



Faculty of Resource Science and Technology

Computational Analysis of Epstein-Barr Virus *BamH1* A Rightward Transcript (BART) MicroRNA (miRNA) Regulation on Messenger and Long Non-Coding RNAs in Nasopharyngeal Cancer

Daphne Olivia Jawai Anak Sadai

**Master of Science
2024**

Computational Analysis of Epstein-Barr Virus *BamH1* A Rightward
Transcript (BART) MicroRNA (miRNA) Regulation on Messenger and Long
Non-Coding RNAs in Nasopharyngeal Cancer

Daphne Olivia Jawai Anak Sadai

A thesis submitted

In fulfillment of the requirements for the degree of Master of Science

(Genetics)

Faculty of Resource Science and Technology
UNIVERSITI MALAYSIA SARAWAK

2024

DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Malaysia Sarawak. Except where due acknowledgements have been made, the work is that of the author alone. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.



.....

Signature

Name: Daphne Olivia Jawai Anak Sadai

Matric No.: 20020007

Faculty of Resource Science and Technology

Universiti Malaysia Sarawak

Date : 19/3/2024

ACKNOWLEDGEMENT

First and foremost, I would like to express my gratitude to my supervisor, Prof. Dr. Edmund Sim Ui Hang, for his invaluable supervision and guidance in completing this study. Besides, I appreciate the knowledge he shared with me and his assistance in writing my thesis throughout this study. Without his help, this study would not have been a success and completed.

Next, special thanks to my fellow labmates and friends, Ms. Adrianne Vong, Ms. Alyaa Syafiqah, and Mr. Alvin Wee, for sharing the knowledge and have consistently and patiently helped and guided me throughout my study. It was soothing to have them to confide in when I felt like giving up and I was also grateful for the emotional support they gave me during the ups and downs in completing this study. I am thankful for the times we have spent together and the safety they offered throughout this journey.

Furthermore, I would like to thank the Faculty of Resource Science and Technology of UNIMAS for providing access to the facilities, particularly the Immunology Laboratory and Animal and Tissue Culture Laboratory, which enabled me to learn and gain new knowledge by giving me the opportunity to conduct laboratory-based experiments.

Lastly, I would like to thank my parents, Mrs. Pepiet and Mr. Sadai, for their endless support and monetary. My mother, especially, for her endless emotional support, understanding, and words of encouragement, which have been fuelling me to complete this study when I feel hopeless. Thank you so much for allowing this journey to happen.

ABSTRACT

The interaction and regulation amongst messenger RNA (mRNA), microRNA (miRNA) and long non-coding RNA (lncRNA) have been identified as causative of nasopharyngeal cancer (NPC) when their expression levels are dysregulated by each other and affecting the biological and molecular functions of affected nasopharynx cells. This interaction is known as competing endogenous RNA (ceRNA) regulatory activity. However, the study of regulation mechanisms of *BamHI A* rightward transcript (BART) miRNAs encoded by Epstein-Barr virus (EBV) on ceRNA in the pathogenesis of NPC is unavailable. According to previous studies, the Epstein-Barr virus was deemed one of NPC causative factors through its regulation of ceRNA activity using BART miRNAs. Therefore, in this study, candidate BART miRNAs, mRNAs, miRNAs and lncRNAs were obtained by integrating multi-level RNAs expressions data through bioinformatics analysis by constructing the cross-regulatory network. The cross-regulatory network was the combination of mRNA-miRNA-lncRNA ceRNA network with EBV miRNAs-mRNAs network. The results revealed six EBV miRNAs (ebv-miR-BART21-3p, ebv-miR-BART19-3p, ebv-miR-BART15, ebv-miR-BART2-5p, ebv-miR-BART20-3p and ebv-miR-BART11-5p) were interacting with four mRNAs (EYA4, EYA1, EBF1 and MACROD2). In addition, these mRNAs were interacting with six miRNAs (hsa-miR-1246, hsa-miR-93-5p, hsa-miR-16-5p, hsa-miR-135b-5p, hsa-miR-211-5p and hsa-miR-1305) where these miRNAs were interacting with three lncRNAs which are CASC2, TPTE2P1 and ARHGEF26-AS1. The data attained showed that these BART miRNAs may deregulate these mRNAs and affect the expressions of both miRNAs and lncRNAs. This gives insight into understanding the regulatory mechanisms of BART miRNAs on ceRNA activity. This study also revealed the potential functions of BART miRNAs and lncRNAs acting similarly as either tumour

suppressors or oncogenes. Furthermore, the data obtained suggested that BART miRNAs affect DNA repair regulation and apoptosis in NPC by down-regulating the mRNAs, which pertains to poor overall survival (OS) in NPC patients. Overall, these findings exhibited the potential role of these BART miRNAs, mRNAs and lncRNAs as diagnostic and prognostic biomarkers in EBV-induced NPC carcinogenesis based on their ability to target and regulate the expressions of each other.

Keywords: Nasopharyngeal Cancer, EBV, Bioinformatics, BART miRNAs, Long Non-Coding RNA

***Analisis Komputasi ke atas Pengawalan MikroRNA (miRNA) Transkrip BamHI A
Rightward (BART) Diekspresikan Oleh Virus Epstein-Barr Ke atas RNA Bukan
Pengekoden yang Panjang Dalam Kanser Nasofarink***

ABSTRAK

Interaksi dan pengawalan di antara pengutus RNA (mRNA), mikroRNA (miRNA) dan RNA bukan pengekoden yang panjang (lncRNA) telah dibuktikan sebagai penyebab dalam pembentukan kanser nasofarink (NPC) apabila tahap ekspresi mereka disregulasi oleh sama sendiri dan hal ini didapati memberi kesan terhadap fungsi-fungsi biologikal dan molekular di dalam sel-sel nasofarink yang terkesan. Interaksi ini dikenali sebagai aktiviti pengawalan persaingan asid ribonukleik yang berasal dari dalam (ceRNA). Walau bagaimanapun, kajian terhadap mekanisme pengawalan oleh miRNA transkrip BamHI A rightward (BART) yang diekspreskan oleh virus Epstein-Barr (EBV) keatas ceRNA didalam pembentukan NPC belum pernah dikaji. Menurut kajian-kajian terdahulu, Epstein-Barr virus merupakan salah satu faktor penyebab kepada NPC melalui pengawalan keatas aktiviti ceRNA dengan menggunakan miRNA BART. Oleh itu, kajian ini calon miRNA BART, mRNA, miRNA dan lncRNA mampu didedahkan dengan mengabungkan data ekspresi berbilang peringkat asid ribonukleik (RNA) melalui analisa bioinformatik dengan membina rangkaian silang kawal selia. Rangkaian silang kawal selia ini merupakan pengabungan antara rangkaian mRNA-miRNA-lncRNA ceRNA dengan rangkaian EBV miRNA-mRNA. Hasil kajian ini menunjukkan bahawa enam miRNA EBV (ebv-miR-BART21-3p, ebv-miR-BART19-3p, ebv-miR-BART15, ebv-miR-BART2-5p, ebv-miR-BART20-3p dan ebv-miR-BART11-5p) berinteraksi dengan empat mRNA (EYA4, EYA1, EBF1 dan MACROD2). Sehubungan dengan itu, mRNA tersebut juga berinteraksi dengan enam miRNA (hsa-miR-1246, hsa-miR-93-5p, hsa-miR-16-5p, hsa-miR-135b-5p, hsa-miR-211-5p dan hsa-miR-1305) dimana miRNA tersebut juga

berinteraksi dengan tiga lncRNA yaitu CASC2, TPTE2P1 dan ARHGEF26-AS1. Data yang diperoleh ini menunjukkan bahwa miRNA BART mendisregulasi mRNA tersebut dan memberi kesan terhadap ekspresi-ekspresi miRNA dan lncRNA. Ini memberi pengertian terhadap kefahaman tentang mekanisme pengawalan miRNA BART ke atas aktiviti pengawalan ceRNA. Kajian ini juga telah mendedahkan potensi miRNA BART dan lncRNA berfungsi menyerupai pembantut tumor dan onkogene. Tambahan pula, data yang diperoleh mencadangkan bahawa miRNA BART memberi kesan terhadap pengawalan pemulihan asid deoksiribonukleik (DNA) dan apoptosis di dalam NPC dengan menurunkan ekspresi mRNA dan hal ini dikaitkan dengan kelangsungan hidup secara keseluruhan yang buruk di dalam pesakit-pesakit NPC. Secara keseluruhannya, penemuan-penemuan ini mempamerkan potensi miRNA BART, mRNA, miRNA dan lncRNA sebagai biomarker diagnostik dan prognostik dalam NPC karsinogenesis yang disebabkan oleh EBV berdasarkan keupayaan mereka dalam mensasarkan dan mengawal ekspresi satu sama lain.

Kata kunci: *Kanser nasofarinks, virus Epstein-Barr, mikroRNA BART, RNA bukan pengkodan yang panjang*

TABLE OF CONTENTS

	Page
DECLARATION	i
ACKNOWLEDGEMENT	ii
ABSTRACT	iii
<i>ABSTRAK</i>	v
TABLE OF CONTENTS	vii
LIST OF TABLES	xii
LIST OF FIGURES	xv
LIST OF ABBREVIATIONS	xx
CHAPTER 1 INTRODUCTION	1
1.1 Study Background	1
1.2 Problem Statement	2
1.3 Objectives and Hypothesis	3
CHAPTER 2 LITERATURE REVIEW	5
2.1 Nasopharyngeal Cancer	5
2.1.1 Histopathology of Nasopharyngeal Cancer	6
2.1.2 Epidemiology of Nasopharyngeal Cancer	7
2.1.3 Etiology of Nasopharyngeal Cancer	8

2.1.4	Diagnosis, Prognosis and Treatment of Nasopharyngeal Cancer	10
2.2	Epstein-Barr Virus	13
2.2.1	Epidemiology of Epstein-Barr Virus	14
2.2.2	Etiology of Epstein-Barr Virus	15
2.2.3	Lytic and Latent Cycle of Epstein-Barr Virus	17
2.2.4	Epstein-Barr Virus Induced Nasopharyngeal Cancer	18
2.2.5	Epstein-Barr Virus MicroRNA	20
2.2.6	<i>Bam</i> H1 Fragment A Rightward Transcript MicroRNA	21
2.3	Competing Endogenous RNA (ceRNA)	23
2.3.1	MicroRNA (miRNA)	26
2.3.2	Long Non-Coding RNA (lncRNA)	28
2.4	Bioinformatics	30
2.4.1	The GEO (Gene Expression Omnibus) Database	30
2.4.2	The GEO2R	30
2.4.3	The miRWalk 2.0 Database	31
2.4.4	MicroRNA Target Interaction Database (miRTarBase)	31
2.4.5	DIANA-LncBase V2 and V3 database	32
2.4.6	StarBase V2	33
2.4.7	miRcode Database	34
2.4.8	TFcheckpoint database	34

2.4.9	ViRBase	35
2.4.10	The Search Tool for Retrieval of Interacting Genes (STRING) Database	35
2.4.11	Cytoscape Software	36
2.4.12	The Database of Annotation, Visualization and Discovery software (DAVID)	38
2.4.13	Kaplan-Meier (KM) Plotter	39
CHAPTER 3 MATERIALS AND METHODS		42
3.1	Microarray Dataset Retrieval	42
3.2	Differential Expressed Gene, lncRNA and miRNAs Identification	42
3.2.1	Differential Expressed Gene, lncRNA, miRNA Visualization	43
3.3	Target mRNA and Target lncRNA of DE miRNA Identification	43
3.3.1	DE miRNA-Target mRNA Interaction Identification	43
3.3.2	DE miRNA-Target lncRNA Interaction Identification	44
3.4	Competing Endogenous RNA (ceRNA) Network Construction	44
3.5	Epstein-Barr Virus miRNA-mRNA Interaction	45
3.5.1	Epstein-Barr Virus miRNAs-mRNAs Interaction Identification	45
3.5.2	Epstein-Barr Virus miRNA-mRNA Interaction Network Construction	45
3.6	Protein-Protein Interaction (PPI)	46
3.6.1	Protein-Protein Interaction (PPI) Network Retrieval and Construction	46
3.6.2	Hub Genes Identification from The PPI	46
3.6.3	Gene Modules Identification	47

3.7	Transcription Factors (TF)-mRNAs Interaction and Network	47
3.7.1	Transcription Factors Identification	47
3.7.2	Transcription Factors (TF)-mRNAs Network Construction	47
3.8	Cross Regulatory Networks Construction	48
3.9	Functional and Pathway Enrichment Analysis of mRNAs	48
3.10	Survival Curve Analysis	48
CHAPTER 4 RESULTS		51
4.1	Microarray Dataset	51
4.2	Differential Expressed Gene, lncRNA and miRNA	52
4.2.1	Differential Expressed Gene, lncRNA and miRNA Visualization	53
4.2.2	Differential Expressed Gene, lncRNA and miRNA Overlapping	56
4.3	Target mRNA and Target lncRNA Interaction with DE miRNA	59
4.3.1	DE miRNA-Target mRNA Interaction	59
4.3.2	DE miRNA-Target lncRNA Interaction	65
4.4	mRNA-miRNA-lncRNA Competing Endogenous RNA (ceRNA) Network	67
4.5	Epstein-Barr Virus miRNA-mRNA Interaction	68
4.5.1	Epstein-Barr Virus miRNAs-mRNAs Interaction	68
4.5.2	Epstein-Barr Virus miRNA-mRNA Interaction Network	69
4.6	Protein-Protein Interaction (PPI)	70
4.6.1	Protein-Protein Interaction (PPI) Network	70

4.6.2	Hub Genes from The Protein-Protein Interaction (PPI)	72
4.6.3	Gene Modules from The Protein-Protein Interaction (PPI)	76
4.7	Transcription Factors (TF)-mRNAs Interaction	78
4.7.1	Transcription Factors (TF)-mRNAs Network	78
4.8	Cross-Regulatory Network (ceRNA and EBV miRNAs-mRNAs network)	79
4.9	Functional and Pathway Enrichment Analysis	82
4.10	Survival Curve	84
CHAPTER 5 DISCUSSION		86
5.1	Competing Endogenous RNA (CeRNA) Network	86
5.2	Cross-Regulatory Network	94
5.2.1	Functional and Pathway Enrichment Analysis of The Four DEGs	98
5.3	Overall Survival	108
CHAPTER 6 CONCLUSIONS AND RECOMMENDATIONS		112
6.1	Conclusion	112
6.2	Recommendations	113
REFERENCES		115
APPENDICES		159

LIST OF TABLES

	Page
Table 2.1 The functions of each EBV protein, miRNAs and snRNAs transcribed during both latent and infection of EBV in host cells (Farrell. 2019)	16
Table 2.2 The latency patterns of gene expression in EBV-infected cells during latent infection (Farrell, 2019)	18
Table 4.1 The information of the NPC RNAs microarray datasets for each of the five mRNAs, miRNAs and lncRNAs datasets.	51
Table 4.2 The number of the up-regulated and down-regulated RNAs in each of the five microarray datasets curated from the GEO database. The number of the up-regulated and down-regulated RNAs in each of the five microarray datasets curated from the GEO database.	52
Table 4.3 The lists of common up-regulated and down-regulated mRNAs between the datasets GSE95166, GSE126683 and GSE61218.	57
Table 4.4 Target mRNAs of up-regulated DE miRNAs which were overlapped with the common down-regulated GEO DEGs from the three mRNA datasets (GSE64634, GSE126683 and GSE61218).	60
Table 4.5 Target mRNAs of down-regulated DE miRNAs which were overlapped with the common up-regulated GEO DEGs from the three mRNA datasets (GSE64634, GSE126683 and GSE61218).	63

Table 4.6	Target lncRNA of up-regulated DE miRNAs which were overlapped with the common down-regulated GEO lncRNAs from the three lncRNAs datasets (GSE95166, GSE126683 and GSE61218).	66
Table 4.7	The top 20 DEGs with connectivity degree more than 5 from the PPI ranked according to connectivity degree by CytoHubba. The score showed the number of connections of each DEG with other DEGs in the PPI. DNAH5 was ranked 1 st with 21 connections to other DEGs in the PPI, and in the contrary DNAAF1 was ranked last with seven connections to other DEGs in the PPI.	73
Table 4.8	The 18 DEGs with connectivity degree more than 5 excluding self-loop among the Top 20 DEGs. The connectivity degree showed the number of connections of each of the 18 DEGs amongst each other. It showed DNAH5 has the most connection with a connectivity degree of 18 with the other DEGs, whereas DNAAF1 and TEK1 have the lowest connection with the other DEGs with only 6 interactions.	75
Table 4.9	The list of DEGs in Module 1 and Module 2, extracted from the PPI network. Module 1 consisted of 9 DEGs with 36 interactions between them, with the highest score of 9.000, whereas Module 2 consisted of 4 DEGs with only six interactions between them, thus a low score of 4.000.	77
Table 4.10	The interaction between BART miRNAs and RNAs in the cross-regulatory network. The down-regulation of the four mRNAs	

(EYA4, EYA1, EBF1 and MACROD2) due to BART miRNAs caused the up-regulation of the six miRNAs. This also caused the down-regulation of the three lncRNAs.

81

LIST OF FIGURES

	Page
Figure 2.1 The location of nasopharynx which is located behind of the nose, in the upper part of pharynx. Adapted from https://www.nhs.uk/conditions/nasopharyngeal-cancer (The National Health Center, 2018).	5
Figure 2.2 The location of the fossa of Rosenmüller which is the origin of NPC. Adopted from https://www.entspecialist.sg/doctors-corner/a-new-screening-tool-for-nose-cancer/ (The ENT Clinic, n.d).	6
Figure 2.3 The structure of enveloped Epstein-Barr virus. The double stranded DNA is encapsulated by nucleuocapsid and the enveloped is surrounded encapsulated by spiked glycoprotein. Retrieved from https://commons.wikimedia.org/wiki/File:Viral_Tegume (Taylor, 2017).	13
Figure 2.4 The pri-miRNAs of the BHRF1 and BART clusters. Adopted from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7062708/ (Lizasa et al., 2020).	21
Figure 2.5 The effect of dysregulated lncRNA (miRNA sponge). The up-regulation of lncRNA (miRNA sponge) causes dysregulation of miRNAs and mRNAs where miRNAs are sequestered by lncRNAs leading to up-regulation of mRNAs. Retrieved from:	

<https://hereditasjournal.biomedcentral.com/articles/10.1186/s41065-021-00208-7> (Shi et al., 2021).

24

Figure 4.1 The volcano plots of all the five GEO RNAs datasets. The red dots represent the significantly up-regulated and down-regulated RNAs whereas the gray dots represent the not significantly deregulated RNAs. The RNAs on the left side are the down-regulated ones and the up-regulated ones are on the right side. The Y axis presents the negative logarithm of the p-value in base 10 and the X axis is the logarithm of the fold change between the up-regulated and down-regulated expression levels.

56

Figure 4.2 The Venn diagram of up-regulated and down-regulated mRNAs between the three mRNA datasets (GSE64634, GSE126683 and GSE61218). The Venn diagrams showed 65 common up-regulated and 123 common down-regulated mRNAs between the three datasets.

57

Figure 4.3 The Venn diagram of up-regulated and down-regulated lncRNAs between the three lncRNA datasets (GSE95166, GSE126683 and GSE61218). The Venn diagrams showed 0 common up-regulated lncRNAs and three common down-regulated lncRNAs between the three datasets.

59

Figure 4.4 The lncRNA-miRNA-mRNA competing endogenous RNA (ceRNA) network. The ceRNA network consisted of 58 down-regulated mRNAs (red eclipse), 17 up-regulated miRNAs (yellow rectangle),

and three down-regulated lncRNAs (purple diamond). It showed the interactions between the RNAs and the deregulation in one of the RNAs affecting the regulation of other RNAs interacting with it. 68

Figure 4.5 The Venn diagram of mRNAs between EBV miRNAs mRNAs with ceRNA DEGs. It exhibited four common mRNAs between the mRNAs targeted by BART miRNAs with the DEGs in the ceRNA, which were EBF1, MACROD2, EYA1 and EYA4. 69

Figure 4.6 The interactions of ceRNA mRNAs with EBV BART miRNAs. The four mRNAs (EBF1, MACROD2, EYA1 and EYA4) which were indicated by the red eclipse revealed to be targeted and down-regulated by six BART miRNAs (ebv-miR-BART19-3p, ebv-miR-BART15, ebv-miR-BART21-3p, ebv-miR-BART12-5p, ebv-miR-BART20-3p and ebv-miR-BART11-5p) which were indicated by green V. 70

Figure 4.7 The protein-protein interaction (PPI) between the down-regulated common DEGs from the three GEO datasets (GSE64634, GSE126683 and GSE61218). constructed using the Cytoscape plugin. The PPI exhibited 123 nodes and 179 edges, with 77 of the DEGs interacting with each other, represented by the red ellipse. 72

Figure 4.8 The top 20 DEGs with connectivity degree more than 5 from the PPI ranked according to connectivity degree by CytoHubba. DNAH5 was ranked 1st indicated by the red colour, and DNAAF1 was ranked last indicated by the yellow colour. The colour gradient with red

indicated the connectivity degree ranks represented DEGs with high connectivity degree, and the colour scheme changes to orange and yellow as the connectivity degree decreases. The yellow represented the lowest connectivity degree.

73

Figure 4.9 The DEGs with connectivity degree more than 5 amongst the Top 20 DEGs. It showed 18 DEGs of the top 20 DEGs were interacting with each other. These DEGs are deemed as hub genes due to their ability to regulate multiple genes, and their expressions are affecting the expressions of those DEGs interacting with them.

75

Figure 4.10 The two modules in the PPI network with scores greater than or equal to 4 and nodes greater than or equal to 4. In Module 1, all the DEGs are hub genes with connectivity degree more than 5, whereas, in Module 2, all the DEGs are not hub genes. Both modules are essential in proper cilia functions and motility.

77

Figure 4.11 The Venn diagram of transcription factors (TFs) from the TFcheckpoint database overlapped with common down-regulated mRNAs from the three GEO datasets (GSE64634, GSE126683 and GSE61218). Two TFs were found amongst the 123 GEO down-regulated mRNAs, which are EBF1 and NR2F2.

78

Figure 4.12 The TFs, EBF1 and NR2F2 in the PPI network. Both TFs indicated by the green eclipse do not interact with other down-regulated mRNAs indicated by the red eclipse and they were also not interacting with each other.

79

Figure 4.13 The cross-regulatory network. It was merged from competing endogenous RNA network (ceRNA) and EBV miRNAs-mRNAs network. Only BART11-5p was targeting and down-regulating TF, which was EBF1 (purple eclipse), and none of the hub genes (green eclipse) was targeted by BART miRNAs.

81

Figure 4.14 The Gene Ontology (GO) enrichment result for EYA4, EYA1, EBF1 and MACROD2 by using DAVID. The count means the total number of DEGs involved in each of the GO terms. Only EYA4 and EYA1 were involved in all the Biological Processes (BP). For Molecular Functions (MF), only EYA4 and EYA1 were involved in protein tyrosine phosphatase activity, whereas in metal ion binding EYA4, EYA1 and EBF1 were involved. All the four DEGs were located in nucleus as indicated by the Cellular Component (CC).

83

Figure 4.15 The Kaplan Meier survival curve for EYA1, EYA4, EBF1 and MACROD2. The Y axis represented the probability of surviving, and the X axis represented the duration of the study or survival.

85

LIST OF ABBREVIATIONS

2D	Two-dimensional
3D	Three-dimensional
$\alpha\beta$	Alpha v beta
ADP-ribosylation	Adenine dinucleotide phosphate-ribosylation
AGO2	Human Argonaute 2
AKR1B1	Aldo-keto reductase family 1 member B1
ARHGEF26-AS1	Rho guanine Nucleotide exchange factor 26 antisense RNA 1
ASOs	Antisense oligonucleotides
ATM	Ataxia-telangiectasia-mutated
BART	<i>Bam</i> HI fragment A rightward
BC	Breast cancer
BHRF1	<i>Bam</i> HI fragment H rightward open reading frame 1
BL	Burkitt lymphoma
BP	Biological Processes
CASC2	Cancer susceptibility candidate 2
CC	Cellular Component
ceRNA	Competing endogenous RNA
ceRNET	ceRNA network
CCDC113	Coiled-Coil Domain Containing 113
CCRT	Chemoradiotherapy
circRNA	circular RNA
CLASH-seq	Cross-linking, ligation, and sequencing of hybrids

CLASH-seq	Cross-linking, ligation, and sequencing of hybrids
CLEAR-CLIP	Covalent ligation and endogenous Argonaute-bound RNA
CLIP-seq	Cross-linking and immunoprecipitation and high-throughput sequencing
CMTM3	CKLF-like MARVEL transmembrane domain containing 3
CT	Computed tomography
CR2	Complement receptor type 2
CRC	Colorectal cancer
CTL	Cytotoxic T-lymphocytes
cvhRNAs	Competitive viral and host RNAs
DAVID	Database of Annotation, Visualization and Discovery software
DBD	DNA-binding domain
DEG	Differentially expressed gene
DEL	Differentially expressed lncRNA
DEmiRNA	Differentially expressed miRNA
DDR	DNA damage response
DICE1	Deleted in cancer 1
DLBCL	Diffuse large B-cell lymphoma
DNA	Deoxyribonucleic acid
DNAH	Dynein axonemal heavy chain
DNAI	Dynein Axonemal Intermediate Chain
DNALI1	Dynein light intermediate polypeptide 1
DSB	Double-stranded break
DYNLRB2	Dynein light chain roadblock-type 2
DYNC2H1	Cytoplasmic Dynein 2 Heavy Chain

E2F3	E2F transcription factor 3
EA	Early antigen
EAd/BMRF1	Early antigen diffuse
EBNA1	Epstein-Barr nuclear antigen-1
EBV	Epstein-Barr virus
EBVaGC	EBV-associated gastric cancer
EGA	European Genome-phenome Archive
ENKUR	Enkurin
EBERs	Epstein-Barr small encoded RNAs
EBF1	Early B-cell factor transcription factor 1
EGFR	Epidermal growth factor receptor
ESCC	Esophageal squamous cell carcinoma
EYA	Coactivator and phosphatase
FANK1	Fibronectin type 3 and ankyrin repeat domains 1
GC	Gastric cancer
GENCODE	Genome ENCyclopedia Of DNA Elements
GEO	Gene Expression Omnibus
GEO2R	Gene Expression Omnibus to R language
gB	Glycoprotein B
gH	Glycoprotein H
gL	Glycoprotein L
gp42	Glycoprotein 42
GC	Gastric cancer
GO	Gene Ontology
GWAS	Genome-wide association studies

HITS-CLIP	High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation
HLA	Human leukocyte antigen
HR	Hazard ratio
HCC	Hepatocellular carcinoma
IFN1	I interferon
IM	Infectious mononucleosis
IMRT	Intensity-modulated radiotherapy
lncRNAs	Long non-coding RNAs
KEGG	Kyoto Encyclopedia of Genes and Genomes
K-M	Kaplan-Meier
LCL	Lymphoblastoid cell line
LMPs	Latent membrane proteins
LRRC23	Leucine rich repeat containing 23
LSCC	Laryngeal squamous cell carcinoma
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
MACROD2	Mono-ADP-Ribosylhydrolase 2
MCODE	Molecular Complex Detection
MDC1	Mediator Of DNA Damage Checkpoint 1
MF	Molecular Function
MIB	Metal ion binding
MICA	Major histocompatibility complex class I chain-related peptide A
MICB	Major histocompatibility complex class I chain-related peptide B
miRNAs	MicroRNAs

miRTarBase	MicroRNA target interaction database
MREs	MiRNA recognition elements
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
MTI	MiRNA-target interaction
NCBI	National Center for Biotechnology Information
ncRNAs	Non-coding RNAs
NF1	Neurofibromin 1
NK	Natural killer
NF-KB	Nuclear factor KB
NGS	Next-generation sequencing
NKG2D	Natural killer group 2, member D
NLP	Natural language processing
NSCLC	Non-small cell lung cancer
NPC	Nasopharyngeal cancer
OME	Otitis media with effusion
OS	Overall survival
oriP	Plasmid origin of replication
PAR-CLIP	Photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation
PARP1	Poly(ADP-ribose) polymerases 1
PNO1	Partner Of Nin one binding
PPI	Protein-protein interaction
PCD	Primary ciliary dyskinesia
PTP	Protein tyrosine phosphatase

PUMA	p53 up-regulated modulator of apoptosis
qPCR	Real time PCR
RT	Radiotherapy
RNAs	Ribonucleic acids
RNAGAPLINC	Gastric adenocarcinoma predictive long intergenic noncoding
RNAi	RNA-mediated interference
RNAP II	RNA polymerase II
SCC	Squamous cell carcinoma
snoRNA	Nucleolar RNA
StarBase	sRNA target base
STRING	The Search Tool for Retrieval of Interacting Genes/Proteins
TCGA	The Cancer Genome Atlas
TFs	Transcription factors
THBS1	Thrombospondin 1
RNAGAPLINC	Gastric adenocarcinoma predictive long intergenic noncoding
RNAi	RNA-mediated interference
TNF- α	Tumor necrosis factor alpha
TNFRSF	Tumor Necrosis Factor Receptor Superfamily
TPTE2P1	Transmembrane phosphoinositide 3-phosphatase and tensin homolog 2 pseudogene 1
UTR	Untranslated region
VCA	Viral capsid antigen
VM	Vasculogenic mimicry
WDR	WD repeat-containing protein
WHO	World Health Organization

X^2

Chi-square

CHAPTER 1

INTRODUCTION

1.1 Study Background

Nasopharyngeal cancer (NPC) is one of the rare malignancies in most parts of the world, which highly causes death in several regions in southern China and has medium to high prevalence in other Southeast Asia countries, including Malaysia (Adham et al., 2012). Its pathogenesis is associated with Epstein-Barr virus (EBV) infection and RNA dysregulation. This virus formed the two NPC subtypes, which are type II (differentiated non-keratinizing carcinoma) and type III (undifferentiated non-keratinizing carcinoma) NPC (Su et al., 2023). Besides, it can regulate its host mRNAs by utilizing its encoded *BamHI* A rightward transcript (BARTs) microRNAs (miRNAs) by binding to the 3' untranslated region (UTR) of host mRNAs (Skalsky & Cullen, 2010). This affects the RNA expression levels in mRNA-miRNA-lncRNA competing endogenous RNA (ceRNA) activity. In recent years, increasing data have demonstrated that ceRNA dysregulation causes tumorigenesis (Lee & Young, 2013). This is due to both mRNAs and lncRNAs competing for the same miRNAs following the ceRNA hypothesis that stated lncRNA and mRNAs regulate each other by competing for binding sites known as miRNA recognition elements (MREs) on shared miRNAs through partially complementary sequences (Salmena et al., 2011).

Nevertheless, upon EBV infection, these lncRNAs are indirectly regulated by BART miRNAs and shift the ceRNA activity balance. As described by He et al. (2016), EBV-miR-BART6-3p down-regulates lncRNA LOC553103 in gastric cancer (GC) and NPC cell lines, where Wang et al. (2020), further verified this mechanism and added that

BART6-3p also dysregulates the mRNA STMN1. There are two miRNA clusters encoded by EBV, which are *Bam*HI A rightward (BARTs) and *Bam*HI fragment H rightward open reading frame 1 (BHRF1) clusters (Cai et al., 2006). In NPC, latency pattern type II is expressed with the presence of BART cluster and the absence of *Bam*HI fragment H rightward open reading frame 1 (BHRF1) cluster (Kang et al., 2015). Cosmopoulos et al. (2009) demonstrated that the BHRF1 miRNAs are absent in NPC tumor samples and are present at a very low level in the C666-1 NPC cell line. However, Cai et al. (2006) reported that BHRF1 miRNAs are absent in C666-1 and C15 NPC cell lines. For BART miRNAs, previous studies exhibited its overexpression in NPC tissue samples and cell lines (Zhang et al., 2022). These miRNAs in NPC often act similarly to oncogenes by down-regulating apoptotic host mRNAs through binding to the 3'UTR (Kang et al., 2015). In a previous study, Jing et al. predicted BART1-3p and BART22 targeting C-X-C motif chemokine ligand 10 (CXCL10) and Growth/differentiation factor 5 (GDF5) causes the up-regulation of miR-6877-3p and down-regulation of lncRNAs RP5-1039K5.19 and TP73-AS1 in gastric cancer (GC) through bioinformatics analysis by constructing the cross regulatory network.

1.2 Problem Statement

The regulatory mechanism of BART miRNAs on ceRNA activity in EBV-associated NPC is unavailable despite their abundant expression in EBV-positive NPC. Through this regulatory mechanism, potential biomarkers, including BART miRNAs, mRNAs and lncRNAs, could be identified, and this can serve as preliminary data. The biomarkers are useful since an early diagnosis of NPC is hard due to the deep location of the nasopharynx and lack of obvious clinical symptoms at an early stage, prompting the high mortality rate contributed by its highly invasive and metastatic nature (Tabuchi et al.,

2011; Wang et al., 2017; Dwijayanti et al., 2020; Tan et al., 2022). The biomarkers are also crucial for early diagnosis, better prognosis and survival rate in NPC patients, where eventually, they are valuable for developing therapeutic targets. In a previous study, Xu et al. (2020) predicted that 2,654 mRNAs, 132 lncRNAs, and 565 miRNAs are deregulating each other in NPC by affecting tumor-related pathways through bioinformatics analysis by constructing a ceRNA network (ceRNET). Nonetheless, their study excluded the BART miRNAs regulation on ceRNA activity.

1.3 Objectives and Hypothesis

Given the importance of ceRNET and EBV miRNA-mRNA network, this study aimed to give insight into understanding BART EBV regulatory mechanisms on ceRNA activity and the biological and molecular processes affected by it in NPC by using bioinformatics analysis through the construction of cross-regulatory network which is a combination of ceRNET and EBV miRNA-mRNA network. From this network, mRNAs and lncRNAs, directly and indirectly, targeted by BART miRNAs, could be revealed. These RNAs could be utilized as potential early biomarkers in EBV-positive NPC, and their functions as either an oncogene or tumor suppressor in NPC development and progression could be deduced. Lastly, the possibility of NPC patients' prognosis could be determined by using the deregulated mRNAs. We hypothesised that (1) BART miRNAs dysregulate mRNAs causing deregulation of miRNAs and lncRNAs, (2) BART miRNAs might target and regulate mRNAs which are responsible for cell apoptosis and proliferation, (3) BART miRNAs and host lncRNAs, could act similar to oncogene and tumor suppressor, (4) the dysregulated mRNAs resulted in poor overall survival (OS) in patients. The objectives of this study were:

- i. To predict and illustrate mRNAs, miRNAs, lncRNAs and EBV BART miRNAs regulation mechanism in EBV-positive NPC pathogenesis.
- ii. To identify the biological and molecular processes and pathways affected by BART miRNAs in NPC.
- iii. To predict the prognosis of dysregulated mRNAs targeted by BART miRNAs in NPC.

CHAPTER 2

LITERATURE REVIEW

2.1 Nasopharyngeal Cancer

Nasopharyngeal cancer (NPC) is a rare type of head and neck cancer which first commonly forms in the squamous cells lining the nasopharynx (National Cancer Institute, 2018). The nasopharynx is located behind the nose, in the upper part of the pharynx, which is the throat (Figure 2.1). The symptoms of NPC include a lump in the neck, hearing loss, usually in one ear, blood discharge in saliva and from the nose, ringing in the ears, nasal congestion, sore throat and headaches (Mayo Clinic, 2018; The National Health Center, 2018).

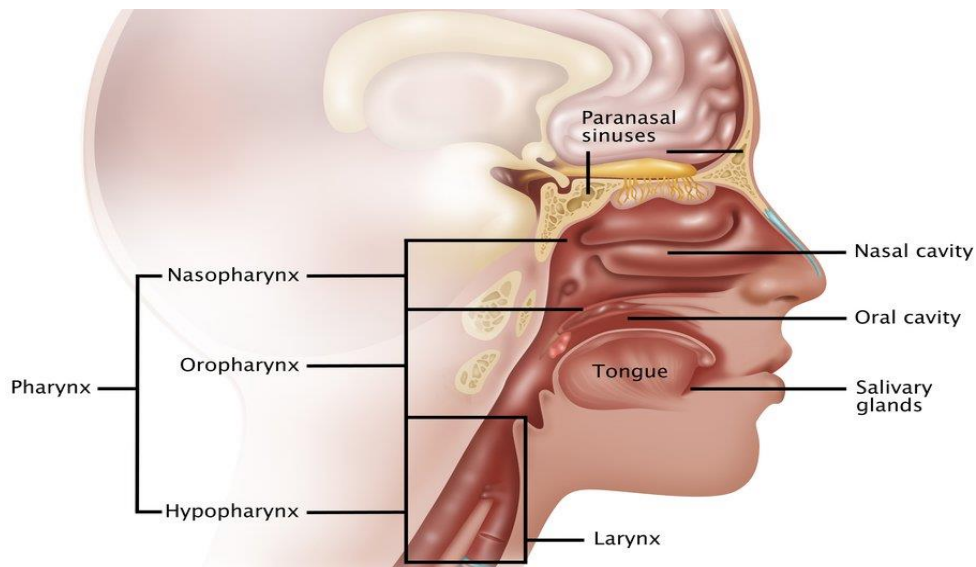


Figure 2.1: The location of nasopharynx which is located behind of the nose, in the upper part of pharynx. Adapted from <https://www.nhs.uk/conditions/nasopharyngeal-cancer> (The National Health Center, 2018).

2.1.1 Histopathology of Nasopharyngeal Cancer

Nasopharyngeal cancer (NPC) is a non-lymphomatous squamous cell carcinoma (SCC) that occurs in the epithelial lining of the nasopharynx and commonly originates from the fossa of Rosenmüller (Tabuchi et al., 2011) (Figure 2.2). The SCC means the majority of the tumor shows obvious squamous differentiation under the light microscope (Thompson, 2007). The histopathological classification of NPC is divided into two main categories according to the World Health Organization (WHO) classification. These are keratinizing squamous cell carcinoma (Type I) and non-keratinizing squamous cell carcinoma (type II and III) (Young & Dawson, 2014). The non-keratinizing type is further subdivided into differentiated non-keratinizing carcinoma (type II) and undifferentiated non-keratinizing carcinoma (type III) (Young & Dawson, 2014).

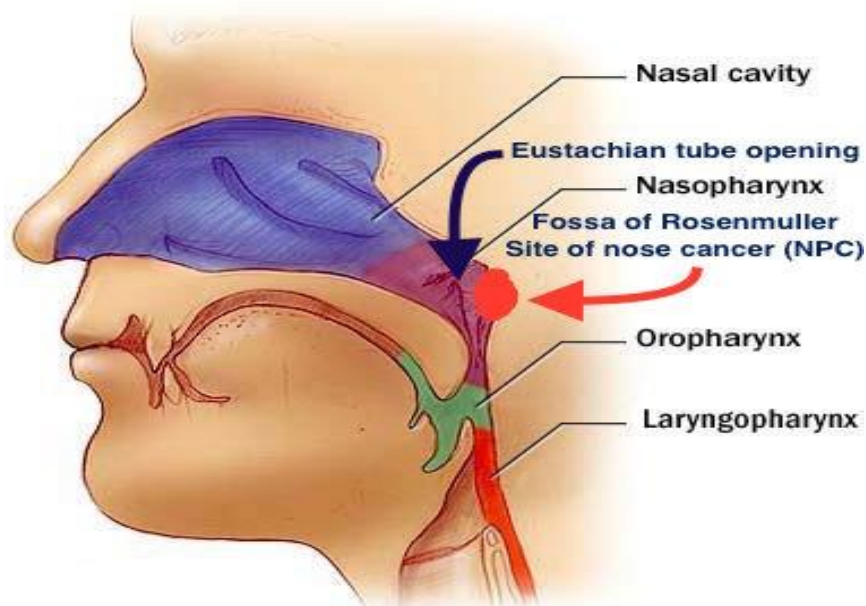


Figure 2.2: The location of the fossa of Rosenmüller which is the origin of NPC.
Adopted from <https://www.entspecialist.sg/doctors-corner/a-new-screening-tool-for-nose-cancer/> (The ENT Clinic, n.d).

Type I NPC is observed in 25% of NPC cases where the tumor cells produce keratin with the presence of intercellular bridges (Gullo et al., 2008). In type II NPC, it is observed in ~20% of cases where the tumor cells exhibit various morphologies with well-defined borders and vague intercellular bridges (Gullo et al., 2008; Chan, 2017). Lastly, type III NPC is observed in 55% of NPC cases with no uniform morphology characteristic with indistinct intercellular bridges and is commonly found with EBV genomes (Gullo et al., 2008). The tumor cells are also surrounded by lymphocytes and plasma cells. The NPC category, which is predominantly EBV positive, is the non-keratinizing squamous cell carcinoma (type II and III) where EBV-positive NPC is found 100% in undifferentiated non-keratinizing carcinoma (type III) compared to differentiated non-keratinizing carcinoma (type II) (Tsang et al., 2014; Young & Dawson, 2014).

2.1.2 Epidemiology of Nasopharyngeal Cancer

The tumorigenesis of NPC is rare among most populations but commonly occurs among natives in southern China, Southeast Asia, the Arctic, and the Middle East/North Africa (Chang & Adami, 2006). From this, it indicates that both environmental factors and genetic traits contribute to the development of NPC. Type I NPC is commonly observed in Caucasians in Westerns, whereas Type II and III are commonly observed in Asians (Wang et al., 2013). The type II and III are commonly observed in Chinese residing in Hong Kong, Taiwan, Macao, Southern China, and Singapore (Wang et al., 2013; Mak et al., 2015). Regardless, other races in Southeast Asia countries, including Singapore and Malaysia, have also been reported to have type II and III NPC.

In Southeast Asia, NPC is ranked the eleventh most common cancer and the 5th most common cancer in Malaysia (Global Cancer Observatory, 2020). Interestingly, NPC

is ranked the fourth most common cancer among males in Malaysia (Global Cancer Observatory, 2020). In Peninsular Malaysia, Prasad and Rampal (1992) revealed that out of 315 NPC patients, 86.3% were of Chinese ethnicity, 12.9% were of Malay ethnic and 0.8% were of Indian ethnicity. In Sarawak, however, Devi et al. (2004) revealed that Sarawak Bidayuh people have the highest rate of NPC among other native people and compared to Chinese people. The patients of NPC were reported to be residents from Kuching and Serian, with Kuching residents showing the highest reports (Tiong & Selva, 2005). A different trend was also observed in Sabah, as stated by Wong et al. (2021); out of 215 cases, 29.8% were of Dusun ethnic, 16.7% were of Chinese ethnic, 9.3% were of Kadazan, 8.8% were of Bajau ethnic.

The occurrence of NPC is also unspecific to any age group and seems to occur more in men, with approximately two to three times as prevalent in men than women (Chang et al., 2021). In low-risk populations commonly with Type 1 NPC, a moderate peak of incidence was observed in young adulthood of ~ ages 15-24 years, with a slight decline in adults of ages 35-39 years and a higher peak of incidence in ~ ages 65-79 years (Chang et al., 2021). On the contrary, a single high peak of incidence was observed in ~ ages 45-59 years in high-risk populations, commonly with Type II and III NPC (Chang et al., 2021).

2.1.3 Etiology of Nasopharyngeal Cancer

The environmental factors contributed to NPC, including consuming salt-preserved fish and other preserved foods, including meats, eggs, fruits and vegetables. These foods are known to be a dietary staple in all populations endemic to NPC. This was stated by Zheng et al. (1994), where NPC was frequently observed in individuals who frequently

consumed salt-preserved meats and fish compared to control individuals. Preserved foods with salt containing nitrites and nitrates cause the formation and accumulation of high nitrosamines, known carcinogens to animals and humans (IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, 2012; Xie et al., 2023). The nitrosamines formed due to exposure to nitrogen oxide in the air during preservation and contact with acidic conditions in the stomach after ingestion, a process called nitrosation (Ward et al., 2000; IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, 2012).

Besides, nitrosamines could induce DNA methylation, which activates proto-oncogenes such as ras or myc and inactive tumor suppressor, including p53 or cyclin-dependent kinase inhibitor 2A (Belinsky et al., 1990). The low consumption of vegetables and fruits containing vitamin C also leads to NPC, as vitamin C can hinder the formation of nitrosamines (Chang & Adami, 2006). This was also stated by Zheng et al. (1994), where the controls in a case-control study revealed they consumed fresh fruit frequently compared to NPC patients despite consuming salt-preserved fish and meat. Another culprit that can induce NPC is the smoke of tobacco and alcohol consumption, as reported by most studies, especially in Type I NPC (Ou et al., 2007; Adham et al., 2012).

Moreover, exposure to Epstein-Barr virus (EBV) also leads to NPC, as most NPC cases are linked to positive expression of multiple viral genes in the tumor cells (Adham et al., 2012). The accumulations of genetic and environmental factors and EBV are believed to contribute to NPC, as EBV infection alone is insufficient to cause NPC. Contributed to the consumption of salt-preserved food, the high concentration of salt was observed to cause EBV lytic reactivation in human lymphoblastoid Raji cells, according to Shao et al. (1988). The expressed EBV miRNAs in cells lead to NPC, especially miRNAs from the

*Bam*H1 A rightward transcript (BART) region of the EBV genome by altering the expression of host's RNAs to facilitate EBV latency, promote cancer cell proliferation and invasion, induce genome instability, prevent apoptosis and impaired immune response (Hau et al., 2020).

For the genetic factors, the polymorphic nature of human leukocyte antigen (HLA) alleles contributed to a high chance of developing NPC in those who inherited specific HLA alleles (Ning et al., 2020). These HLA alleles possessed a reduced ability to present Epstein-Barr virus (EBV) antigens to the immune system by preventing the detection of EBV by cytotoxic T-cells thus, no immune response is elicited on infected cells and causes NPC (Tsao et al., 2014). These HLA alleles are mostly from the Class 1 HLA alleles, including HLA-A, -B and -C (Ning et al., 2020). Furthermore, recent studies suggested the involvement of deregulation between messenger RNAs (mRNAs), microRNAs (miRNAs) and long non-coding RNAs (lncRNAs) expressions level as competing endogenous RNA (ceRNA) in the development and progression of NPC (Xu et al., 2020). This ceRNA interaction follows the hypothesis that states lncRNA and mRNAs regulate each other by competing for binding site on shared miRNAs through partially complementary sequences known as miRNA recognition elements (MREs) (Salmena et al., 2011).

2.1.4 Diagnosis, Prognosis and Treatment of Nasopharyngeal Cancer

Diagnosis for nasopharyngeal cancer (NPC) is often done at mid or late stages due to the deep location of the nasopharynx, and the onset of symptoms occurs after it affects surrounding organs (Wang et al., 2017). Until now, the gold standard diagnostic methods for NPC are nasopharyngoscopy, endoscopic biopsy and magnetic resonance imaging (MRI) (Tang et al., 2021). The MRI is currently used to replace the computed tomography

(CT) method for staging and treatment planning. However, CT can be combined with MRI for accurate diagnosis due to its ability to recognize lymph node metastasis (Tang et al., 2021). Initially, nasopharyngoscopy is used for diagnosis, and if a tumor is suspected in the Rosenmüller's fossa upon nasopharyngoscopy, an endoscopic biopsy will be performed for cell cytology analysis (Bossi et al., 2021; Tang et al., 2021). Nasopharyngoscopy is inserting a thin tube through the mouth or nose to examine the areas of the head and neck. If no initial tumor is visible upon nasopharyngoscopy, MRI will be performed, followed by an endoscopic biopsy if a tumor is detected in the Rosenmüller's fossa (Bossi et al., 2021). Therefore, MRI has been deemed the most sensitive diagnostic tool for NPC detection at an early stage when a tumor is unable to be visualized with an endoscopic examination (King et al., 2019). MRI also can map the sites and boundaries of nodal metastasis useful for staging and treatment planning (King, 2022).

