., and Hedborg, F. (1993b). Revisions of the international criteria for neuroblastoma diagnosis, staging, and response to treatment. J. Clin. Oncol. *11*, 1466–1477.

Brodeur, G.M., Minturn, J.E., Ho, R., Simpson, A.M., Iyer, R., Varela, C.R., Light, J.E., Kolla, V., and Evans, A.E. (2009). Trk Receptor Expression and Inhibition in Neuroblastomas. Clin Cancer Res *15*, 3244–3250.

Cahill, D.P., da Costa, L.T., Carson-Walter, E.B., Kinzler, K.W., Vogelstein, B., and Lengauer, C. (1999). Characterization of MAD2B and other mitotic spindle checkpoint genes. Genomics *58*, 181–187.

Califano, A., Butte, A.J., Friend, S., Ideker, T., and Schadt, E. (2012). Leveraging models of cell regulation and GWAS data in integrative network-based association studies. Nat. Genet. *44*, 841–847.

Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.

Capasso, M., Devoto, M., Hou, C., Asgharzadeh, S., Glessner, J.T., Attiyeh, E.F., Mosse, Y.P., Kim, C., Diskin, S.J., Cole, K.A., et al. (2009). Common variations in BARD1 influence susceptibility to high-risk neuroblastoma. Nat. Genet. *41*, 718–723.

Carén, H., Kryh, H., Nethander, M., Sjöberg, R.-M., Träger, C., Nilsson, S., Abrahamsson, J., Kogner, P., and Martinsson, T. (2010). High-risk neuroblastoma tumors with 11q-deletion display a poor prognostic, chromosome instability phenotype with later onset. PNAS *107*, 4323–4328.

Caron, H., van Sluis, P., de Kraker, J., Bökkerink, J., Egeler, M., Laureys, G., Slater, R., Westerveld, A., Voûte, P.A., and Versteeg, R. (1996). Allelic loss of chromosome 1p as a predictor of unfavorable outcome in patients with neuroblastoma. N. Engl. J. Med. *334*, 225–230.

Carro, M.S., Lim, W.K., Alvarez, M.J., Bollo, R.J., Zhao, X., Snyder, E.Y., Sulman, E.P., Anne, S.L., Doetsch, F., Colman, H., et al. (2010). The transcriptional network for mesenchymal transformation of brain tumours. Nature *463*, 318–325.

Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. Nat Biotech *30*, 413–421.

Castleberry, R.P., Pritchard, J., Ambros, P., Berthold, F., Brodeur, G.M., Castel, V., Cohn, S.L., De Bernardi, B., Dicks-Mireaux, C., Frappaz, D., et al. (1997). The International Neuroblastoma Risk Groups (INRG): a preliminary report. Eur. J. Cancer *33*, 2113–2116.

Cecchetto, G., Mosseri, V., De Bernardi, B., Helardot, P., Monclair, T., Costa, E., Horcher, E., Neuenschwander, S., Tomà, P., Rizzo, A., et al. (2005). Surgical risk factors in primary surgery for localized neuroblastoma: the LNESG1 study of the European International Society of Pediatric Oncology Neuroblastoma Group. J. Clin. Oncol. *23*, 8483–8489.

Chan, H.S., Gallie, B.L., DeBoer, G., Haddad, G., Ikegaki, N., Dimitroulakos, J., Yeger, H., and Ling, V. (1997). MYCN protein expression as a predictor of neuroblastoma prognosis. Clin. Cancer Res. *3*, 1699–1706.

Chao, M.V. (2003). Neurotrophins and their receptors: A convergence point for many signalling pathways. Nat Rev Neurosci *4*, 299–309.

Charron, J., Malynn, B.A., Fisher, P., Stewart, V., Jeannotte, L., Goff, S.P., Robertson, E.J., and Alt, F.W. (1992). Embryonic lethality in mice homozygous for a targeted disruption of the N-myc gene. Genes Dev. *6*, 2248–2257.

Cheadle, C., Vawter, M.P., Freed, W.J., and Becker, K.G. (2003). Analysis of Microarray Data Using Z Score Transformation. J Mol Diagn *5*, 73–81.

Chen, J.C., Alvarez, M.J., Talos, F., Dhruv, H., Rieckhof, G.E., Iyer, A., Diefes, K.L., Aldape, K., Berens, M., Shen, M.M., et al. (2014). Identification of causal genetic drivers of human disease through systems-level analysis of regulatory networks. Cell *159*, 402–414.

Chen, Y., Takita, J., Choi, Y.L., Kato, M., Ohira, M., Sanada, M., Wang, L., Soda, M., Kikuchi, A., Igarashi, T., et al. (2008). Oncogenic mutations of ALK kinase in neuroblastoma. Nature *455*, 971–974.

Chen, Y., Tsai, Y.-H., and Tseng, S.-H. (2013). Inhibition of cyclin-dependent kinase 1-induced cell death in neuroblastoma cells through the microRNA-34a-MYCN-survivin pathway. Surgery *153*, 4–16.

Cheung, N.-K.V., and Dyer, M.A. (2013). Neuroblastoma: developmental biology, cancer genomics and immunotherapy. Nat. Rev. Cancer *13*, 397–411.

Choi, S.H., Wright, J.B., Gerber, S.A., and Cole, M.D. (2010). Myc protein is stabilized by suppression of a novel E3 ligase complex in cancer cells. Genes Dev. *24*, 1236–1241.

Chudnovsky, Y., Kim, D., Zheng, S., Whyte, W.A., Bansal, M., Bray, M.-A., Gopal, S., Theisen, M.A., Bilodeau, S., Thiru, P., et al. (2014). ZFHX4 interacts with the NuRD core member CHD4 and regulates the glioblastoma tumor-initiating cell state. Cell Rep *6*, 313–324.

Cohn, S.L., Pearson, A.D.J., London, W.B., Monclair, T., Ambros, P.F., Brodeur, G.M., Faldum, A., Hero, B., Iehara, T., Machin, D., et al. (2009a). The International Neuroblastoma Risk Group (INRG) Classification System: An INRG Task Force Report. JCO *27*, 289–297.

Cohn, S.L., Pearson, A.D.J., London, W.B., Monclair, T., Ambros, P.F., Brodeur, G.M., Faldum, A., Hero, B., Iehara, T., Machin, D., et al. (2009b). The International Neuroblastoma Risk Group (INRG) classification system: an INRG Task Force report. J. Clin. Oncol. *27*, 289–297.

Cole, K.A., Attiyeh, E.F., Mosse, Y.P., Laquaglia, M.J., Diskin, S.J., Brodeur, G.M., and Maris, J.M. (2008). A Functional Screen Identifies miR-34a as a Candidate Neuroblastoma Tumor Suppressor Gene. Mol Cancer Res *6*, 735–742.

Cole, K.A., Huggins, J., Laquaglia, M., Hulderman, C.E., Russell, M.R., Bosse, K., Diskin, S.J., Attiyeh, E.F., Sennett, R., Norris, G., et al. (2011). RNAi screen of the protein kinome identifies checkpoint kinase 1 (CHK1) as a therapeutic target in neuroblastoma. Proc. Natl. Acad. Sci. U.S.A. *108*, 3336–3341.

Coller, H.A., Grandori, C., Tamayo, P., Colbert, T., Lander, E.S., Eisenman, R.N., and Golub, T.R. (2000). Expression analysis with oligonucleotide microarrays reveals that MYC regulates genes involved in growth, cell cycle, signaling, and adhesion. Proc. Natl. Acad. Sci. U.S.A. *97*, 3260–3265.

Compagno, M., Lim, W.K., Grunn, A., Nandula, S.V., Brahmachary, M., Shen, Q., Bertoni, F., Ponzoni, M., Scandurra, M., Califano, A., et al. (2009). Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. Nature *459*, 717–721.

Cox, D.R. (1972). Regression Models and Life-Tables. Journal of the Royal Statistical Society. Series B (Methodological) *34*, 187–220.

Cristescu, R., Lee, J., Nebozhyn, M., Kim, K.-M., Ting, J.C., Wong, S.S., Liu, J., Yue, Y.G., Wang, J., Yu, K., et al. (2015). Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med *21*, 449–456.

Croft, D., Mundo, A.F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M.R., et al. (2014). The Reactome pathway knowledgebase. Nucleic Acids Res. *42*, D472–D477.

Dalla-Favera, R., Bregni, M., Erikson, J., Patterson, D., Gallo, R.C., and Croce, C.M. (1982). Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. Proc. Natl. Acad. Sci. U.S.A. *79*, 7824–7827.

Dang, C.V. (2013). MYC, Metabolism, Cell Growth, and Tumorigenesis. Cold Spring Harb Perspect Med *3*, a014217.

