Computational Analysis of Multi-omic Data for the Elucidation of Molecular Mechanisms of Neuroblastoma

Abdulazeez Giwa

Supervisor: Dr Hocine Bendou

Co-Supervisors: Dr Azeez Fatai, Prof Junaid Gamieldien

South African National Bioinformatics Institute
Faculty of Natural Sciences
University of the Western Cape

This dissertation is submitted for the degree of

Doctor of Philosophy

September 2021

http://etd.uwc.ac.za/

Declaration

I declare that **Computational analysis of multi-omic data for the elucidation of molecular mechanisms of neuroblastoma** is my work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Abdulazeez Giwa
September 2021

http://etd.uwc.ac.za/

Acknowledgements

I would like to express my appreciation to all those without whom this dissertation would not have been possible.

My deepest appreciation goes to my supportive supervisor, Dr Hocine Bendou for his valuable guidance, support, contributions, and time to make this research project successful. You have improved me in many ways especially my technical and writing skills.

Many thanks to my co-supervisor, Prof Junaid Gamieldien, for your guidance and support. Your valuable insights at the early stages of my program were beneficial. There was always much to learn from you at every interaction.

I appreciate the efforts of my co-supervisor, Dr Azeez Fatai, for supporting me in obtaining admission to SANBI, UWC. Your contributions and guidance throughout the research project is appreciated.

I appreciate my spouse, Zainab Abdulrahman-Giwa for her patience of my absence and support throughout the course of this program.

I sincerely appreciate Prof Alan Christoffels for coming through for me and others at crucial times during this program.

I also acknowledge the support of all SANBI members including Nasr, Maryam, Rumbidzai, Femi, Abiola, Nomlindelo and Catherine. Many thanks to the administrative staff, Fungiwe, Ferial and the SANBI help desk for assisting with the compute cluster usage and other technical issues.

Also, my appreciations go to the Lagos State University for their support. Many thanks to all members of staff of the Department of Zoology and Environmental Biology at the Lagos State University.

Finally, I would like to thank my family in Nigeria and Scotland, as well as my friends in Nigeria, here in UWC, and in South Africa, for their immense support.

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

Abstract

Neuroblastoma is the most common extracranial solid tumor in childhood. The survival rates of patients with neuroblastoma, especially those in the high-risk category, are still low despite varied therapies. The detailed understanding of the molecular mechanisms underlying the pathogenesis of neuroblastoma is essential to develop better therapeutics and improve the poor survival rates. This study provides a multi-omic analysis of neuroblastoma datasets from the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) neuroblastoma project and the Gene Expression Omnibus (GEO) data portals to better understand the molecular mechanisms of neuroblastoma. For this, bioinformatics methods, including differential gene expression, gene and disease enrichment, gene regulatory network, differential methylation, Cox regression, Kaplan-Meier estimates and machine learning (ML), were applied to identify genetic signatures that can serve as biomarkers of survival time and *MYCN* amplification in high-risk neuroblastoma. A 16-gene expression signature was found to be predictive of survival time in high-risk neuroblastoma. This gene expression signature was validated in an external neuroblastoma dataset obtained from the GEO using Support Vector Machine (SVM) and Artificial Neural Networks (ANN) algorithms. The gene regulatory networks of short survival samples revealed the importance of the *SMIM28*, *MAPK15* and *UBC* genes as the origins of most connections. The differentially expressed genes (DEGs) identified are involved in tumor proliferating activities, and some are already associated with neuroblastoma and other cancers, while others are novel and reported here for the first time. In addition, a 25-CpG signature was discovered to be diagnostic of the amplification and non-amplification of *MYCN* in neuroblastoma. This DNA methylation signature was validated in three external neuroblastoma datasets obtained from the GEO using the SVM algorithm. Eight of the 25-CpGs were associated with overall survival, with the methylation of cg13558971 (*ATP2BP4*), cg25310824 (*SEPP1*) and cg07476617 (*CFLAR*) indicating a poor outcome in *MYCN* amplified neuroblastoma. The accuracy of the gene expression and methylation signatures in predicting survival and clinical events

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

demonstrate their utility for clinical use. These signatures may also be combined and translated into clinical use in the immediate as well as be developed into a robust standardized clinical test for neuroblastoma diagnosis and prognosis in the near future.

http://etd.uwc.ac.za/

Table of Contents

Declaration ii
Acknowledgements iii
Abstract iv
Table of Contents vi
List of figures ix
List of tables x
List of Abbreviations xi
Introduction 1
1.1 Pediatric Cancer 2
1.1.1 Neuroblastoma 3
1.2 Justification 5
1.3 Aims and objectives 6
1.4 Scope 6
Literature Review 8
2.1 Multi-omic approach to diseases 8
2.2 Multi-omics Resources 9
2.2.1 UCSC Xena Browser 10
2.2.2 Therapeutically Applicable Research to Generate Effective Treatments (TARGET) 11
2.2.3 The Cancer Genome Atlas (TCGA) 12
2.2.4 Gene Expression Omnibus (GEO) 13
2.3 Data Preprocessing 13
2.4 Bioinformatics tools and methods 15
2.5 Machine Learning (ML) 16
2.5.1 Supervised learning 16
2.5.2 Unsupervised learning 18
2.5.3 Semi-supervised learning 18
2.5.4 Machine learning in multi-omics and cancer 18
2.6 Biomarkers 20
2.6.1 Biomarker discovery 20
2.6.2 Biomarkers in clinic 21
2.7 Neuroblastoma 22
2.7.1 Origin 22
2.7.2 Genomic alterations in neuroblastoma 23
2.7.3 Neuroblastoma and *MYCN* 25
2.7.3.1 Therapeutic targeting of *MYCN* 27
2.7.4 Epigenetics 28

UNIVERSITY of the WESTERN CAPE
http://etd.uwc.ac.za/

2.7.4.1 DNA methylation 28
2.7.4.2 Histone modifications 29
2.7.4.3 Micro RNAs 30
2.7.4.4 The epigenome as a therapeutic target 30
2.7.5 Therapy 31
2.8 Summary 32
Identification of novel prognostic markers of survival time in high-risk neuroblastoma using gene expression profiles 34
3.1 Abstract 34
3.2 Introduction 35
3.3 Methods 36
3.3.1 Datasets 36
3.3.2 Differential Gene Expression analysis 37
3.3.3 Machine learning 38
3.3.4 Regulatory network analysis 38
3.3.5 Availability of data and materials 39
3.4 Results 41
3.5 Discussion 47
3.5.1 Differentially expressed genes 47
3.5.2 Gene and Disease Ontology enrichment analyses 51
3.5.3 Regulatory network analysis 52
3.5.4 Machine Learning 53
3.6 Conclusion 53
Predicting amplification of *MYCN* using CpG methylation biomarkers in neuroblastoma 55
4.1 Abstract 55
4.2 Introduction 56
4.3 Methods 58
4.3.1 Dataset 58
4.3.2 Differential methylation analysis 59
4.3.5 Availability of data and materials 61
4.4 Results 63
4.4.1 Differential methylation analysis 63
4.4.2 Machine learning (ML) 63
4.5 Discussion 75
5.0 Conclusions and further considerations 81
5.1 Summary 81
5.1.1 What was done 81
5.1.2 What was found 82
5.2 Potential future research 82

UNIVERSITY of the WESTERN CAPE
http://etd.uwc.ac.za/

5.3 Potential clinical applications....83
5.3.1 Clinical applications: Immediate....83
5.3.2 Clinical applications: Future....83
Bibliography....85

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

List of figures

2.1 Simplified illustration of the SVM and ANN classification model……..17

2.2 Cancer classification using machine learning……....19

3.1 Workflow outlining the steps and methods undertaken in gene expression study…….40

3.2 Gene and disease enrichment analyses……...44

3.3 Gene regulatory network analysis……..46

4.1 Workflow depicting the steps and methods used in the DNA methylation study……..62

4.2 Heatmap of methylation level of the 25 top significant CpG sites in *MYCN* amplification group……..70

4.3 Clustering of *MYCN* amplification samples using all 663 significant CpGs……....71

4.4 Kaplan-Meier estimates of overall and event-free survival……..73

4.5 Kaplan-Meier estimates of overall and event-free survival of the GSE65306 test dataset……74

5.1 Summary of the potential clinical application of the findings of this research project……84

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

List of tables

3.1 Query criteria for short and long survival sample selection.……37

3.2 Upregulated and downregulated DEGs in SS neuroblastoma samples compared to LS samples……...42

3.3 Machine learning results for the classification of the GSE49711 test set of high-risk neuroblastoma samples with SVM and ANN models…..……...47

4.1 Fourteen highly methylated genes between the *MYCN* amplified and *MYCN* non-amplified groups…….65

4.2 Enrichment analysis of the 369 differentially methylated…….66

4.3 Twenty-five differentially methylated CpGs between the *MYCN* amplified and *MYCN* non-amplified groups selected by RFE…….69

4.4 Results of the machine learning prediction of *MYCN* amplification of the samples from GSE54719, GSE120650, GSE65306 test sets……..…….72

4.5 CpGs selected by the cox-regression model and their respective regression coefficients for overall and event-free survival……72

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

List of Abbreviations

ANN	Artificial Neural Networks
CGI	CpG Island
CIMP	CpG Island Methylator Phenotype
COG	Children's Oncology Group
CRC	Colorectal Cancer
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSC	Cancer Stem Cells
DEG	Differentially Expressed Gene
DGE	Differential Gene Expression
DMG	Differentially Methylated Gene
DNA	Deoxyribonucleic acid
ECM	Extracellular Matrix
EFS	Event-Free Survival
EMT	Epithelial-Mesenchymal Transition
GEO	Gene Expression Omnibus
GO	Gene Ontology
INRG	International Neuroblastoma Risk Group
INSS	International Neuroblastoma Staging System
K-M	Kaplan-Meier
LS	Long Survival
ML	Machine Learning

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

MYCN	MYCN proto-oncogene, bHLH transcription factor
NC	Neural Crest
NCBI	National Center for Biotechnology Information
NCC	Neural Crest Cells
OS	Overall Survival
RFE	Recursive Feature Elimination
RNA	Ribonucleic acid
SS	Short Survival
SVM	Support Vector Machines
TARGET	Therapeutically Applicable Research to Generate Effective Treatments

http://etd.uwc.ac.za/

Chapter 1

Introduction

Cancer is a significant cause of morbidity and mortality globally, irrespective of the level of socio-economic development. It is an important public health problem worldwide with an estimated 19.3 million new cases and almost 10 million deaths in 2020 (Sung et al., 2021) and remains a leading cause of death globally. The burden of cancer grows and continues to exert tremendous emotional, physical and financial strain on individuals, families, communities and health systems.

UNIVERSITY of the
WESTERN CAPE

Cancer is a generic term used to describe a large number of diseases characterized by the development of abnormal cells that divide without constraints and have the ability to spread and negatively impact normal body tissues. It can develop at any stage in life and in any organ of the body. Normal cells are continuously subjected to cues that instruct them to undergo mitosis, maintain basic physiological cellular functions and/or apoptose. Cancer cells develop some pathological autonomy from these signals, thereby resulting in unconstrained growth and proliferation. Cancer genes are genes that, if altered in sequence, copy number or expression level, can initiate and/or contribute to tumorigenesis. Cancer genes can be categorized into two classes, namely oncogenes and tumour suppressor genes (Klein, 1988; Stass and Mixson, 1997; Guo et al., 2014). Oncogene alterations and/or overexpression

http://etd.uwc.ac.za/

results in the activation of cellular events that promote cell growth, proliferation and survival. Some risk factors associated with cancer include tobacco, infectious agents (e.g bacteria and viruses; *Helicobacter pylori*, Epstein-Barr virus, HIV-1, Human Papilloma Virus), alcohol consumption, exposure to ultraviolet and ionizing radiation, chemicals and family history (Vineis and Wild, 2014).

Cancer is a complex, heterogeneous and highly dynamic disease with multiple evolving molecular components (Moses et al., 2018). Hanahan and Weinberg, (2011) defined hallmarks of cancer as functional abilities acquired by cancer cells that allow them to survive, proliferate and invade. These hallmarks include proliferative signalling sustenance, growth suppressors evasion, resistance to cell death, enabling of replicative immortality, angiogenesis induction, activation of invasion and metastasis, cellular energetics deregulation and avoidance of immune destruction (Hanahan and Weinberg, 2011). Pavlova and Thompson (2016) organized metabolic changes due to cancer into six hallmarks, namely: (1) deregulated uptake of glucose and amino acids, (2) use of opportunistic modes of nutrient acquisition, (3) use of glycolysis/TCA cycle intermediates for biosynthesis and NADPH production, (4) increased demand for nitrogen, (5) alterations in metabolite-driven gene regulation, and (6) metabolic interactions with the microenvironment (Pavlova and Thompson, 2016).

UNIVERSITY of the
WESTERN CAPE

1.1 Pediatric Cancer

Cancer is also a major cause of death in children worldwide and the most common cause of death in children in developed nations, with the recorded incidence rate increasing yearly (Kaatsch, 2010; Steliarova-Foucher et al., 2017). Every year, around 300,000 children between infancy to fourteen years of age are diagnosed with cancer worldwide (Steliarova-Foucher et al., 2017). The incidence of cancer in childhood has long been noted to vary by age, sex and ethnicity. The incidence of most types of cancer is lower in Black, Asian and Hispanic children than in white children (Spector et al., 2015). The burden of childhood cancer globally remains poorly qualified, as there are limited data

http://etd.uwc.ac.za/

from middle and low-income countries (Bhakta et al., 2019). Furthermore, since children who survive cancer are at increased risk for early mortality, second malignancies, cardiac impairments, musculoskeletal abnormalities, chronic diseases, and fertility issues, among other problems (Ness and Gurney, 2007; Ward et al., 2014), they are regularly monitored after treatment. The most common pediatric cancers are acute lymphoblastic leukemia (ALL), central nervous system (CNS) tumors, neuroblastoma, soft tissue sarcomas, Wilms tumor, osteosarcoma and retinoblastoma. Of interest to this study is pediatric neuroblastoma.

Sequencing studies have shown that there are salient differences between adult and pediatric cancers. One such difference is the substantially lower number of somatic mutations in most pediatric cancers compared to adult cancers, with the exception of germline mutations in cancer predisposition genes (Lawrence et al., 2013; Vogelstein et al., 2013; Campbell et al., 2017). This lower mutational burden has been proposed to be the result of embryonic genesis of these cancers, alteration of developmental pathways, and a smaller contribution of environmental cancer-causing agents (Sweet-Cordero and Biegel, 2019). Another difference is the diversity observed in the types of genomic alterations likely driving pediatric cancer growth including fusions of genes, alterations in copy number, enhancer hijacking events and chromoplexy (Sweet-Cordero and Biegel, 2019) which are prognostic of many pediatric cancer subtypes.

UNIVERSITY of the
WESTERN CAPE

1.1.1 Neuroblastoma

Neuroblastoma is the most common extracranial solid tumor in childhood (mostly under five years of age). It is an embryonal malignancy of early childhood with a poor prognosis for children diagnosed at over 18 months of age with disseminated disease, accounting for approximately 15% of childhood cancer mortality (Maris, 2010; Smith et al., 2010, Ward et al., 2014). It is a malignancy of

http://etd.uwc.ac.za/

the sympathetic nervous system occurring anywhere in the system and often presenting with wide-spread metastatic disease, resulting in survival rates of less than 50% (Maris et al., 2007; Maris, 2010). Sixty percent of the tumors occur within the abdomen, commonly in the adrenal medulla (Zhang et al., 2018). Clinical presentation of children with neuroblastoma varies depending on tumor location and extent of disease. This may include abdominal mass, abdominal pain, bone pain, inability to walk, polymyoclonic ataxia, opsoclonus, nausea, vomiting, weight loss, nystagmus and Horner's syndrome (Holland et al., 1980; Weinstein et al., 2003).

Diagnosis is made using histology, chemical profiling and imaging (Swift et al., 2018). As neuroblastoma often expresses catecholamine metabolism enzymes, elevated serum and urine levels of vanillylmandelic acid and homovanillic acid may also indicate chemical evidence of the disease (Swift et al., 2018). A very heterogenous clinical course characterizes the disease with the observation of cases presenting with spontaneous regression to relentless progression and eventual death despite extensive and prompt therapy (Baali et al., 2018; Newman and Nuchtern, 2016). The mechanisms for this heterogeneous clinical course are not yet fully understood.

UNIVERSITY of the
WESTERN CAPE

The Children's Oncology Group (COG) uses age at diagnosis, disease stage, *MYCN* amplification status, the International Neuroblastoma Pathology Classification and DNA ploidy to stratify risk groups. High-risk disease presents with poor outcome even with intensive and varied mode of therapy while low risk group have good outcomes (Zhang et al., 2018). However, patients within the same risk group can still show variable prognosis (Baali et al., 2018). A disproportionate number of patients with high-risk neuroblastoma will die or suffer significant treatment-related morbidity with 5-year overall survival rates of approximately 20% (Oeffinger et al., 2006; Yu et al., 2010, Maris, 2010; London et al., 2010).

http://etd.uwc.ac.za/

The genetic repertoire underlying the genesis and progression of neuroblastoma is still not completely understood (Barr and Applebaum, 2018). Analysis of multi-omic datasets from patients with neuroblastoma is therefore imperative to determine and fully understand the mechanisms of the disease and the development of new therapeutic targets and approaches to improve survival while minimizing toxicity. The TARGET initiative (https://ocg.cancer.gov/programs/target/overview) employs extensive molecular characterization to investigate the drivers of hard-to-treat, high-risk pediatric cancer initiation and progression to produce comprehensive genomic profiles. Therefore, it was utilised as one of the main sources of neuroblastoma genetic data for this study.

1.2 Justification

Dissecting the underlying events driving high-risk neuroblastoma remains a major hurdle for clinicians and researchers. In addition, the survival rate of high-risk neuroblastoma is still low despite multimodal therapy. Analysing different genomic data types has the potential to provide new knowledge about the mechanisms of the disease as well as provide novel therapeutic targets. Assessing the changes that occur in the transcriptome and methylome using computational tools and methods may provide accurate prognostic, diagnostic and predictive tools for assessing neoplastic disease.

Genomic biomarkers (transcriptomic and methylomic) have advantages that makes them suitable for broad use in diagnostics. Therefore, finding the appropriate biomarkers that are accurately diagnostic and predictive of disease states is crucial. Finding diagnostic and prognostic markers with the potential to improve the survival of high-risk neuroblastoma patients as well as low and intermediate risk patients is a necessity. This study therefore proposes to address these important gaps, with accurate prognosis being necessary to enable upfront initiation of targeted therapy and ultimately improve survival.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

1.3 Aims and objectives

The main aim of this research project was to find novel genetic signatures that are prognostic and/or diagnostic of *MYCN* amplification and survival time in high-risk neuroblastoma.

The specific objectives were to:

(1) Identify genes that are differentially expressed between short and long survival groups in neuroblastoma.

(2) Perform gene and disease ontology enrichment analysis to functionally annotate the identified differentially expressed genes.

(3) Apply machine learning (ML) techniques to validate the differential gene expression results.

- Apply ML techniques to predict survival time in neuroblastoma using gene expression profiles.

(4) Uncover the genetic regulatory networks underlying survival time in neuroblastoma using gene expression data.

(5) Identify genes differentially methylated between *MYCN* amplified and *MYCN* non-amplified neuroblastoma.

UNIVERSITY of the
WESTERN CAPE

(6) Identify CpG methylation biomarkers diagnostic of *MYCN* amplification using ML techniques.

(7) Identify CpG methylation biomarkers associated with survival and poor prognosis.

- Determine significant CpGs best correlated with patient survival using Cox regression model.
- Compute Kaplan-Meier estimates and Hazard Ratio for overall and event-free survival.

1.4 Scope

This thesis is organised into five chapters. Chapter 1 provides a background to neuroblastoma as well as the aims and objectives of this project. Chapter 2 (literature review) details the multi-omics approach to disease analysis and characterization, available bioinformatics resources, biomarker

http://etd.uwc.ac.za/

discovery, the origin, genomic and epigenetic basis of neuroblastoma as well as therapeutic implications. Chapter 3 outlines the use of transcriptomic data for network analysis and finding predictive gene expression biomarkers of survival time in neuroblastoma. In Chapter 4, ML techniques are applied to tumor DNA methylation data for the discovery of diagnostic biomarkers of *MYCN* amplification and prediction of survival prognosis in neuroblastoma. The project conclusions, potential future research and clinical applications are discussed in chapter 5.

http://etd.uwc.ac.za/

Chapter 2

Literature Review

2.1 Multi-omic approach to diseases

High-throughput technologies have revolutionized biomedical research and are now a standard method of carrying out biological research. Omics is the comprehensive characterization, and quantification of many molecules grouped according to fundamental structural or functional biological similarities that they demonstrate (Perakakis et al., 2018). The currently important omics levels include genomics, epigenomics, transcriptomics, proteomics and metabolomics. Genomics uses sequencing and bioinformatic methods to sequence, assemble and analyze the structure and function of genomes. It allows the identification of genetic variants associated with disease (Hasin et al., 2017) using experimental techniques like genome-wide association studies, genotype arrays (Hirschhorn and Daly, 2005; Ragoussis, 2009), exome sequencing (Ng et al., 2009) and next generation sequencing for whole genome sequencing (Koboldt et al., 2013). Epigenomics characterizes reversible modifications of DNA and histone proteins (Esteller, 2007). Transcriptomics identifies and quantifies all RNA levels in the genome (Blow, 2009; Sager et al., 2015). Proteomics examines and quantifies proteins present in the biological system (Aslam et al., 2017). Metabolomics

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

qualifies and quantifies small molecules such as carbohydrates, fatty acids, amino acids and other products of cellular metabolism (Pinu et al., 2019). Since cancer is a complex disease, it requires investigation at different omics levels.

Multi-omics is a robust and multifaceted approach for probing biological systems, encompassing multiple layers of biological information (Brademan et al., 2020). It allows the understanding of information flow that underlies a disease. Each omics data type can be valuable as markers of disease process and inform about differential pathways between control and disease groups. Omics data have enabled the characterization of the molecular features of cancer, allows for cancer subtyping, and identification and prediction of biomarkers, thereby ultimately facilitating the understanding of oncological mechanisms (Iorio et al., 2016; Subramanian et al., 2020).

A major advantage of multi-omics is the breadth of information it provides. Multi-omics analysis allows the identification of disease factors at different omics levels, thus increasing the likelihood of identifying the underlying cause of a disease. Multi-omics datasets can provide molecular insights beyond the sum of individual omics. The availability of large omics datasets is also enduring, ensuring that they can be reanalyzed repeatedly if needed. Therefore, the development of multi-omics resources and bioinformatics methods and tools to extract knowledge from existing data types is important.

2.2 Multi-omics Resources

The continued growth and complexity of cancer genomics data arising from numerous technological advancements and their continued reduction in costs created a need for large repositories and databases to store this data. Most of these resources are publicly available for researchers to query, use and study without restrictions for purposes of reproducibility, discovery, and validation of

UNIVERSITY of the WESTERN CAPE
http://etd.uwc.ac.za/

findings (Yang et al., 2015; Pavlopoulou et al., 2015). Several types of data are curated in bioinformatics resources globally. Two classes of these data types are omics data, and transcription factors and signal transduction pathways (Chen, 2015). These types of data are stored in different file formats and can be retrieved from relevant cancer data repositories such as UCSC Xena, TCGA, TARGET, and GEO.

2.2.1 UCSC Xena Browser

UCSC Xena (http://xena.ucsc.edu) is a high-performance visualization and analysis tool for multi-omics data of large public repositories and private datasets with a front-end browser and a back end hub (Goldman et al., 2020). Xena hosts over 1600 datasets from more than 50 different cancer types, including datasets from the TCGA, ICGC, TCGA Pan-Cancer Atlas, PCAWG (Pan-Cancer analysis of Whole Genomes) and the Genomic Data Commons (GDC). It also hosts results from the UCSC Toil RNA-Seq Recompute Compendium, which uniformly realigned gene and transcript expression dataset for all TCGA, TARGET and GTEx samples to allow for comparison between samples of these datasets.

UNIVERSITY of the
WESTERN CAPE

The Xena Browser (https://xenabrowser.net) is web-based and enables data exploration across multiple hubs, along with a wide variety of analysis and visualization tools, including survival analyses, scatter plots, bar graphs and statistical tests. Current supported data types are clinical data and sample annotations, phenotype, DNA methylation, RNA-seq gene expression counts and normalized counts, copy number variations, ATAC-Seq peak signals and large structural variants (Goldman et al., 2020). For example, the GDC TARGET neuroblastoma dataset available in Xena includes RNA-Seq gene expression data in log-normalized count/FPKM/FPKM-UQ formats, phenotype information and survival data. These type of datasets in Xena Browser are open access and easy to retrieve and download. Several bioinformatics studies have used data (Zhu et al., 2019; Wu

http://etd.uwc.ac.za/

et al., 2019; Barman et al., 2020; Giwa et al., 2020), published data (Kang et al., 2020), and made visualizations (Barman et al., 2020; Zheng and Fu, 2020; Zhang et al., 2020) from or on the Xena browser.

2.2.2 Therapeutically Applicable Research to Generate Effective Treatments (TARGET)

The TARGET program comprehensively characterizes the molecular changes that drive pediatric cancers at the transcriptome, genome and epigenome level. The program is organized into disease-specific project teams of five cancers (acute lymphoblastic leukemia, acute myeloid leukemia, kidney tumors, neuroblastoma, osteosarcoma) collaborating. Available data types include transcriptome sequences (RNA-Seq, mRNA-Seq, miRNA-Seq), copy number variation, whole genome sequences, whole exome sequences and DNA methylation. The data are organized at four levels, namely level 1 data is raw or low-level data files, level 2 and 3 are normalized and integrated data, and level 4 data are summarized results. These data levels are made available in (https://ocg.cancer.gov/programs/target/data-matrix) and their access is open or restricted depending on the type of data.

UNIVERSITY of the
WESTERN CAPE

Open access data are freely accessible and downloadable, and are typically analyzed data, which cannot be used to identify individual patients. This includes de-identified clinical information, tissue pathology data, chromosome specific (segmented) copy number alterations, loss of heterozygosity, gene expression and methylation arrays. Controlled access data, on the other hand, are data presenting a certain risk of patient re-identification. Controlled access data have specific tumor/patient information and raw molecular data (DNA and RNA sequence files). Therefore, approvals must be submitted and obtained from NCBI's Database of Genotypes and Phenotypes (dbGaP) (Mailman et al., 2007; Tryka et al., 2014; Wong et al., 2016) to access and use this type of data.

http://etd.uwc.ac.za/

The TARGET neuroblastoma cohort comprises nearly 200 high-risk cases, including some low-risk and intermediate-risk cases (Pugh et al., 2013; Oldridge et al., 2015; Eleveld et al., 2015; Wei et al., 2018). The patients are managed through the Children's Oncology Group (COG), a clinical trials group focused on childhood cancers. A fully characterized case consists of nucleic acid samples that have been extracted from primary tumor samples obtained at diagnosis, normal tissue sample(s) and/or peripheral blood or bone marrow (case-matched), as well as relapsed tumor sample(s) (case-matched) when available. Open access data on the data matrix include level 3 gene expression array files, level 3 DNA methylation files (beta values), level 3 copy number array files, level 3 and 4 whole genome sequence files, as well as level 3 whole exome sequencing files, clinical information and sample matrix files. For this study, level 3 gene expression and DNA methylation data were used. Clinical information including gender, first event, event-free survival time, overall survival time, vital status, *MYCN* status, International Neuroblastoma Staging System stage and COG risk group of the TARGET samples were made available (https://target-data.nci.nih.gov/Public/NBL/clinical/harmonized/TARGET_NBL_ClinicalData_Discovery_20170525.xlsx).

UNIVERSITY of the WESTERN CAPE

2.2.3 The Cancer Genome Atlas (TCGA)

The TCGA is one of the largest cancer genomics programs to date and a joint project of the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI). The TCGA provides a comprehensive map of the important genomic changes that occur in the major types and subtypes of cancer. It contains clinical information, genomic characterization data and high-level sequence analysis of tumor genomes (Cancer Genome Atlas Research Network, 2013). The TCGA program has generated and analyzed genomic data on over 11000 individuals and 30 different types of cancer (Wang et al., 2016). Data types are categorized into open access data and controlled data, which require application and approval for access. The data available at the GDC is broadly

http://etd.uwc.ac.za/

categorized into biospecimen and clinical data, molecular analysis data and analysis metadata (Gao et al., 2019). Data levels consist of either raw or processed data. An example of a raw file includes an Affymetrix CEL file, while processed files can be VCF or MAF file types. TCGA data can be accessed directly via the data portal (https://portal.gdc.cancer.gov/) or by using software such as the R packages, TCGAbiolinks (Colaprico et al., 2016), TCGA2STAT (Wan et al., 2016), TCGAIntegrator and xenaPython python package.

2.2.4 Gene Expression Omnibus (GEO)

The Gene Expression Omnibus (GEO) database (https://ncbi.nlm.nih.gov/geo/) is an international public repository of high-throughput gene expression and functional genomics datasets submitted by the research community (Clough and Barrett, 2016). It started out as a resource for gene expression studies only, but has evolved to accept data from newer technologies such as genome methylation, chromatin structure, and genome-protein interactions (Edgar et al., 2002; Barret et al., 2013; Clough and Barrett, 2016). The main goals of GEO are to efficiently store data in the right format and annotations in an easily accessible manner. As of June 2021, there were 4348 dataset records and 22258 platforms. Data on the GEO can be programmatically accessed, and can also be visualized and analyzed using the GEO2R and GEOProfiles tools. Data on GEO are stored in text and spreadsheet formats and raw files are also available in tar file formats, with links provided to navigate to the raw sequence data (if available). Prior to any data analysis, data preprocessing is a required step to ensure efficient and accurate results.

2.3 Data Preprocessing

Data preprocessing is a very important initial step in the analysis of raw data. It is necessary to get data into a format that will be more easily and efficiently processed and to ensure the data quality is

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

maintained for accurate downstream analysis and biological inference. Data preprocessing aims to deal with noise, missing values, high-dimensionality and inconsistent format of data. Preprocessing methods to deal with these issues include sampling, dimensionality reduction, imputation, and data normalization (Roy et al., 2019).

The goal of sampling is to use only a small representative sample of a large dataset, and is obtained by methods such as Simple Random, Stratified and Cluster sampling (Lee et al., 1998; Ali et al., 2019; Mishra et al., 2020). Dimensionality reduction aims to select only relevant attributes by removing unnecessary or insignificant ones from data. A commonly used technique for dimensionality reduction is principal component analysis (PCA). It is a statistical procedure that uses an orthogonal transformation to map a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (Hotelling, 1933). PCA reduces the complexity in high-dimensionality data by transformation into fewer dimensions called principal components aiming to find the best summary of the data using a limited number of principal components (Lever et al., 2017). Missing values could be handled by either ignoring and discarding the missing data or by imputation. The imputation method is the replacement of missing data with estimated values using information present in the whole data. Imputed value could be obtained by mean imputation, regression or k-nearest neighbour methods (Schmitt et al., 2015; Song et al., 2020).

Normalization transforms data from one domain to another to make it roughly normally distributed (Roy et al., 2019). Normalization methods include $\log_2$ transformation, min-max, z-score, and quantile (Baumgartner et al., 2011; Roy et al., 2019). Log_2 transformation is a widely used normalization for gene expression data (Quackenbush, 2002; Anders and Huber, 2010). As such, some catalogues like the Xena Browser allow downloading normalized gene expression data in $\log_2$ transformed format for most cancers, including the TARGET neuroblastoma dataset used in this study (Goldman et al., 2020). Quantile normalization makes two or more distributions identical statistically

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

(Bolstad et al., 2003). In Min-Max normalization, the attribute value is transformed to a new value based on the value of the minimum and maximum values of the data.

$$X = \frac{x-min}{max-min}$$

Finally, for z-score normalization, the values of the variable is transformed based on the mean and standard deviation of the distribution.

$$X = \frac{x-mean}{standarddeviation}$$

Normalization is required to account for systematic biases introduced during sample processing and data generation (Chawade et al., 2014). For numerical data with features of different ranges, normalization changes the values to a common scale without distorting the underlying differences. Failure to account for the biases and scale differences in data could lead to wrong conclusions from quantitative analyses (Chawade et al., 2014).

2.4 Bioinformatics tools and methods

Developing methods that accurately extract knowledge from ever-growing multi-omics datasets is a necessity. Bioinformatics tools and methods used to analyse the different levels of omics data are continually developed and improved upon (Reich et al., 2006; Afgan et al., 2018). For example, Bioconductor (https://bioconductor.org) is one of the largest open source projects for omics data analysis (Huber et al., 2015) with the latest version (v3.13) containing 2042 software packages. Bioinformatics tools and software packages are generally developed in R language (e.g. Bioconductor) or Python environment (e.g. Bioconda). Differential gene expression (DGE) analysis and Differential methylation analysis (DMA) are bioinformatics methods used for extracting biological meaning from transcriptome and DNA methylation data, respectively. The DGE and DMA methods were employed in this study to uncover the molecular mechanism differences in neuroblastoma.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

DGE analysis is a statistical method for evaluating differences in abundance of gene transcripts between experimental groups (Conesa et al., 2016). Some of the tools used for DGE analysis include edgeR (Robinson et al., 2010), DESeq2 (Love et al., 2014), limma (Ritchie et al., 2015), Cuffdiff (Trapnell et al., 2012) and NOISeq (Tarazona et al., 2011). The output of DGE analysis is a list of genes that have significant differences in gene expression between the experimental groups being compared. The output file contains the gene names/ID, log2 fold change value, *p*-value and *p*-adjusted value. DMA identifies genomic loci with different levels of methylation among distinct biological conditions or experimental groups (Park and Wu, 2016). Some of the tools used for DMA include ChAMP (Morris et al., 2014; Tian et al., 2017), RnBeads (Assenov et al., 2014), methylKit (Akalin et al., 2012), and minfi (Aryee et al., 2014). The output of DMA is a list of genes that have significant differences in CpG methylation sites between the compared biological conditions. The output file typically contains gene names/CpG ID, beta value difference (Δβ) and *p*-values. With the growing applicability of ML in genomics, several algorithms could be applied to the outputs of DGE analysis and DMA for prediction and validation studies.

2.5 Machine Learning (ML)

UNIVERSITY of the
WESTERN CAPE

ML focuses on methods for developing computer programs that learn from experience with respect to specific tasks (Mitchell, 1997). It represents a powerful set of algorithms that can characterize, adapt, learn, predict and analyse data (DeGregory et al., 2018). The ultimate goal in many ML tasks is to optimize performance of models built on independent test datasets (Zou et al., 2019). ML can be categorized into supervised learning, unsupervised learning and semi-supervised learning.

2.5.1 Supervised learning

In supervised learning, a classifier is built which then makes predictions on future input to allot their class labels (Kotsiantis, 2007; Yousef et al., 2014). A training dataset (that will be used to build the classifier) and a testing dataset (that will be used to measure the performance of the classifier)

http://etd.uwc.ac.za/

constitute the data parts in supervised learning. Examples of supervised learning algorithms include, decision trees (Yuan and Shaw, 1995), support vector machines (SVMs) (Vapnik, 1995) and artificial neural networks (ANN) (McCulloch and Pitts, 1943). Decision trees are trees that classify instances by sorting them based on feature values. In contrast, SVMs separate classes in training data by producing a separating hyperplane (Fig. 2.1A). Applying SVMs with feature selection methods, like recursive feature elimination (RFE), has been demonstrated to improve classification performance (Guyon et al., 2002). Feature selection generates more robust classification models by eliminating noisy and unimportant features (Krishnapuram et al., 2004), as well as more compact and faster models (Guyon and Elisseeff, 2003). SVMs have recently been widely used in bioinformatics tasks. Another popular classification method, ANN, processes information like the neurons of the human nervous system. It relays information across layers of nodes (neurons) with varying degrees of interconnectedness (Agrawal and Agrawal, 2015; Nagy et al., 2020) (Fig. 2.1B). Information is passed to an input layer, processed by hidden layers and finally to an output layer (Nagy et al., 2020).