For the endemic area of NPC, currently, real-time fluorescent quantitative polymerase chain reaction (PCR) is used to detect and quantify EBV DNA copy numbers targeting the *Bam*HI-W region of the genome from plasma or serum samples (Tang et al., 2021). This technique possessed a 97.1% sensitivity and 98.6% specificity in detecting the presence of EBV compared to previously utilized serological screening, including identifying Immunoglobulin A (IgA) antibody titers to EBV viral capsid antigen (VCA), EBV early antigen (EA) and Epstein-Barr nuclear antigen-1 (EBNA1) IgA using immunofluorescent assays (Tang et al., 2021). Biomarkers such as EBV miRNAs are also used to rule out EBV-related NPC tumorigenesis and patients' prognosis, especially *Bam*HI A rightward transcript (BART) cluster miRNAs because they are found to be abundantly expressed in NPC (Abusalah et al., 2022). Besides, RNA biomarkers, including mRNAs, miRNAs and lncRNAs, are also used as NPC diagnostic methods by identifying

and quantifying RNA expressions in NPC tumor biopsy and cell lines, known as transcriptomic analysis (Zhang et al., 2021). Early detection is possible with biomarkers if incorporated into routine screening in highly susceptible individuals, which is crucial for a better prognosis. RNA biomarkers are desired more than previously used protein biomarkers due to high sensitivity, specificity, and cost efficiency without requiring specific antibodies (Xi et al., 2017).

Due to the location of the NPC tumor, the surgical method is not recommended; thus, the main method for NPC treatment is radiotherapy (RT) for the early stages of NPC, and recently intensity-modulated radiotherapy (IMRT) is opted more than the traditional two-dimensional (2D) or three-dimensional (3D) radiotherapy (Tang et al., 2021). This is due to the improved patient survival rate when using IMRT; its dose depends on the tumor shape and it does not affect patients' adjacent tissue (Tang et al., 2021). Chemotherapy is also used in treating NPC, either alone or combined with radiotherapy, for advanced stages of NPC. This is known as chemoradiotherapy (CCRT) and is linked with improved patients' overall survival (OS) (Tang et al., 2021). For recurrent and metastatic NPC, chemotherapy of platinum-containing two or three drug regimens is administrated to patients (Tang et al., 2021). The drugs include cisplatin, gemcitabine, camrelizumab and toripalimab. However, these treatments manifested undesirable side effects; thus, nucleic acid therapeutics based on gene silencing, including RNA-mediated interference (RNAi) and antisense oligonucleotides (ASOs), are currently in research, especially *in vivo* to replace them (Arun et al., 2018). This employs the identification of RNA biomarkers for NPC in the recent years.

2.2 Epstein-Barr Virus

Epstein-Barr virus (EBV) is a linear double-stranded DNA virus, classified in the family *Herpesviridae* and the subfamily *Gammaherpesviridae* (Sarwari et al., 2016). The life cycle of EBV is classified into two modes, which are lytic and latent (Murata et al., 2014). It is a toroid-enveloped virus, where the viral DNA is wrapped inside of an icosahedral nucleocapsid, and the space between the nucleocapsid and the outer envelope is filled with proteins, and the envelope is spiked glycoprotein (Figure 2.3) (Smatti et al., 2018). The discovery of EBV was made by Anthony Epstein and Yvonne Barr in Burkitt lymphoma (BL) cell lines derived from BL tumor cells of African children in 1964 given by Dennis Burkett, where they discovered virus particles similar to herpesvirus in the BL cell lines (Pagano, 2007; Young et al., 2016). The virus was then deemed as a lymphotropic virus and the first virus discovered as a human tumor virus onwards (Pagano, 2007; Jha et al., 2016).

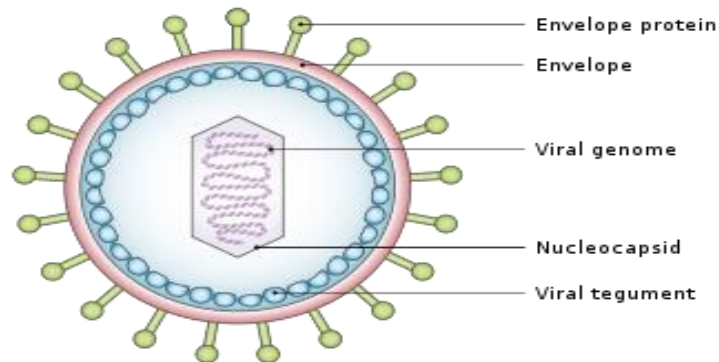


Figure 2.3: The structure of enveloped Epstein-Barr virus. The double stranded DNA is encapsulated by nucleocapsid and the envelope is surrounded encapsulated by spiked glycoprotein. Retrieved from https://commons.wikimedia.org/wiki/File:Viral_Tegume (Taylor, 2017).

2.2.1 Epidemiology of Epstein-Barr Virus

Epstein-Barr virus (EBV) is known to infect 90% of the population worldwide in the modern world, and it is mostly asymptomatic due to the persistent nature of the virus, which eventually leads to benign (Young et al., 2016). This is due to its transmission via saliva through coughing, sharing of foods, and kissing (Fugl & Andersen, 2019). The infection in children occurs from the transmission from carrier family members or caregivers through pre-masticated foods and the cleaning of the face and mouth with saliva (Dunmire et al., 2018). The infection of EBV in communities is geographic, age and socioeconomic dependent such as overcrowded living conditions with poor sanitation. This is shown through infection at an early age occurs mostly in children in underdeveloped countries such as Africa and New Guinea, whereas in Western countries, infection often occurs during childhood to early adulthood (Young, 2008). In Africa, EBV-DNA was detected in at least one-quarter of infants within the first 26 weeks, and some of the infants had EBV DNA before week 14 due to less antibodies transferred from their human immunodeficiency virus (HIV)-positive maternal through breastfeeding in a cohort study conducted by Montoya-Ferrer et al. (2021).

Early-age infection was also common in children in developing countries, including China, as reported by Shi et al. (2022), where EBV-DNA was detected mostly in the plasma of children ages 2-4 years. Interestingly, in a study conducted from the year 2018-2022 by Ye et al. (2023), EBV-DNA was detected mostly in the plasma of children ages more than 10 years, and the highest level of EBV loads was observed in children with infectious mononucleosis (IM). Compared to underdeveloped countries, both studies reported that infants under 1-year-old were detected with the lowest EBV-DNA. In a developed country such as the United Kingdom, Kuri et al. (2020) stated that EBV

seropositive increases with age, with a spike to more than 90% in the 15-19 years old age group and more than 95% EBV seroprevalence in the age group of 21–25 years old out of 1,982 individuals. Interestingly, Dowd et al. (2013) reported a different EBV infection trend in Mexican-American households where children aged 6-9 years old were reported to showed the highest EBV seroprevalence compared to non-Hispanic blacks and whites, which indicated EBV infection was associated with socioeconomic status.

2.2.2 Etiology of Epstein-Barr Virus

Epstein-Barr virus (EBV) is persistent in nature, which eventually leads to benign due to its ability to infect and colonize resting B-cells and transform the B-cells into latent lymphoblastoid cell lines (LCL), resulting in latent infection with occasional reactivation and shedding of virus progeny to infect more cells (Young, 2008; Murata et al., 2021). The ability to infect B-cells is due to the glycoproteins spikes on the envelope enabling the attachment and entry into the host's B-cell through the complement receptor type 2 (CR2) and epithelial cell (Nemerow et al., 1985; Hoover & Higginbotham, 2023). These LCLs bear multiple copies of circular viral DNA (episomes), which can transcribe latent proteins, including the six EBV-encoded nuclear antigens (EBNAs 1, 2, 3A, 3B, 3C, and -LP) and three latent membrane proteins (LMPs 1, 2A, and 2B) (Young & Dawson, 2014). In addition, the LCLs also bare the Epstein-Barr small encoded RNAs (EBERs) and two microRNAs (miRNAs), which are *Bam*HI A rightward transcripts (BARTs) and *Bam*HI H rightward reading frame 1 (BHRF1), which combined with both EBNAs and LMPs proteins to manifest oncogenic properties (Young & Dawson, 2014). The functions of each protein and miRNAs are summarized in Table 2.1.

Table 2.1: The functions of each EBV protein, miRNAs and snRNAs transcribed during both latent and infection of EBV in host cells (Farell. 2019)

Genes	Functions
EBNA1	● Binds to the plasmid origin of replication (oriP) of EBV and recruits co-factors for host cells to replicate hence the viral genome replication during the S phase of the cell cycle. ● Activates LMP1.
EBNA3A and EBNA3C	● Repress the expression of cyclin-dependent kinase inhibitor p16 and apoptotic protein BIM. ● Prevent differentiation into plasma cells during the proliferative lymphoblast stage.
EBNA3B	● Acts as a tumor suppressor in the mouse diffuse large B-cell lymphoma (DLBCL) model.
EBNA-LP	● Cooperates with EBNA2 in preventing host cell innate response to viral genome.
LMP1	● Analogous to CD40 ● Activate nuclear factor KB (NF-KB) transcription factor to induces BCL2 to prevent apoptosis.
LMP2A	● Influence signalling from B-cell receptor and induce PI3K to activate ATK pathway hence preventing apoptosis.
EBER1 and EBER2	● Regulate gene expression and prevent adaptive immune system response.
BART miRNAs	● Dysregulate tumor suppressor genes of host cells. ● Mediates evasion of T and natural killer (NK) cells during the infection of B-cells in peripheral blood lymphocytes.
BHRF1 miRNAs	● Regulates PTEN, p27 and EBV BHRF1 for efficient B-cells transformation during infection.
BHLF1 or BALF1	● The BCL-2 family proteins which prevent apoptosis.

2.2.3 Lytic and Latent Cycle of Epstein-Barr Virus

The lytic cycle of Epstein-Barr virus (EBV) emerges upon primary infection in the oral cavity, where EBV infects and shuttles between epithelial cells and the resting B-cells in the pharyngeal epithelial cells, including nasopharynx which leads to the viral genome and progeny replication (Frappie, 2012; Tsao et al., 2017). This is to ensure its persistence and spread throughout the body via the bloodstream, and its infection of B-cells leads to lifelong latent infection (Kenney, 2007; Fugl & Andersen, 2019). In the lytic cycle, EBV episome replicates via the rolling circle method to generate multiple linearized EBV genomes packed into infectious viruses and transmitted to other susceptible hosts through saliva (Tsao et al., 2017). The viral genome and progeny replication are known as the incubation period, which occurs four to seven weeks before eliciting cytotoxic T-lymphocytes (CTL) and natural killer (NK) cells reaction due to an increase of EBV genomes and progeny (Fugl & Andersen, 2019).

After the incubation period, CTLs will remove EBV; thus, infectious mononucleosis (IM) symptoms occur (Fugl & Andersen, 2019). The exception applies to EBV that resides in B-cells, where a limited number of viral genes are encoded to avoid the recognition by CTL, resulting in lifelong latent infection in B-cells and eventually leading to oncogenesis (Fugl & Andersen, 2019). Repeated glycine and alanine residues region on EBNA1 can also inhibit proteasome degradation, hence avoiding presenting peptides to CTLs in infected cells (Farrell, 2019). However, during this lytic infection, no epithelial cell proliferation and immortalization occur (Tsao et al., 2017).

The latent infection of EBV occurs when B-cells are immortalized prior to infection and after the transformation of resting B-cells into proliferative lymphoblastoid cells (Tsao

et al., 2017). Nevertheless, in healthy individuals, continuous proliferation of the B-cells with the establishment of lifelong infection in the circulating memory B-cell is inhibited (Tsao et al., 2017). As the B-cells proliferate during the S phase, the genome of the EBV in episome form replicates once (Murata, 2014; Murata et al., 2021). During the latent cycle, six EBV nuclear antigen (EBNA) genes, including EBNA1, 2, 3A-C and LP, increase after infection, followed by a gradual increase in LMPs including LMP1, 2A and 2B aided by EBNA2 (Murata et al., 2021). The functions of these genes are described in Table 2.1. The latent infection of EBV is classified into three latency according to the gene expression pattern of the EBV latent genes, which are latency I, II and III (Farrell, 2019) (Table 2.2). The genes are expressed in small amounts to avoid the host immune system (Murata, 2014). In epithelial cells, latent infection is not uncommon, which subsequently leads to oncogenesis, including NPC (Tsang et al., 2014).

Table 2.2: The latency patterns of gene expression in EBV-infected cells during latent infection (Farrell, 2019)

Latency pattern	Genes
Latency I	EBNA1, EBER1, EBER2, and BART miRNAs
Latency II	EBNA1, LMP1, LMP2, EBER1, EBER2, and BART miRNAs
Latency III	All EBNA1s (EBNA 1, 2, 3A, 3B, 3C and -LP), LMP (LMP 1, and 2A), EBER (EBER 1, and 2) and miRNAs (BART and BHRF1)

2.2.4 Epstein-Barr Virus Induced Nasopharyngeal Cancer

Generally, infection of Epstein-Barr virus (EBV) on epithelial cells is inconvenient due to the absence of CR2 receptor on epithelial cells (Tsang et al., 2014). Due to this, latent epithelial infection is believed to be aided by the reduced level of glycoprotein 42

(gp42) protein envelop in EBV and the close proximity between B-cells and epithelial cells (Tsang et al., 2014). EBV produced from B-cells has a reduced level of gp42 due to degradation, where it has a tropism for epithelial cells by binding to the integrins alpha v beta 6 ($\alpha v\beta 6$) and alpha v beta 8 ($\alpha v\beta 8$) on the epithelial cells via the L/H (gL/gH) glycoproteins complex with glycoprotein B (gB) as co-factor (Borza & Hutt-Fletcher, 2002; Chesnokova et al., 2009). This tropism allows EBV to shuttle between epithelial cells and B-cells throughout the life of the infected host to maintain its latent infection in both epithelial and B-cells (Tsang et al., 2014).

Pre-invasive lesions, premalignant and cancerous nasopharyngeal epithelial (NP) cells and NPC-bearing genetic alterations contribute to EBV infection due to altered cellular response to growth arrest and apoptosis upon infection (Pathmanathan et al., 1995; Tsang et al., 2014). These events support latent EBV infection and promptly cause oncogenesis during the reactivation of EBV. Lymphoid infiltration, called lymphoepithelioma, also contributes to the developing epithelial cancer caused by EBV (Takada, 2000). The cell-to-cell contact between infected B-cells and epithelial cells transfers the EBV to the epithelial cells via epithelial stroma or epithelial space (Takada, 2000). It involves transferring EBV miRNAs from infected cells to recipient non-infected cells through the excretion of exosomes to avoid RNases (Pegtel et al., 2010). It commonly occurs in individuals with undifferentiated NPC due to the heavy infiltration of lymphocytes and inflammatory stroma features observed in this type of NPC.

Latency pattern II occurs in EBV-infected epithelial cells, including NPC, whereas latency pattern III occurs in B-cells. This prevents the transformation of epithelial cells into proliferative cells due to the absence of EBNA2 and EBNA3C expressions (Tsang et al.,

2014). The link between EBV and NPC was first discovered when serum antibodies against EBV antigens with viral capsid antigen (VCA) and early antigen diffuse (EAAd/BMRF1) were detected in high titres in BL and NPC cells as observed by Gunvén et al. (1970). The link was further confirmed by Wolf et al. (1975) when they discovered EBV genomes in NPC cells through *in situ* hybridisation.

2.2.5 Epstein-Barr Virus MicroRNA

There are two clusters of miRNAs encoded by Epstein-Barr virus (EBV), which are the *Bam*H1 H rightward facing 1 (BHRF1) cluster and the *Bam*H1 A rightward transcript (BART) cluster, named according to the location on the EBV genome (Polling et al., 2017). The BHRF1 cluster consists of three pre-miRNAs, and the BARTs cluster consists of 22 pre-miRNAs (Kang et al., 2015). The three BHRF1 pre-miRNA encodes four mature miRNAs, whereas the 22 BARTs pre-miRNAs encodes 40 mature miRNAs (Polling et al., 2017; Fan et al., 2018) (Figure 2.4). Hence, to date, 25 pre-miRNA of EBV with 44 mature EBV miRNAs have been discovered (Chen et al., 2010). 21 out of 22 BART miRNAs transcripts were also detected to be expressed in EBV-associated gastric cancer (EBVaGC), and NPC hence indicates the potential of these miRNAs in pathogenesis and act as a biomarker for these cancers (Kang et al., 2015). The BART cluster is present in latency patterns I-III with reduced expression in latency pattern III, whereas the BHRF1 cluster is only expressed in latency pattern III (Cai et al., 2006). The suffixes in the nomenclature of BART miRNAs indicate the order of discovery (Marquitz & Raab-Traub, 2012). The suffixes -5p and -3p in the nomenclature designate the 5' end and 3' end of the hairpin, respectively, with the symbol “*” designated to the more favoured end side (Marquitz & Raab-Traub, 2012). This is due to both strands being able to function occasionally instead of being degraded (Marquitz & Raab-Traub, 2012).

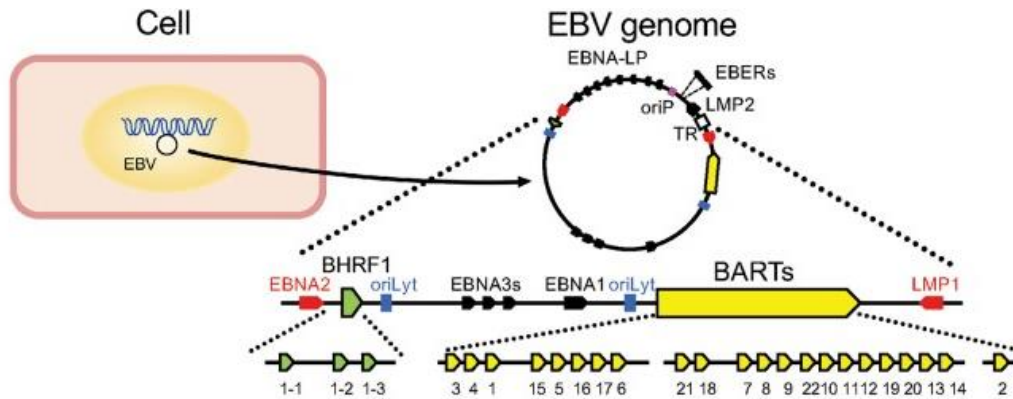


Figure 2.4: The pri-miRNAs of the BHRF1 and BART clusters. Adopted from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7062708/> (Lizasa et al., 2020).

The functions of miRNAs in viruses are to control cellular and viral gene expressions for a conducive replication environment of the virus, including regulating latent-lytic cycle switching, viral replication, promoting cell survival, proliferation and differentiation and modulating host immune responses (Skalsky & Cullen, 2010). This is enabled by the common miRNA binding sites on both viral miRNA and the host's miRNA (Li et al., 2014). Like cellular miRNAs, mature BART miRNAs can regulate both EBV and host cell's gene expressions by repressing the mRNAs protein translation (Pegtel, 2010).

2.2.6 *Bam*H1 Fragment A Rightward Transcript MicroRNA

Generally, cellular repression occurs through the imperfect complementary binding of miRNAs 5' seed region to the 3' untranslated region (UTR) of mRNAs (Skalsky & Cullen, 2010). Similarly, EBV miRNAs utilize this imperfect binding to repress host's mRNAs. This results in the deregulation or translation inhibition of host's mRNAs leading to NPC pathogenesis (Fan et al., 2018). These BARTs miRNAs could regulate RNAs involved in cell proliferation, pro-apoptosis and apoptosis, where this interaction interfered

with signalling pathways involved in apoptosis, proliferation, invasion and metastasis as shown by previous studies (Wang et al., 2019). Kang et al. (2015) revealed seven mRNAs which were CASZ1, DICE1, OCT1, CREBBP, SH2B3, PAK2 and TP53INP1, that encoded pro-apoptotic proteins, were down-regulated by 5 BART miRNAs which were BART3, 6, 8, 16 and 22 to ensure the survival of latent EBV-infected cells by inhibited apoptosis in NPC cell line. Choy et al. (2008) also revealed BART5 binds to 3'UTR of the tumor suppressor p53 up-regulated modulator of apoptosis (PUMA) and repressed its expression, hindering apoptosis in EBV-positive cells and NPC. In addition, Lei et al. (2013) also revealed BART3* was found to be highly expressed in NPC cells and bound the 3'UTR of deleted in cancer 1 (DICE1) tumor suppressor protein, causing its down-regulation and cancer cell proliferation.

Past studies also indicated that BART miRNAs dysregulation on mRNAs could dysregulate the other RNAs in a competing endogenous RNA (ceRNA) activity. As mentioned above, BART miRNAs act similarly to the host's miRNAs, thus competing with the host's miRNAs for mRNA and contributing to EBV-induced malignancies, including gastric cancer (GC) and NPC (Torres et al., 2021). This competition is deemed as competitive viral and host RNAs (cvhRNAs) where BART miRNAs mimic host miRNAs and act as sponges for host mRNA resulting in the deregulation of the ceRNA network (Kincaid & Sullivan, 2012; Li et al., 2014). This interaction gives insight into understanding the molecular mechanisms of BART miRNAs in oncogenesis. In a previous study, Jing et al. (2018) predicted BART1-3p and BART22 were targeting and dysregulating CXCL10 and SMAD5 which caused the dysregulation of miRNAs hsa-miR-4446-3p, hsa-miR-5787, hsa-miR-1915-3p, hsa-miR-335-3p and hsa-miR-6877-3p and the lncRNAs RP5-1039K5.19 and TP73-AS1. They also revealed that both CXCL10 and

SMAD5 expressions were deregulated in EBV-associated GC (EBVaGC) tumours compared to EBV-negative GC (EBVnGC). Although these cvhRNAs helps in understanding the regulatory mechanisms of EBV miRNAs on host's RNAs, the studies on it are mostly identified by using bioinformatic analysis and are only available in GC and are unavailable in NPC. Also based on these, BART miRNAs are suggested to act similarly to an oncogene in NPC tumorigenesis as they possess oncogenic properties by acting similarly to the host's miRNAs and competing with the host's miRNAs for mRNA.

2.3 Competing Endogenous RNA (ceRNA)

The competing endogenous RNA (ceRNA) are various RNA transcripts which able to regulate each other at the post-transcriptional level and rarely at the transcriptional level by competing for shared microRNAs (miRNAs) (Salmena et al., 2011; Gupta, 2014). Generally, ceRNA is essential for normal cell physiology regulation and for cells to adapt to changes in accordance to cell stimulus, cell growth and differentiation owing to miRNA's cell-specific nature and its availability dependent on tissue developmental stages (Salmena et al., 2011; Cesana & Daley, 2013). Its activity is influenced by the abundance of the RNAs where the aberrant expressions in one of the RNAs leading to the misregulation of other RNAs in the network and human diseases, including cancer (Qi et al., 2015). In normal cells, miRNAs are up-regulated, and ceRNA activity is reduced, which is known as a near equimolar equilibrium state as there is enough amount of miRNAs present to repress or degrade target mRNAs (Qi et al., 2015). On the contrary, significant alteration in the expression level of one RNA transcript deemed as miRNA sponge, will sequester the shared miRNAs causing de-repression of the other RNA transcripts (Qi et al., 2015) (Figure 2.5). The miRNA sponge is commonly lncRNAs compared to mRNAs, affecting protein translation of the targeted mRNA (Tang et al.,

2022). This leads to cancer pathogenesis, including NPC, by affecting biological and molecular functions in cells.

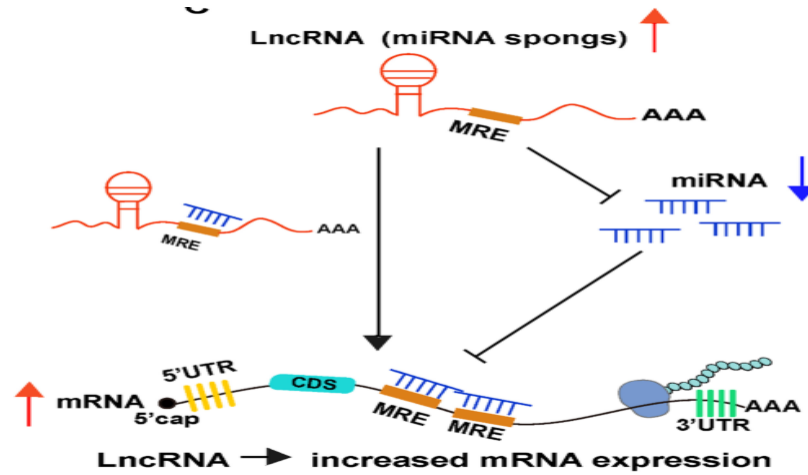


Figure 2.5: The effect of dysregulated lncRNA (miRNA sponge). The up-regulation of lncRNA (miRNA sponge) causes dysregulation of miRNAs and mRNAs where miRNAs are sequestered by lncRNAs leading to up-regulation of mRNAs. Retrieved from: <https://hereditasjournal.biomedcentral.com/articles/10.1186/s41065-021-00208-7> (Shi et al., 2021).

In cancer studies, bioinformatic analysis involving the construction of a ceRNA network (ceRNET) is commonly utilized. This network links protein-coding mRNAs with non-coding RNAs (ncRNAs), including miRNA, long non-coding RNA, pseudogene RNA and circular RNA (Qi et al., 2015). Usually, this network is constructed by curating the deposited microarray or/and next-generation sequencing (NGS) gene expression profiles of mRNAs and ncRNAs from the Gene Expression Omnibus (GEO) database. Next, the target mRNAs and lncRNAs of miRNAs are identified by using computationally predicted and/or experimentally verified databases, including miRWalk V2, miRTarBase, DIANA-LncBase V2 and V3 database, StarBase V2 and miRcode Database (Li et al., 2014; Dweep & Gretz, 2015; Paraskevopoulou et al., 2015; Huang et al., 2020). These were the databases used in this current study. Then, the target mRNAs and lncRNAs were

overlap with the mRNAs and lncRNAs from the GEO database using a Venn diagram. Subsequently, these overlapped mRNAs and lncRNAs were used to construct ceRNET using Cytoscape software. This ceRNET was used to visualise the regulatory mechanisms amongst the RNAs and to screen out potential diagnostic and prognostic biomarkers for identifying therapeutic targets (Gao et al., 2021; Xiao et al., 2021).

Usually, ribonucleic acid (RNA) biomarkers are preferred more than protein and DNA biomarkers because only a small amount of RNA sequences are required for PCR amplification, ensuring the products' sensitivity and specificity (Xi et al., 2017). They are also cost-efficient compared to protein biomarkers, and RNA is present abundantly inside the body (Xi et al., 2017). RNA biomarkers also can provide insight into cellular states and regulatory processes compared to DNA biomarkers (Xi et al., 2017). According to Xu et al. (2020), the ceRNET they constructed revealed 2,654 mRNAs regulated by 132 lncRNAs via 565 corresponding miRNAs, which they suggested could be potential biomarkers for NPC. Regardless, no further experimental verification was done to validate the results.

For non-small cell lung cancer (NSCLC) ceRNET constructed by Xiao et al. (2021), eight mRNAs were dysregulated by 12 dysregulated lncRNAs via five dysregulated miRNAs. They also further verified the expression of a significantly up-regulated lncRNA LINC00968 using qPCR and western blot in NSCLC cell lines where it showed that LINC00968 caused the down-regulation of mRNAs CTCFL and KRT5 and prevented cell apoptosis in NSCLC. On the other hand, Luo et al. (2020) revealed that ceRNET for breast cancer (BC) consisted of interactions of IL6/hsa-miR-182-5p/CDKN2B-AS1 and MMP11/hsa-miR-145-5p/ADAMTS9-AS1. They further verified the expressions of these RNAs where the results demonstrated that ADAMTS9-AS1, IL6

and hsa-miR-145-5p were down-regulated and CDKN2B-AS1, MMP11 and hsa-miR-182-5p were up-regulated similar to the ceRNET result.

2.3.1 MicroRNA (miRNA)

MicroRNAs (miRNAs) are short, non-coding RNAs, approximately 22 nucleotides in length which function in regulating gene expression (Alvarez-Garcia & Miska, 2005). A single miRNA is capable of regulating multiple mRNAs, and conversely, multiple miRNAs cooperatively are capable of regulating one mRNA due to multiple miRNA Response Elements (MRE) on an mRNA (MacFarlane & Murphy, 2010). The regulation is to repress the translation of target messenger RNA (mRNA) into protein through a partially complementary pairing of miRNA seed sequences (6-8 bases long) with the MRE located on the 3' untranslated region (3'UTR) of the mRNA (Doench et al., 2003; Nair et al., 2020). A high degree of sequence complimentary between miRNA and mRNA sequences leads to mRNA degradation by mRNA cleavage, catalyzed by Human Argonaute 2 (AGO2) endonuclease activity (MacFarlane & Murphy, 2010). Nevertheless, all mRNAs are eventually degraded, and mRNAs translated initially may be temporarily repressed when it is unfavorable for cell homeostasis (Shyu et al., 2008). The degradation rate depends on cell stimuli and signals, including hormones, iron, cell cycle progression, cell differentiation and viral infection (Wang et al., 2002).

The regulation of mRNAs is necessary to ensure a proper level of protein expressed at that given time for normal cell function and for mRNA stability in terms of a lifetime to enable correct cell homeostasis and for cells to adapt to changes upon cell stimulus, cell growth and differentiation (Shyu et al., 2008; Bicknell & Ricci, 2017; Sebastian-delaCruz et al., 2021). The stability of mRNA is important in tissues and organs' stress signals,

including starvation, infection, inflammation, toxins or tissue invasion by immune cells (Sebastian-delaCruz et al., 2021). Previous studies show cancers are due to the dysregulation of miRNAs, which lead to mRNAs dysregulation, and mostly, the miRNAs are up-regulated in cancers (Rupaimoole et al., 2016; Ratti et al., 2020). The dysregulation of miRNAs is due to genetic changes of miRNAs, epigenetic alteration including CpG islands methylation of their promoters, defect in miRNAs biogenesis machinery, dysregulation of transcription factors, and long non-coding RNA (lncRNA) dysregulation (Croce, 2009; Peng & Croce, 2016; Bhattacharjee et al., 2023).

In GC, Wang et al. (2016) revealed the up-regulation of miR-107 down-regulated Neurofibromin 1 (NF1) through dual-luciferase reporter assay, whereas Ren et al. (2014) demonstrated the down-regulated miR-449a in lung cancer caused down-regulation of E2F transcription factor 3 (E2F3) and promoted cancer cells proliferation and inhibited senescence. In NPC, Chen et al. (2016) mentioned that p21 was down-regulated in NPC cell lines and tumor tissue when miR-17-5p was up-regulated through dual-luciferase reporter assay and promoted tumor growth. In addition, Zhou et al. (2021) revealed the up-regulation of 2 hub miRNAs, hsa-miR-135b and hsa-miR-141, in the SUNE-1 NPC cell line. Nevertheless, no mRNA targets of the two hub miRNAs were identified in the study. Moreover, Liu et al. (2020) demonstrated that has-miR-452 was up-regulated in the plasma of NPC patients subsequent to predicting the up-regulation of has-miR-452 and down-regulation of its mRNA target, ITGA9 via curation of GEO microarray datasets. Interestingly, they also revealed that let-7d and let-7f miRNAs were down-regulated in the plasma of NPC patients, where both miRNAs were predicted to interact with up-regulated MYC and HMGA2 in NPC. Regardless, the interaction between these miRNAs with respective mRNAs was not verified.

The mechanism of up-regulated miRNAs in causing NPC tumorigenesis is by targeting mRNAs induces overly down-regulated or repression of protein translation coded by the mRNAs via partial binding on the MRE located on the 3' untranslated region (3'UTR) of the mRNA (Doench et al., 2003; Nair et al., 2020). This occurs when up-regulated mature miRNAs transported into the cytoplasm are loaded into AGO2, forming miRNA-induced silencing complex (miRISC), targeting and partially binding to mRNAs on its MRE (O'Brien et al., 2018). This partial binding creates bulges in the RNA duplex, which is important for mRNA repression translation instead of mRNA degradation by preventing AGO2 cleavage (MacFarlane & R Murphy, 2010). Besides, the partial binding also accelerates mRNA deadenylation and decapping for mRNA repression translation, which eventually is degraded by exoribonuclease 1 (XRN1) (MacFarlane & R Murphy, 2010). The deadenylation is initiated by PAN2/3 and completed by the CCR4-NOT complex, whereas the decapping is facilitated by decapping protein 2 (DCP2) (O'Brien et al., 2018). Similarly, this mechanism is also utilized by down-regulated miRNAs in causing target mRNA overexpression in NPC tumorigenesis, resulting in a decreased amount of miRISC complex targeting mRNAs.

2.3.2 Long Non-Coding RNA (lncRNA)

Long non-coding RNAs (lncRNAs) are non-coding RNAs which is longer than 200 nucleotides where the functions are largely unknown (Bhat et al., 2016). Generally, lncRNAs act as key regulators, either as positive or negative regulators, in gene expression at transcriptional and post-transcriptional levels in development, differentiation, and human diseases (Bhat et al., 2016). It is often expressed in a low amount and dependent on tissue and cell type according to developmental stages, and its deregulation is observed in diseases and cancers (Fatica & Bozzoni, 2014; Ratti et al., 2020). Similar to miRNAs,

lncRNAs able to regulate mRNA stability and decay by regulating miRNAs' interaction with mRNAs where it competes with miRNA as miRNA sponge (Zhang et al., 2019; Sebastian-delaCruz et al., 2021). This is due to the multiple miRNA-complementary sites on lncRNAs known as miRNA Response Elements (MRE) for multiple miRNA binding, thus reducing miRNA availability to mRNAs (Jalali et al., 2013; Hombach & Kretz, 2016). Commonly, lncRNAs act as positive regulators of mRNAs translation by preventing mRNA decay and are deemed as either a tumor suppressor or oncogene (Dhanoa et al., 2018; López-Urrutia et al., 2019).

The dysregulation of lncRNA, which leads to miRNA and mRNA deregulation, has been linked with tumorigenesis and tumor progression. It is where lncRNA commonly acts as a miRNA sponge, also known as competing endogenous RNA (ceRNA) activity, and reduces miRNA regulatory effect on mRNA (Paraskevopoulou & Hatzigeorgiou, 2016). This is revealed by Lv et al. (2018) where lncRNA ZEB1-AS1 sponged miR-181a-5p through luciferase reporter due to its significantly up-regulated expression in colorectal cancer (CRC) and this caused the up-regulation of β -catenin and TCF4. This was also revealed by Ku et al. (2021) where up-regulated LINC00240 sponged on miR-7-5p causing up-regulation of epidermal growth factor receptor (EGFR) in lung cancer. A study by Hu et al. (2014) also demonstrated that gastric adenocarcinoma predictive long intergenic non-coding (RNAGAPLINC) was significantly expressed in GC, where it sponged on miR-211-3p causing the up-regulation of CD44 and tumor cell proliferation. Generally, miRNA sponged act as quality control by regulating appropriate levels of miRNA expression to repress mRNA expression in response to cellular stress or development transition (Ebert & Sharp, 2010).

2.4 Bioinformatics

Bioinformatics is defined as conceptual biology in terms of molecules by applying informatics techniques which are computer science, mathematics and statistics, to understand and organize information related to the molecules on a large scale (Luscombe et al., 2001). It also allows the collection, storage, catalogue and analysis of information in various formats, enabling manipulation for future research (Khan, 2018). Bioinformatics applications include sequence and structure alignment, database design and mining, macromolecular geometry, phylogenetic tree construction, protein structures and function prediction, gene finding and expression data clustering (Luscombe et al., 2001). Generally, bioinformatics uses computation tools, including software programs, websites and databases with internet availability (Bayat, 2002).

2.4.1 The GEO (Gene Expression Omnibus) Database

The Gene Expression Omnibus (GEO) database is a database supported by National Centre for Biotechnology Information (NCBI), which is used to retrieve high-throughput gene expression and other functional genomic datasets containing raw and processed data with the experimental design, sample attributes and methodology (Clough & Barrett, 2016). This database is separated into two databases which are GEO Datasets and GEO profiles where GEO datasets function in searching for studies of interest, whereas GEO profiles function in searching for expression profiles of genes (Clough & Barrett, 2016).

2.4.2 The GEO2R

The Gene Expression Omnibus to R language (GEO2R) is an R programming language-based analytical tool to analyse GEO datasets by visualising the comparison of

different expression levels of “Top 250” genes of aberrant gene expression with a control group(s) (Clough & Barrett, 2016).

2.4.3 The miRWalk 2.0 Database

The miRWalk 2.0 database is the updated version of the miRWalk database, which provides predicted miRNA-target interactions and amalgamates 12 existing target mRNAs databases (DIANA-microTv4.0, DIANA-microT-CDS, miRanda-rel2010, mirBridge, miRDB4.0, miRmap, miRNAmap, doRiNA which includes PicTar2, PITA, RNA22v2, RNAhybrid2.1 and TargetScan6.2) (Moore et al., 2015). This is to increase accuracy and improve predicted targets, as each database utilized different methods to predict binding sites (Sticht et al., 2018). Besides, miRWalk 2.0 provides information on the binding sites of miRNAs within the complete sequence of a gene and compares these binding sites between all 12 databases (Dweep & Gretz, 2015). This means it predicted the binding sites of miRNAs on a gene not only on the 3' UTR but also on the promoter, 5' UTR, and CDS (Sticht et al., 2018). It also provides experimentally verified miRNA-target interaction curated from miRTarBase, PhenomiR, miR2Disease and HMDD (Robinson & Henderson, 2018).

2.4.4 MicroRNA Target Interaction Database (miRTarBase)

The microRNA TARget interaction dataBASE or miRTarBase is a curated database that provides experimentally verified miRNA-target interactions (MTIs) using high-throughput techniques, including cross-linking and immunoprecipitation and high-throughput sequencing (CLIP-seq), cross-linking, ligation, and sequencing of hybrids (CLASH-seq) and degradome-seq (Moore et al., 2015; Huang et al., 2020). These high-throughput techniques are validated by the curators using the miRNA-target gene

expression profiles from GEO and The Cancer Genome Atlas (TCGA) databases (Moore et al., 2015). The MTIs are curated manually based on promising experimental evidence using an artificial intelligence approach known as natural language processing (NLP), and the downloaded MTIs provide users with information on the technique used to verify the MTIs and link PubMed references for the MTIs (Moore et al., 2015; Huang et al., 2020).

2.4.5 DIANA-LncBase V2 and V3 database

DIANA-LncBase v2 database is the improved version of the DIANA-LncBase v1 database, whereas DIANA-LncBase V3 is the improved version of the DIANA-LncBase v2 database. These databases provide experimentally supported and computationally identified miRNA Recognition Elements (MREs) on lncRNAs in *Homo sapiens* and *Mus musculus* (Paraskevopoulou et al., 2015). These databases provide the experimental approach of miRNA-lncRNA interactions from low-yield or high-throughput methods such as western blot and CLIP-Seq, respectively (Paraskevopoulou et al., 2015; Karagkouni et al., 2020). The CLIP-Seq approaches, including HITS-CLIP (high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation), PAR-CLIP (photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation) and CLEAR-CLIP (covalent ligation and endogenous Argonaute-bound RNA), which can identify Argonaute (AGO):miRNA binding sites on a targeted RNAs (Karagkouni et al., 2020). This is due to the binding of miRNA on targeted mRNAs involving the loading of miRNAs into protein AGO to induce cleavage, degradation or translational suppression of mRNAs (Karagkouni et al., 2020). Compared to V1, additional information on the miRNA-lncRNA interactions and providing cell types and tissue-specific miRNA-lncRNA interactions regulation for specific conditions in both humans and mouse are provided in V2 and are kept in V3 (Paraskevopoulou et al., 2015).

The miRNA-lncRNA interactions in raw sequencing data and manually curated publications in DIANA-lncBase v2 are added from more than 70,000 interactions to ~240,000 interactions in DIANA-LncBase V3 (Karagkouni et al., 2020). The deposited curated experimental approach interactions in V2 are also updated in V3 by gathering data from re-analysed literature of a similar experimental approach. However, in V3, the curation of *in silico* computational approach for miRNA-lncRNA interactions is excluded, and the focus is shifted to curating more high-throughput datasets. Additional information is also included in V3 such as pseudogene and the curation of lncRNAs transcript expression profiles in different cell types/tissues and cellular compartments, enabling users to choose cell and tissue types and the methodology used. Known short variant information from reference databases such as dbSNP, ClinVar and COSMIC are also incorporated in V3 and annotation of lncRNAs in V3 are updated according to the updated annotation of GENCODE; hence the change in lncRNAs identifiers and sequences observed in V3 (Karagkouni et al., 2020).

2.4.6 StarBase V2

StarBase, which stands for sRNA target base, is a novel RNAs interaction database where in V2, the RNA-RNA and protein-RNA interactions from all datasets of cross-linking and Argonaute (Ago) immunoprecipitation coupled with high-throughput sequencing (CLIP-seq) methods are added for easy identification of regulatory interaction networks among RNAs (Yang et al., 2011; Li et al., 2014). This CLIP-seq techniques includes HITS-CLIP (high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation), and PAR-CLIP (photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation) (Li et al., 2014). Compared to V1, RNA classes interactions, including lncRNAs, circular RNA (circRNA), and pseudogenes, are added in

V2 with the addition of three organisms including *Homo sapiens*, *Mus musculus*, and *Caenorhabditis elegans* (Li et al., 2014). Typically, StarBase is used to identify the interactions of RNAs with miRNAs through the identification of miRNAs conserved sites and other RNAs before confirming it by using the Ago CLIP cluster and the functions of miRNAs are provided through targeted mRNAs enrichment analysis by integrating the data from databases such as KEGG database and GO ontology (Li et al., 2014).

2.4.7 miRcode Database

miRcode database provides predicted miRNA binding sites on the lncRNA transcripts in humans, which are retrieved from GENCODE version 11 (Paraskevopoulou et al., 2015). It also provides a complete prediction of putative target sites across miRNA for lncRNAs curated from the GENCODE (Jeggari et al., 2012). The target sites are identified based on principles such as seed complementary and evolutionary conservation where the seed complementary uses seed region, which is the region of bases 2-8, from the 5'-end of miRNA for putative target sites with lncRNA (Jeggari et al., 2012).

2.4.8 TFcheckpoint database

The TFcheckpoint database is a PubMed literature-based database which contains transcription factors (TFs) candidates of humans, mouse and rats; thus, it acts as a checkpoint for TF candidates due to manual curation of the TF candidates from small-scale and genome-scale experimentally available literature (Chawla et al., 2013). The selected TF consists of tangible experimental evidence of binding to RNA polymerase II (RNAP II) following the function of TFs in guiding RNAP II to core-promoter DNA, forming a pre-initiation transcription complex for DNA transcription (Chawla et al., 2013).

2.4.9 ViRBase

ViRBase is a PubMed literature-based database containing viral non-coding RNAs-associated virus-virus, virus-host, host-virus and host-host interactions from 61 classes of viruses and 24 host species (Li et al., 2015). The non-coding RNAs included in this database are long non-coding RNA (lncRNA), microRNA (miRNA) and small nucleolar RNA (snoRNA) (Li et al., 2015). The interaction networks between the virus and host are linked to viral infection and manually curated from experimentally- or computational-based literature (Li et al., 2015). The experimental-based literature is curated from techniques including cross-linking and immunoprecipitation and high-throughput sequencing (CLIP-seq), quantitative real-time PCR, Western Blot and immunohistochemistry analyses (Li et al., 2015). For predicted virus-host RNA interactions, the predicted binding sites and match scores of the interactions are provided.

2.4.10 The Search Tool for Retrieval of Interacting Genes (STRING) Database

The Search Tool for Retrieval of Interacting Genes/Proteins (STRING) database consists of all the collection of information from publicly available protein-protein interaction (PPI) databases to complement the information with computational predictions (Szklarczyk et al., 2018). The link between the proteins by the STRING database is based on physical (direct) and functional (indirect) interactions (Szklarczyk et al., 2023). All the curated protein interactions in the STRING database are assessed and quantified by benchmarking/calibrating the interactions against common membership in KEGG pathway maps (Szklarczyk et al., 2023). This benchmarking is represented by a ‘confidence score’ for each interaction, scaled between 0 and 1 (Szklarczyk et al., 2023). The ‘confidence score’ corresponds to estimating the likelihood of a given interaction being true, according

to given evidence, where all the scores are integrated into the final combined confidence score (Szkłarczyk et al., 2023).

The protein-protein interaction information in the STRING database integrates seven distinct evidence types for the protein-protein interactions. The first three evidence types are the neighborhood, fusion, and co-occurrence known as genomic context, where protein interactions are identified based on only genome sequences by detecting evolutionary constraints due to functional gene-gene partnership (Szkłarczyk et al., 2023). Next, the co-expression evidence type provides protein interactions through common expression regulation of both proteins and transcripts across a multitude of conditions (Szkłarczyk et al., 2023). These evidence types are predicted protein-protein interactions providing functional interactions between proteins (Szkłarczyk et al., 2023).

For validated evidence type, the first is the experiments evidence type, which provides protein interactions that were experimentally supported and imported from primary databases, including BioGRID, DIP, PDB, IntAct, and its partner databases in the IMEx consortium (Szkłarczyk et al., 2023). Second, the database evidence type imports well-established knowledge from KEGG, Reactome, MetaCyc, EBI Complex Portal, and Gene Ontology Complexes (Szkłarczyk et al., 2023). Lastly, the text mining evidence type parsed articles from the PMC Open Access Subset, PubMed abstracts, and summary texts from OMIM and SGD entry descriptions (Szkłarczyk et al., 2023). These evidence types provide physical interactions between proteins (Szkłarczyk et al., 2023).

2.4.11 Cytoscape Software

Cytoscape is a software used to visualize and integrate interaction networks between bio-molecule with annotations, gene expression profiles and other molecular

states and is also used to link the network to databases of functional annotations (Shannon et al., 2003). The interaction networks are represented where the molecular species are exhibited in nodes and inter-molecular interactions in the form of links which are edges, between nodes (Shannon et al., 2003). It also enables integrating additional data from public repositories or experiments, changing visualization styles, and using algorithms for network layout, clustering and enrichment, and network analysis (Szkarczyk et al., 2016). Besides, it is also used to construct and visualize lncRNA-miRNA-mRNA competing endogenous RNAs (ceRNA) network as well as to visualize and analyze protein-protein interaction (PPI) through its integrated The Search Tool for Retrieval of Interacting Genes/Proteins (STRING) Database software (Szkarczyk et al., 2016; Gao et al., 2019). It also provides Network Analyzer and Molecular Complex Detection (MCODE) tools as its plug-ins to analyze the PPI network further.

The Network Analyzer displays the topology properties and centrality measures for un-directed and directed networks such as the nodes, edges, and connected components numbers, the network diameter, radius, density, centralization, heterogeneity, clustering coefficient, and path length characteristic (Assenov et al., 2007). It also provides charts of the distribution of node degrees, neighborhood connectivity, average clustering coefficients, and shortest path lengths (Assenov et al., 2019). The hub genes are identified using Network Analyzer, which calculates the topological properties of the PPI network and the node degree (Ye et al., 2019). Hub genes are genes that interact with many other genes and can regulate the expression of the other genes; thus, they are deemed tumor suppressor genes or oncogenes (Yu et al., 2017). In addition, transcription factor genes are also regarded as hub genes due to their ability to regulate the expression of a set of genes (Yu et al., 2017). Molecular Complex Detection (MCODE) is used to detect densely

connected regions within a graph of large biological networks, such as protein complexes, by assigning a weight to each vertex (node) in conjunction with the density of the protein complexes' local neighborhood (Angelov et al., 2016).

2.4.12 The Database of Annotation, Visualization and Discovery software (DAVID)

The Database of Annotation, Visualization and Discovery software (DAVID) is commonly used for pathway enrichment analysis where it provides Gene Ontology (GO) for term analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) for pathway enrichment analysis of gene. The functional annotation and analysis of large lists of genes are facilitated by the amalgamation of descriptive data on the gene functions with graphical displays due to its data mining tools. The GO is the project of the Gene Ontology Consortium to provide structured, defined, common, controlled vocabulary terms to describe the roles of genes and gene products in any organism (Ashburner et al., 2000).

The word ontology is defined as describing entities into categories and the relationships between the entities (Ashburner et al., 2000). The vocabulary terms are subdivided into three non-overlapping ontologies, which are Molecular Function (MF), Biological Process (BP), and Cellular Component (CC), where each ontology describes the aspects of a gene or gene product functionality combined with the relationship between the terms (du Plessis et al., 2011). The KEGG provides systematic analysis to identify the functions of genes by linking genomic information with the higher order functional information, and the gene functions are assigned according to the function of the molecules interacting with the gene (Kanehisa & Goto, 2000). KEGG is a project of the Human Genome Program of the Ministry of Education, Science, Sports, and Culture in Japan where the database consists of three databases such as PATHWAY, GENES and LIGAND

(Kanehisa & Goto, 2000). The databases represent higher-order functions based on the interaction network of the molecules, collecting gene catalogues for the completely sequenced and partially sequenced genomes and collecting chemical compounds in the cell, enzyme molecules and enzymatic reactions, respectively (Kanehisa & Goto, 2000).

