IMMUNOPHENOTYPIC AND MOLECULAR FEATURES OF *TP53* IN DIFFUSE LARGE B-CELL LYMPHOMA (DLBCL)

By: Thembi Mashele

Supervised by:

Dr Pumza Magangane

Dr Sugeshnee Pather

A dissertation submitted to the Department of Anatomical Pathology,

School of Pathology, University of the Witwatersrand, Johannesburg

in fulfilment of the requirements for

the degree of Master of Science in Medicine in Anatomical Pathology by research.

Johannesburg, 15 October 2025

DECLARATION

I, Thembi Mashele, declare that this Dissertation; titled: *Immunophenotypic and Molecular features of TP53 in Diffuse Large B-Cell Lymphoma (DLBCL)* is my own, unaided work. It is being submitted for the Degree of Master in Science in Medicine in Anatomical Pathology by research, at the School of Pathology, University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

(Signature of candidate)

15th day of October 2025 in Braamfontein

DEDICATION STATEMENT

I dedicate this dissertation to my mother, Madikate Welhemina Mashele and sister, Thembiso Kodisang, for all the support they have given me throughout this process. Moreover, I also pay tribute to God.

It was God

"I survived another day because of God's grace.

How can I not believe that God is with me?

when I have days that I know I only survived because of his grace?

Thank you God for everything, you never fail."

Unknown

A modimo le badimo continue to shine light throughout my existence.

ABSTRACT

Introduction: Diffuse large B-cell lymphomas (DLBCL) are prevalent among HIV-seropositive patients in Southern Africa. The *TP53* gene is the most frequently mutated gene in many tumours. However, there is paucity of data regarding p53 immunoexpression and *TP53* mutations in DLBCL.

Aims: The study aimed to determine the p53 immunoexpression profile and *TP53* gene mutations associated with DLBCL at Chris Hani Baragwanath Academic Hospital (CHBAH).

Methods: A retrospective cross-sectional study was conducted using confirmed cases of DLBCL from 2013 to 2017. Immunohistochemistry for p53 protein expression and mutational analysis on *TP53* exons 5 to 8 were performed.

Results: Sixty patients (N=60) with DLBCL were included with an age range of 22-88 years and a median age of 45 years (38-54 years). p53 immunoexpression of ≥ 50% was observed in 27 (45%) of cases, of which 19 (70.4%) were HIV seropositive. HIV seropositive patients with p53 protein expression tended to be younger in age, predominantly male with extranodal involvement and also demonstrated a germinal centre (GCB) subtype. *TP53* genetic mutations were documented in 10% of the cases and 83.3% thereof expressed p53. *TP53* catalogued frameshift mutations were observed in 35 of 50 (70%). Majority of DLBCL mutations were on the NDBL/beta-sheets structural motifs.

Conclusion: DLBCL with p53 immunoexpression and *TP53* mutations occurred in younger HIV seropositive patients with extranodal involvement and exhibited the GCB subtype.

ACKNOWLEDGEMENTS

CHBAH and NHLS, Anatomical Pathology department for allowing the use of their facilities and patient materials.

CMJAH and NHLS, Anatomical Pathology department for granting access to utilise their facilities and equipment.

University of the Witwatersrand Postgraduate Masters Merit Award for financing my academic fees.

NHLS- Learning Academy in the Training department for funding my academic fees.

Dr Pumza Magangane and Dr Sugeshnee Pather; my supervisors for the support and valuable input they have given my research and for assisting me to fund this study.

Mrs Mpho Edith Modisane, for assisting me with the retrieval and filling of study blocks and slides.

Ms Madi Nzeba for assisting me with the validation of p53 immunohistochemical stains.

Ms Hellen Nxumalo for helping me with the retrieval of missing reports.

FUNDING

University of the Witwatersrand Postgraduate Masters Merit Award for financing my academic fees.

NHLS- Learning Academy in the Training department for funding my academic fees.

University of the Witwatersrand RINC (Research Incentive) funding, thank you to the Department of Anatomical Pathology for providing funding support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

5’	Five Prime
3’	Three Prime
%	Percentage
µl	Microlitre
µm	Micrometer
A	Adenine
ABC	Activated B-cell
AIDS	Acquired Immunodeficiency Syndrome
bp	Base-Pair
BCL2	B-cell leukemia/lymphoma 2 protein
BCL6	B-cell leukemia/lymphoma 6 protein
C	Cytosine
CD10	Cluster of differentiation 10
CDW	Central Data Warehouse
CHBAH	Chris Hani Baragwanath Academic Hospital
CMJAH	Charlotte Maxeke Johannesburg Academic Hospital
COO	Cell-of-origin
dH2O	Distilled Water
dNTP	Deoxyribonucleotide Triphosphate
DBD	DNA binding domain
DH-HGBL	Double-hit lymphoma
DLBCL	Diffuse large B-cell lymphoma
DLBCL,NOS	Diffuse large B-cell lymphoma, not otherwise specified
DNA	Deoxyribonucleic acid
EBV	Epstein-Barr virus
FFPE	Formalin-fixed paraffin-embedded
G	Guanine

GCB	Germinal centre B-cell
GCET1	Germinal centre B cell-expressed transcript 1
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HHV-8	Human herpes virus-8
HIER	Heat Induce Epitope Retrieval
HIV	Human Immunodeficiency Virus
IARC	International Agency for Research on Cancer
IHC	Immunohistochemical
Ki67	Kiel 67
KSHV	Kaposi sarcoma-associated herpes virus
mmol	Millimoles
MUM-1	Multiple Myeloma Oncogene-1
NHL	Non-Hodgkin Lymphomas
NHLS	National Health Laboratory Service
PCR	Polymerase Chain Reaction
SIR	Standard Incidence Ratio
SNOMED	Systematized Nomenclature of Medicine
T	Thymine
TP53	Tumour protein 53
TH-HGBL	Triple-hit lymphoma
WT	Wild type

CHAPTER ONE: BACKGROUND

PART A: LITERATURE REVIEW OF DIFFUSE LARGE B-CELL LYMPHOMA (DLBCL)

1.1. Definitions and subtypes

Non-Hodgkin lymphoma's (NHL) are the most prevalent haematological malignancies. Diffuse large B-cell lymphoma (DLBCL) is the most common high grade B-cell NHL accounting for approximately 40% of the NHL cases.[1] According to the 2021 World Health Organization (WHO) classification of tumours of hematopoietic and lymphoid tissue, the diagnosis of DLBCL is not otherwise specified (NOS) is based on molecular and morphological features.[2,3]

1.2. Prevalence

DLBCL tends to develop in patients who are 65 years or older.[4] It is more prevalent in males than females,[5] and in the United States of America (USA) a high prevalence is documented in white males.[5]

According to the International Agency for Research on Cancer (IARC), NHL has a global incidence of 544 352, and an African incidence of 50 516, of which 3 500 cases were accounted for within Southern Africa during 2020.[6] A global NHL-associated mortality incidence of 259 793 was recorded with an incidence of 30 960 in Africa and 1 797 in South Africa.[6] NHL is the seventh most common cancer in men and the sixth most common cancer in women in Southern Africa.[6]

In South Africa, NHL is the eighth most common cancer and the ninth leading cause of cancer related mortality.[7] There is a higher incidence and mortality rate of men to women with DLBCL.[7,8]

1.3. Epidemiology and aetiology of DLBCL

The exact cause of DLBCL is unknown but risk factors associated with DLBCL are both genetic and non-genetic. The greatest risk is associated with factors that cause severe immune deficiency, while other risk factors include autoimmune conditions, infectious agents and hereditary diseases.[4,5]

1.3.1. Autoimmune conditions

Long-term inflammation and chronic immune stimulation from autoimmune conditions such as systemic lupus erythematosus (SLE) and rheumatoid arthritis may consequently predispose an individual to DLBCL.[4]

Several other autoimmune conditions have been associated with DLBCL, such as haemolytic anaemia and Sjögren syndrome.[5] The association of primary Sjögren syndrome to NHL was evaluated by Voulgarelis M *et al*. in a cohort of 584 patients, 53 of whom had NHL and DLBCL was the third most common NHL (15%).[9]

1.3.2. Hereditary factors

Personal history of a cancer is a risk factor and NHL is the most common malignancy to arise as a second cancer, of which DLBCL is most frequent.[5] According to research based on the data registry of Sweden and Denmark, a family history of cancer increases the standard incidence ratio (SIR) to 2.3 for NHL and 11.8 for DLBCL, if the cancer was in a first degree relative such as parents, siblings or children.[5] A study on European ancestry established several genes which may be associated with risk for DLBCL. The study found five genomes and four loci, the role of HLA-B*08:01 in DLBCL risk is said to be enhanced by the 6p21.33 loci.[5]

1.3.3. Organ transplants

Immunosuppression during solid organ transplantation has been associated with high risk of developing DLBCL. In the context of organ transplants, the SIR for DLBCL risk was determined to be 12.6 post one year of transplantation.[10] Patients with the highest risk include those who received an organ at a young age, and those who had transplanted a kidney, pancreas or lung.[5,10]

1.3.4. Infections

Conditions affecting the immune system are risks factors associated with DLBCL. These include oncogenic viral infections such as Human Immunodeficiency virus (HIV), Epstein-Barr virus (EBV), hepatitis C virus (HCV) and human herpes virus-8 (HHV8).[4,5] There is also emerging evidence for an association with hepatitis B virus (HBV)[5]. These agents are listed by International Agency for Research on Cancer (IARC) as class I human carcinogens.[11]

1.3.4.1. DLBCL in the setting of Epstein-Barr virus (EBV)

During 2022, the diagnostic entity EBV positive DLBCL was recognised by the World Health Organization (WHO).[3] EBV+DLBCL cases comprise approximately 15% of DLBCL cases.[3,12] In elderly patients, older than 50 years, it is seen in extra-nodal sites such the gastro-intestinal tract (GIT), skin and lungs. It most commonly presents in the lymph nodes of younger patients, especially those in their thirties.[12] The B-cells appear neoplastic, comprising large blasts in monotonous sheets with prominent areas of coagulative geographical necrosis. [12] This lymphoma is positive in CD20, PAX5 and CD19 B-cell markers with CD30 being expressed in 40% of the cases.[12] EBV is detected in majority of the tumour cells and this lymphoma is associated with poor prognosis.[12]

1.3.4.2. DLBCL in the setting of Human Herpes Virus-8 (HHV8)

An oncogenic double-stranded DNA virus, HHV8 also known as Kaposi Sarcoma-Associated Herpes Virus (KSHV) affects various cells including hematopoietic progenitors and B-cells.[13,14] HHV8-positive DLBCL, has been newly categorised in the recent 2021 WHO classification and is commonly associated with the HHV8 positive multicentric Castleman disease and HIV infection.[3] In a retrospective analysis conducted by Wang on patients during the timeframe 2011-2020, 67 patients were diagnosed with HHV8-positive DLBCL, majority of them were Caucasian (65.7%).[13] The patients had a median age of 51.8 years at diagnosis and males represented the majority (79.1%).[13]

1.3.4.3. DLBCL in the setting of Human Immunodeficiency Virus (HIV)

Although the aetiology of DLBCL is undefined, there is an established association with HIV infection.[15]

South Africa is an epicentre of the HIV pandemic, representing the largest AIDS epidemic globally.[4] In the HIV seropositive patients, DLBCL is often documented at least two decades earlier with a mean/median age of 39-42 years.[4] Herewith aligned, DLBCL studies within South Africa by Patel *et al.* and Magangane *et al.* showed a median age of 40 years with a predominance of extra nodal involvement.[16,17]

In contrast, in the HIV-seronegative patients, DLBCL tends to occur at 65 years of age or older.[4] HIV associated DLBCL is the most common type of DLBCL occurring in elderly patients and HIV seropositive patients.[6] It accounts for 80% of all DLBCL.[4]

High-grade B-cell NHL is one of the acquired immunodeficiency syndrome (AIDS) defining malignancies of which DLBCL is the most common subtype that occurs in the context of HIV infection.[18] In the quest to document the global burden of HIV associated NHL, Chen *et al*. sourced 14,929 literature articles of which 39 (0.26%) met their study criteria,[19] and it was established that globally, 7% of the cases were attributed to HIV infection with 44% of those cases detected from the Eastern and Southern African regions.[19]

1.4. Pathophysiology of DLBCL

The classification of DLBCL is based on how the B-cells function in the body. B cell lymphomas are categorised based on the stage of B cell development that has resulted in malignancy. This may be due to the development of genetic mutations, gene amplifications, altered gene expressions and translocations to chromosomal sites.[1,4] These genetic changes occur frequently in *BCL2, BCL6, MYC, MYD88, CD79A, CD79B, PAX5, EZH2 and CREBBP*.[20] Alterations in these genes cause disease development and progression by loss of function and or gain of function resulting in the dysregulation of the cell signalling pathways.[1,4,20]

1.5. Presentation and presentation sites of DLBCL

1.5.1. Clinical symptoms

Symptoms of DLBCL are highly dependent on the site of the cancerous mass. The most common symptoms include: fatigue, night sweats, loss of appetite, weight loss and fever.[20] Depending on the topographic site of the mass, a patient may experience enlarged lymph nodes, abdominal pain, swelling of the face and/or arms, bloody stool, chest pains, shortness of breath, neurological symptoms or skin rash.[20]

DLBCL most commonly affects the lymph nodes and may present in extranodal topographic sites such as the gastrointestinal tract, bone, testis, spleen, thyroid, and lung, especially in HIV seropositive patients.[1–3]

1.5.1.1. <u>Nodal disease</u>

Nodal involvement is the most common site of DLBCL and said to affect 60-70% of patients. The criteria used to classify DLBCL as nodal include the enlarged lymph node, the lymph node as the primary topographic site of involvement, diffuse growth pattern of large B-cell on histopathological examination of sections and immunoexpression of CD20 and CD79a.[3,4]

1.5.1.2. Extra-nodal

Extra-nodal presentation of DLBCL is on the rise with approximately 40% of DLBCL cases presenting in extranodal sites.[21] The gastrointestinal tract is said to be the most common extra-nodal site of origin.[21] In several studies, the GIT was documented to account for approximately 40% of extra-nodal DLBCL cases.[22,23] The small intestine including the terminal ileum, stomach, caecum and the rectum are the prevalent sites, whereas, oesophageal involvement is rare.[23] The activated B-cell (ABC) phenotype is more prominent in gastric DLBCL than it is in intestinal DLBCL.[23]

1.6. Sub-classification of DLBCL

In 2022, the 5th edition of the WHO classification of Haematolymphoid Tumours focussing on lymphoid neoplasms was released.[3] The changes made in large B-cell lymphomas included the revision of;

1. Diffuse large B-cell lymphoma/ high grade B-cell lymphoma with *MYC* and *BCL2*,
2. High-grade B-cell lymphoma with 11q aberrations,
3. EBV-positive diffuse large B-cell lymphoma, and
4. Mediastinal grey zone lymphoma.[3]

Lymphoma subtypes not previously included are:

1. Fibrin-associated large B-cell lymphoma,
2. Fluid overload-associated large B-cell, and
3. Primary large B-cell lymphoma of immune-privileged sites.[3]

1.6.1. T-cell/histiocyte-rich B-cell lymphoma

This type of DLBCL constitutes large atypical B cells with a background filled with white blood cells (T-cell/histiocytes).[6] The histopathological criteria used for diagnosis constitutes a cellular population of less than 10% of malignant B-cells and 90% or more of reactive T-cells with scattered neoplastic B-cells in the background.[24]

1.6.2. Primary DLBCL of the central nervous system (CNS)

DLBCL primary site is in the brain or spinal cord. The biological and pathological features of this subtype are similar to those of DLBCL arising in the testis. This type

of lymphoma is associated with diffuse growth patterns in perivascular spaces and rarely with lymphomatosis cerebri.[6] The prognosis is notably unfavourable.

1.6.3. Primary cutaneous DLBCL, leg type

The primary site of involvement in this DLBCL subtype is the skin on the legs or cutaneous area of the arm or buttocks area.[6] According to Xie *et al.* elderly females are most affected by this disease compared to males.[6] The morphological characteristics of this subtype comprises monotonous proliferation of immunoblasts, sometimes centroblast-like cells, with several admixed reactive cells.[6] The immunophenotypic characteristics are said to mimic the activated B-cell type of nodal DLBCL.[6]

1.7. DLBCL, NOS

DLBCL, NOS do not measure up to be included in specific disease subgroups and is the most common type of DLBCL.[6] This group may comprise three morphologic variants such as centroblastic, immunoblastic and anaplastic. A combination of these variants may also be evident in a tumour.[6]

1.7.1. Classification of DLBCL, NOS

1.7.1.1. Hans algorithm

A malignant proliferation of B-cells leads to the development of B cell lymphomas. According to the cell-of-origin (COO) these lymphomas are subdivided into two subtypes: the germinal centre B-cell (GCB) and ABC subtypes, each displaying different molecular profiles.[1,4,25] Immunohistochemistry can be utilised as a surrogate for the molecular profiles of DLBCL, NOS, by utilising the Hans algorithm.[1,3,4,25]

Has *et al.* recommended the use of CD10, Bcl-6, and MUM1 immunohistochemistry to classify the subgroups based on protein expression, Figure 1.1.[1,4,25] The algorithm is based on immunoexpression of at least 30% of positive tumour cells, the GCB subtype is defined as CD10+ or CD10-, Bcl-6+ and MUM1-. The ABC subtype is defined as CD10- , Bcl-6+ or - and MUM1+.[4,25]

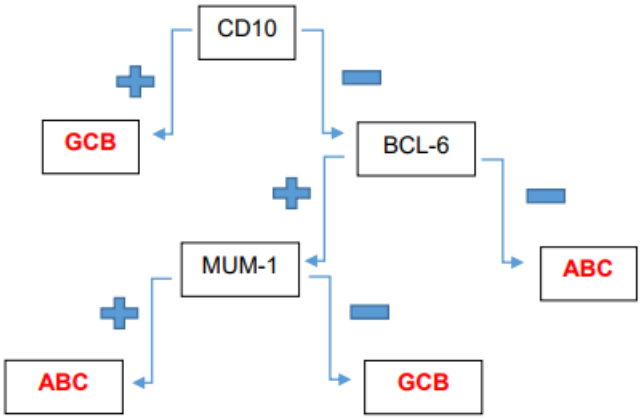


Figure 1.1: Definition of the Hans algorithm subtypes by IHC (adapted from Hans *et al*. 2004)

In the study of 152 patients conducted by Hans *et al*. 58% (n=88) of the patients had non-GCB subtype of DLBCL, with a median age of 64 years and the majority (60%) were male. The GCB subtype of DLBCL (n=64, 42%) had a median age of 60 years and 55% were female.[25]

1.7.1.2. Other reported algorithms

Apart from the Hans algorithm, several other models have been established and reported for determining the COO by IHC. These include the Choi model, the Tally algorithm, the Nyman, the Colomo and the Visco-Young with three [VY3] and four [VY4] antibodies.[6]

The Tally and the Choi algorithms determine the DLBCL, NOS COO subtypes based on the immunoexpression of CD10 (positive ≥30%), BCL-6 (positive ≥30%), MUM1 (positive ≥80%).[6,26,27] They also make use of the immunoexpression of germinal center B cell-expressed transcript 1 (GCET1) (positive ≥80%) and FOXP1 (positive ≥80%) biomarkers.[26] The Tally algorithm has been found to be accurate in 90% cases.[27]

1.8. Morphological variants

Diffuse large B-cell lymphomas are composed of large cells comprising of nuclei that are at least twice the size of a small lymphocyte (usually larger than the nuclei of a

macrophage). The nuclei appear vesicular and have prominent nucleoli with a basophilic cytoplasm.[28] Several variants have been reported.

1.8.1. Centroblastic variant

The centroblastic variant is the most common, accounting for approximately 80% of the DLBCL cases. It is said to have a better overall survival compared to other variants.[29] This variant consists of centroblasts with moderate to scanty cytoplasm, non-cleaved nuclei which are round to oval, located centrally, with multiple peripherally located nucleoli.[29]

1.8.2. Immunoblastic variant

The immunoblastic variant comprises immunoblasts which constitute more than 90% of tumour cells.[30] This variant accounts for 10% of the DLBCL cases. These cells display abundant basophilic cytoplasm with a large round to oval nucleus and a single, prominent and central nucleolus.[30]

1.8.3. Anaplastic variant

The anaplastic variant of DLBCL is very rare, accounting for only 3% of the cases. The cells appear enlarged with markedly pleomorphic and bizarre nuclei.[31]

1.8.4. T-cell/histiocytic-rich variant

The T- cell/histiocytic-rich variant is considered very rare and aggressive, and only occurs in 1-3% of cases.[24] It is characterised by an inflammatory infiltrate of active, small polyclonal T cells and histiocytes with scanty large neoplastic B cells which comprise less than 10%.[24]

1.9. Biomarkers associated with DLBCL

1.9.1. CD10

Cluster of differentiation 10 (CD10) is a germinal center phenotype immuno-marker that stains the membrane and cytoplasm. It is encoded by MME gene. In DLBCL it is generally used as a germinal center phenotype marker.[32] It is located on chromosome 3 at 3q25.2 cytogenetic band and plays a role in the development and differentiation of lymphoid cells.[32]

The biomarker appears upregulated in many malignancies such as lung adenocarcinoma, head and neck squamous cell carcinoma, colorectal cancer, breast

cancer and relapse post radical prostatectomy in prostate cancer. It may also be linked to the severity of the disease and stage.[32]

1.9.2. CD20

Cluster of differentiation 20 (CD20) is membrane bound.[33] It is important for B-cell development and is expressed until they develop into plasma cells. it plays a significant role in B-cell activation, differentiation, and cell-cycle progression.[33]

It is encoded by the MS4A1 gene which is located on chromosome 11q12.2.[34]

In a study conducted by Li *et al*. it was concluded that CD20 negative DLBCL is very rare (1-2%), aggressive and it is associated with poor prognosis.[33] CD20 expression is a key target in the treatment of diffuse large B-cell lymphoma (DLBCL).[33] CD20 negative DLBCL patients may not respond well to the standard CHOP therapy and require a combination of treatment modules.[34]

1.9.3. KI67

The Kiel 67 (Ki-67) is a nuclear marker, and stems from the *MKI67* gene. The protein plays a significant role in cellular proliferation and ribosomal RNA transcription.[35] It is located on chromosome 10 (10q25).

1.9.4. MUM1

Multiple myeloma oncogene 1 (MUM1) is a nuclear marker that is expressed in activated B and T cells, melanocytes and plasma cells. Its gene rearrangement is expressed at 6p25 (*IRF4 / DUSP22*). MUM1 is said to determine poor survival in DLBCL patients.[36] In a cohort study of DLBCL patients who have undergone R-CHOP, Irawan *et al.* concluded that MUM1 with a higher cut-off value of 50%, would be able to predict the death or treatment failure in 24 months.[37] The study aimed to analyse the prognostic value of the Hans algorithm proteins as well as Ki67.[37]

1.9.5. BCL2

BCL2 gene located at 18q21.33 and on the mitochondrial outer membrane is responsible for encoding the Bcl-2 protein. The gene plays a critical role in promoting cellular survival while suppressing the activity of pro-apoptotic proteins. The gene rearrangement is commonly observed, occurring in 20–30% cases.[38]

Over-expression of the gene is associated with oncogenesis whereas the over-expression of the Bcl-2 protein alone is not associated with malignancy. The over expression of *BCL-2* with other oncogenes such as *MYC* could result in a B-cell malignancy.

1.9.6. BCL6

The *BCL-6* gene, which is located at 3q27, regulates the proliferation of T follicular helper cells and produces the Bcl-6 protein.[39] It is commonly found in germinal centres of both neoplastic and normal B-cells,[40] and plays an important role in *MYC* and *BCL2* suppression in these cells.[39] In several studies, *BCL6* genetic rearrangement is said to be commonly detected on chromosome 3q27 and to be seen in approximately 35% of DLBCL cases.[38,40]

1.9.7. MYC

The *MYC* gene is a transcription factor that regulates the cell cycle, plays a significant role in tumorigenesis. It is involved in apoptosis, cell growth and differentiation.[27] The gene is trans-located in 10% of DLBCL cases.[40] The most common translocation is t(8;14)(q24;q32).[38] *MYC* gene rearrangement is characteristic in double and triple-hit lymphomas, which have an aggressive clinical outcome.[38]

1.10. Chromosomal translocations in DLBCL

In the recent WHO classification, high-grade B-cell lymphomas (HGBL) from a heterogeneous group of genetic alterations (chromosomal translocations) have been included in the classification of mature B-cell neoplasms. These are separated into double and triple-hit lymphomas (DTH-HGBL) according to the gene arrangements associated with *MYC, BCL2* and/or *BCL6*.[39] Fluorescence in situ hybridization (FISH) analysis serves as a reliable and precise method for establishing a definitive diagnosis of DTH-HGBL.[41] In a cohort study conducted by Kim *et al*. on 111 patients, DTH-HGBL was prevalent in eight (7.2%) of the cases.[41]

1.10.1. Double-hit lymphoma (DH-HGBL)

Double-hit lymphomas involve the translocation or rearrangement of the *MYC* oncogene in conjunction with either the *BCL2* or *BCL6* gene.[39] These genetic alterations result in *MYC/BCL2* or *MYC/BCL6* rearrangements, leading to highly aggressive tumour biology and poor clinical outcomes. The concurrent dysregulation

of *MYC*, which promotes cellular proliferation, and either *BCL2* or *BCL6*, which influence apoptosis and germinal centre function, contributes to the rapid disease progression. [39]

1.10.2. Triple-hit lymphoma (TH-HGBL)

Triple-hit lymphomas are very rare and their morphologic, biologic, and cytogenetic properties are said to be similar to those of DLBCL and Burkitt lymphoma (BL).[39]

1.11. Overall survival rate and treatment of DLBCL

Treatment of DLBCL is dependant of the disease type, staging and the subtype. The standard treatment regime constitutes of chemo-immunotherapy with rituximab, cyclophosphamide, doxorubicin, vincristine and prednisone (R-CHOP).[2] GCB DLBCL patients are treated every 21 days with 6 cycles of rituximab, prednisone (R-CHOP), vincristine, cyclophosphamide and doxorubicin. ABC DLBCL patients and those with double-hit and triple-hit lymphomas present with poor outcomes when treated with R-CHOP, requiring intensified treatment regiments.[42]

In approximately 50% of the cases DLBCL has an overall survival rate of 5 years (1 826 days) after diagnosis.[43] Patients with low-risk DLBCL showed 3-year progression-free survival (PFS) of 96% after treatment with four cycles of R-CHOP.[44]

PART B: LITERITURE REVIEW OF *Tumour Protein 53* in the development of DLBCL

1.12. *TP53* definitions

TP53 is described as a tumour suppressor gene encoding for the synthesis of protein 53 (p53). p53 protein also known as Tumour protein P53, cellular tumour antigen p53, or transformation-related protein 53 (TRP53) is defined as a protein found within the nucleus of cells and it regulates cell division, apoptosis, tumour suppression and DNA repair.[8,15]

Several studies have documented that the incidence of p53 mutations varies from 5–25% in lymphoid malignancies. Majority of the mutations (~90%) are detected within the DNA-binding domain of the protein.[8],[38],[45]

1.13. *Tumour protein 53* (*TP53*) gene

Tumour protein 53 (*TP53*) gene which is referred to as the guardian of the genome was discovered in 1979 by Crawford *et al*.[15,46] It is the most frequently mutated gene in more than 50% of tumours. Mutations may include gain-of-function (GOS) or loss-of-function (LOS). Gain-of-function is when cellular responses that are affected by WT *TP53* are activated when mutant *TP53* interacts with transcriptional regulators. Loss-of-function results in an aneuploidy phenotype due to genomic instability.[15] In human cancers *TP53* is the most commonly mutated gene, with the majority of these cancers being loss of function mutations.[8]

1.14. *TP53* Structural features

The *TP53* gene, is characterized as a multi-domain gene, is situated on the short arm of chromosome 17 at region 17p13.1 (Figure 1.2). It spans 19,144 base pairs (bp), encodes a protein composed of 393 amino acids, and includes 11 exons.[15,43] *TP53* gene encodes several proteins resulting in p53 isoforms. It plays an important physiological role in immune and inflammatory responses.[8] A mutated or non-functional copy of the *TP53* gene can be inherited from a parent and predisposes an individual to the development of tumours early in life. This rare condition is known as the Li-Fraumeni syndrome.[47]

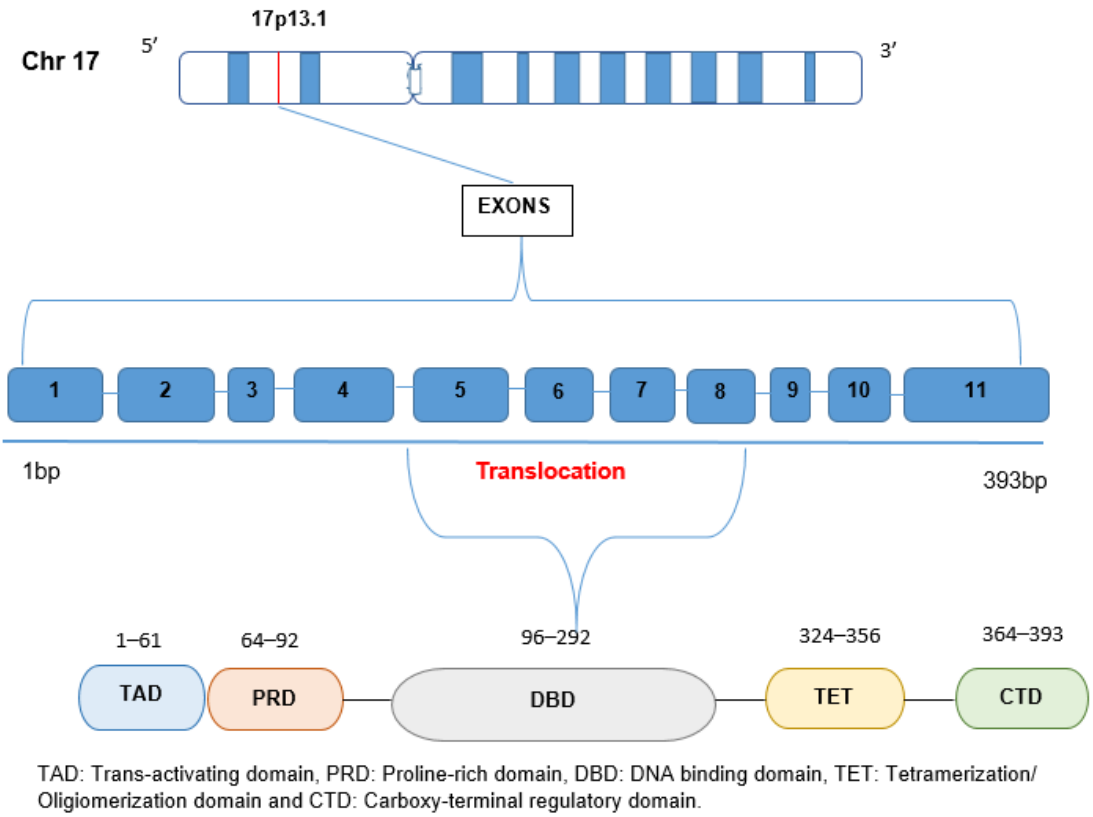


Figure 1.2: The location of the *TP53* gene on chromosome 17, depicting the different domains of the p53 protein and the exons in which they reside (adapted from de Andrade *et al*. 2017)

This gene is commonly divided into domains which work together to ensure that p53 can effectively respond to cellular stress by halting cell division, repairing DNA, or initiating cell death if the damage is irreparable.[48,49] There are five domains (Figure 1.2) each of which functions differently. The RNA polymerase trans-activating domain, located at the N-terminus (TAD, residues at 1–61) is responsible for activating the transcription of target genes involved in cell cycle arrest and apoptosis. Proline-rich domain (PRD, residues 64–92) promotes p53 mediated apoptosis of tumour cells by cooperation with anti-neoplastic agents. DNA binding domain, also known as the central domain (DBD, residues 96–292) allows p53 to bind to specific DNA sequences allowing the regulation of expression of genes that control cell growth and apoptosis. Tetramerization / Oligiomerization domain (TET, residues 324–356) which plays an important role in the formation of active forms of the p53 protein, the tetramers. The Carboxy-terminal regulatory domain (CTD, residues 364–

393) assist in regulating the DNA binding activity of p53 and is involved in post-translational modifications that affect p53's stability and function.[49,50]

1.15. p53 Pathway

Under normal conditions p53 is expressed in very low levels and maintains homeostasis, genomic stability and prevents the development of cancer, in a normal cell by cell cycle arrest, DNA repair and apoptosis. RING-finger type E3 ubiquitin ligase protein, mouse double-minute-2-homolog (MDM2), mediates proteasomal degradation which maintains low levels of p53. MDM2 also acts by transporting p53 from the nucleus to the cytoplasm of the cell.[7]

When a stimulus such as DNA damage, osmotic shock, deregulated oncogene expression, oxidative stress and certain infections are introduced, p53 induces cell cycle arrest in order to either repair and recirculate the surviving cell or induce apoptosis by killing and eliminating the damaged cell.[7]

1.16. Frequency of p53 protein expression and *TP53* gene rearrangements

Genetic mutations affecting tumour suppressor genes and proto-oncogenes may result in B cell lymphomas. Akay *et al.* concluded that the expression of oncogenic proteins, including p53, may be beneficial as prognostic markers in DLBCL cases.[38] *TP53* gene rearrangements were detected in 22.7% of cases and p53 protein was expressed in 69.04% of DLBCL.[38] Young *et al*. contributed that *TP53* mutations were detected in 102 of 477(21.4%) patients.[51] GCB-DLBCL may predict poor overall survival in the presence of *TP53* mutations, while patients with wild-type (WT) *TP53* had a better overall survival rate (p value <0.001).[51] In patients with different types of tumours, the p53 protein was identified to be highly expressed in tumour and transformed cells and very low levels are expressed in non-transformed cells. The protein also possesses the ability to bind to viral oncoproteins from DNA tumour viruses.[8] Therefore we are lead to question whether the detection or expression of p53 protein and or *TP53* gene rearrangements play a major role as a prognostic marker in DLBCL cases?