UNIVERSITY of the WESTERN CAPE

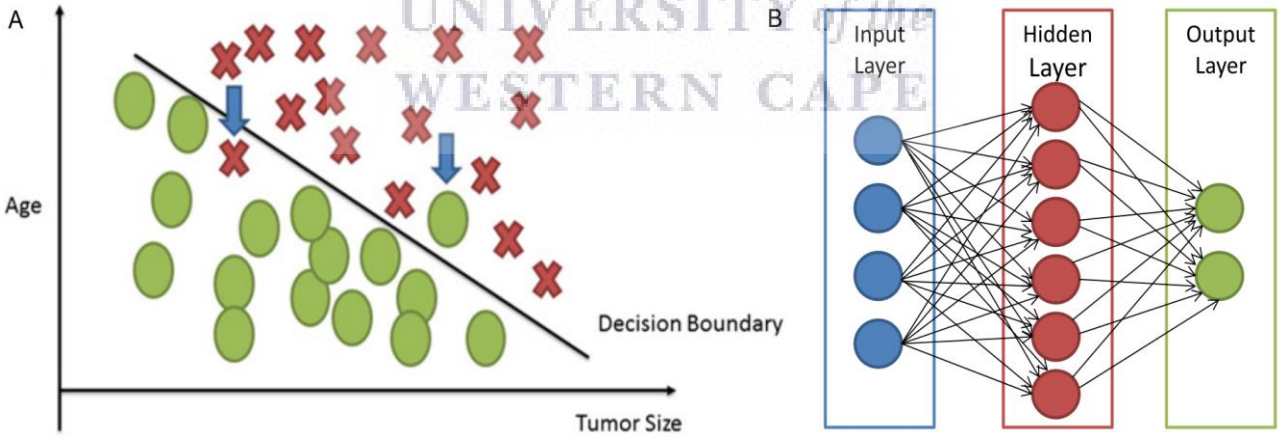


Figure 2.1: Simplified illustrations of the SVM and ANN classification model. (A) SVM classification of input data where tumors are classified according to their size and patient's age. Arrows show the misclassified tumors. (B) An illustration of ANN structure. The arrows connect the output of one node to the input of another. From Kourou et al. (2015).

http://etd.uwc.ac.za/

2.5.2 Unsupervised learning

Unsupervised learning infers inherent patterns in data with unknown labels (Lopez et al., 2018). Clustering and PCA are popular examples of unsupervised learning and are of wide use in bioinformatics research (Xu and Jackson, 2019; Zou et al., 2019). Clustering methods divide input data into clusters based on some common or shared characteristics according to a defined distance measure (Yousef et al., 2014). There are four basic types of clustering, namely exclusive clustering (e.g. K-means algorithm) (MacQueen, 1967), overlapping clustering (e.g. fuzzy c-means algorithm) (Dunn, 1973; Bezdek, 1981), hierarchical clustering (Johnson, 1967) and probabilistic clustering (e.g. expectation-maximization algorithm) (Dempster et al., 1997). Clustering and PCA are powerful tools for data exploration and outlier detection (Witten, 2013).

2.5.3 Semi-supervised learning

Semi-supervised learning is a fusion of both supervised and unsupervised learning and combines unlabelled and labelled data in order to build an accurate learning model (Kourou et al., 2015). It is typically used when there are more unlabelled datasets than labelled sets. Semi-supervised learning models include self-training, mixture models, co-training and multiview learning, graph-based methods and semi-supervised support vector machines (Zhu and Goldberg, 2009).

2.5.4 Machine learning in multi-omics and cancer

ML has been broadly and extensively applied in genomics (Libbrecht and Noble, 2015; Telenti et al., 2018; Koumakis, 2020; Talukder et al. 2021). In addition, different ML approaches were applied in transcriptomics, for example prediction of transcript abundance (Washburn et al., 2019), target gene expression (Chen et al., 2016), gene expression profiles from genotypes (Xie et al., 2017) and tissue

http://etd.uwc.ac.za/

specific expression (Zhou et al., 2018). Other applications of ML in other types of omics data include prediction of regulatory elements and non-coding variant effects from DNA sequence (Zou et al., 2019; Barshai et al., 2020), missing SNPs imputation and DNA methylation states (Angermueller et al., 2017; Sun and Kardia, 2008) and variant calling (Poplin et al., 2018). Prediction in healthcare is very important considering the costs of delayed diagnosis and treatment (Agarwal and Dhar, 2014). Cancer ML techniques are widely applied, becoming a standard approach with a multitude of publications produced as a result (Waddell et al., 2005; Xu et al., 2012; Chen et al., 2014; Gupta et al., 2014; Capper et al., 2018; Li et al., 2021). Fig. 2.2 depicts a classification example in cancer using machine learning. The general application of ML in cancer is in finding and validating potential biomarkers of the underlying pathology that may be useful for diagnosis, prognosis and disease monitoring (Yamada et al., 2019; Matek et al., 2019). The broad aim of this study is to identify gene expression and DNA methylation biomarkers in neuroblastoma.

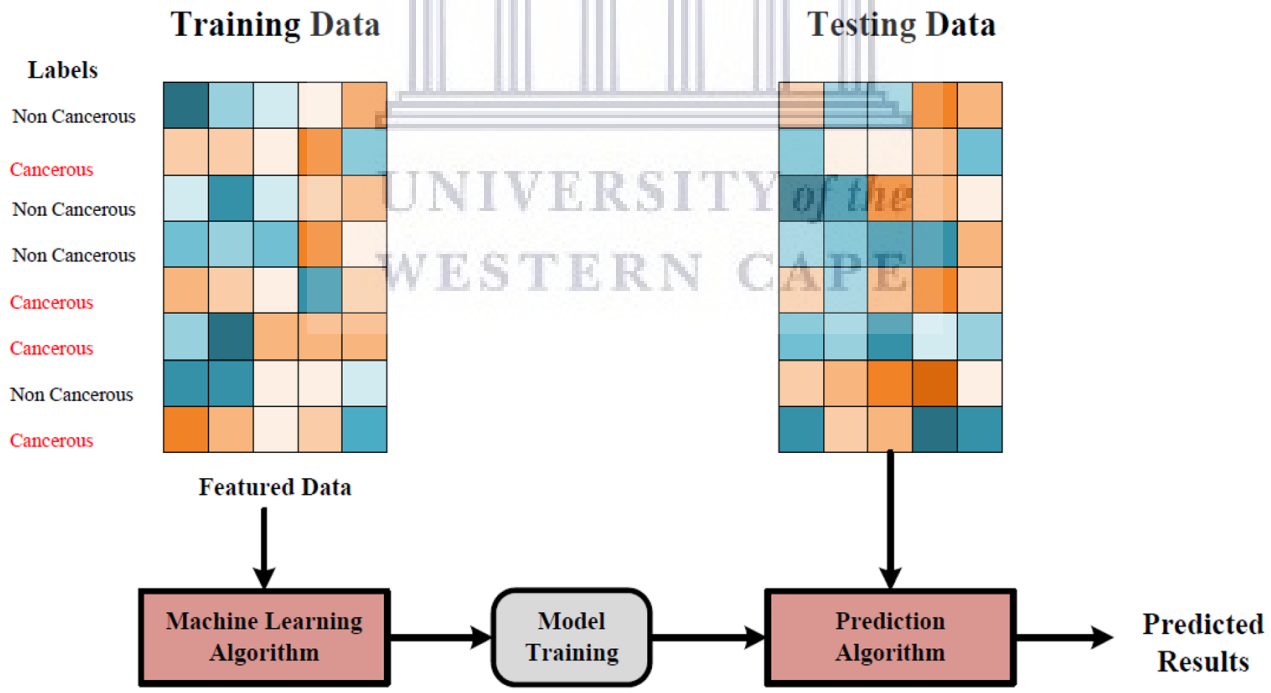


Figure 2.2: Cancer classification using machine learning. A labelled input data is fed to a machine learning algorithm trained to learn the structure inherent in the data (cancerous and non-cancerous). The model then makes prediction of the classes in which items in the testing data belongs. (Sharma and Rani, 2021).

http://etd.uwc.ac.za/

2.6 Biomarkers

Advances in high-throughput sequencing and the molecular characterization of cancer tissue have led to a prominent rise in attempts to measure and target affected pathways at the molecular level (Goossens et al., 2015). These high-throughput techniques allow for the comparison of diseased samples with control samples, which is the most common experimental design for biomarker discovery studies. The processes involved in biomarker development include discovery in research studies, validation and clinical implementation, with the aim of establishing clinically applicable biomarker assays to inform clinical decision-making in improving patient outcomes (Parkinson et al., 2014; Sawyers and van't Veer, 2014).

Bioinformatics plays an important role in the discovery and validation of biomarkers underlying certain abnormal biological states or pharmacological response to a therapeutic intervention (Veytsman and Baranova, 2014). The aim of clinical biomarkers is to empower clinicians with relevant information about the absence or presence of disease, as well as disease and patient characteristics that influence treatment choices (Bock, 2009). Biomarkers are classified into three categories; diagnostic, prognostic and predictive. The diagnostic biomarkers detect presence or absence of disease. Predictive biomarkers provide information about clinical outcome based on treatment decisions, while prognostic biomarkers provide information about the patient's clinical outcome irrespective of therapy, such as disease recurrence or progression (Ballman, 2015).

2.6.1 Biomarker discovery

The quantification of various molecules in biological tissues and fluids is the primary method of finding biomarkers (Veytsman and Baranova, 2014). Cancer biomarkers can be found at the genomic, transcriptomic, proteomic and metabolomic levels. The deregulation of the expression of certain genomic/transcriptomic biomarkers collectively is indicative of a particular underlying

http://etd.uwc.ac.za/

pathophysiology (Bhattacharya and Mariani, 2009). Gene expression technologies (microarray and RNA-Seq) are able to identify the expression of genes at any given disease state, and scientists have employed bioinformatics (see Section 2.4) on gene expression data to discover biomarkers in several types of cancer such as in lung cancer (Li et al., 2019; Sheng et al., 2020), glioblastoma (Fatai and Gamieldien, 2018; Sheng et al., 2020), papillary thyroid carcinoma (Han et al., 2018), nasopharyngeal carcinoma (Ye et al., 2019) and neuroblastoma (Asgharzadeh et al., 2006; He et al., 2020; Wang et al., 2020; Chen et al., 2021). Other technologies used for biomarker discovery include liquid chromatography coupled to tandem mass spectrometry for the search of metabolomic biomarkers, and mass spectrometry for the discovery of protein biomarkers (Dettmer et al., 2007; Lenz and Wilson, 2007). Protein biomarkers reflect the cellular state determined by the expression of a set of common genes (Baumgartner et al., 2011). Generally, the discovery, interpretation and validation of disease biomarkers require innovations in high throughput techniques, biostatistics and bioinformatics.

2.6.2 Biomarkers in clinic

UNIVERSITY of the
WESTERN CAPE

The discovery and validation of biomarkers is an important prerequisite for the development of anticancer drugs (Sarker and Workman, 2007). They are assessed by clinical trials after their validation. Although there have been thousands of potential biomarkers in the published literature, few have been translated to clinical use. Some adduced reasons include the requirement for a large interdisciplinary effort and funding challenges. Clinically approved biomarkers include Mammaprint (70-gene expression) (Van de Vijer et al., 2002) and Oncotype Dx (21-gene expression) (Paik et al., 2004) prognostic biomarker tests for recurrence prediction in breast cancer patients. ClonoSEQ, an NGS clinical test, is used for the detection of minimal residual disease in patients with acute lymphoblastic leukemia and multiple myeloma (Hristova and Chan, 2019). Cologuard is a multigene stool DNA test for colorectal cancer screening (Imperiale et al., 2014; Bering et al., 2017). Some other

http://etd.uwc.ac.za/

biomarkers approved for clinical use include; Prostate Specific Antigen (PSA) for Prostrate cancer screening (Bell et al., 2014; Tikkinen et al., 2018), *MYCN* amplification in neuroblastoma (Brodeur et al., 1984; Seeger et al., 1985), α-fetoprotein (AFP) in liver and testicular cancer (Kew, 1974; Wang and Wang, 2018), *CA125* antigen in ovarian cancer (Felder et al., 2014; Dochez et al., 2019; Bonifacio, 2020), MGMT methylation for Glioblastoma (Iafrate and Louis, 2008; Mikeska et al., 2007) and *BRCA1* and *BRCA2* in breast cancer (Toland et al., 2018). There is currently a paucity of clinically approved biomarkers in neuroblastoma.

2.7 Neuroblastoma

2.7.1 Origin

Neuroblastoma is an embryonal tumor originating from the sympathoadrenal lineage of the neural crest during development. The sympathetic nervous system comprises the sympathetic chain and truncus ganglia, paraganglia and the adrenal gland (De Preter et al., 2006). Neuroblastoma tumors can develop anywhere along the sympathetic trunk from the neck to the pelvis, within the paravertebral sympathetic ganglia and the adrenal gland (Delloye-Bourgeois and Castellani, 2019). Due to the heterogeneity of neuroblastoma, it has been suggested that it is possible that neuroblastoma also originated from earlier neural crest (NC) derivatives before the development of the sympathethoadrenal lineage but after the initial fate specification (Brodeur, 2003; Maris et al., 2007). Kerosuo et al. (2018) noted that for some forms of aggressive neuroblastoma, some priming event may have occurred before neural emigration from the central nervous system and well before sympathoadrenal specification. Neuroblastoma is thought to be a result of a blockage or delay in development of the NC lineage. Neuroblastoma mainly develops in the abdomen (65%), more often in the adrenal medulla, and in other sites such as the chest (20%), pelvis (5%) and neck (5%) (Maris et al., 2007).

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Neural crest cells (NCCs) are derived from the embryonic ectoderm and are unique to vertebrates. They are multipotent, highly migratory cells arising in the neural folds at the border between neural and non-neural ectoderm in early embryos (Theveneau and Mayor, 2012; Gammill and Bronner-Fraser, 2003). One of the genes important for NCC development is *MYCN*, which has been found to be an oncogenic driver in neuroblastoma (Otte et al., 2021). The expression of *MYCN* begins in migrating NCCs during the later stages of embryonic development and its main function is to preserve the proliferative and self-renewal abilities of NCCs (Huang and Weiss, 2013). Its expression increases during sympathoadrenal development and decreases during the differentiation process (Marshall et al., 2014). Olsen et al. (2017) succeeded in transforming primary NCCs into neuroblastoma by overexpression of *MYCN* with the resulting tumors phenotypically and molecularly similar to *MYCN*-amplified neuroblastoma. In addition, Schulte et al. (2013) showed that enforced expression of *MYCN* in NCCs is enough to drive a neuroblastoma-like phenotype. As such, amplification and overexpression of *MYCN* is important for neuroblastoma development and progression. Amplification of *MYCN* allows pre-neuroblastoma cells to escape apoptosis (Otte et al., 2021). Furthermore, it is a strong indicator of poor prognosis in high-risk neuroblastoma (Brodeur et al., 1984; Seeger et al., 1985). Other genomic alterations including gene polymorphisms and mutations are implicated in neuroblastoma.

UNIVERSITY of the
WESTERN CAPE

2.7.2 Genomic alterations in neuroblastoma

Neuroblastoma is mainly a sporadic disease with only about 1-2% of cases being familial/genetic (Knudson and Strong 1972, Kushner et al., 1986, Dodge 1945, Chompret et al., 1998). It is inherited in an autosomal dominant fashion with the *PHOX2B* (Trochet et al., 2004) and *ALK* (Mosse et al., 2008) genes implicated. Induction of neuronal differentiation with retinoic acid has been shown to downregulate expression of *PHOX2B*, confirming the role of *PHOX2B* in the regulation of neuroblastoma differentiation and stemness maintenance (Yang et al., 2016). *ALK* is a receptor

http://etd.uwc.ac.za/

tyrosine kinase expressed in the nervous system (Hurley et al., 2006; Degoutin et al., 2009) for which activating mutations have been discovered for both sporadic and familial neuroblastoma cases (George et al., 2008; Chen et al., 2008; Mosse et al., 2008; Janoueix-Lerosey et al., 2008).

Results from Genome Wide Association Studies (GWAS) have shown that sporadic neuroblastoma ensues from the interaction of multiple common, low penetrance risk alleles (Deyell and Attiyeh, 2011). However, there is a relative paucity of somatic mutations in neuroblastoma. Pugh et al. (2013) reported few recurrently mutated genes in a small proportion of cases, including *ALK*, *ATRX*, *MYCN*, *NRAS* and *PTPN11,* as well as rare potentially pathogenic variants for the *ALK, CHECK2, PINK1* and *BARD1* genes. *ALK* is the most common somatically mutated gene in neuroblastoma seen in approximately 9% and 14% of cases, respectively, in primary neuroblastoma tumors and high-risk neuroblastoma (Mosse et al., 2008; Janoueix-Lerosey et al., 2008; Bresler et al., 2014, Caren et al., 2008; Chen et al., 2008), and such alterations portends a poorer prognosis (Bresler et al., 2014). Also, *TERT* rearrangements (Valentijn et al., 2015; Peifer et al., 2015) and *DAXX* alterations (Cheung et al., 2012; Peifer et al., 2015; Kurihara et al., 2014) have been reported in neuroblastoma, associating with poorer prognosis. In addition, polymorphisms in genes including *BARD1*, *LMO1*, *CASC15*, *MMP20*, *DDX1* have been found to increase susceptibility to neuroblastoma (Capasso et al., 2009; Wang et al., 2011; Russell et al., 2015; Chang et al., 2017; Jin et al., 2020).

Chromosomal instability has been proposed to be the primary driver of neuroblastoma oncogenesis (Ciriello et al., 2013; Tonini, 2017). It has been observed that low-risk and intermediate-risk tumors with good prognosis exhibit several numerical copy number variations (CNVs) while tumors from high-risk patients with poor prognosis display several structural CNVs (Coco et al., 2012; Fusco et al., 2018). The chromosomes of stage 4 tumor cells exhibit a variety of damage, including rearrangements, chromothripsis and deletions (Molenaar et al., 2012; Tonini and Capasso, 2020). The most frequently observed segmental chromosomal alterations in neuroblastoma include losses of

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

chromosome 1p, 3p, 4p, 11q and gains in chromosome 1q, 2p (*MYCN*, *ALK*), 17q (Schleiemacher et al., 2014). These alterations can lead to changes in the expression of several genes possibly causing the loss of cell cycle control as well as other necessary cell functions (Coco et al., 2012; Stigliani et al., 2012). Gain of 17q is the most frequent copy number abnormality and is associated with more aggressive disease and decreased survival (Lastowska et al., 1997; Bown et al., 1999). In addition, the main copy number amplification observed in neuroblastoma is *MYCN* amplification which is a predictor of poor prognosis (Brodeur et al., 1984; Seeger et al., 1985).

2.7.3 Neuroblastoma and *MYCN*

The *MYCN* gene belongs to the MYC transcription factor family of proto-oncogenes (Ruiz-Pérez et al., 2017). It is located in the 2p24.3 locus and was first identified in neuroblastoma by Schwab et al. (1983). Expression of *MYCN* is restricted to the nervous system, lung, kidney and spleen during embryo development (Galderisi et al., 1999). It interacts with DNA through the E-box consensus sequence (CACGTG) and heterodimerizes with *MAX* to facilitate transcription of many target genes (Chayka et al., 2015, Gherardi et al., 2013). These target genes are involved in cell proliferation, growth, cycle and apoptosis (Huang and Weiss, 2013). The amplification of *MYCN* was shown to be linked to progressing disease and poor outcome in neuroblastoma (Brodeur et al., 1984; Seeger et al., 1985) and has remained the most reliable prognostic marker in this disease.

The role of *MYCN* as a direct cause of neuroblastoma was established with the demonstration of development of neuroblastoma in mice through its targeted misexpression in the peripheral neural crest (Weiss et al., 1997). Its inhibition causes regression of neuroblastoma *in vitro* and *in vivo* (Burkhart et al., 2003), reverses stem-like phenotype (Kang et al., 2006) and stops aberrant proliferation (Tweddle et al., 2001, Muth et al., 2010, Bell et al., 2007, Yaari at al., 2005). *MYCN* amplified cells are often undifferentiated or poorly differentiated (Dang, 2012; Adhikary and Eilers, 2005). *MYCN* blocks differentiation pathways and maintains pluripotency (Cotterman and Knoepfler,

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

2009). In addition, it prevents neuronal differentiation by repressing transcription of genes such as *ERα* (Dzieran et al., 2018), *CDKL5* (Valli et al., 2012) and *NLRR3* (Koppen et al., 2007; Akter et al., 2011). The expression of ERα correlates with clinical markers for good prognosis (Dzieran et al., 2018). A 157-gene signature directly correlating with *MYCN* mRNA levels but not with *MYCN* amplification was identified by Valentijn et al. (2012). Twenty-one of these genes were involved in neuronal differentiation and were downregulated (Valentijn et al. 2012).

The tumorigenic effect of *MYCN* promoting cell proliferation and cell cycle progression is well characterized. *MYCN* is involved in all metastatic processes including adhesion, motility, extracellular matrix (ECM) degradation and invasion (Zaizen et al., 1993; Goodman et al., 1997; Cohn et al., 2009). It facilitates endothelial growth by causing the loss of endothelial growth inhibitors which promotes angiogenesis (Fotsis et al., 1999). Furthermore, *MYCN* allows neuroblastoma cells escape the immune system by downregulating Major Histocompatibility Complex (MHC) molecules (Versteeg et al., 1990) and inhibiting Natural Killer T (NKT) cells causing upregulation of tumor associated macrophages (Song et al., 2007, 2009). Tumor-associated macrophages are cells that create an immunosuppressive tumor microenvironment that promotes and aids metastasis (Lin et al., 2019; Yang et al., 2020). Neuroblastoma tumors with *MYCN* amplification are characteristically aggressive and therapy resistant (Hogarty, 2003). This amplification phenotype is also associated with increased expression of *MRP1* gene which is implicated in chemotherapy failure (Haber et al., 1999, 2006).

High expression of Survivin (*BIRC5*) is reported in advanced stage neuroblastoma in conjunction with *MYCN* amplification and is indicative of poor prognosis (Bown et al., 1999; Miller et al., 2006; Bell et al., 2010). *MYCN* increases the expression of *TFAP4* whose higher expression is linked with poorer outcome in neuroblastoma patients (Xue et al., 2016). In addition, *MYCN* upregulates these miRNAs; *miR-221*, *miR-9* and *miR-350-5p*, which have an oncogenic function (Schulte et al., 2008, Swarbrick et al., 2010; Ma et al., 2010). For example, *miR-9* suppresses E-cadherin, thus contributing

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

to epithelial-mesenchymal transition (EMT) (Ma et al., 2010), a process essential for tumor invasion (Yeung and Yang, 2017; Mittal, 2018). RNA interference or use of antisense oligonucleotides to silence expression of *MYCN* causes apoptosis, differentiation and suppression of tumor growth in neuroblastoma (Burkhart et al., 2003; Gustafson and Weiss, 2010). *MYCN* epigenetically regulates global gene expression through DNA methylation as well as the recruitment of histone acetyltransferase complexes that keep chromatins active (Frank et al., 2003; Knoepfler et al., 2006; Beltran, 2014).

2.7.3.1 Therapeutic targeting of *MYCN*

There are no clinical drugs available that directly target *MYCN* despite its discovery in neuroblastoma 38 years ago. This is because *MYCN* is difficult to block directly due to its structural specifications wherein it lacks appropriate surfaces where drugs can bind to (Bielinsky, 2015; Futami and Sakai, 2010) and its tertiary structure which is variable in solution (Moreno et al., 2020). Instead, research has been geared towards indirectly targeting it through its binding partners and downstream targets (Moreno et al., 2020). Zhang et al. (2015) demonstrated that forced overexpression of *miR-375* inhibits the translation of *MYCN's* mRNA through an IRES-dependent mechanism showing that *miR-375* negatively regulates *MYCN*. Schramm and Lode (2016), explored the possibility of targeting *MYCN* through DNA vaccination. Oncolytic viral suppression of *MYCN* amplified neuroblastoma cells was also demonstrated by Tanimoto et al. (2020). Finally, Yoda et al. (2019) developed a novel DNA-alkylating pyrrole-imidazole polyamide (MYCN-A3) which targets *MYCN* by directly binding to and alkylating DNA at homing motifs within the *MYCN* transcript. These novel approaches requires further validation before they can be certified for clinical use.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

2.7.4 Epigenetics

Epigenetics covers all chemical changes that can alter gene expression without changing the DNA sequence. These include DNA methylation, histone modification and RNA-associated silencing. Tumor development is a multi-step process comprising multiple epigenetic and genetic alterations (Macaluso et al., 2003; Coyle et al., 2017). Proteins that regulate the epigenome have been found to have mutations in almost all cancer types including neuroblastoma (Baylin and Jones, 2011; Shen and Laird, 2013; Garraway and Lander, 2013; You and Jones, 2012). The epigenome, during tumorigenesis, undergoes multiple alterations such as global changes in histone modification marks (Kanwal and Gupta, 2012; Fraga et al., 2005; Hosseini and Minucci, 2018), deregulation of non coding RNA networks (Liz and Esteller, 2015; Esteller, 2011), global DNA methylation loss and CpG promoter islands hypermethylation of tumor suppressor genes (Baylin and Jones, 2016; Fernandez et al., 2012). In neuroblastoma, CpG promoter islands hypermethylation acts by regulating the expression of tumor suppressor genes including *RASSF1A* and *BLU*, and correlates with poor outcome (Asada et al., 2013; Abe et al., 2005, 2007, 2008; Banelli et al., 2012, 2013).

2.7.4.1 DNA methylation

UNIVERSITY of the
WESTERN CAPE

This was the first epigenetic alteration identified in cancer (Feinberg and Vogelstein, 1983; Riggs and Jones, 1983) which is now established to be common. It is involved with bioprocesses such as repression of transcription, stem cell differentiation, inactivation of transposons and genomic imprinting (Smith and Meissner, 2013). Methylation of DNA generally results in transcriptional repression while unmethylation of DNA allow for gene transcription. The well known type of DNA methylation is the transfer of a methyl group from D-adenosyl-L-methionine to cytosine at the 5th position of its carbon ring (5-mc) especially at CpG sites (Cheng et al., 2019, Rauluseviciute et al., 2019). This chemical process is catalyzed by enzymes, called DNA methyltransferases (DNMTs) which are four including *DNMT1*, *DNMT3a*, *DNMT3b* and *DNMT3L*. The expression of DNMTs are

http://etd.uwc.ac.za/

altered in neuroblastoma (Jubierre et al., 2018). Several advantages qualify DNA methylation biomarkers for broad use as in vitro diagnostics such as its cell type specificity, its binary measurement (each CpG is methylated or unmethylated), the availability of infrastructure for assaying them in many clinical diagnostics laboratories, ease of integration into routine clinical workflows and DNA methylation patterns are retained for a long time in well-stored samples (Bock et al., 2016). Chapter 4 of this thesis demonstrates the utility of DNA methylation in neuroblastoma tumor segregation and prediction of *MYCN* amplification.

2.7.4.2 Histone modifications

The compaction of large amounts of DNA in eukaryotic cells is mediated by histone proteins in which the DNA is wrapped. The wrapping of the DNA in the histone octamers forms a nucleosome which are assembled to form chromosomes. The nucleosome builds chromatin which can exist as either the transcriptionally active form, euchromatin or the transcriptionally inactive form, heterochromatin. The compaction of chromatin is regulated by modifications such as methylation, acetylation, phosphorylation and sumoylation on the histone tails (Jubierre et al., 2018). The enzymes involved include writers which add these modifications, erasers which remove these modifications and readers which recognize these modifications (Jubierre et al., 2018). These enzymes include histone deacetylases (HDACs), histone acetyltransferases (HATs) and histone demethylases (HDMs). The effect of altered expression of some of these enzymes in neuroblastoma have been investigated by several researchers. Overexpression of *HDAC8* significantly correlated with poor prognosis (Oehme et al., 2009). It was also shown that the neuroblastoma oncogene *MYCN* cooperates with HDACs to downregulate their target genes (Fabian et al., 2016; Lodrini et al., 2013).

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

2.7.4.3 Micro RNAs

MicroRNAs (miRNA) are non-coding RNAs approximately 22 nucleotides in length. They are part of the epigenetic system as they modulate gene expression without changing the DNA sequence. They act by base pairing with the 3'UTRs of mRNAs thereby negatively regulating gene expression. In cancer, miRNA can function as oncogenes or tumor suppressors depending on the function of their target genes (Zhang et al., 2007, Ruan et al., 2009; Su et al., 2015). They play an important role in neuroblastoma pathology (Stallings, 2009). Some have been identified as being epigenetically regulated in malignant neuroblastoma, such as *miR-184*, *let-7*, *miR-9*, *miR-101*, *miR-340*, *miR-34a*, *miR-335* and *miR-202* (Romania et al., 2012; Das et al., 2013). Parodi et al. (2016) identified hypermethylated and downregulated miRNAs (miR-34b-3p, miR-34b-5p, miR-34c-5p and miR-124-2-3p) involved in apoptosis, cell cycle regulation, and in *MYCN* expression regulation. Twenty-one miRNA prognostic marker candidates in high-risk neuroblastoma were identified by Ramraj et al. (2016). Zhao et al. (2018) identified miR-2110 to possess oncosuppressive functions in neuroblastoma. These results therefore, demonstrate the importance of miRNAs as therapeutic targets in neuroblastoma.

UNIVERSITY of the
WESTERN CAPE

2.7.4.4 The epigenome as a therapeutic target

Unlike genetic mutations, epigenetic alterations are potentially reversible to their normal state (Sharma et al., 2010). The aim of epigenetic therapies is therefore to make the affected cells revert to a more normal state. The three main families of epigenetic proteins; readers, erasers and writers can be targeted therapeutically. Epigenetic therapies are an emerging form of therapy for overcoming drug resistance (Strauss and Figg, 2016; Lu et al., 2020; Quagliano et al., 2020; Keyvani-Ghamsari et al., 2021). Most epigenetic drugs act at the level of DNA methylation, histone modifications and blockage of interpretation of epigenetic modifications (Jubierre et al., 2018). DNA methylation can be modulated by targeting DNMTs while histone modifications can be targeted by inhibiting the

http://etd.uwc.ac.za/

enzymes involved in histone modifications. Epigenetic drugs are classified into; Histone deacetylase inhibitors (HDACIs), DNA methyltransferase inhibitors (DNMTIs) and non-coding RNA (ncRNA) based therapeutics (Lu et al., 2020).

2.7.5 Therapy

Treatment for neuroblastoma depends on the risk and disease stage. Low-risk neuroblastoma (Stage 1, 2A, 2B disease) are usually treated with surgery alone, but chemotherapy could also be included for patients with non-metastatic neuroblastoma who exhibit symptoms of their organs being threatened at diagnosis (Park et al., 2008). Treatment of intermediate-risk neuroblastoma involves surgery and moderate multiple dose chemotherapy. Children with tumors that exhibit favorable characteristics have greater than 95% survival chance (Matthay et al., 1998; Schmidt et al., 2000).

High-risk neuroblastoma poses the biggest challenge because despite multimodal and improved therapy, only 30-40% of patients survive long-term (Matthay et al., 1999), which accounts for 12% of pediatric cancer deaths (Smith et al., 2010). Neuroblastoma demonstrates high cell heterogeneity, evades the immune system, thereby limiting the effectiveness of the approaches currently used (Kholodenko et al., 2018). Standard therapy involves three phases; induction, consolidation and post-consolidation therapy. The induction phase aims to reduce the amount of disease not only at the primary sites but also at metastatic sites and includes multi-agent chemotherapy and surgery (Armideo at al., 2017). In the consolidation phase, myeloablative chemotherapy is applied alongside stem cell rescue and external-beam radiation therapy (Greengard et al., 2013). The post consolidation phase aims to eliminate residual disease that might exist after intensive treatment, and includes immunotherapy (ch14.18 monoclonal antibody against GD2), granulocyte macrophage stimulating factor (GM-CSF), interleukin-2 (IL-2)) and a differentiating agent (isotretinoin) (Greengard et al., 2013). Key cytotoxic agents for neuroblastoma include doxorubicin, vinctistine, carboplatin, cyclophosphamide, cisplatin, ifosfamide and etoposide (Berthold et al., 2017; Berlanga et al., 2017).

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

Significant side effects and toxicity of standard neuroblastoma therapy include anemia, internal organ damage, fertility issues and hair loss (Kholodenko et al., 2018). Newer therapeutic approaches such as the use of oncolytic viruses (Tanimoto et al., 2020) are being developed, but their clinical safety and validation still needs to be demonstrated before their acceptance.

2.8 Summary

High-risk neuroblastoma presents with poor outcome even with intensive and varied therapy (Zhang et al., 2018), and patients in the same risk group may have a variable prognosis. Current risk-stratification schemes do not account for an ultra-high-risk subgroup within the high-risk group. Using gene expression data, studies have proposed expression signatures of 238 genes (Liu et al. 2020), 27 genes (Russo et al., 2017), and 18 genes (Formicola et al., 2016) for the identification of ultra-high-risk groups. Finding a gene expression signature of fewer number of genes will have the potential to be cost effective and easy to implement. Furthermore, identifying sub-groups within the high-risk group with the worst prognosis should be a good strategy to optimize treatment decisions (This is the basis of Chapter 3).

UNIVERSITY of the
WESTERN CAPE

MYCN amplification in neuroblastoma is a predictor of poor prognosis (Brodeur et al., 1984; Seeger et al., 1985), although patients without it may have poor outcome (Baali et al., 2018). Its diagnosis is a clinical standard in the management of neuroblastoma. The current diagnostic technique is by fluorescence in situ hybridization technique (FISH) which is technically difficult, expensive and not readily available in low- and middle-income countries (Nwose, 2016). It is therefore important to develop easily accessible and less expensive potential diagnostic techniques to detect *MYCN* amplification. In the future, a simpler and faster blood test to detect *MYCN* amplification could potentially be developed. We attempted to discover DNA methylation biomarkers diagnostic of the amplification of *MYCN* and prognostic of poor outcome which, in future research, could potentially

http://etd.uwc.ac.za/

lead to the development of a simple blood test procedure. As it is also difficult to directly target *MYCN* for therapeutic purposes, it is pertinent to uncover the genomic signatures associated with its amplification as this could reveal the underlying mechanisms and provide possible candidates for indirect targeting of *MYCN*. We attempted to find these possible candidates using DNA methylation data (Chapter 4).

http://etd.uwc.ac.za/

Chapter 3

Identification of novel prognostic markers of survival time in high-risk neuroblastoma using gene expression profiles

Published Paper

UNIVERSITY of the
WESTERN CAPE

Abdulazeez Giwa, Azeez Fatai, Junaid Gamieldien, Alan Christoffels, Hocine Bendou. Identification of novel prognostic markers of survival time in high-risk neuroblastoma using gene expression profiles. Oncotarget. 2020;11(46):4293-4305. https://doi.org/10.18632/oncotarget.27808. [PubMed: 33245713]

3.1 Abstract

Neuroblastoma is the most common extracranial solid tumor in childhood. Patients in high-risk group often have poor outcomes with low survival rates despite several treatment options. This study aimed to identify a genetic signature from gene expression profiles that can serve as prognostic indicators of survival time in patients of high-risk neuroblastoma, and that could be potential therapeutic targets. RNA-seq count data was downloaded from UCSC Xena browser and samples grouped into Short

http://etd.uwc.ac.za/

Survival (SS) and Long Survival (LS) groups. Differential gene expression (DGE) analysis, enrichment analyses, regulatory network analysis and machine learning (ML) prediction of survival group were performed. Forty differentially expressed genes (DEGs) were identified including genes involved in molecular function activities essential for tumor proliferation. DEGs used as features for prediction of survival groups included *EVX2, NHLH2, PRSS12, POU6F2, HOXD10, MAPK15, RTL1, LGR5, CYP17A1, OR10AB1P, MYH14, LRRTM3, GRIN3A, HS3ST5, CRYAB and NXPH3*. An accuracy score of 82% was obtained by the ML classification models. *SMIM28* was revealed to possibly have a role in tumor proliferation and aggressiveness. Our results indicate that these DEGs can serve as prognostic indicators of survival in high-risk neuroblastoma patients and will assist clinicians in making better therapeutic and patient management decisions.

