Identification of Therapeutic Targets of Kaposi Sarcoma and Effectivity of Available Therapeutics

BY

Syeda Samana Zahid

Fall-2020-MSCS&E00000359058

Supervised by

Dr. Rehan Zafar Paracha

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

in

Bioinformatics

September, 2020

School of Interdisciplinary Engineering and Sciences (SINES)

National University of Sciences and Technology (NUST)

This thesis is dedicated to my beloved parents, Mrs. Nusrat Zahid and Mr. Zahid Raza Jaffary for their consistent support and love

DECLARATION

I Syeda Samana Zahid, hereby declared that work presented in this thesis is the result of my own work except specific reference is made to the work of others wherever due. I also declared that the content presented in this thesis is original and has not been submitted in whole or in part to this university or to any other university for any other degree or qualification.

Syeda Samana Zahid

September, 2020

Acknowledgement

All the praise to Allah Almighty for blessing me with the best family members, teachers, colleagues and friends and illuminating me with the best of his knowledge. All the admiration for the Holy Prophet(PBUH) who is an ideogram of patience, guidance, and lenity.

My extensive gratitude to my supervisor Dr. Rehan Zafar for his unmatched effort and encouragement throughout the entire journey of research. He assured his availability every single moment. His way of dealing with issues and handling difficult situations guided me on how to manage things during distress. I am grateful for his constant support, encouragement, and help, which made this research fruitful.

I would like to thank my GEC members Dr. Zamir Hussain, Dr. Zartasha Mustansar and Dr. Mehak Rafiq for their timely availability despite of their tough routines.

Finally, words are not enough to express my gratitute to my parents for providing me with the best of everything in life

Contents

Contents iv

List of Tables viii

List of Figures ix

ABSTARCT 1

1 INTRODUCTION 3
- 1.1 Cancer . . . 3
- 1.2 Kaposi Sarcoma . . . 5
  - 1.2.1 Epidemiology . . . 8
  - 1.2.2 Clinical Presentation . . . 8
  - 1.2.3 Diagnosis and Prognosis . . . 10
  - 1.2.4 Human Herpes Virus 8 . . . 10
  - 1.2.5 Treatment of Kaposi Sarcoma . . . 13
- 1.3 Biomarker . . . 15
- 1.4 Next Generation Sequencing . . . 17
  - 1.4.1 RNA Seq . . . 18
  - 1.4.2 Microarray . . . 18
- 1.5 Docking . . . 19
- 1.6 Objectives . . . 19

2 LITERATURE REVIEW 20
- 2.1 Kaposi Sarcoma . . . 20
- 2.2 Human Herpes Virus 8 . . . 21
- 2.3 Biomarkers in Kaposi Sarcoma . . . 23
  - 2.3.1 RNA Seq Analysis . . . 23
  - 2.3.2 Microarray Analysis of Kaposi Sarcoma . . . 24
- 2.4 Problem Statment . . . 24
- 2.5 Objectives . . . 24

3 MATERIALS AND METHODS 25
- 3.1 Methodology Overview . . . 25
- 3.2 Identification of Therapeutic Targets . . . 25
- 3.3 Agilent microarray . . . 25
  - 3.3.1 Data Retrieval . . . 27
  - 3.3.2 GEO Database . . . 27
  - 3.3.3 Library Calling . . . 28
  - 3.3.4 Preprocessing . . . 29
  - 3.3.5 Preparing Design Matrix . . . 29
  - 3.3.6 Annotation . . . 29
  - 3.3.7 FeatureCut off . . . 30
  - 3.3.8 Plots . . . 30
  - 3.3.9 Linear Model . . . 30

3.4 mRNA Seq . . . 31
3.4.1 Data Retrieval . . . 33
3.4.2 Quality Control . . . 33
3.4.3 Alignment . . . 34
3.4.4 MarkDuplicates and Duplicates Removal . . . 35
3.4.5 Quantification and Transcript Identification . . . 35
3.4.6 Identification of Differentially Expressed Genes . . . 36
3.5 Comparative Analysis . . . 37
3.6 Pathway Analysis . . . 37
3.6.1 Selection of Therapeutic Target . . . 38
3.7 Docking . . . 39
3.7.1 Data Retrieval . . . 40
3.7.2 Preprocessing of Protein Structure . . . 41
3.7.3 Preprocessing of Ligands . . . 43
3.7.4 Docking of Protein and Ligand . . . 44
3.7.5 Boxplot of Binding Affinities . . . 44

4 RESULTS 46
4.1 Microarray . . . 46
4.1.1 Data Retrieval . . . 46
4.1.2 Plots . . . 47
4.2 mRNA Seq . . . 55
4.2.1 mRNA Seq 1 . . . 55
4.2.2 Alignment . . . 57
4.2.3 Identification of DEGs . . . 58
4.2.4 mRNA Seq 2 . . . 76
4.3 Comparative Analysis . . . 83
4.4 Pathway Analysis . . . 84
4.5 Protein Modelling . . . 85
4.5.1 GSTP . . . 85
4.5.2 Modelling . . . 86
4.5.3 Non-Covalent Interactions . . . 90

5 DISCUSSION 99

6 CONCLUSION AND FUTURE PERSPECTIVES 102

REFERENCES 103

Nomenclature

Acronyms / Abbreviations

APCs	Antigen Presenting Cells
BCR	B cell receptor
BRN	Biological Regulatory Network
CAR	Coxsackie-adenovirus receptor
CD4+T cells	Helper T cells
CD8+T cells	Cytotoxic T cells
CNS	Central nervous system
CRS	Cytoreductive surgery
DAMPs	Danger-associated molecular pattern signals
DC	Dendritic Cells
EGF	Epidermal growth factor
FDA	US Food and Drug Administrationa
G-CSF	Granulocyte colony stimulating factor
GM-CSF	Granulocyte monocyte colony stimulating factor
HER2+	Human epidermal growth factor receptor 2+
HN	Hemagglutinin-neuraminidase
IL	Interleukin
INF	Interferon
MAPK	Mitogen-activated protein kinases
MHC	Major Histocompatibility complex
NDV	Newcastle Disease Virus
NK cells	Natural Killer Cells
NSCLC	Non-small cell lung cancer
PAMPs	Pathogen-associated molecular patterns
ROS	Reactive Oxygen specie
SCLC	Small Cell Lung Cancer
SLAM	Signalling lymphocytic activation molecule
STAT3	Signal Transducer and Activator of Transcription 3

SVV	Seneca Valley Virus
T-VEC	Talimogene laherparepvec
TAA	Tumor associated antigens
TAM	Tumor Microenvironment
TCR	T cell receptor
TK	Thymidine kinase
TNF	Tumor Necrosis Factor
VGF	Vaccinia growth factor
VV	Vaccinia Virus
WHO	World Health Organization

List of Tables

1.1 Types of Kaposi Sarcoma 7

2.1 Protein and their Function of Human Herpes Virus 8 22

3.1 Table 3.1: Microarray Libraries and their Activities 28

4.1 Parameters GSE16353 46

4.2 Sample IDs and Phenotype 47

4.3 Attributes of mRNA Seq dataset 55

4.4 Sample IDs and Phenotypes 56

4.5 Alignment Rate of Samples to Reference Genome-Hg38 57

4.6 Parameters of Dataset GSE100684 76

4.7 Phenotype and Sample IDs 76

4.8 Alignment rate of Sample to Reference Genome Hg38 77

4.9 Significant Pathways Among Common Genes 84

4.10 Parameters of the Template 17gs 86

4.11 Binding Affinities of Top 10 Ligands 89

List of Figures

1.1 Clinical Presentation of Kaposi Sarcoma Lesions 9
1.2 Treatment of Kaposi Sarcoma 14
1.3 Workflow of Next Generation Sequencing 18

3.1 Workflow of Microarray 26
3.2 Workflow of mRNA Seq 31
3.3 Workflow of mRNA Seq and Tools 32
3.4 Overview of Docking Methodology 39
3.5 Preprocessing Methodology 41

4.1 Boxplot of Intensities of Data 49
4.2 Heatplot of Intensities of Samples 50
4.3 Histogram of Median Intensities 51
4.4 RLE Plot of Expression Data 52
4.5 Histogram of Variation of Gene Expression 53
4.6 Volcano Plot of Upregulated and Downregulated Genes 54
4.7 Boxplot of FPKM values versus Samples 59
4.8 Transcript Distribution per Gene Count 60
4.9 Transcript Length versus Frequency 61
4.10 Histogram of Differential Expression 62
4.11 Enhanced Volcano Plot 63
4.12 FPKM Values versus Samples 65
4.13 Transcript Frequency per Gene Count 66
4.14 Transcript Length Distribution 67
4.15 Histogram of Differential Expression 68
4.16 Enhanced Volcano Plot 69
4.17 FPKM Values Versus Samples 71

4.18 Transcript Frequency Distribution per Gene Count 72
4.19 Transcript Length Frequency Distirbution 73
4.20 Histogram of Differential Expression 74
4.21 Enhanced Volcano Plot 75
4.22 Distirbution Log2FPKM of Samples 79
4.23 Distribution of Transcript Frequency Per Gene Count 80
4.24 Histogram of Differential Expression 81
4.25 Enhanced Volcano Plot 82
4.26 Common Genes among RNA Seq Dataset 83
4.27 Protein Structure generated through Swiss Model 87
4.28 Boxplot of Binding Affinities between Proteins and Ligands 88
4.29 Non-Covalent Interactions between GSTP1 and BDBM50458517 . 90
4.30 Non-Covalent Interactions between GSTP1 and DB00773 91
4.31 Non-Covalent Interactions between GSTP1 and DB06595 92
4.32 Non-Covalent Interactions between GSTP1 and DB11942 93
4.33 Non-Covalent Interactions between GSTP1 and DB11986 94
4.34 Non-Covalent Interactions between GSTP1 and DB12483 95
4.35 Non-Covalent Interactions between GSTP1 and DB12887 96
4.36 Non-Covalent Interactions between GSTP1 and DB14568 97
4.37 Non-Covalent Interactions between GSTP1 and DB15035 98

ABSTRACT

Kaposi Sarcoma is a malignancy of antiproliferative origin. Its etiology is linked to Kaposi Sarcoma-associated Herpes Virus or Human Herpes Virus 8 abbreviated as HHV8. The causative agent of Kaposi sarcoma is HHV8 consists of double-stranded DNA enclosed in an icosahedral capsid that remains in a latent state and switches to the lytic stage at any point in life with humans being natural hosts to HHV8 and is capable of modulating human immune and signaling pathways. Kaposi sarcoma lesions incorporate mucosal regions and skin that are featured with macules, nodules, and papules of purple, brown, or red color. There are 4 variants of Kaposi Sarcoma based on epidemiology and prognosis of the disease which include classic, endemic, iatrogenic, and epidemic types. This research primarily focuses on the identification of genes involved in the prognosis of Kaposi sarcoma with the help of high throughput sequencing platform and microarray analysis. Therefore, the genes involved at the onset and progression of Kaposi sarcoma are identified. Moreover, differentially expressed genes are identified with the comparison of various datasets and signaling pathways modulated as a result of the disease are determined. Moreover, variety of ligands are docked to the shortlisted therapeutic target to hand-pick the most effective ligand against Kaposi sarcoma. The short-listed differentially expressed genes included CA6, MIA, DCD, STRIP1, WNT4, FXDY1, IKBKE, and EXTL2, AQR, TIE2 and RBP2. These are the differentially expressed genes that were identified as a result of the analysis involved in the oncogenesis. Therefore, the protein Glutathione Transferase Protein 1 abbreviated as GSTP1 is reported to have a major role in the progression of Kaposi sarcoma that hides tumor cells from apoptosis by inhibiting of MAPK pathway by protein-protein interaction. Therefore, the mechanism of action of GSTP1 in inhibition of MAPK pathway is the interaction of GSTP1 with c JUN that delivers anti apoptotic signals that lead to apoptosis inhibition. The inhibition of apoptosis is a key component in carcinogenesis. This pathway is inhibited

in Kaposi sarcoma which inhibits apoptosis and leads to the survival of cancer cells. The second focus of this thesis was virtual screening of ligands to the target protein that was GSTP1.A variety of ligands that incorporated inhibitors of GSTP, anti-cancer drugs and drugs of Kaposi sarcoma were docked to GSTP1. A range of binding affinities were observed between the ligands and the protein that ranged from -4 Kcal/- mol to -10 Kcal/mol . As a result, the top 10 ligands with the maximum binding affinities were BDBM50458517, BDBM50562983, DB06595, DB11942, DB11986, DB12483, DB12887, DB14568 and DB15035. The docking between protein and ligand revealed BDBM50458517 to have a significant role in the treatment of Kaposi sarcoma that was previously just employed as an anti-cancer drug by promoting apoptosis. This analysis was based on international data. In this case we need data at the national level so that the differentially expressed genes from the Pakistani population can be identified and compared with the genes of the international population.

Chapter 1

INTRODUCTION

1.1 Cancer

The human body is formulated of cells. Cells integrate to and fabricate tissues and tissues amalgamate to configure organs. Cells divide to forge new cells as the human body demands them when the preceding one goes old and departs. Sometimes the process of cell division goes imprecise. New cells are formed when they are not required and the old one does not die. For the cells to divide DNA is replicated to be inherited by the succeeding cell. Sometimes this replication accompanies mutations that disrupt the normal life cycle of the cell. The existing cells do not die and continue to divide which develops cancer. Cancer development is the output of interferences of genetic changes in the cell cycle. As a result, cells divide and proliferate uncontrollably forming tumors. Tumors are benign or malignant. Malignant tumors metastasize to other body parts (1)

Genetic and epigenetic alterations participate in carcinogenesis. Abnormal gene expression led to neoplasia. The genetic path to tumorigeneses incorporates mutations in oncogenes and tumor suppressor genes. It involves loss or gain of activity of uncontrol expression. The epigenetic path of cancer is determined by chromatins that involve DNA methylation, nucleosome remodeling, histone modification, and microRNAs. Cancer initiation and progression involve multiple mutations in the epigenome. Hypomethylation, swap in nucleosome occupancy, hypermethylation, and alteration profiles. There is a crosstalk between genetics and epigenetics during carcinogenesis. They intervene and lead to tumor development and progression(2)

Epigenetic machinery aid in developing cellular identities and loss of genuine preservation of epigenetic factors leads to activation or inhibition of signaling

pathways that lead to cancer. Cancer cells harbor epigenetic alteration and initiate tumorigenesis. The epigenome of cancer is defined as alterations in epigenetic layers. It involves epigenetic silencing that results in the loss of activity of genes that facilitates mutations in signaling pathways and overexpression of oncogenic miRNAs. Various epigenetic modifiers are altered in human cancers. Mutations in epigenetic modifiers result in profound epigenetic alterations. It includes histone modifications, DNA methylation, and nucleosome positioning. Epigenetic mutations result in abnormal expression of genes and instability that predisposes to cancer (3).

The immune system plays a pivotal role in the regulation of tumor progression (4). Immune cells respond to tumors in two regards which include

- Tumor-associated antigens
- Tumor-specific antigens

Tumor-associated antigens are presented differently by normal and cancer cells. Whereas tumor-specific antigens are expressed by cancer cells. Cancer cells have the unique ability to suppress immunity systematically or in the tumor's microenvironment. Tumors produce immunosuppressive compounds like transforming growth factor beta, indoleamine-2,3-dioxygenase (immunosuppressive enzyme), and Fas ligand. Regulatory T cells dominate the tumor microenvironment by expressing antitumor cytokines and interleukin-10. Transforming growth factor-beta converts antitumor T cells to regulatory T cells that prevent their destruction by the immune system (5).

Carcinogenesis is a multifactor process. The cellular genome is altered after being exposed to a carcinogenic event. Malignancy can be an outcome of apoptosis prevention or immune surveillance malfunctions to abolish transformed cells. Pathogens possess these characteristics and alter host cells. Viruses linked to human cancers are termed tumor viruses. These viruses can integrate into the host genome and are capable to immortalize the infected cells for their replication. The target

cell expresses viral genes that result in apoptosis prevention and cellular growth and proliferation (6)

Almost 12% of cancer is a result of viral infection. These viral cancers include Epstein-Barr Virus, high-risk Human Papillomavirus, Hepatitis B and C virus, Human T cell lymphotropic virus-1 and Kaposi sarcoma herpesvirus. Human viral carcinogenesis has similar traits that include

- Tumor viruses are necessary but not adequate for oncogenesis
- Viral cancers pop up in the context of persistent infection
- Immune system performs a protective of deleterious role with tumor viruses (7)

1.2 Kaposi Sarcoma

Kaposi sarcoma is defined as an infrequent antiproliferative cancer. Kaposi sarcoma involves lesions that encounter skin and mucosal surfaces. Kaposi sarcoma lesions are characterized by red-blue, purple, or brown macules, papules or nodules which are susceptible to ulceration or bleeding (8).

The first case of Kaposi sarcoma was delineated by Moritz Kaposi in 1872. Moritz Kaposi was a physician and dermatologist by profession. He narrated numerous cases of multifocal and pigmented sarcoma on the skin of elderly men of European origin. They all died within 2 years (9). In 1947 numerous reports cataloged cases of Kaposi sarcoma in Africa (10) In 1981 Kaposi sarcoma came to the vanguard of public consciousness at the inception of the AIDS epidemic. The first case of highly aggressive Kaposi sarcoma was a young man who had sex with men who took place preliminary to the realization that these people are immunodeficient and are infected by opportunistic infections (11). The etiological agent of Kaposi sarcoma was not familiar till 1994. Directed research led to the discovery of Kaposi Sarcoma Herpes

Virus also called Human Herpes Virus-8(12) This virus was recognized by International Agency for Research on Cancer. Based on the epidemiology, symptoms, and prognosis of sarcoma it is divided into five types.

- Classic Kaposi Sarcoma
- Iatogenic Kaposi Sarcoma
- Endemic Kaposi Sarcoma
- Epidemic Kaposi Sarcoma
- Non-Epidemic Kaposi Sarcoma

The type of Kaposi sarcoma discovered by Kaposi is called the classic Kaposi Sarcoma. It occurs in elderly men of Jewish or Mediterranean ancestry. It manifests a sluggish and prolonged clinical course, primarily influencing skin and legs (13). A lymphadenopathic form of Kaposi sarcoma in children is known as endemic Kaposi sarcoma. A type of Kaposi Sarcoma in HIV-infected individuals suffering from AIDS when infected with HHV8 is known as epidemic Kaposi Sarcoma. It is also called as AIDS-related Kaposi sarcoma because it gained public attention during the AIDS epidemic. This form is largely seen in men who have sex with men. Kaposi sarcoma also infects organ transplant recipients known as Iatrogenic Kaposi sarcoma (14). The fifth form of Kaposi sarcoma occurs in men who have sex with men who are not infected with HIV (15). Table 4.1 shows a comparison of the epidemiological forms of Kaposi sarcoma.

Table 1.1. Types of Kaposi Sarcoma

Type	**Target**	**Clinical Appearance**	**Prognosis**
Classic Kaposi Sarcoma	Elderly men of Mediterranean, Eastern, European South American Origin	Cutaneous, Gastrointestinal Tract, Lymph nodes and various organs	Chronic
Iatrogenic Kaposi Sarcoma	Middle-aged men of Sub-Saharan African Origin	Benign in nodules, aggressive in soft tissues, skin, and visceral organs	Indolent to aggressive fatal
Endemic Kaposi Sarcoma	Organ transplant recipients, immunosuppressed individuals	Cutaneous, Gastrointestinal Hemorrhage	Indolent to aggressive, but treatable
Epidemic Kaposi Sarcoma	HIV Positive	Multifocal pigmented lesions in mucosal and visceral organs	Chronic and mostly malignant, treated with highly effective antiretroviral therapy
Non-Epidemic Kaposi Sarcoma	Middle aged, MSM	cutaneous	Indolent

1.2.1 Epidemiology

Classic Kaposi sarcoma is observed in elder men of Eastern Europe, Mediterranean, Jewish, and South America (16). Endemic Kaposi Sarcoma is seen in young, black, and HIV-negative men and 1-5 years aged African children (17). Iatrogenic Kaposi Sarcoma is mostly observed in immunosuppressed individuals who are organ transplant recipients mostly in the US population. Solid organ transplant aggravates the risk of Kaposi Sarcoma in the US population by 500-fold (18) The increasing incidence of Kaposi Sarcoma in young homosexual men in the USA was a harbinger of the AIDS epidemic (19). The seroprevalence of Kaposi Sarcoma is estimated to be 5% to 20% worldwide. Brazilian Amazon and Africa possess the highest rate (>50%). Eastern Europe, the Mediterranean, the Caribbean, and the Middle East have intermediate seroprevalence (5%-20%). Whereas North America and Asia have low seroprevalence (<5%). In a nutshell, the incidence pattern of Kaposi Sarcoma is lower in North America, Asia, and Western Europe compared to the huge incidence in the Mediterranean region and Africa. According to a study conducted by The Cancer Journal of Clinicians, a total of 42,000 new cases of Kaposi Sarcoma were reported with 2000 deaths in 2018 (20).

1.2.2 Clinical Presentation

The clinical presentation of Kaposi Sarcoma is highly variable depending on the epidemiological forms. The clinical symptoms of lesions are shown in fgure 1.1. Kaposi Sarcoma lesions are furious macules, papules and nodules. Thirteen morphological variants have been put forward which include patch, nodular, plaque, exophytic, infiltrative, telangiectatic, keloidal, ulcerative, cavernous, bullous and verrucous types. These lesions may be localized with or without visceral involvement. Oral, lymph nodes, lungs, or gastrointestinal tract may also be involved. Kaposi Sarcoma is aggressive in African patients with AIDs and tends to be more fatal (21).

The clinical outcomes differ between men and women. Females presented

lower CD4 T cell count, possessed recurrent orofacial lesions, and were less prone to have cancer-associated edema or nodular lesions as compared to males. Women also exhibited less clinical refinement than male (22).

Figure 1.1. Clinical Presentation of Kaposi Sarcoma Lesions

1.2.3 Diagnosis and Prognosis

Diagnosis is executed clinically but conformation is concluded through biopsy. Histological patterns divulge spindle cells, irregular blood vessels proliferation with slit-like forms, leukocytic infiltrate with intra and extracellular hyalin globules red blood cells extravasation. Immunohistochemical staining and polymerase chain reaction for latency-associated nuclear antigen for human herpes virus 8 should be positively Depending upon the epidemiological forms the diagnosis of Kaposi Sarcoma varies For Classic Kaposi Sarcoma in regard to a clinical presentation that involves age, local involvement, indolent course, infrequent involvement of internal organs. Examination of lymph nodes and skin is sufficient. If a patient presents symptoms of visceral compromise complementary tests are performed. No staging system is proposed for immunosuppression of Kaposi sarcoma. The lack of consensus recommended applicable criteria and the tests are the same as HIV-associated Kaposi sarcoma. There is no widely accepted staging system for HIV-associated Kaposi Sarcoma. A chest X-ray is suggested. In the era of respiratory abnormalities computed chest tomography or bronchoscopy is recommended (23).

The pathogenesis of Kaposi Sarcoma is complicated. It incorporates Human Herpes Virus 8 infection, host immune suppression, and cytokines. Kaposi Sarcoma-associated Herpes Virus has carcinogenic properties that impersonate oncogenes that contribute to cell division, modulate inflammation, inhibit apoptosis, and induce angiogenesis but do not lead to cell immortalization. Kaposi sarcoma spindle cells are not monoclonal. Human Herpes Virus 8 promotes host and viral cytokine that induces Kaposi Sarcoma cell differentiation and proliferation (24).

1.2.4 Human Herpes Virus 8

The primary cause of Kaposi sarcoma is Human Herpes Virus 8 also known as Kaposi Sarcoma associated Herpes Virus. In addition to Kaposi, Sarcoma HHV8 causes lymphoproliferative disorders that include Primary Effusion Lymphoma and

MultiCentric Castleman disease (25).

Human Herpes Virus 8 possesses a linear double-stranded DNA, an icosahedral capsid i.e 20 faces, a tegument, and an envelope (26).

The target of Kaposi Sarcoma Associated Herpes Virus includes various cell types. B cells, endothelial cells, epithelial cells, dendritic cells, fibroblasts, and monocytes. The viral introduction into endothelial cells is through the binding of KSHV to the host receptors. Integrins, cystine-glutamate transporter, tyrosine kinase receptor, and heparan sulfate are the receptors that facilitate the viral entry into the host cell. The binding introduces a signal transduction cascade that authorizes the virus to set foot in the cellular cytoplasm. Onto the viral entry, the viral envelope sanctions the delivery of the viral genome into the host cell via uncoating its envelope (27). The viral genome enters the host nucleus and forms an episome through circularizing. It will either enter into the latent stage which is the default stage or lytic stage (28).

HHV8 can establish latency in the infected cells by forming a latency locus. Latency locus includes expression of ORF71, ORF72, ORF73, and ORF12 that encode viral FLICE inhibitory protein (vFLIP), vCyclin, latency-associated nuclear protein (LANA), and kaposins respectively (29). Few microRNAs and few genes are expressed in small quantities that include K1 and K15 which encode transmembrane proteins. The body cells infected with HHV8 involve the expression of these genes when the virus is in the latent stage (30). The HHV8 genome replicates along with the host genome because the latent HHV8 genome forms an episome that is attached to the host nucleus with the help of HHV8 LANA. The infected cells are survived with the help of proteins encoded by other latent genes. The nuclear factor-Kb (NF-kB) pathway is activated to increase cellular survival with is simulated by vFLIP. HHV8 mir-K12-10a transforms NIH 3T3 cells. HHV8 miRNAs promote the reprogramming of endothelial cells along with the migration of endothelial cells induced by HHV8 miR-K12-3 activates protein kinase B(AKT). HHV8 miRK9 targets ORF50 that encodes replication and transcription activator (RTA) that prevents reactivation from the

latent stage of the virus (31).

The lytic stage of the virus is the active stage of the virus that involves the progression of the disease. At first, Immediate-Early (IE) genes are expressed that encode for RTA. RTA is a cellular promoter that promotes viral replication (32). Delayed-Early genes are expressed after IE genes that monitor viral replication after the DE phase. It involves a cycle in which viral genomes are produced and packaged in a capsid. It involves the production of infectious viruses in the late lytic phase (33). Lytic genes contribute to tumorigenesis but only a few proportions of tumor cells express lytic genes. HHV8 Vil-6, G-protein Coupled Receptor simulates cellular signaling that expresses angiogenic factors, pro-inflammatory cytokines, vascular endothelial factors, and platelet-derived growth factors (34).