Davis, A.C., Wims, M., Spotts, G.D., Hann, S.R., and Bradley, A. (1993). A null c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice. Genes Dev. *7*, 671–682.

Dent, E.W., and Gertler, F.B. (2003). Cytoskeletal Dynamics and Transport in Growth Cone Motility and Axon Guidance. Neuron *40*, 209–227.

De Preter, K., Vandesompele, J., Heimann, P., Yigit, N., Beckman, S., Schramm, A., Eggert, A., Stallings, R.L., Benoit, Y., Renard, M., et al. (2006). Human fetal neuroblast and neuroblastoma transcriptome analysis confirms neuroblast origin and highlights neuroblastoma candidate genes. Genome Biol. *7*, R84.

De Preter, K., Vermeulen, J., Brors, B., Delattre, O., Eggert, A., Fischer, M., Janoueix-Lerosey, I., Lavarino, C., Maris, J.M., Mora, J., et al. (2010). Accurate outcome prediction in neuroblastoma across independent data sets using a multigene signature. Clin. Cancer Res. *16*, 1532–1541.

Diede, S.J. (2014). Spontaneous regression of metastatic cancer: learning from neuroblastoma. Nat Rev Cancer *14*, 71–72.

Diskin, S.J., Capasso, M., Schnepp, R.W., Cole, K.A., Attiyeh, E.F., Hou, C., Diamond, M., Carpenter, E.L., Winter, C., Lee, H., et al. (2012). Common variation at 6q16 within HACE1 and LIN28B influences susceptibility to neuroblastoma. Nat. Genet. *44*, 1126–1130.

Dominguez-Sola, D., Ying, C.Y., Grandori, C., Ruggiero, L., Chen, B., Li, M., Galloway, D.A., Gu, W., Gautier, J., and Dalla-Favera, R. (2007). Non-transcriptional control of DNA replication by c-Myc. Nature *448*, 445–451.

Dowling, R.J.O., Topisirovic, I., Alain, T., Bidinosti, M., Fonseca, B.D., Petroulakis, E., Wang, X., Larsson, O., Selvaraj, A., Liu, Y., et al. (2010). mTORC1-mediated cell proliferation, but not cell growth, controlled by the 4E-BPs. Science *328*, 1172–1176.

Egan, C.M., Nyman, U., Skotte, J., Streubel, G., Turner, S., O'Connell, D.J., Rraklli, V., Dolan, M.J., Chadderton, N., Hansen, K., et al. (2013). CHD5 is required for neurogenesis and has a dual role in facilitating gene expression and polycomb gene repression. Dev. Cell *26*, 223–236.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. PNAS *95*, 14863–14868.

Emmert-Streib, F., Glazko, G.V., Altay, G., and de Matos Simoes, R. (2012). Statistical Inference and Reverse Engineering of Gene Regulatory Networks from Observational Expression Data. Front Genet *3*.

Evans, L., Chen, L., Milazzo, G., Gherardi, S., Perini, G., Willmore, E., Newell, D.R., and Tweddle, D.A. (2015). SKP2 is a direct transcriptional target of MYCN and a potential therapeutic target in neuroblastoma. Cancer Lett. *363*, 37–45.

Fagan, A.M., Zhang, H., Landis, S., Smeyne, R.J., Silos-Santiago, I., and Barbacid, M. (1996). TrkA, but not TrkC, receptors are essential for survival of sympathetic neurons in vivo. J. Neurosci. *16*, 6208–6218.

Feijoo, C., Hall-Jackson, C., Wu, R., Jenkins, D., Leitch, J., Gilbert, D.M., and Smythe, C. (2001). Activation of mammalian Chk1 during DNA replication arrest a role for Chk1 in the intra-S phase checkpoint monitoring replication origin firing. J Cell Biol *154*, 913–924.

Felsher, D.W., and Bishop, J.M. (1999). Transient excess of MYC activity can elicit genomic instability and tumorigenesis. PNAS *96*, 3940–3944.

Feng, L., Barnhart, J.R., Seeger, R.C., Wu, L., Keshelava, N., Huang, S.-H., and Jong, A. (2008). Cdc6 knockdown inhibits human neuroblastoma cell proliferation. Mol. Cell. Biochem. *311*, 189–197.

Fisher, R.A. (1922). On the Interpretation of $\chi 2$ from Contingency Tables, and the Calculation of P. Journal of the Royal Statistical Society *85*, 87–94.

Franken, N.A.P., Rodermond, H.M., Stap, J., Haveman, J., and van Bree, C. (2006). Clonogenic assay of cells in vitro. Nat Protoc *1*, 2315–2319.

Fridman, W.H., Pagès, F., Sautès-Fridman, C., and Galon, J. (2012). The immune contexture in human tumours: impact on clinical outcome. Nat Rev Cancer *12*, 298–306.

Fujita, T., Igarashi, J., Okawa, E.R., Gotoh, T., Manne, J., Kolla, V., Kim, J., Zhao, H., Pawel, B.R., London, W.B., et al. (2008). CHD5, a tumor suppressor gene deleted from 1p36.31 in neuroblastomas. J. Natl. Cancer Inst. *100*, 940–949.

Gaillard, H., García-Muse, T., and Aguilera, A. (2015). Replication stress and cancer. Nat Rev Cancer *15*, 276–289.

Gajewski, T.F., Schreiber, H., and Fu, Y.-X. (2013). Innate and adaptive immune cells in the tumor microenvironment. Nat Immunol *14*, 1014–1022.

Garcia, I., Mayol, G., Ríos, J., Domenech, G., Cheung, N.-K.V., Oberthuer, A., Fischer, M., Maris, J.M., Brodeur, G.M., Hero, B., et al. (2012). A three-gene expression signature model for risk stratification of patients with neuroblastoma. Clin. Cancer Res. *18*, 2012–2023.

Gardner, T.S., and Faith, J.J. (2005). Reverse-engineering transcription control networks. Physics of Life Reviews *2*, 65–88.

Gartel, A.L., Ye, X., Goufman, E., Shianov, P., Hay, N., Najmabadi, F., and Tyner, A.L. (2001). Myc represses the p21(WAF1/CIP1) promoter and interacts with Sp1/Sp3. Proc. Natl. Acad. Sci. U.S.A. *98*, 4510–4515.

Gehring, M., Berthold, F., Edler, L., Schwab, M., and Amler, L.C. (1995). The 1p deletion is not a reliable marker for the prognosis of patients with neuroblastoma. Cancer Res. *55*, 5366–5369.

Gerlinger, M., Rowan, A.J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., et al. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N. Engl. J. Med. *366*, 883–892.

Goga, A., Yang, D., Tward, A.D., Morgan, D.O., and Bishop, J.M. (2007). Inhibition of CDK1 as a potential therapy for tumors over-expressing MYC. Nat Med *13*, 820–827.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., et al. (1999). Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. Science *286*, 531–537.

Gorgoulis, V.G., Vassiliou, L.-V.F., Karakaidos, P., Zacharatos, P., Kotsinas, A., Liloglou, T., Venere, M., DiTullio, R.A., Kastrinakis, N.G., Levy, B., et al. (2005). Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. Nature *434*, 907–913.

Govek, E.-E., Newey, S.E., and Aelst, L.V. (2005). The role of the Rho GTPases in neuronal development. Genes Dev. *19*, 1–49.

Greenman, C., Stephens, P., Smith, R., Dalgliesh, G.L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C., et al. (2007). Patterns of somatic mutation in human cancer genomes. Nature *446*, 153–158.

Griffiths, A.J., Miller, J.H., Suzuki, D.T., Lewontin, R.C., and Gelbart, W.M. (2000). An Introduction to Genetic Analysis.

Guillemot, F., Lo, L.-C., Johnson, J.E., Auerbach, A., Anderson, D.J., and Joyner, A.L. (1993). Mammalian achaete-scute homolog 1 is required for the early development of olfactory and autonomic neurons. Cell *75*, 463–476.

Gulaya, N.M., Volkov, G.L., Klimashevsky, V.M., Govseeva, N.N., and Melnik, A.A. (1989). Changes in lipid composition of neuroblastoma C1300 N18 cell during differentiation. Neuroscience *30*, 153–164.

Guo, C., White, P.S., Weiss, M.J., Hogarty, M.D., Thompson, P.M., Stram, D.O., Gerbing, R., Matthay, K.K., Seeger, R.C., Brodeur, G.M., et al. (1999). Allelic deletion at 11q23 is common in MYCN single copy neuroblastomas. Oncogene *18*, 4948–4957.

Halazonetis, T.D., Gorgoulis, V.G., and Bartek, J. (2008a). An Oncogene-Induced DNA Damage Model for Cancer Development. Science *319*, 1352–1355.