3.2 Introduction

Neuroblastoma is the most common extracranial solid tumor in childhood accounting for approximately 15% of pediatric cancer death (Maris, 2010; Smith et al., 2010; Ward et al., 2014). It develops anywhere along the sympathetic nervous system, with 60% of the tumors occurring in the abdomen, commonly in the adrenal gland (Zhang et al., 2018; Johnsen et al., 2019). Outcomes ranging from spontaneous regression to relentless progression despite extensive therapies indicate the heterogeneity of neuroblastoma (Baali et al., 2018). The Children's Oncology Group (COG) classifies neuroblastoma patients into low, intermediate and high-risk groups. Patients classified in low risk groups have good outcomes contrary to high-risk groups who present poor outcomes despite extensive therapies (Zhang et al., 2018) and with a disproportionate number dying or suffering profound treatment related morbidities (Oeffinger et al., 2006; Yu et al., 2010). Tumors in high-risk neuroblastoma patients are often metastatic, resulting in survival rates of less than 50% (Maris, 2010).

Genomic studies associated high-risk neuroblastoma with mutations or alterations in a number of genes, such as *ALK*, *ATRX* and *TERT* (Mosse et al., 2008; Cheung et al., 2012; Valentijn et al., 2015;

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Peifer et al., 2015). Furthermore, genome wide association studies have revealed genetic markers, such as *CASC15*, *LMO1*, *DUSP2* and *BARD1* to be neuroblastoma susceptibility genes (Maris et al., 2008; Capasso et al., 2009; Capasso et al., 2014; Diskin et al., 2009; Wang et al., 2011; Bosse et al., 2012; Oldridge et al., 2015; Russell et al., 2015). However, these genes do not provide information about patient survival. The objective of our study is to identify a genetic signature from gene expression data that can serve as prognostic indicators of survival time in high-risk neuroblastoma patients and that could be therapeutic targets in that patient category.

3.3 Methods

The workflow describing the steps and methods undertaken in this study is illustrated in Figure 3.1. It includes five essential steps; dataset retrieval, differential gene expression, disease/gene ontology enrichment, gene regulatory network inference and machine learning.

UNIVERSITY of the
WESTERN CAPE

3.3.1 Datasets

The Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative employed comprehensive molecular characterization of hard-to-treat childhood cancers which included neuroblastoma. TARGET data is accessible via the TARGET data matrix as well as via the Xena browser. The Xena browser is a web-based visualization and exploration tool for multi-omic data large public repositories and private datasets (Goldman et al., 2020). The TARGET neuroblastoma dataset in the Xena database (Vivian et al., 2017; Goldman et al., 2020) is composed of high-risk neuroblastoma samples with available clinical information. Gene expression RNA-Seq read counts of the TARGET neuroblastoma dataset (dataset ID: TARGET-NBL.htseq_counts.tsv) were obtained from the GDC hub in Xena browser using xenaPython package. Fields used in querying the dataset are described in Table 3.1. The dataset itself was composed of 151 samples in total.

http://etd.uwc.ac.za/

Querying the dataset with the above fields returned 32 neuroblastoma samples; 20 of which with an overall survival time of less than 730 days (2 years) and vital status is “dead”, were considered short survival (SS), while 12 samples with overall survival time greater than 2555 days (7 years) and vital status is “alive” were considered long survival (LS).

Table 3.1: Query criteria for short survival (SS) and long survival (LS) sample selection

Field	**SS values**	**LS values**
Diagnostic category	Neuroblastoma	Neuroblastoma
INSS stage	Stage 4	Stage 4
COG risk group	High risk	High risk
Vital status	Dead	Alive
Overall survival time	730 days	2555 days

UNIVERSITY of the
WESTERN CAPE

3.3.2 Differential Gene Expression analysis

Normalized expression counts were converted to raw counts and filtered to remove low expressed genes using the edgeR filterByExpr function in R (Robinson et al., 2010). We then performed a differential gene expression (DGE) analysis between the short and long survival groups using the DESeq2 package in R (Love et al., 2014). Differentially expressed genes (DEGs) were selected by meeting criteria of adjusted *p-value* < 0.05. Gene Ontology (GO) enrichment analysis was carried out to functionally annotate the DEGs using clusterProfiler (Yu et al., 2012) and visualized using enrichplot. The DOSE R library (Yu et al., 2015) was used to detect the diseases enriched by the upregulated and downregulated DEGs.

http://etd.uwc.ac.za/

3.3.3 Machine learning

Scikit-learn (Pedregosa et al., 2011) and LibSVM (Chang and Lin, 2011) libraries were used for machine learning (ML) model creation and classification tasks. Features (genes) not present in the test dataset (GSE49711) were not used for machine learning prediction. The top upregulated and downregulated genes used as features for ML prediction of patient survival were; *EVX2, NHLH2, PRSS12, POU6F2, HOXD10, MAPK15, RTL1, LGR5, CYP17A1, OR10AB1P, MYH14, LRRTM3, GRIN3A, HS3ST5, CRYAB, NXPH3*. The features were extracted from the log-normalized counts data. Algorithms used for training and evaluation of the models include; Support Vector Machines (SVM) and Artificial Neural Networks (ANN). For LibSVM, the feature values of the training and test sets were scaled using a built-in python script. With regards to ANN, the feature values of the training and test sets were scaled with the Scikit-learn MinMaxScaler function. 5-fold cross validation was done to determine the best parameters of the SVM and ANN models which was then applied to our test samples. Both models, created with LibSVM and ANN were applied to predict the classification of samples in an external dataset (GSE49711) with same sample characteristics (overall survival < 730 days with vital status as dead for SS samples, and overall survival > 2555 days with vital status as alive for LS samples). The evaluation metrics for the LibSVM and ANN models were precision, recall and accuracy.

UNIVERSITY of the
WESTERN CAPE

3.3.4 Regulatory network analysis

The GENIE3 package (Huynh-Thu et al., 2010) in R was used for genetic regulatory network inference analysis. The GENIE3 algorithm uses a Random Forest or Extra Randomized Trees approach to infer gene regulatory networks from gene expression data (Huynh-Thu et al., 2010). It outputs a ranked list of each pairwise comparison from the most to the least confident regulatory connection. The library was run on the gene expression counts from the SS and LS samples separately. For the output analysis, only connections involving the DEGs were considered. In addition, a

http://etd.uwc.ac.za/

weighting threshold of 0.00251 was applied to reduce the large number of connections and also to focus on high confident regulatory connections. Cytoscape (Shannon et al., 2003) was used for the visualization of gene regulatory networks.

3.3.5 Availability of data and materials

The R and python scripts implemented for this study are available at https://github.com/SANBI-SA/NBSurvival.

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

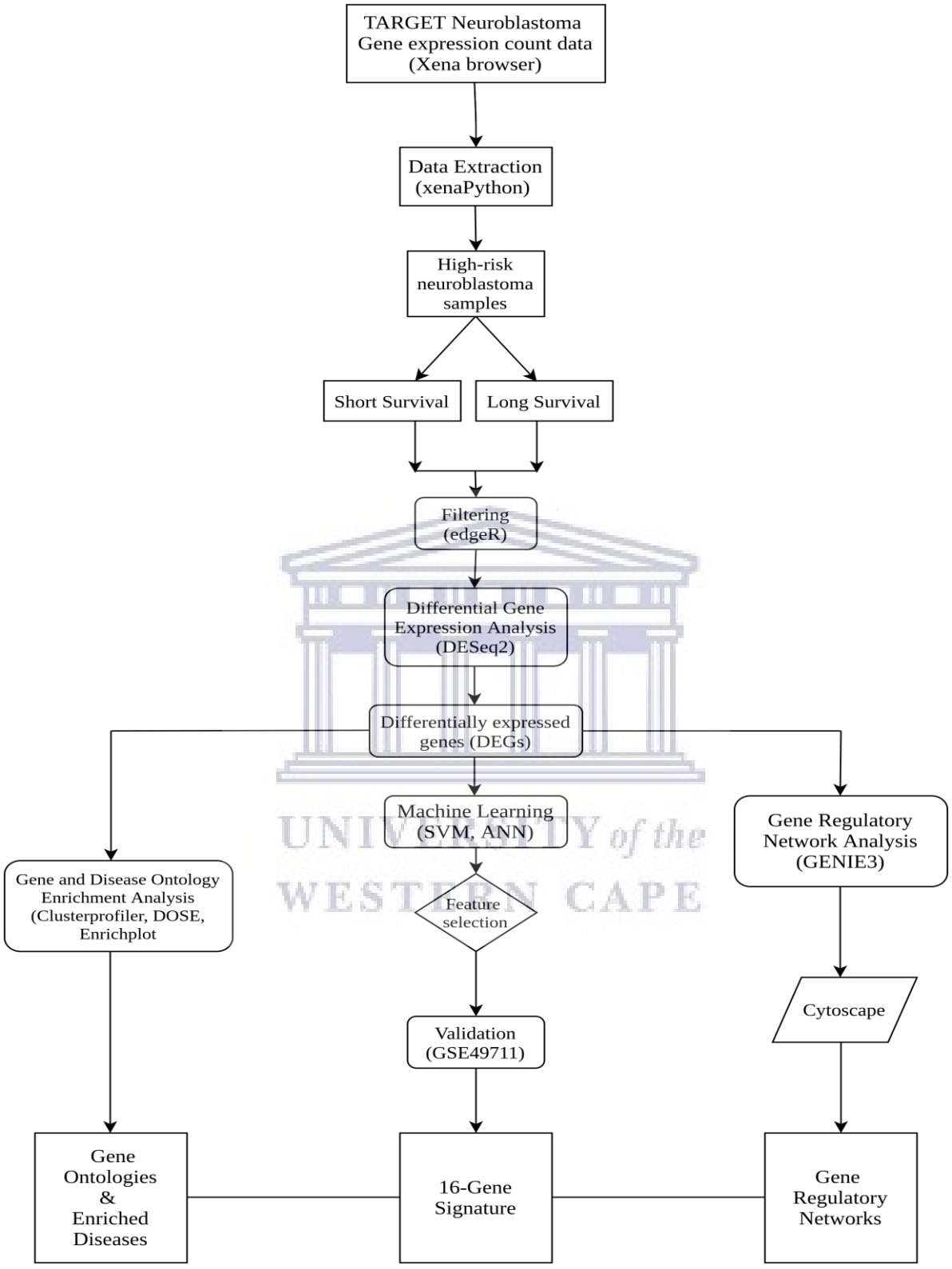


Figure 3.1: Workflow outlining the steps and methods undertaken in this study

http://etd.uwc.ac.za/

3.4 Results

Querying the Xena TARGET dataset returned 20 and 12 SS and LS samples, respectively. Based on the gene expression levels in these samples, the edgeR filterByExpr function removed 35,873 low expressed genes and kept 24,610 genes for downstream analysis. The DGE analysis with DESeq2 identified 40 DEGs between the SS and LS groups, of which 21 genes were upregulated and 19 genes were downregulated. Table 3.2 shows information about the 40 DEGs.

The Gene Ontology (GO) Molecular Function enrichment analysis revealed that upregulated genes were mainly enriched in MAP kinase activity, retinol binding and RNA polymerase II activating transcription factor binding, as well as, in other activities shown in (Fig. 3.2A). No statistically significant results (p-adjusted value < 0.05) were obtained for the downregulated genes as well as for the other GO categories; Biological Process and Cellular Component. In addition, the Disease Ontology enrichment analysis associated upregulated and downregulated genes with several genetic disorders; intellectual disability, cardiac dysfunction, bone development, impaired infertility and pulmonary dysfunction caused by diaphragm-associated abnormalities (Fig. 3.2B-C).

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

Table 3.2: Upregulated and downregulated DEGs in SS neuroblastoma samples compared to LS samples

Symbol	Name	log2FC	p-adj	Status
SMIM28	small integral membrane protein 28	5.4652	0.0146	Up
EVX2	even-skipped homeobox 2	4.9304	0.0014	Up
NHLH2	nescient helix-loop-helix 2	4.2404	0.0046	Up
PRSS12	serine protease 12	3.4788	0.0168	Up
POU6F2	POU class 6 homeobox 2	3.4534	0.0430	Up
HOXD10	homeobox D10	3.3386	0.0357	Up
MAPK15	mitogen-activated protein kinase 15	3.0939	0.0499	Up
RTL1	retrotransposon Gag like 1	2.8233	0.0241	Up
LGR5	leucine rich repeat containing G protein-coupled receptor 5	2.7453	0.0386	Up
DPY19L2P4	DPY19L2 pseudogene 4	2.6439	0.0187	Up
STRA6	signaling receptor and transporter of retinol STRA6	2.5625	0.0437	Up
MEG9	maternally expressed 9	1.9449	0.0146	Up
LINC01410	long intergenic non-protein coding RNA 1410	1.6411	0.0334	Up
CYP17A1	cytochrome P450 family 17 subfamily A member 1	4.1951	0.0005	Down
OR10AB1P	olfactory receptor family 10 subfamily AB member 1 pseudogene	4.0068	0.0146	Down
MYH14	myosin heavy chain 14	3.7783	0.0317	Down
LRRTM3	leucine rich repeat transmembrane neuronal 3	3.6646	0.0168	Down
GRIN3A	glutamate ionotropic receptor NMDA type subunit 3A	-3.1104	0.0445	Down
HS3ST5	heparan sulfate-glucosamine 3-sulfotransferase 5	-2.9968	0.0168	Down
NBAS	NBAS subunit of NRZ tethering complex	-2.8992	0.0146	Down
FNDC9	fibronectin type III domain containing 9	-2.8611	0.0419	Down
H1-4	H1.4 linker histone, cluster member	-2.8427	0.0146	Down
CRYAB	crystallin alpha B	-2.7802	0.0146	Down
NXPH3	neurexophilin 3	-2.5502	0.0348	Down
MYL3	myosin light chain 3	-2.5310	0.0437	Down
CMYA5	cardiomyopathy associated 5	-2.4531	0.0311	Down
AMIGO2	adhesion molecule with Ig like domain 2	-2.2807	0.0499	Down
SIK1B	salt inducible kinase 1B (putative)	-2.2002	0.0446	Down
EDIL3	EGF like repeats and discoidin domains 3	-2.1682	0.0311	Down
UBC	ubiquitin C	-1.1926	0.0499	Down
lnc-FANCC-1		2.6996	0.0437	Up

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

lnc-KLHL28-1		2.5450	0.0311	Up
lnc-TBCCD1-4		2.1511	0.0051	Up
lnc-CDC27-8		2.1123	0.0146	Up
lnc-SPG21-45		2.0232	0.0166	Up
AC137695.1		2.0062	0.0146	Up
Lnc-NSUN6-1		1.5671	0.0499	Up
Lnc-ZNF814-1		1.3445	0.0146	Up
Lnc-METRNL-8		-2.8955	0.0311	Down
Lnc-METRNL-1		-2.5951	0.0437	Down

Shown are gene symbols, gene names, fold change (log2FC), p-adjusted value and gene expression status. Genes without names are non-coding RNAs (LNCipedia GeneIDs) except *AC137695.1*

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

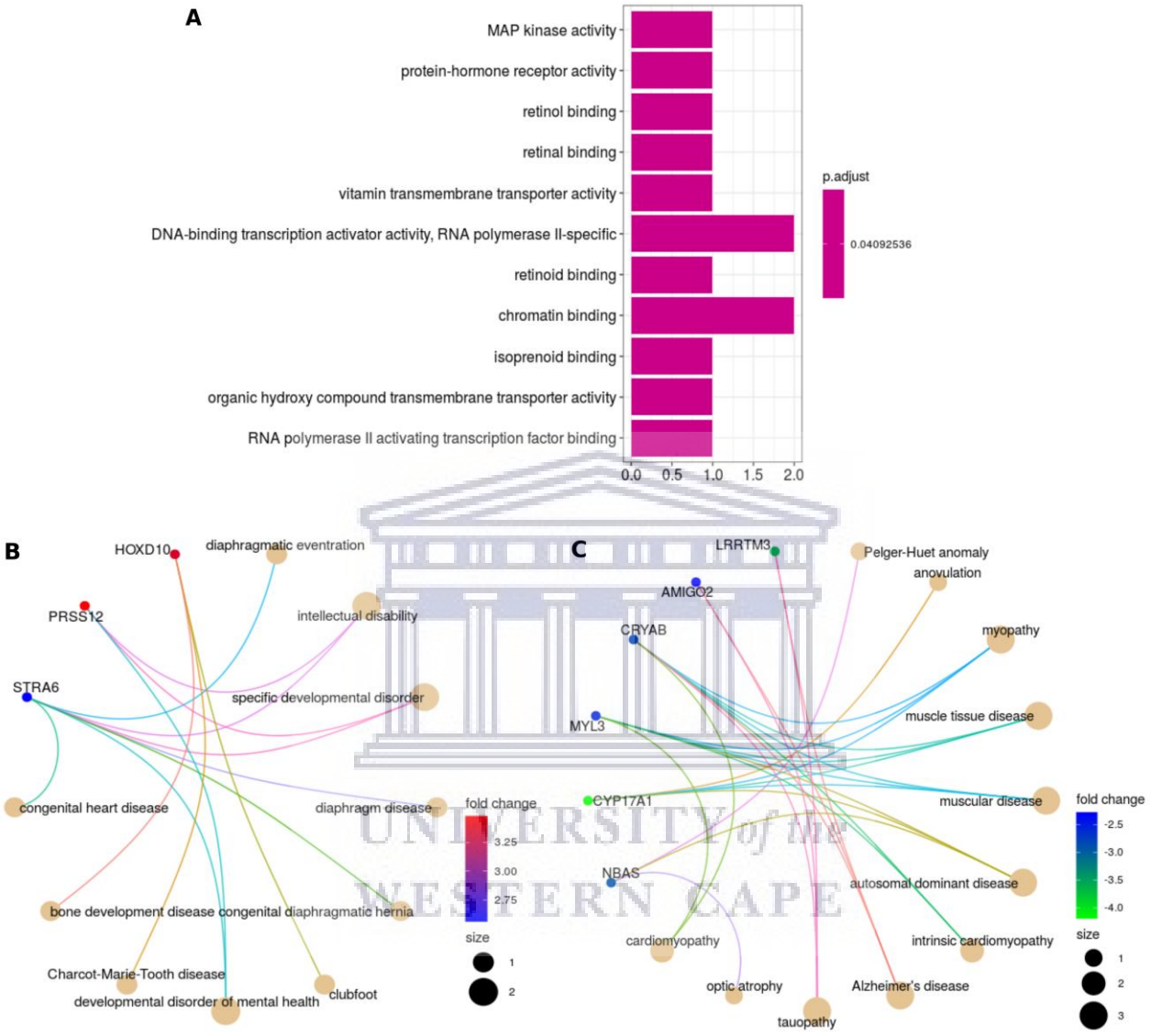


Figure 3.2: Gene and Disease enrichment analyses. Molecular Function activities enriched by the upregulated DEGs (A). Network representations of enriched diseases for upregulated DEGs (B) and downregulated DEGs (C) (absolute value of fold change > 1.5). The size of the circles represents the number of genes that enrich a disease

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Reconstruction of gene regulatory networks, using the GENIE3 algorithm, for the SS and LS samples respectively deduced 1,966,606 and 1,967,020 weighted interactions involving the DEGs. Applying a weight threshold of 0.00251 resulted in 1018 and 650 DEG interactions for the SS and LS groups, respectively. The visualization of the 1018 DEG interactions using Cytoscape enabled the detection of 4 essential regulatory networks (Fig. 3.3). The first network (Fig. 3.3A) involves *SMIM28*, *LGR5*, *PRSS12*, *EVX2*, *NHLH2* and *HOXD10*. All of these DEGs are upregulated, and the last three genes are transcription factors. The following network (Fig. 3.3B) interconnect *MAPK15*, *Lnc-ZNF814-1*, *EDIL3*, *NBAS* and *CYP17A1*. The first two genes are upregulated and the last three genes are downregulated. Most of the DEGs in the third and last networks (Fig. 3.3C-D) are downregulated, with the exception of *MEG9* and *STRA6*, which are upregulated. Interestingly, these interactions between the DEGs are not present in the LS group and the following genes; *SMIM28*, *HOXD10*, *PRSS12*, *NHLH2*, *MEG9*, *MAPK15*, *Lnc-ZNF814-1* and *FNDC9*, have no interactivity with any other DEG.

Filtering the GSE49711 dataset with the query criteria in Table 3.1 yielded 43 SS and 19 LS samples, respectively. DEGs that do not have an associated *NCBI GeneID* were not found in this dataset, particularly, those that are identified as long non-coding RNAs (lncRNAs). Only 25 of the 40 DEGs have expression data in this dataset. Based on the results of feature selection with scikit-learn and several classification tests, the following 16 features were selected for the machine learning construction of the training and test sets; *EVX2, NHLH2, PRSS12, POU6F2, HOXD10, MAPK15, RTL1, LGR5, CYP17A1, OR10AB1P, MYH14, LRRTM3, GRIN3A, HS3ST5, CRYAB and NXPH3*. The evaluation of the Support Vector Machines (SVM) and Artificial Neural Networks (ANN) models using 5-fold cross-validation resulted in an accuracy of 78% and 87% for SVM and ANN, respectively. By testing the ML models on the GSE49711 test set, ANN again achieved better results with an accuracy of 82% of samples correctly classified as SS or LS compared to SVM which obtained an accuracy of 79% (Table 3.3).

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

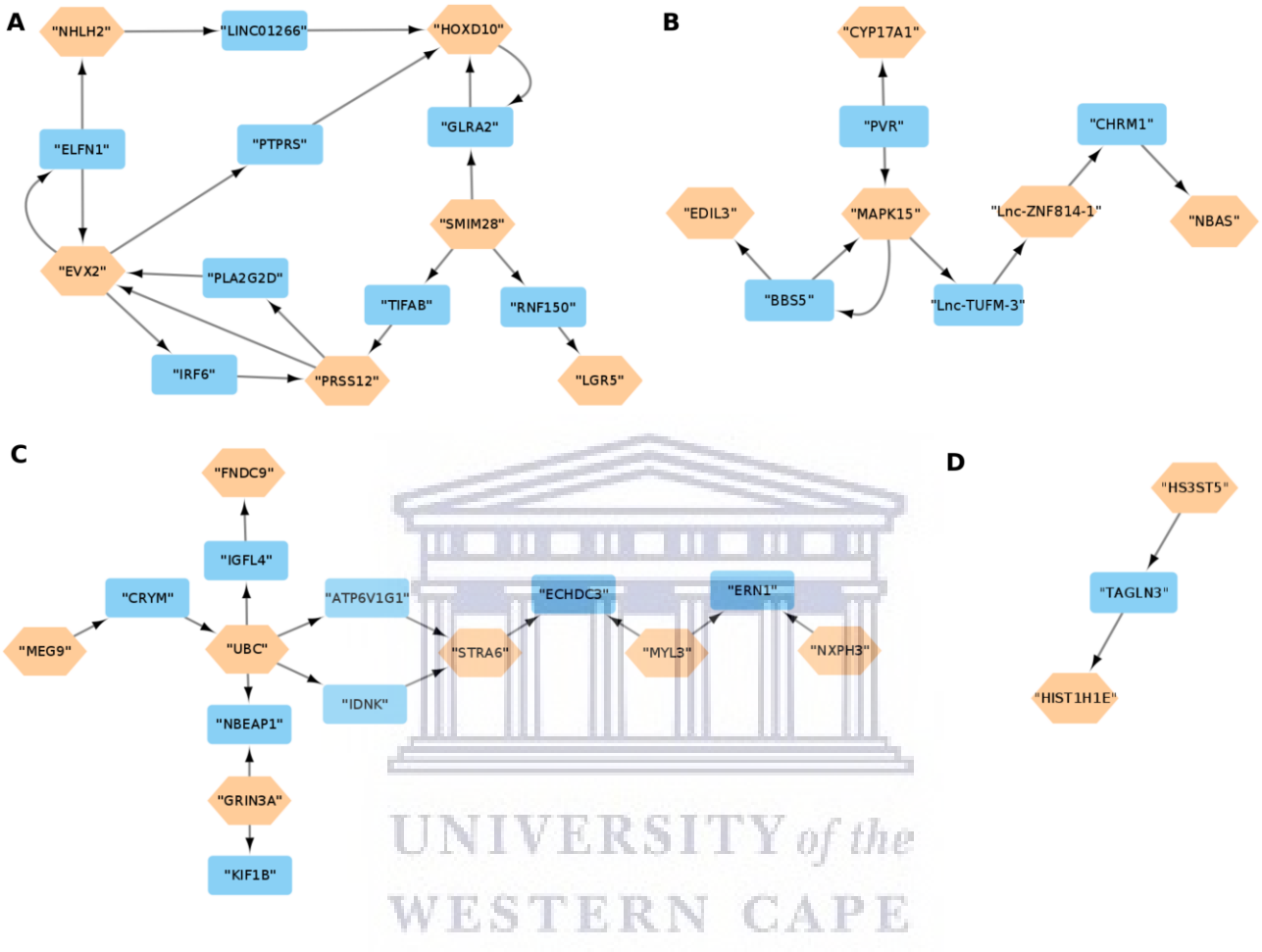


Figure 3.3(A-D): Gene Regulatory Network Analysis. Regulatory networks for the DEG interactions in the SS group. The brown hexagonal nodes are DEGs and the blue rectangles are intermediate genes connecting two DEGs. DEGs that do not interact with other DEGs are omitted.

UNIVERSITY of the WESTERN CAPE

Table 3.3: Machine learning results for the classification of the GSE49711 test set of high-risk neuroblastoma samples with SVM and ANN models.

	SVM		ANN	
Class	Precision	Recall	Precision	Recall
Short Survival	0.87	0.81	0.86	0.88
Long Survival	0.64	0.74	0.72	0.68
Accuracy	79%		82%	

3.5 Discussion

We aimed at identifying genes that are differentially expressed between high-risk SS and LS patients that could be potential prognostic indicators and or therapeutic targets. The results of the DGE analysis between the two groups showed the upregulation and downregulation of genes associated with neuroblastoma and other cancers.

UNIVERSITY of the
WESTERN CAPE

3.5.1 Differentially expressed genes

3.5.1.1 Upregulated genes

The upregulated DEGs included some genes whose overexpression in the SS group have been correlated with poor survival in neuroblastoma and other cancers. Higher expression levels of *NHLH2* were found to be higher in unfavourable neuroblastomas and was significantly associated with a poor prognosis (Aoyama et al., 2005). Additional roles of *NHLH2* in obesity and fertility has also been uncovered (Good and Braun, 2013; Vella et al., 2007). The upregulation of *PRSS12* in this study is similar to the results of Hiyama et al. (2003), which reported the overexpression of *PRSS12* in neuroblastoma tumors with high telomerase activity correlating with unfavourable tumors. Serine

http://etd.uwc.ac.za/

proteases are often altered and significantly upregulated in cancer as malignant cells need proteolytic activities to enable their growth, survival and expansion. Our result support that upregulation of *PRSS12* is indicative of poor survival in neuroblastoma. *HOXD10* is a transcription factor whose expression is altered in many cancers. Its high expression gives cancer cells proliferative and migratory abilities (Hakami et al., 2014). Elevated expression of HOXD genes including *HOXD10* was reported to be associated with unfavourable prognosis and poor outcome in neuroblastoma (Kocak et al., 2013), which supports our results indicating a more aggressive disease in the SS group. *LGR5* is a stem cell marker which is highly expressed and associated with an aggressive phenotype in neuroblastoma (Vieira et al., 2015; Forgham et al., 2015). It has also been associated with pancreatic ductal adenocarcinoma (Kuraishi et al., 2019) and colorectal cancer (Morgan et al., 2018). *LGR5* potentially contributes to stem cell maintenance and self-renewal and is indicative of poor survival in high-risk neuroblastoma. *SMIM28* is a less studied protein whose upregulation is indicative of poor survival in this study. Similar to our results, Jiang et al. (2018) reported the upregulation of *SMIM28* in prostate cancer. *EVX2* is a homeobox transcription factor essential for vertebrate spinal cord neuronal specification (Juarez-Morales et al., 2016). *POU6F2* belongs to the POU class homeobox family whose members are transcriptional regulators and is involved in hereditary predisposition to Wilms tumor, a pediatric malignancy of the kidney (Perotti et al., 2004). Functional studies are required to elucidate the roles of *SMIM28*, *EVX2* and *POU6F2* in high-risk neuroblastoma.

MAPK15, a protein kinase involved in many cellular activities including cell proliferation was upregulated in this study. Highest levels of *MAPK15* was found in aggressive embryonal carcinomas and it acts by sustaining the progression of the cell cycle of embryonal carcinomas by limiting p53 activation and preventing the facilitation of p53 dependent mechanisms that results to the arrest of the cell cycle (Rossi et al., 2016). Neuroblastoma is a malignancy of embryonal origin and upregulation of *MAPK15* would be expected to facilitate tumor progression and indicative of

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

aggressive disease and poor survival as in the SS group. *RTL1*, a paternally expressed imprinted gene highly expressed in the fetus, placenta and brain. It has been reported to be a driver of hepatocellular carcinoma (Riordan et al., 2013). Being a protease, it possibly promotes tumor invasion and metastasis in neuroblastoma tumors. Higher expression of *RTL1* is thus suggestive of poor prognosis in neuroblastoma. Also upregulated is *STRA6*, a plasma membrane protein that transports retinol and is involved in a signalling mediated by *JAK2*, *STAT3* and *STAT5* (Berry et al., 2014). Its upregulation indicates poor survival in our study possibly through its maintenance of cancer stem cells and promotion of tumor formation as reported in colorectal cancer (Berry et al., 2014; Karunanithi et al., 2017).

Ten lncRNAs were upregulated in this study. Three of these (*lnc-SPG21-45*, *lnc-NSUN6-1* and *lnc-KLHL28-1*) are antisense to *ANKDD1A*, *CACNB2* and *C14orf28* genes, respectively, which are associated with other cancers and diseases, possibly by regulating their expression. *C14orf28* has been observed to be overexpressed in colorectal cancer cells, promoting proliferation, migration and invasion (Yang et al., 2017). *ANKDD1A* has been described as a functional tumor suppressor with germline variants predicting poor patient outcomes in low-grade glioma (Chatrath et al., 2019), and is frequently methylated in glioblastoma multiforme (Feng et al., 2019) and in clinically non-functioning pituitary adenomas (Cheng et al., 2019). *CACNB2* is a calcium channel protein linked to diabetic retinopathy (Vuori et al., 2019), bipolar disorder (Liu et al., 2019), hypertension (Niu et al., 2010) and autism spectrum disorders (Breitenkamp et al., 2014). The role of calcium signalling in cancer has been reviewed by (Yang et al., 2010; Stewart et al., 2015; Cui et al., 2017). *MEG9*, is located in an imprinted non-coding RNA genomic region, DLK1-DIO3 (Hagan et al., 2009; Benetatos et al., 2013). *LINC01410* is a lncRNA highly expressed in pancreatic cancer tissues and cell lines (Cai et al., 2019). High expression in cholangiocarcinoma and gastric cancer patients have been associated with poor prognosis and survival (Zhang et al., 2018; Jiang et al., 2020). These lncRNAs may be facilitating the promotion of tumor progression, proliferation and invasion which might have

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

impacted the survival of patients in the short survival group. It is well established that metastasis is the primary cause of cancer mortality. Functional studies are required to evaluate the roles of these lncRNAs in neuroblastoma.

3.5.1.2 Downregulated genes

The downregulated genes include genes such as *AMIGO2*, *LRRTM3*, *GRIN3A*, *MYH14*, *EDIL3*, *FNDC9*, involved in neural function and development, and in ECM organization. *AMIGO2* is a transmembrane molecule expressed in neuronal tissues and participates in their formation (Kuja-Panula et al., 2003). *EDIL3* is an inducer of the epithelial-mesenchymal transition, that promotes angiogenesis and invasion in hepatocellular carcinoma (Xia et al., 2015). *FNDC9* which exhibits biased expression in the brain is an ECM protein involved in tumorigenesis in different cancers (Wang and Hielscher, 2017). *MYH14* is a myosin, an actin-dependent motor protein that plays a role in neuritogenesis. Members of the mysoin superfamily have been known to enhance or suppress tumor progression (Ouderkirk and Krendel, 2014). *MYH14* could be suppressing tumor progression in high-risk neuroblastoma. The downregulation of these ECM associated genes could be promoting the invasion and metastasis of neuroblastoma tumors.

UNIVERSITY of the
WESTERN CAPE

OR10AB1P belongs to the olfactory receptor family of genes. Olfactory receptors are expressed in various human tissues and are involved in different cellular processes such as migration and proliferation. Some are biomarkers for prostate, lung and small intestine carcinoma tissues (Weber et al., 2018). Decreased expression of *CRYAB* indicated its tumor suppressor function in bladder cancer (Ruan et al., 2020). It may thus also be functioning as a tumor suppressor in neuroblastoma. Ubiquitin C (*UBC*) is a polyubiquitin precursor. Ubiquitination has been associated with many cellular processes, which play roles in tumorigenesis.

CYP17A1 is a key enzyme in the steroidogenic pathway with restricted expression in the adrenal gland. Neuroblastoma tumors commonly occur in the adrenal medulla. The downregulation of

http://etd.uwc.ac.za/

CYP17A1 may be an indicator of poor prognosis in neuroblastoma. *LRRTM3* has high expression in the brain and belongs to a group of proteins involved in nervous system development. How it contributes to neuroblastoma remains to be investigated but it is currently a candidate gene for Alzheimer's disease (Reitz et al., 2012; Wang et al., 2014), a neurological disorder. *NBAS* is thought to be involved in golgi-to-ER transport and is typically amplified in *MYCN*-amplified neuroblastoma tumors (Wimmer et al., 1999). *GRIN3A* is a glutamate receptor that promotes nerve outgrowth (Shi et al., 2016). The downregulation of GRIN3A may suggest a higher level of disease because glutamate is a major excitatory neurotransmitter in the CNS that is involved in many neuronal processes. Functional studies are required to investigate how these genes contribute to poor survival in neuroblastoma.

3.5.2 Gene and Disease Ontology enrichment analyses

The molecular function activities; MAP kinase activity, retinol binding and RNA polymerase II activating transcription factor binding, enriched by the upregulated genes *MAPK15*, *STRA16* and *NHLH2* respectively, are activities that promote tumor cell proliferation. Deregulation of the *MAPK* signalling was associated with cancer development, progression and cell proliferation (Dhillon et al., 2007; Jin et al., 2015). Retinol binding through the *STRA6* upregulation activates a signalling cascade that is found to play a role in cancer initiation, maintenance and growth (Karunanithi et al., 2017). Furthermore, the increased global transcription activity (activation of RNA polymerase II) indicates an intensity of a rapid proliferation of cancer cells (Yokoyama, 2019). These activities again demonstrate the aggressiveness of the neuroblastoma tumors in SS patients compared to LS patients.