2.4.13 Kaplan-Meier (KM) Plotter

The Kaplan-Meier (K-M) plotter is a database that provides a large sample used to evaluate the survival of patients with different types of cancers (Ma et al., 2021). This plotter uses databases, including Gene Expression Omnibus (GEO), European Genome-phenome Archive (EGA) and The Cancer Genome Atlas (TCGA) (Liu et al., 2021). Compared to other K-M plotters, this website provides a user-friendly plotter that integrates databases, including Gene Expression Omnibus (GEO), European Genome-phenome Archive (EGA) and The Cancer Genome Atlas (TCGA), which requires no curation or input of expressions datasets unless preferred as it is provided as custom data option. It also integrates clinical data for each sample extracted from original publications and files from databases used. This includes race, histological subtype, stage, grade, mutation status, treatment data, and analysis based on cellular contents and cancer type (Zwyea et al., 2021; Györfy, 2023). It also automated hazard ratio (HR) and 95% confidence with a P-value < 0.05 , which is regarded as statistically significant by using the Log-rank test in calculating the overall survival (OS) (Liu et al., 2021).

Besides, the K-M plotter website corrects the multiple hypothesis testing by computing the false discovery rate (FDR) generated via the Benjamini-Hochberg method (Lánczky & Györfy, 2021). The FDR is presented in percentage where the lowest FDR is equivalent to the highest significance for P-value or HR (Györfy, 2023). This website also

implements Cox regression on proportional hazard assumption test to tackle the censoring, independence of prognosis throughout the desired event, and that the samples participate at different time points in the study should have the same prognosis (Lánczky & Györffy, 2021). The Cox regression was also used to compute univariate survival analysis for all input genes, followed by multivariate survival analysis to include clinical and pathological variables (Györffy, 2023).

Generally, K-M curves are used to correlate the expression of mRNA, miRNA and proteins with patients' survival. The K-M survival curves (K-M curve) are generated using this plotter. These curves are generated in non-continuous steps where the Y-axis represents the survival time probabilities, known as cumulative probability. The X-axis represents the survival time (times-to-event), defined as the interval time of study and, is dependent on the desired event of the study and does not necessarily mean death (Rich et al., 2010). The cumulative probability is the probability at the beginning and throughout the interval, calculated by multiplying the interval probability up to a certain interval (Rich et al., 2010). The lengths of the horizontal lines along the X-axis of the serial times indicate the survival duration for the interval, whereas the vertical lines are only for pleasing observation, and the vertical distances between the horizontals illustrate the change in cumulative probability (Rich et al., 2010).

In cancer gene expression studies, the survival time is defined as death (Rich et al., 2010). However, it varies according to the subjects, where the end could be anywhere along the study's timeline due to censored or the study's endpoint (event of interest) (Rich et al., 2010). Censored means the subjects drop out, are absent for follow-up, or survived a drug/treatment trial indicated by tick marks/black dots in the K-M curve (Rich et al.,

2010). This basic concept of K-M survival analysis considers censored subjects compared to the Mann-Whitney test (Lánczky & Györffy, 2021). To plot the curve, subjects are divided into two groups; thus, the statistical significance between the groups is calculated using the log-rank test method to calculate the chi-square (X^2) for each event time for each group and sum the results (Rich et al., 2010). Next, the hazard ratio (HR) is used to test the statistical significance of the two groups using the Cox regression method, which also provides the confidence interval for the HR. It will indicate the increase rate in the desired event in one curve versus the other (Rich et al., 2010).

The HR is an estimation of the hazard rates ratio in the treated vs control group (Spruance et al., 2004). The hazard rate is the probability that if the event of interest has not occurred, it will occur in the next interval divided by the length of that interval (Spruance et al., 2004). It is calculated following the formula as follow (Blagoev et al., 2012):-

$$HR = h^{arm1} / h^{arm2} \quad \text{Equation 2.1}$$

Where arm 1 and arm 2 are the hazard rate or number of patients experiencing “hazardous event” at any time interval (Blagoev et al., 2012). The arm 1 or 2 is calculated by minus the patients in a previous time with the patients at the desired time period.

CHAPTER 3

MATERIALS AND METHODS

3.1 Microarray Dataset Retrieval

The microarray gene expression profiles of messenger RNAs (mRNAs), microRNAs (miRNAs), and long non-coding RNAs (lncRNAs) between nasopharyngeal cancer (NPC) and nasopharyngeal were screened and retrieved from Gene Expression Omnibus (GEO) database on the National Centre for Biotechnology Information (NCBI) website (<https://www.ncbi.nlm.nih.gov/>) (Chen et al., 2015). The access year of the datasets was between 2020 until 2023. The keyword “nasopharyngeal carcinoma” was input into the search box of GEO datasets, the study type was set as expression profiling by array and the organism filter was set to *Homo sapiens*. The following datasets were excluded: i) the samples are of cell samples; ii) no normal nasopharyngeal controls; ii) submitted before 2014. The datasets were curated only for datasets from 2014 to ensure the expression datasets followed updated microarray guidelines and protocols. Subsequently, all the differentially expressed gene (DEG), differentially expressed lncRNA (DEL), and differentially expressed miRNA (DEmiRNA) datasets were checked for quantile normalization, and quantile normalization was done using an R software called Limma R package in R studio (version 4.0.1). The R software was downloaded to a laptop with an Intel(R) Core(TM) processor with 8.00 GB RAM and a 64-bit operating system.

3.2 Differential Expressed Gene, lncRNA and miRNAs Identification

The differentially expressed gene (DEGs), differentially expressed microRNAs (DEmiRNAs) and differentially expressed lncRNAs (DELs) from the GEO databases were identified by using an online tool GEO2R (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>) (Ye

et al., 2019). The cut-off criterion was set to P-value < 0.05, expression Fold Change (FC) ≥ 2 and the $|\log_2 \text{ fold-change}| \geq 1$ (Liu et al., 2019; Zhu et al., 2020).

3.2.1 Differential Expressed Gene, lncRNA, miRNA Visualization

Next, the visualization of the top 50 differentially expressed genes (DEGs), differentially expressed miRNAs (DEmiRNAs) and differentially expressed lncRNAs (DElNs) datasets were done using volcano plots from a web-based software, VolcanoR (<http://volcanor.bioinf.su>). The cut-off criterion was set to $|\log_2 \text{ fold-change}| \geq 1$ and P-values < 0.05, where the P-value was converted into $-\log_{10}$. Subsequently, the overlapping of differentially expressed RNAs between two or more similar datasets was identified by using a Venn diagram from a web-based Venny tool (<https://bioinfogp.cnb.csic.es/tools/venny/>) (Ye et al., 2019).

3.3 Target mRNA and Target lncRNA of DEmiRNA Identification

3.3.1 DEmiRNA-Target mRNA Interaction Identification

The differentially expressed miRNAs (DEmiRNAs) from the GEO database were used to identify computationally predicted and experimentally validated target mRNAs from miRwalk 2.0 (<http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/>) and miRTarBase (<http://mirtarbase.cuhk.edu.cn/php/search.php>) databases, respectively (Yang et al., 2019; Hu et al., 2020). The predicted target mRNAs with P-value < 0.05 and were common in at least five databases integrated into miRWalk 2.0 were selected for further analysis (Yang et al., 2019). For miRTarBase, the predicted target mRNAs with both “less strong” and “strong evidence” validation methods were selected. The target mRNAs from both databases were combined and overlapped with DEGs from the GEO database by using the Venn diagram from the Venny tool.

3.3.2 DEmiRNA-Target lncRNA Interaction Identification

The differentially expressed miRNAs (DEmiRNAs) from the GEO database were used to identify computationally predicted and experimentally validated target lncRNAs. The computationally predicted databases used were miRcode (<http://www.mircode.org/>) and DIANA-LncBase V2 (www.microrna.gr/LncBase) where for miRcode the parameters were set as gene class for all lncRNAs, site conservation for most mammals and transcript region as any/ncRNA and for DIANA-LncBase V2 the parameters were set to default threshold greater than or equal to 0.07 for high sensitivity and precision (Zhu et al., 2020).

In contrary, for identification of experimentally validated target lncRNAs, databases DIANA-LncBase V3 (<https://diana.e-ce.uth.gr/lncbasev3/home>) and StarBase V2 (<http://www.sysu.edu.cn/>) were used where the parameters for DIANA-LncBase V3 were set as *Homo sapiens* for species and high miRNA confidence level whereas for StarBase V2 the parameters were set to strict stringency (≥ 5) for the CLIP Data to yield the higher sequence similarity in seed regions for binding and low stringency (≥ 1) were set for Degradome Data parameter to yield lower to high near-perfect sequence match of cleavage sites between miRNA and lncRNAs (Liang & Sun, 2019). The target lncRNAs from these databases were combined and overlapped with DELs from the GEO database using the Venn diagram from the Venny tool.

3.4 Competing Endogenous RNA (ceRNA) Network Construction

The construction of lncRNA-miRNA-mRNA competing endogenous RNA (ceRNA) network for both down-regulated and up-regulated RNAs expression was done by using Cytoscape software (Version 3.8.2) (<https://cytoscape.org/>) (Zhou et al., 2019). This software can link and connect the ceRNA activity based on the parameters mentioned

above during the curation of target mRNAs and lncRNAs for miRNAs. The lncRNAs and mRNAs were the overlapped target lncRNAs with differentially expressed lncRNAs (DELs) and the overlapped target mRNAs with the differentially expressed genes (DEGs).

3.5 Epstein-Barr Virus miRNA-mRNA Interaction

3.5.1 Epstein-Barr Virus miRNAs-mRNAs Interaction Identification

The predicted Epstein-Barr Virus (EBV) miRNAs-mRNAs were identified by using ViRBase (<http://www.rna-society.org/virbase/index.html>), and keyword search for Virus Type was set to *Human gammaherpesvirus 4* (Jing et al., 2018). The filter for Interaction Type was set to Virus-Host Interaction, the host Species was set to *Homo sapiens*, and the Detection Method was set to All. The confidence score was set from 0.5 to 1.00 because the higher the confidence score, the more evidence resources for the interactions (ViRBase, 2021). Only EBV miRNAs from *BamH1* fragment A transcript (BART) cluster was selected.

3.5.2 Epstein-Barr Virus miRNA-mRNA Interaction Network Construction

Using a Venn diagram and Cytoscape, the predicted mRNAs interacting with the EBV miRNAs were overlapped with the differentially expressed genes (DEGs) in the ceRNA network. Next, the EBV BART miRNA-mRNA interaction network was constructed and visualized using Cytoscape (Jing et al., 2018). The reason for BART miRNAs was solely focused on in this study due to their overexpression in NPC tumor samples and cell lines compared to the BHRF1 miRNAs (Zhang et al., 2022).

3.6 Protein-Protein Interaction (PPI)

3.6.1 Protein-Protein Interaction (PPI) Network Retrieval and Construction

The physical protein-protein interaction (PPI) between the differentially expressed genes (DEGs) was identified by using the Search Tool for the Retrieval of Interacting Gene (STRING) database (<https://string-db.org/>) and to visualise the indirect and direct interaction types of the PPI due to Cytoscape unable to read and present these interactions (Tang et al., 2018). The cut-off criterion for the PPI pairs retrieved from the STRING database was set to a combined score more than 0.4, a medium or high confidence (Ye et al., 2018). However, the PPI constructed in Cytoscape using the STRING database plug-in was set to combined score more than 0.55, a medium or high confidence to strengthen the interaction evidence. Next, the topological properties of the PPI were calculated using the Network Analyzer plug-in to analyse its network diameter and clustering coefficient (Ye et al., 2019).

3.6.2 Hub Genes Identification from The PPI

The cytoHubba and Network Analyzer plug-ins of the Cytoscape were used to identify the hub genes from the PPI network (Xu et al., 2020). The cytoHubba was used to filter out the top 20 hub genes strongly connected to each other using the degree algorithm (Xu et al., 2020). Subsequently, the Network Analyzer was used to further filter out the hub genes from the top 20 hub genes by calculating the node degree. The parameter for this was genes with a degree more than 5 were considered hub genes due to being deemed as contributor causes of biological processes (Han et al., 2004; Ye et al., 2018). An undirected network parameter was applied to further filter out the hub genes using the Network Analyzer.

3.6.3 Gene Modules Identification

The Molecular Complex Detection (MCODE) plug-in of the Cytoscape was used to further analyse and screen for the gene modules of the protein-protein interaction (PPI) network. Gene modules are densely connected regions in PPI where the genes are closely related and functionally similar (Xu et al., 2020). The parameters for the network scoring were set as degree cutoff = 2, and the self-edges nodes were excluded by turning off Loops. In addition, the parameters for the cluster finding were set as node score cutoff = 0.2, k-core = 2, and max. Depth = 100 (Yan et al., 2019). The Haircut was included to remove singly connected nodes from clusters, and the Fluff was excluded to prevent the expansion of the cluster core (Bader et al., 2020). The gene modules with scores greater than or equal to 4 and nodes greater than or equal to 4 were considered significant.

3.7 Transcription Factors (TF)-mRNAs Interaction and Network

3.7.1 Transcription Factors Identification

The predicted transcription factors (TFs) among the differentially expressed genes (DEGs) were identified by using the TFcheckpoint database (<http://tfcheckpoint.org/>) (Zhong et al., 2019). Only the TFs with experimental evidence were selected. Next, using a Venn diagram, the TFs were overlapped with the DEGs from Step 3.2.

3.7.2 Transcription Factors (TF)-mRNAs Network Construction

The transcription factors (TFs)-mRNAs network was constructed and visualized using Cytoscape, and any of the TFs that were hub genes in the PPI network were considered hub TFs (Jing et al., 2018).

3.8 Cross Regulatory Networks Construction

The overlapped viral and host RNAs in all of the three networks (TF-mRNAs, ceRNA, EBV miRNAs-mRNAs) were identified by using Cytoscape and were used to construct a cross-regulatory network and their connections were visualized by using Cytoscape (Jing et al., 2018).

3.9 Functional and Pathway Enrichment Analysis of mRNAs

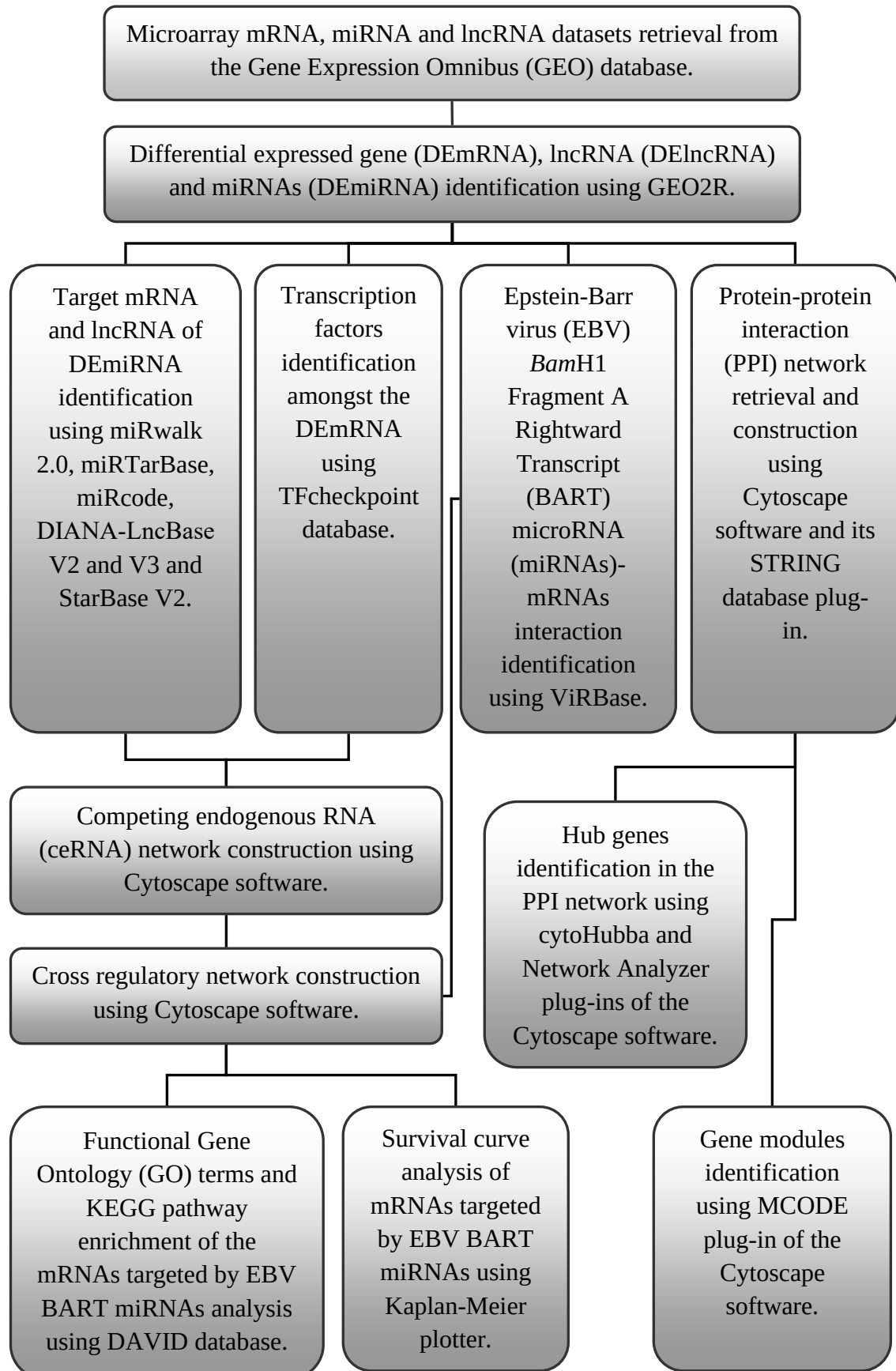
The functional and pathway enrichment analysis of the differentially expressed genes (DEGs) targeted by EBV miRNAs in the cross-regulatory were analysed by using the Database for Annotation, Visualisation and Integrated Discovery (DAVID) database (<https://david.ncifcrf.gov/home.jsp>) (Zhang et al., 2019). This was to identify the cellular functions and pathways affected by EBV miRNAs in tumorigenesis of nasopharyngeal cancer (NPC). The database provides Gene Ontology (GO) terms analysis and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis (Tang et al., 2014). The Gene Ontology (GO) terms analysis was used to analyse the functions of the DEGs through Biological Processes (BP), Molecular Function (MF) and Cellular Component (CC) (Yang et al., 2019). On the contrary, the KEGG pathway enrichment analysis was used to explore the pathway of the targeted DEGs (Yang et al., 2019). The cut-off criteria for both the GO and KEGG pathway were set as P-value < 0.05 (Jing et al., 2018). The enrichment analysis results were visualised using bubble plots in GraphPad Prism (Version 9.1.0) (Zhang et al., 2019).

3.10 Survival Curve Analysis

The predicted overall survival (OS) in dysregulation of the mRNAs targeted by BART miRNAs was done by using survival curve analysis, plotted by using Kaplan-Meier

plotter (K-M plotter) (<http://kmplot.com/analysis/>). It is a user-friendly plotter which requires no curation or input of expressions datasets. This is to identify the deregulated mRNAs' correlation with the patient's survival rate (Xiao et al., 2021). The curve shows both up-regulated and down-regulated expressions of mRNAs (Xiao et al., 2021). The P-value <0.05 was considered statistically significant, which was generated by the plotter calculated using a log-rank test (Wang et al., 2022). Moreover, the hazard ratio (HR) was also generated to measure the magnitude of the difference between the two curves in the K-M plot, and the P-value indicates the statistical significance of the difference (Brody, 2016). The HR is with 95 % confidence intervals. The curves generated was interpreted as the up-regulated group patients have a low survival rate when the $HR > 1$, whereas up-regulated group patients have a high survival when the $HR < 1$ (Mishra et al., 2019). The RNA-seq database from a Head and Neck cancer was selected for the K-M plotter to generate the curves for nasopharyngeal cancer.

Flow Chart of Materials and Methodology



CHAPTER 4

RESULTS

4.1 Microarray Dataset

The datasets which were curated from Gene Expression Omnibus (GEO) were in total of five microarray datasets where one dataset was a miRNAs dataset (GSE70970), three datasets were lncRNA datasets (GSE95166, GSE126683 and GSE61218), and three datasets were mRNA datasets (GSE64634, GSE126683 and GSE61218) as shown in Table 4.1. Both GSE126683 and GSE61218 datasets were a combination of lncRNA and mRNA expression datasets.

Table 4.1: The information of the NPC RNAs microarray datasets for each of the five mRNAs, miRNAs and lncRNAs datasets.

Accession Number	Platform	Sample Number		Reference	PMID	Submitted Year
		Normal	NPC			
GSE70970	GPL20699 nCounter® Human miRNA Assay (v1.0, Nanostring)	17	198	Bruce et al. (2015)	25738365	2015
GSE95166	GPL15314 Arraystar Human LncRNA microarray V2.0 (Agilent-033010 Probe Name version)	4	4	Zhang et al. (2017)	Unknown	2017

Table 4.1 continued

GSE126683	GPL16956 Agilent-045997 Arraystar human lncRNA microarray V3 (Probe Name Version)	3	3	Zheng et al. (2019)	31331909	2019
GSE61218	GPL19061 Agilent-043965 custom human array oelinc_xw	6	10	Fan et al. (2014)	32972383	2014
GSE64634	GPL570 [HG- U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	4	12	Bo et al. (2015)	26246469	2015

4.2 Differential Expressed Gene, lncRNA and miRNA

The information on the number of up-regulated and down-regulated RNAs in each database identified using GEO2R is shown in Table 4.2 and visualized using a volcano plot (Figure 4.1).

Table 4.2: The number of the up-regulated and down-regulated RNAs in each of the five microarray datasets curated from the GEO database.

Accession Number	Differentially expressed RNAs			
	RNAs	Up-regulated	Down-regulated	Total
GSE70970	miRNAs	38	33	71
GSE95166	lncRNAs	1,861	1,775	3,636
GSE126683	mRNA	579	599	1,178

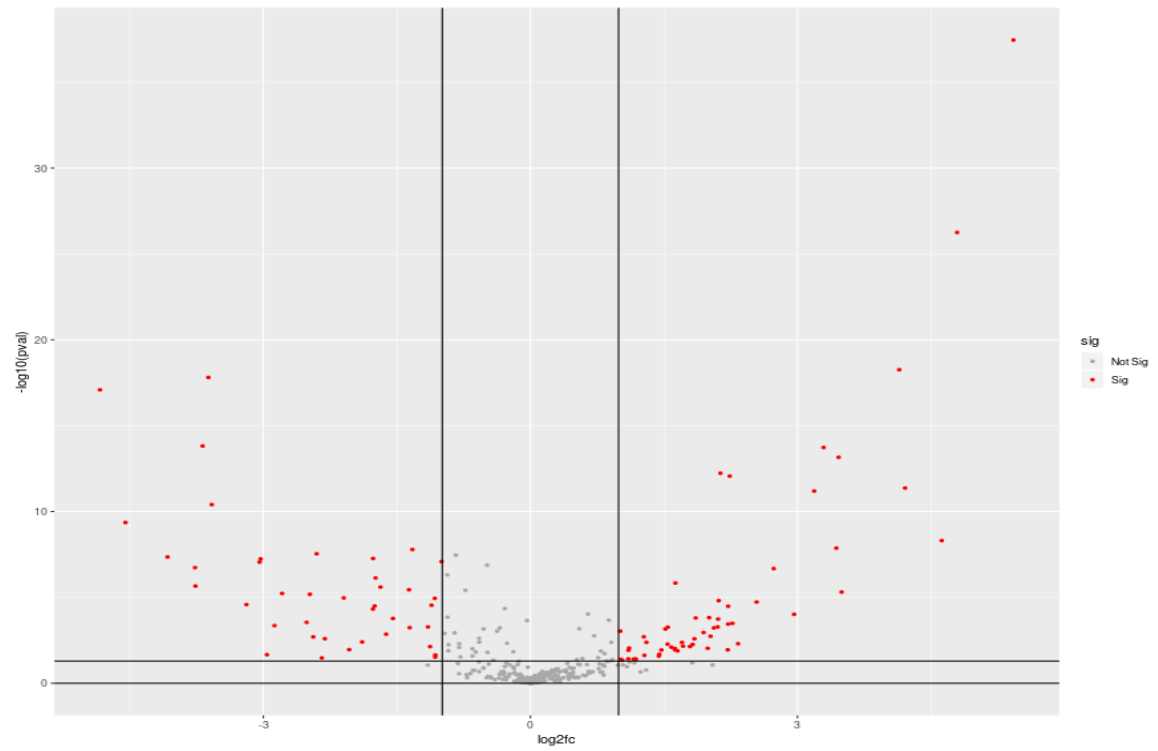
Table 4.2 continued

	lncRNA	264	539	803
GSE61218	lncRNA	789	567	1,356
	mRNA	2,752	4,876	7,628
GSE64634	mRNA	1,424	1,605	3,029

4.2.1 Differential Expressed Gene, lncRNA and miRNA Visualization

Based on the volcano plots in Figure 4.1, the volcano plot of GSE70970 showed 33 down-regulated miRNAs on the left side and 38 up-regulated miRNAs on the right side, indicated by the red dots. The volcano plot GSE95166 showed 1,775 down-regulated lncRNAs on the left side and 1,861 up-regulated lncRNAs on the right side in red dots. For GSE126683, the volcano plot exhibited a total of 1,138 for both down-regulated mRNAs and lncRNAs and a total of 843 up-regulated mRNA and lncRNA in red dots. Next, the volcano plot for GSE61218 showed 5,443 total down-regulated mRNA and lncRNA and 3,541 total up-regulated mRNA and lncRNA. Lastly, the volcano plot GSE64634 exhibited 1,605 down-regulated mRNAs and 1,424 up-regulated mRNAs in red dots.

GSE70970



GSE95166

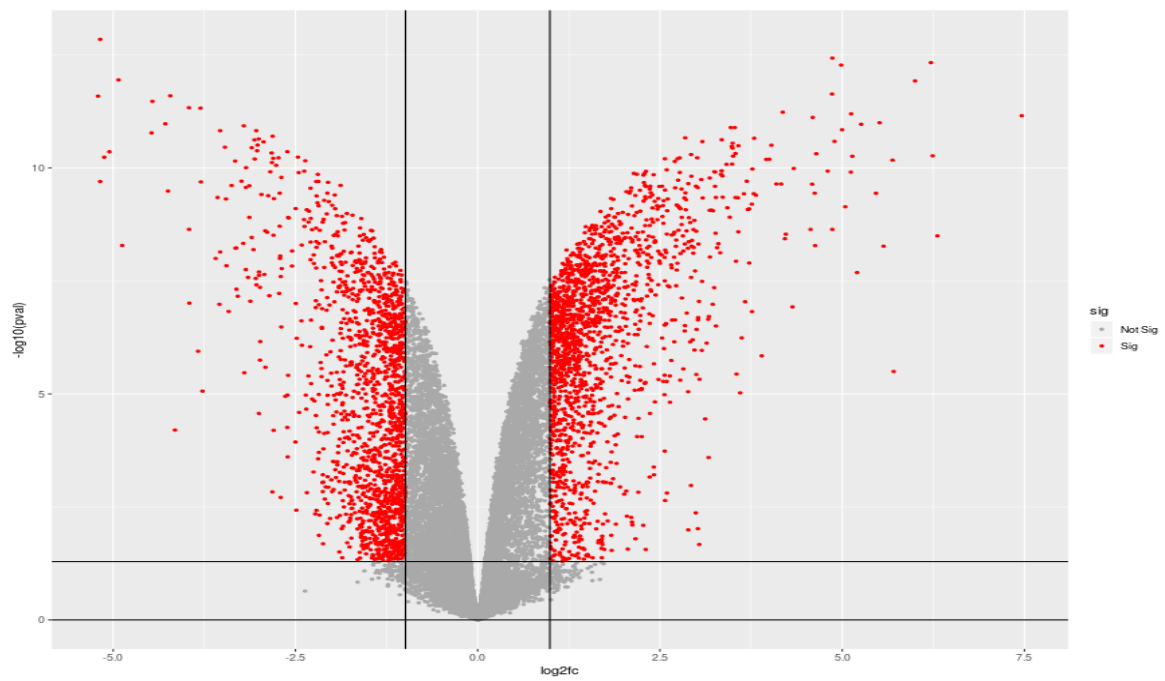
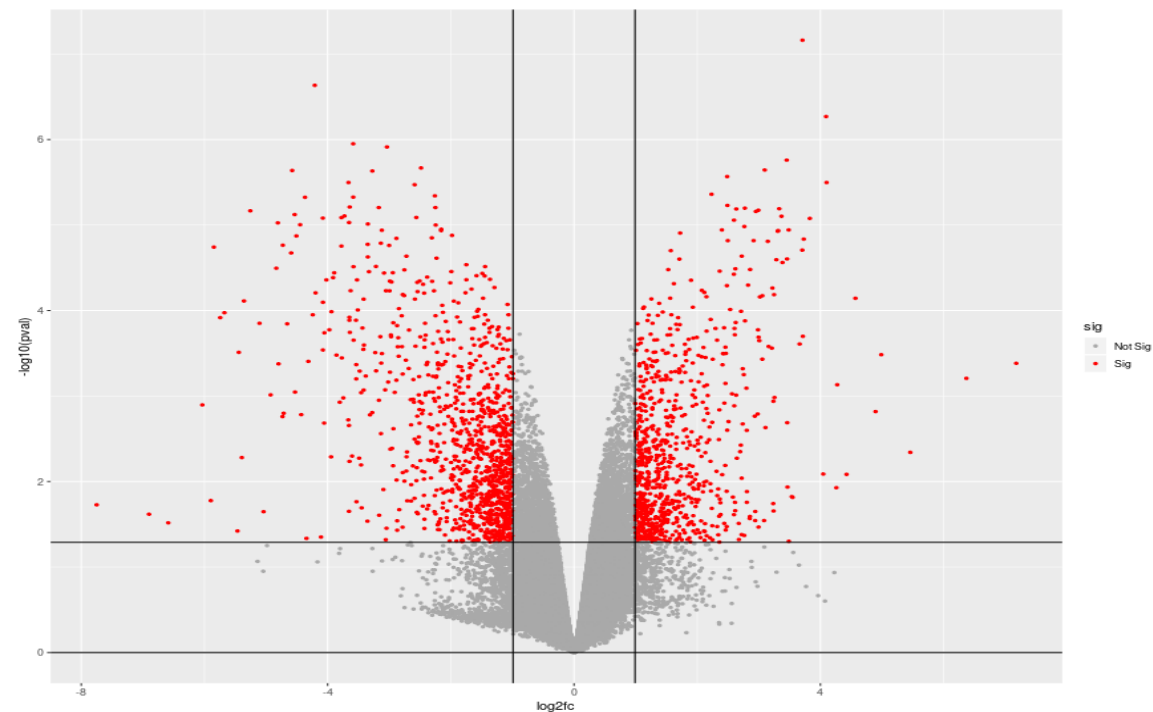


Figure 4.1 continued

GSE126683



GSE61218

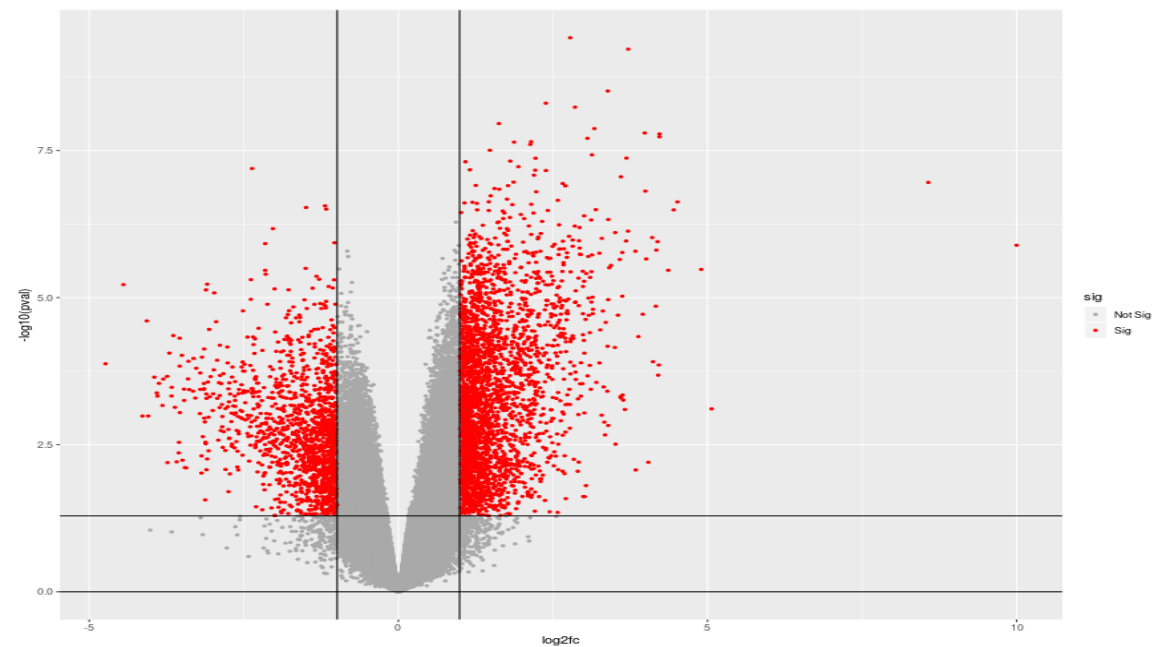


Figure 4.1 continued

GSE64634

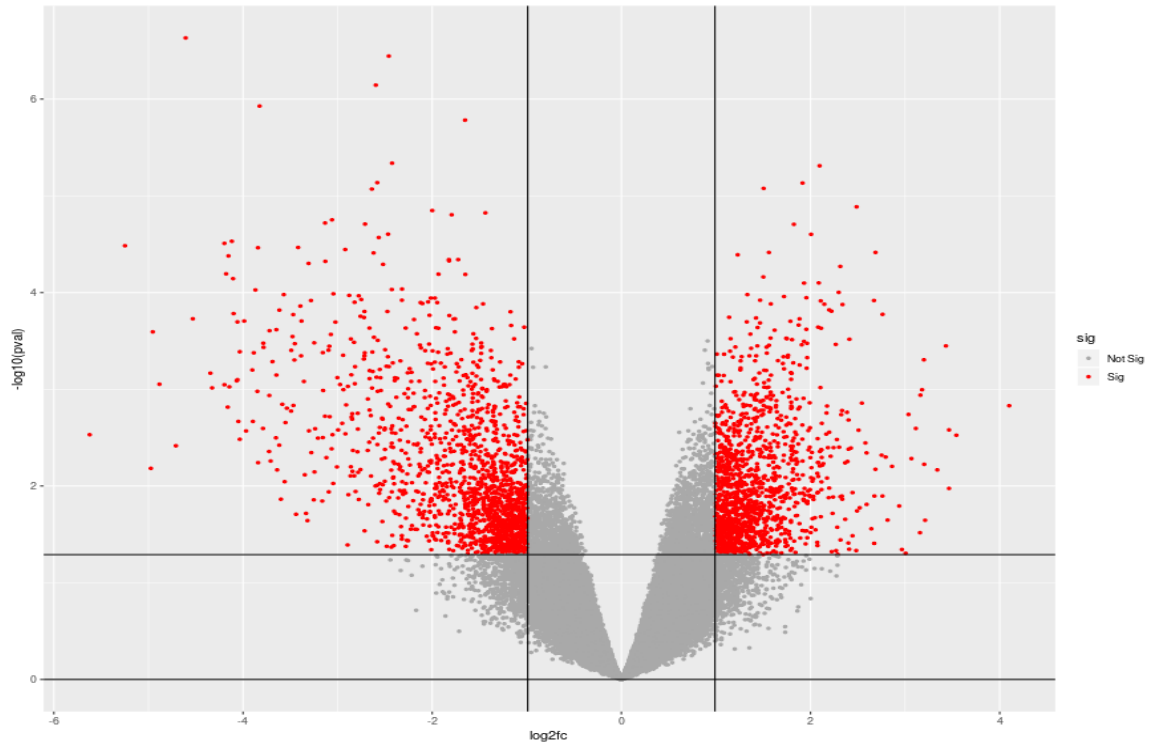


Figure 4.1: The volcano plots of all the five GEO RNAs datasets. The red dots represent the significantly up-regulated and down-regulated RNAs whereas the gray dots represent the not significantly deregulated RNAs. The RNAs on the left side are the down-regulated ones and the up-regulated ones are on the right side. The Y axis presents the negative logarithm of the p-value in base 10 and the X axis is the logarithm of the fold change between the up-regulated and down-regulated expression levels.

4.2.2 Differential Expressed Gene, lncRNA and miRNA Overlapping

The Venn diagram results for the intersection between three of the mRNA datasets (GSE64634, GSE126683 and GSE61218) showed 65 common up-regulated mRNAs and 123 common down-regulated mRNAs (Figure 4.2 and Table 4.3), whereas the overlapping of three lncRNA datasets (GSE95166, GSE126683 and GSE61218) showed no common up-regulated lncRNAs (Figure 4.3). Nonetheless, three common down-regulated lncRNAs

were identified in three of the lncRNAs datasets (Figure 4.3). These lncRNAs were CASC2, TPTE2P1, ARHGEF26-AS1. However, no overlapping step was done on the miRNA dataset (GSE70970) because only one miRNA dataset was submitted after 2014 until 2020.

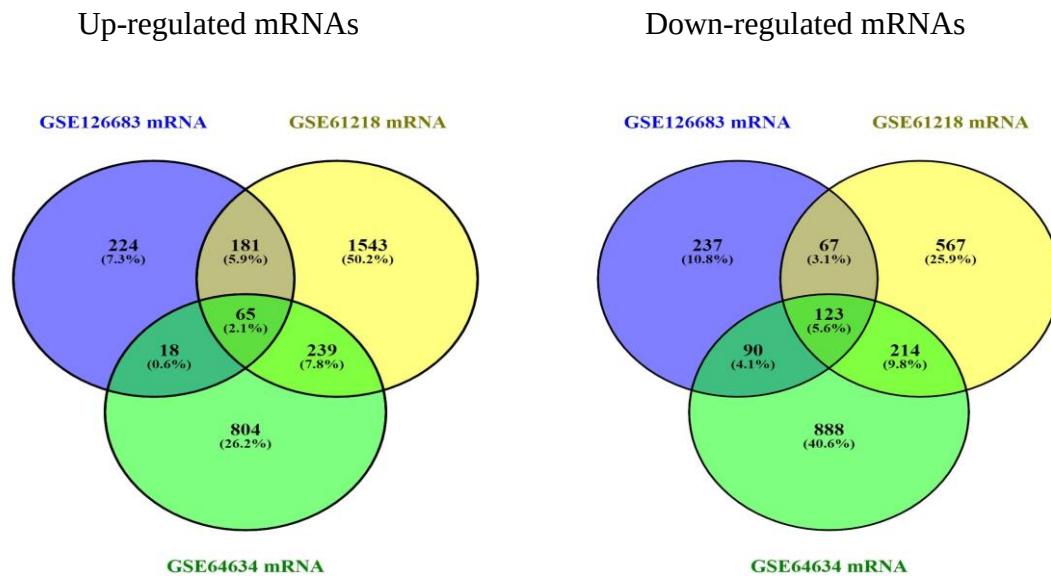


Figure 4.2: The Venn diagram of up-regulated and down-regulated mRNAs between the three mRNA datasets (GSE64634, GSE126683 and GSE61218). The Venn diagrams showed 65 common up-regulated and 123 common down-regulated mRNAs between the three datasets.

Table 4.3: The lists of common up-regulated and down-regulated mRNAs between the datasets GSE95166, GSE126683 and GSE61218.

Expression	Overlapped DEGs
Up-regulated	AHCY, RFC5, ENO1, CHI3L1, OASL, BRIP1, KIF23, NDC80, MSR1, GINS1, CHCHD3, IFIH1, NEIL3, PLK4, GINS4, ATAD2, CENPN, FANCI, TPX2, BIRC5, PBK, CLEC7A, CXCL10, CCNE2, BRCA1, WHSC1, TACC3, MAGOHB, TYMS, CDCA8, KPNA2, CENPA, NLRP7, DMD, PSAT1, AURKA, EXO1, PRC1, MCM10, KIF14, NUF2, CENPW, MKI67, DEPDC1, CENPI, RFC3, RAD51, RAN, CEP152, CDKN3, ASPM, KRT16, ICAM1, OLA1, C1QB, ULBP2,

Table 4.3 continued

	CENPF, RAD54B, DSCC1, KIF18A, OIP5, SSTR2, BUB1, MAD2L1, MCM7
Down-regulated	SLC22A4, SPA17, C11orf63, RSPH4A, SLC44A4, ZMYND10, OSCP1, DNAH6, BBS9, IQCA1, CP, C11orf52, TMEM45B, VILL, KIF6, TPPP3, RSPH1, TMEM190, C1orf115, SPATA18, EFHB, DNALI1, C9orf24, ARMC3, IQCD, RRAD, CHP2, DYNLRB2, AGR3, DNAH12, KCNRG, MAK, SPAG6, LRRC34, CASC1, LRRC23, TTC29, VSIG2, C10orf107, ENKUR, AKAP14, TEK13, FBXL13, MORN2, NEK10, CXCL17, CCDC81, EYA4, C4orf22, MSMB, GULP1, RGS22, SPATA17, FANK1, SELENBP1, WDR78, LEPR, C9orf116, SPATA6, RIBC1, C11orf88, C9orf135, TMEM71, DNAAF1, C4orf47, NQO1, STEAP4, TMC5, IL33, EYA1, DNAH2, ZBBX, NR2F2, BCAS1, DYNC2H1, C5orf49, CCDC153, SORBS2, SRPX, EFCAB1, C11orf70, CCDC113, LRTOMT, CAPS, NEK11, ADH1C, LRRC10B, NEK5, PPP1R36, TTC9, TEK11, PIGR, ST6GALNAC1, TTC25, LRRC46, SLPI, WFDC2, WDR63, DNAH5, TMEM232, LRRC6, CH25H, ADRA2A, C11orf16, SYTL2, PLA2G10, CRYM, CES1, TSPAN1, WDR54, DHRS9, SPAG16, MACROD2, SNTN, ADH7, SPEF2, EBF1, FAM153A, RSPH10B, MORN5, AKR1C3, PRSS23, CCDC65

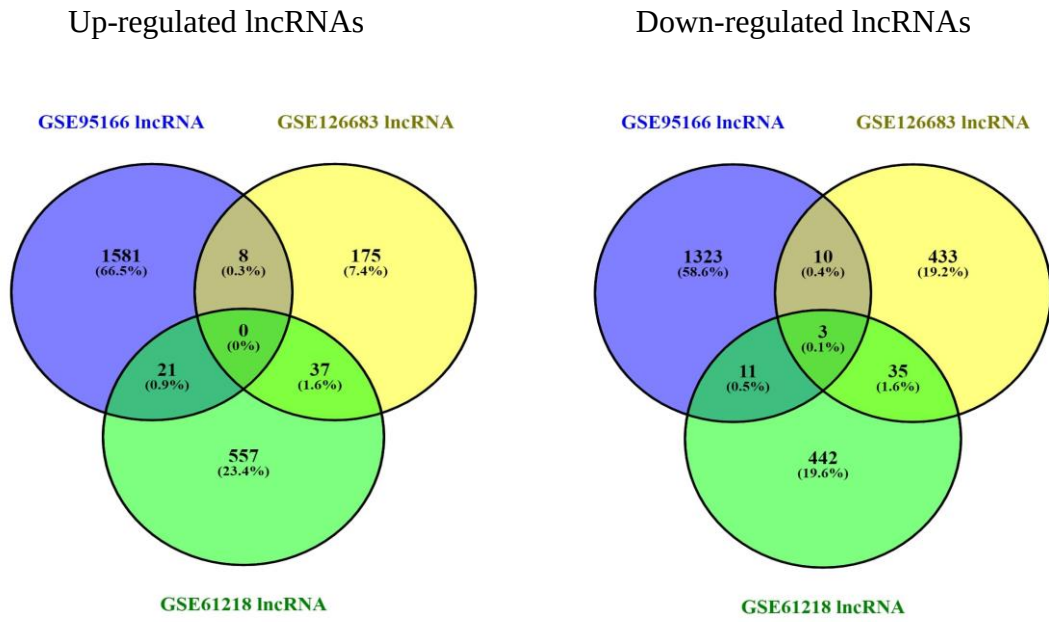


Figure 4.3: The Venn diagram of up-regulated and down-regulated lncRNAs between the three lncRNA datasets (GSE95166, GSE126683 and GSE61218). The Venn diagrams showed 0 common up-regulated lncRNAs and three common down-regulated lncRNAs between the three datasets.

4.3 Target mRNA and Target lncRNA Interaction with DE miRNA

4.3.1 DE miRNA-Target mRNA Interaction

The lists of target mRNAs for both up-regulated and down-regulated miRNAs (GSE70970) curated from miRwalk 2.0 and miRTarBase were overlapped with the common differentially expressed genes (DEGs) from the three GEO datasets (GSE64634, GSE126683 and GSE61218) by using a Venn diagram. The lists of the common mRNAs were tabulated in Table 4.4 and Table 4.5.

Table 4.4: Target mRNAs of up-regulated DE miRNAs which were overlapped with the common down-regulated GEO DEGs from the three mRNA datasets (GSE64634, GSE126683 and GSE61218).