Halazonetis, T.D., Gorgoulis, V.G., and Bartek, J. (2008b). An oncogene-induced DNA damage model for cancer development. Science *319*, 1352–1355.

Halkidi, M., Batistakis, Y., and Vazirgiannis, M. (2001). On Clustering Validation Techniques. Journal of Intelligent Information Systems *17*, 107–145.

Hanahan, D., and Weinberg, R.A. (2011a). Hallmarks of Cancer: The Next Generation. Cell *144*, 646–674.

Hanahan, D., and Weinberg, R.A. (2011b). Hallmarks of cancer: the next generation. Cell *144*, 646–674.

Hansford, L.M., Thomas, W.D., Keating, J.M., Burkhart, C.A., Peaston, A.E., Norris, M.D., Haber, M., Armati, P.J., Weiss, W.A., and Marshall, G.M. (2004). Mechanisms of embryonal tumor initiation: Distinct roles for MycN expression and MYCN amplification. PNAS *101*, 12664–12669.

Hartwell, L.H., and Kastan, M.B. (1994). Cell cycle control and cancer. Science *266*, 1821–1828.

Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W. (1999). From molecular to modular cell biology. Nature *402*, C47–C52.

Harvey, K.F., Zhang, X., and Thomas, D.M. (2013). The Hippo pathway and human cancer. Nat Rev Cancer *13*, 246–257.

Henrich, K.-O., Bauer, T., Schulte, J., Ehemann, V., Deubzer, H., Gogolin, S., Muth, D., Fischer, M., Benner, A., König, R., et al. (2011). CAMTA1, a 1p36 tumor suppressor candidate, inhibits growth and activates differentiation programs in neuroblastoma cells. Cancer Res. *71*, 3142–3151.

Heukamp, L.C., Thor, T., Schramm, A., Preter, K.D., Kumps, C., Wilde, B.D., Odersky, A., Peifer, M., Lindner, S., Spruessel, A., et al. (2012). Targeted Expression of Mutated ALK Induces Neuroblastoma in Transgenic Mice. Science Translational Medicine *4*, 141ra91–ra141ra91.

Hills, S.A., and Diffley, J.F.X. (2014). DNA Replication and Oncogene-Induced Replicative Stress. Current Biology *24*, R435–R444.

Hoehner, J.C., Gestblom, C., Hedborg, F., Sandstedt, B., Olsen, L., and Påhlman, S. (1996). A developmental model of neuroblastoma: differentiating stroma-poor tumors' progress along an extra-adrenal chromaffin lineage. Lab. Invest. *75*, 659–675.

Hoehner, J.C., Hedborg, F., Eriksson, L., Sandstedt, B., Grimelius, L., Olsen, L., and Påhlman, S. (1998). Developmental gene expression of sympathetic nervous system tumors reflects their histogenesis. Lab. Invest. *78*, 29–45.

Hogarty, M.D., White, P.S., Sulman, E.P., and Brodeur, G.M. (1998). Mononucleotide Repeat Instability Is Infrequent in Neuroblastoma. Cancer Genetics and Cytogenetics *106*, 140–143.

Home, P., Saha, B., Ray, S., Dutta, D., Gunewardena, S., Yoo, B., Pal, A., Vivian, J.L., Larson, M., Petroff, M., et al. (2012). Altered subcellular localization of transcription factor TEAD4 regulates first mammalian cell lineage commitment. Proc. Natl. Acad. Sci. U.S.A. *109*, 7362–7367.

Howard, M.J., Stanke, M., Schneider, C., Wu, X., and Rohrer, H. (2000). The transcription factor dHAND is a downstream effector of BMPs in sympathetic neuron specification. Development *127*, 4073–4081.

Huang, R., Cheung, N.-K.V., Vider, J., Cheung, I.Y., Gerald, W.L., Tickoo, S.K., Holland, E.C., and Blasberg, R.G. (2011). MYCN and MYC regulate tumor proliferation and tumorigenesis directly through BMI1 in human neuroblastomas. FASEB J *25*, 4138–4149.

Huber, K. (2006). The sympathoadrenal cell lineage: specification, diversification, and new perspectives. Dev. Biol. *298*, 335–343.

Ikiz, B., Alvarez, M.J., Ré, D.B., Le Verche, V., Politi, K., Lotti, F., Phani, S., Pradhan, R., Yu, C., Croft, G.F., et al. (2015). The Regulatory Machinery of Neurodegeneration in In Vitro Models of Amyotrophic Lateral Sclerosis. Cell Rep *12*, 335–345.

International Cancer Genome Consortium, Hudson, T.J., Anderson, W., Artez, A., Barker, A.D., Bell, C., Bernabé, R.R., Bhan, M.K., Calvo, F., Eerola, I., et al. (2010). International network of cancer genome projects. Nature *464*, 993–998.

Iraci, N., Diolaiti, D., Papa, A., Porro, A., Valli, E., Gherardi, S., Herold, S., Eilers, M., Bernardoni, R., Valle, G.D., et al. (2011). A SP1/MIZ1/MYCN Repression Complex Recruits HDAC1 at the TRKA and p75NTR Promoters and Affects Neuroblastoma Malignancy by Inhibiting the Cell Response to NGF. Cancer Res *71*, 404–412.

Islam, A., Kageyama, H., Takada, N., Kawamoto, T., Takayasu, H., Isogai, E., Ohira, M., Hashizume, K., Kobayashi, H., Kaneko, Y., et al. (2000). High expression of Survivin, mapped to 17q25, is significantly associated with poor prognostic factors and promotes cell survival in human neuroblastoma. Oncogene *19*, 617–623.

Jackson, A.L., and Linsley, P.S. (2010). Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. Nat Rev Drug Discov *9*, 57–67.

Janoueix-Lerosey, I., Novikov, E., Monteiro, M., Gruel, N., Schleiermacher, G., Loriod, B., Nguyen, C., and Delattre, O. (2004). Gene expression profiling of 1p35-36 genes in neuroblastoma. Oncogene *23*, 5912–5922.

Janoueix-Lerosey, I., Lequin, D., Brugières, L., Ribeiro, A., de Pontual, L., Combaret, V., Raynal, V., Puisieux, A., Schleiermacher, G., Pierron, G., et al. (2008a). Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma. Nature *455*, 967–970.

Janoueix-Lerosey, I., Lequin, D., Brugières, L., Ribeiro, A., de Pontual, L., Combaret, V., Raynal, V., Puisieux, A., Schleiermacher, G., Pierron, G., et al. (2008b). Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma. Nature *455*, 967–970.

Jones, S., Zhang, X., Parsons, D.W., Lin, J.C.-H., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A., et al. (2008). Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses. Science *321*, 1801–1806.

Joshi, V.V., Cantor, A.B., Altshuler, G., Larkin, E.W., Neill, J.S., Shuster, J.J., Holbrook, C.T., Hayes, F.A., Nitschke, R., and Duncan, M.H. (1992). Age-linked prognostic categorization based on a new histologic grading system of neuroblastomas. A clinicopathologic study of 211 cases from the Pediatric Oncology Group. Cancer *69*, 2197–2211.

Joshi, V.V., Rao, P.V., Cantor, A.B., Altshuler, G., Shuster, J.J., and Castleberry, R.P. (1996). Modified histologic grading of neuroblastomas by replacement of mitotic rate with mitosis karyorrhexis index. A clinicopathologic study of 223 cases from the Pediatric Oncology Group. Cancer *77*, 1582–1588.

Kaghad, M., Bonnet, H., Yang, A., Creancier, L., Biscan, J.C., Valent, A., Minty, A., Chalon, P., Lelias, J.M., Dumont, X., et al. (1997). Monoallelically expressed gene related to p53 at 1p36, a region frequently deleted in neuroblastoma and other human cancers. Cell *90*, 809–819.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30.

Kaneko, K.J., and DePamphilis, M.L. (1998). Regulation of gene expression at the beginning of mammalian development and the TEAD family of transcription factors. Dev. Genet. *22*, 43–55.

Kaneko, Y., and Knudson, A.G. (2000). Mechanism and relevance of ploidy in neuroblastoma. Genes, Chromosomes and Cancer *29*.

Kang, J.-H., Rychahou, P.G., Ishola, T.A., Qiao, J., Evers, B.M., and Chung, D.H. (2006). MYCN silencing induces differentiation and apoptosis in human neuroblastoma cells. Biochem. Biophys. Res. Commun. *351*, 192–197.

Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C., Silverman, R., and Wu, A.Y. (2000). An Efficient k-Means Clustering Algorithm: Analysis and Implementation.

Kaplan, D.R., and Miller, F.D. (2000). Neurotrophin signal transduction in the nervous system. Curr. Opin. Neurobiol. *10*, 381–391.