The enriched diseases by the upregulated and downregulated genes, particularly the disorders inducing heart failure (Cardiomyopathy and Congenital heart disease) and respiratory illness (Diaphragmatic dysfunction) may have negative impact on survival (Best and Rankin, 2016; Dube

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

and Dres, 2016), and could be the cause of death in the SS neuroblastoma patients. There is no indication in the clinical information accompanying the gene expression counts dataset if the patients suffered from additional disorders in addition to neuroblastoma. However, it is reported that high-risk neuroblastoma survivors treated with intensive multimodal therapy are at risk for a broad variety of treatment-related late effects including cardiac dysfunction, bone development disease, pulmonary dysfunction and impaired fertility (Friedman and Henderson, 2018). These treatment-related morbidities were enriched by the upregulated and downregulated DEGs (Figure 3.1B-C), suggesting that the cause of death in patients with SS neuroblastoma may be due to treatment complications.

3.5.3 Regulatory network analysis

After the application of a weight threshold of 0.00251 to the GENIE3 output, the numbers of interactions involving the DEGs in the SS and LS groups were 1018 and 650, respectively. There is a clear substantial difference in number of interactions between the two groups, indicating a higher cellular/tumor activity in the SS group which could be a sign of tumor aggressiveness in patients with SS neuroblastoma. In addition, it is noticeable from the networks in Figure 3.2 the importance of the genes *SMIM28*, *MAPK15* and *UBC* as origins of most of the interactions. The role of *MAPK15* and *UBC* in tumorigenesis has been reported in many previously discussed scientific works, while *SMIM28* is a less studied gene with an unclear role in cancer. However, observation of the role of the final target genes of *SMIM28* in Figure 3.2A network, which are *LGR5* and *HOXD10*, can shed light on the role of this gene. As reported previously, both *LGR5* and *HOXD10* were associated with cancer cell proliferation and tumor aggressiveness in neuroblastoma. Thus, it is possible that *SMIM28* (Small Integral Membrane Protein 28) is part of a signalling pathway whose role is to accelerate neuroblastoma tumor proliferation, making it a possible new gene therapy target in high-risk neuroblastoma cancer. Further investigations are required to elucidate precisely the role of *SMIM28* in the aggressiveness of neuroblastoma tumors.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

3.5.4 Machine Learning

Although not all the DEGs were included in the training and testing of the machine learning models, the obtained prediction results were significantly good. The ANN model obtained the highest accuracy of 82% for the classification of the external neuroblastoma samples (GSE49711) into short and long survival classes. This high classification accuracy makes it possible to consider that the DEG expression profiles have been preserved in the various high-risk neuroblastoma tumors, although neuroblastoma is known to be heterogeneous. Therefore, the DEG list can serve as prognostic indicators (genetic signature) for survival time in high-risk neuroblastoma patients and can be targets for drug discovery analyses. Relatively similar to our study, other studies have proposed genetic signatures for prognostic stratification of patients with neuroblastoma. Liu et al. (2020) used an unsupervised biclustering machine learning technique to find high-risk neuroblastoma subtypes. They proposed a signature of 238 neuroblastoma-specific immune genes to identify ultra high-risk and high-risk neuroblastoma subtypes. Russo et al. (2017) applied K-means clustering to 27 kinome gene signature to identify ultra high-risk subtypes of high-risk neuroblastoma. Formicola et al. (2016) used Cox regression and Kaplan-Meier analysis methods to propose a 18-gene expression based risk scoring system to predict overall survival of patients with stage 4 neuroblastoma. We demonstrated the use of a shorter signature in a 16-gene expression classifier based on survival time to stratify high-risk neuroblastoma into SS (ultra high-risk) and LS subtypes. The number of genes in our classifier should provide the advantage of being less costly and easier to implement.

3.6 Conclusion

The DGE analysis is a powerful technique for the identification of DEGs in a studied condition. In this study, using DESeq2 we identified 40 DEGs between SS and LS neuroblastoma samples. Many of the DEGs were found to be related to different cancers (including neuroblastoma), thus strengthening their possibility of being associated with neuroblastoma. The ML models based on 16

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

DEGs were capable of stratifying high-risk neuroblastoma samples on the basis of survival time, demonstrating their ability to be used as a genetic signature or prognostic indicators of survival in high-risk neuroblastoma patients. This study furthers our understanding of the molecular mechanisms of neuroblastoma in high-risk patients. We identified *SMIM28* gene to be critical for tumor proliferation making it as a possible gene therapy target. Nevertheless, additional studies are required to elucidate the role of *SMIM28* in the pathogenesis of neuroblastoma. Finally, prognostic stratification of high-risk neuroblastoma patients will help clinicians in making better therapeutic and patient management decisions.

http://etd.uwc.ac.za/

Chapter 4

Predicting amplification of *MYCN* using CpG methylation biomarkers in neuroblastoma

Published Paper

Abdulazeez Giwa, Sophia Catherine Rossouw, Azeez Fatai, Junaid Gamieldien, Alan Christoffels, Hocine Bendou. Predicting amplification of MYCN using CpG methylation biomarkers in neuroblastoma. Future Oncology. 2021;17(34):4769-4783. https://doi.org/10.2217/fon-2021-0522. [PubMed: 34751044]

UNIVERSITY of the
WESTERN CAPE

4.1 Abstract

Neuroblastoma is the most common extracranial solid tumor in childhood. Amplification of *MYCN* is found in 22% of neuroblastomas and is a predictor of poor prognosis. Here, we aimed to identify a DNA methylation signature that is diagnostic of *MYCN* amplification, and prognostic of patient survival. DNA methylation data was downloaded from the TARGET data matrix, and samples were stratified into two groups based on their *MYCN* amplification status, as either amplified or not amplified. Using ChAMP R package, we identified 663 CpGs from 369 genes to be differentially

http://etd.uwc.ac.za/

methylated between 45 *MYCN* amplified and 81 non-amplified samples. Of these 369 genes, 14 were highly methylated, having at least six significant CpGs. Applying recursive feature selection (RFE) on the 663 CpGs, 25 CpGs were selected for machine learning (ML) classification tasks. The beta values of each of the 25 selected CpGs were used to perform clustering analysis and to create a Support Vector Machine (SVM) model for *MYCN* amplification prediction. We achieved 100% clustering accuracy of the 126 TARGET samples into their respective *MYCN* amplification groups. The validation of the SVM model was performed on 3 external datasets, and high accuracy scores of 100%, 97% and 93% were obtained for all three test sets. The Cox regression analysis and Kaplan-Meier estimates identified 8 survival-associated CpGs with the methylation of cg13558971 (*ATP2BP4*), cg25310824 (*SEPP1*) and cg07476617 (*CFLAR*), indicative of poor outcome in *MYCN* amplified neuroblastoma. The results demonstrate the efficacy of the method used here and the utility of the identified CpGs in helping clinicians to make better targeted therapeutic management decisions.

4.2 Introduction

Neuroblastoma is the most commonly occuring extracranial solid tumor in childhood and accounts for approximately 15% of pediatric cancer-related deaths (Maris, 2010; Smith et al., 2010; Ward et al., 2014). It can develop anywhere along the sympathetic nervous system, with 60% of tumors occurring in the abdominal region, of which approximately half are located in the medulla of the adrenal glands (Zhang et al., 2018; Johnsen et al., 2019). Tumors in high-risk neuroblastoma patients are often metastatic, resulting in survival rates of less than 50% (Maris, 2010). Characteristics for high-risk neuroblastoma include age, loss of chromosome 1p or 11q and amplification of *MYCN* (Mueller and Matthay, 2009; Tonini et al., 2012). Amplification of *MYCN* is a well-studied genomic alteration found in approximately 22% of cases (Newman and Nuchtern, 2016) and is a predictor of poor prognosis (Brodeur et al., 1984; Seeger et al., 1985), although patients without *MYCN* amplification may also have a poor outcome (Baali et al., 2018). Neuroblastoma is a heterogeneous

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

disease with outcomes ranging from spontaneous regression, as seen in some tumors, to relentless progression despite extensive and varied therapies (Baali et al., 2018).

Some alterations identified in neuroblastoma include mutations in *ALK*, *ATRX* and *TERT* (Mossé et al., 2008; Cheung et al., 2012; Valentijn et al., 2015; Peifer et al., 2015). Through genomic sequencing, it is now known that pediatric cancers, including neuroblastoma, have lower numbers of mutations compared to adult cancers (Vogelstein et al., 2013; Sweet-Cordero and Biegel, 2019), with many primary neuroblastomas not containing recognisable driver mutations. This suggests the involvement of epigenetic alterations. It has been shown that epigenetic factors, especially alterations in DNA methylation, play a role in the pathogenesis of neuroblastoma (Ratner et al., 2016; Decock et al., 2011). For example, hypermethylation of *TERT* was proposed as a biomarker for poor prognosis in neuroblastoma (Olsson et al., 2016). In addition, DNA methylation of *CASP8* and *RASSF1A* was linked to the development and progression of neuroblastoma (Decock et al., 2011) and were associated with poor prognosis (Hassan et al., 2020).

Another indicator of poor prognosis and survival in cancer is the CpG island methylator phenotype (CIMP) (Decock et al., 2016). It was first established in colorectal (CRC) cancer and then adopted in neuroblastoma (Asada et al., 2013). In a genome-wide study of 140 neuroblastomas, CIMP was defined by methylation of 5 CpG islands (CGIs) in the PCDHB family, and was associated with the methylation of promoter CGIs of several tumor suppressor genes such as *RASSF1A* and *BLU* (Abe et al., 2005). Most of *MYCN* amplified cases exhibited CIMP. However, the presence of CIMP was also detected in many cases without *MYCN* amplification (Abe et al., 2005), therefore emphasising the need for an accurate predictor of *MYCN* amplification.

While some studies have proposed different methods of diagnosing *MYCN* amplification in neuroblastoma (such as PCR based and hybridization laboratory methods) (Hiyama et al., 1999; Oude

http://etd.uwc.ac.za/

Luttikhuis et al., 2000; Mathew et al., 2001; Yagyu et al., 2016), we are unaware of any assessing the use of CpG methylation biomarkers. Therefore, in this study, we are interested in identifying a CpG methylation signature that is diagnostic of *MYCN* amplification. In addition, we aim to identify CpGs that are associated with survival and poor prognosis. The use of methylation biomarkers for the diagnosis of *MYCN* amplification and prognosis of survival has the potential to be clinically useful in deciding treatment strategy and could be cost effective as well.

4.3 Methods

The complete workflow describing the steps and methods undertaken in this study is shown in Figure 4.1. It included five essential steps; data retrieval, data preprocessing (imputation, normalization and batch effect removal), differential methylation analysis, machine learning, and, Cox regression & Kaplan-Meier estimates.

4.3.1 Dataset

The Therapeutically Applicable Research to Generate Effective Treatments (TARGET) (https://ocg.cancer.gov/programs/target) initiative uses comprehensive genomic approaches for the molecular characterization of hard-to-treat childhood cancers. Neuroblastoma methylation data is accessible via the TARGET data matrix portal and is mainly composed of high-risk samples with available clinical information. The level 3 methylation dataset includes beta values from 235 samples (https://target-data.nci.nih.gov/Public/NBL/methylation_array/L3/). Based on the clinical information contained in the metadata file (https://target-data.nci.nih.gov/Public/NBL/clinical/harmonized/TARGET_NBL_ClinicalData_Discovery_20170525.xlsx), International Neuroblastoma Staging System (INSS) stage 4 samples with known *MYCN* status were considered. A total of 126 samples (45 samples with *MYCN* amplification and 81 samples without *MYCN* amplification) were downloaded from the TARGET data matrix portal. Missing values were imputed using methyLImp package (Di Lena et al., 2019) which applies a computationally

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

efficient imputation method based on linear regression. The R-package ChAMP (Morris et al., 2014; Tian et al., 2017) was used for data normalisation and elimination of variability and batch effects between groups. A total of 396,065 CpGs were used for downstream analysis.

4.3.2 Differential methylation analysis

Differential methylation analysis was done using the ChAMP package (Morris et al., 2014; Tian et al., 2017). In finding biologically relevant differentially methylated probes, the ChAMP package uses the limma R-package (Ritchie et al., 2015) to compare two groups. We performed differential methylation analysis between *MYCN* amplified and non-amplified groups as illustrated in Figure 1**.** Differentially methylated CpGs with a p-value < 0.05 and an absolute value of delta-beta > 0.4 were called as significant. Heatmaps were plotted with the gplots R package. Gene Ontology (GO) and Disease enrichment analysis was carried out to functionally annotate the differentially methylated genes (DMGs) using Enrichr (Chen et al., 2013; Kuleshov et al., 2016).

UNIVERSITY of the
WESTERN CAPE

4.3.3 Machine learning

The factoextra R package (Kassambara and Mundt, 2020) was used to create and visualize clustering of samples based on all significant CpGs into their respective groups. A dendrogram plotting function was applied to data object produced from the application of an ensemble of hierarchical and k-means clustering algorithms to the significant CpGs (Kassambara and Mundt, 2020). Recursive feature elimination (RFE) was done to extract the most important features and eliminate those that create bias and negatively contribute to the model performanc**e.** In the RFE, linear support vector machine (SVM) algorithm was used for building ML models which were evaluated with repeated stratified 10-fold cross-validation (3 repeats) to determine the best parameters and features to use in the ML classification.

http://etd.uwc.ac.za/

The CpG features selected by RFE were then used to create an SVM model. LibSVM (Chang and Lin, 2011) was used to create the SVM model with the following parameters; kernel = linear, cost = 10. The SVM model was then tested on tumor methylation samples from two independent GEO datasets, GSE54719 (Gomez et al., 2015) and GSE120650 (Ackermann et al., 2018), as well as on matched primary tumor and relapse samples from another independent GEO dataset, GSE65306 (Schramm et al., 2018). These independent datasets are composed of samples of different INSS stages (Stage 1, 2, 3, 4 and 4s) and with known *MYCN* amplification statuses. The evaluation metrics for the SVM model was precision, recall and accuracy.

4.3.4 Cox regression analysis and Kaplan-Meier estimates

To determine among the most highly significant CpGs those that best correlated with patient survival, we used a Cox regression model based on the LASSO algorithm of the glmnet R package (Friedman et al., 2010; Simon et al., 2011; Tibshirani et al., 2012) and which was evaluated by leave-one-out cross-validation. The model assigns each CpG a regression coefficient value. CpGs with a zero coefficient were considered to have no effect on survival and were therefore eliminated. The method described by Ng et al. (2016) was followed, whereby a CpG score value is calculated for each patient as a linear combination of beta values of the top significant CpGs weighted by their corresponding coefficients obtained from the Cox regression model. A median value was inferred from the patient scores. Each score was then compared to the median, and patients were assigned a status value of 1 or 0, depending on whether the score was above or below the median. Kaplan-Meier (K-M) estimates and Hazard Ratio (HR) were then calculated for the overall survival (OS) and event-free survival (EFS) according to patient status information. Additionally, we performed K-M analysis using the CpGs identified by the Cox regression model on one of the independent test dataset, GSE65306, as it is the only dataset providing survival information. K-M curves were generated using the *ggsurvplot* function from the survminer R package.

UNIVERSITY of the
WESTERN CAPE

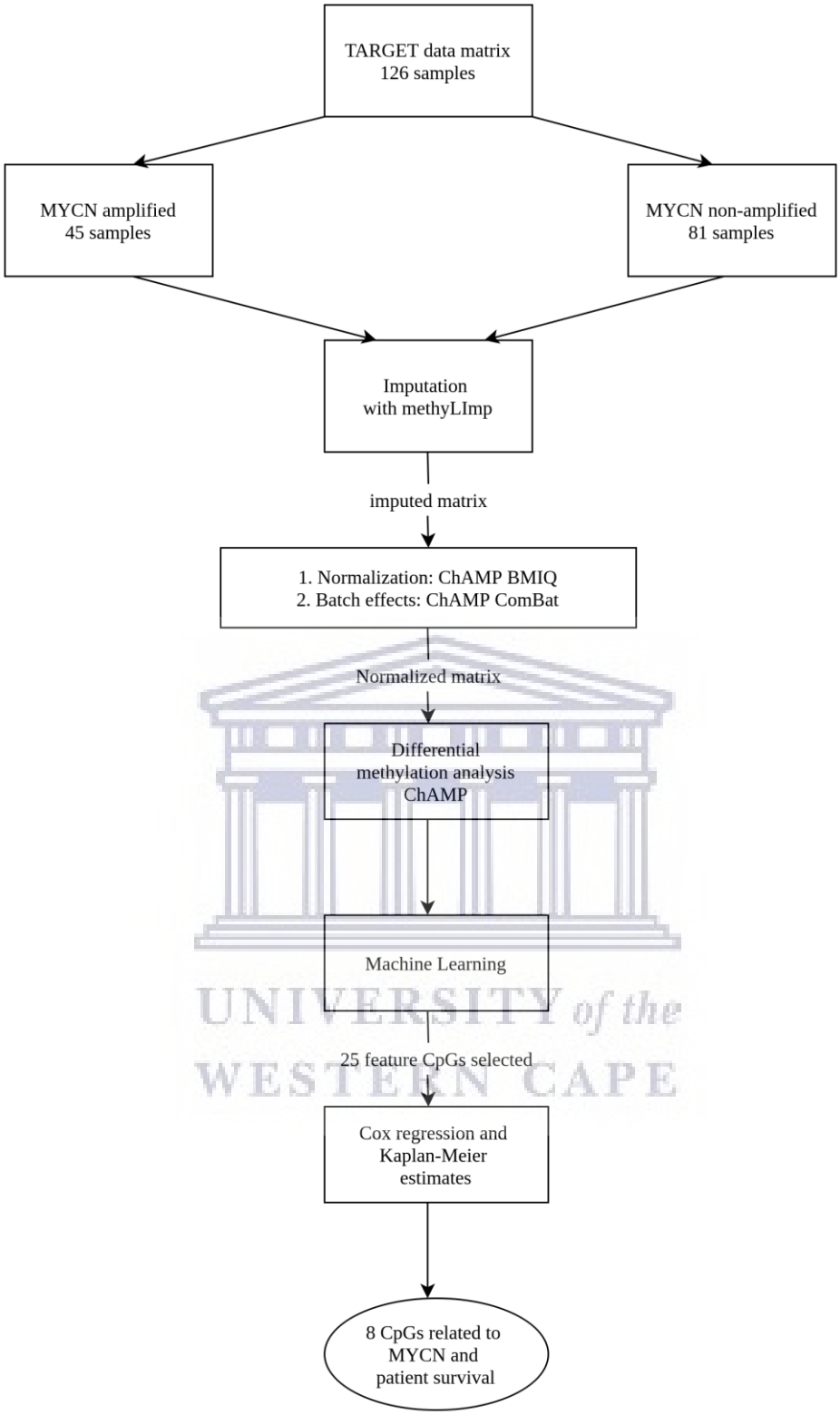
http://etd.uwc.ac.za/

4.3.5 Availability of data and materials

R and python scripts were implemented for this study and are available in https://github.com/SANBI-SA/NBMethyl.git.

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/



UNIVERSITY of the
WESTERN CAPE

Figure 4.1: Workflow depicting the steps and tools used in this analysis. The workflow has 5 principal steps; data retrieval from data matrix portal, data preprocessing using methyLImp and ChAMP R packages, differential methylation analysis with ChAMP, ML training and testing using SVM, and survival analysis using glmnet R package and Kaplan-Meier (K-M) curves.

http://etd.uwc.ac.za/

4.4 Results

4.4.1 Differential methylation analysis

Differential methylation analysis between the *MYCN* amplified and *MYCN* non-amplified groups identified 663 differentially methylated CpGs of 369 DMGs. Of these 369 DMGs, 238 and 131 DMGs had high and low methylation, respectively, in the *MYCN* amplified group compared to the non-amplified group. Fourteen genes were highly methylated between the *MYCN* amplified and non-amplified groups. Table 4.1 shows information about the highly methylated genes, defined as having at least six differentially methylated CpGs with *NXPH1* having the highest number of differentially methylated CpGs.

Pathway enrichment analysis indicate that the DMGs were enriched in pathways of ECM organization, cardiac hypertrophic response and neural crest differentiation (Table 4.2). Concerning molecular function and biological processes, GO analysis revealed that the DMGs were mainly enriched in regulation of transcription and cell differentiation. In addition, for disease enrichment analysis, the DMGs were associated with different cancers (kidney, liver, nasopharynx) as well as heart conduction disease (Table 4.2).

UNIVERSITY of the
WESTERN CAPE

4.4.2 Machine learning (ML)

Recursive feature selection (RFE) was performed which selected 25 CpGs as the most important for classifying samples by their amplification groups (Table 4.3). These 25 CpGs accurately clustered (100% precision) the 126 TARGET samples by their *MYCN* amplification group, with no sample misclassified (Fig. 4.2). The dendrogram plot of hierarchical and k-means applied to the total 663 significant CpGs resulted in correct clustering of 122 (96%) of the 126 TARGET samples (Fig. 4.3) The 25 CpGs selected by RFE from the 663 CpGs as the most informative features in the data were then used in ML training and test set construction (Table 4.3). The training set was built on the 126

http://etd.uwc.ac.za/

TARGET data matrix samples, while the test sets comprised 35, 58 and 28 samples from the GSE54719, GSE120650 and GSE65306 datasets, respectively. Repeated stratified 10-fold cross-validation yielded an accuracy of 96%. Evaluation of the SVM model using the GSE54719, GSE120650, and GSE65306 test sets resulted in high accuracies of 100%, 97%, and 93%, respectively, of correctly predicted samples (Table 4.4).

4.4.3 Cox regression analysis and Kaplan-Meier estimates

A minimal subset of 8 CpGs with a non-zero coefficient was selected by the Cox regression model (Table 4.5). These CpGs are believed to have a role in patient survival. All 8 CpGs were associated with overall survival, while only 5 were associated with event-free survival. The positive coefficient CpGs were methylated in the *MYCN* amplified group, while the negative coefficient CpGs were methylated in the *MYCN* non-amplified group. The CpGs and their coefficients were used to calculate a CpG score and assign a status value of 1 or 0 for each patient. K-M estimates for OS and EFS based on patient statuses were derived and shown in Figure 4.4. *MYCN* status correlated with CpG score and survival (OS: HR=5.11; $P<0.0001$, EFS: HR=4.845; $P<0.0001$). Patients with a CpG score above the median were predominantly *MYCN* amplified and those below were mostly non-amplified. Same significant results were observed in the GSE65306 test dataset (OS: HR=35.87; P<0.0001, EFS: HR=7.99: P<0.00041). The K-M plots for the GSE65306 test dataset are shown in Figure 4.5.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Table 4.1: Fourteen highly methylated genes between the *MYCN* amplified and *MYCN* non-amplified groups. These genes have ≥ 6 significantly methylated CpGs.

Gene	Gene name	Chr	Number of significant CpG
NXPH1	neurexophilin 1	7	14
SOX2-OT	SOX2 overlapping transcript	3	12
DLX5	distal-less homeobox 5	7	10
TFAP2D	transcription factor AP-2 delta	6	10
CAVIN3	caveolae associated protein 3	11	8
VAX2	ventral anterior homeobox 2	2	8
TERT	telomerase reverse transcriptase	5	7
HHEX	hematopoietically expressed homeobox	10	7
KRT19	keratin 19	17	7
RNF207	ring finger protein 207	1	7
MIRLET7BHG	MIRLET7B host gene	22	7
CHRNE	cholinergic receptor nicotinic epsilon subunit	17	6
DLX6-AS1	DLX6 antisense RNA 1	7	6
TMCO3	transmembrane and coiled-coil domains 3	13	6

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

Table 4.2: Enrichment analysis of the 369 differentially methylated genes

Pathways, Disease, Ontology	adj p-value	Number of overlapped genes	Genes
WikiPathways			
Cardiac hypetrophic response WP2795	0.030335	6	*HDAC4, PPP3CA, NPPA, GUCA1A, AKT1, PRKCA*
Neovascularisation processes WP4331	0.030335	4	*FLT4, CXCR4, AKT1, EPHB4*
Neural crest differentiation WP2064	0.030335	8	*HDAC4, TFAP2B, DLX5, TLX2, MPZ, ETS1, BMP7, SOX5*
Reactome			
Extracellular matrix organization Homo sapiens R-HSA-1474244	0.014525	17	*PTPRS, COL22A1, COL23A1, LTBP4, PDGFA, PRKCA, PCOLCE, LTBP2, BMP7, COL5A1, COL4A1, COL4A4, TIMP2, NCAM1, CD44, DDR2, ITGA9*
Jensen Diseases			
Kidney cancer	0.01632	74	*SAMD9L, TRIO, PTPRS, FLT4, CHD5, KNDC1, FRY, AFF3, SIPA1L3, CELSR3, PTPRG, DOCK10, CDH4, HEPHL1, GRM6, AKT1, RNF150, ERC2, MCF2L2, EPHB4, SOX5, ARHGEF10, MEF2C, UNC13A, RIPK4, TET1, EBF3, WDR72, FRMD4A, NAV1, TYK2, DNM2, ADCY9, BANK1, COL4A1, PARD3, COL4A4, ANGPTL2, DDR2, ANKRD11, LTBP2, LRP2, THBS2, ACACA, ARNTL, KIAA0556, GRIN2A, CUX1, PCBP1, CDH22, G3BP1, MAN1C1, TRPM8, NCAM2, RGS22, XDH, DYNC1I1, TMEM132D, FARP1, TFAP2B, DNAH10, TFAP2D, COL22A1, PTCH1, DAB2IP, ATP2B4, NFATC1, TMCO3, SDK1, COL5A1, KCNS2, ZNF536, RGS12, DZIP3*
Heart conduction disease	0.02058	8	*SFRP2, RNF207, CXCR4, FRMD4A, C9ORF3, PTPRG, FOXP1, ITGA9*
Liver cancer	0.03701	24	*RSPH6A, TMEM132D, SAMD9L, TRIO, DNAH10, PTPRS, COL22A1, AGAP2, KNDC1, LRP2, FRY, THBS2, ACACA, SDK1, SNTG2, CUX1, COL5A1, COL4A1, GRM6, PARD3, ZNF536, FAM65B, NCAM1, ERC2*

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

Nasopharynx carcinoma	0.04403	3	*HLA-A, HS3ST4, ITGA9*
GO Biological Process			
Positive regulation of cell differentiation (GO:0045597)	0.0224	14	*IFITM1, MEF2C, ZBTB16, SOX11, BMP7, ARNTL, SFRP2, PDPN, MYF6, AKT1, ASB4, CMKLR1, SOX5, DDR2*
Negative regulation of cell differentiation (GO:0045596)	0.0224	12	*TBX1, TRIO, SFRP2, COL5A1, ZBTB16, PTCH1, DAB2IP, ANP32B, XDH, MEIS2, SMAD7, ARNTL*
Positive regulation of ossification (GO:0045778)	0.0224	7	*IFITM1, MEF2C, ZBTB16, FZD9, SOX11, BMP7, DDR2*
Regulation of transcription from RNA polymerase II promoter (GO:0006357)	0.0224	49	*ZCCHC12, CD40, CHD5, ENO1, ETS1, PPP3CA, HHEX, CHP2, TEAD4, MEF2C, SMARCC1, ZHX2, TET1, EBF3, SOX11, FOXP1, SFRP2, GAL, ETV3L, EZR, SKAP1, HDAC4, DLX5, DOT1L, GATA4, ARNTL, RXRA, CUX1, DEAF1, HSF5, TBX1, WWOX, TFAP2B, TFAP2D, BCL11B, PTCH1, ZBTB16, DAB2IP, ATP2B4, NFATC1, BMP7, MEIS2, FLI1, SMAD7, NFIB, TLX2, NFIC, MYF6, ZNF536*
Cellular response to growth factor stimulus (GO:0071363)	0.0224	11	*TBX1, WWOX, MEF2C, DLX5, FLT4, DAB2IP, AKT1, BMP7, CD44, SOX5, SMAD7*
Positive regulation of transcription DNA-templated (GO:0045893)	0.02774	39	*HDAC4, CD40, DLX5, DOT1L, GATA4, ETS1, ARNTL, PPP3CA, HHEX, RXRA, DEAF1, ATOH8, CHP2, AKT1, TEAD4, TBX1, WWOX, TFAP2B, MEF2C, SMARCC1, BCL11B, ZBTB16, DAB2IP, TET1, EBF3, SOX11, NFATC1, BMP7, MEIS2, FLI1, SMAD7, DNM2, SFRP2, GAL, NFIB, TLX2, NFIC, MYF6, SKAP1*
Positive regulation of transcription from RNA polymerse II promoter (GO:0045944)	0.02774	32	*HDAC4, CD40, DLX5, DOT1L, GATA4, ETS1, ARNTL, PPP3CA, HHEX, RXRA, TERT, CHP2, TEAD4, TBX1, WWOX, TFAP2B, MEF2C, BCL11B, DAB2IP, TET1, EBF3, SOX11, NFATC1, MEIS2, SMAD7, SFRP2, GAL, NFIB, TLX2, NFIC, MYF6, SKAP1*
Regulation of transmembrane receptor protein seine/threonine kinase signaling pathway (GO:0090092)	0.03248	5	*TFAP2B, LTBP4, SOX11, BMP7, SMAD7*
GO: Molecular Function			
RNA polymerase II regulatory region sequence-	0.01956	22	*TFAP2B, MEF2C, SMARCC1, TFAP2D, DLX5, ZBTB16, SOX11, GATA4,*

UNIVERSITY of the WESTERN CAPE
http://etd.uwc.ac.za/

specific DNA binding (GO:0000977)			*NFATC1, ENO1, ETS1, MEIS2, HHEX, RXRA, CUX1, NFIB, DEAF1, NFIC, ZFP42, MYF6, ZNF536, HSF5*
Transcription factor activity RNA polymerase II core promoter proximal region sequence-specific binding (GO:0000982)	0.04652	15	*MEF2C, BCL11B, DLX5, EBF3, SOX11, ENO1, ETS1, MEIS2, ARNTL, HHEX, NFIB, NFIC, ZFP42, ZNF536, ZNF467*

UNIVERSITY of the
WESTERN CAPE

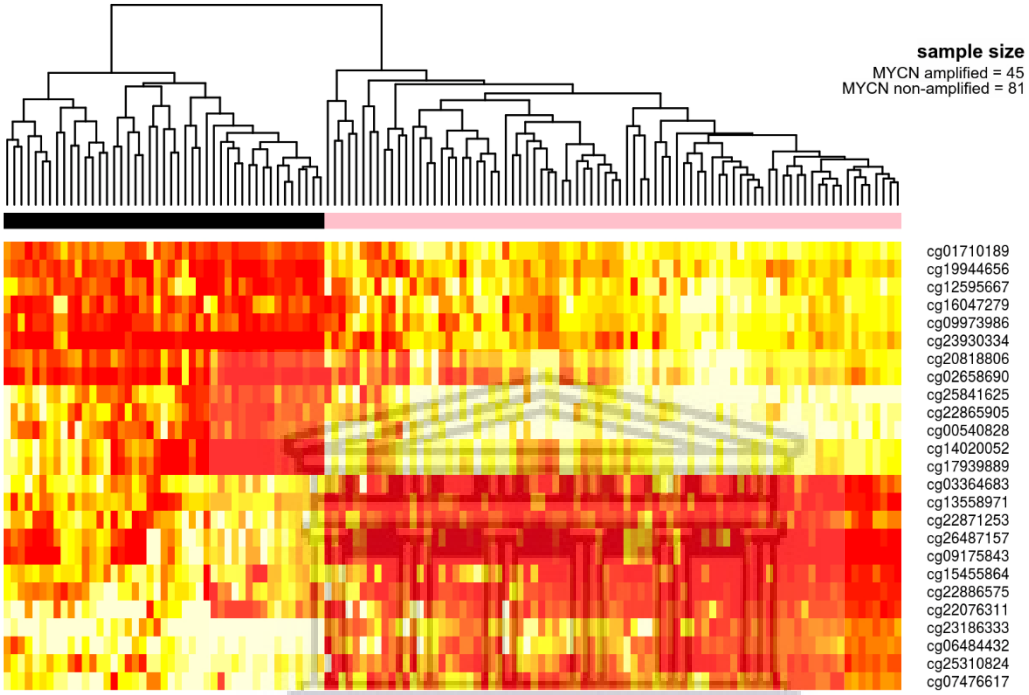
http://etd.uwc.ac.za/

Table 4.3: Twenty-five (25) differentially methylated CpGs between the *MYCN* amplified and *MYCN* non-amplified groups selected by RFE

CpG	adjusted p-value
cg00540828	4.276078e-29
cg01710189	2.392558e-24
cg13558971	5.777789e-23
cg23930334	2.474014e-22
cg03364683	7.279483e-22
cg19944656	6.866947e-20
cg23186333	1.801227e-19
cg25310824	1.964955e-19
cg20818806	3.380732e-19
cg09973986	1.051204e-18
cg06484432	4.085595e-18
cg07476617	2.075058e-17
cg22865905	2.403524e-17
cg09175843	4.275436e-17
cg22886575	1.358736e-16
cg25841625	1.564289e-16
cg26487157	4.648216e-16
cg22871253	1.528965e-15
cg12595667	5.146262e-15
cg17939889	9.408868e-15
cg14020052	1.592441e-14
cg02658690	1.720897e-14
cg15455864	4.358962e-14
cg16047279	1.709752e-13
cg22076311	9.729690e-12

UNIVERSITY of the WESTERN CAPE

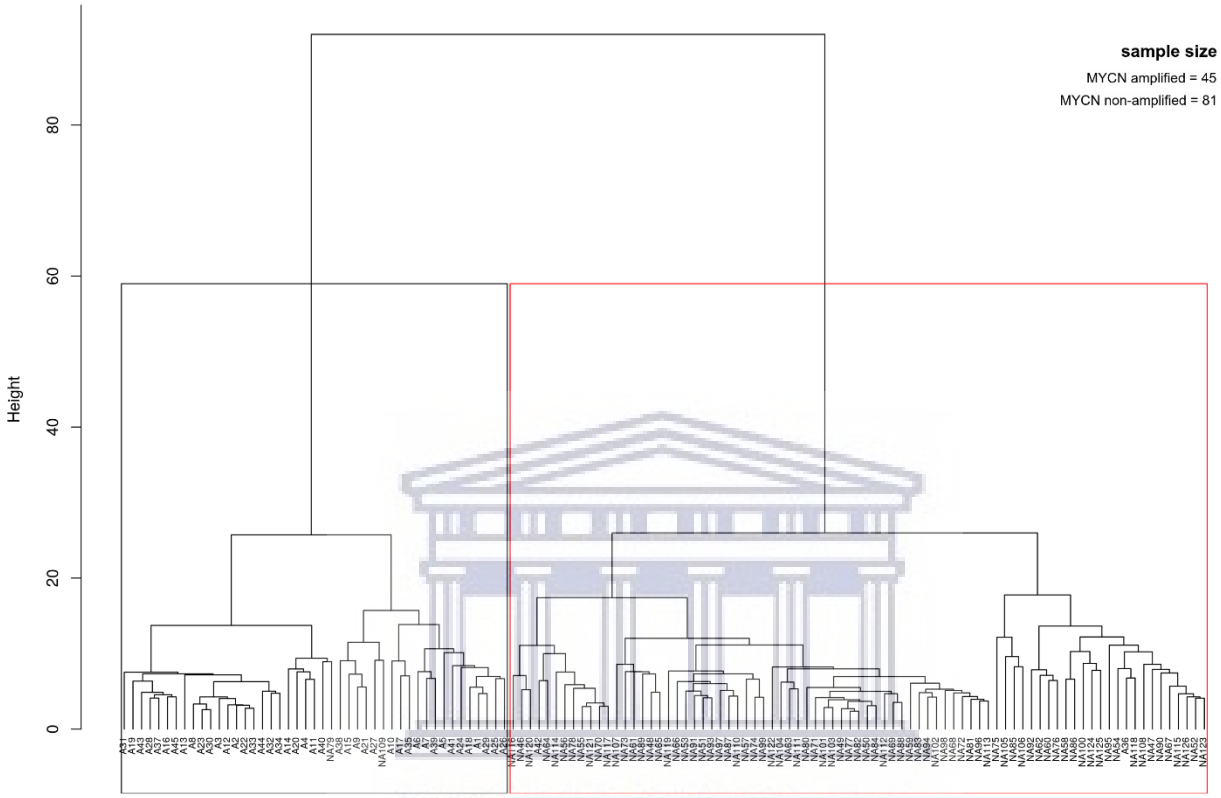
http://etd.uwc.ac.za/



UNIVERSITY of the
WESTERN CAPE

Figure 4.2: Heatmap of methylation level of the 25 top significant CpG sites (selected by RFE) in the *MYCN* amplification groups plotted with the gplots R package. 126 samples were classified including 45 *MYCN* amplified samples and 81 *MYCN* non-amplified samples. Two main clusters were identified with the *MYCN* amplified samples principally clustered in the left cluster while the *MYCN* non-amplified samples were principally clustered in the right cluster. *MYCN* amplifed samples are coded black while *MYCN* non-amplified samples are coded pink. No sample was misclassified. The heatmap colors represent intensity ranging from a lower intensity of red to a higher intensity of yellow.

http://etd.uwc.ac.za/



UNIVERSITY of the
WESTERN CAPE

Figure 4.3: Clustering of *MYCN* amplification samples using all 663 significant CpGs plotted with the factoextra R package. The samples are on the x-axis, with *MYCN* amplified samples labelled with A and *MYCN* non-amplified samples labelled with NA. Two main clusters were identified with the *MYCN* amplified samples principally clustered in the left cluster (black box) while the *MYCN* non-amplified samples were principally clustered in the right cluster (red box). 126 samples were clustered including 45 *MYCN* amplified samples and 81 *MYCN* non-amplified samples. Four (2 *MYCN* amplified and 2 *MYCN* non-amplified) samples were misclassified.

http://etd.uwc.ac.za/

Table 4.4: Results of the machine learning prediction of *MYCN* amplification of the samples from GSE54719, GSE120650 and GSE65306 test sets

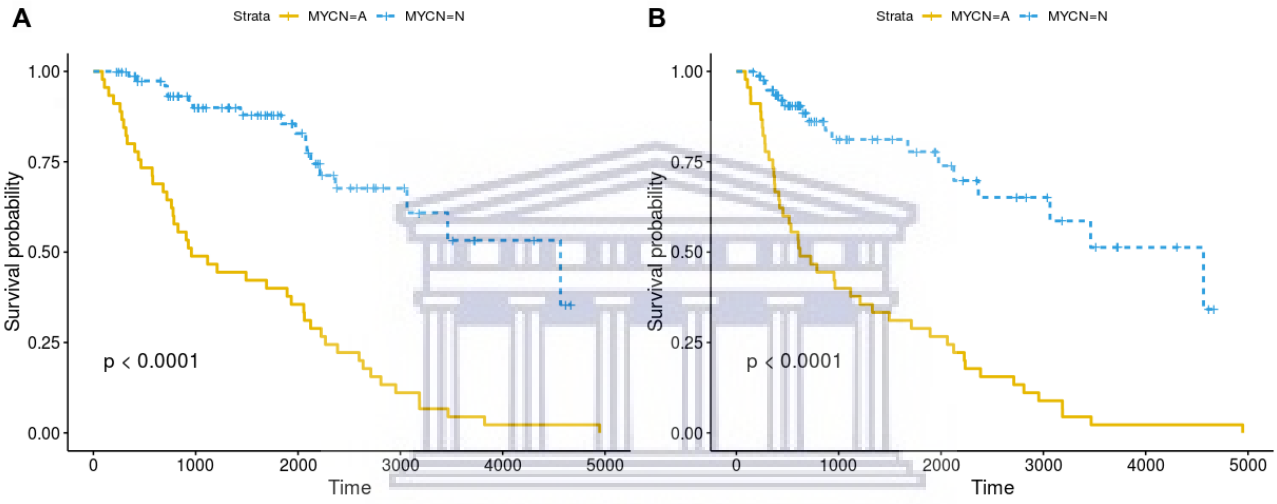
	GSE54719		**GSE120650**		**GSE65306**	
MYCN amplification	Precision	Recall	Precision	Recall	Precision	Recall
Yes	1.0	1.0	1.0	0.87	0.83	1.0
No	1.0	1.0	0.96	1.0	1.0	0.89
Accuracy	100% (35/35)		97% (56/58)		93% (26/28)	

Table 4.5: CpGs selected by the Cox regression model and their respective regression coefficients for overall survival (OS) and event-free survival (EFS). Eight CpGs and 5 CpGs were found to be significantly associated with OS and EFS respectively.