HHV8 modulates the host pathways in a way to increase cellular survival and inhibits apoptosis. The viral proteins that include K15, K1, G-protein-coupled receptor(vGPCR) and IL-6 influence phosphoinositide 3-kinase (PI3K)-Protein kinase B(AKT)-Mechanistic Target of rapamycin(mTOR) pathway (35). This modulation increases cellular survival by promoting protein synthesis. ORF36 further leads to the phosphorylation of ribosomal S6 which will further increase protein synthesis. vGPCR, K1, and K15 also lead to the upregulation of the extracellular signal-regulated kinase 1(ERK1-ERK2) pathway that will lead to the activation of ribosomal S6 kinase to promote protein synthesis. Viral FLICE inhibitory protein(vFLIP)leads to the activation of nuclear factor kB (NF-kB) by the binding of NF-kB essential modulator (NEMO) (36) . This will lead to the degradation of NF-kB inhibitor and translocate the transcription factor of NF-kB in the cellular nucleus which leads to cellular survival (37).

There are multiple proteins encoded by HHV8 that modulate the host immune system. Toll-like receptors are downregulated thus inhibiting the type 1 interferon response from the host (38). HHV8 viral interferon regulatory factor (vIRF1) and replication and transcription factor (RTA) downregulate the toll-like receptors that

lead to the suppression of type 1 interferon response. The interferon response is further inhibited by ORF64 which inhibits retinoic acid-inducible gene 1(RIG1) by ubiquitylation of the RNA sensor. IRF3 and IRF7 are inhibited by viral ORF45. Viral latency-associated nuclear protein (LANA) and ORF52 inhibit the DNA sensing pathway which is GMP-AMP synthase(cGAS)-a simulator of interferon gene protein (STING) (39). IRF3 and IRF7 activate the interferon production but viral IRF1 prevents the link between STING and serine-threonine protein kinase (TBK1) thus dampening the production of interferons from the host in turn preventing host antiviral response. ORF63 inhibits the activation of IL-1B and IL-18 by inhibiting NOD-LRR, pyrin domain containing 1 (NLRP1) which consists of NLRP1, and apoptosis-associated speck-like protein containing a CARD (ASC) and pro-caspase-1 (40).

1.2.5 Treatment of Kaposi Sarcoma

To date, there is no treatment to completely eradicate HHV8 from the body. The types of treatment for Kaposi sarcoma are shown in figure 1.2

So, the main goal of oncologists and doctors is to control symptoms and inhibit the progression of the disease to limit the tumor spread and avoid impairment of organs to relieve the psychological stress (41).

There are two main types of therapies available to treat Kaposi sarcoma

The type of treatment that is confined to a specific region of the body is called local therapy (42). A treatment that uses ionizing radiations to control cancer cells in the body. In Kaposi, sarcoma Radiation is applied to the lesions on the skin to kill the cancerous cells in the lesions. The average response rate for radiotherapy is between 47% to 99%. The side effects of radiotherapy include hyperpigmentation, skin atrophy, telangiectasia, and fibrosis (43).

The removal of part of the skin that possesses the lesion. has a high recurrence rate. Extensive lesions should not be treated with this procedure. It should only be practiced on a few and well-defined lesions. Repeated surgical removals can cause

skin and functional impairments

- Local Therapy
- Systemic Therapy

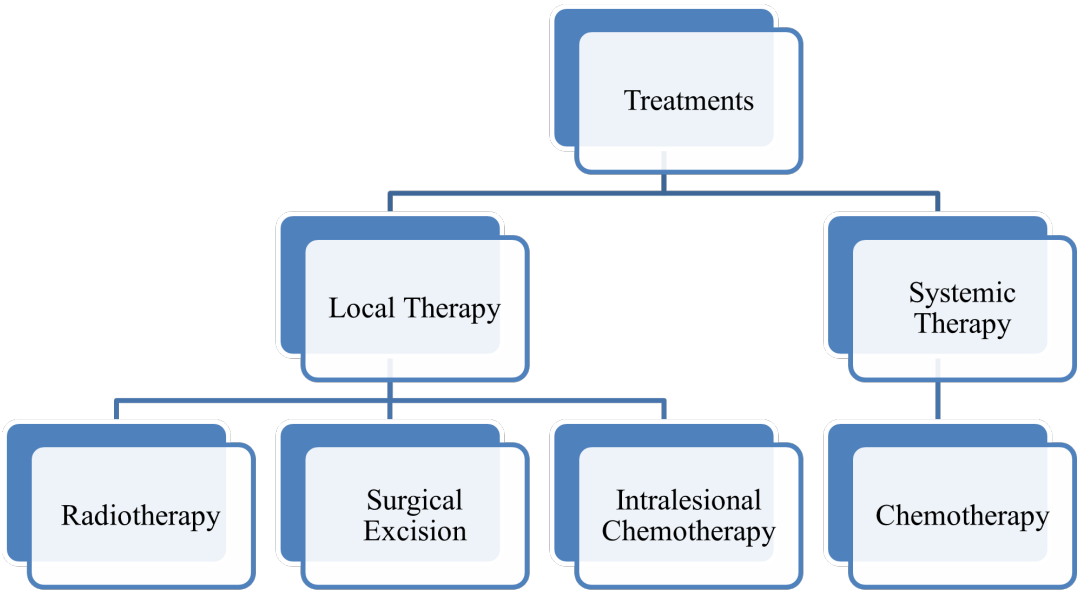

Figure 1.2. Treatment of Kaposi Sarcoma

It involves targeting lesions with chemicals. It is a historical procedure with a response rate of 70% with the use of vinblastine. When local therapy fails to treat Kaposi sarcoma systemic therapy comes into play. It involves treating the disease with the compounds that travel through the bloodstream In chemotherapy cancer cells are treated with drugs. The most used drugs are pegylated liposomal doxorubicin and paclitaxel (44).

The types of Kaposi sarcoma are classified based on prognosis and symptoms of the disease and treatment. The types of Kaposi sarcoma are treated individually based on the differences in symptoms. The treatment of the types is described below

Anti-retroviral therapy (ART) is used to treat patients with epidemic Kaposi sarcoma. Currently 6 classes of antiretroviral drugs are available

- Reverse transcriptase inhibitors (NRTIs)

- Non-nucleoside reverse transcriptase inhibitors
- Protease inhibitors
- Integrase inhibitors
- Fusion inhibitors
- CCR5 antagonists

For patients with an advanced level of disease, cART is an option which is a combination of antiretroviral drugs combined with chemotherapy. Although the incidence of epidemic Kaposi sarcoma has rapidly declined with cART. For patients with rapidly progressive disease systemic therapy is also prescribed in combination with cART. Liposomal anthracyclines and taxanes are mostly used in systemic therapy in combination with cART. Kaposi Sarcoma is not totally cured because tumors may reappear, and additional therapies may also be required (45).

To manage Kaposi sarcoma in organ transplant recipients dealing with immunosuppression is the main goal. To taper down immunosuppressive therapy to the least possible level and swapping to mammalian target of rapamycin(mTOR) inhibitors are important while keeping the allograft functional. In life-threatening conditions along with decreasing immunosuppression systemic therapy is also prescribed (46).

In the aggressive forms of Kaposi sarcoma that involve lymph nodes and visceral organs rapidly progressive disease, and local complications systemic therapy is recommended. Mostly anthracyclines are suggested. Interferon alpha in low doses can also be prescribed (47).

1.3 Biomarker

Biomarkers are the features that are objectively computed and assessed as gauges of standard biological processes, pharmacological response, and pathogenic

processes (48). Cancer biomarkers are elucidated as biochemical characteristics elaborated by tumor cells as an effect of the malignant operation. Cancer biomarkers are elaborated and detected in blood, plasma, serum, and urine. They are usual endogenous substances that are churned out at an elevated rate by cancer cells. However, they can be the output of newly switched genes that were inhibited in normal cells. Biomarkers can be proteins, molecules that are present in the tissue or circulation that designate the presence of cancer. The utility of biomarkers is the potential to provide diagnosis and prognosis of a disease (49).

Biomarkers are applicable for risk evaluation, diagnosis, treatment, and management of tumors. Molecular scanning at miRNA, RNA, DNA, and protein level can bestow the identification of the latest tumor subclasses. Biomarkers allow the characterization of patients and quantification of the compass to which the therapeutic reach the target, adjust suggested pathophysiological agency, and accomplish clinical upshot. The utility of biomarkers is measured through their sensitivity, specificity, reproducibility, and predictability (50).

The classification of biomarkers hangs on various parameters that include function and characteristics. Type 0 biomarkers are characterized based on their function compute the natural disease history and tally with known clinical indicators. The biomarkers linked to the efficacy of therapeutics are type I biomarkers. Whereas type II biomarkers are inclined to reserve clinical endpoints. Ongoing biomarkers are classified into various classes that include RNA, genetic markers, receptors, hormones, oncofetal antigens, glycoproteins, and proteins (51).

Cancer biomarkers are grouped into prediction, diagnosis, prognosis, and pharmacodynamic biomarkers. Prognostic biomarkers differentiate between benign and malignant tumors. Predictive biomarkers are employed in evaluating the potency of the administered drug. Pharmacodynamic biomarkers are utilized in determining the dose of the pharmaceutical agent in tumor patients. Diagnostic biomarkers are there at any stage of tumorigenesis (52).

1.4 Next Generation Sequencing

DNA is characterized as a genetic medium in 1944 by Oswaid and Theodore Avery. DNA is a double helix composed of four nucleotide bases that were discovered by James Watson and Francis Crick in 1953. This paved the way for the central dogma of life. DNA defines life, species, and individuals that decode the mysteries of life. DNA sequencing platforms aid biologists and researchers in molecular, cloning, breeding, pathogenic genes, and evolutionary fields. Next Generation Sequencing enormously parallel or deep sequencing are the phrases that describe the process of decoding DNA sequence (53).

Next Generation Sequencing technology is revolutionizing the potential to sort out cancer at genetic, epigenetic, and transcriptomic levels. Cataloging every mutation, copy number variations and somatic aberration at the base-pair level is a matter of days, massive parallel sequencing platforms are implemented as a means of an unbiased transcriptomic analysis of mRNA and miRNA, high throughput chromatin assays, and genome-wide methylation assay (54).

In 1970 Sanger developed the chain termination method and Maxam-Gilbert developed the fragmentation method to sequence DNA. The overview of the next-generation sequencing technology is given below in figure 1.3.This revolutionized biology by providing a medium to decipher the genome. The method put forward by Sanger is called Sanger sequencing which handles fewer toxic substances and radioisotopes as opposed to Maxam Gilbert sequencing method. It became the dominating DNA sequencing methodology for the next 30 years (55).

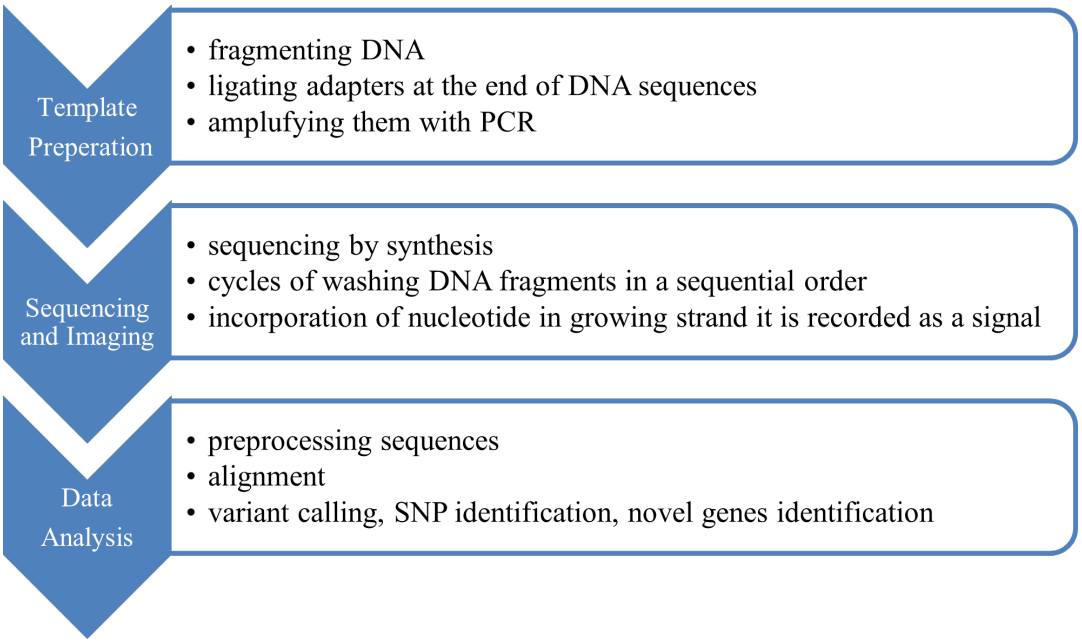

Figure 1.3. Workflow of Next Generation Sequencing

1.4.1 RNA Seq

Transcript identification and quantification of genes is a core unit of molecular biology as RNA act as an intermediate carrier between genome and proteome. RNA sequencing provides twin aspects of quantification and discovery in a single high throughput sequencing process (56).RNA seq has been implemented in several aspects of cancer research that include biomarker discovery and characterization of cancer heterogeneity and evolution, cancer microenvironment and immune therapy, drug resistance, and cancer neoantigens (57).

1.4.2 Microarray

A microarray is a tool that permits to examination expression of several genes simultaneously. DNA microarrays are microscopic glass slides that are dumped with minute spots at positions. Each spot on the glass slide possesses a known DNA or gene. The DNA molecules printed at the spot behave as a probe to investigate gene expression. The process of microarray involves gathering mRNA from the genes to be examined. mRNA is reverse transcribed to cDNA, labeling them with fluorescently

labeled dyes. Producing an amalgam of the samples to be examined and projecting them to the glass slide where the genes of known sequences are fabricated. If the sequences perceive their complementary sequence, they will hybridize and they will be scanned to examine gene expression. Microarray is a platform that allows the analysis of gene targets, expression, diagnosis, or prognosis of a disease (58).

1.5 Docking

Molecular docking has greatly contributed to structural molecular biology and structure-based drug discovery. Docking has been substantially facilitated by tremendous computer growth and ever-growing access to public molecular databases. The central objective of molecular docking is to predict molecular recognition, discover likely binding modes, and predict binding affinity energetically. Docking is dispatched between a protein and a ligand. Molecular docking has tremendous application in drug discovery structure-activity analysis, lead identification and optimization, binding hypothesis to clear the way for mutagenesis, and chemical mechanism studies (59).

1.6 Objectives

The objectives of the current analysis are as follows

- Performing NGS of individuals infected with Kaposi Sarcoma against normal individuals for the identification of potential therapeutic targets.
- Identification of up-regulated and down-regulated genes in Epidemic Kaposi Sarcoma through Next Generation Sequencing Analysis.
- Molecular Docking of existing ligands to hunt out the most effective therapy for Kaposi Sarcoma

Chapter 2

LITERATURE REVIEW

2.1 Kaposi Sarcoma

Kaposi sarcoma is described as multifocal angio proliferative disorganization of vascular endothelium. It is believed that primarily mucocutaneous tissues are affected (60) To date 4 clinical variants have been identified. They include classical, iatrogenic, epidemic, and endemic Kaposi sarcoma. (61)

Classic Kaposi sarcoma is the variant that is rare but mild originally discovered in 19 century. It is a vascular tumor predominantly influencing the lower extremities of the elderly male of Mediterranean origin (62) Traces of this variant have been spotted in other geographical locations (63). Lesions are fundamentally present as several purplish pigmented plaques on the arms, and trunk legs of elderly men primarily older than 50 (64). Lesions originate at extremities and accelerate to proximal regions. Classical Kaposi sarcoma has sluggish progress, mainly infrequent viscera, and do not need hostile therapeutics (65).

Iatrogenic Kaposi sarcoma touches on the form that is analogous to the use of immunosuppressive agents, steroids, and drugs with antitumor necrosis factor design of action in individuals with autoimmune disorders, solid organ transplantation or inflammatory conditions (66) Iatrogenic Kaposi sarcoma incorporates mucocutaneous tissues, viscera or lymph nodes and mainly affects the liver, kidney, or heart allografts (67).

Endemic Kaposi sarcoma is a disease that infects human immune deficiency virus seronegative infants and young adults in Africa. Succeeding the AIDS epidemic, the incidence of this variant distinctly surges in the pediatric community (17). The clinical presentation of this form includes local lesions on extremities, indolent skin

disease, and harsh visceral involvement. Lymphadenopathy is a very general feature of this variant and oral mucosa is not commonly affected (68).

Epidemic Kaposi sarcoma also known as immunodeficiency syndrome associated with Kaposi sarcoma is considered the most hostile form of this disorder (69). The incidence of Kaposi sarcoma is more favorable in the context of immunosuppression. Lesions of this variant multiply in number, enlarge, become more nodular and deteriorate CD4 levels. Mucocutaneous tissues are infected as multifocal plaques (65).

2.2 Human Herpes Virus 8

The causative agent of Kaposi sarcoma is Kaposi Sarcoma Associated Herpes Virus (KSHV) also called Human Herpes Virus Type 8 (HHV8). The virus can cause Multicentric Castleman Disease and Primary Effusion Lymphoma. Multicentric Castleman. This disease is defined as the massive proliferation of lymphoid tissue (70). Primary effusion Lymphoma is described as non-Hodgkin Lymphoma in body cavities (71).

HHV8 is a member of *Gammaherpesvirinae* that is the branch of herpesvirus subfamily (72). HHV8 consists of a linear double-stranded DNA that is enclosed in an icosahedral (20 faces) capsid, a tegument and an envelope (40). The genome of the virus is approximately 145kb in length which is unique and viral expressed genes with approximately 20-30kb flanked by terminal repeats (41). A total of 90 genes are expressed by the virus that produces their respective proteins and multiple noncoding RNAs (42). Given below is the list of some of the important proteins encoded by the virus along with their function in table 2.1(43).

Table 2.1. Protein and their Function of Human Herpes Virus 8

ORF	**Protein Produced**	**Function**
K1		Activation of NF-B
K2	vIL6	Blockage of IFN-a..
K3		Downregulation of MHC class I
K4	vMIP-II	CCR-3 agonist. cytokine receptor.
K5		Downregulation of MHC class I
K6	vMIP-1	Inhibition of NK cell-mediated lysis.
K7	Survivin, vMIP	Inhibits apoptosis of protein
K9	vIRF-1 B	Blockage of IFN-mediated transcriptional activation.
K10.5	vIRF-3 or LANA-2	Decrease in transcription of MHC-1
ORF16	vBcl-2	Inactivation of p53.
0RF71	vFLIP I	Inhibition of Fas-mediated apoptosis. Oncogenic
ORF72	vCyclin	Constitutive activates CDK4 and CDK6 and Cell proliferation
ORF73	LANA-1 or LANA-2	Inhibition of the tumor suppressor genes p53 and pRb.
ORF74	vGPCR	Constitutively actives GPCR. Binds IL-8. Oncogenic

2.3 Biomarkers in Kaposi Sarcoma

Biomarker detection is a core unit in bioinformatic research performed through RNA seq and microarray analysis (73).

2.3.1 RNA Seq Analysis

Over the past few decades, RNA-sequencing (RNA-seq) has been applied in the diagnosis and molecular targeted therapy of numerous diseases. It is important to elucidate the functions of novel genes in diagnosing, treating, and predicting the prognosis for Kaposi sarcoma in clinics. Several biomarkers in Kaposi sarcoma have been reported in recent years (74).

The differential detection of Kaposi Sarcoma encompasses cutaneous angiosarcoma, dermatofibrosarcoma protuberans, spindle cell hemangioma, the vascular transformation of lymph nodes, stasis dermatitis, pilar leiomyoma, and pyogenic granuloma. The epidemiological forms of Kaposi sarcoma are identified by escalating proliferation of spindle cells. HHV8 genome produces many genes that persuade and conserve the lesions. K12, K13, viral FADD like interferon converting enzyme inhibitory proteins, LANA-1 and vCyclin that induces transcription (75).

Friborg, J reported that LANA-1 is encoded by viral ORF 73. LANA-1 causes cell cycle dysfunction by degradation of p53 and inactivation of pRb (76). In a study conducted by Faris, M.etal revealed HIV encoded Tat protein, inflammatory cytokines, and peptide growth factors promoted Kaposi Sarcoma development (77). Ballon, G, and coworkers investigated vCyclin, Vflip, Bcl-2, and vIL-6disrupt cell signaling pathways and disturb apoptosis (78). Elevated concentrations of IL-1,6 and Tumor Necrosis Factor have been discovered in Kaposi Sarcoma patients by Guo, W and fellows (79).

Oncostatin M produced by T lymphocytes and macrophages has proved to be a mitogen for Kaposi sarcoma-derived spindle cells put forward by Cai, J et (27). It

has been reported that IL-1, oncostatin-M, and TNF promote Kaposi sarcoma-infected cell growth (77).

2.3.2 Microarray Analysis of Kaposi Sarcoma

Microarray is a laboratory platform for the analysis of gene expression of numerous genes simultaneously. The implementation of gene expression microarray revealed that the neoplastic cells of Kaposi sarcoma are like lymphatic endothelial cells. HHV8 infects lymphatic endothelial cells in addition to blood vascular endothelial cells. Transcriptional reprogramming of infected cells was observed through microarray profiles. This was investigated in an experiment conducted by Hsei Wei Wang and fellows (80).

2.4 Problem Statment

Kaposi sarcoma is a malignant form of skin cancer that has no effective therapy available claiming numerous lives per annum. Different origins show different behavior and genetic makeup due to differences in lifestyle. No particular treatment is available to cure Kaposi Sarcoma only a symptomatic cure is available. This analysis is based upon the identification of therapeutic targets and Kaposi sarcoma and the analysis of top-rated drugs for Kaposi sarcoma.

2.5 Objectives

Comparison of differentially expressed genes of Kaposi Sarcoma from various regions of the world

Identification of up-regulated and down-regulated genes in Kaposi Sarcoma through Next Generation Sequencing Analysis

Modification of existing therapeutics against Kaposi sarcoma

Chapter 3

MATERIALS AND METHODS

3.1 Methodology Overview

For the fulfillment of objectives illustrated in chapter 1, a workflow was plotted. The pattern of the workflow involves the identification of therapeutic targets of Kaposi sarcoma, and docking of existing ligands against the candidate protein to search for the effective and prime therapeutic against the target.

3.2 Identification of Therapeutic Targets

Identification of therapeutic targets is to distinguish differentially expressed genes. The upregulated or downregulated genes are termed differentially expressed genes. In each disease, there are a few genes whose expression level is divergent in contrary to a normal condition of a human body. Such genes are earmarked as therapeutic targets. For this motive microarray and Next Generation Sequencing are the two profound approaches. Agilent the state-of-the-art microarray, mRNA seq were implemented.

3.3 Agilent microarray

Agilent microarray involves two color sample hybridization of oligonucleotides. Two different fluorescent samples are used for hybridization and measurement of differential expression against the relative abundance of the hybridized oligonucleotide. The workflow for microarray is described in figure 3.1

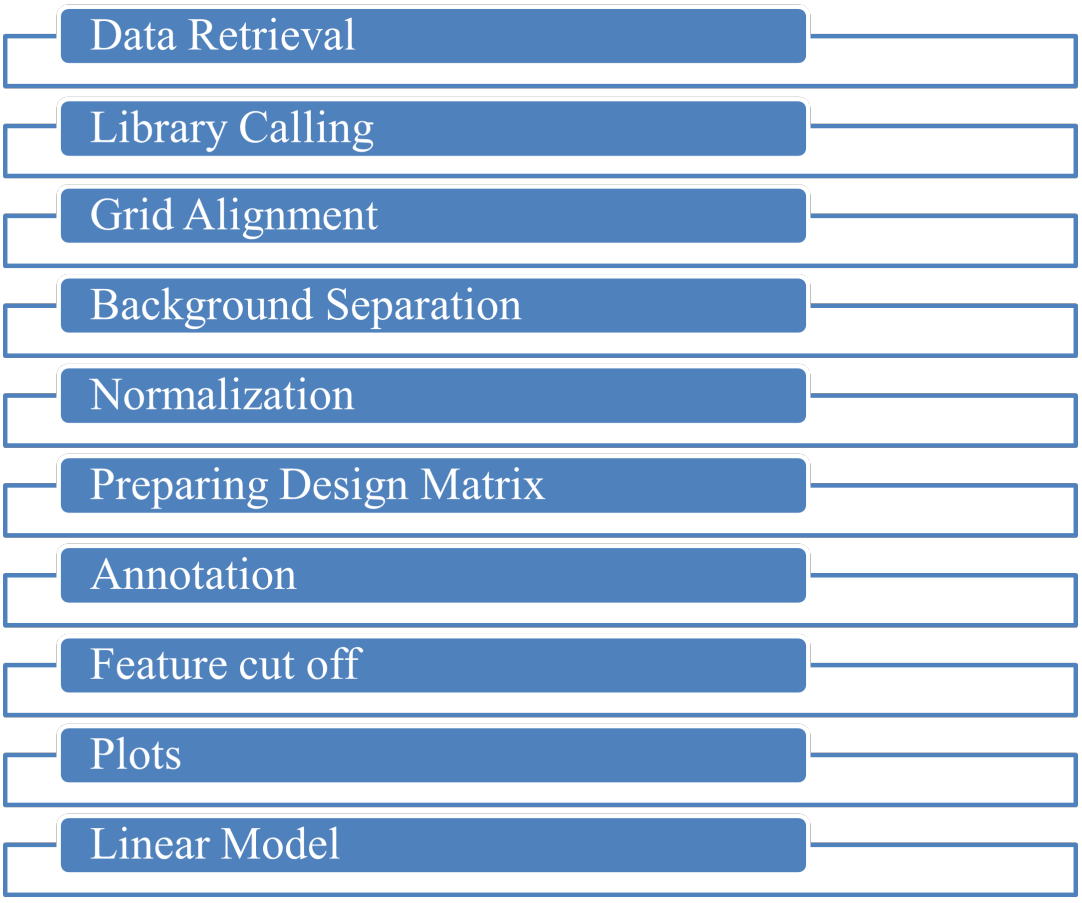

Figure 3.1. Workflow of Microarray

3.3.1 Data Retrieval

The first step for the measurement of differential expression is the retrieval of the dataset. There are two ways to retrieve the data. In wet lab by direct sequencing of organisms that are to be analyzed or by using the databases that contain the required sequences.

In this case, GEO database was used to retrieve the data.

3.3.2 GEO Database

GEO abbreviated as Gene Expression Omnibus is an online, international communal repository. It freely distributes microarray and next-generation sequencing and various forms of high throughput genomics data put forward by the research community. The main objectives of GEO are

- Data Organization
- Submission Guide
- Query Analysis

It has a platform record that is composed of a description of the sequencer or array, and a table describing the array template. Every platform is assigned a unique GEO accession ID (GPLxxx). A sample record defines the conditions in which every individual sample was handled and manipulated and the measurement of abundance it had gone through. Each sample is designated with a unique GEO accession number (GSMxxx). The series record combines related samples and furnishes the focal point and interpretation of the whole analysis. It is composed of tables that describe the summary conclusions. A dataset represents an organized collection of statistically and biologically comparable samples on the basis of the GEO suite of data representation and analysis tools. A Profile provides a measurement of gene expression of every single gene all across samples.GEO provides two platforms GEO datasets and GEO

profiles. GEO datasets provide a study-based platform for the researchers to perceive the study of their interest. However, GEO profiles offer a gene-level database that provides the expression measurement of genes.