1.17. Mutant p53

p53 is known to bind to MDM2, which is induced by p53 protein, at increased levels, the mutant p53 protein does not induce MDM2. Mutant *TP53* results in genomic instability, cell cycle dysregulation and loss of control of proliferation.[7] Mutant p53

protein, especially those with missense mutations, may disrupt normal functioning of p53 such as stability and function.[7]

1.18. Wild-type p53 (association of p53 and *TP53* to *BLC2*)

Wild type (WT) p53 functions as a tumour suppressor gene by suppressing transformation by mutant p53 and *ras*.[8] WT *TP53* comprises folded and unstructured regions that function in a synergistic manner.[51] *TP53* associated with overexpression of *BCL2* at t(14;18)(q32;q21) in DLBCL is considered WT.[7] WT *TP53,* has a better clinical outcome in the apoptosis mode whereas in the autophagy mode it has a worse clinical outcome as it promotes survival of lymphoma cells.[52] In a study conducted by Barrans *et al*. comprising 137 cases, 79 (58%) tested negative for Bcl-2 protein and had a negative t(14;18) FISH test, while 18 (13%) cases tested positive for both the Bcl-2 protein and *BCL2* gene rearrangement. There was a 70% correlation of *BCL2* by FISH and IHC analysis.[53]

Mutant p53 contains dominant negative mutations which act by blocking the activity of WT p53 as an action of transformation by inhibitory proteins.[52]

1.19. Cross talk between p53 and c-myc in DLBCL

c-myc protein is a transcription factor that controls cell cycle and cell growth processes and it may be expressed in 30-50% of DLBCL cases. It is encoded by the proto-oncogene, *MYC*, rearrangement of which is linked to poor overall survival of DLBCL patients who receive the R-CHOP treatment regime.[54]

A high clinical stage may correlate with the expression of c-myc and p53 protein in 58% of patients.[38] In contrast, co-expression of the two markers was absent in patients with low clinical stage. [38] The co-expression of c-myc and p53 is associated with a high-risk of relapse and poor overall survival.[38] In a study conducted by Yu *et al.* it was suggested that the cross-talk pertaining to c-myc and p53 needs to be further studied. Blockage of several pathways which induce apoptosis, result in good therapeutic effects and has been noted in patients with minimal toxicity.[54] Similarly, Bouroumeau *et al.* suggested that MYC, p53, FOXO1, B2M, TNFRSF14, and PIM1 create a pro-apoptotic environment, a tumour suppressor pathway which may serve as an ideal therapeutic target.[55] In the aforementioned study the overexpression of p53 and c-myc with a morphological pattern referred to as starry-sky pattern showed an enhanced survival rate.[55]

1.20. *TP53* gene arrangement and mutations

All the modular regions of the *TP53* gene were originally identified when the central region of p53 protein containing the four conserved regions and the major mutation hot spots (residues 102 ± 292) was recognised as the specific DNA-binding region.[8] The coding sequence gene shows non-coding on exon 1 and contains high degree coding on exons 5, 6, 7 and 8 (DBD).[8]

Mutations in *TP53* results in divergent isoforms of the p53 protein, with the p53α never being expressed on its own. To date twelve p53 isoforms have been identified (p53α, p53β, p53γ, Δ40p53α, Δ40p53β, Δ40p53γ, Δ133p53α, Δ133p53β, Δ133p53γ, Δ160p53α, Δ160p53β, Δ160p53γ).[56] Isoforms play a fundamental role in the regulation of the p53 pathway but their expression is deregulated in cancer.[56]

Voropaeva *et al*. found that *TP53* mutation status is of prognostic value. In their study, the most prevalent mutations accounting for 95% of the cases were detected in the DBD, with codons 275, 155, 272 and 212 as hotspots in the *TP53* gene. Missense mutations accounted for 33 (37%) of the mutations detected.[57] The IARC *TP53* database and the *TP53* UMD mutation database were utilised to assess the detected mutations .[58,59] Majority of the mutations detected were not listed in the IARC *TP53* mutation database with the exception of codon 244.[57] Several mutations, including p.L130F, p.T155I, p.R196Q, p.G244S, p.V272E, and p.A276V, were classified as possessing a high or moderate risk. Additionally, intron mutations such as IVS6-36G>C and IVS5+43G>T were identified and documented as functionally significant.[58]

1.21. Correlation between *TP53* mutational status and p53 overexpression

Rodrigues *et al.* investigated the correlation between p53 immunohistochemical (IHC) staining and missense mutations in the *TP53* gene in mantle cell lymphoma, achieving 100% specificity, 82% accuracy, and 82% sensitivity in their analysis.[60] This study demonstrated that missense mutations occurring within the hotspot regions of the DBD can be reliably detected using p53 IHC staining as a surrogate marker.[60] Conversely, Peroja *et al.* concluded that p53 protein expression could serve as a screening tool for *TP53* mutations. However, it should not be utilised as a replacement for sequencing.[61] There is a paucity of published literature describing the relationship between p53 immunoexpression and *TP53* genetic mutations in DLBCL.

1.22. Overall survival rate for *TP53* DLBCL mutated cases

A poor overall survival rate was documented in DLBCL patients with *TP53* mutations.[51] Conversely, a study by Tamimi *et al.* reported a worse overall survival rate among patients without *TP53* mutations compared to those harbouring such mutations.[62] Ichikawa *et al*. documented a poor overall survival rate in 22 patients with *TP53* mutations compared to 80 patients without mutations.[63]

The p53 protein prevents neoplastic transformation by senescence, apoptosis and quiescence.[64] Limited published data describe the translocation and mutation of *TP53* and immunohistochemical expression of p53 in DLBCL. Akay *et al.* concluded that the expression of oncogenic proteins, including p53, may be beneficial as prognostic markers in DLBCL cases.[38] Mutations in the DBD of the *TP53* gene hold prognostic significance for survival outcomes in DLBCL, with point mutations in *TP53* being linked to an unfavourable prognosis.[52,64] Exploring *TP53* mutations in patients with DLBCL may prove to have prognostic value as treatment with certain drugs can affect the transcription activity of the gene and possibly improve clinical outcome.[52,64]

1.23. Treatment of DLBCL associated with TP53 mutations

In a treatment trial of DLBCL patients with mutant *TP53*, treatment with CHOP or R-CHOP was unfavourable and patients had a diminished survival outcome.[64] The trial (RICOVER-60) constituted eight cycles of CHOP or R-CHOP.[64] In a study conducted by Qin *et al.* on DLBCL patients with mutant *TP53,* it was concluded that incorporating a loco-regional procedure, such as surgery or radiation therapy, into the treatment regimen resulted in improved clinical outcomes for the patients.[64] Overall survival and progression-free survival showed significant improvement, this was supported by p-values less than 0.05.[64]

There is a scarcity of published data on *TP53* translocations, *TP53* mutations, and p53 immunohistochemical expression in DLBCL, with no available data specific to the Southern African region. This study aimed to characterise p53 expression in DLBCL cases within the context of HIV sero-prevalence and to identify associated *TP53* gene mutations, including their specific locations within the DBD. Additionally, survival outcomes were compared across different COO DLBCL subtypes.

Aims and Objectives

1.24. Study aims

- The study aimed to determine the p53 protein expression and *TP53* gene mutations associated with DLBCL at Chris Hani Baragwanath Academic Hospital (CHBAH), among HIV sero-prevalent patients.

1.25. Study objectives

- To document the histopathologic features of DLBCL at CHBAH.
- To determine and correlate p53 IHC and *TP53* mutational profiles with the histopathologic features of DLBCL cases.

2. CHAPTER TWO: METHODS

2.1. Study design

A retrospective cross-sectional study was conducted.

2.2. Ethical Considerations

This study was granted approval from the Human Research Ethics Committee (HREC) at the University of the Witwatersrand (M220385) (Appendix A).

2.3. Sample size calculation

$$\frac{Z^2 p\,(1-p)}{d^2}$$

Z = is standard normal variate (at 5% type 1 error ($p<0.05$) it is 1.96 and at 1% type 1 error ($p<0.01$) it is 2.58). As in majority of studies P values are considered significant below 0.05 hence 1.96 is used in formula.

p = Expected proportion in population based on previous studies (19% TP53 mutations in DLBCL).

d = Absolute error or precision – Decided by researcher (10% precision error).

Using the standard error of 5%, expected proportion of 19% and a precision error of 10%, our required minimum sample size was calculated to be 60 cases.

2.4. Sample selection

Selected cases were extracted from the database of a previous DLBCL study that was conducted at CHBAH (ethics clearance certificates M151017 and M2010122). The previous study conducted a retrospective Systematized Nomenclature of Medicine (SNOMED) search by the Central Data Warehouse (CDW), using the laboratory TrakCare application to detect confirmed cases of DLBCL during the abovementioned five-year time frame. Cases of DLBCL were identified using the following SNOMED codes: M-96803, M-80003, and M-95903. All cases of histopathologically confirmed of DLBCL over a five-year timeframe, from 01 October 2013 to 31 June 2017 were selected, except when the patient was a minor, younger

than 18 years. Cases with c-myc IHC results and survival outcome were prioritised for selection.

The study was conducted using formalin-fixed paraffin-embedded (FFPE) tissue blocks and slides, from the Department of Anatomical Pathology at the CHBAH, University of the Witwatersrand and National Health Laboratory Service, Soweto, South Africa.

Seventy cases were available. Forty-two cases with clinical outcomes were selected and 18 cases were selected at random from the previously researched DLBCL database. One case was eliminated as it belonged to a minor patient and nine other cases were removed from the study due to inconsistencies with data and limited tumour in the tissue block.

2.5. Data collection from pathology reports

Cases were assigned a unique study number so as to preserve patient anonymity. The information was obtained from the histopathology reports and documented. These included age at the time of histopathological diagnosis (in years), sex, topographic site of involvement and HIV status. All data was recorded in the data collection spreadsheet (Appendix B) and summarised in Appendix C. Missing variables such as HIV status were expressed as unknown. All topographic sites which were not nodal are expressed as extra nodal in this report.

Survival outcome data of the patients was extracted from the previously conducted DLBCL study at CHBAH (ethics clearance certificates M151017 and M2010122).

2.5.1. Data analysis

Descriptive Statistics

All the numerical variables such as age were not normally distributed and therefore the interquartile range (IQR) was used to describe the data. The Mann-Whitney (or Wilcoxon-Mann-Whitney) test was then adapted for comparisons. The categorical data were described as frequencies and proportions. Data analysis was performed using the STATA (version 18) software.

2.6. Immunohistochemistry on DLBCL

Immunohistochemistry was performed to detect the expression of various proteins on DLBCL.

The following IHC panel: CD10, bcl-6, MUM1, Bcl-2, c-myc and Ki67 (Roche Diagnostics, USA); was performed using the Ventana Benchmark XT instrument (Roche Diagnostics, USA) (Table 2.1). IHC staining for p53 was conducted utilizing the 48 Equipment systems (Dako, Glostrup, Denmark) for antigen retrieval, followed by automated staining with the DAKO automated instrument (PT Link) (Dako, Glostrup, Denmark), see section on validation of p53.

The tissue blocks were sectioned at 4 µm and picked up on charged slides with a known positive control (Table 2.2). Sections were then incubated at 60°C for 45 minutes in order to melt wax and assist the sections to adhere to slide, thereby, preventing detachment or washing off of sections. The sections on the slide underwent deparaffinization using the EZ prep solution (Dako, Glostrup, Denmark). The slides were pre-treated with Heat Induce Epitope Retrieval (HIER) method using the Envision^{TM} FLEX Target Retrieval Solution, High pH 9 (50x) (Dako, Glostrup, Denmark) for 20 minutes at 97°C for antigen retrieval. In between each step a wash buffer (20x) was used to clear the sections of any excess reagents. Blocking of endogenous peroxide activity was performed using hydrogen peroxidase for 5 minutes. The slides were treated with the primary antibody as per Table 2.1, antibody incubation conditions, and envision for 20 minutes, this step assists biotin from disturbances and increased the dilution of primary antibodies. Ready to use Envision^{TM} FLEX DAB (3,3' Diaminobenzidine) (Dako, Glostrup, Denmark) solution was used for 10 minutes to visualize the antigen-antibody reaction and Mayer's haematoxylin (Clinical Sciences Diagnotics, Aspen Hills, SA) was used to counterstains the sections for 3 minutes, the sections were placed under running tap water to remove excess stain until the water ran clear and the stained sections had a blue colour. The sections were treated with graded alcohols to remove excess water, xylene to remove alcohol and create a base compatible with Entellan (Merck, Germany) for slide mounting.

Validation of the p53 primary antibody was done using 50 in 1000 µl, 100 in 1000 µl and 150 in 1000 µl of p53 antibody into a diluent. A colon tissue was used as a positive control. Test slides were stained using the 150 in 1000 µl dilution for the antibody, following the above stated method. For quality control purposes; a known positive control was placed on each test slide.

Table 2.1: Characteristics of antibodies used for IHC and incubation conditions

Primary Antibody	Species	clone	Reference code	Antibody Incubation conditions
Bcl-2	124	mouse monoclonal	790-4464	16 minutes, 37°C
Bcl-6	GI191E/A8	mouse monoclonal	760-4241	18 minutes, 37°C
CD10	SP67	rabbit monoclonal	790-4506	16 minutes, 37°C
MUM-1	EP190	rabbit monoclonal	760-6082	18 minutes, 37°C
Ki67	30-9	rabbit monoclonal	790-4286	18 minutes, 37°C
c-myc	Y69	rabbit monoclonal	790-4628	16 minutes, 37°C
p53	BP53-12	mouse monoclonal	300215009	20 minutes, 25°C

In order to classify the two different subtypes using the Han's algorithm CD10, Bcl-6, and MUM1 IHC stains are routinely applied to classify the subgroups.[25] This was based on immunoexpression (Table 2.2) of at least 30% positive tumour cells.[25]

Table 2.2: Criteria of positivity of IHC markers

IHC Stains	Positive	Known positive Controls used	Component/s		
			Cytoplasmic	Membranous	Nuclear
Bcl-2	≥50%[#]	Lymph node	✓		
Bcl-6	≥30%*	Lymph node			✓
CD10	≥30%*	Lymph node	✓	✓	
MUM-1	≥30%*	Lymph node			✓
Ki67	≥90%[†]	Lymph node			✓
c-myc	≥40%[$]	Lymph node			✓
p53	≥50%[#]	Bowel			✓

*CD10, Bcl-6 and MUM-1 reference ranges were adapted from Hans *et al.* (2004)

[†]Ki67 reference ranges were adapted from n Huber *et al.* (2021)

[$]c-myc reference range were adapted from Pather *et al.* (2022)

[#]Bcl-2 and p53 reference range were adapted Peroja *et al.* (2018)

2.7. Polymerase chain reaction (PCR)

The polymerase chain reaction was performed to determine which samples contained mutations and to identify the specific exon(s) whereby the mutations were located.

2.7.1. Sectioning of paraffin embedded formalin fixed tissue blocks

DNA was extracted from formalin-fixed paraffin-embedded (FFPE) tissue blocks. Six tissue sections of 10 µm in thickness were cut and placed in a 1.5 ml micro-centrifuge tube. For quality purposes two control measures were undertaken;

Between each new case the microtome and forceps were cleaned with xylene, 70% alcohol and 10% Bleach solution. The second, a negative control of pure paraffin was cut after the cutting procedure.

2.7.2. DNA extraction

To remove wax from the tissue sections, coils were vigorously vortexed in two changes of xylene (1 ml), and the supernatant was carefully removed leaving only residual tissue fragments. To remove excess xylene, the supernatant was washed using two changes of 1 ml absolute alcohol. The tissue fragments were heat treated at 37°C for 20 minutes until excess ethanol had evaporated, DNA was then extracted using the QIAGEN QIAamp DNA kit (QIAGEN, SA) (Pty). For cell lysis, the tissue

fragments were suspended in 180 µl of Buffer ATL and to denature proteins, 20 µl of QIAGEN Proteinase K was added. The microcentrifuge tubes were briefly mixed and incubated at 56°C on a rocking platform overnight.

In order to remove droplets from the inside of the lid, centrifuge micro-centrifuge tubes at 3000 rpm for 15 seconds. 200 µl of Buffer AL was added to completely lyse tissue, mixed briefly and incubated at 70°C for 10 minutes. Tubes were centrifuged at 3000 rpm for 15 seconds to remove droplets from inside the lids.

On addition of AL buffer, a white precipitate may form. It will dissolve during the incubation. The tissue samples that were not completely lysed were centrifuged and the supernatant was transferred to a new micro-centrifuge tube, leaving behind the sediment.

Ethanol of 200 µl is added into the sample, mixed for 15 seconds and centrifuged at 3000 rpm for 15 seconds. Both the supernatant and precipitate were transferred into QIAamp spin columns placed in 2 ml collection tubes. Tubes were centrifuged at 8 000 rpm for 1 minute. The QIAamp spin column was placed into a new 2 ml collection tube and the tube containing the filtrate was discarded. To the QIAamp spin column, 500 µl of Buffer AW1 was added and tubes were centrifuged at 8 000 rpm for 1 minute to wash the DNA. The QIAamp spin column was placed into a new 2 ml collection tube and the tube containing the filtrate was discarded. To the QIAamp spin column, 500 µl of Buffer AW2 was added for a second wash and tubes were centrifuged at 8 000 rpm for 1 minute. The QIAamp spin column was placed into a new 2 ml collection tube and the tube containing the filtrate was discarded. In order to completely dry the membranes, the tubes are centrifuged at 8 000 rpm for 1 minute and 13 000 rpm for 3 minutes and the collection tubes containing the filtrate are discarded. To elute DNA, 100 µl of Buffer AE was added into the QIAamp spin columns without wetting the rim, and they were incubated at room temperature for 1 minute and centrifuged at 8 000 rpm for 1 minute. The filtrate was transferred into a clean 1.5 ml micro-centrifuge tube and DNA was quantified using the NanoDrop One Spectrophotometer (Thermo Fischer Scientific, United States) and recorded on Appendix D. DNA was then stored at 4°C, in order to maintain its stability.

2.7.3. *TP53* PCR Primers for Amplification

The forward (F) and reverse (R) primers were prepared as per specification sheet to create a stock of 100 µM. They were then diluted to a working concentration of 10

µM. The primers used were as per Peroja *et al*. and Appendix E.[61] Primer pairs used for *TP53* exons 5, 6, 7 and 8 were as per Table 2.3. The primers were sourced from Integrated DNA Technologies, USA. In order to adequately analyse fragmented DNA, two sets of primers were utilized for exon 5 and 8, thereby, reducing the PCR product sizes to less than 200 bp.[61]

Table 2.3: Primers used for PCR and sequencing the *TP53* DBD exons 5 to 8

EXON	PRIMER	FORWARD (F) SEQUENCE 5' - 3'	REVERSE (R) SEQUENCE 5' – 3'	AMPLICON SIZE (bp)
5a	TP53 EXON 5A	CCTGACTTTCAACTCTGTCTC	ACTGCTTGTAGATGGCCATG	158
5b	TP53 EXON 5B	CAGCTGTGGGTTGATTCCAC	CTGGGGACCCTGGGCAAC	182
6	TP53 EXON 6	GCCTCTGATTCCTCACTGAT	TTAACCCCTCCTCCCAGAGA	181
7	TP53 EXON 7	AGGCGCACTGGCCTCATCTT	TGTGCAGGGTGGCAAGTGGC	177
8a	TP53 EXON 8A	CCTTACTGCCTCTTGCTTCTC	CTTGCGGAGATTCTCTTCCTC	130
8b	TP53 EXON 8B	TTGTGCCTGTCCTGGGAGAG	CTCCACCGCTTCTTGTCCT	127

2.7.4. PCR run conditions

The extracted DNA (100 ng) was used for *TP53* PCR amplification of exons 5-8. For amplification a master mix of 50 µl was created using the AmpliTaq Gold® (Applied Biosystems) (Thermo Fischer Scientific, USA) with 10X Buffer II kit (Catalogue No. N808-0241). Both forward and reserve primers were diluted to a concentration of 10 µM.

The PCR run consisted of 40 cycles, the conditions used constituted: An initial denaturing at 95°C for 10 minutes, followed by denaturing at 94°C for 55 seconds; in order to break the double-stranded DNA (dsDNA) into a single-stranded DNA (ssDNA) by dissociating the hydrogen bonds between the complementary bases. Annealing was performed at 55°C for 55 seconds for exon 5A, 5B, 6, 8A, 8B and at 60°C for 55 seconds for exon 7. This is a process where ssDNA bind to their complementary sites of the primers beginning from the 3' end of the DNA template. Elongation, a step whereby a dsDNA is created when nucleotide triphosphates (dNTPs) bind to the ends of primers in the 5' to 3' end direction at 72°C for 55 seconds. An extended final elongation time of 7 minutes at 72 °C is performed after the 40 cycles. The PCR cycling was performed using an Applied Biosystems ProFlex PCR System thermocycler (Applied Biosystems) (Thermo Fischer Scientific, USA).

2.7.5. Agarose Gel electrophoresis

PCR products were visualised using a 2% agarose gel (Appendix F).

To visualize the DNA product, 15 µl of the product was mixed with 5 µl of gel loading dye. A 50 bp DNA ladder was included to assist with band sizes. The gel was run at 100 V for 60 minutes. Post electrophoresis, the gel was visualised utilizing the G: Box Vacutec Gel Documentation system with GeneSnap v.6.08 software (Syngene, UK).

2.7.6. PCR Product Clean Up

PCR products from samples with representative bands were cleaned using the Exo/SAP protocol prior to sequencing. This procedure was used to remove excess materials, such as primers and master-mix reagents, from the PCR products.

The Exo/SAP master mix was prepared by adding 10 U Exonuclease I (Catalogue No. NEB M0293L) (Fischer Scientific, USA) and 2 U Shrimp Alkaline Phosphatase (Catalogue No. NEB M0371) (Fischer Scientific, USA) to a 0.6 ml micro-centrifuge tube. The final reaction mixture constituted of 10 µl of amplified PCR product and 2.5 µl of the Exo/SAP mixture. The reaction mixture was well mixed and incubated for 15 minutes at 37°C. In order to stop the reaction, high heat (80°C) was applied to the mixture for 15 minutes.

2.7.7. *TP53* sequencing

Fragments were sequenced using the Nimagen, BrilliantDye™ Terminator Cycle Sequencing Kit V3.1, BRD3-100/1000 (NimaGen, Netherlands). The labelled products were then cleaned with the ZR-96 DNA Sequencing Clean-up Kit (Catalogue No. D4053): 240 µl of Sequencing Binding Buffer was added to 15 µl of the sequencing reaction. The mixture was transferred into a Zymo-Spin™ IB-96 plate (Zymo Research Corporation, USA) mounted onto a collection plate. The mixture was centrifuged for 2 minutes at ≥ 3 000 x g. To each well, 300 µl of sequencing wash buffer was added and the mixture was centrifuged for 5 minutes at ≥ 3 000 x g. To the column matrix on the filter plate, 15 µl of water was added. The Zymo-Spin™ IB-96 plate was mounted on top of the 96-Well PCR plate and these were then mounted onto the collection plate, and centrifuged at ≥ 3 000 x g for 2 minutes to elute ultra-pure DNA.

The purified ultra-pure DNA was loaded onto an Applied Biosystems ABI 3730XL Genetic Analyser (Thermo Fischer Scientific, USA) with a 50cm capillary array and POP7 polymer for electrophoretic separation.

2.7.8. *TP53* Sequencing data analysis and reporting of mutated variants

DNA sequence chromatographs were viewed using Finch TV 1.4.0. software. The forward and reverse chromatographs were inspected using the Basic Local Alignment Search Tool (BLAST) on NCBI database (https://blast.ncbi.nlm.nih.gov/) and the resulting complementary regions were interpreted using the genetic codon wheel and the codon chart, see Appendix G. The resulting mutations were queried and interpreted using the IRAC *TP53* database (https://tp53.cancer.gov/) and the UMD *TP53* database (https://p53.fr/download-the-database).

2.7.9. Inferential Statistics

Data analysis was performed using the STATA 18 software. The Mann-Whitney (or Wilcoxon-Mann-Whitney) test was used to estimate the associations between age and staining. The Fisher Exact and the Chi-Square tests were used to estimate the association between categorical variables. The Cramer's V value was used to estimate the strength of the association between the categorical variables. Weak associations were estimated at <0.10, and a moderate at 0.10 and 0.25. While strong association was above 0.25. The relationship that generated probabilities (p-values) of less than 0.05 ($p<0.05$) was considered a significant association while the relationship with greater than 0.05 ($p>0.05$) were interpreted as non-significant. The data was presented in dummy Tables format showing the interquartile range (IQR), the proportions and frequencies, the p-values and the Cramer's V outcomes.

3. CHAPTER THREE: RESULTS

The results represented focus on an in-depth examination of the immunophenotypic characteristics and mutational profiles of *TP53* in DLBCL. This study aims to clarify the involvement of *TP53* mutations in the molecular mechanisms driving DLBCL pathogenesis and its prognostic significance, thereby contributing to a better understanding of the clinical implications and identifying potential avenues for therapeutic intervention.

3.1. Cohort description

The study consisted of 60 known DLBCL patients, the demographic characteristics are described in Figures 3.1 and Table 3.1. The age range of the patients was from 22 to 88 years with a median (IQR) age of 45 (38-54) years.

The male patients predominated (n=33; 55.0%), and the females accounted for 45.0% (n=27). The distribution of sex had a male to female ratio of 1.6:1.

Other factors evaluated included human immunodeficiency virus (HIV), nodality, cell-of-origin (COO), and other immunohistochemistry markers such as Bcl-2 and Ki67.

Table 3.1: Demographic characteristics of our study cohort

Variables				
AGE				
Range	22-88 years			
Median (IQR)	45 (38-54) years			
	Total			
	n	**%**		
SEX				
Female	27	45.0		
Male	33	55.0		
Total	**60**	**100.0**		
HIV				
Negative	12	20.0		
Positive	45	75.0		
Unknown	3	5.0		
Total	**60**	**100.0**		
SITE				
Nodal	29	48.3		
Extra Nodal	31	51.7		
Total	**60**	**100.0**		
COO PROFILE				
ABC	26	43.3		
GCB	28	46.7		
NOS	6	10.0		
Total	**60**	**100.0**		
SURVIVAL OUTCOME DATA				
Alive	16	26.7		
Demised	26	43.3		
Unknown	18	30.0		
Total	**60**	**100.0**		

	IMMUNOHISTOCHEMISTRY		FISH	
	n	**%**	**n**	**%**
Ki67 (≥90%)				
<90%	9	15.0		
≥90%	50	83.3		
Unavailable	1	1.7		
Total	**60**	**100.0**		
Bcl-2 (≥50%)			***BCL2***	
<50% (FISH negative)	32	53.3	3	33.3
≥50%	26	43.3	-	-
Unavailable	2	3.3	6	66.7
Total	**60**	**100.0**	**9**	**100.0**
c-myc (≥40%)			***MYC***	
<40% (FISH negative)	23	38.3	9	100.0
≥40%	33	55.0	-	-
Unavailable	4	6.7	-	-
Total	**60**	**100.0**	**9**	**100.0**
Bcl-6 (≥30%)			***BCL6***	
<30% (FISH negative)	17	28.3	3	33.3
≥30%	40	66.7	-	-
Unavailable	3	5.0	6	66.7
Total	**60**	**100.0**	**9**	**100.0**
CD10 (≥30%)				
<30%	31	51.7		
≥30%	27	45.0		
Unavailable	2	3.3		
Total	**60**	**100.0**		
MUM1 (≥30%)				
<30%	19	31.7		
≥30%	38	63.3		
Unavailable	3	5.0		
Total	**60**	**100.0**		
p53 (≥50%)				
<50%	33	55.0		
≥50%	27	45.0		
Unavailable	-	-		
Total	**60**	**100.0**		

In this study a sero-positive HIV status was documented in 45 (75.0%) patients and 12 (20.0%) tested negative. The status was unknown for three of the patients, 5.0% (Table 3.1). Amongst the HIV positive patients 23 (51.1%) were male and 22 (48.9%) were female. These patients had an age range of 22-65 years, with a median (IQR) age of 42 years. Association of HIV was assessed against several variables [sex (p=0.14), nodality (p=0.56) and COO (p=0.01)] and a significant association was demonstrated with COO. Of the different immunohistochemical markers used to determine the COO association was not demonstrated on MUM1, [CD10 (p=0.02), Bcl-6 (p=0.01) and MUM1 (p=0.19)].

The patients were confirmed to have DLBCL within 59 topographic sites. Nodal disease includes LNs, spleen, thymus or Waldeyer ring.[65] Nodal disease was present in 29 (48.3%) of cases. The extranodal sites comprised 31 (51.7%), as shown in Figure 3.1. No significant association of nodality and p53 immunoexpression was found (p-value =0.29).