Keim, D.R., Hailat, N., Kuick, R., Reynolds, C.P., Brodeur, G.M., Seeger, R.C., and Hanash, S.M. (1993). PCNA levels in neuroblastoma are increased in tumors with an amplified N-myc gene and in metastatic stage tumors. Clin. Exp. Metastasis *11*, 83–90.

Kelly, T.J., and Brown, G.W. (2000). Regulation of Chromosome Replication. Annual Review of Biochemistry *69*, 829–880.

Klein, R., Smeyne, R.J., Wurst, W., Long, L.K., Auerbach, B.A., Joyner, A.L., and Barbacid, M. (1993). Targeted disruption of the trkB neurotrophin receptor gene results in nervous system lesions and neonatal death. Cell *75*, 113–122.

Kohl, N.E., Kanda, N., Schreck, R.R., Bruns, G., Latt, S.A., Gilbert, F., and Alt, F.W. (1983). Transposition and amplification of oncogene-related sequences in human neuroblastomas. Cell *35*, 359–367.

Kretzner, L., Blackwood, E.M., and Eisenman, R.N. (1992). Myc and Max proteins possess distinct transcriptional activities. Nature *359*, 426–429.

Kushwaha, R., Jagadish, N., Kustagi, M., Tomishima, M.J., Mendiratta, G., Bansal, M., Kim, H.R., Sumazin, P., Alvarez, M.J., Lefebvre, C., et al. (2015). Interrogation of a context-specific transcription factor network identifies novel regulators of pluripotency. Stem Cells *33*, 367–377.

Lamers, F., van der Ploeg, I., Schild, L., Ebus, M.E., Koster, J., Hansen, B.R., Koch, T., Versteeg, R., Caron, H.N., and Molenaar, J.J. (2011). Knockdown of survivin (BIRC5) causes apoptosis in neuroblastoma via mitotic catastrophe. Endocr. Relat. Cancer *18*, 657–668.

Land, H., Parada, L.F., and Weinberg, R.A. (1983). Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. Nature *304*, 596–602.

Lapointe, J., Li, C., Higgins, J.P., van de Rijn, M., Bair, E., Montgomery, K., Ferrari, M., Egevad, L., Rayford, W., Bergerheim, U., et al. (2004). Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proc. Natl. Acad. Sci. U.S.A. *101*, 811–816.

Łastowska, M., Cotterill, S., Bown, N., Cullinane, C., Variend, S., Lunec, J., Strachan, T., Pearson, A.D.J., and Jackson, M.S. (2002). Breakpoint position on 17q identifies the most aggressive neuroblastoma tumors. Genes Chromosomes Cancer *34*, 428–436.

Łastowska, M., Viprey, V., Santibanez-Koref, M., Wappler, I., Peters, H., Cullinane, C., Roberts, P., Hall, A.G., Tweddle, D.A., Pearson, A.D.J., et al. (2007). Identification of candidate genes involved in neuroblastoma progression by combining genomic and expression microarrays with survival data. Oncogene *26*, 7432–7444.

Lavasani, M.A., and Moinfar, F. (2012). Molecular classification of breast carcinomas with particular emphasis on “basal-like” carcinoma: A critical review. J. Biophoton. *5*, 345–366.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biology *15*, R29.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218.

Lefebvre, C., Rajbhandari, P., Alvarez, M.J., Bandaru, P., Lim, W.K., Sato, M., Wang, K., Sumazin, P., Kustagi, M., Bisikirska, B.C., et al. (2010). A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. Mol. Syst. Biol. *6*, 377.

Lim, B., Park, J.-L., Kim, H.-J., Park, Y.-K., Kim, J.-H., Sohn, H.A., Noh, S.-M., Song, K.-S., Kim, W.-H., Kim, Y.S., et al. (2013). Integrative genomics analysis reveals the multilevel dysregulation and oncogenic characteristics of TEAD4 in gastric cancer. Carcinogenesis bgt409.

Lim, K.C., Lakshmanan, G., Crawford, S.E., Gu, Y., Grosveld, F., and Engel, J.D. (2000). Gata3 loss leads to embryonic lethality due to noradrenaline deficiency of the sympathetic nervous system. Nat. Genet. *25*, 209–212.

Liu, Y., Wang, G., Yang, Y., Mei, Z., Liang, Z., Cui, A., Wu, T., Liu, C.-Y., and Cui, L. (2015). Increased TEAD4 expression and nuclear localization in colorectal cancer promote epithelial–mesenchymal transition and metastasis in a YAP-independent manner. Oncogene.

London, W.B., Castleberry, R.P., Matthay, K.K., Look, A.T., Seeger, R.C., Shimada, H., Thorner, P., Brodeur, G., Maris, J.M., Reynolds, C.P., et al. (2005). Evidence for an age cutoff greater than 365 days for neuroblastoma risk group stratification in the Children’s Oncology Group. J. Clin. Oncol. *23*, 6459–6465.

Look, A.T., Hayes, F.A., Nitschke, R., McWilliams, N.B., and Green, A.A. (1984). Cellular DNA content as a predictor of response to chemotherapy in infants with unresectable neuroblastoma. N. Engl. J. Med. *311*, 231–235.

Look, A.T., Hayes, F.A., Shuster, J.J., Douglass, E.C., Castleberry, R.P., Bowman, L.C., Smith, E.I., and Brodeur, G.M. (1991). Clinical relevance of tumor cell ploidy and N-myc gene amplification in childhood neuroblastoma: a Pediatric Oncology Group study. J. Clin. Oncol. *9*, 581–591.

Lowndes, N.F., and Murguia, J.R. (2000). Sensing and responding to DNA damage. Curr. Opin. Genet. Dev. *10*, 17–25.

Luo, J., Solimini, N.L., and Elledge, S.J. (2009a). Principles of Cancer Therapy: Oncogene and Non-oncogene Addiction. Cell *136*, 823–837.

Luo, J., Solimini, N.L., and Elledge, S.J. (2009b). Principles of Cancer Therapy: Oncogene and Non-oncogene Addiction. Cell *136*, 823–837.

Mahoney, W.M., Hong, J.-H., Yaffe, M.B., and Farrance, I.K.G. (2005). The transcriptional co-activator TAZ interacts differentially with transcriptional enhancer factor-1 (TEF-1) family members. Biochem J *388*, 217–225.

Malumbres, M., and Barbacid, M. (2009). Cell cycle, CDKs and cancer: a changing paradigm. Nat Rev Cancer *9*, 153–166.

Malynn, B.A., de Alboran, I.M., O'Hagan, R.C., Bronson, R., Davidson, L., DePinho, R.A., and Alt, F.W. (2000). N-myc can functionally replace c-myc in murine development, cellular growth, and differentiation. Genes Dev. *14*, 1390–1399.

Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and Califano, A. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics *7 Suppl 1*, S7.

Maris, J.M. (2010). Recent Advances in Neuroblastoma. New England Journal of Medicine *362*, 2202–2211.

Maris, J.M., and Matthay, K.K. (1999a). Molecular Biology of Neuroblastoma. JCO *17*, 2264–2264.

Maris, J.M., and Matthay, K.K. (1999b). Molecular biology of neuroblastoma. J. Clin. Oncol. *17*, 2264–2279.

Maris, J.M., Weiss, M.J., Guo, C., Gerbing, R.B., Stram, D.O., White, P.S., Hogarty, M.D., Sulman, E.P., Thompson, P.M., Lukens, J.N., et al. (2000). Loss of heterozygosity at 1p36 independently predicts for disease progression but not decreased overall survival probability in neuroblastoma patients: a Children's Cancer Group study. J. Clin. Oncol. *18*, 1888–1899.

Maris, J.M., Guo, C., Blake, D., White, P.S., Hogarty, M.D., Thompson, P.M., Rajalingam, V., Gerbing, R., Stram, D.O., Matthay, K.K., et al. (2001). Comprehensive analysis of chromosome 1p deletions in neuroblastoma. Med. Pediatr. Oncol. *36*, 32–36.

Marshall, G.M., Carter, D.R., Cheung, B.B., Liu, T., Mateos, M.K., Meyerowitz, J.G., and Weiss, W.A. (2014). The prenatal origins of cancer. Nature Reviews Cancer *14*, 277–289.

Martinsson, T., Sjöberg, R.-M., Hedborg, F., and Kogner, P. (1995). Deletion of Chromosome 1p Loci and Microsatellite Instability in Neuroblastomas Analyzed with Short-Tandem Repeat Polymorphisms. Cancer Res *55*, 5681–5686.

Metelitsa, L.S., Wu, H.-W., Wang, H., Yang, Y., Warsi, Z., Asgharzadeh, S., Groshen, S., Wilson, S.B., and Seeger, R.C. (2004). Natural killer T cells infiltrate neuroblastomas expressing the chemokine CCL2. J. Exp. Med. *199*, 1213–1221.