The link to GEO database is: https://www.ncbi.nlm.nih.gov/geo/

3.3.3 Library Calling

Microarray analysis for differential expression is mainly executed by employing various packages of R language. It is based on BiocManager project built in R language that provides numerous tools for microarray data analysis. The required libraries for the analysis are installed in R language using the command "BiocManager::install("library name")" and called "library(library name)" respectively. Multiple libraries are required for this purpose table 3.1 provides a list of libraries along with their corresponding function. The libraries that will be used for microarray analysis are discussed below

Table 3.1. Table 3.1: Microarray Libraries and their Activities

Library	**Function**
Convert	Coerce method for microarray data
GEOquery	Extracts the data from GEO database
Geneplotter	Graphics functions
Limma	Linear models for microarray data
Xlsx	Programmatic control of excel files
Openxlsx	Simplifies the creation of excel files with a highly interactive interface
Biobase	Provides base functions
oligo	Preprocessing for microarray data
Simpleaffy	Functions for reading phenotypic files
OligoClasses	Classes for high throughput arrays
Ggplot2	Data visualization
ArrayQualityMetrics	Diagnostic plots for microarray data

3.3.4 Preprocessing

Preprocessing of microarray data is regulating raw intensities of microarray so that their biological significance starts to make sense. the steps included in microarray data preprocessing are mentioned below.

• Grid alignment represents microarray data as a 2D array of spots by registering uneven intensities with the 2D image content.

• Background correction adjusts data for ambient intensities revolving around each feature of microarray data.

• Normalization is fine-tuning microarray data that is a result of technical biases as opposed to biological biases

The microarray data is preprocessed to exclude all the technical biases. The biological meaning of the data can be interpreted through statistical processes

3.3.5 Preparing Design Matrix

Microarray data after preprocessing is all set for analysis that involves the construction of a design matrix. Design matrix specified RNA samples have pertained to each channel on every array. The design matrix leads to the construction of a contrast matrix that allows the comparison between RNA samples.

3.3.6 Annotation

Microarray data should be annotated. Annotation is the description of microarray samples. Microarray data is annotated that depicts which biological entities are portrayed on microarray chip. This step is essential for inferring the biological significance of the data.

3.3.7 FeatureCut off

Differentially expressed genes are identified by choosing an appropriate cut-off. The crucial goal of microarray is to distiguish differentially expressed genes. The identification is based on plumping of a threshold value and ruling out all the genes below it because their expression is normal. The genes crossing the cut-off value corresponds to abnormal expressions and they are considered differentially expressed gene.

3.3.8 Plots

The microarray data is plotted in the form of a heat map, histogram, and boxplot. To visualize all microarray intensities of all the genes R provides a suitable package that plots the gene expression and they can be analyzed in the forms of various plots.

3.3.9 Linear Model

After all the above-mentioned steps are performed linear model is built based upon the design and contrast matrix. It helps in the analysis for comparing the intensities.

3.4 mRNA Seq

mRNA Seq is a 6-step process that was performed for the identification of DEGs. The general workflow for the identification of DEGs is shown in Figure 3.2. The steps of mRNA seq include data retrieval, quality control, adapter trimming, mapping genome to the reference, counting reads associated to the genes, and statistical analysis for identification of DEGs.

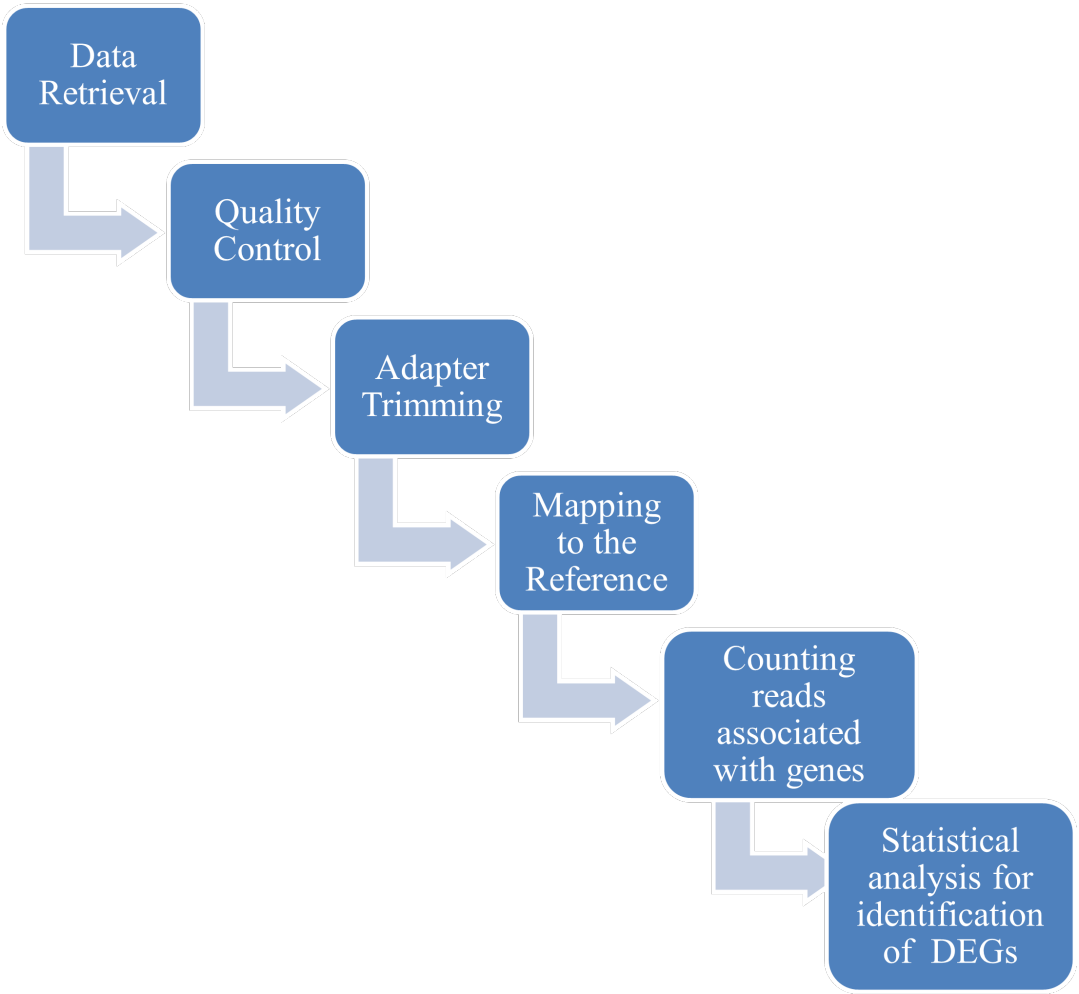

Figure 3.2. Workflow of mRNA Seq

The workflow that was adopted for the identification of DEGs along with the tools used is shown in figure 3.3. The tools employed include GEO and EMBL for data retrieval, FASTQC for Quality control, HISAT2 for alignment, MarkDuplicates for marking duplicates, RmDup to remove duplicates, StringTie for quantification of transcripts and Ballgown for DEGs identification

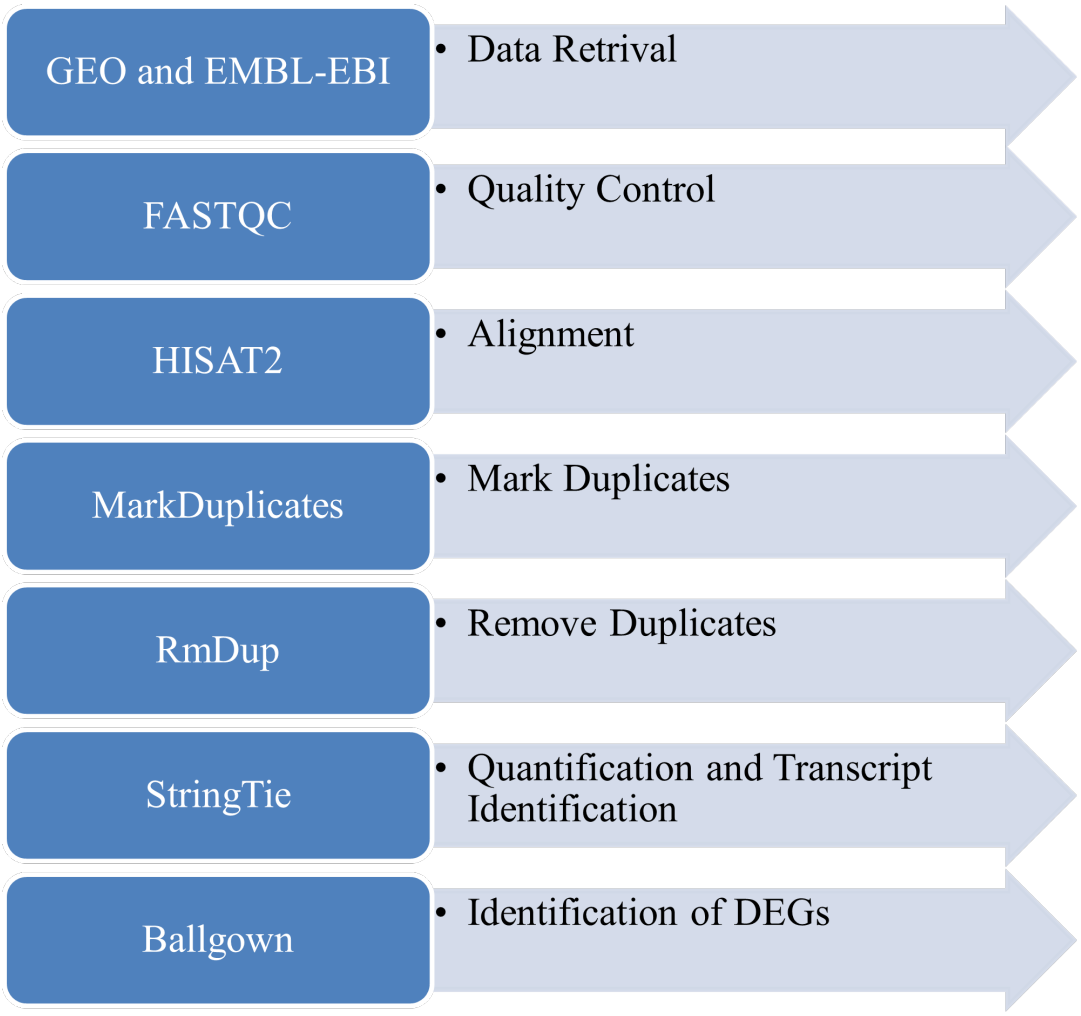

Figure 3.3. Workflow of mRNA Seq and Tools

3.4.1 Data Retrieval

The very first step of mRNA Seq is the retrieval of data. There are numerous ways to get genomic sequences in the wet lab or dry lab. In wet lab direct sequencing is possible on the other hand in a dry lab a variety of databases are available that contain genomic sequences.

For this research, two data sets were retrieved from two databases “GEO” and “EMBL-EBI”.

EMBL-EBI

EMBL European Bioinformatics Institute is an international, interdisciplinary, and innovative source of data for life sciences. EMBL is abbreviated as European Molecular Biology Laboratory. EMBL-EBI provides free bioinformatics services and data, research programs in personal genomics and computational translational bioinformatics, hands-on bioinformatics training, industrial partnership, and ELIXER hub. The link to the EMBL-EBI database is: https://www.ebi.ac.uk/

3.4.2 Quality Control

Posterior to data retrieval, the data is progressed through quality control operation. Quality control is performed to scrutinize the quality of data and remove technical biases. In NGS data there are abundant factors that influence the analysis. The flaws can be in the sequencing platform, the sequencer, and the nucleotides. The major obstacle is the assurance of the reliability of the incorporated nucleotide in the corresponding read. To perform this task FastQC was employed for the investigation of quality of the reads.

FastQC FastQC examines the quality of raw and sequenced reads processed from the sequencing pipelines. It gives a set of analyses that provides insight into the quality of the sequencing process. The main task of FastQC is

- Imports data in a variety of file formats like SAM, BAM and FastQ

- Provides an insight into the problematic region of the data
- Visual representation to assess the reads in the form of graphs and plots
- Exports the quality results into an HTML report
- Automatic report generation is performed offline without processing interactive application

3.4.3 Alignment

After deriving data from preprocessing the succeeding task is to align the reads to the reference genome. This is performed to observe the resemblance between the data that we are analyzing and the reference genome. This is done to reflect light on genetic diversity among the population for the identification of mutations, isoforms, SNPs HISAT2 was employed to perform the alignment between the reads under study and the reference genome which is Hg39.

HISAT2

HISAT2 is a graph-based genome alignment algorithm. It is a splice-aware aligner that employs a graph-based strategy, alignment algorithm, and indexing algorithm. It implements hierarchical indexing which incorporates global index and local index.

- Global index covers the whole genome
- Local indexes encompass 56kb regions covering the whole genome

Initially, HISAT2 uses the global index to map the read to the genome and further uses the local index to map the read that refines the location of the read mapping. The local index comes into play where the splicing strategy comes in. As the local index covers short genomic regions it can easily detect where the read is spliced. This two-step indexing algorithm helps makes HISAT2 a fast and memory-efficient algorithm.

3.4.4 MarkDuplicates and Duplicates Removal

RNA seq depends on PCR for the amplification of oligonucleotides for the sequencing process to initiate. PCR results in the expansion of reads but not every single read is amplified equally. This results in some reads being overrepresented giving an insight that the expression level of the resulting gene is higher in quantity. When it is not the case. Such sequences are termed overrepresented sequences. Such sequences should be excluded from the analysis. Computational removal of these duplicates based on their alignment coordinates is important.These duplicates were marked and then removed by mark duplicates and Rmdup respectively.

MarkDuplicates and RmDup

MarkDuplicates tags duplicate reads in SAM/BAM files. The tools work by comparing sequences at the 5 prime positions of both reads in SAM/BAM files. After the duplicate reads are accumulated, they are distinguished between primary and duplicate reads by ranking the sum of their base quality score. MarkDuplicates locate the duplicated reads and rank them based upon their base quality, Rmdup then removes the low base quality overrepresented reads.

3.4.5 Quantification and Transcript Identification

In the human body, introns are removed and selective exons are retained from pre-mRNA resulting in RNA splicing. This splicing leads to different versions of RNA referred to as transcripts and isoforms. The majority of the human genome is alternatively spliced which leads to the expression of a single gene to different types of proteins. These proteins have unique functions and involve in different pathways and are implicated in numerous diseases involving cancer.

Identification and quantification of transcripts are of paramount importance because they are translated into different proteins. Isoforms' presence and quantity vary in different samples which leads to novel biomarkers. Different biomarkers lead to a level of a biological process that is not detectable at the gene level. Transcript

Identification and quantification was done through String Tie.

StringTie

StringTie identifies gene isoforms in the data and computes their abundance. StringTie assembles transcript fragments using splice map reads and infers isoforms for each sample. StingTie algorithm implements network flow algorithm and denovo approach. SringTie works by mapping the reads to the reference genome and constructing a graph for all the possible isoforms of the gene. The splice graph possesses nodes representing exons and the path between the nodes represents splicing sites. The reads are classified into clusters, constructing a graph for each cluster leading to transcript identification. It creates an individual flow to estimate expression using maximum flow.

3.4.6 Identification of Differentially Expressed Genes

The quantification of the expression level of a gene is followed by a statistical procedure to test the difference between the quantification values of the samples. To consider a gene differentially expressed there should be a statistically significant difference between the read counts or quantification values between two experimental conditions. Statistical methods are employed for gene expression. Differential gene expression patterns are approximated by statistical methods. The genes are short-listed based upon the expression change cutoff and score threshold. Ballgown was used to identify differentially expressed genes.

Ballgown

Ballgown is an R package designed to perform statistical comparisons between transcripts. Its underlying model in general linear model approach. It contains built-in visualization routines for displaying transcript abundance and structure.

3.5 Comparative Analysis

Microarray and mRNA seq results in DEGs. The ultimate goal is to detect common genes among all the datasets of microarray and mRNA seq. The utmost purpose of extracting common genes is pathway analysis and selection of protein. The tool that was employed to extract common genes among all the datasets was Draw Venn Diagram

Draw Venn Diagram

Draw Venn Diagram is an online tool that calculates a list of intersecting elements. A textual output is generated that indicates intersecting and unique elements. When the lists are lesser than 7 a graphical output is also generated in the shape of a Venn diagram. The user can choose between non-symmetric and symmetric diagrams. The intersection between a maximum of 30 lists can be calculated. The graphical output can be downloaded in PNG or SVG format. The link to visit Draw Venn Diagram is: https://bioinformatics.psb.ugent.be/webtools/Venn/

3.6 Pathway Analysis

A schematic representation of an organized, well-characterized segment of physiological machinery at the molecular level. In general, a pathway is described as a model where an extracellular signaling molecule results in the activation of a specific receptor triggering a chain of molecular interactions. The ultimate goal of pathway analysis is the identification of genes within a known pathway corresponding to a specific pathological condition. The set of common genes identified among all the datasets is all set for pathway analysis for the identification of genes/ proteins that can be used for further analysis.

Reactome was for pathway analysis.

Reactome

Reactome is a freely available open access, peer-reviewed pathway database.

The major unit of reactome is the reaction. The fundamental unit of Reactome is a reaction where biological entities participating in a biochemical reaction configure a network of biological interactions and are classified in reaction. The pathways that are possessed by Reactome include intermediary, metabolism, classical, transcriptional, apoptosis, regulation and disease. The link to visit Reactome is: https://reactome.org/When it comes to pathway analysis Reactome a list of DEGs is provided as an input and it gives the output in the form of a pathways list with the entities involved and p-value.

3.6.1 Selection of Therapeutic Target

Statistically, the significant pathway is shortlisted and the entities participating in the pathway are analyzed. The analysis of the genes participating in the pathway results is the selection of 1 gene that has

- Experimentally determined structure with a good resolution
- Domains
- Reference from literature about the involvement of the gene is the corresponding disease

The gene possessing the above-mentioned credentials is brought forward in the analysis

3.7 Docking

Docking is a type of bioinformatics modeling that allows the interaction of two or more two molecules to form a stable adduct. Molecular docking involves a target and a ligand. Docking predicts the 3D structure of the complex. A variety of adducts are formed and they are queued using the scoring function based upon the total energy of the system. The general workflow of docking is represented in Fig 3.4

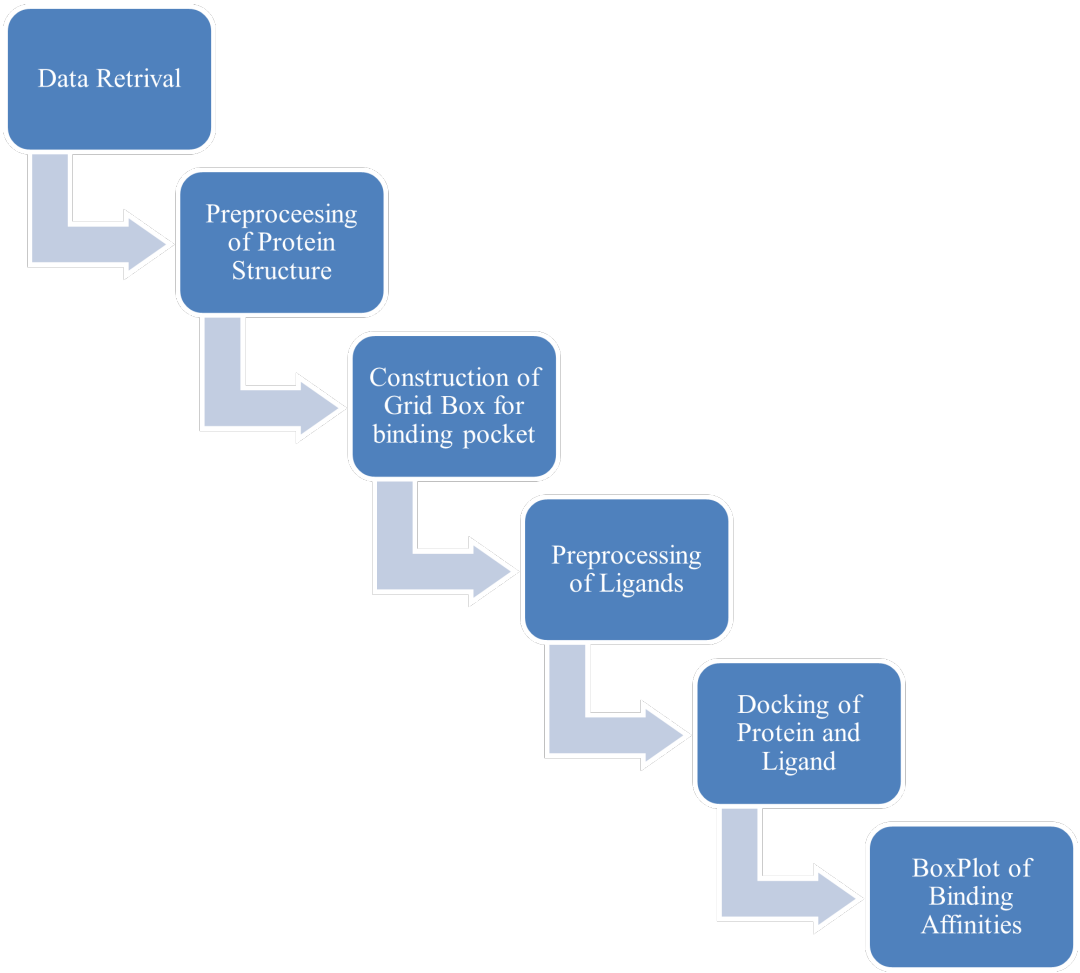

Figure 3.4. Overview of Docking Methodology

3.7.1 Data Retrieval

Molecular docking requires the 3D structure of the target (protein) and ligands in the form of PDBQT. The 3D structure of the protein was extracted from UniProt and ligands from databases like DrugBank and BindingDB.

UniProt

UniProt is also known as Universal Protein Resource is a database for protein sequence and annotation. The UniProt databases are a collection of UniProt KnowledgeBase, UniProt Reference Clusters and UniProt archive. The 3D structure of molecular docking was retrieved from UniProt The link to UniProt is: https://www.uniprot.org/

BindingDB

Binding Database (Binding DB) publics experimental data on the non-covalent interactions of the molecules in solution with the help of the world wide web. The core unit of bimolecular systems, information of host-guest supramolecular systems is added over time. The aim of Binding DB is to ease drug discovery, self-assembling system assembly and predictive compute model development. The ligands smilies were retrieved from Binding DB The link to visit Binding DB is: http://bdb2.ucsd.edu/bind/index.jsp

DrugBank

DrugBank is an online, free, comprehensive database on drugs and drug targets where bioinformatics meets cheminformatics. It possesses chemical, pharmaceutical and pharmacological data. The ligand similes were extracted from DrugBank: https://go.drugbank.com/

3.7.2 Preprocessing of Protein Structure

The preprocessing of the molecular structure of the target is illustrated in fig 3.5.

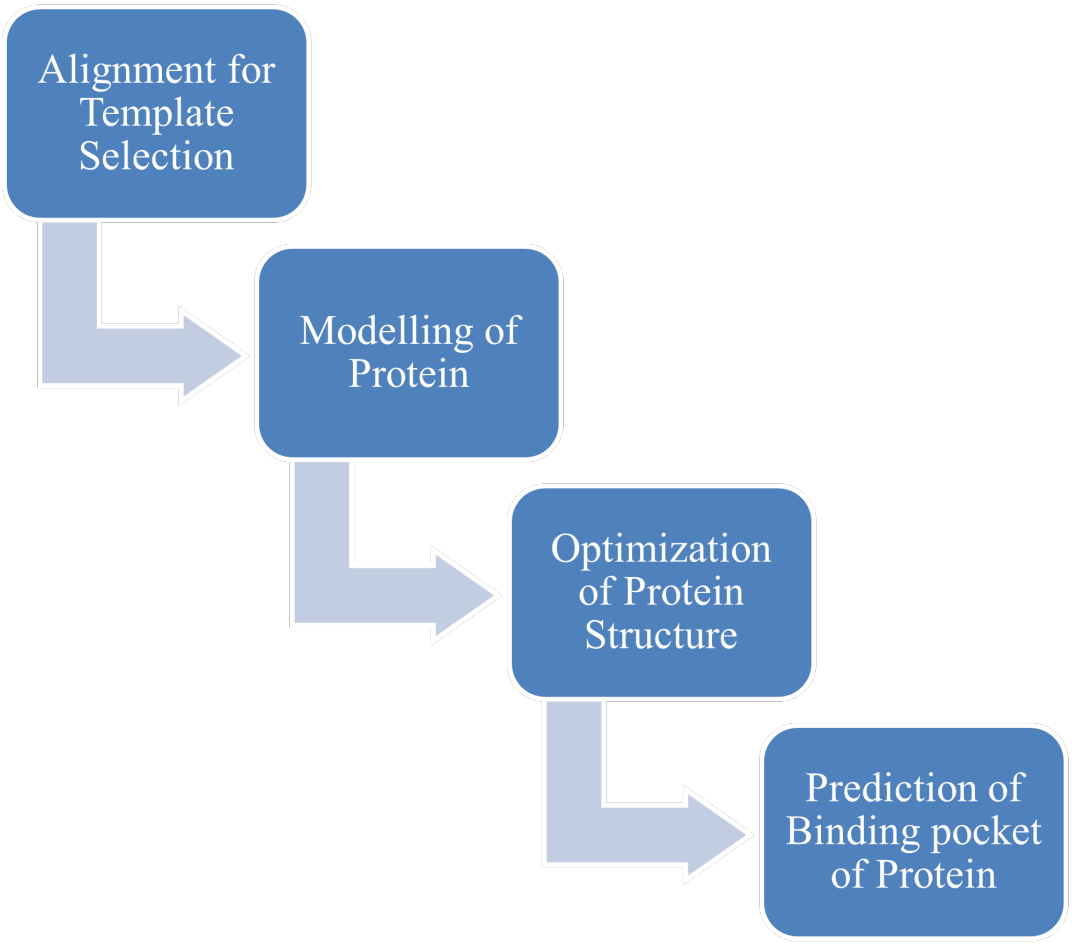

Figure 3.5. Preprocessing Methodology

Alignment of Target for Template Selection

Once the structure for protein is obtained from UniProt the next task is the selection of a suitable template for the target. BLASTp of the protein domain is performed.

BLAST

BLAST investigates regions of similarity between the biological sequences. The tool compares sequences of protein or nucleotide to a sequence database and computes the statistical significance. BLAST searches for a suitable template for the target in the relevant databases and outputs the templates with the maximum identity.

Modeling of Protein

Based on the template sequence the next task is modeling the protein based on the template sequence. The Swiss Model was employed to model protein.