CpG	**Coefficient OS**	**DMG**	**CpG**	**Coefficient EFS**	**DMG**
cg00540828	-0.01461534	CUX1	cg01710189	-1.25426856	PDLIM2
cg01710189	-1.86558620	PDLIM2	cg13558971	0.56167125	ATP2B4
cg13558971	0.36211640	ATP2B4	cg25310824	0.46332724	SEPP1
cg25310824	0.53756846	SEPP1	cg07476617	0.44985062	CFLAR
cg07476617	0.51575895	CFLAR	cg12595667	-0.06647858	CXCR4
cg22886575	0.18633969	HMX2			
cg12595667	-0.21609796	CXCR4			
cg15455864	0.13284453	SYMPK			

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/



UNIVERSITY of the
WESTERN CAPE

Figure 4.4: Kaplan-Meier (K-M) estimates of (A) overall survival (OS) and (B) event-free survival (EFS) according to MYCN amplification and patient status values; 1 or 0 depending on whether CpG scores were above or below the median. Estimates for *MYCN* amplified patients are in yellow and non-amplified in blue.

http://etd.uwc.ac.za/

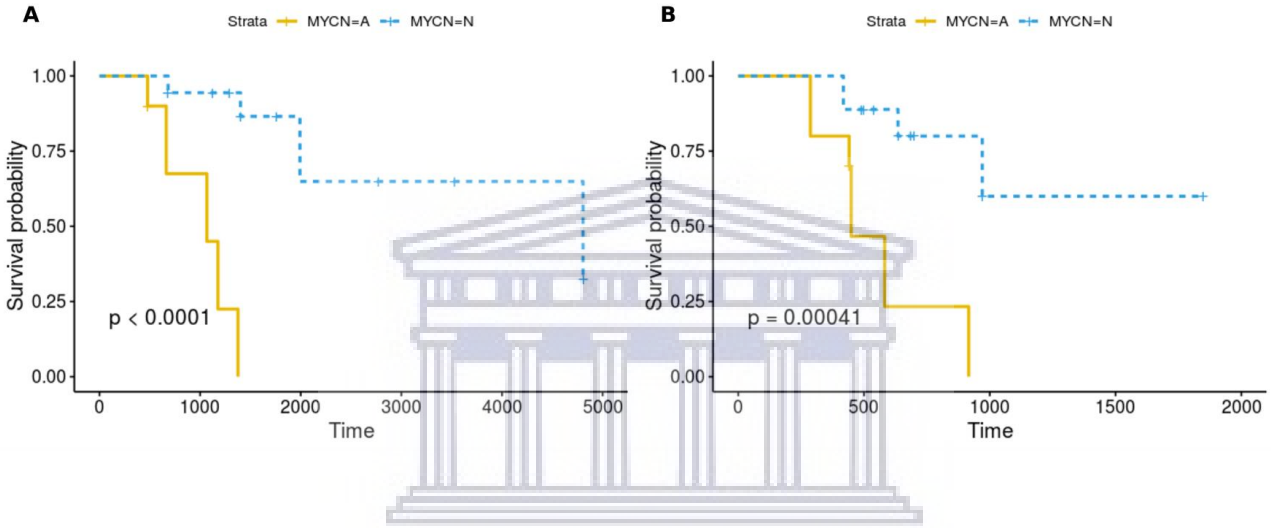


Figure 4.5: Kaplan-Meier (K-M) estimates for the test dataset GSE65306 of (A) overall survival (OS) and (B) event-free survival (EFS) according to *MYCN* amplification and patient status values; 1 or 0 depending on whether CpG scores were above or below the median. Estimates for *MYCN* amplified patients are in yellow and non-amplified in blue.

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

4.5 Discussion

We aimed at identifying predictive methylation biomarkers of the amplification of *MYCN* in neuroblastoma. The differential methylation analysis identified genes that have also been associated with neuroblastoma (Olsson et al., 2016; Decock et al., 2016; Duan et al., 2018) and other cancers (Li et al., 2018; Zhang et al., 2020). The differential methylation analysis and clustering results (Fig. 4.2 and 4.3) also demonstrate that *MYCN* amplification alters the methylation landscape in neuroblastoma and that this methylation landscape differs from that in *MYCN* non-amplified neuroblastoma. The accuracy of the clustering based on all significant CpGs (Fig. 4.3) validated the differential methylation results and demonstrated their possible utility as features for ML prediction.

Transcriptional regulation, which was enriched by the molecular function and biological process ontologies, is an activity that promotes tumor cell proliferation. Therefore, an increase in global transcription activity suggests rapid proliferation of tumor cells. We had also earlier suggested that cardiac disorders may be a cause of mortality in neuroblastoma patients (Giwa et al., 2020), and that this may be a treatment-related late effect (Friedman and Henderson, 2018). In addition, the ECM organization was also enriched in the GO analysis and this correlates with the fact that the ECM is a major structural component of the tumor microenvironment, and also plays an important role in tumor progression and tumor cell signalling (Walker et al., 2018; Poltavets et al., 2018; Eble and Niland, 2019; Nallanthigal et al., 2019).

Regardless of the stage of neuroblastoma cancer (i.e. INSS stage 1, 2, 3, 4 and 4s), the high accuracies obtained by the SVM model (Table 4.4), constructed on the basis of the 25 significant CpGs selected by RFE, demonstrate the validity of these CpGs for the diagnostic identification of *MYCN* amplification in neuroblastoma. The accuracy obtained from the prediction of *MYCN* amplification in both tumor and relapse samples of the GSE65306 dataset, indicates the specificity and accuracy of the CpGs in identifying *MYCN* amplification in neuroblastoma patients. The 2 misclassified samples

http://etd.uwc.ac.za/

were from the paired primary tumor and relapse sample of the same patient. The CpG sites found here can serve as methylation biomarkers of *MYCN* amplification and assist our understanding of tumorigenesis in *MYCN* amplified and *MYCN* non-amplified neuroblastoma. This knowledge could allow for the development of reliable and rapid diagnostic assessment of *MYCN* amplification with potential cost and time saving advantages. In addition, the methylated genes can serve as possible therapeutic targets as the amplification of *MYCN* is a known indicator of poor prognosis in neuroblastoma.

CIMP has also been used to assess the prognosis of patients with neuroblastoma. However, as noted in the introduction, CIMP phenotype can be observed in patients without *MYCN* amplification, which may therefore limit its utility as a biomarker for *MYCN* amplification. In addition, conflicting findings have been reported for the prognostic role of CIMP in CRC cancer due to differences in CIMP definitions (Jia et al., 2016). This issue of defining CIMP can also occur in neuroblastoma.

Loss of *CASP8* expression by methylation was believed to be a predictor of *MYCN* amplification (Teitz et al., 2000). However, in a large-scale study, no correlation was observed between the expression of *CASP8* and the amplification of *MYCN* (Fulda et al., 2006). This may explain why *CASP8* was not among the identified DMGs in our study, although we found that its paralog, *CFLAR,* was differentially methylated and associated with both overall and event-free survival. Being a biomarker for poor outcome, for example methylation of *RASSF1A* and *CASP8*, does not necessarily mean a biomarker for *MYCN* amplification as poor outcome can also be observed in non-amplified cases (Baali et al., 2018). Our method determined a CpG methylation signature specific for predicting *MYCN* amplification phenotype and poor prognosis in neuroblastoma.

By reducing the list of CpGs obtained from the ML analysis, the Cox regression model selected 8 CpGs related to patient overall survival (Table 4.5). The K-M curves (Fig. 4.4 and 4.5) show the

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

ability of these 8 CpGs to distinguish between shorter and longer OS i.e., *MYCN* amplified and non-amplified groups respectively, with a significant p-value<0.0001 and HR>1. Additionally, the results of the survival analysis (Fig. 4.4 and 4.5) validated the poor prognosis in patients with *MYCN* amplified (Brodeur et al., 1984; Seeger et al., 1985), and suggested 8 CpGs for the prognostic diagnosis of *MYCN* amplification. Since a CpG score above the median is associated with a shorter survival, higher methylation of the 3 CpGs with positive coefficients, cg13558971, cg25310824, cg07476617 from *ATP2B4*, *SEPP1* and *CFLAR* genes respectively (Table 4.5), predicts *MYCN* amplification as well as indicates poor prognosis outcome.

CFLAR is an apoptosis regulator, with significantly higher expression in lung cancer tissues (Zheng et al., 2019). Apoptosis resistance is one of the hallmarks of cancer initiation and progression (Hanahan and Weinberg, 2011) and the role of apoptosis in cancer and its potential as a cancer therapy target has been reviewed (Lowe and Lin, 2000; Pfeffer and Singh, 2018). *SEPP1* is involved in selenium transport and has been associated with neuroblastoma (Wang et al., 2020) and some other cancers, including prostate cancer (Gonzalez-Moreno et al., 2011), gastric adenocarcinoma (Wang et al., 2009) and renal cell cancer (Meyer et al., 2012). Similar to our results, Wang et al. (2020) also found *SEPP1* and thirteen other genes to be prognostic for overall survival in neuroblastoma. *ATP2B4* belongs to a family of plasma membrane pumps (Ca^{2+}-ATPases) involved in calcium homeostasis. Cellular processes important for tumorigenesis, such as proliferation, apoptosis, and angiogenesis are influenced by calcium ions (Monteith et al., 2007). The association of cg13558971 (*ATP2B4*) methylation with survival is suggestive of the importance of calcium transport in neuroblastoma. Satheesh and Busselberg (2015) reviewed the role of intracellular calcium in the development and treatment of neuroblastoma and Florea et al. (2017) confirmed the importance of calcium signaling in neuroblastoma cells in response to chemotherapy. The specific roles and mechanisms of these genes in neuroblastoma still need to be fully clarified, however our study shows that their methylation

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

is associated with amplification of *MYCN* and poor outcome. This knowledge could be used to develop a predictive biomarker panel to assess patient survival in *MYCN* amplified groups.

Among the fourteen highly methylated genes were genes related to neuroblastoma, and other cancers. *NXPH1* is a neuronal glycoprotein involved in neuronal differentiation that forms a complex with alpha neurexin proteins that promote adhesion between dendrites and axons. Decock et al. (2016) proposed *NXPH1* as a prognostic methylation biomarker for event-free survival in neuroblastoma. *NXPH1* may be promoting the growth of tumors by stimulating the proliferation of neuroblastoma stem cells. *SOX2-OT* is a long non-coding RNA with restricted expression towards the brain. It regulates *SOX2*, a key regulator of pluripotency. It has been associated with several cancers such as gastric cancer (Farhangian et al., 2018; Qu and Cao, 2018), pancreatic cancer (Li et al., 2018), hepatocellular carcinoma (Shi and Teng, 2015) and Cholangiocarcinoma (Li et al., 2018). *DLX5* is a homeobox-containing transcription factor that promotes neuronal differentiation, neural crest development and is essential for osteogenesis (Merlo et al., 2000; Heo et al., 2017). In acute myeloid leukemia (AML), promoter hypermethylation of *DLX5* led to reduced expression, and correlated with lesser overall survival (Zhang et al., 2020). Higher methylation of *DLX5* has also been reported in colorectal cancer tissue (Mitchell et al., 2014) and breast cancer (Karsli-Ceppioglu et al., 2017). *DLX6-AS1* is a long non-coding RNA that partially enhances the proliferation, migration and invasive abilities of neuroblastoma cells (Li et al., 2020). Olsson et al. (2016) reported similar results with the hypermethylation of *DLX5* and *DLX6-AS1* in aggressive INRG stage M neuroblastoma tumors.

TFAP2D belongs to the activator protein-2 transcription factor family that are essential in cellular processes such as apoptosis, migration and differentiation, and have been implicated in cancer (Sun et al., 2014; Li et al., 2018). Hypermethylation of a member of its protein family, *TFAP2E*, is associated with clinical non-responsiveness to chemotherapy in colorectal cancer (Ebert et al., 2012) and acts as a tumor suppressor in neuroblastoma (Hoshi et al., 2017). *TFAP2D* may also possibly be

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

acting the same role and further studies are required to establish this claim. *TERT* rearrangements was reported to be the most frequent genetic alteration after *MYCN* amplification in high-risk neuroblastoma (Valentijn et al., 2015; Peifer et al., 2015). Consistent with our results, Olsson et al. (2016) also reported hypermethylation of *TERT* in aggressive tumors of INRG stage M neuroblastoma. Cancer cells activate *TERT* to maintain their telomeres (Valentijn et al., 2015). *TMCO3* belongs to a family of transporter proteins involved in the coupling of export of monovalent cations such as potassium or sodium to import protons across the cellular membrane. Mutations in *TMCO3* alongside other genes has been suggested as markers for evaluating the effects of chemotherapy in neuroblastoma patients (Duan et al., 2018). Methylation of *TMCO3* may likely be associated with a more favorable prognosis in neuroblastoma.

CAVIN3 plays an important role in the protein kinase c-delta tumor suppression pathway (Lee et al., 2008). The loss of *CAVIN3* has been observed in different cancers, such as breast (Roy et al., 2003), cervix (Kozlowski et al., 2006), bladder (Panani et al., 2004), lung (Zhao and Bepler, 2001) and stomach (Moskaluk and Rumpel, 1998) cancers. It is commonly altered in colorectal cancer by promoter hypermethylation (Lee et al., 2011). It is also downregulated in breast cancer, possibly due to methylation (Wikman et al., 2012). *KRT19* is highly expressed in multiple cancers serving as a diagnostic marker (Sharma et al., 2019). High *KRT19* expression is associated with clinical progression in lung cancer (Yuan et al., 2021) and correlates with poor prognosis in breast cancer (Saloustros et al., 2011; Kabir et al., 2014). Further studies are necessary to uncover the precise roles of these genes in neuroblastoma.

In this study, we aimed to propose a CpG signature that accurately predicts *MYCN* amplification and is associated with patient survival. Our method, summarized in (Fig. 4.1), generated a CpG signature capable of predicting *MYCN* amplification with high precision. We used the ChAMP R package (Morris et al., 2014; Tian et al., 2017) to discover DMGs between *MYCN* amplified and non-amplified

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

neuroblastoma tumors and the ChAMP.DMP() function was designed to compare methylation between two phenotypes. The high prediction accuracy scores obtained from the SVM model demonstrates the correctness of the obtained CpG signature, with the number of the significant CpGs used in the prediction selected by RFE. DNA methylation based diagnostic methods is increasingly used in clinical practice (Sahm et al., 2018; Capper et al., 2018, Liu et al., 2020). Aberrant DNA methylation patterns have been observed and reported to play a role in neuroblastoma pathogenesis (Abe et al., 2005). It is expected that epigenetic prognostics and therapeutics will be commonplace in the clinic in the near future. This is a purely computational based study. Although some of the DMGs identified here have already been associated with neuroblastoma, functional studies are still needed to discover the role of certain DMGs in the pathogenesis of neuroblastoma.

4.6 Conclusion

This study demonstrated the utility of CpG methylation profiling for subgrouping neuroblastoma tumors and for predicting clinical events. We identified 25 CpGs capable of stratifying neuroblastoma samples on the basis of *MYCN* amplification status, thereby demonstrating their ability to be used as diagnostic indicators of *MYCN* amplification in neuroblastoma. We also evaluated the impact of the 25 CpGs methylation on the survival of patients with neuroblastoma using Cox regression analysis and identified a list of 8 CpGs associated with survival in neuroblastoma. The DMGs reported in this study include some known genes associated with neuroblastoma as well as novel ones. This study furthers our understanding of the mechanisms of tumor progression in neuroblastoma. Therapeutic interventions may need to be targeted at patient subgroups level in order to optimize treatment outcomes and improve survival.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Chapter 5

5.0 Conclusions and further considerations

5.1 Summary

5.1.1 What was done

We requested for and obtained approval to access the TARGET neuroblastoma dataset (phs000218.v22.p8) in NCBI Database of Genotypes and Phenotypes (dbGaP). The data was accessed through https://www.globus.org and downloaded to the SANBI compute cluster. The results were then reviewed in conjunction with the original study, Pugh et al. (2013). TARGET gene expression RNA-Seq counts on the Xena database were then programmatically accessed. Differential gene expression analysis, Gene Ontology enrichment analysis, gene regulatory network analysis and machine learning techniques were applied to the gene expression dataset to uncover a novel genetic signature prognostic of survival time in high-risk neuroblastoma. We also accessed Neuroblastoma tumor methylation data from the TARGET data matrix. Differential methylation analysis, clustering analysis, Cox regression & Kaplan-Meier estimates, and machine learning prediction were applied to the methylation dataset to uncover a novel methylation signature diagnostic of *MYCN* amplification, and prognostic of survival and poor prognosis in neuroblastoma.

http://etd.uwc.ac.za/

5.1.2 What was found

As reported in Pugh et al. (2013) and in the pediatric oncology community, there is relative paucity of mutations in neuroblastoma. Analysis of other omics data types such as the transcriptome and methylome is therefore critical and necessary. A 16-gene (*EVX2, NHLH2, PRSS12, POU6F2, HOXD10, MAPK15, RTL1, LGR5, CYP17A1, OR10AB1P, MYH14, LRRTM3, GRIN3A, HS3ST5, CRYAB and NXPH3*) expression signature was discovered capable of predicting survival time in high-risk neuroblastoma. It was also uncovered that the less studied gene *SMIM28* may possibly have a role in the proliferation and aggressiveness of neuroblastoma tumors. 663 CpGs of 369 genes were found to be differentially methylated between *MYCN* amplified and *MYCN* non-amplified samples. Application of machine learning techniques found a 25-CpG signature to be diagnostic of the amplification and non-amplification of *MYCN* in neuroblastoma. Eight CpGs were also found associated with survival, with methylation of cg13558971 (*ATP2BP4*), cg25310824 (*SEPP1*) and cg07476617 (*CFLAR*), indicative of poor outcome in *MYCN* amplified neuroblastoma. Some of the differentially expressed and methylated genes reported in this research project are associated with neuroblastoma while many are novel, being reported for the first time.

UNIVERSITY of the
WESTERN CAPE

5.2 Potential future research

The novel genes identified in this project provide opportunities for the development of functional validation assays to assess their roles in neuroblastoma. This can be through designing experiments where the expression of these genes are knockdown or silenced using molecular techniques such as RNA interference, small molecule inhibitors and CRISPR-Cas9 on neuroblastoma cell lines and Mouse models. Epigenetic targeting of these novel genes with DNA methyltransferase inhibitors and histone deacetylase inhibitors can also be explored. It also opens the possibility of combinatorial therapy where genes can be targeted at multiple levels as described above, at the same time.

http://etd.uwc.ac.za/

5.3 Potential clinical applications

The results obtained in this project has strong potential to be clinically used to stratify neuroblastoma patients. This knowledge can improve diagnosis and prognosis of clinical events in neuroblastoma patients and can thus be translated to clinical use. It can be used in combination with current clinical methods in follow-up or management of neuroblastoma patients.

5.3.1 Clinical applications: Immediate

Based on the results of this research project, a pediatric oncologist can obtain gene expression and methylation data from a neuroblastoma patient at diagnosis, and using machine learning techniques as reported in this study, determine if that patient has or does not have *MYCN* amplification, as well as make a combined survival time prognosis using both gene expression and methylation data. This information can then influence clinical management decisions by the pediatric oncologist with the goal to ensure survival with possible minimal toxicity (Figure 5.1).

UNIVERSITY of the
WESTERN CAPE

5.3.2 Clinical applications: Future

In the future, standardized laboratory tests and kits can be developed for first line use during diagnosis and management of neuroblastoma patients. Combining the gene expression and methylomic signatures should improve the prognostic performance of such standardized tests. The utility of applying machine learning algorithms to methylation data is fast growing. Liu et al. (2020) developed an exciting methylation-based blood test that detects more than 50 cancer types across stages with high specificity. This test was based on applying machine learning algorithms on methylation data from cell free DNA in the blood of individuals. We are moving to a point in medicine where methylation-based biomarkers and epigenetic therapeutics will be commonplace in the cancer therapeutic space. The results of this research project are in line with this vision.

http://etd.uwc.ac.za/

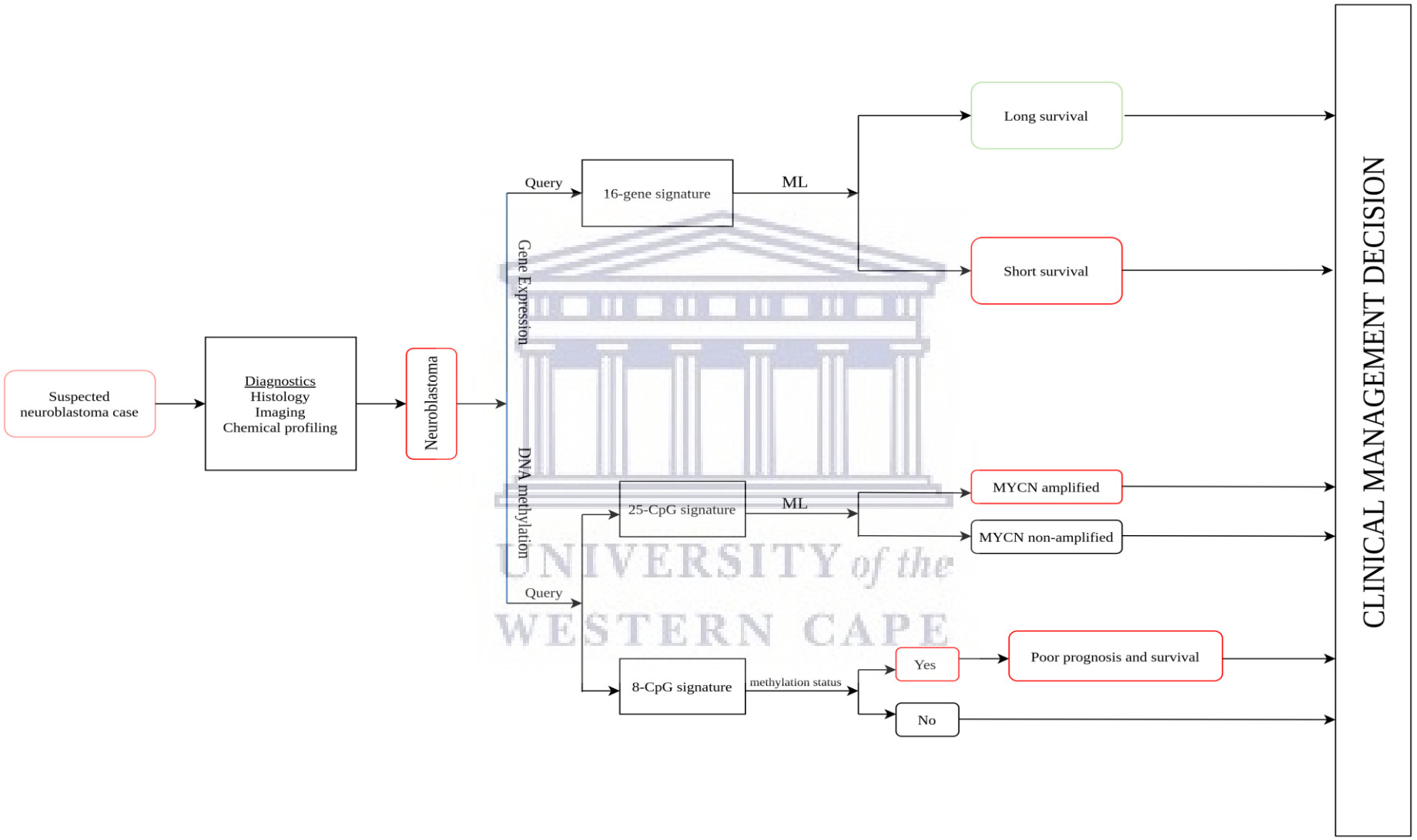


Figure 5.1: Summary of the potential clinical application of the findings of this research project

http://etd.uwc.ac.za/

Bibliography

Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021; 71:209-249. https://doi.org/10.3322/caac.21660

Klein G. Oncogenes and tumor suppressor genes, Acta Oncologica. 1988; 27:427–437, https://doi.org/10.3109/02841868809093569.

Stass SA, Mixson J. Oncogenes and tumor suppressor genes: therapeutic implications. Clin Cancer Res. 1997; 3:2687–2695.

Guo XE, Ngo B, Modrek AS, Lee WH. Targeting tumor suppressor networks for cancer therapeutics. Curr Drug Targets. 2014; 15:2-16. https://doi.org/10.2174/1389450114666140106095151.

Vineis P, Wild CP. Global cancer patterns: causes and prevention. Lancet. 2014; 383:549-557. https://doi.org/10.1016/S0140-6736(13)62224-2.

Moses C, Garcia-Bloj B, Harvey AR, Blancafort P. Hallmarks of cancer: The CRISPR generation. Eur J Cancer. 2018; 93:10 – 18. https://doi.org/10.1016/j.ejca.2018.01.002.

Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell. 2011; 144:646–674. https://doi.org/10.1016/j.cell.2011.02.013.

Pavlova NN, Thompson CB. The emerging hallmarks of cancer metabolism. Cell Metab. 2016; 23:27-47. https://doi.org/10.1016/j.cmet.2015.12.006.

Kaatsch P. Epidemiology of childhood cancer. Cancer Treatment Reviews, 2010; 36:277–285. https://doi.org/10.1016/j.ctrv.2010.02.003.

Steilarova-Foucher E, Colombet M, Reis LAG, Moreno F, Dolya A, Bray F, Hesseling P, Shin HY, Stiller CA, IICC-3 contributors. International incidence of childhood cancer, 2001-10: a population-based registry study. Lancet Oncol. 2017; 18:719–731. https://doi.org/10.1016/S1470-2045(17)30186-9.

Spector LG, Pankratz N, Marcotte EL. Genetic and nongenetic risk factors for childhood cancer. Pediatr Clin North Am. 2015; 62:11-25. https://doi.org/10.1016/j.pcl.2014.09.013.

Bhakta N, Force LM, Allemani C, Atun R, Bray F, Coleman MP, Steliarova-Foucher E, Frazier AL, Robison LL, Rodriguez-Galindo C, Fitzmaurice C. Childhood cancer burden: a review of global estimates. Lancet Oncol. 2019; 20:e42-e53. https://doi.org/10.1016/S1470-2045(18)30761-7.

Ness KK, Gurney JG. Adverse late effects of childhood cancer and its treatment on health and performance. Annu Rev Public Health. 2007; 28:279-302. https://doi.org/10.1146/annurev.publhealth.28.021406.144049.

Ward E, DeSantis C, Robbins A, Kohler B, Jemal A. 2014. Childhood and adolescent cancer statistics, 2014. CA Cancer J Clin. 2014; 64:83–103. https://doi.org/10.3322/caac.21219.

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013; 499:214-218. https://doi.org/10.1038/nature12213

Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. Science. 2013; 339:1546-1558. https://doi.org/10.1126/science.1235122

Campbell BB, Light N, Fabrizio D, Zatzman M, Fuligni F, de Borja R, et al. Comprehensive analysis of hypermutation in human cancer. Cell. 2017; 171:1042-1056. https://doi.org/10.1016/j.cell.2017.09.048.

Sweet-Cordero EA, Biegel JA. The genomic landscape of pediatric cancers: implications for diagnosis and treatment. Science. 2019; 363:1170-1175. https://doi.org/10.1126/science.aaw3535.

Maris JM Recent advances in neuroblastoma. N Engl J Med. 2010; 362:2202–2211. https://doi.org/10.1056/NEJMra0804577

Smith MA, Seibnel NL, Altekruse SF, Ries LA, Melbert DL, O'Leary M, Smith FO, Reaman GH. Outcomes for children and adolescents with cancer: challenges for the twenty-first century. J Clin Oncol. 2010; 28:2625–2634. https://doi.org/10.1200/JCO.2009.27.0421.

Maris JM, Hogarty MD, Bagatell R, Cohn SL. Neuroblastoma. Lancet. 2007; 369:2106-2120. https://doi.org/10.1016/S0140-6736(07)60983-0.

Zhang L, Lv C, Jin Y, Cheng G, Fu Y, Yuan D, Tao Y, Guo Y, Ni X, Shi T. Deep learning-based multi-omics data integration reveals two prognostic subtypes in high-risk neuroblastoma. Front Genet. 2018; 9:1-9. https://doi.org/10.3389/fgene.2018.00477.

Holland T, Donohue JP, Baehner RL, Grosfeld JL. The current management of neuroblastoma. The Journal of Urology. 1980; 124:579-582.

Weinstein JL, Katzenstein HM, Cohn SL. Advances in the diagnosis and treatment of neuroblastoma. The Oncologist. 2003; 8:278-292. https://doi.org/10.1634/theoncologist.8-3-278.

Swift CC, Eklund MJ, Kraveka JM, Alazraki AL. Updates in diagnosis, management, and treatment of neuroblastoma. RadioGraphics. 2018; 38:566–580. https://doi.org/10.1148/rg.2018170132.

Baali I, Acar EA, Aderinwale TW, HafezQorani S, Kazan H. Predicting clinical outcomes in neuroblastoma with genomic data integration. Biology Direct. 2018; 13:20. https://doi.org/10.1186/s13062-018-0223-8.

Newman E, Nuchtern J. Recent biologic and genetic advances in neuroblastoma: implications for diagnostic, risk stratification, and treatment strategies. Semin Pediatr Surg. 2016; 25:257-264. https://doi.org/10.1053/j.sempedsurg.2016.09.007.

Oeffinger KC, Mertens AC, Sklar CA, Kawashima T, Hudson MM, Meadows AT, Friedman DL, Marina N, Hobbie W, Kadan-Lottick NS, Schwartz CL, Leisenring W, Robison LL; Childhood Cancer Survivor Study. Chronic health conditions in adult survivors of childhood cancer. N Engl J Med. 2006; 355:1572-1582. https://doi.org/10.1056/NEJMsa060185.

Yu AL, Gilman AL, Ozkaynak MF, London WB, Kreissman SG, Chen HX, Smith M, Anderson B, Villablanca JG, Matthay KK, Shimada H, Grupp SA, Seeger R, Reynolds CP, Buxton A, Reisfeld

http://etd.uwc.ac.za/

RA, Gillies SD, Cohn SL, Maris JM, Sondel PM. Anti-GD2 antibody with GM-CSF, interleukin-2, and isotretinoin for neuroblastoma. N Engl J Med. 2010; 363:1324–1334. https://doi.org/10.1056/NEJMoa0911123.

London WB, Matthay KK, Ambros PF, Monclair T, Pearson AD, Cohn SL, Castel V. Clinical and biological features predictive of survival after relapse of neuroblastoma: A study from the International Neuroblastoma (NB) Risk Group (INRG) Database. J Clin Oncol. 2010;28:9518.

Barr EK, Applebaum MA. Genetic predisposition to neuroblastoma. Children (Basel). 2018; 5:119. https://doi.org/10.3390/children5090119.