Meyer, N., and Penn, L.Z. (2008). Reflecting on 25 years with MYC. Nat Rev Cancer *8*, 976–990.

Michels, E., Vandesompele, J., De Preter, K., Hoebeeck, J., Vermeulen, J., Schramm, A., Molenaar, J.J., Menten, B., Marques, B., Stallings, R.L., et al. (2007). ArrayCGH-based classification of neuroblastoma into genomic subgroups. Genes Chromosomes Cancer *46*, 1098–1108.

Mohr, S., Bakal, C., and Perrimon, N. (2010). Genomic screening with RNAi: results and challenges. Annu. Rev. Biochem. *79*, 37–64.

Mohr, S.E., Smith, J.A., Shamu, C.E., Neumüller, R.A., and Perrimon, N. (2014). RNAi screening comes of age: improved techniques and complementary approaches. Nat. Rev. Mol. Cell Biol. *15*, 591–600.

Molenaar, J.J., Ebus, M.E., Koster, J., van Sluis, P., van Noesel, C.J.M., Versteeg, R., and Caron, H.N. (2008). Cyclin D1 and CDK4 activity contribute to the undifferentiated phenotype in neuroblastoma. Cancer Res. *68*, 2599–2609.

Molenaar, J.J., Ebus, M.E., Geerts, D., Koster, J., Lamers, F., Valentijn, L.J., Westerhout, E.M., Versteeg, R., and Caron, H.N. (2009). Inactivation of CDK2 is synthetically lethal to MYCN over-expressing cancer cells. Proc. Natl. Acad. Sci. U.S.A. *106*, 12968–12973.

Molenaar, J.J., Koster, J., Zwijnenburg, D.A., van Sluis, P., Valentijn, L.J., van der Ploeg, I., Hamdi, M., van Nes, J., Westerman, B.A., van Arkel, J., et al. (2012). Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. Nature *483*, 589–593.

Monclair, T., Brodeur, G.M., Ambros, P.F., Brisse, H.J., Cecchetto, G., Holmes, K., Kaneko, M., London, W.B., Matthay, K.K., Nuchtern, J.G., et al. (2009). The International Neuroblastoma Risk Group (INRG) staging system: an INRG Task Force report. J. Clin. Oncol. *27*, 298–303.

Monti, S., Tamayo, P., Mesirov, J., and Golub, T. (2003). Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. Machine Learning *52*, 91–118.

Monti, S., Savage, K.J., Kutok, J.L., Feuerhake, F., Kurtin, P., Mihm, M., Wu, B., Pasqualucci, L., Neuberg, D., Aguiar, R.C.T., et al. (2005). Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. Blood *105*, 1851–1861.

Mosse, Y.P., Laudenslager, M., Khazi, D., Carlisle, A.J., Winter, C.L., Rappaport, E., and Maris, J.M. (2004). Germline PHOX2B mutation in hereditary neuroblastoma. Am. J. Hum. Genet. *75*, 727–730.

Mossé, Y.P., Laudenslager, M., Longo, L., Cole, K.A., Wood, A., Attiyeh, E.F., Laquaglia, M.J., Sennett, R., Lynch, J.E., Perri, P., et al. (2008). Identification of ALK as a major familial neuroblastoma predisposition gene. Nature *455*, 930–935.

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., et al. (2008). Identification of ALK as a major familial neuroblastoma predisposition gene. Nature *455*, 930–U22.

Murga, M., Campaner, S., Lopez-Contreras, A.J., Toledo, L.I., Soria, R., Montaña, M.F., D'Artista, L., Schleker, T., Guerra, C., Garcia, E., et al. (2011). Exploiting oncogene-induced replicative stress for the selective killing of Myc-driven tumors. Nat Struct Mol Biol *18*, 1331–1335.

Muth, D., Ghazaryan, S., Eckerle, I., Beckett, E., Pöhler, C., Batzler, J., Beisel, C., Gogolin, S., Fischer, M., Henrich, K.-O., et al. (2010). Transcriptional repression of SKP2 is impaired in MYCN-amplified neuroblastoma. Cancer Res. *70*, 3791–3802.

Negrini, S., Gorgoulis, V.G., and Halazonetis, T.D. (2010). Genomic instability--an evolving hallmark of cancer. Nat. Rev. Mol. Cell Biol. *11*, 220–228.

Nilsson, J.A., and Cleveland, J.L. (2003). Myc pathways provoking cell suicide and cancer. Oncogene *22*, 9007–9021.

Nolan-Stevaux, O., Tedesco, D., Ragan, S., Makhanov, M., Chenchik, A., Ruefli-Brasse, A., Quon, K., and Kassner, P.D. (2013). Measurement of Cancer Cell Growth Heterogeneity through Lentiviral Barcoding Identifies Clonal Dominance as a Characteristic of In Vivo Tumor Engraftment. PLoS ONE *8*, e67316.

Oberthuer, A., Berthold, F., Warnat, P., Hero, B., Kahlert, Y., Spitz, R., Ernestus, K., König, R., Haas, S., Eils, R., et al. (2006). Customized oligonucleotide microarray gene expression-based classification of neuroblastoma patients outperforms current clinical risk stratification. J. Clin. Oncol. *24*, 5070–5078.

Oberthuer, A., Juraeva, D., Hero, B., Volland, R., Sterz, C., Schmidt, R., Faldum, A., Kahlert, Y., Engesser, A., Asgharzadeh, S., et al. (2015). Revised Risk Estimation and Treatment Stratification of Low- and Intermediate-Risk Neuroblastoma Patients by Integrating Clinical and Molecular Prognostic Markers. Clin Cancer Res *21*, 1904–1915.

Ohira, M., Oba, S., Nakamura, Y., Isogai, E., Kaneko, S., Nakagawa, A., Hirata, T., Kubo, H., Goto, T., Yamada, S., et al. (2005). Expression profiling using a tumor-specific cDNA microarray predicts the prognosis of intermediate risk neuroblastomas. Cancer Cell *7*, 337–350.

Otto, T., Horn, S., Brockmann, M., Eilers, U., Schüttrumpf, L., Popov, N., Kenney, A.M., Schulte, J.H., Beijersbergen, R., Christiansen, H., et al. (2009). Stabilization of N-Myc is a critical function of Aurora A in human neuroblastoma. Cancer Cell *15*, 67–78.

Pagès, F., Galon, J., Dieu-Nosjean, M.-C., Tartour, E., Sautès-Fridman, C., and Fridman, W.-H. (2009). Immune infiltration in human tumors: a prognostic factor that should not be ignored. Oncogene *29*, 1093–1102.

Park, J.R., Eggert, A., and Caron, H. (2008). Neuroblastoma: Biology, Prognosis, and Treatment. Pediatric Clinics of North America *55*, 97–120.

Parsons, D.W., Jones, S., Zhang, X., Lin, J.C.-H., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Siu, I.-M., Gallia, G.L., et al. (2008). An Integrated Genomic Analysis of Human Glioblastoma Multiforme. Science *321*, 1807.

Pattyn, A., Morin, X., Cremer, H., Goridis, C., and Brunet, J.-F. (1999). The homeobox gene Phox2b is essential for the development of autonomic neural crest derivatives. Nature *399*, 366–370.

Phillips, H.S., Kharbanda, S., Chen, R., Forrest, W.F., Soriano, R.H., Wu, T.D., Misra, A., Nigro, J.M., Colman, H., Soroceanu, L., et al. (2006). Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. Cancer Cell *9*, 157–173.

Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y., et al. (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat. Genet. *20*, 207–211.

Piovan, E., Yu, J., Tosello, V., Herranz, D., Ambesi-Impiombato, A., Da Silva, A.C., Sanchez-Martin, M., Perez-Garcia, A., Rigo, I., Castillo, M., et al. (2013). Direct reversal of glucocorticoid resistance by AKT inhibition in acute lymphoblastic leukemia. Cancer Cell *24*, 766–776.

Plantaz, D., Mohapatra, G., Matthay, K.K., Pellarin, M., Seeger, R.C., and Feuerstein, B.G. (1997). Gain of chromosome 17 is the most frequent abnormality detected in neuroblastoma by comparative genomic hybridization. Am. J. Pathol. *150*, 81–89.

Prins, R.M., Soto, H., Konkankit, V., Odesa, S.K., Eskin, A., Yong, W.H., Nelson, S.F., and Liau, L.M. (2011). Gene expression profile correlates with T-cell infiltration and relative survival in glioblastoma patients vaccinated with dendritic cell immunotherapy. Clin. Cancer Res. *17*, 1603–1615.