Swiss Model

Swiss Model is an online and automatic homology modeling server accessed through Expasy Web Server. The purpose of Swiss modeling is to make homology modeling of proteins in all life sciences possible. The link to the Swiss Model is: https://swissmodel.expasy.org/

Optimization of Protein Structure

The model of the protein once generated is optimized. Optimization is done by adjusting a suitable pH of the protein structure and removing water molecules. Optimization was done by Play Molecule

Play Molecule

Play Molecule is a freely available, online tool that allows target identification and validation, lead discovery, and identification. Link to visit play molecule website: https://www.playmolecule.com/

Prediction of Binding Pocket

The binding pocket is the cavity on the protein surface or the interior of the protein that possesses suitable characteristics for binding with the ligand. The

prediction of the domain of protein where the ligand will interact is a fundamental step in molecular docking. This step specifies the residues of interaction between the target and the ligand. DoGSiteScorer was used for the prediction of the binding domain of the protein.

DoGSiteScorer

Automated binding domain detection and analysis software that allows the detection of potential pockets in the protein structure. In addition, the global properties, shape, size, and chemical properties of the predicted domains are calculated. The drug ability score is calculated based upon a linear combination of the descriptors that include hydrophobicity, enclosure, and volume. The link to visit DoGSiteScorer is: https://bio.tools/dogsitescorer

3.7.3 Preprocessing of Ligands

The preprocessing of ligands involves ligands similes conversion to mol2, 3D coordinates generation and addition of polar hydrogen atoms. Furthermore, from mol2 format, the ligands are converted to pdbqt format. The ligands are downloaded in the form of similes from the databases. They necessarily should be metamorphosed to mol2 format, 3D coordinates should be generated, and hydrogens should be added. So that they can be put forward for molecular docking for their conversion to pdbqt through autodock. OpenBabel was employed for the preprocessing of ligands.

OpenBabel

OpenBabel is an online and free file format converter that converts the files from one format to nearly all the formats and jumps from one program to another. The link to visit OpenBabel is: http://www.cheminfo.org/Chemistry/Cheminformatics/FormatConverter/inde

AutoDock

AutoDock is a suite that allows automatic docking options. The core task of autodock involves the prediction of small molecules such as drug candidates or substrates, and interaction with a receptor of predicted 3d structure. AutoDock distri-

bution involves two software AutoDock 4 and AutoDock Vina.AutoDock 4 is based upon two main tools, AutoDock allows the docking of the target to a set of ligands to a collection of candidate grids. Autogrid precalculated the grids where the docking will be performed. In addition to docking, the automated affinity grids can be analyzed. AutoDock Vina calculates the grids internally for the atoms that are required and this is performed virtually. AutoDock tools possess GUI that aids to organize bonds that are treated as rotatable in the docking process.

3.7.4 Docking of Protein and Ligand

Docking involves the interaction of proteins and ligand. Where the residues of the active site of the ligand will interact with the binding pocket of targets. In this operation, numerous ligands will be docked to the target. This assessment helps to analyze the binding affinities of various ligands. Docking was performed with the help of autodock vina.

AutoDock Vina

AutoDock Vina is an open-source tool that performs docking between the target and the ligand. AutoDock Vina usually implements AutoDock 4 tools that are much faster, accurate and easy to use. The depiction of the output should not have a statistical bias referring to the conformation of the input. The tools detect the syntactic accuracy of the input and report the errors. It verifies the invariance of covalent bond lengths automatically in the output. Artificial restrictions like a number of atoms in the input, the size of the search space, exhaustiveness of the search and the number of torsions are avoided by the tool.

3.7.5 Boxplot of Binding Affinities

The ligands when docked with the target through autodock vina it yields an output of binding affinities. Lesser the binding affinity of the ligand the better. For the analysis of binding affinities of all the ligands visually boxplot is the foremost course

of action. R provides a package to construct the boxplot of all the ligands through the package called ggplot2. It takes an xlsx files consisting of the binding affinities of all the ligands and displays it in the form of boxplot. R language was used for the construction of boxplot of the binding affinities of the ligands.

Chapter 4

RESULTS

The therapeutic targets of Kaposi sarcoma were identified through microarray, mRNA Seq and miRNA Seq. Therapeutic targets are the differentially expressed genes that are investigated through microarray, mRNA Seq and miRNA Seq approaches. The shortlisted candidate of the therapeutic target was docked to its respective ligand.

4.1 Microarray

Microarray is a procedure for the identification of differentially expressed genes that are termed as therapeutic targets. The outturn of microarray analysis of Kaposi sarcoma is discussed below.

4.1.1 Data Retrieval

The dataset with the GEO accession ID of "GSE16353" was retrieved from GEO database. The parameters of the dataset implemented are shown in table 4.1.

Table 4.1. Parameters GSE16353

Parameter	**Description**
Title	The profile of cellular and KSHV microRNA in AIDS-KS biopsies (and normal skin biopsied)
Organism	Homo Sapiens
Experiment Type	Non-coding RNA profiling by array
Country	United Kingdom
Platform	GPL8617 Agilent-016436 Human miRNA microarray.
No of Samples	11 (3 control and 8 disease)

The sample IDs are shown in Table 4.2

Table 4.2. Sample IDs and Phenotype

Sample ID	Phenotype
GSM410655	Skin-biopsy-patient1-repA
GSM410656	Skin-biopsy-patient2-repA
GSM410657	Skin-biopsy-patient3-repA
GSM410658	AIDS_KS-biospsy-patient4-repA
GSM410659	AIDS_KS-biospsy-patient4-repB
GSM410660	AIDS_KS-biospsy-patient5-repA
GSM410661	AIDS_KS-biospsy-patient6-repB
GSM410662	AIDS_KS-biospsy-patient6-repA
GSM410663	AIDS_KS-biospsy-patient7-repB
GSM410664	AIDS_KS-biospsy-patient7-repA
GSM410665	AIDS_KS-biospsy-patient8-repB

4.1.2 Plots

The data retrieval leads to microarray analysis that was performed in R language. The dominant feature of using R Studio is that it provides multiple features that empower the user to visualize the results in multiple ways. Numerous types of graphs and plots are designed to visually observe the results. The plots built are shown down below.

Boxplot is s systemized strategy of putting on view the dispersal of data on five feature summaries. The five-number summary includes minimum, second quartile, median, third quartile and maximum. Figure 4.1 represents the distribution of genes across the samples in the dataset of Kaposi sarcoma. The first sample GSM410655 is the least distributed sample which implies the minimum and maximum values of the genes are showing the least differences. Whereas the sample GSM410661 is showing the maximum outburst of distribution in regard to gene expression. This leads to the observation of abnormal expression of genes across the sample. The extensive analysis leads to the observation that sample GSM410655 is of normal skin biopsy that shows the least distribution and normal expression. However, GSM41066 is of AIDS-KS biopsy that shows the maximum spread of abnormal DEGs.

Heatplot is a technique to visualize data that represents the immensity of a circumstance in the form of color in 2 dimensions. The disparity in color gives obvious hints to the reader about the clustering of data. In figure 4.2 the distribution of genes in their respective samples The samples GSM410665, GSM410655, and GSM410661 depict the least variation in intensity that represents neutral gene expression. GSM410660, GSM410656, and GSM410657 exhibit fairly hotter colors that are termed as moderately high expression of a few genes across the samples. GSM410662, GSM410663, and GSM410664 produce the darkest hues that reveal the maximum upregulation of genes across these samples.

A histogram represents the distribution of data where the data is classified in continuous ranges. Below in figure 4.3a the distribution of gene expression across the sample is shown in the form of a histogram. The upregulated and downregulated genes can be identified and established on the stretch of the bar that depicts the median of gene expression.

Principal Component Analysis portrays the recommendation of dimensional reduction. It transforms huge datasets into smaller sets still calibrating enormous information. In PCA plot the samples that cluster together represent similar characteristics In figure 4.4 samples lying in close vicinity to each other represent similar gene expressions.

The Relative Log Expression plots are devised for visualizing outbursts of variation in large sets of data of microarray samples. The variation of gene expression across the samples compared to one another is shown in figure 4.5. It represents the variation of expression of GMS410661 is the highest corresponding to every sample analyzed.

Volcano plot is a scatter plot that depicts the statistical significance versus the enormity of change termed as fold change. It empowers the nimble identification of genes with huge fold changes and is statistically significant. Upregulated genes lie in the top right corner whereas downregulated genes originate in the top left corner. The

volcano plot in figure 4.6 represents the extent of differential expression of genes in up and down-regulated ends of the spectrum of microarray

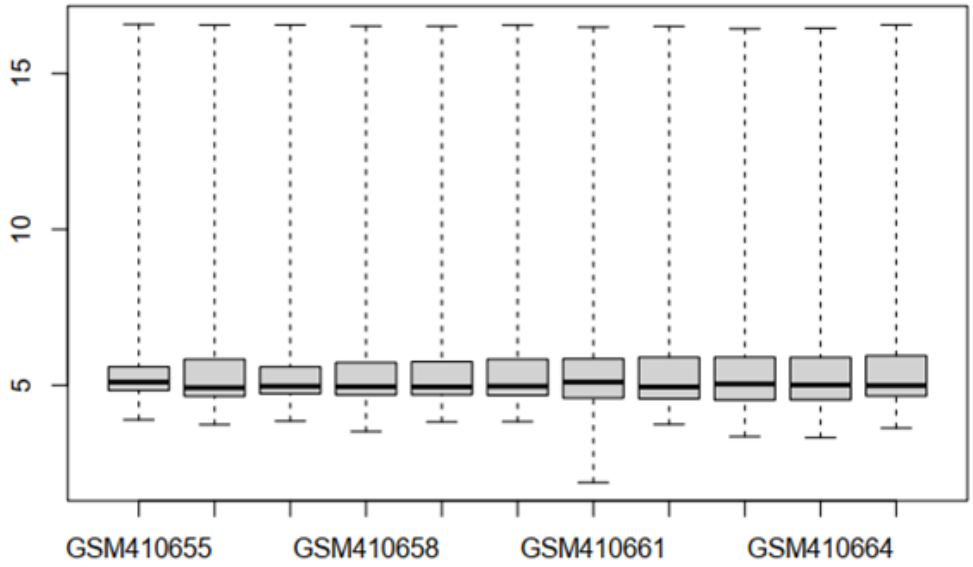

Figure 4.1. Boxplot of Intensities of Data

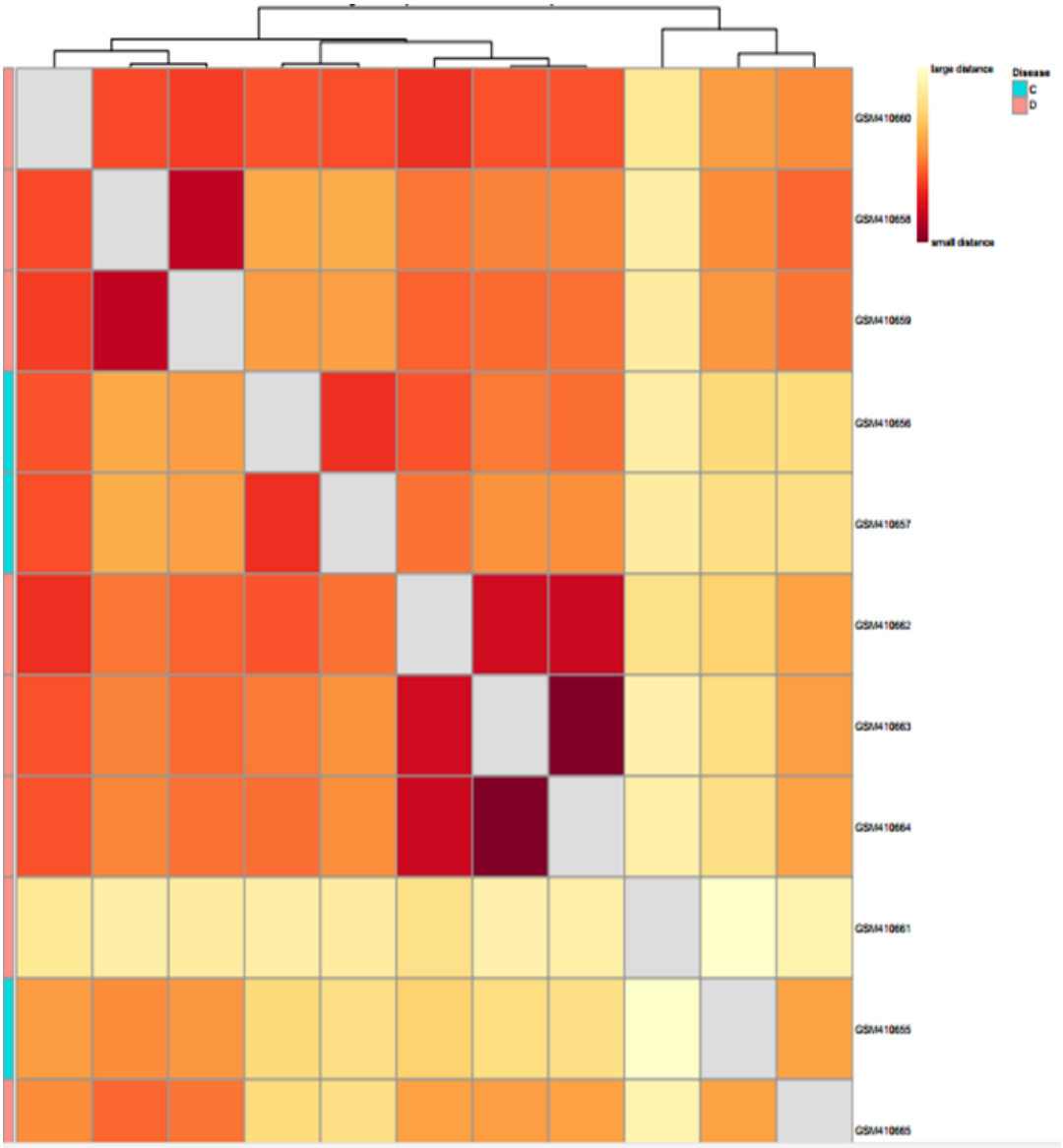

Figure 4.2. Heatplot of Intensities of Samples

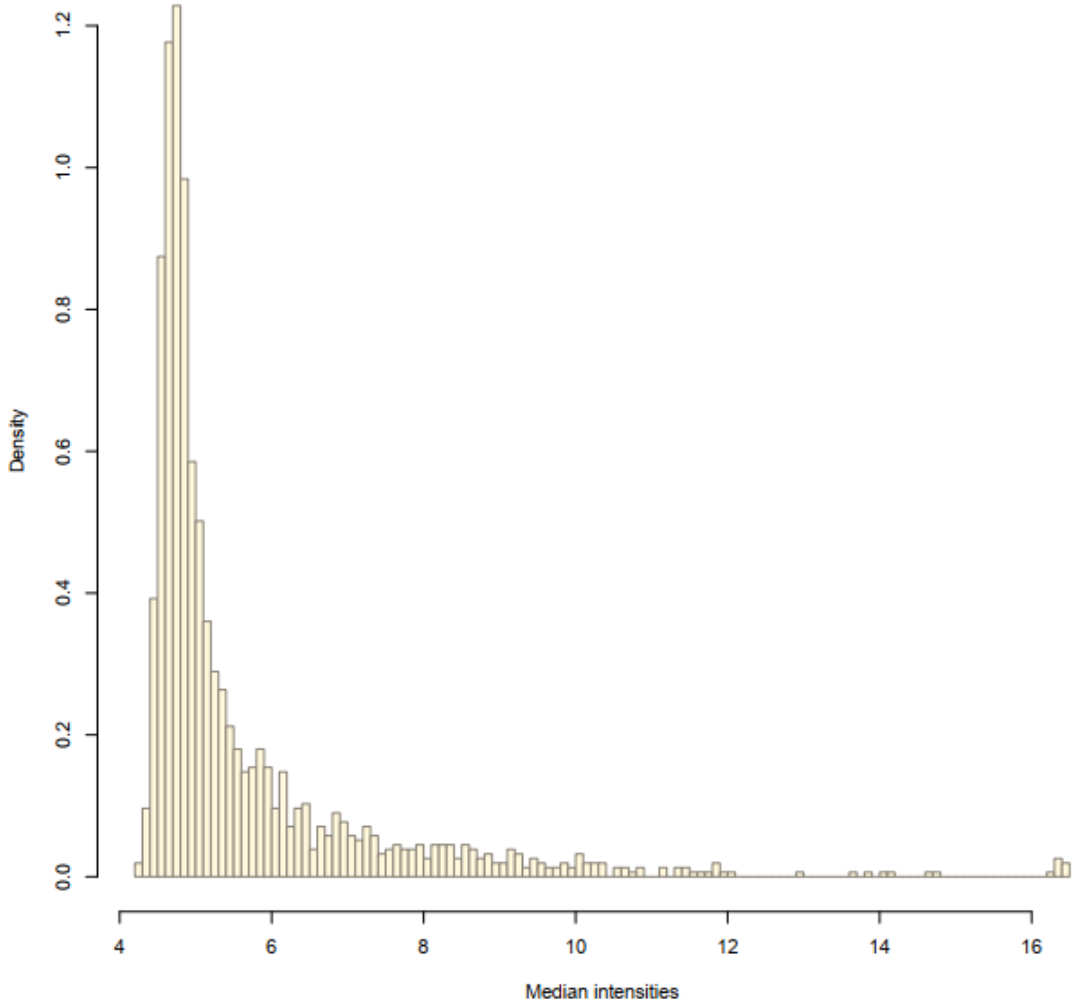

Figure 4.3. Histogram of Median Intensities

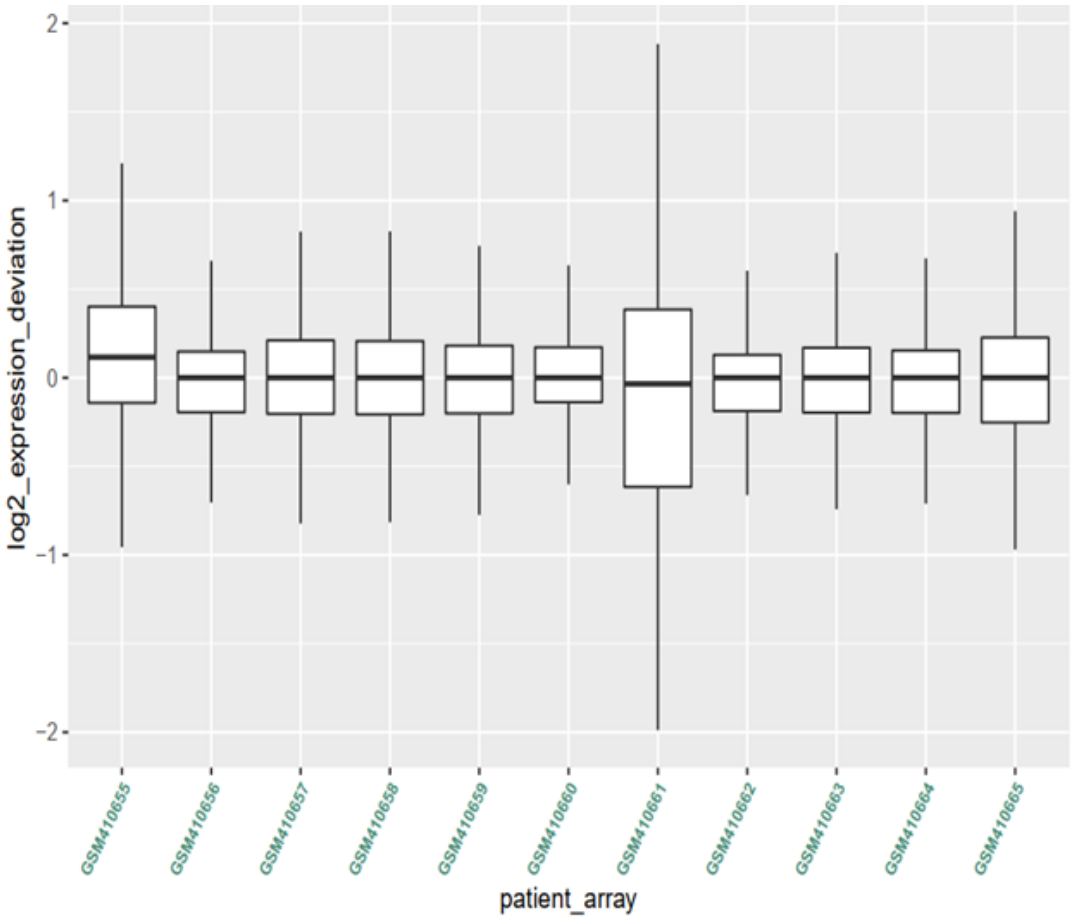

Figure 4.4. RLE Plot of Expression Data

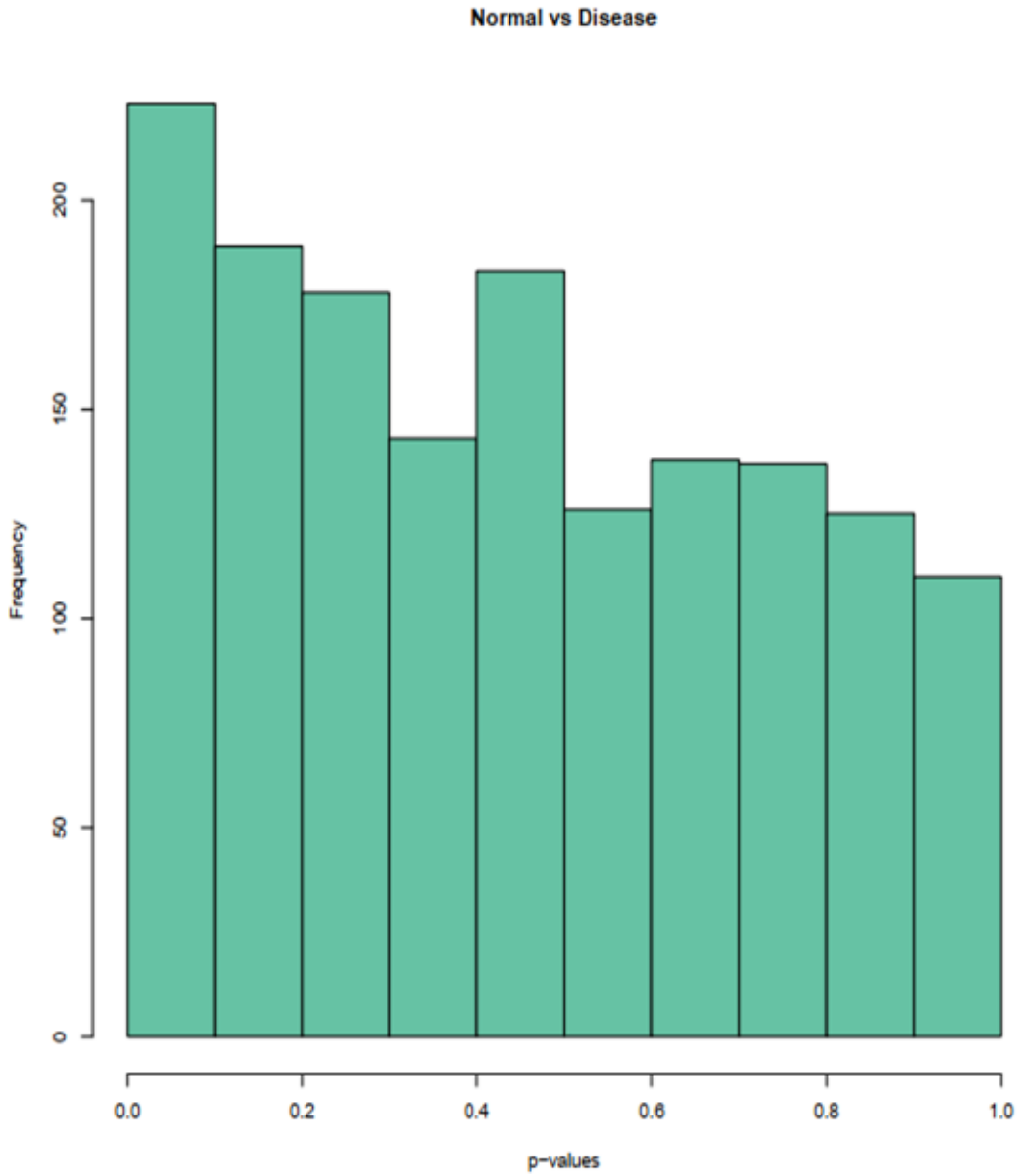

Figure 4.5. Histogram of Variation of Gene Expression

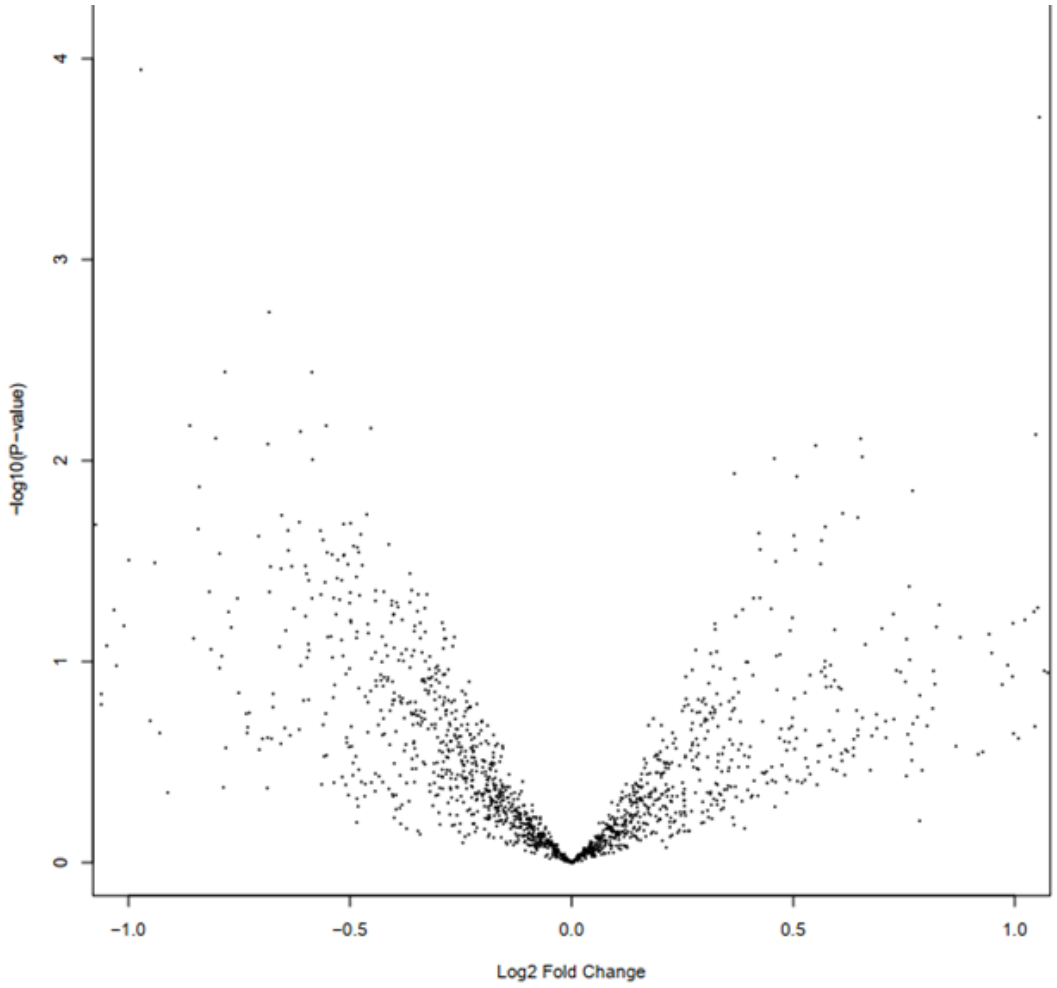

Figure 4.6. Volcano Plot of Upregulated and Downregulated Genes

4.2 mRNA Seq

mRNA seq is a procedure for the identification of differentially expressed genes. In addition to microarray, mRNA seq was also protruded for the identification of therapeutic targets of Kaposi sarcoma. Two datasets RNA seq were manipulated for the identification of therapeutic targets.