Up-regulated DE miRNAs	Overlapped target mRNAs with the GEO Down-regulated DEGs
hsa-miR-1246	EFHB, BCAS1, EYA1, NEK5
hsa-miR-494-3p	MAK, LRRC34, GULP1, ZBBX, EFCAB1, C11orf16, SRPX, NEK5, CH25H
hsa-miR-320a	IQCA1 KIF6, DNALI1, CHP2, MAK, TTC29, FBXL13, LEPR, NQO1, DYNC2H1
hsa-miR-630	CASC1, EFCAB1, C11orf70
hsa-miR-93-3p	CHP2, LRRC34
hsa-miR-93-5p	C11orf63, TMEM45B, KIF6, MAK, FBXL13, EYA4, TMC5, EYA1, BCAS1, SORBS2, C11orf70, TTC9, SPEF2
hsa-miR-493-3p	KCNRG, MAK, EFCAB1, C11orf88
hsa-miR-1290	BBS9, MAK, TTC29, TMC5, BCAS1, NEK11, ADH1C, NEK5
hsa-miR-1973	BCAS1
hsa-miR-21-3p	KIF6, MORN5, CH25H
hsa-miR-21-5p	FBXL13, SPEF2, C11orf70, ST6GALNAC1
hsa-miR-15b-5p	OSCP1, BCAS1, EYA1
hsa-miR-2117	KIF6, LRRC23, STEAP4, FAM153A, PRSS23
hsa-miR-25-3p	ENKUR, GULP1, CCDC113, SNTN
hsa-miR-25-5p	LRTOMT, PIGR
hsa-miR-155-3p	NEK11, CCDC65
hsa-miR-155-5p	SYTL2, SPEF2, AKR1C3, DYNC2H1
hsa-miR-378a-3p	SPA17, CRYM, LEPR, LRRC6

Table 4.4 continued

hsa-miR-378a-5p	SLC44A4, TMEM45B, RRAD, CASC1, FBXL13, TMC5, BCAS1, LRTOMT
hsa-miR-708-3p	IQCA1, SPAG16
hsa-miR-1246	EFHB, BCAS1,EYA1,NEK5
hsa-miR-494-3p	MAK, LRRC34, GULP1, ZBBX, EFCAB1, C11orf16, SRPX, NEK5, CH25H
hsa-miR-320a	IQCA1 KIF6, DNALI1, CHP2, MAK, TTC29, FBXL13, LEPR, NQO1, DYNC2H1
hsa-miR-630	CASC1, EFCAB1, C11orf70
hsa-miR-93-3p	CHP2, LRRC34
hsa-miR-93-5p	C11orf63, TMEM45B, KIF6, MAK, FBXL13, EYA4, TMC5, EYA1, BCAS1, SORBS2, C11orf70, TTC9, SPEF2
hsa-miR-493-3p	KCNRG, MAK, EFCAB1, C11orf88
hsa-miR-1290	BBS9, MAK, TTC29, TMC5, BCAS1, NEK11, ADH1C, NEK5
hsa-miR-1973	BCAS1
hsa-miR-21-3p	KIF6, MORN5, CH25H
hsa-miR-21-5p	FBXL13, SPEF2, C11orf70, ST6GALNAC1
hsa-miR-15b-5p	OSCP1, BCAS1, EYA1
hsa-miR-2117	KIF6, LRRC23, STEAP4, FAM153A, PRSS23
hsa-miR-25-3p	ENKUR, GULP1, CCDC113, SNTN
hsa-miR-25-5p	LRTOMT, PIGR
hsa-miR-155-3p	NEK11, CCDC65
hsa-miR-155-5p	SYTL2, SPEF2, AKR1C3, DYNC2H1
hsa-miR-378a-3p	SPA17, CRYM, LEPR, LRRC6

Table 4.4 continued

hsa-miR-378a-5p	SLC44A4, TMEM45B, RRAD, CASC1, FBXL13, TMC5, BCAS1, LRTOMT
hsa-miR-708-3p	IQCA1, SPAG16
hsa-miR-96-3p	MORN2, ZBBX, SORBS2, FAM153A
hsa-miR-96-5p	MAK, CCDC65
hsa-miR-490-3p	LRTOMT, PIGR
hsa-miR-106a-5p	C10orf107, EFCAB1, NEK11, SORBS2, DYNC2H1
hsa-miR-148b-3p	EYA4, STEAP4
hsa-miR-148b-5p	LEPR, EYA1
hsa-miR-191-3p	RRAD
hsa-miR-421	MAK, ENKUR
hsa-miR-130b-3p	RIBC1, STEAP4, NEK5, FAM153A
hsa-miR-130b-5p	SPATA6, DNAH2, C11orf70, SPEF2
hsa-miR-98-3p	IQCA1, SPATA18, MAK
hsa-miR-98-5p	SPATA18, RRAD, CCDC81, AKR1C3, LRTOMT, NQO1
hsa-miR-1305	SLC22A4, RSPH4A, MAK, C10orf107, GULP1, LEPR, C4orf47, NEK11, DNAH5, LRRC6, DHRS9, SPAG16, MACROD2, SPEF2
hsa-miR-146b-5p	CRYM, LRTOMT
hsa-miR-16-5p	OSCP1, TPPP3, EYA4, EYA1, BCAS1
hsa-miR-574-5p	SLC22A4, BCAS1, CCDC65, SPATA6
hsa-miR-660-3p	SPA17, VSIG2, BCAS1, TSPAN1, MACROD2, SORBS2
hsa-miR-660-5p	MAK, ENKUR, LEPR
hsa-miR-107	OSCP1, TPPP3, KCNRG, MAK, CASC1, ENKUR, CCDC81, BCAS1, CRYM

Table 4.4 continued

hsa-miR-663b	KIF6
hsa-miR-135b-5p	KIF6, DNALI1, KCNRG, GULP1, DNAH2, EBF1
hsa-miR-1275	LRTOMT
hsa-miR-211-3p	C1orf115, EFCAB1, CCDC113, CAPS, DHRS9
hsa-miR-211-5p	BBS9, DHRS9, SPAG16, SPEF2, EBF1

Table 4.5: Target mRNAs of down-regulated DE miRNAs which were overlapped with the common up-regulated GEO DEGs from the three mRNA datasets (GSE64634, GSE126683 and GSE61218).

Down-regulated DE miRNAs	Overlapped target mRNAs with the GEO Up-regulated DEGs
hsa-miR-375	PLK4, TYMS, EXO1, CENPF
hsa-miR-34c-3p	PLK4
hsa-miR-34c-5p	PLK4, CCNE2, ULBP2
hsa-miR-145-3p	BUB1
hsa-miR-145-5p	PSAT1
hsa-miR-216a-5p	TPX2, KIF14, DEPDC1
hsa-miR-485-3p	WHSC1, KPNA2, PSAT1, CENPI, OLA1
hsa-miR-203a-3p	CHCHD3, GINS4, BIRC5, TYMS, PSAT1, CENPI, RAN, ICAM1
hsa-miR-203a-5p	CHCHD3, GINS4, TYMS, CENPI, ICAM1
hsa-miR-99b-3p	MSR1, CENPN, BRCA1
hsa-miR-423-3p	CHCHD3
hsa-miR-1297	BRCA1, KPNA2, PSAT1, DEPDC1, CDKN3, MAD2L1
hsa-miR-31-3p	AURKA, RAN

Table 4.5 continued

hsa-miR-31-5p	DMD, MKI67, DEPDC1, ICAM1
hsa-miR-497-3p	MSR1, KPNA2, DMD, RAN, BUB1
hsa-miR-497-5p	AHCY, KIF23, GINS4, BIRC5, CCNE2, WHSC1, PSAT1, CENPW, C1QB
hsa-miR-100-3p	CLEC7A
hsa-miR-100-5p	KPNA2
hsa-miR-125a-5p	ENO1, CHCHD3, MAGOHB, PRC1
hsa-miR-520e	GINS4, ATAD2, CENPN, PBK, BRCA1, WHSC1, KPNA2, KIF14, RAD51, CEP152, SSTR2, MCM7
hsa-miR-92b-3p	MSR1, NEIL3, AURKA, RFC3, RAD51, RAN, CEP152, ICAM1
hsa-miR-361-5p	CHCHD3, CENPN, TPX2
hsa-miR-30c-5p	AHCY, CHI3L1, BRIP1, BIRC5, CCNE2, DMD, CEP152, SSTR2
hsa-miR-570-5p	RFC5, BRIP1, MSR1, KIF14, NUF2, DEPDC1, OIP5
hsa-miR-125b-5p	CHCHD3, MAGOHB, PSAT1
hsa-miR-30d-3p	BRIP1, KIF23, FANCI, KPNA2
hsa-miR-30d-5p	AHCY, CHI3L1, BRIP1, CCNE2, DMD, SSTR2
hsa-miR-424-3p	CHCHD3, TPX2, BRCA1
hsa-miR-424-5p	KIF23, IFIH1, GINS4, BIRC5, CCNE2, WHSC1, PSAT1, CENPW, C1QB
hsa-miR-23b-3p	WHSC1, CDCA8, RAD51
hsa-miR-23b-5p	FANCI, TPX2
hsa-miR-150-5p	OASL, GINS1, BIRC5, CENPA, NLRP7
hsa-miR-200a-3p	CCNE2, PRC1, CEP152, CDKN3

Table 4.5 continued

hsa-miR-200a-5p	BIRC5, OLA1, OIP5
hsa-miR-342-3p	WHSC1, KPNA2, PSAT1, BUB1
hsa-let-7c-3p	GIN51
hsa-let-7c-5p	MSR1, FANCI, DMD
hsa-let-7b-3p	KIF23, DEPDC1
hsa-let-7b-5p	MSR1, BIRC5, TYMS, CDCA8, DMD, AURKA, MCM7
hsa-miR-34b-3p	BRIP1
hsa-miR-34b-5p	CCNE2, ICAM1, ULBP2
hsa-miR-2278	CLEC7A, CDCA8, KRT16

4.3.2 DE miRNA-Target lncRNA Interaction

The lists of target lncRNAs for up-regulated miRNAs (GSE70970) curated from four combined databases, including miRcode, DIANA-LncBase V2, DIANA-LncBase V3 and StarBase V2, were overlapped with common differentially expressed lncRNAs (DELs) from the three GEO lncRNAs datasets (GSE95166, GSE126683 and GSE61218) by using Venn diagrams. The lists of these mRNAs were tabulated in Table 4.6. However, there was no overlapping of target lncRNAs for down-regulated miRNAs with the GEO lncRNAs due to no common up-regulated lncRNAs in the datasets.

Table 4.6: Target lncRNA of up-regulated DE miRNAs which were overlapped with the common down-regulated GEO lncRNAs from the three lncRNAs datasets (GSE95166, GSE126683 and GSE61218).

Down-regulated DE miRNAs	Overlapped target lncRNAs with the GEO Up-regulated DELs
hsa-miR-1246	ARHGEF26-AS1
hsa-miR-320a-3p	CASC2, ARHGEF26-AS1
hsa-miR-320a-5p	CASC2
hsa-miR-93-5p	CASC2, TPTE2P1, ARHGEF26-AS1
hsa-miR-1290	ARHGEF26-AS1
hsa-miR-21-3p	CASC2
hsa-miR-21-5p	CASC2
hsa-miR-155-3p	CASC2
hsa-miR-155-5p	CASC2
hsa-miR-708-5p	CASC2, ARHGEF26-AS1
hsa-miR-96-5p	ARHGEF26-AS1
hsa-miR-106a-5p	CASC2
hsa-miR-148b-5p	CASC2, ARHGEF26-AS1
hsa-miR-130b-5p	CASC2
hsa-miR-98-5p	CASC2, TPTE2P1, ARHGEF26-AS1
hsa-miR-1305	CASC2, ARHGEF26-AS1
hsa-miR-16-5p	TPTE2P1
hsa-miR-107	CASC2, TPTE2P1
hsa-miR-135b-5p	CASC2
hsa-miR-211-5p	ARHGEF26-AS1

4.4 mRNA-miRNA-lncRNA Competing Endogenous RNA (ceRNA) Network

The constructed competing endogenous RNA (ceRNA) network comprised of 58 down-regulated mRNAs, 17 up-regulated miRNAs and 3 down-regulated lncRNAs. This network illustrated that these RNAs were regulating each other, and the deregulation in one of the RNAs could affect the expression level of other RNAs, causing NPC (Figure 4.4). Only one ceRNA was constructed due to no common up-regulated lncRNAs in the three GEO datasets (GSE95155, GSE126683 and GSE61218).

Enrichment analysis was done on the 58 down-regulated mRNAs, and these genes were revealed to be involved primarily in cilia assembly and motility. The isoforms of the RNAs were disregarded as there is no isoform information for all the RNAs provided in the microarray datasets curated from the GEO database. Besides, the downstream analysis to construct the protein-protein interaction (PPI) network using the STRING database also dismissed the nature of the protein isoforms (Szklarczyk et al., 2023).

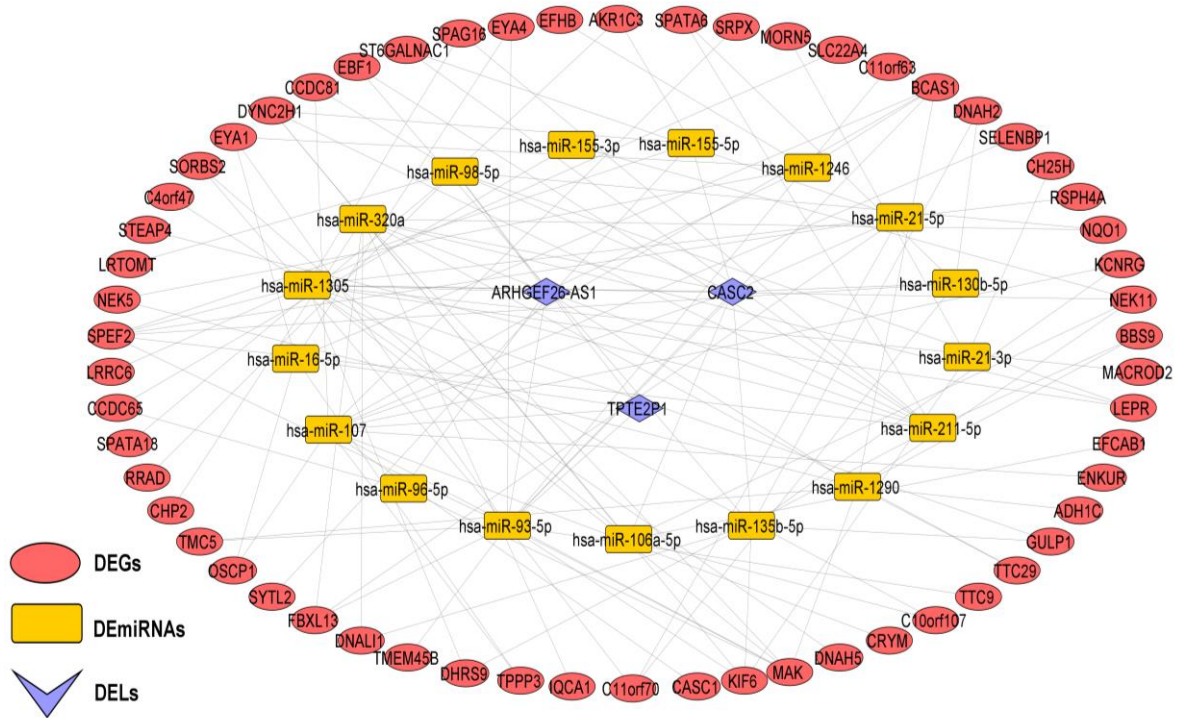


Figure 4.4: The lncRNA-miRNA-mRNA competing endogenous RNA (ceRNA) network. The ceRNA network consisted of 58 down-regulated mRNAs (red eclipse), 17 up-regulated miRNAs (yellow rectangle), and three down-regulated lncRNAs (purple diamond). It showed the interactions between the RNAs and the deregulation in one of the RNAs affecting the regulation of other RNAs interacting with it.

4.5 Epstein-Barr Virus miRNA-mRNA Interaction

4.5.1 Epstein-Barr Virus miRNAs-mRNAs Interaction

The ViRBase database revealed 5,411 predicted Epstein-Barr Virus (EBV) *Bam*H1 A transcript (BART) miRNA-mRNA interactions following the confidence score from 0.5 to 1.00. The overlapping of the predicted mRNAs with differentially expressed (DEGs) in the competing endogenous RNA (ceRNA) network using a Venn diagram revealed four common down-regulated DEGs which are EBF1, MACROD2, EYA1 and EYA4 (Figure 4.5).

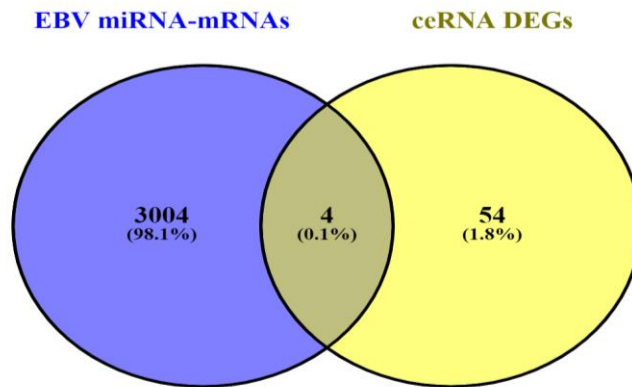


Figure 4.5: The Venn diagram of mRNAs between EBV miRNAs mRNAs with ceRNA DEGs. It exhibited four common mRNAs between the mRNAs targeted by BART miRNAs with the DEGs in the ceRNA, which were EBF1, MACROD2, EYA1 and EYA4.

4.5.2 Epstein-Barr Virus miRNA-mRNA Interaction Network

Four of the commonly predicted mRNAs were used to construct Epstein-Barr Virus (EBV) miRNA-mRNA network to visualize the regulatory mechanisms of *BamH1* A transcript (BART) miRNAs in down-regulating EBF1, MACROD2, EYA1 and EYA4, which inherently up-regulated miRNAs and down-regulated lncRNAs in EBV positive nasopharyngeal cancer (NPC) (Figure 4.6). These mRNAs were targeted by six BART miRNAs which were ebv-miR-BART19-3p, ebv-miR-BART15, ebv-miR-BART21-3p, ebv-miR-BART12-5p, ebv-miR-BART20-3p and ebv-miR-BART11-5p. Specifically, EYA4 was targeted by ebv-miR-BART19-3p and ebv-miR-BART15, EYA1 were targeted by ebv-miR-BART21-3p and ebv-miR-BART12-5p, MACROD2 were targeted by ebv-miR-BART20-3p, and EBF1 was targeted by ebv-miR-BART11-5p.

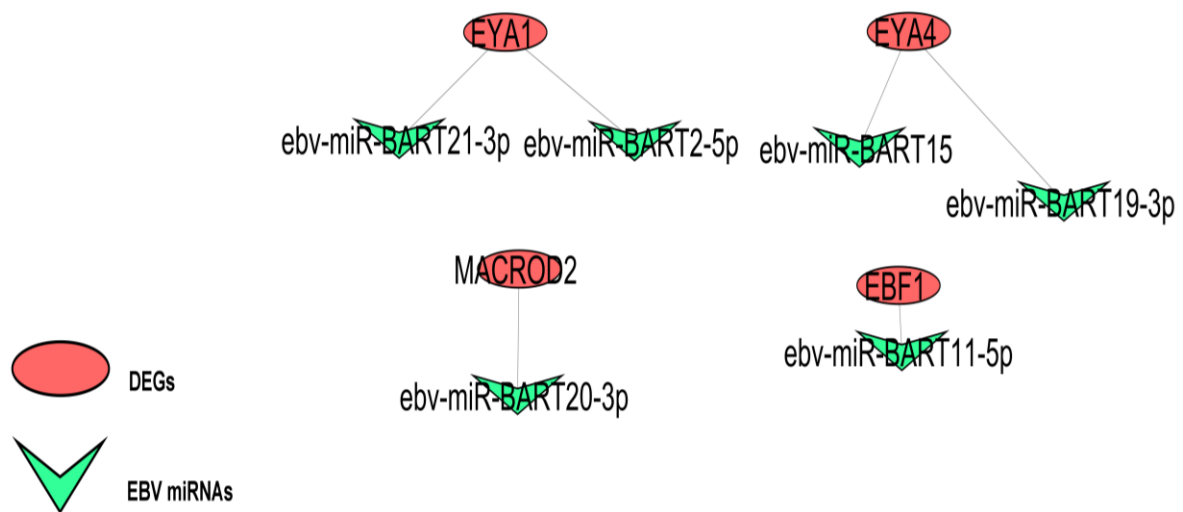


Figure 4.6: The interactions of ceRNA mRNAs with EBV BART miRNAs. The four mRNAs (EBF1, MACROD2, EYA1 and EYA4) which were indicated by the red eclipse revealed to be targeted and down-regulated by six BART miRNAs (ebv-miR-BART19-3p, ebv-miR-BART15, ebv-miR-BART21-3p, ebv-miR-BART12-5p, ebv-miR-BART20-3p and ebv-miR-BART11-5p) which were indicated by green V.

4.6 Protein-Protein Interaction (PPI)

4.6.1 Protein-Protein Interaction (PPI) Network

As indicated by the Search Tool for the Retrieval of Interacting Gene (STRING) database of the Cytoscape plug-in, the down-regulated common mRNAs in three of the GEO datasets (GSE64634, GSE126683 and GSE61218) exhibited 123 nodes and 179 edges. This was further demonstrated in the protein-protein interaction (PPI) network, where 77 of the down-regulated mRNAs interacted, as indicated by the grey lines out of all the 123 DEGs (Figure 4.7). Here, the proteins refer to the expressed proteins of the down-regulated mRNAs.

The node's degree distribution of the PPI network is based on the known and predicted interactions between the genes or proteins and the strength of the interactions evidence compared with Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway maps. This means that 77 of the down-regulated mRNAs interacting in the PPI network were, according to a combined score of more than 0.55, a medium or high confidence. Besides, the 77 down-regulated mRNAs interacting in the PPI network are suggested to share similar biological functions.

The remaining 46 down-regulated mRNAs at the bottom of the PPI network are suggested to have a lower combined score than 0.55, with less interaction evidence (Figure 4.7). Therefore, no grey lines connect them to the network's other mRNAs. Based on the STRING dataset, the distance between the genes/proteins in the PPI network is based on a confidence score, which is the predicted probability of two genes/protein products interacting curated from the KEGG database. Due to Cytoscape software's inability to read the interaction type retrieved from the STRING database plug in, the direct and indirect interactions between these proteins were retrieved from the STRING database website and depicted in Figure 14, Appendices section.

Moreover, based on further analysis by using Network Analyzer, the clustering coefficient of the PPI network is 0.426, indicating the proteins in the PPI network are less connected to each other as proteins which tend to cluster together is indicated by clustering coefficient closer to 1 (Wandora, 2008). Furthermore, the Network Analyzer revealed that the PPI network diameter is 9, which means the longest links for one protein connected to another protein is 9 links (edges) in the PPI network (Max Planck Institute for Informatics, 2018). The other topological properties of the PPI network are depicted in Figure 18, Appendices section.

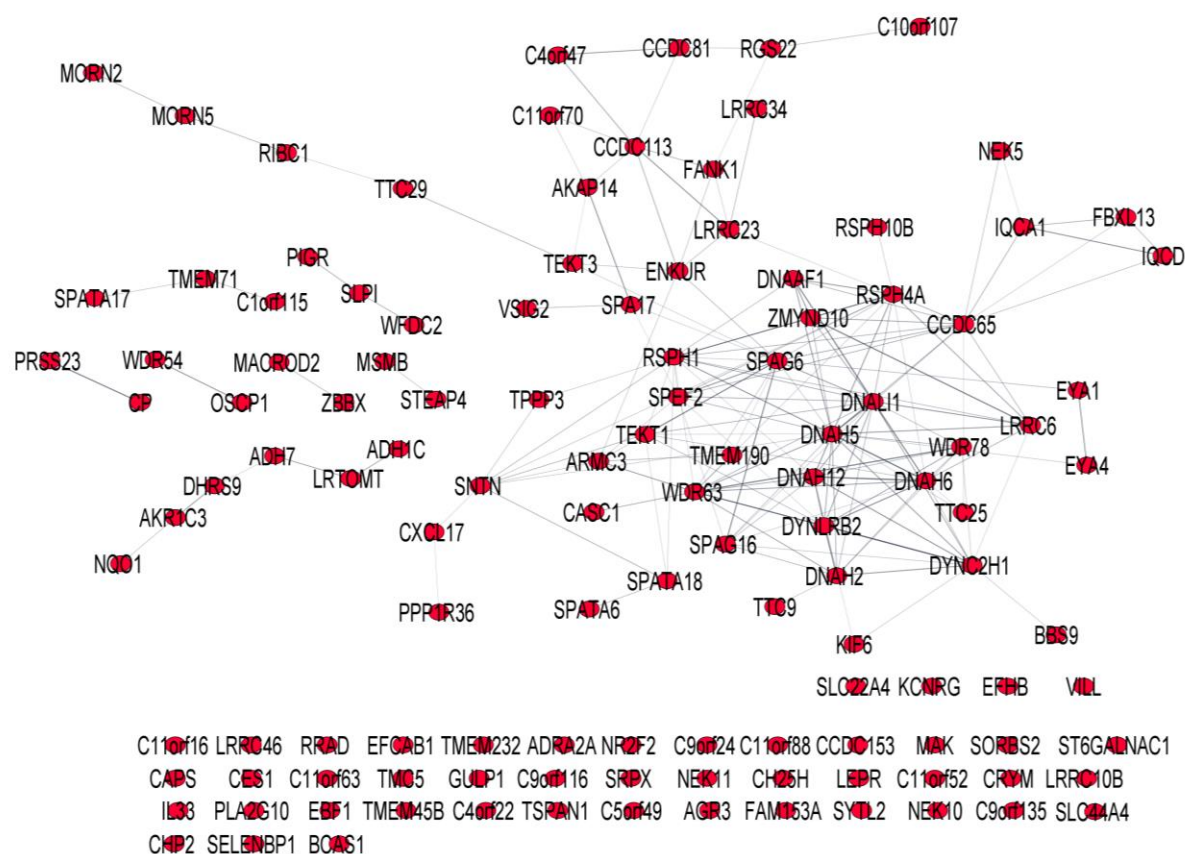


Figure 4.7: The protein-protein interaction (PPI) between the down-regulated common DEGs from the three GEO datasets (GSE64634, GSE126683 and GSE61218). constructed using the Cytoscape plug-in. The PPI exhibited 123 nodes and 179 edges, with 77 of the DEGs interacting with each other, represented by the red ellipse.

4.6.2 Hub Genes from The Protein-Protein Interaction (PPI)

After the top 20 differentially expressed genes (DEGs) with connectivity degree more than 5 were filtered from the PPI network, DNAH5 was ranked first with 21 connectivity degrees, and DNAAF1 was ranked last with seven connectivity degrees by CytoHubba (Figure 4.8 and Table 4.7). From Figure 4.8, DNAH5 has the highest number of DEGs interacting with it in the PPI which is 21 DEGs indicated by the red colour, whereas DNAAF1 has the lowest number of DEGs interacting with it in the PPI, which

was only 7 DEGs indicated by the yellow colour. Only CCDC113 was not connected to the other 19 DEGs.

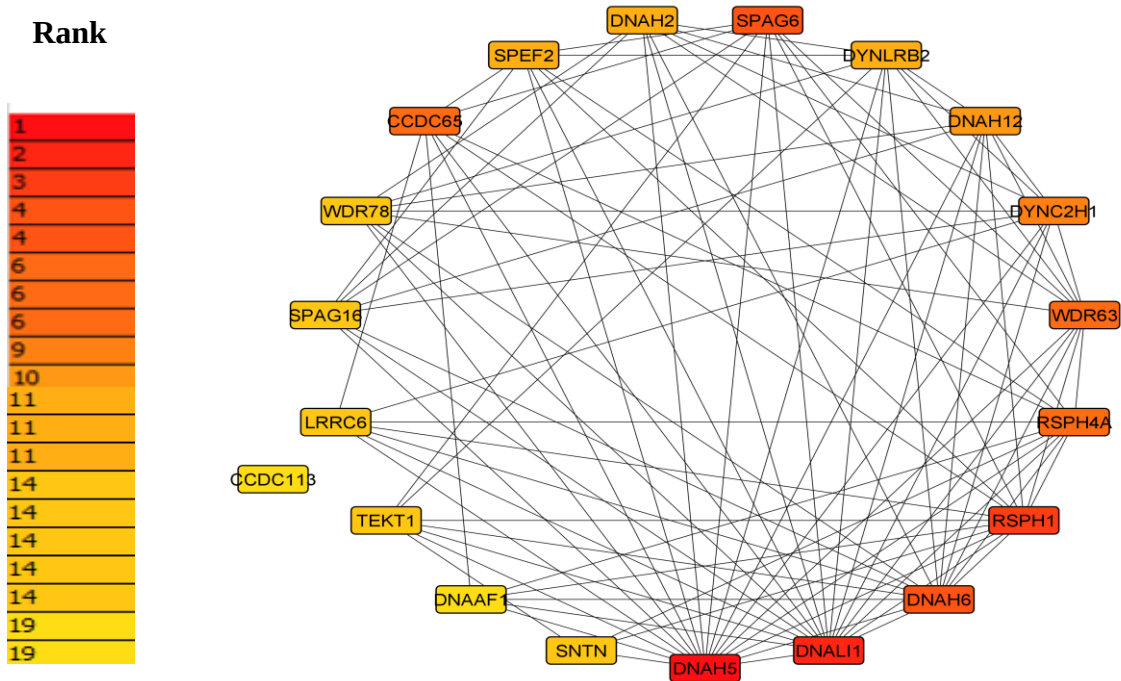


Figure 4.8: The top 20 DEGs with connectivity degree more than 5 from the PPI ranked according to connectivity degree by CytoHubba. DNAH5 was ranked 1st indicated by the red colour, and DNAAF1 was ranked last indicated by the yellow colour. The colour gradient with red indicated the connectivity degree ranks represented DEGs with high connectivity degree, and the colour scheme changes to orange and yellow as the connectivity degree decreases. The yellow represented the lowest connectivity degree.

Table 4.7: The top 20 DEGs with connectivity degree more than 5 from the PPI ranked according to connectivity degree by CytoHubba. The score showed the number of connections of each DEG with other DEGs in the PPI. DNAH5 was ranked 1st with 21 connections to other DEGs in the PPI, and in the contrary DNAAF1 was ranked last with seven connections to other DEGs in the PPI.

Top 20 DEGs in PPI ranked by Degree method		
Rank	DEGs	Score
1	DNAH5	21
2	DNALI1	18

Table 4.7 continued

3	RSPH1	16
4	DNAH6	15
4	SPAG6	15
6	WDR63	14
6	RSPH4A	14
6	CCDC65	14
9	DYNC2H1	12
10	DNAH12	11
11	SPEF2	10
11	DNAH2	10
11	DYNLRB2	10
14	TEKT1	8
14	WDR78	8
14	LRRC6	8
14	SPAG16	8
14	SNTN	8
9	CCDC113	7
9	DNAAF1	7

Moreover, Network Analyzer revealed that 18 of these Top 20 DEGs have a connectivity degree more than 5, excluding self-loop resulting from undirected network parameter. DEGs with a connectivity degree more than 5 are deemed hub genes due to the

ability to regulate the expression of multiple genes (Figure 4.9 and Table 4.8) (Yu et al., 2017).

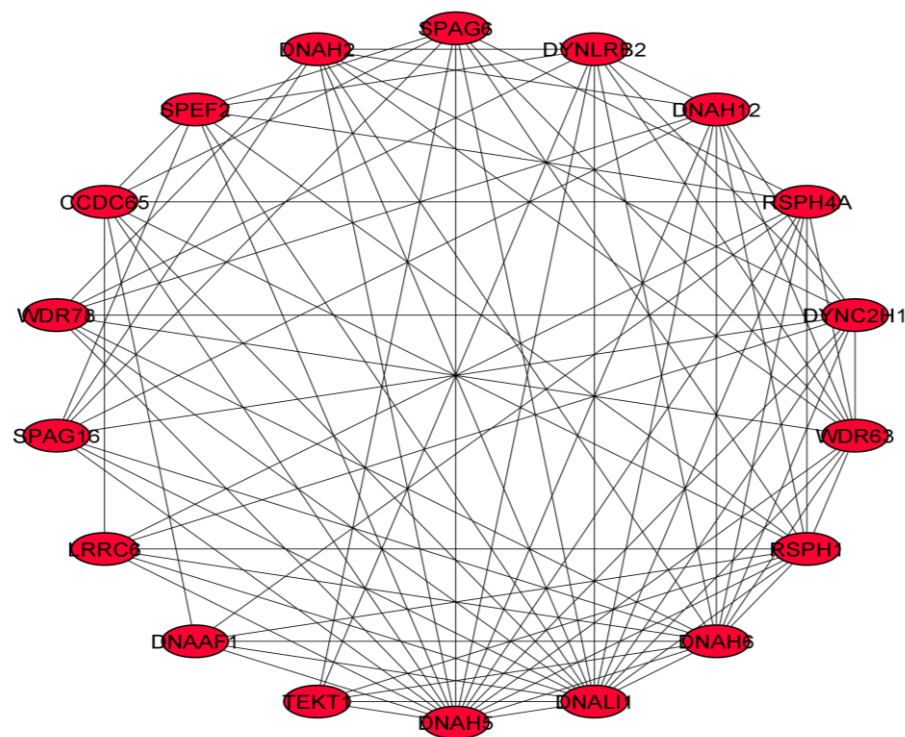


Figure 4.9: The DEGs with connectivity degree more than 5 amongst the Top 20 DEGs. It showed 18 DEGs of the top 20 DEGs were interacting with each other. These DEGs are deemed as hub genes due to their ability to regulate multiple genes, and their expressions are affecting the expressions of those DEGs interacting with them.

Table 4.8: The 18 DEGs with connectivity degree more than 5 excluding self-loop among the Top 20 DEGs. The connectivity degree showed the number of connections of each of the 18 DEGs amongst each other. It showed DNAH5 has the most connection with a connectivity degree of 18 with the other DEGs, whereas DNAAF1 and TEKT1 have the lowest connection with the other DEGs with only 6 interactions.

DEGs	Connectivity Degree
TEKT1	6
DNAAF1	6
LRRC6	7
SPAG16	8

Table 4.8 continued

CCDC65	8
WDR78	8
SPEF2	8
DNAH2	9
SPAG6	9
DYNC2H1	10
DYNLRB2	10
DNAH12	10
RSPH4A	11
WDR63	11
RSPH1	13
DNAH6	14
DNALI1	17
DNAH5	18

4.6.3 Gene Modules from The Protein-Protein Interaction (PPI)

Furthermore, the MCODE revealed two gene modules in the protein-protein interaction (PPI) network, which were Module 1 and Module 2, filtered out by using score greater than or equal to 4 and nodes greater than or equal to 4 in Cytoscape (Figure 4.10 and Table 4.9). Module 1 consisted of DEGs which were the hub genes but not Module 2. The genes in both modules are neighboring genes packed closely together in the PPI network and are highly interconnected with each other compared to other genes in the PPI network. The proximity or distance between the genes in PPI is linked to the high

confidence score of the interactions or strength of data support. This means the genes shared similar biological functions.

All DEGs in each module are closely related and encoded proteins with similar functions, which are proper cilia functions and motility. The down-regulation in these modules is suggested to cause NPC by affecting the nasal ciliary motor of nasopharynx cells for proper flow and clearance of fluids and particles in the nasopharynx (Ringers et al., 2020).

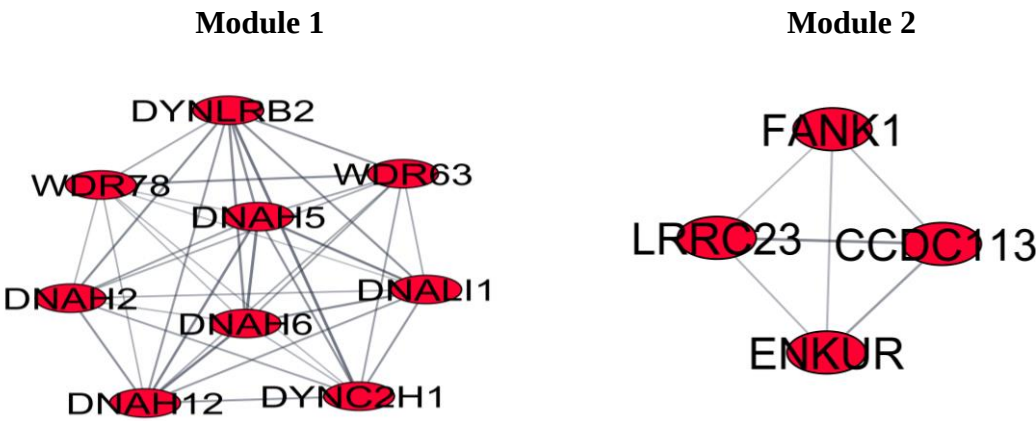


Figure 4.10: The two modules in the PPI network with scores greater than or equal to 4 and nodes greater than or equal to 4. In Module 1, all the DEGs are hub genes with connectivity degree more than 5, whereas, in Module 2, all the DEGs are not hub genes. Both modules are essential in proper cilia functions and motility.

Table 4.9: The list of DEGs in Module 1 and Module 2, extracted from the PPI network. Module 1 consisted of 9 DEGs with 36 interactions between them, with the highest score of 9.000, whereas Module 2 consisted of 4 DEGs with only six interactions between them, thus a low score of 4.000.

Module	Score	Nodes	Edges	DEGs
1	9.000	9	36	WDR78, DNAH2, DNAH12, DYNLRB2, DNAH6, DNAH5, WDR63, DNALI1, DYNC2H1
2	4.000	4	6	FANK1, LRRC23, CCDC113, ENKUR

4.7 Transcription Factors (TF)-mRNAs Interaction

4.7.1 Transcription Factors (TF)-mRNAs Network

The TFcheckpoint database yielded 1,021 transcription factors (TFs), and the overlapping of these TFs with commonly down-regulated mRNAs from three of the GEO datasets (GSE64634, GSE126683 and GSE61218) by using the Venn diagram revealed 2 common mRNAs which were EBF1 and NR2F2 (Figure 4.11). However, when visualized in Cytoscape, both of these TFs were not connected to the other DEGs in the protein-protein interaction (PPI) network or each other hence no Transcription Factors (TF)-mRNAs network was able to be constructed (Figure 4.12).

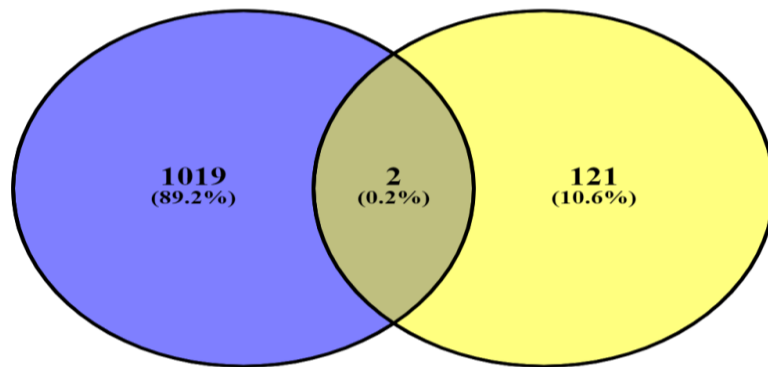


Figure 4.11: The Venn diagram of transcription factors (TFs) from the TFcheckpoint database overlapped with common down-regulated mRNAs from the three GEO datasets (GSE64634, GSE126683 and GSE61218). Two TFs were found amongst the 123 GEO down-regulated mRNAs, which are EBF1 and NR2F2.

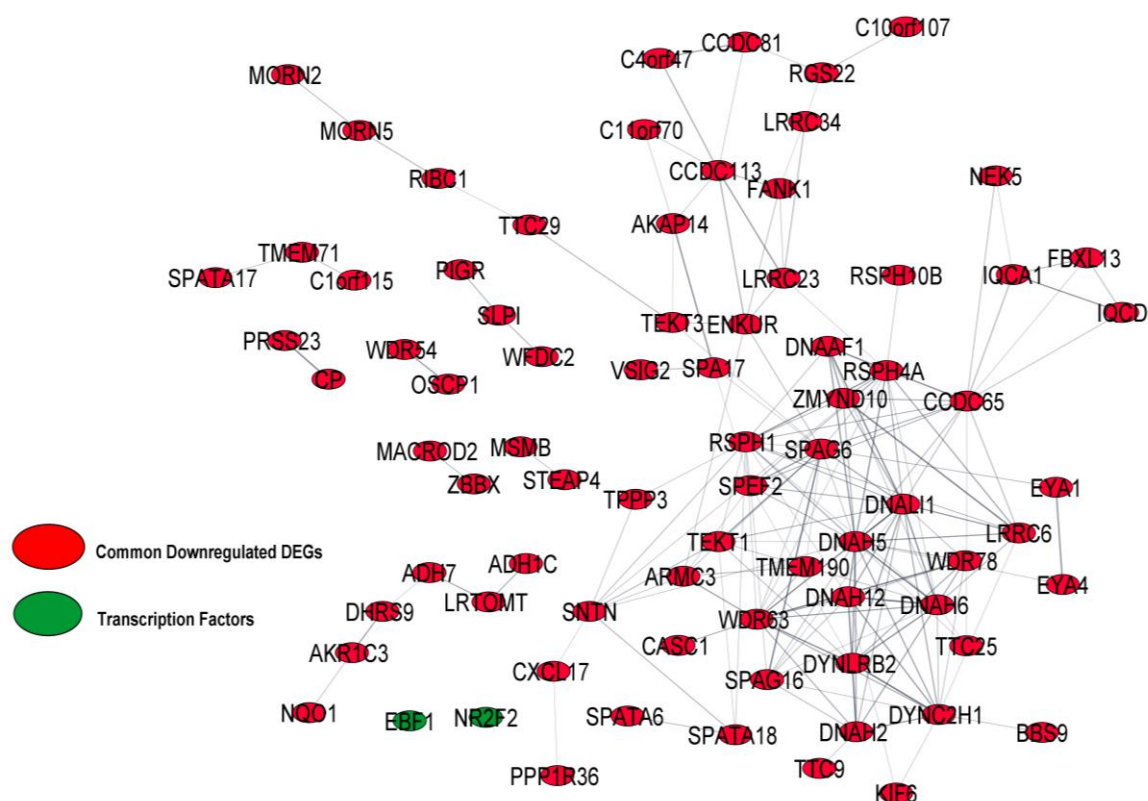


Figure 4.12: The TFs, EBF1 and NR2F2 in the PPI network. Both TFs indicated by the green eclipse do not interact with other down-regulated mRNAs indicated by the red eclipse and they were also not interacting with each other.

4.8 Cross-Regulatory Network (ceRNA and EBV miRNAs-mRNAs network)

The cross-regulatory network which were constructed by merging the two networks, which were the competing endogenous RNA (ceRNA) network and EBV miRNAs-mRNAs network, demonstrated the regulatory mechanisms of six EBV BART miRNAs (ebv-miR-BART21-3p, ebv-miR-BART19-3p, ebv-miR-BART15, ebv-miR-BART2-5p, ebv-miR-BART20-3p and ebv-miR-BART11-5p) targeting and down-regulate four DEGs (EYA4, EYA1, EBF1 and MACROD2) which then causes the up-regulation of six host miRNAs (hsa-miR-1246, hsa-miR-93-5p, hsa-miR-16-5p, hsa-miR-135b-5p, hsa-miR-211-5p and hsa-miR-1305). This was due to the limited availability of mRNAs to be

repressed or degraded where this also caused a down-regulation of the three lncRNAs (CASC2, TPTE2P1 and ARHGEF26-AS1) because no miRNAs were sponged in EBV-positive NPC (Figure 4.13 and Table 4.10). One of the BART miRNAs was targeting and down-regulating the transcription factor EBF1. Nevertheless, EBF1 was not interacting with any other DEGs in the network. Specifically, The MACROD was targeted and down-regulated by BART20-3p; EYA4 was targeted and down-regulated by BART15 and BART19-3p; EYA1 was targeted and down-regulated by BART2-5p and BART21-3p. This in turn caused up-regulation of six miRNAs (hsa-miR-1246, hsa-miR-93-5p, hsa-miR-16-5p, hsa-miR-135b-5p, hsa-miR-211-5p and hsa-miR-1305) and down-regulation of three lncRNAs (CASC2, TPTE2P1 and ARHGEF26-AS1).

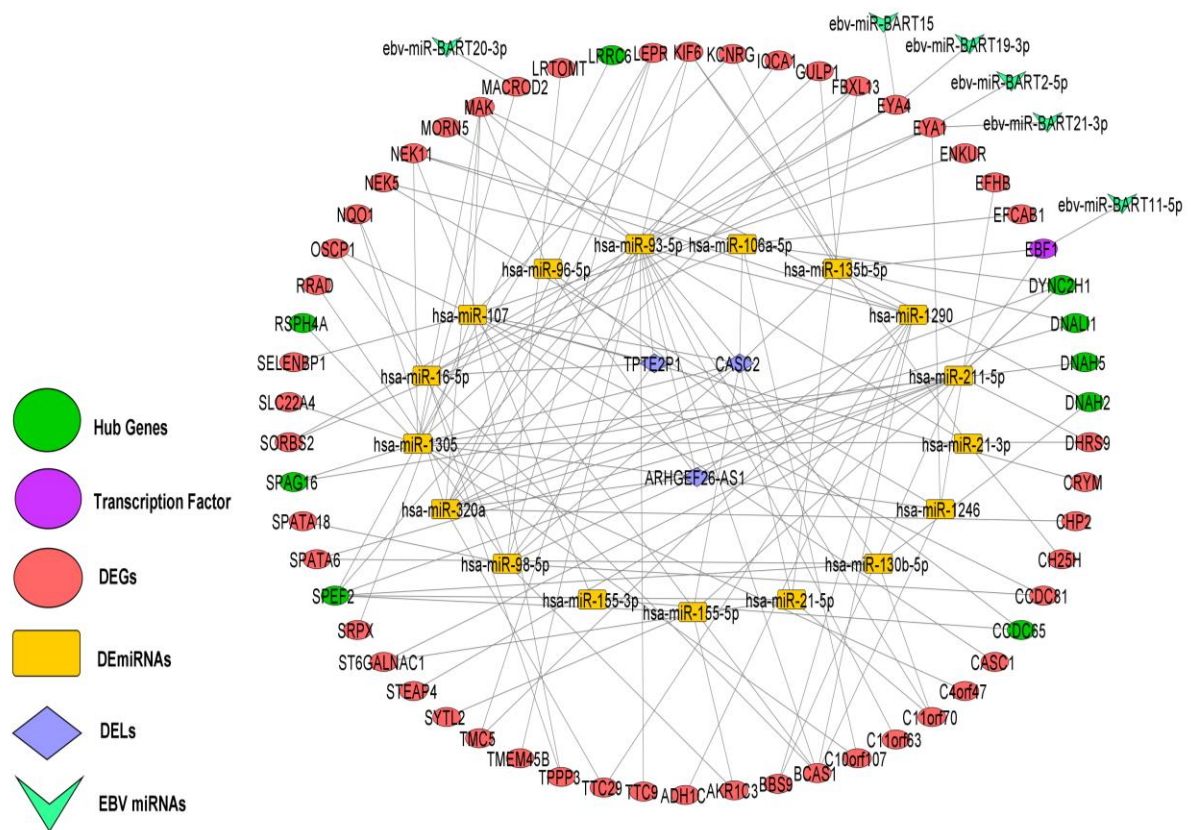


Figure 4.13: The cross-regulatory network. It was merged from competing endogenous RNA network (ceRNA) and EBV miRNAs-mRNAs network. Only BART11-5p was targeting and down-regulating TF, which was EBF1 (purple eclipse), and none of the hub genes (green eclipse) was targeted by BART miRNAs.

Table 4.10: The interaction between BART miRNAs and RNAs in the cross-regulatory network. The down-regulation of the four mRNAs (EYA4, EYA1, EBF1 and MACROD2) due to BART miRNAs caused the up-regulation of the six miRNAs. This also caused the down-regulation of the three lncRNAs.

Up-regulated EBV BART miRNAs	Down-regulated mRNAs	Up-regulated miRNAs	Down-regulated lncRNAs
EBV-miR-BART11-5p	EBF1	hsa-miR-211-5p	ARHGEF26-AS1
		hsa-miR-135b-5p	CASC2
EBV-miR-BART 20-3p	MACROD2	hsa-miR-1305	CASC2 and ARHGEF26-AS1

Table 4.10 continued

EBV-miR-BART21-3p and EBV-miR-BART2- 5p	EYA1	hsa-miR-1246	ARHGEF26-AS1
		hsa-miR-93-5p	ARHGEF26-AS1, CASC2 and TPTE2P1
EBV-miR-BART15 and EBV-miR-BART19-3p	EYA4	hsa-miR-93-5p	ARHGEF26-AS1, CASC2 and TPTE2P1
		hsa-miR-16-5p	TPTE2P1

4.9 Functional and Pathway Enrichment Analysis

The Gene Ontology (GO) enrichment results for the four differentially expressed genes (DEGs) (EYA4, EYA1, EBF1 and MACROD2), which were targeted by the six BART miRNAs, were enriched in 11 GO terms following the cut-off value of P-value < 0.05. These terms were 8 GO terms for Biological Processes (BP), 1 GO term for Cellular Component (CC), and 2 GO terms for Molecular Function (BF) (Figure 4.14). The Biological Processes (BP) revealed these DEGs were involved in regulating DNA repair and apoptosis, and due to down-regulation by BART miRNAs, these DEGs were predicted to be unable to perform their functions, leading to tumour cell proliferation and NPC. Besides, the Molecular Function (BF) exhibited that the proteins encoded by these DEGs were involved in DNA repair and surveillance of immune systems. All of these DEGs were located in the nucleus of the cells, indicated by the Cellular Component (CC). Nonetheless, no result was yielded for KEGG pathway enrichment analysis.