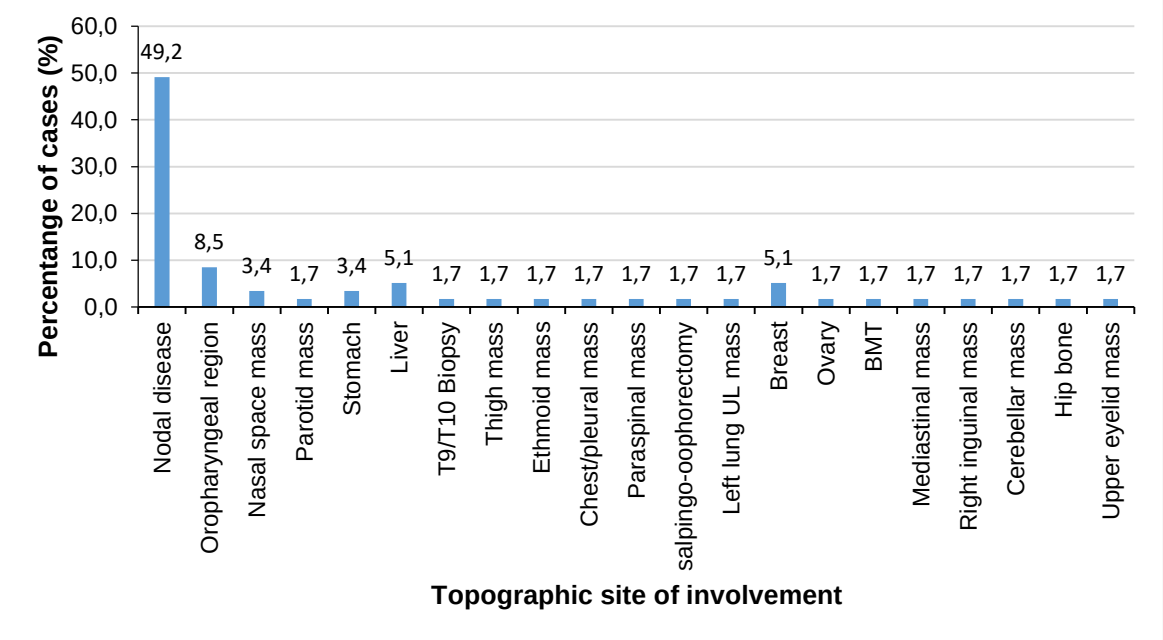


Figure 3.1: Different topographic sites of involvement associated with our study cohort

The IHC profile for the COO classification of DLBCL is summarised in Table 3.1. There were 54 DLBCL cases (90.0%) with successful classification of COO. The GCB subtype comprised 28 cases (46.7%), while 26 cases (43.3%) were classified as the ABC subtype. The age distribution of patients with the GCB subtype ranged between 22 and 76 years, with a median age of 45.5 years (IQR). Patients with the

ABC subtype exhibited an age range of 32 to 88 years, with a median age of 44 years (IQR). Among the ABC subtype cohort, eight patients (30.8%) were alive while 12 (46.2%) had demised. The survival duration among ABC patients who survived ranged from 28 to 804 days, whereas for those who demised, the duration to mortality varied between 11 and 825 days. The study cohort with the GCB subtype constituted eight patients (28.6%) who survived and 12 (42.9%) who had demised. The survival duration among GCB subtype patients who survived ranged from 14 to 1 819 days, whereas for those who demised, the duration to mortality varied between 11 and 849 days. Overall, the duration of survival for patients ranged from 14 to 1,819 days, while the duration to mortality for deceased patients varied between 11 and 849 days.

Sections 3.1.1. and 3.1.2. depict figures of the immunoexpression of cases of ABC and GCB subtypes, respectively.

3.1.1. DLBCL case of ABC:

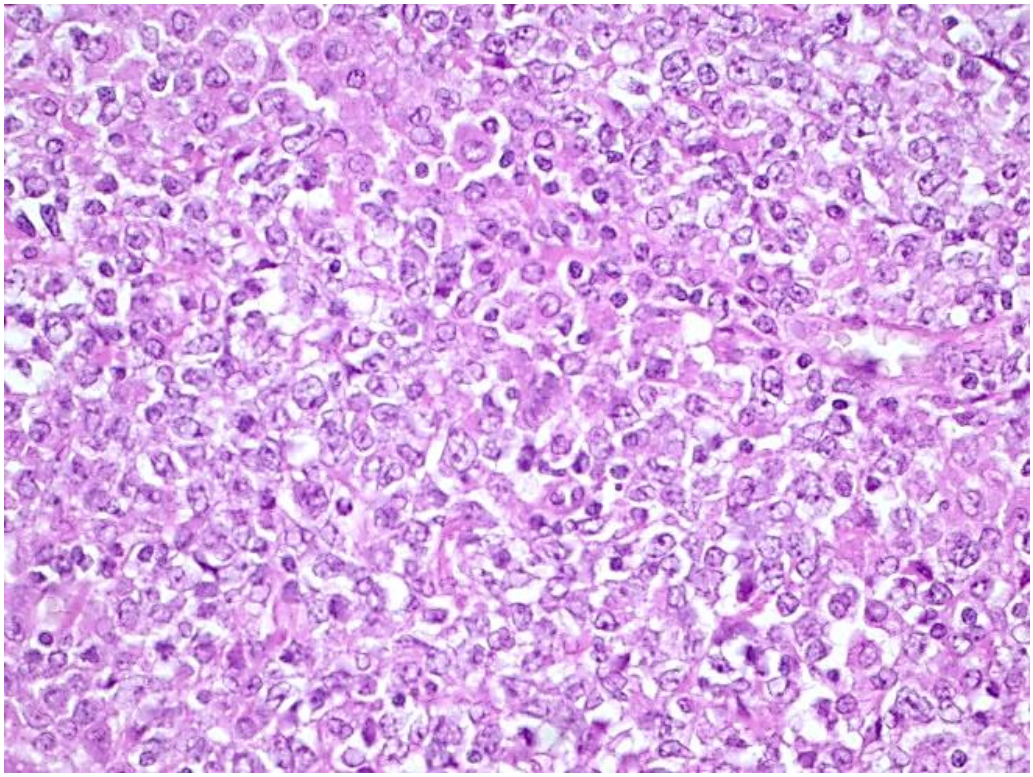


Figure 3.2: Photomicrograph of a lymph node with DLBCL, from a 53-year–old female (H&E staining, X400 magnification)

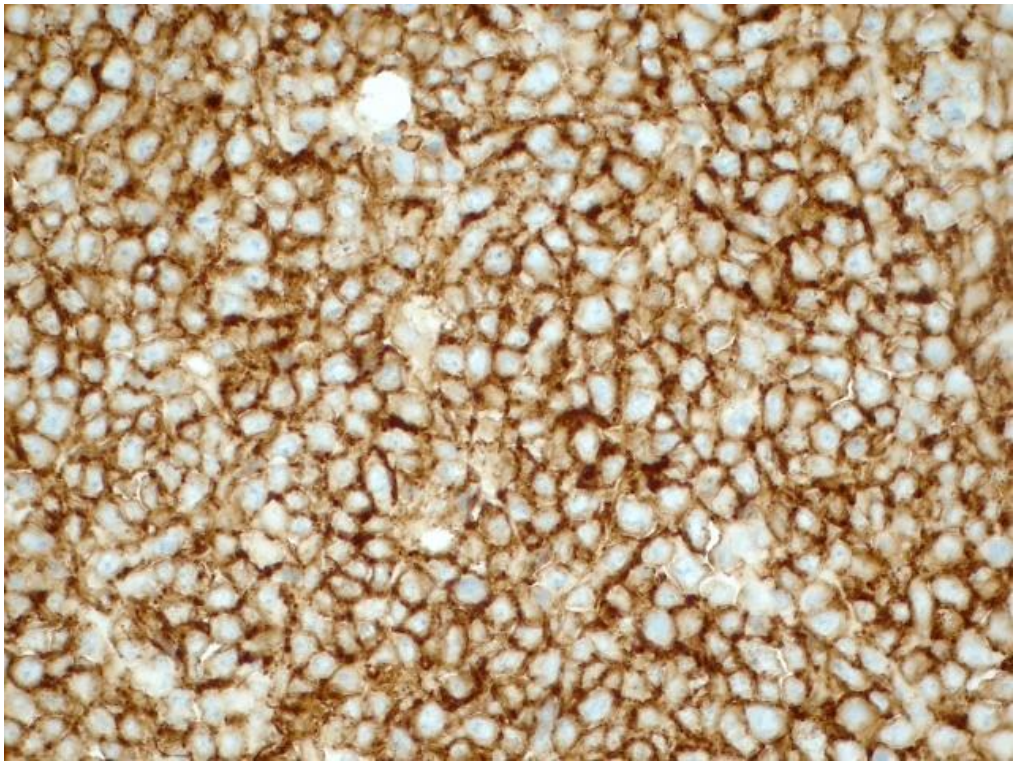


Figure 3.3: Photomicrograph of DLBCL with diffuse membrane expression of CD20 (X400 magnification)

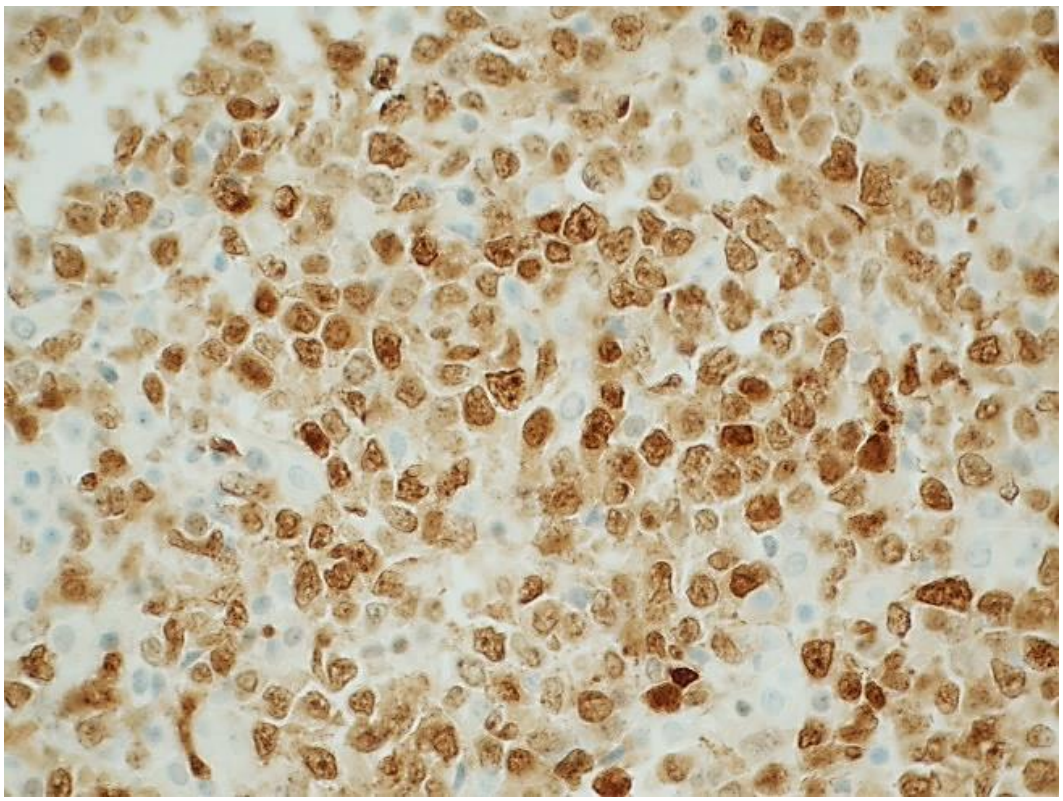


Figure 3.4: DLBCL, ABC subtype from a 53-year–old female, MUM1 nuclear expression (X400 magnification)

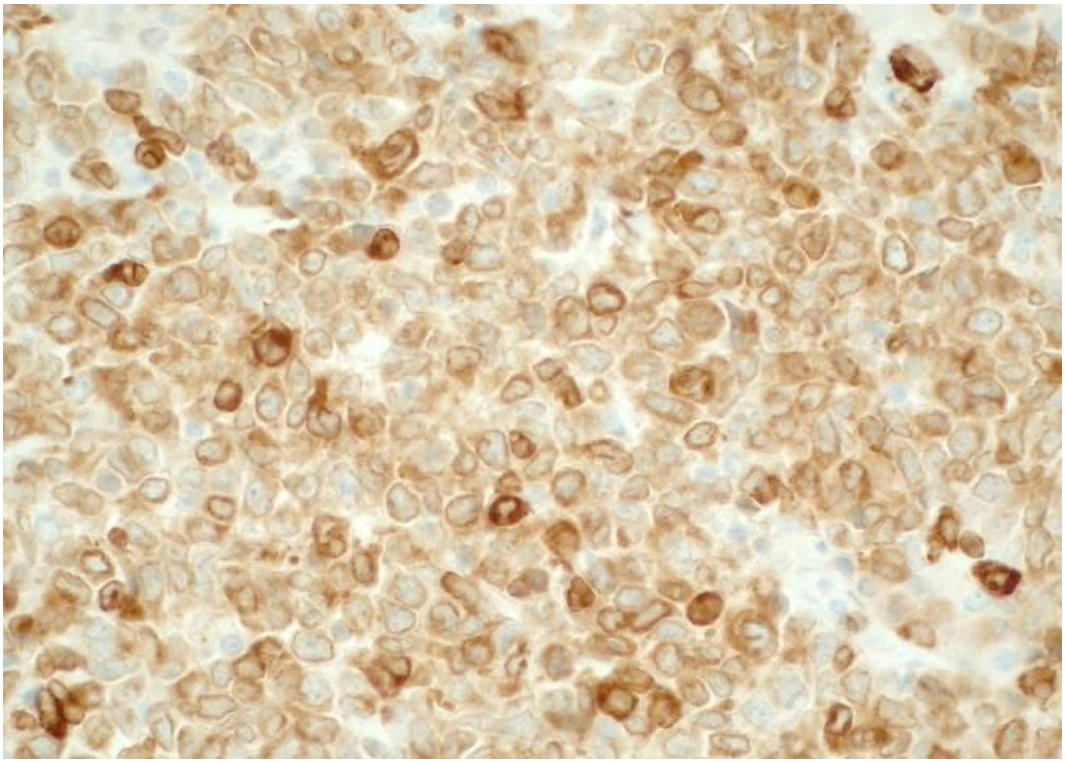


Figure 3.5: DLBCL with cytoplasmic expression of Bcl-2 in >50% of cells (X400 magnification)

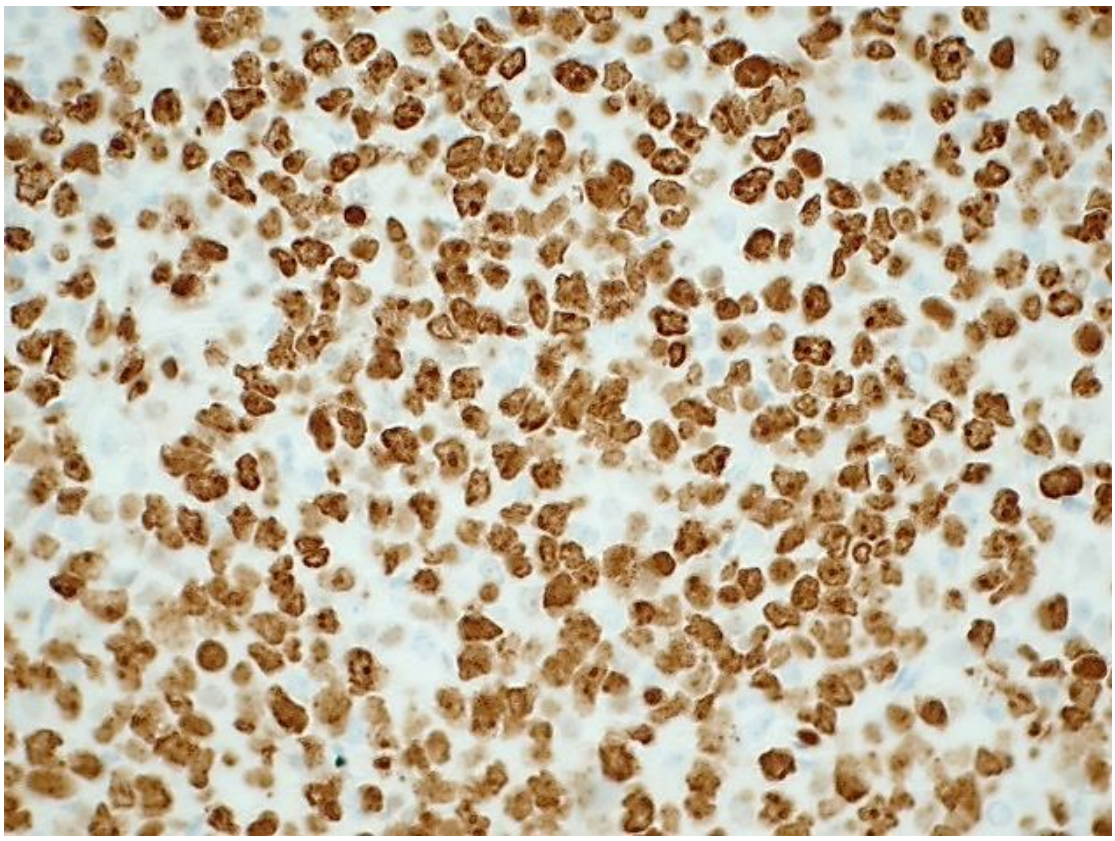


Figure 3.6: DLBCL with high proliferation index, nuclear expression >90% (Ki67, X400 magnification)

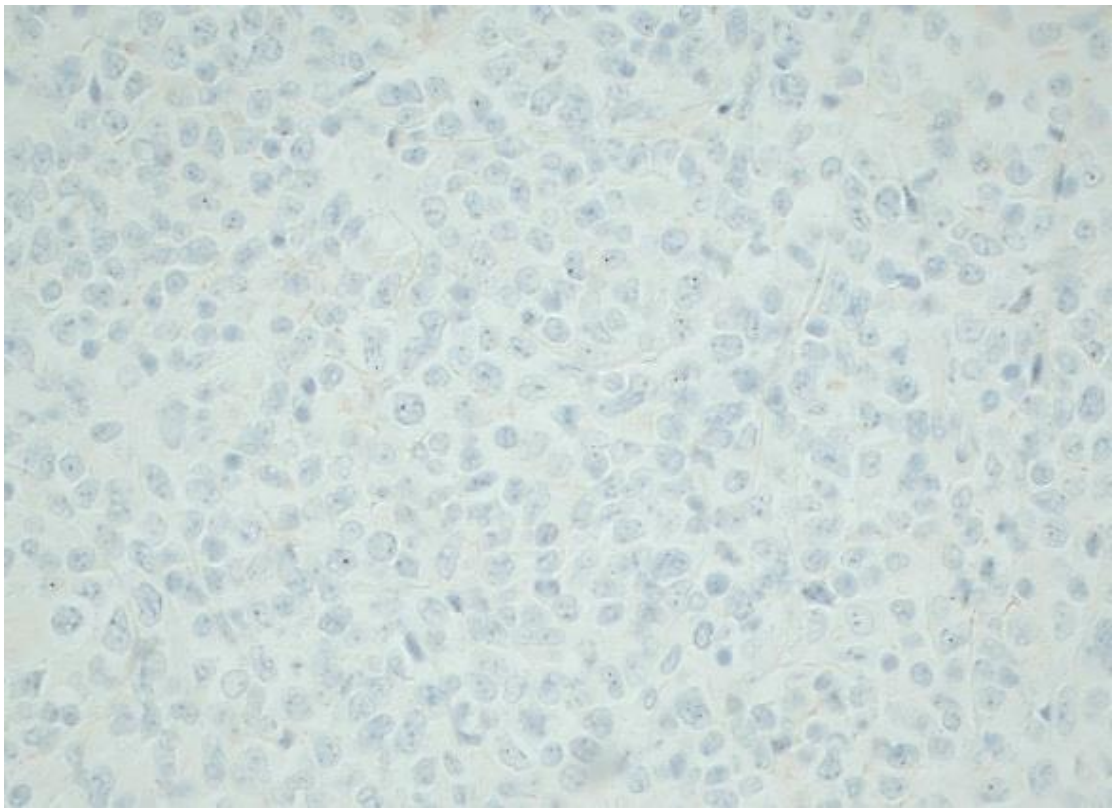


Figure 3.7: DLBCL with negative expression of CD10 (X400 magnification)

3.1.2. DLBCL case of GCB:

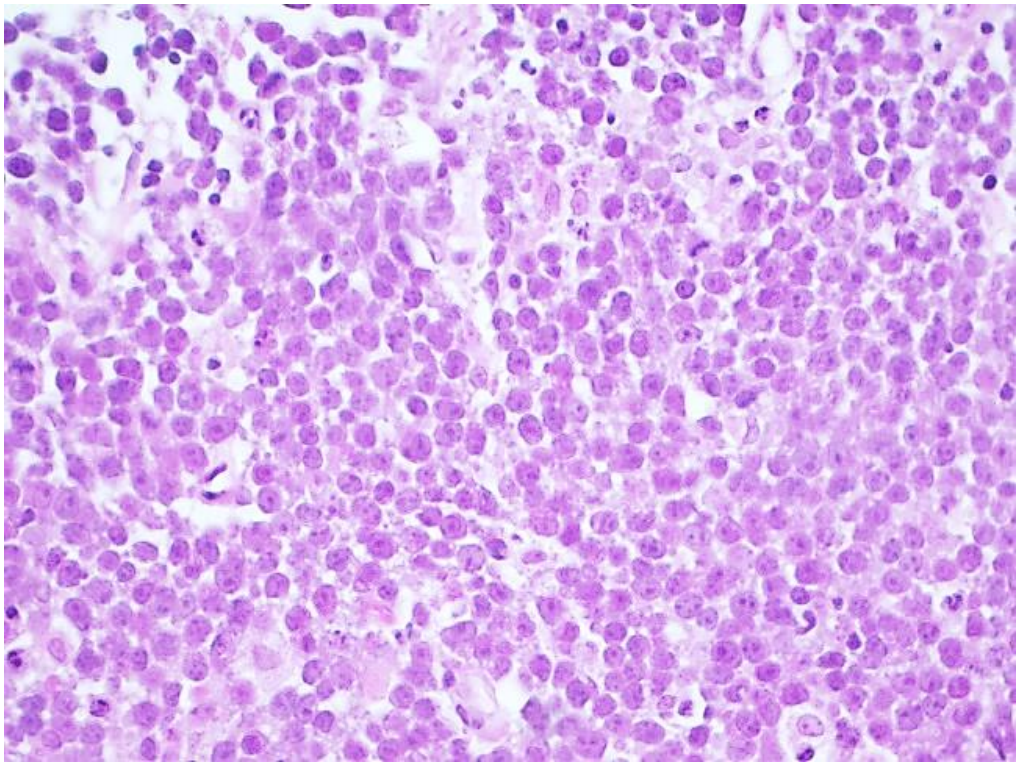


Figure 3.8: Photomicrograph of DLBCL, GCB subtype in a 54-year-old female patient with extra-nodal involvement (H&E staining, X400 magnification)

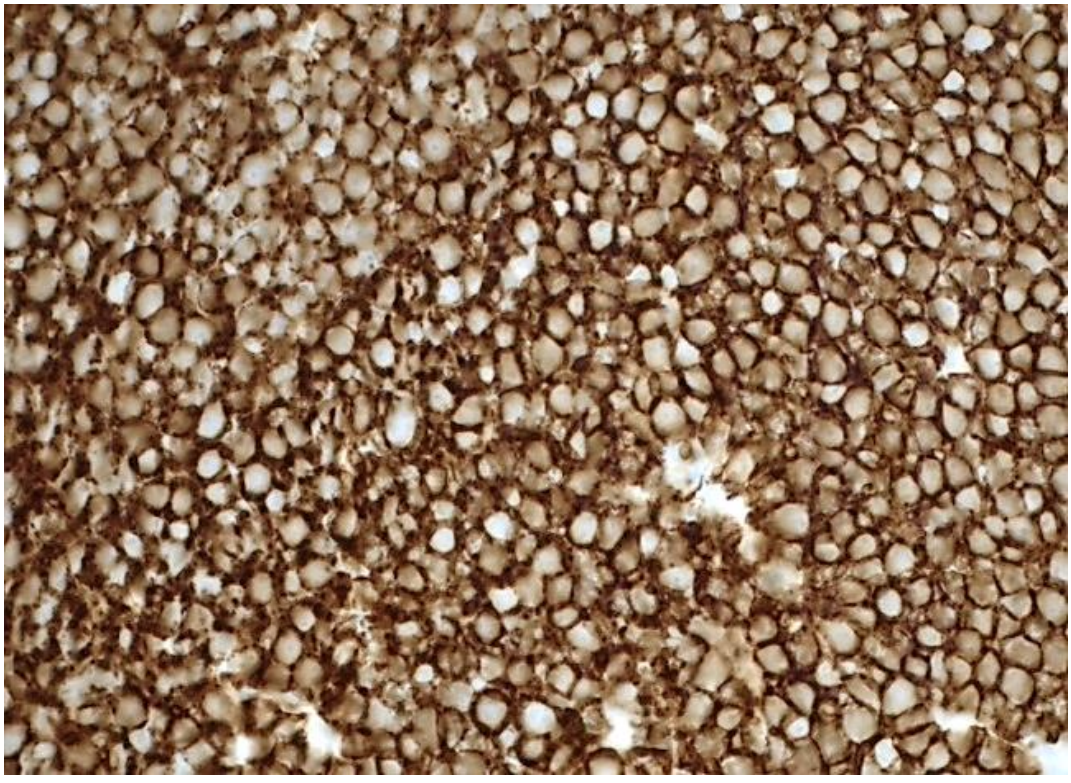


Figure 3.9: Photomicrograph of DLBCL, GCB subtype in a 54-year-old female patient with extra-nodal involvement (CD20 IHC, X400 magnification)

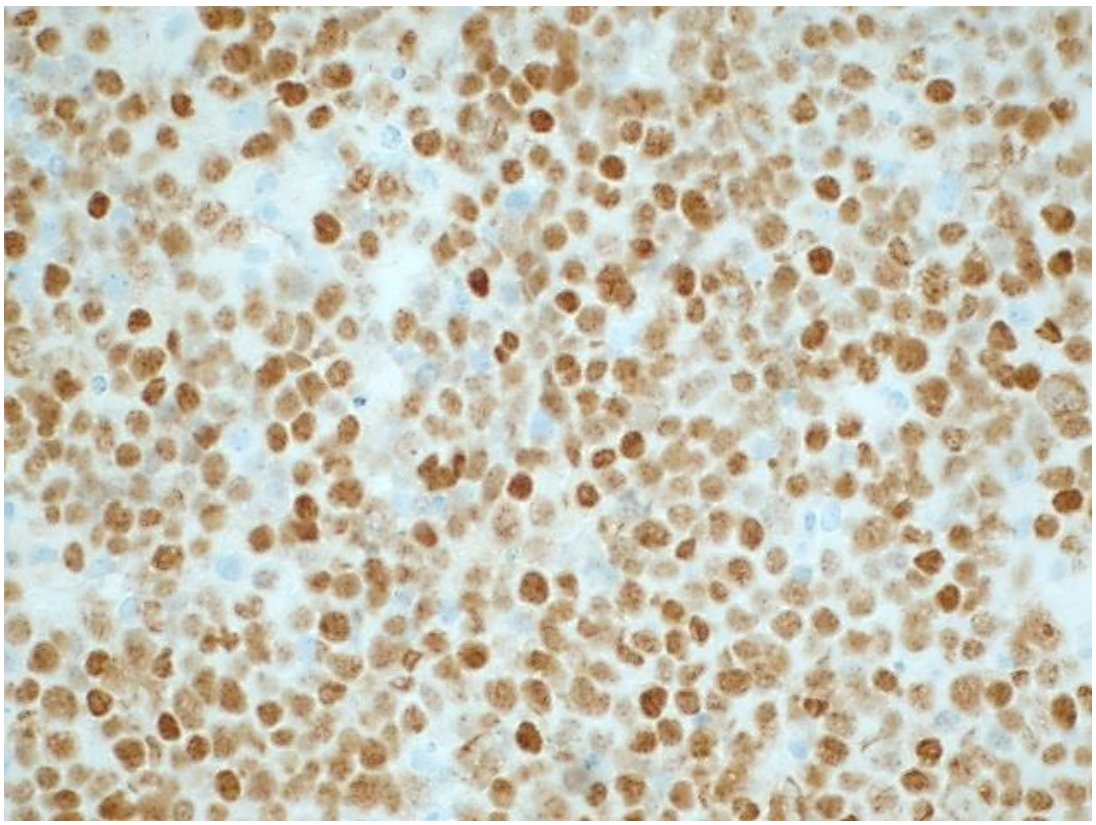


Figure 3.10: Photomicrograph of DLBCL, GCB subtype in a 54-year-old female patient with extra-nodal involvement (Bcl-6 IHC, X400 magnification)

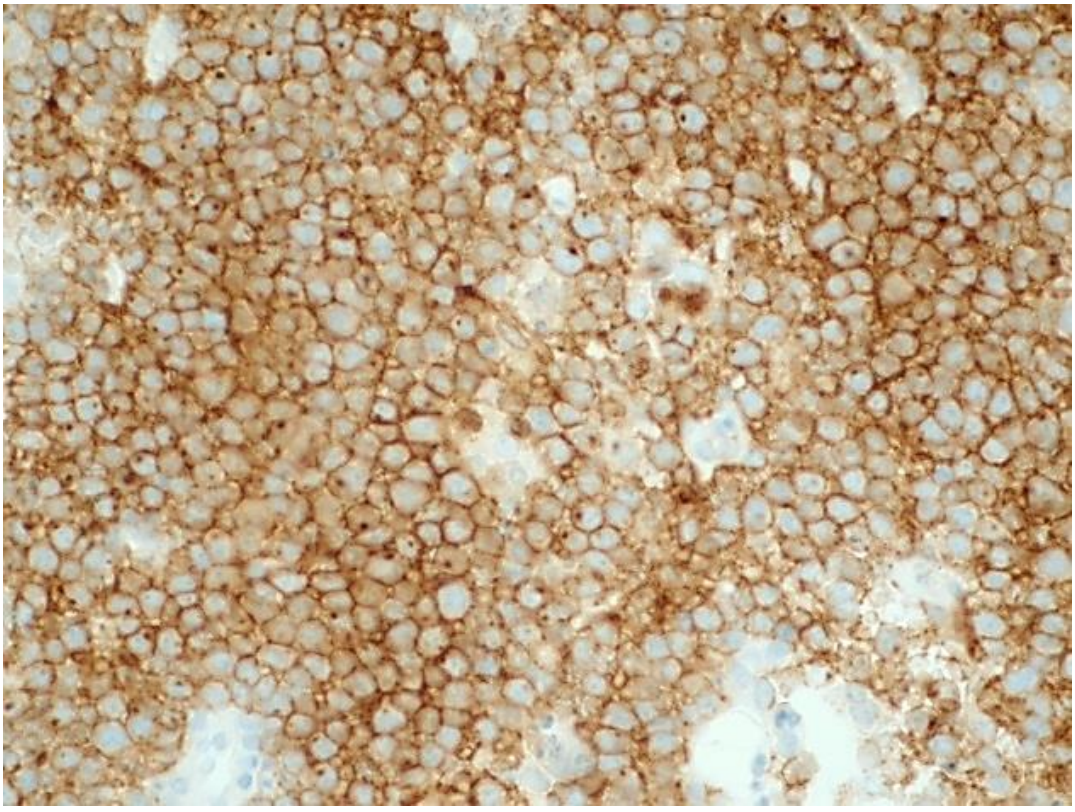


Figure 3.11: Photomicrograph of DLBCL, GCB subtype in a 54-year-old female patient with extra-nodal involvement (CD10 staining, X400 magnification)