4.2.1 mRNA Seq 1

Data Retrieval

GEO database was maneuvered for the extraction of dataset of Kaposi sarcoma. The dataset with the GEO accession ID "GSE147704" was manipulated for analysis. The attributes of the data are shown in table 4.3

Table 4.3. Attributes of mRNA Seq dataset

Parameter	**Description**
Title	comparative transcriptomic analysis of endemic and epidemic Kaposi sarcoma lesions and the secondary role of HIV-1 in KS pathogenesis.
Organism	Homo Sapiens
Experiment Type	Expression profiling by high throughput sequencing
Country	USA
Platform	GPL16791 Illumina HiSeq 2500 (Homo Sapiens) GPL18573 Illumina NextSeq 500 (Homo Sapiens)
No of Samples	51

Out of 51 samples, 15 were used for analysis that exhibited three different phenotypes. The sample IDs put forward in the analysis along with their phenotype are shown in Table 4.4

Table 4.4. Sample IDs and Phenotypes

Sample ID	**Phenotype**
GSM4443482	Normal Skin donor
GSM4443483	Normal Skin donor
GSM4443484	Normal Skin donor
GSM4443485	Control skin from KS patient
GSM4443486	Lesion skin from KS patient
GSM4443487	Control skin from KS patient
GSM4443488	Lesion skin from KS patient
GSM4443489	Control skin from KS patient
GSM4443490	Lesion skin from KS patient
GSM4443497	Control skin from KS patient
GSM4443498	Lesion skin from KS patient
GSM4443499	Control skin from KS patient
GSM4443500	Lesion skin from KS patient
GSM4443501	Control skin from KS patient
GSM4443502	Lesion skin from KS patient

4.2.2 Alignment

15 samples were aligned to hg38 through HISAT2. The maximum alignment was observed for every sample. The range of genome coverage ranged from 95.62% to 98.43%. The alignment rate of every sample included in the analysis is shown in Table 4.5

Table 4.5. Alignment Rate of Samples to Reference Genome-Hg38

Sample	**Alignment Rate**
GSM4443482	98.22%
GSM4443483	97.73%
GSM4443484	98.07%
GSM4443485	98.15%
GSM4443486	97.70%
GSM4443487	98.04%
GSM4443488	97.77%
GSM4443489	97.42%
GSM4443490	97.87%
GSM4443497	97.87%
GSM4443498	95.62%
GSM4443499	98.17%
GSM4443500	92.09%
GSM4443501	98.43%
GSM4443502	98..24%

4.2.3 Identification of DEGs

The DEGs were identified with the help of the R package ballgown. As this dataset has 3 different phenotypes. Ballgown is a package of R language that calculates the differential expression of 2 phenotypes. To work with 3 phenotypes the ballgown is implemented three times. In the first round first two phenotypes are analyzed. In the second shift, the last two phenotypes are examined. In the last round first and the third phenotype is investigated. In this regard, the same strategy was implemented and 3 times ballgown was run each time working with two phenotypes. Below the results of all the 3 ballgown are discussed.

First Phenotype (Normal and Disease)

The disease phenotype depicts lesion skin and the normal phenotype describes the skin of a healthy individual.Figure 4.7 is a boxplot of all the samples' FPKM values versus their log2 values. Each individual plot represents a single sample and the dots represent the outliers.

Figure 4.8 shows the transcript distribution count per gene. It is obvious by looking at the bars of the graph that most of the genes possess 5 to 10 transcripts as opposed to other genes that are composed of less than five transcripts. The distribution of transcript lengths versus their frequency is shown in figure 4.9. The second bar has the highest frequency and the remaining transcripts are of shorter length depicting lower frequency.

Figure 4.10 represents the bar chart of differential expression. In this bar chart, a threshold value is handpicked for up and down-regulated genes. A threshold value was set to 2 on both sides of the spectrum. The genes on the left side of the graph falling below the threshold line indicate downregulated genes and the right side of the plot above the cut-off line represents downregulated genes.

Figure 4.11 shows the enhanced volcano plot of Kaposi sarcoma. As soon as we move away from zero on both sides of the graph the expression of genes transposes towards abnormal expression. The left side of the plot indicates down-regulated genes

and the right side points to up-regulated genes.

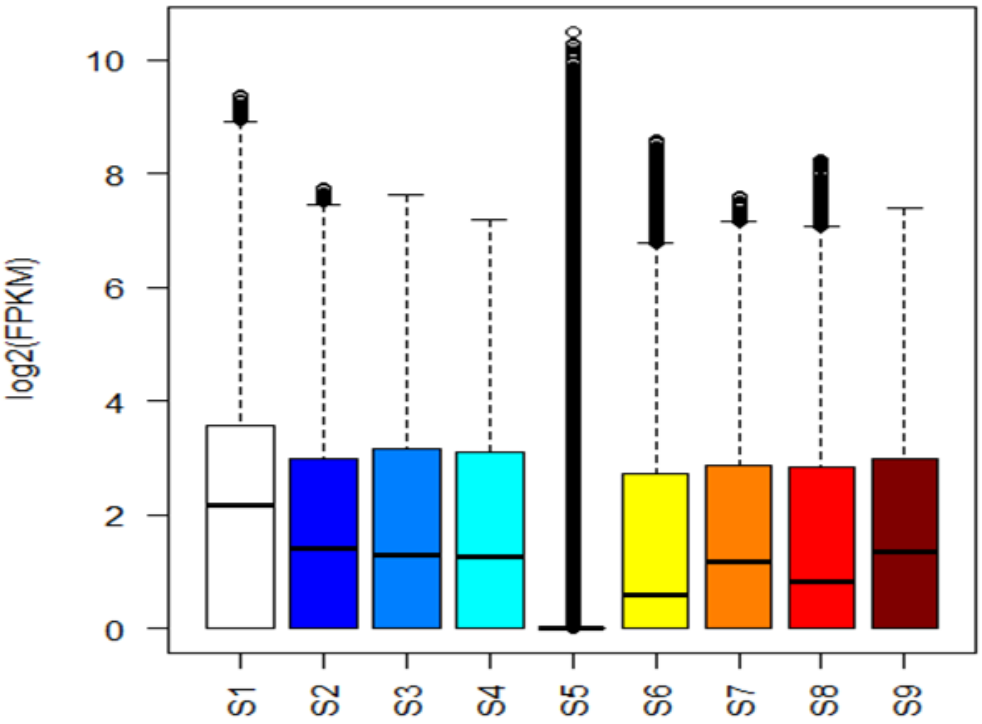

Figure 4.7. Boxplot of FPKM values versus Samples

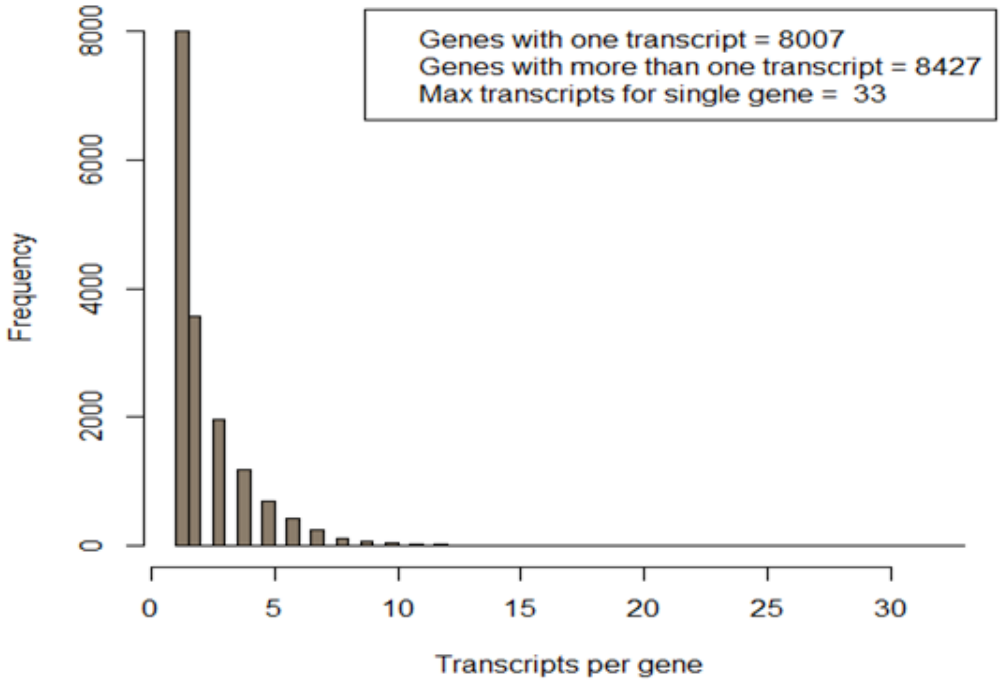

Figure 4.8. Transcript Distribution per Gene Count

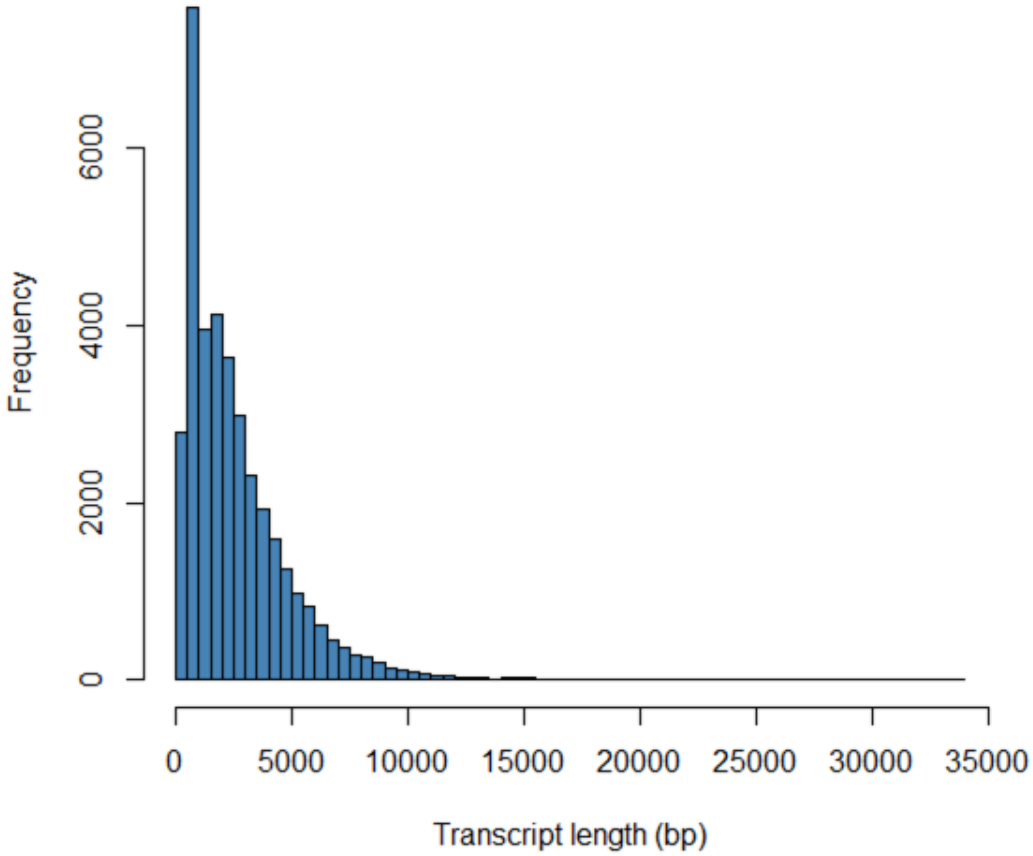

Figure 4.9. Transcript Length versus Frequency

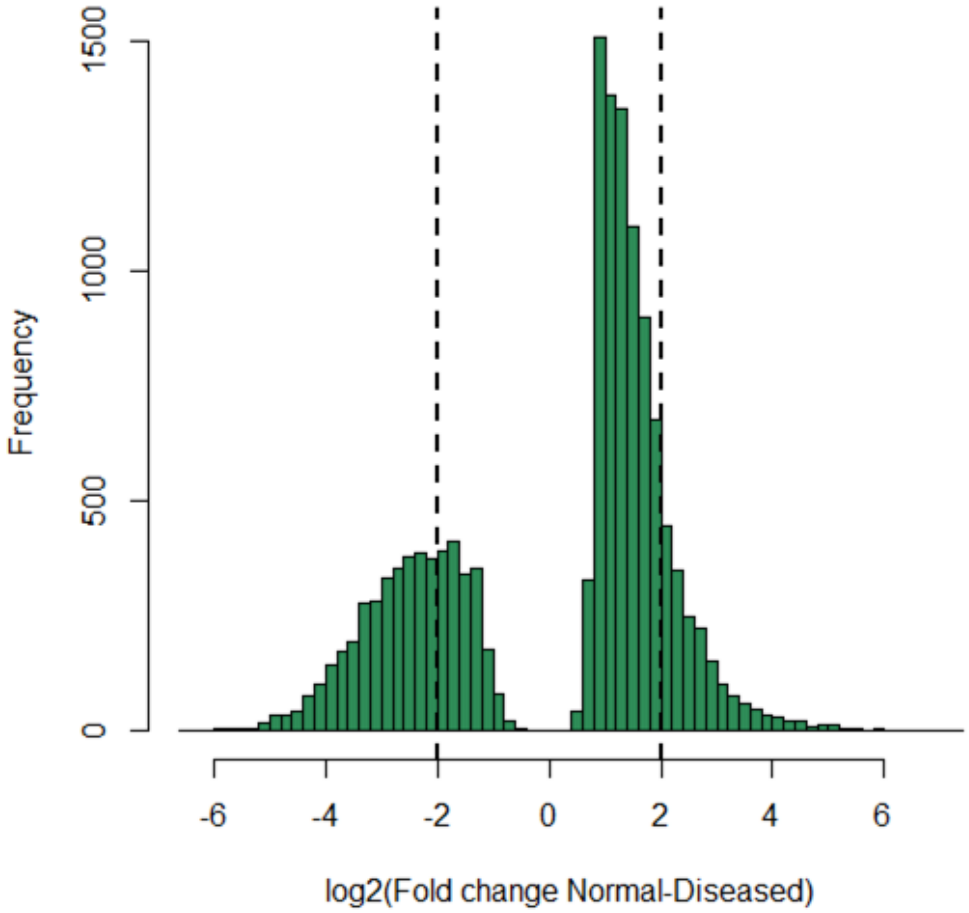

Figure 4.10. Histogram of Differential Expression

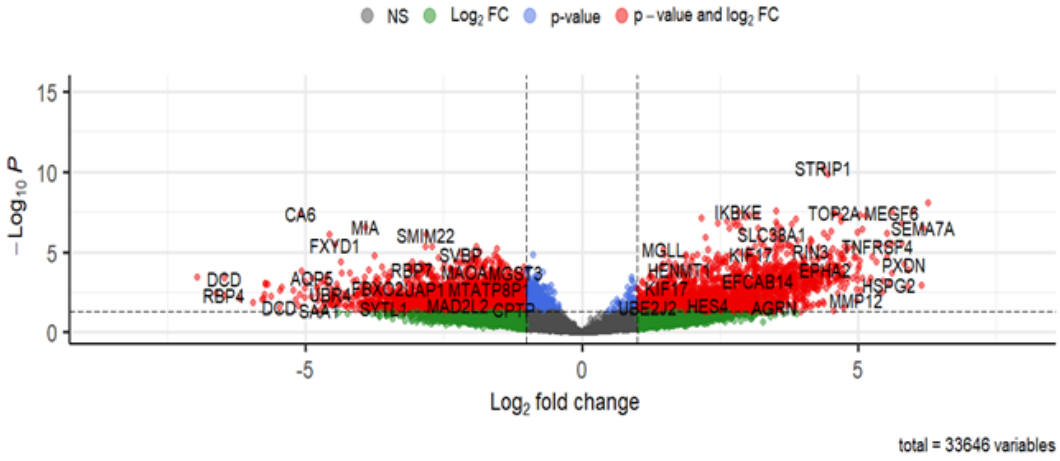

Figure 4.11. Enhanced Volcano Plot

Second Phenotypes (Control and Disease)

Control depicts the biopsy of the normal portion of the skin of an individual that is suffering from Kaposi sarcoma and disease representing the biopsy from the lesion.Figure 4.12 shows the box plot of FPKM values of Kaposi sarcoma samples. Each individual boxplot represents a single sample. Minimum, median, and maximum values of gene expression in the sample are represented by the first, second, and third layers of the plot and the dots above the box represent outliers.

The bar chart depicts the frequency of transcript per gene count as shown in figure 4.13 The highest frequency is as high as 8000. It is evident that maximum genes possess 1 to 5 transcripts.Figure 4.14 shows the bar chart of the distribution of transcript lengths against their frequency. It can be concluded that the second bar represents the highest frequency in terms of length as compared to others.

The histogram of differential expression is shown in figure 4.15. Based on threshold value fold change the genes a classified as up and downregulated genes. A fold change of +2 and -2 for up and down-regulated genes was selected respectively. The left side of the graph falling before the line shows downregulated genes. Whereas the right side of the graph lying above the line corresponds to upregulated genes.Figure 4.16 represents the enhanced volcano plot of Kaposi sarcoma. The extreme left and right side of the graph represents down and upregulated genes in red color.