Pritchard, J., and Hickman, J.A. (1994). Why does stage 4s neuroblastoma regress spontaneously? The Lancet *344*, 869–870.

Pugh, T.J., Morozova, O., Attiyeh, E.F., Asgharzadeh, S., Wei, J.S., Auclair, D., Carter, S.L., Cibulskis, K., Hanna, M., Kiezun, A., et al. (2013). The genetic landscape of high-risk neuroblastoma. Nat. Genet. *45*, 279–284.

Qiao, C., Jiang, Y., Deng, C., Huang, Z., Teng, K., Chen, L., and Liu, X. (2015). Characterization of the transcriptional activation domains of human TEF3-1 (transcription enhancer factor 3 isoform 1). Archives of Biochemistry and Biophysics *569*, 54–61.

Reiff, T., Huber, L., Kramer, M., Delattre, O., Janoueix-Lerosey, I., and Rohrer, H. (2011). Midkine and Alk signaling in sympathetic neuron proliferation and neuroblastoma predisposition. Development *138*, 4699–4708.

Reiter, J.L., and Brodeur, G.M. (1996). High-resolution mapping of a 130-kb core region of the MYCN amplicon in neuroblastomas. Genomics *32*, 97–103.

Remus, D., and Diffley, J.F.X. (2009). Eukaryotic DNA replication control: lock and load, then fire. Curr. Opin. Cell Biol. *21*, 771–777.

Repunte-Canonigo, V., Shin, W., Vendruscolo, L.F., Lefebvre, C., van der Stap, L., Kawamura, T., Schlosburg, J.E., Alvarez, M., Koob, G.F., Califano, A., et al. (2015). Identifying candidate drivers of alcohol dependence-induced excessive drinking by assembly and interrogation of brain-specific regulatory networks. Genome Biol. *16*, 68.

de Ridder, D., van der Linden, C.E., Schonewille, T., Dik, W.A., Reinders, M.J.T., van Dongen, J.J.M., and Staal, F.J.T. (2005). Purity for clarity: the need for purification of tumor cells in DNA microarray studies. Leukemia *19*, 618–627.

van Riggelen, J., Yetil, A., and Felsher, D.W. (2010a). MYC as a regulator of ribosome biogenesis and protein synthesis. Nat Rev Cancer *10*, 301–309.

van Riggelen, J., Yetil, A., and Felsher, D.W. (2010b). MYC as a regulator of ribosome biogenesis and protein synthesis. Nat Rev Cancer *10*, 301–309.

Rodriguez-Barrueco, R., Yu, J., Saucedo-Cuevas, L.P., Olivan, M., Llobet-Navas, D., Putcha, P., Castro, V., Murga-Penas, E.M., Collazo-Lorduy, A., Castillo-Martin, M., et al. (2015). Inhibition of the autocrine IL-6-JAK2-STAT3-calprotectin axis as targeted therapy for HR-/HER2+ breast cancers. Genes Dev. *29*, 1631–1648.

Russell, M.R., Levin, K., Rader, J., Belcastro, L., Li, Y., Martinez, D., Pawel, B., Shumway, S.D., Maris, J.M., and Cole, K.A. (2013). Combination therapy targeting the Chk1 and Wee1 kinases shows therapeutic efficacy in neuroblastoma. Cancer Res. *73*, 776–784.

Saito-Ohara, F., Imoto, I., Inoue, J., Hosoi, H., Nakagawara, A., Sugimoto, T., and Inazawa, J. (2003). PPM1D is a potential target for 17q gain in neuroblastoma. Cancer Res. *63*, 1876–1883.

Sarkanen, J.-R., Nykky, J., Siikanen, J., Selinummi, J., Ylikomi, T., and Jalonen, T.O. (2007). Cholesterol supports the retinoic acid-induced synaptic vesicle formation in differentiating human SH-SY5Y neuroblastoma cells. J. Neurochem. *102*, 1941–1952.

Sato, Y., Sasaki, H., Kondo, S., Fukai, I., Kiriyama, M., Yamakawa, Y., and Fujii, Y. (2001). Expression of the cdc25B mRNA correlated with that of N-myc in neuroblastoma. Jpn. J. Clin. Oncol. *31*, 428–431.

Schleiermacher, G., Mosseri, V., London, W.B., Maris, J.M., Brodeur, G.M., Attiyeh, E., Haber, M., Khan, J., Nakagawara, A., Speleman, F., et al. (2012). Segmental chromosomal alterations have prognostic impact in neuroblastoma: a report from the INRG project. Br J Cancer *107*, 1418–1422.

Schulte, J.H., Lindner, S., Bohrer, A., Maurer, J., De Preter, K., Lefever, S., Heukamp, L., Schulte, S., Molenaar, J., Versteeg, R., et al. (2013). MYCN and ALKF1174L are sufficient to drive neuroblastoma development from neural crest progenitor cells. Oncogene *32*, 1059–1065.

Schwab, M. (2004). MYCN in neuronal tumours. Cancer Lett. *204*, 179–187.

Schwab, M., Alitalo, K., Klempnauer, K.-H., Varmus, H.E., Bishop, J.M., Gilbert, F., Brodeur, G., Goldstein, M., and Trent, J. (1983). Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. Nature *305*, 245–248.

Schwab, M., Varmus, H.E., Bishop, J.M., Grzeschik, K.H., Naylor, S.L., Sakaguchi, A.Y., Brodeur, G., and Trent, J. (1984). Chromosome localization in normal human cells and neuroblastomas of a gene related to c-myc. Nature *308*, 288–291.

Schwab, M., Varmus, H.E., and Bishop, J.M. (1985). Human N-myc gene contributes to neoplastic transformation of mammalian cells in culture. Nature *316*, 160–162.

Sears, R., Nuckolls, F., Haura, E., Taya, Y., Tamai, K., and Nevins, J.R. (2000). Multiple Ras-dependent phosphorylation pathways regulate Myc protein stability. Genes Dev *14*, 2501–2514.

Seeger, R.C., Brodeur, G.M., Sather, H., Dalton, A., Siegel, S.E., Wong, K.Y., and Hammond, D. (1985a). Association of multiple copies of the N-myc oncogene with rapid progression of neuroblastomas. N. Engl. J. Med. *313*, 1111–1116.

Seeger, R.C., Brodeur, G.M., Sather, H., Dalton, A., Siegel, S.E., Wong, K.Y., and Hammond, D. (1985b). Association of multiple copies of the N-myc oncogene with rapid progression of neuroblastomas. N. Engl. J. Med. *313*, 1111–1116.

Seth, A., Alvarez, E., Gupta, S., and Davis, R.J. (1991). A phosphorylation site located in the NH2-terminal domain of c-Myc increases transactivation of gene expression. J. Biol. Chem. *266*, 23521–23524.

Shimada, H., Chatten, J., Newton, W.A., Sachs, N., Hamoudi, A.B., Chiba, T., Marsden, H.B., and Misugi, K. (1984). Histopathologic prognostic factors in neuroblastic tumors: definition of subtypes of ganglioneuroblastoma and an age-linked classification of neuroblastomas. J. Natl. Cancer Inst. *73*, 405–416.

Shimada, H., Ambros, I.M., Dehner, L.P., Hata, J., Joshi, V.V., Roald, B., Stram, D.O., Gerbing, R.B., Lukens, J.N., Matthay, K.K., et al. (1999). The International Neuroblastoma Pathology Classification (the Shimada system). Cancer *86*, 364–372.

Shimada, H., Umehara, S., Monobe, Y., Hachitanda, Y., Nakagawa, A., Goto, S., Gerbing, R.B., Stram, D.O., Lukens, J.N., and Matthay, K.K. (2001). International neuroblastoma pathology classification for prognostic evaluation of patients with peripheral neuroblastic tumors: a report from the Children's Cancer Group. Cancer *92*, 2451–2461.

Shohet, J.M., Hicks, M.J., Plon, S.E., Burlingame, S.M., Stuart, S., Chen, S.-Y., Brenner, M.K., and Nuchtern, J.G. (2002). Minichromosome maintenance protein MCM7 is a direct target of the MYCN transcription factor in neuroblastoma. Cancer Res. *62*, 1123–1128.

da Silva, J.S., and Dotti, C.G. (2002). Breaking the neuronal sphere: regulation of the actin cytoskeleton in neuritogenesis. Nat. Rev. Neurosci. *3*, 694–704.

Simon, T., Hero, B., Benz-Bohm, G., von Schweinitz, D., and Berthold, F. (2008). Review of image defined risk factors in localized neuroblastoma patients: Results of the GPOH NB97 trial. Pediatr Blood Cancer *50*, 965–969.

Sjostrom, S.K., Finn, G., Hahn, W.C., Rowitch, D.H., and Kenney, A.M. (2005). The Cdk1 complex plays a prime role in regulating N-myc phosphorylation and turnover in neural precursors. Dev. Cell *9*, 327–338.