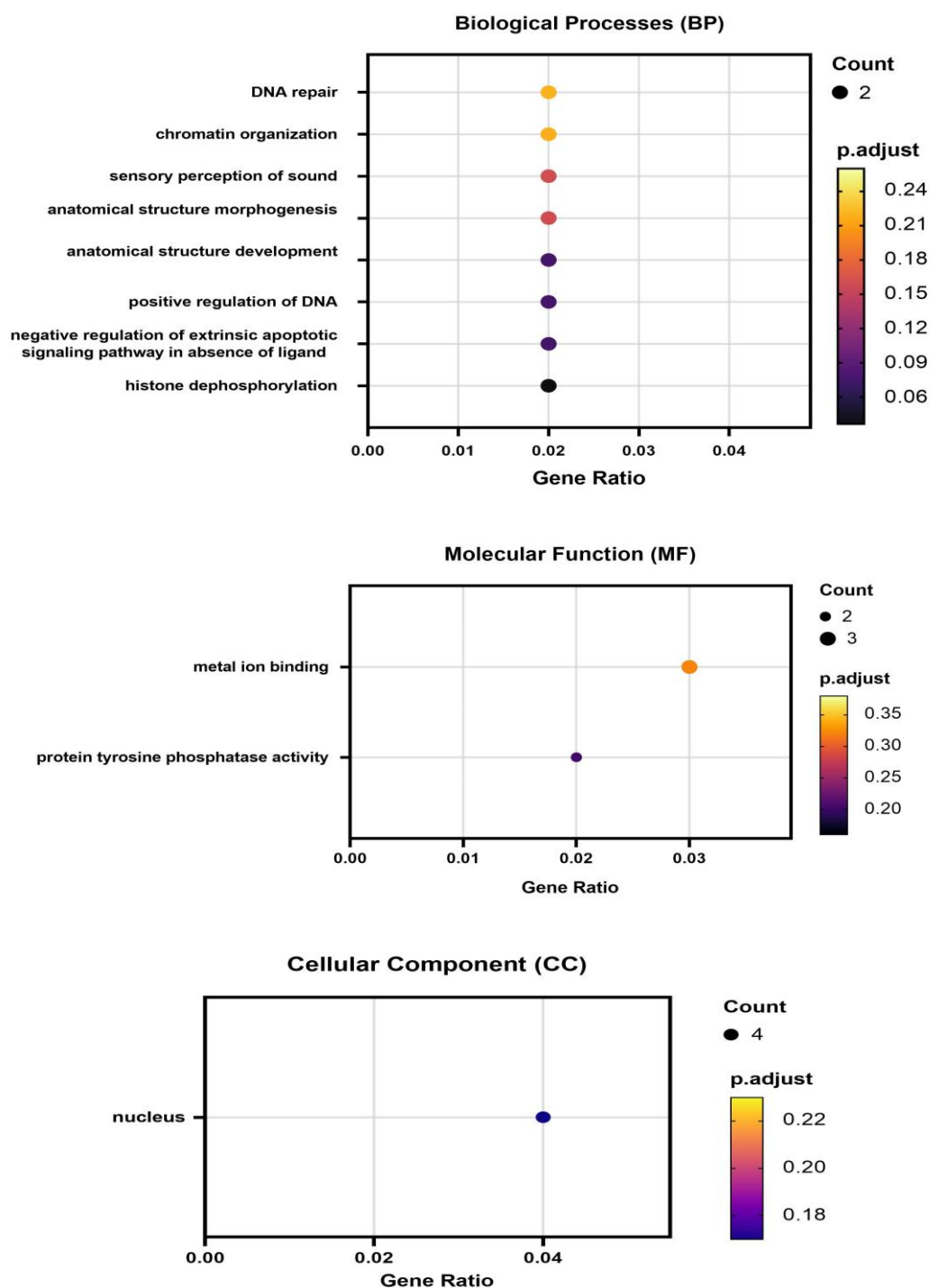


Figure 4.14: The Gene Ontology (GO) enrichment result for EYA4, EYA1, EBF1 and MACROD2 by using DAVID. The count means the total number of DEGs involved in each of the GO terms. Only EYA4 and EYA1 were involved in all the Biological Processes (BP). For Molecular Functions (MF), only EYA4 and EYA1 were involved in protein tyrosine phosphatase activity, whereas in metal ion binding EYA4, EYA1 and EBF1 were involved. All the four DEGs were located in nucleus as indicated by the Cellular Component (CC).

4.10 Survival Curve

The Kaplan-Meier survival curve for EYA1, EYA4 and MACROD2 revealed an inconsistent positive or negative correlation with overall survival (OS) except for EBF1 following P-value <0.05 and FDR $> 50\%$ but less than 100% (Figure 4.15). The OS curves for EYA1 and EYA4 showed a negative correlation with patients' OS; eventually, the curves changed to a positive correlation with patients' OS. On the contrary, both EYA1 and EYA2 exhibited a negative correlation with patients' OS until approaching 100 months the curves changed to positively correlated with patients' OS. This was supported by HR > 1 , which interpreted up-regulated group patients as having a low survival than down-regulated patients and are 1.34 and 1.33 times more likely to die compared to the down-regulated groups in EYA1 and EYA4, respectively.

On the contrary, the OS curves for MACROD2 exhibited a positive correlation with patients' OS and eventually, the curves shifted to negatively correlated with patients' OS. Nevertheless, EBF1 exhibited a positive correlation with patients' OS from the beginning until the end of the timeline. The hazard ratio (HR) < 1 indicated the patients in up-regulated group have a higher chances of survival of 0.71 times more likely to survive compared to down-regulated patients. Similarly, MACROD2 exhibited HR < 1 with 0.66 times more likely to survive compared to down-regulated patients. However, after 150 months the OS curves shifted from positively correlated with patients' OS to negatively correlated with censoring. Similarly, MACROD2 exhibited HR < 1 with 0.66 times more likely to survive compared to down-regulated patients. However, after 150 months, the OS curves shifted from positively correlated with patients' OS to negatively correlated with censoring.

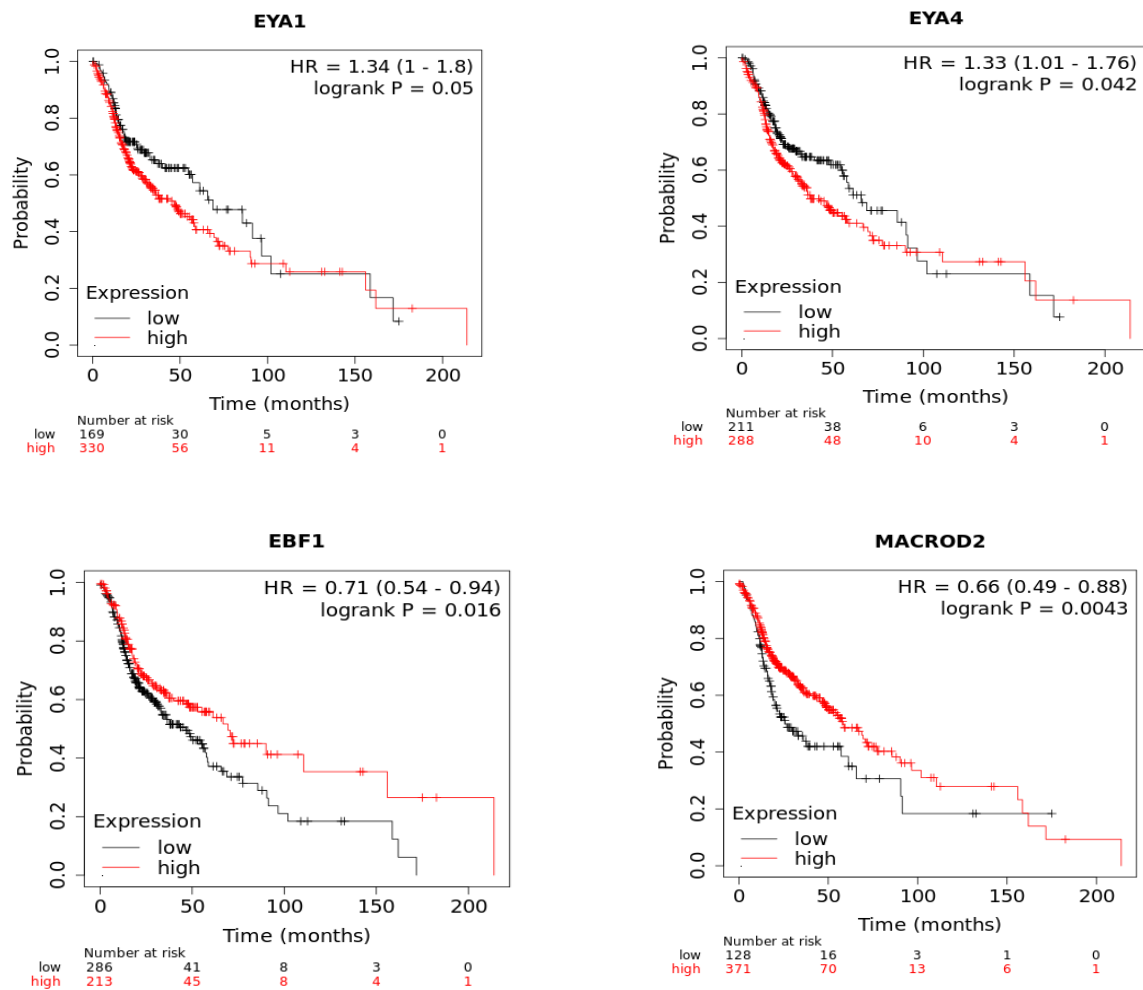


Figure 4.15: The Kaplan Meier survival curve for EYA1, EYA4, EBF1 and MACROD2. The Y axis represented the probability of surviving, and the X axis represented the duration of the study or survival.

CHAPTER 5

DISCUSSION

5.1 Competing Endogenous RNA (CeRNA) Network

In recent years, the deregulation of messenger RNAs (mRNAs), microRNAs (miRNAs) and long non-coding RNAs (lncRNAs) has been linked to tumorigenesis (Lee & Young, 2013). In this current study, a competing endogenous RNAs (ceRNA) network was constructed to predict the interaction between mRNAs, miRNAs and lncRNAs in EBV-positive nasopharyngeal cancer (NPC). This step is crucial to construct the cross-regulatory network to identify the candidate RNAs deregulated by EBV BART miRNAs. The interaction showed that RNAs are controlling the expression levels of each other. Generally, lncRNA controls mRNA expression by binding to a miRNA which is also targeted by a mRNA because both the lncRNA and mRNA have similar miRNA binding sites. This follows the competing endogenous RNAs (ceRNA) hypothesis that states lncRNA and mRNAs regulate each other by competing for the binding site on miRNAs known as miRNA recognition elements (MREs) through partially complementary sequences (Salmena et al., 2011).

Based on the ceRNA network constructed (Figure 4.4), it was suggested that three lncRNAs indirectly deregulate 58 mRNAs via 17 miRNAs targeted by the lncRNAs. The lncRNAs were suggested to be down-regulated and caused the down-regulation of the mRNAs without taking into consideration the presence of EBV BART miRNAs. In a previous study conducted by Xu et al. (2020), their ceRNA network revealed 132 lncRNAs regulate 2,654 mRNAs via 565 miRNAs in NPC using datasets submitted in the GEO database before the year 2019. Comparing our ceRNA network with their ceRNA network,

we shared 14 up-regulated miRNAs with them and shared no similar lncRNAs and mRNAs. These miRNAs were hsa-miR-98-5p, hsa-miR-320a, hsa-miR-135b-5p, hsa-miR-1305, hsa-miR-106a-5p, hsa-miR-211-5p, hsa-miR-1290, hsa-miR-93-5p, hsa-miR-130b-5p, hsa-miR-16-5p, hsa-miR-1246, hsa-miR-21-3p, hsa-miR-155-3p, and hsa-miR-107. Thus, based on the results obtained, these additional mRNAs, miRNAs, and lncRNAs are potential biomarker candidates in NPC tumorigenesis.

The interaction between all the mRNAs with miRNAs and the interactions between miRNAs with lncRNAs in the ceRNA network were unavailable in NPC and other cancers. Indicating these interactions are curated from predicted databases and require further validation. Nevertheless, hsa-miR-106a-5p overexpression was observed to target and down-regulate other mRNA in NPC. This was demonstrated by Li et al. (2021), where hsa-miR-106a-5p was up-regulated in NPC tumor tissue, where it was validated to target and down-regulate ARNT2 in the CNE1 NPC cell line. Besides, for hsa-miR-155-3p and 5p, unspecific hsa-miR-155 interaction with AKT in the CNE-1 NPC cell line was reported by Wu et al. (2019), where they are inversely expressed. Besides, Ou et al. (2014) also mentioned unspecific has-miR-21 overexpression in NPC cell lines and tissue samples by targeting and down-regulating PTEN for both has-miR-21-3p and 5p.

Interestingly, hsa-miR-96-5p, hsa-miR-320a, and hsa-miR-98-5p were down-regulated in NPC tissue samples and cell lines contrary to our ceRNA network. For hsa-miR-96-5p, Luo et al. (2020) reported that it was down-regulated in NPC tissue sample and cell line, leading to up-regulation of its targeted mRNA, CDK1. Similarly, Qi et al. (2014) demonstrated that hsa-miR-320a was down-regulated in NPC tumor samples and cell lines (CNE1 and C-666-1), leading to the up-regulation of its targeted mRNA, BMI-1. Utilizing

similar NPC cell lines, Xing et al. (2021) also demonstrated that the down-regulation of hsa-miR-320a in NPC leads to up-regulation of SOX4 due to the overexpression of TMPO-AS1 lncRNA. In addition, Li et al. (2021) also reported that hsa-miR-98-5p was down-regulated in NPC due to CDKN2B-AS1 lncRNA in NPC tissue and cell lines, where this leads to the up-regulation of hsa-miR-98-5p mRNA target, which was E2F2. Regardless, previous studies also exhibited the up-regulation of these miRNAs in tissue samples and cell lines leading to the carcinogenesis of ovarian cancer and other cancers by down-regulating its target mRNAs (Wang et al., 2018; Liu et al., 2019; Huang et al., 2021). This indicates that these miRNAs can act similarly to tumor suppressors when up-regulated and oncogenes when down-regulated.

On the contrary, hsa-miR-130b-5p, hsa-miR-107, and hsa-miR-1290 were reported to be up-regulated in other cancers. The has-miR-130b-5p was observed to be overexpressed in gastric cancer (GC) cell lines where it targets and down-regulates RASAL1, according to Chen et al. (2018). For hsa-miR-107, Wang et al. (2016) mentioned it was overexpressed in the GC tissue sample, targeting and down-regulating NF1 in the GC cell line, whereas Qin et al. (2019) mentioned that has-miR-1290 was overexpressed in oral squamous cell carcinoma (OSCC) tissue sample and cell line by targeting and down-regulating CCNG2. For the interactions of the three lncRNAs with other miRNAs in NPC and other cancers, the explanation was in section 5.2 in conjunction with the interactions between hsa-miR-1246, hsa-miR-211-5p, hsa-miR-135b-5p, hsa-miR-93-5p, hsa-miR-16-5p and hsa-miR-1305 with their target mRNAs and their expressions in NPC and other cancers for being targeted by EBV BART miRNAs based on cross-regulatory network (Figure 4.13).

Next, 18 hub genes or hub proteins with connectivity degree greater than 5 were identified in the protein-protein interaction (PPI) network (Figure 4.9 and Table 4.8). It means the genes are connected or interacting with more than five other genes in the PPI network and can regulate a set of genes and act similarly to oncogenes or tumor suppressor genes (Seo et al., 2009; Yu et al., 2017). Comparing this to the bioinformatic study by Xu et al. (2020), all of these genes were new potential hub genes in NPC, whereas a bioinformatic study for NPC by Ye et al. (2019), our result shared two similar hub genes with them which are DNAL1 and RSPH4A. Some of these genes share similar biological functions within a module (Chen et al., 2017). A gene module is a set of co-expressed genes requiring the same transcription factors (Bar-Joseph et al., 2003). These genes are densely connected regions in the large PPI network which may represent molecular complexes and coded proteins with similar biological functions. Thus, the deregulation of any of the genes affects the expression of their neighbouring genes and is suggested to cause NPC tumorigenesis.

Our result revealed that only module 1 consisted of the hub genes based on Table 4.8, where these hub genes co-expressed, and their down-regulation affected the expression of the genes connected to them. Consequently, this is suggested to cause NPC tumorigenesis by affecting similar biological functions essential for healthy nasopharynx functions. These genes may also be the down-regulated tumor suppressor genes that cause the dysregulation of many other genes in NPC tumorigenesis unidentified in this study due to limited microarray datasets. Nevertheless, none of the genes in module 2 are hub genes based on Table 4.8. Therefore, the genes in Module 2 are suggested not to be regarded as tumor suppressor genes, and their down-regulation affects similar biological functions

essential for healthy nasopharynx functions. This down-regulation subsequently causes NPC tumorigenesis.

The genes in module 1 are mainly important for motile cilia assembly, motility, and function (Rashid et al., 2006; Zhu et al., 2019; Vig et al., 2020; Braschi et al., 2022). Only the interactions between DNALI1, DNAH6, DNAH5, and DNAH12 down-regulation were observed in humans. As mentioned by Thomas et al. (2020), DNALI1, DNAH6, and DNAH12 were down-regulated simultaneously in PCD patients causing abnormal motile cilia motility. Shoemark et al. (2017) also revealed that the absence of DNAH5 coexisted with the absence of DNALI1 in PCD patients. Thus, the interactions between the genes in Module 1 require further validation.

Out of all these genes, only DNAH5 was validated to be down-regulated in a replicated cohort of NPC patients based on GWAS and gene expression on NPC tissues where the down-regulation is believed to cause loss of respiratory cilia functions, leading to poor mucociliary clearance and causing infections on the upper and lower respiratory tract, including the nasopharynx (Chin et al., 2016). Although no information indicates the other dynein axonemal heavy chain (DNAH) family (DNAH2, DNAH6, and DNAH12) are down-regulated in NPC, previous studies had shown their down-regulation or absence led to the stiff abnormal beating of cilia, decreased in cilia beating or immotile cilia in the human nasal biopsies and caused primary ciliary dyskinesia (PCD) including DNAH5 (Hornef et al., 2006; Berg et al., 2011; Li et al., 2016; Thomas et al., 2020). PCD is the defect in motile cilia lining the epithelium of the respiratory tract causing abnormal cilia structure and beating pattern or absence of cilia (Bergström et al., 2012).

From this, it is plausible to link the down-regulation of these genes with NPC. This is because PCD patients exhibited similar defects in the nasal ciliary motor with NPC leading to nasal congestion, which is also demonstrated by NPC patients (Bush & O'Callaghan, 2002; Hao et al., 2008). Similarly, the down-regulation of Dynein light intermediate polypeptide 1 (DNALI1) also resulted in PCD when either GAS8 or DNAH5 was absent in nasal epithelial and demonstrated stiff abnormal beating of cilia, decrease in cilia beating or immotile cilia (Shoemark et al., 2017; Baz-Redón et al., 2020). However, the study of PCD on WDR78, WDR63, DYNLRB2, and DYNC2H1 in humans is unavailable to our knowledge. Regardless, WD repeat-containing protein 78 (WDR78) down-regulation in mice is known to cause cilia immotility and defect in the cilia assembly of the Dynein-f complex, including WD repeat-containing protein 63 (WDR63) (Zhang et al., 2019). This is due to the interaction of WDR78 with the subunits of the axonemal inner arm I1/f dynein complex (Zhang et al., 2019). These genes are also known as DNAI4 and DNAI3, respectively (Braschi et al., 2022). However, the cilia defects are observed in kidney cell lines thus further verification on nasal cilia is required. For the Dynein light chain roadblock-type 2 (DYNLRB2) and Cytoplasmic Dynein 2 Heavy Chain (DYNC2H1), the information on their down-regulation or defective effects in cilia were limited and was suggested to be further validated.

In cancer studies, only WDR63, DYNLRB2, DYNC2H1, and DNALI1 down-regulation are associated with cancers and metastasis, excluding the other genes. Zhao et al. (2020) revealed that WDR63 was down-regulated in both lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), and its expression was inversely related to metastasis by repressing p53 expression, whereas for DYNLRB2, Tian et al. (2017) revealed it was significantly down-regulated in LUSC with metastasis in the tumor

sample. For both DYNC2H1 and DNALI1, their expressions were down-regulated in premalignant and malignant breast cancer tissue, respectively (Parris et al., 2010; Menzl et al., 2014).

For the genes in module 2, they were reported to be essential for normal cilia differentiation, assembly, and functions (Johnson et al., 2018; Bazan et al., 2021). The studies of their expressions and effect on cilia were limited in human and mostly on mouse models. Besides, the study of regulatory interactions amongst the genes in Module 2 causing cilia defect is unavailable and requires further verification. Johnson et al. (2018) found Fibronectin type 3 and ankyrin repeat domains 1 (FANK1) involved in cilia differentiation in mouse trachea airway epithelial cells, and Zhang et al. (2018) found leucine-rich repeat containing 23 (LRRC23) was essential in radial spoke assembly in sperm flagella compared to in cilia lining the trachea in mice, where its down-regulation exhibited minute changed in cilia motility, beating and length. For Enkurin (ENKUR), Sigg et al. (2017) described its down-regulation and absence in expression exhibited no distinguishable cilia beating and motility in mice trachea epithelial cells compared to control cells. On the contrary, the study of the down-regulation effects in cilia beating, motility, and its function for Coiled-Coil Domain Containing 113 (CCDC113) in the respiratory system was limited. Nevertheless, Hou et al. (2021) revealed that ENKUR was significantly down-regulated in NPC tissue, and its up-regulation inhibited metastasis induced by cinobufotalin. Only ENKUR down-regulation was studied in NPC, and generally, the study of down-regulation in FANK1 and ENKUR in cancer mechanisms was limited except for CCDC113. Zhang et al. (2021) found it is down-regulated in colorectal cancer (Zhang et al., 2021).

Motile cilia are tiny hair-like structure which lines the epithelial cells of the upper and lower respiratory tract, including the nasopharynx (Lee et al., 2021). It is essential for the flow and clearance of fluids and particles in a coordinated beating (Ringers et al., 2020). The down-regulation of the genes in both modules were suggested to cause NPC by affecting the nasal ciliary motor. It is plausible to deduce this because in PCD, similar nasal ciliary motor defects, including abnormal cilia structure and beating pattern or absence of cilia, occurred, leading to nasal congestion, which is a similar symptom in NPC (Hao et al., 2008; Bush & O'Callaghan, 2002; Bergström et al., 2012). Therefore, when EBV infection occurs, it is easy for EBV-positive NPC tumorigenesis to occur due to the mucociliary clearance defect because it is the primary airway defence mechanism against inhaled particles trapped in the mucus, including pathogens like viruses (Gizurason, 2015; Kamiya et al., 2020). The coordinated cilia beating in mucociliary clearance can sweep the mucus covering the nasal epithelium towards the nasopharynx by coughing to remove the particles and pathogens from the respiratory tract (Martin et al., 1997; Kamiya et al., 2020). The defect in this mechanism is suggested to assist EBV infection and colonization of epithelial cells and B-cells in the nasopharynx. This occurred and can be observed in individuals with undifferentiated NPC, aided by the heavy infiltration of lymphocytes and inflammatory stroma features of this type of undifferentiated NPC (Tsao et al., 2017).

Interestingly, six of the BART miRNAs revealed by the EBV miRNA-mRNA interaction and cross-regulatory networks (Figure 4.6 and Figure 4.13) do not target these hub genes. From this, it can be deduced that EBV-positive NPC is not due to the down-regulation of tumor suppressor genes or it could be simply due to the limited mRNA microarray dataset from the GEO database restricting the overlapping results with ViRBase. Recent studies showed EBV miRNAs were linked with tumorigenesis, as

observed in their up-regulated expression in tumor biopsies and cell lines (Zhang et al., 2019). These miRNAs assisted the EBV life cycle by down-regulating host mRNAs (Abusalah et al., 2022). For this study, we were only focusing on the *Bam*HI A rightward transcripts (BART) miRNAs because their up-regulation was common in EBV-positive NPC and was also present in NPC latency pattern II with the absence of the *Bam*H1 fragment H rightward facing 1 (BHRF1) miRNAs (Cosmopoulos et al., 2009; Zhang et al., 2022).

5.2 Cross-Regulatory Network

The construction of a cross-regulatory network was done by integrating the ceRNA network with the EBV miRNA-mRNA interaction network to give insight and understand the regulatory mechanisms of EBV in NPC. This is done by constructing a cross-regulatory network by integrating the ceRNA network with the EBV miRNA-mRNA interaction network. The cross-regulatory network (Figure 4.13) showed six EBV miRNAs (ebv-miR-BART21-3p, ebv-miR-BART19-3p, ebv-miR-BART15, ebv-miR-BART2-5p, ebv-miR-BART20-3p, and ebv-miR-BART11-5p) were interacting with four mRNAs (EYA4, EYA1, EBF1 and MACROD2) and these mRNAs were interacting with six host miRNAs which were hsa-miR-1246, hsa-miR-93-5p, hsa-miR-16-5p, hsa-miR-135b-5p, hsa-miR-211-5p and hsa-miR-1305. In addition, these miRNAs were observed interacting with three lncRNAs (CASC2, TPTE2P1, and ARHGEF26-AS1). In previous studies, these BART miRNAs were highly expressed in NPC tissue samples, cell lines and serum samples (Song et al., 2019; Zhu et al., 2009; Lung et al., 2018; Jiang et al., 2020; Zhou et al., 2022). However, the expression level of BART15 was inconsistent in NPC, where Amoroso et al. (2011) revealed it was barely expressed in the NPC cell line, whereas Cosmopoulos et al. (2009) revealed it was highly expressed in both the NPC cell line and tumor biopsy. Both

of these results contradict Lung et al. (2018) result, where they demonstrated BART15 was undetected in the NPC cell line. Despite the availability of these studies, the regulatory mechanisms of these BART miRNAs in down-regulating mRNAs in cancers and NPC were limited except for in BART2-5p. As described by Jiang et al. (2020), BART2-5p down-regulated RND3, a negative regulator of the cell migration ROCK signalling pathway and promoted metastasis in NPC. The information on the interaction between these BART miRNAs with the four targeted DEGs in NPC and other cancers was unavailable except for BART11-5p interaction with EBF1 in an EBV-transformed lymphoblastoid cell line (LCL) (Ross et al., 2013).

For the six host miRNAs, their expressions were up-regulated in NPC tumor samples, cell lines, and serum, excluding hsa-miR-211-5p and hsa-miR-16-5p (Plieskatt et al., 2014; Tang et al., 2014; Wargasetia & Priyandoko, 2018). Their interactions with the four targeted DEGs in the regulatory network were unavailable to our knowledge, either in NPC or other cancers. Nevertheless, multiple studies have shown their interactions with other mRNAs, causing down-regulation of the mRNAs and cancerogenesis, except for hsa-miR-16-5p, hsa-miR-211-5p, and hsa-miR-1305. These three miRNAs were observed to demonstrate the opposite effect in various cancers, including non-small cell lung cancer (NSCLC) and hepatocellular carcinoma (HCC), where they were down-regulated and caused the up-regulation of their target mRNAs, including XIST, MDM2, and ACSL4 (Cai et al., 2019; Qin et al., 2020; Du et al., 2021). For hsa-miR-16-5p, its unspecific hsa-miR-16 was also observed to be down-regulated in NPC tissue and cell lines, leading to the up-regulation of its target FGF2 mRNA (He et al., 2016). However, no interactions of hsa-miR-211-5p and hsa-miR-1305 with other mRNAs were observed in any NPC types.

Similarly, no interaction of has-miR-1246 was observed in NPC. For has-miR-1246, Wu et al. (2022) demonstrated its up-regulation in laryngeal squamous cell carcinoma (LSCC) cell line and tissue, where it promoted cell proliferation and migration and inhibited apoptosis by down-regulating Thrombospondin 1 (THBS1). On the contrary, unspecific up-regulation of has-miR-93 in NPC tissue samples and cell lines (CNE-1, SUNE-1, CNE-2, HNE-1, C666-1, and HONE-1) was exhibited by Xu et al. (2015) where it was observed to target Dab2 in CNE-1 and SUNE-1 cell lines. Besides, has-miR-93-5p was demonstrated to be up-regulated in gastric cancer (GC) cells by Ma et al. (2017), where it targets and down-regulates type I interferon (IFN1) and causes proliferation, migration, and invasion of GC cells. A similar expression pattern was observed in GC tissues for has-miR-135b-5p, where it targets and down-regulates CKLF-like MARVEL transmembrane domain containing 3 (CMTM3), according to Lu et al. (2018). In NPC, has-miR-135b-5p was overexpressed in NPC tissue and cell lines, causing the down-regulation of Sirtuin 1 (SIRT1) reported by Cheng et al. (2022). Regardless, the interactions of these miRNAs with the three lncRNAs is unavailable to our knowledge. Further investigation using molecular techniques for this multi-level RNA cross-regulatory interaction with BART miRNAs should be done to validate these findings, especially in EBV-positive NPC.

The cancer susceptibility candidate 2 (CASC2) lncRNA was the only lncRNA which was observed to be down-regulated in NPC tissue sample and cell lines, where it promoted cell proliferation and inhibited apoptosis by down-regulating RBBP8 through up-regulation of miR-18a-5p (Miao et al., 2019). On the contrary, its up-regulation significantly inhibited cell proliferation and enhanced apoptosis in lung adenocarcinoma tissue and cell by inhibiting miR-21 expression and up-regulating p53 expression (Wu et

al., 2021). This indicated that CASC2 can act similarly to a tumor suppressor when up-regulated, and its down-regulation leads to carcinogenesis. For both transmembrane phosphoinositide 3-phosphatase and tensin homolog 2 pseudogene 1 (TPTE2P1) and rho guanine Nucleotide exchange factor 26 antisense RNA 1 (ARHGEF26-AS1) lncRNAs, the studies of their deregulation in NPC is unavailable. In other cancers, TPTE2P1 is generally up-regulated, such as in colorectal, gallbladder, and hepatocellular carcinomas (Liu et al., 2019). This contradicted our results and its function as a PTEN tumor suppressor because it is a pseudogene lncRNA of the TPTE2 and has a similar phosphoinositide phosphatase activity to the TPTE2 (Dai et al., 2019). From our data, this suggested that TPTE2P1 can act similarly to both tumor suppressors and not only as oncogenes. When up-regulated, Peng et al. described that it promoted cancer cell proliferation in thyroid carcinoma by sponging and down-regulating miR-520c-3p. Besides, Dai et al. also mentioned that its up-regulation promoted cell viability and tumor progression by inhibiting apoptosis in colorectal cancer. Its oncogene nature was described by Lv et al. (2018), where its down-regulation significantly inhibited the migration and invasion of gallbladder cancer.

The ARHGEF26-AS1 is an antisense RNA that was observed to be down-regulated in various cancers, including lung adenocarcinoma (LUAD) and esophageal carcinoma (Yao et al., 2021; Chen et al., 2021). As described by Yao et al. (2021), ARHGEF26-AS1 was down-regulated in LUAD cell and was assumed to down-regulate iron concentration and inhibits ferroptosis, which is a process of cell death through membrane lipid peroxidation resulting from increased iron concentrations (Yao et al., 2021). This was also reported by Chen et al. (2021), where the down-regulation of ARHGEF26-AS1 led to the up-regulation of miR-372-3p and caused the down-regulation of ADAM23 in esophageal cancer by inhibiting the ferroptosis pathway. Conversely, the up-regulation of

ARHGEF26-AS1 and ADAM23 were observed to inhibit cancer cell proliferation and migration by activating the ferroptosis pathway (Chen et al., 2021). This suggested ARHGEF26-AS1 can act similarly to the tumor suppressor.

5.2.1 Functional and Pathway Enrichment Analysis of The Four DEGs

As shown by the Cellular Component (CC) of the GO term, EYA1, EYA4, EBF1 and MACROD2, which were targeted by the BART miRNAs, are located in the nucleus (Figure 4.14). The DEGs were also enriched in 8 GO terms of the Biological Processes (BP) (Figure 4.14). These were DNA repair, chromatin organization, sensory perception of sound, anatomical structure morphogenesis and development, positive regulation of DNA repair, and negative regulation of extrinsic apoptosis signalling pathway in the absence of ligand and histone dephosphorylation. The infection from EBV can induce DNA damage due to viral genome integration with the host genome via DNA breaks (Katerji & Duerksen-Hughes, 2021). This lytic infection induces double-stranded break (DSB) signalling in host cells (Frappier, 2019). Signalling is a part of DNA repair known as DNA damage response (DDR) to detect and fix DNA damage through apoptosis (Hau & Tsao, 2017). However, EBV can control DDR using viral proteins and BART miRNAs to maintain its lytic and latent cycles; thus, instead of apoptosis, abnormal cell proliferation occurs due to genomic instability, which leads to tumorigenesis (Frappier, 2019; Hau & Tsao, 2017).

This occurrence of DNA damage also leads to chromatin organization, histone phosphorylation, and dephosphorylation. The chromatin organization is initiated by the phosphorylation of the histone variant H2AX and is essential for the binding of factors involved in DDR (Dinant et al., 2008). Generally, upon DNA damage, the H2AX is

phosphorylated on serine 139 by ataxia-telangiectasia-mutated (ATM) kinase, which forms γ H2AX foci to enable retention and increase of 53BP1 repair factor concentration at the damage sites allowing DNA repair (Celeste et al., 2003). This is coupled with the dephosphorylation of H2AX Y142 to recruit Mediator Of DNA Damage Checkpoint 1 (MDC1) to γ H2AX (Xiao et al., 2009). The MDC1 mediates additional ATM binding to γ H2AX and activates cell cycle checkpoints to inhibit the replication of damaged DNA and act as a linkage between γ H2AX and ATM (Stucki et al., 2005; Moon et al., 2010). This regulation is known as a positive feedback loop of DNA repair (Nowsheen et al., 2018).

The dephosphorylation of H2AX Y142 is aided by ZNF506, which recruits phosphatase EYA for the removal of the phosphate group and MDC1 from H2AX Y142 on the lesion region (Nowsheen et al., 2018). Therefore, in EBV-positive NPC, it is suggested that EYA transcriptional coactivator and phosphatase 1 (EYA1) and EYA transcriptional coactivator and phosphatase 4 (EY4) are targeted and down-regulated by BART miRNAs leading to decrease histone desphosphorylation activity and inhibition of DNA repair. This corroborates our findings illustrated through the cross-regulatory network. Both the EYA1 and EYA4, including other EYA family genes, are responsible for encoding the protein tyrosine phosphatase (PTP) where it targets Tyr, Ser and Thr on histone during dephosphorylation (Tadjuidje et al., 2012; Sadatomi et al., 2013). This was shown by the Molecular Function (MF) of the GO term as protein tyrosine phosphatase activity (Figure 4.14). After DNA repair, EYA will dephosphorylate Tyr142 in the C-terminal tail of the H2AX and activate histone phosphorylation on Ser139 (Cook et al., 2009; Sadatomi et al., 2013).

The down-regulation of EYA1 and EYA4 in cancers is limited, and EYA4 is found to be down-regulated in esophageal and EYA1 is found to be down-regulated in gastric cancer (Nikpour et al., 2014; Luo et al., 2018). According to Luo et al. (2018), the down-regulation of EYA4 caused metastatic esophageal squamous cell carcinoma (ESCC) by activating the Akt/GSK-3 β /slug signalling pathway. They also further described that EYA4 promoted DNA repair, inhibited apoptosis in ESCC, and maintained genomic stability when cisplatin was introduced, indicated by higher expression of Ser139 histone phosphorylation. For EYA1, Nikpour et al. (2014) described EYA1 to be significantly down-regulated in gastric cancer tissues and negatively correlated with tumor size and metastasis. Despite all these, the studies of EYA1 and EYA4 down-regulation in NPC and targeted by BART miRNAs were limited and worth studying to validate our result. These EYAs are also metalloproteins, as shown by the Molecular Function (MF) of the GO term as metal ion binding (MIB). Their metalloenzymes activity requires the divalent metal ion Mg²⁺ to catalyse histone dephosphorylation compared to other protein Tyr phosphatases (Tadjuidje et al., 2012; Sadatomi et al., 2013). It is suspected that BART15, BART19-3p, BART21-3p and BART12-5p target EYA1 and EYA4 and consequently inhibits Mg²⁺ binding with EYA1 and EYA4. This leads to defect in positive feedback of DNA repair regulation and EBV-positive NPC. Yet, this hypothesis is to be experimentally tested.

Typically, the defect in histone dephosphorylation leads to the activation of the extrinsic apoptosis pathway through the recruitment of JNK (c-Jun N-terminal protein kinase) (Dhanasekaran & Reddy, 2008; Nowsheen et al., 2018). However, this mechanism is suggested to be altered by the presence of EBV BART miRNAs. These BART miRNAs are suggested to down-regulate the major histocompatibility complex class I chain-related peptide A (MICA) and peptide B (MICB) ligands which are expressed by the infected cells

upon DNA damage and viral infection by targeting EYA1 and EYA4 in EBV positive NPC (Zingoni et al., 2018; Png et al., 2021). As described by Wong et al. (2018), ebv-miR-BART7 targeted and suppressed TGF β 1 expression, causing the down-regulation of MICA ligand presentation on infected cells and evading natural killer (NK) cell surveillance. This was further reported by Nachmani et al. (2009), where they only described ebv-miR-BART2-5p down-regulated MICB ligand presentation without explaining the mechanism. Both the MICA and MICB ligands are essential to elicit tumor necrosis factor-alpha (TNF- α) ligands from NK cells upon binding to the natural killer group 2, member D (NKG2D) receptor expressed on the NK cells surface (Jinushi et al., 2005). Thus, the absence of these MIC ligands causes no TNF- α ligand binding to the Tumor Necrosis Factor Receptor Superfamily (TNFRSF) on the infected cells for JNK signalling pathway activation (Locksley et al., 2001; Dhanasekaran & Reddy, 2017). This further leads to negative regulation of the extrinsic apoptosis signalling pathway in the absence of ligands and escaping NK cells-mediated immune surveillance. Eventually, this is suggested to inhibit apoptosis and leads to cell proliferation, tumorigenesis and metastasis (Wong et al., 2018).

Subsequently, this leads to anatomical structure morphogenesis and development, including angiogenesis and vasculogenic mimicry (VM), which are common in various cancers, including NPC (Luo et al., 2021; Tian et al., 2021). Angiogenesis is the formation of a blood vessel from pre-existing blood vessel to transport oxygen and nutrients in the tumor microenvironment to support tumor proliferation, growth and metastasis (Zuazo-Gaztelu & Casanovas, 2018; Teleanu et al., 2019). It is promoted by hypoxia, which leads to inflammatory signals and cytokines and thus recruits vascular cells to form tumor vessels (Zuazo-Gaztelu & Casanovas, 2018). Similarly, VM is also promoted by hypoxia, a condition of oxygen deprivation in cancerous cells as tumor tissue and cells increase in size

and distance from blood vessels (Muz et al., 2015; Zuazo-Gaztelu & Casanovas, 2018; Fernández-Cortés et al., 2019). VM is the development of a vascular network in infected epithelial cells (Xiang et al., 2018). Compared to angiogenesis, VM is the cancerous cells mimicking endothelial cells by organizing and transforming themselves into elastic vascular-like structures independent of pre-existing endothelial cells (Li et al., 2017; Fernández-Cortés et al., 2019). Its phenotype resembles and acts similarly to endothelial blood vessels (Xiang et al., 2018). Its function is to supply blood for tumor cell growth induced by EBV to obtain nutrients and oxygen independently of blood vessels or angiogenesis (Xiang et al., 2018; Fernández-Cortés et al., 2019).

The increase in the size of NPC tumor cells emerges from the primary location of NPC, known as fossa Rosenmüller covers the adjacent Eustachian tube, causing its dysfunction (Utama et al., 2022). This causes ventilation impairment, middle ear secretion drainage and change in middle ear pressure, which leads to hearing loss, middle ear effusion, and tinnitus (Utama et al., 2022). This is suggested to correlate with the symptoms in NPC, which are hearing loss and tinnitus (Ho et al., 2008). The Eustachian tube is essential in mucociliary clearance to drain ear secretion from the middle ear into the nasopharynx (Casale et al., 2023). The failure in this causes middle ear effusion or Otitis media with effusion (OME) resulting from the decreased in cilia density of the mucociliary epithelium and the increase in goblet cells lining the Eustachian tube (Matsune et al., 1992; Depreux et al., 2008). OME is a condition of fluid buildup in the middle ear with the absence of acute infection compared to otitis media (Searight et al., 2022). The fluid puts pressure on the tympanic membrane, thus preventing proper vibration and causing hearing loss (Searight et al., 2022). The occurrence of OME is one of the well-known symptoms of NPC (Tsunoda et al., 2021). This was reported by Huang et al. (2012), where those

diagnosed with OME were diagnosed with NPC within five years with an incidence rate of 6.41 per 10,000 person-years. Besides, the tumor cells covering the Eustachian tube hinder ventilation of the middle ear from equalizing pressure in the middle ear with atmospheric pressure for proper vibration of the eardrum and sound transmission (Casale et al., 2023). This causes negative pressure in the middle ear, causing tinnitus due to fluid accumulation and exudation (Bal & Deshmukh, 2022; Mayo Clinic, 2022).

In cancer pathogenesis, transcription factors (TFs) are commonly deregulated (Bhagwat & Vakoc, 2015). This is due to the ability of TFs to recognize and binding to the DNA sequence of multiple genes and regulate the expressions (Islam et al., 2021). In this study, the Early B-cell factor transcription factor 1 (EBF1) was known to function as TF in the development of B cells by pairing with E2A and PAX5 genes (Vilagos et al., 2012). It was targeted by ebv-miRNA-BART11-5p in the cross-regulatory network and not interacting with any other DEGs in the ceRNA network (Figure 4.13). Due to this, no TF-mRNAs network was constructed, and its regulation on the other DEGs in NPC could not be deduced. This gene was down-regulated in gastric and colorectal cancers by targeting and regulating other genes (Shen et al., 2020; Cheng et al., 2023). According to Chen et al. (2023), EBF1 targeted the gene Aldo-keto reductase family 1 member B1 (AKR1B1) in gastric cancer, and its over-expression caused the down-regulation of AKR1B1 and suppressed the growth of tumor cells whereas its down-regulation caused up-regulation of AKR1B1 and tumor cells growth and invasion. Shen et al. (2020) also mentioned that EBF1 over-expression caused down-regulation of Partner Of Nin one binding (PNO1) and up-regulation of p53 and p21 in colorectal cancer (CRC).

The information on EBV BART miRNAs interaction and its regulation mechanism with EBF1 is limited in cancer pathogenesis, especially in NPC, compared to EBV viral proteins. Only in an EBV-transformed lymphoblastoid cell line (LCL), BART11-5p binded on the miRNA target site located at 2137-2159 bp after the stop codon of EBF1, revealed by Ross et al. (2013). They further justified that this leads to an abnormal germinal centre reaction in the lymphoid tissues and inhibits mature B-cells development. This affects the proliferation of B-cell to B-cell lymphoblast and inhibits its differentiation into resting memory B-cells upon virus infection recognition (Thorley-Lawson & Gross, 2004). With this, EBV can escape immune surveillance of memory B-cells and its CD21 receptor, which can detect EBV progeny (Thorley-Lawson & Gross, 2004; Hatton et al., 2014). As indicated by the GO analyses of the MF, EBF1 was involved in metal ion binding (MIB), similar to EYA1 and EYA4. Divalent zinc ion (Zn^{2+}) is required by the EBF1 protein to bind to DNA sequences of targeted genes (Hagman et al., 2011). This Zn^{2+} binds the zinc-coordination motif termed zinc knuckles located in the DNA-binding domain (DBD) of EBF1 (Hagman et al., 2011; Boller & Grosschedl, 2014). The importance of MIB on proteins is to assist in conferring the protein's structural, regulatory, and enzymatic activity (Permyakov, 2021). It is suspected that BART11-5p targets EBF1 and inhibits MIB-EBF1 interaction leading to the deterioration of B-cell production. This is a way for EBV to escape the immune surveillance of memory B-cells leading to EBV-positive NPC. Yet, this hypothesis is to be experimentally tested.

Lastly, Mono-ADP-Ribosylhydrolase 2 (MACROD2) is a gene that involves in reversible ADP-ribosylation and acts as a hydrolase (Žaja et al., 2020). One of the functions of ADP-ribosylation is recruiting repair factors to chromatin in DNA repair (Golia et al., 2017). After DNA repair, MACROD2 will remove the terminal autoinhibitory

mono-ADP-ribose from the transferase Poly(ADP-ribose) polymerases 1 (PARP1), which is known as dePARylation or histone dephosphorylation to prevent the trapping of proteins involved in DNA repair at the damage site and causing hypersensitivity to DNA damage (Sakthianandeswaren et al., 2018; Kassab et al., 2020). This also allows the DNA damage machinery to efficiently recognise DNA lesions in different proximity (Chen & Yu, 2019). Based on this, MACROD2 is deemed as a negative feedback loop and restricts the recruitment repair factors and effector proteins to the DNA break site when phosphorylates by ATM and exported out of the nucleus (Golia et al., 2017). This is also demonstrated by Sakthianandeswaren et al. (2018), where its down-regulation led to colorectal cancer by up-regulating PARP1 mono-ADP-ribosylation and reducing its transferase activity which causes impairment in DNA repair and hypersensitivity to DNA damage.

However, no results for pathways affected by EBV miRNAs were obtained from this study. The possible reason was the limited updated microarray datasets curated from the GEO database because this study only involved curating microarray databases dated after 2014 from the accessed year (2020) and excluded the RNA-sequencing (RNA-seq) database. The reason for this was to ensure the expression datasets were curated following updated microarray guidelines and protocols. The exclusion of the RNA-seq database was due to its analysis utilizing different normalization methods with microarray techniques. Its normalization methods only involved reads per kilobase of transcripts where the reads number is proportional to transcripts abundance without using housekeeping genes expression (Liu et al., 2019). If RNA-seq was included, it might have led to a false positive result.

To address more on the limitation of our study, the exclusion of constructing EBV-negative ceRNA to compare to the result might led to a false positive. To justify this, the mRNA and lncRNA microarray datasets curated consist of NPC patients in the endemic area, China, which is highly associated with EBV-positive NPC. Thus, this increased the chances that the obtained result might not be false positive. Although the exact reason did not apply to the miRNA dataset, but the dataset contained the expression levels of EBV BART miRNAs, where some are significantly up-regulated with P-value <0.05, insinuating that the NPC patients in this dataset might be EBV-positive NPC.

Besides, the mRNA and lncRNA targeted by miRNAs and the EBV miRNAs targeting the mRNAs were curated from up-to-date databases, which are experimental-based in combination with computational prediction. For example, the miRTarBase uses high-throughput techniques, including cross-linking and immunoprecipitation and high-throughput sequencing (CLIP-seq), cross-linking, ligation, and sequencing of hybrids (CLASH-seq) and degradome-seq which were validated by using the miRNA-target gene expression profiles from GEO and The Cancer Genome Atlas (TCGA) databases for identification of miRNA-target mRNA (Moore et al., 2015; Huang et al., 2020).

Similarly, the DIANA-LncBase V3 and StarBase V2 were also used to curate miRNA-target lncRNA interactions where these databases were validated by using CLIP-seq, including HITS-CLIP (high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation), PAR-CLIP (photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation) and CLEAR-CLIP (covalent ligation and endogenous Argonaute-bound RNA) which enable the identification of Argonaute (AGO):miRNA binding sites on a targeted RNAs (Karagkouni et al., 2020). Next, the EBV BART

interactions with target mRNAs curated from VirBase were experimentally validated using CLIP-seq, quantitative real-time PCR, Western Blot and immunohistochemistry analysis techniques. A high confidence score was set to curate EBV miRNAs-mRNAs interactions as this is commensurate to more evidence resources for the interactions. Nevertheless, the results of this study need to be validated experimentally in the future despite the promising possible interactions between the RNAs with BART miRNAs.

To decrease the chances of obtaining false positive results, a thorough curation involving RNA sequencing datasets such as from the TCGA database could be done to give insight into understanding the regulatory mechanisms of BART miRNAs in NPC. Using only RNA sequencing datasets is also plausible as it thoroughly quantifies RNA expressions. Furthermore, molecular experimentations that include quantitative real-time PCR, Western Blot and immunohistochemistry analyses are recommended to be done to verify BART miRNAs mechanisms on EYA1, EYA4, EBF1 and MACROD2, and the RNAs in the ceRNA activity by identifying their expression level. Their interactions also can be validated through CLIP-Seq techniques. As these BART miRNAs, DEGs and lncRNAs are potential targets in therapeutic design for EBV-positive NPC. A future and more thorough study would have to include RNA-sequence databases as microarray is less favoured than RNA-sequencing in recent years, granting to the limited datasets in this study. The reason is that RNA-seq provides thorough quantification by sequencing the whole transcriptome and can detect more differentially expressed genes of the higher and lower folds, thus definitely increase result accuracy (Zhao et al., 2014; Rao et al., 2019).