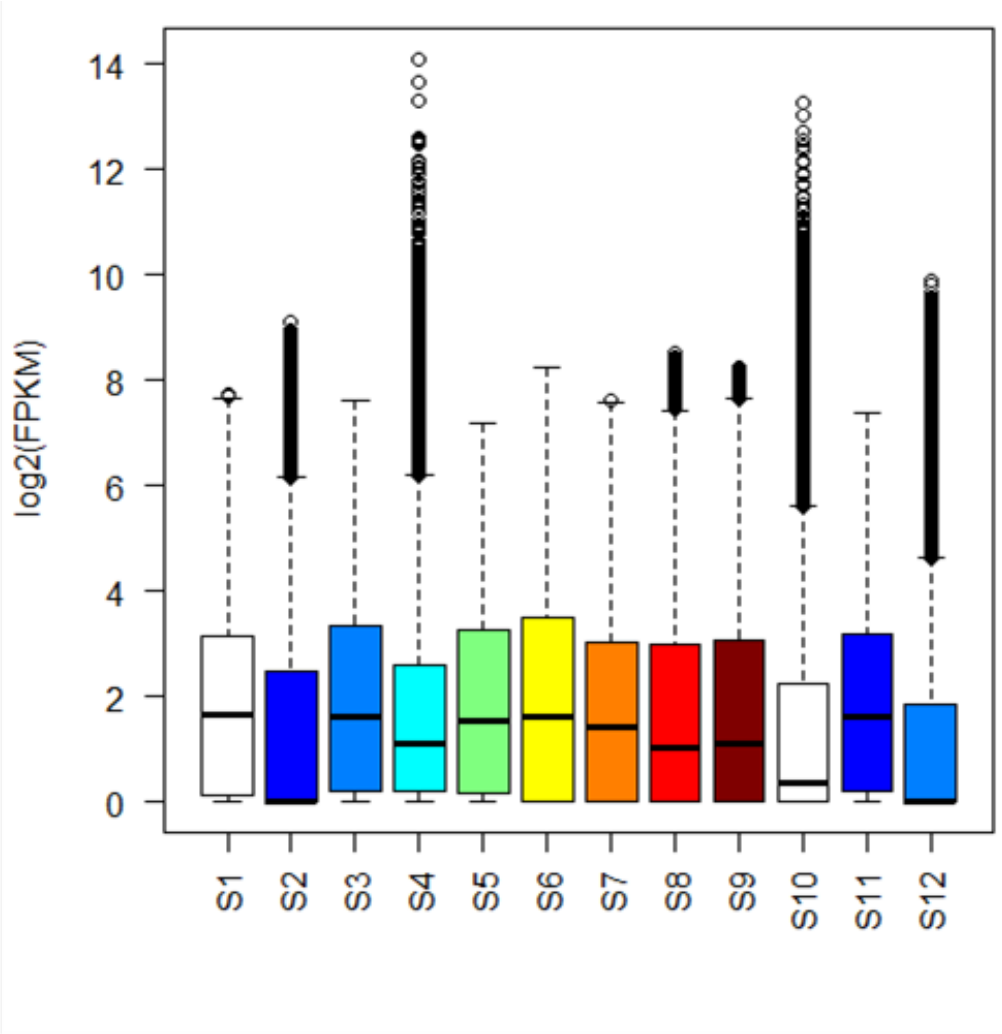

Figure 4.12. FPKM Values versus Samples

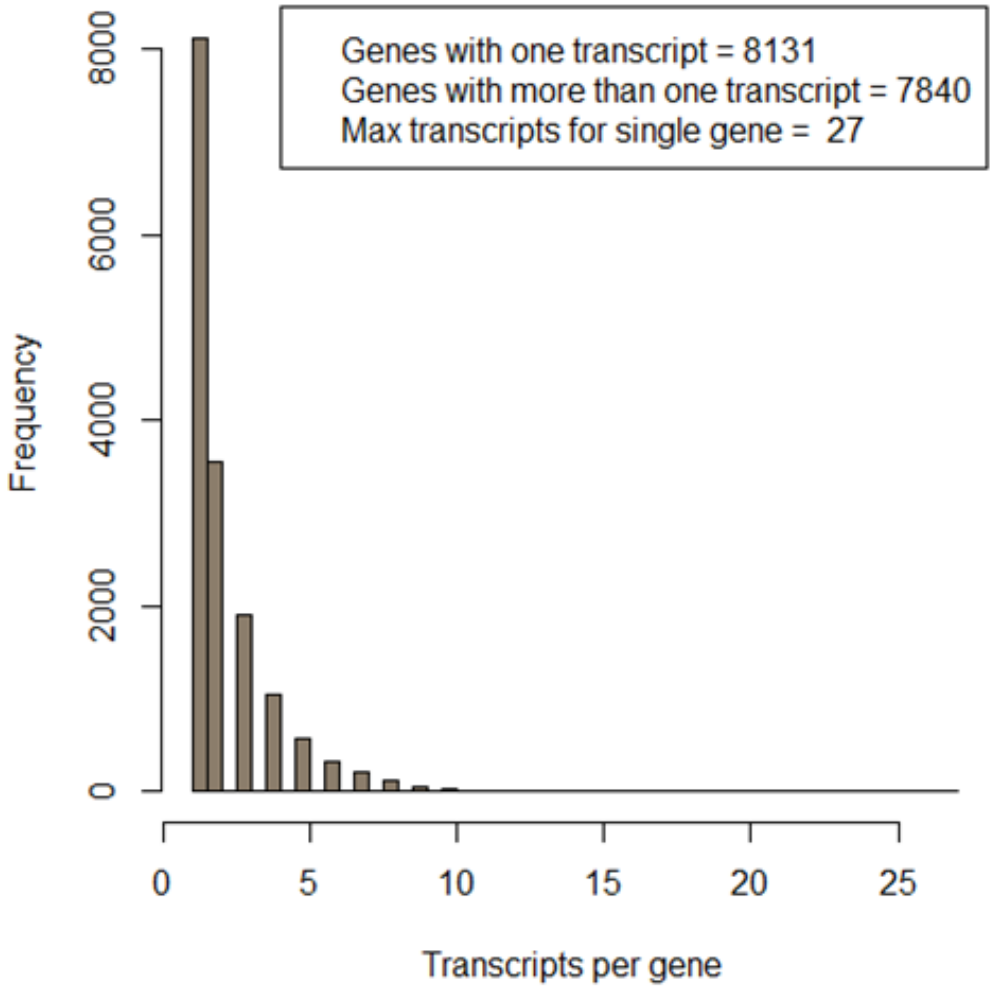

Figure 4.13. Transcript Frequency per Gene Count

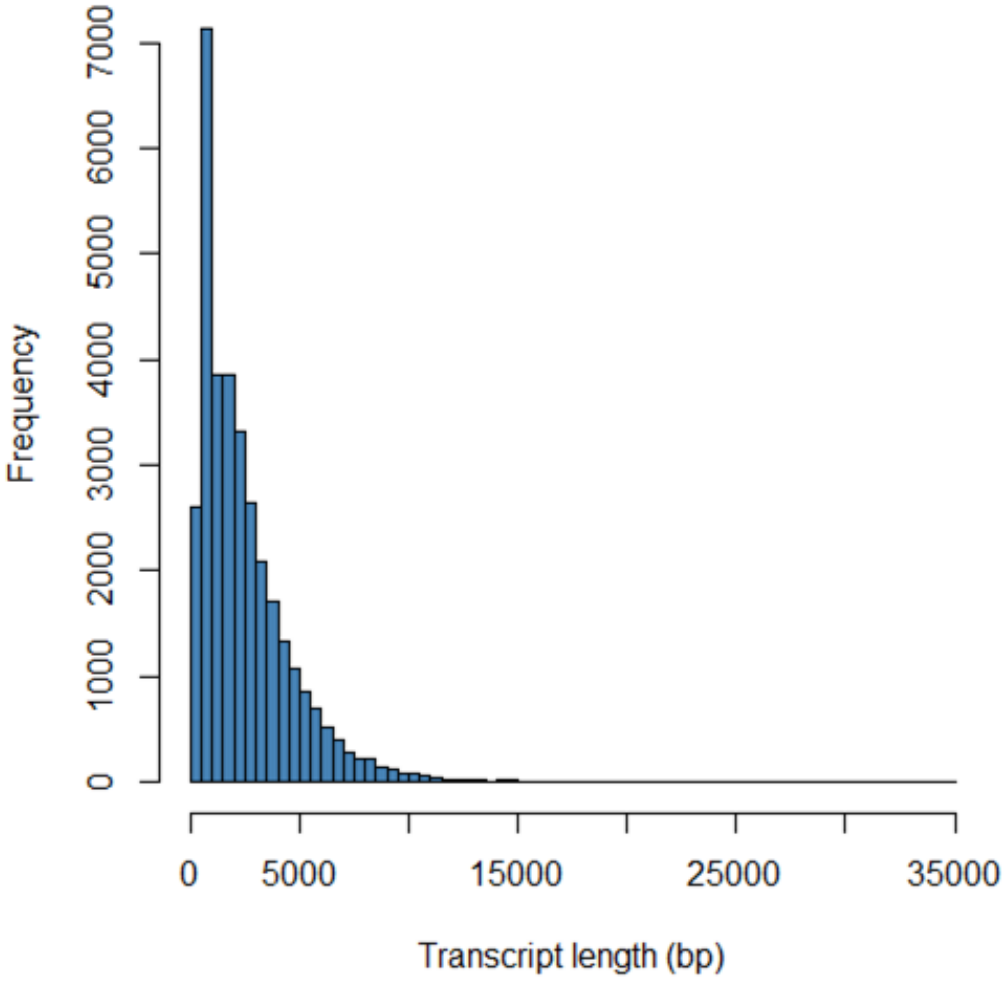

Figure 4.14. Transcript Length Distribution

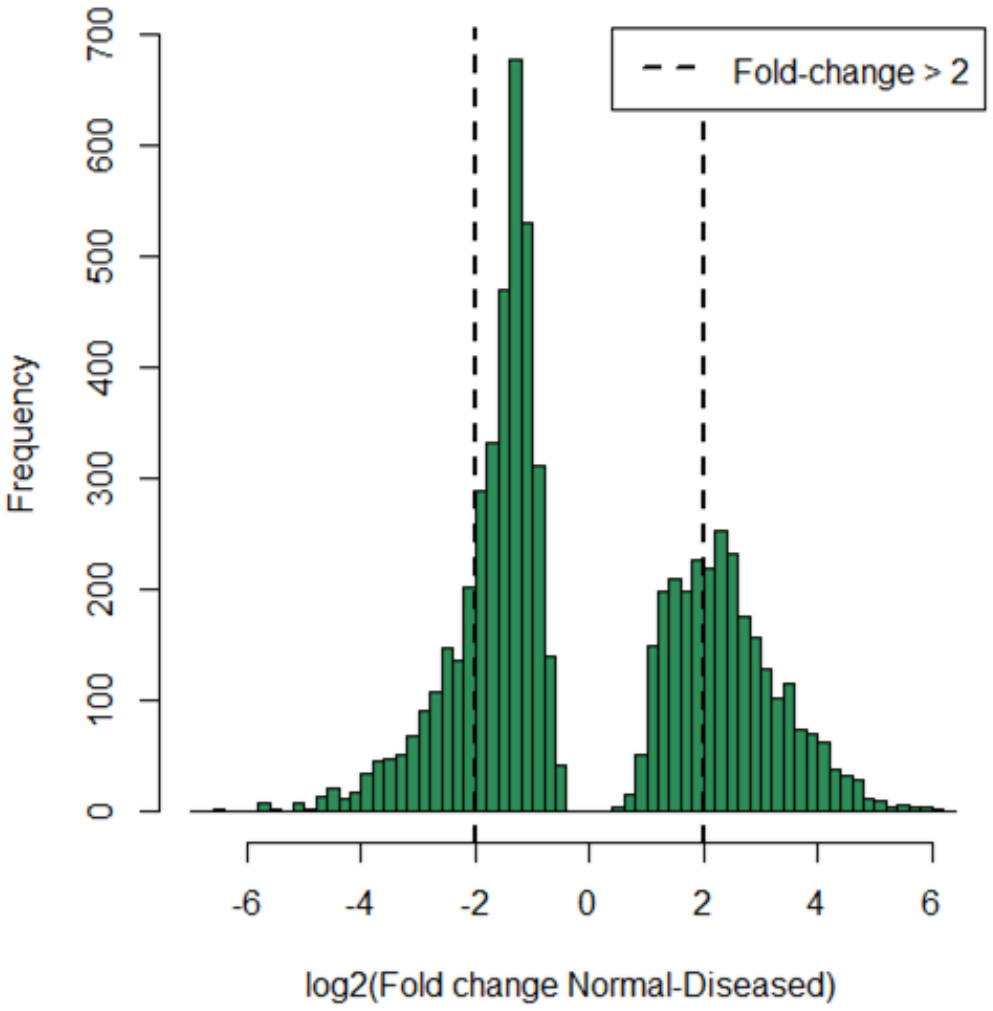

Figure 4.15. Histogram of Differential Expression

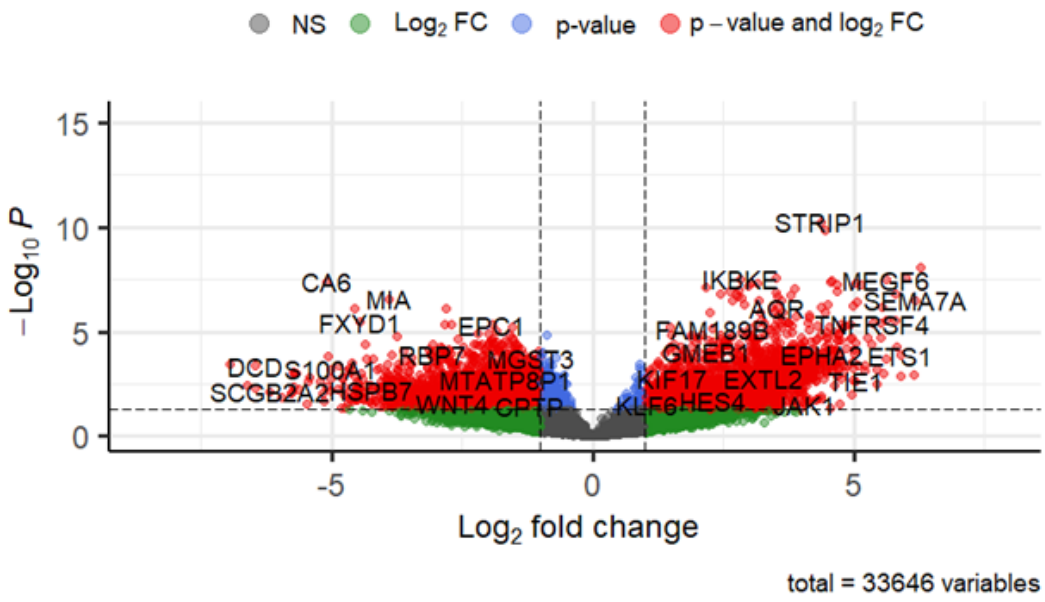

Figure 4.16. Enhanced Volcano Plot

Third Phenotype (Normal and Control)

The third round of ballgown was performed on normal that represents healthy individuals and control in this regard represents the patients suffering from Kaposi sarcoma but the biopsy is from the normal region of the skinThe boxplot of individuals is shown in figure 4.17. Each box corresponds to each sample. Log2 of FPKM values of each sample is shown. The minimum, median, and maximum values of each sample are represented by the respective layers of the boxplot and the dots represent the outlier.

Figure 4.18 shows the distribution of transcript frequency count per gene. The maximum genes possess less than five transcripts. The maximum number crosses 10000. Few genes lie in the range of 5 to 10 transcripts.Figure 4.19 shows the distribution of transcript lengths against their frequency. The second bar represents the highest frequency of transcript length.

Figure 4.20 represents the differential expression. To express the differential expression fold change threshold is selected to pick out the down and upregulated genes. +2 and -2 are the threshold for up and downregulated genes that are shown on the right and left side of the graph respectively.Figure 4.21 shows the enhanced volcano plot of Kaposi sarcoma. The extreme left and right side of the graph shows differentially expressed gene depicted in red color.

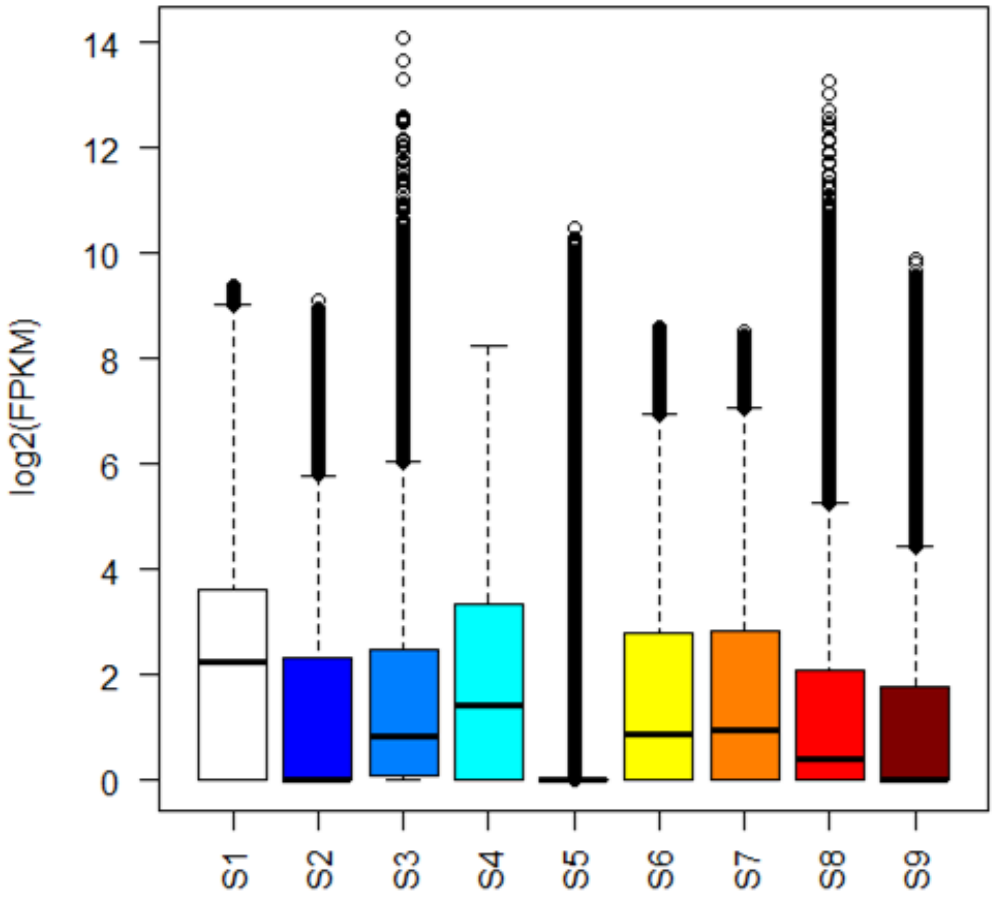

Figure 4.17. FPKM Values Versus Samples

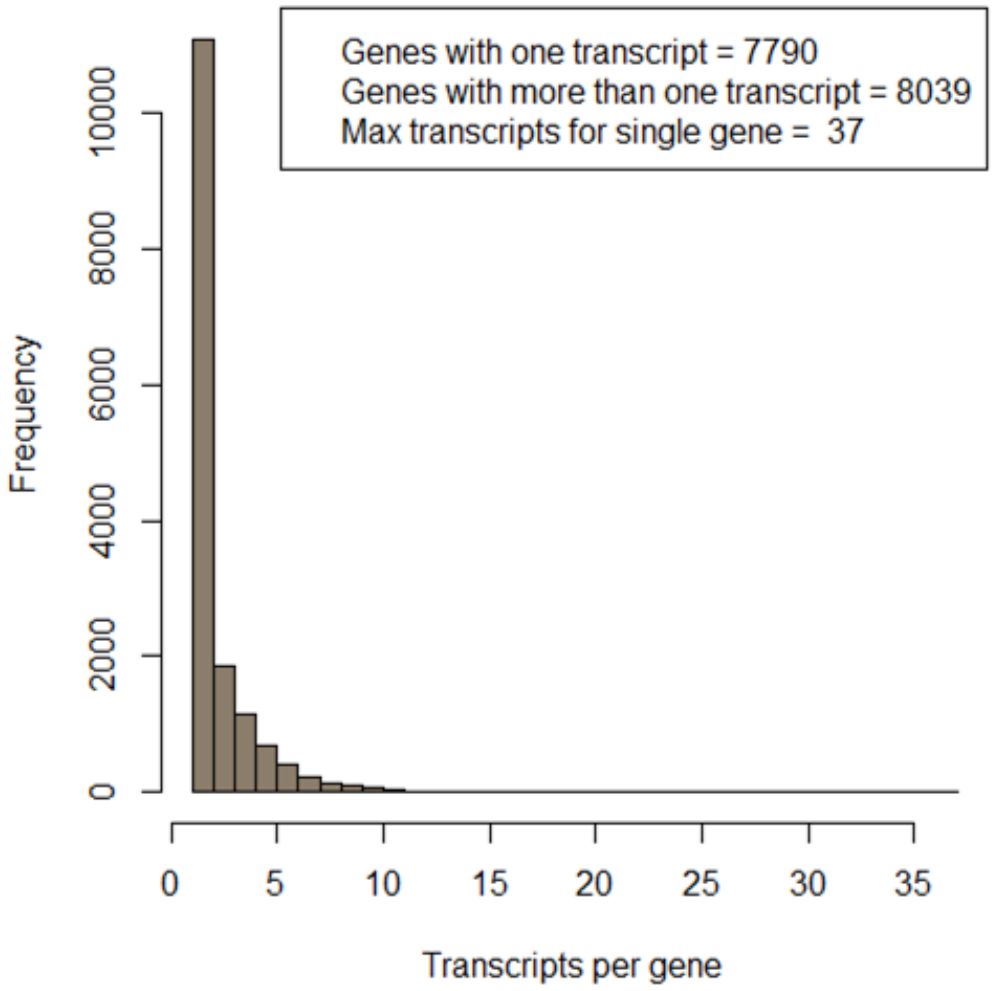

Figure 4.18. Transcript Frequency Distribution per Gene Count

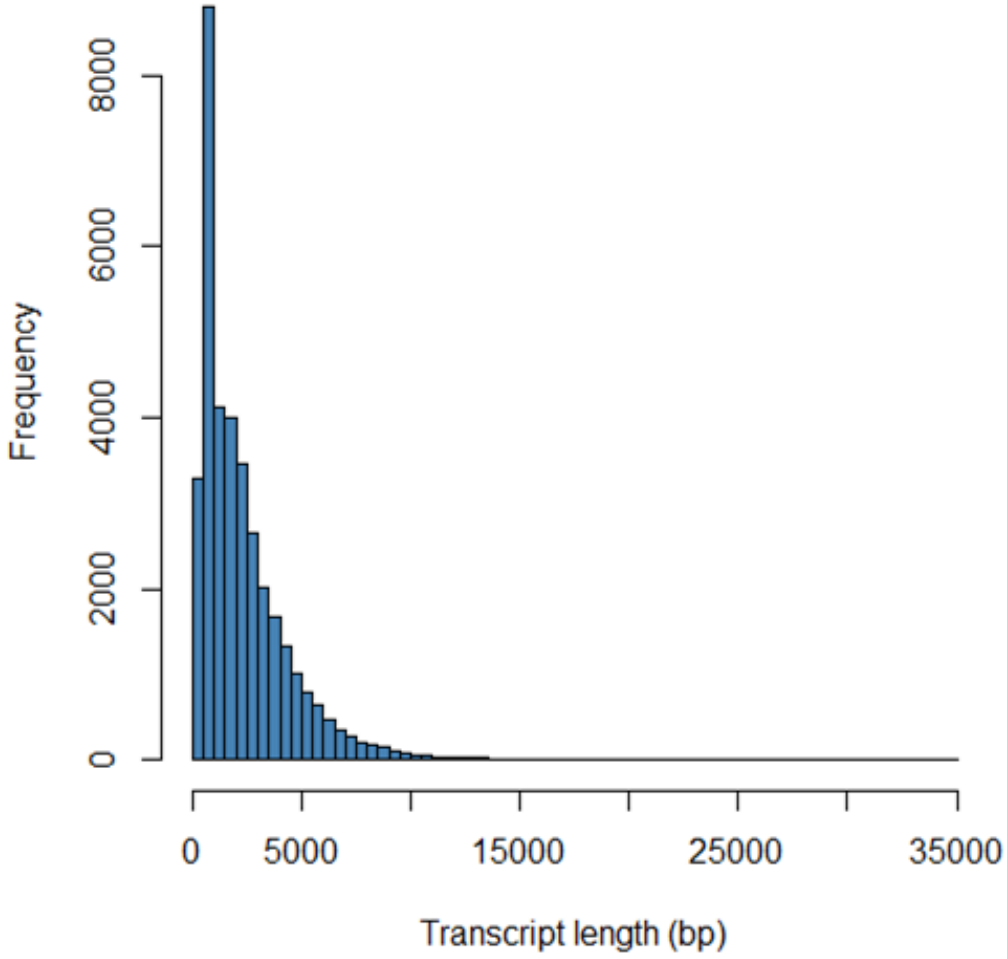

Figure 4.19. Transcript Length Frequency Distirbution

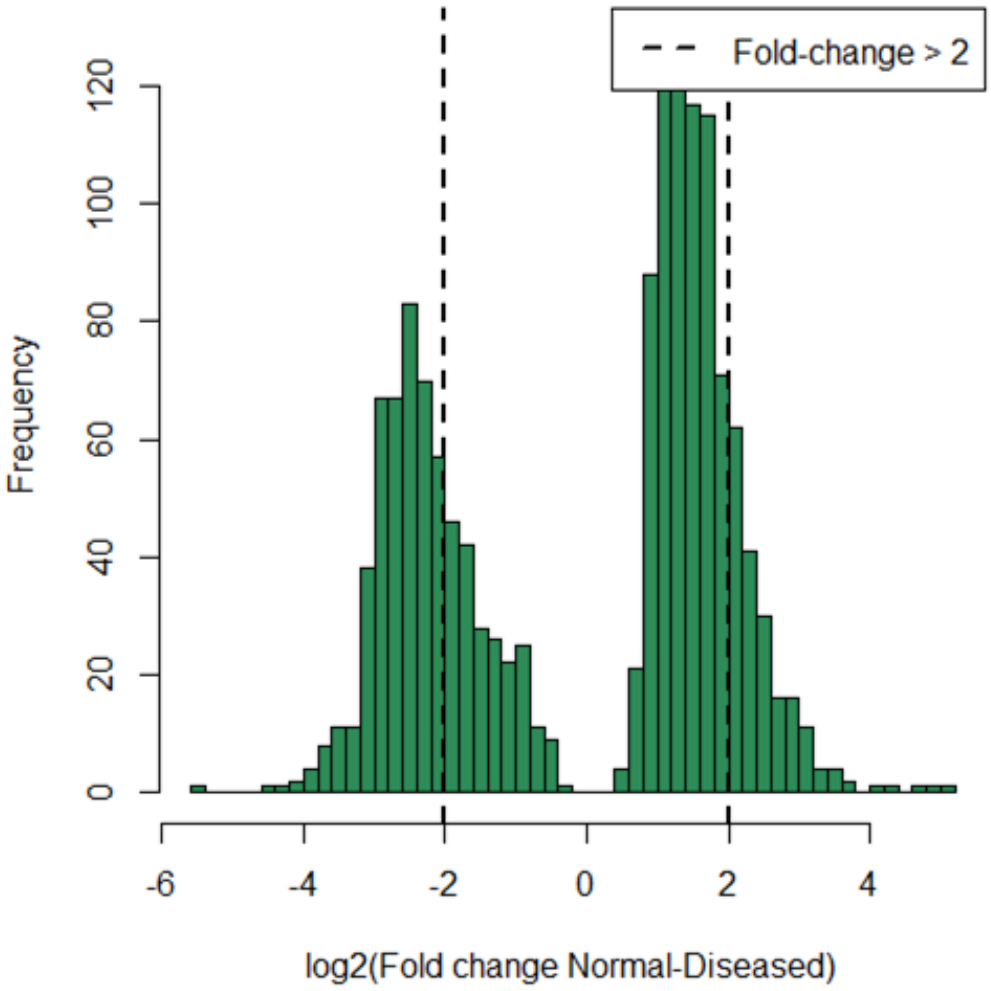

Figure 4.20. Histogram of Differential Expression

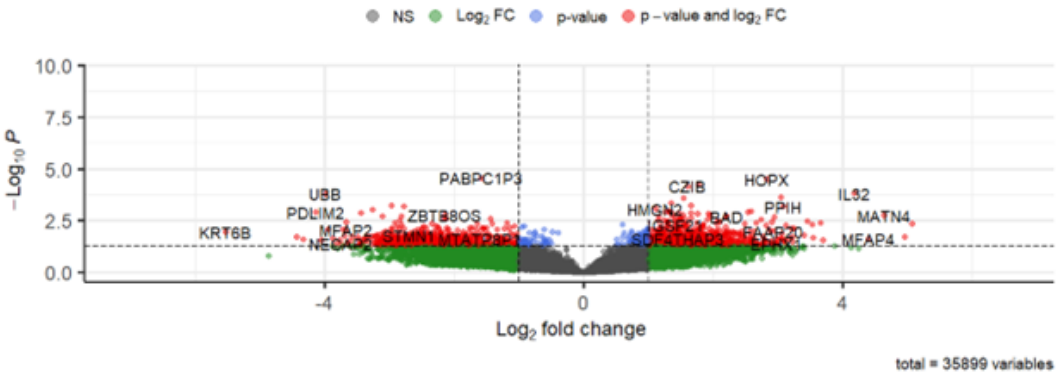

Figure 4.21. Enhanced Volcano Plot

4.2.4 mRNA Seq 2

Two datasets of mRNA seq were manipulated.

Data Retrieval

The second dataset of mRNA seq was also extracted from GEO database. The parameters of the dataset are given in the table below

Table 4.6. Parameters of Dataset GSE100684

Parameter	**Description**
GEO Accession ID	GSE100684
Title	RNA seq of Kaposi sarcoma reveals alteration of glucose and lipid metabolism
Organism	Homo Sapiens
Experiment	Expression profiling by high throughput sequencing
Country	USA
Platform	GPLI6791 Illumina HiSeq 2500
No of Samples	8

The sample id along with the phenotype operated in the analysis are discussed in table 4.7

Table 4.7. Phenotype and Sample IDs

Sample ID	**Phenotype**
GSM2691239	Control tissue of patient p32
GSM2691240	Control tissue of patient p22
GSM2691241	Control tissue of patient p83
GSM2691242	Control tissue of patient p23
GSM2691243	Lesion tissue of patient p32
GSM2691244	Lesion tissue of patient p22
GSM2691245	Lesion tissue of patient p83
GSM2691246	Lesion tissue of patient p23

Alignment

The retrieved data after progressing through the series of quality control procedures were aligned to the reference genome of Homo Sapiens which is hg38. The alignment rate of every sample is shown in table 4.8.

Table 4.8. Alignment rate of Sample to Reference Genome Hg38

Sample ID	**Alignment rate**
GSM2691239	95.05%
GSM2691240	97.52%
GSM2691241	96.74%
GSM2691242	97.35%
GSM2691243	97.52%
GSM2691244	97.52%
GSM2691245	97.60%
GSM2691246	97.04%

Identification of DEGs

Identifying differentially expressed genes is the core objective of RNA seq. Ballgown performs a remarkable job of identifying DEGs. The boxplot in figure 4.22 represents the samples and log2 of their FPKM values. Each box represents the minimum, median and maximum values by its respective layers of each sample. The outliers are shown by the dots above the box

Figure 4.23 represents the distribution of transcript count per gene. Less than five transcripts are owned by the maximum genes in the genome.

The graph of differential expression is shown in figure 4.24. A threshold value of fold change is selected for up and downregulated genes. +2 and -2 is the threshold value for up and down-regulated genes respectively. The left side of the graph represents the down and the right side of the graph represents upregulated gene.

The graph of differential expression is shown in figure 4.25. A threshold value of fold change is selected for up and downregulated genes. +2 and -2 is the threshold value for up and down-regulated genes respectively. The left side of the graph represents the down and the right side of the graph represents upregulated gene.

Figure 4.26 shows the enhanced volcano plot. The left and right extremes of the graph depict the down and upregulated genes respectively shown in red color.