Perakakis N, Yazdani A, Karniadakis GE, Mantzoros C. Omics, big data and machine learning as tools to propel understanding of biological mechanisms and to discover novel diagnostics and therapeutics. Metabolism. 2018; 87:A1-A9. https://doi.org/10.1016/j.metabol.2018.08.002.

Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome Biol. 2017; 18:83. https://doi.org/10.1186/s13059-017-1215-1.

Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet. 2005; 6:95-108. https://doi.org/10.1038/nrg1521.

Ragoussis J. Genotyping techniques for genetic research. Annu Rev Genomics Hum Genet. 2009; 10:117-133. https://doi.org/10.1146/annurev-genom-082908-150116.

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009; 461:272-276. https://doi.org/10.1038/nature08250.

Koboldt, DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The next-generation seuqencing revolution and its impacts in genomics. Cell. 2013; 155:27-38. https://doi.org/10.1016/j.cell.2013.09.006.

Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet. 2007; 8:286-298. https://doi.org/10.1038/nrg2005.

Blow, N. The digital generation. Nature. 2009; 458:239–240. https://doi.org/10.1038/458239a.

Sager M, Yeat NC, Pajaro-Van der Stadt S, Lin C, Ren Q, Lin J. Transcriptomics in cancer diagnostics: developments in technology, clinical research and commercialization. Expert Rev Mol Diagn. 2015; 15:1589-1603. https://doi.org/10.1586/14737159.2015.1105133.

Aslam B, Basit M, Nisar MA, Khurshid M, Rasool MH. Proteomics: Technologies and Their Applications. J Chromatogr Sci. 2017; 55:182-196. https://doi.org/10.1093/chromsci/bmw167.

Pinu FR, Goldansaz SA, Jaine J. Translational Metabolomics: Current Challenges and Future Opportunities. Metabolites. 2019; 9:108. https://doi.org/10.3390/metabo9060108.

Brademan DR, Miller IJ, Kwiecien NW, Pagliarini DJ, Westphall MS, Coon JJ, Shishkova E. Argonaut: a web platform for collaborative multi-omic data visualization and exploration. Patterns. 2020; 1:100122. https://doi.org/10.1016/j.patter.2020.100122.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, et al. A landscape of pharmacogenomic interactions in cancer. Cell. 2016; 166:740-754. https://doi.org/10.1016/j.cell.2016.06.017.

Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell, 2017; 171:1437-1452. https://doi.org/10.1016/j.cell.2017.10.049.

Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics Data Integration, Interpretation, and Its Application. Bioinform Biol Insights. 2020; 14:1177932219899051. https://doi.org/10.1177/1177932219899051.

Yang Y, Dong X, Xie B, Ding N, Chen J, Li Y, Zhang Q, Qu H, Fang X. Databases and web tools for cancer genomics study. Genomics Proteomics Bioinformatics. 2015; 13:46-50. https://doi.org/10.1016/j.gpb.2015.01.005.

Pavlopoulou A, Spandidos DA, Michalopoulos I. Human cancer databases (review). Oncol Rep. 2015; 33:3-18. https://doi.org/10.3892/or.2014.3579.

Chen R. On bioinformatic resources. Genomics Proteomics Bioinformatics. 2015; 13:1-3. https://doi.org/10.1016/j.gpb.2015.02.002.

Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol. 2020; 38:675-678. https://doi.org/10.1038/s41587-020-0546-8.

Zhu Q, Yang H, Cheng P, Han Q. Bioinformatic analysis of the prognostic value of the lncRNAs encoding snoRNAs in hepatocellular carcinoma. BioFactors. 2019; 45:244-252. https://doi.org/10.1002/biof.1478.

Wu J, Li M, Zhang Y. Long noncoding RNA HOXA-AS2 regulates the expression of *SCN3A* by sponging *miR-106a* in breast cancer. J Cell Biochem. 2019; 120:14465-14475. https://doi.org/10.1002/jcb.28706.

Barman UD, Saha SK, Kader MA, Jamal MM, Sharma SP, Samad A, Rahman MS. Clinicopathological and prognostic significance of GPC3 in human breast cancer and its 3D structure predicition. *Netw Model Anal Health Inform Bioinforma*. 2020; 9:24. https://doi.org/10.1007/s13721-020-00234-x.

Giwa A, Fatai A, Gamieldien J, Christoffels A, Bendou H. Identification of novel prognostic markers of survival time in high-risk neuroblastoma using gene expression profiles. Oncotarget. 2020; 11:4293-4305. https://doi.org/10.18632/oncotarget.27808.

Kang T, Yau C, Wong CK, Sanborn JZ, Newton Y, Vaske C, et al. A risk-associated active transcriptome phenotype expressed by histologically normal human breast tissue and linked to a pro-tumorigenic adipocyte population. Breast Cancer Res. 2020; 22:81. https://doi.org/10.1186/s13058-020-01322-6.

Zheng S, Fu Y. Age-related copy number variation and expression levels of F-box protein FBXL20 predict ovarian cancer prognosis. Transl Oncol. 2020; 13:100863. https://doi.org/10.1016/j.tranon.2020.100863.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Zhang M, Chen H, Wang M, Bai F, Wu K. Bioinformatics analysis of prognostic significance of COL10A1 in breast cancer. Biosci Rep. 2020; 40:BSR20193286. https://doi.org/10.1042/BSR20193286.

Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, et al. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet. 2007; 39:1181-1186. https://doi.org/10.1038/ng1007-1181.

Tryka KA, Hao L, Sturcke A, Jin Y, Wang ZY, Ziyabari L, et al. NCBI's Database of Genotypes and Phenotypes: dbGaP. Nucleic Acids Res. 2014; 42(D1):D975–D979, https://doi.org/10.1093/nar/gkt1211.

Wong KM, Langlais K, Tobias GS, Fletcher-Hoppe C, Krasnewich D, Leeds HS, et al. The dbGaP data browser: a new tool for browsing dbGaP controlled-access genomic data. Nucleic Acids Res. 2017; 45:D819-D826. https://doi.org/10.1093/nar/gkw1139.

Pugh TJ, Morozova O, Attiyeh EF, Asgharzadeh S, Wei JS, Auclair D, et al. The genetic landscape of high-risk neuroblastoma. Nat Genet. 2013; 45:279-284. https://doi.org/10.1038/ng.2529.

Oldridge DA, Wood AC, Weichert-Leahey N, Crimmins I, Sussman R, Winter C, et al. Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism. Nature. 2015; 528:418-421. https://doi.org/10.1038/nature15540.

Eleveld TF, Oldridge DA, Bernard V, Koster J, Colmet Daage L, Diskin SJ, et al. Relapsed neuroblastomas show frequent RAS-MAPK pathway mutations. Nat Genet. 2015; 47:864-871. https://doi.org/10.1038/ng.3333.

Wei JS, Kuznetsov IB, Zhang S, Song YK, Asgharzadeh S, Sindiri S, et al. Clinically relevant cytotoxic immune cell signatures and clonal expansion of T-cell receptors in high-risk MYCN-not-amplified human neuroblastoma. Clin Cancer Res. 2018; 24:5673-5684. https://doi.org/10.1158/1078-0432.CCR-18-0599.

Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013; 45:1113-1120. https://doi.org/10.1038/ng.2764.

Wang Z, Jensen MA, Zenklusen JC. A Practical Guide to The Cancer Genome Atlas (TCGA). Methods Mol Biol. 2016; 1418:111-141. https://doi.org/10.1007/978-1-4939-3578-9_6.

Gao GF, Parker JS, Reynolds SM, Silva TC, Wang LB, Zhou W, et al. Before and after: comparison of legacy and harmonized TCGA Genomic Data Commons' data. Cell Syst. 2019; 9:24-34. https://doi.org/10.1016/j.cels.2019.06.006.

Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 2016; 44:e71. https://doi.org/10.1093/nar/gkv1507.

Wan YW, Allen GI, Liu Z. TCGA2STAT: simple TCGA data access for integrated statistical analysis in R. Bioinformatics. 2016; 32:952–954. https://doi.org/10.1093/bioinformatics/btv677.

Clough E, Barrett T. The Gene Expression Omnibus Database. Methods Mol Biol. 2016; 1418:93-110. https://doi.org/10.1007/978-1-4939-3578-9_5.

http://etd.uwc.ac.za/

Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002; 30:207-210. https://doi.org/10.1093/nar/30.1.207.

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. 2013; 41:D991-D995. https://doi.org/10.1093/nar/gks1193.

Roy S, Sharma P, Nath K, Bhattacharyya DK, Kalita JK. Pre-Processing: A Data Preparation Step. In: Guenther R, Steel D. (eds.). Encyclopedia of Bioinformatics and Computational Biology. 2019; 1:463–471. Oxford:Elsevier. https://doi.org/10.1016/B978-0-12-809633-8.20457-3

Lee S, Cheung DW, Kao B. Is Sampling Useful in Data Mining? A Case in the Maintenance of Discovered Association Rules. Data Min Knowl Discov. 1998; 2:233–262. https://doi.org/10.1023/A:1009703019684.

Ali H, Mohd Salleh N, Hussain K, Ahmad A, Ullah A, Muhammad A, Naseem R, Khan M. A review on data preprocessing methods for class imbalance problem. Int J Eng Technol. 2019; 8:390-397.

Mishra S, Mallick PK, Jena L, Chae GS. Optimization of Skewed Data Using Sampling-Based Preprocessing Approach. Front Public Health. 2020; 8:274. https://doi.org/10.3389/fpubh.2020.00274.

Hotelling H. 1933. Analysis of a complex of statistical variables into principal components. J Edu Psych. 1933; 24:498–520. https://doi.org/10.1037/h0070888.

Lever J, Krzywinski M, Altman N. Principal component analysis. Nat Methods. 2017; 14:641–642. https://doi.org/10.1038/nmeth.4346.

Schmitt P, Mandel J, Gued M. A comparison of six methods for missing data imputation. J Biomet Biostat 2015, 6:1. https://doi.org/10.472/2155-6180.1000224.

Song M, Greenbaum J, Luttrell J 4th, Zhou W, Wu C, Shen H, Gong P, Zhang C, Deng HW. A Review of Integrative Imputation for Multi-Omics Datasets. Front Genet. 2020; 11:570255. https://doi.org/10.3389/fgene.2020.570255.

Baumgartner C, Osl M, Netzer M, Baumgartner D. Bioinformatic-driven search for metabolic biomarkers in disease. J Clin Bioinforma. 2011; 1:2. https://doi.org/10.1186/2043-9113-1-2.

Quackenbush J. Microarray data normalization and transformation. Nat Genet. 2002; 32:496-501. https://doi.org/10.1038/ng1032.

Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11:R106. https://doi.org/10.1186/gb-2010-11-10-r106.

Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics. 2003; 19:185-193. https://doi.org/10.1093/bioinformatics/19.2.185.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Chawade A, Alexandersson E, Levander F. Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets. J Proteome Res. 2014; 13:3114-3120. https://doi.org/10.1021/pr401264n.

Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. Nat Genet. 2006; 38:500-501. https://doi.org/10.1038/ng0506-500.

Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 2018; 46(W1):W537-W544. https://doi.org/10.1093/nar/gky379.

Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods. 2015; 12:115-121. https://doi.org/10.1038/nmeth.3252.

Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szcześniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A. A survey of best practices for RNA-seq data analysis. Genome Biol. 2016; 17:13. https://doi.org/10.1186/s13059-016-0881-8

Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139–140. https://doi.org/10.1093/bioinformatics/btp616.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15:550. https://doi.org/10.1186/s13059-014-0550-8.
Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43: e47. https://doi.org/10.1093/nar/gkv007.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43: e47. https://doi.org/10.1093/nar/gkv007.

UNIVERSITY of the
WESTERN CAPE

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012; 7:562-578. https://doi.org/10.1038/nprot.2012.016.

Tarazona S, García-Alcalde F, Dopazo J, Ferrer A, Conesa A. Differential expression in RNA-seq: a matter of depth. Genome Res. 2011; 21:2213-2223. https://doi.org/10.1101/gr.124321.111.

Park Y, Wu H. Differential methylation analysis for BS-seq data under general experimental design. Bioinformatics. 2016; 32:1446-1453. https://doi.org/10.1093/bioinformatics/btw026.

Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. ChAMP: 450k Chip Analysis Methylation Pipeline. Bioinformatics. 2014; 30:428–430. https://doi.org/10.1093/bioinformatics/btt684.

Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, Teschendorff AE. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. Bioinformatics. 2017; 33:3982-3984. https://doi.org/10.1093/bioinformatics/btx513.

http://etd.uwc.ac.za/

Assenov Y, Müller F, Lutsik P, Walter J, Lengauer T, Bock C. Comprehensive analysis of DNA methylation data with RnBeads. Nat Methods. 2014; 11:1138-1140. https://doi.org/10.1038/nmeth.3115.

Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. 2012; 13:R87. https://doi.org/10.1186/gb-2012-13-10-r87.

Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014; 30:1363-1369. https://doi.org/10.1093/bioinformatics/btu049.

Mitchell TM. Machine Learning. McGraw-Hill, New York, 1997.

DeGregory KW, Kuiper P, DeSilvio T, Pleuss JD, Miller R, Roginski JW, Fisher CB, Harness D, Viswanath S, Heymsfield SB, Dungan I, Thomas DM. A review of machine learning in obesity. Obes Rev. 2018; 19:668-685. https://doi.org/10.1111/obr.12667.

Zou J, Huss M, Abid A, Mohammadi P, Torkamani A, Telenti A. A primer on deep learning in genomics. Nat Genet. 2019; 51:12-18. https://doi.org/10.1038/s41588-018-0295-5.

Kotsiantis SB. Supervised machine learning: a review of classification techniques. Informatica. 2007; 31:249-268.

Yousef M, Najami N, Abedallah L, Khalifa W. Computational approaches for biomarker discovery. Journal Intell Learn Syst Appl. 2014; 6:153-161. https://doi.org/10.4236/jilsa.2014.64012

Yuan Y, Shaw MJ. Induction of fuzzy decision trees. Fuzzy Sets and Systems. 1995; 69:125-139.https://doi.org/10.1016/0165-0114(94)00229-Z

Vapnik V. The nature of statistical learning theory. Springer, New York, 1995. https://doi.org/10.1007/978-1-4757-2440-0.

McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophy. 1943; 5:115–133.

Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Machine Learning. 2002; 46:389-422. https://doi.org/10.1023/A:1012487302797

Krishnapuram B, Hartemink AJ, Carin L, Figueiredo MA. A Bayesian approach to joint feature selection and classifier design. IEEE Trans Pattern Anal Mach Intell. 2004; 26:1105-1111. https://doi.org/10.1109/TPAMI.2004.55.

Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res. 2003; 3: 1157–1182.

Agrawal S, Agrawal J. Neural network techniques for cancer prediction: a survey. Proc Comput Sci. 2015; 50:769-774.

Nagy M, Radakovich N, Nazha A. Machine Learning in Oncology: What Should Clinicians Know? JCO Clin Cancer Inform. 2020; 4:799-810. https://doi.org/10.1200/CCI.20.00049.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Lopez C, Tucker S, Salameh T, Tucker C. An unsupervised machine learning method for discovering patient clusters based on genetic signatures. J Biomed Inform. 2018; 85:30-39. https://doi.org/10.1016/j.jbi.2018.07.004.

Xu C, Jackson SA. Machine learning and complex biological data. Genome Biol. 2019; 20:76. https://doi.org/10.1186/s13059-019-1689-0.

MacQueen JB. Some Methods for Classification and Analysis of Multivariate Observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, 1967; 1:281-297.

Dunn JC. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J Cybern. 1973; 3:32-57. https://doi.org/10.1080/01969727308546046.

Bezdek JC. Pattern recognition with fuzzy objective function algoritms. Plenum Press, New York, 1981. http://dx.doi.org/10.1007/978-1-4757-0450-1.

Johnson SC. Hierarchical clustering schemes. Psychometrika. 1967; 32:241-254. http://dx.doi.org/10.1007/BF02289588.

Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data via the EM Algorithm. J R Stat Soc Series B Methodol. 1977; 39:1-38.

Witten DM. Penalized unsupervised learning with outliers. Stat Interface. 2013; 6:211-221. https://doi.org/10.4310/sii.2013.v6.n2.a5.

Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J. 2015; 13:8-17. https://doi.org/10.1016/j.csbj.2014.11.005.

Zhu X, Goldberg AB. Introduction to Semi-Supervised Learning. Morgan and Claypool, USA, 2009, 130pp. https://doi.org/10.2200/S00196ED1V01Y200906AIM006.

Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. Nat Rev Genet. 2015; 16:321-332. https://doi.org/10.1038/nrg3920.

Telenti A, Lippert C, Chang PC, DePristo M. Deep learning of genomic variation and regulatory network data. Hum Mol Genet. 2018; 27(R1):R63-R71. https://doi.org/10.1093/hmg/ddy115.

Koumakis L. Deep learning models in genomics; are we there yet? Comput Struct Biotechnol J. 2020; 18:1466-1473. https://doi.org/10.1016/j.csbj.2020.06.017.

Talukder A, Barham C, Li X, Hu H. Interpretation of deep learning in genomics and epigenomics. Brief Bioinform. 2021; 22:1-16. https://doi.org/10.1093/bib/bbaa177.

Washburn JD, Mejia-Guerra MK, Ramstein G, Kremling KA, Valluru R, Buckler ES, Wang H. Evolutionarily informed deep learning methods for predicting relative transcript abundance from DNA sequence. Proc Natl Acad Sci U S A. 2019; 116:5542-5549. https://doi.org/10.1073/pnas.1814551116.

Chen Y, Li Y, Narayan R, Subramanian A, Xie X. Gene expression inference with deep learning. Bioinformatics. 2016; 32:1832-1839. https://doi.org/10.1093/bioinformatics/btw074.

http://etd.uwc.ac.za/

Xie R, Wen J, Quitadamo A, Cheng J, Shi X. A deep auto-encoder model for gene expression prediction. BMC Genomics. 2017; 18:845. https://doi.org/10.1186/s12864-017-4226-0.

Zhou J, Theesfeld CL, Yao K, Chen KM, Wong AK, Troyanskaya OG. Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. Nat Genet. 2018; 50:1171-1179. https://doi.org/10.1038/s41588-018-0160-6.

Barshai M, Tripto E, Orenstein Y. Identifying regulatory elements via deep learning. Annu. Rev. Biomed. Data Sci. 2020; 3:315–338. https://doi.org/10.1146/annurev-biodatasci-022020-021940.

Angermueller C, Lee HJ, Reik W, Stegle O. DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning. Genome Biol. 2017 Apr 11;18(1):67. https://doi.org/10.1186/s13059-017-1189-z.

Sun YV, Kardia SL. Imputing missing genotypic data of single-nucleotide polymorphisms using neural networks. Eur J Hum Genet. 2008; 16:487–95. https://doi.org/10.1038/sj.ejhg.5201988.

Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, et al. A universal SNP and small-indel variant caller using deep neural networks. Nat Biotechnol. 2018; 36:983-987. https://doi.org/10.1038/nbt.4235.

Agarwal R, Dhar V. Big data, data science, and analytics: the opportunity and challenge for IS research. Inf. Syst. Res. 2014; 25:443–448. https://doi.org/10.1287/isre.2014.0546.

Waddell M, Page D, Shaughnessy Jr J. Predicting cancer susceptibility from single-nucleotide polymorphism data: a case study in multiple myeloma. ACM. 2005; 21-28.

Xu X, Zhang Y, Zou L, Wang M, Li A. A gene signature for breast cancer prognosis using support vector machine. IEEE. 2012; 928-931. https://doi.org/10.1109/BMEI.2012.6513032.

Chen YC, Ke WC, Chiu HW. Risk classification of cancer survival using ANN with gene expression data from multiple laboratories. Comput Biol Med. 2014; 48:1-7. https://doi.org/10.1016/j.compbiomed.2014.02.006.

Gupta S, Tran T, Luo W, Phung D, Kennedy RL, Broad A, et al. Machine learning prediction of cancer survival: a retrospective study using electronic administrative records and a cancer registry. BMJ Open. 2014; 4:e004007. https://doi.org/10.1136/bmjopen-2013-004007.

Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, Sturm D, et al. DNA methylation-based classification of central nervous system tumours. Nature. 2018; 555:469-474. https://doi.org/10.1038/nature26000.

Li J, Zhou Z, Dong J, Fu Y, Li Y, Luan Z, Peng X. Predicting breast cancer 5-year survival using machine learning: A systematic review. PLoS ONE. 2021; 16:e0250370. https://doi.org/10.1371/journal.pone.0250370.

Yamada M, Saito Y, Imaoka H, Saiko M, Yamada S, Kondo H, et al. Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy. Sci Rep. 2019; 9:14465. https://doi.org/10.1038/s41598-019-50567-5.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Matek C, Schwarz S, Spiekermann K, Marr C. Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. Nat Mach Intell. 2019; 1:538–544. https://doi.org/10.1038/s42256-019-0101-9

Sharma A, Rani R. A systematic review of applications of machine learning in cancer prediction and diagnosis. Arch Computat Methods Eng. 2021. https://doi.org/10.1007/s11831-021-09556-z.

Goossens N, Nakagawa S, Sun X, Hoshida Y. Cancer biomarker discovery and validation. Transl Cancer Res. 2015; 4:256-269. https://doi.org/10.3978/j.issn.2218-676X.2015.06.04.

Parkinson DR, McCormack RT, Keating SM, Gutman SI, Hamilton SR, Mansfield EA, et al. Evidence of clinical utility: an unmet need in molecular diagnostics for patients with cancer. Clin Cancer Res. 2014; 20:1428-1444. https://doi.org/10.1158/1078-0432.CCR-13-2961.

Sawyers CL, van 't Veer LJ. Reliable and effective diagnostics are keys to accelerating personalized cancer medicine and transforming cancer care: a policy statement from the american association for cancer research. Clin Cancer Res. 2014; 20:4978-4981. https://doi.org/10.1158/1078-0432.CCR-14-2295.

Veytsman B, Baranova A. High-Throughput Approaches to Biomarker Discovery and the Challenges of Subsequent Validation. In: Preedy V, Patel V. (eds), General Methods in Biomarker Research and their Applications. Springer, Dordrecht. 2014, https://doi.org/10.1007/978-94-007-7740-8_20-1.

Bock C. Epigenetic biomarker development. Epigenomics. 2009; 1:99-110. https://doi.org/10.2217/epi.09.6.

Ballman KV. Biomarker: Predictive or Prognostic? J Clin Oncol. 2015; 33:3968-3971. https://doi.org/10.1200/JCO.2015.63.3651.

Bhattacharya S, Mariani TJ. Array of hope: expression profiling identifies disease biomarkers and mechanism. Biochem Soc Trans. 2009; 37:855-862. https://doi.org/10.1042/BST0370855.

Li Z, Sang M, Tian Z, Liu Z, Lv J, Zhang F, Shan B. Identification of key biomarkers and potential molecular mechanisms in lung cancer by bioinformatics analysis. Oncol Lett. 2019; 18:4429-4440. https://doi.org/10.3892/ol.2019.10796.

Sheng KL, Kang L, Pridham KJ, Dunkenberger LE, Sheng Z, Varghese RT. An integrated approach to biomarker discovery reveals gene signatures highly predictive of cancer progression. Sci Rep. 2020; 10:21246. https://doi.org/10.1038/s41598-020-78126-3.

Fatai AA, Gamieldien J. A 35-gene signature discriminates between rapidly- and slowly-progressing glioblastoma multiforme and predicts survival in known subtypes of the cancer. BMC Cancer. 2018; 18:377. https://doi.org/10.1186/s12885-018-4103-5.

Han J, Chen M, Wang Y, Gong B, Zhuang T, Liang L, Qiao H. Identification of biomarkers based on differentially expressed genes in papillary thyroid carcinoma. Sci Rep. 2018; 8:9912. https://doi.org/10.1038/s41598-018-28299-9.

Ye Z, Wang F, Yan F, Wang L, Li B, Liu T, et al. Bioinformatic identification of candidate biomarkers and related transcription factors in nasopharyngeal carcinoma. World J Surg Oncol. 2019; 17:60. https://doi.org/10.1186/s12957-019-1605-9.

http://etd.uwc.ac.za/

Asgharzadeh S, Pique-Regi R, Sposto R, Wang H, Yang Y, Shimada H, et al. Prognostic significance of gene expression profiles of metastatic neuroblastomas lacking MYCN gene amplification. J Natl Cancer Inst. 2006; 98:1193-1203. https://doi.org/10.1093/jnci/djj330.

He X, Qin C, Zhao Y, Zou L, Zhao H, Cheng C. Gene signatures associated with genomic aberrations predict prognosis in neuroblastoma. Cancer Commun (Lond). 2020; 40:105-118. https://doi.org/10.1002/cac2.12016.

Wang Y, Luo H, Cao J, Ma C. Bioinformatic identification of neuroblastoma microenvironment-associated biomarkers with prognostic value. J Oncol. 2020; 2020:5943014. https://doi.org/10.1155/2020/5943014.

Chen B, Ding P, Hua Z, Qin X, Li Z. Analysis and identification of novel biomarkers involved in neuroblastoma via integrated bioinformatics. Invest New Drugs. 2021; 39:52-65. https://doi.org/10.1007/s10637-020-00980-9.

Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. Mass Spectrom Rev. 2007; 26:51-78. https://doi.org/10.1002/mas.20108.

Lenz EM, Wilson ID. Analytical strategies in metabonomics. J Proteome Res. 2007; 6:443-458. https://doi.org/10.1021/pr0605217.

Sarker D, Workman P. Pharmacodynamic biomarkers for molecular cancer therapeutics. Adv Cancer Res. 2007; 96:213-268. https://doi.org/10.1016/S0065-230X(06)96008-4.

van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med. 2002; 347:1999-2009. https://doi.org/10.1056/NEJMoa021967.

Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med. 2004; 351:2817-2826. https://doi.org/10.1056/NEJMoa041588.

Hristova VA, Chan DW. Cancer biomarker discovery and translation: proteomics and beyond. Expert Rev Proteomics. 2019; 16:93–103. https://doi.org/10.1080/14789450.2019.1559062.

Imperiale TF, Ransohoff DF, Itzkowitz SH. Multitarget stool DNA testing for colorectal-cancer screening. N Engl J Med. 2014; 371:187-188. https://doi.org/10.1056/NEJMc1405215.

Bering J, Kahn, A, Rodriguez E, Ginos B, Ramirez F, Gurudu SR. Outcomes of Cologuard Screening at an Academic Medical Center: 1-Year Results, Am J Gastroenterol. 2017; 112:123-124.

Bell N, Connor Gorber S, Shane A, Joffres M, Singh H, Dickinson J, et al. Canadian Task Force on Preventive Health Care. Recommendations on screening for prostate cancer with the prostate-specific antigen test. CMAJ. 2014; 186:1225-1234. https://doi.org/10.1503/cmaj.140703.

Tikkinen KAO, Dahm P, Lytvyn L, Heen AF, Vernooij RWM, Siemieniuk RAC, et al. Prostate cancer screening with prostate-specific antigen (PSA) test: a clinical practice guideline. BMJ. 2018; 362:k3581. https://doi.org/10.1136/bmj.k3581.

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

Brodeur GM, Seeger RC, Schwab M, Varmus HE, Bishop JM. Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. Science. 1984; 224:1121-1124. https://doi.org/10.1126/science.6719137.

Seeger RC, Brodeur GM, Sather H, Dalton A, Siegel SE, Wong KY, Hammond D. Association of multiple copies of the N-myc oncogene with rapid progression of neuroblastomas. N Engl J Med. 1985; 313:1111-1116. https://doi.org/10.1056/NEJM198510313131802.

Kew M. Alpha-fetoprotein in primary liver cancer and other diseases. Gut. 1974; 15:814-821. https://doi.org/10.1136/gut.15.10.814.

Wang X, Wang Q. Alpha-fetoprotein and hepatocellular carcinoma immunity. Can J Gastroenterol Hepatol. 2018; 2018:9049252. https://doi.org/10.1155/2018/9049252.

Felder M, Kapur A, Gonzalez-Bosquet J, Horibata S, Heintz J, Albrecht R, et al. MUC16 (CA125): tumor biomarker to cancer therapy, a work in progress. Mol Cancer. 2014; 13:129. https://doi.org/10.1186/1476-4598-13-129.

Dochez V, Caillon H, Vaucel E, Dimet J, Winer N, Ducarme G. Biomarkers and algorithms for diagnosis of ovarian cancer: CA125, HE4, RMI and ROMA, a review. J Ovarian Res. 2019; 12:28. https://doi.org/10.1186/s13048-019-0503-7.

Bonifácio VDB. Ovarian cancer biomarkers: moving forward in early detection. Adv Exp Med Biol. 2020; 1219:355-363. https://doi.org/10.1007/978-3-030-34025-4_18.

Iafrate AJ, Louis DN. "MGMT for pt mgmt": is methylguanine-DNA methyltransferase testing ready for patient management? J Mol Diagn. 2008; 10:308-310. https://doi.org/10.2353/jmoldx.2008.080043.

Mikeska T, Bock C, El-Maarri O, Hübner A, Ehrentraut D, Schramm J, et al. Optimization of quantitative MGMT promoter methylation analysis using pyrosequencing and combined bisulfite restriction analysis. J Mol Diagn. 2007; 9:368-381. https://doi.org/10.2353/jmoldx.2007.060167.

Toland AE, Forman A, Couch FJ, Culver JO, Eccles DM, Foulkes WD, et al. Clinical testing of BRCA1 and BRCA2: a worldwide snapshot of technological practices. NPJ Genom Med. 2018; 3:7. https://doi.org/10.1038/s41525-018-0046-7.

De Preter K, Vandesompele J, Heimann P, Yigit N, Beckman S, Schramm A, et al. Human fetal neuroblast and neuroblastoma transcriptome analysis confirms neuroblast origin and highlights neuroblastoma candidate genes. Genome Biology. 2006; 7(9):R84. https://doi.org/10.1186/gb-2006-7-9-r84.

Delloye-Bourgeois C, Castellani V. Hijacking of Embryonic Programs by Neural Crest-Derived Neuroblastoma: From Physiological Migration to Metastatic Dissemination. Front. Mol. Neurosci. 2019; 12:52. https://doi.org/10.3389/fnmol.2019.00052

Brodeur GM. Neuroblastoma: biological insights into a clinical enigma. Nat Rev Cancer. 2003; 3(3):203–216. https://doi.org/10.1038/nrc1014.

Kerosuo L, Neppala P, Hsin J, Mohlin S, Vieceli FM, Török Z, Laine A, Westermarck J, Bronner ME. Enhanced expression of MycN/CIP2A drives neural crest toward a neural stem cell-like fate:

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Implications for priming of neuroblastoma. Proc Natl Acad Sci U S A. 2018; 115:E7351-7360. https://doi.org/10.1073/pnas.1800039115.

Theveneau E, Mayor R. Neural crest delamination and migration: from epithelium-to-mesenchyme transition to collective cell migration. Dev Biol. 2012; 366:34-54. https://doi.org/10.1016/j.ydbio.2011.12.041.

Gammill LS, Bronner-Fraser M. Neural crest specification: Migrating into genomics. Nat Rev Neurosci. 2003; 4:795–805. https://doi.org/10.1038/nrn1219.

Otte J, Dyberg C, Pepich A, Johnsen JI. MYCN function in neuroblastoma development. Front Oncol. 2021; 10:624079. https://doi.org/10.3389/fonc.2020.624079.

Huang M, Weiss WA. Neuroblastoma and MYCN. Cold Spring Harbor Perspectives in Medicine. 2013; 3:e1014415. https://doi.org/10.1101/cshperspect.a014415.

Marshall GM, Carter DR, Cheung BB, Liu T, Mateos MK, Meyerowitz JG, Weiss WA. The prenatal origins of cancer. Nat Rev Cancer. 2014; 14:277–289. https://doi.org/10.1038/nrc3679.

Olsen RR, Otero JH, García-López J, Wallace K, Finkelstein D, Rehg JE, Yin Z, Wang Y-D, Freeman KW. MYCN induces neuroblastoma in primary neural crest cells. Oncogene. 2017; 36:5075–82. https://doi.org/10.1038/onc.2017.128

Schulte JH, Lindner S, Bohrer A, Maurer J, De Preter K, Lefever S, Heukamp L, Schulte S, Molenaar J, Versteeg R, Thor T, Künkele A, Vandesompele J, Speleman F, Schorle H, Eggert A, Schramm A. MYCN and ALKF1174L are sufficient to drive neuroblastoma development from neural crest progenitor cells. Oncogene. 2013; 32:1059-1065. https://doi.org/10.1038/onc.2012.106.

Knudson AG Jr, Strong LC. Mutation and cancer: neuroblastoma and pheochromocytoma. Am J Hum Genet. 1972; 24:514-532.

Kushner BH, Gilbert F, Helson L. Familial neuroblastoma. Case reports, literature review, and etiologic considerations. Cancer. 1986; 57:1887-1893.

Dodge HJBM. Neuroblastoma of the adrenal medulla in siblings. Rocky Mt Med. 1945; 42:35-38.

Chompret A, de Vathaire F, Brugieres L, Abel A, Raquin MA, Hartmann O, Feunteun J, Bonaiti-Pellie C. Excess of cancers in relatives of patients with neuroblastoma. Med Pediatr Oncol. 1998; 31:211.

Trochet D, Bourdeaut F, Janoueix-Lerosey I, Deville A, de Pontual L, Schleiermacher G, Coze C, Philip N, Frébourg T, Munnich A, Lyonnet S, Delattre O, Amiel J. Germline mutations of the paired-like homeobox 2B (PHOX2B) gene in neuroblastoma. Am J Hum Genet. 2004; 74:761–764. https://doi.org/10.1086/383253.

Mosse YP, Laudenslager M, Longo L, Cole KA, Wood A, Attiyeh EF, Laquaglia MJ, Sennett R, Lynch JE, Perri P, Laureys G, Speleman F, Kim C, Hou C, Hakonarson H, Torkamani A, Schork NJ, Brodeur GM, Tonini GP, Rappaport E, Devoto M, Maris JM. Identification of ALK as a major familial neuroblastoma predisposition gene. Nature. 2008; 455:930-935. https://doi.org/10.1038/nature07261.

Yang L, Ke XX, Xuan F, Tan J, Hou J, Wang M, Cui H, Zhang Y. PHOX2B Is associated with neuroblastoma cell differentiation. Cancer Biother Radiopharm. 2016; 31(2):1-8. https://doi.org/10.1089/cbr.2015.1952.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Hurley SP, Clary DO, Copie V, Lefcort F. Anaplastic lymphoma kinase is dynamically expressed on subsets of motor neurons and in the peripheral nervous system. J Comp Neurol. 2006; 495:202–212. https://doi.org/10.1002/cne.20887.

Degoutin J, Brunet-de Carvalho N, Cifuentes-Diaz C, Vigny M. ALK (Anaplastic Lymphoma Kinase) expression in DRG neurons and its involvement in neuron-Schwann cells interaction. Eur J Neurosci. 2009; 29:275–286. https://doi.org/10.1111/j.1460-9568.2008.06593.x.

George RE, Sanda T, Hanna M, Frohling S, Luther W, Zhang J, Ahn Y, Zhou W, London WB, McGrady P, Xue L, Zozulya S, Gregor VE, Webb TR, Gray NS, Gilliland DG, Diller L, Greulich H, Morris SW, Meyerson M, Look AT. Activating mutations in ALK provide a therapeutic target in neuroblastoma. Nature. 2008; 455:975-978. https://doi.org/10.1038/nature07397.

Chen Y, Takita J, Choi YL, Kato M, Ohira M, Sanada M, Wang L, Soda M, Kikuchi A, Igarashi T, Nakagawara A, Hayashi Y, Mano H, Ogawa S. Oncogenic mutations of ALK kinase in neuroblastoma. Nature. 2008; 455:971-974. https://doi.org/10.1038/nature07399.