5.3 Overall Survival

For the survival curve analysis by using Kaplan-Meier (K-M) Plotter, only EBF1 and MACROD2 were suggested to be positively correlated with the overall survival (OS) of the patients, especially for EBF1 as shown by the P value <0.05 (Figure 4.15). This supports the result demonstrated by the cross-regulatory network, where EBF1 and MACROD2 were down-regulated. This implied that the down-regulation of these DEGs was linked with an increased mortality rate. However, only EBF1 survival curves showed a constant positive correlation with OS without censoring of late patients for up-regulated and down-regulated expressions. For MACROD2, the censoring of the late patients in the down-regulated expression group between months 150 to 200 led to inconclusive results for the group, as the patient could have survived after the study ended. This censoring poses limitations to this analysis due to incomplete data collection.

In addition, both EBF1 and MACROD2 exhibited hazard ratio (HR) < 1 , which indicated the up-regulated expression group have a higher survival rate and was estimated to have 0.71 and 0.66 times more likely to survive compared to the down-regulated groups of EBF1 and MACROD2, respectively at any point in time during the study. From this, it could be deduced that these genes can increase the mortality rate with slight down-regulation. This is demonstrated by Shen et al. (2020), where over-expression of EBF1 *in vitro* and *in vivo* suppressed tumor growth by inhibiting cell proliferation and instigating cell apoptosis in colorectal cancer cell lines and mice by inhibiting PNO1-mediated p53/p21 signalling pathway activation. They further described that colorectal cancer patients with down-regulated EBF1 exhibited shorter OS. For MACROD2, Zhou et al. (2020) revealed its over-expression in hepatocellular cell lines significantly reduced cell proliferation *in vitro* and metastasis *in vivo*, demonstrated by xenograft in mice by

inhibiting GSK-3 β / β -catenin signalling. In a different study conducted by Mohseni et al. (2014), they demonstrated that MACROD2 over-expression was observed in metastatic breast cancer cell lines and was associated with worse OS.

On the contrary, the EYA1 and EYA4 curves showed an inconsistent negative correlation with patients' OS due to the censoring of late patients in the down-regulated expression group, and the P-value of EYA1 was not significant (Figure 4.15). This meant the up-regulation of these DEGs was correlated with an increased mortality rate until 150 months, when it showed a positive correlation with patients' OS onward. This shift supported the result demonstrated by the cross-regulatory network, where EYA1 and EYA4 were down-regulated in EBV-positive NPC. This was deduced due to the possibility that the censored patients could have survived longer than those in the down-regulated expression group. Nonetheless, as shown by the KM plot, these DEGs exhibited HR > 1, which indicated the up-regulated group patients have low survival and were estimated to have 1.34 and 1.33 times more likely to die compared to the down-regulated groups of EYA1 and EYA4, respectively, at any point in time during the study. According to this, it can be deduced that a slight up-regulation in these genes can increase the mortality rate.

While this contradicts our findings from the cross-regulatory network, some studies corroborate our findings where the down-regulation of EYA 1 and EYA4 is linked with poor OS. This is demonstrated by Mo et al. (2018), where EYA4 down-regulation was linked with poor OS, and its *in vitro* over-expression inhibited hepatocellular cell (HCC) proliferation, reduced cell invasion and inhibited tumor growth in mice by repressing RAP1. Another study conducted by Gu et al. (2019) also demonstrated that the down-regulation of EYA4 was associated with poor OS, and its up-regulation inhibited

metastasis and reduced cell invasion and angiogenesis by repressing c-JUN and VEGFA expression levels in HCC and mice xenograft. However, the study on EYA1 down-regulation association with poor OS is unavailable to our knowledge. Only a study on its up-regulation associated with poor OS and promotes metastasis and invasion in HCC cells was available, conducted by Kong et al. (2019). Based on these varied results from previous studies, further study is required to validate these DEGs' deregulation association with OS in EBV-positive NPC.

The limitation of this OS analysis is the non-specific NPC database used by the K-M plotter in generating the curves. It uses head and neck cancer expression datasets. As it is known, head and neck cancer encompasses various types of cancers, not only NPC. This could be the reason for the inconsistent expressions correlation with patients' OS observed in MACROD2, EYA1, and EYA 4 when considering the result obtained from the cross-regulatory network. The other reason for this could be that the mRNA expressions used by the K-M plotter were analysed using the RNA-sequencing technique curated from Gene Expression Omnibus (GEO), European Genome-phenome Archive (EGA), and The Cancer Genome Atlas (TCGA) databases. As stated, our study only involved the curation of RNA expression analysed using the microarray technique. In hindsight, the results generated from the plotter might decreased the chance of false positives of our results when taking into consideration comparing microarray expressions with RNA-sequencing expressions.

We also noticed that the curves were constructed by including all patients across the globe regardless of ethnicity and geographical areas, adding a limitation to our result as this study solely aims to identify mRNAs dysregulated by EBV miRNAs to help in understanding regulatory mechanisms of EBV miRNAs on mRNAs in EBV positive NPC.

The inclusion of patients originating from the endemic areas could not be executed by the K-M plotter due to a sample size of less than 300, as stated by the K-M plotter. Due to this, false positives might have occurred, and it is probable to conclude our results were only mere predictions, and further clinical validation of NPC patients' OS, especially EBV-positive NPC, specifically must be done in the future to verify our findings.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

Based on the findings, bioinformatics analysis can reveal the candidate EBV BART miRNAs and the targeted RNAs for EBV-positive nasopharyngeal cancer (NPC). This gave insight into understanding the regulatory mechanisms of RNAs on other RNAs through a competing endogenous RNA (ceRNA) network and the regulatory mechanisms of BART miRNAs on mRNAs through cross-regulatory network. Interestingly, through the cross-regulatory network, six of the candidate BART miRNAs were demonstrated to regulate the expression of three lncRNAs indirectly by mimicking the host miRNAs. Through these constructed networks, it is rational to conclude that BART miRNAs and lncRNAs act similarly to oncogenes and tumor suppressors due to their ability to regulate mRNAs and miRNAs expressions directly and indirectly. From the ceRNA network, lncRNAs were suggested to be the culprits in down-regulating the mRNAs expressions due to one of their functions competing for miRNAs.

In addition, bioinformatics analysis also gave insight into the affected biological and molecular processes in cells, such as DNA repair and apoptosis in EBV-positive NPC due to the presence of BART miRNAs that interact and down-regulate the target mRNAs responsible for regulating these processes. Defects in these processes are common in the pathogenesis of various cancers, including NPC making the down-regulated EYA1, EYA4, EBF1 and MACROD2 mRNAs plausible as candidate biomarkers in EBV-positive NPC.

Apart from this, by using survival curves, NPC patients' overall survival (OS) could be predicted by generating curves based on the four down-regulated mRNAs. The curves revealed that EBF1 down-regulation led to consistently poor prognosis in patients compared to MACROD2, EYA1 and EYA4. These three mRNAs exhibited inconsistent poor overall survival (OS), shown only at the end of the study. These curves were only mere predictions as the databases used by the Kaplan-Meier (K-M) plotter were from head and neck cancer. Further clinical verification is necessary to validate these findings to rule out the use of these mRNAs as EBV-positive NPC biomarkers. In short, these findings served as preliminary data that affirms the potential of these BART miRNAs and lncRNAs as biomarkers for early diagnostic and therapeutic targets in EBV-positive NPC.

6.2 Recommendations

As plausible as the interactions in ceRNA and cross-regulatory networks, these candidates RNAs and BART miRNAs were only mere predicted biomarkers until verified by using experimental and clinical verifications, such as quantitative real-time PCR, western blot, gene knockdown, MTT assay and immunohistochemistry analyses. The availability of RNA microarray datasets was also limited, which contributes to only a few predicted RNA biomarkers obtained. This could be avoided by curating RNA datasets analysed using a more powerful and updated technique such as RNA-sequencing (RNA-seq). This technique quantifies the expressions of RNAs by sequencing the whole transcriptome and can quantify high-fold and low-fold RNA expressions, making it more desirable in recent years. This was seen in the availability of RNA-seq datasets are more than microarray datasets.

Furthermore, using our findings, a possible follow-up study by constructing a 2D or 3D structure of the ceRNA and BART miRNA-mRNA interaction is more valuable to ensure high accuracy in binding as RNA interactions are dynamic in nature. This also allows for clear visualisation of the binding sites using more updated software than Cytoscape. Visualising the binding sites between the RNAs can be used to predict motifs that could dock with therapeutic drugs.

REFERENCES

- Abusalah, M. A. H., Irekeola, A. A., Hanim Shueb, R., Jarrar, M. T., & Yean Yean, C. (2022). Prognostic Epstein-Barr Virus (EBV) miRNA biomarkers for survival outcome in EBV-associated epithelial malignancies: Systematic review and meta-analysis. *PloS One*, 17(4), e0266893.
- Adham, M., Kurniawan, A. N., Muhtadi, A. I., Roezin, A., Hermani, B., Gondhowiardjo, S., Roezin, A., & Middeldorp, J. M. (2012). Nasopharyngeal carcinoma in Indonesia: epidemiology, incidence, signs, and symptoms at presentation. *Chinese Journal of Cancer*, 31(4), 185.
- Alvarez-Garcia, I., & Miska, E. A. (2005). MicroRNA functions in animal development and human disease. *Development*, 132(21), 4653-4662.
- Amoroso, R., Fitzsimmons, L., Thomas, W. A., Kelly, G. L., Rowe, M., & Bell, A. I. (2011). Quantitative studies of Epstein-Barr virus-encoded microRNAs provide novel insights into their regulation. *Journal of Virology*, 85(2), 996-1010.
- Angelov, P., Gegov, A., Jayne, C., & Shen, Q. (2016). *Advances in Computational Intelligence Systems*, Cham: Springer International Publishing.
- Arun, G., Diermeier, S. D., & Spector, D. L. (2018). Therapeutic targeting of long non-coding RNAs in cancer. *Trends in Molecular Medicine*, 24(3), 257-277.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Harris, M. A. (2000). Gene ontology: tool for the unification of biology. *Nature Genetics*, 25(1), 25.

- Assenov, Y., Ramírez, F., Schelhorn, S. E., Lengauer, T., & Albrecht, M. (2007). Computing topological parameters of biological networks. *Bioinformatics*, 24(2), 282-284.
- Bader, G., Pavlovic, V., & Lopes, C. (2020). *MCODE Documentation release 2.0.0*. [online] Available at: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://mcode.readthedocs.io/_/downloads/en/latest/pdf/ [Accessed on 23 March 2021].
- Bal, R., & Deshmukh, P. (2022). Management of eustachian tube dysfunction: A Review. *Cureus*, 14(11).
- Bar-Joseph, Z., Gerber, G. K., Lee, T. I., Rinaldi, N. J., Yoo, J. Y., Robert, F., Gordon, D. B., Fraenkel, E., Jaakkola, T. S., Young, R. A., & Gifford, D. K. (2003). Computational discovery of gene modules and regulatory networks. *Nature Biotechnology*, 21(11), 1337-1342.
- Bayat, A. (2002). Science, medicine, and the future: Bioinformatics. *BMJ: British Medical Journal*, 324(7344), 1018.
- Baz-Redón, N., Rovira-Amigo, S., Fernández-Cancio, M., Castillo-Corullón, S., Cols, M., Caballero-Rabasco, M. A., Asensio, Ó., de Vicente, C. M., Gartner, S., Iglesias-Serrano, I., Díez-Izquierdo, A., Polverino, E., Amengual-Pieras, E., Amaro-Rodríguez, R., Vendrell, M., Mumany, M., Pascual-Sánchez, M. T., Pérez-Dueñas, B., Reula, A., Escribano, A., Dasí, F., Armengot-Carceller, M., Garrido-Pontnou, M., Camats-Tarruella, N., & Moreno-Galdó, A. (2020). Immunofluorescence analysis as a diagnostic tool in a Spanish cohort of patients with suspected primary ciliary dyskinesia. *Journal of Clinical Medicine*, 9(11), 3603.
- Bazan, R., Schröfel, A., Joachimiak, E., Poprzeczko, M., Pigino, G., & Wloga, D. (2021).

- Ccdc113/Ccdc96 complex, a novel regulator of ciliary beating that connects radial spoke 3 to dynein g and the nexin link. *PLoS Genetics*, 17(3), e1009388.
- Belinsky, S. A., Devereux, T. R., & Anderson, M. W. (1990). Role of DNA methylation in the activation of proto-oncogenes and the induction of pulmonary neoplasia by nitrosamines. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 233(1-2), 105-116.
- Berg, J. S., Evans, J. P., Leigh, M. W., Omran, H., Bizon, C., Mane, K., Knowles, M. R., Weck, K. E., & Zariwala, M. A. (2011). Next generation massively parallel sequencing of targeted exomes to identify genetic mutations in primary ciliary dyskinesia: implications for application to clinical testing. *Genetics in Medicine*, 13(3), 218-229.
- Bergström, S. E., King Jr, T. E., & Hollingsworth, H. (2012). *Primary ciliary dyskinesia (immotile-cilia syndrome)*. [online] Available at: <https://www.uptodate.com/contents/primary-ciliary-dyskinesia-immotile-cilia-syndrome> [Accessed on September 8 2016].
- Bhagwat, A. S., & Vakoc, C. R. (2015). Targeting transcription factors in cancer. *Trends in Cancer*, 1(1), 53-65.
- Bhat, S. A., Ahmad, S. M., Mumtaz, P. T., Malik, A. A., Dar, M. A., Urwat, U., Shah, R. A., & Ganai, N. A. (2016). Long non-coding RNAs: Mechanism of action and functional utility. *Non-Coding RNA Research*, 1(1), 43-50.
- Bhattacharjee, R., Prabhakar, N., Kumar, L., Bhattacharjee, A., Kar, S., Malik, S., Kumar, D., Ruokolainen, J., Negi, A., Jha, N. K., & Kesari, K. K. (2023). Crosstalk between long noncoding RNA and microRNA in Cancer. *Cellular Oncology*, 46(4), 1-24.

- Blagoev, K. B., Wilkerson, J., & Fojo, T. (2012). Hazard ratios in cancer clinical trials—a primer. *Nature Reviews Clinical Oncology*, 9(3), 178-183.
- Bicknell, A. A., & Ricci, E. P. (2017). When mRNA translation meets decay. *Biochemical Society Transactions*, 45(2), 339-351.
- Boller, S., & Grosschedl, R. (2014). The regulatory network of B-cell differentiation: a focused view of early B-cell factor 1 function. *Immunological Reviews*, 261(1), 102-115.
- Braschi, B., Omran, H., Witman, G. B., Pazour, G. J., Pfister, K. K., Bruford, E. A., & King, S. M. (2022). Consensus nomenclature for dyneins and associated assembly factors. *Journal of Cell Biology*, 221(2), e202109014.
- Borza, C. M., & Hutt-Fletcher, L. M. (2002). Alternate replication in B cells and epithelial cells switches tropism of Epstein–Barr virus. *Nature Medicine*, 8(6), 594-599.
- Brody, T. (2016). *Study Design Endpoints and Biomarkers Drug Safety and FDA and ICH Guidelines*. [e-book] Chennai: Mica Haley. Available through: https://books.google.com.my/books?id=zgNKCgAAQBAJ&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false [Accessed on 27 May 2023].
- Bossi, P., Chan, A. T., Licitra, L., Trama, A., Orlandi, E., Hui, E. P., Halámková, J., Matthesis, S., Baujat, B., Hardillo, J., Smeele, L., Herpen, V., Castro, A., & Machiels, J. P. (2021). Nasopharyngeal carcinoma: ESMO-EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, 32(4), 452-465.
- Bush, A., & O’Callaghan, C. (2002). Primary ciliary dyskinesia. *Archives of Disease in Childhood*, 87(5), 363-365.

- Cai, Y., Hao, Y., Ren, H., Dang, Z., Xu, H., Xue, X., & Gao, Y. (2019). miR-1305 inhibits the progression of non-small cell lung cancer by regulating MDM2. *Cancer Management and Research*, 11, 9529-9540.
- Cai, X., Schäfer, A., Lu, S., Bilello, J. P., Desrosiers, R. C., Edwards, R., Raab-Traub, N., & Cullen, B. R. (2006). Epstein–Barr virus microRNAs are evolutionarily conserved and differentially expressed. *PLoS Pathogens*, 2(3), e23.
- Casale J, Shumway KR, Hatcher JD. (2023). *Physiology, eustachian tube function*. [e-book] Treasure Island: StatPearls Publishing. Available through: <https://www.ncbi.nlm.nih.gov/books/NBK532284/> [Accessed on 27 May 2023].
- Celeste, A., Fernandez-Capetillo, O., Kruhlak, M. J., Pilch, D. R., Staudt, D. W., Lee, A., Bonner, R. F., Bonner, W. M., & Nussenzweig, A. (2003). Histone H2AX phosphorylation is dispensable for the initial recognition of DNA breaks. *Nature Cell Biology*, 5(7), 675-679.
- Cesana, M., & Daley, G. Q. (2013). Deciphering the rules of ceRNA networks. *Proceedings of The National Academy of Sciences*, 110(18), 7112-7113.
- Chan, J. K. (2017). Virus-associated neoplasms of the nasopharynx and sinonasal tract: diagnostic problems. *Modern Pathology*, 30(s1), S68-S83.
- Chang, E. T., & Adami, H. O. (2006). The enigmatic epidemiology of nasopharyngeal carcinoma. *Cancer Epidemiology and Prevention Biomarkers*, 15(10), 1765-1777.
- Chawla, K., Tripathi, S., Thommesen, L., Lægreid, A., & Kuiper, M. (2013). TFcheckpoint: a curated compendium of specific DNA-binding RNA polymerase II transcription factors. *Bioinformatics*, 29(19), 2519-2520.
- Chen, S. J., Chen, G. H., Chen, Y. H., Liu, C. Y., Chang, K. P., Chang, Y. S., & Chen, H. C. (2010). Characterization of Epstein-Barr virus miRNAome in nasopharyngeal

- carcinoma by deep sequencing. *PloS One*, 5(9), e12745.
- Chen, C., Lu, Z., Yang, J., Hao, W., Qin, Y., Wang, H., Xie, C., & Xie, R. (2016). MiR-17-5p promotes cancer cell proliferation and tumorigenesis in nasopharyngeal carcinoma by targeting p21. *Cancer Medicine*, 5(12), 3489-3499.
- Chen, H., Yang, Y., Wang, J., Shen, D., Zhao, J., & Yu, Q. (2018). miR-130b-5p promotes proliferation, migration and invasion of gastric cancer cells via targeting RASAL1. *Oncology Letters*, 15(5), 6361-6367.
- Chen, F., Shen, C., Wang, X., Wang, H., Liu, Y., & Yu, C. (2017). Identification of genes and pathways in nasopharyngeal carcinoma by bioinformatics analysis. *Oncotarget*, 8(38), 63738.
- Chen, J., Yang, R., Zhang, W., & Wang, Y. (2015). Candidate pathways and genes for nasopharyngeal carcinoma based on bioinformatics study. *International Journal of Clinical and Experimental Pathology*, 8(2), 2026.
- Chen, S. H., & Yu, X. (2019). Targeting dePARylation selectively suppresses DNA repair-defective and PARP inhibitor-resistant malignancies. *Science Advances*, 5(4), eaav4340.
- Chen, C., Zhao, J., Liu, J. N., & Sun, C. (2021). Mechanism and role of the neuropeptide LGI1 receptor ADAM23 in regulating biomarkers of ferroptosis and progression of esophageal cancer. *Disease Markers*, 2021.
- Cheng, Y. (2022). miR-135b-5p Targets SIRT1 to Inhibit Deacetylation of c-JUN and Increase MMP7 Expression to Promote Migration and Invasion of Nasopharyngeal Carcinoma Cells. *Molecular Biotechnology*, 64(6), 693-701.
- Cheng, Y., Ni, Y. J., & Tang, L. M. (2023). ZNF521/EBF1 axis regulates AKR1B1 to promote the proliferation, migration, and invasion of gastric cancer cells. *The*

Kaohsiung Journal of Medical Sciences, 39(3), 244-253.

- Chesnokova, L. S., Nishimura, S. L., & Hutt-Fletcher, L. M. (2009). Fusion of epithelial cells by Epstein–Barr virus proteins is triggered by binding of viral glycoproteins gHgL to integrins $\alpha\text{v}\beta 6$ or $\alpha\text{v}\beta 8$. *Proceedings of the National Academy of Sciences*, 106(48), 20464-20469.
- Choy, E. Y. W., Siu, K. L., Kok, K. H., Lung, R. W. M., Tsang, C. M., To, K. F., Kwong, D. L. W., Tsao, S. W., & Jin, D. Y. (2008). An Epstein-Barr virus–encoded microRNA targets PUMA to promote host cell survival. *The Journal of Experimental Medicine*, 205(11), 2551-2560.
- Chin, Y. M., Tan, L. P., Abdul Aziz, N., Mushiroda, T., Kubo, M., Mohd Kornain, N. K., Tan, G. W., Khoon, A. S. B., Krishnan, G., Pua, K. C., Yap, Y. Y., Teo, S. H., Lim, P. V. H., Nakamura, Y., Lum, C. L., & Ng, C. C. (2016). Integrated pathway analysis of nasopharyngeal carcinoma implicates the axonemal dynein complex in the Malaysian cohort. *International Journal of Cancer*, 139(8), 1731-1739.
- Clough, E., & Barrett, T. (2016). The gene expression omnibus database. *In Statistical Genomics*, 1418, 93-110.
- Cook, P. J., Ju, B. G., Telese, F., Wang, X., Glass, C. K., & Rosenfeld, M. G. (2009). Tyrosine dephosphorylation of H2AX modulates apoptosis and survival decisions. *Nature*, 458(7238), 591-596.
- Cosmopoulos, K., Pegtel, M., Hawkins, J., Moffett, H., Novina, C., Middeldorp, J., & Thorley-Lawson, D. A. (2009). Comprehensive profiling of Epstein-Barr virus microRNAs in nasopharyngeal carcinoma. *Journal of Virology*, 83(5), 2357-2367.
- Croce, C. M. (2009). Causes and consequences of microRNA dysregulation in cancer. *Nature Reviews Genetics*, 10(10), 704-714.

- Dai, X., Xie, Y., Dong, M., Zhao, J., Yu, H., Zhou, B., Xu, Y., Yu, Y., Cao, Y., & Zhang, Y. (2019). The long noncoding RNA TPTE2P1 promotes the viability of colorectal cancer cells. *Journal of Cellular Biochemistry*, 120(4), 5268-5276.
- Depreux, F. F., Darrow, K., Conner, D. A., Eavey, R. D., Liberman, M. C., Seidman, C. E., & Seidman, J. G. (2008). Eya4-deficient mice are a model for heritable otitis media. *The Journal of Clinical Investigation*, 118(2), 651-658.
- Devi, B. C., Pisani, P., Tang, T. S., & Parkin, D. M. (2004). High incidence of nasopharyngeal carcinoma in native people of Sarawak, Borneo Island. *Cancer Epidemiology and Prevention Biomarkers*, 13(3), 482-486.
- Dhanasekaran, D. N., & Reddy, E. P. (2008). JNK signaling in apoptosis. *Oncogene*, 27(48), 6245-6251.
- Dhanasekaran, D. N., & Reddy, E. P. (2017). JNK-signaling: A multiplexing hub in programmed cell death. *Genes & Cancer*, 8(9-10), 682-694.
- Dhanoo, J. K., Sethi, R. S., Verma, R., Arora, J. S., & Mukhopadhyay, C. S. (2018). Long non-coding RNA: its evolutionary relics and biological implications in mammals: a review. *Journal of Animal Science and Technology*, 60, 1-10.
- Dinant, C., Houtsmuller, A. B., & Vermeulen, W. (2008). Chromatin structure and DNA damage repair. *Epigenetics & Chromatin*, 1(1), 1-13.
- Doench, J. G., Petersen, C. P., & Sharp, P. A. (2003). siRNAs can function as miRNAs. *Genes & Development*, 17(4), 438-442.
- Dowd, J. B., Palermo, T., Brite, J., McDade, T. W., & Aiello, A. (2013). Seroprevalence of Epstein-Barr virus infection in US children ages 6-19, 2003-2010. *PloS One*, 8(5), e64921.
- Du, R., Jiang, F., Yin, Y., Xu, J., Li, X., Hu, L., & Wang, X. (2021). Knockdown of

- lncRNA X inactive specific transcript (XIST) radiosensitizes non-small cell lung cancer (NSCLC) cells through regulation of miR-16-5p/WEE1 G2 checkpoint kinase (WEE1) axis. *International Journal of Immunopathology and Pharmacology*, 35, 2058738420966087.
- Dunmire, S. K., Verghese, P. S., & Balfour Jr, H. H. (2018). Primary Epstein-barr virus infection. *Journal of Clinical Virology*, 102, 84-92.
- du Plessis, L., Škunca, N., & Dessimoz, C. (2011). The what, where, how and why of gene ontology—a primer for bioinformaticians. *Briefings in Bioinformatics*, 12(6), 723-735.
- Dweep, H., & Gretz, N. (2015). miRWalk2. 0: a comprehensive atlas of microRNA-target interactions. *Nature Methods*, 12(8), 697.
- Dwijayanti, F., Prabawa, A., Besral, & Herawati, C. (2020). The five-year survival rate of patients with nasopharyngeal carcinoma based on tumor response after receiving neoadjuvant chemotherapy, followed by chemoradiation, in Indonesia: a retrospective study. *Oncology*, 98(3), 154-160.
- Ebert, M. S., & Sharp, P. A. (2010). MicroRNA sponges: progress and possibilities. *Rna*, 16(11), 2043-2050.
- Fan, C., Tang, Y., Wang, J., Xiong, F., Guo, C., Wang, Y., Xiang, B., Zhou, M., Li, X., Wu, X., Li, Y., Li, X., Li, G., Xiong, W., & Zeng, Z. (2018). The emerging role of Epstein-Barr virus encoded microRNAs in nasopharyngeal carcinoma. *Journal of Cancer*, 9(16), 2852-2864.
- Fatica, A., & Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. *Nature Reviews Genetics*, 15(1), 7-21.
- Fernández-Cortés, M., Delgado-Bellido, D., & Oliver, F. J. (2019). Vasculogenic mimicry:

- become an endothelial cell “but not so much”. *Frontiers in Oncology*, 9, 803.
- Frappier, L. (2021). Epstein-Barr virus: Current questions and challenges. *Tumour Virus Research*, 12, 200218.
- Farrell, P. J. (2019). Epstein-Barr virus and cancer. *Annual Review of Pathology: Mechanisms of Disease*, 14, 29-53.
- Frappier, L. (2019). Epstein-Barr virus: overcoming the DNA damage response. *Future Virology*, 14(5), 349-360.
- Fugl, A., & Andersen, C. L. (2019). Epstein-Barr virus and its association with disease-a review of relevance to general practice. *BMC Family Practice*, 20(1), 1-8.
- Gao, Z., Fu, P., Yu, Z., Zhen, F., & Gu, Y. (2019). Comprehensive analysis of lncRNA-miRNA-mRNA network ascertains prognostic factors in patients with colon cancer. *Technology in Cancer Research & Treatment*, 18, 1533033819853237.
- Gao, L., Zhao, Y., Ma, X., & Zhang, L. (2021). Integrated analysis of lncRNA-miRNA-mRNA ceRNA network and the potential prognosis indicators in sarcomas. *BMC Medical Genomics*, 14(1), 67.
- Gizurarson, S. (2015). The effect of cilia and the mucociliary clearance on successful drug delivery. *Biological and Pharmaceutical Bulletin*, 38(4), 497-506.
- Global Cancer Observatory. (2021). *Population fact sheets*. [online] Available at: <https://gco.iarc.fr/today/fact-sheets-populations> [Accessed on 25 May 2023].
- Golia, B., Moeller, G. K., Jankevicius, G., Schmidt, A., Hegele, A., Preißer, J., Tran, M. L., Imhof, A., & Timinszky, G. (2017). ATM induces MacroD2 nuclear export upon DNA damage. *Nucleic Acids Research*, 45(1), 244-254.
- Gu, F., Yuan, S., Liu, L., Zhu, P., Yang, Y., Pan, Z., & Zhou, W. (2019). EYA4 serves as a prognostic biomarker in hepatocellular carcinoma and suppresses tumour

- angiogenesis and metastasis. *Journal of Cellular and Molecular Medicine*, 23(6), 4208-4216.
- Gullo, C. A., Low, W. K., & Teoh, G. (2008). Association of Epstein-Barr virus with nasopharyngeal carcinoma and current status of development of cancer-derived cell lines. *Annals of the Academy of Medicine Singapore*, 37(9), 769-77.
- Gunvén, P., Klein, G., Henle, G., Henle, W., & Clifford, P. (1970). Epstein–Barr Virus in Burkitt's Lymphoma and Nasopharyngeal Carcinoma: Antibodies to EBV associated Membrane and Viral Capsid Antigens in Burkitt Lymphoma Patients. *Nature*, 228(5276), 1053-1056.
- Gupta, P. K. (2014). Competing endogenous RNA (ceRNA): a new class of RNA working as miRNA sponges. *Current Science*, 106(6), 823-830.
- Gyórfy, B. (2023). Discovery and ranking of the most robust prognostic biomarkers in serous ovarian cancer. *Geroscience*, 45(3), 1-10.
- Hagman, J., Ramírez, J., & Lukin, K. (2011). B lymphocyte lineage specification, commitment and epigenetic control of transcription by early B cell factor 1. *Epigenetic Regulation of Lymphocyte Development*, 356, 17-38.
- Han, J. D. J., Bertin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V., Dupuy, D., Walhout, A. J. M., Cusick, M. E., Roth, F. P., & Vidal, M. (2004). Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature*, 430(6995), 88-93.
- Hao, S. P., Ngan, M. T., & Kupferman, M. E. (2008). *Nasopharyngeal carcinoma*. [e-book] San Diego, CA: Plural Publishing. Available through: <https://books.google.ki/books?id=JOEzBwAAQBAJ&printsec=frontcover#v=onepage&q&f=false> [Accessed on 23 March 2023].

- Hau, P. M., Lung, H. L., Wu, M., Tsang, C. M., Wong, K. L., Mak, N. K., & Lo, K. W. (2020). Targeting Epstein-Barr virus in nasopharyngeal carcinoma. *Frontiers in Oncology*, 10, 600.
- Hau, P. M., & Tsao, S. W. (2017). Epstein–Barr virus hijacks DNA damage response transducers to orchestrate its life cycle. *Viruses*, 9(11), 341.
- Hatton, O. L., Harris-Arnold, A., Schaffert, S., Krams, S. M., & Martinez, O. M. (2014). The interplay between Epstein–Barr virus and B lymphocytes: implications for infection, immunity, and disease. *Immunologic Research*, 58(2), 268-276.
- He, B., Li, W., Wu, Y., Wei, F., Gong, Z., Bo, H., Wang, Y., Li, X., Xiang, B., Guo, C., Liao, Q., Chen, P., Zu, X., Zhou, M., Ma, J., Li, X., Li, Y., Li, G., Xiong, W., & Zeng, Z. (2016). Epstein-Barr virus-encoded miR-BART6-3p inhibits cancer cell metastasis and invasion by targeting long non-coding RNA LOC553103. *Cell Death & Disease*, 7(9), e2353-e2353.
- He, Q., Ren, X., Chen, J., Li, Y., Tang, X., Wen, X., Yang, X., Zhang, J., Wang, Y., Ma, J., & Liu, N. (2016). miR-16 targets fibroblast growth factor 2 to inhibit NPC cell proliferation and invasion via PI3K/AKT and MAPK signaling pathways. *Oncotarget*, 7(3), 3047- 3058.
- Ho, K. Y., Lee, K. W., Chai, C. Y., Kuo, W. R., Wang, H. M., & Chien, C. Y. (2008). Early recognition of nasopharyngeal cancer in adults with only otitis media with effusion. *Journal of Otolaryngology--Head & Neck Surgery*, 37(3), 362–365.
- Hombach, S., & Kretz, M. (2016). Non-coding RNAs: classification, biology and functioning. *Non-coding RNAs in Colorectal Cancer*, 937, 3-17.
- Hoover, K., & Higginbotham, K. (2023). Epstein Barr Virus. [e-book] StatPearls: StatPearls Publishing. Available through

<https://www.ncbi.nlm.nih.gov/books/NBK559285/> [Accessed on 23 March 2023].

- Hornef, N., Olbrich, H., Horvath, J., Zariwala, M. A., Fliegauf, M., Loges, N. T., Wildhaber, J., Noone, P. G., Kennedy, M., Antonarakis, S. E., Blouin, J. L., Bartoloni, L., Nüsslein, T., Ahrens, T., Griesse, M., Kuhl, H., Sudbrak, R., Knowles, M. R., Reinhart, R., & Omran, H. (2006). DNAH5 mutations are a common cause of primary ciliary dyskinesia with outer dynein arm defects. *American Journal of Respiratory and Critical Care Medicine*, 174(2), 120-126.
- Hou, R., Liu, X., Yang, H., Chen, C., Liu, J., Li, Y., Zhang, Y., Deng, S., Xie, Y., Liu, X., Liu, Z., & Fang, W. Y. (2021). *Cinobufotalin, a Small Chemical Drug, Suppresses Nasopharyngeal Carcinoma Metastasis by Inducing ENKUR to Depress β -Catenin/C-Jun/MYH9 Signal-Mediated p53 Ubiquitination*. SSRN. <http://dx.doi.org/10.2139/ssrn.3854572>
- Hu, W., Xie, Q., Xu, Y., Tang, X., & Zhao, H. (2020). Integrated bioinformatics analysis reveals function and regulatory network of miR-200b-3p in endometriosis. *BioMed Research International*, 2020, 1-7.
- Hu, Y., Wang, J., Qian, J., Kong, X., Tang, J., Wang, Y., Chen, H., Hong, J., Zou, W., Chen, Y., Xu, J., & Fang, J. Y. (2014). Long noncoding RNA GAPLINC regulates CD44-dependent cell invasiveness and associates with poor prognosis of gastric cancer lncRNA GAPLINC regulates CD44 oncogene. *Cancer Research*, 74(23), 6890-6902.
- Huang, W. Y., Lin, C. C., Jen, Y. M., Lin, K. T., Yang, M. H., Chen, C. M., Chang, Y. N., Sung, F. C., & Kao, C. H. (2012). Association between adult otitis media and nasopharyngeal cancer: a nationwide population-based cohort study. *Radiotherapy and Oncology*, 104(3), 338-342.

- Huang, H. Y., Lin, Y. C. D., Li, J., Huang, K. Y., Shrestha, S., Hong, H. C., Tang, Y., Chen, Y. G., Chen, Jin, C. N., Yu, Y., Xu, J. T., Li, Y. M., Cai, X. X., Zhou, Z. Y., Chen, X. H., Pei, Y. Y., Hu, L., Su, J. J., Cui, S. D., Wang, F. Xie, Y. Y., Ding, S. Y., Luo, M. F., Chou, C. H., Chang, N. W., Chen, K. W., Cheng, Y. H., Wan, X. H., Hsu, W. L., Lee, T. Y., We, F. X., & Huang, H. D. (2020). miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database. *Nucleic Acids Research*, 48(D1), D148-D154.
- Huang, Y., Xu, M., Jing, C., Wu, X., Chen, X., & Zhang, W. (2021). Extracellular vesicle-derived miR-320a targets ZC3H12B to inhibit tumorigenesis, invasion, and angiogenesis in ovarian cancer. *Discover Oncology*, 12(1), 51-63.
- IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. (2012). *Personal Habits and Indoor Combustions*. [e-book] Lyon: International Agency for Research on Cancer. Available through: National Center for Biotechnology Information <<https://www.ncbi.nlm.nih.gov/books/NBK304384/>> [Accessed on 3 January 2023].
- Islam, Z., Ali, A. M., Naik, A., Eldaw, M., Decock, J., & Kolatkar, P. R. (2021). Transcription factors: The fulcrum between cell development and carcinogenesis. *Frontiers in Oncology*, 11, 681377.
- Jalali, S., Bhartiya, D., Lalwani, M. K., Sivasubbu, S., & Scaria, V. (2013). Systematic transcriptome wide analysis of lncRNA-miRNA interactions. *PloS One*, 8(2), 1-9.
- Jeggari, A., Marks, D. S., & Larsson, E. (2012). miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics*, 28(15), 2062-2063.
- Jha, H. C., Pei, Y., & Robertson, E. S. (2016). Epstein–Barr virus: diseases linked to

- infection and transformation. *Frontiers in Microbiology*, 7, 1602.
- Jiang, C., Li, L., Xiang, Y. Q., Lung, M. L., Zeng, T., Lu, J., Tsao, S. W., Zeng, M. S., Yun, J. P., Kwong, D. L. W., & Guan, X. Y. (2020). Epstein–Barr virus miRNA BART2-5p promotes metastasis of nasopharyngeal carcinoma by suppressing RND3BART2-5p promotes metastasis of nasopharyngeal carcinoma. *Cancer Research*, 80(10), 1957-1969.
- Jing, J. J., Wang, Z. Y., Li, H., Sun, L. P., & Yuan, Y. (2018). Key elements involved in Epstein–Barr virus-associated gastric cancer and their network regulation. *Cancer Cell International*, 18(1), 1-12.
- Jinushi, M., Takehara, T., Tatsumi, T., Hiramatsu, N., Sakamori, R., Yamaguchi, S., & Hayashi, N. (2005). Impairment of natural killer cell and dendritic cell functions by the soluble form of MHC class I-related chain A in advanced human hepatocellular carcinomas. *Journal of Hepatology*, 43(6), 1013-1020.
- Johnson, J. A., Watson, J. K., Nikolić, M. Z., & Rawlins, E. L. (2018). Fank1 and Jazf1 promote multiciliated cell differentiation in the mouse airway epithelium. *Biology Open*, 7(4), bio033944.
- Kalyanaraman, B. (2017). Teaching the basics of cancer metabolism: developing antitumor strategies by exploiting the differences between normal and cancer cell metabolism. *Redox Biology*, 12, 833-842.
- Kamiya, Y., Fujisawa, T., Katsumata, M., Yasui, H., Suzuki, Y., Karayama, M., Hozumi, H., Furuhashi, K., Enomoto, N., Nakamura, Y., Inui, N., Setou, M., Itom M., Suzuki, T., Koji, I., & Suda, T. (2020). Influenza A virus enhances ciliary activity and mucociliary clearance via TLR3 in airway epithelium. *Respiratory Research*, 21(1), 1-13.

- Kang, D., Skalsky, R. L., & Cullen, B. R. (2015). EBV BART microRNAs target multiple pro-apoptotic cellular genes to promote epithelial cell survival. *PLoS Pathogens*, 11(6), e1004979.
- Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1), 27-30.
- Karagkouni, D., Paraskevopoulou, M. D., Tastsoglou, S., Skoufos, G., Karavangeli, A., Pierros, V., Zacharopoulou, E., & Hatzigeorgiou, A. G. (2020). DIANA-LncBase v3: indexing experimentally supported miRNA targets on non-coding transcripts. *Nucleic Acids Research*, 48(D1), D101-D110.
- Kassab, M. A., Yu, L. L., & Yu, X. (2020). Targeting dePARylation for cancer therapy. *Cell & Bioscience*, 10(1), 1-9.
- Katerji, M., & Duerksen-Hughes, P. J. (2021). DNA damage in cancer development: special implications in viral oncogenesis. *American Journal of Cancer Research*, 11(8), 3956- 3979.
- Kenney, S. C. (2007). Reactivation and lytic replication of EBV. *Human herpesviruses: biology, therapy, and immunoprophylaxis*. [e-book] Cambridge: Cambridge University Press. Available through: <https://www.ncbi.nlm.nih.gov/books/NBK47442/> [Accessed on 27 May 2023].
- Khan, N. T. (2018). The Emerging Role of Bioinformatics in Biotechnology. *Journal of Biotechnology and Biomedical Science*, 1(3), 13-24.
- Kincaid, R. P., & Sullivan, C. S. (2012). Virus-encoded microRNAs: an overview and a look to the future. *PLoS Pathogens*, 8(12), e1003018.
- King, A. D. (2022). MR imaging of nasopharyngeal carcinoma. *Magnetic Resonance Imaging Clinics*, 30(1), 19-33.

- King, A. D., Woo, J. K. S., Ai, Q. Y., Chan, J. S. M., Lam, W. K. J., Tse, I. O. L., Bhatia, K. S., Zee, B. C. Y., Hui, E. P., Ma, B. B. Y., Chiu, R. W. K., van Hasselt, A. C., Chan, A. T. C., Lo, Y. M. D., & Chan, K. C. A. (2019). Complementary roles of MRI and endoscopic examination in the early detection of nasopharyngeal carcinoma. *Annals of Oncology*, 30(6), 977-982.
- Kong, D., Ma, W., Zhang, D., Cui, Q., Wang, K., Tang, J., Liu, Z., & Wu, G. (2019). EYA1 promotes cell migration and tumor metastasis in hepatocellular carcinoma. *American Journal of Translational Research*, 11(4), 2328- 2338.
- Ku, G. W., Kang, Y., Yu, S. L., Park, J., Park, S., Jeong, I. B., Kang, M. W., Son, J. S., & Kang, J. (2021). LncRNA LINC00240 suppresses invasion and migration in non-small cell lung cancer by sponging miR-7-5p. *BMC Cancer*, 21(1), 1-13.
- Kuri, A., Jacobs, B. M., Vickaryous, N., Pakpoor, J., Middeldorp, J., Giovannoni, G., & Dobson, R. (2020). Epidemiology of Epstein-Barr virus infection and infectious mononucleosis in the United Kingdom. *BMC Public Health*, 20(1), 1-9.
- Lánczky, A., & Györfy, B. (2021). Web-based survival analysis tool tailored for medical research (KMplot): development and implementation. *Journal of Medical Internet Research*, 23(7), e27633.
- Lee, L., & Ostrowski, L. E. (2021). Motile cilia genetics and cell biology: Big results from little mice. *Cellular and Molecular Life Sciences*, 78(3), 769-797.
- Lee, T. I., & Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. *Cell*, 152(6), 1237-1251.
- Lei, T., Yuen, K. S., Xu, R., Tsao, S. W., Chen, H., Li, M., Kok, K. H., & Jin, D. Y. (2013). Targeting of DICE1 tumor suppressor by Epstein–Barr virus-encoded miR-BART3* microRNA in nasopharyngeal carcinoma. *International Journal of*

Cancer, 133(1), 79-87.

- Li, Z., Cai, X., Zou, W., & Zhang, J. (2021). CDKN2B-AS1 promotes the proliferation, clone formation, and invasion of nasopharyngeal carcinoma cells by regulating miR-98-5p/E2F2 axis. *American Journal of Translational Research*, 13(12), 13406-13422.
- Li, J., Hu, C., Chao, H., Zhang, Y., Li, Y., Hou, J., & Huang, L. (2021). Exosomal transfer of miR-106a-5p contributes to cisplatin resistance and tumorigenesis in nasopharyngeal carcinoma. *Journal of Cellular and Molecular Medicine*, 25(19), 9183-9198.
- Li, J. H., Liu, S., Zhou, H., Qu, L. H., & Yang, J. H. (2014). starBase v2. 0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Research*, 42(D1), D92-D97.
- Li, Y., Wang, C., Miao, Z., Bi, X., Wu, D., Jin, N., Wang, L., Wu, H., Qian, K., Li, C., Zhang, T., Zhang, C., Yi, Y., Lai, H., Hu, Y., Cheng, L., Leung, K. S., Li, X., Zhang, F., Li, K., Li, X., & Wang, D. (2015). ViRBase: a resource for virus-host ncRNA-associated interactions. *Nucleic Acids Research*, 43(D1), D578-D582.
- Li, Y., Wu, Z., Yuan, J., Sun, L., Lin, L., Huang, N., Bin, J., & Liao, W. (2017). Long non-coding RNA MALAT1 promotes gastric cancer tumorigenicity and metastasis by regulating vasculogenic mimicry and angiogenesis. *Cancer Letters*, 395, 31-44.
- Li, H., Xie, S., Liu, M., Chen, Z., Liu, X., Wang, L., Li, D., & Zhou, Y. (2014). The clinical significance of downregulation of mir-124-3p, mir-146a-5p, mir-155-5p and mir-335-5p in gastric cancer tumorigenesis. *International Journal of Oncology*, 45(1), 197-208.
- Li, Y., Yagi, H., Onuoha, E. O., Damerla, R. R., Francis, R., Furutani, Y., Tariq, M., King,

- S. M., Hendricks, G., Cui, C., Saymohammed, M., Lee, D. M., Zahid, M., Sami, I., Leatherbury, L., Pazour, G. J., Ware, S. M., Nakanishi, T., Goldmuntz, E., Tsang, M., & Lo, C. W. (2016). DNAH6 and its interactions with PCD genes in heterotaxy and primary ciliary dyskinesia. *PLoS Genetics*, 12(2), e1005821.
- Liu, K., Kang, M., Zhou, Z., Qin, W., & Wang, R. (2019). Bioinformatics analysis identifies hub genes and pathways in nasopharyngeal carcinoma. *Oncology Letters*, 18(4), 3637-3645.
- Liu, X., Li, N., Liu, S., Wang, J., Zhang, N., Zheng, X., Kwong-Sak, L., & Cheng, L. (2019). Normalization methods for the analysis of unbalanced transcriptome data: a review. *Frontiers in Bioengineering and Biotechnology*, 7, 358.
- Liu, Y., Pan, B., Qu, W., Cao, Y., Li, J., & Zhao, H. (2021). *Kaplan–Meier plotter (KM plotter) database for survival analysis*. [online] Available at: <https://bio-protocol.org/exchange/minidetail?type=30&id=8993439> [Accessed on 4 April 2023].
- Liu, L., Wang, H., Yan, C., & Tao, S. (2020). An integrated analysis of mRNAs and miRNAs microarray profiles to screen miRNA signatures involved in nasopharyngeal carcinoma. *Technology in Cancer Research & Treatment*, 19, 1533033820956998.
- Liu, J., Wang, Z., Yin, Y., Li, N., Ye, N., Bao, B., & Gu, H. C. (2019). Long noncoding RNA TPTE2P1 promotes the migration and invasion of hepatocellular carcinoma. *European Review for Medical & Pharmacological Sciences*, 23(9), 3733-3741.
- Liu, B., Zhang, J., & Yang, D. (2019). miR-96-5p promotes the proliferation and migration of ovarian cancer cells by suppressing Caveolae1. *Journal of Ovarian Research*, 12(1), 1-9.