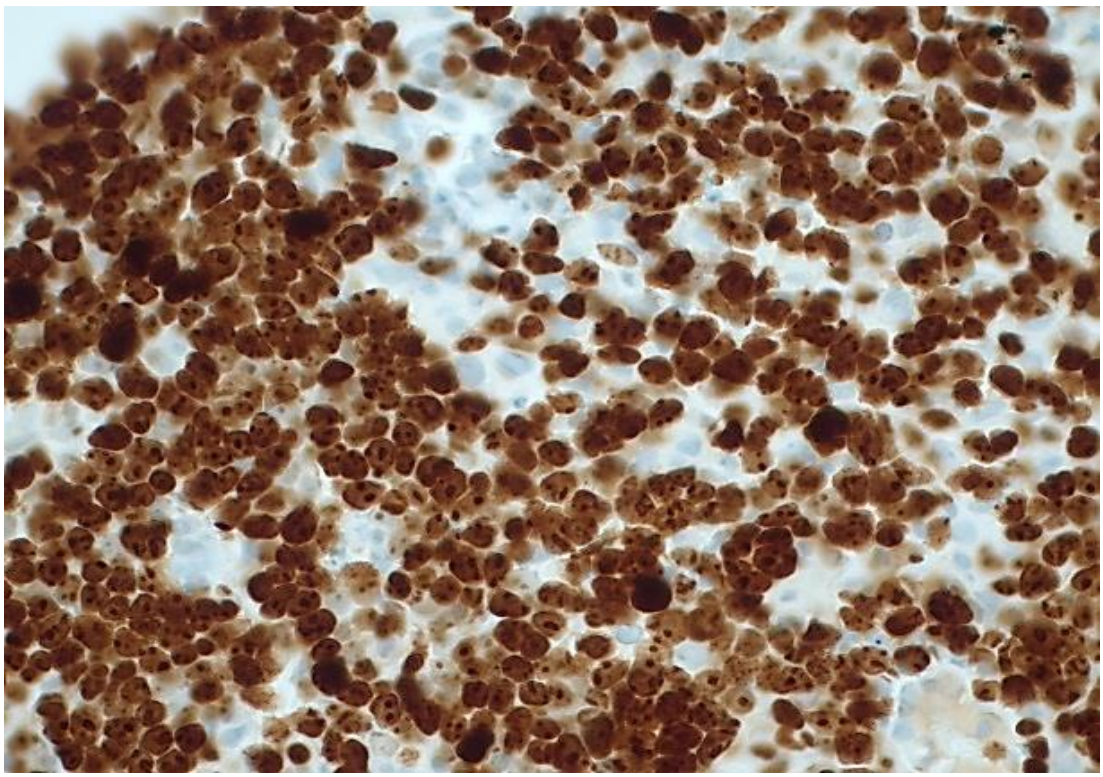


Figure 3.12: Photomicrograph of DLBCL, GCB subtype in a 54-year-old female patient with extra-nodal involvement**,** displaying nucleic overexpression (KI67 staining, X400 magnification)

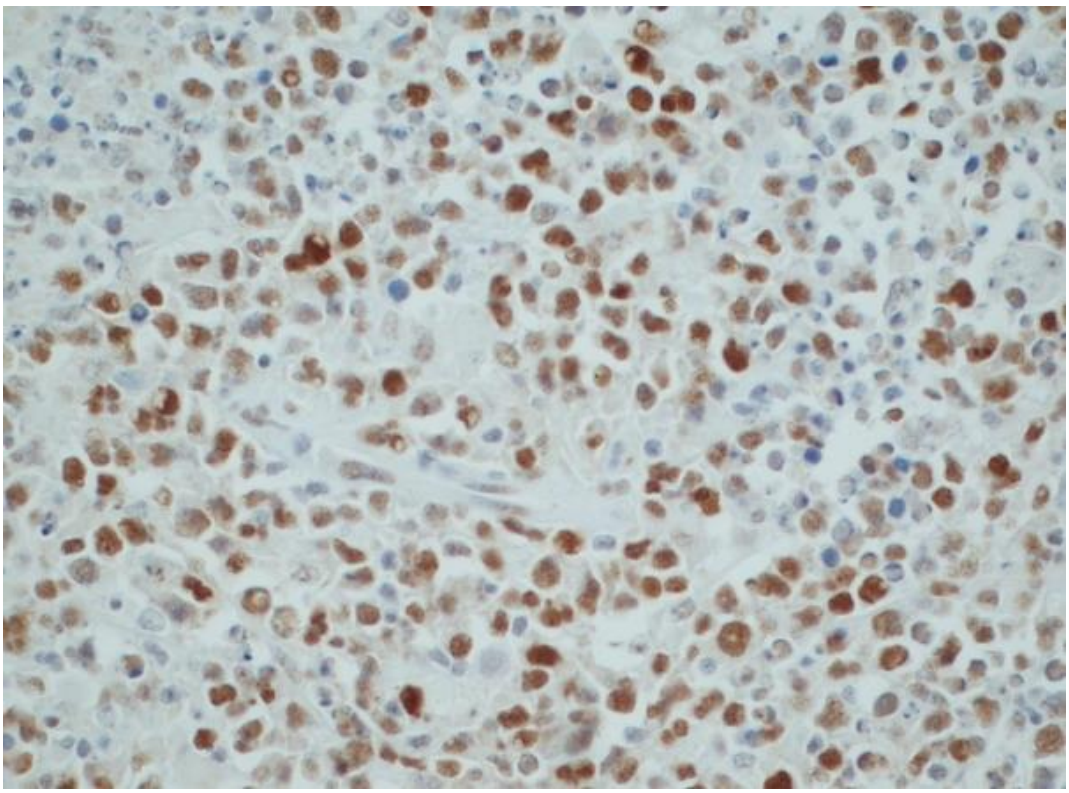


Figure 3.13: Photomicrograph of DLBCL, GCB subtype in a 54-year-old female patient with extra-nodal involvement (c-myc staining, X400 magnification)

Immunohistochemical analysis revealed cytoplasmic expression ≥50% of Bcl-2 in 26 cases (43.3%), 17 (65.4%) were males and nine (34.6%) were females. Among the patients who expressed Bcl-2, 23.1% were alive and the duration of survival ranged from 28 to 1 819 days. For patients who had demised (65.4%), the duration to mortality ranged from 14 to 825 days. There was no significant association of Bcl-2 with p53 (p-value of 0.956). Bcl-2 showed a significant association to DLBCL within lymph nodes (p=0.04).

Immunohistochemical analysis of the proliferative marker Ki67, revealed nuclear immunoexpression ≥90% in 50 cases (83.3%). The cohort comprised 27 male and 23 female patients. Among those alive, the duration of survival ranged from 14 to 1 743 days, while the time to mortality for deceased patients ranged from 11 to 849 days. The Ki67 ≥90% was associated with HIV sero-positivity (p=0.03) and Bcl-6 expression (p=0.01).

FISH analysis for translocation t(8;14) and/or *MYC* rearrangement was available for nine cases (15%), and no translocations were for *BCL-2, MYC* and *BCL-6*. Correlation analysis for FISH results and immunoexpression was not performed due to paucity of the cases and the absence of translocations.

Survival outcome data was available for 42 (70.0%) patients as at 31 June 2017. Table 3.1 depicts the summarized survival data. Overall the minimum days of survival were 11 days and the maximum was 1 819 days.

3.2. Immunohistochemistry of p53 and c-myc

3.2.1. Validation and staining of p53 antibody

Three control slides were stained to determine the optimal staining dilution of the p53 primary antibody (clone: BP53-12) (Sigma, Missouri, USA) using the Dako Envision™ FLEX detection system kit, Figure 3.14. Photomicrograph A depicts a 50 in 1000 µl dilution, which showed over staining of cellular components as well as background staining. Photomicrograph B, which was the 100 in 1000 µl dilution had less background staining but the cytoplasmic and membranous staining were still significant. Photomicrograph C depicts validation staining results for a 150 in 1000 µl of p53 antibody dilution which shows adequate staining of the nuclear components,

nucleoli can also be seen, whereas there is minimal background staining making it easier to differentiate other cellular components (cytoplasm and membrane).

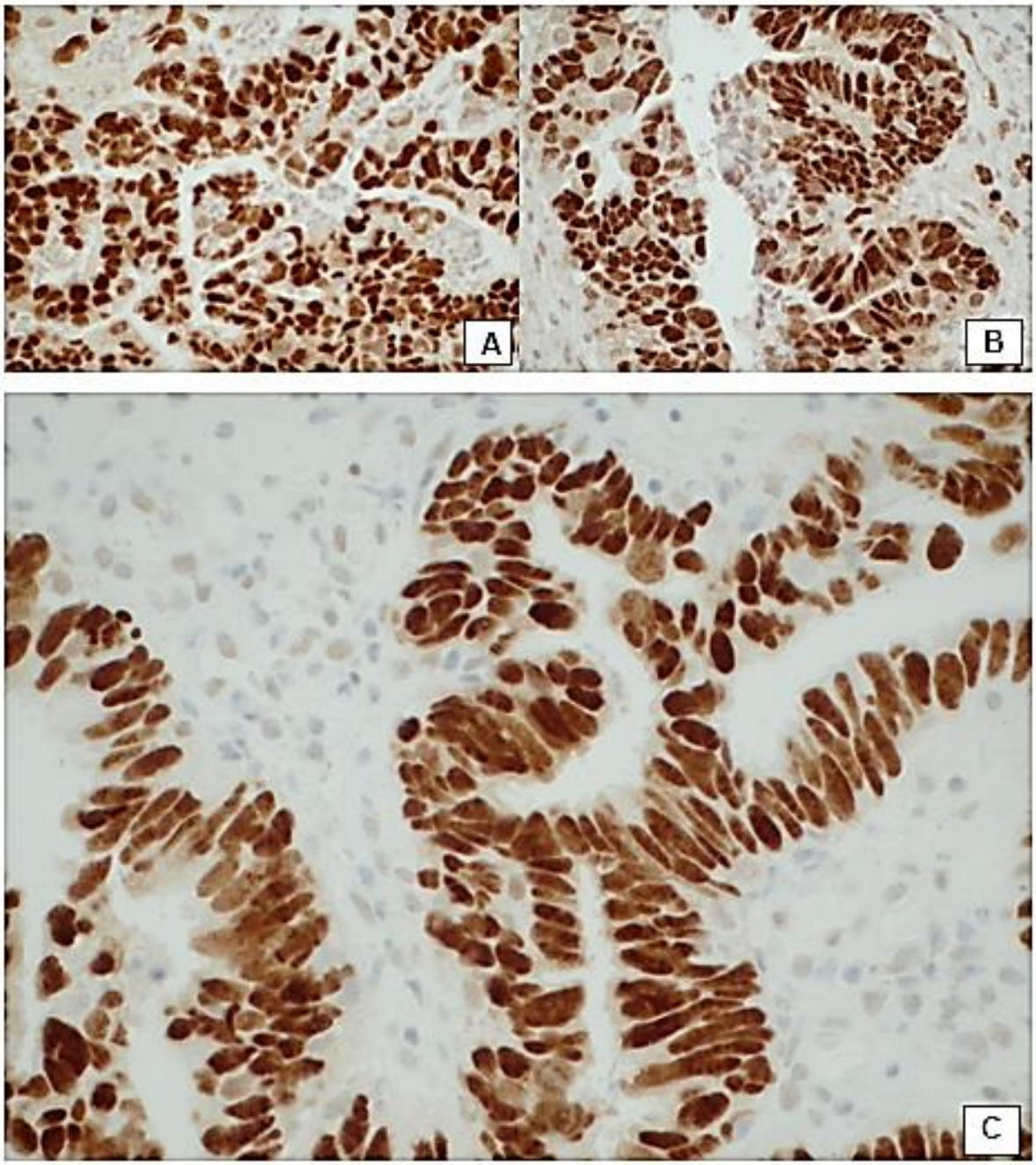


Figure 3.14: Photomicrograph of p53 IHC antibody validation. **A:** 1 in 50 dilution (X400 magnifications), **B:** 1 in 100 dilution (X400 magnifications), and **C:** 1 in 150 dilution (X400 magnifications).

3.3. IHC expression of p53 and c-myc

Figure 3.15 depicts the various staining index profiles analysed for c-myc and p53 protein expression, where there was high incidence of cases not expressing both markers.

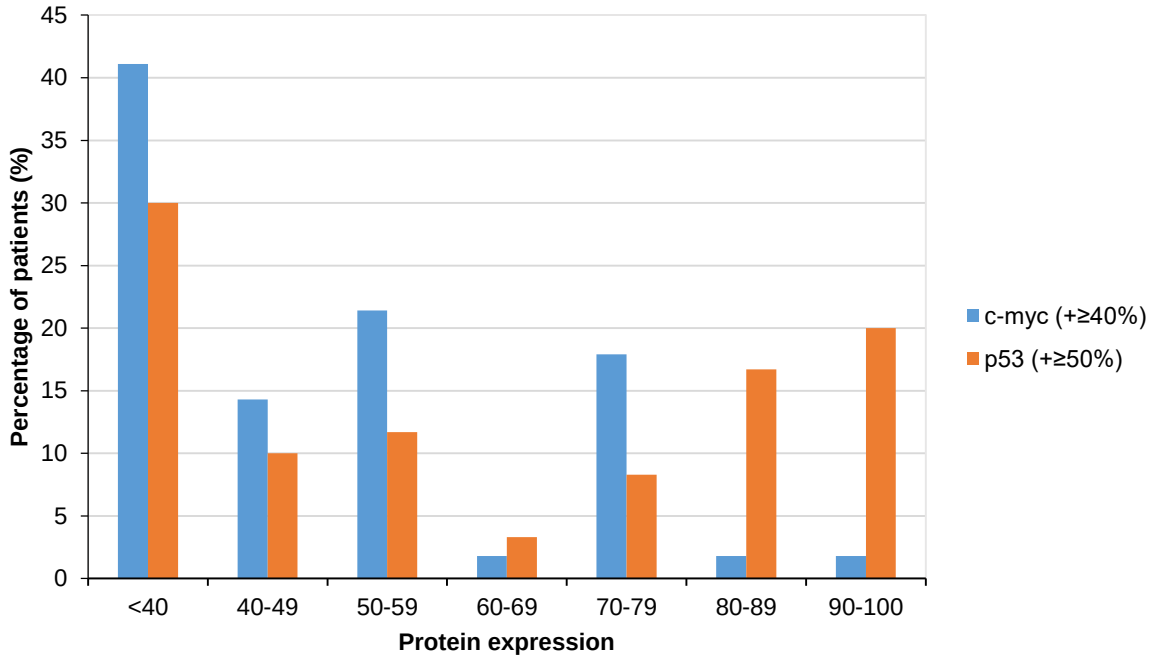


Figure 3.15: Protein expression of c-myc and p53 markers

Staining and evaluation of p53 was possible in all 60 cases (100.0%). Nuclear immunoexpression of ≥50% was detected in 27 cases (45.0%). There were 14 (51.9%) males compared to 13 females (48.1%). The age ranged from 32 to 78 years with median (IQR) age of 48.5 years. HIV sero-positivity was noted in 19 patients (70.4%) who expressed p53 protein. More patients had extranodal involvement (n=16; 59.3%) compared to nodal (n=11), 40.7%. The overall maximum survival duration was 804 days for those who were alive (22.2%) and 825 days for those who had demised (37.0%). The COO classification was possible in 22 (81.5%) of the patients and CD10 was found to significant to p53 immunoexpression, p=0.02. The GCB subtype accounted for 13 (59.1%). Among the two (15.4%) patients who were alive, the duration of survival ranged from 150 to 175 days. For five (38.5%) patients who had demised, the duration to mortality ranged from 11 to 325 days. The ABC subtype accounted for nine (40.9%) cases. Among the four (44.4%) patients who had survived, the duration of survival ranged from 76 to 804 days. For the three (33.3%)

patients who had demised, the duration to mortality ranged from 15 to 325 days. Association of p53 was assessed against several variables and was found to be insignificant, see Table 3.2.

Table 3. 2: Correlation analysis of p53

p53	Gender		HIV		Site		COO		Bcl-2		Ki67		c-myc	
	M	F	+	-	Nodal	Extranodal	ABC	GCB	+	-	+	-	+	-
Expressed	14 (42.4%)	13 (48.1%)	19 (42.2%)	5 (41.7%)	11 (40.7%)	18 (54.5%)	9 (34.6%)	13 (46.4%)	12 (46.2%)	15 (46.9%)	23 (46.0%)	4 (44.4%)	15 (57.7%)	11 (42.3%)
Not expressed	19 (57.6%)	14 (51.9%)	26 (57.8%)	7 (58.3%)	16 (59.3%)	15 (45.5%)	17 (65.4%)	15 (53.6%)	14 (53.8%)	17 (53.1%)	27 (54.0%)	5 (55.6%)	18 (60.0%)	12 (40.0%)
p-value*	0.66		0.97		0.29		0.38		0.96		0.93		0.54	

HIV: Human immunodeficiency, COO: Cell-of-origin, p53: tumour protein 53, p-value: probability value, M: Male, F: Female, +: Positive, -: Negative, ABC: Activated B-cell, GCB: Germinal centre B-cell

* Fisher Exact/ Chi-Squared Test

Protein expression of c-myc was detected in 33 (55.0%) of the cases. Males (n=19; 57.6%) were predominant and females comprised 14 (42.4%). The age ranged from 32 to 77 years with a median (IQR) age of 45.0 years and 26 (78.8%) patients were HIV sero-positive. Nodal involvement occurred in 13 (39.4%) patients, while 20 (60.6%) demonstrated extranodal disease. The GCB subtype accounted for 17 (51.5%) cases. Among the patients who were alive (n=7 41.2%), the duration of survival ranged from 14 to 1 743 days. For patients who had demised (n=7, 41.2%), the duration to mortality ranged from 11 to 325 days. The ABC subtype accounted for 14 (42.4%) cases. Among the patients who had survived (n=4, 28.6%), the duration of survival ranged from 24 to 225 days. For patients who had demised (n=9, 64.3%), the duration to mortality ranged from 11 to 825 days.

Co-expression of p53 and c-myc proteins was evident in 15 (25.0%) cases. There were nine (60.0%) males compared to six females (40.0%). The age ranged from 34 to 73 years with median (IQR) of 46.8 years. HIV sero-positivity was seen in 12 patients, 80.0%, two (13.3%) patients were sero-negative and the other had an unknown status, 6.7%. Patients presenting with nodal involvement accounted for five (33.3%) of the cases compared to extranodal, n=10; 66.7%. The overall maximum survival duration was 225 days for those who were alive (26.7%) and 825 days for those who had demised (46.7%). The COO classification was possible in 13 (86.7%) of the patients. The GCB subtype accounted for eight (61.5%) cases. Among the patients who were alive (n=2, 25.0%), the duration of survival ranged from 150 to 175 days. For patients who had demised (n= 3, 37.5%), the duration to mortality ranged from 11 to 325 days. The ABC subtype accounted for five (38.5%) cases. Among the patients who had survived (n=2; 40.0%), the duration of survival ranged from 76 to 225 days. For patients who had demised (n=3, 60.0%), the duration to mortality ranged from 11 to 825 days. Statistical analysis revealed no significant association of p53 immunoexpression with c-myc (p=0.54).

3.3.1. DLBCL case with p53 and c-myc co-expression (ND-31)

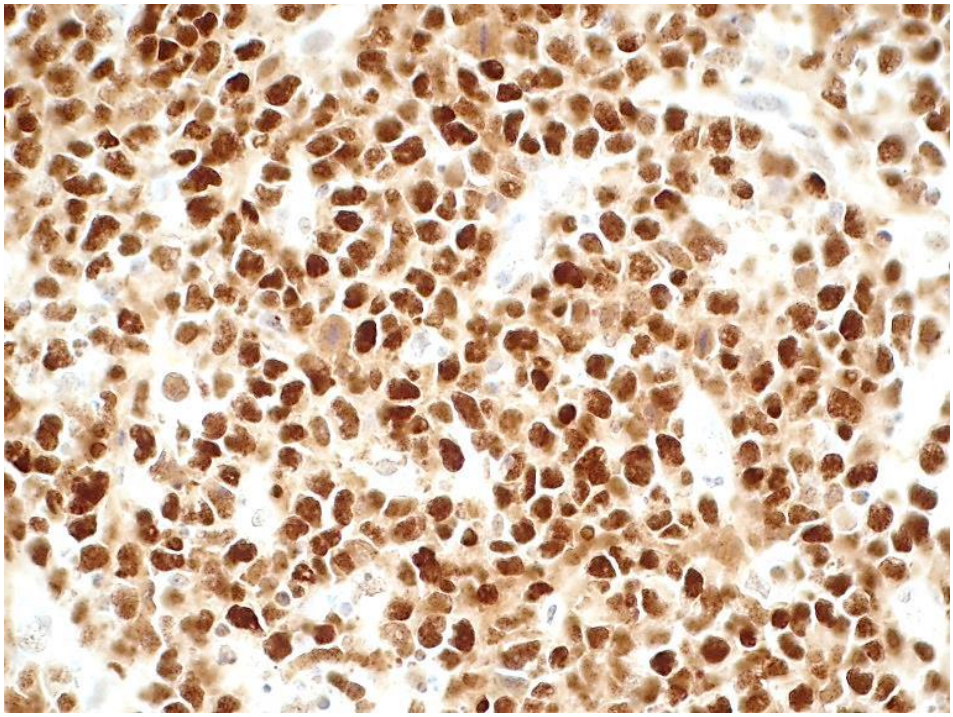


Figure 3.16: Photomicrograph of DLBCL in a 47-year-old male patient with nodal involvement displaying nuclear expression of p53 (X400 magnification)

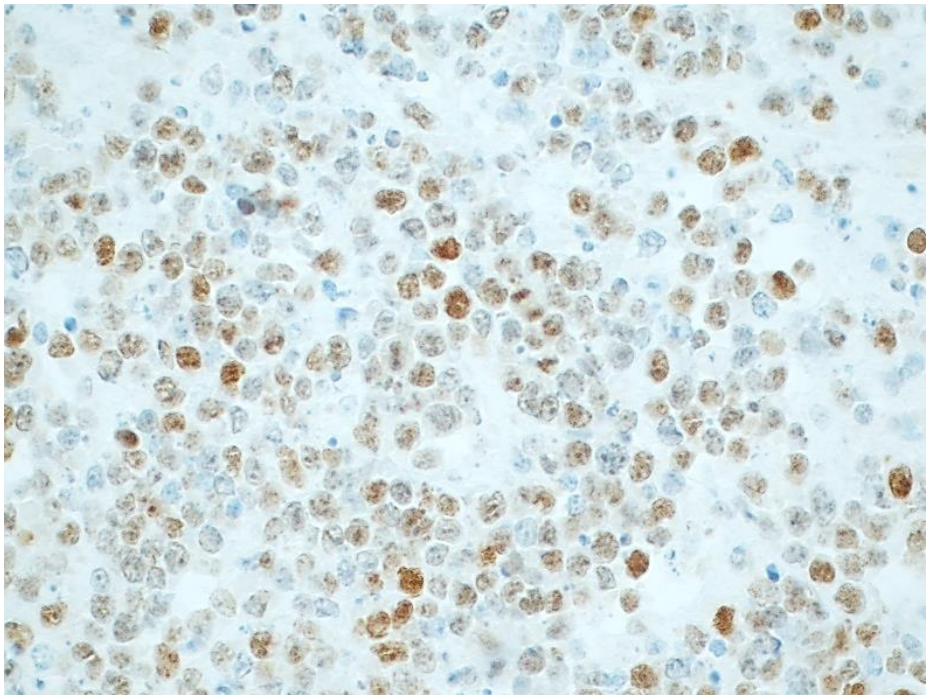


Figure 3.17: Photomicrograph of DLBCL in a 47-year-old male patient with nodal involvement displaying nuclear expression of c-myc protein (X400 magnification)

3.4. Polymerase chain reaction (PCR) and gel electrophoresis

DNA extracted showed a variability in in yield, DNA integrity and purity, these differences may be attributed to different factors such as sample type and size, extraction method and or techniques used, Appendix D. Amplification by conventional PCR was performed on 60 cases to target the *TP53* gene. Amplification bands were detected in six of the cases. Successful amplification of the DBD exons was indicated by a DNA fragment size of 158 bp for exon 5a, 182 bp for exon 5b, 181 bp for exon 6, 177 bp for exon 7, 130 bp for exon 8a and 127 bp for exon 8b. To improve output, exons 5 and 8 were divided into two (a and b) with smaller base pairs. In these cases, overall 16 bands were amplified on gel electrophoresis.

Figure 3.18 depicts gel electrophoresis images by conventional PCR for case A, (ND-05) which exhibited mutations across all exons and case B that had no mutations.

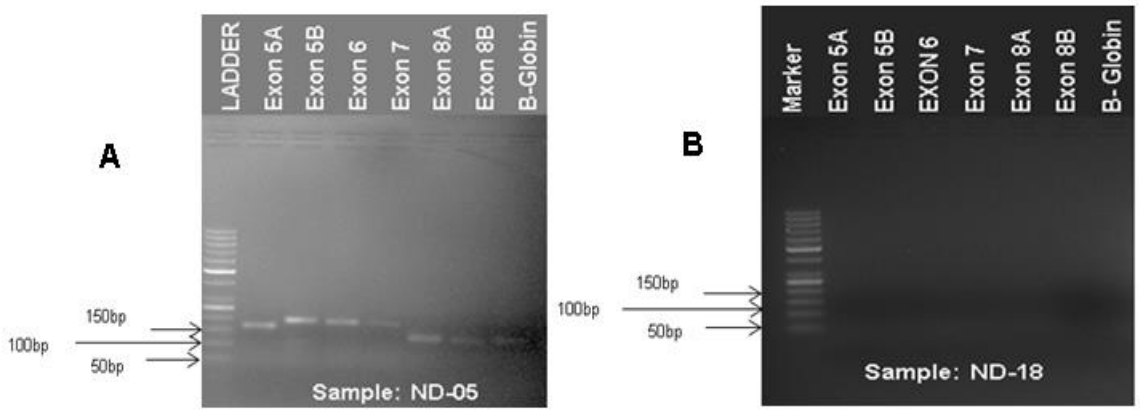


Figure 3.18: Amplification p53 mutations on DBD region by conventional PCR, **A:** Depicts a case with amplification of mutations on all the DBD exons studied. **B:** Depicts a case without any amplification on gel electrophoresis.

Amplification of all exons was successful in one case (ND-05) in this study. The majority of PCR amplification bands were observed in five cases (83.3%) for exon 8, four cases (66.7%) for exon 5, two cases (33.3%) for exon 6, and one case (16.7%) for exon 7, see Figure 3.19A.

3.5. *TP53* Mutational analysis

Among the 60 DLBCL patients, six cases (10.0%) amplified mutational bands on conventional PCR. During sequencing, isolation of genetic mutational data was unsuccessful in exon 8b for a single case. This may be attributed to sequencing fragments smaller than 150 bp utilizing Sanger. DNA quality and quality are impacted

on by PCR reagents especially in samples that have been exposed to formalin fixation.

3.5.1. Demographic characteristics

The demographic characteristics seen in patients with *TP53* mutations in this cohort are described subsequently. The age ranged from 22 to 55 years, with a median (IQR) age of 51 years. Females (66.7%) were more prevalent than males, 33.3%, nodal involvement was in four (66.7%) of cases and four (66.7%) were HIV sero-positive with one (16.7%) sero-negative and one unknown case, 16.7%. The ABC subtype was classified in one case (16.7%), while four cases (66.7%) were categorized as the GCB subtype. The majority of cases (83.3%) exhibited immunoexpression ≥90% of Ki67. p53 was expressed in five (83.3%) of the cases. Survival outcome was available for four cases (66.7%), one patient remained alive with survival duration of 1 743 days, while survival data was unavailable for two patients and three patients had demised with a survival duration ranging from 257 to 825 days.

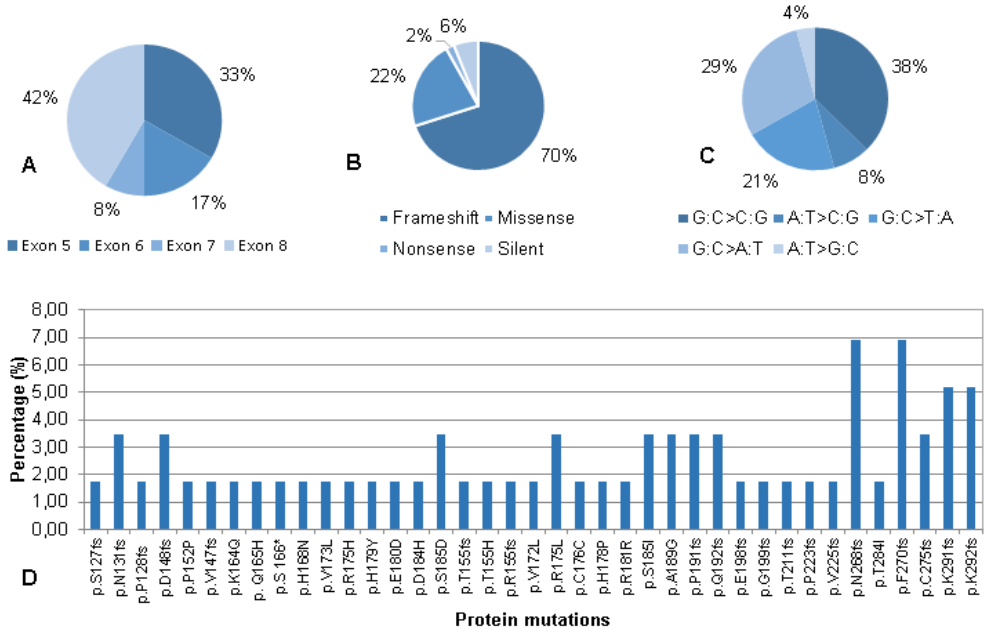


Figure 3.19: A: Proportions of *TP53* mutations in different exons. **B:** Proportions of the class of mutations in sequences. **C:** Proportions of the type of nucleotide changes. **D:** Resultant protein mutations detected in our cohort.

3.5.2. Mutational analysis

Sequencing isolated 67 mutations, Appendix I. Exon 5 harboured the majority of mutations 35 (52.2%), while exon 6 harboured 10 (14.9%) mutations, exon 7 had two mutations (3.0%), and exon 8 had 20 mutations (29.9%). From the mutated cases, successful sequencing of wild-type *TP53* (11.9%) was on exon 7 reverse and 8b reverse, while mutated *TP53* (88.1%) was identified in all the patients and exons. Within the DBD region of the *TP53* gene, 67 mutations were identified. Missense mutations accounted for 22% with seven partially functional, three non-functional and one functional mutations were isolated, as per transactivation class. Frameshift mutations accounted for 70%, silent mutations for 6% and 2% were nonsense mutations. Among these, frame-shift mutations were present in 10 cases (28.6%), missense mutations in eight cases (22.9%), silent mutations in two cases (5.7%), and nonsense mutations in a single case (2.9%). Exon 6 mutations were observed in two cases (33.3%), as illustrated in Figures G.3. A and B. All ten mutations identified within exon 6 were catalogued in the *TP53* database. The majority of mutations harboured in that region were frameshift mutations accounting for eight cases, 80%, followed by missense mutations accounting for 20%. Exon 7 harboured one frameshift mutation, 16.7%. Exon 8 harboured 20 mutations across five of the six mutated cases, 83.3%. Sequencing attempts were unsuccessful for one case, ND-08. Among the mutations identified, 12 cases (60%) involved frameshift mutations, one case (5%) exhibited a missense mutation. Overall 10.4% of the mutations were novel (not previously catalogued) in the TP53 databases were scattered through exons 5, 6 and 8.

3.5.3. Nucleotide changes

The overall single nucleotide variant transition rate was 35.8%. The most dominant single nucleotide variant transitions were G>C accounting for 38.0% followed by G>A accounting for 29.0%, G>T accounting 21.0% and the A>G was the least occurring single nucleotide variant accounting for 4.0% (Figure 3.19C). Deletions were seen as follows, adenine (A) deletions were most common (n=17; 45.9%), followed by thymine (T) deletions (n=12; 32.4%), cytosine (C) deletions (n=7; 18.9%), and guanine (G) deletions (n=1; 2.7%). Insertions were identified in two instances where one involved a thymine insertion and the other a guanine insertion.