Janoueix-Lerosey I, Lequin D, Brugieres L, Ribeiro A, de Pontual L, Combaret V, Raynal V, Puisieux A, Schleiermacher G, Pierron G, Valteau-Couanet D, Frebourg T, Michon J, Lyonnet S, Amiel J, Delattre O. Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma. Nature. 2008; 455:967-970. https://doi.org/10.1038/nature07398.

Deyell RJ, Attiyeh EF. Advances in the understanding of constitutional and somatic genomic alterations in neuroblastoma. Cancer Genet. 2011; 204:113-121. https://doi.org/10.1016/j.cancergen.2011.03.001.

Bresler SC, Weiser DA, Huwe PJ, Park JH, Krytska K, Ryles H, Laudenslager M, Rappaport EF, Wood AC, McGrady PW, Hogarty MD, London WB, Radhakrishnan R, Lemmon MA, Mossé YP. ALK mutations confer differential oncogenic activation and sensitivity to ALK inhibition therapy in neuroblastoma. Cancer Cell. 2014; 26:682-694. https://doi.org/10.1016/j.ccell.2014.09.019.

Carén H, Abel F, Kogner P, Martinsson T. High incidence of DNA mutations and gene amplifications of the ALK gene in advanced sporadic neuroblastoma tumours. Biochem J. 2008; 416:153–159. https://doi.org/10.1042/bj20081834.

Valentijn LJ, Koster J, Zwijnenburg DA, Hasselt NE, van Sluis P, Volckmann R, van Noesel MM, George RE, Tytgat GAM, Molenaar JJ, Versteeg R. TERT rearrangements are frequent in neuroblastoma and identify aggressive tumors. Nat Genet. 2015; 47:1411-1414. https://doi.org/10.1038/ng.3438.

Peifer M, Hertwig F, Roels F, Dreidax D, Gartlgruber M, Menon R, et al. Telomerase activation by genomic rearrangements in high-risk neuroblastoma. Nature. 2015; 526:700-704. https://doi.org/10.1038/nature14980.

Cheung NK, Zhang J, Lu C, Parker M, Bahrami A, Tickoo SK, et al. Association of age at diagnosis and genetic mutations in patients with neuroblastoma. JAMA. 2012; 307:1062-1071. https://doi.org/10.1001/jama.2012.228.

Kurihara S, Hiyama E, Onitake Y, Yamaoka E, Hiyama K. Clinical features of ATRX or DAXX mutated neuroblastoma. Journal of Pediatric Surgery. 2014; 49:1835–1838

http://etd.uwc.ac.za/

Capasso M, Devoto M, Hou C, Asgharzadeh S, Glessner JT, Attiyeh EF, Mosse YP, et al. Common variations in BARD1 influence susceptibility to high-risk neuroblastoma. Nat Genet. 2009; 41:718–723. https://doi.org/10.1038/ng.374.

Wang K, Diskin SJ, Zhang H, Attiyeh EF, Winter C, Hou C, et al. Integrative genomics identifies LMO1 as a neuroblastoma oncogene. Nature. 2011; 469:216-220. https://doi.org/10.1038/nature09609.

Russell MR, Penikis A, Oldridge DA, Alvarez-Dominguez JR, McDaniel L, Diamond M, et al. CASC15-S is a tumor suppressor lncRNA at the 6p22 neuroblastoma susceptibility locus. Cancer Res. 2015; 75:3155-3166. https://doi.org/10.1158/0008-5472.CAN-14-3613.

Chang X, Zhao Y, Hou C, Glessner J, McDaniel L, Diamond MA,et al. Common variants in MMP20 at 11q22.2 predispose to 11q deletion and neuroblastoma risk. Nat Commun. 2017; 8(1):569. https://doi.org/10.1038/s41467-017-00408-8.

Jin Y, Shi J, Wang H, Lu J, Chen C, Yu Y, et al. MYC-associated protein X binding with the variant rs72780850 in RNA helicase DEAD box 1 for susceptibility to neuroblastoma. Sci China Life Sci. 2020; 63. https://doi.org/10.1007/s11427-020-1784-7.

Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nat Genet. 2013; 45:1127-1133. https://doi.org/10.1038/ng.2762.

Tonini GP. Growth, progression and chromosome instability of neuroblastoma: a new scenario of tumorigenesis? BMC Cancer. 2017; 17(1):20. https://doi.org/10.1186/s12885-016-2986-6.

Coco S, Theissen J, Scaruffi P, Stigliani S, Moretti S, Oberthuer A, Valdora F, Fischer M, Gallo F, Hero B, Bonassi S, Berthold F, Tonini GP. Age-dependent accumulation of genomic aberrations and deregulation of cell cycle and telomerase genes in metastatic neuroblastoma. Int J Cancer. 2012; 131:1591-1600. https://doi.org/10.1002/ijc.27432.

Fusco P, Esposito MR, Tonini GP. Chromosome instability in neuroblastoma. Oncol Lett. 2018; 16:6887-6894. https://doi.org/10.3892/ol.2018.9545.

Molenaar JJ, Koster J, Zwijnenburg DA, van Sluis P, Valentijn LJ, van der Ploeg I, et al. Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. Nature. 2012; 483:589-593. https://doi.org/10.1038/nature10910.

Tonini GP, Capasso M. Genetic predisposition and chromosome instability in neuroblastoma. Cancer and Metastasis Reviews. 2020; 39:275–285. https://doi.org/10.1007/s10555-020-09843-4.

Schleiermacher G, Janoueix-Lerosey I, Delattre O. Recent insights into the biology of neuroblastoma. Int J Cancer. 2014; 135:2249–2261. https://doi.org/10.1002/ijc.29077.

Stigliani S, Coco S, Moretti S, Oberthuer A, Fischer M, Theissen J, Gallo F, Garavent A, Berthold F, Bonassi S, Tonini GP, Scaruffi P. High genomic instability predicts survival in metastatic high-risk neuroblastoma. Neoplasia. 2012; 14:823-832. https://doi.org/10.1593/neo.121114.

Lastowska M, Cotterill S, Pearson AD, Roberts P, McGuckin A, Lewis I, Bown N. Gain of chromosome arm 17q predicts unfavourable outcome in neuroblastoma patients. U.K. Children's

http://etd.uwc.ac.za/

Cancer Study Group and the U.K. Cancer Cytogenetics Group. Eur J Cancer. 1997; 33:1627-1633. https://doi.org/10.1016/s0959-8049(97)00282-7.

Bown N, Cotterill S, Lastowska M, O'Neill S, Pearson AD, Plantaz D, et al. Gain of chromosome arm 17q and adverse outcome in patients with neuroblastoma. N Engl J Med. 1999; 340:1954-1961. https://doi.org/10.1056/NEJM199906243402504.

Ruiz-Pérez MV, Henley AB, Arsenian-Henriksson M. The MYCN protein in health and disease. Genes (Basel). 2017; 8:113. https://doi.org/10.3390/genes8040113.

Schwab M, Alitalo K, Klempnauer KH, Varmus HE, Bishop JM, Gilbert F, et al. Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour, Nature. 1983; 305:245–248. https://doi.org/10.1038/305245a0.

Galderisi U, Di Bernardo G, Cipollaro M, Peluso G, Cascino A, Cotrufo R, Melone MA. Differentiation and apoptosis of neuroblastoma cells: role of N-myc gene product, J Cell Biochem. 1999; 73:97–105. https://doi.org/10.1002/(sici)1097-4644(19990401)73:1<97::aid-jcb11>3.0.co;2-m.

Chayka O, D'Acunto CW, Middleton O, Arab M, Sala A. Identification and pharmacological inactivation of the MYCN gene network as a therapeutic strategy for neuroblastic tumor cells. J Biol Chem. 2015; 290:2198-2212. https://doi.org/10.1074/jbc.M114.624056.

Gherardi S, Valli E, Erriquez D, Perini G. MYCN-mediated transcriptional repression in neuroblastoma: the other side of the coin. Frontiers in Oncology. 2013; 3:42. https://doi.org/10.3389/fonc.2013.00042.

Weiss WA, Aldape K, Mohapatra G, Feuerstein BG, Bishop JM. Targeted expression of MYCN causes neuroblastoma in transgenic mice. EMBO J. 1997; 16:2985-2995. https://doi.org/10.1093/emboj/16.11.2985.

Burkhart CA, Cheng AJ, Madafiglio J, Kavallaris M, Mili M, Marshall GM, et al. Effects of MYCN antisense oligonucleotide administration on tumorigenesis in a murine model of neuroblastoma. J Natl Cancer Inst. 2003; 95:1394 –1403. https://doi.org/10.1093/jnci/djg045.

Kang JH, Rychahou PG, Ishola TA, Qiao J, Evers BM, Chung DH. MYCN silencing induces differentiation and apoptosis in human neuroblastoma cells. Biochem Biophys Res Commun. 2006; 351:192-197. https://doi.org/10.1016/j.bbrc.2006.10.020.

Tweddle DA, Malcolm AJ, Cole M, Pearson AD, Lunec J. p53 cellular localization and function in neuroblastoma: evidence for defective G(1) arrest despite WAF1 induction in MYCN-amplified cells. Am J Pathol. 2001; 158:2067–2077. https://doi.org/10.1016/S0002-9440(10)64678-0.

Muth D, Ghazaryan S, Eckerle I, Beckett E, Pöhler C, Batzler J, Beisel C, Gogolin S, Fischer M, Henrich KO, Ehemann V, Gillespie P, Schwab M, Westermann F. Transcriptional repression of SKP2 is impaired in MYCN-amplified neuroblastoma. Cancer Res. 2010; 70:3791-3802. https://doi.org/10.1158/0008-5472.CAN-09-1245.

Bell E, Lunec J, Tweddle DA. Cell cycle regulation targets of MYCN identified by gene expression microarrays. Cell Cycle. 2007; 6:1249–1256. https://doi.org/10.4161/cc.6.10.4222.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Yaari S, Jacob-Hirsch J, Amariglio N, Haklai R, Rechavi G, Kloog Y. Disruption of cooperation between Ras and MycN in human neuroblastoma cells promotes growth arrest. Clin Cancer Res. 2005; 11:4321-4330. https://doi.org/10.1158/1078-0432.CCR-04-2071.

Dang CV. MYC on the path to cancer. Cell. 2012; 149:22–35. https://doi.org/10.1016/j.cell.2012.03.003.

Adhikary S, Eilers M. Transcriptional regulation and transformation by Myc proteins. Nat Rev Mol Cell Biol. 2005; 6:635–645. https://doi.org/10.1038/nrm1703.

Cotterman R, Knoepfler PS. N-Myc regulates expression of pluripotency genes in neuroblastoma including lif, klf2, klf4, and lin28b. PLoS One. 2009; 4:e5799. https://doi.org/10.1371/journal.pone.0005799.

Dzieran J, Rodriguez Garcia A, Westermark UK, Henley AB, Sánchez EE, Träger C, Johansson HJ, Lehtiö J, Arsenian-Henriksson. MYCN-amplified neuroblastoma maintains an aggressive and undifferentiated phenotype by deregulation of estrogen and NGF signaling. PNAS. 2018; 115:E1229-E1238. https://doi.org/10.1073/pnas.1710901115.

Valli E, Trazzi S, Fuchs C, Erriquez D, Bartesaghi R, Perini G, Ciani E. CDKL5, a novel MYCN-repressed gene, blocks cell cycle and promotes differentiation of neuronal cells. Biochim Biophys Acta. 2012; 1819:1173-1185. https://doi.org/10.1016/j.bbagrm.2012.08.001.

Koppen A, Ait-Aissa R, Hopman S, Koster J, Haneveld F, Versteeg R, Valentijn LJ. Dickkopf-1 is down-regulated by MYCN and inhibits neuroblastoma cell proliferation. Cancer Lett. 2007; 256:218-228. https://doi.org/10.1016/j.canlet.2007.06.011.

Akter J, Takatori A, Hossain MS, Ozaki T, Nakazawa A, Ohira M, Suenaga Y, Nakagawara A. Expression of NLRR3 orphan receptor gene is negatively regulated by MYCN and Miz-1, and its downregulation is associated with unfavorable outcome in neuroblastoma. Clin Cancer Res. 2011; 17:6681-6692. https://doi.org/10.1158/1078-0432.CCR-11-0313.

Valentijn LJ, Koster J, Haneveld F, Aissa RA, van Sluis P, Broekmans MEC, Molenaar JJ, van Nes J, Versteeg R. Functional MYCN signature predicts outcome of neuroblastoma irrespective of MYCN amplification. PNAS. 2012; 109:19190–19195. https://doi.org/10.1073/pnas.1208215109.

Zaizen Y, Taniguchi S, Noguchi S, Suita S. The effect of N-myc amplification and expression on invasiveness of neuroblastoma cells. J Pediatr Surg. 1993; 28:766–769. https://doi.org/10.1016/0022-3468(93)90321-b.

Goodman LA, Liu BC, Thiele CJ, Schmidt ML, Cohn SL, Yamashiro JM, Pai DS, Ikegaki N, Wada RK. Modulation of N-myc expression alters the invasiveness of neuroblastoma. Clin Exp Metastasis. 1997; 15:130–139. https://doi.org/10.1023/a:1018448710006.

Cohn SL, Pearson AD, London WB, Monclair T, Ambros PF, Brodeur GM, et al. The International Neuroblastoma Risk Group (INRG) classification system: an INRG task force report. J Clin Oncol. 2009; 27(2):289–297. https://doi.org/10.1200/JCO.2008.16.6785.

Fotsis T, Breit S, Lutz W, Rössler J, Hatzi E, Schwab M, Schweigerer L. Down-regulation of endothelial cell growth inhibitors by enhanced MYCN oncogene expression in human neuroblastoma cells. Eur J Biochem. 1999; 263:757–764. https://doi.org/10.1046/j.1432-1327.1999.00575.x.

http://etd.uwc.ac.za/

Versteeg R, van der Minne C, Plomp A, Sijts A, van Leeuwen A, Schrier P. N-myc expression switched off and class I human leukocyte antigen expression switched on after somatic cell fusion of neuroblastoma cells. Mol Cell Biol. 1990; 10:5416-5423. https://doi.org/10.1128/mcb.10.10.5416-5423.1990.

Song L, Ara T, Wu Hong-wei, Woo CW, Reynolds CP, Seeger RC, DeClerck YA, Thiele CJ, Sposto R, Metelitsa LS. Oncogene MYCN regulates localization of NKT cells to the site of disease in neuroblastoma. J Clin Invest. 2007; 117:2702-2712. https://doi.org/10.1172/JCI30751.

Song L, Asgharzadeh S, Salo J, Engell K, Wu HW, Sposto R, Ara T, Silverman AM, DeClerck YA, Seeger RC, Metelitsa LS. Valpha24-invariant NKT cells mediate antitumor activity via killing of tumor-associated macrophages. J Clin Invest. 2009; 119:1524-1536. https://doi.org/10.1172/JCI37869.

Lin Y, Xu J, Lan H. Tumor-associated macrophages in tumor metastasis: biological roles and clinical therapeutic applications. J Hematol Oncol. 2019; 12:76. https://doi.org/10.1186/s13045-019-0760-3.

Yang Q, Guo N, Zhou Y, Chen J, Wei Q, Han M. The role of tumor-associated macrophages (TAMs) in tumor progression and relevant advance in targeted therapy. Acta Pharm Sin B. 2020; 10:2156-2170. https://doi.org/10.1016/j.apsb.2020.04.004.

Hogarty MD. The requirement for evasion of programmed cell death in neuroblastomas with MYCN amplification, Cancer Lett. 2003; 197:173–179. https://doi.org/10.1016/s0304-3835(03)00103-4.

Haber M, Bordow SB, Gilbert J, Madafiglio J, Kavallaris M, Marshall GM, Mechetner EB, Fruehauf JP, Tee L, Cohn SL, Salwen H, Schmidt ML, Norris MD. Altered expression of the MYCN oncogene modulates MRP gene expression and response to cytotoxic drugs in neuroblastoma cells. Oncogene. 1999; 18:2777-2782. https://doi.org/10.1038/sj.onc.1202859.

Haber M, Smith J, Bordow SB, Flemming C, Cohn SL, London WB, Marshall GM, Norris MD. Association of high-level MRP1 expression with poor clinical outcome in a large prospective study of primary neuroblastoma. J Clin Oncol. 2006; 24:1546-1553. https://doi.org/10.1200/JCO.2005.01.6196.

Miller MA, Ohashi K, Zhu X, McGrady P, London WB, Hogarty M, Sandler AD. Survivin mRNA levels are associated with biology of disease and patient survival in neuroblastoma: a report from the children's oncology group. J Pediatr Hematol Oncol. 2006; 28:412–417. https://doi.org/10.1097/01.mph.0000212937.00287.e5.

Bell E, Chen L, Liu T, Marshall GM, Lunec J, Tweddle DA. MYCN oncoprotein targets and their therapeutic potential. Cancer Lett. 2010; 293:144–157.

Xue C, Yu DM, Gherardi S, Koach J, Milazzo G, Gamble L, Liu B, Valli E, Russell AJ, London WB, Liu T, Cheung BB, Marshall GM, Perini G, Haber M, Norris MD. MYCN promotes neuroblastoma malignancy by establishing a regulatory circuit with transcription factor AP4." Oncotarget. 2016; 7:54937-54951. https://doi.org/10.18632/oncotarget.10709.

Schulte JH, Horn S, Otto T, Samans B, Heukamp LC, Eilers UC, Krause M, Astrahantseff K, Klein-Hitpass L, Buettner R, Schramm A, Christiansen H, Eilers M, Eggert A, Berwanger B. MYCN regulates oncogenic MicroRNAs in neuroblastoma. Int J Cancer. 2008; 122:699-704. https://doi.org/10.1002/ijc.23153.

http://etd.uwc.ac.za/

Swarbrick A, Woods SL, Shaw A, Balakrishnan A, Phua Y, Nguyen A, et al. miR-380-5p represses p53 to control cellular survival and is associated with poor outcome in MYCN-amplified neuroblastoma. Nat Med. 2010; 16:1134-4110. https://doi.org/10.1038/nm.2227.

Ma L, Young J, Prabhala H, Pan E, Mestdagh P, Muth D, et al. miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. Nat Cell Biol. 2010; 12:247-256. https://doi.org/10.1038/ncb2024.

Yeung KT, Yang J. Epithelial-mesenchymal transition in tumor metastasis. Mol Oncol. 2017; 11:28-39. https://doi.org/10.1002/1878-0261.12017.

Mittal V. Epithelial mesenchymal transition in tumor metastasis. Annu Rev Pathol. 2018; 13:395-412. https://doi.org/10.1146/annurev-pathol-020117-043854.

Gustafson WC, Weiss WA. Myc proteins as therapeutic targets. Oncogene. 2010; 29:1249–1259. https://doi.org/10.1038/onc.2009.512.

Frank SR, Parisi T, Taubert S, Fernandez P, Fuchs M, Chan HM, Livingston DM, Amati B. MYC recruits the TIP60 histone acetyltransferase complex to chromatin. EMBO Rep. 2003; 4:575-580. https://doi.org/10.1038/sj.embor.embor861.

Knoepfler PS, Zhang X-Y, Cheng PF, Gafken PR, McMahon SB, Eisenman RN. 2006. Myc influences global chromatin structure. EMBO J. 2006; 25:2723–2734. https://doi.org/10.1038/sj.emboj.7601152.

Beltran H. The N-myc oncogene: maximizing its targets, regulation, and therapeutic potential. Mol Cancer Res. 2014; 12:815–822. https://doi.org/10.1158/1541-7786.MCR-13-0536.

Bielinsky A. Penetrating enemy territory: Soluble PCNA-peptides stress out MYCN-overexpressing neuroblastomas. EbioMedicine. 2015; 2:1844–1845. https://doi.org/10.1016/j.ebiom.2015.12.012.

Futami H, Sakai R. All-trans retinoic acid downregulates ALK in neuroblastoma cell lines and induces apoptosis in neuroblastoma cell lines with activated ALK. Cancer Lett. 2010; 297:220–225. https://doi.org/10.1016/j.canlet.2010.05.014.

Moreno L, Barone G, DuBois SG, Molenaar J, Fischer M, Schulte J, et al. Accelerating drug development for neuroblastoma: Summary of the Second Neuroblastoma Drug Development Strategy forum from Innovative Therapies for Children with Cancer and International Society of Paediatric Oncology Europe Neuroblastoma. Eur J Cancer. 2020; 136:52-68. https://doi.org/10.1016/j.ejca.2020.05.010.

Zhang H, Liu T, Yi S, Gu L, Zhou M. Targeting MYCN IRES in MYCN-amplified neuroblastoma with miR-375 inhibits tumor growth and sensitizes tumor cells to radiation. Mol Oncol. 2015; 9:1301-1311. https://doi.org/10.1016/j.molonc.2015.03.005.

Schramm A, Lode H. MYCN-targeting vaccines and immunotherapeutics. Human Vaccines & Immunotherapeutics. 2016; 12:2257-2258. https://doi.org/10.1080/21645515.2016.1171430.

Tanimoto T, Tazawa H, Ieda T, Nouso H, Tani M, Oyama T, Urata Y, Kagawa S, Noda T, Fujiwara T. Elimination of MYCN-Amplified Neuroblastoma Cells by Telomerase-Targeted Oncolytic Virus via MYCN Suppression. Mol Ther Oncolytics. 2020; 18:14-23. https://doi.org/10.1016/j.omto.2020.05.015.

http://etd.uwc.ac.za/

Yoda H, Inoue T, Shinozaki Y, Lin J, Watanabe T, Koshikawa N, Takatori A, Nagase H. Direct Targeting of MYCN Gene Amplification by Site-Specific DNA Alkylation in Neuroblastoma. Cancer Res. 2019; 79:830-840. https://doi.org/10.1158/0008-5472.CAN-18-1198.

Macaluso M, Paggi M, Giordano A. Genetic and epigenetic alterations as hallmarks of the intricate road to cancer. Oncogene. 2003; 22:6472–6478. https://doi.org/10.1038/sj.onc.1206955.

Coyle KM, Boudreau JE, Marcato P. Genetic Mutations and Epigenetic Modifications: Driving Cancer and Informing Precision Medicine. Biomed Res Int. 2017; 2017:9620870. https://doi.org/10.1155/2017/9620870.

Baylin SB, Jones PA. A decade of exploring the cancer epigenome — biological and translational implications. Nat Rev Cancer. 2011; 11:726–734. https://doi.org/10.1038/nrc3130.

Shen H, Laird PW. Interplay between the cancer genome and epigenome. Cell. 2013; 153:38–55. https://doi.org/10.1016/j.cell.2013.03.008.

Garraway LA, Lander ES. Lessons from the cancer genome. Cell. 2013; 153:17–37. https://doi.org/10.1016/j.cell.2013.03.002.

You JS, Jones PA. Cancer genetics and epigenetics: two sides of the same coin? Cancer Cell. 2012; 22:9–20. https://doi.org/10.1016/j.ccr.2012.06.008.

Kanwal R, Gupta S. Epigenetic modifications in cancer. Clin Genet. 2012; 81:303–311. https://doi.org/10.1111/j.1399-0004.2011.01809.x

Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, et al. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nat Genet. 2005; 37:391–400. https://doi.org/10.1038/ng1531.

Hosseini A, Minucci S. Alterations of histone modifications in cancer. In: Epigenetics in Human Disease, 2018, 2nd ed; 6:pp141–217.

Liz J, Esteller M. lncRNAs and microRNAs with a role in cancer development. Biochim Biophys Acta. 2015; 1859:169–176. https://doi.org/10.1016/j.bbagrm.2015.06.015.

Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011; 12:861–874. https://doi.org/10.1038/nrg3074.

Baylin SB, Jones PA. Epigenetic determinants of cancer. Cold Spring Harb Perspect Biol. 2016; 8:a019505. https://doi.org/10.1101/cshperspect.a019505.

Fernandez AF, Assenov Y, Martin-Subero JI, Balint B, Siebert R, et al. A DNA methylation fingerprint of 1628 human samples. Genome Res. 2012; 22:407-419. https://doi.org/10.1101/gr.119867.110.

Asada K, Abe M, Ushijima T. Clinical application of the CpG island methylator phenotype to prognostic diagnosis in neuroblastomas. J Hum Genet. 2013; 58:428–433.

Abe M, Ohira M, Kaneda A, Yagi Y, Yamamoto S, Kitano Y, et al. CpG island methylator phenotype is a strong determinant of poor prognosis in neuroblastomas. Cancer Res. 2005; 65:828–834

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Abe M, Westermann F, Nakagawara A, Takato T, Schwab M, Ushijima T. Marked and independent prognostic significance of the CpG island methylator phenotype in neuroblastomas. Cancer Lett. 2007; 247:253–258. https://doi.org/10.1016/j.canlet.2006.05.001.

Abe M, Watanabe N, McDonell N, Takato T, Ohira M, Nakagawara A, Ushijima T. Identification of genes targeted by CpG island methylator phenotype in neuroblastomas, and their possible integrative involvement in poor prognosis. Oncology. 2008; 74:50-60. https://doi.org/10.1159/000139124.

Banelli B, Brigati C, Di Vinci A, Casciano I, Forlani A, Borzi L, Allemanni G, Romani M. A pyrosequencing assay for the quantitative methylation analysis of the PCDHB gene cluster, the major factor in neuroblastoma methylator phenotype. Lab Invest. 2012; 92:458–465. https://doi.org/10.1038/labinvest.2011.169.

Banelli B, Merlo DF, Allemanni G, Forlani A, Romani M. Clinical potentials of methylator phenotype in stage 4 high-risk neuroblastoma: an open challenge. PLoS One. 2013; 8(5):e63253. https://doi.org/10.1371/journal.pone.0063253.

Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature. 1983; 301:89-92. https://doi.org/10.1038/301089a0.

Riggs AD, Jones PA. 5-methylcytosine, gene regulation, and cancer. Adv Cancer Res. 1983; 40:1-30. https://doi.org/10.1016/s0065-230x(08)60678-8.

Smith ZD, Meissner A. DNA methylation: Roles in mammalian development. Nat Rev Genet. 2013; 14:204–220. https://doi.org/10.1038/nrg3354.

Cheng Y, He C, Wang M, Ma X, Mo F, Yang S, Han J, Wei X. Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. Signal Transduct Target Ther. 2019; 4:62. https://doi.org/10.1038/s41392-019-0095-0.

Rauluseviciute I, Drabløs F, Rye MB. DNA methylation data by sequencing: experimental approaches and recommendations for tools and pipelines for data analysis. Clinical Epigenetics. 2019; 11:193. https://doi.org/10.1186/s13148-019-0795-x.

Jubierre L, Jiménez C, Rovira E, Soriano A, Sábado C, Gros L, Llort A, Hladun R, Roma J, de Toledo JS, Gallego S, Segura MF. Targeting of epigenetic regulators in neuroblastoma. Experimental & Molecular Medicine. 2018; 50:51. https://doi.org/10.1038/s12276-018-0077-2.

Bock C, Halbritter F, Carmona FJ, Tierling S, Datlinger P, Assenov Y, et al. Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. Nat Biotechnol. 2016; 34:726-737. https://doi.org/10.1038/nbt.3605.

Oehme I, Deubzer HE, Wegener D, Pickert D, Linke JP, Hero B, Kopp-Schneider A, Westermann F, Ulrich SM, von Deimling A, Fischer M, Witt O. Histone deacetylase 8 in neuroblastoma tumorigenesis. Clin Cancer Res. 2009; 15:91–99. https://doi.org/10.1158/1078-0432.CCR-08-0684.

Fabian J, Opitz D, Althoff K, Lodrini M, Hero B, Volland R, et al. MYCN and HDAC5 transcriptionally repress CD9 to trigger invasion and metastasis in neuroblastoma. Oncotarget. 2016; 7:66344–66359. https://doi.org/10.18632/oncotarget.11662.

Lodrini M, Oehme I, Schroeder C, Milde T, Schier MC, Kopp-Schneider A, et al. MYCN and HDAC2 cooperate to repress miR-183 signaling in neuroblastoma. Nucleic Acids Res. 2013; 41:6018–6033

http://etd.uwc.ac.za/

Zhang B, Pan X, Cobb GP, Anderson TA. microRNAs as oncogenes and tumor suppressors. Dev Biol. 2007; 302:1-12. https://doi.org/10.1016/j.ydbio.2006.08.028.

Ruan K, Fang X, Ouyang G. MicroRNAs: novel regulators in the hallmarks of human cancer, Cancer Lett. 2009; 285:116–126.

Su Z, Yang Z, Xu Y, Chen Y, Yu Q. MicroRNAs in apoptosis, autophagy and necroptosis. Oncotarget. 2015; 6:8474-8490. https://doi.org/10.18632/oncotarget.3523.

Stallings RL. MicroRNA involvement in the pathogenesis of neuroblastoma: potential for microRNA mediated therapeutics. Curr Pharm Des. 2009; 15:456–462. https://doi.org/10.2174/138161209787315837.

Romania P, Bertaina A, Bracaglia G, Locatelli F, Fruci D, Rota R. Epigenetic deregulation of microRNAs in rhabdomyosarcoma and neuroblastoma and translational perspectives. Int J Mol Sci. 2012; 13:16554-16579.

Das S, Bryan K, Buckley PG, Piskareva O, Bray IM, Foley N, et al. Modulation of neuroblastoma disease pathogenesis by an extensive network of epigenetically regulated microRNAs. Oncogene. 2013; 32:2927-2936. https://doi.org/10.1038/onc.2012.311.

Parodi F, Carosio R, Ragusa M, Di Pietro C, Maugeri M, Barbagallo D, et al. Epigenetic dysregulation in neuroblastoma: A tale of miRNAs and DNA methylation. Biochim Biophys Acta. 2016; 1859:1502-1514. https://doi.org/10.1016/j.bbagrm.2016.10.006.

Ramraj SK, Aravindan S, Somasundaram DB, Herman TS, Natarajan M, Aravindan N. Serum-circulating miRNAs predict neuroblastoma progression in mouse model of high-risk metastatic disease. Oncotarget. 2016; 7:18605-18619. https://doi.org/10.18632/oncotarget.7615.

Zhao Z, Partridge V, Sousares M, Shelton SD, Holland CL, Pertsemlidis A, Du L. microRNA-2110 functions as an onco-suppressor in neuroblastoma by directly targeting Tsukushi. PLoS One. 2018; 13(12): e0208777. https://doi.org/10.1371/journal.pone.0208777.

Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. Carcinogenesis. 2010; 31:27–36. https://doi.org/10.1093/carcin/bgp220.

Strauss J, Figg WD. Using epigenetic therapy to overcome chemotherapy resistance. Anticancer Res. 2016; 36:1-4.

Lu Y, Chan YT, Tan HY, Li S, Wang N, Feng Y. Epigenetic regulation in human cancer: the potential role of epi-drug in cancer therapy. Mol Cancer. 2020; 19:79. https://doi.org/10.1186/s12943-020-01197-3.

Quagliano A, Gopalakrishnapillai A, Barwe SP. Understanding the mechanisms by which epigenetic modifiers avert therapy resistance in cancer. Front Oncol. 2020; 10:992. https://doi.org/10.3389/fonc.2020.00992.

Keyvani-Ghamsari S, Khorsandi K, Rasul A, Zaman MK. Current understanding of epigenetics mechanism as a novel target in reducing cancer stem cells resistance. Clin Epigenetics. 2021; 13:120. https://doi.org/10.1186/s13148-021-01107-4.

http://etd.uwc.ac.za/

Park JR, Eggert A, Caron H. Neuroblastoma: biology, prognosis, and treatment. Pediatr Clin North Am. 2008; 55:97–120. https://doi.org/10.1016/j.pcl.2007.10.014.

Matthay KK, Perez C, Seeger RC, Brodeur GM, Shimada H, Atkinson JB, Black CT, Gerbing R, Haase GM, Stram DO, Swift P, Lukens JN. Successful treatment of stage III neuroblastoma based on prospective biologic staging: a Children's Cancer Group study. J Clin Oncol. 1998; 16:1256-1264. https://doi.org/10.1200/JCO.1998.16.4.1256.

Schmidt ML, Lukens JN, Seeger RC, Brodeur GM, Shimada H, Gerbing RB, Stram DO, Perez C, Haase GM, Matthay KK. Biologic factors determine prognosis in infants with stage IV neuroblastoma: A prospective Children's Cancer Group study. J Clin Oncol. 2000; 18:1260-1268. https://doi.org/10.1200/JCO.2000.18.6.1260.

Matthay KK, Villablanca JG, Seeger RC, Stram DO, Harris RE, Ramsay NK, Swift P, Shimada H, Black CT, Brodeur GM, Gerbing RB, Reynolds CP. Treatment of high-risk neuroblastoma with intensive chemotherapy, radiotherapy, autologous bone marrow transplantation, and 13-cis-retinoic acid. Children's Cancer Group. N Engl J Med. 1999; 341:1165-1173. https://doi.org/10.1056/NEJM199910143411601.

Kholodenko IV, Kalinovsky DV, Doronin II, Deyev SM, Kholodenko RV. Neuroblastoma origin and therapeutic targets for immunotherapy. J Immunol Res. 2018; 2018:7394268. https://doi.org/10.1155/2018/7394268.

Armideo E, Callahan C, Madonia L. Immunotherapy for high-risk neuroblastoma: management of side effects and complications. J Adv Pract Oncol. 2017; 8:44–55. https://doi.org/10.6004/jadpro.2017.8.1.4.

Greengard E, Hill-Kayser C, Bagatell R. Treatment of high-risk neuroblastoma in children: recent clinic trial results. Clin Invest. 2013; 3:1071–1081.

Berthold F, Spix C, Kaatsch P, Lampert F. Incidence, survival, and treatment of localized and metastatic neuroblastoma in Germany 1979-2015. Paediatr Drugs. 2017; 19:577-593. https://doi.org/10.1007/s40272-017-0251-3.

Berlanga P, Canete A, Castel V. Advances in emerging drugs for the treatment of neuroblastoma. Expert Opin Emerg Drugs. 2017; 22:63-75. https://doi.org/10.1080/14728214.2017.1294159.

Liu Z, Grant CN, Sun L, Miller BA, Spiegelman VS, Wang HG. Expression patterns of immune genes reveal heterogeneous subtypes of high-risk neuroblastoma. Cancers (Basel). 2020; 12:1739. https://doi.org/10.3390/cancers12071739.

Russo R, Cimmino F, Pezone L, Manna F, Avitabile M, Langella C, Koster J, Casale F, Raia M, Viola G, Fischer M, Iolascon A, Capasso M. Kinome expression profiling of human neuroblastoma tumors identifies potential drug targets for ultra high-risk patients. Carcinogenesis. 2017; 38:1011–1020. https://doi.org/10.1093/carcin/bgx077.

Formicola D, Petrosino G, Lasorsa VA, Pignataro P, Cimmino F, Vetrella S, Longo L, Tonini GP, Oberthuer A, Iolascon A, Fischer M, Capasso M. An 18 gene expression-based score classifier predicts the clinical outcome in stage 4 neuroblastoma. J Transl Med. 2016; 14:142. https://doi.org/10.1186/s12967-016-0896-7.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Nwose EU. Comparison of immunohistochemistry within situ hybridization methods: discrepancy resolution and implications for low-mid incomecountries, J Histotechnol. 2016; 39:21-25, https://doi.org/10.1080/01478885.2015.1115587.

Johnsen JI, Dyberg C, Wickström M. Neuroblastoma - A Neural Crest Derived Embryonal Malignancy. Front Mol Neurosci. 2019; 12:9. https://doi.org/10.3389/fnmol.2019.00009.