Smyth, G.K. (2005). limma: Linear Models for Microarray Data. In Bioinformatics and Computational Biology Solutions Using R and Bioconductor, R. Gentleman, V.J. Carey, W. Huber, R.A. Irizarry, and S. Dudoit, eds. (Springer New York), pp. 397–420.

Song, L., Ara, T., Wu, H.-W., Woo, C.-W., Reynolds, C.P., Seeger, R.C., DeClerck, Y.A., Thiele, C.J., Sposto, R., and Metelitsa, L.S. (2007). Oncogene MYCN regulates localization of NKT cells to the site of disease in neuroblastoma. J. Clin. Invest. *117*, 2702–2712.

Spitz, R., Oberthuer, A., Zapatka, M., Brors, B., Hero, B., Ernestus, K., Oestreich, J., Fischer, M., Simon, T., and Berthold, F. (2006). Oligonucleotide array-based comparative genomic hybridization (aCGH) of 90 neuroblastomas reveals aberration patterns closely associated with relapse pattern and outcome. Genes Chromosomes Cancer *45*, 1130–1142.

Stanke, M., Junghans, D., Geissen, M., Goridis, C., Ernsberger, U., and Rohrer, H. (1999). The Phox2 homeodomain proteins are sufficient to promote the development of sympathetic neurons. Development *126*, 4087–4094.

Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A., et al. (2011). Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. Cell *144*, 27–40.

Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009). The cancer genome. Nature *458*, 719–724.

Stricker, T.P., La Madrid, A.M., Chlenski, A., Guerrero, L., Salwen, H.R., Gosiengfiao, Y., Perlman, E.J., Furman, W., Bahrami, A., Shohet, J.M., et al. (2014). Validation of a Prognostic Multi-gene Signature in High-risk Neuroblastoma using the High Throughput Digital NanoString nCounter™ System. Mol Oncol *8*, 669–678.

Strieder, V., and Lutz, W. (2003). E2F proteins regulate MYCN expression in neuroblastomas. J. Biol. Chem. *278*, 2983–2989.

Su, T.T., and O'Farrell, P.H. (1998). Size control: Cell proliferation does not equal growth. Current Biology *8*, R687–R689.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. PNAS *102*, 15545–15550.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S., and Golub, T.R. (1999). Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. PNAS *96*, 2907–2912.

Tamborero, D., Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Kandoth, C., Reimand, J., Lawrence, M.S., Getz, G., Bader, G.D., Ding, L., et al. (2013). Comprehensive identification of mutational cancer driver genes across 12 tumor types. Scientific Reports *3*, 2650.

Thompson, P.M., Gotoh, T., Kok, M., White, P.S., and Brodeur, G.M. (2003). CHD5, a new member of the chromodomain gene family, is preferentially expressed in the nervous system. Oncogene *22*, 1002–1011.

Toledo, L.I., Murga, M., and Fernandez-Capetillo, O. (2011). Targeting ATR and Chk1 kinases for cancer treatment: A new model for new (and old) drugs. Molecular Oncology *5*, 368–373.

Torres, E.M., Sokolsky, T., Tucker, C.M., Chan, L.Y., Boselli, M., Dunham, M.J., and Amon, A. (2007). Effects of Aneuploidy on Cellular Physiology and Cell Division in Haploid Yeast. Science *317*, 916–924.

Tothill, R.W., Tinker, A.V., George, J., Brown, R., Fox, S.B., Lade, S., Johnson, D.S., Trivett, M.K., Etemadmoghadam, D., Locandro, B., et al. (2008). Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. Clin. Cancer Res. *14*, 5198–5208.

Tsai, H.-Y., Hsi, B.-L., Hung, I.-J., Yang, C.-P., Lin, J.-N., Chen, J.-C., Tsai, S.-F., and Huang, S.-F. (2004). Correlation of MYCN amplification with MCM7 protein expression in neuroblastomas: a chromogenic in situ hybridization study in paraffin sections. Hum. Pathol. *35*, 1397–1403.

Tweddle, D.A., Pearson, A.D.J., Haber, M., Norris, M.D., Xue, C., Flemming, C., and Lunec, J. (2003). The p53 pathway and its inactivation in neuroblastoma. Cancer Lett. *197*, 93–98.

Vafa, O., Wade, M., Kern, S., Beeche, M., Pandita, T.K., Hampton, G.M., and Wahl, G.M. (2002). c-Myc Can Induce DNA Damage, Increase Reactive Oxygen Species, and Mitigate p53 Function: A Mechanism for Oncogene-Induced Genetic Instability. Molecular Cell *9*, 1031–1044.

Valentijn, L.J., Koster, J., Haneveld, F., Aissa, R.A., Sluis, P. van, Broekmans, M.E.C., Molenaar, J.J., Nes, J. van, and Versteeg, R. (2012). Functional MYCN signature predicts outcome of neuroblastoma irrespective of MYCN amplification. PNAS *109*, 19190–19195.

Valentijn, L.J., Koster, J., Zwijnenburg, D.A., Hasselt, N.E., van Sluis, P., Volckmann, R., van Noesel, M.M., George, R.E., Tytgat, G.A.M., Molenaar, J.J., et al. (2015). TERT rearrangements are frequent in neuroblastoma and identify aggressive tumors. Nat. Genet.

Van Maerken, T., Vandesompele, J., Rihani, A., De Paepe, A., and Speleman, F. (2009). Escape from p53-mediated tumor surveillance in neuroblastoma: switching off the p14ARF-MDM2-p53 axis. Cell Death Differ *16*, 1563–1572.

Van Roy, N., Laureys, G., Cheng, N.C., Willem, P., Opdenakker, G., Versteeg, R., and Speleman, F. (1994). 1;17 translocations and other chromosome 17 rearrangements in human primary neuroblastoma tumors and cell lines. Genes Chromosomes Cancer *10*, 103–114.

Vassilev, A., Kaneko, K.J., Shu, H., Zhao, Y., and DePamphilis, M.L. (2001a). TEAD/TEF transcription factors utilize the activation domain of YAP65, a Src/Yes-associated protein localized in the cytoplasm. Genes Dev. *15*, 1229–1241.

Vassilev, A., Kaneko, K.J., Shu, H., Zhao, Y., and DePamphilis, M.L. (2001b). TEAD/TEF transcription factors utilize the activation domain of YAP65, a Src/Yes-associated protein localized in the cytoplasm. Genes Dev. *15*, 1229–1241.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A.M., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530–536.

Venet, D., Dumont, J.E., and Detours, V. (2011). Most random gene expression signatures are significantly associated with breast cancer outcome. PLoS Comput. Biol. *7*, e1002240.

Vitale, I., Galluzzi, L., Castedo, M., and Kroemer, G. (2011). Mitotic catastrophe: a mechanism for avoiding genomic instability. Nat Rev Mol Cell Biol *12*, 385–392.

Wang, C., Nie, Z., Zhou, Z., Zhang, H., Liu, R., Wu, J., Qin, J., Ma, Y., Chen, L., Li, S., et al. (2015a). The interplay between TEAD4 and KLF5 promotes breast cancer partially through inhibiting the transcription of p27Kip1. Oncotarget *6*, 17685–17697.

Wang, K., Diskin, S.J., Zhang, H., Attiyeh, E.F., Winter, C., Hou, C., Schnepp, R.W., Diamond, M., Bosse, K., Mayes, P.A., et al. (2011). Integrative genomics identifies LMO1 as a neuroblastoma oncogene. Nature *469*, 216–220.

Wang, L.L., Suganuma, R., Ikegaki, N., Tang, X., Naranjo, A., McGrady, P., London, W.B., Hogarty, M.D., Gastier-Foster, J.M., Look, A.T., et al. (2013). Neuroblastoma of undifferentiated subtype, prognostic significance of prominent nucleolar formation, and MYC/MYCN protein expression: a report from the Children's Oncology Group. Cancer *119*, 3718–3726.

Wang, L.L., Teshiba, R., Ikegaki, N., Tang, X.X., Naranjo, A., London, W.B., Hogarty, M.D., Gastier-Foster, J.M., Look, A.T., Park, J.R., et al. (2015b). Augmented expression of MYC and/or MYCN protein defines highly aggressive MYC-driven neuroblastoma: a Children's Oncology Group study. Br J Cancer *113*, 57–63.

Wang, Q., Diskin, S., Rappaport, E., Attiyeh, E., Mosse, Y., Shue, D., Seiser, E., Jagannathan, J., Shusterman, S., Bansal, M., et al. (2006). Integrative genomics identifies distinct molecular classes of neuroblastoma and shows that multiple genes are targeted by regional alterations in DNA copy number. Cancer Res. *66*, 6050–6062.