3.5.4. Structural motifs

Analysis of the structural motif of the mutations showed, 24 (35.8%) of the mutations were located in the NDBL/beta-sheets, 20 (29.9%) in the L2/L3 and 7 (10.4%) in the L1/S/H2, Appendix J. There were two different mutations present in the NDBL/Beta-sheet, 23 were frameshift and one was silent. Within the L1/S/H2 two mutations were present, 13 were missense and six were frameshift. All the mutations isolated were present in the L2/L3 structural motif, four were frameshift, two were silent and, missense and nonsense accounted for one each. The structural functional class data was not available for majority of the mutations, 83.6%. The mutations with a non-functional structural class were six (9.0%), one was within the L1/S/H2 motif and five were in the L2/L3 motif and those determined to have a functional class accounted for five (7.5%) and they were all in the L2/L3 motif. All these motifs were characterized as damaging. Missense mutations accounted for six non-functional and five functional structural mutations were isolated. Frameshift, silent and nonsense mutations did not contain any structural functions. Pathogenicity was determined on 29.9% of the mutations, 6.0% (n=4) were found to be pathogenic, 6.0% (n=4) likely pathogenic, 4.5% (n=3) possibly pathogenic, 3.0% (n=2) were benign and 10.4% (n=7) were VUS. VUS refers to when insufficient evidence exists to classify pathogenicity.

3.5.5. Protein (Amino acid) mutations [NM_000546.5]

The most common protein mutations consisted of four cases with amino acid changes at Asn268 (p.N268fs), four with Phe270 (p.F270fs), three with Lys291 (p.K291fs) and three cases with Lys292 (p.k292fs), see Figure 3.19D for all the amino acid changes identified. The most common mutations did not comprise any structural or functional data. There were 26 protein mutations that occurred only once amongst all the exons, and nine had occurred twice. Of the 67 mutation, catalogued mutations accounted for 79.1% of the mutations and overall 10.4% of the mutations were novel (not previously catalogued) in the *TP53* databases and were scattered throughout the DBD exons.

3.5.5.1. MUT *TP53* with p.N268.fs and p.N270.fs (ND-5, ND-8, ND-37, ND-64)

All these patients expressed p53 immunoexpression. Three were females and one was male. They were all younger than 60 years. Two had nodal involvement and another two were HIV sero-positive.

3.5.5.2. MUT *TP53* with p.K291.fs and p.K292.fs (ND-5, ND-37, ND-59)

These patients possessed WT *TP53* and were all females. ND-37 and ND-59 had extranodal involvement, GCB subtype and did not express c-myc protein. Their ages were 22 and 54 years.

4. CHAPTER FOUR: DISCUSSION

4.1. Demographic characteristics in DLBCL

Diffuse large B-cell lymphoma (DLBCL) may be associated with a poor prognosis.[1–3] The analysis of multiple biomarkers aids in early detection and developing improved treatment plans, potentially increasing overall survival rates.[3]

This study analysed 60 cases from CHBAH, NHLS Anatomical Pathology department between 01 October 2013 to 31 June 2017.

The study revealed that the mean age of the patients was 45 years (IQR, 22-88 years). Majority of our cohort fell within the 40 to 49 years age range. Although DLBCL may be diagnosed in a large age range, it is usually reported in older patients aged 65 years and older.[4] According to the World Health Organization (WHO), DLBCL primarily affects more males than females.[5] Consistent with the published literature, our study also showed a higher incidence among males 33 (55.0%) compared to females. The younger age of this study cohort may be attributed to the high HIV sero-prevalence rate of 16.7% in South Africa among patients aged 15 to 49 years.[66] NHL ranks as the second most common cancer among patients with advanced HIV disease, with DLBCL as one of the defining conditions.[26] In seropositive settings, the mean age is typically between 39 and 42 years.[4,17,18] Our findings align, as majority of our patients were documented as HIV sero-positive, and the patients had a median age of 42.1 years (IQR, 32-73 years).

We demonstrated that DLBCL predominantly involved the extranodal compared to nodal sites. In contrast, the primary site of involvement in cases of DLBCL is the lymph nodes, which are affected in approximately 60-70% of cases where the HIV sero-status of the patients was not studied.[4]The extranodal topographic distribution of DLBCL is documented include the gastrointestinal tract as the most commonly involved, accounting for 40% of the cases. [21–23] In our cohort the most prevalent extranodal sites were oropharyngeal region, liver and breast. Although the LNs constitutes as the most prominent site, an overall comparison of nodal involvement to the extra nodal involvement as an independent variable contrasts with literature. These discrepancies in topographic sites may be attributed to sample size, HIV infection and the selection criteria of the samples.

4.2. Immunohistochemistry: A diagnostic tool for cancer

Immunohistochemical techniques are invaluable in the laboratory for diagnosing and grading various cancers. This approach tends to be more cost-effective compared to PCR and sequencing methods. The diagnosis of DLBCL is based on the utilisation of immunohistochemistry as a surrogate for the molecular profiles.[1,3,4,25]

The Hans algorithm was utilised for the COO classification as it provides valuable insights into the specific B-cell that underwent the initial mutation leading to cancer.[1,3,4,25] The most common COO subtype is the GCB subtype.[51] Similarly, in our cohort of DLBCL patients, we were able to classify the COO into the GCB ABC subtypes for the majority 54 (90.0%) of the cases. The GCB subtype emerged as most prevalent in 46.7% of the patients, although there was a marginal difference compared to the ABC subtype, which was found in 43.3%. The GCB subgroup is associated with a better overall survival in DLBCL patients compared to the ABC subgroup.[51] In our study, the mortality rate for ABC subtype was 46.2% and the GCB subtype was 42.9%. In the study conducted by Hans *et al*. 58% of the patients had non-GCB subtype of DLBCL with a median age of 64 years and the majority were male. The GCB subtype of DLBCL accounted for the remaining percentage of the patients with a median age of 60 years and a higher incidence of females.[25] In our study cohort the median age for the ABC subtype patients was 44 years and 45.5 years for the GCB subtype, males comprised the majority for both the GCB and the ABC subtypes. The differences noted in the median age may be attributed to the higher number younger patients seen in our study cohort.

Bcl-2 protein plays a significant role in the regulation of proliferation and tumour formation. The Bcl-2 protein expression is observed in approximately half of DLBCL cases.[55] The ABC subtype patients with the overexpression of Bcl-2 have a poor prognosis.[55] Bcl-2 expression with c-myc and p53 is associated with poor prognosis.[55] Our findings are similar, patients with the overexpression of Bcl-2 presented with the ABC subtype compared to the GCB subtype. Equivalent to the study, unfavourable outcome was associated with the co-expression of Bcl-2, c-myc and p53 proteins, as they were evident in 10.0% of the patients and of those, 66.7% had demised. The association between p53 and Bcl-2 immunoexpression did not demonstrate significance (p=0.96). Bcl-2 had a significant association to lymph node

involvement (p=0.04). Bcl-2 plays a key role in influencing survival and proliferation of lymphoid cells within nodes. Bcl-2 overexpression leads to apoptosis resistance which allows for the persistence and accumulation of malignant cell within lymph nodes.[38]

Ki67 serves as a crucial marker for assessing tumour cell proliferative activity and is associated with poor overall survival outcome.[27] In our study cohort, Ki67 ≥90% was noted in the majority of the cases (83.3%). A statistical analysis of Ki67 and p53 showed no significant correlation, p>0.93. This is similar to published literature where Lui *et al.* and Patrascu *et al*. noted that cell proliferation of normal B-cells is not affected by *TP53* mutations.[27,67] Increased levels of Ki67 protein are an indicator of the rapid development of lymphoma. The Ki67 ≥90% was associated with HIV sero-positivity (p=0.03) and Bcl-6 expression (p=0.01). This may be because Ki67 is expressed in the active proliferation phase and certain factors such as infections cause suppression of the immune system, thereby influencing the advancement of DLBCL. The co-expression of Bcl-6 and Ki67 is said to suggest poor prognosis in ABC subtype.

The p53 protein is essential for numerous cellular functions, such as halting the cell cycle, triggering cell death (apoptosis), repairing DNA, maintaining genomic stability, influencing cell differentiation, and contributing to cellular aging (senescence).[61,62,68] Mutant p53 is associated with the dysregulation of the cell cycle arrest, genomic instability and uncontrolled proliferation.[51,61] The overexpression of p53 is associated with a poor prognosis. In various studies the immunoexpression of p53 was recommended to be a predictive diagnostic marker.[27,61] Expression of p53 is associated with poor prognosis and it affects 20.0% of DLBCL patients.[27,55] A higher incidence of p53 protein expression was observed in our cohort (45%), of which 70.4% were HIV sero-positive. p53 expression was more prevalent in males than in females. This may be attributed to the possible late stage of the tumour and high rate of HIV-sero-positive patients. The over expression of p53 has been linked to lymphomagenesis and treatment resistance. When p53 protein loses its function to suppress and kill tumour cells, it tends to play a key role in disease progression.[52] Of the p53 positive cases, majority (40.7%) had demised hereby suggesting that patients with p53 immunoexpression have an unfavourable prognosis.

There was a similar distribution of GCB and ABC subtypes. CD10 which is a biomarker that plays a role in the development and differentiation of lymphoid cells

was significantly associated with p53 (p=0.02), a protein that prevents the transformation of diseased lymphoid cells. This association explains the high incidence of GCB subtype in p53 as CD10 protein expression is the primary marker of the GCB subtype. Due to the paucity of published literature showing correlation of p53 immunoexpression to CD10 in DLBCL cases, the relationship between the two biomarkers needs to be further studied as it may possibly give an insight in tumour progression and treatment.

c-myc plays an essential role in tumorigenesis, impacting apoptosis, cellular growth and differentiation.[27] The protein expression of c-myc is reported in 30 to 50% of DLBCL cases.[27,55] In our study, 55.0% of the DLBCL tumours expressed c-myc (≥40% tumour cells), aligning with existing literature. For the COO grouping, GCB subtype accounted for the greater majority compared to the ABC subtype. Noteworthy is the co-expression of p53 and c-myc, which has been previously reported to have an improved prognosis in R-CHOP treated patients.[55] In this study they concluded that the rate of apoptosis has a protective impact thus creating a tumour suppressor pathway. Our findings revealed that the co-expression occurred in 25.0% of the cases. However, no significant correlation could be established, p=0.86. A study conducted by Chang *et al.* reported that the co-expression of p53 and c-myc occurred more frequently amongst patients at high clinical stages, with a median survival of seven months.[69]

A study by Patrascu *et al*. found that specific molecular markers (Bcl-2, Ki67, c-myc and p53), not included in the Hans algorithm, are associated with a low prognostic index and poor survival rates in DLBCL.[27] In our cohort, prognosis was poor for the 6,7% of the patients who had showed immunoexpression of the above molecular markers, 75% had demised with one patient having unknown survival data. These markers are valuable as diagnostic and therapeutic targets, emphasizing the importance of incorporating their routine evaluation into a new IHC algorithm for better disease management.[27]

4.3. FISH analysis

The most common gene rearrangement is that of *BCL6* (3q27) documented in approximately 35% cases of DLBCL.[38] The *BCL2* t(14;18)(q32;q21) rearrangement is documented in 20-30% cases of DLBCL and is linked to poor survival.[38] Gene rearrangement of *MYC* by FISH is said to have a poor prognostic outcome in patients with aggressive p53 IHC staining.[70] The over expression of *MYC* with t(8;14)(q24;q32) rearrangement is present in 15% of DLBCL cases.[69] In the current study, *MYC* rearrangements was negative for 9 of our patients irrespective of the p53 results.

Analysis of *TP53* by FISH is available but not feasible in a low-income population. Xu-Monette *et al.* successfully amplified chromosome 17p13.1 deletion in 87.0% of their samples.[52] Stefancikova *et al*. reported that p53 aberrations had a negative effect on survival.[71] Akay *et al.* concluded that *TP53* mutations do not control the expression of *TP53* rearrangement, which was associated with better survival.[38]

4.4. Mutational analysis

The p53 protein is encoded by the *TP53* gene, which acts as a transcription factor and key regulator in the cell cycle.[38] The gene plays a significant role in cell cycle regulation, proliferation, apoptosis and genomic integrity.

Limited literature exists on the demographic characteristics of patients with *TP53* genetic mutations. Qin *et al*. described a Chinese cohort of 214 patients with a median age of 56 years, with 63.6% over 60 years, females were predominant accounting for 52.3%.[64] Conversely, our study identified *TP53* genetic mutations in 10% of the patients, who are younger with median (IQR) age of 45.1 years. Similarly, we demonstrated that females were predominant (71.4%). Additionally, 66.7% of these patients were HIV sero-positive. Young *et al.* found that younger patients at diagnosis are associated with *TP53* mutations and they further documented a median age in mutant *TP53* to be 51 years and 64 years in patients with WT *TP53*.[51] The patients in our cohort harboured mixed mutations. The patients with WT *TP53* were all females, 22, 48 and 54 years of age. Given our small sample size, further research may be required to fully explore the demographic characteristics which may be associated with *TP53* mutations, to enhance clinical management and address existing knowledge gaps.

Notably, the GCB subtype patients exhibit more *TP53* mutations than patients with the ABC subtype. However, both subtypes are associated with poor survival outcomes.[52] Similarly in our cohort, we found that majority of the patients with *TP53* mutations had the GCB-DLBCL subtype, compared to the ABC subtype. Xu-Monette *et al.* queried whether p53 is not fully activated by hypermutation during chemotherapy, which may explain the decreased number of ABC subtype patients harbouring *TP53* mutations.[72]

The presence of *TP53* mutations is associated with poor prognosis.[51,73] Patients with *TP53* mutations are documented to have a diminished overall survival of 1.3 years compared to DLBCL patients without *TP53* mutations, who have an OS of approximately 5 years.[51] Given our low rate of cases with *TP53* mutations, it is challenging to make conclusive statements. Patients with p53 protein expression had high mortality.

TP53, known as the guardian of the human genome, is the most mutated tumour suppressor gene found in the majority of human cancers. In DLBCL it is mutated in 20% of the cases.[51,52,61] Whereas, some have documented mutations to occur in 18-30% of DLBCL patients.[26] *TP53* genetic mutations were documented in only 10% of our cohort and four of these patients were HIV sero-positive. In addition, of the cases that demonstrated *TP53* mutations, 83.3% cases were positive for p53 immunoexpression (≥50% of tumour cells). A reduced incidence of *TP53* mutations has been reported in DLBCL malignancies, which are characterized by rapid disease progression, poor treatment response, and shortened survival in other cancers.[51] This discrepancy may also be due to sample collection and the use of FFPE samples. We employed Sanger sequencing to identify various types of mutations (missense, deletions, substitutions, and insertions) linked to p53 in our study cohort. Patients exhibited mutations in the DBD region, specifically from exon 5 to exon 8, with mutations in exon 8 being more prevalent than in other exons. These exons have been reported to contain mutations greater than 90% in DLBCL compared to other lymphomas.[51] The mutations in the DBD are said to be associated with a poor OS.[51]

The mutations we detected include frameshift, silent, missense and nonsense mutations. A study by Stengel *et al.* in acute lymphoblastic leukemia (ALL) reported various mutation types (frameshift, in-frame, nonsense, and splice-site), with

missense mutations being the most common.[68] Conversely, in our cohort frameshift mutations accounted for the majority followed by missense mutations.

Frameshift mutations catalogued on the *TP53* database account for 11% of all mutations and produce a truncated protein.[74] Frameshift mutations were on all the DBD exons, and detected in 70.0% of all the mutations in our cohort. The most common frameshift variants were p.N268fs, p.F270fs, p.K291fs and p.K292fs. All these mutations were on exon 8 and their functional effects are unknown. The most common amino acid mutations (p.N268fs, p.F270fs, p.K291fs and p.K292fs) are documented to affect a few cases in a variety of carcinomas. There is a paucity of literature documenting these mutations, therefore not much is known about their effects. In our cohort, three female patients expressed both p.K291fs and p.K292fs, these patients also expressed WT *TP53* in exons 5 and 8. Amino acid mutations, p.N268fs and p.F270fs, were identified in patients who are younger of age and express p53 protein. In this study there was an even number of nodal and extranodal cases. Similarly, these mutations were previously reported to involve extranodal sites such as the breast, mouth, GIT and ovary.[75] According to the Catalogue of somatic mutations in cancer these mutations affect several pathways, these are: Androgen Receptor Signaling Pathway, AMPK signaling, DNA damage response, miRNAs involved in DDR, Fluoropyrimidine Activity, Folate, Cell cycle, Apoptosis, MAPK signaling pathway, TGF-beta Receptor Signaling Pathway. Wnt Signaling, Pathway and Pluripotency, G1 to S cell cycle control, Delta-Notch Signaling Pathway, Senescence and Autophagy, ErbB signaling pathway, DNA damage response and DNA damage response (only ATM dependent). Codons catalogued as prevalent frameshift mutations are 151 and 152. These mutations are said to be associated with the absence of protein expression and are commonly found in exon 4, 9 and 10.[74] They are related with poor prognosis such as aggressive disease, resistance to treatment and high relapse rates. The disparity may be attributed to our study taking a focus on analysing only the DBD region consisting of exon 5 to 8. These mutations are found on the NDBL/Beta-sheet, however the pathogenicity, functional and SIFT class of these mutations are not catalogued.

Missense mutations are reported to be most common type of mutations in the DBD of the *TP53* gene.[76] Unlike other tumour suppressor genes, *TP53* is primarily inactivated through missense mutations, with over 80% of alterations resulting in a stable full-length protein. In contrast, most tumour suppressor genes are typically

inactivated by mutations that either prevent protein synthesis or produce truncated proteins.[61] Missense mutations occur due to a single nucleotide change that alters the amino acid sequence of a protein, leading to structural and functional problems. In our cohort, missense mutations accounted for 22.0% of the variants. Among the variants with a single nucleotide change, the mutation classification G>C accounted for the highest number of proportion, followed by G>A and G>T. Methylation of the cytosine in the *TP53* gene is associated with transition mutations found in human cancers and genetic diseases.[74] The transition occurs when a mutation is translated from G:C to A:T. The most frequently affected areas are hotspot mutations, primarily located in exon 8, notably codon 273 as well as codons 248 and 282 have been reported as hotspots.[74,77] The variants responsible for loss of function at this certain site (p.R248Q, p.R248W, p.R273H, and p.R273C) were not detected in our analysis. Similarly, the transition at the CpG site is a hotspot for the p.R175C (G>A) mutational variant, which is found in 30% of tumours.[74] This variant was not detected in this study. Kennedy *et al.* previously identified R175, R213, G245, R248, R273, and R282 as hotspot codons responsible for 25% of *TP53* missense mutations.[76] Du *et al.* reported missense mutations to have a poor survival outcome compared to non-missense mutations, $p<0.05$.[78] The missense mutations in this study were predominantly found in the L1/S/H2 motif and majority of these mutations were partially functional.

Nonsense mutations occur when a codon is altered to a stop codon.[75] A single nonsense mutation (p.S166*) was detected on the L2/L3 motif and it was catalogued as pathogenic. There is paucity of literature documenting these mutations in DLBCL. A study by Rossi *et al.* documented *TP53* nonsense mutations in 8% of their cohort of chronic lymphocytic leukemia patients.[79] This mutation is known but its functional effects are unknown. The mutation was recorded on codon 497 with a codon mismatch of C>G, where TCA>TGA (stop codon). These mutations produced incomplete, truncated and non-functional proteins and were associated with the termination of protein synthesis.[75]

Silent mutations are regarded as wild-type *TP53* since they do not result in any phenotypic changes. The *TP53* database has catalogued silent mutations at codons 375, 477 and 672 as the most frequently mutated silent SNVs.[74] We identified three silent mutational variants in our cohort which were Pro > Pro, Cys > Cys and Arg > Arg. All these variants were found on exon 5. These variants are catalogued in the

database and have a benign pathogenicity. The disparity in the detection of frequently mutated silent variants may be due to the different demographic characteristics of the patients as well as how the mutations are associated to the tumour. Additionally to silent mutations, other wild-type *TP53* variants were detected in exons 7 and 8, where no changes appeared during sequencing. Wild-type *TP53* variants are known to have a shorter half-life of approximately 20 minutes, in contrast to mutant *TP53*, which have a half-life ranging from 1.4 to 7 hours.[67]

Exon 5 and 8 mutations are associated with a poor prognosis in patients with lung cancer. *TP53* mutations occur with a frequency of 20% or more in cases of DLBCL.[57,80] Whereas our study reported a frequency of only 10%, this may be due to sample size differences and case selection criteria. Mutations classified as those with a high or moderate degree of danger were; p.L130F, p.T155I, p.R196Q, p.G244S, p.V272E, and p.A276V. IVS6-36G > C and IVS5 + 43G > T, and were therefore, documented as functionally significant mutations.[57] None of these mutations were found in our study cohort.

There are several mechanisms affecting *TP53* in DLBCL. In normal cells the *TP53* levels remain low through the negative regulators such as MDM2. The functions of mutant *TP53* and *WT TP53* may be affected by the dysregulation of signalling pathways, as they modify *TP53* in tumours by over regulating the expression of MDM2.[81] A structural motif in proteins is a recurring three-dimensional configuration of secondary structural elements that are connected in a specific pattern. Secondary structural elements constitute alpha-helices and beta-sheets. The motifs frequently serve critical functional roles in proteins, these may include mediating interactions with other proteins, small molecules and DNA. Only a small fraction of our mutations had functional data. Those with non-functional structural motifs are unable to perform their biological roles due to alterations. These motifs are associated with aggressive tumour phenotypes.[81] Loop-sheet-helix motifs play a critical role in protein-protein interactions.[82] Majority of the mutations in our cohort were from the NDBL/ beta-sheets motif. It has not yet been established whether this motif correlates with inferior survival. The L1/S/H structural motif binds to the DNA major groove and is associated with predicting poor survival.[51,61] The DNA minor binds Loop-L3 which is a poor survival predictor.[51,61] Under physiological conditions, the L2 motif supplements *TP53* binding activity with the DNA helix and its association to prognosis has not been established.[51] More than half of the mutations in this cohort were found in the L1/S/H

and L2/L3 motifs, thereby indicating that these mutations may predict poor survival. Although the frequency of VUS is high, several mutations have been linked to being pathogenic or likely pathogenic. According to Soussi *et al.* this information plays a critical role in formulating clinical outcome probability scores.[83] The phosphorylation of serines and threonines in mutant *TP53*, especially Ser392, as information on tumorigenesis and impacting activities is available. Phosphorylation is associated with poor overall prognosis, altering the half-life of proteins and affecting activities associated with how mutant *TP53* and other proteins respond to anticancer drugs.[81] It also increases the potency of oncoproteins by enhancing certain tetramers. The acetylation of *TP53* is often associated with stabilization, whereas in mutant *TP53* lysines, especially Lys120 and Lys163 located in the DBD are often mutated even though they are not altered in majority of cancers. The acetylation of mutated lysines is associated with the activation of *TP53* target gene, which suggest that these lysines may play a critical physiological role.[81] None of our mutations constituted Ser392, Lys120 or Lys163. The mutations in our study cohort are suggested to predict poor survivals as majority are found within the L1/S/H2 and the L2/L3 structural motifs.

The prognostic significance of *TP53* mutations in DLBCL remains inconsistent. Although they are typically viewed as a poor prognostic factor, this variability may arise from the heterogeneity of the mutations, limitations in detection methods, or the diverse functions of *TP53* mutations.[62] In our study, p53 immunoexpression was evident in 43.3% of the cases. Of the cases that demonstrated *TP53* mutations, 83.3% cases were positive for p53 immunoexpression (≥50% of tumour cells). Despite the inefficiency of molecular studies like PCR and sequencing, it would be beneficial to have an alternative method to screen for and detect *TP53* mutations. Immunohistochemical nuclear staining for p53 (>50%) is suggested as a substitute for identifying *TP53* mutations. [73,70] However, our results contradicted this hypothesis, as only 19.2% of p53 immuno-expressive DLBCL cases had mutations. The disparities in the figures may be attributed to the demographic characteristics of the cohort groups, high HIV infection rate and possibly sample selection criteria used. Patients harbouring *TP53* mutations tend to have poor overall survival compared to those without these mutations. These mutations occur in the DBD region of the gene.[77] Majority (90%) of these mutations within the DBD account for over 50% of epithelial cancers and 5-25% of lymphoid malignancies.[61,62] *TP53* inactivation can

lead to tumour formation. Analysing gene mutations in cancer patients is essential, as this can significantly impact diagnosis, prognosis, and treatment.[64] The *TP53* gene plays a crucial role in determining how tumours respond to radiation and chemotherapy, influencing patients' sensitivity to these treatments.[51] Patients with p53 protein expression and those who harbour *TP53* mutations are treated with the standard CHOP or R-CHOP treatment and patients tend to have poor survival outcomes.[64] Studies on *TP53* mutations are pivotal in order to tailor different treatment options that will have a favourable effect on the patients quality of life.

4.5. STUDY LIMITATIONS

A number of limitations existed in this study. These included incomplete records and tissue inadequacy due to tumour being cut away in some the cases. Some of the immunohistochemical stains were not available for review and in that context, data was extracted from the reports. For three of the patients' HIV status was unknown.

5. CHAPTER FIVE: CONCLUSIONS

In this study, the DLBCL patients were predominantly HIV sero-positive with a median (IQR) age of 48.5 years and male patients predominated. In our setting, extranodal involvement was more common. There was a relatively even distribution of GCB and ABC COO subtypes. p53 was documented in 45.0% DLBCL and 70.4% in the HIV sero-positive patients. HIV sero-positive patients with p53 protein expression tend to be younger in age, predominantly male, with extranodal involvement and with GCB subtype. *TP53* mutations were detected in 10.0% of the tumours and included frameshift, missense, nonsense and silent. The G>C was the most common codon change, exon 8 had the highest number of mutations and majority of the mutations were located on NDBL/Beta-sheet motif. Patients with these mutations were younger, HIV sero-positive, predominantly female and with GCB subtype. These patients had an even distribution of nodality. There was no association detected for p53 and c-myc protein expression. Future research based on the utilisation of PCR/ sequencing and FISH techniques is warranted.

Drawing a definitive conclusion regarding *TP53* mutations in DLBCL patients remains challenging due to the limited availability of literature and research on this topic. Addressing these knowledge gaps through further studies is essential, particularly within the Southern African context, to enhance understanding and improve clinical management by designing personalised treatment trials for favourable outcomes. Given the high cost of sequencing, systematic documentation of these mutations is essential for designing PCR assays to enable efficient genetic analysis, particularly in disadvantaged demographic groups.

REFERENCES

1. Trembath D. *Rosai and Ackerman's Surgical Pathology Review*. Vol 33. 11th ed. (Goldblum JR, Lamps LW, McKenney JK, Myers JL, eds.). Elsevier; 2008. doi:10.1097/pas.0b013e31817fb3e1

2. Susanibar-adaniya S, Barta SK. Update on Diffuse large B cell lymphoma: A review. *Am J Hematol*. 2021;96(5):617-629. doi:10.1002/ajh.26151.2021

3. Alaggio R, Amador C, Anagnostopoulos I, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. *Leukemia*. 2022;36(7):1720-1748. doi:10.1038/s41375-022-01620-2

4. Swerdlow SH, Campo E, Pileri SA, et al. WHO Classification of Tumours of the Haematopoietic and Lymphoid Tissues Review Series THE UPDATED WHO CLASSIFICATION OF HEMATOLOGICAL MALIGNANCIES The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*. 2016;127(20). doi:10.1182/blood-2016-01-643569.The

5. Wang SS. Epidemiology and etiology of diffuse large B-cell lymphoma. *Semin Hematol*. 2023;60(5):255-266. doi:10.1053/j.seminhematol.2023.11.004

6. Xie. Yi. PS and JSE. The Histological Classification of Diffuse Large B-cell Lymphomas. *Semin Hematol*. 2015;52(2):57-66. doi:10.1053/j.seminhematol.2015.01.006.The

7. Lu TX, Young KH, Xu W, Li JY. TP53 dysfunction in diffuse large B-cell lymphoma. *Crit Rev Oncol Hematol*. 2016;97:47-55. doi:10.1016/j.critrevonc.2015.08.006

8. May P, May E. Corrigendum: Twenty years of p53 research: Structural and functional aspects of the p53 protein (Oncogene (1999) 18 (7621-7636)). *Oncogene*. 2000;19(13):1734.

9. Voulgarelis M, Ziakas PD, Papageorgiou A, Baimpa E, Tzioufas AG, Moutsopoulos HM. Prognosis and outcome of non-hodgkin lymphoma in primary sjögren syndrome. *Medicine (Baltimore)*. 2012;91(1):1-9. doi:10.1097/MD.0b013e31824125e4

10. Gibson TM, Engels EA, Clarke CA, Lynch CF, Weisenburger DD, Morton LM. Risk of diffuse large B-cell lymphoma after solid organ transplantation in the United States. *Am J Hematol*. 2014;89(7):714-720. doi:10.1002/ajh.23726

11. International Agency for Research on Cancer. https://gco.iarc.fr/causes/infections/home

12. Ross AM, Leahy CI, Neylon F, et al. Epstein–Barr Virus and the Pathogenesis of Diffuse Large B-Cell Lymphoma. *Life*. 2023;13(2):1-18. doi:10.3390/life13020521

13. Wang F. Characteristics and Prognosis of HHV8-Positive Diffuse Large B-Cell Lymphoma, Not Otherwise Specified: A Population-Based Study. Published online 2023.

14. Vega F, Miranda RN, Medeiros LJ. KSHV/HHV8-positive large B-cell lymphomas and associated diseases: a heterogeneous group of lymphoproliferative processes with significant clinicopathological overlap. *Mod Pathol*. 2020;33(1):18-28. doi:10.1038/s41379-019-0365-y

15. Kumar, V., Abbas, A.K., Aster, J.C. et al. *Robbins Basic Pathology*. 10th ed. Elsevier; 2018. https://studentconsult.inkling.com/

16. Patel M, Philip V, Omar T, et al. The Impact of Human Immunodeficiency Virus Infection (HIV) on Lymphoma in South Africa. *J Cancer Ther*. 2015;06(06):527-535. doi:10.4236/jct.2015.66057

17. Magangane PS, Mohamed Z, Naidoo R. Diffuse large B-cell lymphoma in a high human immunodeficiency virus (HIV) prevalence, low-resource setting. *South African J Oncol*. 2020;4:1-7. doi:10.4102/sajo.v4i0.104

18. Berham A, Bayleyegn B, Getaneh Z. HIV/AIDS Associated Lymphoma: Review. Blood Lymphat Cancer. 2022. doi:10.2147/BLCTT.S361320

19. Chen Y, Zhao J, Sun P, et al. Estimates of the global burden of non-Hodgkin lymphoma attributable to HIV: a population attributable modeling study. *eClinicalMedicine*. 2024;67(December 2023):1-15. doi:10.1016/j.eclinm.2023.102370

20. Sukswai N, Lyapichev K, Khoury JD, Medeiros LJ. Diffuse large B-cell

lymphoma variants: an update. *Pathology*. 2020;52(1):53-67. doi:10.1016/j.pathol.2019.08.013

21. Vitolo U, Seymour JF, Martelli M, et al. Extranodal diffuse large B-cell lymphoma (DLBCL) and primary mediastinal B-cell lymphoma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2016;27(July):v91-v102. doi:10.1093/annonc/mdw175

22. Candelaria M, Onate-Ocana LF, Corona-Herrera J, et al. Clinical Characteristics of Primary Extraodal versus Nodal Diffuse Large B-cell Lymphoma: Aretrospective cohort study in a cancer center. *Rev Investig Clin*. 2019;71:349-358.

23. Chen SY, Ji MM, Xu PP, Zhao WL. Primary extranodal diffuse large B-cell lymphoma: Molecular features, treatment, and prognosis. *Aging and Cancer*. 2022;3(3-4):133-146. doi:10.1002/aac2.12058

24. Abramson JS. T-Cell/Histiocyte-Rich B-Cell Lymphoma: Biology, Diagnosis, and Management. *Oncologist*. 2006;11(4):384-392. doi:10.1634/theoncologist.11-4-384

25. Hans CP, Weisenburger DD, Greiner TC, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood*. 2004;103(1):275-282. doi:10.1182/blood-2003-05-1545

26. Perry AM, Mitrovic Z, Chan WC. Biological prognostic markers in diffuse large B-cell lymphoma. *Cancer Control*. 2012;19(3):214-226. doi:10.1177/107327481201900306

27. Pătraşcu AM, Rotaru I, Olar L, et al. The prognostic role of Bcl-2, Ki67, c-MYC and p53 in diffuse large B-cell lymphoma. *Rom J Morphol Embryol*. 2017;58(3):837-843.