- Liang, W., & Sun, F. (2019). Identification of pivotal lncRNAs in papillary thyroid cancer using lncRNA–mRNA–miRNA ceRNA network analysis. *PeerJ*, 7, e7441.
- Locksley, R. M., Killeen, N., & Lenardo, M. J. (2001). The TNF and TNF receptor superfamilies: integrating mammalian biology. *Cell*, 104(4), 487-501.
- López-Urrutia, E., Bustamante Montes, L. P., Ladrón de Guevara Cervantes, D., Pérez-Plasencia, C., & Campos-Parra, A. D. (2019). Crosstalk between long non-coding RNAs, micro-RNAs and mRNAs: deciphering molecular mechanisms of master regulators in cancer. *Frontiers in Oncology*, 9, 669.
- Lu, M., Huang, Y., Sun, W., Li, P., Li, L., & Li, L. (2018). miR-135b-5p promotes gastric cancer progression by targeting CMTM3. *International Journal of Oncology*, 52(2), 589-598.
- Lung, R. W. M., Hau, P. M., Yu, K. H. O., Yip, K. Y., Tong, J. H. M., Chak, W. P., Chan, A. W. H., Lam, K. H., Lo, A. K. F., Tin, E. K. Y., Chau, S. L., Pang, J. C. S., Kwan, J. S. H., Busson, P., Young, L. S., Yap, L. F., Tsao, S. W., To, K. F., & Lo, K. W. (2018). EBV-encoded miRNAs target ATM-mediated response in nasopharyngeal carcinoma. *The Journal of Pathology*, 244(4), 394-407.
- Luo, X., He, X., Liu, X., Zhong, L., & Hu, W. (2020). miR-96-5p suppresses the progression of nasopharyngeal carcinoma by targeting CDK1. *OncoTargets and Therapy*, 13, 7467–7477.
- Luo, Z. B., Lai, G. E., Jiang, T., Cao, C. L., Peng, T., & Liu, F. E. (2020). A competing endogenous RNA network reveals novel lncRNA, miRNA and mRNA biomarkers with diagnostic and prognostic value for early breast cancer. *Technology in Cancer Research & Treatment*, 19, 1533033820983293.
- Luo, M., Li, Y., Shi, X., Yang, W., Zhou, F., Sun, N., & He, J. (2018). Aberrant

- methylation of EYA4 promotes epithelial-mesenchymal transition in esophageal squamous cell carcinoma. *Cancer Science*, 109(6), 1811-1824.
- Luo, Y., Wang, J., Wang, F., Liu, X., Lu, J., Yu, X., Ma, X., Peng, X., & Li, X. (2021). Foxq1 promotes metastasis of nasopharyngeal carcinoma by inducing vasculogenic mimicry via the EGFR signaling pathway. *Cell Death & Disease*, 12(5), 411.
- Luscombe, N. M., Greenbaum, D., & Gerstein, M. (2001). What is bioinformatics? A proposed definition and overview of the field. *Methods of Information in Medicine*, 40(04), 346-358.
- Lv, S. Y., Shan, T. D., Pan, X. T., Tian, Z. B., Liu, X. S., Liu, F. G., Sun, X. G., Xue, H. G., Li, X. H., Han, Y., Sun, L. J., Chen, L., & Zhang, L. Y. (2018). The lncRNA ZEB1-AS1 sponges miR-181a-5p to promote colorectal cancer cell proliferation by regulating Wnt/ β -catenin signaling. *Cell Cycle*, 17(10), 1245-1254.
- Ma, D. H., Li, B. S., Liu, J. J., Xiao, Y. F., Yong, X., Wang, S. M., Wu, Y. Y., Zhu, H. B., Wang, D. X., & Yang, S. M. (2017). miR-93-5p/IFNAR1 axis promotes gastric cancer metastasis through activating the STAT3 signaling pathway. *Cancer Letters*, 408, 23-32.
- Ma, H., Liu, Z., Li, H., Guo, X., Guo, S., Qu, P., & Wang, Y. (2021). Bioinformatics analysis reveals MCM3 as an important prognostic marker in cervical cancer. *Computational and Mathematical Methods in Medicine*, 2021, 8494260.
- MacFarlane, L. A., & R Murphy, P. (2010). MicroRNA: biogenesis, function and role in cancer. *Current Genomics*, 11(7), 537-561.
- Mak, H. W., Lee, S. H., Chee, J., Tham, I., Goh, B. C., Chao, S. S., Ong, Y. K., Loh, K. S., & Lim, C. M. (2015). Clinical outcome among nasopharyngeal cancer patients in a multi-ethnic society in Singapore. *PloS One*, 10(5), e0126108.

- Marquitz, A. R., & Raab-Traub, N. (2012, April). The role of miRNAs and EBV BARTs in NPC. *Seminars in Cancer Biology*, 22(2), 166-172.
- Martin, E., Schipper, N. G., Verhoef, J. C., & Merkus, F. W. (1998). Nasal mucociliary clearance as a factor in nasal drug delivery. *Advanced Drug Delivery Reviews*, 29(1-2), 13-38.
- Matsune, S., Sando, I., & Takahashi, H. (1992). Distributions of eustachian tube goblet cells and glands in children with and without otitis media. *Annals of Otolaryngology & Rhinology & Laryngology*, 101(9), 750-754.
- Max Planck Institute for Informatics. (2018). *NetworkAnalyzer Online Help*. [online] Available at: <https://med.bioinf.mpi-inf.mpg.de/netanalyzer/help/2.7/index.html#interpret> [Accessed on 5 January 2023].
- Mayo Clinic. (2018). *Nasopharyngeal carcinoma*. [online] Available at: https://www.mayoclinic.org/diseases_conditions/nasopharyngeal_carcinoma/symptoms-causes/syc-20375529 [Accessed on 17 October 2022].
- Mayo Clinic. (2022). *Tinnitus*. [online] Available at: https://www.mayoclinic.org/diseases-conditions/tinnitus/symptoms-causes/syc_20350156#:~:text=Your%20ear%20canals%20can%20become,brain%20function%20linked%20to%20hearing. [Accessed on 27 May 2023].
- Menzl, I., Lebeau, L., Pandey, R., Hassounah, N. B., Li, F. W., Nagle, R., Weihs, K., & McDermott, K. M. (2014). Loss of primary cilia occurs early in breast cancer development. *Cilia*, 3(1), 1-17.
- Miao, W. J., Yuan, D. J., Zhang, G. Z., Liu, Q., Ma, H. M., & Jin, Q. Q. (2019). lncRNA CASC2/miR-18a-5p axis regulates the malignant potential of nasopharyngeal carcinoma by targeting RBBP8. *Oncology Reports*, 41(3), 1797-1806.

- Mishra, N. K., Southekal, S., & Guda, C. (2019). Survival analysis of multi-omics data identifies potential prognostic markers of pancreatic ductal adenocarcinoma. *Frontiers in Genetics*, 10, 624.
- Mo, S. J., Hou, X., Hao, X. Y., Cai, J. P., Liu, X., Chen, W., Chen, D., & Yin, X. Y. (2018). EYA4 inhibits hepatocellular carcinoma growth and invasion by suppressing NF- κ B-dependent RAP1 transactivation. *Cancer Communications*, 38(1), 1-15.
- Mohseni, M., Cidado, J., Croessmann, S., Cravero, K., Cimino-Mathews, A., Wong, H. Y., Scharpf, R., Zabransky, D. J., Abukhdeir, A. M., Garay, J. P., Wang, G. M., Beaver, J. A., Cochran, R. L., Blair, B. G., Rosen, D. M., Erlanger, B., Argani, P., Hurley, P. J., Lauring, J., & Park, B. H. (2014). MACROD2 overexpression mediates estrogen independent growth and tamoxifen resistance in breast cancers. *Proceedings of The National Academy of Sciences*, 111(49), 17606-17611.
- Montoya-Ferrer, A., Sanosyan, A., Fayd'herbe de Maudave, A., Pisoni, A., Bollore, K., Molès, J. P., Peries, M., Tylleskar, T., Tumwine, J., K., Ndeezi, G., Gorgolas, M., Nagot, N, van de Perre, P., & Tuaillon, E. (2021). Clinical and biological factors associated with early Epstein-Barr virus infection in human immunodeficiency virus–exposed uninfected infants in Eastern Uganda. *Clinical Infectious Diseases*, 72(6), 1026-1032.
- Moon, S. H., Nguyen, T. A., Darlington, Y., Lu, X., & Donehower, L. A. (2010). Dephosphorylation of γ -H2AX by WIP1: an important homeostatic regulatory event in DNA repair and cell cycle control. *Cell Cycle*, 9(11), 2092-2096.
- Moore, A. C., Winkjer, J. S., & Tseng, T. T. (2015). Bioinformatics resources for microRNA discovery. *Biomarker Insights*, 10(Suppl 4), 53-58.

- Murata, T., Sato, Y., & Kimura, H. (2014). Modes of infection and oncogenesis by the Epstein–Barr virus. *Reviews in Medical Virology*, 24(4), 242-253.
- Murata, T., Sugimoto, A., Inagaki, T., Yanagi, Y., Watanabe, T., Sato, Y., & Kimura, H. (2021). Molecular basis of Epstein–Barr virus latency establishment and lytic reactivation. *Viruses*, 13(12), 2344.
- Muz, B., de la Puente, P., Azab, F., & Azab, A. K. (2015). The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. *Hypoxia (Auckland, N.Z.)*, 3, 83-92.
- Nachmani, D., Stern-Ginossar, N., Sarid, R., & Mandelboim, O. (2009). Diverse herpesvirus microRNAs target the stress-induced immune ligand MICB to escape recognition by natural killer cells. *Cell Host & Microbe*, 5(4), 376-385.
- Nair, A. A., Tang, X., Thompson, K. J., Vedell, P. T., Kalari, K. R., & Subramanian, S. (2020). Frequency of MicroRNA response elements identifies pathologically relevant signaling pathways in triple-negative breast cancer. *iScience*, 23(6), 101249.
- National Cancer Institute. (2018). *Nasopharyngeal Cancer Treatment (Adult) (PDQ®) - Patient Version*. [online] Available at: <https://www.cancer.gov/types/head-and-neck/patient/adult/nasopharyngeal-treatment-pdq> [Accessed on 17 October 2022].
- Nemerow, G. R., Wolfert, R., McNaughton, M. E., & Cooper, N. R. (1985). Identification and characterization of the Epstein-Barr virus receptor on human B lymphocytes and its relationship to the C3d complement receptor (CR2). *Journal of Virology*, 55(2), 347-351.
- Nikpour, P., Emadi-Baygi, M., Emadi-Andani, E., & Rahmati, S. (2014). EYA1 expression in gastric carcinoma and its association with clinicopathological characteristics: a

- pilot study. *Medical Oncology*, 31(5), 1-5.
- Ning, L., Ko, J. M. Y., Yu, V. Z., Ng, H. Y., Chan, C. K. C., Tao, L., Lam, S. Y., Man-Long, M., Ngan, R. K. C., Kwong, D. L. W., Lee, A. W. M., Ng, W. T., Cheng, A., Tung, S., Lee, V. H. F., Lam, K. O., Kwan, C. K., Li, W. S., Yau, S., Bei, J. X., & Lung, M. L. (2020). Nasopharyngeal carcinoma MHC region deep sequencing identifies HLA and novel non-HLA TRIM31 and TRIM 39 loci. *Communications Biology*, 3(1), 759.
- Nowsheen, S., Aziz, K., Luo, K., Deng, M., Qin, B., Yuan, J., Jeganathan, K. B., Yu, J., Zhang, H., Ding, W., van Deursen, J. M., & Lou, Z. (2018). ZNF506-dependent positive feedback loop regulates H2AX signaling after DNA damage. *Nature Communications*, 9(1), 1-11.
- O'Brien, J., Hayder, H., Zayed, Y., & Peng, C. (2018). Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Frontiers in Endocrinology*, 9, 402.
- Ou, H., Li, Y., & Kang, M. (2014). Activation of miR-21 by STAT3 induces proliferation and suppresses apoptosis in nasopharyngeal carcinoma by targeting PTEN gene. *PloS One*, 9(11), e109929.
- Ou, S. H., Zell, J. A., Ziogas, A., & Anton-Culver, H. (2007). Epidemiology of nasopharyngeal carcinoma in the United States: improved survival of Chinese patients within the keratinizing squamous cell carcinoma histology. *Annals of Oncology*, 18(1), 29-35.
- Pagano, J. S. (2007). Is Epstein-Barr virus transmitted sexually? *The Journal of Infectious Diseases*, 195(4), 469-470.
- Paraskevopoulou, M. D., & Hatzigeorgiou, A. G. (2016). Analyzing miRNA-lncRNA interactions. *Long Non-Coding RNAs: Methods and Protocols*, 1402, 271-286.

- Paraskevopoulou, M. D., Vlachos, I. S., Karagkouni, D., Georgakilas, G., Kanellos, I., Vergoulis, T., Zagganas, K., Tsanakas, P., Floros, E., Dalamagas, T., & Hatzigeorgiou, A. G. (2015). DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. *Nucleic Acids Research*, 44(D1), D231-D238.
- Parris, T. Z., Danielsson, A., Nemes, S., Kovács, A., Delle, U., Fallenius, G., Möllerström, E., Karlsson, P., & Helou, K. (2010). Clinical implications of gene dosage and gene expression patterns in diploid breast carcinoma integrative genomics in diploid breast cancer. *Clinical Cancer Research*, 16(15), 3860-3874.
- Pathmanathan, R., Prasad, U., Sadler, R., Flynn, K., & Raab-Traub, N. (1995). Clonal proliferations of cells infected with Epstein–Barr virus in preinvasive lesions related to nasopharyngeal carcinoma. *New England Journal of Medicine*, 333(11), 693-698.
- Pegtel, D. M., Cosmopoulos, K., Thorley-Lawson, D. A., van Eijndhoven, M. A., Hopmans, E. S., Lindenberg, J. L., de Gruijl, T. D., Würdinger, T., & Middeldorp, J. M. (2010). Functional delivery of viral miRNAs via exosomes. *Proceedings of the National Academy of Sciences*, 107(14), 6328-6333.
- Peng, Y., & Croce, C. M. (2016). The role of MicroRNAs in human cancer. *Signal Transduction and Targeted Therapy*, 1(1), 1-9.
- Peng, L., Gao, S., Bai, F., An, Y., & Wan, F. (2019). LncRNA TPTE2P1 promotes the proliferation of thyroid carcinoma by inhibiting miR-520c-3p. *Panminerva Medica*, 62(1), 69-71.
- Permyakov, E. A. (2021). Metal binding proteins. *Encyclopedia*, 1(1), 261-292.
- Plieskatt, J. L., Rinaldi, G., Feng, Y., Levine, P. H., Easley, S., Martinez, E., Hashmi, S., Sadeghi, N., Brindley, P. J., Bethony, J. M., & Mulvenna, J. P. (2014). Methods

- and matrices: approaches to identifying miRNAs for Nasopharyngeal carcinoma. *Journal of Translational Medicine*, 12(1), 1-20.
- Png, Y. T., Yang, A. Z. Y., Lee, M. Y., Chua, M. J. M., & Lim, C. M. (2021). The role of NK cells in EBV infection and EBV-associated NPC. *Viruses*, 13(2), 300.
- Poling, B. C., Price, A. M., Luftig, M. A., & Cullen, B. R. (2017). The Epstein-Barr virus miR-BHRF1 microRNAs regulate viral gene expression in cis. *Virology*, 512, 113-123.
- Prasad, U., & Rampal, L. (1992). Descriptive epidemiology of nasopharyngeal carcinoma in Peninsular Malaysia. *Cancer Causes & Control*, 3(2), 179-182.
- Qi, X., Li, J., Zhou, C., Lv, C., & Tian, M. (2014). MicroRNA-320a inhibits cell proliferation, migration and invasion by targeting BMI-1 in nasopharyngeal carcinoma. *FEBS Letters*, 588(20), 3732-3738.
- Qi, X., Zhang, D. H., Wu, N., Xiao, J. H., Wang, X., & Ma, W. (2015). ceRNA in cancer: possible functions and clinical implications. *Journal of Medical Genetics*, 52(10), 710-718.
- Qin, W. J., Wang, W. P., Wang, X. B., Zhang, X. T., & Du, J. D. (2019). MiR-1290 targets CCNG2 to promote the metastasis of oral squamous cell carcinoma. *European Review for Medical & Pharmacological Sciences*, 23(23), 10332-10342.
- Qin, X., Zhang, J., Lin, Y., Sun, X. M., Zhang, J. N., & Cheng, Z. Q. (2020). Identification of MiR-211-5p as a tumor suppressor by targeting ACSL4 in Hepatocellular Carcinoma. *Journal of Translational Medicine*, 18(1), 326.
- Rao, M. S., Van Vleet, T. R., Ciurlionis, R., Buck, W. R., Mittelstadt, S. W., Blomme, E. A., & Liguori, M. J. (2019). Comparison of RNA-Seq and microarray gene expression platforms for the toxicogenomic evaluation of liver from short-term rat

- toxicity studies. *Frontiers in Genetics*, 9, 636.
- Rashid, S., Breckle, R., Hupe, M., Geisler, S., Doerwald, N., & Neesen, J. (2006). The murine Dnali1 gene encodes a flagellar protein that interacts with the cytoplasmic dynein heavy chain 1. *Molecular Reproduction and Development: Incorporating Gamete Research*, 73(6), 784-794.
- Ratti, M., Lampis, A., Ghidini, M., Salati, M., Mirchev, M. B., Valeri, N., & Hahne, J. C. (2020). MicroRNAs (miRNAs) and long non-coding RNAs (lncRNAs) as new tools for cancer therapy: first steps from bench to bedside. *Targeted Oncology*, 15(3), 261-278.
- Ren, X. S., Yin, M. H., Zhang, X., Wang, Z., Feng, S. P., Wang, G. X., Luo, Y. J., Liang, P. Z., Yang, X. Q., He, J. X., & Zhang, B. L. (2014). Tumor-suppressive microRNA-449a induces growth arrest and senescence by targeting E2F3 in human lung cancer cells. *Cancer Letters*, 344(2), 195-203.
- Rich, J. T., Neely, J. G., Paniello, R. C., Voelker, C. C., Nussenbaum, B., & Wang, E. W. (2010). A practical guide to understanding Kaplan-Meier curves. *Otolaryngology—Head and Neck Surgery*, 143(3), 331-336.
- Ringers, C., Olstad, E. W., & Jurisch-Yaksi, N. (2020). The role of motile cilia in the development and physiology of the nervous system. *Philosophical Transactions of The Royal Society B*, 375(1792), 20190156.
- Robinson, J. M., & Henderson, W. A. (2018). Modelling the structure of a ceRNA-theoretical, bipartite microRNA–mRNA interaction network regulating intestinal epithelial cellular pathways using R programming. *BMC Research Notes*, 11(1), 1-7.
- Ross, N., Gandhi, M. K., & Nourse, J. P. (2013). The Epstein-Barr virus microRNA

- BART11-5p targets the early B-cell transcription factor EBF1. *American Journal of Blood Research*, 3(3), 210-224.
- Rupaimoole, R., Calin, G. A., Lopez-Berestein, G., & Sood, A. K. (2016). miRNA deregulation in cancer cells and the tumor microenvironment. *Cancer Discovery*, 6(3), 235-246.
- Sadatom, D., Tanimura, S., Ozaki, K. I., & Takeda, K. (2013). Atypical protein phosphatases: emerging players in cellular signaling. *International Journal of Molecular Sciences*, 14(3), 4596-4612.
- Sakthianandeswaren, A., Parsons, M. J., Mouradov, D., MacKinnon, R. N., Catimel, B., Liu, S., Palmieri, M., Love, C., Jorissen, R. N., Li, S., Whitehead, L., Putoczki, T. L., Preaudet, A., Tsui, C., Nowell, C. J., Ward, R. L., Hawkins, N. J., Desai, J., Gibbs, P., Ernst, M., Street, I., Buchert, M., & Sieber, O. M. (2018). MACROD2 Haploinsufficiency Impairs Catalytic Activity of PARP1 and Promotes Chromosome Instability and Growth of Intestinal Tumors. *MACROD2 Loss impairs PARP1 activity and promotes CIN in colorectal cancer. Cancer Discovery*, 8(8), 988-1005.
- Salmena, L., Poliseno, L., Tay, Y., Kats, L., & Pandolfi, P. P. (2011). A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*, 146(3), 353-358.
- Sarwari, N. M., Khoury, J. D., & Hernandez, C. M. R. (2016). Chronic Epstein Barr virus infection leading to classical Hodgkin lymphoma. *BMC Hematology*, 16(1), 1-6.
- Searight, F. T., Singh, R., & Peterson, D. C. (2022). *Otitis Media With Effusion*. [e-book] Treasure Island: StatPearls Publishing. Available through: <https://www.ncbi.nlm.nih.gov/books/NBK538293/> [Accessed on 27 May 2023].
- Sebastian-de-laCruz, M., Gonzalez-Moro, I., Olazagoitia-Garmendia, A., Castellanos-

- Rubio, A., & Santin, I. (2021). The role of lncRNAs in gene expression regulation through mRNA stabilization. *Non-Coding RNA*, 7(1), 3.
- Seo, C. H., Kim, J. R., Kim, M. S., & Cho, K. H. (2009). Hub genes with positive feedbacks function as master switches in developmental gene regulatory networks. *Bioinformatics*, 25(15), 1898-1904.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11), 2498-2504.
- Shao, Y. M., Poirier, A., Ohshima, H., Malaveille, C., Zeng, Y., Thé, G. D., & Bartsch, H. (1988). Epstein–Barr virus activation in Raji cells by extracts of preserved food from high risk areas for nasopharyngeal carcinoma. *Carcinogenesis*, 9(8), 1455-1457.
- Shen, Z., Chen, Y., Li, L., Liu, L., Peng, M., Chen, X., Wu, X., Sferra, T. J., Wu, M., Lin, X., Ying, C., Chu, J., Shen, A., & Peng, J. (2020). Transcription factor EBF1 overexpression suppresses tumor growth in vivo and in vitro via modulation of the PNO1/p53 pathway in colorectal cancer. *Frontiers in Oncology*, 10, 1035.
- Shi, T., Huang, L., & Tian, J. (2022). Prevalence of Epstein-Barr Viral DNA among children at a single hospital in Suzhou, China. *Jornal de Pediatria*, 98(2), 142-146.
- Shi, Y., Liu, J. B., Deng, J., Zou, D. Z., Wu, J. J., Cao, Y. H., Yin, J., Ma, Y. S., Da, F & Li, W. (2021). The role of ceRNA-mediated diagnosis and therapy in hepatocellular carcinoma. *Hereditas*, 158(1), 1-14.
- Shoemark, A., Frost, E., Dixon, M., Ollosson, S., Kilpin, K., Patel, M., Scully, J., Rogers, A. V., Mitchison, H. M., Bush, A., & Hogg, C. (2017). Accuracy of

- immunofluorescence in the diagnosis of primary ciliary dyskinesia. *American Journal of Respiratory and Critical Care Medicine*, 196(1), 94-101.
- Sigg, M. A., Menchen, T., Lee, C., Johnson, J., Jungnickel, M. K., Choksi, S. P., Garcia, G., Busengdal, H., Dougherty, G., Pennekamp, P., Werner, C., Rentzsch, F., Florman, H. M., Krogan, N., Wallingford, J. B., Omran, H., & Reiter, J. F. (2017). Evolutionary proteomics uncovers ancient associations of cilia with signaling pathways. *Developmental Cell*, 43(6), 744-762.
- Shyu, A. B., Wilkinson, M. F., & Van Hoof, A. (2008). Messenger RNA regulation: to translate or to degrade. *The EMBO Journal*, 27(3), 471-481.
- Skalsky, R. L., & Cullen, B. R. (2010). Viruses, microRNAs, and host interactions. *Annual Review of Microbiology*, 64, 123-141.
- Smatti, M. K., Al-Sadeq, D. W., Ali, N. H., Pintus, G., Abou-Saleh, H., & Nasrallah, G. K. (2018). Epstein–Barr virus epidemiology, serology, and genetic variability of LMP-1 oncogene among healthy population: an update. *Frontiers in Oncology*, 8, 211.
- Song, Y., Li, X., Zeng, Z., Li, Q., Gong, Z., Liao, Q., Li, X., Chen, P., Xiang, B., Zhang, W., Xiong, F. Zhou, Y., Zhou, M., Ma, J., Li, Y., Chen, X., Li, G., & Xiong, W. (2016). Epstein-Barr virus encoded miR-BART11 promotes inflammation-induced carcinogenesis by targeting FOXP1. *Oncotarget*, 7(24), 36783-36799.
- Spruance, S. L., Reid, J. E., Grace, M., & Samore, M. (2004). Hazard ratio in clinical trials. *Antimicrobial Agents and Chemotherapy*, 48(8), 2787-2792.
- Sticht, C., De La Torre, C., Parveen, A., & Gretz, N. (2018). miRWalk: an online resource for prediction of microRNA binding sites. *PloS One*, 13(10), e0206239.
- Stucki, M., Clapperton, J. A., Mohammad, D., Yaffe, M. B., Smerdon, S. J., & Jackson, S. P. (2005). MDC1 directly binds phosphorylated histone H2AX to regulate cellular

- responses to DNA double-strand breaks. *Cell*, 123(7), 1213-1226.
- Su, Z. Y., Siak, P. Y., Leong, C. O., & Cheah, S. C. (2023). The role of Epstein–Barr virus in nasopharyngeal carcinoma. *Frontiers in Microbiology*, 14, 1116143.
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Mering, C. V., & Jensen, L. J. (2018). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47(D1), D607-D613.
- Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & von Mering, C. (2023). The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, 51(D1), D638-D646.
- Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., von Mering, C., & Jensen, L. J. (2016). The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Research*, 45(D1), D362–D368.
- Tabuchi, K., Nakayama, M., Nishimura, B., Hayashi, K., & Hara, A. (2011). Early detection of nasopharyngeal carcinoma. *International Journal of Otolaryngology*, 2011, 638058.
- Tadjuidje, E., Wang, T. S., Pandey, R. N., Sumanas, S., Lang, R. A., & Hegde, R. S. (2012). The EYA tyrosine phosphatase activity is pro-angiogenic and is inhibited by benzbromarone. *PLoS One*, 7(4), e34806.
- Takada, K. (2000). Epstein-Barr virus and gastric carcinoma. *Molecular Pathology*, 53(5),

255.

- Tan, Y., Jiang, C., Jia, Q., Wang, J., Huang, G., & Tang, F. (2022). A novel oncogenic seRNA promotes nasopharyngeal carcinoma metastasis. *Cell Death & Disease*, 13(4), 401.
- Tang, L. L., Chen, Y. P., Chen, C. B., Chen, M. Y., Chen, N. Y., Chen, X. Z., Du, X. J., Fang, W. F., Feng, M., Gao, J., Han, F., He, X., Hu, C. S., Hu, D. S., Hu, G. Y., Jiang, H., Jiang, W., Jin, F., Lang, J. Y., Li, J. G., Lin, S. J., Liu, X., Liu, Q. F., Ma, L., Mai, H. Q., Qin, J. Y., Shen, L. F., Sun, Y., Wang, P. G., Wang, R. S., Wang, R. Z., Wang, X. S., Wang, Y., Wu, H., Xia, Y. F., Xiao, S. W., Yang, K. Y., Yi, J. L., Zhu, X. D., Ma, J., & Ma, J. (2021). The Chinese Society of Clinical Oncology (CSCO) clinical guidelines for the diagnosis and treatment of nasopharyngeal carcinoma. *Cancer Communications*, 41(11), 1195-1227.
- Tang, J., Kong, D., Cui, Q., Wang, K., Zhang, D., Yuan, Q., Liao, X., Gong, Y., & Wu, G. (2018). Bioinformatic analysis and identification of potential prognostic microRNAs and mRNAs in thyroid cancer. *PeerJ*, 6, e4674.
- Tang, S., Liao, K., Shi, Y., Tang, T., Cui, B., & Huang, Z. (2022). Bioinformatics analysis of potential Key lncRNA-miRNA-mRNA molecules as prognostic markers and important ceRNA axes in gastric cancer. *American Journal of Cancer Research*, 12(5), 2397-2418.
- Tang, J. F., Yu, Z. H., Liu, T., Lin, Z. Y., Wang, Y. H., Yang, L. W., He, H. J., Huang, H. L., & Liu, G. (2014). Five miRNAs as novel diagnostic biomarker candidates for primary nasopharyngeal carcinoma. *Asian Pacific Journal of Cancer Prevention*, 15(18), 7575-7581.
- Taylor, B. (2017). *The structure of a virus with a tegument*. [online] Available at:

https://commons.wikimedia.org/wiki/File:Viral_Tegument_nn.svg [Accessed on 1 June 2023].

Teleanu, R. I., Chircov, C., Grumezescu, A. M., & Teleanu, D. M. (2019). Tumor angiogenesis and anti-angiogenic strategies for cancer treatment. *Journal of Clinical Medicine*, 9(1), 84.

The ENT Clinic. (n.d). *A NEW Screening Tool for Nose Cancer*. [online] Available at: <https://www.entspecialist.sg/doctors-corner/a-new-screening-tool-for-nose-cancer/> [Accessed on 1 June 2023].

The National Health Center. (2018). *Nasopharyngeal cancer*. [online] (updated on 4 November 2021) Available at: <https://www.nhs.uk/conditions/nasopharyngeal-cancer/> [Accessed on 17 October 2022].

Thomas, L., Bouhouche, K., Whitfield, M., Thouvenin, G., Coste, A., Louis, B., Szymanski, C., Bequignon, E., Papon, J. F., Castelli, M., Lemullois, M., Dhalluin, X., Drouin-Garraud, V., Montantin, G., Tissier, S., Duquesnoy, P., Copin, B., Dastot, F., Couvet, S., Barbotin, A. L., Faucon, C., Honore, I., Maitre, B., Beydon, N., Tamalet, A., Rives, N., Koll, F., Escudier, E., Tassin, A. M., Touré, A., Mitchell, V., Amselem, S., & Legendre, M. (2020). TTC12 loss-of-function mutations cause primary ciliary dyskinesia and unveil distinct dynein assembly mechanisms in motile cilia versus flagella. *The American Journal of Human Genetics*, 106(2), 153-169.

Thompson, L. D. (2007). Update on nasopharyngeal carcinoma. *Head and Neck Pathology*, 1(1), 81-86.

Thorley-Lawson, D. A., & Gross, A. (2004). Persistence of the Epstein–Barr virus and the origins of associated lymphomas. *New England Journal of Medicine*, 350(13),

1328-1337.

- Tian, X., Liu, Y., Wang, Z., & Wu, S. (2021). miR-144 delivered by nasopharyngeal carcinoma-derived EVs stimulates angiogenesis through the FBXW7/HIF-1 α /VEGF-A axis. *Molecular Therapy-Nucleic Acids*, 24, 1000-1011.
- Tian, F., Zhao, J., Fan, X., & Kang, Z. (2017). Weighted gene co-expression network analysis in identification of metastasis-related genes of lung squamous cell carcinoma based on the Cancer Genome Atlas database. *Journal of Thoracic Disease*, 9(1), 42-53.
- Tiong, T. S., & Selva, K. S. (2005). Clinical presentation of nasopharyngeal carcinoma in Sarawak Malaysia. *Medical Journal of Malaysia*, 60(5), 624-628.
- Tsang, C. M., Deng, W., Yip, Y. L., Zeng, M. S., Lo, K. W., & Tsao, S. W. (2014). Epstein-Barr virus infection and persistence in nasopharyngeal epithelial cells. *Chinese Journal of Cancer*, 33(11), 549-555.
- Tsao, S. W., Tsang, C. M., & Lo, K. W. (2017). Epstein–Barr virus infection and nasopharyngeal carcinoma. *Philosophical Transactions of The Royal Society B: Biological Sciences*, 372(1732), 20160270.
- Tsao, S. W., Yip, Y. L., Tsang, C. M., Pang, P. S., Lau, V. M. Y., Zhang, G., & Lo, K. W. (2014). Etiological factors of nasopharyngeal carcinoma. *Oral Oncology*, 50(5), 330-338.
- Tsunoda, A., Suzuki, M., Kishimoto, S., Anzai, T., Matsumoto, F., Ikeda, K., & Terasaki, O. (2021). Otitis media with effusion caused by a parapharyngeal tumor showing normal nasopharyngeal findings. *Ear, Nose & Throat Journal*, 100(7), 543-545.
- Torres, K., Landeros, N., Wichmann, I. A., Polakovicova, I., Aguayo, F., & Corvalan, A. H. (2021). EBV miR-BARTs and human lncRNAs: Shifting the balance in

- competing endogenous RNA networks in EBV-associated gastric cancer. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1867(4), 166049.
- Utama, D. S., Ralahayu, P., & Bahar, E. (2022). Relationship between primary tumors of nasopharyngeal carcinoma with the degree of conductive hearing loss in Dr. Mohammad Hoesin Hospital Palembang. *Bioscientia Medicina: Journal of Biomedicine and Translational Research*, 6(3), 1423-1434.
- Vig, A., Poulter, J. A., Ottaviani, D., Tavares, E., Toropova, K., Tracewska, A. M., Mollica, A., Kang, J., Kehelwathugoda, O., Paton, T., Maynes, J. T., Wheway, G., Arno, G., Khan, K. N., McKibbin, M., Toomes, C., Ali, M., Scipio, M. D., Li, S., Ellingford, J., Black, G., Webster, A., Rydzanicz, M., Stawiński, P., Ploski, R., Vincent, A., Cheetham, M. E., Inglehearn, C. F., Roberts, A., & Heon, E. (2020). DYNC2H1 hypomorphic or retina-predominant variants cause nonsyndromic retinal degeneration. *Genetics in Medicine*, 22(12), 2041-2051.
- Vilagos, B., Hoffmann, M., Souabni, A., Sun, Q., Werner, B., Medvedovic, J., Bilic, I., Minnich, M., Axelsson, E., Jaritz, M., & Busslinger, M. (2012). Essential role of EBF1 in the generation and function of distinct mature B cell types. *Journal of Experimental Medicine*, 209(4), 775-792.
- ViRBase V3.0. (2021). *Tutorial*. [online] Available at: <http://www.rna-society.org/virbase/help.html> [Accessed on 23 August 2021].
- Wandora. (2018). *Clustering coefficient*. [online] Available at: http://www.wandora.org/wiki/Clustering_coefficient#:~:text=Clustering%20coefficient%20is%20a%20property,any%20connections%20in%20the%20neighborhood. [Accessed on 5 January 2023].

- Wang, J., Abhinav, P., Xu, Y. J., Li, R. G., Zhang, M., Qiu, X. B., Di, R. M., Qiao, Q. Li, X. M., Huang, R. T., Xue, S., & Yang, Y. Q. (2019). NR2F2 loss-of-function mutation is responsible for congenital bicuspid aortic valve. *International Journal of Molecular Medicine*, 43(4), 1839-1846.
- Wang, K. H., Austin, S. A., Chen, S. H., Sonne, D. C., & Gurushanthaiah, D. (2017). Nasopharyngeal carcinoma diagnostic challenge in a nonendemic setting: Our experience with 101 patients. *Permanente Journal*, 21(3), 16-180.
- Wang, Y., Bao, W., Liu, Y., Wang, S., Xu, S., Li, X., Li, Y., & Wu, S. (2018). miR-98-5p contributes to cisplatin resistance in epithelial ovarian cancer by suppressing miR-152 biogenesis via targeting Dicer1. *Cell Death & Disease*, 9(5), 447.
- Wang, S., Ma, G., Zhu, H., Lv, C., Chu, H., Tong, N., Wu, D., Qiang, F., Gong, W., Zhao, Q., Tao, Q., Zhou, J., Zhang, Z., & Wang, M. (2016). miR-107 regulates tumor progression by targeting NF1 in gastric cancer. *Scientific Reports*, 6(1), 36531.
- Wang, Y., Liu, C. L., Storey, J. D., Tibshirani, R. J., Herschlag, D., & Brown, P. O. (2002). Precision and functional specificity in mRNA decay. *Proceedings of The National Academy of Sciences*, 99(9), 5860-5865.
- Wang, Y., Zhang, Y., & Ma, S. (2013). Racial differences in nasopharyngeal carcinoma in the United States. *Cancer Epidemiology*, 37(6), 793-802.
- Wang, L., Zhao, J., Zhu, C., Yang, K., Zhu, L., & Liu, Y. (2022). Construction of a ceRNA network and comprehensive analysis of lncRNA in hepatocellular carcinoma. *Genes*, 13(5), 785.
- Wang, D., Zeng, Z., Zhang, S., Xiong, F., He, B., Wu, Y., Li, W., Tang, L., Wei, F., Xiang, B., Li, Z., Zhou, Y., Zhou, M., Li, X., Li, Y., Li, G., Xiong, W., & Guo, C. (2020). Epstein-Barr virus-encoded miR-BART6-3p inhibits cancer cell proliferation

- through the LOC553103-STMN1 axis. *The FASEB Journal*, 34(6), 8012-8027.
- Ward, M. H., Pan, W. H., Cheng, Y. J., Li, F. H., Brinton, L. A., Chen, C. J., Hsu, M. M., Chen, I. H., Levine, P. H., Yang, C. S., & Hildesheim, A. (2000). Dietary exposure to nitrite and nitrosamines and risk of nasopharyngeal carcinoma in Taiwan. *International Journal of Cancer*, 86(5), 603-609.
- Wargasetia, T. L., & Priyandoko, D. (2018). Up-regulation of two human microRNAs which might be involved in the carcinogenesis of nasopharyngeal cancer and that are promising for biomarker development. *Comparative Clinical Pathology*, 27, 795-801.
- Wolf, H., zur Hausen, H., Klein, G., Becker, V., Henle, G., & Henle, W. (1975). Attempts to detect virus-specific DNA sequences in human tumors. *Medical Microbiology and Immunology*, 161(1), 15-21.
- Wong, E. H. C., Tan, H. Y., Dompok, T. M., Mohamad Ishak, L. A., & Loong, S. P. (2021). The first report on incidence of nasopharyngeal carcinoma in Sabah, Borneo. *ORL; Journal for Oto-Rhino-Laryngology and Its Related Specialties*, 83(4), 258-262.
- Wong, T. S., Chen, S., Zhang, M. J., Chan, J. Y. W., & Gao, W. (2018). Epstein-Barr virus-encoded microRNA BART7 downregulates major histocompatibility complex class I chain-related peptide A and reduces the cytotoxicity of natural killer cells to nasopharyngeal carcinoma. *Oncology Letters*, 16(3), 2887-2892.
- Wu, Z. H., Zhou, J., Hu, G. H., Liu, J., Li, W. C., Lai, X. H., & Liu, M. (2021). LncRNA CASC2 inhibits lung adenocarcinoma progression through forming feedback loop with miR-21/p53 axis. *The Kaohsiung Journal of Medical Sciences*, 37(8), 675-685.

- Wu, L., Zuo, N., Pan, S., Wang, Y., Wang, Q., & Ma, J. (2022). miR-1246 promotes laryngeal squamous cell carcinoma progression by interacting with THBS1. *Journal of Environmental Pathology, Toxicology and Oncology*, 41(3), 65-75.
- Wu, S., Xie, D. L., & Dai, X. Y. (2019). Down-regulation of miR-155 promotes apoptosis of nasopharyngeal carcinoma CNE-1 cells by targeting PI3K/AKT-FOXO3a signaling. *European Review for Medical & Pharmacological Sciences*, 23(17), 7391-7398.
- Xi, X., Li, T., Huang, Y., Sun, J., Zhu, Y., Yang, Y., & Lu, Z. J. (2017). RNA biomarkers: frontier of precision medicine for cancer. *Non-Coding RNA*, 3(1), 9.
- Xiang, T., Lin, Y. X., Ma, W., Zhang, H. J., Chen, K. M., He, G. P., Zhang, X., Xu, M., Feng, Q. S., Chen, M. Y., Zeng, M. S., Zeng, Y. X., & Feng, L. (2018). Vasculogenic mimicry formation in EBV-associated epithelial malignancies. *Nature Communications*, 9(1), 1-15.
- Xiao, A., Li, H., Shechter, D., Ahn, S. H., Fabrizio, L. A., Erdjument-Bromage, H., Ishibe-Murakami, S., Wang, B., Tempst, P., Hofmann, K., Patel, D. J., Elledge, S. J., & Allis, C. D. (2009). WSTF regulates the H2A. X DNA damage response via a novel tyrosine kinase activity. *Nature*, 457(7225), 57-62.
- Xiao, K., Wang, Y., Zhou, L., Wang, J., Wang, Y., Tong, D., Zhu, Z., & Jiang, J. (2021). Construction of ceRNA network to identify the lncRNA and mRNA related to non-small cell lung cancer. *PLoS One*, 16(10), e0259091.
- Xie, Y., Geng, Y., Yao, J., Ji, J., Chen, F., Xiao, J., Hu, X., Ma, L., & Ma, L. (2023). N-nitrosamines in processed meats: Exposure, formation and mitigation strategies. *Journal of Agriculture and Food Research*, 13, 100645.
- Xing, B., Qiao, X. F., Qiu, Y. H., & Li, X. (2021). TMPO-AS1 regulates the

- aggressiveness-associated traits of nasopharyngeal carcinoma cells through sponging miR-320a. *Cancer Management and Research*, 13, 415-425.
- Xu, Y., Huang, X., Ye, W., Zhang, Y., Li, C., Bai, P., Lin, Z., & Chen, C. (2020). Comprehensive analysis of key genes associated with ceRNA networks in nasopharyngeal carcinoma based on bioinformatics analysis. *Cancer Cell International*, 20(408), 1-14.
- Xu, Y. F., Mao, Y. P., Li, Y. Q., Ren, X. Y., He, Q. M., Tang, X. R., Sun, Y., Liu, N., & Ma, J. (2015). MicroRNA-93 promotes cell growth and invasion in nasopharyngeal carcinoma by targeting disabled homolog-2. *Cancer Letters*, 363(2), 146-155.
- Yan, X., Liu, X. P., Guo, Z. X., Liu, T. Z., & Li, S. (2019). Identification of hub genes associated with progression and prognosis in patients with bladder cancer. *Frontiers in Genetics*, 10, 408.
- Yang, J. H., Li, J. H., Shao, P., Zhou, H., Chen, Y. Q., & Qu, L. H. (2011). starBase: a database for exploring microRNA–mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Research*, 39(suppl_1), D202-D209.
- Yang, Y., Liu, W., Zhang, Y., Wu, S., & Luo, B. (2019). *Integrated analysis of the miRNA-mRNA network associated with LMP1 gene in nasopharyngeal carcinoma*. BioRxiv. <https://doi.org/10.1101/615237>
- Yao, J., Chen, X., Liu, X., Li, R., Zhou, X., & Qu, Y. (2021). Characterization of a ferroptosis and iron-metabolism related lncRNA signature in lung adenocarcinoma. *Cancer Cell International*, 21(1), 1-14.
- Ye, Z., Chen, L., Zhong, H., Cao, L., Fu, P., & Xu, J. (2023). Epidemiology and clinical characteristics of Epstein-Barr virus infection among children in Shanghai, China, 2017-2022. *Frontiers in Cellular and Infection Microbiology*, 13, 1139068.

- Ye, Y., Li, S. L., & Wang, S. Y. (2018). Construction and analysis of mRNA, miRNA, lncRNA, and TF regulatory networks reveal the key genes associated with prostate cancer. *PloS One*, 13(8), e0198055.
- Ye, Z., Wang, F., Yan, F., Wang, L., Li, B., Liu, T., Hu, F., Jiang, M., Li, W., & Fu, Z. (2019). Bioinformatic identification of candidate biomarkers and related transcription factors in nasopharyngeal carcinoma. *World Journal of Surgical Oncology*, 17(1), 60.
- Young, L. S. (2008). Epstein–Barr virus: general features. In B. W. J. Mahy & M. H. V. Van Regenmortel (Eds.), *Desk Encyclopedia of Human and Medical Virology* (pp. 532-540). London: Academic Press.
- Young, L. S., & Dawson, C. W. (2014). Epstein-Barr virus and nasopharyngeal carcinoma. *Chinese Journal of Cancer*, 33(12), 581.
- Young, L. S., Yap, L. F., & Murray, P. G. (2016). Epstein–Barr virus: more than 50 years old and still providing surprises. *Nature Reviews Cancer*, 16(12), 789-802.
- Yu, D., Lim, J., Wang, X., Liang, F., & Xiao, G. (2017). Enhanced construction of gene regulatory networks using hub gene information. *BMC Bioinformatics*, 18(1), 186.
- Žaja, R., Aydin, G., Lippok, B. E., Feederle, R., Lüscher, B., & Feijis, K. L. H. (2020). Comparative analysis of MACROD1, MACROD2 and TARG1 expression, localisation and interactome. *Scientific Reports*, 10(1), 1-16.
- Zhang, T., Chen, Z., Deng, J., Xu, K., Che, D., Lin, J., Jiang, P., Gu, X., & Xu, B. (2022). Epstein–Barr virus-encoded microRNA BART22 serves as novel biomarkers and drives malignant transformation of nasopharyngeal carcinoma. *Cell Death & Disease*, 13(7), 1-14.
- Zhang, Y., Chen, Y., Zheng, J., Wang, J., Duan, S., Zhang, W., Yan, X., & Zhu, X. (2019).

- Vertebrate Dynein-f depends on Wdr78 for axonemal localization and is essential for ciliary beat. *Journal of Molecular Cell Biology*, 11(5), 383-394.
- Zhang, S. Q., Pan, S. M., Liang, S. X., Han, Y. S., Chen, H. B., & Li, J. C. (2021). Research status and prospects of biomarkers for nasopharyngeal carcinoma in the era of high-throughput omics. *International Journal of Oncology*, 58(4), 9.
- Zhang, X., Sun, J., Lu, Y., Zhang, J., Shimada, K., Noda, T., Zhao, S., Koyano, T., Matsuyama, M., Zhou, S., Wu, J., Ikawa, M., & Liu, M. (2021). LRRC23 is a conserved component of the radial spoke that is necessary for sperm motility and male fertility in mice. *Journal of Cell Science*, 134(20), jcs259381.
- Zhang, X., Wang, W., Zhu, W., Dong, J., Cheng, Y., Yin, Z., & Shen, F. (2019). Mechanisms and functions of long non-coding RNAs at multiple regulatory levels. *International Journal of Molecular Sciences*, 20(22), 5573.
- Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K., & Liu, X. (2014). Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PloS One*, 9(1), e78644.
- Zhao, M., Kim, P., Mitra, R., Zhao, J., & Zhao, Z. (2015). TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. *Nucleic Acids Research*, 44(D1), D1023-D1031.
- Zhao, K., Wang, D., Zhao, X., Wang, C., Gao, Y., Liu, K., Wang, F., Wu, X., Wang, X., Sun, L., Zang, J., & Mei, Y. (2020). WDR63 inhibits Arp2/3-dependent actin polymerization and mediates the function of p53 in suppressing metastasis. *EMBO Reports*, 21(4), e49269.
- Zheng, C. X., Yan, L., Nilsson, B., Eklund, G., & Drettner, B. (1994). Epstein-Barr virus infection, salted fish and nasopharyngeal carcinoma: A case-control study in

- southern. *Acta Oncologica*, 33(8), 867-872.
- Zhong, Y., Zhu, Y., Dong, J., Zhou, W., Xiong, C., Xue, M., Shi, M., & Chen, H. (2019). Identification of key transcription factors AP-1 and AP-1-dependent miRNAs forming a co-regulatory network controlling PTEN in liver ischemia/reperfusion injury. *BioMed Research International*, 2019, 8962682.
- Zhou, Z. J., Luo, C. B., Xin, H. Y., Hu, Z. Q., Zhu, G. Q., Li, J., & Zhou, S. L. (2020). MACROD2 deficiency promotes hepatocellular carcinoma growth and metastasis by activating GSK-3 β / β -catenin signaling. *NPJ Genomic Medicine*, 5(1), 15-28.
- Zhou, R. S., Zhang, E. X., Sun, Q. F., Ye, Z. J., Liu, J. W., Zhou, D. H., & Tang, Y. (2019). Integrated analysis of lncRNA-miRNA-mRNA ceRNA network in squamous cell carcinoma of tongue. *BMC Cancer*, 19(1), 1-10.
- Zhou, J., Zhang, B., Zhang, X., Wang, C., & Xu, Y. (2022). Identification of a 3-miRNA signature associated with the prediction of prognosis in nasopharyngeal carcinoma. *Frontiers in Oncology*, 11, 5969.
- Zhu, H. M., Fei, Q., Qian, L. X., Liu, B. L., He, X., & Yin, L. (2019). Identification of key pathways and genes in nasopharyngeal carcinoma using bioinformatics analysis. *Oncology Letters*, 17(5), 4683-4694.
- Zhu, X., Jiang, L., Yang, H., Chen, T., Wu, X., & Lv, K. (2020). Analyzing the lncRNA, miRNA, and mRNA-associated ceRNA networks to reveal potential prognostic biomarkers for glioblastoma multiforme. *Cancer Cell International*, 20(1), 1-12.
- Zhu, J. Y., Pfuhl, T., Motsch, N., Barth, S., Nicholls, J., Grässer, F., & Meister, G. (2009). Identification of novel Epstein-Barr virus microRNA genes from nasopharyngeal carcinomas. *Journal of Virology*, 83(7), 3333-3341.
- Zingoni, A., Molfetta, R., Fionda, C., Soriani, A., Paolini, R., Cippitelli, M., Cerboni, C.,

- & Santoni, A. (2018). NKG2D and its ligands:“one for all, all for one”. *Frontiers in Immunology*, 9, 476.
- Zuazo-Gaztelu, I., & Casanovas, O. (2018). Unraveling the role of angiogenesis in cancer ecosystems. *Frontiers in Oncology*, 8, 248.
- Zwyea, S., Naji, L., & Almansouri, S., 2021. Kaplan-Meier plotter data analysis model. In: Journal of Physic, *The International Conference of Chemistry 2020*. Bagdad, Iraq, 17-18 December 2020. IOP Publishing.