Maris JM, Mosse YP, Bradfield JP, Hou C, Monni S, Scott RH, et al. Chromosome 6p22 locus associated with clinically aggressive neuroblastoma. N Engl J Med. 2008; 358:2585-2593. https://doi.org/10.1056/NEJMoa0708698.

Capasso M, Diskin S, Cimmino F, Acierno G, Totaro F, Petrosino G, Pezone L, Diamond M, McDaniel L, Hakonarson H, Iolascon A, Devoto M, Maris JM. Common genetic variants in NEFL influence gene expression and neuroblastoma risk. Cancer Res. 2014. 74:6913–6924. https://doi.org/10.1158/0008-5472.CAN-14-0431.

Diskin SJ, Hou C, Glessner JT, Attiyeh EF, Laudenslager M, Bosse K, et al. Copy number variation at 1q21.1 associated with neuroblastoma. Nature. 2009; 459:987-991. https://doi.org/10.1038/nature08035.

Bosse KR, Diskin SJ, Cole KA, Wood AC, Schnepp RW, Norris G, et al. Common variation at BARD1 results in the expression of an oncogenic isoform that influences neuroblastoma susceptibility and oncogenicity. Cancer Res. 2012; 72:2068-78. https://doi.org/10.1158/0008-5472.CAN-11-3703.

Vivian J, Rao AA, Nothaft FA, Ketchum C, Armstrong J, Novak A, et al. Toil enables reproducible, open source, big biomedical data analyses. Nat Biotechnol. 2017; 35:314-316. https://doi.org/10.1038/nbt.3772.

Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS. 2012; 16:284–287. https://doi.org/10.1089/omi.2011.0118.

Yu G, Wang LG, Yan G-R, He Q-Y. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. Bioinformatics. 2015; 31:608–609. https://doi.org/10.1093/bioinformatics/btu684.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011; 12:2825-2830

Chang CC, Lin CJ. LIBSVM: A library for support vector machines. ACM Trans Intell Syst Technol. 2011; 2(3):1–27. https://doi.org/10.1145/1961189.1961199.

Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. PLoS One. 2010; 5(9):e12776. https://doi.org/10.1371/journal.pone.0012776.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13:2498–2504. https://doi.org/10.1101/gr.1239303.

http://etd.uwc.ac.za/

Aoyama M, Ozaki T, Inuzuka H, Tomotsune D, Hirato J, Okamoto Y, Tokita H, Ohira M, Nakagawara A. LMO3 interacts with neuronal transcription factor, HEN2, and acts as an oncogene in neuroblastoma. Cancer Res. 2005; 65:4587–4597. https://doi.org/10.1158/0008-5472.CAN-04-4630.

Good DJ, Braun T. NHLH2: At the intersection of obesity and fertility. Trends Endocrinol Metab. 2013; 24:385–390. https://doi.org/10.1016/j.tem.2013.04.003.

Vella KR, Burnside AS, Brennan KM, Good DJ. Expression of the Hypothalamic Transcription Factor Nhlh2 is Dependent on Energy Availability. J Neuroendocrinol. 2007; 19:499–510. https://doi.org/10.1111/j.1365-2826.2007.01556.x.

Hiyama E, Hiyama K, Nishiyama M, Reynolds CP, Shay JW, Yokoyama T. Differential gene expression profiles between neuroblastomas with high telomerase activity and low telomerase activity. J Pediatr Surg. 2003; 38:1730–1734. https://doi.org/10.1016/j.jpedsurg.2003.08.042.

Hakami F, Darda L, Stafford P, Woll P, Lambert DW, Hunter KD. The roles of HOXD10 in the development and progression of head and neck squamous cell carcinoma (HNSCC). Br J Cancer. 2014; 111:807–816. https://doi.org/10.1038/bjc.2014.372.

Kocak H, Ackermann S, Hero B, Kahlert Y, Oberthuer A, Juraeva D, et al. Hox-C9 activates the intrinsic pathway of apoptosis and is associated with spontaneous regression in neuroblastoma. Cell Death Dis. 2013; 4:e586. https://doi.org/10.1038/cddis.2013.84.

Vieira GC, Chockalingam S, Melegh Z, Greenhough A, Malik S, Szemes M, et al. LGR5 regulates pro-survival MEK/ERK and proliferative Wnt/β-catenin signalling in neuroblastoma. Oncotarget. 2015; 6:40053–40067. https://doi.org/10.18632/oncotarget.5548.

Forgham H, Johnson D, Carter N, Veuger S, Carr-Wilkinson J. Stem Cell Markers in Neuroblastoma-An Emerging Role for LGR5. Front Cell Dev Biol. 2015; 3:77. https://doi.org/10.3389/fcell.2015.00077.

Kuraishi Y, Uehara T, Kobayashi Y, Nakajima T, Watanabe T, Shimizu A, Ota H, Tanaka E. Correlation of clinicopathological features and leucine-rich repeat-containing G-protein-coupled receptor 5 expression in pancreatic ductal adenocarcinoma. Pathol Res Pract. 2019; 215:152623. https://doi.org/10.1016/j.prp.2019.152623.

Morgan R, Mortensson E, Williams A. Targeting LGR5 in Colorectal Cancer: therapeutic gold or too plastic? Br J Cancer. 2018; 118:1410–1418. https://doi.org/10.1038/s41416-018-0118-6.

Jiang T, Guo J, Hu Z, Zhao M, Gu Z, Miao S. Identification of Potential Prostate Cancer-Related Pseudogenes Based on Competitive Endogenous RNA Network Hypothesis. Med Sci Monit. 2018; 24:4213–4239. https://doi.org/10.12659/MSM.910886.

Juárez-Morales JL, Schulte CJ, Pezoa SA, Vallejo GK, Hilinski WC, England SJ, de Jager S, Lewis KE. Evx1 and Evx2 specify excitatory neurotransmitter fates and suppress inhibitory fates through a Pax2-independent mechanism. Neural Dev. 2016; 11:5. https://doi.org/10.1186/s13064-016-0059-9.

Perotti D, De Vecchi G, Testi MA, Lualdi E, Modena P, Mondini P, Ravagnani F, Collini P, Di Renzo F, Spreafico F, Terenziani M, Sozzi G, Fossati-Bellani F, Radice P. Germline mutations of the POU6F2 gene in Wilms tumors with loss of heterozygosity on chromosome 7p14. Hum Mutat. 2004; 24(5):400-407. https://doi.org/10.1002/humu.20096.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Rossi M, Colecchia D, Ilardi G, Acunzo M, Nigita G, Sasdelli F, Celetti A, Strambi A, Staibano S, Croce CM, Chiariello M. MAPK15 upregulation promotes cell proliferation and prevents DNA damage in male germ cell tumors. Oncotarget. 2016; 7:20981–20998. https://doi.org/10.18632/oncotarget.8044.

Riordan JD, Keng VW, Tschida BR, Scheetz TE, Bell JB, Podetz-Pedersen KM, Moser CD, Copeland NG, Jenkins NA, Roberts LR, Largaespada DA, Dupuy AJ. Identification of rtl1, a retrotransposon-derived imprinted gene, as a novel driver of hepatocarcinogenesis. PLoS Genet. 2013; 9(4):e1003441. https://doi.org/10.1371/journal.pgen.1003441.

Berry DC, Levi L, Noy N. Holo-retinol-binding protein and its receptor STRA6 drive oncogenic transformation. Cancer Res. 2014; 74:6341–6351. https://doi.org/10.1158/0008-5472.CAN-14-1052.

Karunanithi S, Levi L, DeVecchio J, Karagkounis G, Reizes O, Lathia JD, Kalady MF, Noy N. RBP4-STRA6 Pathway Drives Cancer Stem Cell Maintenance and Mediates High-Fat Diet-Induced Colon Carcinogenesis. Stem Cell Reports. 2017; 9:438–450. https://doi.org/10.1016/j.stemcr.2017.06.002.

Yang X, Hu Y, Liu Y, Liu W, Zhao X, Liu M, Tang H. C14orf28 downregulated by miR-519d contributes to oncogenicity and regulates apoptosis and EMT in colorectal cancer. Mol Cell Biochem. 2017; 434:197–208. https://doi.org/10.1007/s11010-017-3049-2.

Chatrath A, Kiran M, Kumar P, Ratan A, Dutta A. The Germline Variants rs61757955 and rs34988193 are Predictive of Survival in Lower Grade Glioma Patients. Mol Cancer Res. 2019; 17: 1075–1086. https://doi.org/10.1158/1541-7786.MCR-18-0996.

Feng J, Zhang Y, She X, Sun Y, Fan L, Ren X, Fu H, Liu C, Li P, Zhao C, Liu Q, Liu Q, Li G, Wu M. Hypermethylated gene ANKDD1A is a candidate tumor suppressor that interacts with FIH1 and decreases HIF1α stability to inhibit cell autophagy in the glioblastoma multiforme hypoxia microenvironment. Oncogene. 2019; 38:103-119. https://doi.org/10.1038/s41388-018-0423-9.

Cheng S, Xie W, Miao Y, Guo J, Wang J, Li C, Zhang Y. Identification of key genes in invasive clinically non-functioning pituitary adenoma by integrating analysis of DNA methylation and mRNA expression profiles. J Transl Med. 2019; 17:407. https://doi.org/10.1186/s12967-019-02148-3.

Vuori N, Sandholm N, Kumar A, Hietala K, Syreeni A, Forsblom C, Juuti-Uusitalo K, Skottman H, Imamura M, Maeda S, Summanen PA, Lehto M, Groop PH; FinnDiane Study. *CACNB2* Is a Novel Susceptibility Gene for Diabetic Retinopathy in Type 1 Diabetes. Diabetes. 2019; 68:2165-2174. https://doi.org/10.2337/db19-0130.

Liu F, Gong X, Yao X, Cui L, Yin Z, Li C, Tang Y, Wang F. Variation in the CACNB2 gene is associated with functional connectivity of the Hippocampus in bipolar disorder. BMC Psychiatry. 2019; 19:62. https://doi.org/10.1186/s12888-019-2040-8.

Niu Y, Gong Y, Langaee TY, Davis HM, Elewa H, Beitelshees AL, Moss JI, Cooper-DeHoff RM, Pepine CJ, Johnson JA. Genetic variation in the beta 2 subunit of the voltage-gated calcium channel (CACNB2) and pharmacogenetic association with adverse cardiovascular outcomes in the International Verapamil SR-Trandolapril Study-Genetic Substudy (INVEST-GENES). Circ Cardiovasc Genet. 2010; 3:548–555. https://doi.org/10.1161/CIRCGENETICS.110.957654.

Breitenkamp AFS, Matthes J, Nass RD, Sinzig J, Lehmkuhl G, Nürnberg P, Herzig S. Rare Mutations of CACNB2 Found in Autism Spectrum Disease-Affected Families Alter Calcium Channel Function. PLoS One, 2014, 9(4):e95579. https://doi.org/10.1371/journal.pone.0095579.

http://etd.uwc.ac.za/

Yang H, Zhang Q, He J, Lu W. Regulation of calcium signaling in lung cancer. J Thorac Dis. 2010; 2:52–56.

Stewart TA, Yapa KTDS, Monteith GR. Altered calcium signaling in cancer cells. Biochim Biophys Acta. 2015; 1848:2502–2511. https://doi.org/10.1016/j.bbamem.2014.08.016.

Cui C, Merritt R, Fu L, Pan Z. Targeting calcium signaling in cancer therapy. Acta Pharm Sin B. 2017; 7:3–17. https://doi.org/10.1016/j.apsb.2016.11.001.

Hagan JP, O'Neill BL, Stewart CL, Kozlov SV, Croce CM. At least ten genes define the imprinted Dlk1-Dio3 cluster on Mouse chromosome 12qF1. PLoS One. 2009; 4(2):e4352 https://doi.org/10.1371/journal.pone.0004352.

Benetatos L, Hatzimichael E, Londin E, Vartholomatos G, Loher P, Rigoutsos I, Briasoulis E. The microRNAs within the DLK1-DIO3 genomic region: involvement in disease pathogenesis. Cell Mol Life Sci. 2013; 70:795–814. https://doi.org/10.1007/s00018-012-1080-8.

Cai M, Xu L, Shen L, Zhang J. The expression of long non-coding RNA-LINC01410 in pancreatic cancer and its effect on proliferation and migration of pancreatic cancer cells. Zhonghua Yi Xue Za Zhi. 2019; 99:1406–1411. https://doi.org/10.3760/cma.j.issn.0376-2491.2019.18.010.

Zhang JX, Chen ZH, Chen DL, Tian XP, Wang CY, Zhou ZW, Gao Y, Xu Y, Chen C, Zheng ZS, Weng HW, Ye S, Kuang M, Xie D, Peng S. LINC01410-miR-532-NCF2-NF-kB feedback loop promotes gastric cancer angiogenesis and metastasis. Oncogene. 2018; 37:2660-2675. https://doi.org/10.1038/s41388-018-0162-y.

Jiang T, Wang C, Zhu Y, Han H. LINC01410 promotes cell proliferation and migration of cholangiocarcinoma through modulating miR-124-3p/SMAD5 axis. J Gene Med. 2020; 22:e3162. https://doi.org/10.1002/jgm.3162.

Kuja-Panula J, Kiiltomäki M, Yamashiro T, Rouhiainen A, Rauvala H. AMIGO, a transmembrane protein implicated in axon tract development, defines a novel protein family with leucine-rich repeats. J Cell Biol. 2003; 160:963–973. https://doi.org/10.1083/jcb.200209074.

Xia H, Chen J, Shi M, Gao H, Sekar K, Seshachalam VP, Ooi LLPJ, Hui KM. EDIL3 is a novel regulator of epithelial-mesenchymal transition controlling early recurrence of hepatocellular carcinoma. J Hepatol. 2015; 63:863–873. https://doi.org/10.1016/j.jhep.2015.05.005.

Wang JP, Hielscher A. Fibronectin: How its aberrant expression in tumors may improve therapeutic targeting. J Cancer, 2017, 8:674–82. https://doi.org/10.7150/jca.16901.

Ouderkirk JL, Krendel M. Non-muscle myosins in tumor progression, cancer cell invasion and metastasis. Cytoskeleton (Hoboken). 2014; 71:447–463. https://doi.org/10.1002/cm.21187.

Weber L, Maßberg D, Becker C, Altmüller J, Ubrig B, Bonatz G, Wölk G, Philippou S, Tannapfel A, Hatt H, Gisselmann G. Olfactory receptors as biomarkers in human breast carcinoma tissues. Front Oncol. 2018; 8:33. https://doi.org/10.3389/fonc.2018.00033.

Ruan H, Li Y, Wang X, Sun B, Fang W, Jiang S, Liang C. CRYAB inhibits migration and invasion of bladder cancer cells through the PI3K/AKT and ERK pathways. Jpn J Clin Oncol. 2020; 50:254–260. https://doi.org/10.1093/jjco/hyz172.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Reitz C, Conrad C, Roszkowski K, Rogers RS, Mayeux R. Effect of genetic variation in LRRTM3 on risk of Alzheimer disease. Arch Neurol. 2012; 69:894–900. https://doi.org/10.1001/archneurol.2011.2463.

Wang J, Yu JT, Jiang T, Tan MS, Wang HF, Tan L, Hu N, Sun L, Zhang W, Tan L. Association of LRRTM3 polymorphisms with late-onset Alzheimer's disease in Han Chinese. Exp Gerontol. 2014; 52:18–22. https://doi.org/10.1016/j.exger.2014.01.013.

Wimmer K, Zhu XX, Lamb BJ, Kuick R, Ambros PF, Kovar H, Thoraval D, Motyka S, Alberts JR, Hanash SM. Co-amplification of a novel gene, NAG, with the N-myc gene in neuroblastoma. Oncogene. 1999; 18:233–238. https://doi.org/10.1038/sj.onc.1202287.

Shi X, Lu L, Jin X, Liu B, Sun X, Lu L, Jiang Y. GRIN3A and MAPT stimulate nerve overgrowth in macrodactyly. Mol Med Rep. 2016; 14:5637–5643. https://doi.org/10.3892/mmr.2016.5923.

Dhillon AS, Hagan S, Rath O, Kolch W. MAP kinase signalling pathways in cancer. Oncogene. 2007; 26:3279–3290. https://doi.org/10.1038/sj.onc.1210421.

Jin DH, Lee J, Kim KM, Kim S, Kim DH, Park J. Overexpression of MAPK15 in gastric cancer is associated with copy number gain and contributes to the stability of c-Jun. Oncotarget. 2015; 6: 20190–20203.

Yokoyama A. RNA polymerase II-dependent transcription initiated by selectivity factor 1: a central mechanism used by MLL fusion proteins in leukemic transformation. Front Genet. 2019; 9:722. https://doi.org/10.3389/fgene.2018.00722.

Best KE, Rankin J. Long-term survival of individuals born with congenital heart disease: a systematic review and meta-analysis. J Am Heart Assoc. 2016; 5(6):e002846. https://doi.org/10.1161/JAHA.115.002846.

Dubé BP, Dres M. Diaphragm dysfunction: diagnostic approaches and management strategies. J Clin Med. 2016; 5(12):113. https://doi.org/10.3390/jcm5120113.

Friedman DN, Henderson TO. Late effects and survivorship issues in patients with neuroblastoma. Children (Basel). 2018; 5(8):107. https://doi.org/10.3390/children5080107.

Mueller S, Matthay KK. Neuroblastoma: biology and staging. Curr Oncol Rep. 2009; 11:431-438. https://doi.org/10.1007/s11912-009-0059-6.

Tonini G, Nakagawara A, Berthold F. Towards a turning point of neuroblastoma therapy. Cancer Lett. 2012; 326:128-134. https://doi.org/10.1016/j.canlet.2012.08.017.

Ratner N, Brodeur GM, Dale RC, Schor NF. The "neuro" of neuroblastoma: neuroblastoma as a neurodevelopmental disorder. Ann Neurol. 2016; 80:13-23. https://doi.org/10.1002/ana.24659.

Decock A, Ongenaert M, Vandesompele J, Speleman F. Neuroblastoma epigenetics: from candidate gene approaches to genome-wide screenings. Epigenetics. 2011; 6:962-970. https://doi.org/10.4161/epi.6.8.16516.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Olsson M, Beck S, Kogner P, Martinsson T, Carén H. Genome-wide methylation profiling identifies novel methylated genes in neuroblastoma tumors. Epigenetics. 2016; 11:74-84. https://doi.org/10.1080/15592294.2016.1138195.

Hassan W, Bakry M, Siepmann T, Illigens B. Association of RASSF1A, DCR2, and CASP8 methylation with survival in neuroblastoma: a pooled analysis using reconstructed individual patient data. Biomed Res Int. 2020; 2020:7390473. https://doi.org/10.1155/2020/7390473.

Decock A, Ongenaert M, Cannoodt R, Verniers K, De Wilde B, Laureys G, et al. Methyl-CpG-binding domain sequencing reveals a prognostic methylation signature in neuroblastoma. Oncotarget. 2016; 7:1960–1972. https://doi.org/10.18632/oncotarget.6477.

Hiyama E, Hiyama K, Yokoyama T, Fukuba I, Yamaoka H, Shay JW, et al. Rapid detection of MYCN gene amplification and telomerase expression in neuroblastoma. Clin Cancer Res. 1999; 5:601-609.

Oude Luttikhuis M, Iyer VK, Dyer S, Ramani P, McConville, CM. Detection of MYCN amplification in neuroblastoma using competitive PCR quantitation. Lab Invest. 2000; 80:271–273. https://doi.org/10.1038/labinvest.3780030.

Mathew P, Valentine MB, Bowman LC, Rowe ST, Nash MB, Valentine VA, et al. Detection of MYCN gene amplification in neuroblastoma by fluorescence in situ hybridization: a pediatric oncology group study. Neoplasia. 2001; 3:105-109. https://doi.org/10.1038/sj.neo.7900146.

Yagyu S, Iehara T, Tanaka S, Gotoh T, Misawa-Furihata A, Sugimoto T, et al. Serum-Based quantification of MYCN gene amplification in young patients with neuroblastoma: potential utility as a surrogate biomarker for neuroblastoma. PLoS One. 2016; 11:e0161039. https://doi.org/10.1371/journal.pone.0161039.

Di Lena P, Sala C, Prodi A, Nardini C. Missing value estimation methods for DNA methylation data. Bioinformatics. 2019; 35:3786–3793. https://doi.org/10.1093/bioinformatics/btz134.

Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013; 14:128. https://doi.org/10.1186/1471-2105-14-128.

Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016; 44: W90-W97. https://doi.org/10.1093/nar/gkw377.

Kassambara A, Mundt F. Factoextra R package. https://cran.r-project.org/. 2020.

Gómez S, Castellano G, Mayol G, Queiros A, Martín-Subero JI, Lavarino C. DNA methylation fingerprint of neuroblastoma reveals new biological and clinical insights. Genom Data. 2015; 5:360-363. https://doi.org/10.1016/j.gdata.2015.07.016.

Ackermann S, Cartolano M, Hero B, Welte A, Kahlert Y, Roderwieser A, et al. A mechanistic classification of clinical phenotypes in neuroblastoma. Science. 2018; 362:1165-1170. https://doi.org/10.1126/science.aat6768.

Schramm A, Koster J, Assenov Y, Althoff K, Peifer M, Mahlow E, et al. Mutational dynamics between primary and relapse neuroblastomas. Nat Genet. 2015; 47:872-877. https://doi.org/10.1038/ng.3349.

http://etd.uwc.ac.za/

Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010; 33:1–22.

Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. J Stat Softw. 2011; 39:1-13. https://doi.org/10.18637/jss.v039.i05.

Tibshirani R, Bien J, Friedman J, Hastie T, Simon N, Taylor J, et al. Strong rules for discarding predictors in lasso-type problems. J R Stat Soc Series B Stat Methodol. 2012; 74:245-266. https://doi.org/10.1111/j.1467-9868.2011.01004.x.

Ng SW, Mitchell A, Kennedy JA, Chen WC, McLeod J, Ibrahimova N, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. Nature. 2016; 540:433–437. https://doi.org/10.1038/nature20598.

Duan C, Wang H, Chen Y, Chu P, Xing T, Gao C, et al. Whole exome sequencing reveals novel somatic alterations in neuroblastoma patients with chemotherapy. Cancer Cell Int. 2018; 18:21. https://doi.org/10.1186/s12935-018-0521-3.

Li Z, Li J, Ji D, Leng K, Xu Y, Huang L, Jiang X, Cui Y. Overexpressed long noncoding RNA Sox2ot predicts poor prognosis for cholangiocarcinoma and promotes cell proliferation and invasion. Gene. 2018; 645:131-136. https://doi.org/10.1016/j.gene.2017.12.017.

Zhang TJ, Xu ZJ, Gu Y, Wen XM, Ma JC, Zhang W, et al. Identification and validation of prognosis-related DLX5 methylation as an epigenetic driver in myeloid neoplasms. Clin Transl Med. 2020; 10:e29. https://doi.org/10.1002/ctm2.29.

Giwa A, Fatai A, Gamieldien J, Christoffels A, Bendou H. Identification of novel prognostic markers of survival time in high-risk neuroblastoma using gene expression profiles. Oncotarget. 2020; 11:4293-4305. https://doi.org/10.18632/oncotarget.27808.

Walker C, Mojares E, Del Río Hernández A. Role of extracellular matrix in development and cancer progression. Int J Mol Sci. 2018; 19:3028. https://doi.org/10.3390/ijms19103028.

Poltavets V, Kochetkova M, Pitson SM, Samuel MS. The role of the extracellular matrix and its molecular and cellular regulators in cancer cell plasticity. Front Oncol. 2018; 8:431. https://doi.org/10.3389/fonc.2018.00431.

Eble JA, Niland S. The extracellular matrix in tumor progression and metastasis. Clin Exp Metastas. 2019; 36:171-198. https://doi.org/10.1007/s10585-019-09966-1.

Nallanthigal S, Heiserman JP, Cheon D. The role of the extracellular matrix in cancer stemness. Front Cell Dev Biol. 2019; 7:86. https://doi.org/10.3389/fcell.2019.00086.

Jia M, Gao X, Zhang Y, Hoffmeister M, Brenner H. Different definitions of CpG island methylator phenotype and outcomes of colorectal cancer: a systematic review. Clin Epigenetics. 2016; 8:25. https://doi.org/10.1186/s13148-016-0191-8.

Teitz T, Wei T, Valentine MB, Vanin EF, Grenet J, Valentine VA, et al. Caspase 8 is deleted or silenced preferentially in childhood neuroblastomas with amplification of MYCN. Nat Med. 2000; 6:529–535. https://doi.org/10.1038/75007.

Fulda S, Poremba C, Berwanger B, Häcker S, Eilers M, Christiansen H, Hero B, Debatin KM. Loss of caspase-8 expression does not correlate with MYCN amplification, aggressive disease, or

http://etd.uwc.ac.za/

prognosis in neuroblastoma. Cancer Res. 2006; 66:10016–10023. https://doi.org/10.1158/0008-5472.CAN-05-4079.

Zheng H, Zhang Y, Zhan Y, Liu S, Lu J, Wen Q, Fan S. Expression of DR5 and c-FLIP proteins as novel prognostic biomarkers for non-small cell lung cancer patients treated with surgical resection and chemotherapy. Oncol Rep. 2019; 42:2363-2370. https://doi.org/10.3892/or.2019.7355.

Lowe SW, Lin AW. Apoptosis in cancer. Carcinogenesis. 2000; 21:485–495. https://doi.org/10.1093/carcin/21.3.485.

Pfeffer CM, Singh ATK. Apoptosis: a target for anticancer therapy. Int J Mol Sci. 2018; 19:448. https://doi.org/10.3390/ijms19020448.

Wang Y, Luo H, Cao J, Ma C. Bioinformatic identification of neuroblastoma microenvironment-associated biomarkers with prognostic value. J Oncol. 2020; 2020:5943014. https://doi.org/10.1155/2020/5943014.

Gonzalez-Moreno O, Boque N, Redrado M, Milagro F, Campion J, Endermann T, Takahashi K, Saito Y, Catena R, Schomburg L, Calvo A. Selenoprotein-P is down-regulated in prostate cancer, which results in lack of protection against oxidative damage. The Prostate. 2011; 71:824-834. https://doi.org/10.1002/pros.21298.

Wang Q, Gong L, Dong R, Qiao Q, He XL, Chu YK, Du XL, Yang Y, Zang L, Nan J, Lin C, Lu JG. Tissue microarray assessment of selenoprotein P expression in gastric adenocarcinoma. J Int Med Res. 2009; 37:169-174. https://doi.org/10.1177/147323000903700120.

Meyer HA, Endermann T, Stephan C, Stoedter M, Behrends T, Wolff I, Jung K, Schomburg L. Selenoprotein P status correlates to cancer-specific mortality in renal cancer patients. PLoS One. 2012; 7:e46644. https://doi.org/10.1371/journal.pone.0046644.

Monteith GR, McAndrew D, Faddy HM, Roberts-Thomson SJ. Calcium and cancer: targeting Ca2+ transport. Nat Rev Cancer. 2007; 7:519-530. https://doi.org/10.1038/nrc2171.

Satheesh NJ, Büsselberg D. The role of intracellular calcium for the development and treatment of neuroblastoma. Cancers (Basel). 2015; 7:823-848. https://doi.org/10.3390/cancers7020811.

Florea AM, Varghese E, McCallum JE, Mahgoub S, Helmy I, Varghese S, Gopinath N, Sass S, Theis FJ, Reifenberger G, Büsselberg D. Calcium-regulatory proteins as modulators of chemotherapy in human neuroblastoma. Oncotarget. 2017; 8:22876-22893. https://doi.org/10.18632/oncotarget.15283.

Farhangian P, Jahandoost S, Mowla SJ, Khalili M. Differential expression of long non-coding RNA SOX2OT in gastric adenocarcinoma. Cancer Biomark. 2018; 23:221-225. https://doi.org/10.3233/CBM-181325.

Qu F, Cao P. Long noncoding RNA SOX2OT contributes to gastric cancer progression by sponging miR-194-5p from AKT2. Exp Cell Res. 2018; 369:187-196. https://doi.org/10.1016/j.yexcr.2018.05.017.

Li Z, Jiang P, Li J, Peng M, Zhao X, Zhang X, et al. Tumor-derived exosomal lnc-Sox2ot promotes EMT and stemness by acting as a ceRNA in pancreatic ductal adenocarcinoma. Oncogene. 2018; 37:3822-3838. https://doi.org/10.1038/s41388-018-0237-9.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Shi XM, Teng F. Up-regulation of long non-coding RNA Sox2ot promotes hepatocellular carcinoma cell metastasis and correlates with poor prognosis. Int J Clin Exp Pathol. 2015; 8:4008-4014.

Li Z, Li J, Ji D, Leng K, Xu Y, Huang L, Jiang X, Cui Y. Overexpressed long noncoding RNA Sox2ot predicts poor prognosis for cholangiocarcinoma and promotes cell proliferation and invasion. Gene. 2018; 645:131-136. https://doi.org/10.1016/j.gene.2017.12.017.

Merlo GR, Zerega B, Paleari L, Trombino S, Mantero S, Levi G. Multiple functions of Dlx genes. Int J Dev Biol. 2000; 44:619-626.

Heo JS, Lee SG, Kim HO. Distal-less homeobox 5 is a master regulator of the osteogenesis of human mesenchymal stem cells. Int J Mol Med. 2017; 40:1486-1494. https://doi.org/10.3892/ijmm.2017.3142.

Mitchell SM, Ross, JP, Drew, HR, Ho T, Brown GS, Saunders NF, Duesing KR, Buckley MJ, Dunne R, Beetson I, Rand KN, McEvoy A, Thomas ML, Baker RT, Wattchow DA, Young GP, Lockett TJ, Pedersen SK, Lapointe LC, Molloy PL. A panel of genes methylated with high frequency in colorectal cancer. BMC Cancer. 2014; 14:54. https://doi.org/10.1186/1471-2407-14-54.

Karsli-Ceppioglu S, Dagdemir A, Judes G, Lebert A, Penault-Llorca F, Bignon YJ, Bernard-Gallon D. The epigenetic landscape of promoter genome-wide analysis in breast cancer. Sci Rep. 2017; 7:6597. https://doi.org/10.1038/s41598-017-06790-z.

Li C, Wang S, Yang C. Long non-coding RNA DLX6-AS1 regulates neuroblastoma progression by targeting YAP1 via miR-497-5p. Life Sci. 2020; 252:117657. https://doi.org/10.1016/j.lfs.2020.117657.

Sun L, Zhao Y, Gu S, Mao Y, Ji C, Xin X. Regulation of the HMOX1 gene by the transcription factor AP-2δ with unique DNA binding site. Mol Med Rep. 2014; 10:423-428. https://doi.org/10.3892/mmr.2014.2196.

Li Z, Xu X, Luo M, Hao J, Zhao S, Yu W, et al. Activator Protein-2β Promotes Tumor Growth and Predicts Poor Prognosis in Breast Cancer. Cell Physiol Biochem. 2018; 47:1925–1935. https://doi.org/10.1159/000491463.

Ebert MPA, Tänzer M, Balluff B, Burgermeister E, Kretzschmar AK, Hughes DJ, et al. TFAP2E–DKK4 and chemoresistance in colorectal cancer. N Engl J Med. 2012; 366:44-53. https://doi.org/10.1056/NEJMoa1009473.

Hoshi R, Watanabe Y, Ishizuka Y, Hirano T, Nagasaki-Maeoka E, Yoshizawa S, Uekusa S, Kawashima H, Ohashi K, Sugito K, Fukuda N, Nagase H, Soma M, Ozaki T, Koshinaga T, Fujiwara K. Depletion of TFAP2E attenuates adriamycin-mediated apoptosis in human neuroblastoma cells. Oncol Rep. 2017; 37: 2459-2464. https://doi.org/10.3892/or.2017.5477.

Lee JH, Byun DS, Lee MG, Ryu BK, Kang MJ, Chae KS, et al. Frequent epigenetic inactivation of hSRBC in gastric cancer and its implication in attenuated p53 response to stresses. Int J Cancer 2008; 122:1573–1584. https://doi.org/10.1002/ijc.23166.

Roy D, Calaf G, Hei TK. Allelic imbalance at 11p15.5–15.4 correlated with c-Ha-ras mutation during radiation-induced neoplastic transformation of human breast epithelial cells. Int J Cancer 2003; 103:730–737. https://doi.org/10.1002/ijc.10895.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Kozlowski L, Filipowski T, Rucinska M, Pepinski W, Janica J, Skawronska M, Poznanski J, Wojtukiewicz MZ. Loss of heterozygosity on chromosomes 2p, 3p, 18q21.3 and 11p15.5 as a poor prognostic factor in stage II and III (FIGO) cervical cancer treated by radiotherapy. Neoplasma. 2006; 53:440–443.

Panani AD, Ferti AD, Raptis SA, Roussos C. Novel recurrent structural chromosomal aberrations in primary bladder cancer. Anticancer Res. 2004; 24:2967–2974.

Zhao B, Bepler G. Transcript map and complete genomic sequence for the 310 kb region of minimal allele loss on chromosome segment 11p15.5 in non-small-cell lung cancer. Oncogene. 2001; 20:8154–8164. https://doi.org/10.1038/sj.onc.1205027.

Moskaluk CA, Rumpel CA. Allelic deletion in 11p15 is a common occurrence in esophageal and gastric adenocarcinoma. Cancer. 1998; 83:232–239.

Lee JH, Kang MJ, Han HY, Lee MG, Jeong SI, Ryu BK, et al. Epigenetic alteration of PRKCDBP in colorectal cancers and its implication in tumor cell resistance to TNFα-induced apoptosis. Clin. Cancer Res. 2011; 17:7551-7562. https://doi.org/10.1158/1078-0432.CCR-11-1026.

Wikman H, Sielaff-Frimpong B, Kropidlowski J, Witzel I, Milde-Langosch K, Sauter G, et al. Clinical relevance of loss of 11p15 in primary and metastatic breast cancer: association with loss of PRKCDBP expression in brain metastases. PLoS One. 2012; 7:e47537. https://doi.org/10.1371/journal.pone.0047537.

Sharma P, Alsharif S, Bursch K, Parvathaneni S, Anastasakis DG, Chahine J, Fallatah A, Nicolas K, Sharma S, Hafner M, Kallakury B, Chung BM. Keratin 19 regulates cell cycle pathway and sensitivity of breast cancer cells to CDK inhibitors. Sci Rep. 2019; 9:14650. https://doi.org/10.1038/s41598-019-51195-9.

Yuan X, Yi M, Dong B, Chu Q, Wu K. Prognostic significance of KRT19 in Lung Squamous Cancer. J Cancer. 2021; 12:1240-1248. https://doi.org/10.7150/jca.51179.

Saloustros E, Perraki N, Apostolaki S, Kallergi G, Xyrafas A, Kalbakis K, et al. Cytokeratin-19 mRNA-positive circulating tumor cells during follow-up of patients with operable breast cancer: prognostic relevance for late relapse. Breast Cancer Res. 2011; 13:R60. https://doi.org/10.1186/bcr2897.

Kabir NN, Rönnstrand L, Kazi JU. Keratin 19 expression correlates with poor prognosis in breast cancer. Mol Biol Rep. 2014; 41:7729-7735. https://doi.org/10.1007/s11033-014-3684-6.

Sahm F, Schrimpf D, Stichel D, Jones DTW, Hielscher T, Schefzyk S et al.DNA methylation-based classification and grading system for meningioma: a multicentre, retrospective analysis. Lancet Oncol. 2017; 18(5):682–694. https://doi.org/10.1016/S1470-2045(17)30155-9.

Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV, & on behalf of the CCGA Consortium. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. Ann Oncol. 2020; 31(6):745-759. https://doi.org/10.1016/j.annonc.2020.02.011.

http://etd.uwc.ac.za/