Wanzel, M., Herold, S., and Eilers, M. (2003). Transcriptional repression by Myc. Trends in Cell Biology *13*, 146–150.

Weinstein, I.B., and Joe, A.K. (2006). Mechanisms of Disease: oncogene addiction—a rationale for molecular targeting in cancer therapy. Nature Clinical Practice Oncology *3*, 448–457.

Weiss, W.A., Aldape, K., Mohapatra, G., Feuerstein, B.G., and Bishop, J.M. (1997). Targeted expression of MYCN causes neuroblastoma in transgenic mice. EMBO J *16*, 2985–2995.

Welcker, M., Orian, A., Jin, J., Grim, J.E., Grim, J.A., Harper, J.W., Eisenman, R.N., and Clurman, B.E. (2004). The Fbw7 tumor suppressor regulates glycogen synthase kinase 3 phosphorylation-dependent c-Myc protein degradation. Proc. Natl. Acad. Sci. U.S.A. *101*, 9085–9090.

Westermann, F., Muth, D., Benner, A., Bauer, T., Henrich, K.-O., Oberthuer, A., Brors, B., Beissbarth, T., Vandesompele, J., Pattyn, F., et al. (2008). Distinct transcriptional MYCN/c-MYC activities are associated with spontaneous regression or malignant progression in neuroblastomas. Genome Biol. *9*, R150.

White, P.S., Maris, J.M., Beltinger, C., Sulman, E., Marshall, H.N., Fujimori, M., Kaufman, B.A., Biegel, J.A., Allen, C., and Hilliard, C. (1995). A region of consistent deletion in neuroblastoma maps within human chromosome 1p36.2-36.3. PNAS *92*, 5520–5524.

White, P.S., Thompson, P.M., Seifried, B.A., Sulman, E.P., Jensen, S.J., Guo, C., Maris, J.M., Hogarty, M.D., Allen, C., Biegel, J.A., et al. (2001). Detailed molecular analysis of 1p36 in neuroblastoma. Med. Pediatr. Oncol. *36*, 37–41.

Wood, L.D., Parsons, D.W., Jones, S., Lin, J., Sjöblom, T., Leary, R.J., Shen, D., Boca, S.M., Barber, T., Ptak, J., et al. (2007). The genomic landscapes of human breast and colorectal cancers. Science *318*, 1108–1113.

Xiao, J.H., Davidson, I., Matthes, H., Garnier, J.-M., and Chambon, P. (1991). Cloning, expression, and transcriptional properties of the human enhancer factor TEF-1. Cell *65*, 551–568.

Xiong, W., and Ferrell, J.E. (2003). A positive-feedback-based bistable “memory module” that governs a cell fate decision. Nature *426*, 460–465.

Yada, M., Hatakeyama, S., Kamura, T., Nishiyama, M., Tsunematsu, R., Imaki, H., Ishida, N., Okumura, F., Nakayama, K., and Nakayama, K.I. (2004a). Phosphorylation-dependent degradation of c-Myc is mediated by the F-box protein Fbw7. EMBO J *23*, 2116–2125.

Yada, M., Hatakeyama, S., Kamura, T., Nishiyama, M., Tsunematsu, R., Imaki, H., Ishida, N., Okumura, F., Nakayama, K., and Nakayama, K.I. (2004b). Phosphorylation-dependent degradation of c-Myc is mediated by the F-box protein Fbw7. EMBO J. *23*, 2116–2125.

Yagi, R., Kohn, M.J., Karavanova, I., Kaneko, K.J., Vullhorst, D., DePamphilis, M.L., and Buonanno, A. (2007). Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. Development *134*, 3827–3836.

Yamamoto, K., Hanada, R., Kikuchi, A., Ichikawa, M., Aihara, T., Oguma, E., Moritani, T., Shimanuki, Y., Tanimura, M., and Hayashi, Y. (1998). Spontaneous regression of localized neuroblastoma detected by mass screening. J. Clin. Oncol. *16*, 1265–1269.

Yancopoulos, G.D., Nisen, P.D., Tesfaye, A., Kohl, N.E., Goldfarb, M.P., and Alt, F.W. (1985). N-myc can cooperate with ras to transform normal cells in culture. Proc Natl Acad Sci U S A *82*, 5455–5459.

Yang, H., Ou, C.C., Feldman, R.I., Nicosia, S.V., Kruk, P.A., and Cheng, J.Q. (2004). Aurora-A Kinase Regulates Telomerase Activity through c-Myc in Human Ovarian and Breast Epithelial Cells. Cancer Res *64*, 463–467.

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun *4*, 2612.

Yu, A.L., Gilman, A.L., Ozkaynak, M.F., London, W.B., Kreissman, S.G., Chen, H.X., Smith, M., Anderson, B., Villablanca, J.G., Matthay, K.K., et al. (2010). Anti-GD2 antibody with GM-CSF, interleukin-2, and isotretinoin for neuroblastoma. N. Engl. J. Med. *363*, 1324–1334.

Yu, J., Silva, J., and Califano, A. (2015). ScreenBEAM: a Novel Meta-Analysis Algorithm for Functional Genomics Screens via Bayesian Hierarchical Modeling. Bioinformatics btv556.

Yuan, X., Jin, M., Xu, X., Song, Y., Wu, C., Poo, M., and Duan, S. (2003). Signalling and crosstalk of Rho GTPases in mediating axon guidance. Nat. Cell Biol. *5*, 38–45.

Zanconato, F., Forcato, M., Battilana, G., Azzolin, L., Quaranta, E., Bodega, B., Rosato, A., Bicciato, S., Cordenonsi, M., and Piccolo, S. (2015). Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. Nat. Cell Biol. *17*, 1218–1227.

Zeller, K.I., Jegga, A.G., Aronow, B.J., O'Donnell, K.A., and Dang, C.V. (2003). An integrated database of genes responsive to the Myc oncogenic transcription factor: identification of direct genomic targets. Genome Biol *4*, R69.

Zhang, H., Liu, C.-Y., Zha, Z.-Y., Zhao, B., Yao, J., Zhao, S., Xiong, Y., Lei, Q.-Y., and Guan, K.-L. (2009). TEAD transcription factors mediate the function of TAZ in cell growth and epithelial-mesenchymal transition. J. Biol. Chem. *284*, 13355–13362.

Zhao, B., Ye, X., Yu, J., Li, L., Li, W., Li, S., Yu, J., Lin, J.D., Wang, C.-Y., Chinnaiyan, A.M., et al. (2008a). TEAD mediates YAP-dependent gene induction and growth control. Genes Dev. *22*, 1962–1971.

Zhao, B., Ye, X., Yu, J., Li, L., Li, W., Li, S., Yu, J., Lin, J.D., Wang, C.-Y., Chinnaiyan, A.M., et al. (2008b). TEAD mediates YAP-dependent gene induction and growth control. Genes Dev. *22*, 1962–1971.

Zhao, X., Heng, J.I.-T., Guardavaccaro, D., Jiang, R., Pagano, M., Guillemot, F., Iavarone, A., and Lasorella, A. (2008c). The HECT-domain ubiquitin ligase Huwe1 controls neural differentiation and proliferation by destabilizing the N-Myc oncoprotein. Nat. Cell Biol. *10*, 643–653.

Zhu, S., Lee, J.-S., Guo, F., Shin, J., Perez-Atayde, A.R., Kutok, J.L., Rodig, S.J., Neuberg, D.S., Helman, D., Feng, H., et al. (2012). Activated ALK Collaborates with MYCN in Neuroblastoma Pathogenesis. Cancer Cell *21*, 362–373.

Zimmerman, K.A., Yancopoulos, G.D., Collum, R.G., Smith, R.K., Kohl, N.E., Denis, K.A., Nau, M.M., Witte, O.N., Toran-Allerand, D., Gee, C.E., et al. (1986). Differential expression of myc family genes during murine development. Nature *319*, 780–783.

Attiyeh, E. F., Diskin, S. J., Attiyeh, M. A., Mosse, Y. P., Hou, C., Jackson, E. M., Kim, C., Glessner, J., Hakonarson, H., Biegel, J. A., and Maris, J. M. (2009). Genomic copy number determination in cancer cells from single nucleotide polymorphism microarrays based on quantitative genotyping corrected for aneuploidy. Genome research *19*, 276-283.

Michels, E., Vandesompele, J., De Preter, K., Hoebeeck, J., Vermeulen, J., Schramm, A., Molenaar, J. J., Menten, B., Marques, B., Stallings, R. L.*, et al.* (2007). ArrayCGH-based classification of neuroblastoma into genomic subgroups. Genes Chromosomes Cancer *46*, 1098-1108.