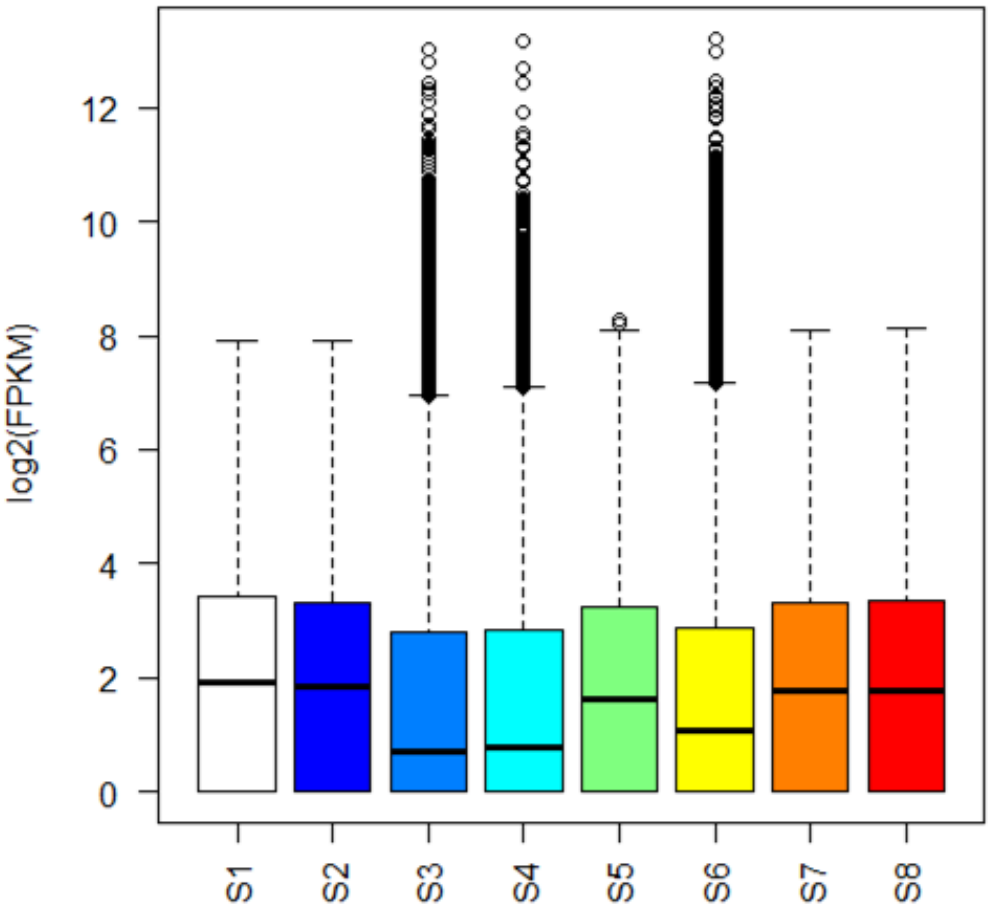

Figure 4.22. Distirbution Log2FPKM of Samples

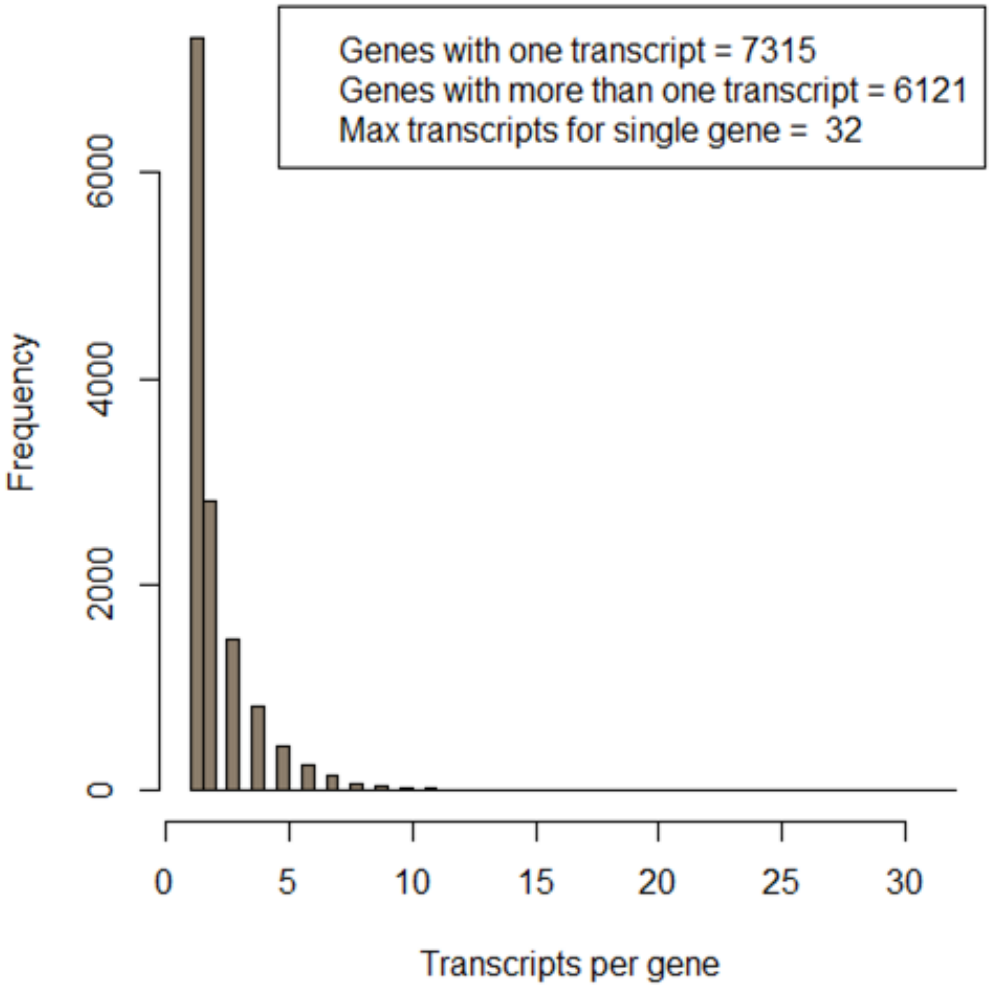

Figure 4.23. Distribution of Transcript Frequency Per Gene Count

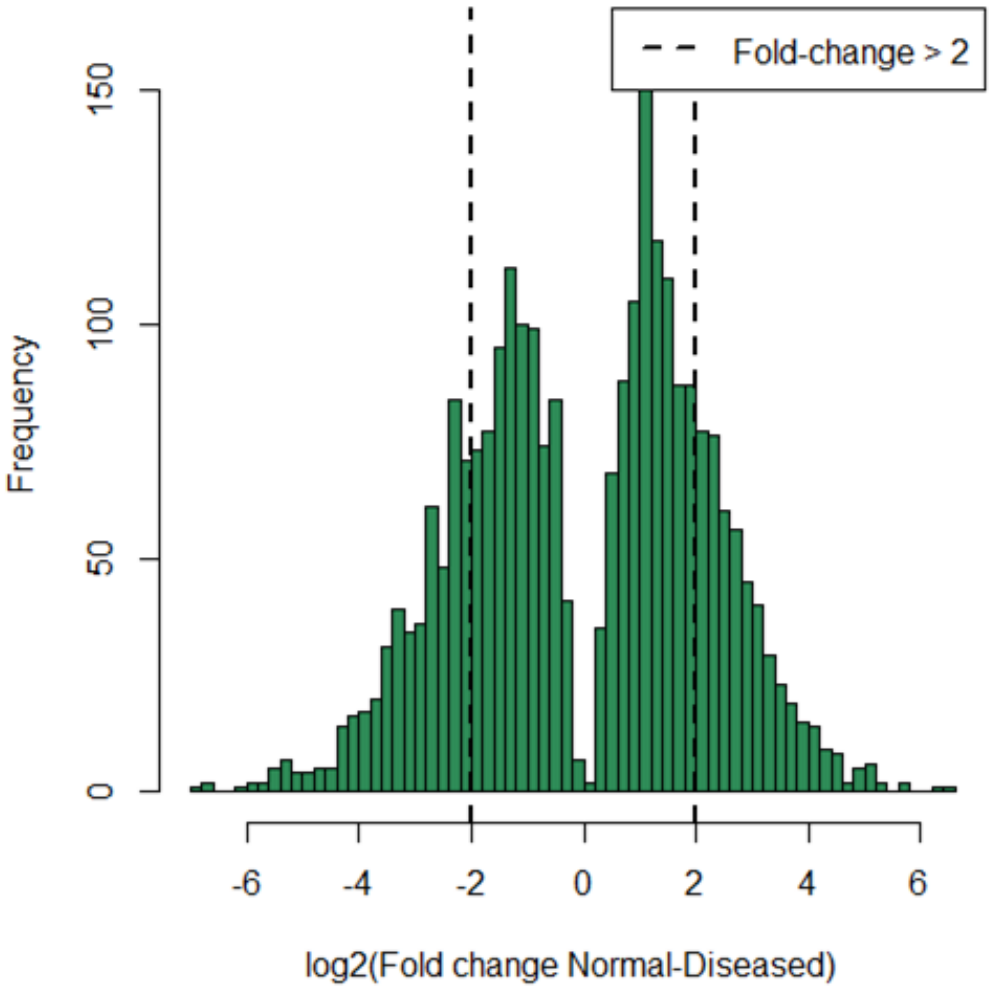

Figure 4.24. Histogram of Differential Expression

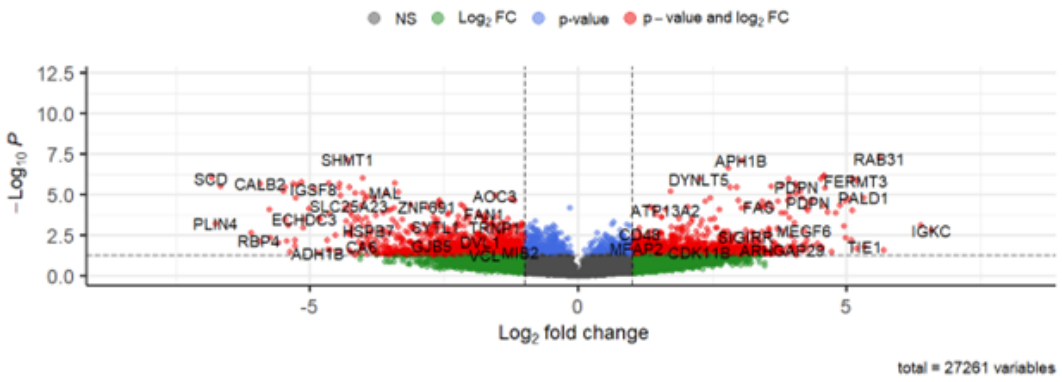

Figure 4.25. Enhanced Volcano Plot

4.3 Comparative Analysis

After the identification of DEGs in both the dataset of RNASeq and microarray comparative analysis was performed for the identification of common genes in all the datasets. 'Draw Venn Diagram" was used for this purpose. 4 different lists of DEGs were given as input. 3 lists were given from GSE147704 because it has 3 phenotypes. So, 3 different lists of DEGs were generated and 1 list of DEGs were given from GSE100684. A total of 88 genes were found to be common among the 2 datasets. Figure 4.26 shows the venn diagram of common gene

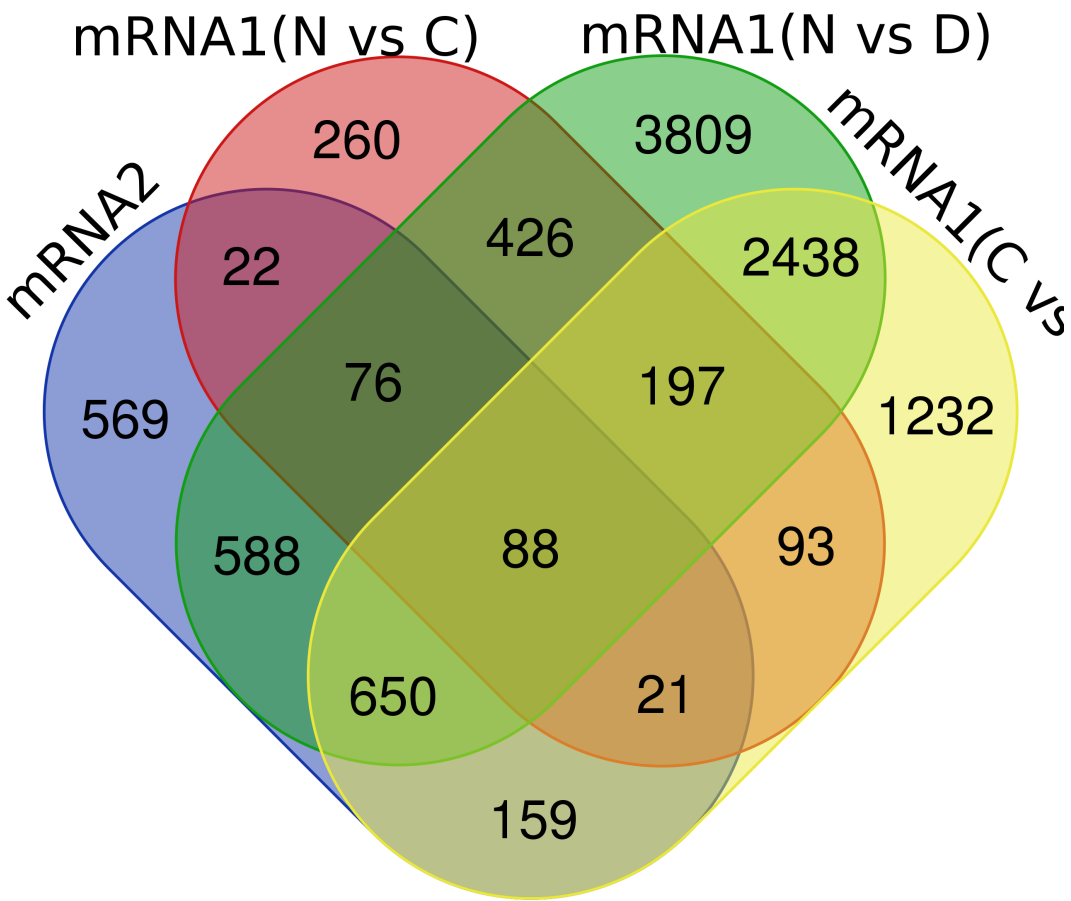

Figure 4.26. Common Genes among RNA Seq Dataset

4.4 Pathway Analysis

Pathway analysis was performed on 88 common genes obtained because of comparative analysis. This was done to find out the pathways in which these genes are involved. The topmost significant pathways with respect to p-value and entities found are shown in table 4.8

Table 4.9. Significant Pathways Among Common Genes

Pathway	**p-value**	**Entities found**
Regulation of expression of SLITs and ROBOs	5.13E-05	8
Signalling by ROBO receptors	2.81E-04	8
Axon guidance	9.08E-04	12
Nervous system development	0.001506008	12
Metabolism	0.009180822	37
Cellular responses to stimuli	0.013122977	14

4.5 Protein Modelling

The genes were analyzed from the most significant pathways based on p-value and number of entities. The examination of individual genes was performed on the following characteristics

- Upregulated in disease
- Experimentally determined structure on PDB

Glutathione S Transferase P abbreviated as GSTP possessed the attributes mentioned above. So GSTP was established to put be put forward in the analysis.

4.5.1 GSTP

GSTP is involved in conjugating reduced glutathione to various exogenous and endogenous hydrophobic electrophiles. It is called for the genesis of glutathione conjugates of prostaglandin A2 and J2. It prevents neurodegeneration by translocation of p25 via regulating CDK5. It is involved in the formation of hepoxilin regioisomers. GSTP modulates signal transduction pathways. It manipulates MAPK family that is involved in cellular survival and apoptosis. GSTP protects tumor cells by apoptosis by phosphorylation of C-Jun-N-terminal kinase. This binding acts as an inhibitor by transmitting negative signals to modulate apoptosis. GSTP is upregulated in Kaposi sarcoma by promoting oncogenesis and inhibiting apoptosis of cancer cells.

4.5.2 Modelling

A variety of experimentally determined structures of GSTP were available on UniProt through NMR and X-ray crystallography. The domain of GSTP subjected to BLAST was between 83-204 at the C terminal. 17gs was the template finalized for the model generation of protein structure by casting around through the Swiss Model. The parameters of 17gs obtained via Swiss modeling are given in Table 4.9

Table 4.10. Parameters of the Template 17gs

Parameter	**Value**
GMQE	0.92
Method	X-ray Crystallography
Resolution	1.9Å
Identity	100
Oligo State	Homo dimer

The model of protein generated through the Swiss Model is shown in figure 4.27

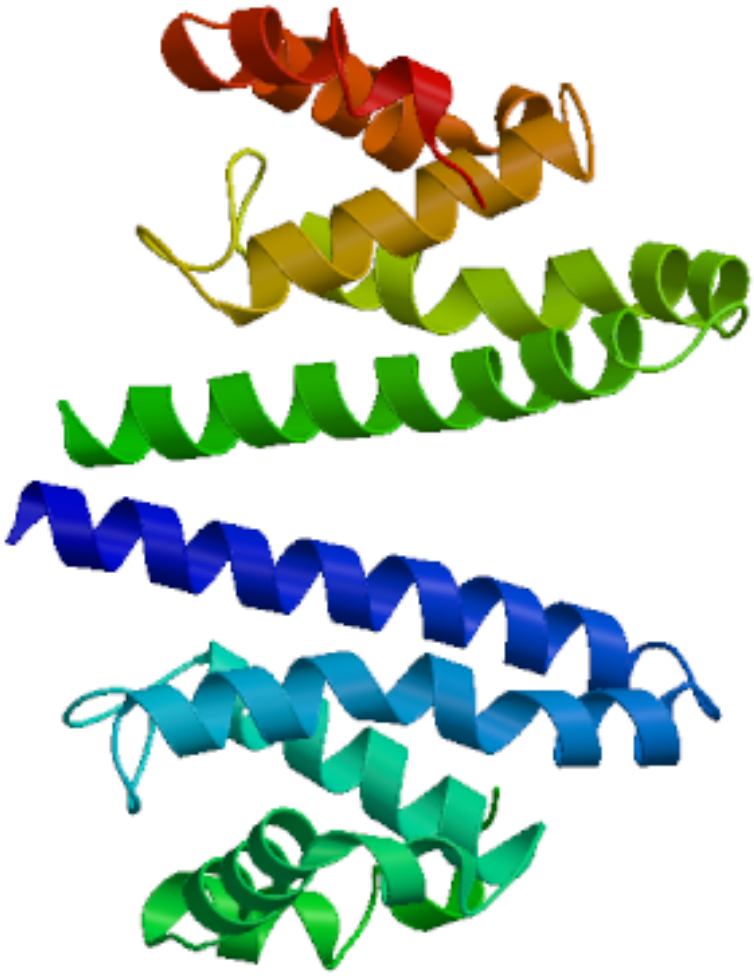

Figure 4.27. Protein Structure generated through Swiss Model

The optimization of the model generated was performed through Play Molecule by removing water molecules and maintaining the pH at 7.4. DoGSiteScorer was employed to determine the binding pocket of the prepared structure and the drug score of the prepared protein. The drug score of the prepared protein was 0.8.

187 ligands for protein GSTP were prepared. 100 ligands were extracted from bindingDB and 87 ligands were downloaded from DrugBank. The smiles format of the proteins was converted to mol2 through OpenBabel and further into pdbqt with the help of AutoDock. The binding Grid of GSTP was set in AutoDock. The docking of 187 ligands to GSTP was performed through AutoDock Vina. The ligands were docked to the protein's binding domain in 9 different conformations that generated different binding energies. The overview of binding affinities between the ligands and protein is shown in figure 4.26

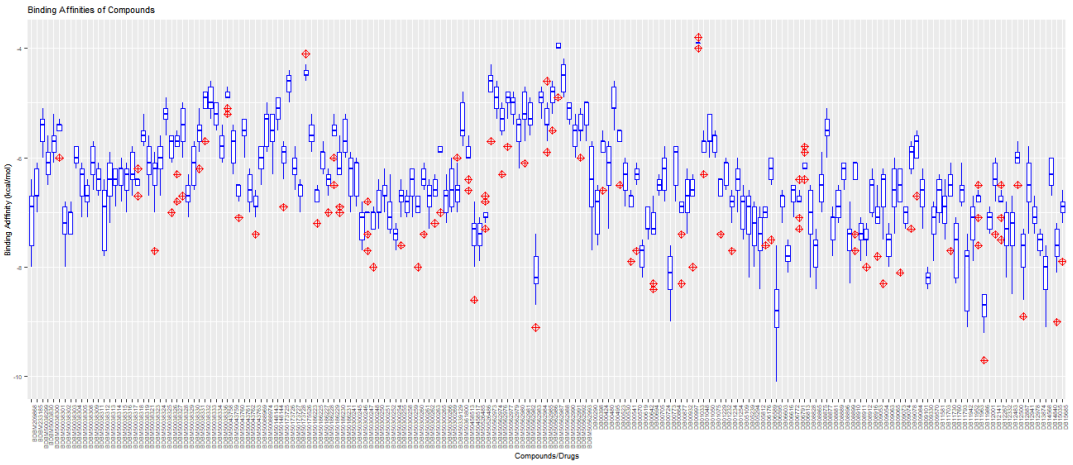

Figure 4.28. Boxplot of Binding Affinities between Proteins and Ligands

The top 10 ligands were ranked the best candidate to stop the progression of Kaposi Sarcoma. The best-selected ligands were based on minimum binding affinities between protein and ligand. The binding affinities of the top selected ligands are given in table 4.10.

Table 4.11. Binding Affinities of Top 10 Ligands

Ligand	**Binding Affinity**
BDBM50458517	-8.6
BDBM50562983	-9.1
DB00773	-9
DB06595	-10.1
DB11942	-9.1
DB11986	-9.6
DB12483	-8.5
DB12887	-8.9
DB14568	-9.1
DB15035	-9

4.5.3 Non-Covalent Interactions

The top 10 ligands that manifested paramount binding affinity with the protein were shortlisted. PLIP was used to discern the non-covalent interaction between the ligand and the protein. The non-covalent interaction between the protein and the short-listed ligands is discussed below.

BDBM50458517

In the figure below the interactions are shown between GSTP1 and BDBM50458517 are shown. In the figure, different colors depict different aspects of the exchanges. Protein is represented through blue, ligand through orange, water through lilac, charge center through yellow, aromatic ring center through white, hydrophobic interactions through dotted lines and hydrogen bonds through sticks.

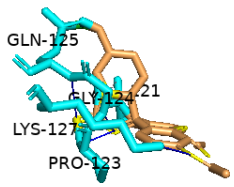

Figure 4.29. Non-Covalent Interactions between GSTP1 and BDBM50458517

BDBM50562983

No non-covalent interactions were observed between the GSTP1 and BDBM50562983.

DB00773

In the figure below the interactions are shown between GSTP1 and DB00773 are shown. In the figure different colors depict different aspects of the interactions. Protein is represented through blue color, ligand through orange, water through lilac, charge center through yellow, aromatic ring center through white, hydrophobic interactions through dotted lines and hydrogen bonds through sticks.

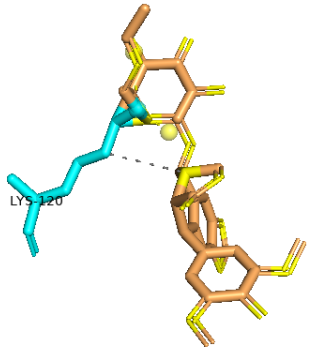

Figure 4.30. Non-Covalent Interactions between GSTP1 and DB00773

DB06595

In the figure below the interactions are shown between GSTP1 and DB06595 are shown. In the figure different colors depict different aspects of the interactions. Protein is represented through blue color, ligand through orange, water through lilac, charge center through yellow, aromatic ring center through white, hydrophobic interactions through dotted lines and hydrogen bonds through sticks.

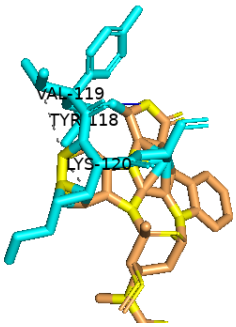

Figure 4.31. Non-Covalent Interactions between GSTP1 and DB06595

DB11942

In the figure below the interactions are shown between GSTP1 and DB11942 are shown. In the figure different colors depict different aspects of the interactions. Protein is represented through blue color, ligand through orange, water through lilac, charge center through yellow, aromatic ring center through white, hydrophobic interactions through dotted lines, and hydrogen bonds through sticks.

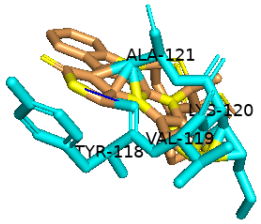

Figure 4.32. Non-Covalent Interactions between GSTP1 and DB11942

DB11986

In the figure below the interactions are shown between GSTP1 and DB11986 are shown. In the figure, different colors depict different aspects of the interactions. Protein is represented through blue color, ligand through orange, water through lilac, charge center through yellow, aromatic ring center through white, hydrophobic interactions through dotted lines and hydrogen bonds through sticks.

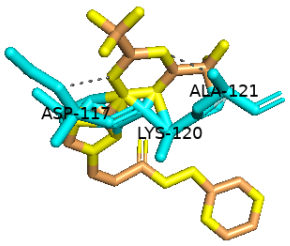

Figure 4.33. Non-Covalent Interactions between GSTP1 and DB11986

DB12483

In the figure below the interactions are shown between GSTP1 and DB12483 are shown. In the figure different color depict different aspects of the interactions. Protein is represented through blue color, ligand through orange, water through lilac, charge center through yellow, aromatic ring center through white, hydrophobic interactions through dotted lines and hydrogen bonds through sticks.

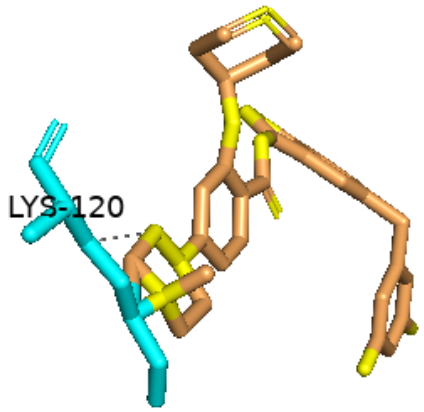

Figure 4.34. Non-Covalent Interactions between GSTP1 and DB12483

DB12887

In the figure below the interactions are shown between GSTP1 and DB12887 are shown. In the figure, different colors depict different aspects of the interactions. Protein is represented through blue color, ligand through orange, water through lilac, charge center through yellow, aromatic ring center through white, hydrophobic interactions through dotted lines, and hydrogen bonds through sticks.

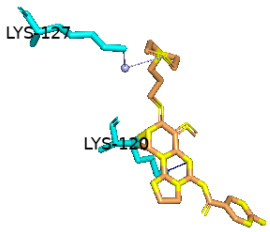

Figure 4.35. Non-Covalent Interactions between GSTP1 and DB12887

DB14568

In the figure below the interactions are shown between GSTP1 and DB14568 are shown. In the figure, different colors depict different aspects of the interactions. Protein is represented through blue color, ligand through orange, water through lilac, charge center through yellow, aromatic ring center through white, hydrophobic interactions through dotted lines, and hydrogen bonds through sticks.

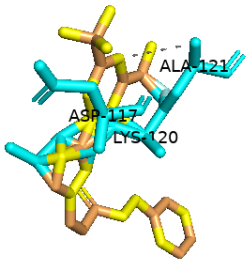

Figure 4.36. Non-Covalent Interactions between GSTP1 and DB14568

DB15035

In the figure below the interactions are shown between GSTP1 and DB15035 are shown. In the figure, different colors depict different aspects of the interactions. Protein is represented through blue color, ligand through orange, water through lilac, charge center through yellow, aromatic ring center through white, hydrophobic interactions through dotted lines, and hydrogen bonds through sticks.

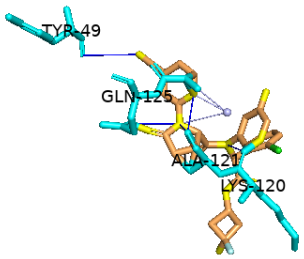

Figure 4.37. Non-Covalent Interactions between GSTP1 and DB15035

Chapter 5

DISCUSSION

Kaposi Sarcoma is a genre where cancer cells are grounded in mucous membranes, skin, and the lining of the gastrointestinal tract, intestine, and stomach. This malignancy pops up as purple patches on the skin and escalates to the lungs and lymph nodes. Based upon the location, clinical stage, and epidemiology four variants of Kaposi Sarcoma have been identified. The most common is classical Kaposi sarcoma in elderly men, in younger individuals, it is described as endemic. Whereas it is described as Iatrogenic Kaposi sarcoma in organ transplant recipients and an epidemic in AIDS-infected individuals. All forms of Kaposi sarcoma have been caused by Human Herpes Virus 8. HHV8 possesses a double-stranded DNA that encodes more than 80 proteins with an extensive number of microRNAs. HHV8 expresses the genes that help the virus escape the primary immune response. These escapes include inhibition of apoptotic factors, autophagy factors, and Natural Killer cells.

The identification of biomarkers and therapeutic targets has gained enormous public insight in the past decade. Microarray analysis and RNA seq usually serve on these grounds. The therapeutic targets are the genes whose expression is beyond normal and disrupt the regular metabolism of the body. The goal of the identification of such genes in different ailments is to inhibit upregulated genes or activate downregulated genes.