28. Chen BJ, Fend F, Campo E, Quintanilla-Martinez L. Aggressive B-cell lymphomas—from morphology to molecular pathogenesis. *Ann Lymphoma*. 2019;3:1-1. doi:10.21037/aol.2018.12.02

29. Nayak P, Desai D, Pandit S, Rai N. Centroblastic variant of diffuse large B-cell

lymphoma: Case report and review of literature. *J Oral Maxillofac Pathol*. 2013;17(2):261-265. doi:10.4103/0973-029X.119749

30. Dominis M, Débro S, Gašparov S, Pešut A, Kušec R. Diffuse large B-cell lymphoma and its variants. *Croat Med J*. 2002;43(5):535-540. doi:10.1201/b16685-13

31. Singh N, Sood R, Agrawal N, Pasricha S, Mehta A. Anaplastic Diffuse Large B Cell Lymphoma: A Single Center Experience. *Indian J Hematol Blood Transfus*. 2019;35(3):557-560. doi:10.1007/s12288-019-01096-8

32. Mishra D, Singh S, Narayan G. Role of B Cell Development Marker CD10 in Cancer Progression and Prognosis. *Mol Biol Int*. 2016;2016:1-9. doi:10.1155/2016/4328697

33. Li YJ, Li ZM, Rao HL, et al. CD20-negative de novo diffuse large B-cell lymphoma in HIV-negative patients: A matched case-control analysis in a single institution. *J Transl Med*. 2012;10(1):1-11. doi:10.1186/1479-5876-10-84

34. Katchi T, Liu D. Diagnosis and treatment of CD20 negative B cell lymphomas. *Biomark Res*. 2017;5(1):1-5. doi:10.1186/s40364-017-0088-5

35. M. R. Ki67. PathologyOutlines.

36. Pather S, Patel M. HIV-associated DLBCL: Clinicopathological factors including dual-colour chromogenic in situ hybridisation to assess MYC gene copies. *Ann Diagn Pathol*. 2022;58(February). doi:10.1016/j.anndiagpath.2022.151913

37. Irawan C, Iskandar M, Harahap AS, Rumende CM, Ham MF. MUM1 Expression versus Hans Algorithm to Predict Prognosis in Indonesian Diffuse Large B-Cell Lymphoma Patients Receiving R-CHOP. *Cancer Manag Res*. 2022;14(March):925-935. doi:10.2147/CMAR.S345745

38. Akay OM, Aras BD, Isiksoy S, et al. BCL2, BCL6, IGH, TP53, and MYC protein expression and gene rearrangements as prognostic markers in diffuse large B-cell lymphoma: A study of 44 Turkish patients. *Cancer Genet*. 2014;207(3):87-93. doi:10.1016/j.cancergen.2014.02.001

39. Almeida R, Abrantes C, Gigliano D, et al. Clinical and Pathological Features of Double-Hit and Triple-Hit High-Grade B-Cell Lymphomas: A Retrospective

Study from Three Portuguese Tertiary Centers. *Int J Hematol Stem Cell Res*. 2022;16(2):94-102. doi:10.18502/ijhoscr.v16i2.9202

40. Kluk MJ, Chapuy B, Sinha P, et al. Immunohistochemical detection of MYC-driven diffuse large B-cell lymphomas. *PLoS One*. 2012;7(4). doi:10.1371/journal.pone.0033813

41. Kim H, Kim HJ, Kim SH. Diagnostic approach for double-hit and triple-hit lymphoma based on immunophenotypic and cytogenetic characteristics of bone marrow specimens. *Ann Lab Med*. 2020;40(5):361-369. doi:10.3343/alm.2020.40.5.361

42. Nowakowski GS, Czuczman MS. ABC, GCB, and Double-Hit Diffuse Large B-Cell Lymphoma: Does Subtype Make a Difference in Therapy Selection? *Am Soc Clin Oncol Educ B*. 2015;(35):e449-e457. doi:10.14694/edbook_am.2015.35.e449

43. Levine AJ, Vosburgh E. P53 mutations in lymphomas: Position matters. *Blood*. 2008;112(8):2997-2998. doi:10.1182/blood-2008-07-167718

44. Roschewski M, Phelan JD, Wilson WH. Molecular classification and treatment of diffuse large B-Cell lymphoma and primary mediastinal B-cell lymphoma. *Cancer J (United States)*. 2020;26(3):195-205. doi:10.1097/PPO.0000000000000450

45. Besenval M, Sessa C, Marty M, et al. Phase I clinical and pharmacokinetic studies of the taxoid derivative RPR 109881A administered as a 1-hour or a 3-hour infusion in patients with advanced solid tumors. *Ann Oncol*. 2002;13(7):1140-1150. doi:10.1093/annonc/mdf174

46. Ananiev J, Tchernev G, Patterson JW, Gulubova M. P 53 – “the guardian of genome.” 2011;(February).

47. Malkin D. Li-fraumeni syndrome. *Genes and Cancer*. 2011;2(4):475-484. doi:10.1177/1947601911413466

48. Monti P, Menichini P, Speciale A, et al. Heterogeneity of TP53 Mutations and P53 Protein Residual Function in Cancer: Does It Matter? *Front Oncol*. 2020;10(October):1-8. doi:10.3389/fonc.2020.593383

49. Pavletich NP, Chambers KA, Pabo CO. The DNA-binding domain of p53 contains the four conserved regions and the major mutation hot spots. *Genes Dev*. 1993;7(12 PART B):2556-2564. doi:10.1101/gad.7.12b.2556

50. Kulesza M, Agnieszka DM, Pieńkowska-Grela B. Repair or perish - The role of p53 protein in a cell's life. *Nowotwory*. 2019;69(5-6):168-178. doi:10.5603/NJO.2019.0031

51. Young KH, Leroy K, Møller MB, et al. Structural profiles of TP53 gene mutations predict clinical outcome in diffuse large B-cell lymphoma: An international collaborative study. *Blood*. 2008;112(8):3088-3098. doi:10.1182/blood-2008-01-129783

52. Xu-Monette ZY, Wu L, Visco C, et al. Mutational profile and prognostic significance of TP53 in diffuse large B-cell lymphoma patients treated with R-CHOP: Report from an International DLBCL Rituximab-CHOP Consortium Program Study. *Blood*. 2012;120(19):3986-3996. doi:10.1182/blood-2012-05-433334

53. Barrans SL, Evans PAS, O'Connor SJM, et al. The t(14;18) is associated with germinal center-derived diffuse large B-cell lymphoma and is a strong predictor of outcome. *Clin Cancer Res*. 2003;9(6):2133-2139.

54. Yu L, Yu TT, Young KH. Cross-talk between Myc and p53 in B-cell lymphomas. *Chronic Dis Transl Med*. 2019;5(3):139-154. doi:10.1016/j.cdtm.2019.08.001

55. Bouroumeau A, Bussot L, Bonnefoix T, et al. c-MYC and p53 expression highlight starry-sky pattern as a favourable prognostic feature in R-CHOP-treated diffuse large B-cell lymphoma. *J Pathol Clin Res*. 2021;7(6):604-615. doi:10.1002/cjp2.223

56. Surget S, Khoury MP, Bourdon JC. Uncovering the role of p53 splice variants in human malignancy: A clinical perspective. *Onco Targets Ther*. 2013;7:57-67. doi:10.2147/OTT.S53876

57. Voropaeva EN, Pospelova TI, Voevoda MI, Maksimov VN, Orlov YL, Seregina OB. Clinical aspects of TP53 gene inactivation in diffuse large B-cell lymphoma. *BMC Med Genomics*. 2019;12(Suppl 2):35-44. doi:10.1186/s12920-019-0484-9

58. de Andrade, K.C., Lee, E.E., Tookmanian EM et al. The TP53 Database. Cell Death Differ. doi:https://doi.org/10.1038/s41418-022-00976-3

59. UMD TP53 MUTATION DATABASE. 2017. https://p53.fr/download-the-database

60. Rodrigues JM, Hassan M, Freiburghaus C, et al. p53 is associated with high-risk and pinpoints TP53 missense mutations in mantle cell lymphoma. *Br J Haematol*. 2020;191(5):796-805. doi:10.1111/bjh.17023

61. Peroja P, Pedersen M, Mantere T, et al. Mutation of TP53, translocation analysis and immunohistochemical expression of MYC, BCL-2 and BCL-6 in patients with DLBCL treated with R-CHOP. *Sci Rep*. 2018;8(1):1-9. doi:10.1038/s41598-018-33230-3

62. Tamimi Y, Al-Harthy S, Al-Haddabi I, et al. The p53 Mutation/Deletion Profile in a Small Cohort of the Omani Population with Diffuse Large B-Cell Lymphoma. *Sultan Qaboos Univ Med J*. 2014;14(February):50-58. doi:10.12816/0003336

63. ATSUSHI I, TOMOHIRO K, TAKASHI W, et al. m u tat i o n s o f t h e p 5 3 g e n e as a p ro g no st i c fac to r i n ag g r e s s i v e b - c e l l ly m p h o m a mutations of the p53 gene as a prognostic factor. *N Engl J Med*. 1997;337(8):529-534. doi:https://doi.org/10.1056/nejm199708213370804

64. Qin Y, Jiang S, Liu P, et al. Characteristics and management of tp53-mutated diffuse large b-cell lymphoma patients. *Cancer Manag Res*. 2020;12:11515-11522. doi:10.2147/CMAR.S269624

65. Bobillo S, Joffe E, Lavery JA, et al. Clinical characteristics and outcomes of extranodal stage I diffuse large B-cell lymphoma in the rituximab era. *Blood*. 2021;137(1):39-48. doi:10.1182/blood.2020005112

66. Release S. Mid-year population estimates. 2025;(July).

67. Liu XF, Zhang H, Zhu SG, et al. Correlation of p53 gene mutation and expression of P53 protein in cholangiocarcinoma. *World J Gastroenterol*. 2006;12(29):4706-4709. doi:10.3748/wjg.v12.i29.4706

68. Stengel A, Schnittger S, Weissmann S, et al. TP53 mutations occur in 15.7% of ALL and are associated with MYC-rearrangement, low hypodiploidy, and a

poor prognosis. *Blood*. 2014;124(2):251-258. doi:10.1182/blood-2014-02-558833

69. Chang CC, Liu YC, Cleveland RP, Perkins SL. Expression of c-Myc and p53 correlates with clinical outcome in diffuse large B-cell lymphomas. *Am J Clin Pathol*. 2000;113(4):512-518. doi:10.1309/yhfe-r65b-d3lk-3ggv

70. Chan A, Dogan A. Prognostic and Predictive Biomarkers in Diffuse Large B-cell Lymphoma. *Surg Pathol Clin*. 2019;12(3):699-707. doi:10.1016/j.path.2019.03.012

71. Stefancikova L, Moulis M, Fabian P, et al. Prognostic impact of p53 aberrations for R-CHOP-treated patients with diffuse large B-cell lymphoma. *Int J Oncol*. 2011;39(6):1413-1420. doi:10.3892/ijo.2011.1170

72. Xu-Monette ZY, Jeffrey Medeiros L, Li Y, et al. Dysfunction of the TP53 tumor suppressor gene in lymphoid malignancies. *Blood*. 2012;119(16):3668-3683. doi:10.1182/blood-2011-11-366062

73. Xie Y, Ajaz Bulbul M, Ji L, et al. P53 expression is a strong marker of inferior survival in de novo diffuse large B-cell lymphoma and may have enhanced negative effect with MYC coexpression: A Single Institutional Clinicopathologic Study. *Am J Clin Pathol*. 2014;141(4):593-604. doi:10.1309/AJCPPHMZ6VHF0WQV

74. Leroy B, Anderson M, Soussi T. TP53 mutations in human cancer: Database reassessment and prospects for the next decade. *Hum Mutat*. 2014;35(6):672-688. doi:10.1002/humu.22552

75. Human mutation. Published online 2017. https://p53.fr/tp53-database/mutation-database

76. Kennedy MC, Lowe SW. Mutant p53: it's not all one and the same. *Cell Death Differ*. 2022;29(5):983-987. doi:10.1038/s41418-022-00989-y

77. Kristensen AT, Bjørheim J, Minarik M, Giercksky KE, Ekstrøm PO. Detection of mutations in exon 8 of TP53 by temperature gradient 96-capillary array electrophoresis. *Biotechniques*. 2002;33(3):650-653. doi:10.2144/02333pf01

78. Murray J, Davies T, Rees D. Occupational lung disease in the South African

mining industry: Research and policy implementation. *J Public Health Policy*. 2011;32(SUPPL. 1). doi:10.1057/jphp.2011.25

79. Rossi D, Khiabanian H, Spina V, et al. Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia. *Blood*. 2014;123(14):2139-2147. doi:10.1182/blood-2013-11-539726

80. Young KH, Weisenburger DD, Dave BJ, et al. Mutations in the DNA-binding codons of TP53, which are associated with decreased expression of TRAIL receptor-2, predict for poor survival in diffuse large B-cell lymphoma. *Blood*. 2007;110(13):4396-4405. doi:10.1182/blood-2007-02-072082

81. Ahn Y, Jun Y. Measurement of pain-like response to various NICU stimulants for high-risk infants. *Early Hum Dev*. 2007;83(4):255-262. doi:10.1016/j.earlhumdev.2006.05.022

82. Ahn Y, Jun Y. Measurement of pain-like response to various NICU stimulants for high-risk infants. *Early Hum Dev*. 2007;83(4):255-262. doi:10.1016/j.earlhumdev.2006.05.022

83. Pandey R, Johnson N, Cooke L, et al. Tp53 mutations as a driver of metastasis signaling in advanced cancer patients. *Cancers (Basel)*. 2021;13(4):1-21. doi:10.3390/cancers13040597

84. Soussi T, Leroy B, Taschner PEM. Recommendations for analyzing and reporting TP53 gene variants in the high-throughput sequencing era. *Hum Mutat*. 2014;35(6):766-778. doi:10.1002/humu.22561

Appendix A: Ethical clearance certificate

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R49 Ms T Mashele

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M220365

NAME: (Principal Investigator)	Ms T Mashele
DEPARTMENT:	School of Pathology Department of Anatomical Pathology Medical School University
PROJECT TITLE:	*Immunophenotypic and molecular features of TP53 in diffuse large B-cell lymphoma (DLBCL).*
DATE CONSIDERED:	2022/03/25
DECISION:	Approved unconditionally
CONDITIONS:	
NOTE:	If contact information regarding student study participants is required, please contact the Registrar's office - <Nicoleen.Potgieter@wits.ac.za>
SUPERVISOR:	Drs P Magangane and S Pather
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	2022/10/20

This Clearance Certificate is valid for 5 years from the date of approval. An extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office secretariat on the 3rd floor, Phillip Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated from the research protocol as approved, I/we undertake to submit details to the Committee. **I agree to submit a yearly progress report**. When a funder requires annual re-certification, the application date will be one year after the date when the study was initially reviewed. In this case, the study was initially reviewed in **March** and therefore reports and re-certification will be due in the month of **March** each year. Unreported changes to the study may invalidate the clearance given by the HREC (Medical).

Signature of Principal Investigator — Date: 21/10/2022

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG | HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

Office of the Deputy Vice-Chancellor (Research and Innovation)

TO:	Ms T Mashele School of Pathology Department of Anatomical Pathology Medical School University E-mail: thembi.hl.mashele@gmail.com
CC:	Supervisor: Drs P Magangane and S Pather <Pumza.Magangane@wits.ac.za> and <HREC-Medical Research Office@wits.ac.za>
FROM:	Mr Iain Burns Human Research Ethics Committee (Medical) Tel: 011 717 1252 E-mail: Iain.Burns@wits.ac.za
DATE:	2022/10/20
REF:	R14/49
PROTOCOL NO:	**M220365** (This is your ethics application reference number. Please quote it in all enquiries, oral or written, relating to this study.)
PROJECT TITLE:	*Immunophenotypic and molecular features of TP53 in diffuse large B-cell lymphoma (DLBCL).*

Please find attached the Clearance Certificate for the above project. I hope it goes well and that an article in a recognized publication comes out of it. This will reflect well on your professional standing and contribute to Government funding of the University.

MSWorks2000/Iain0007/Clearscan.wps

Appendix B: Demographic and immunoexpression details

Study #	AGE (YEARS)	SEX	SITE	HIV CODE	BCL2 (≥50%)	BCL6 (≥30%)	CD10 (≥30%)	KI67 (≥90%)	MUM1 (≥30%)	COO	C-myc (≥50%)	p53 (≥50%)	SURV OUT(A/D/U)	DURATION OF SURVIVAL
ND-01	58	Female	Nodal	Positive	3	3	2	1	1	ABC	2	2	Missing	Missing
ND-02	44	Male	Nodal	Positive	1	1	1	2	1	NOS	2	1	Demised	31 days
ND-03	78	Female	Extranodal	Negative	2	2	1	2	1	NOS	2	1	Missing	Missing
ND-04	34	Female	Extranodal	Positive	2	1	2	1	1	ABC	1	1	Alive	225 days
ND-05	48	Female	Nodal	Negative	1	2	2	1	1	ABC	1	1	Demised	825 days
ND-07	38	Male	Extranodal	Unknown	1	1	2	1	1	ABC	2	1	Missing	Missing
ND-08	54	Female	Extranodal	Unknown	2	1	1	2	1	NOS	2	1	Missing	Missing
ND-09	32	Female	Nodal	Positive	2	1	1	1	2	GCB	2	1	Missing	Missing
ND-10	32	Male	Nodal	Negative	2	1	2	1	1	ABC	1	2	Missing	Missing
ND-11	43	Male	Extranodal	Positive	2	1	1	1	2	GCB	1	1	Demised	11 days
ND-12	34	Male	Extranodal	Positive	2	1	2	1	1	ABC	1	2	Alive	64 days
ND-14	40	Male	Extranodal	Positive	2	1	2	1	1	ABC	3	2	Missing	Missing
ND-15	65	Male	Nodal	Negative	1	3	2	1	1	ABC	3	1	Alive	252 days
ND-16	42	Female	Extranodal	Positive	2	1	2	1	1	ABC	2	1	missing	Missing
ND-18	73	Female	Extranodal	Unknown	2	2	1	1	1	NOS	1	1	Missing	Missing
ND-19	88	Male	Extranodal	Negative	1	2	2	1	1	ABC	2	2	Alive	28 days
ND-20	49	Male	Nodal	Negative	1	1	1	1	2	GCB	1	1	Missing	Missing
ND-21	41	Female	Nodal	Positive	1	1	1	1	2	GCB	2	2	alive	1 819 days
ND-23	65	Male	Extranodal	Positive	1	1	2	1	1	ABC	1	1	Demised	15 days
ND-24	56	Male	Extranodal	Negative	1	1	2	1	1	ABC	2	2	Demised	593 days
ND-25	36	Female	Extranodal	Positive	2	1	1	1	2	GCB	1	1	missing	Missing
ND-26	43	Male	Extranodal	Positive	2	1	2	1	1	ABC	1	2	Demised	101 days
ND-28	45	Female	Nodal	Positive	1	1	1	1	2	GCB	1	2	Demised	63 days

ND-29	46	Male	Extranodal	Positive	2	1	2	1	2	GCB	1	2	Demised	16 days
ND-30	46	Male	Nodal	Positive	1	2	2	1	3	GCB	1	1	Demised	14 days
ND-31	47	Male	Nodal	Positive	2	1	1	1	2	GCB	1	1	Missing	Missing
ND-34	34	Female	Extranodal	Positive	2	1	1	1	2	GCB	2	2	Missing	Missing
ND-35	52	Male	Nodal	Positive	2	1	1	1	2	GCB	2	2	Missing	Missing
ND-36	36	Male	Nodal	Positive	1	1	1	1	1	GCB	2	2	Missing	Missing
ND-37	22	Female	Extranodal	Positive	2	1	1	1	1	GCB	2	1	Missing	Missing
ND-40	38	Female	Nodal	Positive	2	3	3	3	3	NOS	3	2	missing	Missing
ND-41	38	Female	Extranodal	Positive	1	2	1	1	2	ABC	1	2	Demised	398 days
ND-42	44	Male	Extranodal	Positive	2	1	2	1	2	GCB	1	2	Demised	149 days
ND-44	41	Female	Extranodal	Positive	1	1	1	2	1	GCB	1	2	alive	763 days
ND-45	48	Male	Nodal	Positive	2	2	1	1	2	GCB	1	2	alive	1 743 days
ND-46	47	Male	Extranodal	Positive	2	2	1	1	1	GCB	1	1	alive	175 days
ND-47	76	Male	Nodal	Negative	1	2	2	1	2	GCB	2	2	Demised	217 days
ND-48	42	Female	Extranodal	Positive	2	1	2	1	2	GCB	2	2	Demised	849 days
ND-50	51	Male	Nodal	Positive	1	1	2	1	1	GCB	1	2	Demised	320 days
ND-51	45	Male	Nodal	Negative	2	2	2	2	3	ABC	3	2	Missing	Missing
ND-52	56	Male	Nodal	Positive	3	2	1	2	2	ABC	2	2	Alive	172 days
ND-53	68	Male	Nodal	Negative	1	2	2	2	1	ABC	2	1	Alive	804 days
ND-55	42	Female	Nodal	Positive	1	1	1	1	2	GCB	2	1	Demised	109 days
ND-56	38	Male	Extranodal	Positive	1	1	2	1	1	ABC	1	1	Alive	76 days
ND-58	55	Male	Nodal	Positive	1	2	2	2	1	ABC	2	2	Demised	530 days
ND-59	54	Female	Extranodal	Positive	2	1	1	1	2	GCB	2	1	Demised	257 days
ND-60	34	Male	Extranodal	Positive	1	2	2	1	1	ABC	1	1	Demised	74 days
ND-61	46	Male	Extranodal	Positive	2	1	1	1	1	GCB	1	1	Demised	325 days
ND-62	40	Male	Nodal	Positive	2	1	1	1	1	GCB	1	2	Alive	1 382 days
ND-63	53	Female	Nodal	Positive	1	1	2	1	1	ABC	2	2	Demised	96 days

ND-64	55	Male	Nodal	Positive	1	1	1	1	1	GCB	2	1	Demised	752 days
ND-65	35	Female	Extranodal	Positive	1	2	2	1	1	ABC	1	2	Demised	94 days
ND-66	46	Female	Extranodal	Positive	2	1	3	1	1	GCB	1	2	Alive	14 days
ND-67	42	Female	Nodal	Positive	2	1	2	1	2	GCB	1	2	Alive	283 days
ND-68	58	Female	Nodal	Positive	2	1	1	1	1	GCB	1	1	Alive	150 days
ND-69	38	Female	Extranodal	Negative	2	1	2	1	1	ABC	1	2	Alive	196 days
ND-70	77	Male	Extranodal	Negative	1	2	2	2	1	ABC	1	2	Demised	180 days
ND-71	45	Female	Nodal	Positive	1	1	2	1	1	ABC	1	2	Demised	27 days
ND-72	39	Male	Nodal	Positive	2	2	2	1	1	ABC	1	2	Demised	11 days
ND-73	38	Female	Extranodal	Positive	2	1	1	1	1	NOS	1	1	Demised	20 days

1=Expressed, 2=Not expressed and 3=Unavailable

Appendix C: DNA quantity using a Nanodrop

STUDY NUMBER	ng/UL	A260/A280	A260/A230
ND-01	0.6	-0.64	0.02
ND-02	123.5	1.91	1.83
ND-03	158.6	2.01	1.84
ND-04	19.8	1.94	0.81
ND-05	2.7	22.5	0.09
ND-07	89.2	1.92	1.55
ND-08	963.2	1.61	0.62
ND-09	243.2	2.01	1.95
ND-10	1.9	2.91	0.15
ND-11	29.3	2.01	0.88
ND-12	8.4	2.14	0.50
ND-14	2.2	2.55	0.11
ND-15	47.9	2.09	1.14
ND-16	84.1	1.97	1.33
ND-18	21.7	2.31	0.69
ND-19	17.2	2.23	0.57
ND-20	-	-	-
ND-21	2.7	5.71	0.08
ND-23	11.2	2.04	0.66
ND-24	1.4	1.96	0.07
ND-25	-	-	-
ND-26	2.4	2.46	0.10
ND-28	352.1	2.00	1.97
ND-29	-	-	-
ND-30	-	-	-
ND-31	133.4	2.07	1.63
ND-34	-	-	-
ND-35	-	-	-
ND-36	-	-	-
ND-37	3.5	2.61	0.16
ND-40	512.4	2.00	2.15
ND-41	1.1	1.13	0.05

ND-42	-	-	-
STUDY NUMBER	**ng/UL**	**A260/A280**	**A260/A230**
ND-44	3.1	2.55	0.13
ND-45	15.0	2.07	0.57
ND-46	-	-	-
ND-47	9.3	2.69	3.72
ND-48	8.2	1.91	0.40
ND-50	294.6	2.02	2.00
ND-51	-	-	-
ND-52	1.9	7.25	0.07
ND-53	405.1	2.02	2.14
ND-55	146.8	1.96	1.79
ND-56	4.6	2.29	0.27
ND-58	936.5	1.96	2.19
ND-59	414.2	2.01	2.13
ND-60	13.9	1.97	0.70
ND-61	751.6	1.94	2.12
ND-62	2.9	2.11	0.15
ND-63	1 206.8	1.97	2.21
ND-64	1.0	1.11	0.04
ND-65	1.1	8.15	0.05
ND-66	19.0	1.92	0.69
ND-67	808.7	1.99	2.21
ND-68	9.8	2.16	0.40
ND-69	-	-	-
ND-70	4.2	2.37	0.26
ND-71	11.2	2.12	0.43
ND-72	992.6	1.99	2.17
ND-73	10.2	2.04	0.46

Appendix D: Primer blast information of the complete TP53 gene

>U94788.1 Human p53 (TP53) gene, complete cds
TTCCCATCAAGCCCTAGGGCTCCTCGTGGCTGCTGGGAGTTGTAGTCTGAACGCTTCTATCTTGGCGAGAAGCGCCTACG
CTCCCCCTACCGAGTCCCGCGGTAATTCTTAAAGCACCTGCACCGCCCCCCCGCCGCCTGCAGAGGGCGCAGCAGGTCTT
GCACCTCTTCTGCATCTCATTCTCCAGGCTTCAGACCTGTCTCCCTCATTCAAAAAATATTTATTATCGAGCTCTTACTT
GCTACCCAGCACTGATATAGGCACTCAGGAATACAACAATGAATAAGATAGTAGAAAAATTCTATATCCTCATAAGGCTT
ACGTTTCCATGTACTGAAAGCAATGAACAAATAAATCTTATCAGAGTGATAAGGGTTGTGAAGGAGATTAAATAAGATGG
TGTGATATAAAGTATCTGGGAGAAAACGTTAGGGTGTGATATTACGGAAAGCCTTCCTAAAAAATGACATTTTAACTGAT
GAGAAGAAAGGATCCAGCTGAGAGCAAACGCAAAAGCTTTCTTCCTTCCACCCTTCATATTTGACACAATGCAGGATTCC
TCCAAAATGATTTCCACCAATTCTGCCCTCACAGCTCTGGCTTGCAGAATTTTCCACCCCAAAATGTTAGTATCTACGGC
ACCAGGTCGGCGAGAATCCTGACTCTGCACCCTCCTCCCCAACTCCATTTCCTTTGCTTCCTCCGGCAGGCGGATTACTT
GCCCTTACTTGTCATGGCGACTGTCCAGCTTTGTGCCAGGAGCCTCGCAGGGGTTGATGGGATTGGGGTTTTCCCCTCCC
ATGTGCTCAAGACTGGCGCTAAAAGTTTTGAGCTTCTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGG
CTCCGGGGACACTTTGCGTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCCTGGATTGGGTAAGCTCCTG
ACTGAACTTGATGAGTCCTCTCTGAGTCACGGGCTCTCGGCTCCGTGTATTTTCAGCTCGGGAAAATCGCTGGGGCTGGG
GGTGGGGCAGTGGGGACTTAGCGAGTTTGGGGGTGAGTGGGATGGAAGCTTGGCTAGAGGGATCATCATAGGAGTTGCAT
TGTTGGGAGACCTGGGTGTAGATGATGGGGATGTTAGGACCATCCGAACTCAAAGTTGAACGCCTAGGCAGAGGAGTGGA
GCTTTGGGGAACCTTGAGCCGGCCTAAAGCGTACTTCTTTGCACATCCACCCGGTGCTGGGCGTAGGGAATCCCTGAAAT
AAAAGATGCACAAAGCATTGAGGTCTGAGACTTTTGGATCTCGAAACATTGAGAACTCATAGCTGTATATTTTAGAGCCC
ATGGCATCCTAGTGAAAACTGGGGCTCCATTCCGAAATGATCATTTGGGGGTGATCCGGGGAGCCCAAGCTGCTAAGGTC
CCACAACTTCCGGACCTTTGTCCTTCCTGGAGCGATCTTTCCAGGCAGCCCCCGGCTCCGCTAGATGGAGAAAATCCAAT
TGAAGGCTGTCAGTCGTGGAAGTGAGAAGTGCTAAACCAGGGGTTTGCCCGCCAGGCCGAGGAGGACCGTCGCAATCTGA
GAGGCCCGGCAGCCCTGTTATTGTTTGGCTCCACATTTACATTTCTGCCTCTTGCAGCAGCATTTCCGGTTTCTTTTTGC
CGGAGCAGCTCACTATTCACCCGATGAGAGGGGAGGAGAGAGAGAGAAAATGTCCTTTAGGCCGGTTCCTCTTACTTGGC
AGAGGGAGGCTGCTATTCTCCGCCTGCATTTCTTTTTCTGGATTACTTAGTTATGGCCTTTGCAAAGGCAGGGGTATTTG
TTTTGATGCAAACCTCAATCCCTCCCCTTCTTTGAATGGTGTGCCCCACCCCCCGGGTCGCCTGCAACCTAGGCGGACGC
TACCATGGCGTAGACAGGGAGGGAAAGAAGTGTGCAGAAGGCAAGCCCGGAGGCACTTTCAAGAATGAGCATATCTCATC
TTCCCGGAGAAAAAAAAAAAAAGAATGGTACGTCTGAGAATGAAATTTTGAAAGAGTGCAATGATGGGTCGTTTGATAATT
TGTCGGGAAAAACAATCTACCTGTTATCTAGCTTTGGGCTAGGCCATTCCAGTTCCAGACGCAGGCTGAACGTCGTGAAG
CGGAAGGGGCGGGCCCGCAGGCGTCCGTGTGGTCCTCCGTGCAGCCCTCGGCCCGAGCCGGTTCTTCCTGGTAGGAGGCG
GAACTCGAATTCATTTCTCCCGCTGCCCCATCTCTTAGCTCGCGGTTGTTTCATTCCGCAGTTTCTTCCCATGCACCTGC
CGCGTACCGGCCACTTTGTGCCGTACTTACGTCATCTTTTTCCTAAATCGAGGTGGCATTTACACACAGCGCCAGTGCAC
ACAGCAAGTGCACAGGAAGATGAGTTTTGGCCCCTAACCGCTCCGTGATGCCTACCAAGTCACAGACCCTTTTCATCGTC
CCAGAAACGTTTCATCACGTCTCTTCCCAGTCGATTCCCGACCCCACCTTTATTTTGATCTCCATAACCATTTTGCCTGT
TGGAGAACTTCATATAGAATGGAATCAGGATGGGCGCTGTGGCTCACGCCTGCACTTTGGCTCACGCCTGCACTTTGGGA
GGCCGAGGCGGGCGGATTACTTGAGGATAGGAGTTCCAGACCAGCGTGGCCAACGTGGTGAATCCCCGTCTCTACTAAAA
AATACAAAAATTAGCTGGGCGTGGTGGGTGCCTGTAATCCCAGCTATTCGGGAGGGTGAGGCAGGAGAATCGCTTGAACC
CGGGAGGCAGAGGTTGCAGTGAGCCAAGATCGTGCCACTACACTCCAGCCTGGGCGACAAGAACGAAACTCCGTCTCAAA
AAAAAGGGGGGAATCATACATTATGTGCTCATTTTTGTCGGGCTTCTGTCCTTCAATGTACTGTCTGACATTCGTTCATG
TTGTATATATCAGTATTTTGCTCCTTTTCATTTAGTATAGTCCATCGATTGTATATCCGTCCTTTTGATGGCCTTTTGAG
TTGTTTCCCATTTGCGGTTATGAAATAAAGCTGCTATAAACATTCTTGTACAATTCTTTTTGTGATCATATGTTTTCGTG
TTTCTTGGAGAAATACTTAGGAGGGGAATTGTGGAGGAAGTAAAAAGTAGCTGTATTTTGAACTTTTTCAGAAGCTCTGA
GTTTTCCAGAGCGGTTGTACCATTTTACACTCCAACTAGCAAGGTATGGGAGTTATTATGGTTGTGCCACAGCCTTCCGG
ACATTAGGTATGTCAGTCTTTCTAATGTGGTATATCCTTGTGGTTGTAATTTACAGTTCTCTATTGACTAAGGATGTTCA
GCATTTTTTCATGTGCCTATTGGCCATTCGTATTTTGTTTGTAAAGTAGCTCTTCGAGTCTTTTACCTGTTATTTTGGTT
TTTTTGTTTGTTTTTATTGTTCAGTTGTGGGACTGCTTTATACTTTCTGGATACAAGTCCTTATCAGATCCATGAGTCGT
GAATGTTTTCTTCTGATCTGTTGCGGGCCTATTTGTTTGCTTTACAGAGTTTACAGAATCTTAAGAGGAGTGGATTAATC
TTTTTTATGTTCAGTATTTGCCTTGTCCTGTTTAGGACATCTTTTTTTTTTTTTTTTAACCCCAGGGTCATGAAGATATAA
TCTTACATTTTCTTTTAGGACCTTTATGGTGGTAAGTTTTACAGTAAGGTCCTTAAGCCATTAATTAATTCTTAAAATTA
ATTGTTTATGGTGTGAGGTGTAGGAGTCAGTCTCTGGTATCTTTCCTGTATGGAAATCCAGTTATTCTGTCTCCACTTGT
TGAAATAGGCTTCCTTTCTCTACTGAATGCTTTTAATTTTAATTATTTTACAGTTGGAGTATAGGGCTACCATTTTAGTG