APPENDICES

Appendix A: Gene Expression Omnibus (GEO) datasets information and Target RNAs of miRNAs

Table 1: Up-regulated and down-regulated miRNAs in GSE70970 dataset following the cut-off criterion $|\log_2 \text{fold-change}| \geq 1$ and adjusted P-values < 0.05 curated from GEO2R.

Expression	microRNA
Up-regulated	hsa-miR-1308, hsa-miR-1246, hsa-miR-494-3p, hsa-miR-494-5p, hsa-miR-320a, hsa-miR-630, hsa-miR-93-3p, hsa-miR-93-5p, hsa-miR-493-3p, hsa-miR-493-5p, hsa-miR-1290, hsa-miR-1978, hsa-miR-378a-3p, hsa-miR-378a-5p, hsa-miR-1973, hsa-miR-21-3p, hsa-miR-21-5p, hsa-miR-15b-3p, hsa-miR-15b-5p, hsa-miR-2117, hsa-miR-25-3p, hsa-miR-25-5p, hsa-miR-155-3p, hsa-miR-155-5p, hsa-let-7i-3p, hsa-let-7i-5p, hsa-miR-106b-3p, hsa-miR-106b-5p, hsa-miR-1975, hsa-miR-708-3p, hsa-miR-708-5p, hsa-miR-96-3p, hsa-miR-96-5p, hsa-miR-1977, hsa-miR-490-3p, hsa-miR-22-3p, hsa-miR-22-5p, hsa-miR-106a-5p, hsa-miR-148b-3p, hsa-miR-148b-5p, hsa-miR-15a-3p, hsa-miR-15a-5p, hsa-miR-191-3p, hsa-miR-191-5p, hsa-miR-421, hsa-miR-1274a, hsa-miR-130b-3p, hsa-miR-130b-5p, hsa-miR-30a-3p, hsa-miR-30a-5p, hsa-miR-98-3p, hsa-miR-98-5p, hsa-miR-1305, hsa-miR-146b-5p, hsa-miR-16-3p, hsa-miR-16-5p,

Table 1 continued

	hsa-miR-574-5p, hsa-miR-660-3p, hsa-miR-660-5p, hsa-miR-1979, hsa-miR-151-5p, hsa-miR-301b, hsa-miR-107, hsa-miR-19b-3p, hsa-miR-19b-5p, hsa-miR-374a-3p, hsa-miR-374a-5p, hsa-miR-374b-3p, hsa-miR-374b-5p, hsa-miR-331-3p, hsa-miR-196b-3p, hsa-miR-196b-5p, hsa-miR-455-3p, hsa-miR-663b, hsa-miR-135b-3p, hsa-miR-135b-5p, hsa-miR-192-3p, hsa-miR-192-5p, hsa-miR-19a-3p, hsa-miR-19a-5p, hsa-miR-1275, hsa-miR-181b-5p, hsa-miR-454-3p, hsa-miR-454-5p, hsa-miR-340-3p, hsa-miR-340-5p, hsa-miR-30e-3p, hsa-miR-30e-5p, hsa-miR-425-3p, hsa-miR-425-5p, hsa-miR-27a-3p, hsa-miR-27a-5p, hsa-miR-29b-3p, hsa-miR-29b-5p, hsa-miR-301a-3p, hsa-miR-301a-5p, hsa-miR-20a-5p, hsa-miR-455-5p, hsa-miR-211-3p, hsa-miR-211-5p
Down-regulated	hsa-miR-375, hsa-miR-34c-3p, hsa-miR-145-3p, hsa-miR-145-5p, hsa-miR-34c-3p, hsa-miR-216a-3p, hsa-miR-216a-5p, hsa-miR-574-3p, hsa-miR-204-3p, hsa-miR-204-5p, hsa-miR-199b-5p, hsa-miR-485-3p, hsa-miR-203a, hsa-miR-99b-3p, hsa-miR-99b-5p, hsa-miR-423-3p, hsa-miR-1297, hsa-miR-31-3p, hsa-miR-31-5p, hsa-miR-497-3p, hsa-miR-497-5p, hsa-miR-100-3p, hsa-miR-100-5p, hsa-miR-199a-5p, hsa-miR-125a-5p, hsa-miR-520e, hsa-miR-92b-3p,

Table 1 continued

hsa-miR-92b-5p, hsa-miR-361-5p, hsa-miR-30c-5p, hsa-miR-570-3p, hsa-miR-570-5p, hsa-miR-125b-5p, hsa-miR-30d-3p, hsa-miR-30d-5p, hsa-miR-424-3p, hsa-miR-424-5p, hsa-miR-23b-3p, hsa-miR-23b-5p, hsa-miR-150-3p, hsa-miR-150-5p, hsa-miR-200a-3p, hsa-miR-200a-5p, hsa-miR-342-3p, hsa-let-7c-3p, hsa-let-7c-5p, hsa-let-7b-3p, hsa-let-7b-5p, hsa-miR-34b-3p, hsa-miR-34b-5p, hsa-miR-2278

Figure 2: The common DEGs between targeted mRNAs by up-regulated miRNAs with down-regulated mRNAs in the datasets GSE64634, GSE126683 and GSE61218. (https://drive.google.com/file/d/1YZ79ybPvzcoMuu5fk3j8J7ZjaN3KBhqu/view?usp=drive_link).

Figure 3: The common DEGs between targeted mRNAs by down-regulated miRNAs with up-regulated mRNAs in the datasets GSE64634, GSE126683 and GSE61218 (https://drive.google.com/file/d/1DZoE1KYDiJrqNfWKxTXhRx_ziW9URcrl/view?usp=drive_link).

Figure 4: The common DELs between targeted miRNAs by up-regulated miRNAs with down-regulated lncRNAs in the datasets GSE95166, GSE126683 and GSE61218 (https://drive.google.com/file/d/1XWNQUsXodFBVHaJHI11YPIo22LqroW7T/view?usp=drive_link)

Table 5: The experimental methods for determining target mRNAs for up-regulated miRNAs curated from miRTarBase.

miRNA	Target Gene	Experiments	Support Type	References (PMID)
hsa-miR-494-3p	EFCAB1	PAR-CLIP	Functional MTI (Weak)	23446348
hsa-miR-320a	FBXL13	PAR-CLIP	Functional MTI (Weak)	22012620
hsa-miR-93-3p	LRRC34	CLASH	Functional MTI (Weak)	23622248
hsa-miR-93-5p	KIF6, TTC9	PAR-CLIP	Functional MTI (Weak)	23446348
hsa-miR-21-3p	KIF6	PAR-CLIP	Functional MTI (Weak)	22291592
hsa-miR-21-5p	FBXL13	PAR-CLIP	Functional MTI (Weak)	23446348
hsa-miR-25-3p	CCDC113	HITS-CLIP	Functional MTI	23824327
	GULP1	PAR-CLIP	(Weak)	22012620
hsa-miR-25-5p	LRTOMT	HITS-CLIP	Functional MTI (Weak)	23706177

Table 5 continued

hsa-miR-320a	FBXL13	PAR-CLIP	Functional MTI (Weak)	22012620
hsa-miR-93-3p	LRRC34	CLASH	Functional MTI (Weak)	23622248
hsa-miR-93-5p	KIF6, TTC9	PAR-CLIP	Functional MTI (Weak)	23446348
hsa-miR-21-3p	KIF6	PAR-CLIP	Functional MTI (Weak)	22291592
hsa-miR-21-5p	FBXL13	PAR-CLIP	Functional MTI (Weak)	23446348
hsa-miR-25-3p	CCDC113	HITS-CLIP	Functional MTI	23824327
	GULP1	PAR-CLIP	(Weak)	22012620
hsa-miR-25-5p	LRTOMT	HITS-CLIP	Functional MTI (Weak)	23706177
hsa-miR-155-5p	AKR1C3, DYNC2H1	Proteomics	Functional MTI (Weak)	18668040
hsa-miR-378a-3p	SPA17	PAR-CLIP	Functional MTI (Weak)	22012620
hsa-miR-155-5p	AKR1C3, DYNC2H1	Proteomics	Functional MTI (Weak)	18668040

Table 5 continued

hsa-miR-378a-3p	SPA17	PAR-CLIP	Functional MTI (Weak)	22012620
hsa-miR-378a-5p	RRAD	HITS-CLIP	Functional MTI (Weak)	23824327
hsa-miR-148b-3p	STEAP4	Microarray	Functional MTI (Weak)	17612493
	EYA4			17612493
hsa-miR-98-3p	RRAD	PAR-CLIP	Functional MTI (Weak)	22012620
hsa-miR-98-5p	AKR1C3	Microarray	Functional MTI (Weak)	19088304
hsa-miR-16-5p	TPPP3	Luciferase	Functional MTI	15131085
		reporter assay//Reporter assay;Other		
hsa-miR-574-5p	BCAS1, SPATA6	HITS-CLIP	Functional MTI (Weak)	24906430
hsa-miR-660-3p	SPA17,	HITS-CLIP	Functional MTI (Weak)	19536157
	VSIG2,			23824327
	SORBS2, TSPAN1,			

Table 5 continued

MACROD2				
hsa-miR-135b-5p	KIF6	PAR-CLIP	Functional MTI (Weak)	22100165
hsa-miR-1275	SPATA6	PAR-CLIP	Functional MTI (Weak)	23592263
hsa-miR-211-5p	STEAP4, BBS9	HITS-CLIP	Functional MTI (Weak)	23824327

Table 6: The predicted target mRNAs for up-regulated miRNAs were curated from the computational mirWalk 2.0 database (https://drive.google.com/file/d/15ACd5g5XGODtmc3UGkTg89SVPMB0jSSz/view?usp=drive_link).

Table 7: The experimental methods for determining target mRNAs for down-regulated miRNAs curated from miRTarBase (https://drive.google.com/file/d/1YJsqY6-azmN6g0HLBy6C1oE62OceupJO/view?usp=drive_link).

Table 8: The predicted target mRNAs for down-regulated miRNAs curated from the computational mirWalk 2.0 database (https://drive.google.com/file/d/1j4VhrYYv8R-G7Hgo0nMEVJBlZtpEItDv/view?usp=drive_link).

Table 9: The predicted target lncRNAs of up-regulated miRNAs curated from DIANA-lncBASE V2.

MiRNA Name (miRBase version)	Target lncRNAs	Threshold
hsa-miR-1246	ARHGEF26-AS1	0.908
hsa-miR-320a	CASC2	0.849
hsa-miR-1290	ARHGEF26-AS1	0.812
hsa-miR-155-5p	CASC2	0.994
hsa-miR-708-5p	CASC2	0.717
hsa-miR-96-5p	ARHGEF26-AS1	0.839
hsa-miR-148b-5p	ARHGEF26-AS1	0.738
	CASC2	0.711
hsa-miR-130b-5p	CASC2	0.985
hsa-miR-1305	ARHGEF26-AS1	0.756
	CASC2	0.755
hsa-miR-107	CASC2	0.732
hsa-miR-135b-5p	CASC2	0.727
hsa-miR-211-5p	ARHGEF26-AS1	1

Table 10: The experimental methods to determine target lncRNAs of up-regulated miRNAs curated from DIANA-lncBASE V3.

MiRNA	Target lncRNAs	Experimental Method
hsa-miR-320a	ARHGEF26-AS1	HITS-CLIP
	CASC2	
hsa-miR-93-5p	CASC2	HITS-CLIP
	TPTE2P1	
	ARHGEF26-AS1	
hsa-miR-21-5p	CASC2	qPCR
hsa-miR-708-5p	ARHGEF26-AS1	HITS-CLIP
hsa-miR-106a-5p	CASC2	HITS-CLIP
hsa-miR-98-5p	ARHGEF26-AS1	HITS-CLIP
	CASC2	
	TPTE2P1	HITS-CLIP POSITIVE
hsa-miR-16-5p	TPTE2P1	HITS-CLIP
hsa-miR-107	CASC2	HITS-CLIP
	TPTE2P1	

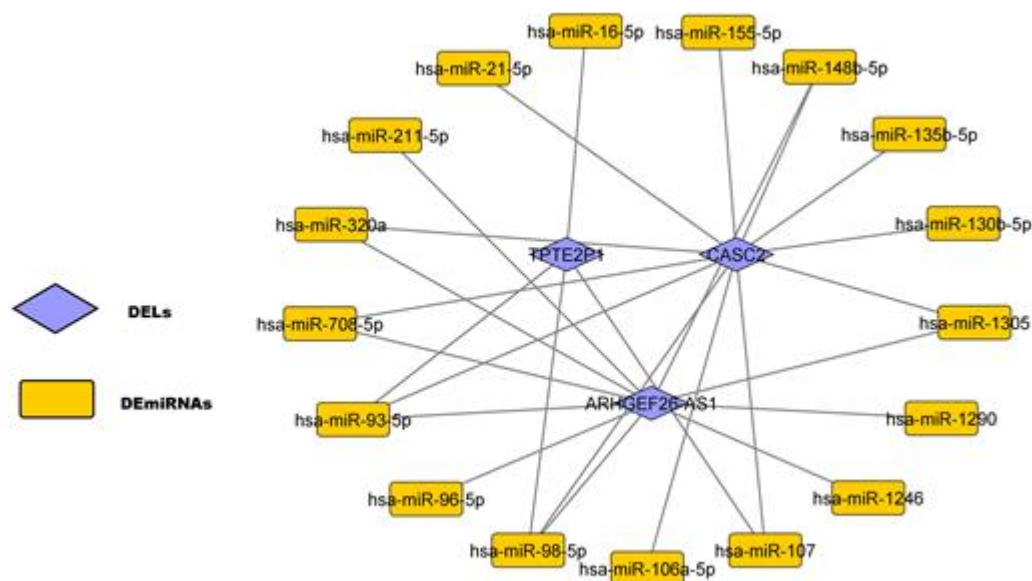


Figure 12: The interaction of up-regulated miRNAs with down-regulated target lncRNAs was visualized by using Cytoscape.

Table 13: The interaction of EBV BART miRNAs with down-regulated mRNAs retrieved from VirBase.

Virus miRNA	Target mRNA	Weak Method	Strong Method	Confidence Score	PMID
Luciferase					
ebv-miR-BART11-5p	EBF1	-	reporter	0.9035	23997984
assay//qPCR					
ebv-miR-BART2-5p	EYA1	PAR-	-	0.5831	22100165
ebv-miR-BART21-3p					
ebv-miR-BART5	EYA4	CLIP	-	0.6048	
ebv-miR-BART19-3p					

Table 13 continued

ebv-miR-BART20-3p	MACROD2	0.6087
-------------------	---------	--------

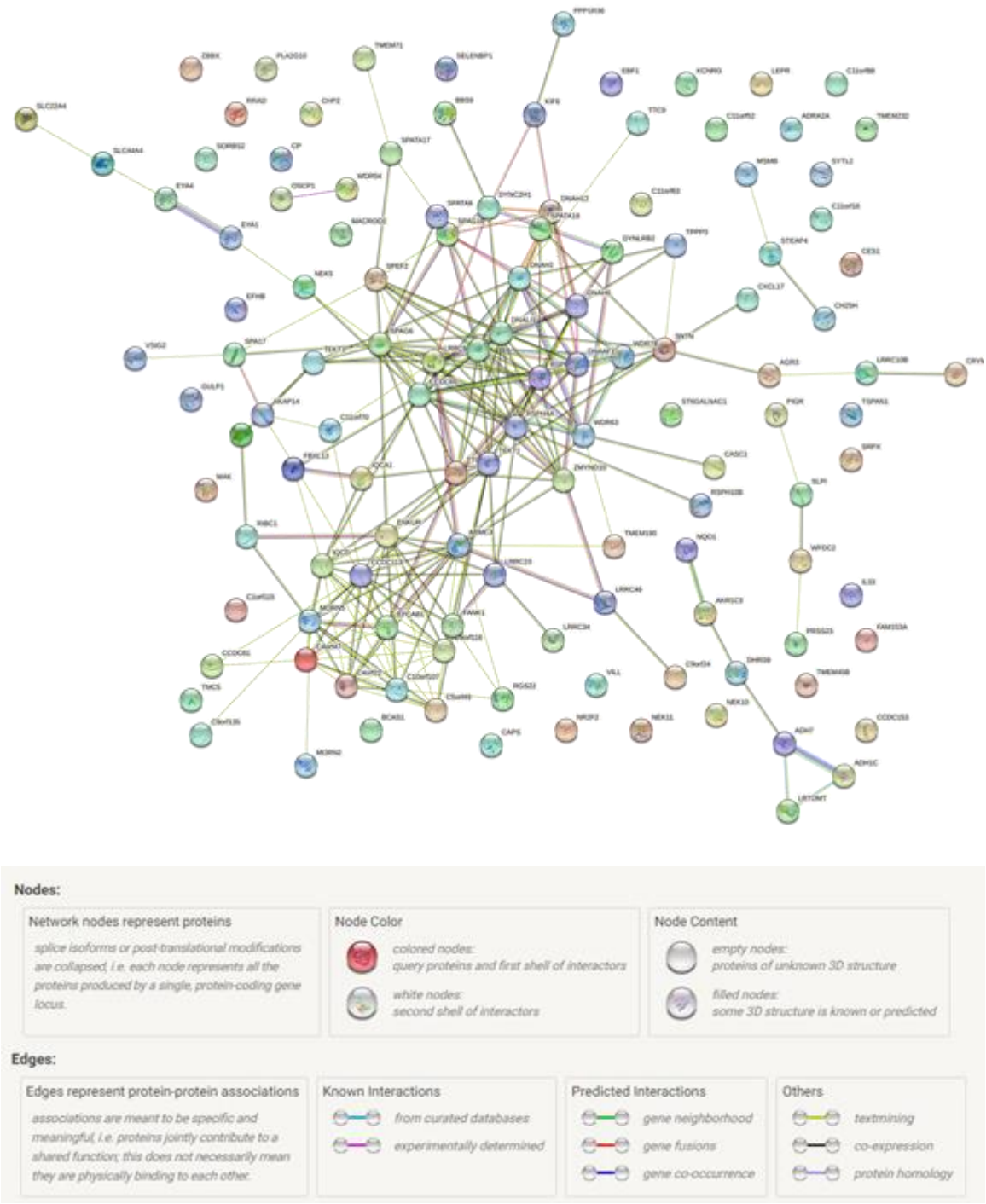


Figure 14 continued

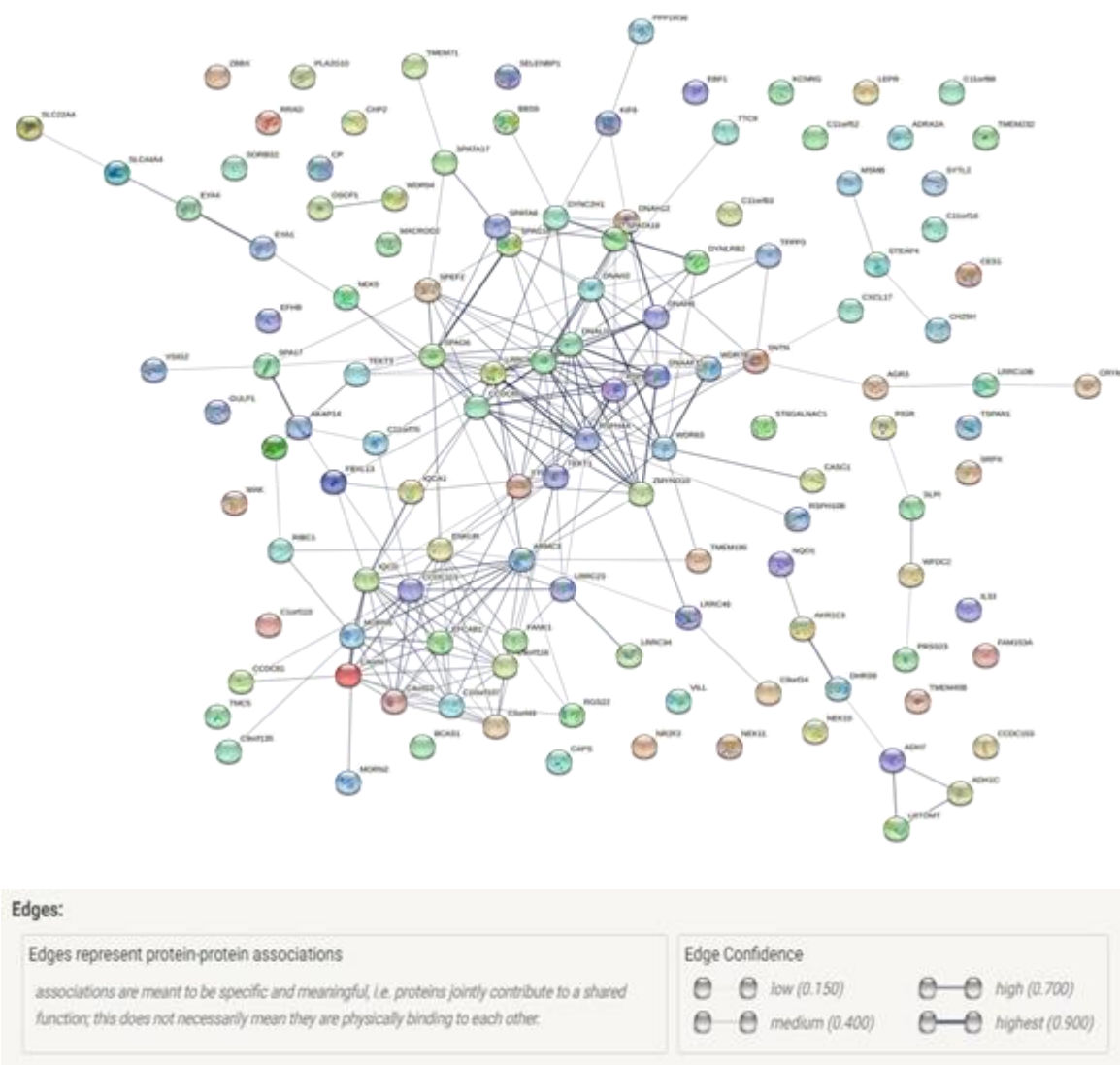


Figure 14: The protein-protein interactions (PPI) between the down-regulated mRNAs retrieved from the STRING database.

Table 15: The protein-protein interactions of common down-regulated mRNAs from the datasets GSE95166, GSE126683 and GSE61218 (https://drive.google.com/file/d/1pyF_nOwhtyMEwGUmBzKeoZrzJGCPqds/view?usp=drive_link).

Table 16: The interactions between the Top 20 DEGs were filtered out using Cytohubba (https://drive.google.com/file/d/1fIGYFboe89IDJatO4oB_EhG6gqCC79IR/view?usp=drive_link).

Table 17: The betweenness centrality, closeness centrality, and node degree of the 18 DEGs with interactions of more than 5 amongst themselves, including self-loop, filtered out from the Top 20 DEGs using the Network Analyzer plug-in provided by the Cytoscape software.

The 18 DEGs with interactions more than 5			
Node Degree	DEGs	Betweenness Centrality	Closeness Centrality
18	DNAH5	0.14600062247121	1
17	DNALI1	0.097525677	0.947368421
14	DNAH6	0.054326175	0.818181818
13	RSPH1	0.052023031	0.782608696
11	RSPH4A	0.030018674	0.72
11	WDR63	0.020821662	0.72
10	DYNC2H1	0.0118892	0.692307692
10	DYNLRB2	0.017538126	0.692307692
10	DNAH12	0.008683473	0.692307692
9	DNAH2	0.002956738	0.666666667
9	SPAG6	0.014114535	0.666666667
8	SPAG16	0.011173358	0.642857143
8	CCDC65	0.006318083	0.642857143
8	WDR78	0	0.642857143
8	SPEF2	0.010690943	0.642857143

Table 17 continued

7	TEKT1	0.011344538	0.620689655
7	LRRC6	0.005664488	0.620689655
6	DNAAF1	0.001089325	0.6

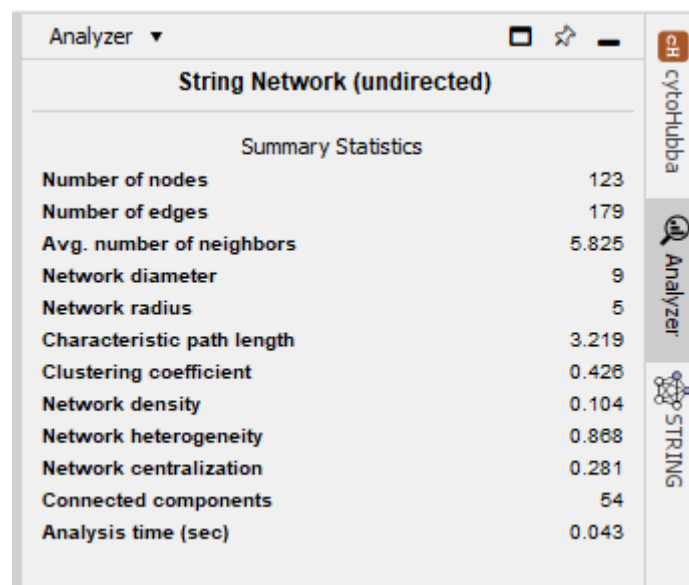


Figure 18: The topology properties of the protein-protein interaction (PPI) network analyzed using the Network Analyzer plug-in provided by the Cytoscape software.

Table 19: The PPI network topology properties analyzed using the Network Analyzer plug-in provided by the Cytoscape software (https://docs.google.com/spreadsheets/d/10la9nmgEGYs8JfVqxQvEAc3rWNI6Fs0C/edit?usp=drive_link&ouid=109467567390471199227&rtpof=true&sd=true).

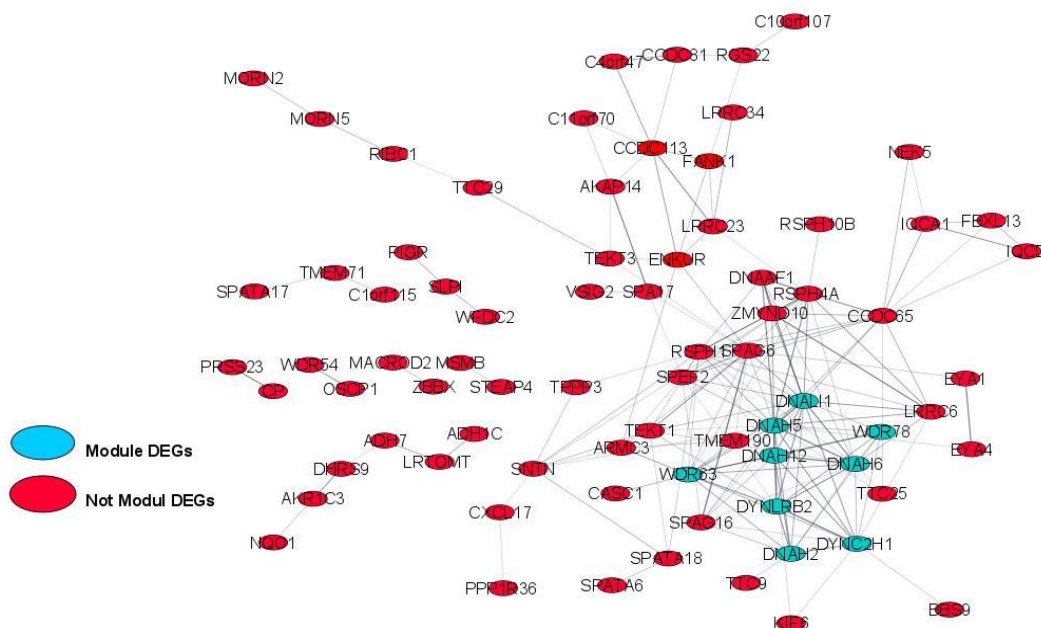


Figure 20: The position of Modules 1 in the protein-protein interaction (PPI) network.

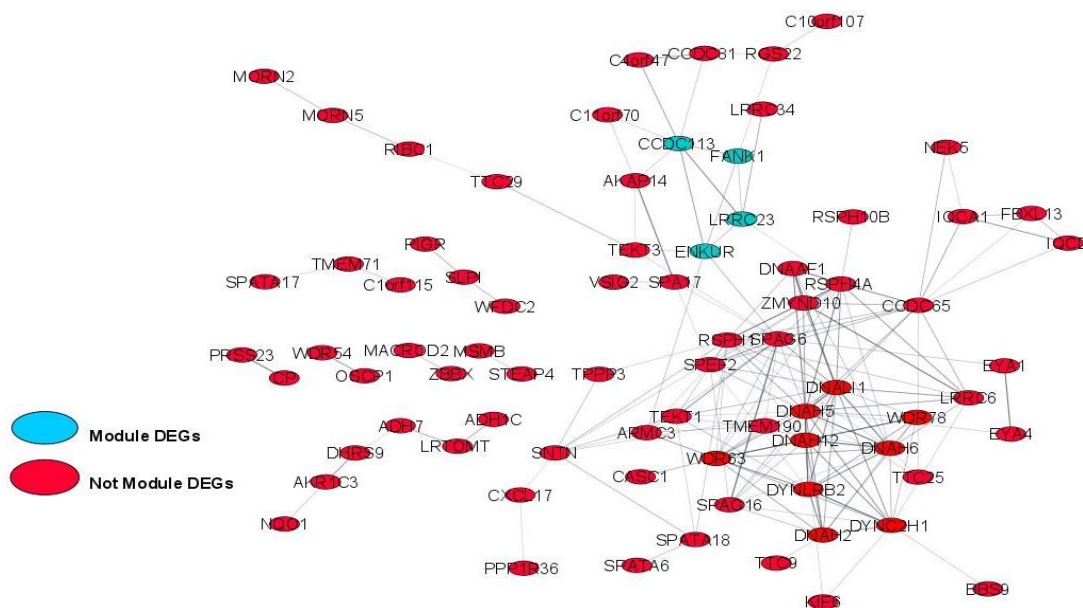


Figure 21: The position of Modules 2 in the protein-protein interaction (PPI) network.

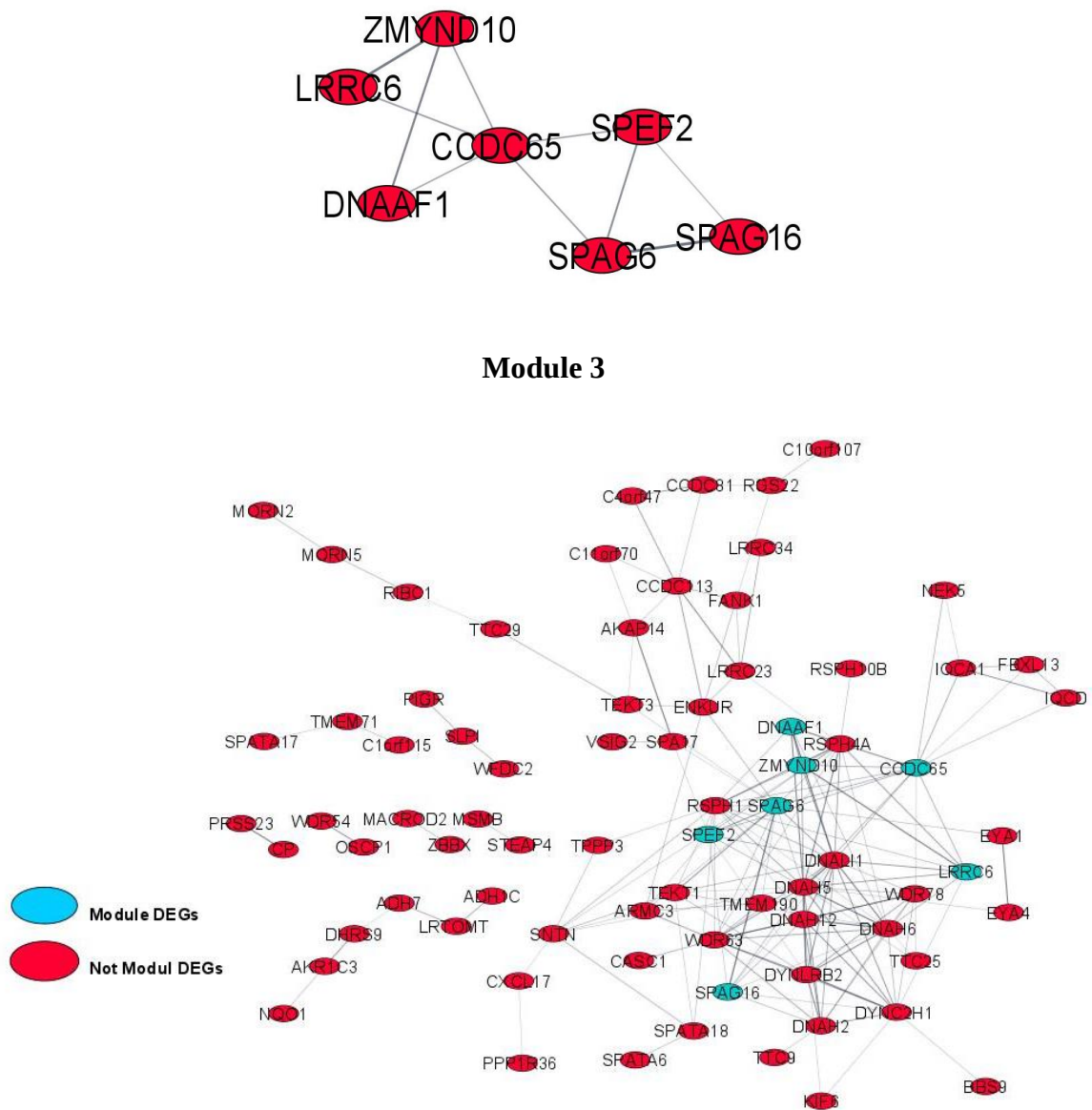


Figure 22: Modules 3 and its position in the protein-protein interaction (PPI) network.

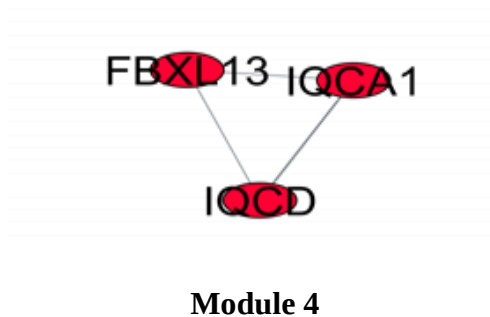


Table 25 continued

absence of ligand			
Positive regulation of DNA repair	0.006529492	0.075757236	EYA1, EYA4
Anatomical structure development	0.007769973	0.075757236	EYA1, EYA4
Anatomical structure morphogenesis	0.021039992	0.159697107	EYA1, EYA4
Sensory perception of sound	0.024568786	0.159697107	EYA1, EYA4
Chromatin organization	0.039054395	0.21758877	EYA1, EYA4
DNA repair	0.045110439	0.219913388	EYA1, EYA4

Table 26: Molecular Functions (MF) of the down-regulated mRNAs targeted by BART miRNAs curated from the DAVID database.

Term	P-Value	FDR	Genes
Protein tyrosine phosphatase activity	0.016968499	0.203621988	EYA1, EYA4
Metal ion binding	0.05362348	0.32174088	EYA1, EYA4, EBF1

Table 27: Cellular Components (CC) of the down-regulated mRNAs targeted by BART miRNAs curated from the DAVID database.

Term	P-Value	FDR	Genes
Nucleus	0.024295574	0.170069018	EYA1, EYA4, EBF1 and MACROD2

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Fold Enrichment	Bonferroni	Benjamini	FDR	Fisher Exact
☐	GOTERM_BP_DIRECT	cilium assembly	RT		7	13.7	2.1E-5	12.1	5.7E-3	4.0E-3	4.0E-3	1.7E-6
☐	GOTERM_BP_DIRECT	cilium movement involved in cell motility	RT		4	7.8	2.9E-5	64.8	8.0E-3	4.0E-3	4.0E-3	4.0E-7
☐	GOTERM_BP_DIRECT	epithelial cilium movement	RT		3	5.9	1.2E-3	57.9	2.7E-1	8.9E-2	8.8E-2	1.8E-5
☐	GOTERM_BP_DIRECT	microtubule-based movement	RT		4	7.8	1.3E-3	18.2	3.0E-1	8.9E-2	8.8E-2	6.8E-5
☐	GOTERM_BP_DIRECT	cilium movement	RT		3	5.9	6.3E-3	24.8	8.2E-1	3.1E-1	3.0E-1	2.4E-4
☐	GOTERM_BP_DIRECT	sensory perception of sound	RT		4	7.8	7.5E-3	9.8	8.7E-1	3.1E-1	3.0E-1	7.4E-4
☐	GOTERM_BP_DIRECT	retinol metabolic process	RT		3	5.9	7.8E-3	22.1	8.8E-1	3.1E-1	3.0E-1	3.4E-4
☐	GOTERM_BP_DIRECT	histone dephosphorylation	RT		2	3.9	1.9E-2	101.3	1.0E0	6.6E-1	6.5E-1	1.7E-4
☐	GOTERM_BP_DIRECT	progesterone metabolic process	RT		2	3.9	2.6E-2	73.7	1.0E0	8.0E-1	7.9E-1	3.2E-4
☐	GOTERM_BP_DIRECT	cilium-dependent cell motility	RT		2	3.9	3.8E-2	50.7	1.0E0	1.0E0	1.0E0	7.0E-4
☐	GOTERM_BP_DIRECT	sperm axoneme assembly	RT		2	3.9	6.6E-2	28.9	1.0E0	1.0E0	1.0E0	2.2E-3
☐	GOTERM_BP_DIRECT	axoneme assembly	RT		2	3.9	6.8E-2	27.9	1.0E0	1.0E0	1.0E0	2.3E-3
☐	GOTERM_BP_DIRECT	negative regulation of extrinsic apoptotic signaling pathway in absence of ligand	RT		2	3.9	8.6E-2	21.9	1.0E0	1.0E0	1.0E0	3.8E-3
☐	GOTERM_BP_DIRECT	protein localization to cilium	RT		2	3.9	8.8E-2	21.3	1.0E0	1.0E0	1.0E0	4.0E-3

Figure 28: The Biological Processes (BP) of the down-regulated mRNAs in the ceRNA network curated from the DAVID database.

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Fold Enrichment	Bonferroni	Benjamini	FDR	Fisher Exact
☐	GOTERM_MF_DIRECT	ATP-dependent microtubule motor activity, minus-end-directed	RT		3	5.9	7.2E-4	73.5	8.3E-2	5.4E-2	5.3E-2	8.6E-6
☐	GOTERM_MF_DIRECT	retinol dehydrogenase activity	RT		3	5.9	8.9E-4	66.2	1.0E-1	5.4E-2	5.3E-2	1.2E-5
☐	GOTERM_MF_DIRECT	dynein light intermediate chain binding	RT		3	5.9	2.1E-3	42.7	2.3E-1	8.6E-2	8.5E-2	4.7E-5
☐	GOTERM_MF_DIRECT	dynein intermediate chain binding	RT		3	5.9	3.2E-3	34.8	3.2E-1	9.7E-2	9.5E-2	8.7E-5
☐	GOTERM_MF_DIRECT	ATPase activity	RT		5	9.8	1.6E-2	5.0	8.6E-1	2.4E-1	2.4E-1	3.2E-3
☐	GOTERM_MF_DIRECT	androsterone dehydrogenase activity	RT		2	3.9	1.8E-2	110.3	8.8E-1	2.4E-1	2.4E-1	1.4E-4
☐	GOTERM_MF_DIRECT	testosterone dehydrogenase (NAD+) activity	RT		2	3.9	1.8E-2	110.3	8.8E-1	2.4E-1	2.4E-1	1.4E-4
☐	GOTERM_MF_DIRECT	ATP binding	RT		9	17.6	1.8E-2	2.6	8.9E-1	2.4E-1	2.4E-1	6.7E-3
☐	GOTERM_MF_DIRECT	alcohol dehydrogenase (NAD) activity	RT		2	3.9	2.0E-2	98.0	9.1E-1	2.4E-1	2.4E-1	1.8E-4
☐	GOTERM_MF_DIRECT	androstane-3-alpha,17-beta-diol dehydrogenase activity	RT		2	3.9	2.0E-2	98.0	9.1E-1	2.4E-1	2.4E-1	1.8E-4
☐	GOTERM_MF_DIRECT	oxidoreductase activity	RT		3	5.9	9.4E-2	5.7	1.0E0	1.0E0	9.9E-1	1.6E-2

Figure 29: The Molecular Functions (MF) of the down-regulated mRNAs in the ceRNA network curated from the DAVID database.

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Fold Enrichment	Bonferroni	Benjamini	FDR	Fisher Exact
☐	GOTERM_CC_DIRECT	motile cilium	RT		9	17.6	6.5E-10	28.9	5.9E-8	5.9E-8	5.5E-8	2.1E-11
☐	GOTERM_CC_DIRECT	9+2 motile cilium	RT		6	11.8	7.5E-9	84.3	6.8E-7	3.4E-7	3.2E-7	7.4E-11
☐	GOTERM_CC_DIRECT	axoneme	RT		8	15.7	1.3E-8	27.4	1.2E-6	4.1E-7	3.9E-7	4.6E-10
☐	GOTERM_CC_DIRECT	sperm flagellum	RT		6	11.8	1.3E-6	30.8	1.2E-4	3.0E-5	2.8E-5	4.0E-8
☐	GOTERM_CC_DIRECT	dynein complex	RT		4	7.8	3.3E-5	62.4	3.0E-3	6.0E-4	5.7E-4	4.7E-7
☐	GOTERM_CC_DIRECT	cytoplasm	RT		24	47.1	2.5E-3	1.8	2.1E-1	3.4E-2	3.3E-2	1.4E-3
☐	GOTERM_CC_DIRECT	cilium	RT		5	9.8	2.6E-3	8.5	2.1E-1	3.4E-2	3.3E-2	3.0E-4
☐	GOTERM_CC_DIRECT	microtubule	RT		5	9.8	7.2E-3	6.4	4.8E-1	8.2E-2	7.8E-2	1.1E-3
☐	GOTERM_CC_DIRECT	9+0 motile cilium	RT		2	3.9	1.2E-2	168.5	6.5E-1	1.2E-1	1.1E-1	5.5E-5
☐	GOTERM_CC_DIRECT	axonemal dynein complex	RT		2	3.9	3.7E-2	52.7	9.7E-1	3.3E-1	3.1E-1	6.5E-4
☐	GOTERM_CC_DIRECT	axonemal microtubule	RT		2	3.9	9.3E-2	20.1	1.0E0	7.7E-1	7.3E-1	4.5E-3

Figure 30: The Cellular Component (CC) of the down-regulated mRNAs in the ceRNA network curated from the DAVID database.

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Fold Enrichment	Bonferroni	Benjamini	FDR	Fisher Exact
☐	GOTERM_BP_DIRECT	microtubule-based movement	RT		8	88.9	3.3E-17	218.6	9.9E-16	9.9E-16	8.9E-16	1.4E-19
☐	GOTERM_BP_DIRECT	cilium movement involved in cell motility	RT		5	55.6	7.4E-11	486.3	2.2E-9	1.1E-9	1.0E-9	1.3E-13
☐	GOTERM_BP_DIRECT	cilium-dependent cell motility	RT		2	22.2	5.7E-3	304.0	1.6E-1	4.6E-2	4.1E-2	1.8E-5
☐	GOTERM_BP_DIRECT	inner dynein arm assembly	RT		2	22.2	6.1E-3	286.1	1.7E-1	4.6E-2	4.1E-2	2.0E-5
☐	GOTERM_BP_DIRECT	cilium movement	RT		2	22.2	1.8E-2	99.2	4.1E-1	1.1E-1	9.5E-2	1.7E-4
☐	GOTERM_BP_DIRECT	determination of left/right symmetry	RT		2	22.2	2.4E-2	72.6	5.2E-1	1.2E-1	1.1E-1	3.2E-4
☐	GOTERM_BP_DIRECT	cilium assembly	RT		2	22.2	8.2E-2	20.7	9.2E-1	3.5E-1	3.1E-1	3.9E-3

Figure 31: The Biological Processes (BP) of the down-regulated mRNAs in the Module 1 curated from the DAVID database.

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Fold Enrichment	Bonferroni	Benjamini	FDR	Fisher Exact
☐	GOTERM_MF_DIRECT	dynein intermediate chain binding	RT		6	66.7	1.4E-12	332.8	1.5E-11	1.5E-11	1.5E-11	3.6E-15
☐	GOTERM_MF_DIRECT	ATP-dependent microtubule motor activity, minus-end-directed	RT		5	55.6	4.0E-11	585.4	4.4E-10	2.2E-10	2.2E-10	5.3E-14
☐	GOTERM_MF_DIRECT	dynein light intermediate chain binding	RT		5	55.6	4.1E-10	339.9	4.5E-9	1.5E-9	1.5E-9	1.0E-12
☐	GOTERM_MF_DIRECT	dynein heavy chain binding	RT		3	33.3	1.9E-5	395.1	2.0E-4	4.2E-5	4.2E-5	4.1E-8
☐	GOTERM_MF_DIRECT	ATPase activity	RT		5	55.6	1.9E-5	23.8	2.1E-4	4.2E-5	4.2E-5	7.9E-7
☐	GOTERM_MF_DIRECT	ATP binding	RT		5	55.6	2.3E-3	6.8	2.5E-2	4.2E-3	4.2E-3	3.3E-4
☐	GOTERM_MF_DIRECT	dynein light chain binding	RT		2	22.2	4.2E-3	421.5	4.5E-2	6.6E-3	6.6E-3	9.0E-6
☐	GOTERM_MF_DIRECT	microtubule motor activity	RT		2	22.2	2.5E-2	70.2	2.4E-1	3.4E-2	3.4E-2	3.5E-4

Figure 32: The Molecular Functions (MF) of the down-regulated mRNAs in the Module 1 curated from the DAVID database.

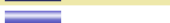
Sublist	Category	Term	RT	Genes	Count	%	P-Value	Fold Enrichment	Bonferroni	Benjamini	FDR	Fisher Exact
☐	GOTERM_CC_DIRECT	dynein complex	RT		6	66.7	1.4E-13	509.8	3.2E-12	2.8E-12	2.8E-12	2.3E-16
☐	GOTERM_CC_DIRECT	9+2 motile cilium	RT		6	66.7	2.5E-13	458.8	5.6E-12	2.8E-12	2.8E-12	4.6E-16
☐	GOTERM_CC_DIRECT	axoneme	RT		7	77.8	1.1E-12	130.6	2.4E-11	8.0E-12	8.0E-12	8.0E-15
☐	GOTERM_CC_DIRECT	axonemal dynein complex	RT		5	55.6	1.7E-11	716.9	3.7E-10	9.2E-11	9.2E-11	1.8E-14
☐	GOTERM_CC_DIRECT	microtubule	RT		6	66.7	5.5E-8	41.6	1.2E-6	2.4E-7	2.4E-7	1.3E-9
☐	GOTERM_CC_DIRECT	motile cilium	RT		5	55.6	1.1E-7	87.6	2.3E-6	3.9E-7	3.9E-7	1.2E-9
☐	GOTERM_CC_DIRECT	cilium	RT		5	55.6	1.4E-6	46.1	3.1E-5	4.4E-6	4.4E-6	3.0E-8
☐	GOTERM_CC_DIRECT	inner dynein arm	RT		2	22.2	1.9E-3	917.6	4.2E-2	5.3E-3	5.3E-3	1.7E-6
☐	GOTERM_CC_DIRECT	cytoplasmic dynein complex	RT		2	22.2	8.9E-3	199.5	1.8E-1	2.2E-2	2.2E-2	4.3E-5
☐	GOTERM_CC_DIRECT	ciliary tip	RT		2	22.2	1.8E-2	95.6	3.4E-1	4.1E-2	4.1E-2	1.9E-4
☐	GOTERM_CC_DIRECT	sperm flagellum	RT		2	22.2	3.1E-2	56.0	5.0E-1	6.3E-2	6.3E-2	5.5E-4
☐	GOTERM_CC_DIRECT	cytoplasm	RT		6	66.7	4.3E-2	2.4	6.2E-1	8.0E-2	8.0E-2	1.8E-2

Figure 33: The Cellular Component (CC) of the down-regulated mRNAs in the Module 1 curated from the DAVID database.

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Fold Enrichment	Bonferroni	Benjamini	FDR	Fisher Exact
☐	GOTERM_CC_DIRECT	ciliary basal body	RT		2	50.0	2.5E-2	58.7	3.4E-1	4.1E-1	4.1E-1	4.3E-4

Figure 34: The Cellular Component (CC) of the down-regulated mRNAs in the Module 2 curated from the DAVID database.