In this analysis, the therapeutic targets of Kaposi sarcoma were identified. RNA seq and microarray spotted such targets. The dataset of microarray and 2 datasets of RNA seq was analyzed and produced the genes with abnormal expression. Comparative analysis between the upregulated and downregulated genes of all three datasets didn't even give a single common gene. However, the comparative analysis of the 2 datasets of RNA seq put forward 88 common genes. The results of microarray analy-

sis were conflicting with RNA seq output. On these grounds, the dataset of microarray was dropped and the RNA seq dataset was analyzed further in the analysis because they yielded similar results.

Pathway analysis of the common genes presented various pathways from which the most significant pathways were shortlisted. The candidate pathways were analyzed and the entities participating were also examined. The candidate genes in the pathways that were shortlisted included MT2A, TUBB6, GSTP1, PSMF1, TKT, CRAYB.

MT2A synchronizes homeostasis, the immune system, regulation of cell differentiation and proliferation, and angiogenesis. MT2A is upregulated in Kaposi Sarcoma. TUBB6 is believed to play a role in microtubule cytoskeleton organization and mitotic cell cycle. The job of TUBB6 establishes a clear ink with Kaposi sarcoma progression. The formation of spindle cells, microfilaments, and tumor cells in lesions involves TUBB6 expression. GSTP1 a gene that is upregulated in Kaposi sarcoma incorporates detoxification, elimination of foreign agents and carcinogens and protection of cells from RNA damage. GSTP1 is involved in oncogenesis is Kaposi Sarcoma. TKT gene is involved in the channeling of sugar-phosphate for glycolysis in pentose phosphate pathway. TKT is reported to have a muted expression in Kaposi sarcoma. CRYAB is the gene that is downregulated in Kaposi sarcoma and is involved in cell differentiation.

The above-mentioned genes have a pivotal role and are involved in the progression of Kaposi sarcoma. GSTP1 was the gene that was finalized for analysis. The selection of genes was based upon various components. Upregulation in disease, experimentally determined structure on PDB with a decent resolution and availability of domains. All the mentioned genes were having a prominent activity in Kaposi sarcoma development, but some genes were downregulated, a few genes didn't have experimentally determined structure, and the rest were not have identified domains. GSTP1 was the one that possessed all three characteristics due to which it was final-

ized as a major therapeutic target.

GSTP1 is reported to have a major role in the progression of Kaposi sarcoma. GSTP1 hides the tumor cells from apoptosis by inhibiting of MAPK pathway by protein-protein interaction. MAPK is a pathway that promotes apoptosis. GSTP1 interacts with c JUN that delivers anti-apoptotic signals that lead to apoptosis inhibition. The inhibition of apoptosis is a key component in carcinogenesis. This pathway is inhibited in Kaposi sarcoma which inhibits apoptosis and leads to the survival of cancer cells.

The identification of therapeutic targets landed to the preceding step which was to examine the effectiveness of available therapeutics for Kaposi sarcoma and inhibitors of GSTP1. 187 ligands were investigated for their effectiveness against the shortlisted candidate that was GSTP1. 100 ligands were inhibitors of GSTP, 36 ligands were drugs for the treatment of Kaposi Sarcoma and 51 were anti-tumor drugs. The binding affinity of these therapeutics was analyzed against GSTP1. Molecular docking was performed between GSTP1 and the ligands to predict the binding affinities. A variety of binding energy was observed between various ligands. The inhibitors of GSTP1 exhibited comparatively less energy as compared to drugs. The maximum binding energy was 4.1Kcal/ mol between BDBM50175526 and GSTP1 whereas the least was 10.1 kcal/mol which was exhibited between DB06595 and GSTP1. The drug that marks the top of the list is DB06595 with -10.1kcal/mol.

DB06595 is known as Midostaurin. It is an antineoplastic component that is employed for the treatment of acute myeloid leukemia. Acute myeloid leukemia is used to handle mutations, systemic mastocytosis, and mast cell leukemia. It was approved by FDA on 28 April 2017 and appeared to peak survival in patients with Acute Myeloid Leukemia.

Chapter 6

CONCLUSION AND FUTURE PERSPECTIVES

Kaposi sarcoma is still a paramount cause of death in AIDS-defining malignancies. To date, there is no designated treatment for Kaposi Sarcoma. The symptomatic treatment is given to the patients with a 72% survival rate with the crucial likelihood of relapse in the patients.

In this project, GSTP1 was designated as the therapeutic target for Kaposi Sarcoma. It was analyzed that GSTP1 promotes oncogenesis by inhibiting apoptosis by truncating MAPK pathways. Phosphorylation of c-JUN with GSTP1 releases anti-apoptotic signals that safeguards cancer cells from apoptosis. Inhibition of GSTP1 will lead to the activation of the pathway specified for apoptosis. GSTP1 will not be able to provide shelter to tumor cells from apoptosis.

187 ligands were docked to GSTP1 which yielded a list of binding affinities of all the ligands. the ligands that were docked included various inhibitors of GSTP1, drugs for Kaposi sarcoma, and anti-tumor drugs. The maximum binding affinity observed on docking was -10.1 kcal/mol by DB06595. This binding affinity can be further improved by modification of DB06595 by changing the functional group and replacing it with a functional group that produces less stearic hindrance between GSTP1 and ligand.

GSTP1 marks itself as a primary contributor to tumorigenesis by upregulation in Kaposi sarcoma. It can be targeted with for inhibition with existing therapeutics and composition of new therapeutics for better results.

REFERENCES

[1] G. P. Dunn, L. J. Old, and R. D. Schreiber, "The immunobiology of cancer immunosurveillance and immunoediting," *Immunity*, vol. 21, no. 2, pp. 137–148, 2004.

[2] J. S. You and P. A. Jones, "Cancer genetics and epigenetics: two sides of the same coin," *Cancer cell*, vol. 22, no. 1, pp. 9–20, 2012.

[3] H. Zhao, J. Wang, Y. Han, Z. Huang, J. Ying, X. Bi, J. Zhao, Y. Fang, H. Zhou, J. Zhou *et al.*, "Arid2: a new tumor suppressor gene in hepatocellular carcinoma," *Oncotarget*, vol. 2, no. 11, p. 886, 2011.

[4] S. Ogino, J. Galon, C. S. Fuchs, and G. Dranoff, "Cancer immunology—analysis of host and tumor factors for personalized medicine," *Nature reviews Clinical oncology*, vol. 8, no. 12, pp. 711–719, 2011.

[5] O. J. Finn, "Cancer immunology," *New England Journal of Medicine*, vol. 358, no. 25, pp. 2704–2715, 2008.

[6] C. Carrillo-Infante, G. Abbadessa, L. Bagella, and A. Giordano, "Viral infections as a cause of cancer," *International journal of oncology*, vol. 30, no. 6, pp. 1521–1528, 2007.

[7] E. A. Mesri, M. A. Feitelson, and K. Munger, "Human viral oncogenesis: a cancer hallmarks analysis," *Cell host & microbe*, vol. 15, no. 3, pp. 266–282, 2014.

[8] R. Vangipuram and S. K. Tyring, "Epidemiology of kaposi sarcoma: review and description of the nonepidemic variant," *International Journal of Dermatology*, vol. 58, no. 5, pp. 538–542, 2019.

[9] M. Kaposi, "Idiopathic multiple pigmented sarcoma of the skin," *CA: A Cancer Journal for Clinicians*, vol. 32, no. 6, pp. 342–347, 1982.

[10] K. Ariyoshi, P. Cook, D. Whitby, T. Corrah, S. Jaffar, F. Cham, S. Sabally, D. O'Donovan, R. Weiss, T. Schulz *et al.*, "Kaposi's sarcoma in the gambia, west africa is less frequent in human immunodeficiency virus type 2 than in human immunodeficiency virus type 1 infection despite a high prevalence of human herpesvirus 8." *Journal of human virology*, vol. 1, no. 3, pp. 193–199, 1998.

[11] G. J. Gottlieb, A. Ragaz, J. V. Vogel, A. Friedman-Kien, A. M. Rywlin, E. A. Weiner, and A. B. Ackerman, "A preliminary communication on extensively disseminated kaposi's sarcoma in young homosexual men," *The American Journal of Dermatopathology*, vol. 3, no. 2, pp. 111–114, 1981.

[12] A. Jary, M. Veyri, A. Gothland, V. Leducq, V. Calvez, and A.-G. Marcelin, "Kaposi's sarcoma-associated herpesvirus, the etiological agent of all epidemiological forms of kaposi's sarcoma," *Cancers*, vol. 13, no. 24, p. 6208, 2021.

[13] T. H. Kim, S. Y. Wee, H. G. Jeong, and H. J. Choi, "Misdiagnosis of human herpes virus-8-associated kaposi's sarcoma as adverse drug eruptions," *Archives of Plastic Surgery*, vol. 49, no. 03, pp. 457–461, 2022.

[14] E. Cesarman, B. Damania, S. E. Krown, J. Martin, M. Bower, and D. Whitby, "Kaposi sarcoma," *Nature reviews Disease primers*, vol. 5, no. 1, pp. 1–21, 2019.

[15] R. Swali, A. Limmer, and S. K. Tyring, "Kaposi sarcoma of the medial foot in an msm, hiv-negative man: A fifth clinical variant," *The Journal of Clinical and Aesthetic Dermatology*, vol. 13, no. 10, p. 42, 2020.

[16] E. Ruocco, V. Ruocco, M. L. Tornesello, A. Gambardella, R. Wolf, and F. M. Buonaguro, "Kaposi's sarcoma: etiology and pathogenesis, inducing factors, causal associations, and treatments: facts and controversies," *Clinics in dermatology*, vol. 31, no. 4, pp. 413–422, 2013.

[17] J. De Waal and W. Dreyer, "Oral medicine case book 2," *South African Dental Journal*, vol. 62, no. 9, p. 406, 2007.

[18] G. Stallone, A. Schena, B. Infante, S. Di Paolo, A. Loverre, G. Maggio, E. Ranieri, L. Gesualdo, F. P. Schena, and G. Grandaliano, "Sirolimus for kaposi's sarcoma in renal-transplant recipients," *New England Journal of Medicine*, vol. 352, no. 13, pp. 1317–1323, 2005.

[19] C. for Disease Control *et al.*, "Kaposi's sarcoma and pneumocystis pneumonia among homosexual men-new york city and california," *mmwr*, vol. 30, pp. 305–308, 1981.

[20] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[21] W. Phipps, F. Ssewankambo, H. Nguyen, M. Saracino, A. Wald, L. Corey, J. Orem, A. Kambugu, and C. Casper, "Gender differences in clinical presentation and outcomes of epidemic kaposi sarcoma in uganda," *PloS one*, vol. 5, no. 11, p. e13936, 2010.

[22] I. Altunay, A. Kucukunal, G. T. Demirci, and B. Ates, "Variable clinical presentations of classic kaposi sarcoma in turkish patients," *Journal of dermatological case reports*, vol. 6, no. 1, p. 8, 2012.

[23] C. Requena, M. Alsina, D. Morgado-Carrasco, J. Cruz, O. Sanmartín, C. Serra-Guillén, and B. Llombart, "Kaposi sarcoma and cutaneous angiosarcoma: guidelines for diagnosis and treatment," *Actas Dermo-Sifiliográficas (English Edition)*, vol. 109, no. 10, pp. 878–887, 2018.

[24] P. Volkow, G. Cesarman-Maus, P. Garciadiego-Fossas, E. Rojas-Marin, and P. Cornejo-Juárez, "Clinical characteristics, predictors of immune reconstitution inflammatory syndrome and long-term prognosis in patients with kaposi sarcoma," *AIDS research and therapy*, vol. 14, no. 1, pp. 1–9, 2017.

[25] T. S. Uldrick, V. Wang, D. O'Mahony, K. Aleman, K. M. Wyvill, V. Marshall, S. M. Steinberg, S. Pittaluga, I. Maric, D. Whitby *et al.*, "An interleukin-6-related systemic inflammatory syndrome in patients co-infected with kaposi sarcoma-associated herpesvirus and hiv but without multicentric castleman disease," *Clinical Infectious Diseases*, vol. 51, no. 3, pp. 350–358, 2010.

[26] K. Ueda, "Kshv genome replication and maintenance in latency," *Human Herpesviruses*, pp. 299–320, 2018.

[27] J. Cai, P. Gill, R. Masood, P. Chandrasoma, B. Jung, R. Law, and S. Radka, "Oncostatin-m is an autocrine growth factor in kaposi's sarcoma." *The American journal of pathology*, vol. 145, no. 1, p. 74, 1994.

[28] B. Abere, T. M. Mamo, S. Hartmann, N. Samarina, E. Hage, J. Rückert, S.-K. Hotop, G. Büsche, and T. F. Schulz, "The kaposi's sarcoma-associated herpesvirus (kshv) non-structural membrane protein k15 is required for viral lytic replication and may represent a therapeutic target," *PLoS pathogens*, vol. 13, no. 9, p. e1006639, 2017.

[29] A. Ray, V. Marshall, T. Uldrick, R. Leighty, N. Labo, K. Wyvill, K. Aleman, M. N. Polizzotto, R. F. Little, R. Yarchoan *et al.*, "Sequence analysis of kaposi sarcoma–associated herpesvirus (kshv) micrornas in patients with multicentric castleman disease and kshv-associated inflammatory cytokine syndrome," *The Journal of infectious diseases*, vol. 205, no. 11, pp. 1665–1676, 2012.

[30] P. Bellare and D. Ganem, "Regulation of kshv lytic switch protein expression by a virus-encoded microrna: an evolutionary adaptation that fine-tunes lytic reactivation," *Cell host & microbe*, vol. 6, no. 6, pp. 570–575, 2009.

[31] D. M. Knipe, P. Raja, and J. Lee, "Viral gene products actively promote latent infection by epigenetic silencing mechanisms," *Current opinion in virology*, vol. 23, pp. 68–74, 2017.

[32] L. E. Cavallin, Q. Ma, J. Naipauer, S. Gupta, M. Kurian, P. Locatelli, P. Romanelli, M. Nadji, P. J. Goldschmidt-Clermont, and E. A. Mesri, "Kshv-induced ligand mediated activation of pdgf receptor-alpha drives kaposi's sarcomagenesis," *PLoS pathogens*, vol. 14, no. 7, p. e1007175, 2018.

[33] S. Gramolelli, M. Weidner-Glunde, B. Abere, A. Viejo-Borbolla, K. Bala, J. Rückert, E. Kremmer, and T. F. Schulz, "Inhibiting the recruitment of plcγ1 to kaposi's sarcoma herpesvirus k15 protein reduces the invasiveness and angiogenesis of infected endothelial cells," *PLoS pathogens*, vol. 11, no. 8, p. e1005105, 2015.

[34] J. Suthaus, C. Stuhlmann-Laeisz, V. S. Tompkins, T. R. Rosean, W. Klapper, G. Tosato, S. Janz, J. Scheller, and S. Rose-John, "Hhv-8–encoded viral il-6 collaborates with mouse il-6 in the development of multicentric castleman disease in mice," *Blood, The Journal of the American Society of Hematology*, vol. 119, no. 22, pp. 5173–5181, 2012.

[35] D. Avey, S. Tepper, W. Li, Z. Turpin, and F. Zhu, "Phosphoproteomic analysis of kshv-infected cells reveals roles of orf45-activated rsk during lytic replication," *PLoS pathogens*, vol. 11, no. 7, p. e1004993, 2015.

[36] G. Ballon, G. Akar, and E. Cesarman, "Systemic expression of kaposi sarcoma herpesvirus (kshv) vflip in endothelial cells leads to a profound proinflammatory phenotype and myeloid lineage remodeling in vivo," *PLoS pathogens*, vol. 11, no. 1, p. e1004581, 2015.

[37] J. West and B. Damania, "Upregulation of the tlr3 pathway by kaposi's sarcoma-associated herpesvirus during primary infection," *Journal of virology*, vol. 82, no. 11, pp. 5440–5449, 2008.

[38] N. Kerur, M. V. Veettil, N. Sharma-Walia, V. Bottero, S. Sadagopan, P. Otageri, and B. Chandran, "Ifi16 acts as a nuclear pathogen sensor to induce the inflammasome in response to kaposi sarcoma-associated herpesvirus infection," *Cell host & microbe*, vol. 9, no. 5, pp. 363–375, 2011.

[39] W. Li, D. Avey, B. Fu, J.-j. Wu, S. Ma, X. Liu, and F. Zhu, "Kaposi's sarcoma-associated herpesvirus inhibitor of cgas (kicgas), encoded by orf52, is an abundant tegument protein and is required for production of infectious progeny viruses," *Journal of virology*, vol. 90, no. 11, pp. 5329–5342, 2016.

[40] P. Zhang, J. Wang, X. Zhang, X. Wang, L. Jiang, and X. Gu, "Identification of aids-associated kaposi sarcoma: A functional genomics approach," *Frontiers in Genetics*, vol. 10, p. 1376, 2020.

[41] N. Sallah, A. L. Palser, S. J. Watson, N. Labo, G. Asiki, V. Marshall, R. Newton, D. Whitby, P. Kellam, and I. Barroso, "Genome-wide sequence analysis of kaposi sarcoma-associated herpesvirus shows diversification driven by recombination," *The Journal of infectious diseases*, vol. 218, no. 11, pp. 1700–1710, 2018.

[42] G. Broussard and B. Damania, "Kshv: immune modulation and immunotherapy," *Frontiers in Immunology*, vol. 10, p. 3084, 2020.

[43] G. Golas, S. J. Jang, N. G. Naik, J. D. Alonso, B. Papp, and Z. Toth, "Comparative analysis of the viral interferon regulatory factors of kshv for their requisite for virus production and inhibition of the type i interferon pathway," *Virology*, vol. 541, pp. 160–173, 2020.

[44] M. Ceccarelli, A. Facciolà, R. Taibi, G. Pellicanò, G. Nunnari, and E. Venanzi Rullo, "The treatment of kaposi's sarcoma: Present and future options, a review of the literature," *Eur. Rev. Med. Pharmacol. Sci*, vol. 23, pp. 7488–7497, 2019.

[45] E. Régnier-Rosencher, B. Guillot, and N. Dupin, "Treatments for classic kaposi sarcoma: a systematic review of the literature," *Journal of the American Academy of Dermatology*, vol. 68, no. 2, pp. 313–331, 2013.

[46] C. Lebbe, C. Garbe, A. J. Stratigos, C. Harwood, K. Peris, V. Del Marmol, J. Malvehy, I. Zalaudek, C. Hoeller, R. Dummer *et al.*, "Diagnosis and treatment of kaposi's sarcoma: European consensus-based interdisciplinary guideline (edf/eado/eortc)," *European Journal of Cancer*, vol. 114, pp. 117–127, 2019.

[47] J. W. Schneider and D. P. Dittmer, "Diagnosis and treatment of kaposi sarcoma," *American journal of clinical dermatology*, vol. 18, no. 4, pp. 529–539, 2017.

[48] T. Malati, "Tumour markers: An overview," *Indian Journal of Clinical Biochemistry*, vol. 22, no. 2, pp. 17–31, 2007.

[49] B. T. Flepisi, P. Bouic, G. Sissolak, and B. Rosenkranz, "Biomarkers of hiv-associated cancer," *Biomarkers in cancer*, vol. 6, pp. BIC–S15 056, 2014.

[50] P. R. Srinivas, B. S. Kramer, and S. Srivastava, "Trends in biomarker research for cancer detection," *The lancet oncology*, vol. 2, no. 11, pp. 698–704, 2001.

[51] I. Sereti, A. J. Rodger, and M. A. French, "Biomarkers in immune reconstitution inflammatory syndrome: signals from pathogenesis," *Current Opinion in HIV and AIDS*, vol. 5, no. 6, p. 504, 2010.

[52] R. F. Ambinder, K. Bhatia, O. Martinez-Maza, and R. Mitsuyasu, "Cancer biomarkers in hiv patients," *Current Opinion in HIV and AIDS*, vol. 5, no. 6, p. 531, 2010.

[53] T. S. Uldrick and D. Whitby, "Update on kshv epidemiology, kaposi sarcoma pathogenesis, and treatment of kaposi sarcoma," *Cancer letters*, vol. 305, no. 2, pp. 150–162, 2011.

[54] S. Behjati and P. S. Tarpey, "What is next generation sequencing?" *Archives of Disease in Childhood-Education and Practice*, vol. 98, no. 6, pp. 236–238, 2013.

[55] J. S. Reis-Filho, "Next-generation sequencing," *Breast cancer research*, vol. 11, no. 3, pp. 1–7, 2009.

[56] M. Hong, S. Tao, L. Zhang, L.-T. Diao, X. Huang, S. Huang, S.-J. Xie, Z.-D. Xiao, and H. Zhang, "Rna sequencing: new technologies and applications in cancer research," *Journal of hematology & oncology*, vol. 13, no. 1, pp. 1–16, 2020.

[57] S. Marguerat and J. Bähler, "Rna-seq: from technology to biology," *Cellular and molecular life sciences*, vol. 67, no. 4, pp. 569–579, 2010.

[58] R. B. Stoughton, "Applications of dna microarrays in biology," *Annu. Rev. Biochem.*, vol. 74, pp. 53–82, 2005.

[59] G. M. Morris and M. Lim-Wilby, "Molecular docking," in *Molecular modeling of proteins*. Springer, 2008, pp. 365–382.

[60] L. Lynen, M. Zolfo, V. Huyst, F. Louis, P. Barnardt, A. Van de Velde, C. De Schacht, and R. Colebunders, "Management of kaposi's sarcoma in resource-limited settings in the era of haart," *AIDS Rev*, vol. 7, no. 1, pp. 13–21, 2005.

[61] R. A. Schwartz, G. Micali, M. R. Nasca, and L. Scuderi, "Kaposi sarcoma: a continuing conundrum," *Journal of the American Academy of Dermatology*, vol. 59, no. 2, pp. 179–206, 2008.

[62] W. Dreyer and J. De Waal, "Oral medicine case book 21," *South African Dental Journal*, vol. 64, no. 8, p. 362, 2009.

[63] T. Bottler, J. Kuttenberger, N. Hardt, H.-P. Oehen, and M. Baltensperger, "Non-hiv-associated kaposi's sarcoma of the tongue: case report and review of the literature," *International journal of oral and maxillofacial surgery*, vol. 36, no. 12, pp. 1218–1220, 2007.

[64] J. E. Naschitz and M. Lurie, "Macular palmo-plantar eruption," *European Journal of Internal Medicine*, vol. 20, no. 5, pp. e118–e119, 2009.

[65] Z. Papagatsia, J. Jones, P. Morgan, and A. R. Tappuni, "Oral kaposi sarcoma: a case of immune reconstitution inflammatory syndrome," *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, vol. 108, no. 1, pp. 70–75, 2009.

[66] W. Louthrenoo, N. Kasitanon, P. Mahanuphab, L. Bhoopat, and S. Thongprasert, "Kaposi's sarcoma in rheumatic diseases," in *Seminars in arthritis and rheumatism*, vol. 32, no. 5. Elsevier, 2003, pp. 326–333.

[67] D. Serraino, C. Angeletti, M. P. Carrieri, B. Longo, M. Piche, P. Piselli, E. Arbustini, P. Burra, F. Citterio, V. Colombo *et al.*, "Kaposi's sarcoma in transplant and hiv-infected patients: an epidemiologic study in italy and france," *Transplantation*, vol. 80, no. 12, pp. 1699–1704, 2005.

[68] V. Ramírez-Amador, G. Anaya-Saavedra, and G. Martínez-Mata, "Kaposi's sarcoma of the head and neck: a review," *Oral oncology*, vol. 46, no. 3, pp. 135–145, 2010.

[69] F. Martellotta, M. Berretta, E. Vaccher, O. Schioppa, E. Zanet, and U. Tirelli, "Aids-related kaposi's sarcoma: state of the art and therapeutic strategies," *Current HIV research*, vol. 7, no. 6, pp. 634–638, 2009.

[70] J. Herrada, F. Cabanillas, L. Rice, J. Manning, and W. Pugh, "The clinical behavior of localized and multicentric castleman disease," *Annals of Internal Medicine*, vol. 128, no. 8, pp. 657–662, 1998.

[71] Y.-B. Chen, A. Rahemtullah, and E. Hochberg, "Primary effusion lymphoma," *The Oncologist*, vol. 12, no. 5, pp. 569–576, 2007.

[72] J. Fröhlich and A. Grundhoff, "Epigenetic control in kaposi sarcoma-associated herpesvirus infection and associated disease," in *Seminars in Immunopathology*, vol. 42, no. 2. Springer, 2020, pp. 143–157.

[73] M. Pulitzer, "Molecular diagnosis of infection-related cancers in dermatopathology," in *Seminars in Cutaneous Medicine and Surgery*, vol. 31, no. 4. WB Saunders, 2012, pp. 247–257.

[74] X. Cai, S. Lu, Z. Zhang, C. M. Gonzalez, B. Damania, and B. R. Cullen, "Kaposi's sarcoma-associated herpesvirus expresses an array of viral microrna in latently infected cells," *Proceedings of the National Academy of Sciences*, vol. 102, no. 15, pp. 5570–5575, 2005.

[75] P. M. Ojala, M. Tiainen, P. Salven, T. Veikkola, E. Castanos-Velez, R. Sarid, P. Biberfeld, and T. P. Makela, "Kaposi's sarcoma-associated herpesvirus-encoded v-cyclin triggers apoptosis in cells with high levels of cyclin-dependent kinase 6," *Cancer research*, vol. 59, no. 19, pp. 4984–4989, 1999.

[76] J. Friborg, W.-p. Kong, M. O. Hottiger, and G. J. Nabel, "p53 inhibition by the lana protein of kshv protects against cell death," *Nature*, vol. 402, no. 6764, pp. 889–894, 1999.

[77] M. Faris, B. Ensoli, N. Kokot, and A. E. Nel, "Inflammatory cytokines induce the expression of basic fibroblast growth factor (bfgf) isoforms required for the growth of kaposi's sarcoma and endothelial cells through the activation of ap-1 response elements in the bfgf promoter," *Aids*, vol. 12, no. 1, pp. 19–27, 1998.

[78] G. Ballon, K. Chen, R. Perez, W. Tam, E. Cesarman *et al.*, "Kaposi sarcoma herpesvirus (kshv) vflip oncoprotein induces b cell transdifferentiation and tumorigenesis in mice," *The Journal of clinical investigation*, vol. 121, no. 3, pp. 1141–1153, 2011.

[79] W.-X. Guo, T. Antakly, M. Cadotte, Z. Kachra, L. Kunkel, R. Masood, and P. Gill, "Expression and cytokine regulation of glucocorticoid receptors in kaposi's sarcoma." *The American journal of pathology*, vol. 148, no. 6, p. 1999, 1996.

[80] H.-W. Wang, M. W. Trotter, D. Lagos, D. Bourboulia, S. Henderson, T. Mäkinen, S. Elliman, A. M. Flanagan, K. Alitalo, and C. Boshoff, "Kaposi sarcoma herpesvirus–induced cellular reprogramming contributes to the lymphatic endothelial gene expression in kaposi sarcoma," *Nature genetics*, vol. 36, no. 7, pp. 687–693, 2004.