TAGGGGCTTGGTTCTGGGACTCCTCACACACCAAACCCACAGATGTCTAAGTCCCATATATAAGACGGAATAGTATTTAA
CCTACACATATCCTCCCATATAGTTTAAATTATCTAGATTACTTACATTACCCCCATACAATGAAAATGCTAATGTACAT
GCAAGTATGTATGTAAGTACTTGTACTATATTGTTTAGGGAATCACTGGACAGATAGGCCTTCAAGACTGATACCAGCAG
CCACTGTTAAGATTCTGGTCAGGCCTGCCCCTGTTTGGGGTCTCAGTTGATCTCATTGCCTTCCCACCCAGCCAAGGGCA
CCTGCATTTCTCTTGGCTCCCTGGCCATTTGGAAGGCCTAGTTCAGCCTGGCACATTTGTATCCTGGCCCACTGATGCTG
GTACCCCTGGGAAGGTCCTGCTCTGAAAAACACGGAGATTTTAGTTGCTACTGAAGATTTGAGAGATAAAGACAGGGAGA
CCTGTCTGTAGACCTGTGTCCCTCCAAGTGGGATTGAGACTTTGGGCCCCCCATTTCAGGACAGCACCTCCTGGCCTGTT
GACTGAATAGATCCCTGAAGGAGGTGTAGTTGCATTTTAGGAGTGGGGGTGGGAGCAGTACCACTGATCCGCACTAACAA
TCACACAGTTCTCTCTAGAATAATAATATAGAACAAGTGAAATAGAACAATTGCAGAAAGAGCTAACCTTTGTTGAGCTC
TTACTGTGTGCCCAGCACTTTCCTCAACTCTACATTTCCCATAATACATAGAGTACTAGGTAGGCGGGGCTTGGGGGCTC
ACGCCTGTAATCCCAGCACTTTAGGAGGCCAAGGGGGGTGGATCACCTGAGGTCGGGAGTTCAAGACCAGCCTGACTAAC
ATGGTGAAACCCCGTCTCTACTAGAAGTACAAAATTAGCCAGGTGTGGTGGCACATGCTTGTAGTCCTAGCTACTCAGCA
GGCTGAGGCAGGAGAATCATTTGAATCCGGGAGGAGGTTGCAGTAAGCGGAGATAGTGCCACTGTACTCCAGCCTGGGCA
ATAAGAGCTGAGACTCCGTCTCAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAAAAAGAAAAGAGCCT
GCCATTAAAGGAGCTGTTTGGTAGGGGATGTTTTGTCAGTGCAAACAACAGAAAAGTGGGCTGGGCACAGTGGTTCATGC
CTGTAATCCCAGCACTTTGGGAGGCCAAGGCGGGCGGATCACCTGAAGTTGGGAGTTCAAGACCAGCCTGACCAATATGG
AGAAACCCCGTCTCTACTAAAAATACAAAATTAGCCGGGCGCAGTGGCCGATGCCTGTAATCCCAGCTACTCGGGAGGCT
GAGGCAGGAGAATCGCTTGAACCTGGGAGGCAGAGGTTGCGGTGAGCCGAGATCGCACCATTGCACTCCAGCCTGGACGA
GAGCAAAACTCTGTCTCAAAAAAAAAAAAAAAACAGAAAAGTGTAACAAACACTTACAGTAGGCATGTTTCTTAGCAAATC
TGATGACAAATTTGGCATAAAGAAAGAGAGCATCCCTGAAAAAAAAAAAAAAGAAAAAGAAAGAGAGCATCCTGCCTGGGC
AACATAGTGAAACCCTGCCTCTACAAAAAAACTCAAAAATTGGCCGGGTGCAGTGGCTCACACCTGTAATCCCAGCACTT
TGGGAGTCGGAGGCGGGAGGATCACCTGAGGTCAGGAGTTCGAAACCAGCCTGGCCAACATGGCAAAACCCCATCTCTAC
TAAAAATACAAAAAATTAATCAGGCGCATTGGTGGGCGCCTGTAATCCCAGCTACTCAGGAAGTTGAGGCAAGAGGATCG
CTTGATACTGGGAGGTGGAGGTTACAGTGAGTCGAGATCACACCACTGCACTCTAGCCTGGGTGACAGGGCGAGACTCCG
TCTCCAAAAAAAAAAAGAAAAAGAAAAAGACTAAAAAATTAGCCAGGCAGGCCTCTGTGGTCCCAGCTACTTGGGAGGCT
GAGGCAGGAGAATCACTGAGCCCAGGAGTGGCAGGCTGTAGTGAGCCATGATTGCACCACTGTACCCTAGCTTGGGCTTC
AAAGCAAGACCCTGCCTCAAAAGAAAAAAGAAAGAAAGAAAGAACATGGCGGGCCAGGCACAGTGGCTCACACCTGTAAT
CCCAGCGCTTTGAGAGGCCGAGGCAGGTGGATCACAAGGTCAGGAGTTCCACACCAGCCTGGCCAACATGGTGAAACCCT
GTCTCTACTAAAAATACAAAAAATCAGCAGGCAGGGTGGTAGGGGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGG
AGAATTGCTTGAAACCAGAAGGCAGAGGTTGCAGTGAGCCTAGACTGCACCACTGCACTCCAGCCTGGGCGAAAAGAGCC
AAACTCCATCTCAAAAAACAAACAAAAAAACAAAACAAAAAAAAACATGGCAGCCTTTGAAAGCTTGTCTGGGAGAAGGT
GCGATGATGGTTGCATAACTTCGTGCAAGATGCTGGTCCACACAGGGGCTGCCCCTTGCTCTTTCTCGCTCTCTTAACCT
CTCATATAACAGGCTTGTGTGTTATGCACATTTATTGAGCCCAAGCAGGTGCAAGGCATTGTGATCTAATACTTTGGTCA
GCAAGACAACAAGATAGATCACTGCCCTGCCCTTAGGAAGTGTATATGCTATTAGAGGAAACAGATAAAATAAACAAGGA
AAAGTATCAGACAATGTAAGTGCTATGAGAATGCAAATGAGGTGATGTGAATTAAAATAGGATGACTTAAGTCTGCACGG
AAGGCCCCTACCCCCATGTTCCTGGCTAGCCAAGGAACCACCAGTTGATTAGCAGAGAAGGGCAGCCCGTCTAGCTAGAG
CTTTTGGGGAAGAGGGAGTGGTTGTTAAGAGATGAGATTAAAGAAGCCGAGACGGGCCCTTCGTGAGGGGGGGTTGTAAT
GCAGGGCTGAGGAGTGTCCGAAGAGAATGGGCAGGTGAGCGGTGAGACAGTTGTTCTTCCAGAAGCTTTGCAGTGAAAGG
AATCAAAGAAATGGAGCCGTGTATCAGGTGGGGAAGGGTGGGGGCCAAGGGGGTGTCCTTCCCCATACAGAGATTGCAGG
CTGAGAATGACTATATCCTTGTTAACAGGAGGTGGGAGCAGGGCACGGTAGCTCACACCTGTAATCTTGGCACTTTAGGA
GGCGGAGGCGGGCCGATCACCTGAAGTAAGGAGTTCGAGACCAGCCTGGCCAACATGCAAAGCCCTGTCTCTACTAAAAA
TACAAAAATTAGCTGGGTGTGGTGGTACTCGCCTGTAATCCCAGCTACTCGGGAGACTGAGGCAGGAGAATGGCTTGAAC
CCGGAAGGTAGAGGTTGCAGTGAGCTGAGATCATGCCACTGTGCTCCAGCCTAGGTGACAGAGAGAGACTCCATCTCAAA
AAAAAAAAAAAATACAGGAAGGGAGTTGGGAATAGGGTGCACATTTAGGAAGTCTTGGGGATTTAGTGGTGGGAAGGTTG
GAAGTCCCTCTCTGATTGTCTTTTCCTCAAAGAAGTGCATGGCTGGTGTGGGGTGGGGCAGGAGTGCTTGGGTTGTGGTG
AAACATTGGAAGAGAGAATGTGAAGCAGCCATTCTTTTCCTGCTCCACAGGAAGCCGAGCTGTCTCAGACACTGGCATGG
TGTTGGGGGAGGGGGTTCCTTCTCTGCAGGCCCAGGTGACCCAGGGTTGGAAGTGTCTCATGCTGGATCCCCACTTTTCC
TCTTGCAGCAGCCAGACTGCCTTCCGGGTCACTGCCATGGAGGAGCCGCAGTCAGATCCTAGCGTCGAGCCCCCTCTGAG
TCAGGAAACATTTTCAGACCTATGGAAACTGTGAGTGGATCCATTGGAAGGGCAGGCCACCACCCCGACCCCAACCCCAG
CCCCCTAGCAGAGACCTGTGGGAAGCGAAAATTCATGGGACTGACTTTCTGCTCTTGTCTTTCAGACTTCCTGAAAACAA

CGTTCTGGTAAGGACAAGGGTTGGGCTGGGGACCTGGAGGGCTGGGGGGCTGGGGGGCTGAGGACCTGGTCCTCTGACTG
CTCTTTTCACCCATCTACAGTCCCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCCGGACGATATTGAACAA
TGGTTCACTGAAGACCCAGGTCCAGATGAAGCTCCCAGAATGCCAGAGGCTGCTCCCCGCGTGGCCCCTGCACCAGCAGC
TCCTACACCGGCGGCCCCTGCACCAGCCCCCTCCTGGCCCCTGTCATCTTCTGTCCCTTCCCAGAAAACCTACCAGGGCA
GCTACGGTTTCCGTCTGGGCTTCTTGCATTCTGGGACAGCCAAGTCTGTGACTTGCACGGTCAGTTGCCCTGAGGGGCTG
GCTTCCATGAGACTTCAATGCCTGGCCGTATCCCCCTGCATTTCTTTTGTTTGGAACTTTGGGATTCCTCTTCACCCTTA
GGCTTCCTGTCAGTGTTTTTTTATAGTTTACCCACTTAATGTGTGATCTCTGACTCCTGTCCCAAAGTTGAATATTCCCC
CCTTGAATTTGGGCTTTTATCCATCCCATCACACCCTCAGCATCTCTCCTGGGGATGCAGAACTTTTCTTTTTCTTCATC
CACGTGTATTCCTTGGCTTTTGAAAATAAGCTCCTGACCAGGCTTGGTGGCTCACACCTGCAATCCCAGCACTCTCAAAG
AGGCCAAGGCAGGCAGATCACCTGAGCCCCAGGAGTTCAAGACCAGCCTGGGTAACATGATGAAACCTCGTCTCTACAAA
AAAATACAAAAAATTAGCCAGGCATGGTGGTGCACACCTATAGTCCCAGCCACTCAGGAGGCTGAGGTGGGAAGATCACT
TGAGGCCAGGAGATGGAGGCTGCAGTGAGCTGTGATCACACCACTGTGCTCCAGCCTGAGTGACAGAGCAAGACCCTATC
TCAAAAAAAAAAAAAAAAGAAAAGCTCCTGAGGTGTAGACGCCAACTCTCTCTAGCTCGCTAGTGGGTTGCAGGAGGTGCT
TACACATGTTTGTTTCTTTGCTGCCGTGTTCCAGTTGCTTTATCTGTTCACTTGTGCCCTGACTTTCAACTCTGTCTCCT
TCCTCTTCCTACAGTACTCCCCTGCCCTCAACAAGATGTTTTGCCAACTGGCCAAGACCTGCCCTGTGCAGCTGTGGGTT
GATTCCACACCCCCGCCCGGCACCCGCGTCCGCGCCATGGCCATCTACAAGCAGTCACAGCACATGACGGAGGTTGTGAG
GCGCTGCCCCCACCATGAGCGCTGCTCAGATAGCGATGGTGAGCAGCTGGGGCTGGAGAGACGACAGGGCTGGTTGCCCA
GGGTCCCCAGGCCTCTGATTCCTCACTGATTGCTCTTAGGTCTGGCCCCTCCTCAGCATCTTATCCGAGTGGAAGGAAAT
TTGCGTGTGGAGTATTTGGATGACAGAAACACTTTTCGACATAGTGTGGTGGTGCCCTATGAGCCGCCTGAGGTCTGGTT
TGCAACTGGGGTCTCTGGGAGGAGGGGTTAAGGGTGGTTGTCAGTGGCCCTCCGGGTGAGCAGTAGGGGGGCTTTCTCCT
GCTGCTTATTTGACCTCCCTATAACCCCATGAGATGTGCAAAGTAAATGGGTTTAACTATTGCACAGTTGAAAAAACTGA
AGCTTACGAGGCTAAGGGCCTCCCCTGCTTGGCTGGGCGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAA
GGCAGGCGGATCACGAGGTTGGGAGATCGAGACCATCCTGGCTAACGGTGAAACCCCGTCTCTACTGAAAAATACAAAAA
AAAATTAGCCGGGCGTGGTGCTGGGCACCTGTAGTCCCAGCTACTCGGGAGGCTGAGGAAGGAGAATGGCGTGAACCTGG
GCGGTGGAGCTTGCAGTGAGCTGAGATCACGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGATTCCATCTCAAAAAAA
AAAAAAAAAGGCCTCCCCTGCTTGCCACAGGTCTCCCCAAGGCGCACTGGCCTCATCTTGGGCCTGTGTTATCTCCTAGG
TTGGCTCTGACTGTACCACCATCCACTACAACTACATGTGTAACAGTTCCTGCATGGGCGGCATGAACCGGAGGCCCATC
CTCACCATCATCACACTGGAAGACTCCAGGTCAGGAGCCACTTGCCACCCTGCACACTGGCCTGCTGTGCCCCAGCCTCT
GCTTGCCGCTGACCCCTGGGCCCACCTCTTACCGATTTCTTCCATACTACTACCCATCCACCTCTCATCACATTTCCGGC
GGGAATCTCCTTACTGCTCCCACTCAGTTTCCTTTTCTCTGGCTTTGGGACCTCTTAACCTGTGGCTTCTCCTCCCACCT
CCTGGAGCTGGAGCTTAGGCTCCAGAAAGGACAAGGGTGGTTGGGAGTAGATGGAGCCTGGTTTTTTAAATGGGACAGGT
AGGACCTGATTTCCTTACTGCCTCTTGCTTCTCTTTTCCTATCCTGAGTAGTGGTAATCTACTGGGACGGAACAGCTTTG
AGGTGCGTGTTTGTGCCTGTCCTGGGAGAGACCGGCGCACAGAGGAAGAGAATCTCCGCAAGAAAGGGGAGCCTCACCAC
GAGCTGCCCCCAGGGAGCACTAAGCGAGGTAAGCAAGCAGGACAAGAAGCGGTGGAGGAGACCAAGGGTGCAGTTATGCC
TCAGATTCACTTTTATCACCTTTCCTTGCCTCTTTCCTAGCACTGCCCAACAACACCAGCTCCTCTCCCCAGCCAAAGAA
GAAACCACTGGATGGAGAATATTTCACCCTTCAGGTACTAAGTCTTGGGACCTCTTATCAAGTGGAAAGTTTCCAGTCTA
ACACTCAAAATGCCGTTTTCTTCTTGACTGTTTTACCTGCAATTGGGGCATTTGCCATCAGGGGGCAGTGATGCCTCAAA
GACAATGGCTCCTGGTTGTAGCTAACTAACTTCAGAACACCAACTTATACCATAATATATATTTTAAAGGACCAGACCAG
CTTTCAAAAAGAAAATAGTTAAAGAGAGCATGAAAATGGTTCTATGACTTTGCCTGATACAGATGCTACTTGACTTACGA
TGGAGTTACTTCTGATAACTCGTCGTAAGTTGAAATATTGAAATATTGTAAGTTGAAAATGGATTTAATACACCTAATCT
AAGGAACATCATAGCTTAGCCTAGCCTGCTTTTTTTTTTTTTTTTTTTTGGAGACAGAGTCTCACTCTGCTACCCAGGC
TGGAGTGCAGTGGCGGGATCTCGGCTCACTGCAACCTCCGCCTTCTGGGTTCAAGCGATTCTCCTGCCTCAGCCCACTGA
GTAGCTGGGATTACAGGCACCTGCCCCGACGCCCAGCTAATTTTTTGTTATTTATTTCCTTTTTTTTTAGTAGAGATAGA
ATTTCACCATGTTGGCCAGGCTAGTCTCGAACTCCTGACCTTGTGATCTGCCTGCCTTGGCCTCCCAAAGTGCTGGGATT
ACAGGCGTGAGCCACCGCACCTGGCCTGCCTAGCCTACTTTTATTTTATTTTTAATGGAGACAGCATCTTGCTCTGTTGC
CCAGGCTGGATTACAGTGATGTGATCATAGCTCATTATACCCTCCTGGGCTCAAGCAATCCCCCTAACTCTGCCTCCCCA
GTAGCTAGGACCACAGGCATACACCACCATACCCAGCTAATTTTTAAAATTTTTTGTAGATAGATAGAGTCTCACTATGT
TGCCCAGGCTGGTCTCTAGCCTACTTTTTTGAGACAAGGTCTTGCTCTGTCACCCAGGCTGGATAGAGTGCAGTAGTGCA
GTCACAGCTCACTGCAGCCTCCACCTCCCAGGCTCCATCCATCCTCCCAGCTCAGCCTCCCAAGTTGCTTCAACTACAGG
CCTGCACCACCATGCCTGGCTAATTTTTATTTATTTATTTTTATTTTATTTTATTTTATTTTTTGAGACTCAGTCTCACT

CTGTCGCCTTAGGCTGGAGTGCAGTGGCATGATCTCGGCTCACTGCTAACCTCTGCCTCCTGGGTTTCAAGTGATTCTCC
TGCCTCAGCCTCCCGAATAGCTAGGACTACAAGCGCCTGCTACCACGCCCGGCTAATTTGTGTATTTTTAGTAGAGACAG
GGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTTCTGACCATGTGATCGCCGCCTCGGCCTCCCAAAGTGCTGGGATTA
CAGGTGTGAGCCACCACGCCCGGCTAATTTTTATTTATTTATTTAAAGACAGAGTCTCACTCTGTCACTCAGGCTAGAGT
GCAGTGGCACCATCTCAGCTCACTGCAGCCTTGACCTCCCTGGGCTCCGGTGATTTCACCCTCCCAAGTAGCTAGGACTA
CAGGCACATGCCACGACACCCAGCTAATTTTTTATTTTCTGTGAAGTCAAGGTCTTGCTACGTTGCCCATGCTGGTATCA
AACCCCTGGGCTCAATCAATCCTTCCACCTCAGCCTCCCCAAGTATTGGGGTTACAGGCATGAGCTACCACACTCAGCCC
TAGCCTACTTGAAACGTGTTCAGAGCATTTAAGTTACCCTACAGTTGGGCAAAGTCATCTAACACAAAGCCCTTTTTATA
GTAATAAAATGTTGTATATCTCATGTGATTTATTAGATATTGTTACTAAAAGTGAGAAACAGCATGGTTGCATGAAAGGA
GGCACAGTCGAAGCCAGGCACAGCCTGGGCGCAGAGCGAGACTCAAAAAAAGAAAAGGCCAGGCGCACTCTCACGCCTGT
AATCCCAGCATTTCGGGAGGCTGAGGCGGGTGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTAGCCAACATGGTGAA
ACCCCGTCTCTACTAAAATACAAAAATTAACCGGGCGTGATGGCAGGTGCCTGTAATCCCAGCTACTTGGGAGGCTGAGG
CAGGAGAATCGCTTGAACCAGGAGGCGGAGGTTGCAGGGAGCCAAGACGGCGCCACTGCACTCCAGCCTGGGCGATAGAG
TGAGACTCCGTCTCAGAAAAAAAAGAAAAGAAACGAGGCACAGTCGCATGCACATGTAGTCCCAGTTACTTGAGAGGCTA
AGGCAGGAGGATCTCTTGAGCCCAAGAGTTTGAGTCCAGCCTGAACAACATAGCAAGACATCATCTCTAAAATTTAAAAA
AGGGCCGGGCACAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGTGGAGGTGGGTAGATCACCTGACGTCAGGAGT
TGGAAACCAGCCTGGCTAACATGGTGAAGCCCCATCTCTACTAAAAACACAAAAATTAGCCAGTGTGAGACACGTTGAGT
CCACGTACTCGGAGGCTGAGGCACAAGAATCACTTGAACCCCAGAGGCGGAGATTCGAATCAGCCAAGATTGCACCATTG
CACTCCCGCCTGGGCGACGAGAGTGAGACCCCATCTCAAAATAAATAAATAAATATTTTTAAAAGTCAGCTGTATAGGTA
CTTGAAGTGCAGTTTCTACTAAATCGATGTTGCTTTTGATCCGTCATAAAGTCAAACAATTGTAACTTGAACCATCTTTT
AACTCAGGTACTGTGTATATACTTACTTCTCCCCCTCCTCTGTTGCTGCAGATCCGTGGGCGTGAGCGCTTCGAGATGTT
CCGAGAGCTGAATGAGGCCTTGGAACTCAAGGATGCCCAGGCTGGGAAGGAGCCAGGGGGGAGCAGGGCTCACTCCAGGT
GAGTGACCTCAGCCCCTTCCTGGCCCTACTCCCCTGCCTTCCTAGGTTGGAAAGCCATAGGATTCCATTCTCATCCTGCC
TTCATGGTCAAAGGCAGCTGACCCCATCTCATTGGGTCCCAGCCCTGCACAGACATTTTTTTAGTCTTCCTCCGGTTGAA
TCCTATAACCACATTCTTGCCTCCACGTAGTATCCACAGAACATCCAAACCCAGGGACGAGTGTGGATACTTCTTTGCCA
TTCTCCGCCAACTCCCCAGCCCAGAGCTGGAGGGTCTCAAGGGGCCTAATAATTGTGTAATACTGAATACAGCCAGAGTT
TCAGGTCATATACTCAGCCCTGCCATGCACCGGCAGGTCCTAGGTGACCCCCGTCAAACTCAGTTTCCTTATATATAAAA
TGGGGTAAGGGGGCCGGGCGCAGTGGCTCACGAATCCCACACTCTGGGAGGCCAAGGCGAGTGGATCACCTGAGGTCGGG
AGTTTGAGCCCAGCCTGACCAACATGGAGAAACCCCATCTCTACTAAAAATACAAAAGTAGCCGGGCGTGGTGATGCATG
CCTGTAATCCCAGCTACCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCAGAGGTTGCGGTGAGCTG
AGATCTCACCATTACACTCCAGCCTGGGCAACAAGAGTGAAACTCCGTCTCAAAAAAGTATAATAAAGTAAAATGGGGTA
AGGGAAGATTACGAGACTAATACACACTAATACTCTGAGGTGCTCAGTAAACATATTTGCATGGGGTGTGGCCACCATCT
TGATTTGAATTCCCGTTGTCCCAGCCTTAGGCCCTTCAAAGCATTGGTCAGGGAAAAGGGGCACAGACCCTCTCACTCAT
GTGATGTCATCTCTCCTCCCTGCTTCTGTCTCCTACAGCCACCTGAAGTCCAAAAAGGGTCAGTCTACCTCCCGCCATAA
AAAACTCATGTTCAAGACAGAAGGGCCTGACTCAGACTGACATTCTCCACTTCTTGTTCCCCACTGACAGCCTCCCACCC
CCATCTCTCCCTCCCCTGCCATTTTGGGTTTTGGGTCTTTGAACCCTTGCTTGCAATAGGTGTGCGTCAGAAGCACCCAG
GACTTCCATTTGCTTTGTCCCGGGGCTCCACTGAACAAGTTGGCCTGCACTGGTGTTTTGTTGTGGGGAGGAGGATGGGG
AGTAGGACATACCAGCTTAGATTTTAAGGTTTTTACTGTGAGGGATGTTTGGGAGATGTAAGAAATGTTCTTGCAGTTAA
GGGTTAGTTTACAATCAGCCACATTCTAGGTAGGTAGGGGCCCACTTCACCGTACTAACCAGGGAAGCTGTCCCTCATGT
TGAATTTTCTCTAACTTCAAGGCCCATATCTGTGAAATGCTGGCATTTGCACCTACCTCACAGAGTGCATTGTGAGGGTT
AATGAAATAATGTACATCTGGCCTTGAAACCACCTTTTATTACATGGGGTCTAAAACTTGACCCCCTTGAGGGTGCCTGT
TCCCTCTCCCTCTCCCTGTTGGCTGGTGGGTTGGTAGTTTCTACAGTTGGGCAGCTGGTTAGGTAGAGGGAGTTGTCAAG
TCTTGCTGGCCCAGCCAAACCCTGTCTGACAACCTCTTGGTCGACCTTAGTACCTAAAAGGAAATCTCACCCCATCCCAC
ACCCTGGAGGATTTCATCTCTTGTATATGATGATCTGGATCCACCAAGACTTGTTTTATGCTCAGGGTCAATTTCTTTTT
TCTTTTTTTTTTTTTTTTTTCTTTTTCTTTGAGACTGGGTCTCGCTTTGTTGCCCAGGCTGGAGTGGAGTGGCGTGATCT
TGGCTTACTGCAGCCTTTGCCTCCCCGGCTCGAGCAGTCCTGCCTCAGCCTCCGGAGTAGCTGGGACCACAGGTTCATGC
CACCATGGCCAGCCAACTTTTGCATGTTTTGTAGAGATGGGGTCTCACAGTGTTGCCCAGGCTGGTCTCAAACTCCTGGG
CTCAGGCGATCCACCTGTCTCAGCCTCCCAGAGTGCTGGGATTACAATTGTGAGCCACCACGTGGAGCTGGAAGGGTCAA
CATCTTTTACATTCTGCAAGCACATCTGCATTTTCACCCCACCCTTCCCCTCCTTCTCCCTTTTTATATCCCATTTTTAT
ATCGATCTCTTATTTTACAATAAAACTTTGCTGCCACCTGTGTGTCTGAGGGGTGAACGCCAGTGCAGGCTACTGGGGTC

AGCAGGTGCAGGGGTGAGTGAGGAGGTGCTGGGAAGCAGCCACCTGAGTCTGCAATGAGTGTGGACTGGGGGGCCCAGTG
CCCGGGTTCCGGGAGGGGAACAAAGGCTGGAGACTGGGTCAGTCTGCGGGCTGCATGACAACAAGGGAGGGGGTGGCTCC
ATTCATAACTCAGGAACCAACCGTCCCTCCTCCCCTCCGCCCACGGCTGGCACAAGGTTCTCTGCCTCCCCTGCTTCTAG
GATTGGGCTGCTTCCCCCTCGGCAGCCTCTCACCAAGGATTACGGGATTTAAATGTCGTGATTTACGAAGGCTGAGCCTC
CAGGGTGGCCATCTTCGTCCATCAGAAGTGGCAGGATACCTGGGTTCCAAGGGAACAGGGTGG

Appendix E: Gel electrophoresis

2%Agarose Gel

4g Agarose powder

200ml 1X TBE

4 ul Ethidium Brommde

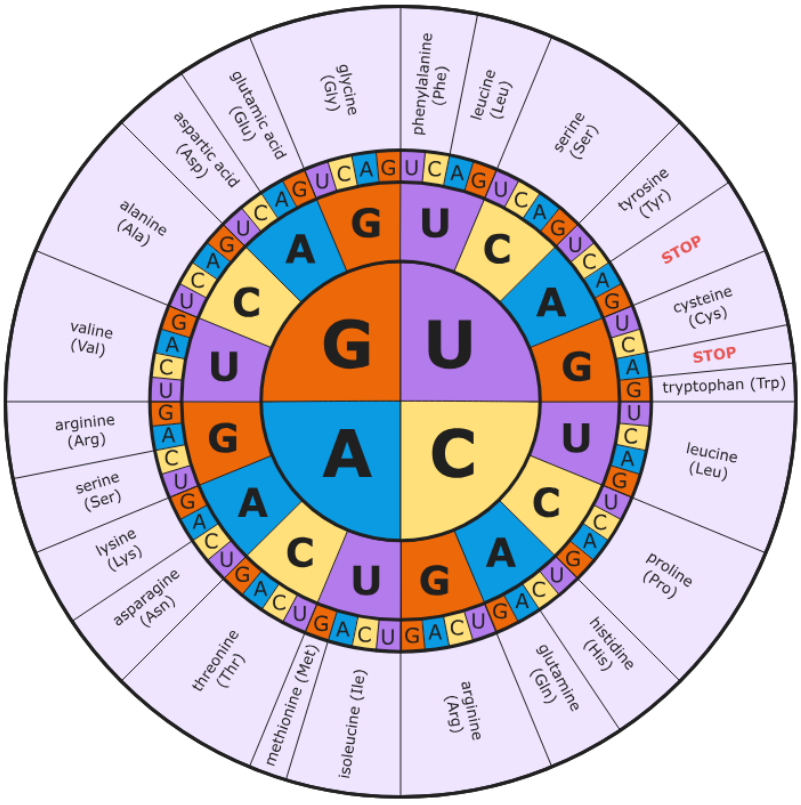
Appendix F: Genetic codon wheel and sheet

Codon-Amino Acid Abbreviations

Codon	Full Name	Abbreviation (3 Letter)	Abbreviation (1 Letter)
TTT	Phenylalanine	Phe	F
TTC	Phenylalanine	Phe	F
TTA	Leucine	Leu	L
TTG	Leucine	Leu	L
TCT	Serine	Ser	S
TCC	Serine	Ser	S
TCA	Serine	Ser	S
TCG	Serine	Ser	S
TAT	Tyrosine	Tyr	Y
TAC	Tyrosine	Tyr	Y
TAA	Termination (ochre)	Ter	X
TAG	Termination (amber)	Ter	X
TGT	Cysteine	Cys	C
TGC	Cysteine	Cys	C
TGA	Termination (opal or umber)	Ter	X
TGG	Tryptophan	Trp	W
CTT	Leucine	Leu	L
CTC	Leucine	Leu	L
CTA	Leucine	Leu	L
CTG	Leucine	Leu	L
CCT	Proline	Pro	P
CCC	Proline	Pro	P
CCA	Proline	Pro	P
CCG	Proline	Pro	P
CAT	Histidine	His	H

CAC	Histidine	His	H
CAA	Glutamine	Gln	Q
CAG	Glutamine	Gln	Q
CGT	Arginine	Arg	R
CGC	Arginine	Arg	R
CGA	Arginine	Arg	R
CGG	Arginine	Arg	R
ATT	Isoleucine	Ile	I
ATC	Isoleucine	Ile	I
ATA	Isoleucine	Ile	I
ATG	Methionine	Met	M
ACT	Threonine	Thr	T
ACC	Threonine	Thr	T
ACA	Threonine	Thr	T
ACG	Threonine	Thr	T
AAT	Asparagine	Asn	N
AAC	Asparagine	Asn	N
AAA	Lysine	Lys	K
AAG	Lysine	Lys	K
AGT	Serine	Ser	S
AGC	Serine	Ser	S
AGA	Arginine	Arg	R
AGG	Arginine	Arg	R
GTT	Valine	Val	V
GTC	Valine	Val	V
GTA	Valine	Val	V
GTG	Valine	Val	V
GCT	Alanine	Ala	A
GCC	Alanine	Ala	A

GCA	Alanine	Ala	A
GCG	Alanine	Ala	A
GAT	Aspartate	Asp	D
GAC	Aspartate	Asp	D
GAA	Glutamate	Glu	E
GAG	Glutamate	Glu	E
GGT	Glycine	Gly	G
GGC	Glycine	Gly	G
GGA	Glycine	Gly	G
GGG	Glycine	Gly	G
n/a	Aspartate or Asparagine	n/a	B
n/a	Glutamate or Glutamine	n/a	Z



Write your answer as the three-letter amino acid abbreviation. Write STOP *for a stop codon.*

Appendix G: Different types of mutations isolated during sequencing

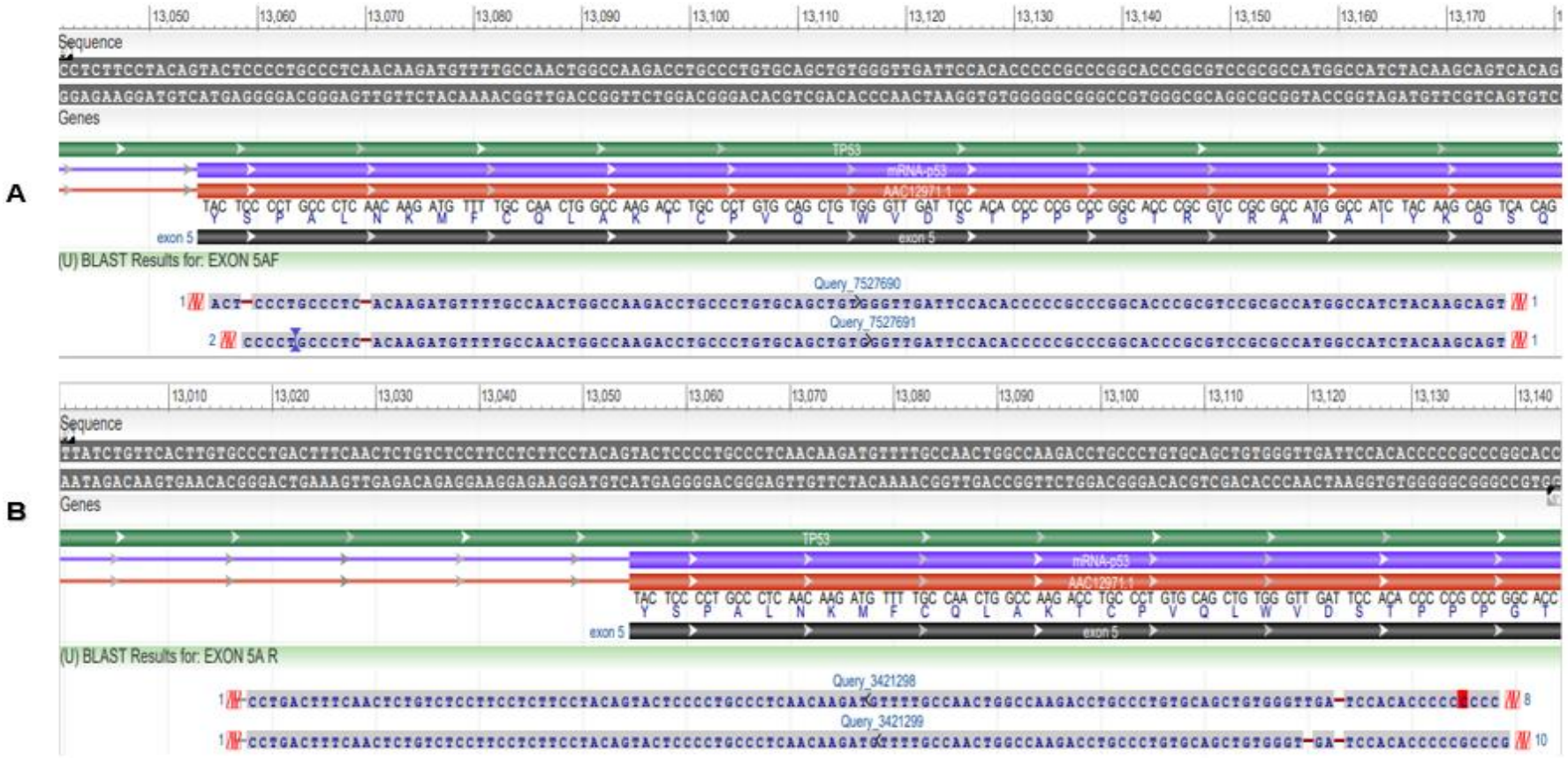


Figure G.1: Different types of mutations sequenced on exon 5A with A representing forward primer mutations and B: reverse primer mutations

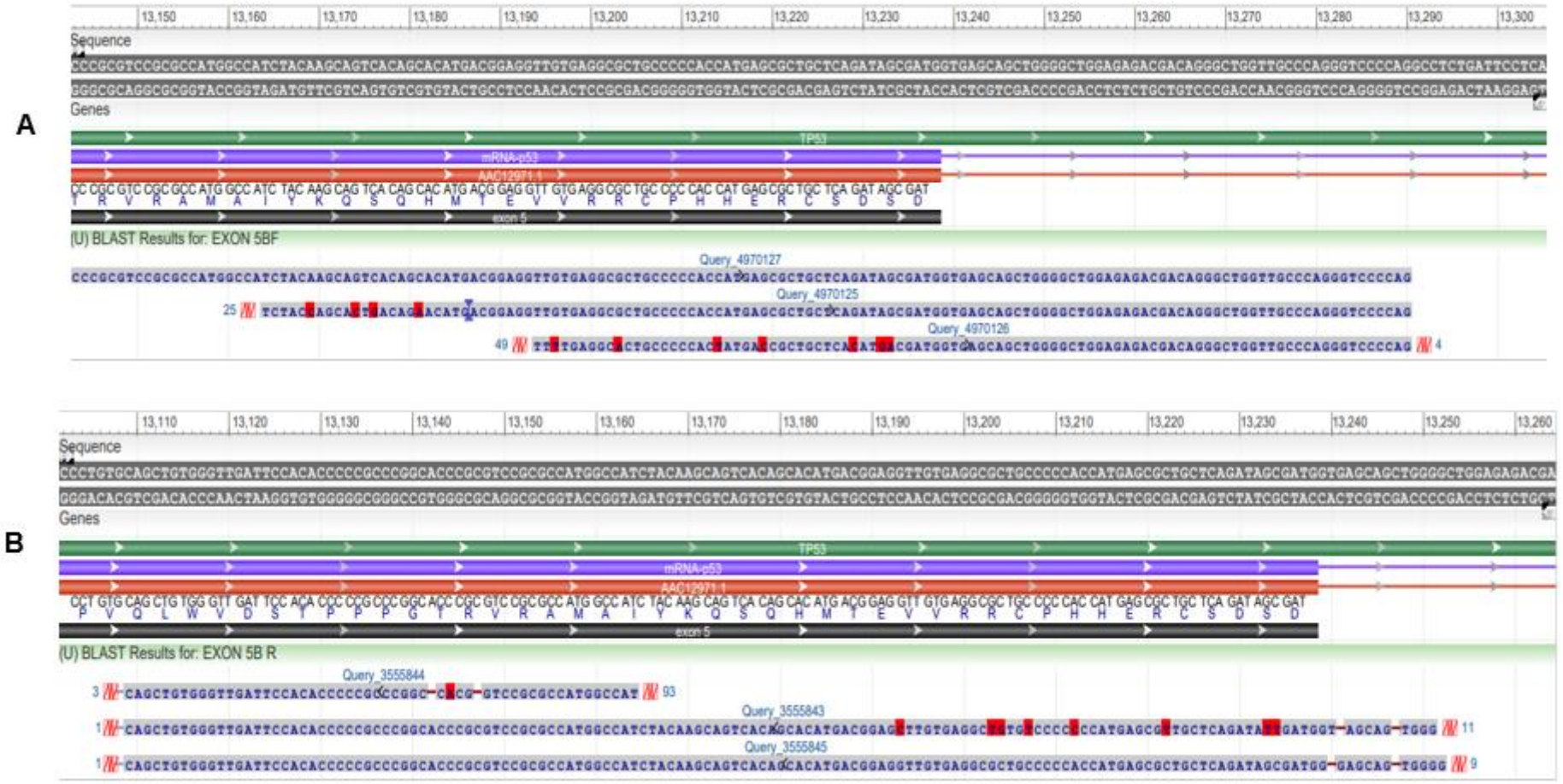


Figure G.2: Different types of mutations sequenced on exon 5B with A representing forward primer mutations and B: reverse primer mutations

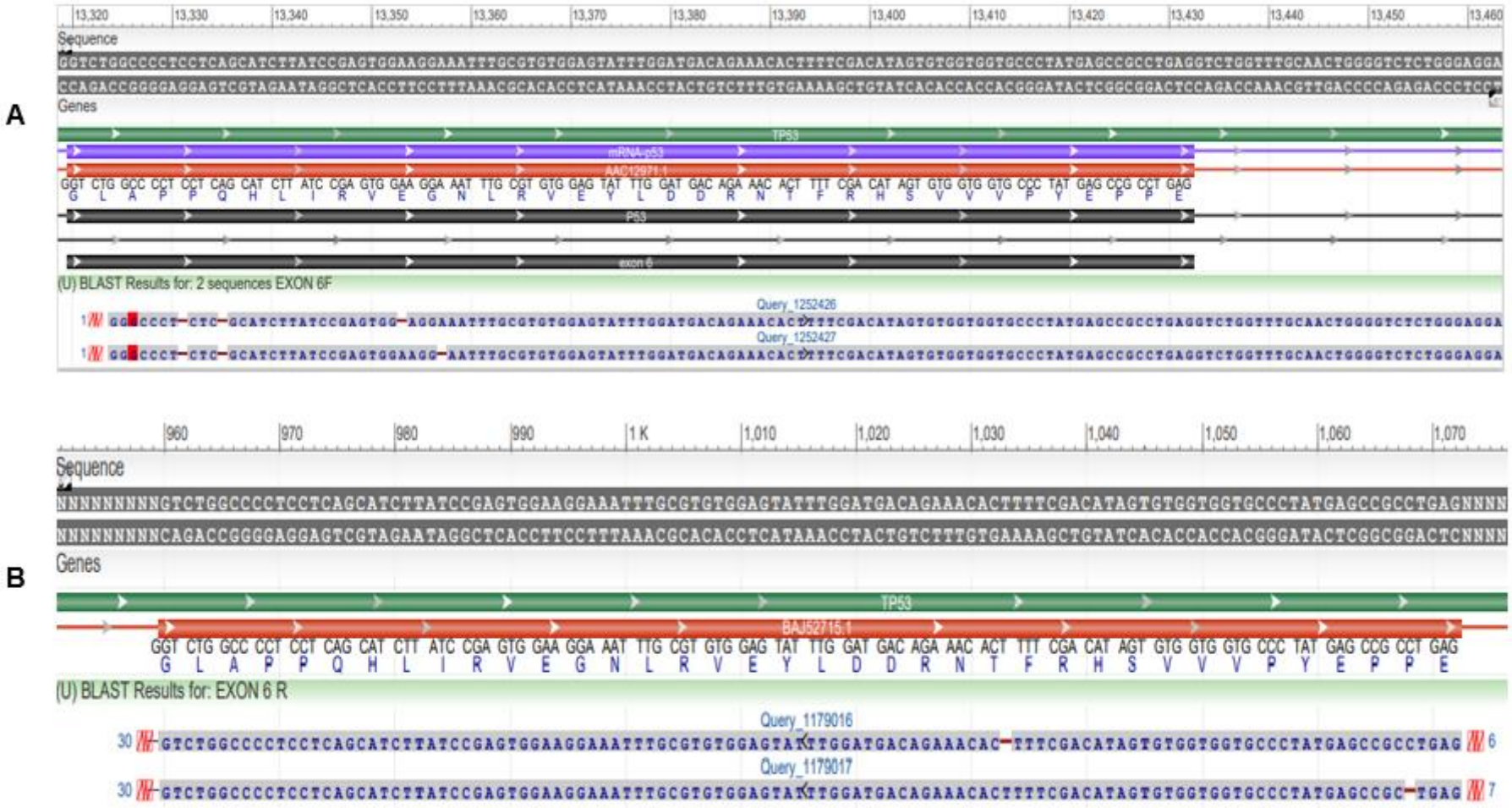


Figure G.3: Different types of mutations sequenced on exon 6 with A representing forward primer mutations and B: reverse primer mutations

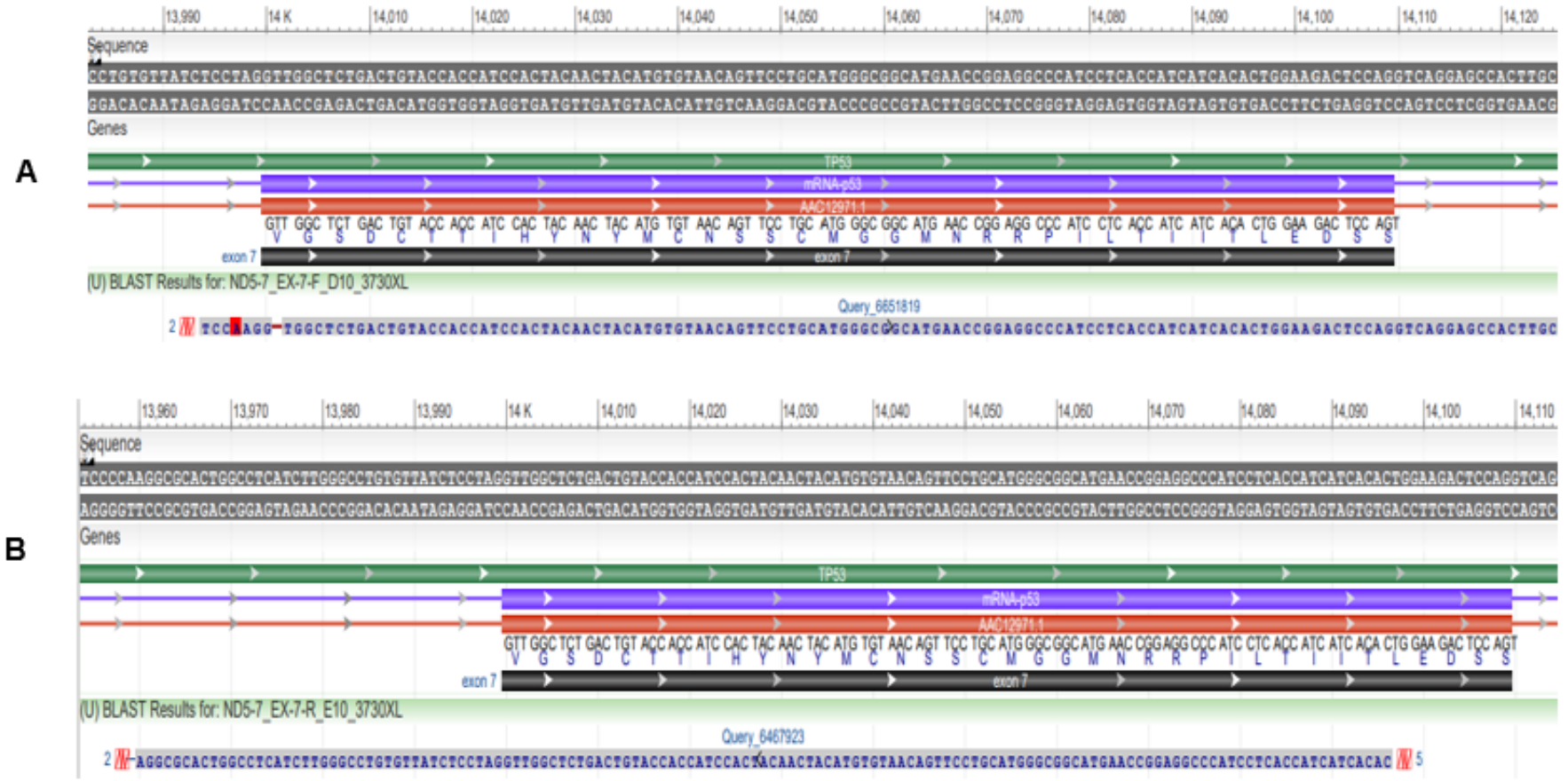


Figure G.4: Different types of mutations sequenced on exon 7 with A representing forward primer mutations and B: reverse primer mutations

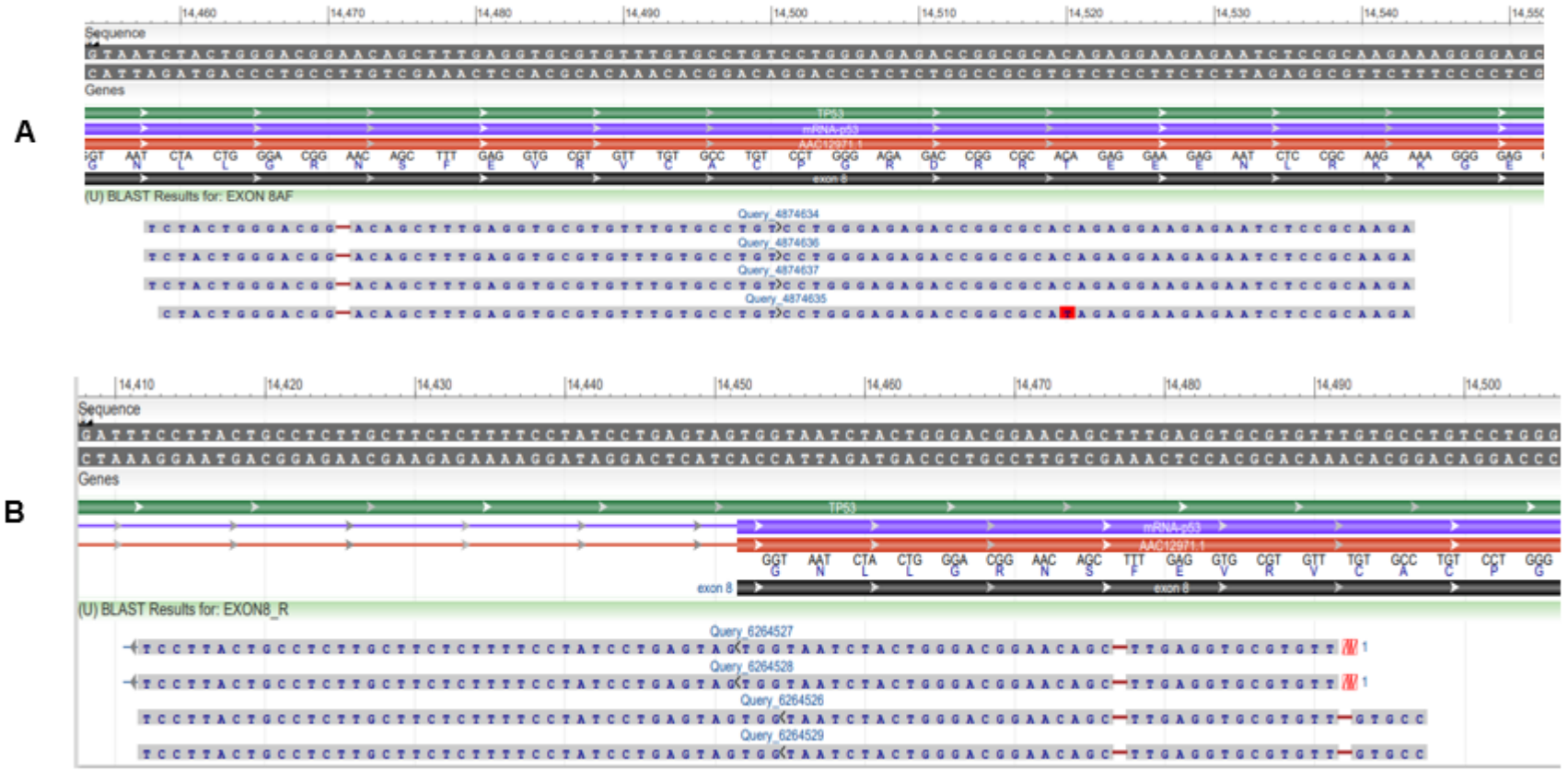


Figure G.5: Different types of mutations sequenced on exon 8A with A representing forward primer mutations and B: reverse primer mutations

Figure G.6: Different types of mutations sequenced on exon 8B with A representing forward primer mutations and B: reverse primer mutations

Appendix H: Summary of mutated DNA and proteins

Exon	Case #	DNA Mutation	Mutated Protein	Mutation type	Change/ Mutation	type	Structural motif	Structure Function Class	SIFT Class	Transactivation Class	Pathogenicity
EXON 5AF	ND-05	c.380 Del C	p.S127fs Ser > -	Frameshift			L1/S/H2	n/a	n/a	n/a	
		c.391 Del A	p.N131fs Asn > -	Frameshift			L1/S/H2	n/a	n/a	n/a	
	ND-59	c.384 Ins G	p.P128fs Pro >	Frameshift			L1/S/H2	n/a	n/a	n/a	
		c.391 Del A	p.N131fs Asn > -	Frameshift			L1/S/H2	n/a	n/a	n/a	
EXON 5AR	ND-05	c.444 Del T	p.D148fs Asp > -	Frameshift	GAT		NDBL/beta-sheets	n/a	n/a	n/a	
		c.456 MM. G > C	p.P152P Pro > Pro	Silent	CCG > CCC	G:C>C:G	NDBL/beta-sheets	n/a	Tolerated	n/a	Benign
	ND-59	c.441 Del T	p.V147fs Val > -	Frameshift	GTT		NDBL/beta-sheets	n/a	n/a	n/a	Likely pathogenic
		c.444 Del T	p.D148fs Asp > -	Frameshift	GAT		NDBL/beta-sheets	n/a	n/a	n/a	
EXON 5BF	ND-05	NONE									
	ND-37	c.490 MM. A > C	p.K164Q Lys > Gln	missense	AAG > CAG	A:T>C:G	L2/L3	non-functional	Damaging	partially functional	VUS
		c.495 MM. G > C	p. Q165H Gln > His	missense	CAG > CAC	G:C>C:G	L2		Damaging		VUS
		c.497 MM. C > G	p.S 166* Ser >	nonsense	TCA> TGA	G:C>C:G	L2/L3	n/a	n/a	n/a	Pathogenic

			STOP(Ter)								
Exon	**Case #**	**DNA Mutation**	**Mutated Protein**	**Mutation type**	**Change/ Mutation**	**type**	**Structural motif**	**Structure Function Class**	**SIFT Class**	**Transactivation Class**	**Pathogenicity**
		c.502 MM. C > A	p.H168N His > Asn	missense	CAC> AAC	G:C>T:A	L2/L3	functional	Damaging	partially functional	VUS
		c.508 Ins T	p.T170	Frameshift							
	ND-45	c.517 MM. G > T	p.V173L Val > Leu	missense	GTG > TTG	G:C>T:A	L2/L3	non-functional	Damaging	non-functional	Pathogenic
		c.524 MM. G > A	p.R175H Arg > His	missense	CGC > CAC	G:C>A:T	L2/L3	non-functional	Damaging	non-functional	Pathogenic
		c.535 MM. C > T	p.H179Y Hist > Tyr	missense	CAT > TAT	G:C>A:T	L2/L3	non-functional	Damaging	partially functional	Pathogenic
		c.540 MM. C > G	p.E180D Glu > Asp	missense	GAG > GAC	G:C>C:G	L2/L3	functional	Damaging	partially functional	Possibly Pathogenic
		c.550 MM. G > C	p.D184H Asp > His	missense	GAT > CAT	G:C>C:G	L2/L3	functional	Damaging	partially functional	
		c.553 MM. A > G	p.S185D Ser > Asp		AGC > GAC	A:T>G:C					
		c.554 MM. G > A	p.S185D Ser > Asp		AGC > GAC	G:C>A:T					
EXON 5BR	**ND-05**	c.463 Del A	p.T155fs Thr > -	Frameshift	ACC > _CC		NDBL/beta-sheets	n/a	n/a	n/a	Likely Pathogenic
		c.465 MM. C > A	p.T155H Thr > His		ACC > CAC	G:C>T:A					
		c.468 Del C	p.R155fs Arg > -	Frameshift	CGC > CG_						
	ND-37	c.514 MM. G > C	p.V172L Val > Leu	missense	GTT > CTT	G:C>C:G	L2		Damaging		VUS

		c.524 MM. G > T	p.R175L Arg > Leu	missense	CGC > CTG	G:C>T:A	L2		Damaging		Possibly pathogenic
		c.525 MM. C > G	p.R175L Arg > Leu		CGC > CTG	G:C>C:G					
		c.528 MM. C > T	p.C176C Cys > Cys	Silent	TGC > TGT	G:C>A:T	L2/L3	n/a	Tolerated	n/a	Benign
		c.533 MM. A > C	p.H178P His > Pro	missense	CAC > CCC	A:T>C:G	L2/L3	non-functional	Damaging	non-functional	Possibly pathogenic
		c.543 MM. C > T	p.R181R Arg > Arg	Silent	CGC > CGT	G:C>A:T	L2/L3	n/a	Tolerated	n/a	Benign
		c.554 MM. G > T	p.S185I Ser > Ile	missense	AGC > ATT	G:C>T:A					VUS
		c.555 MM. C > T	p.S185I Ser > Ile		AGC > ATT	G:C>A:T					
		Del G									
		Del C									
	ND-45	Del T									
		Del C									
EXON 6F	**ND-05**	c.566 MM.C > G	p.A189G Ala > Gly	missense	GCC > GGC	G:C>C:G	L2/L3	functional	Damaging	partially functional	VUS
		c.571 Del C	p.P191fs Pro > -	Frameshift	CCT > _CT		L2/L3	n/a	n/a	n/a	
		c.574 Del A	p.Q192fs Gln > -	Frameshift	CAG > C_G		L2/L3	n/a	n/a	n/a	
		c.593 Del A	p.E198fs Glu	Frameshift	GAA > G_A		NDBL/beta-sheets	n/a	n/a	n/a	
	ND-37	c.566 MM.C > G	p.A189G Ala > Gly	missense	GCC > GGC	G:C>C:G	L2/L3	functional	Damaging	partially functional	
		c.571	p.P191fs	Frameshift	CCT > _CT		L2/L3	n/a	n/a	n/a	

		Del C	Pro > -								
		c.574 Del A	p.Q192fs Gln > -	Frameshift	CAG > C_G		L2/L3	n/a	n/a	n/a	
		c.597 Del A	p.G199fs Gly > -	Frameshift	GGA > GG_		NDBL/beta-sheets	n/a	n/a	n/a	
EXON 6R	**ND-05**	c.633 Del T	p.T211fs Thr > -	Frameshift	ACT > AC_		NDBL/beta-sheets	n/a	n/a	n/a	
	ND-37	c.688 Del C	p.P223fs Pro > -	Frameshift	CCT > C_T		NDBL/beta-sheets	n/a	n/a	n/a	
EXON 7F	**ND-05**	C.674 Del T	p.V225fs Val > -	Frameshift	GTT > G_T		NDBL/beta-sheets	n/a	n/a	n/a	
		c.743 MM.G	p.248								
EXON 7R	**ND-05**	NONE									
EXON 8AF	**ND-05**	c.802 Del A	p.N268fs Asn > -	Frameshift	AAC > _AC		NDBL/beta-sheets	n/a	n/a	n/a	
	ND-08	c.802 Del A	p.N268fs Asn > -	Frameshift	AAC > _AC		NDBL/beta-sheets	n/a	n/a	n/a	
	ND-37	c.802 Del A	p.N268fs Asn > -	Frameshift	AAC > _AC		NDBL/beta-sheets	n/a	n/a	n/a	
	ND-64	c.802 Del A	p.N268fs Asn > -	Frameshift	AAC > _AC		NDBL/beta-sheets	n/a	n/a	n/a	
		c.851 MM. C > T	p.T284I Thr > Ile	missense	ACA > ATA	G:C>A:T	L1/S/H2	non-functional	Damaging	functional	VUS
EXON 8AR	**ND-05**	c.808 Del T	p.F270fs Phe > -	Frameshift	TTT > _TT		NDBL/beta-sheets	n/a	n/a	n/a	

	ND-08	c.808 Del T	p.F270fs Phe > -	Frameshift	TTT > _TT		NDBL/beta-sheets	n/a	n/a	n/a	
	ND-37	c.808 Del T	p.F270fs Phe > -	Frameshift	TTT > _TT		NDBL/beta-sheets	n/a	n/a	n/a	
		c.823 Del T	p.C275fs Cys > -	Frameshift	TGT > _GT		L1/S/H2	n/a	n/a	n/a	Likely Pathogenic
	ND-64	c.808 Del T	p.F270fs Phe > -	Frameshift	TTT > _TT		NDBL/beta-sheets	n/a	n/a	n/a	
		c.823 Del T	p.C275fs Cys > -	Frameshift	TGT > _GT		L1/S/H2	n/a	n/a	n/a	Likely Pathogenic
EXON 8BF	**ND-05**	c.871 Del A	p.K291fs Lys > -	Frameshift	AAG > _AG		NDBL/beta-sheets	n/a	n/a	n/a	
		c.874 Del A	p.K292fs Lys > -	Frameshift	AAA		NDBL/beta-sheets	n/a	n/a	n/a	
		MM. C > A									
		MM. C > A									
	ND-08	UNSUCCESSFUL									
	ND-37	c.871 Del A	p.K291fs Lys > -	Frameshift	AAG > _AG		NDBL/beta-sheets	n/a	n/a	n/a	
		c.874 Del A	p.K292fs Lys > -	Frameshift	AAA		NDBL/beta-sheets	n/a	n/a	n/a	
	ND-59	c.871 Del A	p.K291fs Lys > -	Frameshift	AAG > _AG		NDBL/beta-sheets	n/a	n/a	n/a	
		c.874 Del A	p.K292fs Lys > -	Frameshift	AAA		NDBL/beta-sheets	n/a	n/a	n/a	
		MM. A > G									
EXON 8BR		NONE ON ALL									

Pink: Novel mutations, orange: unknown occurrences (outside the exon)

Appendix I: TURNITIN REPORT

turnitin

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Thembi Mashele
Assignment title:	Final submission (all, regardless of course/programme)
Submission title:	2057354 MSc Dissertation 30042025-4.docx
File name:	2057354_MSc_Dissertation_30042025.docx
File size:	275.41K
Page count:	58
Word count:	14,654
Character count:	79,674
Submission date:	30-Apr-2025 08:27PM (UTC+0200)
Submission ID:	3662240192